IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 June 2023, accepted 20 June 2023, date of publication 23 June 2023, date of current version 28 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3288852

RESEARCH ARTICLE

Fault Classification in Distribution Systems Using Deep Learning With Data Preprocessing Methods Based on Fast Dynamic Time Warping and Short-Time Fourier Transform

NIEN-CHE YANG, (Member, IEEE), AND JEN-MING YANG
Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei 10607, Taiwan

Corresponding author: Nien-Che Yang (ncyang@mail.ntust.edu.tw)

This work was supported in part by the National Science and Technology Council (NSTC) in Taiwan, in part by NSTC under Grant NSTC 111-2622-8-011-014-SB, and in part by the Delta Electronics (DELTA)–National Taiwan University of Science of Technology (NTUST) Joint Research Center.

ABSTRACT Traditional fault classification methods typically rely on manual feature extraction and the application of machine-learning algorithms. However, these approaches encounter difficulties when extracting features and handling large-scale datasets. This study proposes a data preprocessing method for accurately detecting various types of short-circuit faults in power systems, which can lead to more effective power repair and maintenance processes. The proposed method involves converting the measured voltage and current signals into time and frequency domains using the short-time Fourier transform (STFT) to produce a time-frequency energy map. A convolutional neural network (CNN) is subsequently trained and tested to classify the short-circuit faults. However, overfitting may occur during the CNN training process owing to the large volume of data with similar features. To address this issue, this study proposes a data reduction method based on the fast dynamic time warping (Fast-DTW) algorithm, which compares waveform features and eliminates highly similar data regarded as redundant data from the dataset. The simulation results show that the proposed method can improve the model training performance and its adaptability to different power system topologies, as tested in two simulation environments: power systems computer-aided design (PSCAD)/electromagnetic transients, including DC (EMTDC), and the real-time digital simulator (RTDS). The STFT transformation is implemented in MATLAB. The simulation results demonstrate that the proposed method reduces redundant data by 40.2%, while decreasing the model training time. Consequently, the overall accuracy, precision, recall and F1 score of the fault classification reaches 99.37%, 99.36%, 99.35% and 99.35%, confirming the effectiveness of the proposed method for fault classification.

INDEX TERMS Convolutional neural network (CNN), fast dynamic time warping (Fast-DTW), fault classification, power systems computer-aided design/electromagnetic transients including DC (PSCAD/EMTDC), power distribution system, real-time digital simulator (RTDS), short-time Fourier transform (STFT), time-frequency analysis.

I. INTRODUCTION

Driven by the increasing demand for electricity in modern society, ensuring a consistent supply of electricity and energy by improving grid resilience and supply quality has become a critical issue in national energy strategies. In a complex electrical environment, short-circuit faults in power distribution systems may occur owing to various factors such as extreme weather, human errors, and environmental factors like flora and fauna. Faults with high fault currents can harm

The associate editor coordinating the review of this manuscript and approving it for publication was Ehab Elsayed Elattar.

residential and industrial customers, resulting in significant financial losses. Therefore, efficient and accurate detection and resolution of short-circuit faults have gained significant attention.

Efficient and precise fault detection is crucial for ensuring timely repairs. Recent research has demonstrated that load prediction or pattern recognition of fault signals can enable early-stage detection of abnormal nonlinear behaviors in power systems [1], [2]. However, it is challenging to detect short-circuit faults in power distribution systems with arc-suppression coils. The arc-suppression coil can restrict grounding fault currents to maintain the normal operation of the power system during a single-phase-to-ground fault, thereby making the detection of short-circuit faults difficult [3].

Detection of high-resistance ground faults can be challenging owing to their negligible fault current [4], [5]. To overcome this issue, instantaneous zero-sequence signal characteristics can be used for anomaly detection in transmission lines and for short-circuit fault detection. Furthermore, precise discrimination and classification, in addition to fast fault detection, are essential for balanced or unbalanced faults that may occur in the power system [6], [7], [8]. The use of deep learning principles and techniques has become widely popular in fault classification. Among the various methods, the convolutional neural network (CNN) is one of the most commonly employed. In power system fault classification, CNN is highly favored for its exceptional capability in extracting features and classifying data [9], [10], [11], [12].

To enhance recognition precision, effective data preprocessing is crucial for extracting signal features before training and testing neural networks. The wavelet transform (WT) is a commonly used method for data preprocessing. Specifically, the discrete wavelet transform (DWT) decomposes a signal into several frequency bands from which the signal features are extracted. Furthermore, the signal acquisition and transmission may be subject to noise interference. Filtering the higher frequency band of the signal can mitigate the effects of signal noise in subsequent analyses [13], [14], [15]. Time-frequency analysis is also frequently used in deep learning as a data preprocessing technique. The continuous wavelet transform (CWT) can convert a time-domain signal into a time-frequency domain signal and extract the frequency characteristics of the signal in the local time interval for subsequent analysis [16]. However, the WT for data preprocessing is limited by the wavelet functions based on different signal characteristics and contexts. To overcome this limitation, the Hilbert-Huang transform (HHT) is used with empirical mode decomposition (EMD) to decompose the signal into intrinsic mode functions (IMFs).

Furthermore, the HHT bandpass filters have been used to obtain components with different frequency bands [17], [18]. To avoid over-decomposition and feature distortion, the signals must be appropriately evaluated and adjusted based on different scenarios. One widely used method for analyzing nonstationary signals is the short-time Fourier transform (STFT)-based time-frequency analysis. This method divides a long-time signal into short time intervals and applies discrete Fourier transform (DFT) to each interval to quickly obtain the signal spectrum. Studies have demonstrated that the STFT is more efficient and reliable than the CWT and HHT [19], [20], [21].

In deep learning, data preprocessing using time-frequency transformations enables neural networks to extract data features effectively, generating highly accurate identification outcomes. Their applications have proven valuable in various research domains. For instance, Mel-spectrograms, Gammatone spectrograms, and CWT generate spectrograms for speech processing. These spectrograms are then utilized in subsequent tasks, such as detecting speech deficits in cochlear implant users and recognizing phoneme classes. To achieve these goals, CNN and recurrent neural networks with convolutional layers have been employed [22]. The effective classification of noisy non-stationary time-series signals has been achieved by integrating various time-frequency representation methods from Cohen's class with deep learning techniques, improving the detection of non-stationary gravitational wave signals in noisy environments [23].

To achieve accurate recognition, deep learning requires a dataset with an extensive range of features. However, large datasets can lead to problems concerning high data similarity. A high similarity can increase the model training time and lead to overfitting, affecting the training results. Therefore, efficient data reduction is crucial to enhance the performance of neural networks [24]. During data reduction, it is essential to retain valuable and rich features that can be applied in a wide range of contexts [25]. To measure the similarity between time series, the dynamic time warping (DTW) algorithm is commonly used to identify similar time series by comparison [26], [27].

The method proposed in this study combines the Fast-DTW and STFT for data preprocessing. First, a dataset comprising a significant amount of data is generated from signals in various fault scenarios. The Fast-DTW is subsequently utilized to measure the similarity between the data in the dataset. This step eliminates excessively similar data, achieving data reduction while preserving the feature richness of the dataset. Finally, the dataset is trained and tested using CNN for fault classification. The experimental results demonstrate that the proposed method can significantly reduce the model training time and improve the model fitness. Moreover, the proposed method performs remarkably well across different system architectures and scenarios with varying load levels in power distribution systems.

This study presents several contributions and innovations. First, a data preprocessing method is introduced to effectively extract valuable features while reducing data volume. Second, the challenge of overfitting during CNN training is overcome by eliminating redundant data with similar features. Finally, the proposed method is rigorously tested in

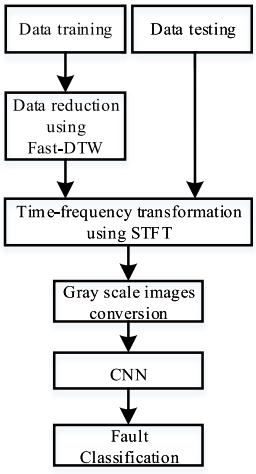

FIGURE 1. Block diagram of the proposed fault classification method in distribution systems.

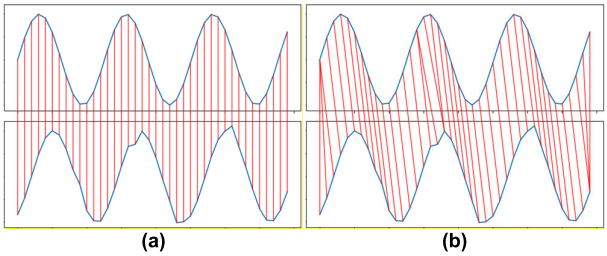

FIGURE 2. Wave comparison with a) ED and b) DTW.

two different environments. The fault classification results demonstrate that the efficiency and reliability of the proposed method in fault classification with an accuracy, precision, recall and F1 score of up to 99.37%, 99.36%, 99.35% and 99.35%, respectively.

The remainder of this paper is organized as follows: Section II provides a comprehensive overview of the theoretical background and outlines the underlying principles. Section III introduces the topology and simulation environment of the sample power system. Section IV discusses simulation results obtained using the proposed method. Section V validates the effectiveness and reliability of the proposed neural network model under different scenarios. Finally, Section VI summarizes the research findings and concludes the paper.

II. THEORETICAL BACKGROUND

This study proposes a fault classification technique based on deep learning. To achieve precise fault classification, fault signals are gathered when a fault occurs in the distribution system to construct a fault dataset for data preprocessing and deep learning. FIGURE 1 illustrates the proposed method for fault classification in the power distribution system.

A. FAST-DTW ALGORITHM

To achieve data reduction and eliminate a massive quantity of data with similar characteristics from the dataset, the similarities between the time-domain signals measured under different fault scenarios are evaluated. Although the Euclidean distance (ED) is commonly used to compare time series, the DTW algorithm is more effective and reliable in reflecting time-series similarity by considering the offset on the time axis and the distance at each time. DTW can also be used more comprehensively and flexibly for time series similarity comparisons. FIGURE 2 illustrates the comparison between ED and DTW. The ED algorithm highlights minor amplitude differences and displacements, resulting in a lower waveform similarity, whereas the DTW algorithm demonstrates a higher waveform similarity despite these differences. Therefore, DTW can more accurately identify feature similarities between different data in applications such as speech recognition, handwritten character recognition, and stock price prediction.

The Fast-DTW algorithm was proposed in [28] to improve the efficiency of time series comparisons. This algorithm optimizes the computational efficiency of the traditional DTW by introducing three additional steps to shorten the computation time. The steps are as follows:

1) Coarsening: Coarsening involves sampling small segments with representative features from the original time series. This transformation simplifies the data representation by reducing the number of dimensions, leading to a reduction in the sequence length and computational complexity while retaining the essential features of the original data. Thus, the DTW performs better in terms of computation speed.
2) Projection: The distance matrix obtained from the coarsening step undergoes DTW computation to identify the minimum warping path, which starts subsequent refinement.
3) Refinement: During the refinement step, the minimum warping path obtained from the projection step is refined to a higher dimensional time-series data. Subsequently, the approximate warping path is interpolated back to the original series space to obtain the exact warping path.

The Fast-DTW algorithm has been shown to rapidly compute time-series similarities and enhance computational efficiency. Furthermore, it requires less memory storage for computation than the traditional DTW, making it a valuable tool for comparing two large time series.

The Fast-DTW algorithm accelerates the computation of the DTW algorithm by implementing specific optimization steps. The DTW algorithm creates a two-dimensional distance matrix that maps two time series $\boldsymbol{X}$ and $\boldsymbol{Y}$, of lengths I and J, respectively, as shown in (1) and (2).

$$\boldsymbol{X} = x_1, x_2, \cdots, x_i, \cdots, x_I \tag{1}$$

$$\boldsymbol{Y} = y_1, y_2, \cdots, y_j, \cdots, y_J \tag{2}$$

where x_i and y_j denote the data in the two series, respectively. A distance matrix of dimension $I \times J$ can be generated based on these two series, as shown in FIGURE 3.

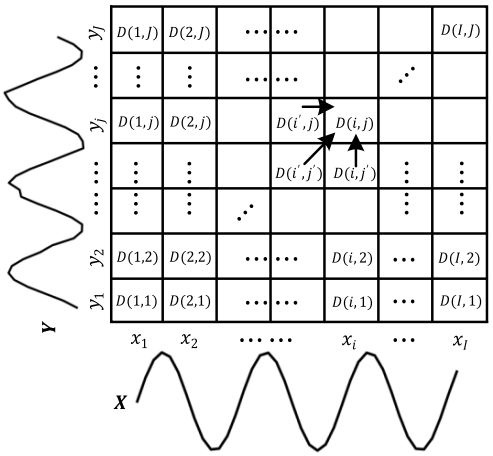

FIGURE 3. Distance matrix of DTW.

The (i, j) entries in the distance matrix denote the warping distance between the corresponding data, x_i and y_j, which is calculated using (3).

$$D(i,j) = dist\left(x_i, y_j\right) + \min\left\{D\left(i',j\right), D\left(i,j'\right), D\left(i',j'\right)\right\} \tag{3}$$

where $i' = i - 1$, $j' = j - 1$, and $dist(x_i, y_j)$ denotes the ED between x_i and y_j, as shown in (4).

$$dist(x_i, y_j) = \left\|x_i - y_j\right\| \tag{4}$$

All entries in the distance matrix can be calculated using (3), and the distance matrix can then be constructed. The element $D(I, J)$ represents the minimum warping distance between sequences $\boldsymbol{X}$ and $\boldsymbol{Y}$, as shown in (5).

$$\mathrm{WD} = D(I, J) \tag{5}$$

where WD is the minimum warping distance between any two time series via Fast-DTW.

To assess the similarity of feature signals, the voltage and current signals are processed using Fast-DTW to determine the WD. The resulting WD values are averaged according to the signal lengths, yielding the average warping distance (AWD), which is used in this study to evaluate the feature similarity between two sets of data. The AWD is calculated using (6).

$$\mathrm{AWD} = \frac{\mathrm{WD}}{L_{sig}}, sig \in \{\boldsymbol{u}_a, \boldsymbol{u}_b, \boldsymbol{u}_c, \boldsymbol{i}_a, \boldsymbol{i}_b, \boldsymbol{i}_c\} \tag{6}$$

where L_{sig} is the length of signal.

FIGURE 4 illustrates the data-reduction process accomplished using the Fast-DTW algorithm. This process can be organized into the following steps:

Step 1. Data collection: Simulate various fault scenarios in the power systems computer-aided design/electromagnetic transients, including DC (PSCAD/EMTDC) environment. Subsequently, collect the three-phase voltage signals ($\boldsymbol{u}_a$, $\boldsymbol{u}_{\mathrm{b}}$, $\boldsymbol{u}_{\mathrm{c}}$) and three-phase current signals ($\boldsymbol{i}_a, \boldsymbol{i}_b, \boldsymbol{i}_{\mathrm{c}}$) from each scenario.

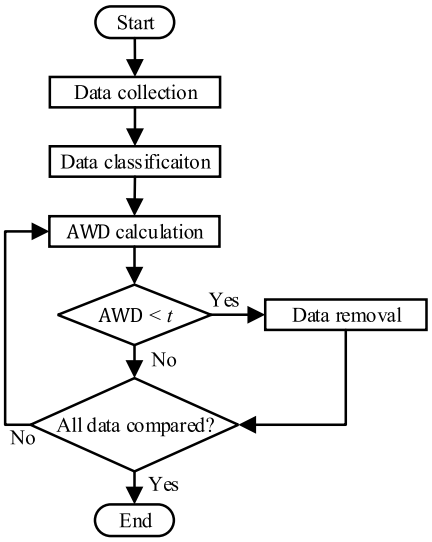

FIGURE 4. Flow chart of the proposed Fast-DTW-based data reduction method.

Step 2. Data classification: Classify all data into different fault types, including three-phase fault (ABC), single-phase-to-ground fault (AG, BG, CG), phase-to-phase fault (AB, BC, CA), and two-phases-to-ground fault (ABG, BCG, CAG).

Step 3. AWD calculation: Calculate the AWD by applying Fast-DTW separately to the three-phase voltage and three-phase current signals from two data in the dataset.

Step 4. Signal similarity confirmation: Evaluate the AWD of the three-phase voltage and current signals from the two compared data. If the AWD is less than the threshold t, one of the data points is considered redundant owing to its high similarity with the other and is removed from the dataset. The remaining data is used for a comprehensive dataset comparison until the process is complete.

Step 5. Comparison completeness confirmation: If all data in the dataset have not yet been compared, repeat Steps 3-4 until all data in the dataset have been compared.

In this study, a threshold value of $t = 0.01$ is used, indicating that the AWD of each sampling point between the two signals is less than 0.01. If two signals possess nearly identical characteristics in the time domain, one of them is removed from the dataset. The effectiveness of the proposed data preprocessing method is validated via simulations in Section IV.

B. STFT BASED TIME-FREQUENCY TRANSFORM

When faults occur, the voltage and current signals may experience changes in both amplitude and frequency. Using time-frequency analysis makes it possible to analyze signals in both the time and frequency domains, facilitating the effective identification of signal characteristics over various time intervals and frequency bands. In this study, the STFT is

used to convert the filtered time-domain signals into time-frequency-domain spectral maps, which then serve as input data for subsequent CNN analysis.

The STFT is a signal processing technique that examines the frequency structure of nonstationary signals. It decomposes a signal using a short-time window function and performs a DFT on each window to obtain the signal spectrum within the time window. The rectangular window function in the time domain is used for the STFT operation. All values in the window are treated equally without any weight adjustment. This window function is the most frequently used function [29], [30], which is given by (7).

$$w(n) = \begin{cases} 1, & 0 \le n \le L_w \\ 0, & else \end{cases} \tag{7}$$

where the integer n signifies the discrete time series of the signal representing the time point of the signal sampling, $w(n)$ is the rectangular window function, and L_w is the window length.

Suppose the signal function is denoted as $s(n)$, and the signal is shifted by $w(n)$ at the interval of R sampling points; the window function has an overlap between two adjacent periods, and the length can be denoted as $L_{overlap}$, as shown in (8).

$$L_{overlap} = L_w - R \tag{8}$$

By sliding the window function $w(n)$ over the signal $s(n)$ and subsequently using the DFT, a spectral signal corresponding to a specific time interval can be obtained. The DFT values obtained at each time interval can be accumulated into a time-frequency matrix for the signal $s(n)$. The resulting matrix contains C rows. The C can be determined as follows:

$$C = \left\lfloor \frac{L_s - L_{overlap}}{R} \right\rfloor \tag{9}$$

where L_s represents the length of the signal $s(n)$ and the $\lfloor\ \rfloor$ symbol denotes the floor function, which rounds off a real number to the nearest integer. The number of columns in the time-frequency matrix corresponds to the frequency component of the DFT.

The STFT time-frequency matrix $\boldsymbol{T}_s(f)$ with different frequency components f is given by (10), where f represents the frequency component obtained from the DFT of the window function. Furthermore, the cth row element of the STFT matrix $T_c(f)$, presented in (11), corresponds to the DFT of the window function $w(n)$ centered on time cR.

$$\boldsymbol{T}_s(f) = [T_1(f)\, T_2(f) \cdots T_c(f) \cdots T_C(f)] \tag{10}$$

$$T_c(f) = \sum_{n=-\infty}^{\infty} s(n) w(n - cR) e^{-j2\pi fn} \tag{11}$$

where $\{c \in \mathbb{Z} | 1 \le c \le C\}$.

The signal can be transformed into a spectrum by constructing the STFT time-frequency energy matrix, as shown in FIGURE 5. The measured waveform of the phase-a voltage signal during a phase-a short-circuit fault in one of the feeders

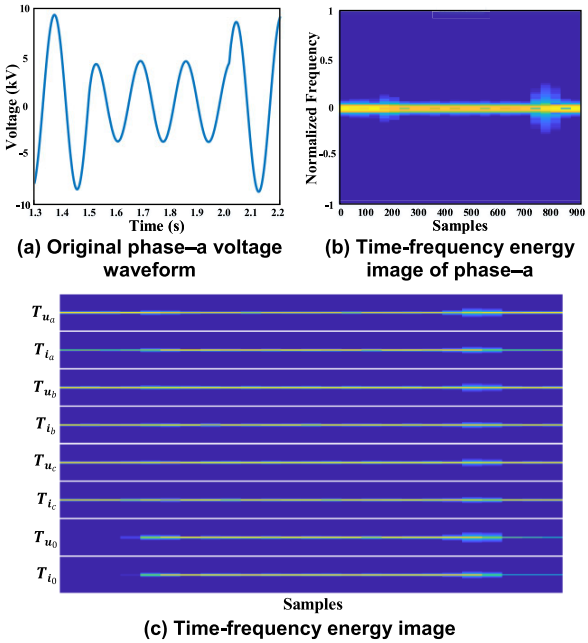

FIGURE 5. STFT-based time-frequency transform.

is illustrated in FIGURE 5 (a), whereas FIGURE 5 (b) displays the time-frequency energy map obtained by the STFT time-frequency transformation. In addition to collecting the three-phase voltage and current signals, Park's transformation is used for the voltage and current signals to extract the zero-sequence component caused by short-circuit faults. The zero-sequence components of the three-phase voltage and current signals are expressed in (12) and (13).

$$\boldsymbol{u}_0 = \frac{1}{3}(\boldsymbol{u}_\mathrm{a} + \boldsymbol{u}_\mathrm{b} + \boldsymbol{u}_\mathrm{c}) \tag{12}$$

$$\boldsymbol{i}_0 = \frac{1}{3}(\boldsymbol{i}_\mathrm{a} + \boldsymbol{i}_\mathrm{b} + \boldsymbol{i}_\mathrm{c}) \tag{13}$$

where $\boldsymbol{u}_0$ is the zero-sequence component of the voltage signal and $\boldsymbol{i}_0$ is the zero-sequence component of the current signal.

This study collects signals, including voltage and current signals, from all three phases and the zero-sequence components to ensure the precise classification of various faults in different phases of transmission lines. To obtain the time-frequency energy matrix $\boldsymbol{E}$, all the signals are transformed into the STFT time-frequency matrix, as shown in (14).

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{T}_{u_a} \\ \boldsymbol{T}_{i_a} \\ \boldsymbol{T}_{u_b} \\ \boldsymbol{T}_{i_b} \\ \boldsymbol{T}_{u_c} \\ \boldsymbol{T}_{i_c} \\ \boldsymbol{T}_{u_0} \\ \boldsymbol{T}_{i_0} \end{bmatrix} \tag{14}$$

where $\boldsymbol{T}_s$ represents the STFT time-frequency energy matrix for signal s. $s \in \{\boldsymbol{u}_a, \boldsymbol{u}_\mathrm{b}, \boldsymbol{u}_\mathrm{c}, \boldsymbol{u}_0, \boldsymbol{i}_\mathrm{a}, \boldsymbol{i}_\mathrm{b}, \boldsymbol{i}_\mathrm{c}, \boldsymbol{i}_0\}$. By converting the $\boldsymbol{E}$ matrix into a spectral map and scaling it, a time-frequency

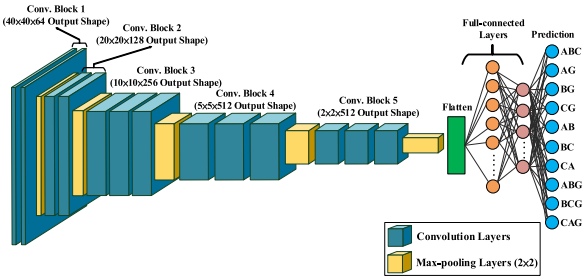

FIGURE 6. Structure of VGG-16 model.

energy map, as shown in FIGURE 5 (c), can be obtained, which can be used as input data for subsequent training and testing of the neural network.

C. CONVOLUTIONAL NEURAL NETWORK (CNN)

CNN is a popular deep learning model, particularly for image recognition, object detection, image segmentation, and speech recognition. The basic principle of CNN involves using a convolution kernel that slides over an image to extract distinctive features. These features undergo a dimensionality reduction process such as pooling. Subsequently, a fully connected layer maps these features to the output layer, which generates prediction outcomes. In this study, we adopt the VGG-16 model, which is a classical CNN model equipped with a deep network layer. VGG-16 has succeeded in image classification and object detection tasks owing to its enhanced ability to extract various image features, resulting in superior classification accuracy. The architecture of the VGG-16 model is illustrated in FIGURE 6.

In CNN-based fault classification, STFT time-frequency energy maps can be used as input data. Using the visualization process, grayscale images with dimensions of $40 \times 40 \times 1$ are inputted into the CNN. The VGG-16 model used in this study employed a 3×3 convolution kernel, and the convolution process is expressed in (15).

$$\boldsymbol{M}_h^l = Act(\sum_{h=1}^{H} \boldsymbol{M}_h^{l-1} \cdot \boldsymbol{K}^l + \boldsymbol{B}_h^l) \tag{15}$$

In this study, the term "step" refers to the number of implement of a convolution kernel or pooling window on an image during the convolution process. The convolution process described in (15) includes several terms, where h denotes the step, H is the number of steps required to convolve the image, l stands for the number of layers in the network, $\boldsymbol{M}_h^l$ denotes the h-th feature map in the l-th layer, $\boldsymbol{M}_h^{l-1}$ represents the h-th feature map in the $(l-1)$-th layer, $\boldsymbol{K}^l$ is the convolution kernel in the l-th layer, and $\boldsymbol{B}_h^l$ represents the bias function. The activation function used to enhance the nonlinearity of the neural network is denoted by $Act()$. In this study, the rectified linear unit (ReLU) function represented by (16) is used as the activation function.

$$Act\ (v) = \max(0, v) \tag{16}$$

where v is an arbitrary real number.

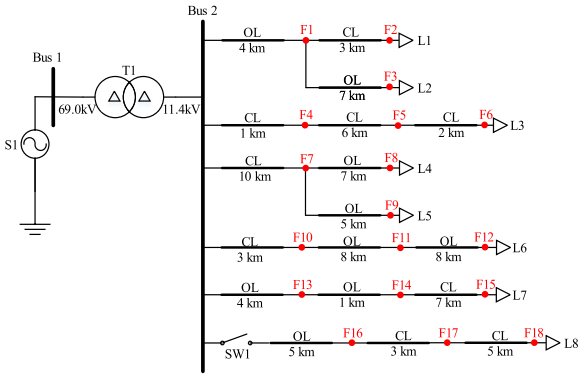

FIGURE 7. System structure in PSCAD/EMTDC environment.

After the convolution process, the pooling layer is used to extract significant feature information and concurrently rescue the image size, thereby decreasing the number of parameters and computing requirements. In the VGG-16 model, the pooling layer employs a 2×2 pooling window size and max pooling is used to accomplish the pooling process, as shown in (17).

$$\boldsymbol{M}_h^l = Act(\max\{\boldsymbol{M}_h^{l-1}\} + \boldsymbol{B}_h^l) \tag{17}$$

After passing through a series of convolutional and pooling layers, the input data is transformed from a three-dimensional vector to a one-dimensional vector. The resultant data is subsequently fed to a fully connected layer that classifies the features extracted from the preceding convolutional and pooling layers, resulting in the final prediction output.

III. CASE STUDY

In this section, distribution systems constructed in PSCAD/EMTDC and real-time digital simulator (RTDS) are used to simulate short-circuit faults and collect fault signals under various fault scenarios to form the dataset for data processing and fault classification.

A. SIMULATION MODEL

A test system was developed in the PSCAD/EMTDC environment to simulate short-circuit faults in a power system, as shown in FIGURE 7. The system operates at a frequency of 60 Hz. Moreover, it consists of an AC voltage source, S1, providing a 69-kV AC root-mean-square (RMS) voltage, and a two-winding transformer, T1, with a capacity of 60 MVA and a turn ratio of 69 kV /11.4 kV, connected to Bus 1 and Bus 2. This power system includes two different types of transmission lines: an overhead line (OL) and a cable line (CL) with zero-sequence impedance (R_0, X_{C0}, X_{L0}) and positive-sequence impedance (R_1, X_{C1}, X_{L1}), as listed in TABLE 1. The possible locations for short-circuit faults in the various scenarios are F1 to F18. The load demand of each feeder, denoted by L1 to L8, is $0.75 + j0.25$ (MVA).

B. VERIFICATION SYSTEM

When simulating power systems, it is imperative to ensure that the simulation parameters accurately reflect actual power

TABLE 1. Parameters of transmission line in power system.

Item	OL	CL
R_0	0.275 (Ω/km)	2.7 (Ω/km)
X_{C0}	491.22 (MΩ ·m)	9.4735 (MΩ ·m)
X_{L0}	1.734 (Ω/km)	0.038415 (Ω/km)
R_1	0.125 (Ω/km)	0.27 (Ω/km)
X_{C1}	276.31 (MΩ ·m)	0.09613 (MΩ ·m)
X_{L1}	0.4901 (Ω/km)	7.824 (Ω/km)

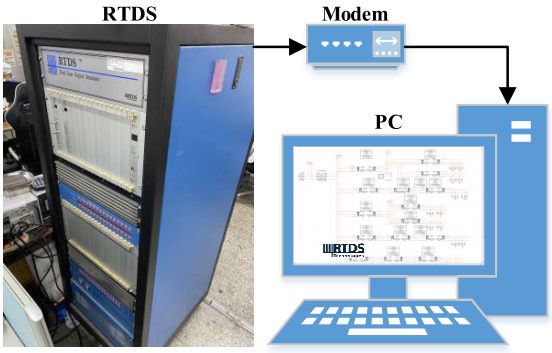

FIGURE 8. RTDS environmental facility architecture.

systems. To validate the reliability of the electrical signals sampled in the PSCAD/EMTDC simulation environment, we used an RTDS to set up the power system in a real-time environment and test the accuracy and reliability of the simulated data. The RTDS is a high-performance computer system platform designed for power system simulations, which can simulate large power systems in real time under various scenarios. Its real-time operation and control capabilities make it ideal for studying power system stability, control, protection, and automation. The RTDS provides a reliable benchmark for real system applications owing to its robust real-time computation capability. Therefore, in this study, the RTDS is considered as a practical power system condition to test the feasibility of the proposed method. The signal data obtained from the real-time simulation platform are shown in FIGURE 8. Using the RTDS environment is regarded as field measurements to verify the performance of the proposed method.

IV. SIMULATION RESULT

In this section, the performance of the proposed method for fault classification is evaluated. The efficiency and generalization ability of the CNN model are assessed to verify the effectiveness of the proposed method.

A. DATASET FOR TRAINING AND TESTING

After obtaining the voltage and current signals from the short-circuit faults, Fast-DTW was used to eliminate redundant data that exhibited similar characteristics in the dataset. Subsequently, a time-frequency energy map was generated via the STFT-based time-frequency transformation, which served as the input to the CNN.

TABLE 2. Settings of fault cases for training dataset.

Fault type	Fault resistance (ohm)	Fault phase angle (degree)	Data size
ABC	0, 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200	0, 30, 60, 90, 120, 150	1800
AG, BG, CG			5400
AB, BC, CA			5400
ABG, BCG, CAG			5400
	Total		18000

TABLE 3. Variation in the size of training data.

Fault type	Original training data size	New training data size	Data reduction rate
ABC	1800	695	38.6%
AG	1800	1486	82.5%
BG	1800	1350	75.0%
CG	1800	1327	73.7%
AB	1800	683	37.9%
BC	1800	596	33.1%
CA	1800	595	33.1%
ABG	1800	1319	73.3%
BCG	1800	1406	78.1%
CAG	1800	1309	72.7%
Total	18000	10766	59.8%

TABLE 4. Settings of fault cases for testing dataset.

Fault type	Fault resistance (ohm)	Fault phase angle (degree)	Data size
ABC	Random [0, 200]	Random [0, 180]	200
AG, BG, CG			600
AB, BC, CA			600
ABG, BCG, CAG			600
	Total		2000

A power distribution system was constructed in the PSCAD/EMTDC environment to generate the initial training dataset. Various fault scenarios, including different fault types, short-circuit fault resistances, fault phase angles, and fault locations, were considered. The specific parameter settings and dataset sizes are listed in TABLE 2.

TABLE 3 lists the data volume and reduction rate of the training data after applying the Fast-DTW algorithm to the original training dataset. The data reduction rate was calculated using (18).

$$\text{Data reduction rate} = \frac{\text{Number of new training data}}{\text{Number of original training data}} \times 100\% \quad (18)$$

The results in TABLE 3 demonstrate that the total data size in the new training dataset is reduced to 59.8% of that in the original dataset using the Fast-DTW algorithm.

In this study, the randomness of fault occurrences was considered to construct the testing dataset. Fault occurrence locations were randomly selected from F1 to F15. The fault

Confusion matrix		True class	
		Positive	Negative
Predicted class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

FIGURE 9. Confusion matrix.

impedances and fault phase angles were generated randomly within the ranges of [0, 200] and [0, 180], respectively. The specific parameters of the test set are listed in TABLE 4. During the construction of the training dataset, Fast-DTW was used for data reduction to mitigate redundant data. However, in the testing dataset, a Fast-DTW-based data reduction was not applied. This decision was based on the requirement to investigate the stochastic nature of fault occurrences and preserve variability in the testing data.

B. PERFORMANCE INDICATORS

In the analysis phase, various metrics are used to assess the effectiveness of the proposed method for model training. A confusion matrix is used to evaluate the predictive performance of the model, as shown in FIGURE 9. The confusion matrix provides a comprehensive view of the classification results and consists of the following key elements: True Positive (TP), True Negatives (TN), False Positive (FP), and False Negative (FN). TP denotes the number of positive samples accurately classified as positive by a classifier. TN represents the number of negative samples correctly classified as negative by the classifier. FP indicates the number of negative samples mistakenly classified as positive by a classifier. FN signifies the number of positive samples incorrectly classified as negative by the classifier.

Among numerous evaluation metrics, accuracy, precision, recall, and F1-score are widely used to assess the performance of deep learning models. These metrics were used to investigate the performance of the proposed method during model training.

1) Accuracy: Accuracy represents the overall correctness of the predictions, indicating the proportion of correctly classified samples, as expressed in (19).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \tag{19}$$

2) Precision: Precision measures the precision of positive predictions, indicating the proportion of correctly predicted positive samples among all positive predictions, as expressed in (20).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{20}$$

3) Recall: Recall, also known as the TP rate, quantifies the ability to correctly identify positive samples. It is used to evaluate the detection ability of TP samples, as expressed in (21).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{21}$$

4) F1 score: The F1 score is a combined metric that considers both precision and recall, providing a balanced measure of model performance, as expressed in (22).

$$\text{F1 score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{22}$$

TABLE 5. Execution time on model training.

Case	Execution time (s/epoch)
Training VGG-16 with D0	34
Training VGG-16 with D1	20

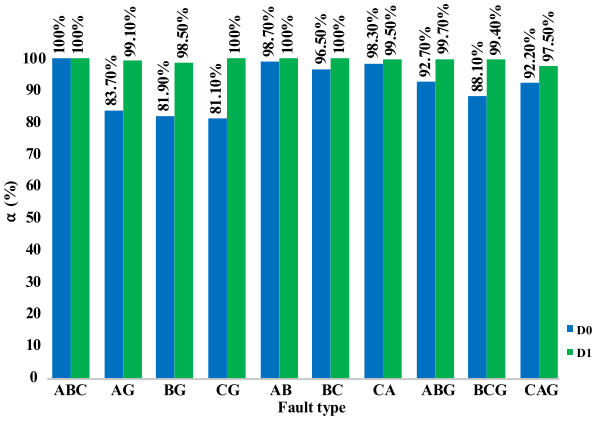

FIGURE 10. Accuracy comparison for various fault types.

C. TEST RESULT

In this study, the effectiveness of fault classification was verified using the datasets listed in TABLE 2 and TABLE 4 for training and testing the CNN-based model. During the training process of the CNN-based model, 10% of the training dataset was used as the validation dataset to evaluate the generalization capability and performance of the proposed method.

Two training datasets were used to evaluate the effectiveness of the proposed method. The first dataset, D0, listed in TABLE 2,was collected without data pre-processing. The second dataset, D1, was generated after using the proposed Fast DTW-based algorithm to reduce the dataset size. Both datasets were used to train and test the VGG-16 model. The results are depicted in FIGURE 10, where the test results for the proposed model trained with D0 and D1 are indicated by green and blue bars, respectively. FIGURE 10 shows that the D1 training dataset can achieve an accuracy of up to 97.5% for fault classification, outperforming the D0 training dataset without data reduction in all categories. The accuracy rate α

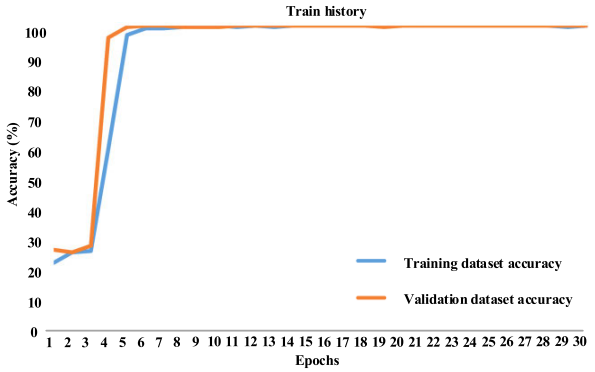

FIGURE 11. Training curve of CNN.

for each fault type is defined in (23).

$$\alpha = \frac{\text{Number of correct predicted data for each fault type}}{\text{Total number of data for each fault type}} \times 100\% \quad (23)$$

As shown in FIGURE 10, the CNN model trained with the D0 dataset performs poorly in predicting single-phase-to-ground and two-phase-to-ground faults. This may be attributed to overfitting caused by the abundance of data with similar characteristics in the D0 training dataset. Overfitting occurs when a machine-learning model learns the training data exceedingly well, leading to poor recognition performance on the testing data. In deep learning, an excessive amount of data with similar features may prevent the CNN model from overly relying on these features during training, resulting in the CNN model identifying particular patterns in the data and failing to generalize to unknown data. Therefore, when testing data are generated using random parameters, overfitting may cause the model to be less accurate in fault classification.

The D1 training dataset exhibits outstanding performance, achieving high accuracy for each fault type by eliminating redundant data with similar features. FIGURE 11 shows the accuracy of the D1 training and the validation dataset during the CNN model training procedure for each epoch. The accuracy of both datasets improves significantly as the CNN model is trained, indicating that the proposed method provides the CNN model with excellent generalization capability. Even for a testing dataset with new and unfamiliar data, the trained model using the proposed data preprocessing method can still effectively identify faults and achieve reliable fault classification.

The proposed data reduction method effectively reduces the dataset size, reducing the training and computation times. The CNN model was trained in a hardware environment with NVIDIA GeForce RTX-3050 GPU. The training times for the D0 and D1 training datasets are compared in TABLE 5. Data reduction significantly reduces the CNN model training time and improves the overall performance.

In addition to evaluating the accuracy for each fault type, generalization capability, and computational efficiency, performance indicators were used to comprehensively evaluate model performance. Furthermore, a comparative analysis was conducted using the model architectures based on [16] and [18] to validate the robustness and effectiveness of the neural network model used in this study. The corresponding performance indicators for each case are listed in TABLE 6.

TABLE 6. Performance indicators for comparison cases.

Case	Accuracy	Precision	Recall	F1 score
Reference [16]	78.89%	84.00%	79.20%	79.42%
Reference [18]	85.29%	86.31%	85.25%	85.25%
D0 dataset	91.32%	90.74%	88.95%	88.15%
D1 dataset (Proposed method)	**99.37%**	**99.36%**	**99.35%**	**99.35%**

TABLE 7. Settings of fault cases for testing dataset in RTDS environment.

Fault type	Fault resistance (ohm)	Fault location	Data size
ABC	0, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 50, 100, 150, 200	F1, F2, F4, F6, F7, F10, F13, F14	120
AG, BG, CG			360
AB, BC, CA			360
ABG, BCG, CAG			360
	Total		1200

Based on the results presented in TABLE 6, it can be confirmed that the CNN model employed in this study outperforms those reported in the existing literature in terms of accuracy, precision, recall, and F1 score. These evaluation indicators offer a comprehensive understanding of the model performance.

The results of fault classification indicate that the proposed method significantly enhances the accuracy of identifying short-circuit faults in power systems. Model overfitting was avoided using the proposed method. Moreover, the generalization ability of the CNN model is enhanced, allowing for more comprehensive fault identification in unpredictable fault scenarios.

V. VERIFICATION

In this section, the performance of the proposed method was evaluated for different system architectures and load levels. Different systems were constructed in the PSCAD/EMTDC environment to assess the capability and adaptability of the proposed method for various applications. Furthermore, tests were performed in the RTDS environment to examine the feasibility and reliability of the proposed method in real-world power distribution systems. The cases used for verification purposes are listed as follows:

Case 0: Original system shown in FIGURE 7
Case 1: Original system with an additional feeder
Case 2: Load level reduced to 50% of the original one
Case 3: Load level increased to 150% of the original one
Case 4: RTDS real-time data verification

The test cases listed above are introduced in subsequent sections. These cases are used to verify the reliability of the

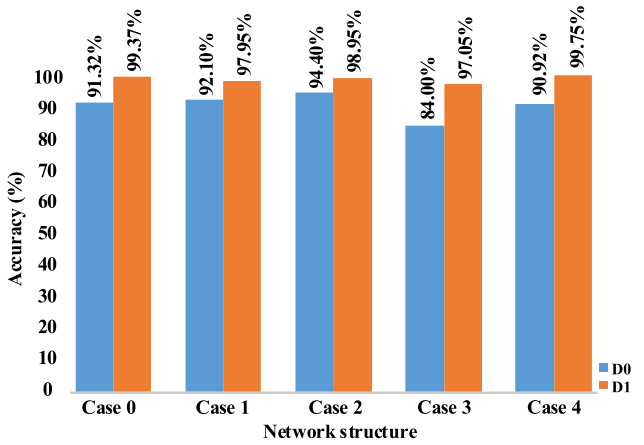

FIGURE 12. Performance verification for various scenarios.

proposed method and compare the effect of various scenarios with the D0 training set. The accuracy of the proposed method is depicted in FIGURE 12.

A. DIFFERENT SYSTEM ARCHITECTURE

The performance of the proposed data processing method was evaluated by extracting the features of fault signals under various system architectures. Specifically, a new feeder with loads was added to the original power system shown in FIGURE 7 by closing SW1, and the testing data were generated using this modified architecture. Only faults occurring at fault locations F1 to F15 were considered in the original test data. However, after closing SW1, the faults that occurred at F16, F17, and F18 were also considered. A new testing dataset was generated to evaluate the adaptability of the proposed method to various system architectures. The results shown in FIGURE 12 demonstrate that the proposed method can effectively identify fault-signal features with an accuracy rate of 97.95% even when an additional feeder is added to the power system, indicating its applicability to distribution systems with various architectures.

B. DIFFERENT LOAD LEVEL

The effectiveness of the proposed fault classification method at different load levels was examined to address potential load variations in power systems. The original power system was used as a benchmark, and the load demands along the feeders shown in FIGURE 7 were adjusted to 50% and 150% of the original loads. The accuracy of the proposed method was as high as 98.95% for Case 2 and 97.05% for Case 3, as shown in FIGURE 12. Based on the test results of the D0 training set, when the load increases, the current in the power system also increases. The features of sudden changes in the fault current become more obscure, making it more challenging for the training model that may have already been overfitted to correctly identify the faults. However, the proposed method can overcome the differences between various load levels to achieve excellent fault identification results.

C. RTDS REAL-TIME DATA VERIFICAITON

The effect of real-time data on the proposed method was also investigated to validate its applicability in real-power systems. The system architecture shown in FIGURE 7 was implemented in the RTDS environment, and a testing dataset of real-time data was generated according to TABLE 7.

The testing dataset generated in the real-time environment was evaluated in the D1 training dataset and achieved a high accuracy rate of 99.75%. The results indicate that the proposed method can be reliably applied to actual distribution systems. The power system simulation software can be used to build various fault scenarios for numerous simulations, allowing the rapid construction of a comprehensive dataset with minimal hardware requirements, thereby overcoming the challenges of operating and frequently accessing data in real power systems. The proposed data reduction method can effectively reduce the dataset size while preserving the richness of the features and achieving accurate fault classification in power distribution systems.

VI. CONCLUSION

In this study, the Fast-DTW was used to mitigate the issue of feature repetition in an extensive dataset while concurrently enhancing signal features via STFT time-frequency transformation. The transformed signals were converted into a time-frequency energy map, which was used as input data for the CNN model. Subsequently, the VGG-16 model was utilized for both feature extraction and classification to identify ten possible faults in the power distribution system. The test results demonstrate that with data preprocessing using the proposed method, the training efficiency and adaptability of the CNN model are enhanced. Finally, real-time simulations in RTDS corroborated the reliability and credibility of the proposed CNN model in real power systems, consequently providing a reference for improving the operation and maintenance of distribution systems. However, it is essential to acknowledge that the threshold of 0.01 used in this study to identify redundant data is yet to be validated as the optimal threshold for fault analysis in all scenarios. Further research is required to determine the optimal threshold and refine the criteria to obtain more accurate and effective fault analysis.

Based on the findings of this study, several perspectives should be considered for future studies. One possibility is to expand the application of the proposed method to address more complex scenarios in power systems, such as those involving harmonic pollution or ferromagnetic resonance. Advancing fault analysis in power systems can contribute to the development of more efficient and reliable operation and maintenance practices.

REFERENCES

[1] M. Ali, K. Kaur, M. Adnan, and S. Nisar, "Adaptive fuzzy controller based early detection and prevention of asymmetrical faults in power systems," *Control Eng. Pract.*, vol. 130, Jan. 2023, Art. no. 105380, doi: 10.1016/j.conengprac.2022.105380.

[2] M.-F. Guo, N.-C. Yang, and L.-X. You, "Wavelet-transform based early detection method for short-circuit faults in power distribution networks," *Int. J. Electr. Power Energy Syst.*, vol. 99, pp. 706–721, Jul. 2018, doi: 10.1016/j.ijepes.2018.01.013.
[3] M.-F. Guo and N.-C. Yang, "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.*, vol. 93, pp. 97–108, Dec. 2017, doi: 10.1016/j.ijepes.2017.05.014.
[4] A. Ghaderi, H. A. Mohammadpour, H. L. Ginn, and Y. Shin, "High-impedance fault detection in the distribution network using the time-frequency-based algorithm," *IEEE Trans. Power Del.*, vol. 30, no. 3, pp. 1260–1268, Jun. 2015, doi: 10.1109/tpwrd.2014.2361207.
[5] Q. Xiao, M. Guo, and D. Chen, "High-impedance fault detection method based on one-dimensional variational prototyping-encoder for distribution networks," *IEEE Syst. J.*, vol. 16, no. 1, pp. 966–976, Mar. 2022, doi: 10.1109/jsyst.2021.3053769.
[6] Y. Xi, W. Zhang, F. Zhou, X. Tang, Z. Li, X. Zeng, and P. Zhang, "Transmission line fault detection and classification based on SA-MobileNetV3," *Energy Rep.*, vol. 9, pp. 955–968, Dec. 2023, doi: 10.1016/j.egyr.2022.12.043.
[7] Y. Wang, X. Wang, Y. Wu, and Y. Guo, "Power system fault classification and prediction based on a three-layer data mining structure," *IEEE Access*, vol. 8, pp. 200897–200914, 2020, doi: 10.1109/access.2020.3034365.
[8] C. Hong, Z. Y. Zeng, Y. Z. Fu, and M. F. Guo, "Deep-belief-networks based fault classification in power distribution networks," *IEEJ Trans. Electr. Electron. Eng.*, vol. 15, no. 10, pp. 1428–1435, Oct. 2020, doi: 10.1002/tee.23213.
[9] P. Rai, N. D. Londhe, and R. Raj, "Fault classification in power system distribution network integrated with distributed generators using CNN," *Electric Power Syst. Res.*, vol. 192, Mar. 2021, Art. no. 106914, doi: 10.1016/j.epsr.2020.106914.
[10] P. Srikanth and C. Koley, "A novel three-dimensional deep learning algorithm for classification of power system faults," *Comput. Electr. Eng.*, vol. 91, May 2021, Art. no. 107100, doi: 10.1016/j.compeleceng.2021.107100.
[11] A. Said, S. Hashima, M. M. Fouda, and M. H. Saad, "Deep learning-based fault classification and location for underground power cable of nuclear facilities," *IEEE Access*, vol. 10, pp. 70126–70142, 2022, doi: 10.1109/access.2022.3187026.
[12] S. Fuada, H. A. Shiddieqy, and T. Adiono, "A high-accuracy of transmission line faults (TLFs) classification based on convolutional neural network," *Int. J. Electron. Telecommun.*, vol. 66, no. 4, pp. 655–664, 2020, doi: 10.24425/ijet.2020.134024.
[13] P. Avagaddi and B. Edward J, "Fault classification in transmission systems using wavelet transform," *Gazi Univ. J. Sci.*, vol. 32, no. 3, pp. 884–893, Sep. 2019, doi: 10.35378/gujs.393451.
[14] C. Lin, W. Gao, and M. Guo, "Discrete wavelet transform-based triggering method for single-phase earth fault in power distribution systems," *IEEE Trans. Power Del.*, vol. 34, no. 5, pp. 2058–2068, Oct. 2019, doi: 10.1109/tpwrd.2019.2913728.
[15] T. S. Abdelgayed, W. G. Morsi, and T. S. Sidhu, "A new harmony search approach for optimal wavelets applied to fault classification," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 521–529, Mar. 2018, doi: 10.1109/tsg.2016.2555141.
[16] M. Guo, X. Zeng, D. Chen, and N. Yang, "Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1291–1300, Feb. 2018, doi: 10.1109/jsen.2017.2776238.
[17] F. Alvarez-Gonzalez, A. Griffo, and B. Wang, "Permanent magnet synchronous machine stator windings fault detection by Hilbert–Huang transform," *J. Eng.*, vol. 2019, no. 17, pp. 3505–3509, Jun. 2019, doi: 10.1049/joe.2018.8173.
[18] M. Guo, N. Yang, and W. Chen, "Deep-learning-based fault classification using Hilbert–Huang transform and convolutional neural network in power distribution systems," *IEEE Sensors J.*, vol. 19, no. 16, pp. 6905–6913, Aug. 2019, doi: 10.1109/jsen.2019.2913006.
[19] O. R. Seryasat, M. A. Shoorehdeli, M. Ghane, J. Haddadnia, and M. Zeinali, "Intelligent fault detection of ball bearing using FFT, STFT energy entropy and RMS," *Life Sci. J.*, vol. 9, no. 3, pp. 1781–1786, 2012.
[20] R. F. R. Junior, I. A. D. S. Areias, M. M. Campos, C. E. Teixeira, L. E. B. Da Silva, and G. F. Gomes, "Fault detection and diagnosis in electric motors using convolution neural network and short-time Fourier transform," *J. Vibrat. Eng. Technol.*, vol. 10, no. 7, pp. 2531–2542, Oct. 2022, doi: 10.1007/s42417-022-00501-3.
[21] A. Maqsood, D. Oslebo, K. Corzine, L. Parsa, and Y. Ma, "STFT cluster analysis for DC pulsed load monitoring and fault detection on naval shipboard power systems," *IEEE Trans. Transport. Electrific.*, vol. 6, no. 2, pp. 821–831, Jun. 2020, doi: 10.1109/tte.2020.2981880.
[22] T. Arias-Vergara, P. Klumpp, J. C. Vasquez-Correa, E. Nöth, J. R. Orozco-Arroyave, and M. Schuster, "Multi-channel spectrograms for speech processing applications using deep learning methods," *Pattern Anal. Appl.*, vol. 24, no. 2, pp. 423–431, May 2021, doi: 10.1007/s10044-020-00921-5.
[23] N. Lopac, F. Hržic, I. P. Vuksanovic, and J. Lerga, "Detection of non-stationary GW signals in high noise from Cohen's class of time-frequency representations using deep learning," *IEEE Access*, vol. 10, pp. 2408–2428, 2022, doi: 10.1109/access.2021.3139850.
[24] H. Jung and J. Ju, "Training data reduction for deep learning-based image classifications using random sample consensus," *J. Electron. Imag.*, vol. 31, no. 1, Feb. 2022, Art. no. 010501, doi: 10.1117/1.Jei.31.1.010501.
[25] M. T. Zeegers, D. M. Pelt, T. van Leeuwen, R. van Liere, and K. J. Batenburg, "Task-driven learned hyperspectral data reduction using end-to-end supervised deep learning," *J. Imag.*, vol. 6, no. 12, p. 132, Dec. 2020, doi: 10.3390/jimaging6120132.
[26] H. Li, J. Liu, Z. Yang, R. W. Liu, K. Wu, and Y. Wan, "Adaptively constrained dynamic time warping for time series classification and clustering," *Inf. Sci.*, vol. 534, pp. 97–116, Sep. 2020, doi: 10.1016/j.ins.2020.04.009.
[27] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu, "Weighted dynamic time warping for time series classification," *Pattern Recognit.*, vol. 44, no. 9, pp. 2231–2240, Sep. 2011, doi: 10.1016/j.patcog.2010.09.022.
[28] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intell. Data Anal.*, vol. 11, no. 5, pp. 561–580, Oct. 2007, doi: 10.3233/ida-2007-11508.
[29] X. Zhang, Z. Li, N. Nan, and X. Wang, "Depth-resolved dispersion compensation method for optical coherence tomography imaging based on rectangular window function optimization," *J. Modern Opt.*, vol. 69, no. 15, pp. 827–837, Sep. 2022, doi: 10.1080/09500340.2022.2090629.
[30] W. Astuti, W. Sediono, A. M. Aibinu, R. Akmeliawati, and M. J. E. Salami, "Adaptive short time Fourier transform (STFT) analysis of seismic electric signal (SES): A comparison of Hamming and rectangular window," in *Proc. IEEE Symp. Ind. Electron. Appl.*, Sep. 2012, pp. 372–377.

NIEN-CHE YANG (Member, IEEE) was born in Keelung, Taiwan, in 1977. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 2002, 2004, and 2010, respectively. Since 2018, he has been the Faculty Member of the National Taiwan University of Science and Technology, where he is currently an Associate Professor in electrical engineering. His research interests include micro-grid state estimation, harmonic three-phase power flow analysis, probabilistic three-phase power flow analysis, energy loss computation in low voltage networks, micro-grids, smart grids, and electric vehicles. He is a member of the Phi Tau Phi Scholastic Honor Society.

JEN-MING YANG was born in Yilan, Taiwan, in 2000. He received the B.S. degree from the National Taiwan University of Science and Technology (NTUST), Taipei, Taiwan, in 2022, where he is currently pursuing the M.S. degree with the Electrical Engineering Department.

His research interests include power system analysis and fault detection.