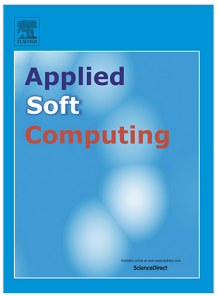
Journal Pre-proof



Faulty feeder detection based on image recognition of current waveform superposition in distribution networks

Jiawei Yuan, Zaibin Jiao

PII: S1568-4946(22)00712-8
DOI: https://doi.org/10.1016/j.asoc.2022.109663
Reference: ASOC 109663

To appear in: *Applied Soft Computing*

Received date : 1 November 2021
Revised date : 20 September 2022
Accepted date : 26 September 2022

Please cite this article as: J. Yuan and Z. Jiao, Faulty feeder detection based on image recognition of current waveform superposition in distribution networks, *Applied Soft Computing* (2022), doi: https://doi.org/10.1016/j.asoc.2022.109663.

Faulty feeder detection based on image recognition of current waveform superposition in distribution networks

Jiawei Yuan [1], Zaibin Jiao [1*]

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

[*] Corresponding author at: School of Electrical Engineering, Xi'an Jiaotong University, Xianning West Road No. 28, Xi'an 710000, China

E-mail address: jiaozaibin@mail.xjtu.edu.cn (Zaibin Jiao)

A B S T R A C T

Faulty feeder detection is essential for maintaining the security and stability of energy supply in distribution networks. However, it is rather difficult to identify a specific faulty feeder owing to small fault currents and complex fault transients. To improve the detection accuracy, this study proposes a faulty-feeder detection method based on image recognition of superimposed zero-sequence currents. A convolutional neural network (CNN) is utilized to recognize the superimposed currents in the same plot, rather than a raw single current, which can realize correlation comparisons between the currents. In addition, the zero-sequence currents of different feeders are superimposed according to a specific sequence, and the CNN can adapt to the changing topologies of distribution networks while conducting correlation comparisons. Because zero-sequence currents decay rapidly over time, an attention learning block is embedded into the CNN to enhance the discriminative capability. A total of 14,718 sets of experimental data obtained from simulations and practical distribution networks were collected to verify the effectiveness of the proposed method. Comparisons with other traditional methods and learning-based methods adopted in previous studies justify the superiority of the proposed method in terms of detection accuracy and detection efficiency. Therefore, the proposed method can be implemented in real distribution networks for faulty feeder detection.

Keywords: attention strategy, correlation comparison, waveform superposition, faulty feeder detection, topology adaptability.

1. Introduction

Minimizing the period of power interruption is essential for improving power quality and customer satisfaction in the routine operation of electrical power grids. Recent research has shown that most interruptions are caused by faults in distribution networks [1], where single line-to-ground (SLG) faults account for 80% of all faults [2]. Serious faults may occur, resulting in equipment damage, casualties, and power outages if an SLG faulty feeder cannot be reliably identified [3]. However, it is difficult to determine a specific faulty feeder owing to the small SLG fault current and complex fault transients [4]. Consequently, it is important to propose SLG faulty-feeder identification methods with high detection accuracy and reliability to enhance the stable operation of distribution networks.

Many research studies have been conducted on SLG faulty feeder identification, ranging from fault characteristic analysis to identification algorithms. For fault characteristic analysis, digital signal processing (DSP) techniques are extensively used to identify specific fault characteristics from bus voltages and feeder currents. The fast Fourier transform (FFT) [5], discrete wavelet transform [6], and S-transform [7] algorithms are the most popular DSP methods for the identification of fault characteristics, which can help in the identification criteria. In addition, mathematical morphology [8], Hilbert transform [9], and variational mode decomposition (VMD) [10] have been used in this field to obtain exact fault characteristics from weak signals. For the identification algorithm, steady-state fault characteristics, such as amplitude, polarity, power, and admittance [11], are extensively used to propose identification criteria in the early stages. Subsequently, transient components, such as differential faulty energy [12], spectral characteristics [13], and traveling waves [14], have recently become more popular for improving reliability. However, most DSP techniques adopt fixed basis functions [15], which can easily result in inadaptability of the extracted characteristics. In fact, the adaptability of the extracted characteristics is usually an inherent limitation for improving detection accuracy [16].

Recently, artificial intelligence (AI) has been introduced to propose learning-based methods in combination with multiple fault characteristics to improve identification accuracy. In [17], an integrated faulty-feeder detection method was proposed in conjunction with an adaptive network-based fuzzy inference system, and decaying direct current (DC) components, waveform similarity coefficients, and energy entropy were combined. In addition, other intelligent algorithms, such as fuzzy theory [18], ensemble learning algorithms [19], artificial neural networks [20], long short-term memory neural networks (LSTM) [21], and convolutional neural networks (CNN) [22] have been used as classifiers, which lead to a higher detection accuracy based on the fusion of multiple fault characteristics. However, the aforementioned integrated schemes have a few drawbacks. Commonly, the existing learning-based methods are expected to mine the effective fault characteristics based on the zero-sequence current of a single feeder owing to the limitation of changing topologies, which implies that they fail to conduct correlation comparisons between feeders, thus leading to poor generalization capability under complex fault scenarios. Therefore, to enhance the generalization capability, it is necessary to conduct the correlation comparison while ensuring topology adaptability.

This study proposes a faulty-feeder identification method based on image recognition of superimposed zero-sequence current waveforms. The zero-sequence currents are superimposed in the same plot to facilitate the CNN to conduct correlation comparisons between the currents. In addition, there are different combinations of current superposition to adapt to changing topologies. For a distribution system with N feeders, $N + 1$ superimposed images are generated, including one image with zero-sequence current superposition of all feeders, called completed feeder (CF) image, and N images with superimposed zero-

sequence current waveforms of N - 1 feeders, called one-feeder missing (OFM) images. Among them, the CF image is utilized to determine the fault type: bus or line faults, and the OFM images are used to select the faulty feeder under a line fault. To improve the discriminative capability of image recognition, a CNN with attention strategy is designed to distinguish images with and without faulty-feeder current superposition. To verify the reliability and generalization of the proposed method, digital simulations using power system computer-aided design (PSCAD) and real-time digital simulation (RTDS) simulators are conducted based on considerations of different topologies, grounding modes, feeder parameters, and fault conditions of the distribution networks. In addition, practical fault data in real distribution systems are employed in the test process. The results reveal that the proposed method can significantly improve the accuracy of fault-feeder detection, thereby demonstrating its strong generalization capability and good applicability.

The remainder of this paper is organized as follows. Section 2 introduces background theories related to the proposed method. In Section 3, the proposed strategy based on the image recognition of superimposed zero-sequence current waveforms is discussed in detail. The conducted case verifications are discussed in Section 4. Furthermore, the simulation models, cases, and practical fault data are presented along with some necessary analyses and comparisons. Finally, Section 5 concludes the study.

2. Background theories

2.1 Convolutional neural network

As a state-of-the-art method in image processing and recognition, the CNN is extensively used in the field of computer vision. The CNN comprises convolutional, pooling, and fully-connected (FC) layers. The convolutional and pooling layers are alternately stacked for feature extraction, and the FC layers are finally utilized for classification based on the extracted features.

The core of the CNN is the convolutional layer, which imitates the visual perception mechanism using a series of filters with receptive fields. Multiple filters are used to capture the notion of left/right and up/down of the feature maps generated by the previous layers. The convolution process can be described as

$$x_j^l = f\left(\sum_t k_{t,j}^l * x_t^{l-1} + b_j^l\right), \tag{1}$$

where x_j^l denotes the output of j^{th} filter in the l^{th} convolutional layer, $x_t^{l\text{-}1}$ denotes the output of t^{th} feature map in the $(l\text{-}1)^{\text{th}}$ convolutional layer, $k_{t,j}^l$ denotes the learned weights of j^{th} filter in the l^{th} convolutional layer, and b_j^l is the bias term. f (.) denotes the non-linear activation function.

Subsequently, the pooling layer is utilized to reduce the learning parameters and avoid over-fitting. There are three types of pooling methods: average pooling, max pooling, and norm pooling. The max pooling method is the most common and it can be expressed as

$$p_j^l = \max_{r \in M_j}\left(x_j^l(r)\right), \tag{2}$$

where p_j^l denotes the output of j^{th} filter in the l^{th} pooling layer and M_j is the area of the pooling operation.

After stacking several convolutional and pooling layers, the learned features of each filter are flattened and input to the FC layer for classification. The Softmax function is commonly applied here, and its output can be expressed as

$$P(y = i \mid z) = \frac{exp(z_i)}{\sum_i exp(z_i)}, \tag{3}$$

where z is the input of the FC layer and $P(y = i \mid z)$ is the probability value of the i^{th} category.

2.2 Faulty feeder detection based on superimposed waveforms

According to the transition impedances, there are four different types of SLG faults that occur in distribution networks [23], as shown in Fig. 1. These include low impedance and under-damped (LIU) faults, low impedance and over-damped (LIO) faults, high impedance and under-damped (HIU) faults, and high impedance and over-damped (HIO) faults. Typical zero-sequence currents when an LIU fault occurs are illustrated in Fig. 2. It can be seen that the zero-sequence current of the faulty feeder has the largest magnitude, and it flows in the direction opposite to that of the healthy feeder.

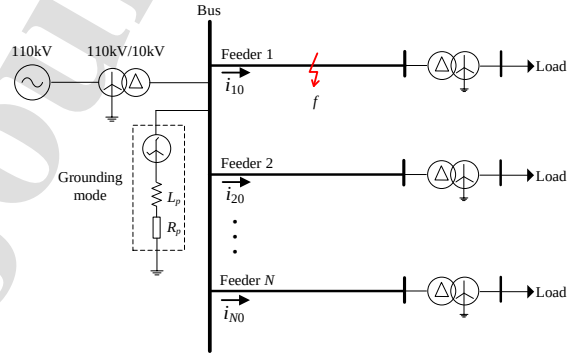


Fig. 1. Non-effectively grounded distribution network.

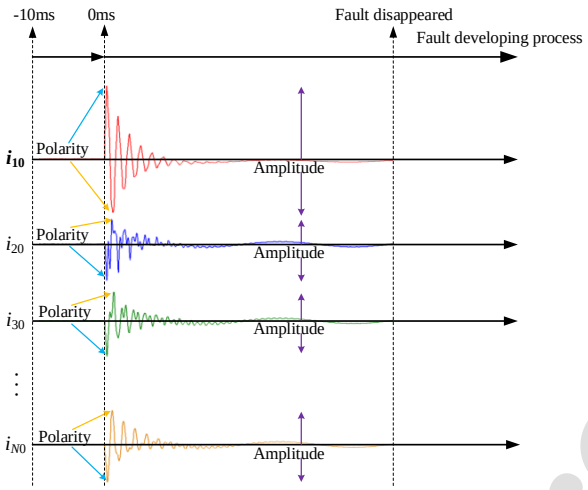


Fig. 2. Zero-sequence currents when an LIU fault occurs in feeder 1.

Because the topologies, parameters, and fault scenarios vary under different distribution networks, the zero-sequence currents of the faulty feeder exhibit changing fault characteristics under various fault scenarios. This is the same as the zero-sequence currents of healthy feeders. However, when zero-sequence currents are superimposed in a specific sequence, the differences between the superimposed waveforms with and without the zero-sequence current of the faulty feeder are always obvious. In addition, the superposition form ensures adaptability to changing topologies.

For a distribution system with N feeders, as shown in Fig. 1, $N + 1$ superimposed images are created, as shown in Fig. 3. Among them, one CF image is generated by superimposing the zero-sequence current waveforms of all feeders, and N OFM images are generated by superimposing the zero-sequence current waveforms of $N - 1$ feeders. The OFM images can be transformed from the CF image by sequentially removing the waveform of each feeder. As shown in Fig. 3, the superimposed images with the faulty-feeder current, such as the CF, OFM 2, and OFM 3, are similar to each other, but they have large differences with the OFM 1 image that is superimposed without the faulty-feeder current. Notably, the superimposed images without the faulty-feeder current can be distinguished from those with the faulty-feeder current owing to the differences between the faulty-feeder and healthy-feeder currents, and they can be captured by conducting correlation comparisons between the currents, which are not limited to specific fault features.

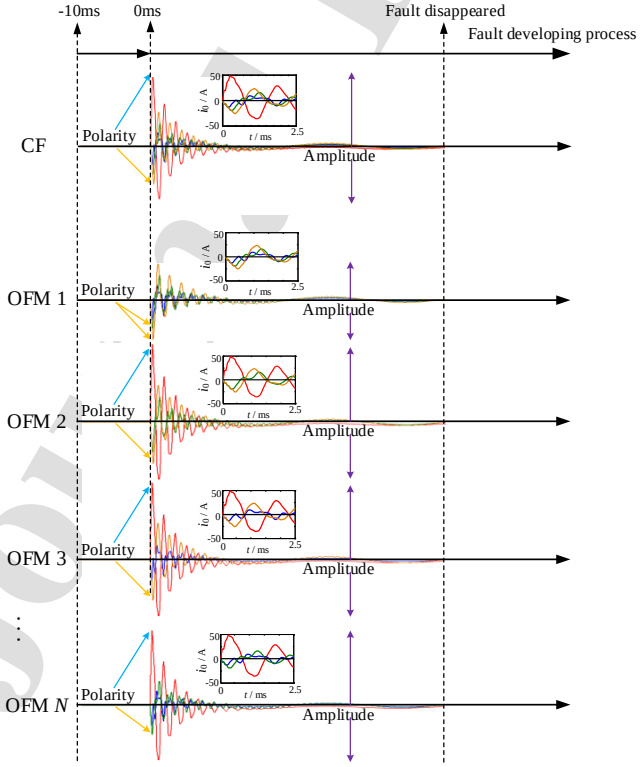


Fig. 3. Superimposed images generated in sequence.

The OFM images of the zero-sequence currents for SLG faults under different fault conditions are shown in Fig. 4. Evidently, the OFM 1 image without the faulty-feeder current can be easily identified from the other OFM images through a correlation comparison between the superimposed currents. Owing to limited space, the comparison between two OFM images are shown in Fig. 4.

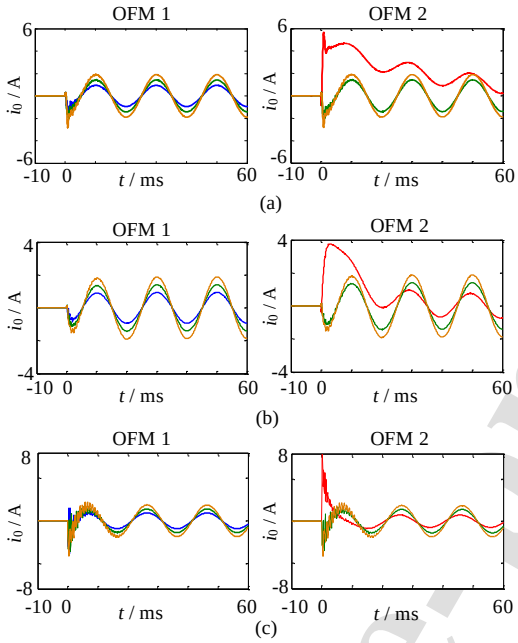


Fig. 4. OFM images when SLG faults occur in feeder 1 under different fault conditions: (a) LIO fault; (b) HIO fault; and (c) HIU fault.

For feeder faults, only one image among the N OFM images does not include the faulty feeder current. The feeder corresponding to the missing current waveform in the OFM image is the faulty feeder.

Furthermore, for bus faults, no fault current waveform is included in the superimposed CF image, whereas the fault current can be found in the CF image during feeder fault conditions.

In summary, the fact that superimposed images with and without the faulty-feeder current are different can be used as an identification criterion to detect faulty feeders in distribution networks. Among them, the correlation comparison between currents can be conducted in the generated superimposed images, and topology adaptability can also be realized by recognizing the $N + 1$ images. The correlation comparison based on the superposition form can always work under different fault scenarios, and a detailed detection scheme is proposed in the next section.

3. Methodology

3.1 Superimposed image creation

There are three modules for generating the images: sampling, start-up, and image creation modules. The main responsibility of the sampling module is to sample the raw zero-sequence voltage and currents. The start-up module is responsible for judging whether an SLG fault occurs or not, and then the image creation module will be implemented to obtain the images for recognition. The entire image creation procedure is illustrated in Fig. 5.

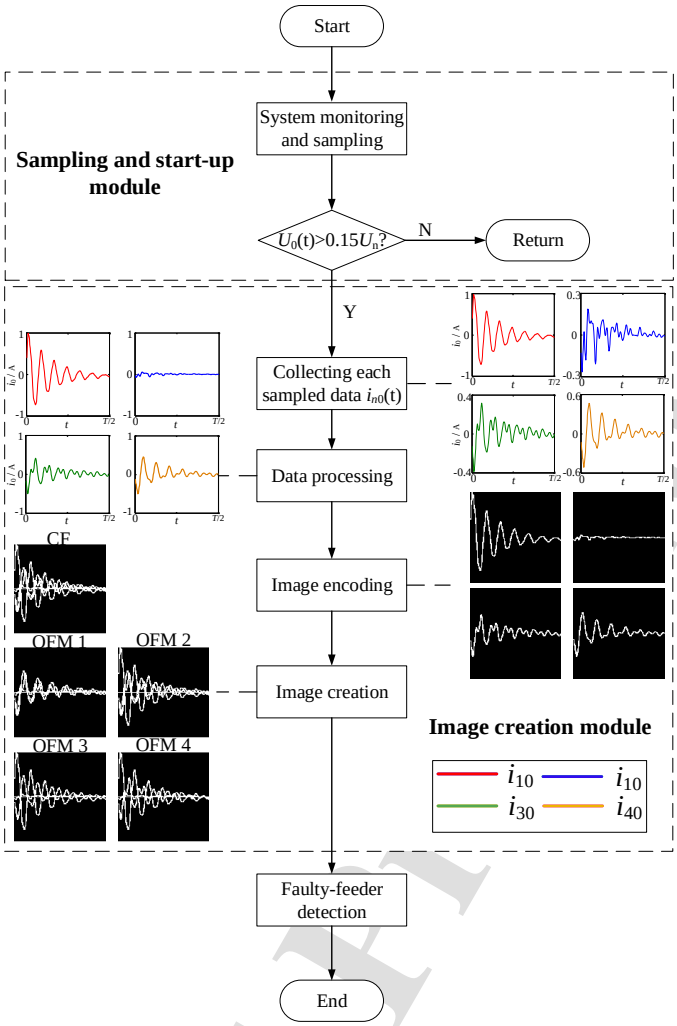


Fig. 5. Image creation (Consider the zero-sequence current waveforms in a distribution network composed of four feeders as an example).

The analog zero-sequence voltage and currents are obtained using a potential transformer (PT) and current transformers (CT), respectively. Subsequently, the analog signals are converted into digital signals after sampling by the protection devices. The devices further determine whether an SLG fault has occurred. The root-mean-square (RMS) value of the sampled voltage is calculated using FFT and compared with the amplitude of the rated voltage U_n. The SLG faults are identified if the calculated RMS value is greater than $0.15U_n$; otherwise, the devices would not perform further detection algorithms.

When an SLG fault occurs, the sampled first half-cycle zero-sequence current of each feeder is collected and the maximum absolute values ($i'_M(n)$) of the currents are obtained using (4).

$$i'_M(n) = Max(|i_{n0}(t)|), \quad n \in [1, N], t \in [0, \frac{T}{2}], \tag{4}$$

where T is the cycle of the zero-sequence current, N is the number of feeders, and i_{n0} is the sampled value of the zero-sequence current of feeder n.

Given that the zero-sequence current waveforms of the feeders are superimposed proportionally, it is necessary to conduct normalization as given below:

$$i'_n(t) = i_{n0}(t) \times \frac{i'_M(n)}{(Max(i'_M(n)))^2}, \quad n \in [1, N], t \in [0, \frac{T}{2}], \tag{5}$$

To conduct a correlation comparison between the currents, the normalized zero-sequence currents are subsequently superimposed in the same plot. However, the correlation comparison fails when most normalized zero-sequence currents are close to zero, which may result from a zero-sequence current with a much larger amplitude than the other currents. Therefore, it is necessary to enhance several zero-sequence currents by increasing their normalized amplitudes.

We can use η, as expressed in (6), to describe the amplitude relationship of zero-sequence currents.

$$\eta_n = \frac{i'_M(n)}{Max(i'_M(n))}, \quad n \in [1, N], \tag{6}$$

If η_n with the second largest or third largest amplitude is smaller than 0.5, the amplitude of the corresponding zero-sequence current will be increased by $0.5/\eta_n$; otherwise, it will remain at the raw amplitude.

After data processing of the zero-sequence currents, the processed currents are then used for image creation. To accelerate the image creation process, a novel image creation method is proposed as follows:

When a processed zero-sequence current in numerical form is converted into an image, the horizontal coordinate in the image denotes the sampling time and the vertical coordinate denotes its amplitude. Suppose the required image size is $H \times W$, the horizontal coordinate is divided into H sections and the vertical coordinate is divided into W sections, as shown in Fig. 6. Subsequently, image creation can be simplified by determining the sections through which the current waveform passes. For a section $e_{i,j}$, the encoding value is 1 if the current waveform passes through the section, which can be easily determined based on its amplitude; otherwise, the encoding value is 0. Finally, each section ($e_{1,1}$–$e_{H,W}$) is encoded as 0 or 1, and an image of size $H \times W$ can be generated. In this study, both H and W are set as 128.

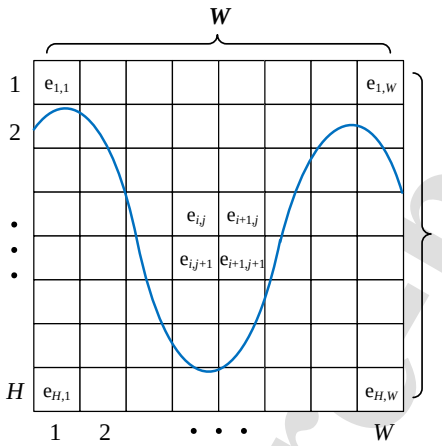


Fig. 6. Encoding for a zero-sequence current.

To obtain the superimposed image generated by different zero-sequence currents, the final encoding value in each section can be obtained by taking the union of each current value in the same section. Furthermore, the 0^{th}, 1^{st}, and -1^{st} axes are added to the image, which implies that the values of sections ($e_{1,1}$–$e_{1,W}$), ($e_{H/2,1}$–$e_{H/2,W}$), and ($e_{H,1}$–$e_{H,W}$) are each equal to one. Compared with the image generated using 'plot' function in Python, the proposed image creation method can be processed quickly. In addition, the proposed method does not require additional image cropping or binary operations, thus significantly improving the efficiency of image creation.

3.2 Proposed Image-recognition-based Framework

After the image creation as explained above, the generated $N + 1$ images are recognized using a CNN to determine whether the faulty-feeder current is superimposed on the images. The CF image is used to determine the fault type (bus or feeder), and the N OFM images are employed for faulty feeder detection. The entire procedure of the proposed image-recognition-based detection scheme is summarized in Algorithm 1.

Algorithm 1 Single-line-to-ground (SLG) fault detection scheme

Step 1: Collect the sampled values of the first half-cycle zero-sequence current of each feeder.

Step 2: Generate one completed feeder (CF) image, input the image to the convolutional neural network (CNN), then infer the following:

If the output of CNN is zero, return results (the fault type is bus fault) and terminate; otherwise go to ***Step 3***

Step 3: Generate N OFM images, input N images to CNN in sequence, and then infer the following:

For sequence in [1, N]:

If the output of CNN is 0, return the sequence number.

end for

The proposed detection scheme is performed in two steps: fault type identification and faulty feeder detection. First, the CF image is input to the CNN, and the fault type is determined according to the recognition result. For a feeder fault, the faulty-feeder current is superimposed on the image. Consequently, the fault type is the line fault. Subsequently, the N OFM images are input to the same CNN in sequence, and the outputs of the CNN are determined. Finally, the OFM image whose CNN output is zero is found, and the feeder that is not superimposed on the image is selected as the faulty feeder.

Regarding the bus fault, the faulty-feeder current is not superimposed on the CF image according to the recognition result from the CNN. Consequently, the fault type is directly judged as a bus fault.

3.3 Attention-based CNN Design

For image recognition using CNNs, it is necessary to emphasize local features of interest and weaken the influence of other objects and backgrounds [22]. In particular, the discriminative features of the zero-sequence currents may concentrate on the transient process in the initial fault time and decay rapidly in a short time, which implies that CNNs need to focus on the representative features and suppress the unnecessary features of the currents. Typical first half-cycle zero-sequence currents after the occurrence of SLG faults are shown in Fig. 7.

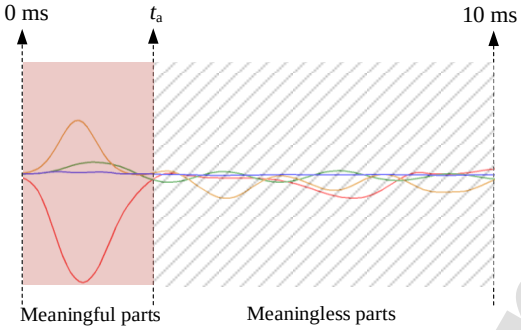


Fig.7. Image recognition for the superimposed waveforms.

Evidently, the zero-sequence currents decayed sharply over time. In addition, it is rather difficult for CNNs to conduct a correlation comparison after the currents decay close to 0^{th} axis. As shown in Fig. 7, it is difficult to conduct image recognition in the time range of [t_a, 10], which implies that CNNs would fail if they focus on the characteristics in this area. Therefore, to conduct image recognition effectively, CNNs must pay more attention to the meaningful parts [0, t_a] in the image.

Traditional CNNs cannot effectively focus on the meaningful features of images owing to the training dataset having few samples and low quality. Furthermore, it may focus on unnecessary features that can easily result in overfitting and unsatisfactory performance in practical applications. Thus, it is the key to improving the recognition accuracy by inserting a particular architecture called 'attention' into the CNNs.

To improve the representation of interests on the meaningful features of images, a particular architecture design called 'attention' [24, 25] has been proven to enable CNNs to focus on meaningful features and suppress unnecessary features. Traditional CNNs equipped with attention modules, called convolutional block attention modules (CBAM) [25], have achieved state-of-the-art performance in many image recognition tasks. The CBAM comprises two modules: channel attention module and spatial attention module, which can adaptively learn 'what' and 'where' to focus in the channel and spatial aspects. Given an intermediate feature map $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$ as input, the channel attention module performs pooling operations, including average pooling and max-pooling, on the spatial area of the feature map. Subsequently, they are passed to a shared network and merged using element-wise summation, as shown in (7).

$$\mathbf{M}_c(\mathbf{F}) = \sigma(W_2\delta(W_1(\mathbf{F}^c_{avg})) + W_2\delta(W_1(\mathbf{F}^c_{max}))), \tag{7}$$

where $\mathbf{F}^c_{avg}$ and $\mathbf{F}^c_{max}$ denote the features in the channel axes after average pooling and max-pooling operations, respectively, and W_1 and W_2 are the learned weights of the FC layers. δ denotes the rectifier linear unit (ReLU) function, σ denotes the sigmoid function, and $\mathbf{M}_c(\mathbf{F}) \in \mathbb{R}^{C\times 1\times 1}$ is the final channel attention map.

For the spatial attention module, pooling operations are performed on the channel area of the feature map, and subsequently, they are concatenated and convolved using a convolution layer, as shown in (8).

$$\mathbf{M}_s(\mathbf{F}) = \sigma(f[\mathbf{F}^s_{avg} \oplus \mathbf{F}^s_{max}]), \tag{8}$$

where $\mathbf{F}^s_{avg}$ and $\mathbf{F}^s_{max}$ denotes the features in the spatial axes after average pooling and max-pooling operations, respectively, $\oplus$ is the concatenation operator, f denotes the convolution operation, σ denotes the sigmoid function, and $\mathbf{M}_s(\mathbf{F}) \in \mathbb{R}^{1\times H\times W}$ is the final spatial attention map.

Typically, a channel attention module and spatial attention module are sequentially stacked to form the final CBAM, as shown in (9).

$$\mathbf{F}' = (\mathbf{F} \otimes \mathbf{M}_c(\mathbf{F})) \otimes \mathbf{M}_s(\mathbf{F} \otimes \mathbf{M}_c(\mathbf{F})), \tag{9}$$

where $\otimes$ denotes element-wise multiplication and $\mathbf{F}'$ is the output of the CBAM.

Finally, a five-layer CNN composed of convolutional layers, CBAMs, a global average pooling (GAP) layer, and two-way FC classification layer is established, as shown in Fig. 8.

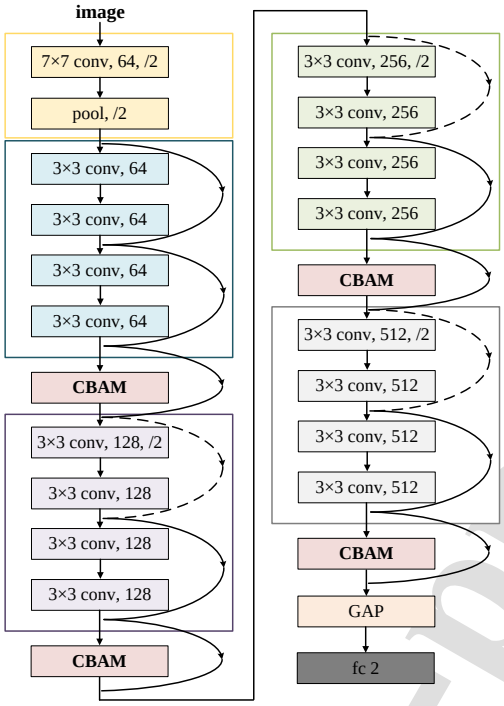


Fig. 8. Structure of attention-based CNN.

In summary, the attention-based CNN can focus on meaningful fault features in the images generated by waveform superposition, and the proposed detection criterion based on image recognition can reliably detect the faulty feeder.

3.4 CNN Training

A large dataset should be used to train the attention-based CNN to ensure that the proposed method performs as expected. Considering that it is difficult to obtain sufficient practical fault data, simulation data were utilized to train the established model in this study. Three training models with different grounding modes and feeder types were established using PSCAD to simulate the SLG faults that occur in feeders and buses. The structures of the simulation models are shown in Fig. 1. The parameters concerning the overhead lines are as follows: $R_1 = 0.17$ Ω/km, $L_1 = 1.21$ mH/km, $C_1 = 0.0097$ μF/km, $R_0 = 0.23$ Ω/km, $L_0 = 5.48$ mH/km, and $C_0 = 0.006$ μF/km. Similarly, the parameters of the cable lines include $R_1 = 0.098$ Ω/km, $L_1 = 0.274$ mH/km, $C_1 = 0.351$ μF/km, $R_0 = 0.246$ Ω/km, $L_0 = 0.955$ mH/km, and $C_0 = 0.166$ μF/km. Furthermore, the sampling frequency is 20 kHz and the current transformer ratio is 50:1. In addition, fault locations, fault time, and fault impedance are considered in the data-production processes, and the detailed fault scenarios are summarized in Table 1.

Table 1
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%	/
	Feeder type	Overhead	Cable / Overhead	Overhead
	Length (km)	10 / 20 / 30 / 40	Cable:10 / 20, Overhead: 30 / 40	10 / 20 / 30 / 40
Scenarios	Fault location	10% / 50% / 90% / Bus		
	Fault inception angle	0°~345.6° per 21.6°(Line fault) / 0°~351° per 0.702°(Bus fault)		
	Grounding resistances	20 Ω / 100 Ω / 500 Ω / 1000 Ω		
	Noise	50 dB		
Sample Number	Simulation data	816 (Line fault) / 2,004 (Bus fault)	816 (Line fault) / 2,004 (Bus fault)	816 (Line fault) / 2,004 (Bus fault)
	Image	8,460 (Label 0)		9,792 (Label 1)
	Image (Data augmentation)	16,920 (Label 0)		19,584 (Label 1)

It should be noted that the arcing effect is not considered in the training dataset, whereas arc fault data from practical fault data and RTDS test system are used in the test process. This is mainly because of the following reasons: (1) there are large differences between the arcs obtained by PSCAD simulations and practical real faults, and (2) it is impossible for the training dataset to include all possible fault scenarios. Therefore, the generalization capability and robustness of the proposed method

can be demonstrated.

When an SLG fault occurs in a feeder in a distribution network with four feeders, the label for the CF image is 1, whereas the labels for the OFM images depend on whether the faulty feeder current waveform is superimposed. If the faulty feeder waveform is superimposed, the label is 1; otherwise, it is 0. Consequently, five images are obtained in this fault scenario, in which the labels for the four images are 1 and the label for the remaining image is 0.

When an SLG fault occurs in a bus in a distribution network, only one CF image is generated and its label is 0.

The training dataset comprises 2,448 sets of fault data with line faults and 6,012 sets of fault data with bus faults, among which the labels for 9,792 (2,448 × 4) images are 1, and the labels for 8,460 (6,012 + 2,448) images are 0. To expand the training dataset and avoid overfitting of CNN, the aforementioned images are flipped vertically. Thus, 36,504 (9,792 × 2 + 8,460× 2) images are obtained, where 90% of the images are randomly selected for training data and the rest are used for validation data.

The learning environment is implemented on a personal computer with an Intel® Core™ processor i7-6700 (CPU), 3.40 GHz, 16.00 GB RAM, and NVIDIA GeForce GTX 1050. Furthermore, Python (version 3.7) and Pytorch (version 1.11) in a Jupyter notebook are applied as the software implementation. The Adam optimizer is used in the training process, and the batch size is set to 64. The learning rate is initialized to 0.001 and drops by 80% after every 20 epochs. During 100 epochs of training, the model with minimum loss in the validation data is selected as the final model. Finally, the trained CNN model achieved 100% detection accuracy for both the training and validation data.

4. Results and discussion

To verify the generalization of the proposed method, 14,637 sets of data and 69 sets of recorded data, generated from different distribution networks in the PSCAD simulations and RTDS hardware-in-the-loop (HIL) test system, respectively, are employed. The PSCAD simulation considers common fault conditions, and RTDS test system includes the simulated arc grounding faults and actual measurement errors. The topologies, parameters, feeder types, and fault conditions in these distribution networks are completely different from those in the training dataset. Moreover, to verify practical applications, 12 sets of practical fault data in real distribution systems are collected during the test process, and the collected zero-sequence currents have huge intermittence, randomness, and asymmetry due to the real arc grounding events.

In total, 14,718 sets of fault data are used as the test dataset. The test dataset considers more complex fault scenarios, particularly for the fault data collected from the RTDS test system and real distribution systems. It should be noted that the test dataset is used only for verifying the detection performance of the trained model, and is unavailable during the training process.

4.1 Simulation-based Test

Five test models were designed for the PSCAD simulations, as summarized in Table 2. The feeder parameters of the overhead lines are as follows: $R_1 = 0.33$ Ω/km, $L_1 = 1.31$ mH/km, $C_1 = 0.007$ μF/km, $R_0 = 1.041$ Ω/km, $L_0 = 3.96$ mH/km, and $C_0 = 0.004$ μF/km. The parameters of the cable lines are as follows: $R_1 = 0.0791$ Ω/km, $L_1 = 0.264$ mH/km, $C_1 = 0.373$ μF/km, $R_0 = 0.227$ Ω/km, $L_0 = 0.926$ mH/km, and $C_0 = 0.166$ μF/km.

Table 2
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
	Compensated degree	/	6%	/	10%	10%
	Feeder numbers	6		3		5
	Length (km)	5/ 15 / 25 / 35 / 45 / 50		5 / 20 / 80		10 / 15 / 25 / 35 / 45
	Feeder type	Overhead	Overhead	Overhead	Overhead	Cable / hybrid/ Overhead
	Cable ratio	/	/	/	/	100% / 20% / 60% / 0% / 0%
Scenarios	Fault location	5% / 80% / Bus		5% / 40% / Bus		5% / 60% / Bus
	Initial phases (θ_f)	0°~360° per 9°				
	Resistances (R_g)	10 Ω / 80 Ω / 350 Ω / 780 Ω / 1350 Ω / 1700 Ω / 2000 Ω				
	Noise	30 dB				
Sample number		3,444 (Line fault)/ 287 (Bus fault)	3,444 (Line fault) / 287 (Bus fault)	1,722 (Line fault) / 287 (Bus fault)	1,722 (Line fault) / 287 (Bus fault)	2,870 (Line fault) / 287 (Bus fault)

Compared with the training dataset, the electromagnetic environment is more complex, and Gaussian white noise with a signal-to-noise-ratio of 30 dB was added to the zero-sequence current. The feeder lengths varied considerably, especially in test models 3 and 4; feeder types were more diverse, and hybrid lines (overhead lines + cable lines) were considered in test model 5. Furthermore, the fault conditions were different, where fault impedance increased and the fault time varied. These various fault scenarios are helpful for comprehensively verifying the generalization capabilities of the proposed method.

Typical zero-sequence currents and the corresponding class activation maps (CAM) are shown in Fig. 9, where the red color in the CAMs represents a greater weight of interest and the blue color indicates a smaller weight. Evidently, there are varied fault

characteristics of zero-sequence currents under different fault scenarios, thus leading to changing oscillation and attenuation characteristics. Notably, the zero-sequence currents would decay rapidly, particularly under test models 2, 4, and 5 (compensation system). In addition, the healthy-feeder current would have large zero-sequence currents when SLG faults occur in test models 3 and 4 with extreme topologies, which makes it difficult to identify the faulty feeder. Despite these changing characteristics, the proposed attention-based CNN can always focus on the discriminative parts of superimposed currents, which correspond to the differences between the faulty and healthy-feeder currents. Therefore, the superposition form of zero-sequence currents helps to distinguish the faulty feeder from the healthy feeder, and the proposed CNN has a strong discriminative capability, thereby further improving the detection accuracy.

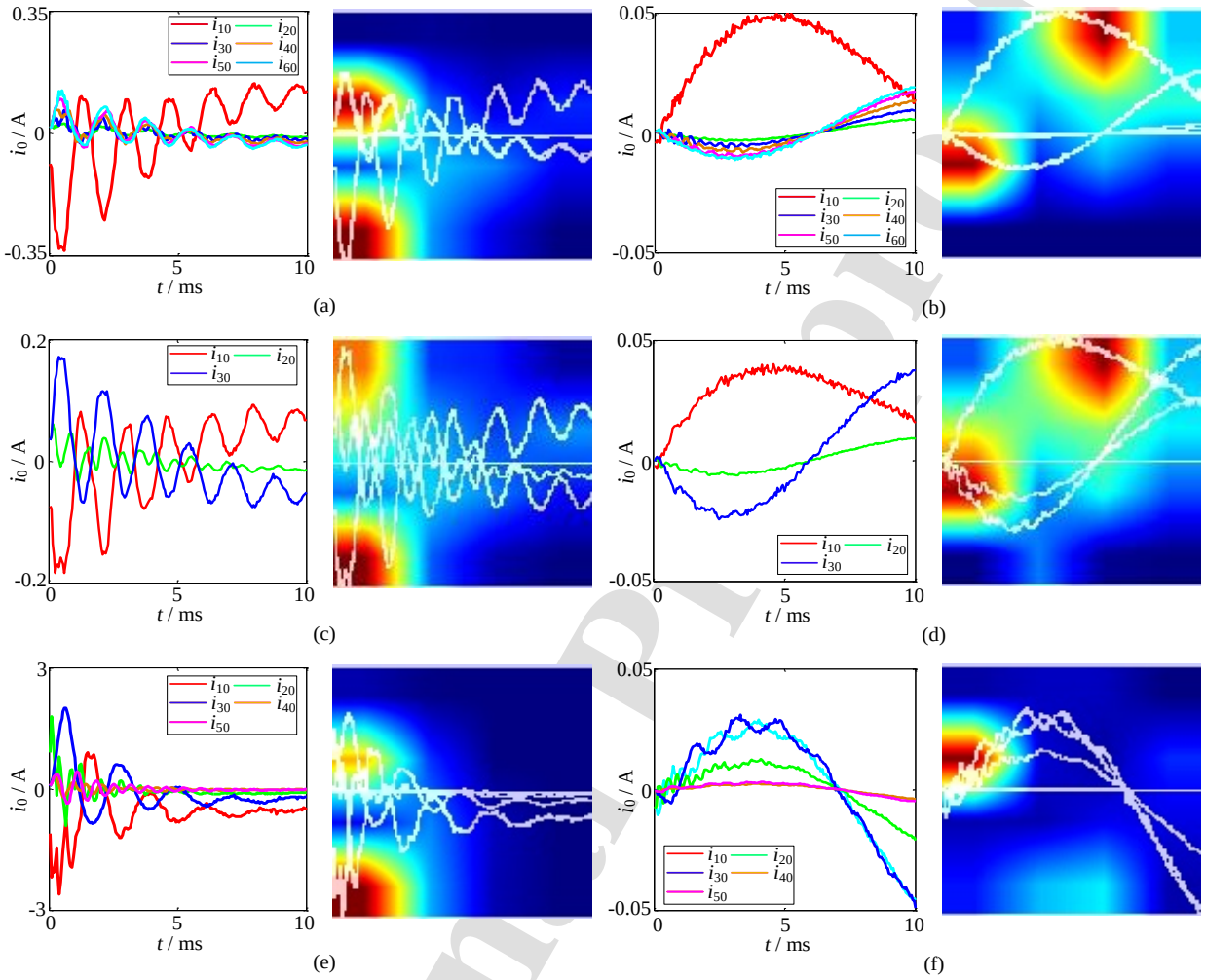


Fig. 9. Zero-sequence currents and the corresponding CAMs in PSCAD simulation: (a) R_g = 10 Ω, θ_f = 225°, fault in feeder 1, test model 1; (b) R_g = 2000 Ω, θ_f = 0°, fault in feeder 1, test model 2; (c) R_g = 10 Ω, θ_f = 225°, fault in feeder 1, test model 3; (d) R_g = 2000 Ω, θ_f = 0°, fault in feeder 1, test model 4; (e) R_g = 10 Ω, θ_f = 225°, fault in feeder 1, test model 5; and (f) R_g = 1700 Ω, θ_f = 171°, fault in bus, test model 5.

Furthermore, three other detection methods are used for comparison. Among them, [26] extracted the detailed characteristics with specific physical meanings and constructed a detection criterion based on multi-feature fusion by using a multiple evidence estimation method. Similarly, [27] extracted intrinsic mode functions (IMF) from zero-sequence currents using VMD, which corresponded to the decaying DC, power-frequency, and high-frequency components, and the extracted components at different times were used as the input of LSTM. Furthermore, [28] utilized a CNN to recognize the time-frequency images generated using a continuous wavelet transform (CWT), and did not consider the correlation between different feeders. The detailed detection performance of the four methods is summarized in Table 3, and the symbol '×' means that the method is not applicable under the corresponding conditions.

Table 3
Detection performance in PSCAD simulations.

Detection	Model	Fault type	Paper [26]	Paper [27]	Paper [28]	Proposed method
	Test model 1	Line fault	0.028 s	0.75 s	0.096 s	0.033 s

Average detection time for an SLG fault		Bus fault	×	0.80 s	0.098 s	0.012 s
	Test model 2	Line fault	0.027 s	0.75 s	0.096 s	0.033 s
		Bus fault	×	0.82 s	0.095 s	0.013 s
	Test model 3	Line fault	0.014 s	0.60 s	0.047 s	0.018 s
		Bus fault	×	0.60 s	0.047 s	0.008 s
	Test model 4	Line fault	0.014 s	0.61 s	0.050 s	0.019 s
		Bus fault	×	0.62 s	0.047 s	0.008 s
	Test model 5	Line fault	0.023 s	0.64 s	0.080 s	0.028 s
		Bus fault	×	0.70 s	0.078 s	0.011 s
Detection accuracy	Test model 1	Line fault	100%	77.58%	97.97%	100%
		Bus fault	×	42.51%	83.62%	100%
	Test model 2	Line fault	100%	98.61%	99.91%	100%
		Bus fault	×	72.47%	84.32%	100%
	Test model 3	Line fault	100%	71.49%	64.81%	100%
		Bus fault	×	8.71%	76.31%	98.26%
	Test model 4	Line fault	80.43%	94.25%	77.87%	99.89%
		Bus fault	×	40.77%	77.00%	98.61%
	Test model 5	Line fault	87.87%	98.99%	90.84%	100%
		Bus fault	×	95.82%	69.04%	98.61%

The method in [26] needs 22.71 ms + 20 ms on average to detect an SLG fault, where 20 ms is the time window of the sampled data. For the method in [27], because the feature extraction using VMD is quite time-consuming, it requires 689.13 ms + 40 ms to perform faulty feeder detection. Furthermore, the method in [28] requires 79.53 ms + 10 ms, where 95.87% of the detection time is consumed for CWT and image creation. However, the proposed method takes just 26.43 ms (18.11 ms for image creation and 8.32 ms for image analysis) + 10 ms to identify the faulty feeder owing to the proposed image creation method. Moreover, the detection accuracy in [26] decreases sharply under test models 4 and 5, and it cannot detect bus faults. Owing to the small number of decaying DC components in ungrounded networks, the method in [27] has poor detection performance, especially under test models 1 and 3. Because the method in [28] does not consider the correlation between different feeders, the extreme topologies of the distribution network (test models 3 and 4) would significantly affect the detection performance. In contrast, the proposed method outperforms the other three methods under different fault scenarios, and has a strong generalization capability and higher faulty-feeder detection accuracy.

Furthermore, two RTDS models with different grounding modes were established in a HIL test system, as shown in Fig. 10. The sampled data for the zero-sequence currents were collected using a fault recorder. The different fault scenarios are listed in Table 4.

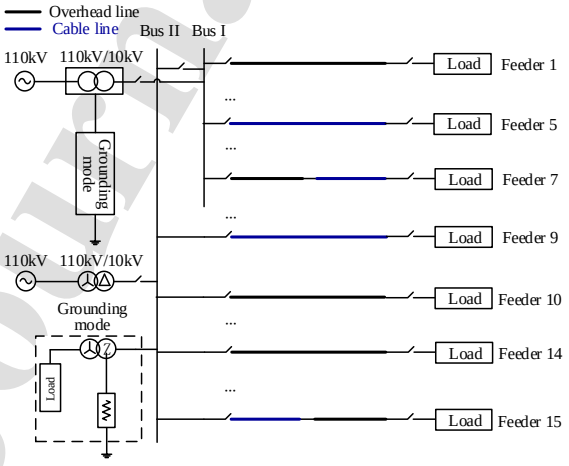


Fig. 10. RTDS model of a 10 kV distribution system.

Table 4
Fault scenarios in RTDS simulations.

RTDS Simulation	Test Samples

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
	Initial phases (θ_f)	30° / 90°/ 120°/ 123°/ 150°/ 180°/ 210°/ 270°/ 330°	30°/ 60°/ 90°/ 123°/ 150°/ 210°/ 243°/ 270°/ 330°
	Grounding resistances (R_g)	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

Considering practical fault conditions, such as unbalanced voltages and arc grounding events, as summarized in Table 4, the performance of faulty-feeder detection methods in the RTDS HIL test system is more convincing. The zero-sequence currents when an SLG fault with arc grounding occurs are shown in Fig. 11, and the waveform distortions are noticeable.

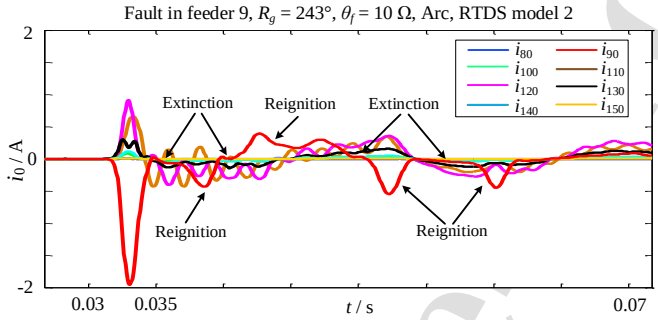


Fig. 11. Zero-sequence current when an SLG fault with arc grounding occurs in the RTDS HIL test system.

In the RTDS HIL test system, the sampling frequency is 12 kHz, which is different from the sampling frequency (20 kHz) in the training dataset. In fact, compared with fault data in numerical form, the size of the images is immune from the changing sampling frequency, which implies that the trained CNN can be directly applied for faulty feeder detection. Several zero-sequence currents and their corresponding CAMs in the RTDS HIL test are shown in Fig. 12. Notably, the zero-sequence currents would have distortions under arc grounding faults, as shown in Fig. 12 (a) and 12 (f), and there are complex transients in the initial fault time. However, the trained CNN can still focus on meaningful parts in images and conduct correlation comparisons accordingly.

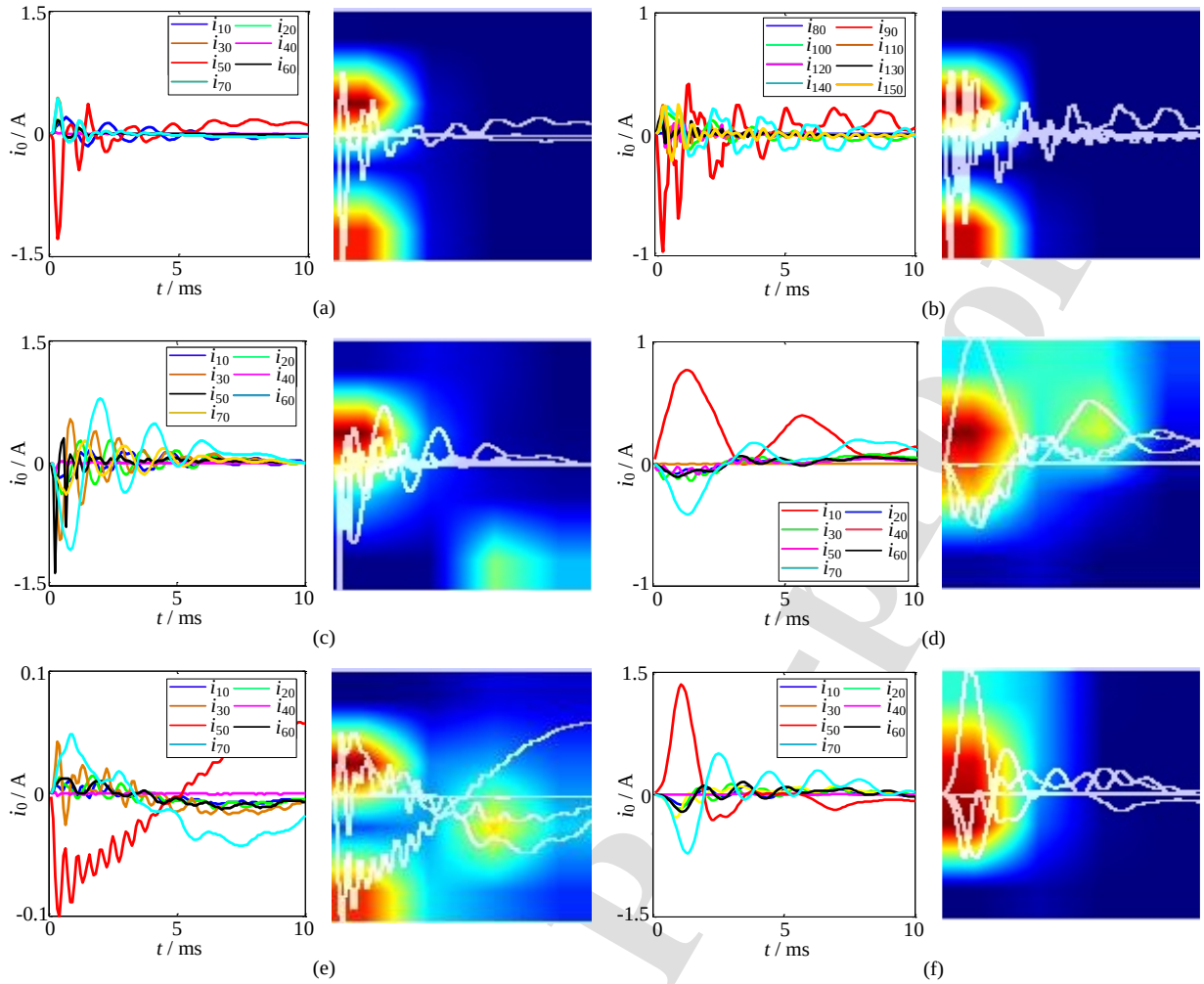


Fig. 12. Zero-sequence currents and their corresponding CAMs in RTDS HIL test: (a) R_g = 10 Ω, θ_f = 210°, fault in feeder 5, arc grounding, RTDS model 1; (b) R_g = 5.82 Ω, θ_f = 210°, fault in feeder 9, RTDS model 1; (c) R_g = 10 Ω, θ_f = 90°, fault in bus, RTDS model 1; (d) R_g = 10 Ω, θ_f = 30°, fault in feeder 1, RTDS model 2; (e) R_g = 1000 Ω, θ_f = 150°, fault in feeder 5, RTDS model 2; and (f) R_g = 10 Ω, θ_f = 123°, fault in feeder 5, arc grounding, RTDS model 2.

The detailed detection results for the four methods are summarized in Table 5. It can be seen that the other three methods would have misjudgments under certain fault scenarios. For instance, [26] may misjudge bus faults as line faults, and [27] has poor detection performance when SLG faults with arc grounding faults occur in practical distribution networks. Because the time-frequency images obtained using CWT are affected by sampling frequency, [28] would inevitably have some misjudgments in the RTDS HIL test under the new sampling frequency (12 kHz). However, the proposed method can always detect the faulty feeder correctly, and it can reliably distinguish bus faults from line faults, demonstrating its strong generalization capability and extensive application prospects.

Table 5
Detection accuracy in RTDS simulations.

Model	Paper [26]	Paper [27]	Paper [28]	Proposed method
RTDS model 1	93.55%	83.87%	83.87%	100%
RTDS model 2	94.74%	100%	68.42%	100%

To comprehensively compare the performance in different aspects, a detailed comparison of the proposed method and the other three methods is shown in Table 6, including whether training is needed or not, adaptability to different fault types, detection efficiency, and detection accuracy. Evidently, the proposed method is a learning-based method, and it needs to be pre-trained before application, which is its main disadvantage when compared with traditional analysis-based methods [26]. Furthermore, although the proposed encoding method can significantly improve the efficiency of image creation, it may need graphics processing unit (GPU) computing for image analysis in practical applications, which would require hardware updating

in devices. Despite these two limiting factors, the proposed method has high adaptability and detection accuracy under different fault scenarios, and has good application prospects in practical distribution networks.

Table 6
Detailed comparison of different methods.

Method	Training	Adaptability	Detection efficiency (GPU)	Detection efficiency (CPU)	Detection accuracy
Paper [26]	No	Medium	High	High	Medium
Paper [27]	Yes	High	Low	Low	Low
Paper [28]	Yes	High	Medium	Medium	Medium
Proposed method	Yes	High	High	Medium	High

4.2 Practical-fault-data-based Test

To verify the application prospects of the proposed method, we collected 12 sets of practical SLG fault data from the Chenzhou, Shaoyang, Xiangxi, and Xi'an substations in 2020. Additional fault data were recorded using real fault recorders sampled at 12 kHz. Among them, most SLG faults occurred with obvious arc grounding events, especially for the SLG faults in the Xi'an substation, and the currents and their corresponding CAMs are shown in Fig. 13.

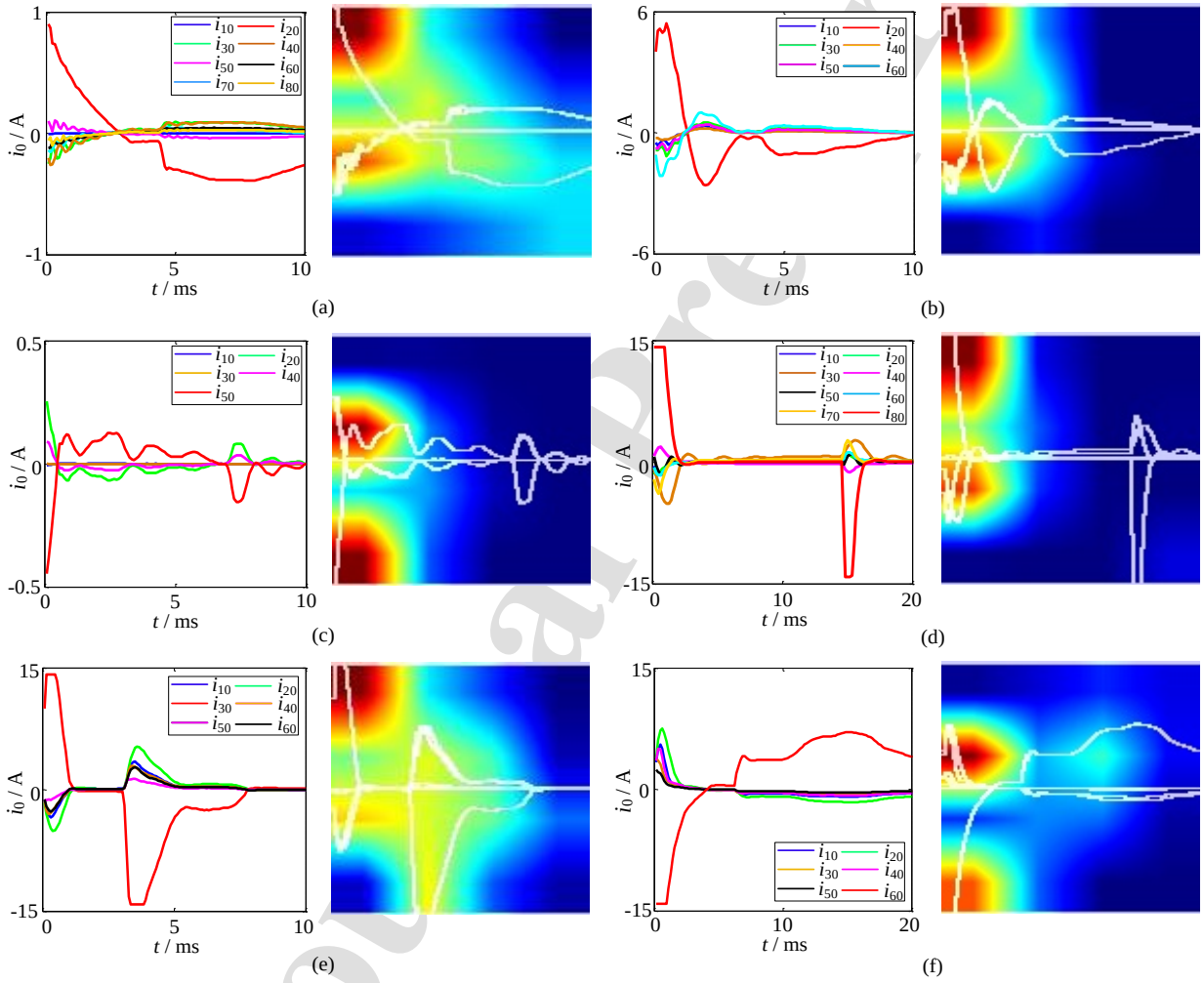


Fig. 13. Zero-sequence currents and their corresponding CAMs in practical distribution networks: (a) an SLG fault occurred in feeder 2, Chenzhou substation; (b) an SLG fault occurred in feeder 2, Shaoyang substation; (c) an SLG fault occurred in feeder 5, Xiangxi substation; (d) an SLG fault occurred in feeder 8, Bus I, Xi'an substation; (e) an SLG fault occurred in feeder 3, Bus II, Xi'an substation; and (f) an SLG fault occurred in feeder 6, Bus II, Xi'an substation.

It can be seen that the actual zero-sequence currents are associated with distinct arc extinction and reignition processes. In addition, when the arc is extinguished, it takes a long time for the current to become zero, which is different from the zero-sequence currents in the training dataset. However, the attention-based CNN can focus on the meaningful and discriminative parts of the waveforms, which ensures that the proposed method adapts to faulty feeder detection under extreme fault conditions.

The test results obtained using the practical fault data are listed in Table 7. Compared with the other three methods, the proposed detection method has 100% detection accuracy in these 12 sets of practical fault data owing to the image recognition of the superimposed currents and the attention-based mechanism. Therefore, the proposed method trained with simulation data in PSCAD can be directly applied for faulty feeder detection in a practical distribution system, which confirms its extensive application prospects.

Table 7
Detection results in practical distribution systems.

Method	Paper [26]	Paper [27]	Paper [28]	Proposed method
Accuracy	91.67%	33.33%	75.00%	100%

In summary, the superposition form of zero-sequence currents helps the CNN to conduct correlation comparisons between feeders, and the generated images obtained from superimposed currents are immune from changing topologies and sampling frequencies. In addition, the established CNN with attention strategy can always focus on meaningful parts in images, thereby enhancing its discriminative capability for image recognition. Large amounts of fault data demonstrate that the proposed method considerably improves the faulty-feeder detection accuracy and possesses a strong generalization capability.

4.3 Statistical Tests

The above comparisons verify the superiority of the proposed method in terms of the detection performance. To further evaluate the statistical differences between the proposed method and the other three methods, the Friedman and Nemenyi tests are conducted for statistical comparisons. The Friedman test [29, 30] is applied to determine if there are significant differences among the analyzed methods, where the null-hypothesis (H0) is that all methods are equivalent. If H0 is rejected, the Nemenyi test [31, 32] can be performed as a post-hoc test to compare the pairwise differences between the performances of the selected methods.

The Friedman statistic can be obtained by using (10).

$$\begin{cases} \chi_F^2 = \dfrac{12Q}{P(P+1)}\left[\sum_j R_j^2 - \dfrac{P(P+1)^2}{4}\right] \\ F_F = \dfrac{(Q-1)\chi_F^2}{Q(P-1)-\chi_F^2} \end{cases}, \tag{10}$$

where Q denotes the total number of datasets, P is the total number of compared methods, and R_j is the average ranking for j-th method in each dataset.

According to the conducted experiments in Section 4.1 and 4.2, there are a total of 13 sets of datasets used for performance comparison, where the bus faults and line faults are considered as two statistical datasets in Table 2. Furthermore, the four methods are ranked based on their detection accuracy in descending order. After ranking the four methods, the average ranking for each method can be calculated. The calculated average rankings of the methods in [26], [27], [28], and the proposed method are 2.962, 2.923, 2.962, and 1.154, respectively. The Friedman statistic values are then calculated by solving (10): χ_F^2 is 18.85 and F_F is 11.230. If the selected significance level α is 0.05, then the critical value $F_{(4, 13)}$ is 3.179. Therefore, the null-hypothesis H0 can be rejected, implying that the four methods are statistically different.

Subsequently, the Nemenyi test is conducted to determine whether there are significant differences between the compared methods. The critical difference (CD) in the Nemenyi test is defined as follows:

$$CD = q_\alpha \sqrt{\frac{P(P+1)}{6Q}}, \tag{11}$$

where q_α is the critical value, which is 2.569 when α is set as 0.05.

In this study, the corresponding CD for the 13 datasets and four methods is 1.301. Based on the obtained average rankings in the Friedman test, the differences between the proposed method and the three compared methods are 1.808, 1.769, and 1.808, respectively. Because the calculated differences are greater than CD, we can conclude that the proposed method has superior performance compared to the other three methods.

Therefore, the aforementioned detection performance and statistical tests demonstrate that the proposed method significantly improves faulty feeder detection and is more suitable for practical applications.

5. Conclusion

This study proposes a novel image-recognition-based method for faulty feeder detection in distribution networks. Owing to the superposition form of zero-sequence currents, the established CNN with the attention strategy can conduct a comprehensive comparison between currents, which is not limited to specific fault characteristics. In fact, the correlation comparison enables it to be used for varied fault scenarios, and the adaptability to changing topologies enables faulty feeder detection under different distribution networks. Additionally, the proposed image creation method can ensure fast computational speed and good detection efficiency. Numerous experimental results confirm the robustness of the proposed method to various topologies, parameters, grounding modes, fault location, fault time, transition resistances, and arc grounding events. Notably, the proposed method only utilizes the current signals available from current transformers already existing in substations, which implies that there is no

additional cost involved in the implementation. Therefore, the proposed method exhibits substantial application prospects in real installations.

Reference

[1] F.A.D. Souza, M.F. Castoldi, A. Goedtel, M.D. Silva, A cascade perceptron and Kohonen network approach to fault location in rural distribution feeders, Appl. Soft Comput. 96 (2020) 106627.

[2] X.W. Wang, J. Gao, X.X. Wei, Z.H. Zeng, Y.F. Wei, M. Kheshti, Single line to ground fault detection in a non-effectively grounded distribution network, IEEE Trans. Power Delivery. 33(2018) 3173-3186.

[3] A. Abdali, K. Mazlumi, R. Noroozian, High-speed fault detection and location in DC microgrids systems using Multi-Criterion System and neural network, Appl. Soft Comput. 79(2019) 341-353.

[4] X.W. Wang, X.X. Wei, D.C. Yang, G.B. Song et al., Fault feeder detection method utilized steady state and transient components based on FFT backstepping in distribution networks, Int. J. Electr. Power Energy Syst. 114(2020) 105391.

[5] M.Y. Chen and B.T. Chen, Online fuzzy time series analysis based on entropy discretization and a Fast Fourier Transform, Appl. Soft Comput. 14(2014) 156-166.

[6] A.C. Adewole, R. Tzoneva and S. Behardien, Distribution network fault section identification and fault location using wavelet entropy and neural networks, Appl. Soft Comput. 46(2016) 296-306.

[7] P. Li, Q.S. Zhang, G.L. Zhang, W. Liu, F.R. Chen, Adaptive S transform for feature extraction in voltage sags, Appl. Soft Comput. 80 (2019) 438-449.

[8] W.H. Yu, A mathematical morphology based method for hierarchical clustering analysis of spatial points on street networks, Appl. Soft Comput. 85 (2019) 105785.

[9] P. Konar, P. Chattopadhyay, Multi-class fault diagnosis of induction motor using Hilbert and Wavelet Transform, Appl. Soft Comput. 30 (2015) 341-352.

[10] M. Sahani, P.K. Dash, Fault location estimation for series-compensated double-circuit transmission line using parameter optimized variational mode decomposition and weighted P-norm random vector functional link network, Appl. Soft Comput. 85 (2019) 105860.

[11] W. Huang and R. Kaczmarek, SLG fault detection in presence of strong capacitive currents in compensated networks, IEEE Trans. Power Delivery. 22(2007) 2132-2135.

[12] J. Gao, X.H. Wang, X.W. Wang, A.J. Yang, H. Yuan, X.X. Wei, A high-impedance fault detection method for distribution systems based on empirical wavelet transform and differential faulty energy, IEEE Trans. Smart Grid. 13(2022) 900-912.

[13] S. Kar and S.R. Samantaray, Time-frequency transform-based differential scheme for microgrid protection, IET Gener. Transm. Distrib, 8(2014) 310-320.

[14] L.W. Xie, L.F. Luo, Y. Li, Y. Zhang, Y.J. Cao, A traveling wave-based fault location method employing VMD-TEO for distribution network, IEEE Trans. Power Delivery. 35(2020) 1987-1998.

[15] X.W. Wang, J. Gao, X.X. Wei, G.B. Song, L. Wu, J.W. Liu, Z.H. Zeng, M. kheshti, High impedance fault detection method based on variational mode Decomposition and teager–kaiser energy operators for distribution network, IEEE Trans. Smart Grid, 10(2019) 6041-6054.

[16] J.W Yuan and Z.B. Jiao, Faulty feeder detection based on fully convolutional network and fault trust degree estimation in distribution networks, Int. J. Electr. Power Energy Syst. 141(2022) 108264.

[17] J. W. Yuan, Z. B. Jiao et al., Study on fault line detection methods based on multi-feature fusion in distribution systems, IET Gener. Transm. Distrib, 15(2021) 860-869.

[18] M.F.S.V. D'Angelo, R.M. Palhares, L.B. Cosme, L.A. Aguiar, F.S.Fonseca, W.M. Caminhas, Fault detection in dynamic systems by a Fuzzy/Bayesian network formulation, Appl. Soft Comput. 21 (2014) 647-653.

[19] A. Vinayagam, V. Veerasamy, P. Radhakrishnan et al., An ensemble approach of classification model for detection and classification of power quality disturbances in PV integrated microgrid network, Appl. Soft Comput. 106(2021) 107294.

[20] M. Jamil, A. Kalam, A.Q. Ansari, M. Rizwan, Generalized neural network and wavelet transform based approach for fault location estimation of a transmission line, Appl. Soft Comput. 19(2014) 322-332.

[21] J. B. Yu and X.F. Yan, Deep unLSTM network: Features with memory information extracted from unlabeled data and their application on industrial unsupervised industrial fault detection, Appl. Soft Comput. 108(2021) 107382.

[22] Y.W. Chen, B. Song, Y. Zeng, X.J. Du, M. Guizani, Fault diagnosis based on deep learning for current-carrying ring of catenary system in sustainable railway transportation, Appl. Soft Comput. 100(2021) 106907.

[23] J.W. Yuan, Y.J. Hu, Z.B. Jiao, Faulty feeder detection method for SLG faults in distribution networks based on comprehensive fault characteristics across entire frequency spectrum, Int. J. Electr. Power Energy Syst. 140(2022) 107835.

[24] K. Xu, J. L. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. S. Zemel, Y. Bengio, Show, attend and tell: neural image caption generation with visual attention, 2016, arXiv:1502.03044.

[25] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, CBAM: Convolutional block attention module, 2018, arXiv:1807.06521.

[26] X.X. Wei, D. C. Yang, X. W. Wang, B. Wang, J. Gao, K. W. Wei, Faulty feeder detection based on fundamental component shift and multiple-transient-feature fusion in distribution networks, IEEE Trans. Smart Grid. 12(2021) 1699–1711.

[27] E.J. Zhai, Z.Y. Shu, J. Wang, Z.P. Huang, Fault line selection method of small current grounding system based on VMD-LSTM, Advanced Technology of Electrical Engineering and Energy. 40(2021) 70-80.

[28] M.F. Guo, X.D. Zeng, D.Y. Chen, N.C. Yang, Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems, IEEE Sensors Journal. 18(2018) 1291-1300.

[29] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Ann. Math. Stat. 11 (1940) 86–92, http://dx.doi.org/10.1214/aoms/1177731944.
[30] H.Y. Pan, L. S, H.F. Xu, J.Y. Tong, J.D. Zheng, Q.Y. Liu, Pinball transfer support matrix machine for roller bearing fault diagnosis under limited annotation data, Appl. Soft Comput. 125 (2022) 109209.
[31] P.B. Nemenyi, Distribution-free multiple comparisons (doctoral dissertation, princeton university, 1963), Diss. Abstr. Int. 25 (1963) 1233.
[32] H. Pan and J. Zheng, An intelligent fault diagnosis method for roller bearing using symplectic hyperdisk matrix machine, Appl. Soft Comput. 105 (2021) 107284.

Highlights

- Waveforms are superimposed to conduct correlation comparison
- Topology adaptability can be ensured by recognizing superimposed waveforms
- A fast image creation method is proposed to improve the detection efficiency
- Images are recognized by an established CNN with attention strategy

CRediT authorship contribution statement

Jiawei Yuan: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing.

Zaibin Jiao: Writing - original draft, Writing - review & editing, Supervision.

Declaration of interests

☒ 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: