

Received August 12, 2020, accepted September 2, 2020, date of publication September 8, 2020, date of current version September 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3022639

A Transfer Learning-Based High Impedance Fault Detection Method Under a Cloud-Edge Collaboration Framework

YONGJIE ZHANG¹, (Graduate Student Member, IEEE), **XIAOJUN WANG**, (Member, IEEE), **JINGHAN HE**¹, **YIN XU**¹, (Senior Member, IEEE), **FANG ZHANG**¹, (Member, IEEE), **AND YIPING LUO**¹, (Graduate Student Member, IEEE)

School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Xiaojun Wang (xjwang1@bjtu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB0902800, and in part by the Science and Technology Project of the State Grid Corporation of China under Grant 52094017003D.

ABSTRACT High impedance faults (HIFs) in distribution networks are hard to describe and be detected precisely because of the complexity and randomness of their features. Therefore, traditional feature analysis methods may lack sufficient reliability and generalization, which makes data-based methods a more appropriate option. However, according to previous statistical analyses, in practical scenarios, only a small quantity of historical HIF data (less than 20%) can be recorded and utilized. In this article, a transfer learning-based HIF detection method is proposed under a cloud-edge collaboration framework of the Internet of Things, which can solve the problem of insufficient data by integrating historical data from multiple distribution networks. Through the cloud-edge collaboration framework, all features from different distribution networks are first integrated to form a basic cloud convolutional neural network model for HIF detection. The features are extracted and updated by edge computers based on the accurate synchronous measurements provided by distribution-level phasor measurement units. To uniform the data scales of the different distribution networks, principal component analysis is adopted during feature extraction. Specific to each distribution network, the target HIF detection model is transferred from the basic cloud model by fine-tuning. Furthermore, a data augmentation method based on locality sensitive hashing is proposed to improve the performance of the transferred model. The proposed HIF detection method can be operated in both online and offline modes. The performance was verified by seven different distribution networks in numerical simulations and one practical experimental distribution network.

INDEX TERMS High impedance faults, cloud-edge collaboration, distribution-level phasor measurement units, transfer learning.

I. INTRODUCTION

High impedance faults (HIFs) are a common type of fault in distribution networks. They always occur when distribution network conductors break and touch highly resistive surfaces, such as soil or tree branches. HIFs often act as arcing grounded faults with unstable and low-fault currents. Normally, HIFs have higher than 600Ω resistance and produce fault current levels in the 0 to 50 ampere range in distribution

networks, which are difficult to protect with common protective devices, such as conventional overcurrent relays. However, HIFs may lead to equipment damage, significant fire hazards, and even threats to human lives [1].

HIF detection methods can be essentially divided into three approaches: 1) model-based methods, 2) feature-based methods, and 3) data-based methods. The model-based methods directly analyze and describe the arc process of HIFs to achieve HIF detection. However, different contact surfaces can affect the arcing phenomenon of HIFs. Therefore, existing methods have mainly been proposed for detecting

The associate editor coordinating the review of this manuscript and approving it for publication was Wuhui Chen¹.

one specific type of HIF, such as vegetation HIFs [2], [3], soil-grounding HIFs [4], underwater cable HIFs [5], etc. However, an accurate model representing all types of HIFs is almost impossible. To expand the scope of application of detection methods, feature-based methods have been proposed. This type of method extracts the general features of HIFs using mathematical algorithms and determines HIFs by setting fixed thresholds. Fourier transform [6] and wavelet transform [7] are the two most common algorithms used for fault feature extraction. Additionally, other time-frequency analyses [8] and mathematical morphology [9], [10] algorithms have been used to extract fault features. However, the discrimination threshold is usually difficult to determine, and fixed thresholds also have poor reliability. By contrast, threshold-free data-based HIF detection methods are more reliable and have strong generalization ability and thus are receiving increased attention from researchers.

In previous studies of data-based HIF detection methods, most have adopted supervised learning (SL) methods, such as decision trees [2], random forest [10], support vector machine (SVM) [11], [12], neural networks [13], deep learning algorithms [14], etc. The performance of these methods is mainly affected by the differentiation of extracted features and the scale of the labeled data. In fact, however, only a small amount of HIF data can be recorded and utilized. Approximately 25% to 30% of distribution network faults are considered HIFs [1], but less than 20% of HIFs can be recorded and cleared by conventional protection methods [15]. Recently, some researchers have proposed different methods to solve these problems. In [16], a graph convolutional network (GCN)-based fault detection and location method was proposed by taking system topology into account, which can reduce the required amount of training data to some extent. In [17], a semi-supervised learning (SSL)-based HIF detection method was proposed, which can achieve fault detection using a combination of a small set of labeled data and additional, unlabeled data.

Most of the existing data-based HIF detection methods only utilize the labeled data generated by the target distribution network itself. However, HIF events occurring in a single distribution network are not sufficient to train a reliable detection model. Meanwhile, HIF events occurring in different distribution networks share similar but not identical features. Therefore, in this article, we proposed a transfer learning-based high impedance fault detection method. Our approach establishes a basic detection model by integrating available HIF event data from multiple distribution networks, identifying common features among different HIF events. Then, specific models are obtained by combining the basic model and local data, reflecting the unique features of each distribution network. As a result, the proposed method can achieve high accuracy even though the data collected from each distribution network are limited. The cloud-edge collaboration framework brought by the development of the power distribution Internet of Things (IoT) provides a suitable deployment platform for the proposed HIF detection method.

The measurement data are generated by distribution-level phasor measurement units (D-PMUs). In the edge, synchronized transient HIF characteristics are introduced as the features, which are extracted from the zero-sequence current by discrete wavelet transform (DWT). Meanwhile, principal component analysis (PCA) is adopted to uniform different scales of features from multiple distribution networks. Then, in the cloud, a basic convolutional neural network (CNN) model is trained by these features. For one target distribution network, a special HIF detection model is transferred from the basic cloud model by fine-tuning through edge computing. To improve the performance of deployment in the target distribution network, a locality sensitive hashing (LSH)-based data augmentation method is proposed.

Contributions of this article are as follows:

(1) Propose a transfer learning-based HIF detection structure under the cloud-edge collaboration framework, which can integrate and utilize data from multiple distribution networks.

(2) Introduce PCA to uniform different data scales and improve the CNN model based on the requirements of practical application.

(3) Propose a LSH-based data augmentation method to improve the performance of the deployment from cloud model to the target distribution network.

This article is organized as follows: An overview of the detection structure is presented in Section II. The HIF feature extraction method and the cloud CNN model are introduced in Section III. In Section IV, the deployment from the cloud to the target distribution network is explained in detail. Section V gives the setting of the verification scenarios, and the verification results are shown, compared, and explained. Finally, conclusions are drawn and future prospects are suggested in Section VI.

II. THE CLOUD-EDGE COLLABORATION FRAMEWORK AND HIF DETECTION STRUCTURE

Power distribution IoT has become a global development trend. According to the definition from the State Grid Corporation of China (SGCC), power distribution IoT is a new operational pattern for power distribution networks based on the integration of traditional power industry technology and next-generation information technologies, such as the IoT, the cloud, big data analysis, and artificial intelligence [18]. Many institutions have already studied and developed related IoT platforms. Schneider provides a power distribution IoT platform called EcoStruxure Power for medium- and low-voltage electrical distribution networks [19]. Siemens and General Electric have also produced open industrial IoT systems, MindSphere [20] and Predix [21], respectively. In 2019, the SGCC published a white paper about the IoT in electricity in China. In distribution networks, the power distribution IoT and the cloud model can be introduced for many applications, including topology identification, fault handling, power quality management, electric vehicle charging management, and power line loss monitoring.

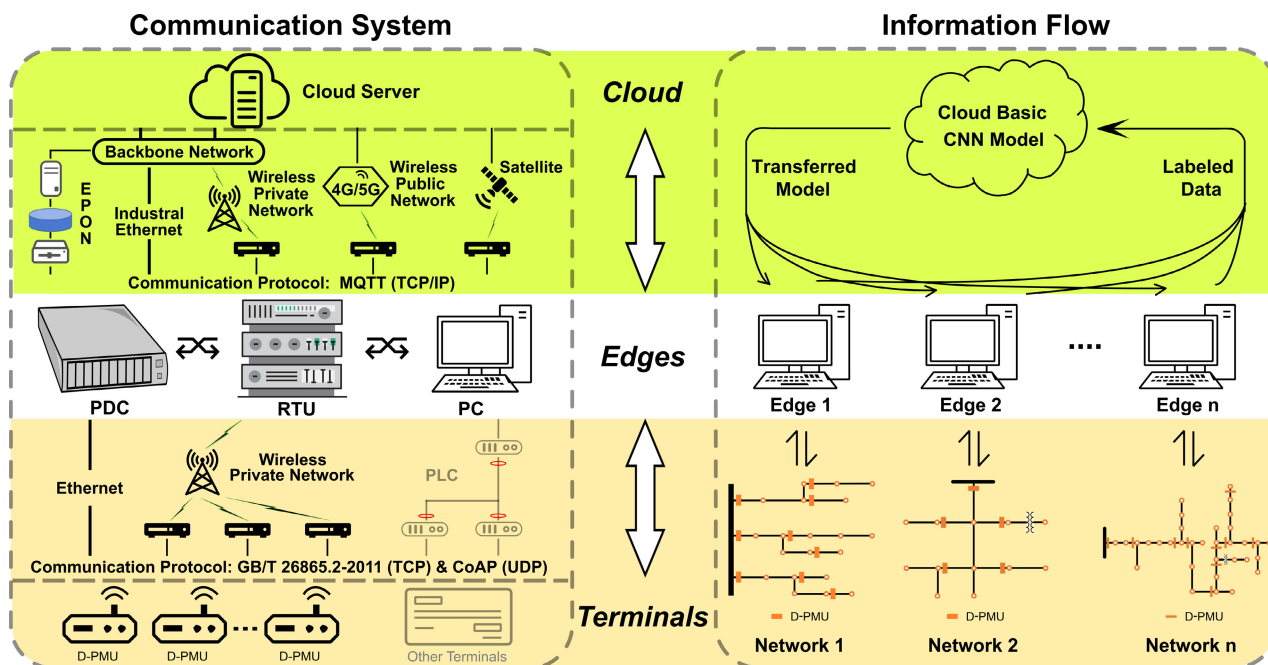


FIGURE 1. The cloud-edge collaboration framework of the proposed HIF detection method.

In this article, a cloud-edge collaboration framework based on the power distribution IoT is employed to utilize data from multiply distribution networks for HIF detection. The whole framework includes three layers: terminals, edges, and the cloud. The terminals provide awareness about the operation status and execute grid commands and controls. The edge is an edge computing platform close to the data source, which is an extension of the cloud server. The cloud is the master platform, which adopts technologies such as cloud computing, big data analysis, and machine learning. Fig. 1 shows the schematic of the cloud-edge collaboration framework and information flow of the proposed HIF detection method. In the cloud-edge system, the edge node and cloud server can achieve effective collaboration through hierarchical computing. The tasks processed at the edge node are targeted to achieve lower latency and better efficiency. And the offloaded, computationally intensive tasks will be processed in the cloud server to take advantage of its abundant computation capacity [22].

The communication system is important for the actual deployment of the cloud-edge collaboration framework. Based on the demonstration project of the supported program and recommended architecture of the SGCC, the communication system is shown in Fig. 1. The communication system can be divided into two parts: edge-terminal communication and cloud-edge communication.

Edge-terminal communication: In this article, the terminals mainly consist of D-PMUs. There are two ways to communicate between D-PMUs and edges. One common way is to adopt the phasor data concentrator (PDC) to collect the data from D-PMUs in a distribution network. According to the

application of the demonstration project, the communication protocol is the National Standard of the People’s Republic of China/Recommended (GB/T) 26865.2-2011 [23], which operates over the Transmission Control Protocol (TCP). Through the PDC, the data from the D-PMUs can be transmitted to the remote terminal unit (RTU) and PCs in the edge computing platform. Another way to transmit the data is to establish direct communication from the D-PMUs to edges. The D-PMUs can bypass the PDC and transmit the data to the edge computers by using the same communication protocol as that used for transmission from the D-PMUs to PDCs. Meanwhile, the D-PMUs can also transmit the data through the constrained application protocol (CoAP), which is a lightweight communication IoT protocol recommended for edge-terminal communication by the SGCC [24]. CoAP uses the user datagram protocol (UDP) as the underlying network protocol. The communication between the D-PMUs and edges mainly relies on Ethernet and wireless private networks. There are also other terminals in the cloud-edge collaboration framework, such as smart meters, temperature humidity sensors, and line circuit breakers. Some communication can rely on power line communication (PLC).

Cloud-edge communication: Based on the recommendation of the SGCC, cloud-edge communication is performed with the message queue telemetry transport (MQTT) communication protocol, which is suitable for communication between multiple edges and the cloud server [25]. MQTT operates over the TCP/Internet Protocol (IP). Through the MQTT broker, the framework can easily realize communication from one sender to multiple clients and from multiple publishers to a single subscriber. There are many

communication methods between the cloud and edges, including Ethernet Passive Optical Network (EPON), industrial Ethernet, electric wireless private networks, wireless public networks, and satellite communication.

Based on the cloud-edge collaboration framework, we propose a transfer learning-based HIF detection method by integrating data from multiply distribution networks. The whole structure is shown in Fig. 1. A basic model is trained in the cloud server using HIF data from multiple distribution networks and sent to the edges associated with each individual distribution network. Then, the basic model is cast into specific models in the edges using local data. Pre-existing labeled features need to be uploaded to the cloud at first as the initial training data. The proposed method has a relatively low communication system requirement. The uploading of labeled data from the edges to the cloud and cloud model deployment from the cloud to the edges do not require real-time communication. Given the various communication conditions and possible restrictions of different distribution networks, two different operating modes, online and offline, are provided. In online operation mode, the edge nodes will detect the HIF in real time. If a period of data is detected as a fault or HIF and verified in an actual site, these data will be labeled and uploaded to the cloud, whether they are faults, disturbances or normal situations, to adjust the parameters of the basic model. Meanwhile, after a period of time, such as one or two weeks, the target models of all distribution networks can update the parameters of the basic cloud model and retransfer the model. If some distribution networks are operated in offline mode, they will only need to regularly update the target models and not upload the operational data.

In such a computational structure, duplicated computation is avoided because the cloud server takes care of common computations and shares the results. As a result, the edges remain computationally inexpensive and cost effective. The integration of data from multiply distribution networks not only solves the shortage of insufficient training data effectively but also covers more types of HIFs than one distribution network. Along with continuous operation, the proposed method will gradually improve the ability to identify general HIF features and the level of adaptation to the target distribution network.

III. FROM EDGE TO CLOUD: CLOUD CNN MODEL TRAINING

A. HIF FEATURE EXTRACTION

HIF is a typical type of weak-feature fault. Given the limitations of the sampling frequency and accuracy of traditional distribution network measurement devices, it has been difficult to make headways with traditional HIF detection methods. The development of D-PMUs has resulted in the acquisition of high sampling-rate three-phase voltage and current phasor data with less error. All the measurements are GPS time stamped to provide time-synchronized observability, which means that the transient information from

multiple D-PMUs can be used uniformly as a feature [26]. D-PMU devices provided by different manufacturers may vary in terms of measurement variables, sampling rate, and data accuracy and precision. The D-PMUs employed in this article were developed by a project supported by the Chinese government. The sampling rate is 6400 samples/s and the zero-sequence currents are provided.

The transient features of HIFs can contain more effective information than steady-state features, especially when affected by disturbances of and noise in the distribution network. In this article, synchronous transient HIF feature matrices are extracted in the edge. The transient features are extracted from the zero-sequence currents by DWT. For convenience of integrating the information from different distribution networks, PCA is utilized here to unify the data scales. The feature extraction can be divided into 3 steps as follows:

- (1) Extract the transient features from the zero-sequence current of each D-PMU by DWT.
- (2) Reduce the dimensionality of the extracted transient features under the same decomposition level of all D-PMUs to the same scale by PCA.
- (3) Combine the dimension-reduced features as the fixed-scale feature matrix of the distribution network.

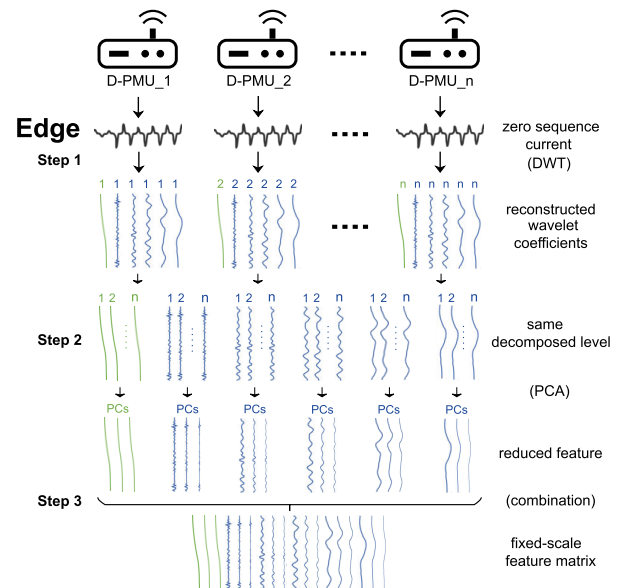


FIGURE 2. The whole process of HIF feature extraction.

The whole process is shown in Fig. 2. After extraction, the transient information from the corresponding time window can be formed as a fixed-scale feature matrix, which can be easily integrated into a cloud model. The details are as follows.

1) STEP 1: EXTRACT THE WAVELET COEFFICIENTS

Assume that a distribution network includes N D-PMUs and the time window is T . Therefore, for D-PMU n_i ($i = 1, \dots, N$), because the D-PMUs adopted here can

generate zero-sequence current directly, the original transient measurement is $\mathbf{I}O^i = [IO_1^i, IO_2^i, \dots, IO_i^i, \dots, IO_T^i]$. The central idea of DWT is to decompose a time series γ into levels of multiple resolutions. At different resolutions, the details of a signal can characterize different physical structures. The low-resolution details can generally characterize large structures of information, and as the resolution increases, finer details are obtained.

The multiple resolution analysis proposed by Mallat can quickly achieve wavelet decomposition and reconstruction. Detailed descriptions of the algorithm can be found in [27], [28]. In DWT decomposition, two factors should be considered: the wavelet function and the number of decomposition levels M . For HIF detection, db4 is considering as one of the most suitable choice of wavelet [7], which is also adopted in this article. For the number of decomposition levels (M), γ can be decomposed into detail coefficients $D_m[\gamma]$, $m \in [1, M]$ and approximation coefficients $A_M[\gamma]$. The corresponding frequency bands are $f_s/2^{m+1} \sim f_s/2^m$, $m \in [1, M]$ and $0 \sim f_s/2^{M+1}$. Considering that disturbances may produce low-frequency harmonics, including second harmonics and third harmonics, the lower frequency spectrum should be subdivided as much as possible to better distinguish HIFs and disturbances. Therefore, in this article, M is set to 5, resulting in an acceptable calculation complexity. Finally, the decomposition coefficients are reconstructed as $recD_m[\gamma]$ and $recA_M[\gamma]$. To retain comprehensive information on faults or disturbances, all reconstructed coefficients are kept as features. For an individual D-PMU, the extracted feature matrix $\mathbf{I}O_{WT}^i$ is shown below, where $J = M + 1$:

$$\mathbf{I}O_{WT}^i(j, :) = \begin{cases} rec A_{J-1} [IO^i], & j = 1 \\ rec D_{j-1} [IO^i], & j \in [2, J] \end{cases} \quad (1)$$

2) STEP 2: REDUCE THE DIMENSIONALITY BY PCA

In different distribution networks, the topologies and the number of D-PMUs are different. To unify the data scales, PCA is utilized in this article. PCA can be thought of as a method that reveals internal structure of the data in a way that best explains its variance [29]. By maximizing the variance in the data, PCA captures the dominant features in an N -dimensional dataset in descending order through an orthogonal transformation. Thus, the transformed data are linearly independent and are referred to as the principal components (PCs). In this article, we adopt PCA to reduce the dimensionality of the wavelet coefficients under the same decomposition level of all D-PMUs $\mathbf{X}^j = [\mathbf{I}O_{WT}^1(j, :); \mathbf{I}O_{WT}^2(j, :); \dots; \mathbf{I}O_{WT}^i(j, :); \dots; \mathbf{I}O_{WT}^N(j, :)]$ to the same scale. The PCs are obtained through singular value decomposition (SVD) of the covariance matrix \mathbf{S}^j ($\mathbf{S}^j = \mathbf{X}^j \mathbf{X}^{jT}$). The transformed PCs, \mathbf{Z}^j , are calculated from the covariance matrix \mathbf{S}^j , where it satisfies

$$\mathbf{W}' \mathbf{S}^j \mathbf{W} = \mathbf{L}^j, \quad (2)$$

where \mathbf{L}^j ($l_1^j, l_2^j, \dots, l_p^j$) are the eigenvalues of \mathbf{S}^j , which can be solved from the characteristic equation $|\mathbf{S}^j - \lambda \mathbf{I}| = 0$.

The eigenvalues $l_1^j, l_2^j, \dots, l_p^j$ are also the variances of each PC, and the sum of \mathbf{L}^j equals the sum of the variance of the original variables. After obtaining these eigenvalues, the corresponding eigenvectors $\mathbf{W} = \{\mathbf{w}_i\}$, where \mathbf{w}_i is a column of \mathbf{W} , $\mathbf{w}_i = (w_{1i}, w_{2i}, \dots, w_{Ni})$, $i = 1, \dots, N$, can be calculated. The eigenvectors \mathbf{W} are referred to as the loadings, representing the correlations between the variables and PCs. The relationship between the PCs, \mathbf{Z}^j ($z_1^j, z_2^j, \dots, z_N^j$), and the dataset $\mathbf{X}^j(T \times N)$ is expressed as $\mathbf{Z}^j = \mathbf{W} \mathbf{X}^j$. By retaining the first σ ($\sigma < N$) PCs, the dimensionality of the data can be reduced significantly, with only minor data variability being sacrificed.

3) STEP 3: COMBINE THE DIMENSION-REDUCED FEATURES AS A MATRIX

Finally, the dimension-reduced features of the same wavelet decomposition level can be expressed as $[\mathbf{P}C_1^j, \mathbf{P}C_2^j, \dots, \mathbf{P}C_\sigma^j]$. For a distribution network, a $(\sigma \times J) \times T$ fixed-scale feature matrix is created by combining the dimension-reduced features to represent the situation during the corresponding time window. The parameter σ is related to the scale of the distribution network, electrical parameters, and HIF features. Based on the results of simulation experiments, when σ is set to 3, 95% of the original data can be explained under most scenarios. In this article, the time window T is set to 64 (half cycle under 50 Hz). With a moving time window, the feature matrix can be generated in real time and then detected.

B. BASIC CLOUD CNN MODEL

After HIF feature extraction, we adopt the CNN to establish the basic model in the cloud. CNNs are a class of deep neural networks in deep learning. The architecture of a CNN can take advantage of the 2D structure of the input data, such as an image or a matrix. A CNN is easier to train and has fewer parameters than a fully connected network with the same number of hidden units [30], [31]. For a fundamental CNN model, convolution layers, pooling layers, activation function, and fully connected layers (FCs) are the four main components.

For the basic cloud CNN model, it is important to prevent overfitting, meaning that the trained model works well on the training set but not on the test set. To prevent overfitting and enhance the generalization ability of the model, some improvements, including L2 regularization and dropout, are adopted accordingly to adapt the model to practical HIF detection scenarios. L2 regularization is a common form of regularization that can be implemented by penalizing the squared magnitude of all parameters directly in the objective function [32]. The fitting degree can be impacted by modifying the scale factor. Dropout is a common and the easiest-to-implement method to address overfitting by randomly dropping a percentage of units (along with their connections) from the neural network, which can prevent units from co-adapting too much.

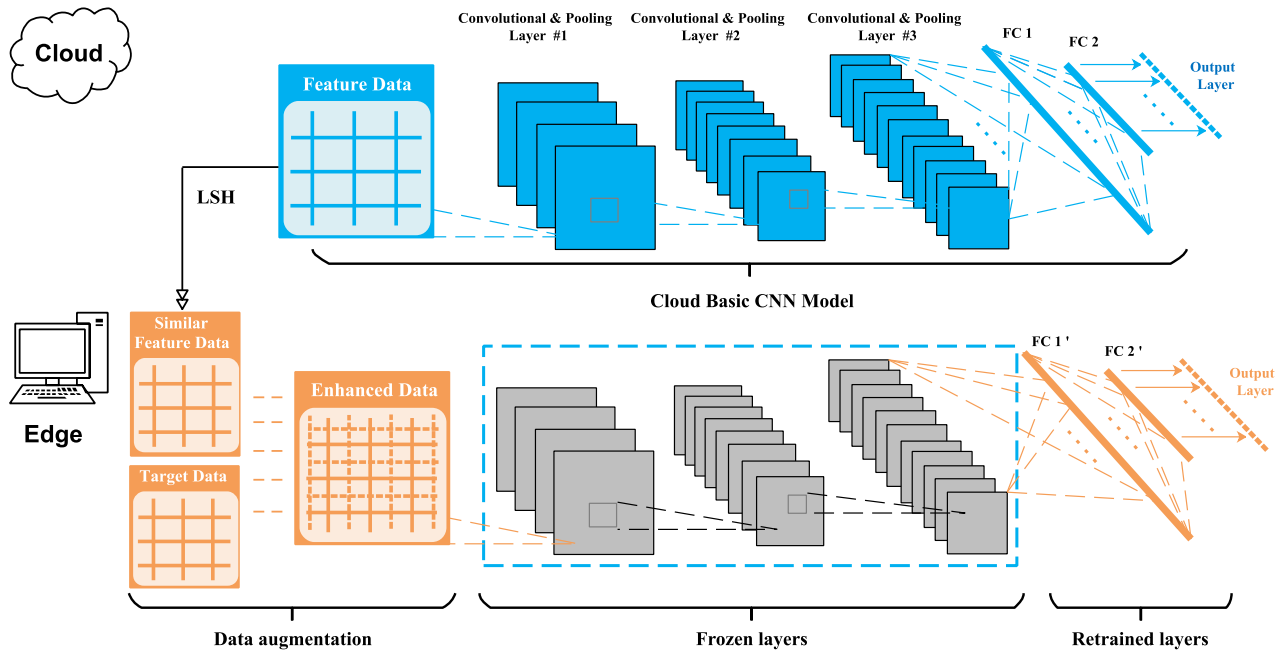


FIGURE 3. Structure of basic cloud CNN model and fine-tuning.

Here, the CNN model includes three convolutional & pooling layers and two FCs. The rectified linear unit (ReLU) is chosen as the nonlinear activation function to address nonlinear problems. Max pooling is chosen for the pooling layer, and the softmax function is chosen for the loss function. Detailed introductions, improvements, and parameter settings for CNNs can be found in our previous research [33]. In this method, the tags of the features are set as N (normal situation), F (fault situation), and D (disturbance). The establishment of the D tag can better distinguish HIFs from disturbances.

IV. FROM CLOUD TO EDGE: DATA AUGMENTATION AND DEPLOYMENT

Direct application of the basic cloud CNN model to a target distribution network will probably fail to obtain good results because of the differences between different distribution networks. To obtain better performance, we perform fine-tuning to deploy the cloud model from the cloud to the edge. Fine-tuning is a process of retaining the parameters of some layers (frozen layers) in the pretrained model and retraining other layers (retrained layers), which is a feasible method for transferring the pretrained model to other scenarios. In a CNN, the earlier layers contain more generic features, and later layers become progressively more specific to the details of the classes. The last FC layers can be regarded as a classifier. Therefore, in this article, we freeze all the hidden layers and fine-tune the last two FCs in the target edge. The fine-tuning architecture is shown in Fig. 3.

To further improve the performance of fine-tuning, it is necessary to adopt some data augmentation tricks to expand the data. In the field of HIF detection, only a small amount

of the data in the target distribution network is available, but the extracted feature matrices may have high similarity. Here, the data argumentation is divided into two steps: search for similar feature data in the cloud server and expand the data by a proportional coefficient in the target edge. Along with this detection system movement, the data in the cloud will become a large-scale dataset. Therefore, choosing traversal searches will result in high time and computational resource costs, especially for high-dimensional feature searches.

In this article, we adopt LSH to efficiently find similar data in the cloud. LSH is an approximation technique for the similarity search problem, which uses a set of specific hash functions to build hash tables for a set of data objects [34]. LSH can make the mapping of similar objects to the same area more likely than that of non-similar objects under a certain similarity measure. Specifically, LSH can map similar objects into the same bucket with high probability. When performing a data search, data objects in the same bucket are used as candidate objects, and the distance between the candidate object and the query object is calculated sequentially. To prevent the candidate object being hashed into the bucket from being different from its nearest neighbor, LSH maintains multiple hash tables by using different hash functions. It is an effective algorithm for dealing with high-dimensional data approximation problems.

The LSH scheme relies on the existence of locality-sensitive hash functions. Consider a family of hash functions \mathcal{H} mapping \mathbb{R}^d to some universe U . For any two points p and q , consider a process in which we choose a function h from \mathcal{H} uniformly at random and analyze the probability such that $h(p) = h(q)$. The definition of LSH is given below:

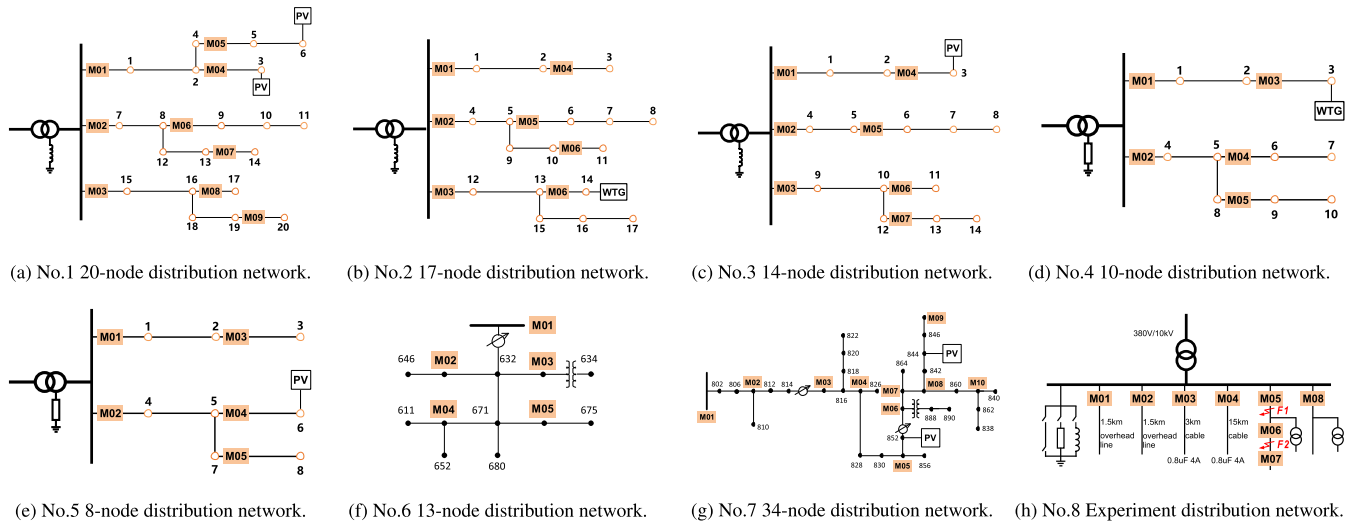


FIGURE 4. The topologies of eight distribution networks and the locations of D-PMUs.

A family \mathcal{H} is called (R, cR, p_1, p_2) -sensitive if for any $p, q \in \mathbb{R}^d$ ($p_1 > p_2$).

$$\text{if } \|p - q\| \leq R \text{ then } \Pr_{\mathcal{H}}[h(q) = h(p)] \geq p_1,$$

$$\text{if } \|p - q\| \geq cR \text{ then } \Pr_{\mathcal{H}}[h(q) = h(p)] \leq p_2.$$

Therefore, in the cloud, LSH will be used first to establish several hash tables with buckets to save the hash values of all features. When the target distribution network sends introductions to the cloud, the cloud server will search a certain number of similar features by LSH and send them to the edge computer. Then, in the target edge, all the feature matrices, including the original data and similar data downloaded through LSH, are multiplied by a proportional coefficient k between 0.94 and 1.05 in steps of 0.01. This step can expand the data easily and effectively by multiplying by k (roughly equal to 1) to improve the generalization ability to different fault locations and impedances.

V. SIMULATION AND VERIFICATION

A. SIMULATION SETTING

In this article, we verified the proposed HIF detection method with eight different distribution networks, including seven distribution networks in PSCAD/EMTDC and one actual 10 kV experimental distribution network. The eight distribution networks are numbered 1-8. No. 1, 2, and 3 are 10 kV resonant ground distribution networks. The overcompensation rate is 8%. No. 4 and 5 are two 380 V low-resistance ground distribution networks with two outgoing lines. No. 6 is an IEEE 13-node 4.16 kV distribution network with heavy load. No. 7 is a typical IEEE 34-node distribution network with two PVs added at nodes 844 and 852. No. 8 is an actual experimental 10 kV distribution network that is used to test several general faults and HIFs in contact with different surfaces under three different grounding methods. The topologies of all eight distribution networks and the locations of the D-PMUs are shown in Fig. 4.

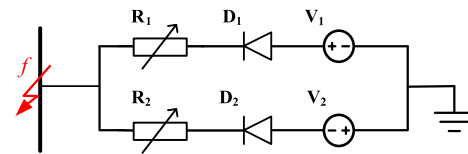


FIGURE 5. HIF model structure.

The HIF model in this article is based on the anti-parallel DC-source model [9], as shown in Fig. 5. An anti-parallel connection of two sets of variable resistors and DC voltage sources along with diodes was considered to simulate the real HIF characteristics. For the purpose of increasing nonlinearity, the resistor and DC source values vary with a frequency of 1 kHz between predefined values, creating a nonlinear current waveform. The HIF model is simple but contains a large portion of possible HIFs due to the randomly and rapidly varying parameters.

We set ten different fault locations in the seven non-experimental distribution networks (i.e., all except the No. 8 distribution network). At each fault location, we simulate ten types of faults, including six HIFs and four general faults. We change the parameters of the HIF model to simulate the six different types of HIF by comparing their V-I characteristics with those obtained from the real field tests in [35]. Take the No. 1 20-node distribution network, for example. The V-I characteristics of the No. 1 distribution network (at a fault location 50% of line 2) are shown in Fig. 6. The three-phase voltage and current waveforms of a dry tile-grounding HIF are shown in Fig. 7. Meanwhile, one metallic fault and three nonmetallic faults (fault impedances 5 Ω , 10 Ω , and 30 Ω) are set in each location. As a convenience, these four types of faults are represented as general faults. All faults are single-phase ground faults. We also set 10 different disturbances for each distribution network, including capacitors, DGs, and load switching. All the faults and disturbances

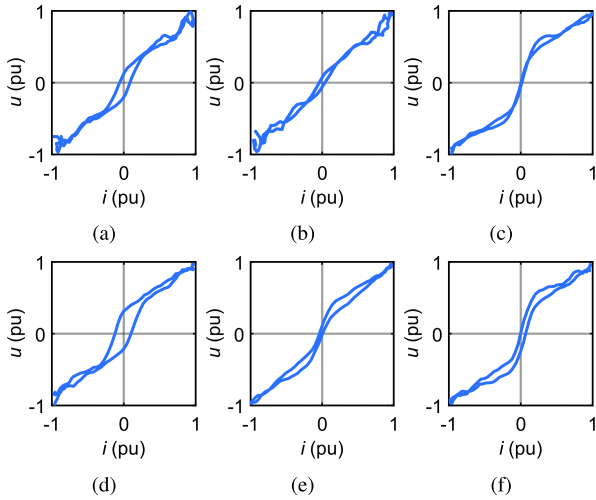


FIGURE 6. The V-I characteristics of different types of HIFs in No.1 distribution network: (a) dry tiles, (b) damp tiles, (c) damp sand, (d) dry rubber (e) damp rubber, and (f) dry gravel.

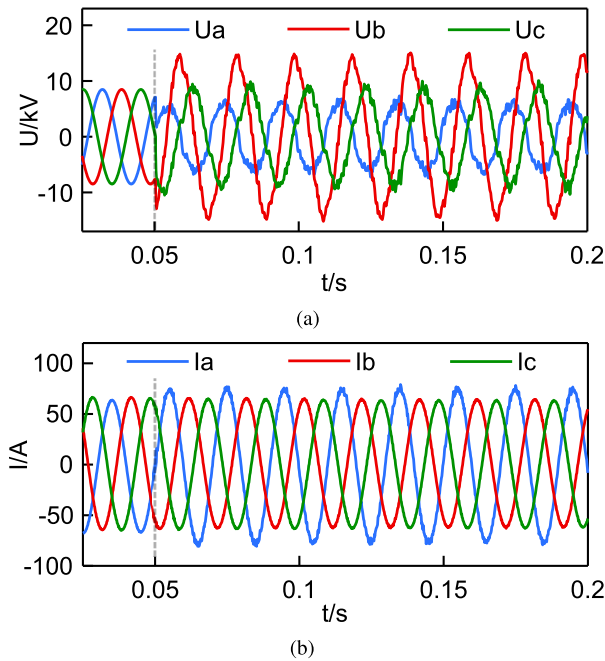


FIGURE 7. Three-phase voltage and current waveforms of a dry tile-grounding HIF in No.1 distribution network: (a) voltage waveforms, (b) current waveforms.

occur at 0.6 s/0.602 s/0.605 s/0.61 s to simulate the different initial phase angles. The feature extraction time window is a half circle (64 points), and the step size is 10 points. Each scenario records 0.3 s from 0.1 s before fault/disturbance occurrence. Note that tag D is listed separately, but the disturbance in a distribution network occurs in a normal situation.

B. VERIFICATION IN SIMULATION

1) ESTABLISHMENT OF BASIC CLOUD MODEL

According to the settings of all simulations, each distribution network obtains 410 scenarios, including different

fault/disturbance types, locations, and initial phase angles. To imitate the actual scene, the basic cloud models are trained by random chosen scenarios from No. 1 to No. 6 distribution networks (no more than 50 scenarios of each distribution network). Three basic cloud models, Cloud Model I, II, and III, are trained by different number of initial data. The proportion of chosen scenarios from each distribution network is listed in Table 1. No.7 and No.8 distribution networks are remained to verify the offline mode.

TABLE 1. The proportion of chosen scenarios from each distribution network for the basic cloud model establishment.

	HIFs	General Faults	Disturbances
Cloud Model I	10	10	1
Cloud Model II	15	15	2
Cloud Model III	20	20	4

In order to quantify performance of the proposed method, the accuracy is defined in Equation (3), with $|TP|$ being the amount of true positive classifications; $|TN|$, the amount of true negative classifications; $|FP|$, the amount of false positive classifications, e.g., a false alarm, and $|FN|$, the amount of false negative classifications, like missed faults.

$$accuracy = \frac{|TP| + |TN|}{|TP| + |FP| + |FN| + |TN|} \tag{3}$$

After training three basic cloud models, the accuracy rates are all higher than 98%, which shows the cloud models have good performances for the initial chosen scenarios.

2) HIF DETECTION PERFORMANCE IN ONLINE MODE

In online mode, the basic cloud model is trained by data from target distribution networks and adjusted by the updated data after a period of operation. For most of distribution networks, the online mode will be the first priority operation mode. In this section, a comprehensive verification of the online model is conducted from the following three aspects: effectiveness, robustness, and accuracy.

Firstly, we established four comparable CNN models for the No. 1 20-node distribution network to verify the effectiveness of online mode. The first one is individually trained by data of the No. 1 distribution network (chosen scenarios are same as the basic cloud model I). And the CNN structure is also same as the cloud CNN model. The second one is direct-transferred from the basic cloud model I, which means just fine-tuning the detection model by the data of the target distribution network without any data argumentation. To verify the performance of proposed LSH-based data argumentation, the third and fourth CNN models were transferred with exhaustive search (ES)-based and LSH-based data argumentation, respectively. The ES algorithm is enumerating all the distance of features and choosing the features of minimum distance. In this article, the chosen similar data account for 5% of cloud data. The transferred models with ES-based and

LSH-based data argumentation are all the same except the data search methods.

TABLE 2. Comparisons of HIF detection models under different training methods.

Type of CNN model	Detection accuracy		
	HIFs	General faults	Disturbances
Individual-trained model	73.28%	93.12%	91.18%
Direct-transferred model	91.72%	100%	100%
Transferred model (ES)	95.83%	100%	100%
Transferred model (LSH)	95.70%	100%	100%

Table 2 shows the details of comparison. The individual-trained model has bad performance, which shows that direct training by small sample data can hardly get a high accuracy rate. Comparing with the direct-transferred model, the proposed data argumentation can improve the accuracy rate by about 6%. The transferred models with ES-based and LSH-based data argumentation both have great performance with over 95% accuracy rate. But the LSH search just costs 0.64s, which can increase the search speed by around ten times comparing with 5.87s by ES search. When several different edges execute searching task in the cloud at the same time, the LSH-based data argumentation will contribute more benefits than ES-based in computational efficiency.

To verify the robustness of the proposed method, we tested the accuracy rate of transferred model of No.1 distribution network in scenarios with different measurement noises and load uncertainty. The measurement noises are generated by adding different signal-to-noise ratios (SNRs) Gaussian white noises to the original data. For validation of load uncertainty, we simulated two groups of data under 50% capacity of phase A load loss at node 2 and node 14, respectively. The verification results are shown as Table 3. It is observed that the proposed method can keep over 90% accuracy rate under the setting scenarios. For D-PMUs, the amplitude error is usually less than $\pm 0.2\%$, which corresponds to about 54dB. Therefore, the proposed method has good performance in robustness to normal D-PMU measurement noises and load uncertainty.

TABLE 3. The accuracy rate of target model I under different measurement noises and load uncertainty.

Noise / Load fluctuation	Accuracy rate
60dB noise	95.70%
50dB noise	94.87%
40dB noise	91.93%
Load fluctuation at node 2	95.44%
Load fluctuation at node 14	95.17%

To verify the accuracy of the proposed method under online mode, 18 target HIF detection models of No. 1 to No. 6 distribution networks are transferred from 3 cloud models.

The performances are verified by data of all scenarios of each distribution network. The detection accuracy rates of different types of faults and disturbances of all target transferred models are shown in Table 4.

As the table shows, the detection accuracy rates of the target models transferred from basic cloud model I, II, and III can reach 94%, 95%, and 97% on average, respectively. The results show that this method is not affected by the initial angles, fault locations/impedances, and time windows. Meanwhile, the proposed method has good performance in robustness to disturbances, which is important for fault detection in distribution networks. In general, the proposed method under online mode has great performance.

3) HIF DETECTION PERFORMANCE IN OFFLINE MODE

The No. 7 34-node distribution network is adopted to verify the offline mode. As the instruction in Section II, the distribution network which works at offline mode just updates the target model without uploading its own data. Based on the cloud model III, three target models are transferred by data of randomly chosen scenarios of the No. 7 distribution network (same proportion as Table 1). The verification results are shown in Table 5.

The result shows that the offline mode can also achieve HIF detection with a good performance. For general faults, the proposed method under offline mode can almost achieve 100% fault detection. Compared with the online mode, the robustness of disturbances under offline mode is relatively poor. And the accuracy rates of HIF detection under offline mode have a higher sensitivity to the number of available data.

C. VERIFICATION IN PRACTICAL EXPERIMENT DISTRIBUTION NETWORK

To verify the proposed method, we obtain 32 groups of fault data under three grounding modes in a practical experimental distribution network. The fault locations are set in F1 and F2 shown in Fig. 3 (h). For the HIF, the contact surfaces have five different types, including dry/damp soil, dry/damp cement, and asphalt concrete. And other 17 general fault scenarios include different fault impedances from 0 to 5000Ω and different initial phase angles. A target model is transferred from the cloud model III by data of five HIFs and five general faults. For comparison, the cloud model III is directly adopted as another target model (without fine-tuning).

For the general faults under 3000Ω, both models can achieve 100% fault detection. And the transferred model can also achieve fault detection with 100% accuracy rate under 5000Ω grounding fault in the resonant grounding mode, which is 9.14% higher than the cloud model III. For HIFs, the transferred model has a 91.67% accuracy rate, which is just 83.41% for the cloud model III. Fig. 8 shows the zero sequence current waveform of the HIF contacted with the dry soil surface occurred at 0s. The transferred model can detect the HIF during the whole progress except the last two cycles of the first unstable arcing period. By contrast, the cloud

TABLE 4. HIF detection accuracy rates of No.1 to No.6 distribution networks.

Number of DNs and target CNN models	HIF_1	HIF_2	HIF_3	HIF_4	HIF_5	HIF_6	General Faults	Disturbances	
#1 Distribution Network	Target Model I	93.84%	93.55%	93.85%	94.51%	93.43%	93.83%	100.00%	100.00%
	Target Model II	94.83%	94.73%	94.29%	94.60%	95.43%	96.10%	100.00%	100.00%
	Target Model III	97.26%	96.23%	96.77%	97.02%	96.70%	97.83%	100.00%	100.00%
#2 Distribution Network	Target Model I	93.47%	93.66%	93.56%	92.76%	93.94%	93.96%	100.00%	98.71%
	Target Model II	95.86%	95.91%	94.99%	96.65%	93.35%	95.67%	100.00%	100.00%
	Target Model III	97.38%	97.07%	97.86%	97.48%	96.24%	97.21%	100.00%	100.00%
#3 Distribution Network	Target Model I	94.34%	93.59%	94.06%	94.79%	92.79%	93.49%	100.00%	100.00%
	Target Model II	94.93%	94.08%	94.72%	95.62%	95.53%	95.17%	100.00%	100.00%
	Target Model III	97.23%	96.81%	97.41%	96.84%	97.67%	97.28%	100.00%	100.00%
#4 Distribution Network	Target Model I	93.94%	94.26%	94.45%	93.57%	93.27%	95.02%	100.00%	100.00%
	Target Model II	94.78%	94.57%	93.60%	94.31%	95.06%	95.26%	100.00%	100.00%
	Target Model III	97.09%	96.96%	97.21%	97.36%	96.50%	97.29%	99.84%	100.00%
#5 Distribution Network	Target Model I	93.82%	94.22%	93.80%	94.40%	93.99%	94.88%	100.00%	100.00%
	Target Model II	94.89%	94.41%	94.85%	93.88%	94.80%	93.80%	100.00%	100.00%
	Target Model III	97.25%	96.17%	96.54%	96.79%	97.27%	96.62%	100.00%	100.00%
#6 Distribution Network	Target Model I	93.78%	93.48%	94.49%	93.79%	94.07%	94.35%	100.00%	100.00%
	Target Model II	96.38%	95.11%	94.61%	96.44%	94.68%	95.61%	100.00%	100.00%
	Target Model III	96.88%	96.71%	97.06%	96.60%	96.75%	96.67%	100.00%	100.00%

TABLE 5. HIF detection accuracy rates under offline mode of No.7 distribution network.

Target model number	Detection accuracy		
	HIFs	General faults	Disturbances
Target Model I	84.31%	99.45%	98.52%
Target Model II	87.35%	100.00%	99.77%
Target Model III	92.47%	100.00%	100.00%

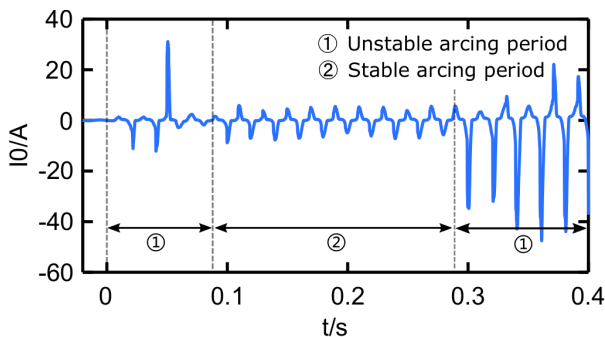


FIGURE 8. The zero sequence current of HIF contacted with dry soil surface in No.8 practical experiment distribution network.

model can not detect the HIF during the stable arcing period until it progressing to the second unstable arcing period with a higher level of fault current after about 0.3s. The result shows

that the proposed transferring method can effectively improve the reliability of the HIF detection.

VI. CONCLUSION

In this article, a transfer learning-based HIF detection method is proposed with the application of D-PMUs. The verification results illustrate that the proposed method can achieve high accurate HIF detection by just a small amount of available data and not affected by measurement noises, load uncertainty, disturbances, time windows, and fault initial angles/locations/impedances. Through the adoption of this cloud-edge collaboration framework, the proposed method can integrate the data from different distribution networks to address the issue of unavailable actual data and insufficient HIF types. The whole information flow of the proposed method and the recommend communication system of the cloud-edge collaboration framework are given in this article. For a specific distribution network, we design online and offline mode to adapt to the requirements of different distribution networks. The online mode has better performance because the proposed LSH-based data argumentation method can effectively improve the performance of deployment from a cloud model to the target distribution network. Along with the continuous operation, the proposed method can benefit from its self-learning ability. Meanwhile, the offline mode provides another alternative for various communication conditions or possible restrictions of different distribution networks. With the development of IoT, more new functions can be integrated into the edge-cloud collaboration framework.

The proposed method can detect the HIFs effectively, and the detection results can be used as the start-up criterion of further location algorithms. For decision makers, appropriate protection and recovery measures can usually be implemented only if the exact location is determined. The accuracy HIF location is considered as the crucial future work. Distribution networks are allowed to operate with the existence of HIFs, which means that a second fault may occur before an existing HIF is cleared. The proposed method cannot be directly applied to detect the consecutive or simultaneous faults. Further research on the detection of consecutive or simultaneous faults and distinguish them from a single HIF is required to deal with such scenarios. Finally, methods based on data-knowledge fusion may be the best solution of HIFs, which worth more attention and need to be further studied.

REFERENCES

- [1] A. Ghaderi, H. L. Ginn, and H. A. Mohammadpour, "High impedance fault detection: A review," *Electr. Power Syst. Res.*, vol. 143, pp. 376–388, Feb. 2017.
- [2] D. P. S. Gomes, C. Ozansoy, and A. Ulhaq, "High-sensitivity vegetation high-impedance fault detection based on signal's high-frequency contents," *IEEE Trans. Power Del.*, vol. 33, no. 3, pp. 1398–1407, Jun. 2018.
- [3] N. Bahador, F. Namdari, and H. R. Matinfar, "Modelling and detection of live tree-related high impedance fault in distribution systems," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 3, pp. 756–766, Feb. 2018.
- [4] M. Mitolo, R. Musca, and G. Zizzo, "A cost-effective solution for clearing high-impedance ground faults in overhead low-voltage lines," *IEEE Trans. Ind. Appl.*, vol. 55, no. 2, pp. 1208–1213, Mar. 2019.
- [5] Z. Zhang, X. Zhou, X. Wang, and L. Wang, "A novel diagnosis and location method of short-circuit grounding high-impedance fault for a mesh topology constant current remote power supply system in cabled underwater information networks," *IEEE Access*, vol. 7, pp. 121457–121471, 2019.
- [6] É. M. Lima, C. M. dos Santos Junqueira, N. S. D. Brito, B. A. de Souza, R. de Almeida Coelho, and H. Gayoso Meira Suassuna de Medeiros, "High impedance fault detection method based on the short-time Fourier transform," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 11, pp. 2577–2584, Jun. 2018.
- [7] J. Chen, E. Ambikairajah, D. Zhang, T. Phung, and T. Blackburn, "Detection of high impedance faults using current transformers for sensing and identification based on features extracted using wavelet transform," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 12, pp. 2990–2998, Sep. 2016.
- [8] A. Ghaderi, H. A. Mohammadpour, H. L. Ginn, and Y.-J. 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.
- [9] S. Gautam and Brahma, "Detection of high impedance fault in power distribution systems using mathematical morphology," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1226–1234, May 2013.
- [10] K. Sekar and N. K. Mohanty, "A fuzzy rule base approach for high impedance fault detection in distribution system using morphology gradient filter," *J. King Saud Univ.-Eng. Sci.*, vol. 32, no. 3, pp. 177–185, Mar. 2020.
- [11] Y. Mohammadnian, T. Amraee, and A. Soroudi, "Fault detection in distribution networks in presence of distributed generations using a data mining-driven wavelet transform," *IET Smart Grid*, vol. 2, no. 2, pp. 163–171, Jun. 2019.
- [12] B. K. Chaitanya, A. Yadav, and M. Pazoki, "An intelligent detection of high-impedance faults for distribution lines integrated with distributed generators," *IEEE Syst. J.*, vol. 14, no. 1, pp. 870–879, Mar. 2020.
- [13] T. Hubana, M. Saric, and S. Avdaković, "Approach for identification and classification of HIFs in medium voltage distribution networks," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 5, pp. 1145–1152, Mar. 2018.
- [14] S. Tharmakulasingam, S. Lu, B. T. Phung, D. Zhang, and E. Ambikairajah, "Sustainable deep learning at grid edge for real-time high impedance fault detection," *IEEE Trans. Sustain. Comput.*, early access, Nov. 7, 2018, doi: 10.1109/TSUSC.2018.2879960.
- [15] PSRC Working Group D15, "High impedance fault detection technology," PSRC, Tech. Rep., 1996. Accessed: Aug. 12, 2020. [Online]. Available: https://www.pes-psrc.org/kb/published/reports/High_Impedance_Fault_Detection_Technology.pdf
- [16] K. Chen, J. Hu, Y. Zhang, Z. Yu, and J. He, "Fault location in power distribution systems via deep graph convolutional networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 119–131, Jan. 2020.
- [17] Q. Cui and Y. Weng, "Enhance high impedance fault detection and location accuracy via μ -PMUs," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 797–809, Jan. 2020.
- [18] J. Lu, W. Luan, R. Liu, P. Wang, and J. Lin, "Architecture of distribution Internet of Things based on widespread sensing software defined technology," *Dianwang Jishu/Power Syst. Technol.*, vol. 42, no. 10, pp. 3108–3115, 2018.
- [19] Schneider EcoStruxure Power. Accessed: Aug. 10, 2020. [Online]. Available: <https://www.se.com/ww/en/work/campaign/innovation/power-distribution.jsp>
- [20] Siemens MindSphere. Accessed: Aug. 10, 2020. [Online]. Available: <https://siemens.mindsphere.io/en>
- [21] GE Predix Platform. Accessed: Aug. 10, 2020. [Online]. Available: <https://www.ge.com/digital/iiot-platform>
- [22] J. Ren, G. Yu, Y. He, and G. Y. Li, "Collaborative cloud and edge computing for latency minimization," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 5031–5044, May 2019.
- [23] W. Tao, M. Ma, M. Ding, W. Xie, and C. Fang, "A priority-based synchronous phasor transmission protocol extension method for the active distribution network," *Appl. Sci.*, vol. 9, no. 10, p. 2135, May 2019.
- [24] CoAP: RFC 7252 Constrained Application Protocol. Accessed: Aug. 12, 2020. [Online]. Available: <http://coap.technology/>
- [25] MQTT Version 5.0. Accessed: Aug. 12, 2020. [Online]. Available: <https://docs.oasis-open.org/mqtt/mqtt/v5.0/mqtt-v5.0.html>
- [26] A. von Meier, E. Stewart, A. McEachern, M. Andersen, and L. Mehrmanesh, "Precision micro-synchrophasors for distribution systems: A summary of applications," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2926–2936, Nov. 2017.
- [27] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, 1990.
- [28] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [29] D. Antory, G. W. Irwin, U. Kruger, and G. McCullough, "Improved process monitoring using nonlinear principal component models," *Int. J. Intell. Syst.*, vol. 23, no. 5, pp. 520–544, 2008.
- [30] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, Nov. 2016.
- [31] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 98–113, Jan. 1997.
- [32] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. Accessed: May 21, 2020. [Online]. Available: <http://www.deeplearningbook.org>
- [33] Y. Zhang, X. Wang, Y. Luo, J. He, B. Hua, and Q. Xu, "High impedance fault detection in distribution network using convolutional neural network based on distribution-level PMU data," in *Proc. 8th Renew. Power Gener. Conf. (RPG)*, 2019, pp. 1–8.
- [34] A. Andoni and P. Indyk, "Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions," in *Proc. 47th Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Oct. 2006, pp. 459–468.
- [35] B. Wang, J. Geng, and X. Dong, "High-impedance fault detection based on nonlinear voltage-current characteristic profile identification," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3783–3791, Jul. 2018.



YONGJIE ZHANG (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Beijing Jiaotong University, Beijing, China, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. His research interests include weak fault detection and application of machine learning in distribution networks.



XIAOJUN WANG (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from North China Electric Power University, Beijing, China, in 2001, 2004, and 2008, respectively. From 2014 to 2015, he went abroad in ALSTOM GRID Ltd., U.K., with the BOND Project. He is currently an Associate Professor with Beijing Jiaotong University, Beijing. His main research interests include integrated energy systems and application of artificial intelligence in electrical engineering.



JINGHAN HE received the M.Sc. degree in automation from Tianjin University, Tianjin, China, in 1994, and the Ph.D. degree in electrical engineering from Beijing Jiaotong University, Beijing, China, in 2007. She is currently a Professor with Beijing Jiaotong University. Her main research interests include protective relaying, fault distance measurement, and location in power systems.



YIN XU (Senior Member, IEEE) received the B.E. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2008 and 2013, respectively.

From 2013 to 2016, he was an Assistant Research Professor with the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, USA. He is currently a Professor with Beijing Jiaotong University, Beijing. His research interests include power system resilience, urban electricity-transportation integrated systems, and power system high-performance simulation. He is currently serving as the Chair of the Energy Internet Resilience Working Group under the IEEE PES Energy Internet Coordinating Committee and a Secretary of the Distribution Test Feeder Working Group under the IEEE PES Distribution System Analysis Subcommittee.



FANG ZHANG (Member, IEEE) received the B.Sc. and Ph.D. degrees from Tsinghua University, Beijing, China, in 2010 and 2015, respectively. He is currently an Associate Professor with the School of Electrical Engineering, Beijing Jiaotong University. His research interests include power system dynamic measurement and control based on WAMS.



YIPING LUO (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from Beijing Jiaotong University, Beijing, China, in 2016, where she is currently pursuing the Ph.D. degree. Her research interests include DC grid fault analysis, protection, and fault isolation.

...