Intelligent Fault Detection Scheme for Microgrids With Wavelet-Based Deep Neural Networks

James J. Q. Yu, *Member, IEEE*, Yunhe Hou, *Senior Member, IEEE*, Albert Y. S. Lam, *Senior Member, IEEE*, and Victor O. K. Li, *Fellow, IEEE*

***Abstract*—Fault detection is essential in microgrid control and operation, as it enables the system to perform fast fault isolation and recovery. The adoption of inverter-interfaced distributed generation in microgrids makes traditional fault detection schemes inappropriate due to their dependence on significant fault currents. In this paper, we devise an intelligent fault detection scheme for microgrid based on wavelet transform and deep neural networks. The proposed scheme aims to provide fast fault type, phase, and location information for microgrid protection and service recovery. In the scheme, branch current measurements sampled by protective relays are pre-processed by discrete wavelet transform to extract statistical features. Then all available data is input into deep neural networks to develop fault information. Compared with previous work, the proposed scheme can provide significantly better fault type classification accuracy. Moreover, the scheme can also detect the locations of faults, which are unavailable in previous work. To evaluate the performance of the proposed fault detection scheme, we conduct a comprehensive evaluation study on the CERTS microgrid and IEEE 34-bus system. The simulation results demonstrate the efficacy of the proposed scheme in terms of detection accuracy, computation time, and robustness against measurement uncertainty.**

***Index Terms*—Fault detection, fault location, microgrid protection, wavelet transform, deep neural network.**

I. Introduction

MICROGRIDS are gathering attention from the industry and research community, due to advances in distributed generation (DG) development [1]. They are expected to bring benefits to the modern power system control thanks to the improved power efficiency, reliability, and quality. Microgrids can operate either in grid-connected mode, where the external grid supports part of the power consumption, or in islanded mode in case the external grid suffers from disturbances such as frequency deviations and voltage fluctuations, etc. Contributed by load-side DGs, critical loads in microgrids can be supplied in islanded mode without the external grid.

Manuscript received June 22, 2017; revised October 10, 2017; accepted November 15, 2017. Date of publication November 22, 2017; date of current version February 18, 2019. This work was supported in part by the Theme-Based Research Scheme of the Research Grants Council of Hong Kong under Grant T23-701/14-N, and in part by the National Natural Science Foundation of China under Grant 51707170. Paper no. TSG-00884-2017. *(Corresponding author: James J. Q. Yu.)*

J. J. Q. Yu, Y. Hou, and V. O. K. Li are with the Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong (e-mail: jqyu@eee.hku.hk; yhhou@eee.hku.hk; vli@eee.hku.hk).

A. Y. S. Lam is with the Fano Labs and the Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong (e-mail: albert@fano.ai).

Digital Object Identifier 10.1109/TSG.2017.2776310

Meanwhile, the protection of microgrids is one of the major and critical operational challenges [1]–[3]. With the gradual adoption of renewable energy sources in modern power systems, microgrids are commonly integrated with inverter-interfaced DGs (IIDG), such as photovoltaic DGs (PVDG) and battery energy storage systems (BESS). Traditional protective relays for distribution system fault detection depend on large fault currents. However, IIDGs can only contribute insignificant fault currents such that the protection schemes are not activated [4]. Thus these relays may fail to protect microgrids. [5] provides a thorough analysis on the current and voltage dynamics in such microgrids.

Fault detection in microgrid generally has three objectives. If there is a fault in the system, a fault detection scheme should determine the fault type (e.g., single-phase-to-ground, three-phase-short-circuit, etc.), fault phase in unbalanced faults, and fault location. The former two enable subsequent fault isolation operations, while the latter can benefit service restoration. According to [5], modern microgrids should maintain the operation of sound phases in unbalanced short-circuit faults, which advocates the integration of single-phase protective devices. Given accurate fault type and fault phase information, selective phase tripping can be achieved [6]. As a result, system reliability can be significantly improved [7], and utilities are gradually adopting this protection scheme [8]. Moreover, accurate detection of fault location can remarkably reduce the effort in service restoration operations [9], and this becomes increasingly essential if the restoration involves underground operations.

In recent years, a growing body of research employs data-driven and digital signal processing approaches for microgrid fault type and/or phase detection. For instance, decision tree and random forest are widely employed to detect faults in both grid-connected and islanded microgrids (see [4], [10]–[12]). Other machine learning techniques, e.g., support vector machine and k-nearest neighbors algorithm, have also been utilized for fault detection in [3] and [12]. Thanks to the high computation speed of these data-driven approaches, satisfactory fault classification can be developed in near real-time. In addition, digital signal processing approaches such as discrete Fourier transform and discrete wavelet transform (DWT) are widely adopted to "pre-process" the input signals to better extract the time-frequency characteristics for analysis [3], [4], [12]. Interested readers may refer to [1] and [4] for surveys on microgrid fault detection schemes in the literature.

However, there remains a research gap in the development of microgrid fault detection schemes. Some existing investigations cannot provide fault type information, thus cannot be properly adopted in the single-phase tripping paradigm (see [12], [13]). Moreover, existing work in fault location detection focuses on low-voltage DC microgrids, e.g., [14] and [15]. Fault location detection in AC distribution networks can typically be achieved through traveling-wave or injection-based algorithms [9], [16]. However, traveling-wave algorithms suffer from reflected wave detection and discrimination issues [14], [17], and some of them require synchronized data over communication links [18]. None of them demonstrated fault location detection performance on islanded microgrids or loop/ring-topology networks. Meanwhile, injection-based algorithms are limited to phase-to-ground faults and are only applicable in radial networks [14].

To bridge the research gap in fault detection for IIDG-enabled microgrids, in this paper we present an intelligent fault detection scheme based on DWT and recent development of deep neural networks (DNN), which is a class of data-driven machine learning techniques. While DWT is prone to noise and power disturbances, DNN is introduced to enhance its robustness thanks to DNN's outstanding capability of handling data with noise [19], [20]. Fig. 1 presents a schematic diagram of the proposed fault detection scheme. The scheme takes branch current magnitudes of three phases in one cycle sampled by protective relays as input data. The measurements are processed by DWT to extract the time-frequency domain features. Afterwards, the features, together with the measurements, are input into three DNNs for fault type classification, fault phase identification, and fault location detection. Eventually, the fault information is developed, which can be employed in later protective and remedial control actions.

The contributions of this work are summarized as follows:

- We propose a fault detection mechanism for AC microgrids that can provide accurate and timely fault type, phase, and location information;
- DNN is combined with DWT to solve the microgrid fault detection problem from the data-driven viewpoint;
- We perform comprehensive case studies to analyze the performance of the proposed mechanism and compare the results with the state-of-the-art.

Compared with existing fault detection mechanisms, the proposed one can develop accurate and fast fault detection result (type, phase, and location) without communications. It can adapt to different operating modes and network topologies, i.e., grid-connected and islanded mode, and radial and loop topology. In addition, power dynamics with both conventional synchronous generators and IIDG can be handled.

The rest of this paper is organized as follows. In Section II we describe the microgrid system employed to analyze the proposed scheme. Section III introduces the DWT technique, and Section IV elaborates on the formulation of the proposed fault detection scheme. We perform a series of simulations to demonstrate the efficacy of the proposed scheme in Section V. Finally, this paper is concluded in Section VI.

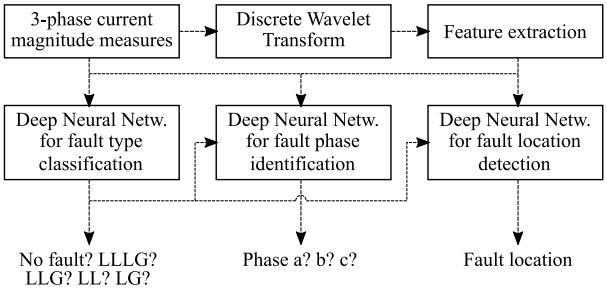


Fig. 1. Schematic diagram of the proposed microgrid fault detection scheme.

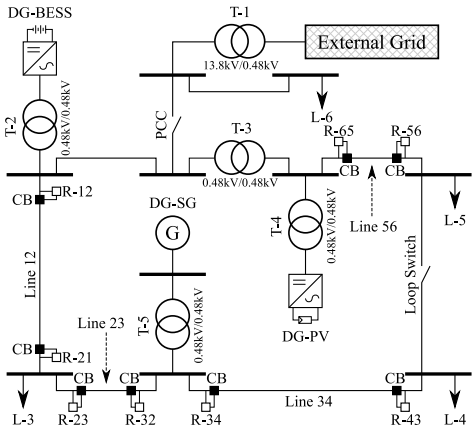


Fig. 2. Modified CERTS microgrid system diagram.

II. Investigated Microgrid System

In this paper, we focus on a modified microgrid system based on the CERTS microgrid [21], which has been employed in many previous investigations on microgrid fault detection, see [3], [5]. As shown in Fig. 2, the employed microgrid operates at 0.48 kV, 60 Hz and can support loads in either grid-connected or islanded mode, controlled by the state of Point of Common Coupling (PCC) switch. In addition, the loop switch enables the system to be operated in radial or loop topology. To test the performance of the proposed scheme on various types of DGs in microgrids, we employ three DGs in the system, namely, a battery-energy storage system (DG-BESS in Fig. 2), a photovoltaic power source (DG-PV), and a typical diesel synchronous generator (DG-SG). DG-BESS is interfaced to the system by a current controlled voltage source inverter, whose control strategy is switched to frequency controlled inverter in islanded mode. DG-PV is interfaced through a current controlled voltage source converter, which maintains its control strategy on grid-connected and islanded modes.

We follow [3] to set the parameters for the loads, transformers, and transmission lines, as shown in Fig. 2. Specifically, four loads are distributed with the system, whose load values are (90 kW, 45 kVAr) for L-3 and L-4, (90 kW, -40 kVAr) for L-5, and (90 kW, -20 kVAr) for L-6 under the nominal operating condition. Line 12, 34, 56 are of AWG2 type with lengths equal to 68.58 meters. Line 23 is of AWG00 type, whose length is 22.86 meters. Other parameters are set according to the record in [3, Table I].

Similar to previous work [3], [5], we focus on four power lines in the system, i.e., Line 12, 23, 34, and 56 as shown in Fig. 2. Since multiple DGs are deployed in the system, and the microgrid can operate in a loop topology, digital protective relays are installed on both ends of the power lines. With the assistance of attached current transformers, these relays sample the branch current magnitudes at 3.84 kHz rate, which accords with the configurations (see [22], [23]). This microgrid system is modelled in DIgSILENT PowerFactory [24]. The software package simulates the current measurements in the fault detection process. The data is employed in the proposed fault detection scheme to detect the fault type and location.

III. Discrete Wavelet Transform Analysis

DWT is a digital signal processing technique which transforms a time-series into mutually orthogonal set of data. It can extract the hidden time-frequency domain characteristics of the fault current. In the proposed fault detection scheme, DWT serves an important role in pre-processing the input data for DNNs in the scheme. In this section, we first briefly introduce DWT and its properties. Then we discuss the DWT-based time-frequency domain features to be computed, which will be later utilized in DNNs.

A. Continuous and Discrete Wavelet Transform

Wavelets are zero-mean functions over time. A wavelet $\psi_{a,b}(t)$ can be derived from its mother wavelet $\psi(t)$ by scaling and shifting as follows:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), \tag{1}$$

where a and b are scaling and shifting parameters, respectively. Utilizing this relation, the continuous wavelet transform of a signal $s(t)$ with scale a and shift b is defined as

$$C(a, b, s(t), \psi(t)) = \int_{-\infty}^{+\infty} s(t)\psi^*_{a,b}(t)dt = \langle s(t), \psi_{a,b}(t)\rangle, \tag{2}$$

where $\psi^*_{a,b}(t)$ is the complex conjugate of $\psi_{a,b}(t)$, and $\langle\cdot,\cdot\rangle$ is the inner product. With different values of a and b, a family of wavelet coefficients $C(a, b, s(t), \psi(t))$ can be developed.

DWT is performed over the continuous wavelet transform by discretizing a and b. Typically, these parameters are set to powers of two: $a = 2^j$, $b = 2^j \times k$, $j, k \in \mathbb{Z}$. Substituting the discrete values into (1), we have

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t}{2^j} - k\right), \tag{3}$$

and DWT is derived by

$$d_{j,k} = \int_{-\infty}^{+\infty} s(t)\psi^*_{j,k}(t)dt = \left\langle s(t), \psi_{j,k}(t)\right\rangle, \tag{4}$$

where $d_{j,k}$ is known as the wavelet detail coefficient at level j and location k [25].

For most signals $s(t)$, however, the analytical solution cannot be solved [26], [27]. Meanwhile, Mallet developed a technique to decompose the multi-resolution signal in [26], which is widely recognized as a standard method to calculate DWT. Given an arbitrary signal $s(t)$, its multi-resolution decomposition at level M is defined by

$$\begin{aligned} s(t) &= \sum\nolimits_k a_{M,k}\frac{1}{\sqrt{2^M}}\varphi\left(\frac{t}{2^M} - k\right) \\ &\quad + \sum\nolimits_j^M \sum\nolimits_k d_{j,k}\frac{1}{\sqrt{2^j}}\psi\left(\frac{t}{2^j} - k\right) \\ &\triangleq A_M(t) + \sum\nolimits_j D_j(t), \end{aligned} \tag{5}$$

where $a_{M,k}$ are the approximation coefficients at level M such that $a_{M,k} = \langle s(t), \varphi_{M,k}(t)\rangle$, and $\varphi(t)$ is a companion scaling function [26]. By this transformation, $s(t)$ is decomposed into an approximation coefficient $A_M(t)$ and a sequence of detail coefficients $D_j(t)$ at level M. Interested readers can refer to [26] for the detailed algorithm.

Following (5), DWT represents input signals in the time-frequency domain [4]. It is widely adopted in fault detection schemes since it can provide features with the optimal time-frequency resolution in all frequency ranges, which results in a better feature extraction ability [28]. Comparing with Fourier transform and its variants (e.g., fast Fourier transform, short-time Fourier transform, etc.), wavelet transform can reveal the time support of frequencies efficiently, and is more computationally efficient [29].

B. Mother Wavelet and Decomposition Level

Different wavelets have unique time-frequency domain characteristics, which can influence the feature extraction ability of DWT [26]. Many wavelet families have been adopted in previous work (see [3], [4]) for DWT in microgrid fault detection, e.g., `coif` (coiflets), `db` (daubechies), `dmey` (discrete meyer), `haar`, `bior` (biorthogonal), and `sym` (symlets). While there must exist an optimal set of wavelet members in these families that will lead to the optimal performance for fault detection, it is impractical to test all combinations of wavelets. They should be selected strategically according to the properties of the sampled data. Specifically, this selection is based on the characteristics of the analyzed data. When the data contains sufficient samples, the `db` and `sym` families are generally preferred thanks to their robustness regardless of data properties [27], e.g., length and number of samples. In such cases, the level of decomposition M has more impact on the system performance than the choice of mother wavelets. Comparing with `db` and `sym`, other mother wavelets may suffer from their longer filter lengths, resulting in low levels of decomposition [27]. This can potentially lead to bad feature extraction capability. Therefore, in this work we employ nine wavelet members in the two wavelet families `db` and `sym` as the mother wavelets to transform the input signal, i.e., branch current measurements.

Besides the mother wavelets, decomposition level is another important parameter that may impact the signal decomposition performance. A larger level can provide more detailed description on the input signal, but the computational cost will increase. There is a maximum level of decomposition in theory for each wavelet member, which is jointly determined by

TABLE I
MOTHER WAVELETS AND DECOMPOSITION LEVELS USED

Wavelet	F	$M = L$	Wavelet	F	$M = L$
db2/sym2	4	4	db4/sym4	8	3
db6/sym6	12	2	db8/sym8	16	2
db10	20	1			

the size of input signal and mother wavelet:

$$L = \left\lfloor \log_2 \frac{N}{F-1} \right\rfloor, \tag{6}$$

where L is the maximum decomposition level, N is the length of the input signal, and F is the filter size of the mother wavelet [26]. According to the heuristic in [27], the decomposition level M should be set to its maximum value L to handle branch current measurements. The length of the input signal is the number of measurements in one cycle, i.e., $3840/60 = 64$. Hence the decomposition levels for the mother wavelets can be calculated using (6), which are listed in Table I. For a mother wavelet with decomposition level M, the input signal can be decomposed into one approximation coefficient and M detail coefficients according to (5). Consequently, 32 coefficients can be calculated from one input signal sequence using the listed mother wavelets.

C. Feature Selection and Extraction

Using each of the mother wavelets, DWT can decompose a sequence of input signal into a series of coefficients $a_{M,k}$ and $d_{j,k}$. Choosing suitable features to represent the characteristics of the input signal is critical for fault detection [4]. In the proposed scheme, we select a series of statistical features of the coefficients to construct the input feature vector for DNN, which contains vital information of the investigated fault. Specifically, we calculate the following features of each decomposed coefficient (here we use s to represent the coefficient $a_{M,k}$ or $d_{j,k}$):

- The maximum value of the coefficient: $\max\{s\}$;
- The minimum value of the coefficient: $\min\{s\}$;
- The mean value of the coefficient: $\mu_s = \mathrm{E}[s]$;
- The standard deviation of the coefficient: $\sigma_s = \mathrm{E}[(s-\mu)^2]^{1/2}$;
- The skewness of the coefficient: $\mathrm{E}\left[\left((s-\mu_s)/\sigma_s\right)^3\right]$;
- The energy of the coefficient: $\sum s^2$.

These features have demonstrated their efficacy in the previous literature which employs DWT for classification tasks (see [4], [27]).

Consequently, for each cycle in the microgrid system, 32(coefficients) $\times$ 6(features) $\times$ 3(phases) $=$ 576 features can be calculated, which constitute a feature vector representing the power dynamics in the cycle. This feature vector is later input into DNN to develop the fault detection results, which will be introduced next.

IV. DEEP NEURAL NETWORK-BASED FAULT DETECTION

DNN is a type of artificial neural networks (ANN) with multiple hidden layers of neurons between the input and output. It is widely adopted to model complex non-linear systems in engineering research [19]. In addition, the computation of DNN only involves simple algebraic equations, rendering a fast computation speed. This characteristic makes DNN capable of handling problems in real-time.

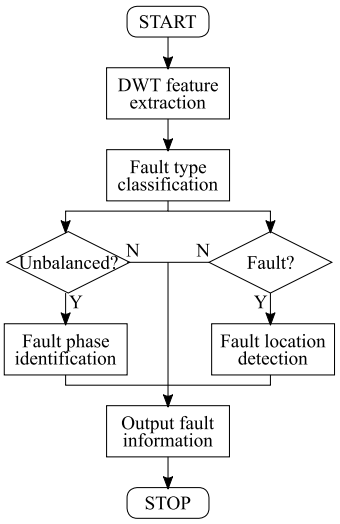


Fig. 3. Flow chart of the proposed intelligent fault detection scheme.

The proposed DNN-based fault detection scheme is based on the hypothesis that the branch current measurements can immediately indicate fault occurrence in the system. This hypothesis is widely recognized in [2]–[5]. In addition, the temporal data dependency in sampled current magnitudes can also help identify the fault. Utilizing the time-series branch current measurements and the extracted wavelet features introduced in Section III, we construct an intelligent fault detection scheme in this section.

The fault detection problem is divided into three sub-problems, namely, fault type classification, fault phase identification, and fault location detection. Each sub-problem is handled by a standalone DNN. The flow chart of the proposed fault detection scheme is depicted in Fig. 3. At an arbitrary time, a protective relay in the microgrid samples the 3-phase branch current magnitudes, which are input into the proposed fault detection scheme. The measurements are first processed by DWT to extract the features. Then the features and measurements are input into a fault type classification DNN. If a fault is detected, the fault location detection DNN is employed to determine the location of the fault. At the same time, if the fault is classified as an unbalanced one, the fault phase identification DNN is used to develop the fault phase. Finally, the generated information can be employed in later control operation decision-making processes, e.g., fault isolation and recovery.

In the remainder of this section, we first introduce the technique employed to construct the three DNNs. Then we present the structures of these DNNs, training data preparation method, and training optimization configuration.

A. Gated Recurrent Unit and Dense Layer

In this work we mainly use gated recurrent unit (GRU) to construct DNNs to handle the three sub-problems in

the microgrid fault detection problem. GRU [30] is a modern variant of ANN, which is among the most commonly used data-mining and machine learning techniques. ANN has been employed in many disciplines [31] due to its model-independent and computationally efficient properties [31], [32]. ANN tries to simulate the model of control systems by learning from the mathematical relationship between system input and output.

However, a typical ANN overlooks the input data correlation in the time domain. GRU, along with some other neural networks, is designed to overcome this drawback. By introducing additional recurrent connections in the hidden layers of a neural network, GRU is capable of maintaining previous information for later use and capturing the temporal dependencies in the input data [30].

Given a time series $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T]$, GRU can develop a sequence of output values $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_T]$ where each output value $\mathbf{h}_t$ is calculated using all input values from $\mathbf{x}_1$ to $\mathbf{x}_t$. This is achieved by its internal structure, which can be expressed as follows:

$$\mathbf{h}_t = \mathbf{z}_t * \mathbf{h}_{t-1} + (\mathbf{1} - \mathbf{z}_t) * \tilde{\mathbf{h}}_t, \tag{7a}$$

$$\mathbf{z}_t = \text{sigm}(\mathbf{w}_{xz}\mathbf{x}_t + \mathbf{w}_{hz}\mathbf{h}_{t-1} + \mathbf{b}_z), \tag{7b}$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{w}_{xh}\mathbf{x}_t + \mathbf{w}_{hh}(\mathbf{r}_t * \mathbf{h}_{t-1}) + \mathbf{b}_h), \tag{7c}$$

$$\mathbf{r}_t = \text{sigm}(\mathbf{w}_{xr}\mathbf{x}_t + \mathbf{w}_{hr}\mathbf{h}_{t-1} + \mathbf{b}_r), \tag{7d}$$

where $*$ is the element-wise product, and all $\mathbf{w}$ and $\mathbf{b}$ matrices are the learning parameters of GRU. From (7) it can be observed that GRU simulates the relationship between output $\mathbf{h}_t$ and input $\mathbf{x}_1, \ldots, \mathbf{x}_t$ using the learning parameters, whose values are initially unknown. In practice, we use known input and output data to tune these parameters to reflect the relationship. This process is called training, and is typically performed off-line. After training, the value of the learning parameters can be used to calculate the predicted output values given a new set of input, even if the actual output is unknown.

Besides GRU, another important neural network layer employed in the structure is the fully connected layer, or Dense layer. Dense layers are composed of multiple artificial neurons, each of which calculates its output using the input data as follows:

$$y = \text{actv}(\mathbf{w}_{\text{dense}} * \mathbf{x} + \mathbf{b}_{\text{dense}}), \tag{8}$$

where $\mathbf{x}$ and y are the input and output, respectively, $\mathbf{w}_{\text{dense}}$ and $\mathbf{b}_{\text{dense}}$ are the learning parameters of the Dense layer, and $\text{actv}(\cdot)$ is the activation function [20].

B. Deep Neural Networks Structures

In the proposed intelligent fault detection scheme, three DNNs are employed to classify the fault type, identify the fault phase, and locate the fault position, respectively. All these networks are constructed using GRU as well as standard fully-connected neuron layers in ANN. Meanwhile, as their outputs are different, their schemata vary slightly.

We first construct the fault type classification DNN. This networks accepts the 3-phase time-series current measurements and the DWT-extracted features as input, aiming to tell if a given sequence of dynamics indicates a fault and its type. Similar to previous work [3]–[5], we consider four types of faults: 1) single-phase-to-ground (LG), 2) double-phase (LL), 3) double-phase-to-ground (LLG), and 4) three-phase-to-ground (LLLG). The constructed DNN have four 0-1 indicators in the output, each of which represents one fault type. In addition, as this DNN should distinguish faults from no-fault cases, an extra no-fault indicator is introduced in the output, making it of length five.

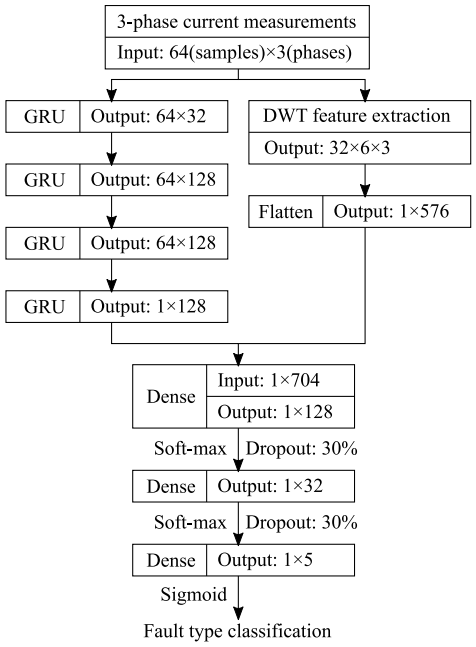
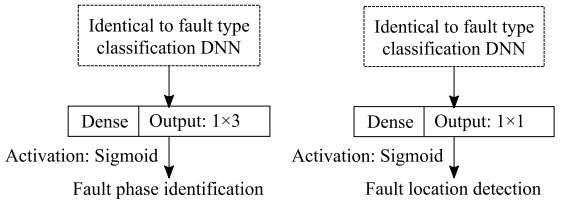


Fig. 4. Schema of fault type classification DNN.

The schema of fault type classification DNN is depicted in Fig. 4. In this DNN, we employ four GRU layers and three Dense layers to process the current measurements and DWT features. Specifically, the first GRU layer is used to map the time-series into a higher dimension space, and the next two GRU layers are utilized to extract the temporal dependency of input data. The last GRU layer expresses the dependency in a 1-D vector, which is combined with the extracted DWT features as the input of subsequent dense neuron layers. Finally, the fault type characteristics are abstracted in the first two Dense layers, and the last one transforms the abstracted features into a human-readable form, i.e., fault type. Since the activation function for the last layer is sigmoid, values in the output 1×5 vector are in $(0, 1)$. The fault type is determined by the maximum value in the vector. Suppose that the third element is the largest in the output vector, then we can tell that a fault of the type represented by the third output, e.g., LLG, occurred in the system.

The schemata for fault phase identification and fault location detection DNNs are similar to that presented in Fig. 4, except for the last dense layer. If the fault is classified as unbalanced (LG, LL, or LLG), the second fault phase identification DNN is utilized to detect the fault phase. This DNN has an output vector of length three, in which each element represents the fault state of a phase. For an LG fault, the phase whose corresponding value is the largest in the output vector is considered



(a) Fault phase identification. (b) Fault location detection.

Fig. 5. The last dense neuron layers of the fault phase identification and fault location detection DNNs.

as the fault phase. For LL or LLG faults, the phases with the two largest values are the fault phases.

Finally, the fault location detection DNN only outputs one value from the last dense layer, which indicates the relative (percentage) position on the transmission line from the sampling protective relay. The last layers of the fault phase identification and fault location detection DNNs are presented in Fig. 5.

Note that the proposed DNN only takes the local three-phase current magnitudes as the input. In some real-world applications, it is possible that extra information, e.g., fault resistance values, is available before the detection is executed. In such cases, such information can be considered as new input of the first dense layer in Fig. 4, and thus the model can be further enhanced by learning from the additional information. In this work, we present a generalized model as depicted in Fig. 4. Variants of the model incorporated with additional information may lead to better system performance, which will be further studied in future work.

C. Time-Series Simulation and Training

In order to use the constructed DNNs for fault detection, their learning parameters need to be trained offline. The training data, composed of input current magnitude measurements and output fault information, can be obtained from either historical data or time-series simulation of different operational events. The input current magnitude measurements are arranged in the form of

$$\mathbf{X} = \begin{bmatrix} I_{a,1} & I_{a,2} & I_{a,3} & \cdots & I_{a,64} \\ I_{b,1} & I_{b,2} & I_{b,3} & \cdots & I_{b,64} \\ I_{c,1} & I_{c,2} & I_{c,3} & \cdots & I_{c,64} \end{bmatrix}^{\mathsf{T}}, \tag{9}$$

where $I_{a,t}$, $I_{b,t}$, and $I_{c,t}$ are the t-th current magnitude measurements on phase a, b, and c in the cycle, respectively. As the fault information will be revealed in three DNNs, different output $\mathbf{Y}^{\text{type}} = (\mathbf{y}_i^{\text{type}}) \in \mathbb{B}^{1\times 5}$, $\mathbf{Y}^{\text{phase}} = (\mathbf{y}_i^{\text{phase}}) \in \mathbb{B}^{1\times 3}$, and $\mathbf{Y}^{\text{loc}} \in (0, 1)$ should be constructed for the three DNNs, respectively:

$$\mathbf{y}_i^{\text{type}} = \begin{cases} 1 & \text{if fault type is represented by } i \\ 0 & \text{otherwise} \end{cases}, \tag{10a}$$

$$\mathbf{y}_i^{\text{phase}} = \begin{cases} 1 & \text{if phase } i \text{ is short circuited} \\ 0 & \text{otherwise} \end{cases}, \tag{10b}$$

$$\mathbf{Y}^{\text{loc}} = \text{relative fault location on the line.} \tag{10c}$$

Given a collection of R training data $\{\mathbf{X}_{(r)}, \mathbf{Y}_{(r)}^{\text{type}}, \mathbf{Y}_{(r)}^{\text{phase}}, \mathbf{Y}_{(r)}^{\text{loc}}\}_{r=1}^{R}$, many gradient descent algorithms can be employed to train the learning parameters in DNNs [19], [32]. In this paper, we use the Adam optimizer [33] to find the optimal values for these parameters. Since the fault type classification is a multinomial classification problem, we use the multinomial cross entropy loss function as the optimization objective. Similarly, the binary cross entropy loss function is assigned to the fault phase identification DNN due to its multi-label classification nature. The loss function for the fault location DNN is mean squared error loss function. Interested readers may refer to [20] for the detailed implementation and theory of these loss functions.

In our DNN model, there are hundreds of thousands of learning parameters to be trained. It is impractical to simultaneously adjust their values without overfitting [34]. Instead, an effective technique, called "dropout", is employed to address this issue [35]. This technique randomly sets the output value(s) of a neuron to zero at a user-defined probability (30% in this work). Therefore, the dropped neurons do not contribute to the result calculation, which can effectively prevent the network from heavily relying on specific neurons to develop output. Thus more robust features can be extracted in the learning process [34], [36]. This "dropout" scheme is only employed in the training process. When the trained parameters are used to develop testing results, all neurons contribute to generate the network output [35].

TABLE II
CONFIGURATIONS FOR FAULT CASES SIMULATION

Parameter	Possible Configuration	Count
Topology	Radial or loop	2
Operating mode	Grid-connected or islanded	2
Fault type	(A/B/C)G, AB, AC, BC, (AB/AC/BC)G, or ABCG	10
Fault resistance (Ω)	0.01, 1, 10, or 100	4
Fault line	Line 12, 23, 34, or 56	4
Fault location	10%, 20%, ⋯, or 90% on fault line	9
L-3 and L-4 Load	(90 kW, 45 kVAr) or (45 kW, 25 kVAr)	2
L-5 Load	(90 kW, -40 kVAr) or (45 kW, -20 kVAr)	2
L-6 Load	(90 kW, -20 kVAr) or (45 kW, -10 kVAr)	2

TABLE III
CONFIGURATIONS FOR NON-FAULT CASES SIMULATION

Parameter	Possible Configuration	Count
Topology	Radial or loop	2
Operating mode	Grid-connected or islanded	2
L-3 and L-4 Load	(90 kW, 45 kVAr) or (45 kW, 25 kVAr)	2
L-5 Load	(90 kW, -40 kVAr) or (45 kW, -20 kVAr)	2
L-6 Load	(90 kW, -20 kVAr) or (45 kW, -10 kVAr)	2
Event	Topology change, operating mode change, L-3, L-4, L-5, or L-6 load change by ±5%, ±10%, ±15%, or ±20%	34

D. Discussion

In this work, we employ DWT and DNN to construct an AC microgrid fault detection mechanism. An advantage for DNN over other conventional computation techniques is that the time-consuming parameter tuning process can be conducted offline. The online testing process involves minimal linear algebraic calculations, which are computationally efficient. In the meantime, the DWT employed in this work does not have a pre-processing step to reduce the testing time. As will be

TABLE IV
FAULT DETECTION ACCURACY ON CERTS MICROGRID SYSTEM

Relay	Fault vs. Non-fault Accuracy		Fault Type Accuracy		Fault Phase Accuracy		Location Error	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
R-12	99.50%	99.56%	98.38%	97.88%	98.39%	97.37%	2.49%	5.63%
R-21	99.99%	99.31%	99.28%	97.46%	97.84%	98.13%	2.09%	4.17%
R-23	99.26%	99.24%	97.96%	98.28%	98.08%	99.03%	8.33%	10.59%
R-32	99.54%	99.70%	97.93%	96.34%	96.87%	96.75%	8.64%	8.70%
R-34	99.21%	99.06%	95.50%	97.32%	96.75%	98.86%	5.68%	4.42%
R-43	99.46%	99.11%	97.27%	97.73%	98.68%	97.39%	6.80%	5.94%
R-56	99.99%	99.20%	97.61%	97.11%	96.83%	97.23%	4.39%	1.93%
R-65	99.76%	99.31%	97.61%	98.67%	97.74%	98.59%	4.50%	5.85%
Average	99.60%	99.31%	97.70%	97.60%	97.64%	97.92%	5.24%	5.90%

illustrated in Section V-A, in normal settings, the employed DWT can be computed fast enough to keep the overall system response time satisfactory. With the recent rapid development of signal processing techniques, there may also exist wavelet transform algorithms which are re-constructed in an online manner to further reduce the computational time. Such online DWT algorithms can be easily incorporated into our proposed fault detection mechanism and the resulting detection accuracy is not influenced as long as the DWT results remain identical.

V. CASE STUDIES

We perform a series of simulations to assess the fault detection performance of the proposed scheme. We first evaluate the fault type/phase classification and location detection accuracy, and compare the results with the state-of-the-art schemes in the literature. Then we investigate the impact of noisy measurements on the fault detection performance. Finally, we test the proposed scheme on another microgrid system besides the CERTS system to assess the generalization of the proposed scheme. All time-series simulations and numerical calculations are conducted on a computer with an Intel Core i7-7700 CPU and an nVidia GTX 1080 GPU. The time-series simulations are performed using DIgSILENT PowerFactory [24], and the DNNs are constructed with TensorFlow [37].

To train the learning parameters in DNN with supervised learning, sufficient previous knowledge that represents post-fault/event power dynamics is critical. The training data should include adequate information to lead the tuning of these parameters approximating the system behavior, subject to different operating conditions and events. In this work, we generate the training data using CERTS microgrid system with time-series simulations under different network topologies, grid-connected modes, types of balanced and unbalanced faults, fault resistances, fault lines, locations on the line, and loads. Details of these configurations are listed in Table II. In total, 46,080 fault cases are generated and simulated. Moreover, multiple non-fault cases are constructed to train the parameters. Details of such cases are listed in Table III, and 1088 non-fault cases are generated and simulated. In each generated case, the simulated branch currents are employed as the measurements sampled by the protective relays. In practice, it is also feasible for utilities to use the historical operation data for training.

TABLE V
SUMMARY OF COMPUTATION TIME

Process	DNN training	DWT & features	DNN classification
Avg. Time	423.94 s/DNN	0.11 ms/wavelet	0.35 ms/DNN

TABLE VI
COMPARISON WITH OTHER FAULT DETECTION SCHEMES

Scheme	Accuracy			Error
	Fault	Type	Phase	Location
Proposed Scheme	**99.31%**	**97.60%**	**97.92%**	**5.90%**
Decision Tree [3]	90.40%+	90.40%+	90.40%	-
K-nearest Neighbors [3]	95.63%+	95.63%+	95.63%	-
Support Vector Machine [3]	93.30%+	93.30%+	93.30%	-
Naive Bayes [3]	94.24%+	94.24%+	94.24%	-
Decision Tree [4]	97.00%	85.00%	-	-
Random Forest [4]	99.00%	94.00%	-	-
Over-current relay (in [4])	56.00%	-	-	-
Differential relay (in [4])	96.00%	-	-	-

For cross validation and prevention of the over-fitting problem, we randomly divide the generated fault and non-fault cases into training and testing sets by the ratio of 3:1, which accords with the common practice, see [4], [38]. The training cases are employed to train the learning parameters of DNN, and the testing cases are used to assess the fault detection accuracy of the trained scheme. Thus, over-fitting problem (machine learning models learn from both data characteristics and random noise, rendering poor generalization performance over new unknown data) can be avoided.

A. Fault Detection Accuracy and Computation Time

Table IV presents the fault detection performance of the proposed scheme. We summarize the accuracy of fault type classification and fault phase identification, which can be directly obtained from the DNNs in the scheme. In addition, we further present the accuracy on distinguishing fault cases from the non-fault ones, and the error of predicted fault locations.

From the simulation results, we can see that the proposed intelligent fault detection scheme can successfully develop the correct information in most training and testing cases. For the typical fault/non-fault detection performance, each relay can achieve more than 99% accuracy, rendering an overall 99.60% accuracy for training cases, and 99.31% for testing ones. For the fault type and phase classification, the performance is

TABLE VII
FAULT TYPE DETECTION ACCURACY AT R-43 WITH DIFFERENT NUMBERS OF NEURAL NETWORK LAYERS

Number of Dense Layers	Number of GRU Layers									
	1		2		3		4		5	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
1	95.02%	94.92%	96.57%	95.35%	96.33%	96.21%	97.71%	96.05%	98.06%	96.19%
2	95.77%	95.89%	96.08%	95.94%	97.53%	97.50%	97.27%	97.73%	98.29%	96.17%
3	95.15%	95.37%	96.01%	95.86%	96.82%	96.75%	97.20%	97.33%	97.43%	95.61%

slightly worse. Despite this, the proposed scheme can still provide satisfactory 97.60% and 97.92% classification accuracies in these tests using testing cases.

The errors of predicted fault locations can lead to some interesting observations. In general, most relays (except R-23 and R-32) can achieve quite accurate prediction on fault locations, resulting in 4.32% and 4.65% error for training and testing cases, respectively. However, the prediction accuracies of R-23 and R-32 are relatively worse. This is because Line 23 (22.86 m) is significantly shorter than the others (68.58 m). In such case, if the fault resistance is large enough, the power dynamics for different fault locations on Line 23 are similar. Meanwhile, the proposed scheme can still locate the fault position with an error of around two meters. This is acceptable when the fault recovery involves underground operations, which is quite common in microgrid systems.

Last but not least, we record the computational time for the proposed scheme in Table V. It can be concluded that the proposed scheme can be executed in real time. With proper parallelization, the scheme can develop fault information in around 0.35 millisecond after the sampling time.[1] Even in the worst case scenario where all calculations are sequential, the fault detection time is around 1.34 ms (non-fault cases) or 1.69 ms (fault cases).

B. Comparison With Other Fault Detection Schemes

We compare the performance of the proposed scheme with existing state-of-the-art schemes for microgrid fault detection. The results are summarized in Table VI, and the best performing items are bolded. Note that [3] only provided fault classification accuracy including fault phase information. Hence it can be inferred that the performance for fault vs. non-fault and fault type classification should be superior than the value presented. These inferred accuracy values are appended with a "+" sign in the table.

From the comparison it can be concluded that the proposed scheme can outperform existing state-of-the-art microgrid fault detection schemes. In addition, the proposed scheme can provide predicted fault locations, which are unavailable in the compared schemes.

C. Deep Neural Network Structure

In this work we adopt a DNN structure as shown in Fig. 1, which contains four GRU layers and two Dense layers.[2]

[1]Assume that DWT feature extraction for nine wavelets is calculated in parallel, which happens when DNN is executing the calculation for the first few GRU layers. DNNs also calculate their output values in parallel.

[2]Here the last Dense layer in Fig. 1 is excluded, since it is compulsory and fixed-size.

TABLE VIII
IMPACT OF MEASUREMENT UNCERTAINTY ON FAULT DETECTION PERFORMANCE

SNR	Accuracy			Error
	Fault	Type	Phase	Location
Perfect	99.31%	97.60%	97.92%	5.90%
40 dB	99.29%	97.64%	97.78%	5.93%
35 dB	99.24%	97.55%	97.86%	5.99%
30 dB	99.20%	97.57%	97.69%	6.01%

The number of layers is one of the most critical hyper-parameters that significantly influence the system performance. In this section, we perform a test on the number of layers in the mechanism to investigate the relationship between fault detection accuracy and DNN structure. Specifically, we test different DNN structures with one to five GRU layers and one to three Dense layers and compare their fault type detection accuracy. All structures are trained with the same training data, and all other simulation configurations are identical. The fault type detection accuracy values at a randomly selected relay (R-43) are presented in Table VII. From the results, it is clear that four GRU layers with two Dense layers yield the most accurate fault type detection information on the testing data. While more GRU layers may lead to better training accuracy, the extra layers also potentially introduce over-fitting issue to the model, rendering a worse testing performance. Note that it is possible to develop more complex DNN structures for even better results. How to configure the structure for microgrid fault detection to further improve the accuracy performance is out of the scope of this work.

D. Measurement Uncertainty

We investigate the impact of noisy branch current measurements on the fault detection accuracy of the proposed scheme. Similar to previous work [3], [12], the current time-series data is distorted with white Gaussian noise. We conduct simulations on three test cases, namely, (i) 40 dB signal-to-noise ratio (SNR), (ii) 35 dB SNR, and (iii) 30 db SNR, which accord with the configuration in [12]. DNNs in these test cases are trained with the distorted data. The performance on testing set cases is summarized in Table VIII.

From the simulation results we have the following observation. While the noise in current measurements does have an impact on the performance, the influence is insignificant. In the worse case scenario (30 dB), the accuracy decrease is only around 0.1% compared with perfect measurements. To conclude, the proposed scheme can achieve almost the same performance considering measurement uncertainties.

TABLE IX
PERFORMANCE ON MODIFIED IEEE 34-BUS MICROGRID SYSTEM

Accuracy			Error
Fault	Type	Phase	Location
99.06%	98.02%	97.71%	6.43%

TABLE X
COMPARING DB AND SYM WAVELETS WITH COIF AND BIOR

Wavelets	Accuracy			Error
	Fault	Type	Phase	Location
db and sym	99.31%	97.60%	97.92%	5.90%
coif and bior	99.04%	97.53%	97.48%	6.11%

E. Performance on Modified IEEE 34-Bus Microgrid System

All previous analyses and simulations are conducted on the CERTS microgrid system. It is of interest to assess the generalization of the proposed scheme. In this sub-section, we employ a modified IEEE 34-bus system and test the fault detection performance of our scheme. The system is constructed according to the descriptions in [39], and we assume that protective relays are installed on transmission lines 808-812, 816-824, 834-842, and 846-848. Using a similar approach as we conducted on CERTS microgrid system, we develop 11,792 cases for training and testing the proposed scheme. The simulation results are illustrated in Table IX. It can be observed that the performance deviation of this modified IEEE 34-bus system from CERTS system is quite trivial, despite the much larger test system. This result leads us to conclude that the proposed fault detection scheme can be generalized and applied to microgrid systems with various sizes.

VI. CONCLUSION

In this paper, we propose a new intelligent fault detection scheme for microgrid systems based on wavelet transform and deep learning approaches. The branch current magnitude measurements sampled by protective relays are input into the scheme, which can develop the detailed information of the fault type, phase, and location for microgrid protection and service recovery. Specifically, the measurement data are pre-processed using DWT and statistical features are extracted from the result. Then the measurements and features are input into tailor-made DNNs to develop fault information. Different from previous work, the proposed scheme can provide a predicted fault location along the transmission line, besides accurately classifying the fault type. In addition, due to the computationally efficient nature of DNN, the whole fault detection process can be conducted in real-time.

To assess the performance of our proposed scheme, we conducted a series of simulations. We first test the fault detection accuracy on a CERTS microgrid system, and compare it with state-of-the-art fault detection schemes in the literature. The results demonstrate that the proposed scheme can provide more accurate fault type classification results, and can discover the fault locations which are unavailable in other approaches. In addition, we evaluate the influence of noisy measurements on the fault detection performance. The simulation shows that measurement uncertainty has a trivial impact on the performance of the scheme. Last but not least, we also test the proposed scheme on a modified IEEE 34-bus system and the fault detection result remains satisfactory. This indicates that the proposed scheme is practical to be adopted in real-world microgrids of different sizes and topologies.

APPENDIX
WAVELET FAMILY SELECTION TEST

In this Appendix, we present the fault detection results on the proposed mechanism with `coif` and `bior` families instead of `db` and `sym`. Specifically, the three-phase current magnitude measurements are processed by wavelets `coif1` ($M = 3$), `coif2` ($M = 2$), `coif3` ($M = 1$), `coif4` ($M = 1$), `coif5` ($M = 1$), `bior1.1` ($M = 6$), `bior1.3` ($M = 3$), `bior1.5` ($M = 2$), `bior2.2` ($M = 3$), and `bior2.4` ($M = 2$). The simulation results are demonstrated in Table X. The presented accuracy values accord with our previous analysis in Section III-B. Therefore, in our design, `db` and `sym` wavelets are employed.

REFERENCES

[1] S. Parhizi, H. Lotfi, A. Khodaei, and S. Bahramirad, "State of the art in research on microgrids: A review," *IEEE Access*, vol. 3, pp. 890–925, 2015.
[2] R. F. Arritt and R. C. Dugan, "Distribution system analysis and the future smart grid," *IEEE Trans. Ind. Appl.*, vol. 47, no. 6, pp. 2343–2350, Nov./Dec. 2011.
[3] T. Abdelgayed, W. Morsi, and T. Sidhu, "A new approach for fault classification in microgrids using optimal wavelet functions matching pursuit," *IEEE Trans. Smart Grid*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/7864451/
[4] D. P. Mishra, S. R. Samantaray, and G. Joos, "A combined wavelet and data-mining based intelligent protection scheme for microgrid," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2295–2304, Sep. 2016.
[5] A. Hooshyar, E. F. El-Saadany, and M. Sanaye-Pasand, "Fault type classification in microgrids including photovoltaic DGs," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2218–2229, Sep. 2016.
[6] T. S. Fahey and N. V. Burbure, "Single-phase tripping," *IEEE Power Energy Mag.*, vol. 6, no. 2, pp. 46–52, Mar./Apr. 2008.
[7] J. R. Agüero, J. Wang, and J. J. Burke, "Improving the reliability of power distribution systems through single-phase tripping," in *Proc. IEEE PES T D*, New Orleans, LA, USA, Apr. 2010, pp. 1–7.
[8] R. M. Cheney, J. T. Thorne, and G. Hataway, "Distribution single-phase tripping and reclosing: Overcoming obstacles with programmable recloser controls," in *Proc. 62nd Annu. Conf. Protect. Relay Eng.*, Austin, TX, USA, Mar. 2009, pp. 214–223.
[9] A. Borghetti, M. Bosetti, C. A. Nucci, M. Paolone, and A. Abur, "Integrated use of time-frequency wavelet decompositions for fault location in distribution networks: Theory and experimental validation," *IEEE Trans. Power Del.*, vol. 25, no. 4, pp. 3139–3146, Oct. 2010.
[10] S. Kar, S. R. Samantaray, and M. D. Zadeh, "Data-mining model based intelligent differential microgrid protection scheme," *IEEE Syst. J.*, vol. 11, no. 2, pp. 1161–1169, Jun. 2017.
[11] E. Casagrande, W. L. Woon, H. H. Zeineldin, and N. H. Kan'an, "Data mining approach to fault detection for isolated inverter-based microgrids," *IET Gener. Transm. Distrib.*, vol. 7, no. 7, pp. 745–754, Jul. 2013.
[12] E. Casagrande, W. L. Woon, H. H. Zeineldin, and D. Svetinovic, "A differential sequence component protection scheme for microgrids with inverter-based distributed generators," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 29–37, Jan. 2014.
[13] L. Che, M. E. Khodayar, and M. Shahidehpour, "Adaptive protection system for microgrids: Protection practices of a functional microgrid system," *IEEE Electrific. Mag.*, vol. 2, no. 1, pp. 66–80, Mar. 2014.
[14] R. Mohanty, U. S. M. Balaji, and A. K. Pradhan, "An accurate non-iterative fault-location technique for low-voltage DC microgrid," *IEEE Trans. Power Del.*, vol. 31, no. 2, pp. 475–481, Apr. 2016.

[15] S. Dhar, R. K. Patnaik, and P. K. Dash, "Fault detection and location of photovoltaic based DC microgrid using differential protection strategy," *IEEE Trans. Smart Grid*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/7820172/

[16] Z. He, J. Zhang, W. H. Li, and X. Lin, "Improved fault-location system for railway distribution system using superimposed signal," *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1899–1911, Jul. 2010.

[17] P. Jafarian and M. Sanaye-Pasand, "A traveling-wave-based protection technique using wavelet/PCA analysis," *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 588–599, Apr. 2010.

[18] Y.-H. Lin, C.-W. Liu, and C.-S. Yu, "A new fault locator for three-terminal transmission lines using two-terminal synchronized voltage and current phasors," *IEEE Trans. Power Del.*, vol. 17, no. 2, pp. 452–459, Apr. 2002.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[20] R. D. Reed and R. J. Marks, II, *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks* (A Bradford Book). Cambridge, MA, USA: MIT Press, Feb. 1999.

[21] R. H. Lasseter *et al.*, "CERTS microgrid laboratory test bed," *IEEE Trans. Power Del.*, vol. 26, no. 1, pp. 325–332, Jan. 2011.

[22] D. Guillén, H. Esponda, E. Vázquez, and G. Idárraga-Ospina, "Algorithm for transformer differential protection based on wavelet correlation modes," *IET Gener. Transm. Distrib.*, vol. 10, no. 12, pp. 2871–2879, Sep. 2016.

[23] J.-A. Jiang, C.-W. Liu, and C.-S. Chen, "A novel adaptive PMU-based transmission-line relay-design and EMTP simulation results," *IEEE Trans. Power Del.*, vol. 17, no. 4, pp. 930–937, Oct. 2002.

[24] *PowerFactory—DIgSILENT Germany*. Accessed: Feb. 1, 2017. [Online]. Available: http://www.digsilent.de/index.php/products-powerfactory.html

[25] E. Magosso, M. Ursino, A. Zaniboni, and E. Gardella, "A wavelet-based energetic approach for the analysis of biomedical signals: Application to the electroencephalogram and electro-oculogram," *Appl. Math. Comput.*, vol. 207, no. 1, pp. 42–62, Jan. 2009.

[26] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[27] D. Chen, S. Wan, and F. S. Bao, "Epileptic focus localization using discrete wavelet transform based on interictal intracranial EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 5, pp. 413–425, May 2017.

[28] E. D. Übeyli, "Combined neural network model employing wavelet coefficients for EEG signals classification," *Digit. Signal Process.*, vol. 19, no. 2, pp. 297–308, Mar. 2009.

[29] D. H. Evans and W. N. McDicken, *Doppler Ultrasound: Physics, Instrumentation and Signal Processing*, 2nd ed. Chichester, U.K.: Wiley, Mar. 2000.

[30] K. Cho *et al.*, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Doha, Qatar, Oct. 2014, pp. 1724–1734.

[31] R. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Mag.*, vol. ASSPM-4, no. 2, pp. 4–22, Apr. 1987.

[32] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 1, pp. 4–27, Mar. 1990.

[33] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, Jul. 2015, pp. 1–15.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1106–1114.

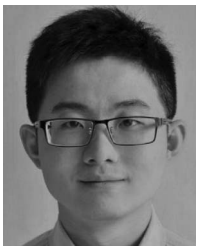
[35] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *Comput. Res. Repository*, vol. abs/1207.0580, 2012. [Online]. Available: http://arxiv.org/abs/1207.0580

[36] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Comput. Res. Repository*, vol. abs/1409.0575, 2014. [Online]. Available: http://arxiv.org/abs/1409.0575

[37] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," *Comput. Res. Repository*, vol. abs/1603.04467, 2016. [Online]. Available: http://arxiv.org/abs/1603.04467

[38] J. J. Q. Yu, D. J. Hill, A. Y. S. Lam, J. Gu, and V. O. K. Li, "Intelligent time-adaptive transient stability assessment system," *IEEE Trans. Power Syst.*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/7932885/

[39] O. N. Faqhruldin, E. F. El-Saadany, and H. H. Zeineldin, "A universal islanding detection technique for distributed generation using pattern recognition," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1985–1992, Jul. 2014.



James J. Q. Yu (S'11–M'15) received the B.Eng. and Ph.D. degrees in electrical and electronic engineering from the University of Hong Kong, Hong Kong, in 2011 and 2015, respectively, where he is currently a Post-Doctoral Fellow with the Department of Electrical and Electronic Engineering. His research interests include smart city technologies, power stability analysis, evolutionary algorithm design and analysis, and deep neural network applications.

Yunhe Hou (M'08–SM'15) received the B.E. and Ph.D. degrees in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1999 and 2005, respectively. He was a Post-Doctoral Research Fellow with Tsinghua University, Beijing, China, from 2005 to 2007, and a Post-Doctoral Researcher with Iowa State University, Ames, IA, USA, and the University College Dublin, Dublin, Ireland, from 2008 to 2009. He was also a Visiting Scientist with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA, in 2010. Since 2017, he has been a Guest Professor with the Huazhong University of Science and Technology, China. He joined the faculty of the University of Hong Kong, Hong Kong, in 2009, where he is currently an Associate Professor with the Department of Electrical and Electronic Engineering. He is an Editor of the IEEE TRANSACTIONS ON SMART GRID and the *Journal of Modern Power Systems and Clean Energy*.

Albert Y. S. Lam (S'03–M'10–SM'16) received the B.Eng. degree (Hons.) in information engineering and the Ph.D. degree in electrical and electronic engineering from the University of Hong Kong (HKU), Hong Kong, in 2005 and 2010, respectively. He was a Post-Doctoral Scholar with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA, from 2010 to 2012. He is currently the Chief Scientist and the acting Chief Technology Officer with Fano Labs, and a Honorary Assistant Professor with the Department of Electrical and Electronic Engineering, HKU. He is a Croucher Research Fellow. His research interests include optimization theory and algorithms, evolutionary computation, smart grid, and smart city.

Victor O. K. Li (S'80–M'81–F'92) received the S.B., S.M., E.E., and Sc.D. degrees in electrical engineering and computer science from MIT in 1977, 1979, 1980, and 1981, respectively. He is a Chair Professor of information engineering and the Head of the Department of Electrical and Electronic Engineering, University of Hong Kong (HKU). He has also served as the Associate Dean of Engineering and the Managing Director of Versitech Ltd., the technology transfer and commercial arm of HKU. He served on the Board of China.com Ltd., and is currently serving on the board of Sunevision Holdings Ltd., listed on the Hong Kong Stock Exchange. He was a Professor of electrical engineering with the University of Southern California (USC), Los Angeles, CA, USA, and the Director of the USC Communication Sciences Institute. Sought by government, industry, and academic organizations, he has lectured and consulted extensively around the world. He was a recipient of numerous awards, including the PRC Ministry of Education Changjiang Chair Professorship at Tsinghua University, the U.K. Royal Academy of Engineering Senior Visiting Fellowship in Communications, the Croucher Foundation Senior Research Fellowship, and the Order of the Bronze Bauhinia Star, Government of the Hong Kong Special Administrative Region, China. He is a Registered Professional Engineer and a fellow of the IAE, and the HKIE.