Single Line-to-Ground Faulted Line Detection of Distribution Systems With Resonant Grounding Based on Feature Fusion Framework

Ying Du, Yadong Liu, Qingzhu Shao, Lingen Luo, Jindun Dai, Gehao Sheng, and Xiuchen Jiang

***Abstract*—Faulted line detection is a key step of intelligent fault diagnosis of distribution systems, laying the foundation for the further fault location and service restoration. A novel single line-to-ground (SLG) faulted line detection method based on the feature fusion framework is proposed. In the proposed framework, one-dimensional convolutional neural network is employed as a powerful tool to extract more effective features. In addition, there is an imbalance phenomenon between data of the faulted line and healthy lines when a data-driven model is used in the faulted line detection. The proposed framework offers an avenue for overcoming it and improves the accuracy of detection. Considering the limited data of SLG faults in actual power systems, prior knowledge of SLG fault detection is integrated into the data-driven model, which proves useful in reducing dependence on the training data quantity. The experiments verified the superior performance of the proposed feature fusion framework-based method.**

***Index Terms*—Feature fusion framework, single line-to-ground fault, one-dimensional convolutional neural network, distribution systems with resonant grounding, prior knowledge.**

I. Introduction

DISTRIBUTION systems are the last link between electric utilities and customers. Many customers' perceived faults usually occur in distribution systems, so the reliability of which is crucial. The reliability is challenged by various faults, where single line-to-ground (SLG) faults are the most prevalent, accounting for 80% of all faults [1]. SLG faults result mainly from wind, falling trees, animals, birds and unfavorable wet weather conditions [2]. When a fault occurs, faulted line detection is the basis for further fault location, isolation and service restoration [3]. In distribution systems with resonant grounding, due to the reverse ground capacitive currents induced by arc suppression coils [4], the SLG fault currents are relatively small. Therefore, it is difficult to detect the SLG faulted line, especially with high impedance faults.

Manuscript received October 31, 2018; revised March 12, 2019 and May 31, 2019; accepted June 8, 2019. Date of publication June 12, 2019; date of current version July 23, 2019. This work was supported by the National Natural Science Foundation of China under Grant 51707117. Paper no. TPWRD-01336-2018. *(Corresponding author: Yadong Liu.)*

The authors are with the Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: duying@sjtu.edu.cn; liuyadong0916@163.com; 731536902@qq.com; llg523@sjtu.edu.cn; Jddai1993@qq.com; shenghe@sjtu.edu.cn; XCjiang@sjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRD.2019.2922480

Although distribution systems with grounding resonant are allowed to operate 1∼2 hours after occurrence of SLG fault [5], it is necessary for us to detect the faulted line immediately to prevent the fault from further deterioration. According to the report of China Southern Power Grid, in actual distribution systems, automation of detecting SLG faulted line with high accuracy has not yet been realized, and manual fault isolation is still necessary. Consequently, SLG faulted line detection of distribution systems with resonant grounding needs further study.

Transient zero sequence current of the faulted line is several times larger than the steady-state zero sequence current after occurrence of SLG fault, containing a lot of fault information. Many detection methods based on features of transient zero sequence currents and pattern classification methods have been proposed. Current waveform energy and normalized joint time-frequency moments are used to describe the features of waveforms of high impedance fault in [6] and support vector machine (SVM) is used for detection. On the basis of wavelet analysis, a more detailed and specific faulted line detection rule is presented in [7]. In [8], empirical mode decomposition (EMD) is used to select energy of high frequency band of transient zero sequence currents as features and ADABOOST is used as classifier. Multi-resolution morphological gradient is used to extract time-based features of signals in [9]. All of the above mentioned methods are based on artificially designed criteria to extract the features from data of lines and then using classification algorithms to recognize the faulted line.

However, transient signals based faulted line detection methods are significantly affected by the network electrostatic asymmetry, fault impedance and inception angle [10]. In addition, sometimes, there are some reverse installations of zero sequence current transformers in engineering, exacerbating the difficulty of detection. The SLG faulted line detection is faced with the dilemma that extracted artificial features are not effective enough, leading to lower reliability of detection under unfavorable conditions.

In addition to the way of feature extraction, there are other factors affecting the performance of detection. When a SLG fault occurs in a distribution system with N lines, the zero sequence currents data of N lines will be collected. In general, there is only one faulted line, so the fault data quantity is only 1/N of the total data quantity. N is more than ten in many areas. When a data-driven model processes these data of N lines, the imbalance phenomenon between data of the faulted line and healthy lines will make algorithms encounter difficulties. The data of healthy lines will be paid more attention to, while the most important information is concentrated in the data of the faulted line with smaller data quantity [11], [12].

Besides, the quantity of training data is of great importance for the performance of data-driven models. In a typical distribution network, the number of SLG faults of one line is difficult to reach one hundred per year. In fact, smaller data quantity is a big bottleneck for the application of data-driven models in power systems. In addition to spending considerable effort on simulating faults in actual systems, and given that we have mastered some prior knowledge based on the mechanism of power systems, integrating prior knowledge into data-driven model should be considered to reduce the dependency on data quantity and achieve better performance of detection [13].

In response to these three problems above, a novel SLG faulted line detection method based on feature fusion framework is proposed in this paper. The proposed feature fusion framework mainly includes three modules: (1) Data preprocessing module; (2) Feature extraction module; (3) Feature fusion module. These three modules address the three aforementioned problems respectively.

In the data preprocessing module, we collect healthy data and fault data of each line as training samples, and then balance the distribution of two-class data to avoid the unfavorable effect of the imbalance phenomenon on the performance of algorithms, which will be introduced in Section II in detail.

With the development of deep learning, adaptive feature extraction has become possible. Convolutional Neural Network (CNN), as one of the representatives of deep learning, is widely used in many pattern classification problems in various fields. For processing one dimensional signals, one dimensional Convolutional Neural Network (1D CNN) is widely adopted [14], [15]. In the feature extraction module, 1D CNN is employed as a tool to extract more effective features of data of each line.

In the feature fusion module, the combination of a data-driven model with prior knowledge is achieved by using the feature fusion method. The prior knowledge based features and features of data of N lines extracted by 1D CNNs are fused as a new feature vector.

Based on the feature fusion framework, the performance of SLG faulted line detection has significantly improved. The capability of feature extraction of CNN has proved to be more powerful. Overcoming imbalance phenomenon improves the accuracy of detection, and integrating prior knowledge into the data-driven model has proved useful in reducing the dependency on data quantity, which is promising in further application of data driven model in power systems.

The rest of this paper is organized as follows. In Section II, two previous options of faulted line detection based on data-driven model are analyzed, pointing out the direction of improvement. The proposed feature fusion framework is overviewed. Section III presents the implementation process of the proposed method in detail. In Section IV, the superiority of our method is verified and experimental results are discussed carefully. Section V concludes this paper.

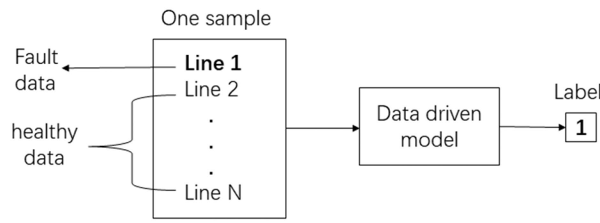


Fig. 1. Option 1 of data-driven model based faulted line detection methods. If line1 is the faulted line, the label is set as '1'. If line2 is the faulted line, the label is set as '2', and so on.

II. Proposed Method

A. Related Analysis

Previous faulted line detection methods based on data-driven models can be concluded into two options, which are named as "Option 1" and "Option 2" in this paper. The two options are analyzed in this section, and their limitations are discussed.

- Option 1

If there are N lines from the same bus in a distribution system, Option 1 selects relevant data of all N lines at fault moment as one training sample. Relevant data of each line are generally zero sequence currents collected after fault moment, i.e., there are zero sequence currents of all N lines in one training sample. The data-driven model can be chosen as neural network, SVM and so on. The label of a training sample is set as a digital mark of the faulted line. Option 1 is depicted in Fig. 1.

There is an imbalance between fault data and healthy data in Option 1. At most failures, there is only one SLG faulted line, which means inside a sample, fault data are only 1/N, and the rest are healthy data. This imbalance phenomenon is special, arising in the context of faulted line detection problems. It is different from typical imbalanced classification problems that the number of training samples is unevenly distributed among the classes [16]. In other words, the number of samples in some classes is much smaller than those in other classes. However, there is always an imbalance phenomenon between fault data and healthy data in Option 1, which exists in the distribution of fault data and healthy data inside each sample.

Apparently, this imbalance phenomenon has some unfavorable influences on the performance of algorithms. The healthy data of samples will catch more attention due to the superiority of data quantity in the learning process of algorithms, leading to insufficient learning of fault data. Consequently, this imbalance phenomenon needs to be overcome urgently. Furthermore, combining data of all lines into a sample makes Option 1 difficult to consider the difference among lines.

- Option 2

In Option 2, the data-driven model trains data of each line separately. Faulted line detection problem is transformed from an N-class problem of Option 1 to a two-class problem. The label of this model is set to be 1 or 0. Label 1 represents a faulted line and label 0 represents a healthy line. When a fault occurs, by comprehensive consideration on the output of the model of each line, the faulted line can be selected. Option 2 is shown as Fig. 2.

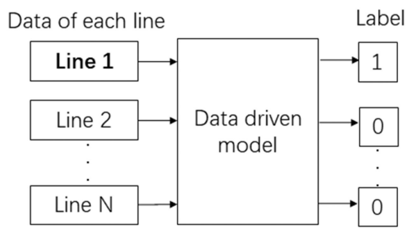


Fig. 2. Option 2 of data driven based faulted line detection methods. If line1 is a faulted line, the label is set as '1'. If line2 is a healthy line, the label is set as '0', and so on.

Option 2 attempts to train a classifier which could tell whether a waveform is normal or not. This work is very challenging and almost infeasible since actual conditions of distribution network are unpredictable, including but not limited to the change of fault resistance, fault distance and fault phase. These factors have a great influence on the collected waveforms after fault moment. Moreover, out of randomness, some factors are difficult to consider, such as changes in the meteorological environment. It is impossible for us to select data under all conditions as training data, consequently, the result of Option2 has lower credibility. We think that the correlation of lines should be considered, and it is sensible to detect the faulted line by associating and comparing the features of data of all lines.

In addition, there is some prior domain knowledge in faulted line detection of distribution systems, which can be combined with features extracted by the data-driven model to further improve the performance of faulted line detection method and reduce the dependency on the data quantity.

By analyzing the two options above of faulted line detection methods, we can conclude that the imbalance phenomenon, the difference and the correlation of lines, and integrating prior knowledge all should be considered for a more desirable faulted line detection method.

B. Proposed Feature Fusion Framework

In this paper, a feature fusion framework based SLG faulted line detection method is proposed. The proposed feature fusion framework comprises four modules: (1) Data preprocessing; (2) Feature extraction; (3) Feature fusion; (4) Classification. The two most important modules are feature extraction and feature fusion. The simplified diagram of the feature fusion framework is exhibited in Fig. 3.

As can be seen in Fig. 3, the features of data of each line are extracted separately. In the data preprocessing module, fault data and healthy data of each line are collected after fault moment, which constitute the training data set of each line. On this basis, the imbalance phenomenon inside each sample in Option 1 can be transformed to imbalance between two classes (faulted line or healthy line). Because of the transformation, we can handle the imbalance by using methods widely used in typical imbalanced classification problems. In the proposed framework, the synthetic minority oversampling technique (SMOTE) algorithm [17] is chosen to overcome the imbalance between two classes by generating new fault data.

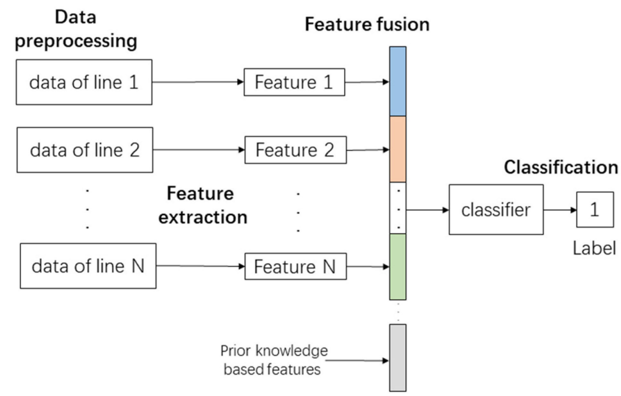


Fig. 3. The diagram of the proposed feature fusion framework. The label of the training set is a digital mark of the faulted line as Option 1.

In the feature extraction module, because features of data of each line are extracted respectively, the difference of lines can be considered. We use 1D CNNs to extract features of data of each line. If there are N lines in a distribution system, we need to train N 1D CNNs. In the training process, transfer learning is adopted as the method of initializing parameters, which benefits network convergence and reduces computational work, thereby simplifying the training process.

Feature fusion module plays two roles. First, correlation of lines is considered. Second, prior knowledge is integrated into the data-driven model. Recently, feature fusion methods are widely used in many pattern classification problems. Fusing diverse features can provide more effective information for the data-driven model and then improve the performance of classification [18], [19]. Because it is essential for us to pay attention to the correlation of lines, features extracted from data of each line are fused as 1D CNN based features. Then, in order to combine the data-driven model with prior knowledge, prior knowledge based features and 1D CNN based features are fused as a new feature vector, which is input to the classification module. In the classification module, the neural network is chosen as the classifier.

We can conclude that in the proposed feature fusion framework, there are three significant advantages:

- The difference and correlation between lines are considered at the same time.
- It provides a way for overcoming the imbalance phenomenon inside samples.
- It achieves the integration of prior knowledge into the data-driven model.

The implementation of it will be further illustrated in the next section in detail.

III. Implementation

When a SLG fault occurs, the symmetry of three-phase voltage is destroyed, so the zero sequence voltage and the zero sequence current of lines are increased, which can be used as the detection signals. However, the changes of current and voltage caused by normal switching operation may make detection system start to detect the faulted line mistakenly. In order to avoid

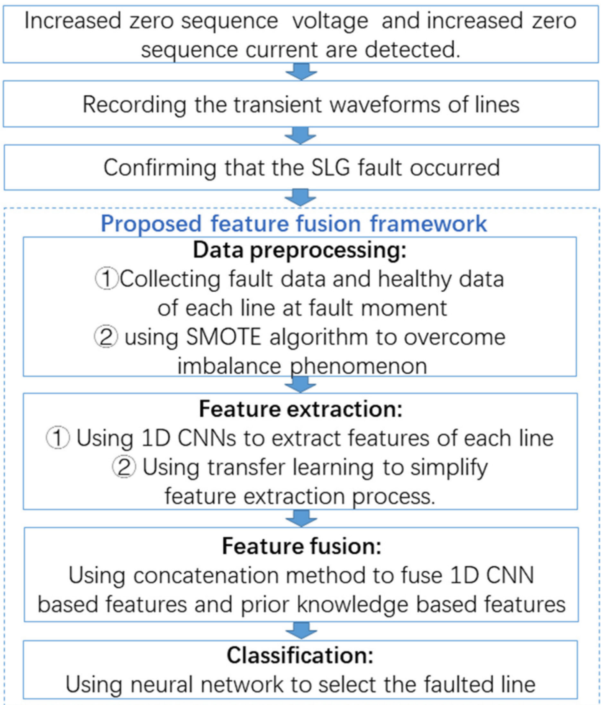


Fig. 4. The flow diagram of the proposed detection method.

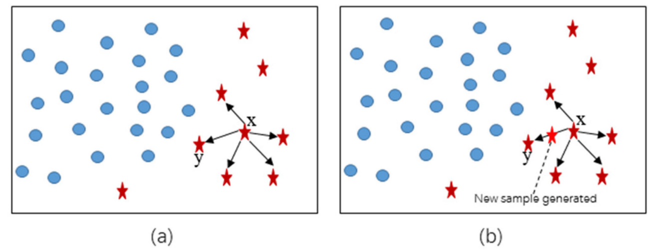


Fig. 5. New sample generation process of SMOTE algorithm. (a) Finding the same K-nearest neighbor samples of the main sample x ($K = 5$, in this example). (b) New samples generation.

TABLE I
LIST OF NOTATIONS USED IN THIS SUBSECTION

Notations	Meaning
x_t	A discrete sequence
w_k	Convolution kernel
h_t	Output of the first convolution layer
$ReLu$	Activation function
a_i^l	Input of the *i-th* neuron of the *l*-th layer
b	Bias parameter
R_k	The region divided by a pooling layer

the influence of transient disturbance, we set three cycles delay for confirmation of SLG fault occurrence because most transient disturbance will disappear in three cycles. After three cycles delay, if the zero sequence voltage is still an abnormally increased value, the proposed feature fusion framework starts to detect the faulted line. After the detection signal occurs, the transient waveforms of lines will be recorded first, so fault detection will not be affected by the delay. The flow diagram of the proposed method is shown in Fig. 4.

Firstly, fault data and healthy data of zero sequence current of each line are collected. Then, the SMOTE algorithm is used to generate enough fault data to balance data distribution of each line, making a contribution to the subsequent feature extraction processes. Next, 1D CNN is employed as a powerful tool to extract features of data of each line. By using feature fusion, the 1D CNN based features and prior knowledge based features are fused as a new feature vector. Finally, the fused feature vector is input to the neural network to recognize the faulted line.

A. Data Preprocessing

In the data preprocessing module, the SMOTE algorithm is used to generate a large number of new samples with certain strategies to balance the distribution of training samples. In general, in addition to the noise factor, some samples tend to appear in the neighborhood space of samples from the same type, which is as the criterion for SMOTE algorithm to generate new samples in the neighborhood space of original minority samples. In SMOTE algorithm, k-nearest neighbor algorithm is used to ensure neighborhood space [20]. Firstly, select a main sample x in minority samples and then find its k-nearest neighbor samples in all remaining minority samples, and then randomly select a main neighbor from them, which is named as y. Next, generate a new sample at a random location of the line connecting the main sample x and its main neighbor sample y. The new sample generation process of SMOTE algorithm is as shown in Fig. 5.

B. Feature Extraction

CNN is considered as a breakthrough in computer vision [21] and has become the research hotspot in pattern classification problems of many fields. In [22], the transient zero sequence current waveforms are converted into time-frequency images by continuous wavelet transform and then CNN is used to extract features of images. However, the rising dimension process will bring some redundancy inevitably, complicating the feature extraction. In this paper, 1D CNN is employed to extract the features of data of each line by the movement of the convolution kernels on the time axis.

1) Architecture of CNN: Typical CNN is a hierarchical model, including an input layer, convolution layers, activation function, pooling layers, fully connected layers and output layer mainly. The architecture of the 1D CNN is as shown in Fig. 6.

In the convolution layers, there are a set of convolution kernels with a certain size. The essence of the convolution kernel is the local weight matrix, which can extract linear features effectively by convolutional operations and experts in acquiring local information. Different convolution kernel can extract the features from diverse aspects.

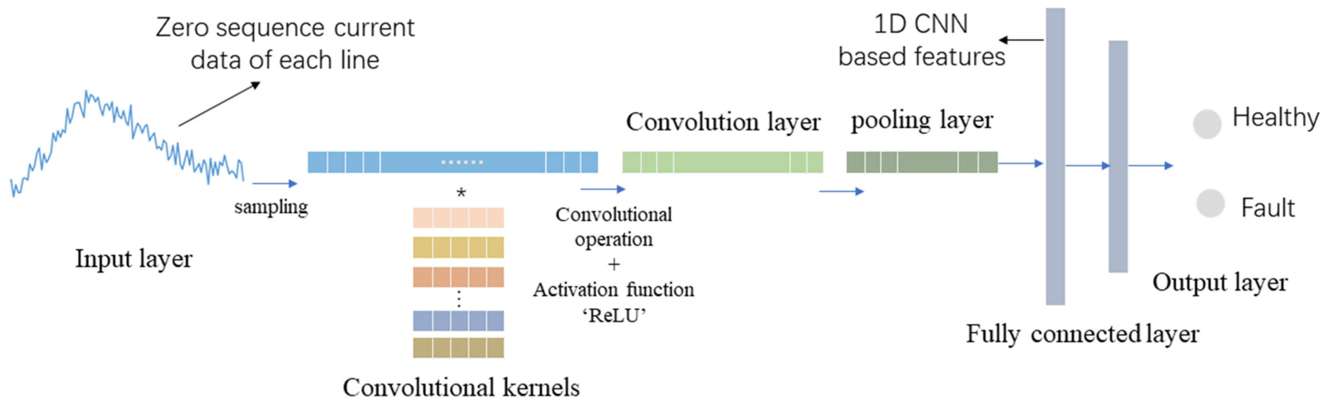


Fig. 6. The structure of 1D CNN.

The convolutional operation of 1D CNN can be described as follows.

$$h_t = \sum_{k=1}^{m} w_k \times x_{t-k+1} \tag{1}$$

The output of convolution layer is represented as h_t. A discrete sequence x_t, where t = *1, 2,…, n*, is a one dimensional time-series signal with n time steps. A convolution kernel is represented as w_k, where k = *1, 2, …,m.* In general, *m* is much smaller than *n*.

Because convolutional operations only extract linear features, the nonlinear activation function is used to add nonlinear mapping after convolutional operations. In this paper, the Rectified Linear Unit (ReLU) [23] is used as activation function, which can avoid the occurrence of gradient saturation and help the convergence of the stochastic gradient descent method, and the speed of it is about 6 times faster than that of others [21]. The definition of the ReLU is as follows:

$$ReLU\,(h) = \max\{0, h\} \tag{2}$$

The definition of the input of the *i-th* neuron of the *l*-th layer is as follows.

$$\begin{aligned} a_i^l &= \mathrm{Re}LU\left(\sum_{j=1}^{m} w_j^l \times a_{i-j+m}^{l-1} + b^l\right) \\ &= \mathrm{Re}LU\left(w^l \times a_{(i+m-1):i}^{l-1} + b^i\right) \end{aligned} \tag{3}$$

In equation (3), $a^l_{(i-m+1):i} = [a^l_{(i+m-l)}, \ldots, a^l_i]^T$. w^l is the convolution kernel with *m* dimensions. b^i is the bias parameter, where *i* = *1, 2,…, n*.

Next, pooling layers simplify the output of previous layers, obtaining feature maps with lower resolution. Actually, the operation of pooling layers is the down-sampling process, which can avoid over-fitting by reducing parameters to some extent. It pays more attention on whether there are some features not the location of some features. Therefore, pooling operations can enhance the robustness of CNNs.

Max-pooling is used in this paper. The output of the previous convolutional layer can be divided into several regions R_k, k $= 1, 2, \ldots, K$. The output of the pooling layer is

$$pool_{\max}\,(R_k) = \max_{i \in R_k} a_i. \tag{4}$$

The output of pooling layers can be regarded as features extracted by 1D CNN. Then, the features are input to the fully connected layers, which work as a classifier, mapping extracted features to the label.

2) The Training of CNN and Transfer Learning: The train of CNN is more complicated than the artificial neural network and back propagation algorithm is also used. To reduce the complexity of computation in the training phase, weight sharing [21] is adopted, which reduces the parameters of network extremely.

Data normalization is an essential step before beginning of the training process, which can cancel errors caused by different dimension, self-variation or large difference in values. In order to prevent the over-fitting phenomenon and enhance the generalization capability of the model, the regularization technology is usually used.

The value of initial weights of model has a significant influence on the training process, greatly determining the final performance of the network. Transfer learning is a good choice in the weight initialization, using the parameters of the pre-training model, which is simple and effective. Transfer learning moves the knowledge of one domain to another related domain so that the learning process of the related domain can perform better. The upper layer response of deep CNN is more general and can be used for different classification task [18]. Usually, the transfer learning approach is to train a base network, and then we can use its first several layers to initialize the first several layers of the target network. Next, the remaining layers of the target network are trained toward the target task with randomly initialized weights.

In this paper, transfer learning is used in weight initialization. In a distribution system with N lines, we need to train N 1D CNNs to extract features of each line. Given that these N networks belong to the same domain (all are zero-sequence transient current signals), we can choose one of N networks as a base network, and then copy its first several layers to the first several layers of other N-1 networks to initialize weights of the N-1 networks. In this way, these N-1 networks do not need to learn starting from the beginning but on the basis of the base network, improving the efficiency of feature extraction and simplifying the training process.

The Adam optimization algorithm is employed in the CNNs. It is an extension of the stochastic gradient descent algorithm, which has recently been widely used in deep learning applications to iteratively update neural network weights based on training data. It is suitable for solving optimization problems with large-scale data and parameters [24].

C. Feature Fusion and Classification

In order to associate 1D CNN based features of each line and integrate prior knowledge into the data-driven model, the feature fusion method is used in the proposed framework. Feature fusion has been applied in various pattern classification problems to improve the accuracy of classification. In [18], anatomical features and features extracted by transfer learning are fused to enhance the detection of automatic eye type, which maximizes the classification accuracy. In [19], the feature fusion method has been employed to process video, audio and text data to do multimodal deception detection, achieving better accuracy. Different issues can use different ways to fuse the related features. In the proposed framework, a simple but effective method is chosen. We concatenated the 1D CNN based features of each line and prior knowledge based features to fuse them, which is also used in [18], [19], [25], [26].

Deep neural network architectures such as CNN, are commonly trained using a large number of supervised examples. However, the data of actual SLG faults are limited. The prior knowledge of SLG faulted line detection should be used effectively to reduce the dependency on the data quantity in the data-driven model. In general conditions, the amplitude of zero sequence transient current of the faulted line is greater than that of other healthy lines, and the polarity of zero sequence transient current of the faulted line is opposite to the polarity of other healthy lines. Therefore, the amplitude and polarity of zero sequence transient currents can be used as criterion of faulted line detection [27], which are used as prior knowledge based features in the framework. The 1D CNN based features of each line and prior knowledge based features are concatenated as a new feature vector. The diagram of feature fusion and classification is as shown in Fig. 7.

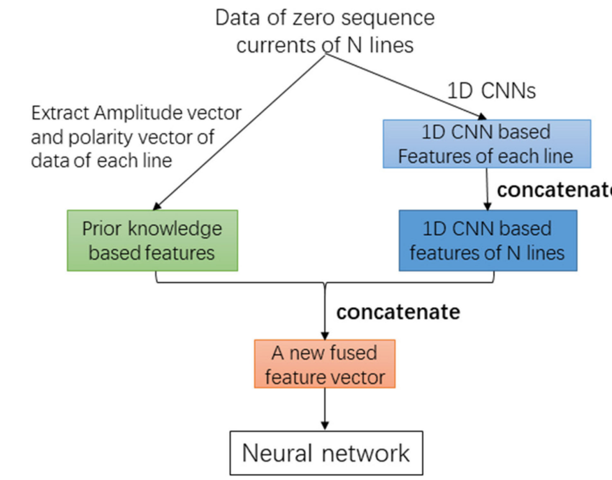


Fig. 7. The diagram of feature fusion and classification.

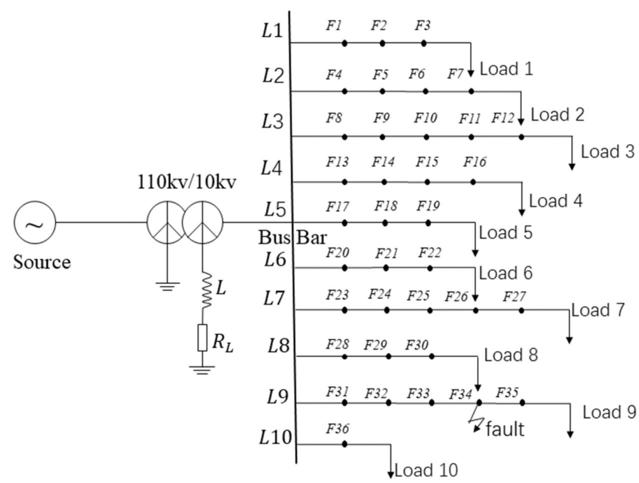


Fig. 8. The structure of simulation model.

TABLE II
TYPE AND LENGTH OF TEN LINES

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
Type	*O*	*O*	*O*	*O*	*U*	*U*	*U*	*U*	*U*	*U*
Length (km)	20	25	30	25	20	20	30	20	30	10

IV. EXPERIMENTS AND DISCUSSION

A. Simulation

1) Simulation Model of Single Line-to-Ground Fault: A model has been established in MATLAB/SIMULINK to simulate single line-to-ground faults in the typical 10-kV distribution system with resonant grounding. The structure of the simulation model is as shown in Fig. 8. This is a hybrid distribution system, including overhead lines and underground lines.

The simulation model consists of ten feeder lines L1-L10, where L1-L4 are overhead lines; L5-L10 are underground lines. The type and length of each line are as shown in Table II. O denotes the overhead line and U denotes the underground line. F1-F36 in Fig. 8 denote different fault location. The load of each line is set as (1+j0.4) MVA. The arc suppression coil is 10% overcompensated. The capacitance to ground of the underground line is much larger than that of the overhead line, which increases the difficulty of faulted line detection in hybrid distribution networks. The parameters of lines are as shown in Table III.

2) Simulation of SLG Faults Under Different Conditions: SLG faults are simulated under different fault conditions, involving different faulted lines, different fault phase, different fault resistance and different fault locations (F1-F36). The fault resistance is set from 5 to 23 kΩ, taking the high impedance faults into account. The fault phase means the phase to ground voltage angle when fault occurs.

TABLE III
PARAMETERS OF TEN FEEDER LINES

Type	Phase-sequence	R (Ω/km)	L (mH/km)	C (μF/ km)
underground line	Positive-sequence	0.265	0.249	0.339
	Zero-sequence	2.700	1.012	0.280
Overhead line	Positive-sequence	0.127	1.300	0.0096
	Zero-sequence	0.275	4.600	0.0054

TABLE IV
PARAMETERS OF DIFFERENT FAULT CONDITIONS

Faulted line	Fault location	Fault phase	Fault resistance
L1 L5 L6 L8	F1~F3 F17~F19 F20~F22 F28~F30	0° 90°	5,100,1k ,3k, 5k, 7k,10k,11k,12k, 13k,15k,17k,19k, 20k,21k,22k,23k
L2 L4	F4~F7 F13~F16	0° 90°	5,100,1k ,3k, 5k, 7k,10k,12k,13k, 15k,17k,19k,21k, 22k,23k
L3 L7 L9	F8~F12 F23~F27 F31~F35	0° 90°	5,100,1k ,3k,5k, 10k,12k,16k, 18k,20k
L10	F36	0° 30° 60°90°	5,100,1k,2k,3k,4k,5k, 6k,7k,8k,9k,10k,11k, 12k,13k,14k,15k,16k, 17k,18k,19k,20k,21k, 22k,23k

We perform 1000 simulations to generate experimental data of next subsection. One of ten lines faulted in each simulation. Each line faulted with 100 times. After the SLG fault occurs, the first half cycle transient zero sequence current waveform data of ten lines are sampled under sampling rate of 10000 Hz. The parameters of different fault conditions are as shown in Table IV.

B. Experiments

In order to explore the effect of data quantity on SLG faulted line detection methods, experiments are carried out under data quantity of 1000 and 500 respectively. The data under 500 faults come from a random sampling of data under 1000 faults. The number of faults in each line is also the same.

In each experiment, the original data set is randomly shuffled and then the training data set and testing data are split by the ratio 7:3. In order to simulate environmental disturbance and enhance the robustness of models, Gaussian white noise with power spectral density of 0.1, 0.0001 and 0.000001 is added to training data respectively. The Gaussian white noise with power spectral density of 0.000001 is added to the testing data.

In actual power systems, sometimes, there are some zero sequence current transformers (CTs) installed reversely due to human error. Due to the change of the polarity of the original zero sequence current, reverse CT installation will increase the difficulty of faulted line detection. In order to simulate the actual power system and enhance the robustness of detection models, we consider this unfavorable condition in our simulation. There are 30% reverse zero sequence CTs installation in the faulted line. For example, among 100 failures of L1, zero sequence CT of L1 is reversely installed with 30 times. The distribution of training data and testing data is as shown in Table V.

For the purpose of verifying the performance of the proposed feature fusion framework based faulted line detection method, we compare it with the other four methods. Method 1 is one of the traditional methods, which is depicted as Fig. 9(a). In Method 1, the prior knowledge based features are extracted manually, which are input to the neural network. In our experiments, the extracted prior knowledge based features are the amplitude and polarity vectors of ten lines in the selected frequency bands [27]. The selected frequency is from 150 Hz to the minimum series resonant frequency in ten lines themselves. We calculate the peak value of the half-cycle zero sequence current data of each line. The peak value of each line is set as the amplitude feature of each line. We take five time points t_i, $i = 1, 2, 3, 4, 5$, around the peak value of the half-cycle data of each line. $A(t_i)$ represents the amplitude of zero sequence current at time t_i. The polarity vector of one line is set as $\boldsymbol{x}$, a five dimensional vector, where $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)$. If $A(t_i) > 0$, $x_i = 1$, otherwise, $x_i = 0$.

Method 2 is as depicted in Fig. 9(b). CNN is used as the feature extractor and classifier. Method 3, shown in Fig. 9(c), integrates the prior knowledge based features and CNN based features. In Method 4, as shown in Fig. 9(d), the features of lines are extracted respectively by 1D CNN after the data preprocessing for overcoming data imbalance. And then, the features of lines are fused to input to neural network to detect the faulted line.

The definition of the accuracy of faulted line detection is the ratio of the samples detecting the faulted line correctly to total samples of testing data set. The accuracy of the proposed method and four comparison methods under data quantity of 1000 and 500 are exhibited in Table VI.

C. Discussion

As shown in Table VI, with the increase of data quantity, the accuracy of detection of all methods has increased. The reason is that the data-driven model is mining rules from data. More data bring more utilizable information.

Regardless of the data quantity, Method 2, Method 3, Method 4 and the proposed method all outperformed Method1. This result fully demonstrates the superiority of features extracted by CNN. CNN is able to extract features from data adaptively, which proved to be more effective and reliable than the features extracted manually. As mentioned above, 30% reverse zero sequence CTs are considered in the simulation, which will bring a great challenge to Method 1, causing the polarity criterion to fail. In addition, sometimes, the amplitude criterion could be affected by high fault resistance and environmental noise.

Two cases are exhibited to verify above inference about the reason why Method 1 has the lowest accuracy of detection. In order to be seen clearly, only zero sequence current waveforms of L1, L2 and L7 are exhibited in the two cases. In both cases,

TABLE V
THE DISTRIBUTION OF TRAINING AND TESTING DATASET

Data quantity	The number of training data	The number of samples with reverse CTs in training data	The number of testing data	The number of samples with reverse CTs in testing data
1000	700	210	300	90
500	350	105	150	45

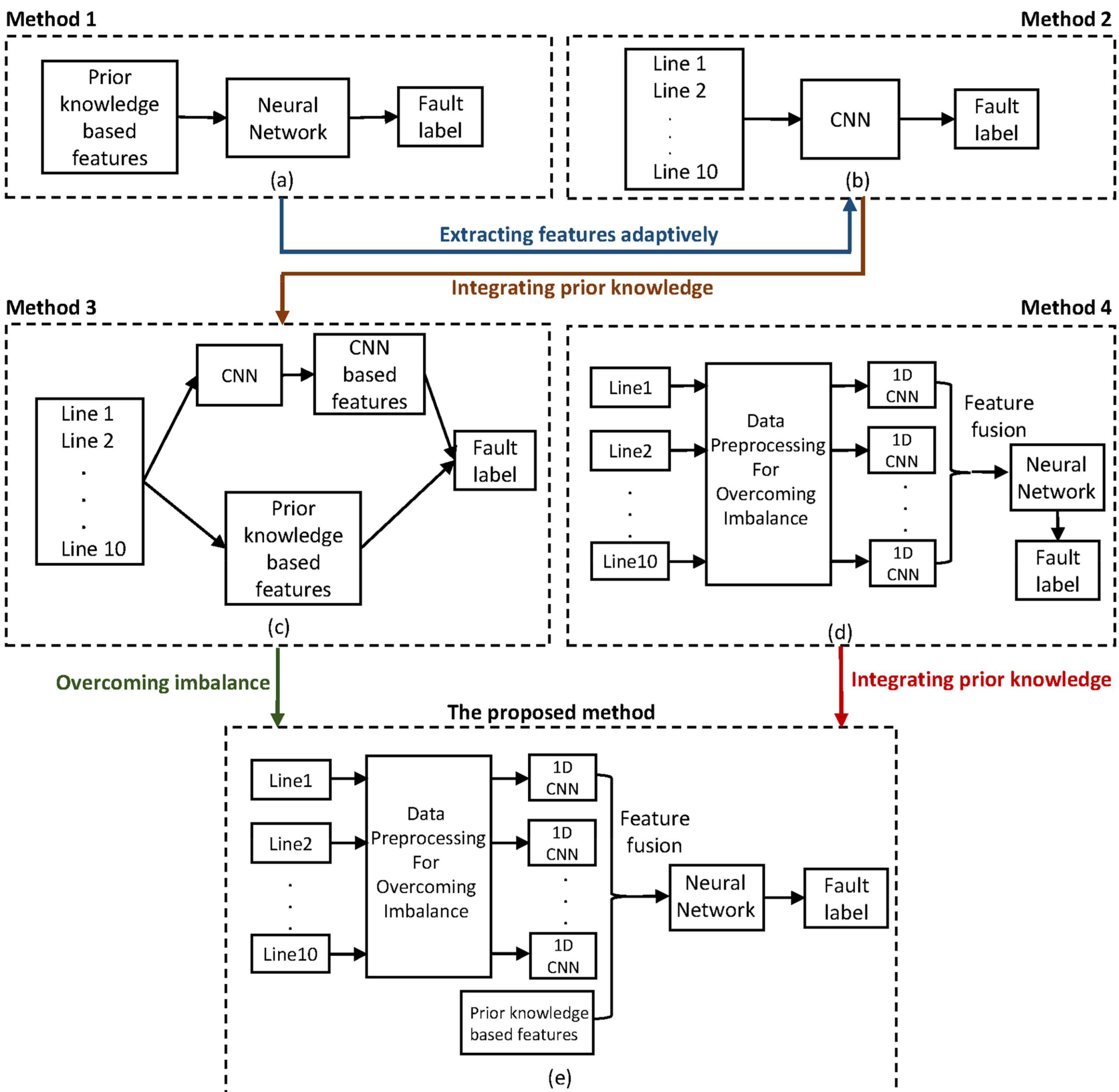


Fig. 9. The proposed method and contrast methods. The texts next to the arrows are used to describe the advantages of the methods pointed by the arrows.

TABLE VI
THE ACCURACY OF DETECTION OF DIFFERENT METHODS UNDER DIFFERENT DATA QUANTITY

Data quantity	Method 1	Method 2	Method 3	Method 4	The proposed method
1000	0.843	0.950	0.963	0.977	0.983
500	0.713	0.833	0.867	0.860	0.880

TABLE VII
THE ACCURACY REDUCTION FROM DATA QUANTITY OF 1000 TO 500

	Method 1	Method 2	Method 3	Method 4	The proposed method
Accuracy reduction	0.130	0.117	0.096	0.117	0.103

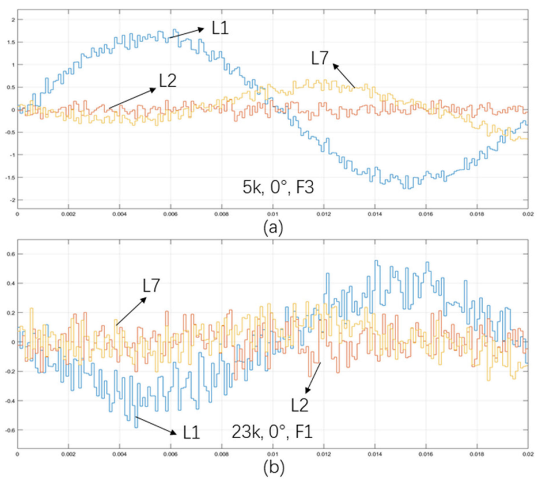


Fig. 10. Two cases for discussing when Method 1 fails. Method 1 works in the Fig. 10(a) and fails in Fig. 10(b). The fault condition is as shown in the bottom of pictures. The CT of L1 is reversed in Fig. 10(b).

the faulted lines are L1. In Fig. 10(a), Method 1 works. The amplitude and polarity characteristics of L1 are obvious, i.e., the amplitude of current of L1 is much larger than those of other lines and the polarity of L1 is opposite to those of other lines. In Fig. 10(b), Method 1 fails. Note that the zero sequence CT of L1 is installed reversely. In addition, the amplitudes of zero sequence currents of all lines are all quite low due to the high resistance. We can conclude that the artificial features based methods are prone to fail in high resistance fault detection or when the zero sequence CT is installed reversely. In contrast, these unfavorable conditions have less influence on deep learning models.

Method 4 has better performance than Method 2 under the data quantity of 1000 and 500, indicating the framework of Method 4 is more effective. Method 4 overcame the imbalanced data distribution of fault data and healthy data, which brings an increase in accuracy. At the same time, the result shows that the imbalance phenomenon of Method 2 has an unfavorable effect on the performance of the algorithm.

The performance of Method 3 is better than that of Method 2 under two data quantity, which indicates integrating prior knowledge into the data-driven model can further improve the accuracy of detection. The improved accuracy is closely related to the data quantity. Under data quantity of 500, the accuracy has improved strongly with growth rate up to 3.4%. While under data quantity of 1000, the growth rate is only 1.3%. We can conclude that when the data quantity is small, what a data-driven model has learned is limited, thus prior knowledge can be an effective supplement.

When the data quantity is reduced from 1000 to 500, the accuracy of all methods reduces. The accuracy reduction of all methods is as shown in Table VII. Due to the integration of prior knowledge, the accuracy reduction of Method 3 and the proposed method is the lowest. It means that these two methods are minimally affected by the reduction of data quantity. It can be concluded that integrating prior knowledge into data-driven models can reduce their dependency on the data quantity. This result is consistent with our previous analysis.

The proposed feature fusion framework based method has the best performance, due to more effective feature extraction, overcoming the imbalance phenomenon and integrating prior knowledge. The proposed method not only improves the performance of faulted line detection remarkably but also reduces the dependency of CNN based model on data quantity, which presents a wide range of possibilities for the further application of deep learning in the analysis of the power systems.

V. CONCLUSION

In this paper, a novel feature fusion framework based SLG faulted line detection method has been proposed. Considering the difference of lines, 1D CNN is used to extract features of zero sequence current data of each line. The extracted features proved to be more effective and robust. The imbalance phenomenon between fault data and healthy data is overcome by the proposed framework, improving the accuracy of detection. The feature fusion method is used, associating 1D CNN based features of each line, and integrating prior knowledge into the data-driven model. Integrating prior knowledge not only improved the accuracy of detection but also reduced the dependency on data quantity.

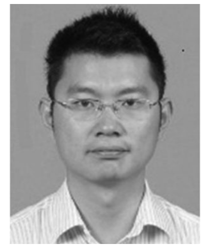
By extracting more effective features, avoiding the unfavorable effect of imbalance phenomenon, and integrating prior knowledge, the proposed method has better performance of SLG faulted line detection, providing a way for more effective use of deep learning technology in power systems.

REFERENCES

[1] S. Hänninen and M. Lehtonen, "Characteristics of earth faults in electrical distribution networks with high impedance earthing," *Elect. Power Syst. Res.*, vol. 44, no. 3, pp. 155–161, 1998.
[2] S. S. Gururajapathy, H. Mokhlis, and H. A. Illias, "Fault location and detection techniques in power distribution systems with distributed generation: A review," *Renewable Sustain. Energy Rev.*, vol. 74, pp. 949–958, 2017.
[3] A. Zidan *et al.*, "Fault detection, isolation, and service restoration in distribution systems: State-of-the-art and future trends," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2170–2185, Sep. 2017.
[4] A. Kalyuzhny, "Analysis of temporary overvoltages during open-phase faults in distribution networks with resonant grounding," *IEEE Trans. Power Del.*, vol. 30, no. 1, pp. 420–427, Feb. 2015.
[5] K. J. Sagastabeitia, I. Zamora, A. J. Mazon, Z. Aginako, and G. Buigues, "Phase asymmetry: A new parameter for detecting single-phase earth faults in compensated MV networks," *IEEE Trans. Power Del.*, vol. 26, no. 4, pp. 2251–2258, Oct. 2011.
[6] A. Ghaderi, H. A. Mohammadpour, H. L. Ginn, and Y. Shin, "High-impedance fault detection in the distribution network using the time-frequency-based algorithm," *IEEE Trans. Power Del.*, vol. 30, no. 3, pp. 1260–1268, Jun. 2015.

[7] T. S. Sidhu and Z. Xu, "Detection of incipient faults in distribution underground cables," *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1363–1371, Jul. 2010.
[8] K. Chen *et al.*, "Fault line detection using sampled data processing and ADABOOST for small current grounding system," in *Proc. Chin. Soc. Elect. Eng.*, 2014, vol. 34, pp. 6228–6236.
[9] M. Sarlak and S. M. Shahrtash, "High impedance fault detection using combination of multi-layer perceptron neural networks based on multi-resolution morphological gradient features of current waveform," *IET Gener., Transmiss. Distrib.*, vol. 5, no. 5, pp. 588–595, 2011.
[10] K. Pandakov, H. K. Høidalen, and J. I. Marvik, "Misoperation analysis of steady-state and transient methods on earth fault locating in compensated distribution networks," *Sustain. Energy, Grids Netw.*, vol. 15, pp. 34–42, 2018.
[11] L. Bao-Liang *et al.*, "Learning from imbalanced data sets with a min-max modular support vector machine," *Frontiers Elect. Electron. Eng. China*, vol. 6, no. 1, pp. 56–71, 2011.
[12] Y. Sun, A. K. Wong, and S. K. Mohamed, "Classification of imbalanced data: A review," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 4, pp. 687–719, 2009.
[13] M. Diligenti, S. Roychowdhury, and M. Gori, "Integrating prior knowledge into deep learning," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl.*, Cancun, Mexico, 2017, pp. 920–923.
[14] L. Eren, "Bearing fault detection by one-dimensional convolutional neural networks," *Math. Problems Eng.*, vol. 2017, 2017, Art. no. 8617315.
[15] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.
[16] Z. F. Ye, Y. M. Wen, and B. L. Lu, "A survey of imbalanced pattern classification problems," *CAAI Trans. Intell. Syst.*, vol. 4, no. 2, pp. 148–156, 2009.
[17] N. V. Chawla *et al.*, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
[18] P. K. Roy, R. Chakravorty, S. Sedai, D. Mahapatra, and R. Garnavi, "Automatic eye type detection in retinal fundus image using fusion of transfer learning and anatomical features," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl.*, Gold Coast, QLD, Australia, 2016, pp. 1–7.
[19] G. Krishnamurthy *et al..*, "A deep learning approach for multimodal deception detection," 2018, arXiv:1803.00344.
[20] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy K-nearest neighbor algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 4, pp. 580–585, Jul./Aug. 1985.
[21] A. Krizhevsky, I. Sutskever, and E. H. Geoffrey, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2012.
[22] M. Guo, X. Zeng, D. Chen, and N. Yang, "Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1291–1300, 2018.
[23] V. Nair and E. H. Geoffrey, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, arXiv:1412.6980.
[25] X. Lu *et al.*, "Feature extraction and fusion using deep convolutional neural networks for face detection," *Math. Problems Eng.*, vol. 2017, 2017, Art. no. 1376726.
[26] E. Boyaci and M. Sert, "Feature-level fusion of deep convolutional neural networks for sketch recognition on smartphones," in *Proc. IEEE Int. Conf. Consum. Electron.*, Las Vegas, NV, USA, 2017, pp. 466–467.
[27] Y. Xue *et al.*, "Earth fault protection in non-solidly earthed network based on transient zero sequence current comparison," *Autom. Elect. Power Syst.*, vol. 27, no. 9, pp. 48–53, 2003.

Ying Du received the B.E. degree in electrical engineering from Central South University, Changsha, China, in 2017. She is currently working toward the Ph.D. degree with Shanghai Jiao Tong University. Her work centers on condition monitoring and fault detection in distribution systems.



Yadong Liu was born in Wuhan, China. He received the B.E. degree in electronic and information from the China University of Geosciences, Wuhan, China, in 2004, and the M.D. and Ph.D. degrees in electrical engineering from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2008 and 2012, respsectively. He is currently a Lecturer with the School of Electronic Information and Electrical Engineering, SJTU. His research interests include detecting fault location and diagnosis of power equipment.

Qingzhu Shao, photograph and biography not available at the time of publication.

Lingen Luo, photograph and biography not available at the time of publication.

Jindun Dai, photograph and biography not available at the time of publication.

Gehao Sheng was born in Hunan, China. He received the B.E. and Ph.D. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 1996 and 2003, respectively. He is currently a Full Professor in Electrical Engineering with Shanghai Jiao Tong University, Shanghai, China. His research interests include control theory and its applications to automatic voltage control in power systems.

Xiuchen Jiang was born in Shandong, China. He received the Graduate degree from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 1987, the M.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 1992, and the Ph.D. degree in electrical engineering from SJTU in 2001. He is currently a Full Professor with SJTU, acting as the Director of Electrical Power Department, SJTU, since 2002. His main research fields are in electrical measuring technology and electrical apparatus automation.