Received 21 October 2023, accepted 14 November 2023, date of publication 28 November 2023, date of current version 1 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3333939

RESEARCH ARTICLE

Transmission Line Fault Diagnosis Method Based on Improved Multiple SVM Model

PEICHUAN SUN[1], XUEFEI LIU[1], MENG LIN[1], JIE WANG[1], TAO JIANG[1], AND YIBO WANG[2]
[1]State Grid Tianjin Electric Power Company, High Voltage Branch, Tianjin 300250, China
[2]Wuhan Kemov Electric Company Ltd., Wuhan 430223, China

Corresponding author: Yibo Wang (15927463005@163.com)

ABSTRACT The accurate diagnosis of transmission line fault types is a prerequisite for quickly removing faults and restoring power supply, as well as the key to effectively reducing user economic losses, ensuring stable operation of the power system. The rapid development of artificial intelligence technology has been a promising way for fault diagnosis. However, the existing methods are still limited by small simples and accuracy of generalization. To overcome these problems, a transmission line fault diagnosis method based on an improved multiple SVM (MSVM) model is proposed in this paper. Firstly, the transmission line was selected as the research object, and its fault types and causes were analyzed in detail. Then, typical fault information are selected and corresponding MSVM model is established. Meanwhile, genetic algorithm (GA) is adopted to optimize model parameters to improve the accuracy of generalization. Finally, an improved IEEE-30 node test system and a real-world testing data are adopted to verify the accuracy and feasibility of the proposed method. Through analysis, fault diagnosis accuracy of the proposed method can be improved by up to 11% with better fitness value.

INDEX TERMS Transmission line, fault diagnosis, improved multiple SVM model.

I. INTRODUCTION

In the current context of the development of the power grid, the power system is moving towards intelligence and complexity. The scale of the power grid is gradually expanding, and the electrical coupling of internal equipment is becoming closer [1], [2]. In order to meet the increasing demand for load electricity, a large number of substations and transmission lines have been put into operation, resulting in a sharp increase in the number of maintenance objects and workload of the power grid. Due to the safe and reliable operation of the power grid, it has always been prioritized by the power grid, Therefore, it is necessary to increase the maintenance cost and capital investment of the power grid, which reduces the economic and social benefits of enterprises [3].

Scientific research has shown that the long service life of power transmission equipment are the main factors leading to faults [5]. The development of faults follows certain objective laws. Generally speaking, the longer the equipment is put into use, the more severe the aging degree of the equipment, and the corresponding probability of failure is also higher. The types and severity of faults exhibited by power transmission equipment vary over different time periods. However, the current methods of troubleshooting are mostly fixed time intervals, which leads to unreasonable and uncoordinated allocation of manpower and time resources. In addition, the power system contains a large number of mechanical switches and power electronic devices [6].

Transmission lines play the role of energy transmission links, which is of great significance for ensuring the stable and reliable supply of electricity. As early as the 1990s, Japanese power companies began to attempt to compare the recording differences of different fault causes before and after flashover to achieve fault cause identification for transmission lines [7]. Based on spectrum analysis, rough conclusions such as an increase in high-order harmonic content before pollution flashover and a low harmonic content in metal line collision faults were obtained, which has considerable enlightenment significance for using waveform features to identify the cause of line faults. The fault features that can be directly obtained from the recording file generally include timestamp, phase feature, and instantaneous value

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

feature. The timestamp feature is divided into fault season and fault day period for characterization, and the phase feature mainly includes fault phase number and fault insertion phase angle (FIPA). The instantaneous value feature needs to be selected in combination with statistical analysis. On this basis, the reference [8] supplements the sequence component characteristics of fault waveform, extracts the zero sequence components of three-phase current and voltage using Karenbauer transform, forms the fault feature set, and uses CN2 rule algorithm to mine the fault cause classification rules. References [9] and [10] analyze the fault characteristics and discusses the feasibility of line faults caused by lightning strikes, trees, and animals from aspects such as timestamp, FIPA, number of fault phases, voltage drop, current transformation rate, arc voltage, and fault impedance estimation. After analyzing the transient current traveling wave in reference [11], the judgment basis for lightning faults is that the tail time of the traveling wave is less than 40 microseconds. Subsequently, scholars have observed the impedance time-varying characteristics based on this waveform to achieve the judgment of non-lightning faults such as external force damage, mountain fire, bird damage, etc. [12]. Malaysian scholars proposed voltage drop and custom fitting coefficients in reference [13] to distinguish transmission line faults caused by lightning strikes and tree barriers based on the waveform shapes of voltage and current, and have successively extracted corresponding waveform features based on more fault cause categories [14], [15]. Reference [16] only starts from the waveform data in the recording file and develops fault features from three perspectives: time domain, frequency domain, and arc. Fault phases, fault duration, fault current component, current attenuation degree, frequency domain energy ratio, and arc voltage are extracted as feature quantities to develop a complete recognition logic for fault classification. Reference [17] analyzed the characteristics of transition resistance based on waveform recording and fault line parameters, and calculated the evaluation resistance and mathematical expectation as the basis for identifying the cause of the fault. Reference [18] is based on the sensing monitoring waveform of Foshan Lightning Location Systems (LLSs), extracting many features of wave head shapes on the horizontal and vertical axes of the time domain, such as peak value, rise time, span time, etc., and exploring and analyzing the correlation between waveform characteristic values and line lightning faults. In reference [19], seven time-frequency features were extracted from the perspectives of fault time, waveform shape, and frequency domain energy spectrum to classify a total of 9 types of fault causes.

When the power transmission line fails, the electrical quantity information on the corresponding line changes before the switching value information, and the electrical quantity information has more advantages than the switching value information in accuracy, Completeness and fault tolerance [20], [21]. With the promotion and application of the wide area measurement system (WAMS), the phasor measurement unit (PMU) used determines the sampling reference based on real-time global positioning system timing, synchronously collects voltage, current, and important switch protection information of units and lines throughout the network, and obtains information such as voltage and current phasor, frequency and frequency change rate, unit and line power, thus recording the transient triggered by disturbance [22], [23]. Reference [24] proposes a fault diagnosis method based on WAMS temporal information, which establishes a reference vector based on known fault types, extracts features from unknown faults for temporal benchmark analysis, and identifies fault types. Reference [25] considers the presence of poor data or noise in WAMS measurement data, and proposes a data fusion method based on temporal data correlation mining to improve the application effectiveness of WAMS data. Making full use of the redundancy of various data information of WAMS and the logical relationship between them, the real-time of PMU synchronous measurement data and the connection between electrical quantity changes and faults can effectively reduce the misjudgment and missed judgment of online fault diagnosis, and improve the efficiency of dispatchers in handling electrical fault.

The rapid development of artificial intelligence technology has brought new ideas to the identification of faults. In previous studies, some mature intelligent neural networks have been utilized in analysis circuit and lithium-ion battery fault diagnosis, and have achieved good results. For example, in [26], a multiple kernel extreme learning machine (MKELM) based diagnosing model is given. A novel scheme for analog circuit fault diagnosis utilizing features extracted from the time-frequency representations of signals and an improved vector-valued regularized kernel function approximation is built in [27]. Similarly, a novel method with XWSE-based feature extractor and SVM is proposed in [28], which possesses a good capability to restrain the environment noise. In [29], a rapid multi-fault diagnosis method for the lithium-ion battery pack is developed, which significantly improved the operational reliability of electric vehicles. Reference [30] introduces semi supervised learning to achieve fault recognition in unlabeled datasets. Reference [31] uses a variational prototype autoencoder to extract signal features and trained a decision tree to determine fault types. The success of fault diagnosis in transmission lines primarily relies on the alarm information obtained from the monitoring devices stationed at various points in the power system. The transmission line section is safeguarded by protective relaying systems, including protective relays (PR), circuit breakers (CB), and communication equipment [32]. Conventional diagnostic methods are typically based on the alarm information obtained from the remote terminal units (RTU) of a supervisory control and data acquisition (SCADA) system. However, when considering failed or malfunctioning PRs or CBs, multiple faults occurring simultaneously, or a combination of these contingencies, the complexity will increase significantly. It is these scenarios that give rise to the incomplete and uncertain characteristics of fault information [32], [33]. As a result of these incompleteness and uncertainties, the accuracy of fault diagnosis can be reduced, or even lead to misdiagnosis. On the other hand, the above fault identification methods based on artificial intelligence technology have to some extent overcome the difficulty of threshold selection, but are limited by the short length of distribution network lines and the single operating position of various closing conditions. In some fault scenarios, it is difficult to construct a sufficient and balanced dataset that can meet the training requirements of neural networks, resulting in difficulties in obtaining sufficient training for such models, limitations in classification performance when applied to fault identification, and poor recognition performance for small sample datasets.

To overcome these challenges, this paper proposes a transmission line fault diagnosis method based on improved multiple SVM model. The innovation points of this paper are as follow:

(1) At the model level, compared with Al-based algorithm mentioned above, classification based SVM used in this paper adopts the principle of structural risk minimization, which guarantees better performance and accuracy of generalization. In addition, its advantages also lie in solving small-sample, nonlinear or high-dimensional pattern recognition problems, with the ability to overcome the problems of curse of dimensionality and over-fitting. Meanwhile, in order to further improve the accuracy of fault diagnosis, multiple SVM are combined for fault diagnosis. Compared to traditional single SVM models, the diagnosis accuracy of the proposed strategy has been significantly improved.

(2) At the parameter level, the direct application of traditional SVM models in transmission line fault classification and diagnosis has significant limitations, with many problems such as low classification recognition accuracy and algorithm efficiency. The construction method of SVM kernel functions and parameter optimization methods are still worth exploring. In this paper, genetic algorithm is adopted for parameter optimization, which significantly improves the accuracy of diagnosis and is easy to promote in practical applications.

This paper is based on PMU synchronous measurement data, decomposing electrical quantity data through symmetrical component method to obtain voltage phasor features, current phasor features, and component features of faults, forming the corresponding fault feature set; simultaneously constructing a multi support vector machine (MSVM) diagnostic model to solve the problem of inaccurate diagnostic results caused by the similarity of fault features among different types of faults. This method analyzes and diagnoses fault information features, which can accurately identify fault types and improve the accuracy of fault diagnosis in power systems. The analysis of simulation experimental results shows that this diagnosis can correctly and effectively identify fault types.

II. INTRODUCTION TO COMMON FAULT TYPES OF TRANSMISSION LINES

There are many types of faults in the power grid. However, from the perspective of fault feature analysis, most typical faults in the power grid including high resistance fault, single phase ground fault and other intermittent faults. In this section, these common fault types of transmission lines are introduced in detail.

A. HIGH RESISTANCE FAULT

In recent years, with the rise of global temperature and the occurrence of various extreme weather events, the aging rate of power lines has gradually accelerated. On the other hand, unpredictable events such as intentional damage to transmission lines and external force impacts by human factors also pose a serious threat to the safe operation of transmission lines. When a transmission line suddenly breaks due to external forces and environmental influences, and comes into contact with surrounding high impedance objects, it often leads to high resistance faults in the transmission line. After a fault occurs, the current level at this time is lower than the current level detected by short-circuit fault detection, which often brings difficulties and challenges to online monitoring and fault identification of transmission line status. If the fault cannot be identified and removed in time, it may cause more serious accidents such as fire, further expand the scope of Electrical fault, and bring greater economic losses.

High resistance faults usually occur in medium voltage distribution networks, and are typically caused by lightning strikes, strong winds, and other events that cause overhead lines to break and fall to the ground, causing them to come into contact with high resistance grounding media. Typically, single-phase grounding faults are the main type.

The main form of high resistance fault is the grounding and sagging of the line, which will not directly cause the protection device of the power grid to trip. However, the contact and discharge of the wire with the ground and tree branches can easily cause fire and damage to facilities such as roads, and there is a risk of expanding the accident hazards. For example, in early 2016, a 10kV overhead insulated cable in central China broke and fell to the ground after friction with a branch of the French tree, melting out large pits on the asphalt pavement. The sparks from the combustion caused a fire inside the surrounding residential courtyard walls; in the winter of 2018, a cable grounding occurred in a certain city in the western region, causing a grounding arc to cause a fire in the cable trench, resulting in power outages for tens of thousands of users [34]; in addition, according to the Western Power Coordination Commission in the United States, fires caused by power failures account for approximately 7.8%-9.6% of forest fires, of which 54% are related to medium voltage distribution network lines coming into contact with vegetation [35]. At the beginning of 2009, several serious ignition points of the ''Black Saturday'' wildfire in Victoria, Australia were caused by a tree collision on the line, resulting in 173 deaths and nearly $10 billion in damage [36].

B. SINGLE PHASE GROUND FAULT

Compared to high resistance faults, single-phase to ground faults occur in a higher proportion and frequency in the power system. Single phase grounding faults often occur in humid and rainy weather. There are many reasons for single-phase grounding faults, and the most common ones include insulator breakdown, single-phase lines, and contact with tall trees. When a single-phase ground fault occurs in the power system, it is often accompanied by the generation of overvoltage. This not only affects the normal electricity consumption of users, but also poses great harm to the safety of electrical loads, and even causes phase to phase short circuits and burns out electrical equipment.

C. OTHER INTERMITTENT FAULTS

This type of fault includes faults caused by line icing, pollution flashover, bird damage, and lightning strikes, which have the characteristics of randomness and unpredictability. When intermittent fault occurs, it is often accompanied by intermittent discharge, and arc light is often generated during discharge. In addition, the duration of intermittent faults is not fixed, ranging from a few seconds to several hours. In recent years, with the introduction of new materials, drones, and bird driving robots, the proportion and scope of such failures have gradually decreased.

III. PRINCIPLES OF FAULT DIAGNOSIS METHODS FOR TRANSMISSION LINES

The fault diagnosis of transmission lines mainly relies on the monitoring information fed back by the measurement devices installed in the power system. According to the type of data used, the operating status of transmission lines, and the monitoring method, fault diagnosis methods for transmission lines can be classified into single terminal distance measurement and double terminal distance measurement, online distance measurement and offline distance measurement, etc. In addition, when a fault occurs in the power system, some parameters of the power system tend to have corresponding sudden changes, such as voltage sag and current surge, and the measurement of voltage and current is relatively easy, so this section selects the phasor of voltage, the phasor of current, and the phase angle as the fault information elements of the system, uses multi SVM model to classify the fault information, and then judges the fault type of the transmission line. As shown in Fig. 1, SVM has unique advantages in processing small samples, nonlinear, and high-dimensional information, and is now widely used in the fields of classification recognition and fault diagnosis. The detailed principles are as follows.

Firstly, according to the symmetrical component method, the electrical components obtained at the time of fault can be decomposed into three components: positive sequence, negative sequence, and zero sequence, represented as P_+, $P-$, and P_0, respectively. The expression is as follows:

$$\begin{bmatrix} P_0 \\ P_+ \\ P_- \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix}\begin{bmatrix} R \\ T \\ S \end{bmatrix} \tag{1}$$

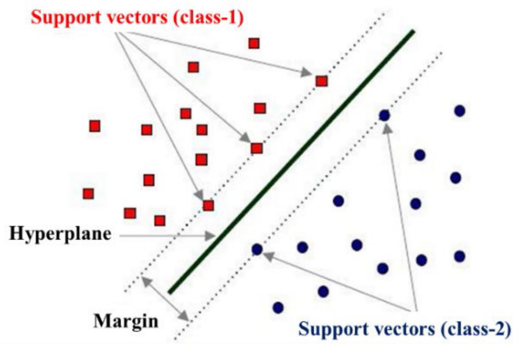

FIGURE 1. The structure of SVM model.

In equation (1), a represents 120 ° phase advance; R, T and S are the electrical quantities measured on the three-phase circuit. When a fault occurs, its zero sequence component is usually 0, but under asymmetric faults such as single-phase grounding faults, the zero sequence component is not 0, and the decomposed quantity should meet the following requirements (2)-(3):

$$|U_-| \ll |1 - U_+| \tag{2}$$

$$|U_0| < 0.004U_{pu} \tag{3}$$

Here, U_{pu} is the standard unit value of voltage, with a reference voltage of 220kV and a reference capacity of 334MVA. According to the above component method, different fault information components can be extracted and decomposed. Then, the method of multiple support vector machines is used to diagnose and classify the fault feature information obtained above. When a fault occurs in power system, the corresponding electrical information will change, such as voltage dips, current surges, etc. According to the definition of electrical phasor in IEEE Standard 1159:2009, voltage phasor, current phasor, and phase angle are selected as fault information to perform fault diagnosis. In theory, after the decomposition of fault voltage phasors other than single phase to ground faults, their zero sequence components should be zero. But to avoid interference from weak zero sequence components, the value is set as 0.004 in eq. (3)

The basic idea of SVM is to select a hyperplane in the data space. After the plane is selected, the collected data will be divided into two parts. In order to enable the obtained plane to correctly reflect the category of samples, we can apply the idea of optimization to transform the problem of constructing the optimal hyperplane into an optimization problem with certain constraints, which can be expressed as follows (4):

$$L(w, \theta) = \frac{1}{2}w^T w + \mu \sum_{i=1}^{n} \theta_i \tag{4}$$

Here, L represents the classification interval between the sample and the Hyperplane; w represents the weight value of the sample when separated; N is the number of samples; μ is a penalty factor, its value is usually greater than 0, which can adjust the loss caused by singular points on the system, balance the complexity of the algorithm, and sample error; θ_i representing relaxation variables and can ensure the accuracy of classification even when the sample cannot be linearly segmented in an ideal state. Kernel function is an important concept in SVM, which is used to map nonlinear separable data into high-dimensional space. Commonly used kernel functions include linear kernel function, polynomial kernel function, RBF kernel function, and sigmoid kernel function. Different kernel functions correspond to different mapping methods and have different processing capabilities. In practical applications, it is necessary to select appropriate kernel functions based on the characteristics of the problem. Linear kernel function is the simplest kernel function and can map raw data into linear space, which is suitable for linearly separable problems. For linearly indivisible problems, the performance of linear kernel function is poor; polynomial kernel function has high nonlinear mapping ability and is suitable for problems with complex nonlinear structures, but its computational complexity is relatively high; RBF kernel function is a widely used kernel function, which has strong nonlinear mapping ability and relatively low computational complexity, making it excellent in many practical problems; Sigmoid kernel function maps the raw data into a space close to the activation function of the neural network, which is suitable for problems with certain nonlinear structures. However, in some cases, sigmoid kernel function may cause the kernel matrix to be not positively definite, thereby affecting the training and prediction of SVM models. Therefore, the kernel function adopted in this article is the RBF kernel function and its expression is shown in (5):

$$F\left(l_i, l_j\right) = \exp\left(-\gamma \left\| l_i - l_j \right\|^2\right) \tag{5}$$

γ represents the parameter that determines the control range of the RBF function, which is called the width parameter; l_i and l_j are the basic eigenvectors used in the function. In this paper, a "1 to N" strategy is adopted to build a diagnostic model for multi-SVM (MSVM). The MSVM fault diagnosis box designed in this paper includes two SVM models, which are trained using different fault samples. The first SVM model is a binary classification model, which is used to determine whether there is a fault occurring at the current moment; the second SVM model is a three class model, which is used to determine the type of fault (single-phase ground fault, interphase short circuit fault, and three-phase ground fault). Considering the possibility of abnormal status, missed/false alarms of relay protection devices and data collection devices in the power system, this paper compares and fuses the diagnostic results of the two diagnostic boxes to obtain the final diagnostic result.

During the training process, the different values of parameters μ, γ, θ_i have a significant impact on the diagnostic accuracy of each SVM model. To improve the accuracy of fault diagnosis results, this paper uses genetic algorithm (GA) to optimize the parameter configuration of each SVM model, and classify the fault diagnosis results. The specific optimization steps are as follows:

(1) Encode the parameters μ, γ, θ_i that need to be optimized, and then generate chromosomes, where unknown variables x_1, x_2, and x_3 replace the aforementioned parameters to be optimized;

(2) In order to avoid the phenomenon of insufficient or overfitting in GA-MSVM model, this paper adopts a cross validation method to select the most suitable parameters for SVM. In the cross validation process, the obtained training set is divided into five equally sized subsets. During each validation, the first four subsets are used as the training set, and the fifth subset is used as the validation set to check whether each SVM is correct. Meanwhile, select f_{MAPE} as the indicator to measure the suitability of the validation, and its expression is shown in equations (6):

$$f_{MAPE} = \frac{\sum_{i=1}^{m} \left| \frac{RE_i - PR_i}{RE_i} \right|}{m} \times 100\% \tag{6}$$

Here, m is the number of samples in the training set; RE_i is the actual value; PR_i is the predicted value.

(3) Then, generate 20 sets of chromosomes and calculate the degree of adaptation of each set of chromosomes to the sample according to equation (6).

(4) After continuous selection, crossover, and mutation operations, replacing the old population with a new one, the probability of generating new chromosomes is adjusted to 0.8, and the probability of mutation is adjusted to 0.05.

(5) Repeat steps (3) and (4) until the upper limit of the number of iterations is reached.

The detailed flowchart of optimizing SVM using GA is shown in Fig. 2.

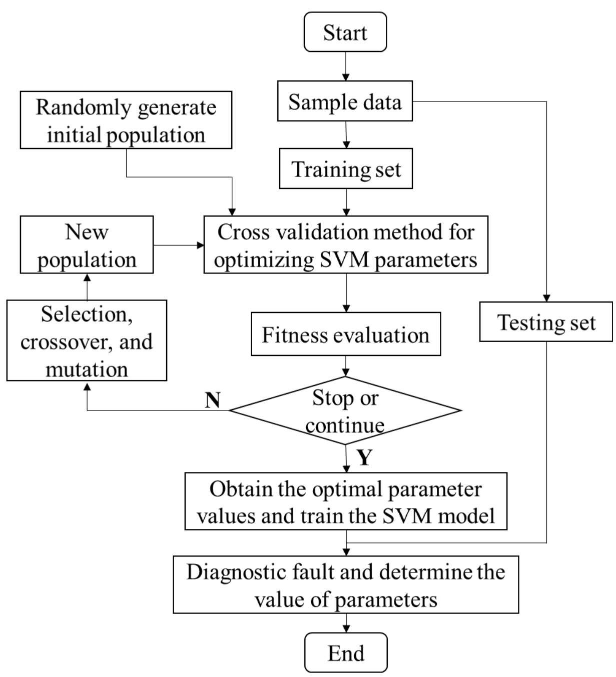

FIGURE 2. Flowchart of optimizing SVM using GA.

Finally, to gain a better understanding, a practical oriented research diagram is given. As shown in Fig. 3, when the proposed method is applied in practice, it can be divided into the following steps: (1) Data collection: Measurement equipment collects electrical data of transmission lines under normal working conditions and various fault conditions; (2) Data transmission: Transfer the collected data to the power system control center; (3) Raw data preprocessing: Adopt methods such as denoising and filtering to reduce the impact of noise on feature extraction; (4) Training and testing: Use symmetric component method to extract features from processed data and obtain electrical feature parameters that reflect fault information; input the filtered features into the SVM for training, and obtain a classifier that can distinguish different fault types; (5) Decision making support: Input the features of the data to be diagnosed into a trained SVM classifier to recognize fault types; (6) Line maintenance: Based on the fault diagnosis results, workers conduct line maintenance.

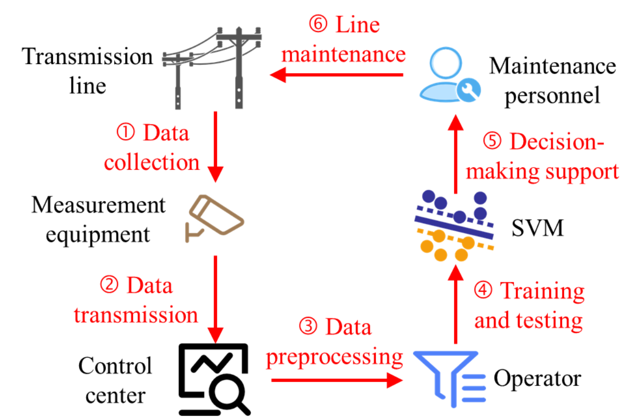

FIGURE 3. A practical oriented research diagram.

IV. CASE STUDY

In this section, two cases are given to verify the accuracy and feasibility of the proposed method. The first case uses the IEEE-30 standard testing system and the second case uses data from the actual power grid. This paper selects voltage, current, and phase angle as fault information. In these two cases, the data sources are collected by simulation and real power grid operation data from SCADA, which can provide normal and accident power flow information for the power grid. The fault information of transmission lines presents characteristics such as multiple sources, heterogeneous information, large quantity, and numerous attributes, making it difficult to ensure data integrity, effectiveness and consistency. In order to construct an accurate transmission line fault diagnosis model, it is necessary to effectively clean multi-source heterogeneous data and apply statistical and clustering methods to eliminate abnormal data. It should be noted that voltage and current data are usually recorded in the form of standard unit values, and there is a significant deviation between the voltage and current values under fault conditions and normal values. Therefore, standard normalization is used to preprocess the data, bringing the order of magnitude of each attribute in the input data closer to improve the accuracy of fault diagnosis. In addition, the location of the fault can be obtained based on the action of measuring equipment and relay protection equipment.

A. THE INTRODUCTION OF TEST SYSTEM

To verify the accuracy and feasibility of the proposed method to identify faults, this section firstly constructs an IEEE-30 node testing system simulation model in ATP-EMTP, sets corresponding faults, and analyzes fault feature information.

Fig. 4 shows the topology of IEEE-30 node testing system, in which faults are set for fault experiments. Table 1 shows the setting parameters of the simulation when three different single-phase to ground short circuit faults occur. The fault duration refers to the time from the system recording the start of the fault to its removal; the fault location refers to the distance between the location where the fault occurred and the monitoring station.

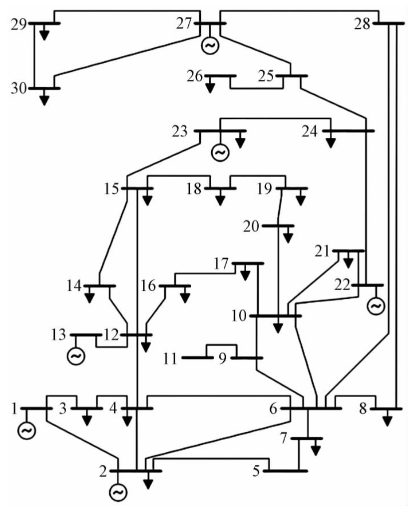

FIGURE 4. Topology of IEEE-30 testing system.

TABLE 1. ATP EMTP simulation test parameter settings.

Parameter	Type-1	Type-2	Type-3
Fault Type	F_{AG}	F_{BG}	F_{CG}
Fault occurrence time /ms	81.0	84.0	92.0
Fault duration time /ms	53.0	65.0	43.0
Fault Distance /km	74.58	25.96	103.97
Maximum fault current /kA	2.87	5.63	7.09
Maximum fault voltage /kV	54.71	109.71	205.51

Set a fault on the line between node 9 and node 10 in the system shown in Fig. 4. The fault type is single-phase ground short circuit fault. After using GA, the value of μ and γ are 6.7619 and 10.5438, respectively.

B. THE FEASIBLITY OF THE PROPOSED METHOD

After a fault occurs, the voltage and current waveforms of the line between two points are shown in Fig. 5 and 6.

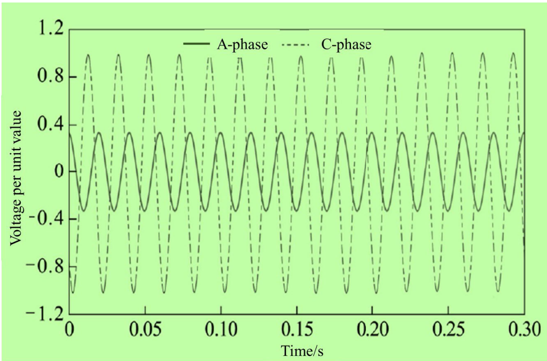

FIGURE 5. Voltage waveform changes of transmission lines under single-phase ground fault.

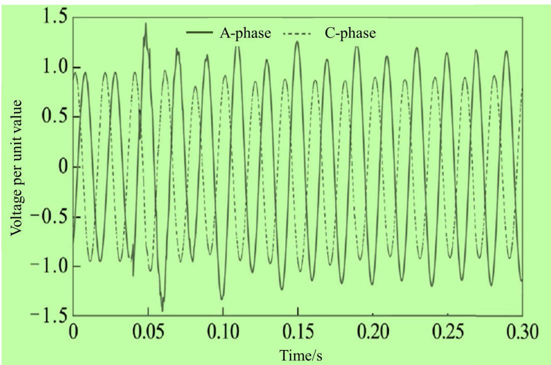

FIGURE 6. Current waveform changes of transmission lines under single-phase ground fault.

It can be observed that after a fault occurs, the voltage amplitude rapidly decreases while the current amplitude rapidly increases, indicating that electrical variables are very sensitive to faults. Almost at the moment of fault occurrence, voltage and current undergo corresponding changes. Therefore, using electrical information for fault diagnosis can greatly improve the system's response speed.

Extract the components of the voltage and current obtained in the above fault scenario, and perform fault diagnosis on them. The calculation accuracy comparison between traditional SVM model and the proposed model is shown in Table 2. It should be noted that the data sample contains a total of 200 sets of data, of which the first 100 sets are used for training and the last 100 sets are used for testing. By analyzing the test results, it can be seen that the designed fault diagnosis box has a correct rate of 86% and 85% for this group of faults, respectively. This indicates that the unoptimized setting of parameters has poor resolution ability for similar faults. Select appropriate optimization value of parameters and construct SVM, the results of testing using the optimized GA-MSVM model are also shown in Table 2. The accuracy rates of the two diagnostic boxes are 97% and 96%, respectively. From the above comparison, it can be clearly seen that the accuracy of the optimized GA-MSVM will be greatly improved, and the feasibility and effectiveness of this method have been verified.

TABLE 2. Comparison of fault diagnosis accuracy between the method proposed in this paper and traditional SVM models.

Different model	Fault diagnosis accuracy
Traditional SVM model to determine if the fault has occurred	86%
Traditional SVM model to determine the type of fault	85%
GA-SVM model to determine if the fault has occurred	97%
GA-SVM model to determine the type of fault	96%

In order to further demonstrate the accuracy of the method proposed in this paper, the method used in this section is compared with conventional methods. In the comparison of examples, the PSO-SVM model is used for effectiveness comparison. The PSO-SVM fault diagnosis box only contains one four classification SVM diagnostic model, which is used to determine the state of the line (normal operation, single-phase ground fault, interphase short circuit fault, and three-phase ground fault). The algorithm proposed in this paper can achieve an accuracy of 97% for the diagnosis of transmission line faults. The accuracy of fault diagnosis using conventional SVM models is 86%, and the accuracy of fault diagnosis using PSO-SVM models is 93%. It is not difficult to find that the fault diagnosis accuracy of the method proposed in this article is significantly higher than that of conventional SVM and PSO-SVM models, and the convergence time is approximately unchanged. In summary, the method proposed in this paper can significantly improve computational accuracy and is more suitable for on-site applications.

In addition, the original ratio of training and testing data is 5:5. When the ratio of training and testing data are 7:3 and 8:2, the fault diagnosis accuracy are 93.2% and 94.0%, which are slightly greater than that with original ratio. In fact, it is generally believed that the larger the training dataset, the better the performance of the model and the higher the diagnostic accuracy. The method proposed in this paper can meet the requirements of online computing and conduct relevant accident analysis. On the other hand, although relay protection equipment can achieve rapid fault isolation, due to some fault situations such as high resistance faults where the amplitude of various signals is small and the fault characteristics are weak, traditional relay protection equipment has difficulty in feature extraction, poor flexibility in threshold selection, and some extreme fault scenarios such as line ending faults and thousands of ohm level faults, which result in missed judgments and the low reliability of fault identification.

C. REAL-WORLD OPERATION DATA TEST

To further demonstrate the effectiveness and feasibility of the proposed method in this paper, a real-world case is given. Relevant real-world operation data in substation are adopted and detailed information are given in Table 3.

TABLE 3. Detailed information of real-world operation data in substation.

Type	Situation description	Number of samples
0	No faults occur.	2640
1	A-phase ground faults	1230
2	Faults between phase B and phase C	642
3	A-B phase ground faults	425
4	Faults between phase A, phase B and phase C	324
5	A-B-C phase ground faults	156

The method proposed in this paper is compared with the methods proposed in references [26] and [27] for case study. The selection results of the number of testing set and training set are shown in Table 4. The results of using different methods to diagnose transmission line faults are shown in Fig. 7 and Table 5. For the power system, it is mostly in a normal operating state and rarely experiences faults. The most common type of fault is single-phase grounded fault, and the frequency of three-phase grounded fault occurring is the lowest. Therefore, sample values are different for different situations of substation.

TABLE 4. Selection results of the number of training and testing set.

Type	Total number of samples	Number of training samples	Number of testing samples
0	2640	2112	528
1	1230	984	246
2	642	514	128
3	425	340	85
4	324	260	64
5	156	125	31

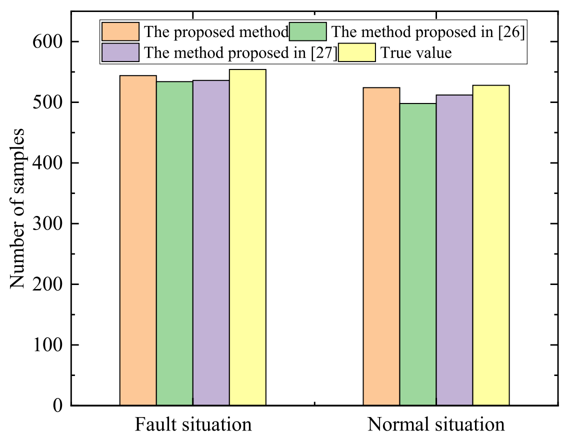

FIGURE 7. Diagnostic results of fault and normal status using different methods.

It can be observed from Figure. 7 and Table 5 that the method proposed in this paper has significant advantages in determining the normal or fault status of the transmission line. Compared with the above methods, the proposed method i.e., SVM model can guarantee better performance and accuracy of generalization. In addition, its advantages also lie in solving small samples, non-linear or high dimensional pattern recognition problems, with the ability to overcome the problems of course of dimensionality and over fitting. Through analysis, the accuracy of the method proposed in this paper is 98.2%, while the accuracy of the methods proposed in references [26] and [27] is 96.4% and 96.8%, respectively. The proposed method has higher accuracy. This is because the classifier used in references [26] and [27] essentially performs a normalization operation on the final classification result that conforms to the probability distribution. Its classification performance is easily affected by the number of samples and the distribution of sample categories. When the number of samples is small, the model is difficult to obtain sufficient training, resulting in classification results biased towards a larger number of categories.

TABLE 5. Type judgment results using different methods under different scenarios.

Fault type	Number of faults	Number of correct diagnoses by the proposed method	Number of correct diagnoses by method proposed in [26]	Number of correct diagnoses by method proposed in [27]
1	246	242	239	237
2	128	126	124	125
3	85	83	80	81
4	64	63	60	63
5	31	30	30	30

In addition, the issue of whether a fault has occurred is a binary classification problem. We choose four indicators to judge: true positive rate, true negative rate, false positive rate, false negative rate The specific definitions of each indicator are as follows:

Result 1: The transmission line is in normal operation.

Result 2: The transmission line is in a faulty state.

True positive rate: Judge result 1 correctly.

True negative rate: Judge result 2 correctly.

False positive rate: Judge result 2 as result 1.

False negative rate: Judge result 1 as result 2.

It can be observed from Table 6 that the indicators of the proposed method in this paper have made significant progress compared to the existing methods proposed by [26] and [27]. In fact, the improved MSVM model proposed in this article includes two SVMs. The first SVM is used to determine whether a fault has occurred; the second SVM is used to determine the type of fault. True negative rates are a common

TABLE 6. Comparison of indicators for different methods.

	The proposed method	The method proposed in [26]	The method proposed in [27]
True positive rate	99.24%	96.39%	96.97%
True negative rate	98.19%	94.32%	96.75%
False positive rate	1.81%	5.68%	3.25%
False negative rate	0.76%	3.61%	3.03%

indicator used to evaluate the effectiveness of SVM classification. Here, we use true negative rates to reflect the accuracy of the first SVM model in determining whether a fault has occurred. At the same time, the true negative rate is generally lower than the true positive rate, because compared to normal conditions, the electrical information of transmission lines in fault conditions will be more complex, making it more difficult to correctly identify faults. The computational time of the proposed method is 468 6s, which is less than the time required for on-site troubleshooting and transmission line maintenance. It should be noted that 468.6s mentioned in this paper is total calculation time instead of protection time. In practical application, the protection time usually takes tens of milliseconds. The total calculation time consists of the following components: data processing time, training time, parameter optimization time, and fault diagnosis time. It should be noted that 468.6s is the total calculation time for the first use of the GA-MSVM model. After the model training is completed, the subsequent calculation time will be shorter than the initial calculation time. Therefore, the proposed method can provide auxiliary decision-making for operators and meet practical engineering needs.

Finally, we also compared the fitness performance of the proposed method with the PSO-SVM model, and the results are shown in Figure 8. It can be clearly observed that the convergence of the method proposed in this paper is better, creating good conditions for online application and analysis. GA and PSO are both typical heuristic algorithms that attempt to simulate the adaptability of individual populations based on natural characteristics. They both use certain transformation rules to solve complex problems through search space, so PSO and GA are usually used to optimize SVM's model parameters. In the GA algorithm, chromosomes share information with each other, so the movement of the entire population is relatively uniform towards the optimal region. The particles in PSO only share information by searching for the current optimal solution, so this is largely a single item information sharing mechanism, and the entire search and update process follows the current optimal solution. In addition, GA already has mature convergence analysis methods, and can estimate the convergence speed.

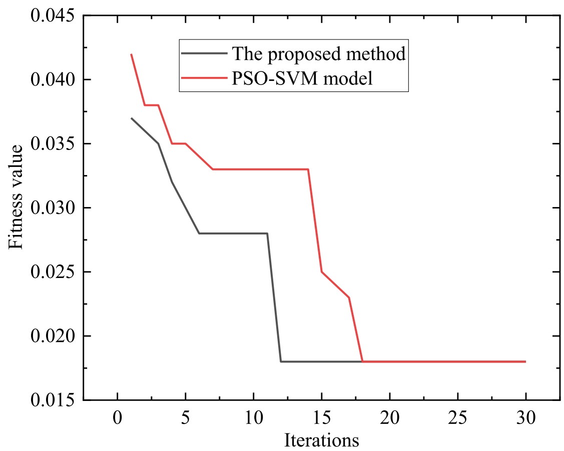

FIGURE 8. Fitness values comparison between different methods.

V. CONCLUSION

Focusing on how to fully utilize existing historical monitoring data for fault diagnosis of transmission lines, this paper first provides a detailed introduction to the causes and hazards of common faults such as high resistance grounding, single-phase grounding faults, and other intermittent faults, as well as fault diagnosis methods represented by the GA-MSVM model. Finally, the given cases demonstrate the accuracy and feasibility of the proposed method. Through analysis, the following conclusions can be summarized:

(1)The proposed method can fully extract and utilize fault feature information, especially with small samples, to ensure better performance and accuracy of generalization. Compared to traditional methods, the fault diagnosis accuracy of the proposed method can be maintained at over 95%, and can be improved by up to 11% compared to traditional methods. On the other hand, the proposed method has good convergence in terms of fitness values, which creates conditions for online applications and data analysis.

(2)The new principles of fault analysis for power systems with high consumption of renewable energy are potential next research directions. The theoretical basis of existing fault analysis is based on the assumption of the dominant fault state system variation characteristics of infinite synchronous generator units. Therefore, it is necessary to explore whether the basic principles of fault analysis will change when the DG is large-scale connected to the power grid.

(3) The combination of new algorithms such as artificial intelligence and traditional methods in the field of protection still needs to be studied. The complexity of fault features makes it difficult for diagnostic algorithms based on a single feature quantity and a single criterion to ensure reliability in various scenarios. With the continuous maturity of artificial intelligence algorithms and decision theory in engineering applications, it is worth further research on how to combine them with traditional methods.

REFERENCES

[1] H. Yu, L. Zhang, P. Zhao, Z. Liu, Z. Yang, M. Jin, and B. Hou, "Fault diagnosis of power transmission line based on elgamal encryption algorithm," in *Proc. IEEE 2nd Int. Conf. Electron. Technol., Commun. Inf. (ICETCI)*, Changchun, China, May 2022, pp. 953–957.

[2] J. Ma, X. Zhang, and X. Peng, "Simulation evaluation of lightning and non lightning fault identification of transmission line," in *Proc. 5th Int. Conf. Energy, Electr. Power Eng. (CEEPE)*, Chongqing, China, Apr. 2022, pp. 452–458.

[3] S. Lahiri, C. Abhijnan, and M. A. De, "Fault diagnosis in power transmission line using decision tree and random forest classifier," in *Proc. IEEE 6th Int. Conf. Condition Assessment Techn. Electr. Syst. (CATCON)*, Durgapur, India, Dec. 2022, pp. 57–61.

[4] S. Agarwal, A. Swetapadma, C. Panigrahi, and A. Dasgupta, "Fault detection in direct current transmission lines using discrete Fourier transform from single terminal current signals," in *Proc. 1st Int. Conf. Electron., Mater. Eng. Nano-Technol. (IEMENTech)*, Kolkata, India, Apr. 2017, pp. 1–5.

[5] F. Hao, X. Yang, G. Wang, and Y. Feng, "Transmission line fault diagnosis based on machine learning," in *Proc. 3rd Int. Conf. Consum. Electron. Comput. Eng. (ICCECE)*, Guangzhou, China, Jan. 2023, pp. 847–850.
[6] H. Yu, M. Liu, and S. Wang, "Research on fault diagnosis in the railway power transmission line based on the modern mathematical methods," in *Proc. 2nd IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Beijing, China, Oct. 2018, pp. 1–5.
[7] X. Peng, C. Li, and G. Zhang, "Research on electromagnetic transient simulation of transmission line fault identification," *High Voltage App.*, vol. 49, no. 8, pp. 8–15, 2013.
[8] V. B. Nuñez, J. Meléndez, S. Kulkarni, and S. Santoso, "Feature analysis and automatic classification of short-circuit faults resulting from external causes," *Int. Trans. Electr. Energy Syst.*, vol. 23, no. 4, pp. 510–525, May 2013.
[9] S. Kulkarni, D. Lee, A. J. Allen, S. Santoso, and T. A. Short, "Waveform characterization of animal contact, tree contact, and lightning induced faults," in *Proc. IEEE PES Gen. Meeting*, Jul. 2010, pp. 1–7.
[10] R. Hongdilokkul and C. Banmongkol, "Classification of transmission line faults with waveform characterization," in *Proc. 13th Int. Conf. Electr. Eng./Electron., Comput., Telecommun. Inf. Technol. (ECTI-CON)*, Jun. 2016, pp. 1–5.
[11] X. Peng, X. Li, and S. Yao, "Identification of transmission line fault causes based on transient characteristics of traveling wave current," *Southern Power Grid Technol.*, vol. 6, no. 5, pp. 43–47, 2012.
[12] Y. Chen, "Discussion on identification of transmission line fault causes based on fault traveling wave characteristics," *Guangxi Electric Power*, vol. 42, no. 6, pp. 59–63, 2019.
[13] A. A. M. Zin and S. P. Abdul Karim, "Protection system analysis using fault signatures in Malaysia," *Int. J. Electr. Power Energy Syst.*, vol. 45, no. 1, pp. 194–205, Feb. 2013.
[14] S. H. Asman, N. F. A. Aziz, M. Z. A. A. Kadir, and U. A. U. Amirulddin, "Fault signature analysis based on digital fault recorder in Malaysia overhead line system," in *Proc. IEEE Int. Conf. Power Energy (PECon)*, Dec. 2020, pp. 188–193.
[15] S. H. Asman, N. F. Ab Aziz, M. Z. A. Abd Kadir, and U. A. U. Amirulddin, "Fault signature analysis for medium voltage overhead line based on new characteristic indices," in *Proc. IEEE Int. Conf. Power Energy (PECon)*, Dec. 2020, pp. 198–203.
[16] X. Qin, P. Wang, Y. Liu, L. Guo, G. Sheng, and X. Jiang, "Research on distribution network fault recognition method based on time-frequency characteristics of fault waveforms," *IEEE Access*, vol. 6, pp. 7291–7300, 2018.
[17] H. Zhang and R. Che, "Fault cause identification based on characteristics of transition resistances for transmission lines," in *Proc. 5th Int. Conf. Electric Utility Deregulation Restructuring Power Technol. (DRPT)*, Nov. 2015, pp. 1405–1409.
[18] L. Cai, X. Zou, J. Wang, Q. Li, M. Zhou, Y. Fan, and W. Yu, "Lightning electric-field waveforms associated with transmission-line faults," *IET Gener., Transmiss. Distribution*, vol. 14, no. 4, pp. 525–531, Feb. 2020.
[19] H. Liang, Y. Liu, G. Sheng, and X. Jiang, "Fault-cause identification method based on adaptive deep belief network and time–frequency characteristics of travelling wave," *IET Gener., Transmiss. Distribution*, vol. 13, no. 5, pp. 724–732, Mar. 2019.
[20] W. Shi, L. Huang, and J. Xu, "Fault detection identification and location based on measurement data from finite synchronized phasor measurement," *Power Syst. Technol.*, vol. 39, no. 4, pp. 1067–1074, 2015.
[21] A. Salehi-Dobakhshari and A. M. Ranjbar, "Application of synchronised phasor measurements to wide-area fault diagnosis and location," *IET Gener., Transmiss. Distribution*, vol. 8, no. 4, pp. 716–729, Apr. 2014.
[22] S. Das and T. S. Sidhu, "Application of compressive sampling in synchrophasor data communication in WAMS," *IEEE Trans. Ind. Informat.*, vol. 10, no. 1, pp. 450–460, Feb. 2014.
[23] Q. Dong, Y. Zhao, and Z. Liu, "A locating method of earth faults in large-scale power grid by using wide area measurement system," *Proc. CSEE*, vol. 33, no. 31, pp. 140–146, 2013.
[24] W. Chen, Y. Chen, and W. Yao, "A fault diagnosis method using WAMS time series information and its application," *Power Syst. Technol.*, vol. 41, no. 6, pp. 1956–1963, 2017.
[25] D. Wei and D. Liu, "WAMS/SCADA data fusion method based on time-series data correlation mining," *High Voltage Eng.*, vol. 42, no. 1, pp. 315–320, 2016.
[26] C. Zhang, Y. He, T. Yang, B. Zhang, and J. Wu, "An analog circuit fault diagnosis approach based on improved wavelet transform and MKELM," *Circuits, Syst., Signal Process.*, vol. 41, no. 3, pp. 1255–1286, Mar. 2022.
[27] W. He, Y. He, Q. Luo, and C. Zhang, "Fault diagnosis for analog circuits utilizing time-frequency features and improved VVRKFA," *Meas. Sci. Technol.*, vol. 29, no. 4, Apr. 2018, Art. no. 045004.
[28] W. He, Y. He, B. Li, and C. Zhang, "Analog circuit fault diagnosis via joint cross-wavelet singular entropy and parametric t-SNE," *Entropy*, vol. 20, no. 8, p. 604, Aug. 2018.
[29] C. Zhang, S. Zhao, Z. Yang, and Y. He, "A multi-fault diagnosis method for lithium-ion battery pack using curvilinear Manhattan distance evaluation and voltage difference analysis," *J. Energy Storage*, vol. 67, Sep. 2023, Art. no. 107575.
[30] C. Qiushi and W. Yang, "Enhance high impedance fault detection and location accuracy via μ-PMUs," *Trans. Smart Grid*, vol. 11, no. 1, pp. 797–809, 2020.
[31] Q.-M. Xiao, M.-F. Guo, and D.-Y. Chen, "High-impedance fault detection method based on one-dimensional variational prototyping-encoder for distribution networks," *IEEE Syst. J.*, vol. 16, no. 1, pp. 966–976, Mar. 2022.
[32] W. A. dos Santos Fonseca, U. H. Bezerra, M. V. A. Nunes, F. G. N. Barros, and J. A. P. Moutinho, "Simultaneous fault section estimation and protective device failure detection using percentage values of the protective devices alarms," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 170–180, Feb. 2013.
[33] V. H. Ferreira, R. Zanghi, M. Z. Fortes, G. G. Sotelo, R. B. M. Silva, J. C. S. Souza, C. H. C. Guimarães, and S. Gomes, "A survey on intelligent system application to fault diagnosis in electric power system transmission lines," *Electric Power Syst. Res.*, vol. 136, pp. 135–153, Jul. 2016.
[34] W. Mingjie, "High resistance fault diagnosis technology and application in distribution networks based on broadband synchronous measurement," Shandong Univ., China, Tech. Rep., 2022.
[35] Western Electricity Coordinating Council (WECC). (2020). *Wildfire Webinar Series–Wildfire in the West*. [Online]. Available: https://www.wecc.org/Lists/WECCMeetings/DispForm.aspx?ID=15087&Source=https%3A%2F%2Fwww%2Ewecc%2Eorg%2FLists%2FWECCMeetings%2Fcalendar%2Easpx
[36] S. Kandanaarachchi, N. Anantharama, and M. A. Muñoz, "Early detection of vegetation ignition due to powerline faults," *IEEE Trans. Power Del.*, vol. 36, no. 3, pp. 1324–1334, Jun. 2021.

PEICHUAN SUN is currently an Engineer with State Grid Tianjin Electric Power Company, High Voltage Branch, Tianjin. His main research interest includes power system control and protection.

XUEFEI LIU is currently an Engineer with State Grid Tianjin Electric Power Company, High Voltage Branch, Tianjin. His main research interest includes power system control and protection.

MENG LIN is currently an Engineer with State Grid Tianjin Electric Power Company, High Voltage Branch, Tianjin. His main research interest includes power system control and protection.

JIE WANG is currently an Engineer with State Grid Tianjin Electric Power Company, High Voltage Branch, Tianjin. His main research interest includes power system control and protection.

TAO JIANG is currently an Engineer with State Grid Tianjin Electric Power Company, High Voltage Branch, Tianjin. His main research interest includes power system control and protection.

YIBO WANG is currently an Engineer with Wuhan Kemov Electric Company Ltd., Wuhan, China. His main research interests include power system control and protection and DC distribution network control and protection.