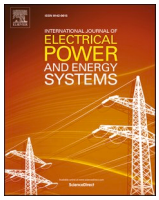
Electrical Power and Energy Systems 140 (2022) 107835

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes



Faulty feeder detection method for SLG faults in distribution networks based on comprehensive fault characteristics across entire frequency spectrum

Jiawei Yuan, Yanjie Hu, Zaibin Jiao *

School of Electrical Engineering, Xi'an Jiaotong University, Xianning West Road No. 28, Xi'an 710000, China

ARTICLE INFO

Keywords:
Comprehensive fault characteristic
Entire frequency spectrum
Faulty-feeder detection
XGBoost algorithm

ABSTRACT

In the event of a single line-to-ground fault in a non-effectively grounded distribution network, reliable detection and isolation of faulty feeders aid in ensuring safe network operation. To improve the accuracy and reliability of faulty-feeder detection, this paper proposes a faulty-feeder-detection method based on the entire-frequency-domain fault characteristics. First, the zero-sequence current variation characteristics of feeders are analyzed using a wavelet packet algorithm; subsequently, the frequency-spectrum energy and zero-sequence current direction are employed to build a detection criterion. Second, the frequency-spectrum energy and direction criteria are combined to obtain a comprehensive detection criterion across the entire frequency spectrum. Third, the ensemble-learning algorithm is utilized to construct the proposed comprehensive criterion. An extreme-gradient-boosting algorithm is employed for efficient modeling along with large amounts of simulated data that are used as training dataset. To verify the reliability and generalization capability of the proposed method, PSCAD simulation, RTDS, and practical fault data are employed considering different topologies, parameters, and fault conditions of various distribution networks. The results obtained in this study reveal the proposed method to improve the accuracy of faulty-feeder detection significantly relative to conventional approaches, which demonstrates that the proposed method shows considerable reliability and generalization potential for practical utility.

1. Introduction

Non-effectively grounded networks are widely used in medium- and low-voltage distribution networks. Moreover, the safe operation of a distribution network is affected by various faults, of which 80% of single line-to-ground (SLG) faults have the greatest impact. When an SLG fault occurs, healthy phase voltages rise and an intermittent arc in the fault point threatens the insulation of the system, which may result in serious faults and wide impacts. However, it is difficult to determine the specific faulty feeder due to smaller SLG fault current and complex fault transient. Consequently, it is of great significance to accurately detect and isolate an SLG fault in distribution networks to ensure safe operation of the entire network.

Several methods were proposed to detect various SLG faults in non-effectively grounded distribution networks, ranging from specific fault feature-based schemes [1–5] to a comprehensive strategy combined with multi-fault features [6–16]. Fault characteristics such as magnitude and phase of the zero-sequence current, the fifth harmonic of the current, and real power were normally used as criteria to detect an SLG fault in conventional distribution networks. These methods were reviewed in [1]. However, because of the complex fault conditions, the reliability of the abovementioned criteria is not acceptable. Therefore, a transient component-based scheme was proposed in [2]. Transient power calculated using the zero-sequence voltage and current was used to detect faulted feeders. However, the power in only a specific frequency band could satisfy the criteria. In [3–4], a wavelet algorithm was used to analyze the zero-sequence currents and voltages spectrum of the feeders. Subsequently, spectrum comparison was conducted to detect the faulty feeders. In [5], the compositions of residual admittances were decomposed, and the faulty feeder was detected by comparing the calculated zero-sequence admittances. However, various factors such as different grounding types, feeder parameters, bypass resistances, and fault locations will result in different spectrum features, which make the detection strategies based on the specific fault characteristics fail.

* Corresponding author.
E-mail address: jiaozaibin@mail.xjtu.edu.cn (Z. Jiao).

https://doi.org/10.1016/j.ijepes.2021.107835
Received 28 May 2021; Received in revised form 16 October 2021; Accepted 24 November 2021
Available online 26 February 2022

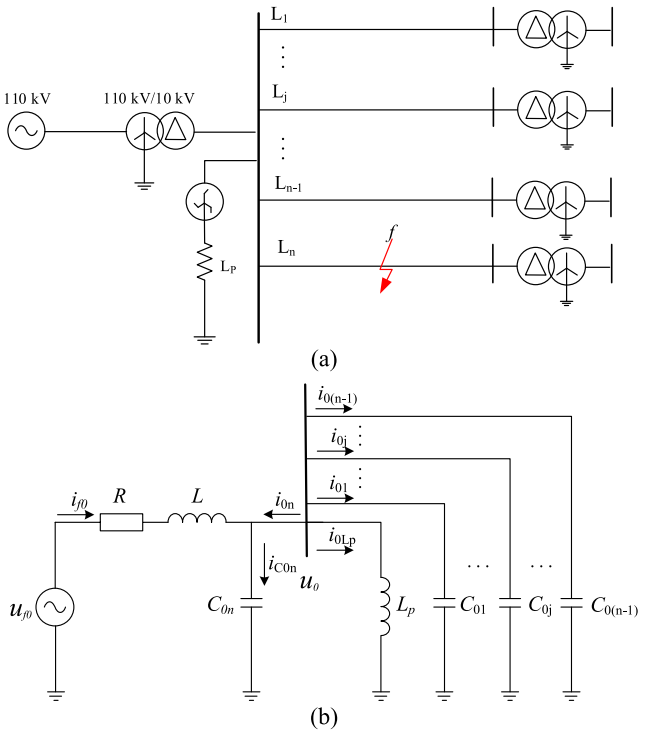


Fig. 1. (a) Diagram of the compensation system; (b) Zero-sequence equivalent circuit with an SLG fault in the compensation system.

Comprehensive methods, which combine different fault characteristics, have been proposed to increase the reliability of faulted feeder detection in distribution networks. Moreover, data driven methods and artificial intelligence (AI) based strategies have been widely used. In [6], fault features such as transient energy, kurtosis, and cross-correlation distance were combined to build a faulty-feeder-detection method using multiple evidence estimation. In [7], the decaying DC component, waveform correlation component, and energy entropy component of zero-sequence currents were extracted to represent the states of feeders, and adaptive network-based fuzzy inference system was employed to combine those components for detecting the faulty feeder. In addition, other intelligent algorithms, such as fuzzy measure fusion criterion [8], rough set theory [9], D-S evidence theory [10], optimal Bayes algorithm [11], support vector machine [12], neural networks [13], mathematical morphology [14], and convolutional neural networks [15–16], are widely used in the field, which lead to higher detection reliability compared to the strategies that only depend on specific fault features. However, generalization problems hinder these methods from being applied practically. Generalization problems usually require AI-based algorithms to be either retrained or subjected to special transfer learning when they are applied to distribution networks with changing topologies, parameters, and grounding modes. Moreover, some particular fault features are used to build a comprehensive criterion, while the others are filtered out. Therefore, a method is required to enable combination of fault features across the entire frequency spectrum, instead of in a specific range, and high applicability in distribution networks.

This paper proposes a comprehensive method, which combines the zero-sequence current variation characteristics with frequency-spectrum energy and direction across the entire frequency spectrum to detect the SLG faulty feeder under different topologies, feeder parameters, and operational conditions. Faulty characteristics are first analyzed using the wavelet packet algorithm. Subsequently, the frequency-spectrum energy and zero-sequence current direction are identified to define the detection criteria for each frequency band. Next, these criteria are combined with the spectrum energy and direction in the respective frequency bands prior to being assembled by an extreme-gradient-boosting (XGBoost) algorithm to construct a classifier that distinguishes faulty feeders from healthy ones. Finally, to verify the generalization of the proposed method, digital simulations for different topologies, feeder parameters, connections, and operational conditions of the distribution networks are conducted. In addition, the verification is conducted using the real-time digital simulation (RTDS) in the hardware-in-the-loop test system and the practical fault data. Based on the entire-frequency-domain characteristics of the faults, the performance of the SLG faulty-feeder detection is improved significantly, demonstrating the strong generalization potential and applicability of the proposed approach.

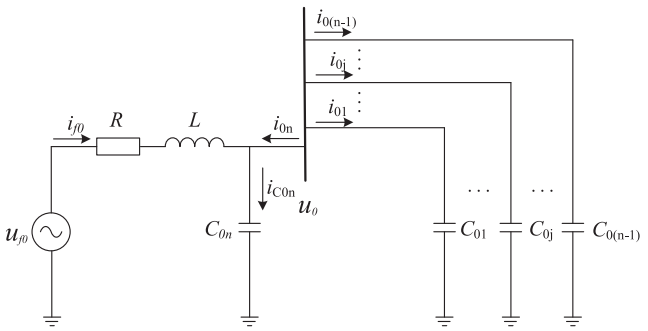


Fig. 2. Simplified zero-sequence equivalent circuit (a) low impedance fault; (b) high impedance fault.

The remainder of the paper is organized as follows. The fault characteristics with frequency-spectrum energy and direction are analyzed in Section 2. In Section 3, the proposed strategy to combine frequency-spectrum energy criterion and direction criterion to achieve a comprehensive detection criterion across the entire frequency spectrum is introduced. In addition, the construction of the XGBoost-based classifier is described, and the training dataset is discussed. The verifications carried out are discussed in Section 4. Further, the PSCAD simulation models, RTDS cases, and practical fault data are introduced along with some necessary analysis and comparisons. Finally, this paper is concluded in Section 5.

2. Spectrum analysis

A typical compensation system with an SLG fault is considered to analyze the fault characteristics in distribution networks. The diagram of the compensation system grounding through Petersen coil with an SLG fault occurred in feeder n is shown in Fig. 1(a), and its zero-sequence equivalent circuit is shown in Fig. 1(b). In Fig. 1(b), R denotes the equivalent resistance of the faulty feeder plus the fault resistance, L denotes the inductance of the faulty feeder, L_p denotes the inductance of the Petersen coil, and C_{0k} is the equivalent capacitance of feeder k. u_{f0} is the zero-sequence voltage at the fault point, u_0 is the zero-sequence voltage on bus, i_{0k} is the zero-sequence current of the healthy feeder k, and i_{0n} is the zero-sequence current of the faulty feeder n.

As shown in Fig. 1 (b), the equivalent circuit is a third-order circuit and needs to be simplified during theoretical analysis. According to the analysis in [17], the resonant frequency of the zero-sequence current is high for an SLG fault with low fault impedance, and the compensation of the Petersen coil can be neglected. Furthermore, for an SLG fault with high fault impedance, the resonant frequency of the zero-sequence current is low, and the compensation of the Petersen coil cannot be neglected, while the inductance of the faulty feeder can be neglected. Consequently, the third-order circuit can be simplified to a second-order circuit, and the zero-sequence equivalent circuit of low impedance fault and high impedance fault are shown in Fig. 2.

The following differential equations and initial conditions are used to analyze the characteristics of the SLG fault with low fault resistance (Fig. 2(a)):

Table 1
Zero-sequence current in different feeders.

Fault type	Feeders	The free component of zero-sequence current
LIUF	Healthy feeder	$i_{0k} = C_{0k}e^{-\delta_1 t}[(-A_1\delta_1 + A_2\omega_{f1})cos\omega_{f1}t + (-A_2\delta_1 - A_1\omega_{f1})sin\omega_{f1}t]$
	Fault feeder	$\boldsymbol{i_{0n}} = -(C_{0\sum} - C_{0n})e^{-\delta_1 t}[(-A_1\delta_1 + A_2\omega_{f1})cos\omega_{f1}t + (-A_2\delta_1 - A_1\omega_{f1})sin\omega_{f1}t] = -[(\boldsymbol{C_{0\sum}} - \boldsymbol{C_{0n}})/\boldsymbol{C_{0k}}]\ \boldsymbol{i_{0k}}$
LIOF	Healthy feeder	$i_{0k} = C_{0k}(B_1p_1e^{p_1 t} + B_2p_2e^{p_2 t})$
	Fault feeder	$\boldsymbol{i_{0n}} = -(C_{0\sum} - C_{0n})(B_1p_1e^{p_1 t} + B_2p_2e^{p_2 t}) = -[(\boldsymbol{C_{0\sum}} - \boldsymbol{C_{0n}})/\boldsymbol{C_{0k}}]\ \boldsymbol{i_{0k}}$
HIUF	Healthy feeder	$i_{0k} = L_PC_{0k}e^{-\delta_2 t}[(A_3\delta_2^2 - A_3\omega_{f2}^2 - 2A_4\delta_2\omega_{f2})cos\omega_{f2}t + (A_4\delta_2^2 - A_4\omega_{f2}^2 + 2A_3\delta_2\omega_{f2})sin\omega_{f2}t]$
	Fault feeder	$\boldsymbol{i_{0n}} = -L_P(C_{0\sum} - C_{0n})e^{-\delta_2 t}[(A_3\delta_2^2 - A_3\omega_{f2}^2 - 2A_4\delta_2\omega_{f2})cos\omega_{f2}t + (A_4\delta_2^2 - A_4\omega_{f2}^2 + 2A_3\delta_2\omega_{f2})sin\omega_{f2}t] - (A_3cos\omega_{f2}t + A_4sin\omega_{f2}t)e^{-\delta_2 t}$ $= -[(\boldsymbol{C_{0\sum}} - \boldsymbol{C_{0n}})/\boldsymbol{C_{0k}}]\ \boldsymbol{i_{0k}} \underline{-(\boldsymbol{A_3cos\omega_{f2}t + A_4sin\omega_{f2}t})\boldsymbol{e^{-\delta_2 t}}}$
HIOF	Healthy feeder	$i_{0k} = B_3e^{\lambda_1 t}\lambda_1^2L_PC_{0k} + B_4e^{\lambda_2 t}\lambda_2^2L_PC_{0k}$
	Fault feeder	$\boldsymbol{i_{0n}} = -(C_{0\sum} - C_{0n})(B_3e^{\lambda_1 t}\lambda_1^2L_P + B_4e^{\lambda_2 t}\lambda_2^2L_P) - B_3e^{\lambda_1 t} - B_4e^{\lambda_2 t} = -[(\boldsymbol{C_{0\sum}} - \boldsymbol{C_{0n}})/\boldsymbol{C_{0k}}]\ \boldsymbol{i_{0k}} \underline{-\boldsymbol{B_3e^{\lambda_1 t}} - \boldsymbol{B_4e^{\lambda_2 t}}}$

Table 2
List of the notations.

Notations	Meaning
$u_{f0}(t)$	$U_msin(\omega_0 t + \theta)$
A	$U_m/\sqrt{(1 - \omega_0^2LC_{0\sum})^2 + (\omega_0RC_{0\sum})^2}$
φ	$\theta - arctan[\omega_0RC_{0\sum}/(1 - \omega_0^2LC_{0\sum})]$
δ_1	$R/2L$
ω_{f1}^2	$1/LC_{0\sum} - \delta_1^2$
p_1	$-R/2L + \sqrt{(R/2L)^2 - 1/LC_{0\sum}}$
p_2	$-R/2L - \sqrt{(R/2L)^2 - 1/LC_{0\sum}}$
A_1	$-Asin\varphi$
A_2	$-(A\delta_1 sin\varphi + A\omega_0 cos\varphi)/\omega_{f1}$
B_1	$(Ap_2 sin\varphi - A\omega_0 cos\varphi)/(p_1 - p_2)$
B_2	$(A\omega_0 cos\varphi - Ap_1 sin\varphi)/(p_1 - p_2)$
B	$U_m/\sqrt{(R - \omega_0^2RL_PC_{0\sum})^2 + (\omega_0L_P)^2}$
η	$\theta - arctan[\omega_0L_P/(1 - \omega_0^2L_PC_{0\sum})R]$
δ_2	$1/2RC_{0\sum}$
ω_{f2}^2	$1/L_PC_{0\sum} - \delta_2^2$
λ_1	$-1/2RC_{0\sum} + \sqrt{(1/2RC_{0\sum})^2 - 1/L_PC_{0\sum}}$
λ_2	$-1/2RC_{0\sum} - \sqrt{(1/2RC_{0\sum})^2 - 1/L_PC_{0\sum}}$
A_3	$-Bsin\eta$
A_4	$-(B\delta_2 sin\eta + B\omega_0 cos\eta)/\omega_{f2}$
B_3	$(B\lambda_2 sin\eta - B\omega_0 cos\eta)/(\lambda_1 - \lambda_2)$
B_4	$(B\omega_0 cos\eta - B\lambda_1 sin\eta)/(\lambda_1 - \lambda_2)$

$$\begin{cases} LC_{0\sum}\dfrac{d^2u_0(t)}{dt^2} + RC_{0\sum}\dfrac{du_0(t)}{dt} + u_0(t) = u_{f0}(t) \\ i_{0k} = C_{0k}\dfrac{du_0(t)}{dt} \\ i_{0n} = -\sum_{k=1}^{n-1} i_{0k} \\ u_0(0_-) = i_{0k}(0_-) = 0 \end{cases} \tag{1}$$

where $C_{0\sum}$ is the sum of the equivalent capacitance of the feeders. $u_{f0} = U_m\sin(\omega_0 t + \theta)$, and U_m, ω_0, and θ are its amplitude, angular frequency, and initial angle, respectively.

To solve (1), we can rewrite the characteristics equation:

$$LC_{0\sum}p^2 + RC_{0\sum}p + 1 = 0 \tag{2}$$

where p is the solution of the characteristic equation, which takes imaginary root in the following discussion.

The characteristics of the SLG fault with a low fault resistance are determined by the feeder parameters, as shown in (2). The zero-sequence currents of the faulty feeder and healthy feeders can be expressed as follows:

$$\begin{cases} i_{0k} = C_{0k}e^{-\delta_1 t}[(-A_1\delta_1 + A_2\omega_{f1})cos\omega_{f1}t + (-A_2\delta_1 - A_1\omega_{f1})sin\omega_{f1}t] \\ i_{0n} = -[(C_{0\sum} - C_{0n})/C_{0k}]\ i_{0k} \end{cases} \tag{3}$$

where $\delta_1 = R/2L$, $\omega_{f1}^2 = 1/LC_{0\sum} - \delta_1^2$. The detailed parameters A_1 and A_2 are related to the system parameters and fault conditions.

Evidently, the zero-sequence current magnitude of the faulty feeder is larger than that of the healthy feeders because $(C_{0\sum}\text{-}\ C_{0n})/C_{0k}$ is greater than 1, and the zero-sequence current of the faulty feeder flows in the direction opposite to that of the healthy ones. In addition, the zero-sequence current spectrum (ω_{f1}) is determined by the parameters R, L, and $C_{0\sum}$. The high frequency component of the zero-sequence current can be used to detect the faulty feeder based on (3) because the main frequency-spectrum energy is located in high frequencies according to the typical parameters in this scenario. However, different distribution networks are characterized by different topologies and feeder parameters, thereby making it difficult to determine the optimum characteristic-frequency band.

The SLG fault discussed above is considered as a low impedance and under-damped fault (LIUF) according to the fault analysis results. Further, three other SLG faults exist, called low impedance and over-damped fault (LIOF), high impedance and under-damped fault (HIUF), and high impedance and over-damped fault (HIOF). The zero-sequence currents of different types of SLG faults are summarized in Table 1, and the detailed parameters are provided in Table 2.

Table 1 indicates that the characteristics of the SLG fault are different because of fault resistance, fault location, feeder parameters, and the topology of distribution networks. The relationship between faulty and healthy feeders' currents of the LIOF is similar to that for the LIUF. However, the main frequency-spectrum energy is located in low frequencies, and a large amount of decaying DC is included in this scenario. The relationship between the faulty and healthy feeders' currents in the case of the HIUF can be expressed as follows:

$$i_{0n} = -[(C_{0\sum} - C_{0n})/C_{0k}]\ i_{0k} - (\boldsymbol{A_3cos\omega_{f2}t + A_4sin\omega_{f2}t})\boldsymbol{e^{-\delta_2 t}} \tag{4}$$

As shown in (4), the characteristics of the zero-sequence current are different for the HIUF. Because of the second term in (4), the magnitude of the faulty feeder and healthy feeders' currents are no longer proportional, and the current direction relationships are uncertain. The same reasoning is valid for the HIOF according to the zero-sequence currents listed in Table 1.

Thus, SLG fault characteristics may differ despite being part of the same distribution network. These characteristics are determined based on feeder parameters, topologies, and fault conditions. Therefore, it is very difficult to propose a general criterion and use its corresponding frequency band to detect the faulty feeder. This is a main reason for several reasonable faulty-feeder-detection methods failing to fulfill their duties properly under certain conditions.

Since the zero-sequence currents in Table 1 are obtained by simplifying the line model using lumped parameter, they only contain fault

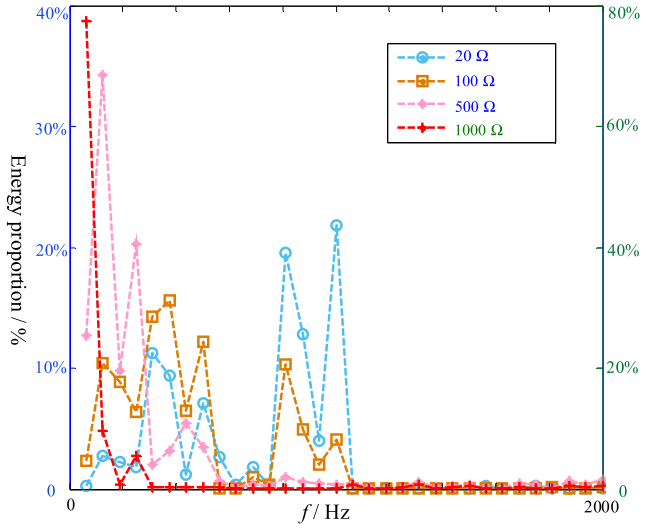


Fig. 3. Frequency-spectrum energy analysis of the faulty feeder's current for SLG faults with different fault resistances.

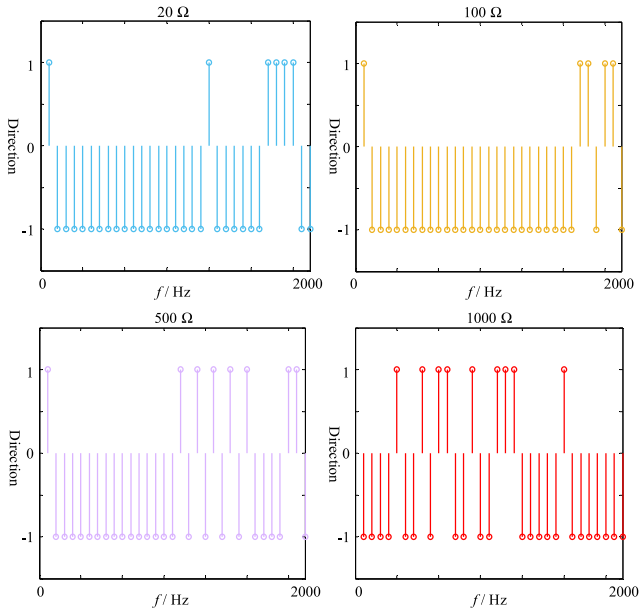


Fig. 4. Direction characteristics of the faulty feeder's current for SLG faults with different fault resistances.

components at resonant frequency. To analyze the fault characteristics across the entire frequency spectrum, a typical distribution network with four feeders, shown in Fig. 1(a), is considered to simulate SLG faults under different fault conditions. It is assumed that SLG faults occur in feeder 1, the initial phase is 247.5°, and the fault resistances are set to 20 Ω, 100 Ω, 500 Ω, and 1000 Ω, respectively. The zero-sequence currents are obtained under different fault scenarios. The spectrum energy analysis of the faulty feeder current is shown in Fig. 3, and the current directions of each frequency are shown in Fig. 4. In Fig. 3, the horizontal coordinate denotes the frequencies of the zero-sequence current, and the vertical coordinate denotes the energy proportion across the entire frequencies, where the left vertical coordinate refers to the energy proportion under fault conditions with resistances of 20 Ω, 100 Ω, and 500 Ω and the right vertical coordinate refers to the energy proportion under fault conditions with resistance of 1000 Ω. In Fig. 4, the vertical coordinate denotes the direction characteristics, where '1' means that the current directions of the faulty feeder and healthy feeders are the same, and '-1' means they flow in the opposite directions.

As shown in Fig. 3, frequency-spectrum energy distributions differ for different cases. As for the SLG fault with 20 Ω fault resistance, approximately 60% frequency-spectrum energy of the zero-sequence current of the faulty feeder is distributed in the range of 800–1000 Hz. When the SLG fault resistance is 100 Ω, 50% of the frequency-spectrum energy is distributed in the range of 300–500 Hz, where only 20% of energy is distributed in the range of 800–1000 Hz. When the fault resistance is 1000 Ω, 80% of the frequency-spectrum energy is distributed in the range of 0–60 Hz, while only 0.2% of the frequency-spectrum energy is in the range of 800–1000 Hz, and 1% of the energy is distributed in the range of 300–500 Hz.

Therefore, with an inappropriate frequency range selection this criterion will fail in detecting the faulty feeder.

Similar results could be obtained in terms of the current directions in the faulty feeder. As shown in Fig. 4, when the relationship between the zero-sequence current directions is used, these directions can accurately detect the faulty feeder in the high-frequency band (800–1000 Hz) for low fault resistances. Furthermore, these directions would fail in high-fault-resistance instances, wherein the main frequency- spectrum energy is located in the low-frequency band (0–60 Hz). It is noteworthy that the frequency bands specified in this study apply exclusively to the case discussed in this paper.

Because the distribution of the frequency-spectrum energy and zero-sequence current directions are determined based on distribution-network parameters and fault conditions, it is very difficult to determine a general criterion corresponding to a fixed frequency or frequency band to realize faulty-feeder detection for all SLG fault types. Therefore, to facilitate high-accuracy faulty-feeder detection, we select fault characteristics across the entire frequency spectrum under the detection criterion for different fault scenarios. In this paper, we propose a comprehensive method, which can (1) combine the criteria with entire frequency spectrum together instead of a specific range; and, (2) combine the criteria based on the spectrum energy and current direction in the corresponding frequency bands. Among them, the entire frequency spectrum in this paper refers to all information of frequency bands, which can be obtained by frequency spectrum analysis according to specific sampling frequency. For instance, the range of entire frequency spectrum is [0, 2000] Hz in Fig. 3 and Fig. 4. Hence, the comprehensive fault characteristics are used across the entire frequency spectrum, and the accuracy of the faulty-feeder detection is increased drastically.

3. Proposed Fault-feeder-detection method

3.1. Frequency spectrum analysis of the zero-sequence current

When an SLG fault occurs, the zero-sequence current of each feeder can be sampled directly or calculated indirectly based on the phase current. Subsequently, the spectrum analysis is conducted using quarter cycle waveform by wavelet packet decomposition [18]. The frequency-spectrum energy in sub-bands can be expressed as follows:

$$e_i^j = \sum_{k=1}^{n} \left(c_k^j\right)^2 \tag{5}$$

where e_i^j is the frequency-spectrum energy of zero-sequence current from the feeder i in the j-th sub-band. c_k^j is the k-th decomposition coefficient in the j-th sub-band, which can be calculated using general algorithm shown in [18]. Additionally, n denotes the number of coefficients in the sub-band.

Since the frequency-spectrum energy is affected by feeder parameters and fault conditions, it is necessary to conduct normalization in the j-th sub-band as follows:

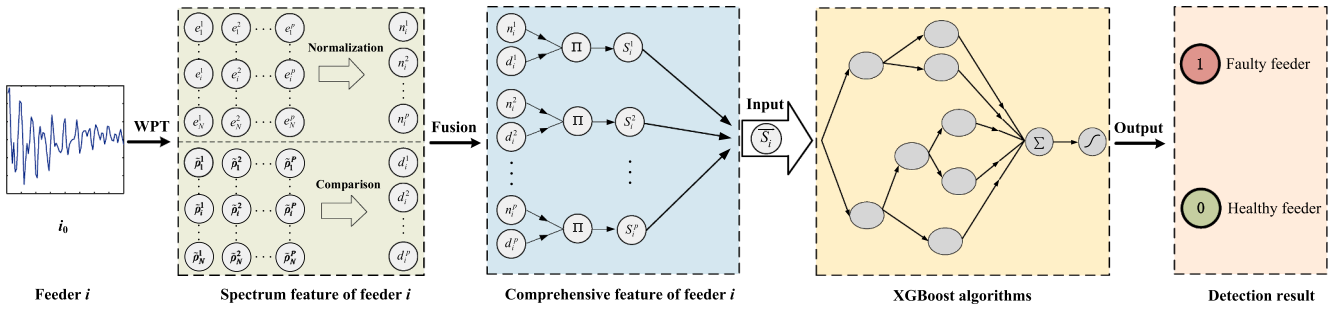


Fig. 5. Flow framework of the proposed SLG faulty-feeder-detection method.

$$n_i^j = \frac{e_i^j}{\sum_{i=1}^{N}\sum_{j=1}^{P} e_i^j} \tag{6}$$

where P is the number of sub-frequency bands and N is the number of feeders.

After the spectrum analysis, the zero-sequence current is reconstructed for each individual frequency. Consequently, the direction-based criteria can be proposed in the corresponding frequency band. As for the direction-based criteria, the correlation coefficient can be used to describe the relationship between the signals in time domain. Based on the principle of the waveform correlation analysis, the correlation coefficient between the zero-sequence currents in each frequency can be calculated using the following equation:

$$\rho_{i,m}^j = \frac{\sum g_{i,j}(s) g_{m,j}(s)}{\left[\sum g_{i,j}^2(s) g_{m,j}^2(s)\right]^{1/2}} \tag{7}$$

where $g_{i,j}$(s) is the wavelet packet reconstruction coefficient of the zero-sequence current of the feeder i in the j-th sub-band. $g_{m,j}$(s) is the wavelet packet reconstruction coefficient of the zero-sequence current sampled from the feeder m in the j-th sub-band. $\rho_{i,m}^j$ is the signal correlation coefficient between the feeder i and feeder m in the j-th sub-band.

The correlation coefficients between the feeder i and other feeders can be calculated using (7), and they can be rewritten as a column vector as follows:

$$\widetilde{\rho}_i^j = [\rho_{i,1}^j, \cdots, \rho_{i,i-1}^j, \rho_{i,i+1}^j, \cdots, \rho_{i,N}^j]^T \tag{8}$$

We can use d_i^j, shown in (9), to describe the direction relationship of the zero-sequence currents between the i-th feeder and other feeders.

$$d_i^j = \begin{cases} -1, & if\ u > v \\ 1, & if\ u \leqslant v \end{cases} \tag{9}$$

where $u = count\left(\rho_{i,m}^j < 0 \middle| \widetilde{\rho}_i^j\right)$, $v = count\left(\rho_{i,m}^j \geqslant 0 \middle| \widetilde{\rho}_i^j\right)$, and $count(.)$ describe the elements member for which specific conditions can be satisfied.

As shown in (9), the direction of the zero-sequence current in j-th sub-band among the i-th feeder and other feeders is negative if d_i^j is equal to -1, otherwise the direction is positive.

3.2. Comprehensive fault characteristic criterion

To comprehensively describe the fault characteristics, the frequency-spectrum energy and direction characteristics in the corresponding frequency can be combined together to construct an integrated criterion for faulty-feeder detection.

$$\begin{cases} f(S_i^j) \\ S_i^j = n_i^j \times d_i^j \end{cases} \tag{10}$$

where $f(S_i^j)$ describes the criterion based on the S_i^j. The detailed criterion f is constructed using XGBoost algorithm, which is a data-driven ensemble learning [19] method.

3.3. XGBoost-based detection algorithm

As the most applied ensemble-learning algorithm, the XGBoost [20] is widely used in classification, regression, feature extraction, and outlier detection. The XGBoost algorithm utilizes multiple base learners (Decision Tree is usually employed) to construct the detection scheme, and optimizes the loss function by fitting the negative gradient value at the previous iteration. Furthermore, it introduces the complexity of the learner model to the objective function, which helps to smooth the final learnt weights to avoid overfitting. The objective function is minimized using (11).

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{11}$$

where l is a loss function that measures the differences between the prediction $\hat{y}_i$ and the target y_i, and Ω denotes the penalty function of the tree model complexity.

For the objective function at the t-th iteration, it will be minimized using (12).

$$L^{(t)} = \sum_i l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{12}$$

where x_i is the features of i-th sample, $f_t(x_i)$ is the function of tree model to be learned at the t-th iteration.

Perform second-order Taylor expansion on the loss function, and equation (12) can be rewritten as (13).

$$L^{(t)} = \sum_i [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \tag{13}$$

where g_i and h_i are the first and second order gradient statistics on the loss function, respectively.

Owing to the second-order Taylor expansion and penalty function, the XGBoost algorithm has achieved state-of-art performance in a number of machine learning and data mining challenges. Moreover, it supports parallel processing of multi-core processors and improves their optimization speed.

The comprehensive scheme with the XGBoost algorithm based on the fusion of the integrated fault characteristics with different frequencies across the entire frequency spectrum is exhibited in Fig. 5.

As shown in Fig. 5, with the feeder i taken as an example, the spectrum analysis of the zero-sequence current is conducted. Subsequently, the frequency-spectrum energy and direction characteristics of the zero-sequence current across the entire frequency are calculated using (6) and (9). Next, the S_i^j value for each feeder is obtained using (10); these are used as the inputs of XGBoost algorithm. The XGBoost algorithm is trained by a dataset, which comprises a large amount of fault data from

simulation. The detailed training process will be discussed later in this paper. Finally, we can achieve the faulty-feeder-detection result from the output of the XGBoost algorithm. The feeder whose label is '1' is determined as faulty feeder, and the feeders with a label of '0' are healthy feeders. Furthermore, if the labels of all feeders are '0', the fault has occurred in the bus bar.

A large fault dataset should be used to train the XGBoost algorithm to ensure the proposed method performs as expected and is ready for use in practical applications. To obtain sufficient simulation data, three training models with four feeders are established using the PSCAD software package to simulate SLG faults that occur in feeders and buses. The structure of these simulation models is shown in Fig. 1(a). Overhead and cable lines are considered in these simulations.

In fact, it is usually accompanied by arc grounding events when SLG faults occur in distribution networks. However, there are large differences between the arc simulation conducted by PSCAD simulations and the practical real faults. In this paper, the arc fault scenario is not considered in the training process of XGBoost model. Meanwhile, arc fault data from practical distribution networks and RTDS simulations are used in the test process, which will be shown in Section IV. In fact, the test results show the good performance of the proposed method for SLG faults accompanied by arc grounding events in real distribution systems. The parameters concerning the overhead lines are as follows: $R_1 = 0.17\,\Omega$/km, $L_1 = 1.21$ mH/km, $C_1 = 0.0097$ μF/km, $R_0 = 0.23\,\Omega$/km, $L_0 = 5.48$ mH/km, and $C_0 = 0.006$ μF/km. Likewise, the parameters concerning cable lines include $R_1 = 0.098\,\Omega$/km, $L_1 = 0.274$ mH/km, $C_1 = 0.351$ μF/km, $R_0 = 0.246\,\Omega$/km, $L_0 = 0.955$ mH/km, and $C_0 = 0.166$ μF/km. Furthermore, Gaussian white noise with a signal-to-noise ratio (SNR) of 50 dB is considered, and the sampling frequency is 4000 Hz, which implies that the range of entire frequency spectrum is [0, 2000] Hz in the following test. The db10 wavelet function is used since it has good orthogonal characteristics and decomposition effect for the non-periodic transient attenuation signals [21], and the wavelet packet adopts five-layer decomposition according to the sampling frequency. In addition, different fault locations, fault time, and fault resistances are also considered in the data production processes, and 3060 sets of fault data are generated. The detailed fault scenarios are summarized in Table 3.

For the training of the XGBoost algorithm, the obtained fault data are firstly split into training dataset and validation dataset, where 90% of the fault data are used for training and 10% are used for validation purposes. Subsequently, the XGBoost algorithm is trained using the packaged function 'XGBClassifier' of xgboost base in python, and scikit-learn tools are utilized for data processing, which are the available machine learning tools for predictive data analysis. Furthermore, grid search in scikit-learn tools is employed for parameters optimization. After parameters optimization of the XGBoost algorithm using the training and validation datasets, the main parameters in 'XGBClassifier' function are as follows: the learning rate is 0.11, the number of Decision Tree is 100, the max depth of Decision Tree is 3, and the minimum child weight is 1. Finally, a detection accuracy of up to 100% could be realized for the training dataset and validation dataset using the trained XGBoost-based detection scheme.

Fig. 6. Flowchart of the SLG faulty-feeder detection.

4. Case study

To verify the generalization of the proposed method, large amounts of data are generated from the simulations of different distribution networks with different topologies, parameters, connections, feeder types, and fault conditions using PSCAD software and RTDS hardware-in-the-loop (HIL) test system. Furthermore, some practical fault data in real power system are collected to test the performance of the proposed method.

For SLG faulty-feeder detection in the test process, zero-sequence currents of feeders are collected after SLG faults. Subsequently, the spectrum energy and direction are extracted and combined in each frequency band, and the comprehensive fault characteristics across the entire frequency spectrum are input into the trained XGBoost-based detection model. Finally, the feeder with an output of '1' is determined as faulty feeder, and the feeder with an output of '0' is healthy feeder. The whole flowchart of the proposed SLG faulty-feeder detection is shown in Fig. 6.

It is noteworthy that the detection performance of the proposed method is tested under new fault scenarios that is not considered in training. Moreover, the trained XGBoost detection model does not need to be retrained when it is used in different distribution systems, and it can be directly applied to the faulty-feeder detection in the following test, which confirms its strong generalization ability.

4.1. PSCAD simulation

Three cases are designed in PSCAD simulations to cover different fault scenarios considering more complex conditions in distribution networks, such as larger grounding resistance, increased system noise, and overhead lines & cables hybrid. Furthermore, varied topologies

Table 3
Parameters and fault scenarios in training set.

PSCAD Simulation		Training Samples		
Parameters	Model	Training model 1	Training model 2	Training model 3
	Grounding mode	Compensation system	Compensation system	Ungrounded system
	Compensation degree	8%	8%	/
	Length (km)/ Feeder type	10/20/30/40 (Overhead)	10/20 (Cable) 30/40 (Overhead)	10/20/30/40 (Overhead)
Scenarios	Fault location	10%/50%/90% (Line fault)/Bus (Bus fault)		
	Fault inception angle	0–360° per 22.5° (Line fault)/0–360° per 7.2° (Bus fault)		
	Grounding resistances	20 Ω/100 Ω/500 Ω/1000 Ω		
	Noise	50 dB		
Sample number		1020	1020	1020

Table 4
Parameters and fault scenarios in test set.

PSCAD Simulation		Test Samples				
Parameters	Model	Test model 1	Test model 2	Test model 3	Test model 4	Test model 5
	Grounding mode	Ungrounded system	Compensation system	Ungrounded system	Compensation system	Compensation system
	Compensation degree	/	6%	/	10%	10%
	Feeder numbers	6	6	3	3	5
	Length (km)	5/15/25/35/45/50		5/20/80		5/15/25/35/45
	Feeder type	Overhead		Overhead		Cable/ hybrid/ Overhead
	Cable ratio	/		/		100%/20%/60%/0%/0%
Scenarios	Fault location	20% (L1)/30% (L2)/40% (L3) /50% (L4)/ 60% (L5)/70% (L6)/Bus		20% (L1)/40% (L2)/60% (L3)/Bus		80% (L1)/60% (L2)/40% (L3) /20% (L4)/10% (L5)/Bus
	Initial phases	0–360° per 9°				
	Grounding resistances	10 Ω/80 Ω/350 Ω/780 Ω/1350 Ω/1700 Ω/2000 Ω				
	Noise	30 dB				
Sample number		2009	2009	1148	1148	1722

Table 5
Detection results in PSCAD simulation.

Method	Test results on the basis of data set from PSCAD simulation				
	Case I		Case II		Case III
	Test model 1	Test model 2	Test model 3	Test model 4	Test model 5
C	**100.00%**	**100.00%**	**100.00%**	**98.61%**	**97.16%**
C1	85.71%	85.71%	75.00%	45.56%	56.39%
C2	97.71%	42.83%	97.91%	48.61%	47.50%
C3	90.49%	94.30%	65.25%	65.77%	82.35%
C4	97.61%	95.52%	80.05%	74.65%	78.75%

from system to system and various feeder parameters are completely different from those in the training set.

In *CASE I*, an ungrounded and a compensation system with six overhead feeders are considered, whose topologies are different from the distribution networks in training dataset. The feeder parameters are as follows: $R_1 = 0.33\,\Omega$/km, $L_1 = 1.31$ mH/km, $C_1 = 0.007$ μF/km, $R_0 = 1.041\,\Omega$/km, $L_0 = 3.96$ mH/km, and $C_0 = 0.004$ μF/km. Gaussian white noise with the SNR of 30 dB is added to the zero-sequence current.

In *CASE II*, an ungrounded and a compensation system with three feeders are considered. To verify the performance of the proposed method in distribution networks with extreme topologies, significantly different feeder lengths of 5 km, 20 km, and 80 km are considered. In this case, there would be a large zero-sequence current in the 80 km feeder when an SLG fault occurs in the 5 km or 20 km feeder, thereby posing great challenges to faulty-feeder detection.

In *CASE III*, a compensation system composed of overhead lines, cable lines, and hybrid lines is considered. The feeder parameters of the cable line are as follows: $R_1 = 0.0791\,\Omega$/km, $L_1 = 0.2642$ mH/km, $C_1 = 0.373$ μF/km, $R_0 = 0.2273\,\Omega$/km, $L_0 = 0.9263$ mH/km, and $C_0 = 0.166$ μF/km.

Five test models are built and 8036 sets of data are generated, as summarized in Table 4.

Table 5 summarizes the faulty-feeder-detection results obtained using the proposed and four other detection methods. Among them, C1 constructs the detection scheme using (14) [22], and C2 constructs the detection scheme using (15) [23]. They are widely used in practice.

$$e_f^{j_M} > e_i^{j_M}, i = 1, 2, \cdots, N, i \neq f \tag{14}$$

$$d_f^{j_M} = -d_i^{j_M}, i = 1, 2, \cdots, N, i \neq f \tag{15}$$

where f is the faulty feeder, j_M is the frequency band with the maximum frequency-spectrum energy, e denotes the spectrum energy, and d denotes the direction characteristic.

Table 6
Detection results based on different intelligent algorithms.

PSCAD Model	**XGBoost-based detection method**	GBDT-based detection method	SVM-based detection method
Test model 1	**100%**	100%	100%
Test model 2	**100%**	100%	100%
Test model 3	**100%**	100%	100%
Test model 4	**98.61%**	97.30%	98.34%
Test model 5	**97.16%**	96.52%	95.88%

C3 detects the faulty feeder based on $f(n_i^j)$ using XGBoost algorithm, where $f(n_i^j)$ describes the criterion based on the n_i^j across the entire frequency spectrum. C4 detects the faulty feeder based on $f(d_i^j)$ using XGBoost algorithm, where $f(d_i^j)$ describes the criterion based on the d_i^j across the entire frequency spectrum. Furthermore, C denotes the proposed method wherein the spectrum energy and direction characteristics across the entire frequency spectrum are combined to detect faulty feeders in distribution networks.

The accuracy of the proposed faulty-feeder-detection method is improved when compared with that of the other methods that work based on a specific single fault characteristic, as summarized in Table 5.

In *CASE I*, the simulation models comprise six feeders with small differences in their lengths. With regard to the detection results of the ungrounded system in *CASE I*, the detection accuracies of C2 and C4 equal approximately 98% while those of C1 and C3 remained relatively low. With regard to the detection results of the compensation system in *CASE I*, the detection accuracies of C3 and C4 equal approximately 95% while that of C2 equal 42.83%. Compared to the four existing methods, the proposed method demonstrates 100% faulty-feeder-detection accuracy for both the ungrounded and compensation systems.

In *CASE II*, the simulation models comprise three feeders, the lengths of which differ significantly, thereby posing a great challenge with regard to faulty-feeder detection. With regard to the simulation results obtained for the ungrounded system in *CASE II*, the C2 demonstrates a high detection accuracy while C3 and C4 demonstrate poor faulty-feeder-detection performance. With regard to the simulation results of the compensation system in *CASE II*, the detection accuracies of C1 and C2 are observed to reduce rapidly while C3 and C4 demonstrate a significant improvement. Compared to the four existing methods, the proposed method is observed to be less influenced by the feeder structure, and the detection accuracy achieved equal 100% and 98.61% for

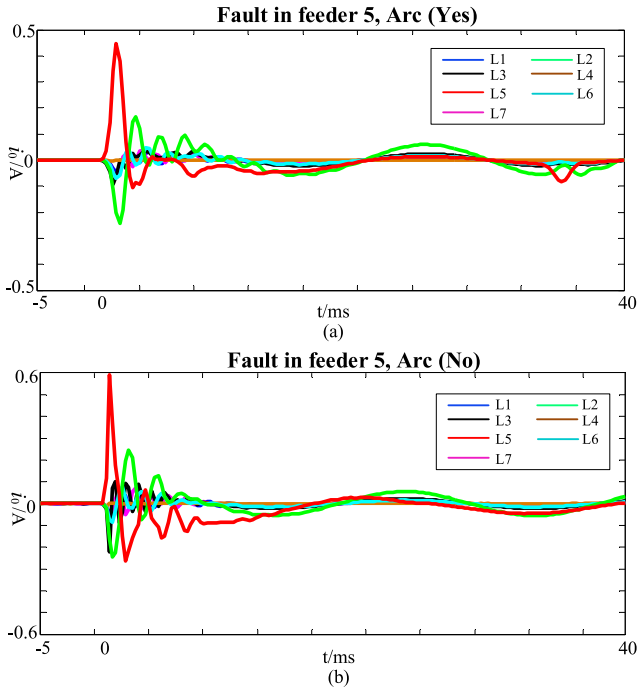


Fig. 7. RTDS model of 10-kV distribution system.

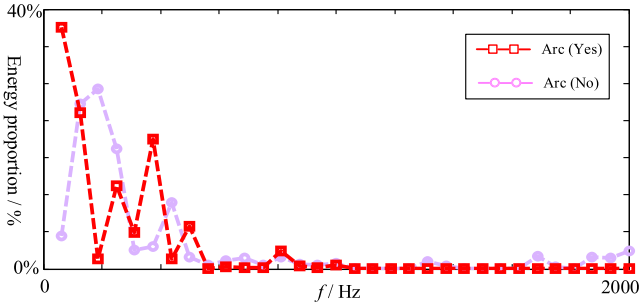


Fig. 8. RTDS-based hardware-in-the-loop simulation system.

the ungrounded and compensation systems, respectively.

In *CASE III*, the fault characteristics of the zero-sequence current are more complex owing to the consideration of multiple feeder types. This results in the observed low detection accuracies obtained using the four existing methods—C1, C2, C3, and C4. In contrast, in the case of the proposed method, the detection scheme is constructed using the comprehensive fault characteristic across the entire frequency spectrum, and finally a detection accuracy of 97.16% is yielded.

Furthermore, since intelligent algorithms with different learning abilities may affect the fusion performance, gradient boosting decision tree (GBDT) [24] and support vector machine (SVM) [25] are selected as classifiers to construct the detection schemes for comparison, respectively. The detection accuracy based on different algorithms is summarized in Table 6. It can be seen that the XGBoost-based detection method has the optimal detection performance, which demonstrates that the XGBoost algorithm is suitable for combining the fault characteristics across the entire frequency spectrum.

In summary, the conventional methods C1 and C2 only employ fault characteristics within a specific frequency band, and consequently, they cannot adapt to faulty-feeder detection in specific fault scenarios, such as high fault resistances, small initial phases, extreme topologies, and hybrid systems. For C3 and C4, the detection accuracy has a significant improvement owing to the combination of fault characteristics across the entire frequency spectrum. However, C3 and C4 construct the detection scheme only based on a single type of characteristic (spectrum energy or direction) over the entire frequency spectrum, and they cannot utilize the comprehensive fault characteristics in each frequency band. In contrast, owing to the comprehensive utilization of the frequency-spectrum-energy and direction characteristics across all frequency bands, the proposed method significantly improves the faulty-feeder-detection accuracy, as confirmed by the above-described simulation results.

4.2. RTDS simulation

The RTDS-based HIL test system is usually employed to test the performance of the detection schemes before they are put into practice. To verify the performance of the proposed method, two RTDS simulation models, shown in Fig. 7, are designed in the HIL test system, as shown in Fig. 8.

The established network contains two grounding models, including neutral point ungrounded and neutral point compensated, and 15 feeders are connected to two Buses. Furthermore, considering the transfer characteristics of CT, the data of zero-sequence currents are collected from practical fault recorder using RTDS-based HIL test system. The feeder parameters and fault scenarios used in RTDS simulation are summarized in Table 7.

Since SLG faults are usually accompanied with arc grounding events, it is necessary to verify the detection performance in fault scenarios with arc faults. The Mayr arc model [26] is employed to simulate the arc faults, and it can be expressed as (16).

$$\frac{dq(t)}{dt} = e_{arc}(t)i_{arc}(t) - p_{loss} \tag{16}$$

where q(t) is the are energy, e_{arc}(t) is the arc voltage per unit length, i_{arc}(t) is the arc current, and p_{loss} is the energy loss per unit length.

Equation (16) can be further expressed as:

$$\frac{1}{g(t)}\frac{dg(t)}{dt} = \frac{1}{\tau}\left(\frac{e_{arc}(t)i_{arc}(t)}{p_{loss}} - 1\right) \tag{17}$$

Table 7
Parameters and fault scenarios in RTDS simulation.

RTDS Simulation		Test Samples					
Parameters	Type	Overhead line			Cable line		
	Phase-sequence	R(Ω/km)	L (mH/km)	C (μF/km)	R (Ω/km)	L (mH/km)	C (μF/km)
	Positive- sequence	0.33/0.45	1.305/1.305	0.007/0.0068	0.0791/0.098	0.264/0.274	0.373/0.351
	Zero-sequence	1.041/1.443	3.963/4.046	0.004/0.0039	0.2273/0.2462	0.926/0.955	0.166/0.166
Scenarios	Model	RTDS model 1			RTDS model 2		
	Grounding mode	Ungrounded system			Compensation system		
	Faulted location	L1/L5/L7/L9/Bus			L1/L5/L7/L9/L10/Bus		
	Fault inception angle	30°/ 90°/ 120°/ 123°/ 150°/ 180°/ 210°/ 270°/ 330°			30°/ 60°/ 90°/ 123°/ 150°/ 210°/ 243°/ 270°/ 330°		
	Grounding resistances	1.1 Ω/5.82 Ω/10 Ω/1000 Ω			1.1 Ω/10 Ω/240 Ω/1000 Ω		
	Unbalance voltage	0 V/2.68 V/5.67 V			0 V/2.68 V/5.67 V		
	Arc grounding	Yes/No			Yes/No		
Sample number		31			38		

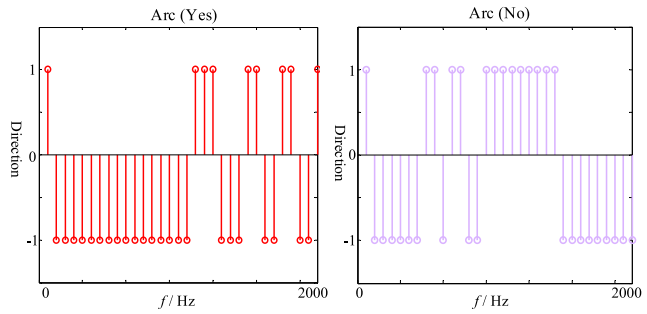


Fig. 9. Zero-sequence currents when SLG faults occur in RTDS.

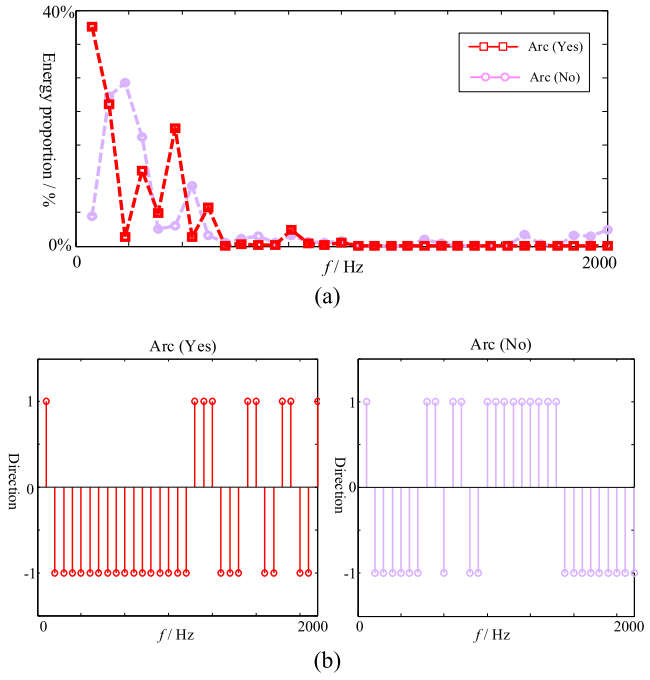


Fig. 10. Spectrum analysis of the faulty feeder's current for SLG faults in RTDS (a) frequency-spectrum energy; (b) direction characteristics.

where $g(t)$ is the arc conductance, and τ is the time constant, as shown in (18).

$$\tau = \frac{g(t)}{p_{loss}} \frac{dq(t)}{dg(t)} \tag{18}$$

Considering the arc length, equation (17) can be rewritten as (19).

$$\frac{1}{g(t)} \frac{dg(t)}{dt} = \frac{1}{\tau} \left(\frac{u_{arc}(t) i_{arc}(t)}{p_0} - 1 \right) \tag{19}$$

where $u_{arc}(t) = l \times e_{arc}(t)$, $p_0 = l \times p_{loss}$, and l is the arc length.

The arc length l is set to 10 ~ 400 cm in RTDS simulation, and the detailed fault conditions are shown in Table 7. The zero-sequence current of SLG faults with and without arc grounding events are shown in Fig. 9.

Compared with the zero-sequence current shown in Fig. 9 (b), the zero-sequence current of the faulty feeder in the case of arc grounding shown in Fig. 9 (a) has noticeable distortions, which would present a challenge to faulty-feeder detection. To investigate the spectrum differences between SLG faults with and without arc, the spectrum energy analysis of the faulty-feeder zero-sequence current is shown in Fig. 10 (a), and the current direction of each frequency is shown in Fig. 10(b).

As shown in Fig. 10, approximately 40% frequency-spectrum energy

Table 8
Detection results in RTDS simulation.

Method	C	C1	C2	C3	C4
Model 1	**100%**	93.55%	54.84%	67.74%	80.65%
Model 2	**100%**	89.47%	42.11%	94.74%	57.89%

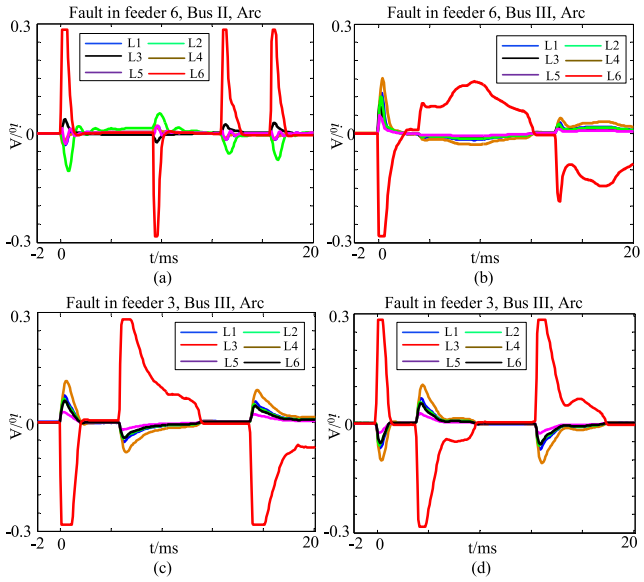


Fig. 11. Zero-sequence currents when SLG faults with arc grounding occur in the practical distribution system.

of the faulty feeder current is distributed in the range of 0–60 Hz when an SLG fault with arc grounding occurs, while only 5% frequency-spectrum energy is distributed in the same range in the fault scenario without arc grounding. Furthermore, the direction criterion would fail in the low-frequency band (0–60 Hz) and most of the high-frequency band due to the direction misjudgment.

To test the performance of the proposed method, the typical distribution systems with an ungrounded system (model 1) and with a compensation system (model 2) are modeled. Totally 69 sets of fault data are collected from RTDS simulation, which contains 16 sets of arc grounding events, as summarized in Table 7. The detection results of the methods are summarized in Table 8.

As for the simulation results obtained for RTDS model 1, C1 demonstrates 93.55% detection accuracy, whereas C2 demonstrates a corresponding accuracy of only 54.84%. This could be attributed to the direction misjudgment of the healthy feeders caused by the electromagnetic environment. Furthermore, methods C3 and C4 demonstrate a less satisfactory faulty-feeder-detection performance. With regard to the simulation results obtained for RTDS model 2, C1 and C3 demonstrate high detection accuracies, whereas C2 and C4 demonstrate very low detection accuracies.

In contrast, the proposed method presents a detection accuracy of 100% under all fault scenarios, and it is not affected by the system structure and feeder parameters, which confirms its strong generalization ability and considerable application prospects.

4.3. *Test by using practical fault data*

To test the detection performance of the proposed method in the practical distribution systems, this paper collects practical SLG fault data from Dongjia Village substation from 15: 8: 47 on September 29, 2020 to 11: 35: 40 on October 27, 2020. The distribution network contains fourteen feeders, where eight feeders are connected to Bus I and six feeders are connected to Bus II. Furthermore, totally 44 sets of fault data

Table 9
Detection results in the practical distribution system.

Data	Arc grounding	Number	Accuracy
Practical fault data	Yes	4	100%
	No	40	100%

are collected from practical fault recorder, where 20 sets of SLG faults occur in feeders connected to Bus I and 24 sets of SLG faults occur in feeders connected to Bus II. The practical fault data set contains 4 sets of fault data with obvious arc grounding events, and the current waveforms are shown in Fig. 11.

As shown in Fig. 11, the zero-sequence currents have distinct process of arc extinguishing and arcing, which is obviously different from other practical fault data from the same substation. And these practical fault data also have different characteristics comparing to the training dataset.

For the faulty-feeder detection using the proposed method, the zero-sequence currents of feeders are collected from fault recorder, while the detailed parameters of the practical distribution system are not recorded. However, the proposed method detects the faulty feeder only based on the fault characteristics of zero-sequence currents, and does not require the detailed parameters of the distribution system. Finally, the faulty feeder can be detected according to the detection flowchart, as shown in Fig. 6, and the test results by using practical fault data are shown in Table 9.

As shown in Table 9, the proposed method based on the XGBoost algorithms trained with simulation data in PSCAD can be directly applied to the faulty-feeder detection in the practical distribution system, and it has 100% detection accuracy in these 44 sets of practical fault data including 4 sets of arc fault data.

In summary, the comprehensive utilization of the frequency-spectrum-energy and direction characteristics across all frequency bands guarantees the proposed method less influenced by the arcing effect, which confirms its considerable application prospects in engineering.

5. Conclusions

The spectrum characteristics of the zero-sequence current vary according to the feeder parameters and fault scenarios when the SLG faults occur in distribution networks. This results in sub-optimal performance of faulty-feeder-detection methods based on fault characteristics existing within specific frequency bands. This paper proposes a comprehensive SLG faulty-feeder detection method based on the integrated fault characteristics of the zero-sequence current across the entire frequency spectrum. The integrated fault characteristics are obtained by combining frequency-spectrum energy and the zero-sequence current direction in the corresponding bands, and the faulty-feeder-detection method is constructed using the XGBoost algorithm as a classifier. To verify the generalization capability of the proposed method, various cases with different topologies, parameters, feeder types, and fault conditions are built in PSCAD and RTDS software packages, and practical fault data collected from real distribution networks are employed to verify the performance of proposed method. The high detection accuracy of the proposed method can be attributed to the consideration of comprehensive fault characteristics across the entire frequency spectrum. In contrast, other detection methods are based on the consideration of single fault characteristics in the maximum-energy-frequency or all frequency bands. The results obtained in this study confirm that the proposed method considerably improves the faulty-feeder-detection accuracy owing to the comprehensive utilization of the frequency-spectrum energy and direction characteristics across the entire frequency spectrum. Furthermore, the proposed method demonstrates a strong generalization capability for use in practical applications.

Despite its high detection accuracy and strong generalization capability, the proposed approach can be further improved in terms of its hybrid-system (overhead lines + cables) performance and self-learning of fault characteristics. The authors intend to consider these objectives in future research.

CRediT authorship contribution statement

Jiawei Yuan: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Yanjie Hu:** Writing – original draft, Writing – review & editing. **Zaibin Jiao:** Writing – original draft, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Yuan S, Zhao J, Song Y. Analysis and comparison of several fault line selective methods in small current grounding power system. In: 2008 China International Conference on Electricity Distribution; 2008. p. 1–7.
[2] Zhang XH, Ha HX, Pan ZC, Xu BY. Grounding faulty line selection in non-solidly grounded systems using transient energy. In: 2007 International Power Engineering Conference (IPEC 2007), Singapore; 2007. p. 1147–50.
[3] Dong X, Shi S. identifying single-phase-to-ground fault feeder in neutral noneffectively grounded distribution system using wavelet transform. IEEE Trans Power Delivery 2008;23(4):1829–37.
[4] Michalik M, Rebizant W, Lukowicz M, Lee SJ, Kang SH. High-impedance fault detection in distribution networks with use of wavelet-based algorithm. IEEE Trans Power Delivery 2006;21(4):1793–802.
[5] Liu BW, Ma HZ, Xu HH, Ju P. Single-phase-to-ground fault detection with distributed parameters analysis in non-direct grounded systems. CSEE J Power Energy Syst 2019;5(1):139–47.
[6] Wei XX, Yang DC, Wang XW, Wang B, Gao J, Wei KW. Faulty feeder detection based on fundamental component shift and multiple-transient-feature fusion in distribution networks. IEEE Trans. Smart Grid 2021;12(2):1699–711.
[7] Yuan JW, Jiao ZB, Feng G, Chen M, Xu MM. Study on fault line detection methods based on multi-feature fusion in distribution systems. IET Gener Transm Distrib 2020:1–10.
[8] Yu K, Zou H, Zeng XJ, Li YY, Li H, Zhuo C, Wang Z. Faulty feeder detection of single phase-earth fault based on fuzzy measure fusion criterion for distribution networks. Int J Electr Power Energy Syst 2021;125:106459.
[9] Hu CH. Distribution network fault location based on rough set and data fusion. In: 2019 IEEE 4th advanced information technology, electronic and automation control conference (IAEAC), Chengdu, China; 2019. p. 1872–6.
[10] Jin N, Xing JW, Liu Y, Li ZT, Lin XN. A novel Single-phase-to-ground fault identification and isolation strategy in wind farm collector line. Int J Electr Power Energy Syst 2018;94(7):15–26.
[11] Faiz J, Lotfi-fard S, Shahri SH. Prony-based optimal Bayes fault classification of overcurrent protection. IEEE Trans Power Delivery 2007;22(3):1326–34.
[12] Ghaderi A, Mohammadpour HA, Ginn HL, Shin Y. High-impedance fault detection in the distribution network using the time-frequency-based algorithm. IEEE Trans Power Delivery 2015;30(3):1260–8.
[13] Routray P, Mishra M, Rout PK. High Impedance Fault detection in radial distribution system using S-Transform and neural network. In: 2015 IEEE power, communication and information technology conference (PCITC), Bhubaneswar; 2015. p. 545–51.
[14] Gautam S, Brahma SM. Detection of high impedance fault in power distribution systems using mathematical morphology. IEEE Trans Power Syst 2013;28(2):1226–34.
[15] Du Y, Liu YD, Shao QZ, Luo LG, Dai JD, Sheng GH, et al. Single line-to-ground faulted line detection of distribution systems with resonant grounding based on feature fusion framework. IEEE Trans Power Delivery 2019;34(4):1766–75.
[16] Guo M, Zeng X, Chen D, Yang N. Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems. IEEE Sens J 2018;18(3):1291–300.
[17] Xue YD, Li J, Xu BY. Transient equivalent circuit and transient analysis of single-phase earth fault in arc suppression coil grounded system. Proc CSEE 2015;35(22):5703–14.
[18] Arvind P, Maheshwari RP. A wavelet packet transform approach for locating faults in distribution system. In: 2012 IEEE Symposium on Computers & Informatics (ISCI), Penang. p. 113–18.
[19] Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. A Review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans Syst Man Cybernet 2012;42(4):463–84.

[20] Chen TQ, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. p. 785–94.
[21] He ZY, Fu L, Lin S, Bo ZQ. Fault detection and classification in EHV transmission line based on wavelet singular entropy. IEEE Trans Power Delivery 2010;25(4):2156–63.
[22] Wang YN, Huo BL, Wang H, Wang X. A new method for earth fault line selection based on wavelet packets in small current neutral grounding system. Proc CSEE 2004;24(6):54–8.
[23] Dai JF, Zhang YX. Study on adaptively choosing fault line under singel-phase to ground fault based on analysis of multi-frequency bands. Proc CSEE 2003;23(5):44–7.
[24] Friedman J. Greedy function approximation: a gradient boosting machine. Ann Stat 2001;29(5):1189–232.
[25] Sanchez VD. Advanced support vector machines and kernel methods. Neurocomputing 2003;55:5–20.
[26] Lee CJ, Park JB, Shin JR. A new two-terminal numerical algorithm for fault location, distance protection, and arcing fault recognition. IEEE Trans Power Syst 2006;21(3):1460–2.