Feature-clustering-based single-line-to-ground fault section location using auto-encoder and fuzzy C-means clustering in resonant grounding distribution systems

Jian-Hong Gao1,2 | Mou-Fa Guo1 | Xiang Shao1 | Duan-Yu Chen2

1 Department of Electric Power Engineering, Fuzhou University, Fuzhou, China
2 Department of Electrical Engineering, Yuan Ze University, Chung Li, Taiwan

Abstract
Many sensors like digital fault indicators (DFIs) have been applied and promoted in distribution systems. The sensors can provide a technical mean for single-line-to-ground (SLG) fault section location, but there are still some feature extraction and fault diagnosis problems. A novel SLG fault section location method utilizing auto-encoder (AE) and fuzzy C-means (FCM) clustering is presented in this work. Taking advantage of abundant information provided by DFIs, striking features can be extracted by the AE network, which is different from the artificially designed features that rely on prior knowledge. Compared with the learning-based methods requiring massive training data, the proposed method only requires the data from one SLG fault. By applying the AE network to the zero-sequence current measured by DFIs, the SLG fault section location's striking features could be obtained. Through feature classification by FCM clustering without setting threshold, the positional relationship between each detection node and the fault point would be distinguished to locate the fault section. Considering the abnormal communication of DFIs, the experiment proves that the proposed method can work effectively under various fault conditions.

1 | INTRODUCTION

Statistically speaking, most faults that occurred in distribution systems are single-line-to-ground (SLG) faults, which would threaten the system's insulation, protection tripping, and equipment safety [1]. Resonant grounding distribution systems are adopted by most electric utilities in China, Japan, and some European countries [2]. The technical standards permit the power supply to run for a specific time after the faults [3]. Unfortunately, because the SLG fault characteristics are not prominent, the traditional protection devices may not detect SLG failure in some special situations. They cannot isolate the fault section in time, prolonging the failure’s duration or leading to a more severe fault [4]. Recently, the state-of-the-art techniques for SLG fault location have developed rapidly. However, the SLG fault section location methods deserve more attention because most of the SLG fault locations are specified for detecting the fault line. Therefore, it is critical to realize the SLG fault section location, which is an essential aspect of enhancing self-healing and automated fault management systems [5].

To address the problem of fault section location, the scholars divided the related methods roughly into several categories, such as the traveling waveform method [6], injection signal method [7], impedance-based method, etc. Each method has its advantages and disadvantages. The method in [6] can measure the traveling wave to locate the fault position precisely. The traveling wave methods can be sorted into single-ended methods and double-ended methods both requiring the specified devices to measure the traveling waves. There is difficulty in wave-head detection for single-ended methods due to the complex network structure with multiple branches. Meanwhile, the slight measurement error would affect the time synchronization for double-ended methods. The injection signal method [7] utilized a specially designed equipment to inject a characteristic signal into the distribution systems and usually work offline to avoid
affecting its regular operation. The first two methods need new devices with high investment installed in primary substations. Some researchers focused on the impedance-based techniques [8, 9], generally calculating the equivalent impedance to locate the fault point based on the voltage and current measured at the primary substation. The proposed method in [8] utilized the bus impedance matrix to express the substation voltage and current as a function of the fault location and fault resistance (FR) to yield the fault point. The fault location could be identified by comparing the calculated reactance between the fault point and the critical node with the known line parameter, the fault location involving the high-frequency transient information in [9]. However, the impedance-based methods can rarely solve the problem of multiple fault sections or fault points with the same electrical distance in industrial applications. Although the above techniques adopted the information of the primary substation for SLG fault section location, the accuracy and speed of such methods would be affected by the limited availability of data due to the line with complicated laterals and different parameters of the line in urban distribution systems yet [10]. Therefore, in addition to the information from the primary substation, more distribution system data is conducive to realize SLG fault section location.

Recently, many sensors, such as digital fault indicators (DFIs), have been widely installed in distribution systems due to their lower cost [11–13]. Although the traveling wave-based method and injection-based method can work well during SLG fault, both methods have their limitations. Traveling wave-based method requires a high sampling rate, and it is difficult to adapt to the multi-branch structure in distribution systems. The injection-based method needs additional signal injection equipment installed, and the cooperation of the detection equipment on the line is required to locate the fault section. In theory, the more the DNs are installed in a line, the higher the accuracy of locating the fault section will be. However, it is neither economical nor necessary to install too many DNs at each line in distribution systems, balancing the relationship between the cost of customer service outage and device investment cost [14]. The DFIs with optimal configuration would reduce operating costs and service interruptions by identifying the fault section. The electrical quantities at DNs could be recorded, of which the issue of fault section location may benefit from the abundant information. In general, the sensors such as DFIs could indicate the line section through which the fault current flows, depending on whether the fault current exceeds the threshold [15]. Because of existing blind zones and misjudgements when using a single criterion, two methods are fused based on fuzzy theory to realize fault section location in [16], which require feature extraction and threshold setting by human. The fault section could be identified by constructing a relationship matrix between possible fault location and line current in [11], which may be affected when receiving incomplete information due to communication loss. The method based on setting threshold cannot provide satisfactory results under the various conditions in industrial applications. Therefore, extracting useful features from redundant information for SLG fault section location are the primary problem for further research.

Many methods [17, 18] utilized the transient signals of SLG fault to extract features artificially by signal processing technologies. Then, the artificial features were used to establish the judgment criterion for fault diagnosis. In [18], the dynamic time warping (DTW) distance is calculated as an artificial feature, measuring the similarity of two zero-sequence currents from two adjacent DNs. The fault section would be located depending on whether the pre-set threshold of DTW is achieved. Although the suitable threshold could be determined by massive data in the simulation platform to find the fault section, it could be hard to make high accuracy due to the complicated industrial environment. The data-driven techniques [19–22] would extract features adaptively and realize the fault location, simultaneously. In [20], the fault point can be predicted by the trained support vector machine, which maps low-dimensional data into high-dimensional space through kernel function to separates the samples. An SLG fault indicator is proposed in [21], estimating SLG fault based on k-nearest neighbour in the regression model. Because there exists a difference between the transient zero-sequence current waveforms of the fault line and the sound lines, continuous wavelet transformation and convolutional neural network (CNN) were adopted in [22] to extract useful features and locate fault line simultaneously, which could also be utilized to realize SLG fault section location incorporating sensors such as DFIs. Even so, a large number of samples with hand-crafted labels are needed for model construction, and the high computational cost has become a bottleneck of such methods. Besides, as for electrical utilities, it is difficult to collect enough SLG fault data to build a suitable sample library for training models in practical applications.

The method in [23] for fault line detection utilizes the correlation between the outlet zero-sequence current of the fault line and the sound lines, whose idea could also be used for SLG fault section location. When SLG fault occurs in the distribution systems, the transient zero-sequence currents through two-ends of the sound line section are almost the same. Meanwhile, the transient zero-sequence currents through two ends of the fault line section are quite different. The DFIs with communication interfaces could transfer electric quantities to the control data, by which the zero-sequence current at different detection nodes (DNs) could be measured. However, the communication unit of DFIs usually blocks channels, loss packets, and then retransmit, thus, which cannot cooperate with the control centre in real-time for better performance of fault location [24].

To overcome the literature’s drawbacks, we proposed a novel fault section location method using auto-encoder (AE) and fuzzy C-means (FCM) clustering in this work. Incorporating with sufficient information provided by DFIs, the AE network can extract useful features from fault signal adaptively, which is different from the artificially designed features based on conventional methods. The proposed method only requires data from the current event of SLG fault, thereby
overcoming the problem that it is challenging to collect on-site recorded data of SLG fault. By applying the AE network to the zero-sequence current of different DNs, the compelling features could be obtained and classified by FCM clustering without setting the threshold. Based on the features classification, the positional relationship between each DN and the fault point would be distinguished to locate the fault section.

The remaining of this work is organized as follows. Section 2 introduces a character analysis of SLG fault and the proposed method, including feature extraction and fault classification. In Section 3, it provided a simulation model based on PSCAD/EMTDC and verification using simulation data and on-site recorded data. In Section 4, the proposed algorithm in this work is compared with other existing methods to assess its performance. Finally, the conclusion is given in Section 5.

2 PRINCIPLE AND METHOD

Figure 1 shows a resonant grounding distribution system with lines divided into some sections by DNs where $r_L$ and $L$ are the equivalent resistance and inductance of the Peterson coil. Supposing that an SLG fault occurs at the $m$-th section between DN $m$ and DN $m+1$ of the line $L_j$, the zero-sequence network is depicted in Figure 2.

After an SLG fault occurs, the zero-sequence current of the fault point consists of the capacitance currents and the inductive current via the Peterson coil. In the transient process after the SLG fault occurs, it can be considered that the zero-sequence virtual voltage source at the fault point charges the Peterson coil, and the grounding capacitance is in a state of rapid impulse discharge. At this time, the frequency of the transient capacitance current’s frequency and the transient inductor current are very different, so the two cannot be compensated. The general transient process lasts for several cycles, during which the role of the Peterson coil can be ignored [25].

Figure 2 shows the zero-sequence network of the fault line, which neglects the Peterson coil. $u_f$ and $R_f$ indicate the virtual source of fault point and FR, respectively. $G_{jk}$ and $B_{jk}$ are the ground conductance and ground admittance of the $k$th section. $i_{0-m}$ is the zero-sequence current measured from DN $m$ in the $m$th section of the line $L_j$. $i_{0-i}(i \neq j)$ is the zero-sequence current measured at the bus outlet in the sound lines, also termed as its leakage currents generated by the total of the ground conductance and ground admittance of the $k$th section. $\sum i_{0-k}$ is the sum of the zero-sequence current flowing through the sound lines. Meanwhile, the zero-sequence current measured from the DNs in the downstream of fault point is determined by the capacitance currents mainly.

The zero-sequence current measured from the DNs in the upstream of fault point depends on the capacitance currents in the downstream line and the leakage currents in all sound lines, as shown in Figure 2. It is inferred that there exists a difference in zero-sequence current between the upstream DNs and the downstream DNs of the fault point; moreover, there are similarities between the zero-sequence currents of the DNs located on the same side of the fault point. However, the transient zero-sequence currents from the DNs located on the same side of the fault point are almost the same. Therefore, the characteristic of SLG fault is the basic principle to determine the SLG fault section in this work. The flow chart of the proposed method can be shown in Figure 3. When an SLG fault occurs, the zero-sequence currents at all DNs would be recorded by DFIs. The AE network is established to extract the features of the normalized transient zero-sequence current automatically. Features classification is used to divide the DNs.
into two groups by FCM clustering without the set threshold to determine whether the DN is at the downstream or upstream of the fault point. Because the amplitude of the zero-sequence current at the upstream of the fault point is more significant than that at the downstream of the point, therefore, the positional relationship between DNs and the fault point would be obtained and used to locate the fault section that SLG fault occurs.

2.1 Measurement and pre-processing

When an SLG fault occurs in resonant grounding distribution systems, the zero-sequence voltage \( U_0 \) is adopted as the triggering criterion depending on whether \( U_0 \) is more significant than 15% of the phase voltage. Then, the transient zero-sequence currents at all DNs would be recorded by the sensors such as DFIs. The first-half cycle waveforms of the transient zero-sequence currents will be obtained for normalization. Because there exists a difference in the transient zero-sequence current at the different DNs in actual distribution systems, the transient zero-sequence currents should be normalized as

\[
\left( x_j^f \right)^* = \frac{x_j^f}{\max \left\{ \max \left( |X^1| \right), \max \left( |X^2| \right), ..., \max \left( |X^N| \right) \right\}},
\]

where \( X^j (j = 1, 2, ..., N) \) is the transient zero-sequence current gathered from DN\(_j\), \( N \) is the number of DNs, \( x_j^f \) is the \( j \)th data point of \( X^j \).

2.2 Feature extraction with AE

It is challenging to extract compelling features from the SLG fault to establish a suitable judgment criterion for fault section location. Generally speaking, a single characteristic quantity of the fault signals such as amplitude/polarity, energy, waveform correlation coefficient, etc., is taken as artificial features to establish the existing method’s criterion. The methods based on supervised-learning algorithms could extract the features of fault signals adaptively and realize intelligent classification for fault diagnosis simultaneously. However, it requires many samples for the training model, which could not be practical because the on-site recorded data is hard to obtain in practical applications. Therefore, AE is utilized to learn the fault signal features, which is an unsupervised learning algorithm and may be an effective solution for fault diagnosis.

As shown in Figure 4, the basic structure of the AE network is composed of an encoder and decoder. The input layer and output layer own the same dimension to make the reconstructed input infinitely close to the original data. Meanwhile, the hidden layer’s size is smaller than the two layers to reduce the dimension of the input data. The hidden layer’s output feature would be regarded as the compressive representation of the input data [26]. The encoder maps the original signal \( \mathbf{x} = [x_1, x_2, ..., x_n]^T \) from the input layer into the hidden layer, and the output of the hidden layer \( \mathbf{h} = [h_1, h_2, ..., h_n]^T \) can be obtained as

\[
\mathbf{h} = f(\mathbf{x}) = \tanh \left( \sum_{i=1}^{n} w_i x_i + b \right).
\]

Then, the decoder would utilize the hidden vector \( \mathbf{h} \) to get the reconstructive signal \( \hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, ..., \hat{x}_n]^T \)

\[
\hat{\mathbf{x}} = g(\mathbf{h}) = \tanh \left( \sum_{i=1}^{n} w'_i h_i + b' \right)
\]

where \( w, w' \), and \( b, b' \) are the weight coefficients and bias coefficients of the AE network in the encoder and decoder processes, respectively. Multiple forward iterations and back-propagation algorithms adjust the weights and bias of the network. \( f(\cdot) \) and \( g(\cdot) \) are the activation functions in encoder and decoder processes, where tanh function is utilized instead of the general sigmoid function or ReLU function, which cannot reflect the polarity of the transient zero-sequence currents.

The AE network’s training termination condition is that the number of epochs exceeds 100 times, or the loss function’s output in (4) is less than the threshold set to \( 10^{-3} \)

\[
E = E^{(j)} + \lambda D
= \frac{1}{2N} \sum_{j=1}^{N} \left\| \hat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)} \right\|^2 + \lambda \sum_{i,j} w_{i,j}^2
\]

where \( E^{(j)} \) represents the loss function’s output of the \( j \)th training sample measuring the difference between the original signal \( \mathbf{x}^{(j)} \) and reconstructive signal \( \hat{\mathbf{x}}^{(j)} \), and \( N \) is the number of training samples. The second item \( D \) is the L2 regularization, and \( \lambda \) is the weight attenuation coefficient set to
TABLE 1 Hyperparameters of AE network

| Description                  | Value |
|------------------------------|-------|
| Number of neurons in the input layer | 100   |
| Number of neurons in the hidden layer | 10    |
| Number of neurons in the output layer | 100   |
| Learning rate                | 0.1   |
| Maximum epochs               | 100   |

Abbreviation: AE, auto-encoder.

2.3 Feature classification with FCM

FCM clustering is a useful pattern recognition method, which divides the input signal into a cluster centre based on the membership matrix to realize flexible fuzzy partitioning. Supposing that \(Z = [z_1, z_2, ..., z_N]\) be the sample space, \(N\) is the total number of samples. Each sample is an \(n\)-dimensional vector represented by \(z_i = [z_{i1}, z_{i2}, ..., z_{in}]\). The \(Z\) is divided into \(c\) classes (\(2 \leq c \leq m\)), and each sample is divided into different classes based on the membership matrix \(U_{i\in C_j} = [u_{ij}]\) to some extent. The \(i\)th sample in \(Z\) belongs to the membership degree of \(i\)th class represented by \(u_{ij}\), which satisfies (5)

\[
\begin{align*}
    \sum_{i=1}^{c} u_{ij} &= 1 \\
    u_{ij} &= 1 / \sum_{i=1}^{c} \left( \frac{d_{ij}}{d_{ij}} \right)^m \\
    &\in [0, 1],\quad (i = 1, 2, ..., c) \\
    &j = 1, 2, ..., n
\end{align*}
\]

(5)

where \(m\) is the fuzzy weighted index, set here as \(m = 2\); and \(d_{ij} = \| z_i - v_j \|\) represents Euclidean distance of the sample \(z_j\) to the cluster centre \(v_j\). \(V = [v_1, v_2, ..., v_c]\) is cluster centre; \(v_i\) (\(i = 1, 2, ..., c\)) is the cluster centre of the \(i\)th class, set here as \(c = 2\). The calculation formula for \(v_i\) and the iterative formula to minimize the objective function are shown as

\[
f_i = \frac{\sum_{j=1}^{c} u_{ij} z_j}{\sum_{j=1}^{c} u_{ij}} (i = 1, 2, ..., c)
\]

(6)

\[
\min J_{FCM}(U, V) = \sum_{i=1}^{m} \sum_{j=1}^{c} u_{ij} \| z_i - v_j \|_2^2.
\]

(7)

After iterative convergence, the optimal clustering centre \(V = [v_j]\) and the membership matrix \(U_{i\in C_j} = [u_{ij}]\) are obtained. The features extracted from the fault data are input into FCM clustering in order to achieve feature classification.

2.4 Fault section location based on AE and FCM

In engineering applications, the AE network is applied to the transient zero-sequence currents to extract useful features. The feature classification is then achieved by FCM clustering to obtain a membership matrix to determine whether the DN is in the upstream of the fault point or the downstream. The membership matrix is as follows:

\[
U = \begin{bmatrix}
    u_{11} & u_{12} & \cdots & u_{1N} \\
    u_{21} & u_{22} & \cdots & u_{2N}
\end{bmatrix}
\]

(8)

where \(u_{ij}\) indicates the membership degree that the DN\(_j\) attaches to the category \(i\), and \(N\) equals the number of DN\(_s\). The DN\(_s\) attached to category 1 or category 2 could be confirmed through the membership matrix \(U\). FCM clustering can divide DN\(_s\) into the two categories, but it fails to determine whether the DN belongs to the upstream or downstream of the fault point because the amplitude of the transient zero-sequence current at the upstream section of the fault point is higher than that at the downstream section. Therefore, the average amplitude of the transient zero-sequence current is obtained by (9) revealing the different categories’ content

\[
Ar_i = \left( \frac{\sum_{x_k \in C_j} \sum_{k=1}^{n} |x_k|}{N_j} \right)
\]

(9)

where \(Ar_i(j = 1, 2, ..., c)\) indicates the average amplitude of the zero-sequence current measured by the DN\(_s\) grouped in the \(j\)th category, \(|\cdot|\) means the absolute value operation, \(N_j\) indicates the total number of the DN\(_s\) grouped in the \(j\)th category.

In this work, utilizing the sensors such as DFIs installed in the distribution system, the proposed method for SLG fault section location was presented combining with AE and FCM clustering. When an SLG fault occurs, the data of the transient zero-sequence current would be collected. After normalization operation, the first-half waveforms of the transient zero-sequence current would be regarded as the AE network’s original input. Through the encoder and the decoder, the reconstructed waveform is obtained to infinitely approximate the original waveform. The transient zero-sequence current’s output feature would be extracted from the hidden layer of the AE network once the training termination condition is satisfied. The FCM clustering is adopted to the extracted features. The extracted features are sorted into two categories, and the clustering result shows the relationship between the fault point and the DN\(_s\) to locate the fault section. The visualization process of the proposed method is shown in Figure 5. Due to problems like data shortage and communication problems existing in the real environment, the proposed method can overcome the difficulty in model training requiring massive training data and
minimizing data. The corresponding experiments are carried in the next section.

## 3 | EXPERIMENTAL VERIFICATION

### 3.1 | Simulation test

The simulation model of a resonant grounding distribution system based on PSCAD/EMTDC® is established to verify the proposed method, where $O_L$ indicates the length of the overhead line, $C_L$ suggests the length of the cable line, $R_f$ indicates FR. The main transformer, overhead line, and cable line parameters could be found in Tables 2 and 3, respectively. What’s more, adopt the impedance $Z$ to equalize each line’s load, which is set as $100+40 \Omega$. The compensation degree of the Peterson coil is temporarily taken to 10%. The corresponding equivalent inductance $L$ is equal to 0.5265 H. Because the active loss of the Peterson coil is about 2.5%–5% of inductive losses, 3% of inductive losses is considered as the active loss for calculation, thus, $r_L$ is $4.95 \Omega$. Due to limited space, the line $L_1$ with one branch is set as the fault line, and there are seven DNs installed with DFIs along the line, as shown in Figure 6. The sampling frequency is set as 10 kHz.

The SLG faults are set to occur at the line $L_1$, which is simulated under various conditions. Table 4 shows typical faults under different fault points, fault initial angles (FIA), and FR). The fault sections are all located correctly. For each fault with a verified condition, for instance, the FR would change randomly, but there is always a certain difference between the transient zero-sequence current waveforms. The AE network can extract the adaptive characteristics from the waveforms. Therefore, the proposed method is suitable for the typical SLG fault with different FIAs and FRs. Besides, the influence of noise is

### TABLE 2 Parameters of main transformer

| Capacity (MVA) | Voltage high/low (kV) | No-load loss (kW) | Short-circuit loss (kW) | No-load current percentage (%) | Short-circuit voltage percentage (%) |
|---------------|-----------------------|------------------|------------------------|-------------------------------|-------------------------------------|
| 31.5          | 110/10 (YNd11)        | 19.5             | 118.9                  | 0.12                          | 14.99                               |

### FIGURE 6 The model of a resonant grounding distribution system
TABLE 4 Location result for typical faults

| Fault point | SNR (dB) | FIA (°) | FR (Ω) | Location results |
|-------------|----------|---------|--------|------------------|
| $f_1$       | 20       | 0       | 100    | S\textsubscript{1} |
|             | 60       | 300     |        |                  |
| $f_2$       | 10       | 30      | 10     | S\textsubscript{2} |
|             | 60       | 500     |        |                  |
| $f_3$       | 5        | 45      | 500    | S\textsubscript{3} |
|             | 90       | 1000    |        |                  |

TABLE 5 Location result for different voltage levels

| Case | Fault point | Voltage level (kV) | Location results |
|------|-------------|--------------------|------------------|
| 1    | $f_2$       | 6                  | S\textsubscript{1} |
| 2    | 10          |                    | S\textsubscript{2} |
| 3    | 35          |                    | S\textsubscript{3} |

TABLE 6 Location result for different compensation degrees

| Case | Fault point | Compensation degree (%) | Location results |
|------|-------------|-------------------------|------------------|
| 1    | $f_2$       | −8                      | S\textsubscript{1} |
| 2    | −5          |                         | S\textsubscript{1} |
| 3    | 5           |                         | S\textsubscript{2} |
| 4    | 8           |                         | S\textsubscript{2} |

TABLE 7 Location result for different topologies

| Case | Fault point | Description | Location results |
|------|-------------|-------------|------------------|
| 1    | $f_2$       | No change   | S\textsubscript{2} |
| 2    | $l_5$ (deleted) |            | S\textsubscript{2} |
| 3    | $l_6$ (added) |             | S\textsubscript{2} |

TABLE 8 Location result for different zero-sequence parameters

| Case | Fault point | Parameters          | Location results |
|------|-------------|---------------------|------------------|
| 1    | $f_2$       | No change           | S\textsubscript{2} |
| 2    | $l_6$       | 90% inductance      | S\textsubscript{2} |
| 3    | $l_6$       | 110% inductance     | S\textsubscript{2} |
| 4    | $l_6$       | 90% capacitance     | S\textsubscript{2} |
| 5    | $l_6$       | 110% capacitance    | S\textsubscript{2} |

TABLE 9 Location result for different loads

| Case | Fault point | Load condition     | Location result |
|------|-------------|--------------------|-----------------|
| 1    | $f_2$       | No change          | S\textsubscript{2} |
| 2    |              | Loads removed      | S\textsubscript{2} |
| 3    |              | Unbalanced loads   | S\textsubscript{2} |

FIGURE 7 Feature extraction and classification under noise interference (a) original waveform, (b) polluted waveform, (c) extracted features, (d) clustering result

also taken into consideration. The SLG fault in the bolded line in Table 4 is taken as an example to illustrate the result in detail. After adding 5 dB of noise, the original and polluted waveform of the transient zero-sequence current is shown in Figure 7(a) and (b), respectively. The extracted feature of transient zero-sequence current is shown in Figure 7(c). And the clustering result based on FCM is shown in Figure 7(d), which gives a good discrimination degree between the fault line section and sound line sections.

Tables 5 and 6 show the fault section results for different voltage levels and compensation degrees. It shows that fault line sections can be correctly identified in a wide range of voltage levels and compensation degrees.

With the rapid development of distribution systems, changes in topology and parameters should be considered. The topology changes are simulated by adding and deleting one line in the model of the distribution system, illustrated in Table 7, which shows that the proposed method is suitable for use in different network structures. The zero-sequence parameters of the system are changed to verify the robustness of the proposed method. When an SLG fault occurs at $f_2$, the FIA, FR, and SNR are $60^\circ$, 100 $\Omega$, and 20 dB, respectively. There is no change in the zero-sequence parameters of Case 1 in Table 8. In Case 2 and Case 3, the zero-sequence inductance of the systems is set to 90% and 110% of the initial inductance. Meanwhile, the zero-sequence capacitance is set to 90% and 110% of the initial capacitance in Case 4 and Case 5. According to the results of the location presented in Table 8, it can be observed that there is no influence on the location results when the zero-sequence parameter of the systems fluctuates by 20%.

The effect of different load types is evaluated, as shown in Table 9. The load conditions considered in Case 1 is that all loads are three-phase balanced constant loads. In Case 2, the loads in line $L_2$ are removed. In Case 3, the loads connected to the end of the first branch of the line $L_1$ are replaced with one-phase constant loads. Meanwhile, the loads connected to the
end of the second branch are replaced with two-phase constant loads. As the location results are shown in Table 9, the proposed method could work correctly for different load types, even if unbalanced loads are connected to the systems.

An SLG fault signals in engineering applications, such as the zero-sequence currents, are obtained through wired or wireless transmission from sensors. However, asynchronous sampling is unavoidable due to low communication quality. Table 10 shows two cases used to verify the proposed method. In Case 1, the signals from DN1 and DN2 lag five sampling points behind the signals from DN3 and DN4. In Case 2, the signal collected from DN1 and DN2 lead ten sampling points ahead of the signals of DN3 and DN4. The location results in Table 10 show that the proposed method can overcome the influence of little asynchronous sampling and effectively locate the fault section.

### 3.2 On-site recorded data test

Nowadays, there are many DFIs installed in real resonant grounding distribution systems in China. Once the SLG fault occurs, the fault signal would be detected and measured by the detection unit. The communication unit then transfers the data into the control centre via a general packet radio service [24]. Although the slight communication problem like asynchronous sampling has been discussed in the simulation test, when the problem becomes worse, the fault signals of some DNs cannot be sent into the control centre. According to the on-site recorded data, the phenomena of data missing is more common. Therefore, the reliable detection node is introduced to express that the corresponding DFI could capture and send the fault signals into the control centre successfully. The range of the fault zone would be expanded due to the problem above. However, the fault section could be located correctly by the proposed method still.

The fault data of various SLG faults is collected in a real 10 kV distribution system in China. The structure of the system is shown in Figure 8(a). As shown in Figure 8(b), each DFI consists of a fault detection unit and a communication unit. There are four cases taken as examples to illustrate the robustness of the proposed method. In Case 1, the SLG fault occurs at the fault point f1. Meanwhile, only eight EDNs are sending the data back. Figure 9(d) shows the on-site recorded waveforms for Case 1. Figure 9(b) and (c) are the original transient zero-sequence currents at the double-end of the fault section and the extracted feature. The clustering result is shown in Figure 9(d). It can be seen that the amplitude of the transient zero-sequence current from DFI12 is more significant than that of other DFIs. The zero-sequence current’s polarity is also opposite to that in the sound line or the downstream of the fault line. In Case 2, an SLG fault occurred at the fault point f2 accompanied by an intermittent arc. In Case 3 and Case 4, the fault points are f3 and f5, respectively. The location results are illustrated in detail in Table 11. The results show that although the fault zone range has been expanded, the fault sections are still located correctly, indicating that the proposed method could affect on-site recorded data.

### 4 COMPARISONS WITH OTHER METHODS

The proposed method based on AE and FCM can adaptively extract the essential features without manual pre-construction and extraction. Besides, the AE belongs to the unsupervised learning method, which is practical for limited sample scenarios. FCM clustering can also realize the feature classification automatically without human intervention. To present the proposed method’s reliability advantage, we compared this method with two methods introduced in [27] and [18], which have been verified in distribution systems.

### 4.1 Method based on 1-D CNN

The authors in literature [27] utilized 1-D CNN to detect the fault line in distribution systems. First, the 1-D CNN model is trained to extract each line’s features to distinguish whether the line is the fault line or sound line. Several trained 1-D CNN models are then used to construct a new model for detecting which line is the fault line in the substation with several outlets.

This idea of fault line detection can also be used in the fault section location. The general 1-D CNN would extract feature by the pooling layer or convolutional layer; then, the features
are input to the fully connected layers, which maps the label’s extracted feature. Because there is a difference between the upstream and the downstream of the fault point, we first train a 1-D CNN model to determine whether the fault signal is upstream or downstream. Additionally, considering that the correlation of fault signals at the double-end of the line sections should be considered, we concatenated the fully connected layer of two trained 1-D CNN models to construct a new framework to distinguish whether the line section is the fault section or not. Therefore, we input the transient zero-sequence currents at the double-end of the line section into a fused framework, which concatenates the two trained 1-D CNN models’ fully connected layers to achieve feature fusion and classification. Such a model based on 1-D CNN can locate the fault section effectively. The structure of the proposed 1-D CNN is shown in Figure 10.

### 4.2 Method based on DTW

The authors in literature [18] adopted the DTW distance as a feature to locate the fault section, which uses the dynamic programming idea to measure the difference between two two-series to obtain an optimal path to get the smallest DTW distance between the two sequences along the path. The DTW distance is taken as an artificially designed feature to establish a judgment threshold based on massive simulation experiments. When an SLG fault occurs, the transient zero-sequence currents measured from the DNs would be obtained and used to calculate DTW’s distance as the features to represent the line sections. The resulting matrix $\text{DTW}^*$ is illustrated as

$$\text{DTW}^* = \left[ d_{ij} \right]$$

where $d_{ij}$ indicates the distance of DTW between the two adjacent DNs such as $\text{DN}_i$ and $\text{DN}_j$. The line section $S_{ij}$ between $\text{DN}_i$ and $\text{DN}_j$ could be determined as a fault section if the matrix component exceeds the set threshold $D_{\text{set}}^{\text{DTW}}$, which is generally taken as 0.4 based on massive simulation experiments in [18]. Because the DTW distance is manually extracted as artificial features relying on prior knowledge, and the thresholds need to be set based on simulation experiments. Thus, it is difficult to adapt to all fault conditions, especially to some extreme conditions.

Three SLG fault section location methods mentioned above are used to identify the fault section when the SLG fault occurs. Because the method based on 1-D CNN needs many samples for the training model, the training samples were provided by simulating SLG fault in the lines, except the line L1, shown in Figure 8 with different parameters, including FR, fault angle, and fault point. Due to limited space, the training process of 1-D CNN is not shown here.
Since the methods are mainly used to locate the fault section rather than the fault line, the simulation is mainly performed on the line $L_1$. The results of the three methods under regular conditions (e.g. different FR, different fault angle, and different fault point) was shown in Table 12, and $\varphi$ is the accuracy of fault section location, the ratio of the number of correctly located samples ($C_n$) to the total number of samples ($T_n$), which is defined in (11)

$$\varphi = \frac{C_n}{T_n} \times 100\%.$$  

(11)

The simulation data's comparison results have improved that the proposed method based on AE-FCM and the method based on 1-D CNN perform better than the method based on DTW, which directly calculated the similarity between fault signals. Because of the complexity of SLG fault, harsh environment, and various parameters, the conventional method like the method based on DTW adopted artificial features cannot accomplish the fault section location correctly in some special conditions. The proposed method and the method based on 1-D CNN both belong to machine learning. Although the machine learning algorithms can automatically extract fault features, the method based on 1-D CNN requires a large amount of historical data to train the model. The accuracy of the model is related to the amount of training data. The more is training data, the higher is the model accuracy. On the contrary, the proposed method based on AE-FCM only needs the data of the current event of SLG fault for self-learning and extracts a practical feature for discriminating the fault section. Therefore, the proposed method can be more suitable and practical in engineering applications than the other two comparison methods.

5 | CONCLUSION

A novel SLG fault section location method based on the AE network and FCM clustering have been proposed in this work. Our contributions are three-fold mainly.

1. Incorporating the sensors such DFIs installed at distribution systems, the AE network is utilized to extract features from the transient zero-sequence currents after the SLG fault's occurrence. The extracted features would be classified by FCM clustering without setting threshold into different categories, representing the position of DNs relative to the fault point approximately. Combining the characteristic of SLG fault, each DN's more accurate position relative to the fault point could be distinguished effectively. Finally, the SLG fault section location would be implemented.

2. AE network is adopted to learn features automatically based on the data from the single event of SLG fault event without choosing artificially designed features, which has advantages over supervised learning methods that need lots of training samples. Considering that the on-site recorded data is hard to collect, the method could be more feasible in engineering applications.

3. Since the sensors exist, the communication problem results in data missing. Simulation experiments have proved that the proposed method could correctly locate the fault section
| Case | Fault point | Valid detection node (DFIᵢ) | Method | U / DTW<sup>a</sup> | Location result<sup>a</sup> |
|------|-------------|-----------------------------|--------|---------------------|--------------------------|
| 1    | f₁₂        | DFI₁₁, DFI₁₀, DFI₁₃, DFI₁₄, DFI₁₆, DFI₁₇ | AE-FCM | \[
\begin{align*}
\text{\(a₁, a₁₀, a₁₁, a₁₂, a₁₃, a₁₄, a₁₆, a₁₇\)}
\end{align*}
| S₁₂ |
|      |             |                            | 1-D CNN | \[
\begin{align*}
\text{\(a₁₉, \ldots, a₁₉₄, a₁₆₇\)}
\end{align*}
| S₁₂ |
|      |             |                            | DTW     | \[
\begin{align*}
\text{\(d_{9,10}, d_{12,13}, d_{13,14}, d_{16,17}\)}
\end{align*}
| S₁₂, S₁₆ |
| 2    | f₅         | DFI₁₁, DFI₁₃, DFI₁₄, DFI₁₆, DFI₁₈ | AE-FCM | \[
\begin{align*}
\text{\(a₁, a₁₃, a₁₅, a₁₇, a₁₉, a₁₁₂, a₁₄, \ldots, a₁₆\)}
\end{align*}
| S₅ |
|      |             |                            | 1-D CNN | \[
\begin{align*}
\text{\(a₁₁₄, a₁₃₅, \ldots, a₁₆₇\)}
\end{align*}
| / |
|      |             |                            | DTW     | \[
\begin{align*}
\text{\(d₁₃, d₁₅, d₂₈, d₂₁₃, d₃₁₄\)}
\end{align*}
| / |
|      |             |                            |         | \[
\begin{align*}
\text{\(0.110, 0.050, 0.460, 0.580, 0.52\)}
\end{align*}
| |
| 3    | f₅         | DFI₁₁, DFI₁₃, DFI₁₄, DFI₁₆, DFI₁₈ | AE-FCM | \[
\begin{align*}
\text{\(a₁, a₁₃, a₁₅, a₁₇, a₁₉, a₁₁₂, a₁₄, \ldots, a₁₆\)}
\end{align*}
| S₁₄ |
|      |             |                            | 1-D CNN | \[
\begin{align*}
\text{\(a₁₁₄, a₁₃₅, \ldots, a₁₆₇\)}
\end{align*}
| S₁₄ |
|      |             |                            | DTW     | \[
\begin{align*}
\text{\(d₁₃, d₃₁₃, d₁₅, d₁₅₁₄, d₁₂₄\)}
\end{align*}
| S₁₃, S₁₅ |
|      |             |                            |         | \[
\begin{align*}
\text{\(0.640, 0.550, 0.590, 0.620, 0.490, 0.50\)}
\end{align*}
| |
| 4    | f₅         | DFI₁₁, DFI₁₃, DFI₁₄, DFI₁₆, DFI₁₈ | AE-FCM | \[
\begin{align*}
\text{\(a₁, a₁₃, a₁₅, a₁₇, a₁₉, a₁₁₂, a₁₄, \ldots, a₁₆\)}
\end{align*}
| S₅ |
|      |             |                            | 1-D CNN | \[
\begin{align*}
\text{\(a₁₉, a₁₉₂, a₁₉₄, a₁₉₆, a₁₆₇\)}
\end{align*}
| S₅, S₁₆ |
|      |             |                            | DTW     | \[
\begin{align*}
\text{\(d₁₃, d₂₈, d₂₁₃, d₃₁₄\)}
\end{align*}
| S₅ |
|      |             |                            |         | \[
\begin{align*}
\text{\(0.116, 0.68, 0.41, 0.42\)}
\end{align*}
| |
under various conditions. Furthermore, the on-site recorded data has confirmed the effectiveness and feasibility of the proposed method.

ACKNOWLEDGEMENTS
This work was sponsored by the National Natural Science Foundation of China under the project of Research of Flexible and Adaptive Arc-suppression Method for Single-phase Grounding Fault in Distribution Networks. (Project number: 51677030).

REFERENCES
1. Wang, X., et al.: Location of single phase to ground faults in distribution networks based on synchronous transients energy analysis. IEEE Trans. Smart Grid. 11(1), 774–785 (2019)
2. Griffel, D., et al.: A new deal for safety and quality on Mv networks. IEEE Trans. Power Delivery. 12(4), 1428–1433 (1997)
3. Chen, H., et al.: Intelligent transient overvoltages location in distribution systems using wavelet packet decomposition and general regression neural networks. IEEE Trans. Ind. Inf. 12(5), 1726–1735 (2016)
4. Topolaneck, D., et al.: Earth fault location based on evaluation of voltage sag at secondary side of medium Voltage/Low voltage transformers. IET Gener., Transm. Distrib. 9(14), 2069–2077 (2015)
5. Elsadd, M.A., et al.: Incorporating earth fault location in management-control scheme for distribution networks. IET Gener., Transm. Distrib. 10(10), 2389–2398 (2016)
6. Zewen, L., et al.: Power grid fault traveling wave network protection scheme. Int. J. Electr. Power Energy Syst. 33(4), 875–879 (2011)
7. He, Z., et al.: Improved fault-location system for railway distribution system using superimposed signal. IEEE Trans. Power Delivery. 25(3), 1899–1911 (2010)
8. Liao, Y., Generalized fault-location methods for overhead electric distribution systems. IEEE Trans. Power Delivery. 26(1), 53–64 (2010)
9. Jia, K., et al.: A new single-ended fault-location scheme for utilization in an integrated power system. IEEE Trans. Power Delivery. 28(1), 38–46 (2012)
10. Kezunovic, M.: Smart fault location for smart grids. IEEE Trans. Smart Grid. 2(1), 11–22 (2011)
11. Teng, J.-H., et al.: Automatic and fast faulted line-section location method for distribution systems based on fault indicators. IEEE Trans. Power Syst. 29(4), 1653–1662 (2014)
12. Zanjani, M.G.M., et al.: Application of pmus for adaptive protection of overcurrent relays in microgrids. IET Gener. Transm. Distrib. 12(18), 9 (2018)
13. Liu, Y., et al.: D-PMU based applications for emerging active distribution systems: A review. Electr. Power Syst. Res. 179, 106063 (2020)
14. Li, B., et al.: Optimal placement of fault indicator and sectionalizing switch in distribution networks. IEEE Access. 8, 17619–17631 (2020)
15. Horowitz, S., et al.: Adaptive transmission system relaying. IEEE Trans. Power Delivery. 3(4), 1436–1445 (1988)
16. Guo, M.F., et al.: Fault section location based on fuzzy information fusion for Nes. Power Syst. Prot. Control. 04(010), 65–70 (2012)
17. Goudarzi, M., et al.: Improved fault location algorithm for radial distribution systems with discrete and continuous wavelet analysis. Int. J. Electr. Power Energy Syst. 67, 423–430 (2015)
18. Liu, P., Huang, C.: A fault section location method for small-current grounding fault in distribution network based on dynamic time warping distance. Power Syst. Technol. 40(3), 952–957 (2016)
19. Yu, K., et al.: A fault section location technique for distribution systems based on wide-area fault information fuzzy clustering analysis. Paper presented at the 2014 International Conference on Power System Technology, 2014
20. Deng, X., et al.: Fault location in loop distribution network using SVM technology. Int. J. Electr. Power Energy Syst. 65, 254–261 (2015)
21. Farshad, M., Saleh, J.: Accurate single-phase fault-location method for transmission lines based on k-nearest neighbor algorithm using one-end voltage. IEEE Trans. Power Delivery. 27(4), 2360–2367 (2012)
22. Guo, M.F., et al.: Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems. IEEE Sens. J. 18(3), 1291–1300 (2018)
23. Wang, Y., et al.: Faulty feeder detection of single phase-earth fault using grey relation degree in resonant grounding system. IEEE Trans. Power Delivery. 32(1), 55–61 (2017)
24. Wang, P., et al.: Fault location method in resonant grounded networks based on distributed modulation and compensation adjustment. IEEE Trans. Power Delivery. 34(5), 1938–1947 (2019)
25. Guo, M.F., Yang, N.C.: Features-clustering-based earth fault detection using singular-value decomposition and fuzzy c-means in resonant grounding distribution systems. Int. J. Electr. Power Energy Syst. 93, 97–108 (2017)
26. Ma, S., et al.: High-voltage circuit breaker fault diagnosis using a hybrid feature transformation approach based on random forest and stacked autoencoder. IEEE Trans. Ind. Electron. 66(12), 9777–9788 (2018)
27. Du, Y., et al.: Single line-to-ground faulted line detection of distribution systems with resonant grounding based on feature fusion framework. IEEE Trans. Power Delivery. 34, (4), 1766–1775 (2019)

How to cite this article: Gao J-H, Guo M-F, Shao X, Chen D-Y. Feature-clustering-based Single-line-to-ground fault section location using auto-encoder and fuzzy C-means clustering in Resonant Grounding Distribution Systems. IET Gener Transm Distrib. 2021;15:938–949. https://doi.org/10.1049/gtd2.12070