IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 13, 2021, accepted May 31, 2021, date of publication June 7, 2021, date of current version June 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3087131

Faulty Line Identification Method Based on Bayesian Optimization for Distribution Network

JIRAN ZHU[1,2], LONGHUA MU[1], DING MA[1], AND XIN ZHANG[1]
[1]College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China
[2]State Grid Hu'nan Electric Power Corporation Research Institute, Changsha 410007, China

Corresponding author: Xin Zhang (ei_zx@tongji.edu.cn)

This work was supported in part by the Science and Technology Project of State Grid Hunan Electric Power Company Ltd., under Grant 5216A5200008, and in part by the Science and Technology Project of State Grid Corporation of China under Grant 5216A019000R.

ABSTRACT Selective ground-fault protection is greatly valued for the safe and reliable operation of power systems. With the wide adoption of fault indicator in distribution network, the amount of available fault data increases dramatically. The in-depth investigation of fault recording data helps improve the accuracy of faulty line identification. To perform fault data analysis with higher efficiency, a single-phase-to-ground fault identification model based on the k-Nearest Neighbor (kNN) classification algorithm is proposed in the paper. In this model, the eigenvectors consist of wavelet energy ratio, wavelet coefficients variance and wavelet power obtained by the decomposition of transient components. Furthermore, through the theoretical analysis and experimental comparison of three parameter adjustment algorithms, Bayesian Optimization algorithm is selected to find the optimal parameters of fault identification model, and realize the adaptive adjustment of model parameters. Finally, the validity and feasibility of the model are verified by the experimental data, and the accuracy and efficiency of fault identification are improved by using Bayesian Optimization algorithm.

INDEX TERMS Distribution network, faulty line identification, k-Nearest Neighbor method (kNN), single-phase-to-ground fault, Bayesian optimization.

I. INTRODUCTION

The neutral non-effectively grounding mode is mostly adopted in medium-voltage distribution network [1], [2]. When a single-phase-to-ground (SPG) fault occurs in the distribution network, the fault characteristics are not obvious. Therefore, the detection and faulty line identification of the SPG fault is a lasting yet unsolved task for safe operation of the distribution network [3], [4].

The faulty line identification schemes of the SPG fault can be classified into three categories: transient characteristics-based, node information-based and fault waveform comparison-based methods. The magnitude of the steady-state fault current is low, but there is plenty of transient fault information when the SPG faults occur [5], [6]. Santos *et al.* [7] used the discrete wavelet transform to monitor high-frequency voltage components and presented a transient-based algorithm for faulty line identification in distribution network. Lin *et al.* [8] proposed a discrete wavelet transform based triggering algorithm, which takes wavelet coefficients variations of the zero-sequence voltage as the judge basis. Guo *et al.* [9] proposed a novel method of fault-section location using the transient zero-sequence currents at double-ends of the line section. Cui and Weng [10] proposed a high impedance fault (HIF) monitoring and location scheme using the combination of transient- and steady-state fault characteristics for distribution network. In order to analyze transient signals, the new filtering algorithms, such as wavelet packet transform, Teager–Kaiser energy operators are used to extract the fault transient characteristics [11], [12], so as to identify the faulty line. However, the relationship between fault transient signal and fault condition is not linear, so it is difficult to give quantitative analytical expression of fault characteristics. Therefore, most of the existing transient characteristics-based methods generally use equivalent, optimization and other means to simplify the characteristics solving process, and the fault characteristics extracted by these methods have limitations in application.

For the faulty line identification method based on node information, I. Dzafic *et al.* [13] proposed a downstream marking approach for locating faults based on the status of fault nodes that are telemetered to the distribution con-

The associate editor coordinating the review of this manuscript and approving it for publication was Arturo Conde.

trol center. Hossan and Chowdhury [14] proposed a fault location method based on the measured data from advanced distribution management system (ADMS) software platform. Shi *et al.* [15] proposed a fault location scheme using network topology information and reclosure-generating traveling waves. Majidi and Etezadi-Amoli [16] proposed a new impedance-based technique to locate all fault types in distribution networks. These schemes use the information of fault direction elements and trip signals provided by the distribution terminal equipment to compare the differences of node equipment information along the line, and then identify the faulty line. Regardless of fault direction element or trip signal, this kind of method requires all node information to be complete. If the information of some points is missing, there is strong possibility of misjudgment. In order to solve the problem of lack of information and distortion, some researches have adopted optimization algorithms to improve the fault-tolerant performance of the fault identification system [17], [18]. However, when the number of nodes is large, the number of iterations and convergence speed of these algorithms are difficult to be guaranteed.

For the method based on fault waveform, the waveform distribution characteristics and similarity degree between healthy lines and faulty line are compared, and then faulty line is identified. Chakraborty and Das [19] utilizes the voltage measurements of smart meters to address the HIF detection. Shi *et al.* [20] proposed a SPG fault section identification method based on feature extraction of the synchronous waveforms and the calculation of the eigenvalues for the time-sequenced features. Bhandia *et al.* [21] presented an advanced distortion detection technique (ADDT), based on waveform analytics to distinguish SPG fault and detect HIF. These methods usually take various waveforms of the initial traveling wave and waveform distortions as the fault characteristics. However, due to the short power supply distance, the complex topology of distribution network, and the complex fault conditions, the fault waveforms have differentiated features for different SPG faults. It is difficult to establish an accurate and effective SPG fault identification model by distribution characteristics and similarity degree of waveform data. This kind of method usually only relies on a small number of simulation waveforms under typical fault conditions for analysis and comparison, so the method lacks universality.

Fault indicators mounted on distribution line sections and used to indicate the fault current flowing through the line sections have been widely installed in distribution networks due to their lower cost [22]. Fault indicators can generally improve reliability of distribution network and reduce outage duration by identifying the faulty line section [23]. With the wide application of fault indicator with the fault waveforms recording function and the improvement of information storage technology, the amount of available SPG fault data in distribution networks increases dramatically [24]–[26]. Except the real-time recorded fault data, there are also a lot of historical fault data, which are completely consistent with the characteristics of big data. Under the background of big data, combining with a variety of fault types to analyze the characteristics of ground fault signal and then identifying the faulty line becomes a practical strategy to solve the problem of ground fault detection [27]–[29]. However, the existing intelligence-based ground fault detection methods extract the fault characteristics from a single source, and hence can easily result in misjudgment [30]. The k-Nearest Neighbor (kNN) classification algorithm [31], [32] and Bayesian Optimization algorithm [33] have been widely used for power system fault identification. In this paper, they are introduced into the faulty line identification of distribution network.

This paper proposes a new faulty line identification scheme based on kNN classification algorithm. A SPG fault identification model based on kNN is presented, which can be trained and analyzed by using data of actual fault waveforms. Specifically, in order to improve the accuracy of faulty line identification, Bayesian Optimization is used to adjust the model parameters. The actual recorded data of the SPG fault in the experimental platform are used to verify the effectiveness of the faulty line identification model. The experimental results show that the scheme can break the shackles of fault feature threshold and identify SPG fault through big data.

The organization of this paper is as follows. Section II discusses the SPG fault characteristics selection. Section III presents the faulty line identification model based on kNN. Section IV presents the adaptive parameter adjustment algorithm of SPG fault diagnosis system. The experimental results to prove the effectiveness of SPG fault identification scheme are demonstrated in Section V. Finally, Section VI concludes the paper.

II. SPG FAULT CHARACTERISTICS SELECTION

A. SELECTION OF FAULT CHARACTERISTICS

When a SPG fault occurs in distribution network, the electrical transient quantity generated by the system is much larger than the steady state quantity, and the existence of arc-suppression coil hardly affects the distribution of transient current. However, it is difficult to compare fault transient quantities directly, so it is necessary to process fault data and pick out useful components from transient components. The db4 wavelet is an orthogonal wavelet, which has good time-frequency locating performance and fast attenuation ability. Therefore, the db4 wavelet is adopted to extract the transient characteristic quantities.

As a time-frequency analysis method, wavelet decomposition has a good application effect in the ground fault detection. In this paper, the number of decomposition layers is 5, the fault data is decomposed and reconstructed, and three transient characteristic quantities are extracted: wavelet energy ratio, wavelet coefficients variance and wavelet power.

1) WAVELET ENERGY RATIO

The amplitude of transient components is extracted with a specific time window and frequency window. Discrete wavelet transform is adopted in this paper. Let $f(t)$ be a

signal, and its discrete wavelet transform is presented as follows [34], [35]:

$$WT_f(m,k) = \langle f, \psi_{m,k} \rangle = \frac{1}{\sqrt{a_0^m}} \int_{-\infty}^{+\infty} f(t)\psi^*(a_0^{-m}t - kb_0)dt \tag{1}$$

where $\psi_{m,k}$ is the mother wavelet, ψ^* is the complex conjugate of $\psi_{m,k}$, m is the scale factor, k is the displacement factor, b_0 is the sampling step, and a_0 is a positive integer, with $a_0 \geq 2$. Changing b_0 will change the signal area to be analyzed, and we usually let it cover the entire fault signal; changing a_0 will affect the time and frequency resolution of the signal. The db4 wavelet is adopted to extract the transient characteristic quantities. The db4 wavelet means the decomposition order is 4, and the decomposition order is a concept in discrete wavelet transformation (DWT). The scale factor a_0 is usually set as the smallest whole power of 2, and a_0 should be larger than the decomposition order, so the scale factor is 8 in continue wavelet transform (CWT).

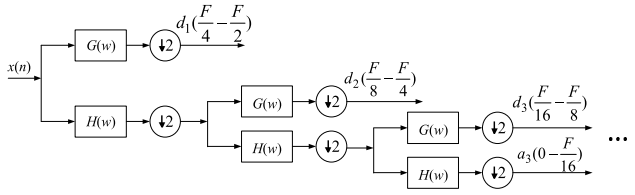

FIGURE 1. Schematic diagram of binary decomposition by wavelet transform.

The schematic diagram of the wavelet transform decomposition is illustrated in Fig. 1. Assuming a signal $x(n)$ to be analyzed, the high-pass filter $G(\omega)$ (wavelet function of Meyer wavelet) and the low-pass filter $H(\omega)$ (scale function of Meyer wavelet) are applied to perform the signal decomposition. F represents the sampling frequency. Firstly $x(n)$ is decomposed to $d_1(\frac{F}{4}, \frac{F}{2})$ and $a_1(0, \frac{F}{2})$, where $d_1(\frac{F}{4}, \frac{F}{2})$ represents the signal information within frequency band $(\frac{F}{4}, \frac{F}{2})$, and $a_1(0, \frac{F}{4})$ represents the one within frequency band $(0, \frac{F}{4})$. Then $a_1(0, \frac{F}{4})$ is decomposed to $d_2(\frac{F}{8}, \frac{F}{4})$ and $a_2(0, \frac{F}{8})$. Similarity, a_i is decomposed to d_{i+1} and a_{i+1} by the same scheme, where i represents the decomposition scale. In this way, the signal is decomposed to the signal information within the expected frequency band in the end. When $F = 250$ kHz and $i = 3$, the corresponding frequency band of d_1, d_2, d_3 and a_3 is 62.5-125 kHz, 31.25-62.5 kHz, 15.625-31.25 kHz, 0-15.625 kHz respectively.

The wavelet energy ratio represents the energy accumulation of the transient zero-sequence current when a fault occurs. According to Parseval Theorem [36], the spectrum energy of transient components is defined as:

$$E_{hj} = \sum_{m=1}^{N} |D_j(m)|^2 \tag{2}$$

$$E_{lS} = \sum_{m=1}^{N} |A_S(m)|^2 \tag{3}$$

where E_{hj} is the high frequency signal energy at the scale j; E_{lS} is the low frequency signal energy at the maximum decomposition scale; m is the sampling points; N is the number of sampling points; S is the maximum scale of wavelet decomposition; $D_j(m)$ is the wavelet coefficient at each scale; and $A_S(m)$ is the scale coefficient at the maximum decomposition scale.

In order to reduce dimensionality, energy ratio ρ_{ic} is defined as the fault characteristic by (4).

$$\rho_{ic} = E_{i\_hc}/E_{i\_lS} \tag{4}$$

where i is the line number; c is the scale of faulty line identification; E_{i_hc} is the high frequency energy of line i at the scale c; and E_{i_lS} is the low frequency energy of line i after the decomposition of maximum scale S.

2) WAVELET COEFFICIENTS VARIANCE

The wavelet coefficients variance is used to represent the change of zero-sequence transient current after SPG fault. The variance is a measure of the difference between the source data and the expected value. The wavelet coefficients variance reflects the fluctuation of wavelet coefficient at the scale of faulty line identification, which can be expressed as:

$$D(d_c) = \frac{1}{N} \sum_{m=1}^{N} (d_c(m) - d_{cav})^2 \tag{5}$$

where N is the sampling points; $d_c(m)$ is the wavelet coefficient at scale c; and d_{cav} is the average value of $d_c(m)$.

3) WAVELET POWER

The wavelet coefficients of the zero-sequence transient current at scale c are denoted as d_{Ic}; the derivative of the wavelet coefficient of the zero-sequence transient voltage at this scale is denoted as d_{Vc}. The power amplitude composed of the zero-sequence transient current and voltage of line i at scale c is defined as:

$$|W_i| = \left| \frac{1}{N} \sum_{m=1}^{N} d_{Ic}(m) d_{Vc}(m) \right| \tag{6}$$

where m is the sampling points, and N is the number of sampling points.

B. NORMALIZED FAULT CHARACTERISTICS

In order to eliminate incoherence in different fault characteristic values and prevent small data from being ignored, this paper chooses a unified expression of these characteristic values. The normalized characteristic values of data are obtained by the linear normalization method, and all characteristics are normalized by (7).

$$X_i' = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{7}$$

where X_i' is the normalized data, X_i is the original data, $X_{\max}$ and $X_{\min}$ are the maximum and minimum values of the original data set, respectively.

III. FAULTY LINE IDENTIFICATION MODEL BASED ON kNN

For the k nearest neighbor algorithm (kNN), k-nearest neighbor means the nearest k neighbors to the data point, and the neighbor represents the nearest point. kNN algorithm does not rely on the class field discrimination to process data points in classification, so kNN algorithm is suitable for machine learning model training with poor class field or more overlaps. In this paper, kNN algorithm is used to train the faulty line identification model of SPG fault.

The most common application of kNN algorithm is to calculate the distance between data points through Euclidean distance. The Euclidean distance is very simple and useful, but its most significant feature is that it equates the differences between different features. Therefore, in order to consider the differences among the three characteristics of wavelet energy ratio, wavelet coefficient variance and wavelet power, this paper adopts weighted Euclidean distance, which means that the distance on each coordinate axis is weighted, so as to realize the differential treatment of the three characteristics in three-dimensional space.

A. IMPROVEMENT OF EUCLIDEAN DISTANCE

The wavelet energy ratio, wavelet coefficient variance and wavelet power are selected as eigenvectors. For Euclidean distance, the weights w between [0, 1] are introduced into each dimension, and the improved Euclidean distance calculation formula is as follows:

$$d = w_1 d_1 + w_2 d_2 + w_3 d_3 \tag{8}$$

where, $w_1 \sim w_3$ are the weights of each dimension; $d_1 \sim d_3$ is the wavelet energy ratio, wavelet coefficient variance and wavelet power of Euclidean distance from the data point to the training point.

B. EVALUATION CRITERIA OF FAULTY LINE IDENTIFICATION MODEL

The accuracy and misreporting rate are used as the evaluation criteria of faulty line identification model. The basis of the criteria is to establish a confusion matrix, as shown in Table 1. In the confusion matrix, true positive (TP) represents the number of faults that are correctly classified as actual faults; true negative (TN) represents the number of normal cases that are classified as normal; false positive (FP) and false negative (FN) represent the number of faults and normal cases that are misclassified, respectively.

TABLE 1. Confusion matrix.

Classification	Fault (Actually)	Normal (Actually)
Fault	TP	FP
Normal	FN	TN

According to the basic requirements of power system relay protection, misreporting normal conditions as abnormal conditions is more acceptable than ignoring an abnormal condition relatively, so it is more practical to select accuracy rate and misreporting rate as the standard of discrimination.

The definition of accuracy and misreporting rate are as follows:

Accuracy:

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Misreporting rate:

$$mis = \frac{FP}{TP + TN + FP + FN} \tag{10}$$

C. FAULTY LINE IDENTIFICATION MODEL BASED ON IMPROVED kNN

Combined with the information above, a faulty line identification model of the SPG fault is established by using the majority voting system as the classification criterion:

$$y = \arg\max_{c_j} \sum_{x_i \in N_k(x)} I(y_i = c_j) \tag{11}$$

where I is the indicator function, which is 1 when the equation in parentheses holds, otherwise it's 0; $i = 1, 2, \ldots, N$, N is the number of all classes; $j = 1, 2, \ldots, k$, k is the number of points adjacent to x. From the equation above, it can be seen clearly that y represents the class with the most x among its k neighbors.

The calculation process of the model is as follows:

(1) The initial fault recorded data were transformed in order to extract eigenvectors. The training data and test data are selected for normalization processing to generate training sample set and test sample set;

(2) Set the value k;

(3) Calculate the weighted distance between the test sample and each training sample, and the descending order of the distance was used to obtain k training data values that are the nearest sample points to be tested;

(4) Count k training data points in step (3), find the class that occurred most frequently, and classify the data points to be tested as this class;

(5) Calculate the accuracy and misreporting rate of classification results of test data.

IV. PARAMETER ADJUSTMENT BASED ON BAYESIAN OPTIMIZATION

When the traditional kNN algorithm deals with the classification problem of unbalanced sample data (that is, the data volume of some classes is much smaller than that of other classes), the results tend to favor the majority classes and ignore the minority classes. The ground fault data of distribution network are unbalanced data, and the size of faulty line data is much smaller than the size of healthy lines data. The weight of each dimension in the distance formula of kNN algorithm is the input parameter of the faulty line identification model, which has a critical influence on the merits of the model. Therefore, the parameter adjustment algorithm is used to assign different weights to the majority classes and

the minority classes. If the parameter setting does not match with the model, the model will end up with low accuracy.

Common parameter adjustment algorithms are Grid Search [37] and Random Search [38], [39]. Grid Search is an exhaustive search that iterates all possibilities and is a stable full-space scan. Random Search is a widely used super parameter adjustment algorithm due to its fast speed and large randomness of parameter adjustment process. At the same time, it is not likely to fall into the trap of local optimization.

The common disadvantage of the above two parameter adjustment methods is that the training of each set of hyperparameters requires a complete call to the objective function. In the faulty line identification model based on kNN algorithm, the process of calling objective function is slow. If the objective function is computed from scratches every time, it will definitely lead to a slow parameter adjustment. Therefore, an improved parameter adjustment method based on Bayesian Optimization is proposed instead. After comprehensive comparison, a more effectively adaptive parameter adjustment method is selected for the faulty line identification.

The distance formula used in the SPG fault diagnosis model is the improved Euclidean distance formula shown in (8). The amount to be adjusted is $w_1 \sim w_3$ in (8). It is expected that the accuracy of the SPG fault detection will be improved by adjusting the weights of the three eigenvalues adaptively.

A. GRID SEARCH ALGORITHM

Grid Search is a commonly used parameter adjustment method, which is well-known for its exhaustive search. The principle is to try every possibility through loop traversal in all candidate parameter selections, and the best parameter is the final result. Take a model with two parameters a and b as an example. There are 3 possibilities for parameter a and 4 possibilities for parameter b. Use Grid Search algorithm to list the traversal table, as shown in Table 2.

TABLE 2. Grid Search traversal.

Parameters	b=1	b=2	b=3	b=4
a=1	a=1, b=1	a=1, b=2	a=1, b=3	a=1, b=4
a=2	a=2, b=1	a=2, b=2	a=2, b=3	a=2, b=4
a=3	a=3, b=1	a=3, b=2	a=3, b=3	a=3, b=4

As can be seen from Table 2, Grid Search algorithm lists all parameter possibilities to form a 3*4 table, and then performs a cyclic process to traverse through the parameter possibility. The flowchart of Grid Search algorithm is shown in Fig. 2.

Grid Search parameter adjustment algorithm is a method of exhaustively searching in the parameter list and checking each case to find the optimal parameters. It has the advantage of producing highly optimal adjustment results, and it is an adjustment algorithm with high popularity. The disadvantage is that it takes a long time to adjust the parameters, and it is easy to get trapped in a local optimum.

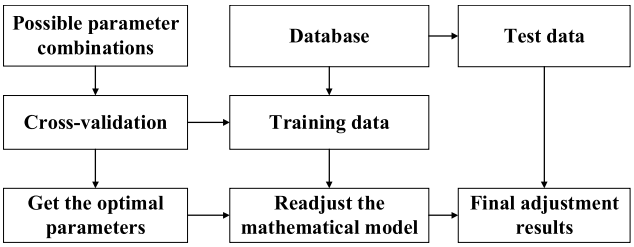

FIGURE 2. Flowchart of Grid Search algorithm.

B. RANDOM SEARCH ALGORITHM

Random Search is a faster adjustment method. Random Search algorithm samples a fixed number of possible parameter configurations from a specified distribution. The theoretical basis is that when the random sample point set is large enough, the global maximum or its approximate value can be found. The flowchart of the Random Search algorithm is shown in Fig. 3.

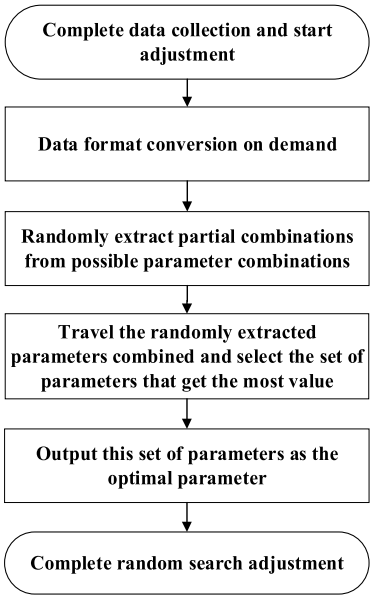

FIGURE 3. Flowchart of Random Search algorithm.

The procedure of Random Search algorithm is simple, the code is easy to implement and has high efficiency. It is often used for quick parameter adjustment of simple models. When the model does not need to get the exact optimal value, Random Search parameter adjustment algorithm is a very efficient method. The disadvantage of Random Search algorithm is that the parameter adjustment process is too random. Generally, this means that only a small part of the data group will be trained by the parameter adjustment method, making more group training less effective and wasting data resources.

C. IMPROVED BAYESIAN OPTIMIZATION ALGORITHM

Bayesian Optimization is a highly efficient machine learning parameter adjustment method, whose main function is to optimize a given objective function [40]–[42]. Bayesian Optimization applies generalized functions, which only need

to specify the inputs and outputs, and the internal structure and mathematical properties is not needed. Based on that, an improved Bayesian Optimization algorithm is described as follows:

The posterior distribution of the objective function is updated by continuously adding sample points until the posterior distribution basically fits the real distribution. Each time the parameter is updated, the information of the previous parameter is considered and the current parameters is adjusted accordingly.

The internal structure of the faulty line identification model is complicated, and a large amount of data will be continuously added in practical applications. The improved Bayesian Optimization algorithm for adaptive parameter adjustment can increase the speed of parameter adjustment and optimize parameters in real time based on new data points.

The key of Bayesian Optimization algorithm is the Gaussian process. The Gaussian processes can be used for nonlinear regression, nonlinear classification, and parameter optimization. The usual parameter adjustment is shown in (12), and only the current data is considered. This kind of adjustment method that only considers the data once usually costs more to achieve adjustment. In (12), y is the parameter adjustment result, and X is the data set for parameter adjustment.

$$p = (y_{N+1}|X_{N+1}) \tag{12}$$

Equation (13) is the parameter adjustment of the Gaussian process. The Gaussian process not only considers the relationship between the current input and output, but also refers to the results of the previous set of parameters.

$$p = (y_{N+1}|X_{N+1}, y_N) \tag{13}$$

The P in (12) and (13) is an intermediate variable in machine learning. The meaning of P is that the parameters have been iteratively tuned for one time.

The Gaussian process considers the relationship between y_N and y_{N+1} by assuming that the value of y follows a joint normal distribution. Since there are conditions that assume a joint normal distribution, the parameters including the mean vector and the covariance matrix need to be given, as shown in (14).

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} \sim N(0, \begin{bmatrix} k(x_1,x_1), k(x_1,x_2), \ldots, k(x_1,x_n) \\ k(x_2,x_1), k(x_2,x_2), \ldots, k(x_2,x_n) \\ \cdots \\ k(x_n,x_1), k(x_n,x_2), \ldots, k(x_n,x_n) \end{bmatrix}) \tag{14}$$

where $k(x_i, x_j)$ is the element of the covariance matrix, N represents the result of the previous iteration.

Among them, the covariance matrix is also called the kernel matrix, and denoted by K, which is only related to the feature x and not to y.

The priori of Gaussian process assumes that y follows the normal distribution of high dimension, and the optimal kernel matrix is obtained according to the training set, so the posteriori is obtained to estimate the test set y_*. The posteriori can be expressed as follows:

$$p = (y_*|y \sim N(K_* K^{-1} y, K_{**} - K_* K^{-1} K_*^T)) \tag{15}$$

where K_* is the kernel vector of the training set, which has the relationship shown in (16).

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K, K_*^T \\ K_*, K_{**} \end{bmatrix}\right) \tag{16}$$

As shown in (15), the parameter adjustment result is only related to the mean. The variance is only related to the kernel matrix, that is, the test set and the training set X, and is not related to the training set y. The training method based on Gaussian process is usually estimated by the square exponential kernel, i.e.

$$k(x_1, x_2) = \sigma_f^2 \exp\left(\frac{-(x_1 - x_2)^2}{2l^2}\right) \tag{17}$$

where σ_f is the standard deviation, l is the distance between two data points in the feature space.

The hyperparameter that needs to be determined during training is θ, and its relationship is shown in (18). Since y obeys the multidimensional normal distribution, the likelihood function shown in (19) can be obtained. Because K is determined by θ, the hyperparameter θ can be obtained by gradient descent.

$$\theta = \left|\sigma_f^2, l\right| \tag{18}$$

$$\begin{aligned} L(\theta|y) &= \log p(y|x, \theta) \\ &= -n\frac{\log 2\pi}{2} - \frac{\log|K|}{2} - \frac{(y-\mu)^T K^{-1}(y-\mu)}{2} \end{aligned} \tag{19}$$

The basic idea of Bayesian Optimization is a Gaussian process. The posterior distribution of the objective function is estimated based on the data, and then the next sampling combination of hyperparameters is selected based on the distribution. Its computing advantage is that it makes full use of the information provided by the previous sampling point, and finds the parameters that maximize the result to the global by learning the shape of the objective function. The Gaussian process plays a role in modeling the objective function in Bayesian Optimization and the posterior distribution is obtained.

After modeling of the Gaussian process, Bayesian Optimization started sampling for sample calculation. Because Bayesian Optimization is easy to continuously sample on the local optimal solution, the balance between development and exploration needs to be considered.

In order to improve sampling efficiency and balance development and exploration, Acquisition Function is needed. Acquisition Function is used to find the next unknown quantity X, and a common simple Acquisition Function is:

$$\begin{aligned} \mathrm{POI}(X) &= P\left(f(X) \geq f(X^+) + \xi\right) \\ &= \Phi\left(\frac{\mu(X) - f(X^+) - \xi}{\sigma(X)}\right) \end{aligned} \tag{20}$$

In (20), $f(X)$ is the value of the objective function of X, $f(X^+)$ is the value of the objective function that is the optimal value of the adjustment so far, and $\mu(X)$, $\sigma(X)$ are the mean and variance of the objective function obtained in the Gaussian process, namely the posterior distribution of $f(X)$. ξ is the tradeoff coefficient. Without the tradeoff coefficient, the POI function tend to be more inclined to value around X^+, which would make the tuning process more prone to development rather than exploration. In the process of parameter adjustment, considering the great cost of calculating $f(X)$, the program keeps trying new X to maximize the value of POI function. When the maximum value is obtained, the value of X is adopted as the optimal parameter.

The POI function shown in (20) is a probability function, mainly considering the probability that $f(X)$ is greater than $f(X^+)$. In the automatic tuning process, the difference between $f(X)$ and $f(X^+)$ needs to be considered, so the expected function *EI* is introduced:

$$EI(X) = \begin{cases} (\mu(X) - f(X^+))\Phi(Z) + \sigma(X)\Phi(Z), & \sigma(X) > 0 \\ 0, & \sigma(X) = 0 \end{cases} \tag{21}$$

The framework of Bayesian Optimization can be constructed by establishing POI function and *EI* function based on Gaussian process. The advantage of Bayesian Optimization is that there is no need to know the internal structure of the objective function, only the input value and output value are needed to adjust the objective function efficiently.

In this paper, Bayesian Optimization algorithm is used to adjust the parameters. The objective function is the faulty line identification model. The objective function value is the diagnostic accuracy of the diagnosis. The adjusted parameter is the weight of the three dimensions in the eigenvector space. That is, the weight of the three eigenvectors of wavelet energy ratio, wavelet coefficient variance and wavelet power. After the first accuracy calculation with the initial parameters, the parameters of the objective function are adjusted by Bayesian Optimization for several times. When a high enough diagnostic accuracy is obtained, the parameter value and diagnostic accuracy are output. The parameter adjustment of the model is finished. The establishment process of fault identification model is shown in Fig. 4.

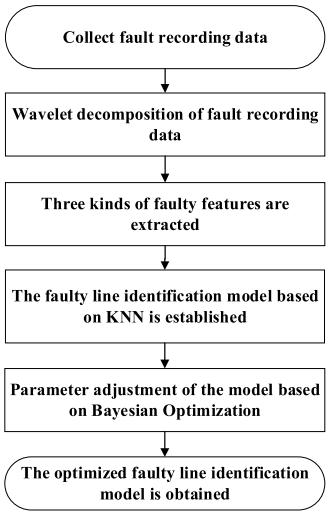

FIGURE 4. The establishment process of fault identification model.

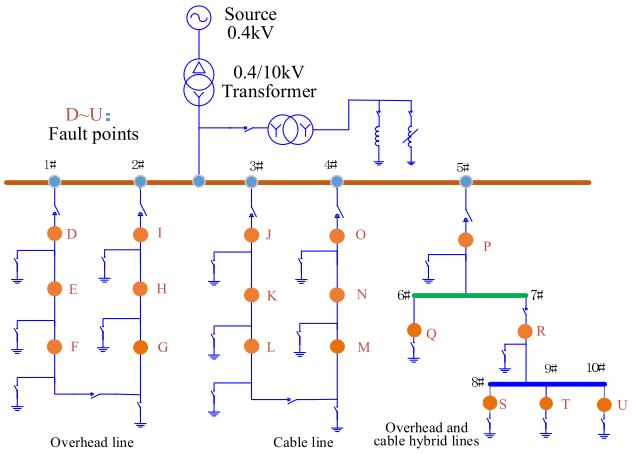

FIGURE 5. Distribution network structure of experimental platform.

V. EXPERIMENTAL COMPARISON OF SPG FAULT IDENTIFICATION

In order to evaluate the proposed faulty line identification scheme based on improved kNN, the recorded data of the SPG fault in the experimental platform are used to verify the effectiveness of the faulty line identification model. All the data are obtained from actual experiments, which can make the results more reliable. The experimental platform is shown in Fig. 5. The experimental platform contains 10 feeders in the system, and the capacitance current is used to simulate the feeder to ground capacitance. The capacitance currents of feeders 1# and 2# are 0.9A and 1.5A respectively, simulating the overhead feeder; the capacitance currents of feeders 3# and 4# are 3A and 4.5A respectively, simulating the cable feeder; the total capacitance current of feeders 5#∼10# is 9.6A, simulating overhead-cable mixed feeder and multi-stage bus. The actual distribution network may have a more complex structure, but the overall structure of experimental platform is similar to that of the actual distribution network.

The ground fault points in Fig. 5 are D∼U, and each fault point is configured with a fault indicator that contains recording function. To verify the feasibility of the SPG fault detection system, a total of 521 experiments were conducted, including two operating conditions: neutral point ungrounded and neutral point grounded by arc-suppression coil (resonant grounded system). A total of 9378 pieces of experimental data were obtained after the end of the experiments. In the 521 groups of experimental data, each group contains the recorded data of 18 (D∼U) fault points, including different fault types:

(1) High impedance fault (1200Ω, 1700Ω, 2100Ω, 2500Ω, 3000Ω, 3600Ω);
(2) Metallic ground fault;
(3) Intermittent arc grounding fault.

A. DATA CONVERSION AND FILTERING

According to the neutral point grounded method, the recorded data are divided into arc-suppression coil system (ASCS) data and ungrounded system (UGS) data, and all the experimental data constitute mixed data. Several groups of experimental data were randomly selected from each group of experiments to constitute test data, while the rest are training data. The detailed data information is shown in Table 3.

TABLE 3. Experimental data.

Data type	Training data group	Test data group	Training data	Test data
Mixed	521	20	9378	360
ASCS	135	13	2430	234
UGS	339	13	6102	234

Format conversion and data classification are carried out for the recorded fault data, and the recorded data of 5 cycles after the fault and 1 cycle before the fault are intercepted. Through (2)-(6), the wavelet energy ratio, wavelet coefficient variance and wavelet power of each line are calculated. According to (7), the linear normalization value is obtained, and then the 3 ground fault eigenvectors are constructed.

B. COMPARISON OF EXPERIMENTAL RESULTS OF PARAMETER ADJUSTMENT ALGORITHM

In order to evaluate the proposed parameter adjustment algorithm, Grid Search algorithm, Random Search algorithm, and the improved Bayesian Optimization algorithm are adopted respectively to obtain the global optimal solution of parameters with the maximum classification accuracy as the criterion. The experiment procedure is to adjust the parameters for 10 times of the same data by using three automatic parameter adjustment algorithms. The results are shown in Table 4 to Table 6, including the optimal k value, the distance weight $w_1 \sim w_3$ value, classification accuracy (*acc*) and misreporting rate (*mis*).

Table 4 shows the results of Grid Search algorithm after 10 automatic parameter adjustments for the SPG fault diagnosis system. It can be seen that the diagnostic accuracy of the three types of data after the adjustment is basically 89%, and the lowest misreporting rate is 2.43%. So this algorithm is not effective. And it can be seen the time spent on adjusting the parameters is long, because Grid Search algorithm must call the objective function completely every time when the parameters are adjusted. While the objective functions of the SPG fault diagnosis system is complicated, and hence Grid Search algorithm would take long time to adjust.

TABLE 4. Experimental results for performance evaluation of Grid Search algorithm.

Data type	w_1	w_2	w_3	t/s	k	*acc*	*mis*
Mixed	0.87	0.55	0.12	9.7	158	89.33%	3.28%
ASCS	0.38	0.74	0.06	5.6	49	90.44%	2.43%
UGS	0.64	0.72	0.23	8.5	42	88.15%	2.71%

TABLE 5. Experimental results for performance evaluation of Random Search algorithm.

Data type	w_1	w_2	w_3	t/s	k	*acc*	*mis*
Mixed	0.69	0.65	0.05	9.6	178	91.36%	2.28%
ASCS	0.36	0.44	0.66	5.8	42	93.46%	2.17%
UGS	0.54	0.15	0.05	8.1	50	91.65%	1.83%

TABLE 6. Experimental results for performance evaluation of Bayesian Optimization algorithm.

Data type	w_1	w_2	w_3	t/s	k	*acc*	*mis*
Mixed	0.69	0.86	0.07	4.8	240	98.06%	0.83%
ASCS	0.81	0.89	0.90	2.2	129	97.44%	0%
UGS	0.87	0.08	0.07	3.9	42	95.30%	1.71%

Table 5 shows the results of Random Search algorithm after 10 automatic parameter adjustments for the SPG fault diagnosis system. Compared with Grid Search algorithm, the time required for the Random Search algorithm to perform 10 times is similar. By comparing *acc* and *mis* of two algorithms, it can be seen that Random Search algorithm has a higher accuracy and a lower misreporting rate within the same parameter adjustment times. According to the theoretical analysis, Grid Search algorithm has stronger development performance and can find the true optimal solution when the amount of data is small. But it easily falls into a local optimal solution when the amount of data is large, thereby affecting its adjustment effect. Meanwhile, Random Search algorithm has a stronger exploration performance. In the case of a large amount of data, it selects data from more regions for training and does not fall into a local optimum. The experiment results are consistent with the theoretical analysis.

In the calculation process of Bayesian Optimization algorithm, the fault eigenvectors are substituted into the Bayesian Optimization algorithm. In the Gaussian process, not only the relationship between the input and output, but also the results of the previous set of parameters are considered. In Section IV, the parameters w_1, w_2 and w_3 for the three eigenvectors in the THREE-DIMENSIONAL space are set according to (8). These three parameters are put into the Bayesian Optimization algorithm. On the premise that the prior obeys the high-dimensional normal distribution, the parameter adjustment results are obtained according to the training set through the posterior. After the accumulation of data, we estimate the posterior distribution of the objective function, and then select the next sampling parameter combination according to the distribution.

Table 6 shows the results of Bayesian Optimization algorithm after 10 automatic parameter adjustments for the SPG fault diagnosis system. Compared with Grid Search algorithm and Random Search algorithm, Bayesian Optimization algorithm significantly reduces the time required for parameter adjustment. In terms of *acc* and *mis*, the diagnostic accuracy after 10 automatic parameter adjustments can reach more than 95%, and the misreporting rate is also significantly

reduced compared to Grid Search algorithm and Random Search algorithm.

Taking the above mixed data as an example, the distribution of the classification results of the training data and test data is shown in Fig. 6.

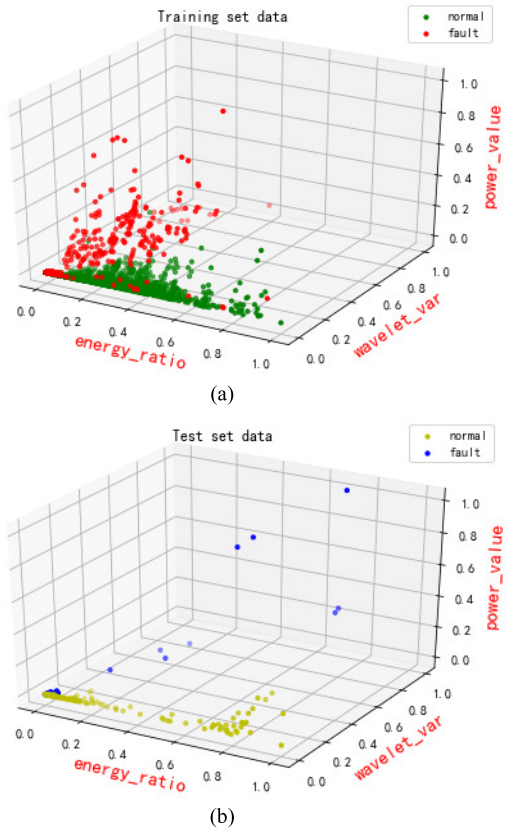

FIGURE 6. Classification distribution map of fault data based on three-dimensional features. (a) Distribution of training data. (b) Distribution of test data.

Fig. 6 clearly shows the distribution of fault and non-fault data. The fault and non-fault data can be basically distinguished in the figure. But some fault data overlap with the non-fault data, so there is a risk of misdiagnosis in theory. After the Bayesian Optimization parameter adjustment procedure, it can be seen that the points that are originally indistinguishable in three-dimensional space can be correctly classified, and the misreporting rate is controlled at 0.83%.

Comparing the experiment results in Table 4, Table 5 and Table 6, it can be seen that Bayesian Optimization algorithm can significantly improve the diagnostic accuracy and reduce diagnostic misreporting rate. So Bayesian Optimization algorithm can benefit fault diagnosis. Because Bayesian Optimization algorithm has the characteristic that it is not necessary to call the objective function completely multiple times, the time required for parameter adjustment can be greatly reduced when adjusting the parameters of the SPG fault diagnosis system. This advantage can improve the training efficiency of the faulty line identification model and make the real-time online training of the model closer to realize.

VI. CONCLUSION

This paper presents a new faulty line identification method of SPG fault in the distribution network. The method takes the SPG fault characteristics, actual fault waveforms and big data of distribution networks into consideration. Based on Bayesian Optimization and adaptive parameters adjustment algorithm, a SPG fault identification model based on kNN algorithm is proposed, which can be trained and analyzed by using a large number of field waveform data. In this model, wavelet energy ratio, wavelet coefficient variance and wavelet power are selected to generate the eigenvectors, and the weight coefficient is introduced on the basis of Euclidean distance.

An improved Bayesian Optimization algorithm is used to find the optimal parameter solution of the model, which improves the accuracy of fault identification and reduces the parameter adjustment time. A distribution network of experimental platform is built and the effectiveness of the fault identification model is verified by the recording data of SPG fault. The experimental results show that the model can meet the requirement of online SPG fault detection of distribution network and has good practical value.

REFERENCES

[1] P. Liu and C. Huang, "Detecting single-phase-to-ground fault event and identifying faulty feeder in neutral ineffectively grounded distribution system," *IEEE Trans. Power Del.*, vol. 33, no. 5, pp. 2265–2273, Oct. 2018.

[2] A. Nikander and P. Jarventausta, "Identification of high-impedance earth faults in neutral isolated or compensated MV networks," *IEEE Trans. Power Del.*, vol. 32, no. 3, pp. 1187–1195, Jun. 2017.

[3] R. Liang, N. Peng, L. Zhou, X. Meng, Y. Hu, Y. Shen, and X. Xue, "Fault location method in power network by applying accurate information of arrival time differences of modal traveling waves," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3124–3132, May 2020.

[4] Z. Wei, Y. Mao, Z. Yin, G. Sun, and H. Zang, "Fault detection based on the generalized S-transform with a variable factor for resonant grounding distribution networks," *IEEE Access*, vol. 8, pp. 91351–91367, May 2020.

[5] Y. Wang, X. Yin, W. Xu, X. Yin, M. Wen, and L. Jiang, "Fault line selection in cooperation with multi-mode grounding control for the floating nuclear power plant grid," *Protection Control Mod. Power Syst.*, vol. 5, no. 1, pp. 164–173, Jul. 2020.

[6] Y. Xiao, J. Ouyang, X. Xiong, Y. Wang, and Y. Luo, "Fault protection method of single-phase break for distribution network considering the influence of neutral grounding modes," *Protection Control Mod. Power Syst.*, vol. 5, no. 1, pp. 111–123, Apr. 2020.

[7] W. C. Santos, F. V. Lopes, N. S. D. Brito, and B. A. Souza, "High-impedance fault identification on distribution networks," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 23–32, Feb. 2017.

[8] C. Lin, W. Gao, and M.-F. Guo, "Discrete wavelet transform-based triggering method for single-phase earth fault in power distribution systems," *IEEE Trans. Power Del.*, vol. 34, no. 5, pp. 2058–2068, Oct. 2019.

[9] M.-F. Guo, J.-H. Gao, X. Shao, and D.-Y. Chen, "Location of single-line-to-ground fault using 1-D convolutional neural network and waveform concatenation in resonant grounding distribution systems," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021, doi: 10.1109/TIM.2020.3014006.

[10] Q. Cui and Y. Weng, "Enhance high impedance fault detection and location accuracy via μ-PMUs," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 797–809, Jan. 2020.

[11] D. A. Gadanayak and R. K. Mallick, "Interharmonics based high impedance fault detection in distribution systems using maximum overlap wavelet packet transform and a modified empirical mode decomposition," *Int. J. Electr. Power Energy Syst.*, vol. 112, pp. 282–293, Nov. 2019.

[12] X. Wang, J. Gao, X. Wei, G. Song, L. Wu, J. Liu, Z. Zeng, and M. Kheshti, "High impedance fault detection method based on variational mode decomposition and Teager–Kaiser energy operators for distribution network," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6041–6054, Nov. 2019.

[13] I. Dzafic, R. A. Jabr, S. Henselmeyer, and T. Donlagic, "Fault location in distribution networks through graph marking," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1345–1353, Mar. 2018.
[14] M. S. Hossan and B. Chowdhury, "Data-driven fault location scheme for advanced distribution management systems," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5386–5396, Sep. 2019.
[15] S. Shi, B. Zhu, A. Lei, and X. Dong, "Fault location for radial distribution network via topology and reclosure-generating traveling waves," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6404–6413, Nov. 2019.
[16] M. Majidi and M. Etezadi-Amoli, "A new fault location technique in smart distribution networks using synchronized/nonsynchronized measurements," *IEEE Trans. Power Del.*, vol. 33, no. 3, pp. 1358–1368, Jun. 2018.
[17] T. Zhang, C. Wang, F. Luo, P. Li, and L. Yao, "Optimal design of the sectional switch and tie line for the distribution network based on the fault incidence matrix," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4869–4879, Nov. 2019.
[18] Y. Jiang, "Toward detection of distribution system faulted line sections in real time: A mixed integer linear programming approach," *IEEE Trans. Power Del.*, vol. 34, no. 3, pp. 1039–1048, Jun. 2019.
[19] S. Chakraborty and S. Das, "Application of smart meters in high impedance fault detection on distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3465–3473, May 2019.
[20] F. Shi, L. Zhang, H. Zhang, K. Xu, and T. Vladimir, "Diagnosis of the single phase-to-ground fault in distribution network based on feature extraction and transformation from the waveforms," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 25, pp. 6079–6086, Dec. 2020.
[21] R. Bhandia, J. de Jesus Chavez, M. Cvetković, and P. Palensky, "High impedance fault detection using advanced distortion detection technique," *IEEE Trans. Power Del.*, vol. 35, no. 6, pp. 2598–2611, Dec. 2020.
[22] J.-H. Teng, W.-H. Huang, and S.-W. Luan, "Automatic and fast faulted line-section location method for distribution systems based on fault indicators," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1653–1662, Jul. 2014.
[23] M. Farajollahi, M. Fotuhi-Firuzabad, and A. Safdarian, "Deployment of fault indicator in distribution networks: A MIP-based approach," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 2259–2267, May 2018.
[24] Q. Jia, X. Dong, and S. Mirsaeidi, "A traveling-wave-based line protection strategy against single-line-to-ground faults in active distribution networks," *Int. J. Electr. Power Energy Syst.*, vol. 107, pp. 403–411, May 2019.
[25] I. Kiaei and S. Lotfifard, "Fault section identification in smart distribution systems using multi-source data based on fuzzy Petri nets," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 74–83, Jan. 2020.
[26] M. Wei, F. Shi, H. Zhang, Z. Jin, V. Terzija, J. Zhou, and H. Bao, "High impedance arc fault detection based on the harmonic randomness and waveform distortion in the distribution system," *IEEE Trans. Power Del.*, vol. 35, no. 2, pp. 837–850, Apr. 2020.
[27] W. Fan and Y. Liao, "Wide area measurements based fault detection and location method for transmission lines," *Protection Control Mod. Power Syst.*, vol. 4, no. 1, pp. 53–64, Mar. 2019.
[28] I. M. Karmacharya and R. Gokaraju, "Fault location in ungrounded photovoltaic system using wavelets and ANN," *IEEE Trans. Power Del.*, vol. 33, no. 2, pp. 549–559, Apr. 2018.
[29] V. Veerasamy, N. I. Abdul Wahab, R. Ramachandran, M. Thirumeni, C. Subramanian, M. L. Othman, and H. Hizam, "High-impedance fault detection in medium-voltage distribution network using computational intelligence-based classifiers," *Neural Comput. Appl.*, vol. 31, no. 12, pp. 9127–9143, Aug. 2019.
[30] X. Wang, J. Gao, X. Wei, Z. Zeng, Y. Wei, and M. Kheshti, "Single line to ground fault detection in a non-effectively grounded distribution network," *IEEE Trans. Power Del.*, vol. 33, no. 6, pp. 3173–3186, Dec. 2018.
[31] Y. Zhang, J. Wu, J. Wang, and C. Xing, "A transformation-based framework for KNN set similarity search," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 3, pp. 409–423, Mar. 2020.
[32] P. Zhao and L. Lai, "Analysis of KNN information estimators for smooth distributions," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3798–3826, Jun. 2020.
[33] K.-R. Kim, Y. Kim, and S. Park, "A probabilistic machine learning approach to scheduling parallel loops with Bayesian optimization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 32, no. 7, pp. 1815–1827, Jul. 2021.
[34] A. P. S. Meliopoulos and C.-H. Lee, "An alternative method for transient analysis via wavelets," *IEEE Trans. Power Del.*, vol. 15, no. 1, pp. 114–121, Jan. 2000.
[35] S. K. Mishra and L. N. Tripathy, "A critical fault detection analysis & fault time in a UPFC transmission line," *Protection Control Mod. Power Syst.*, vol. 4, no. 1, pp. 24–33, Feb. 2019.
[36] A. Iwasaki, "Deriving the variance of the discrete Fourier transform test using Parseval's theorem," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 1164–1170, Feb. 2020.
[37] L. Shi, Y. He, B. Li, T. Cheng, Y. Huang, and Y. Sui, "Tilt angle monitoring by using sparse residual LSTM network and grid search," *IEEE Sensors J.*, vol. 19, no. 19, pp. 8803–8812, Oct. 2019.
[38] J. Pillans, "Improved analog filter design by random search," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 6, pp. 2350–2360, Jun. 2019.
[39] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.
[40] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using Bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 188–201, Jan. 2020.
[41] M. Bozorg, A. Bracale, P. Caramia, G. Carpinelli, M. Carpita, and P. D. Falco, "Bayesian bootstrap quantile regression for probabilistic photovoltaic power forecasting," *Protection Control Mod. Power Syst.*, vol. 5, no. 1, pp. 218–229, Dec. 2020.
[42] Y. Chen, F. Liu, G. He, and S. Mei, "A seidel-type recursive Bayesian approach and its applications to power systems," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1710–1711, Aug. 2012.

JIRAN ZHU received the B.S. and M.S. degrees in electrical engineering from Tongji University, Shanghai, China, in 2008 and 2011, respectively, where he is currently pursuing the Ph.D. degree with the College of Electronics and Information Engineering.

His current research interests include protective relaying of power systems, fault indicator for distribution automation systems, and operation and monitoring of distribution systems.

LONGHUA MU received the B.S., M.E., and Ph.D. degrees in electrical engineering from the China University of Mining and Technology, Xuzhou, China, in 1986, 1988, and 1998, respectively.

He is currently a Professor with the Department of Electrical Engineering, Tongji University, Shanghai, China. His current research interests include power system protection and control, microgrid technology, integrated energy systems, power electronics and its applications in the power systems, distribution automation, and power quality analysis.

DING MA received the B.S. degree in electrical engineering from Tongji University, Shanghai, China, in 2019, where he is currently pursuing the M.S. degree with the College of Electronics and Information Engineering.

His current research interests include power system protection and control, and dc distribution systems.

XIN ZHANG received the B.S. and M.S. degrees in electrical engineering from the China University of Mining and Technology, in 2002 and 2005, respectively, and the Ph.D. degree in electrical engineering from Tongji University, Shanghai, China, in 2012.

He is currently a Research Associate with Tongji University. His current research interests include power system protection and control, and smart distribution systems.