Sag Source Location and Type Recognition via Attention-based Independently Recurrent Neural Network

Yaping Deng, Xinghua Liu, Rong Jia, Qi Huang, Gaoxi Xiao, and Peng Wang

***Abstract*—Accurate sag source location and precise sag type recognition are both essential to verifying the responsible party for the sag and taking countermeasures to improve power quality. In this paper, an attention-based independently recurrent neural network (IndRNN) for sag source location and sag type recognition in sparsely monitored power system is proposed. Specially, the given inputs are voltage waveforms collected by limited meters in sparsely monitored power system, and the desired outputs simultaneously contain the following information: the located lines where sag occurs; the corresponding sag types, including motor starting, transformer energizing and short circuit; and the fault phase for short circuit. In essence, the responsibility of the proposed method is to automatically establish a nonlinear function that relates the given inputs to the desired outputs with categorization labels as few as possible. A favorable feature of the proposed method is that it can be realized without system parameters or models. The proposed method is validated by IEEE 30-bus system and a real 134-bus system. Experimental results demonstrate that the accuracy of sag source location is higher than 99% for all lines, and the accuracy of sag type recognition is also higher than 99% for various sag sources including motor starting, transformer energizing and 7 different types of short circuits. Furthermore, a comparison among different monitor placements for the proposed method is conducted, which illustrates that the observability of power networks should be ensured to achieve satisfactory performance.**

***Index Terms*—Independently recurrent neural network, sag source location, sag type recognition, voltage sag, attention mechanism.**

I. Introduction

VOLTAGE sag, as one of the most critical power quality issues, is attracting extensive attention from both industry and academia [1], [2]. The sag may cause great damage to high-tech manufacturers. Statistically, more than 80% of the complaints about power quality problems are due to voltage sags [3]. Voltage sags bring not only significant economic losses, but also adverse social impacts [4]-[6]. Both accurate sag source location and precise sag type recognition are therefore essential to verifying the responsible party for the sag and taking corresponding countermeasures to improve power quality.

An analysis on the existing research dealing with sag source locations shows that most of the existing methods can be divided into two categories.

In the first category, the voltage is measured by a single power quality monitor for upstream-downstream detection. In detail, the direction of energy flow and the magnitude of disturbance energy, or the changes of the current, voltage and impedance are computed and compared before and after sag occurrence [7]. However, such approaches can only determine which side of the monitor the sag originates on. In other words, the above methods are incapable to locate the line where the sag occurs.

With the construction of smart grid, a large number of power quality monitors are being installed [8], [9]. Recent research consequently shifts the focus to be on accurately locating the line of sag source in a power system with multi-monitors [10], [11], which forms up the second category. In [10], the path or a confined area where the source likely occurs is tracked by computing each branch current deviation index, voltage-disturbance energy and current phase-angle variation. In [11], the faulty line is located by a set of rules utilizing the magnitude of three-phase voltage, voltage deviation of each phase and total average voltage deviation. However, all these methods are not suitable for locating sag source caused by motor starting or transformer energizing. Specifically, the above handcraft-established indices or rules have difficulties in capturing features for different sag sources, since ① handcrafted-feature selection greatly relies on prior domain knowledge, leading to complicated process; and ② handcrafted-feature design may ignore some potential or essential characteristics under different conditions, leading

Manuscript received: July 27, 2020; accepted: February 19, 2021. Date of CrossCheck: February 18, 2021. Date of online publication: June 22, 2021.

This work was partly supported by National Natural Science Foundation of China (No. 61903296), Key Project of Natural Science Basic Research Plan in Shaanxi Province of China (No. 2019ZDLGY18-03), Thousand Talents Plan of Shaanxi Province for Young Professionals, Project of Shaanxi Science and Technology (No. 2019JQ-329), and Doctoral Scientific Research Foundation of Xi'an University of Technology (No. 103-451116012).

Y. Deng, X. Liu (corresponding author), and R. Jia are with the School of Electrical Engineering, Xi'an University of Technology, Xi'an 710048, China (e-mail: xautdyp@xaut.edu.cn; liuxh@xaut.edu.cn; jiarong@xaut.edu.cn).

Q. Huang is with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: hwong@uestc.edu.cn).

G. Xiao and P. Wang are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore (e-mails: egxxiao@ntu.edu.sg; epwang@ntu.edu.sg).

DOI: 10.35833/MPCE.2020.000528

to limited accuracy.

To deal with the shortcomings of the above methods, data-driven approaches have been introduced to realize sag source location relying only on measured data [12]-[14]. For example, in [12], a decision-tree based approach is adopted to determine the faulty segment by using the measured voltage and current.

Furthermore, it is necessary to point out that sag type recognition is also essential in industrial applications, which is of significant value to help estimate its damage on equipment. Existing techniques, such as S-transform and Clarke transform, have been adopted to recognize sag types [15]-[17]. However, the characteristics captured in time domain and frequency domain by these techniques cannot sufficiently describe sag, resulting in low accuracy.

More significantly, it is worth mentioning that very limited research has been carried out to realize all of the following features simultaneously: ① locate the line where sag occurs in the whole power system; ② recognize the sag type for the located sag source including not only short circuits but also the causes with motor starting and transformer energizing; ③ identify the faulty phase for sag source caused by short circuits, including single-line-to-ground fault, line-to-line fault and three-phase short circuit.

Deep learning, as a representative of data-driven method, has the following favorable properties: ① it enables a machine or model to automatically learn deep and multi-aspect features from data [18]; ② feature learning can be achieved without knowledge of the power system model or human expert experiences [19]. Thus, deep learning provides a powerful solution to the above problems. Specifically, the recurrent neural network (RNN), which adopts a layer-by-layer feature extraction structure, can learn essential features from time series data [20], such as the monitored voltage waveforms. Recently, the potential of employing deep learning in power systems starts to be explored, supported by the developments of high-performance hardware [21], [22]. The standard RNN owns the capacity to retain historical information from an arbitrarily long past. However, in practical application, it faces vanishing or exploding gradient challenge. To overcome the deficiency of standard RNN, the long short-term memory (LSTM), gated recurrent unit (GRU) and independently recurrent neural network (IndRNN), as the varieties of standard RNN, have been proposed [23], [24]. Moreover, the GRU further simplifies the LSTM with higher execution efficiency [25]. The effectiveness of the proposed method therefore shall be compared against that of GRU.

In this paper, a deep learning architecture using attention-based IndRNN is proposed to simultaneously realize sag source location and type recognition. The main contributions of this paper are concluded as follows.

1) A deep learning architecture based on IndRNN, which can process longer sequences and construct deeper network, is designed to fully learn the deep and multi-aspect features from measured voltage in sparsely monitored power systems. This is beneficial to improving accuracy and preventing the vanishing or exploding gradients.

2) Attention mechanism is further combined with IndRNN to simultaneously realize sag source location, sag type recognition and faulty phase identification with as few categorization labels as possible.

3) The effectiveness and practicability of the proposed attention-based IndRNN are verified and compared against those of GRU-based model.

The remainder of this paper is organized as follows. In Section II, the reasons for choosing attention mechanism and IndRNN-based deep learning method are introduced, and then, the overall architecture of the proposed model is illustrated. In Section III, experiments are carried out on the IEEE 30-bus system and a real 134-bus system to verify the performance and feasibility. Finally, the conclusion is drawn in Section IV.

II. Model Construction and Solution

In this section, the definition of the sag source location and sag type recognition problems studied in this paper is presented. Then a brief introduction to IndRNN and attention mechanism is given. Finally, the proposed attention-based IndRNN model is illustrated.

A. Problem Formulation

The problem of sag source location and sag type recognition is defined to determine sag source and sag type simultaneously by using the information from limited meters in sparsely monitored power systems. In detail, sag source location aims at finding out the line in the whole power system where the sag occurs. Meanwhile, sag type recognition intends to detect the type of the located sag source, where monitors may not be installed. In particular, all the following sag types can be recognized: sag caused by short circuits, sag caused by motor starting and sag caused by transformer energizing. Furthermore, the faulty phase also can be identified if the sag type belongs to short circuits. In essence, this problem is to establish a nonlinear function between the given inputs and the desired outputs, where the given inputs are voltage waveforms of the monitored buses and the desired outputs are the lines of sag source and its corresponding sag type. That is to say, by using only the monitored voltage waveforms, sag source and sag type can be estimated directly without system parameters and handcrafted features.

B. Brief Introduction of IndRNN

The IndRNN, as an improved version of RNN, is proposed in [26]. It can be described as:

$$\boldsymbol{h}_t = \sigma(\boldsymbol{W}\boldsymbol{x}_t + \boldsymbol{U} \odot \boldsymbol{h}_{t-1} + \boldsymbol{b}) \tag{1}$$

where $\odot$ is an operational symbol denoting Hadamard product; $\boldsymbol{W}$ and $\boldsymbol{U}$ are the matrices to show the connection weights for the current inputs and the recurrent inputs, respectively; $\boldsymbol{b}$ is the bias vector; $\boldsymbol{x}_t$ and $\boldsymbol{h}_t$ are the inputs and hidden states at time step t, respectively; and σ is a nonlinear activation function such as sigmoid or tanh functions.

In IndRNN, the processing of the inputs is independent at different time steps and can be implemented in parallel. Besides, for IndRNN, each neuron in one layer is also independent from others. Aiming at the n^{th} neuron, the hidden state $\boldsymbol{h}_{n,t}$ can be calculated as:

$$\boldsymbol{h}_{n,t}=\sigma(\boldsymbol{w}_n\boldsymbol{x}_t+\boldsymbol{u}_n\boldsymbol{h}_{n,t-1}+\boldsymbol{b}_n) \tag{2}$$

where $\boldsymbol{w}_n$, $\boldsymbol{u}_n$, and $\boldsymbol{b}_n$ are the input weights, the recurrent weights, and the bias vector of the n^{th} neuron, respectively.

Since IndRNN addresses the gradient exploding and vanishing problems over time, the gradient can be efficiently propagated over different time steps. Therefore, the IndRNN can be substantially deeper and longer. Compared with the standard RNN and its variants, such as LSTM and GRU, the IndRNN owns the following advantages, as demonstrated in [26].

1) The IndRNN can be easily regulated to prevent vanishing or exploding gradients.

2) Long-term memory can be kept with IndRNNs to process long sequences.

3) Multiple layers of IndRNNs can be efficiently stacked to increase the depth of the network.

In brief, the IndRNN is able to process longer sequences and construct deeper networks, which enables the IndRNN to fully extract the features of the inputs.

In our study, to fully learn deep and multi-aspect features from the monitored voltage waveforms, IndRNN is designed to construct deep learning architecture. It helps improve the accuracy of sag source location and sag type recognition, as we will demonstrate later.

C. Attention Mechanism

Learning from the selective attention mechanism in human visual system, the idea of attention mechanism has been proposed to assign different weights to different inputs according to their respective importances [27]. Such a mechanism can focus on significant information while ignoring the insignificant information. In fact, not all features extracted by deep learning are equally meaningful for establishing the relationship between inputs and outputs. Hence, it is an effective method to pay more attention to the most relevant and important information [28]. Recently, attention mechanism has been successfully employed in a wide variety of tasks, such as sign language recognition, object detection, action recognition, defect detection and so on [29], [30]. In general, attention mechanism is usually employed to further optimize the extracted features [31]. In our work, the attention mechanism is combined with IndRNN to explore the optimized features of monitored voltage.

D. Attention-based IndRNN Model

As described in Fig. 1, a deep learning architecture employing attention-based IndRNN is proposed for sag source location and sag type recognition, where J is the total number of monitors, and L is the total number of layers. The whole model is characterized by the following modules: input layer, IndRNN layer, attention layer and output layer. In particular, several IndRNN layers are stacked together for the deep and global feature extraction. Meanwhile, the attention mechanism is adopted to capture their respective features for source location and type recognition. For clarity, we give an example to illustrate the detailed process of the proposed model as follows.

1) Input Layer

The input layer is the bottom component of this model, whose outputs are directly sent to the IndRNN layer. The layer is responsible for the measured data pre-processing. Then, the processed data $\boldsymbol{X}$ are introduced to the model. The shape of the inputs $\boldsymbol{X}$ is designed as [*Batch_size*, *Time_steps*, *Dimensions*], where the *Batch_size* is the number of periods for measured voltage in one batch; the *Time_steps* is the product of the number of sampling points in one cycle and the number of periods in one batch; and the *Dimensions* is the dimension of the input data.

Due to the fact that three-phase voltages at each metered bus are needed in our model, the measured data $\boldsymbol{X}_j$ for monitor j can be presented as:

$$\boldsymbol{X}_j=\begin{bmatrix} x^j_{a,1} & x^j_{a,2} & \cdots & x^j_{a,t} & \cdots & x^j_{a,Time\_steps-1} & x^j_{a,Time\_steps} \\ x^j_{b,1} & x^j_{b,2} & \cdots & x^j_{b,t} & \cdots & x^j_{b,Time\_steps-1} & x^j_{b,Time\_steps} \\ x^j_{c,1} & x^j_{c,2} & \cdots & x^j_{c,t} & \cdots & x^j_{c,Time\_steps-1} & x^j_{c,Time\_steps} \end{bmatrix} \tag{3}$$

where $x^j_{a,t}$, $x^j_{b,t}$, and $x^j_{c,t}$ are the voltage amplitudes of phases A, B, C measured by monitor j at time step t, respectively; and $t=1,2,\ldots,Time\_steps$.

Here, we take $x^j_{a,t}$ as an example to explain the meaning of the element in the above matrix. For batch m, the measured data can be given as:

$$\boldsymbol{X}^m=[\boldsymbol{x}^m_1 \quad \boldsymbol{x}^m_2 \quad \cdots \quad \boldsymbol{x}^m_t \quad \cdots \quad \boldsymbol{x}^m_{Time\_steps-1} \quad \boldsymbol{x}^m_{Time\_steps}]^{\mathrm{T}}=$$
$$\begin{bmatrix} x^{1,m}_{a,1} & x^{1,m}_{a,2} & \cdots & x^{1,m}_{a,t} & \cdots & x^{1,m}_{a,Time\_steps-1} & x^{1,m}_{a,Time\_steps} \\ x^{1,m}_{b,1} & x^{1,m}_{b,2} & \cdots & x^{1,m}_{b,t} & \cdots & x^{1,m}_{b,Time\_steps-1} & x^{1,m}_{b,Time\_steps} \\ x^{1,m}_{c,1} & x^{1,m}_{c,2} & \cdots & x^{1,m}_{c,t} & \cdots & x^{1,m}_{c,Time\_steps-1} & x^{1,m}_{c,Time\_steps} \\ x^{2,m}_{a,1} & x^{2,m}_{a,2} & \cdots & x^{2,m}_{a,t} & \cdots & x^{2,m}_{a,Time\_steps-1} & x^{2,m}_{a,Time\_steps} \\ x^{2,m}_{b,1} & x^{2,m}_{b,2} & \cdots & x^{2,m}_{b,t} & \cdots & x^{2,m}_{b,Time\_steps-1} & x^{2,m}_{b,Time\_steps} \\ x^{2,m}_{c,1} & x^{2,m}_{c,2} & \cdots & x^{2,m}_{c,t} & \cdots & x^{2,m}_{c,Time\_steps-1} & x^{2,m}_{c,ime\_steps} \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ x^{J,m}_{a,1} & x^{J,m}_{a,2} & \cdots & x^{J,m}_{a,t} & \cdots & x^{J,m}_{a,Time\_steps-1} & x^{J,m}_{a,Time\_steps} \\ x^{J,m}_{b,1} & x^{J,m}_{b,2} & \cdots & x^{J,m}_{b,t} & \cdots & x^{J,m}_{b,Time\_steps-1} & x^{J,m}_{b,Time\_steps} \\ x^{J,m}_{c,1} & x^{J,m}_{c,2} & \cdots & x^{J,m}_{c,t} & \cdots & x^{J,m}_{c,Time\_steps-1} & x^{J,m}_{c,Time\_steps} \end{bmatrix}^{\mathrm{T}} \tag{4}$$

where $\boldsymbol{x}^m_t$ is the three-phase voltage amplitude at time step t for batch m, which is shown as (5); $x^{j,m}_{a,t}$, $x^{j,m}_{b,t}$, and $x^{j,m}_{c,t}$ are the voltage amplitudes of phases A, B, C measured by monitor j at time step t for batch m, respectively; and $t=1,2,\ldots,Time\_steps$.

$$\boldsymbol{x}^m_t=[x^{1,m}_{a,t} \quad x^{1,m}_{b,t} \quad x^{1,m}_{c,t} \quad x^{2,m}_{a,t} \quad x^{2,m}_{b,t} \quad x^{2,m}_{c,t} \quad \cdots \quad x^{J,m}_{a,t} \quad x^{J,m}_{b,t} \quad x^{J,m}_{c,t}]^{\mathrm{T}} \tag{5}$$

Finally, the inputs $\boldsymbol{X}$ for different batches m ($m\in\{1, 2, \ldots, Batch\_size\}$) is given as (6) and it is further taken into the next IndRNN layer.

$$\boldsymbol{X}=[\boldsymbol{X}^1 \quad \boldsymbol{X}^2 \quad \cdots \quad \boldsymbol{X}^m \quad \cdots \quad \boldsymbol{X}^{Batch\_size-1} \quad \boldsymbol{X}^{Batch\_size}] \tag{6}$$

The sketch of the shape of inputs $\boldsymbol{X}$ is demonstrated in Fig. 2 for further understanding. It is observed that the shape of inputs $\boldsymbol{X}$ is [*Batch_size*, *Time_steps*, 3*J*].

2) IndRNN Layer

In the IndRNN layer, multiple IndRNNs with batch normalization, are stacked together for global and deep characteristics extraction. The inputs $\boldsymbol{X}$ at each time step t are continuously sent to the IndRNN layer.

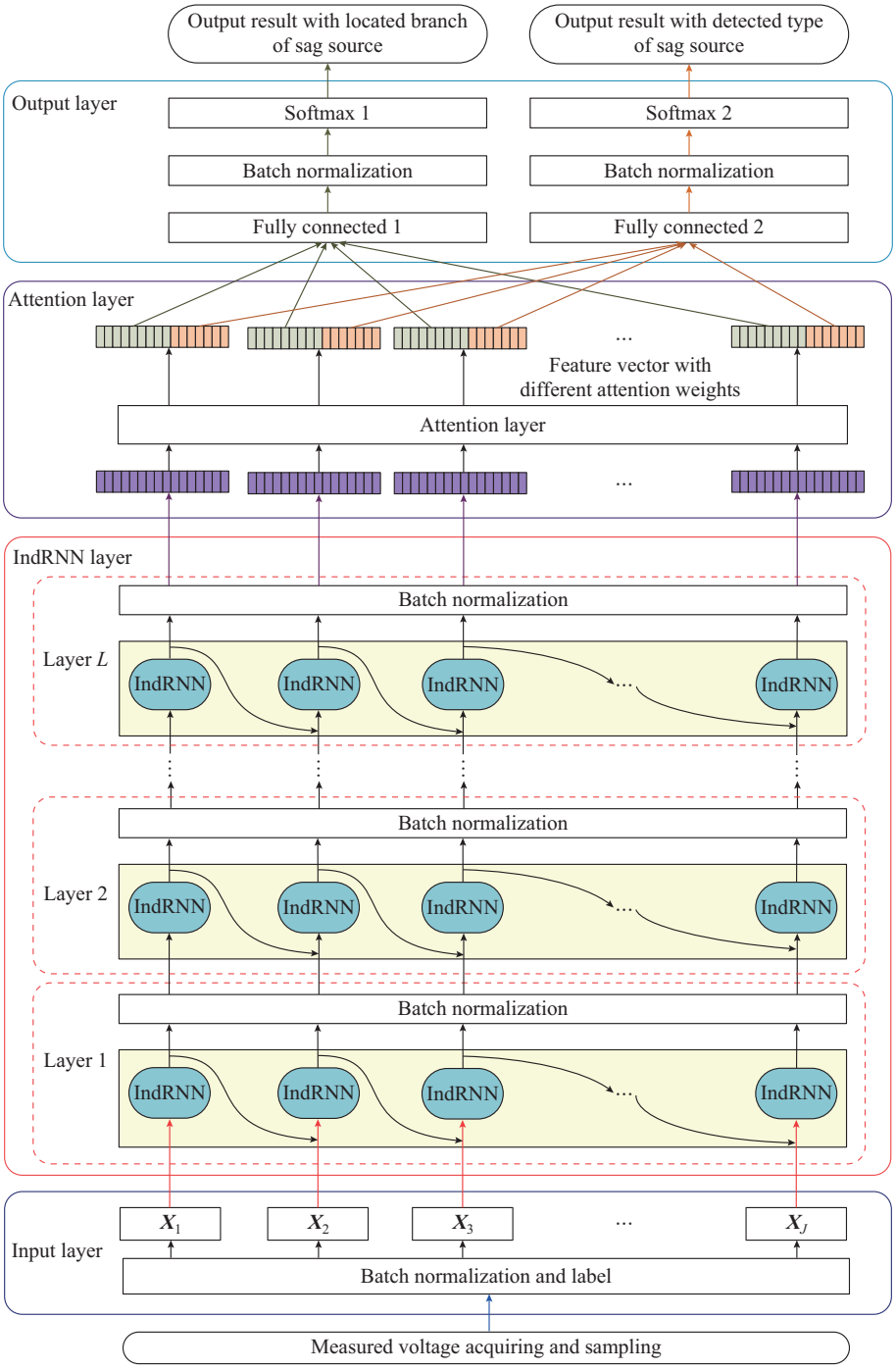

Fig. 1. Overall architecture of proposed model.

Here, let the number of neurons in each IndRNN denoted as $n_{neurons}$, the shape of the hidden states $\boldsymbol{h}_t$ and $\boldsymbol{h}_{t-1}$ in each IndRNN is [*Batch_size*, $n_{neurons}$], and the shapes of $\boldsymbol{W}$, $\boldsymbol{U}$ and $\boldsymbol{b}$ in each IndRNN are [$n_{neurons}$, *Dimensions*], [$n_{neurons}$] and [$n_{neurons}$], respectively. The batch normalization follows each IndRNN to further zero-center and normalize the data from IndRNN. The whole operation of batch normalization is summarized in (7)-(10) [32].

$$\boldsymbol{\mu}_{\beta} = \frac{1}{Batch\_size}\sum_{t=1}^{Batch\_size}\boldsymbol{h}_t \tag{7}$$

$$\sigma_{\beta}^2 = \frac{1}{Batch\_size}\sum_{t=1}^{Batch\_size}(\boldsymbol{h}_t - \boldsymbol{\mu}_{\beta})^2 \tag{8}$$

$$\hat{\boldsymbol{h}}_t = \frac{\boldsymbol{h}_t - \boldsymbol{\mu}_{\beta}}{\sqrt{\sigma_{\beta}^2 + \varsigma}} \tag{9}$$

$$\boldsymbol{y}_t = \gamma\hat{\boldsymbol{h}}_t + \boldsymbol{\beta} \tag{10}$$

where $\boldsymbol{\mu}_{\beta}$ is the empirical mean; σ_{β}^2 is the empirical standard deviation; γ and $\boldsymbol{\beta}$ are the scaling parameter and shifting parameters to be optimized, respectively; ς is a smoothing term to avoid division by zero (typically 10^{-3}); $\hat{\boldsymbol{h}}_t$ is the zero-cen-

tered and normalized inputs; and $\boldsymbol{y}_t$ is the outputs of batch normalization operation, which is a scaled and shifted version of its inputs.

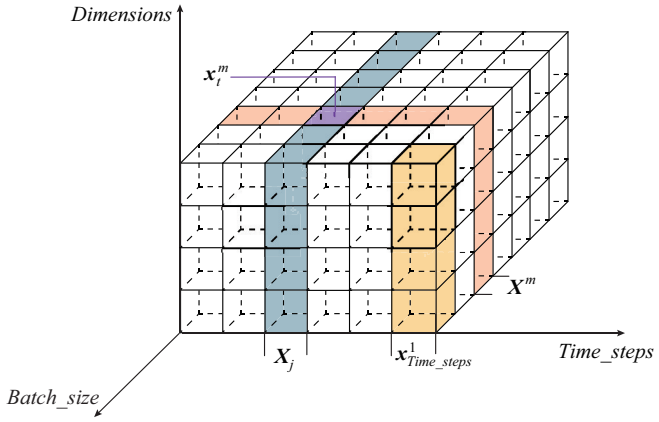

Fig. 2. Sketch of shape of inputs $\boldsymbol{X}$.

In this layer, at time step t, if the outputs of last IndRNN are denoted as $\boldsymbol{h}_t^L$, and the outputs of the last batch normalization are computed as $\boldsymbol{y}_t^L$, the final outputs of the IndRNN layer $\boldsymbol{Y}$ can be written as:

$$\boldsymbol{Y}=\left[\boldsymbol{y}_1^L \quad \boldsymbol{y}_2^L \quad \cdots \quad \boldsymbol{y}_t^L \quad \cdots \quad \boldsymbol{y}_{Time\_steps-1}^L \quad \boldsymbol{y}_{Time\_steps}^L\right] \tag{11}$$

where t = 1, 2, …, *Time_steps*.

The shape of $\boldsymbol{y}_t^L$ is [*Batch_size*, $n_{neurons}$], and the shape of $\boldsymbol{Y}$ is [*Batch_size*, *Time_steps*, $n_{neurons}$]. In other words, for one of the batch inputs, there are a total of $n_{neurons}$ feature vectors at each time step to express $\boldsymbol{X}$, which come from the input layer. In general, aiming at the IndRNN layer, $\boldsymbol{X}$ is the inputs, and $\boldsymbol{Y}$ is the outputs. Hence, $\boldsymbol{Y}$ can be regarded as the extracted characteristics from $\boldsymbol{X}$ through the IndRNN layer.

3) Attention Layer

According to the problem formulation in Section II-A, sag source location and sag type recognition can be regarded as a classification problem, or more specifically, a supervised learning classification task. Then, the number of categorization labels should be the product of numbers of lines and sag types. For IEEE 30-bus system with 37 transmission lines, the number of categorization labels should be set as 333 if the sag type is 9 ($37\times9=333$). Obviously, the number of categorization labels explodes for more complicated networks. To effectively reduce the number of categorization labels, an attention layer is introduced into our model, where the number of categorization labels can be reduced to be the sum of numbers of lines and sag types. That is to say, the number of categorization labels decreases from 333 to 46 ($37+9=46$) with the attention layer.

In detail, from the IndRNN layer, the above extracted characteristics $\boldsymbol{Y}$ are sent to the attention layer, where the feature $\boldsymbol{Y}$ is further analyzed. Note that the feature $\boldsymbol{Y}$ contains two parts. One part intends to reflect the relationship between input voltage waveforms and lines in the whole power system, which is helpful for sag source location. The other part aims to establish the relationship between input voltage waveforms and sag types, which is useful to recognize the type of located sag sources. Here, the attention layer is combined with the previous IndRNN layer, where the extracted global features from the IndRNN layer are divided into two parts. One part of the extracted global features is adopted to locate sag sources, while the other part is applied to recognize sag types.

As presented in (12), an attention function is adopted to calculate the attention weight.

$$Attention(\boldsymbol{y}_t^L,\boldsymbol{q})=\frac{\exp(s(\boldsymbol{y}_t^L,\boldsymbol{q}))}{\sum_{t=1}^{Time\_steps}\exp(s(\boldsymbol{y}_t^L,\boldsymbol{q}))} \tag{12}$$

where $Attention(\boldsymbol{y}_t^L,\boldsymbol{q})$ is the attention function; $s(\boldsymbol{y}_t^L,\boldsymbol{q})$ is the score function which can be calculated by the additive attention or dot-product attention; and $\boldsymbol{q}$ is a query vector, which can be regarded as an indirect expression of the output results of sag type and sag source.

Specifically, $\boldsymbol{q}$ can be taken as referring to the states of fully connected 1 and fully connected 2 in the output layer in Fig. 1. The additive attention $s_{additive}(\boldsymbol{y}_t^L,\boldsymbol{q})$ and dot-product attention $s_{dot\text{-}product}(\boldsymbol{y}_t^L,\boldsymbol{q})$ are the two most frequently used attention functions, and correspondingly, their definitions can be described as:

$$s_{additive}(\boldsymbol{y}_t^L,\boldsymbol{q})=\boldsymbol{v}^{\mathrm{T}}\tanh(\boldsymbol{W}_y\boldsymbol{y}_t^L+\boldsymbol{W}_q\boldsymbol{q}) \tag{13}$$

$$s_{dot\text{-}product}(\boldsymbol{y}_t^L,\boldsymbol{q})=(\boldsymbol{y}_t^L)^{\mathrm{T}}\boldsymbol{q} \tag{14}$$

where $\boldsymbol{W}_y$, $\boldsymbol{W}_q$, and $\boldsymbol{v}$ are the weight matrices.

Then, the attention mechanism is presented as:

$$Attention(\boldsymbol{Y},\boldsymbol{q})=\sum_{t=1}^{Time\_steps}\boldsymbol{y}_t^L Attention(\boldsymbol{y}_t^L,\boldsymbol{q}) \tag{15}$$

Finally, features with different attention weights for line location and type recognition of sag sources are achieved, as described in Fig. 1. Hence, for one batch input, the $n_{neurons}$ feature of $\boldsymbol{Y}$, which has been received from the IndRNN layer at each time step, can be separated into two parts: θ_{line} and ψ_{type}. The θ_{line} feature is employed for sag source location, and the ψ_{type} feature is adopted for sag type recognition of the located sag source.

4) Output Layer

The output layer is the top component of this model and it is responsible for determining the line and type of sag source according to their feature vectors from the attention layer, respectively. Assuming that the number of lines in the whole system is n_{line}, it can be concluded that the shape of fully connected 1, as presented in Fig. 1, is [*Batch_size*, n_{line}]. And the shape of fully connected 2, as presented in Fig. 1, is [*Batch_size*, n_{type}], where n_{type} is the number of sag types. And then, the above results can be directly fed to their batch normalization layer for further data regulation. Finally, the above outputs are sent to their Softmax respectively, where the corresponding probability indicates the line and type of sag source.

Using the proposed model as presented in Fig. 1, the parameters or hyper-parameters can be well learned with a large amount of training data, where the weights and parameters or hyper-parameters of the model are optimized to match the given inputs with their desired outputs. Then, the well-trained model can be used for sag source location and sag type recognition in the practical field. Figure 3 demonstrates the whole flow chart for sag source location and sag

type recognition in a real application. Firstly, the measured three-phase voltage waveforms at each metered bus are acquired and sampled. Secondly, the above sampled data are directly fed into the proposed model as presented in Fig. 1 once the sag is detected. Here, the sampled data are batched, normalized, feature extracted and optimized. Finally, the line on which the sag occurs is located. Meanwhile, the type of the located sag source is also recognized. Figure 4 illustrates how the proposed method is adopted to realize sag source location and sag type recognition.

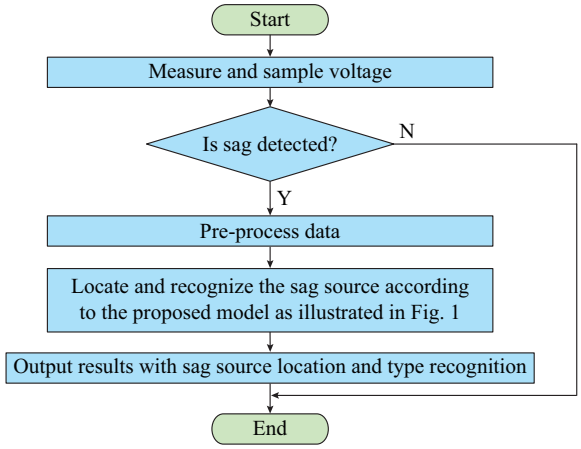

Fig. 3. Flow chart for sag source location and sag type recognition.

III. Experimentation and Results

In this section, the validation of the proposed method with IEEE 30-bus system and a real 134-bus system is presented, and the performance is also analyzed.

A. Dataset Description

The effectiveness of the proposed method has been firstly tested with IEEE 30-bus system, as presented in Fig. 5. The system comprises 6 sources, 30 buses, and 37 transmission lines. Here, it is noted that the number and placement of monitors play an important role for sag source location and sag type recognition. According to the method presented in [33], 4 meters are located at bus 2, bus 15, bus 21, and bus 25, respectively, as shown in Fig. 5. These meters are distributed in the network to ensure that the whole system is observable.

For preparation, the dataset should be generated to train the model and test the performance of the proposed method. Specially, a total of 9 mutually exclusive types of voltage sag causes, including phase A grounding, phase B grounding, phase C grounding, phase A and phase B fault, phase A and phase C fault, phase B and phase C fault, three-phase fault, motor starting, and transformer energizing, are considered in this paper. Moreover, the faulty resistance variation, the transition resistance variation, the sag duration change, the active power variation, and the reactive power variation are also considered, where the range of faulty resistance is between 0.5 Ω and 10 Ω, the range of ground resistance is between 0.01 Ω and 5 Ω, and the range of the active power and reactive power is between 0.95 p.u. and 1.05 p.u., respectively.

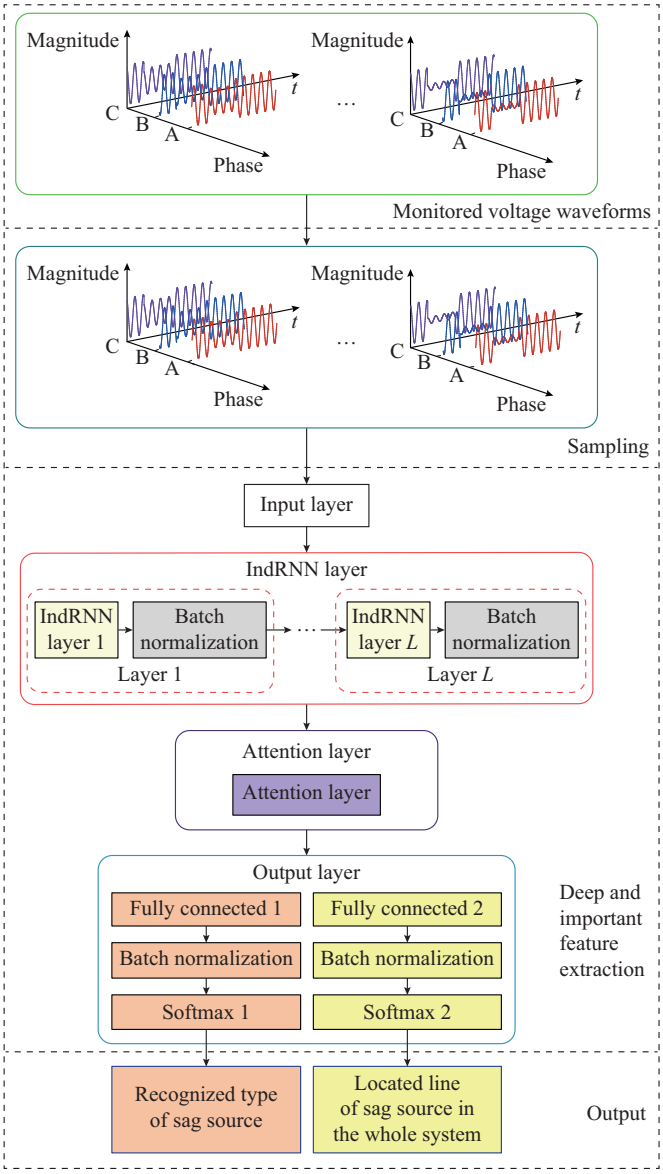

Fig. 4. Procedure for sag source location and type recognition.

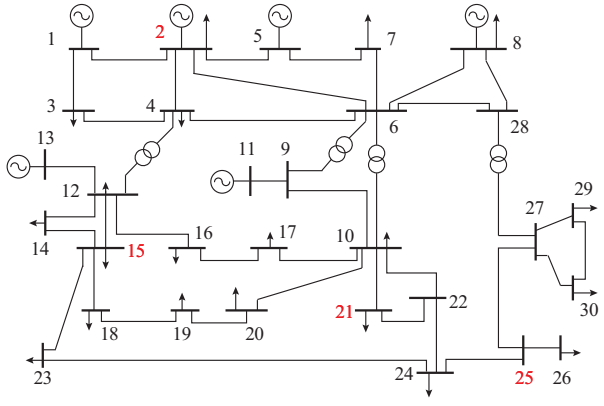

Fig. 5. IEEE 30-bus system.

As illustrated in Table I, numerous simulations are operated to generate the dataset, where n_{IndRNN} is the number of IndRNN layers. All data are collected through MATLAB/Simulink simulation.

TABLE I
PARAMETERS AND HYPER-PARAMETERS OF MODEL

Parameter and hyper-parameter	Value
Time_steps	62
Dimensions	12
Batch_size	36
Learning rate	0.00015
n_{IndRNN}	8
$n_{neurons}$	512
θ_{line}	406
ψ_{type}	106
n_{line}	37
n_{type}	9

Figure 6 demonstrates the flow chart for dataset generation. All simulated data are complementarily split into training set and testing set. Here, 80% of the simulated data are used for training, and 20% of the simulated data are used for testing (without being used in training process). For each sag type at each line, there are 400 samples for training and 100 samples for testing.

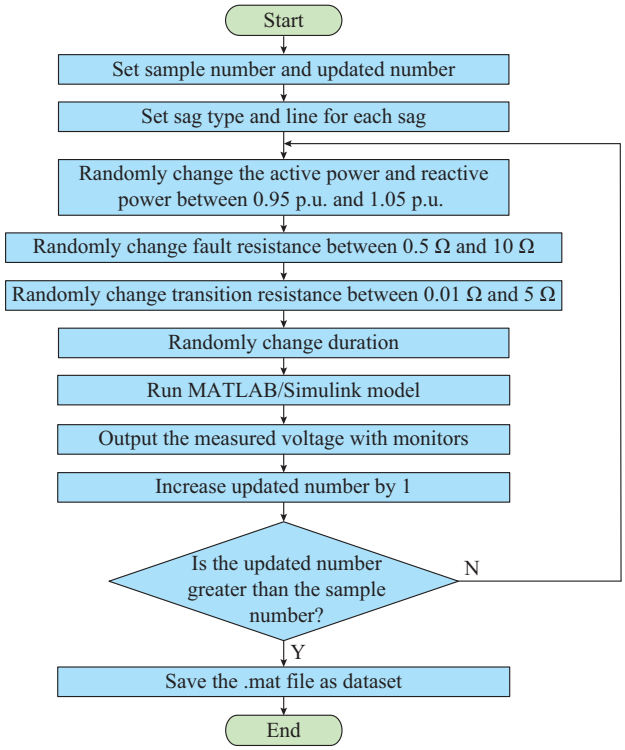

Fig. 6. Flow chart for dataset generation.

The detailed procedure for data separation can be described as follows.

Step 1: for each sag type at each line, a total of 500 samples are generated according to Fig. 6. Then, 400 samples are randomly extracted from these 500 samples to form up the training set. The remaining 100 samples are taken as the testing set.

Step 2: the final training set is wholly formed with 333 parts (9×37=333), which are generated according to *Step 1*. Similarly, the final testing set is also wholly formed with 333 parts. It is obvious that the final training set and testing set are comprehensive and balanced.

To test the feasibility of our method, we conduct simulations for the following different cases. The first case is that a motor starting occurs between bus 15 and bus 18. Correspondingly, Fig. 7(a)-(d) illustrates the monitored voltages through 4 meters at bus 2, bus 15, bus 21 and bus 25, respectively. The second case is for a scenario that a three-phase short circuit occurs at the same line as that in the first case, and the monitored voltages are demonstrated in Fig. 8(a)-(d). The third case is that a three-phase short circuit occurs on another line between bus 8 and bus 9, where the monitored voltages are shown in Fig. 9(a)-(d).

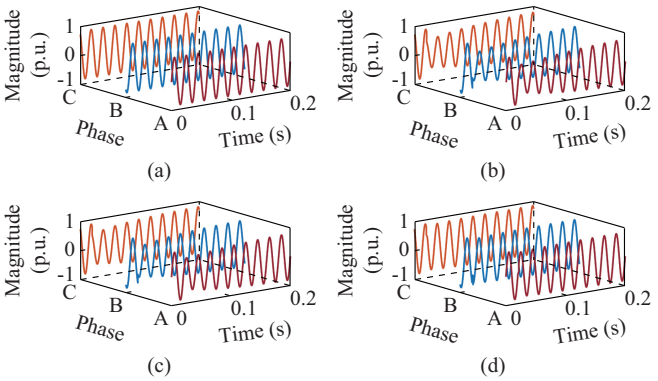

Fig. 7. Monitored voltage with motor starting between bus 15 and bus 18. (a) Monitored voltage of bus 2. (b) Monitored voltage of bus 15. (c) Monitored voltage of bus 21. (d) Monitored voltage of bus 25.

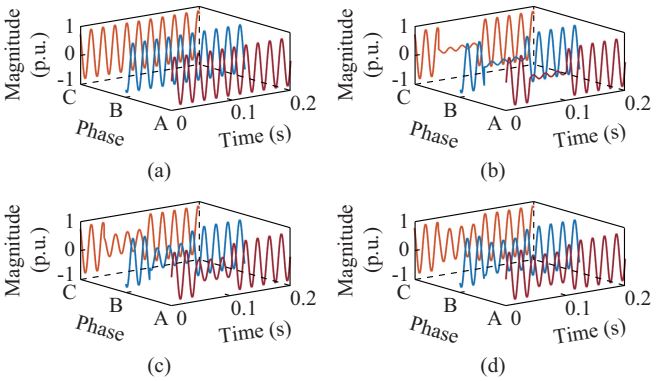

Fig. 8. Monitored voltage with three-phase short circuit between bus 15 and bus 18. (a) Monitored voltage of bus 2. (b) Monitored voltage of bus 15. (c) Monitored voltage of bus 21. (d) Monitored voltage of bus 25.

As clearly validated by Fig. 7 and Fig. 8, it can be observed that on the same line, different sag types will lead to different voltage waveforms. Furthermore, compared Fig. 8 with Fig. 9, it is noted that the different sag sources (sag lines) with the same sag type also lead to different voltage waveforms. Therefore, there is a one-to-one correspondence between the monitored voltage waveforms and sag sources or sag types. Conversely, sag source and sag type can be detected through monitored voltage waveforms.

Obviously, the relationship or function between the monitored voltage waveforms and sag sources or sag types is non-

linear. The proposed model, as presented in Fig. 1, is adopted to automatically associate the monitored voltage waveforms with sag source and sag type.

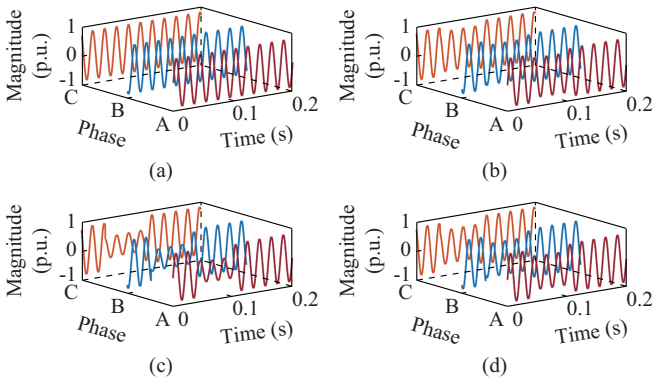

Fig. 9. Monitored voltage with three-phase short circuit between bus 6 and bus 7. (a) Monitored voltage of bus 2. (b) Monitored voltage of bus 15. (c) Monitored voltage of bus 21. (d) Monitored voltage of bus 25.

B. Design of Parameters and Hyper-parameters

It should be pointed out that the proper selections of parameters or hyper-parameters are necessary for the proposed architecture illustrated in Fig. 1 to achieve satisfactory performance. There are mainly the following parameters or hyper-parameters to be determined, including *Time_steps*, *Dimensions*, *Batch_size*, *Learning rate*, n_{IndRNN}, $n_{neurons}$, θ_{line}, ψ_{type}, n_{line} and n_{type}. Detailed parameters or hyper-parameters are given in Table I.

Firstly, it is noticed that *Time_steps* is related to sampling frequency, accuracy and computational burden. In our work, the frequency of fundamental voltage is 50 Hz, and thus, according to Shannon's theorem, the sampling frequency is designed as 1050 Hz to guarantee a better restitution performance. Then, the *Time_steps* is chosen as 62 to achieve a good balance between accuracy and complexity.

Secondly, as presented in Section III-A, the number of monitors J is 4, and three-phase voltages at each monitor are acquired. The *Dimensions* should be chosen as 12. That is, $3\times4=12$.

Thirdly, the n_{type} is 9 since 9 mutually exclusive sag sources are considered, and the n_{line} is 37 since there are 37 transmission lines in the IEEE 30-bus system.

Finally, it is worth mentioning that *Batch_size* is closely related to the real-time performance. Figure 10 demonstrates the relationship between the *Batch_size* and the algorithm execution time, and that between the *Batch_size* and the data acquisition time, respectively.

In Fig. 10, the algorithm execution time is strongly associated with calculation ability of computer hardware. It is tested by changing *Batch_size* from 1 to 69 on our experimental platform as described in Section III-C. The data acquisition time denotes the product of *Bacth_size* and the fundamental period, as presented in Fig. 10, where the fundamental period is 20 ms. The intersection point of the above two curves happens when the *Batch_size* is between 35 and 36. That is to say, the *Batch_size* should be greater than or equal to 36 to guarantee the on-line performance for our current hardware, and the corresponding execution time is greater than or equal to 0.72 s. Hence, the proposed method can realize real-time source location and type recognition under this condition. Since a too large *Batch_size* may lead to high memory resource occupancy, we chose the *Batch_size* as 36, where the corresponding time is 0.72 s.

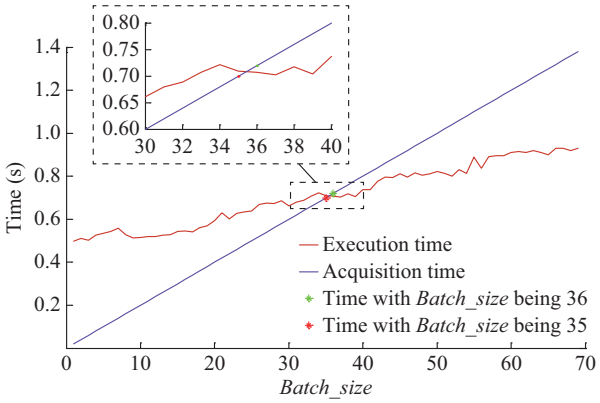

Fig. 10. Relationship between *Batch_size* and algorithm execution time, and that between *Batch_size* and data acquisition time.

Besides, it is necessary to note that the *Learning rate*, n_{IndRNN}, $n_{neurons}$, θ_{line} and ψ_{type} also should be designed carefully and balanced well to gain satisfactory results. However, there is no effective way to theoretically figure out the best values of these hyper-parameters. Fortunately, for each of the above-mentioned hyper-parameters, the qualitative analysis on the relationship between hyper-parameters and accuracy, model complexity, computational cost has been developed with a lot of experimental results. As for n_{IndRNN} and $n_{neurons}$, more features or information will be captured with larger numbers of n_{IndRNN} and $n_{neurons}$, while the computation burden and time cost also increase correspondingly. Therefore, the above-mentioned hyper-parameters are selected to achieve a good balance between the accuracy and complexity by the cut-and-trial method.

C. Verification of Effectiveness of Proposed Method

The overall algorithm is implemented by using the TensorFlow software. The software environment adopted is Anaconda Python 3.6.8. All the experiments have been carried out on a workstation equipped with an Intel i7-8700K processor and a GTX 1080 Ti×2 graphics processing unit.

Since cross-validation can effectively avoid overfitting, and it is helpful for checking how well a model has generality for new data [34], the 5-fold cross validation technique has been adopted to evaluate the performance of the proposed model. In detail, the above-mentioned training set is further separated into complementary subsets, and each model is trained against a different combination of these subsets and validated against the remaining parts. The single test is finally measured against the testing set to confirm the performance of the proposed method. Moreover, two indices, including the respective *Accuracy* for sag source location and type recognition, the respective *Loss* based on cross-entropy for sag source location and type recognition, are adopted to evaluate the performance of the proposed model. The defini-

tions of *Accuracy* and *Loss* are presented as shown in (16) and (17), respectively.

$$Accuracy = \frac{N_{correct}}{N_{total}} \times 100\% \tag{16}$$

$$Loss = -\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{K} y_k^i \ln \hat{y}_k^i \tag{17}$$

where $N_{correct}$ is the number of correct results; N_{total} is the number of total tests; $\hat{y}_k^i$ is the outputs of Softmax in output layer; y_k^i is the 0-1 parameter, and it is equal to 1 if the target class for the i^{th} instance is k, otherwise, it is equal to 0; M is the total number of instances; and K is the total number of target classes.

Figure 11 demonstrates the attention weights in the attention layer, as presented in Fig. 1. Here, all these attention weights are dynamically decided by the attention layer in Fig. 1, according to the measured voltage. The measured three-phase voltages corresponding to the attention weights in Fig. 11 are illustrated in Fig. 12.

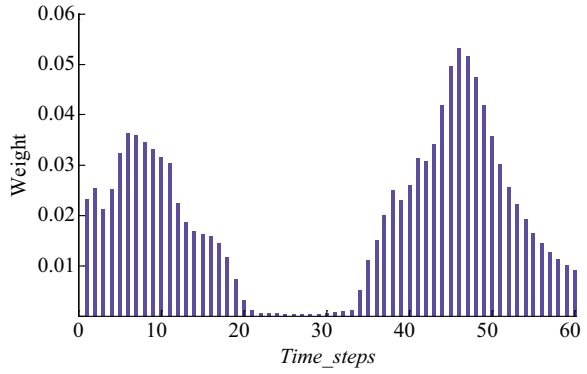

Fig. 11. Relationship between *Time_steps* and attention weight.

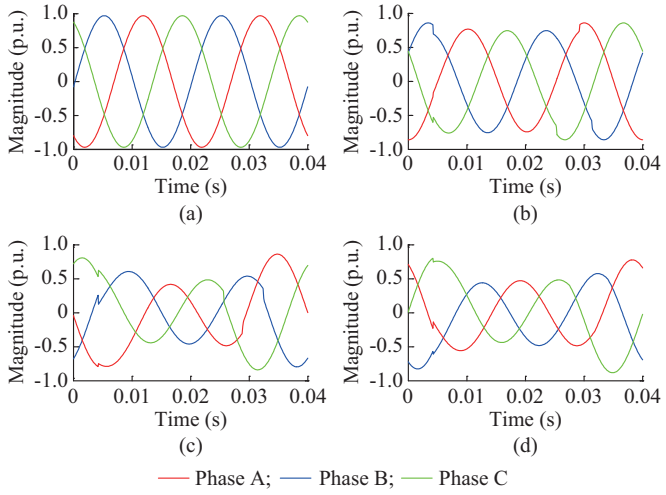

Fig. 12. Measured three-phase voltages corresponding to Fig. 11. (a) Measured three-phase voltages of bus 2. (b) Measured three-phase voltages of bus 15. (c) Measured three-phase voltages of bus 21. (d) Measured three-phase voltages of bus 25.

From Fig. 11 and Fig. 12, the following conclusions can be obtained.

1) The attention weights are different at different time steps. This means that the attention layer in Fig. 1 enables the extracted features ***Y*** from the IndRNN layer to have different importance for sag source location and sag type recognition. The greater attention weight, the more important it is for sag source location and sag type recognition. It is also noticed that attention weights tend to be larger during changing time such as starting and ending moments in Fig. 12, while attention weights tend to be smaller for stable situation in Fig. 12.

2) The extracted global features ***Y*** can be automatically divided into two parts: the first half part is mainly adopted to locate the sag source, and the latter half part is mainly used to recognize the sag type. Note that compared with the sag source location, the sag type recognition requires more detailed information during changing time. Therefore, the attention weights for time steps 43-53 are relatively higher than those for other time steps in Fig. 11.

In short, the attention layer does play a significant role in capturing their separate information for sag source location and type recognition.

The relationship between the epochs and loss for sag source location and type recognition is indicated in Fig. 13. Further, Fig. 14 shows the relationship between the epoch and the accuracy for the sag source location and sag type recognition, respectively.

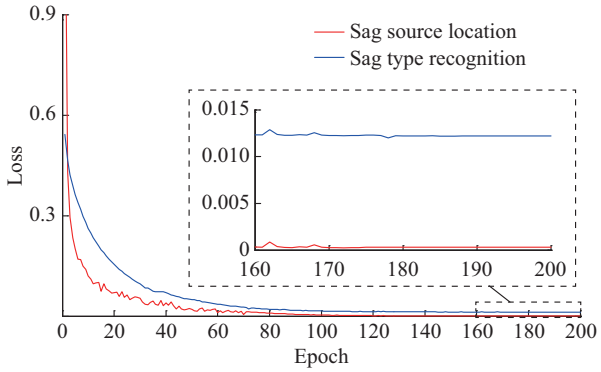

Fig. 13. Relationship between epoch and loss.

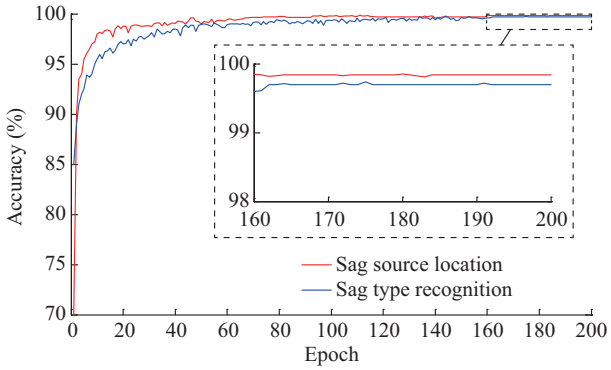

Fig. 14. Relationship between epoch and accuracy.

From Fig. 13, it can be observed that the model converges gradually with steady decreasing loss along the training process, where around 40 epochs are needed for the proposed deep learning architecture to achieve a loss less than 10^{-1}.

And then, the model provides the best system performance with the loss value close to 10^{-9}. This indicates that the IndRNN can be trained robustly and also with a small value of loss.

However, it also can be concluded that the convergence speed of the trained model is not so fast. Actually, there exists a balance between convergence speed and model performance, where a faster convergence may be reached at a cost of degraded performance. For example, according to the literature [34], although the learning rate can be increased so as to further speed up the convergence rate, the convergence is now fast but the network accuracy is suboptimal. Another influencing factor, namely *Batch_size*, is not only related to convergence rate, but also relevant to real-time performance. The detailed analysis about the selection of *Batch_size* is demonstrated in Section III-B.

Although a sophisticated method shall be able to reach faster convergence speed, it usually requires complicated operations. For example, the *Learning rate* may be set as an exponential decay function to allow a changeable *Learning rate* rather than a constant *Learning rate*. Consequently the convergence speed may become faster, but additional parameters need to be tuned.

Since an offline training method is adopted in this paper, the convergence speed is not a very critical concern.

Due to the considerations as discussed above, we choose a relatively simple method to set the *Learning rate* with a relatively low speed but reliable performance.

Moreover, as illustrated in Fig. 14, it can be further concluded that the testing accuracies for both the sag source location and the sag type recognition are higher than 99%. Specifically, sag sources including not only short circuits but also causes with motor starting and transformer energizing are considered in this paper. It is noticed that the features of motor starting cases and transformer energizing causes are similar, where only one transient process (also namely the starting process) can be used. Therefore, it is not easy to achieve 100% accuracy resulting in sag type recognition.

D. Comparison Among Different Monitor Placements

For the proposed method, the basic requirement is that the measured inputs $\boldsymbol{X}$ through monitors, as presented in (4), should reflect each sag in the considered power system. In other words, monitor placement should enable each sag in the considered network to be detected. Monitor placement therefore can affect the performance as different monitor placements generate different observabilities for the whole power system. Ideally, monitors can be installed at all buses in the power system, in whose case any sag can be detected. Such an approach, however, is not practical for economic reasons. A reasonable approach is to install only a limited number of monitors at the selected buses where the whole power system is observable. Here, we take the method proposed in [33] as reference for monitor selection, where uncertainties associated with transition resistance are considered. It has been proven that this method achieves more applicable and satisfactory performance compared with the traditional methods [33].

For the proposed method in this paper, the influences of the number and allocation of monitors on accuracy have been verified, as illustrated in Table II and Fig. 15.

TABLE II
ACCURACY FOR DIFFERENT MONITOR PLACEMENTS

Monitor placement	Number of monitors	Accuracy of sag source location (%)	Accuracy of sag type recognition (%)
18, 24, 25	3	56.36	55.28
15, 22, 25	3	80.41	79.25
15, 22, 27	3	88.61	89.65
2, 15, 21, 25	**4**	**99.78**	**99.67**
2, 15, 18, 21, 25	5	99.81	99.70
2, 15, 16, 21, 25	5	99.82	99.71
2, 5, 15, 16, 21, 25	6	99.85	99.75

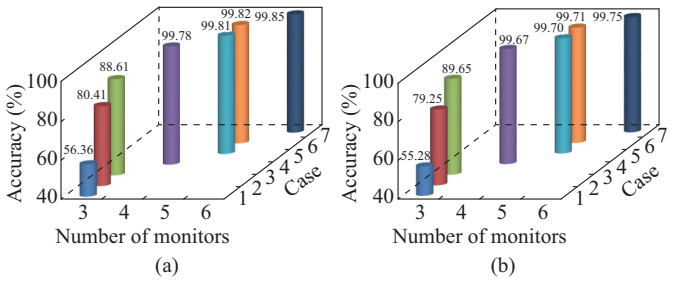

Fig. 15. Accuracy with different monitor placement. (a) Accuracy of sag source location. (b) Accuracy of sag type recognition.

From Table II and Fig. 15, the following statements and conclusions can be summarized.

1) The accuracy of sag source location and the accuracy of sag type recognition are improved effectively with an increased number of monitors. It is because that it is beneficial by more monitors for data redundancy.

2) When the number of monitors reaches a certain value, the accuracy is improved slowly when the number further increases. A good balance among installation costs, computational burden, accuracy, costs and complexity needs to be considered. In this paper, the number of monitors is set as a moderate value of 4.

3) Not only the number of monitors, but also the allocation of them is of significant importance. For placements at buses 18, 24, 25 and placements at buses 15, 22, 25, different allocations with the same number of monitors generate different results. However, for placements at buses 2, 15, 16, 21, 25 and placements at buses 2, 15, 18, 21, 25, different locations with the same number of monitors achieve the similar accuracy. The main issue is that the allocation of monitors needs to ensure a certain level of observability of the network.

4) For monitor placements at buses 18, 24, 25, the accuracy for sag source location and the accuracy for sag type recognition are unsatisfactory. This is because that the monitor placements cannot fulfill the observability of the network, leading to insufficient data and information.

E. Comparison Between Proposed Model and GRU-based Model

To further test the performance of the proposed model, we compare it with the GRU-based model. Here, the monitor placements at buses 2, 15, 21, 25, as illustrated in [33], are adopted for fair comparison. We observe that as follows.

1) As clearly shown in Fig. 16, the loss of the proposed model is less than that of the GRU-based model. Specifically, the minimum loss of our model can achieve 1×10^{-9}, whereas, that of GRU-based model can only achieve 3×10^{-2}, as illustrated in Table III. This shows that the nonlinear relationship or function between the monitored voltages and sag sources or sag types can be more accurately captured by the proposed model, whereby achieving higher accuracy for the sag source location and sag type recognition. The main reason of this phenomenon lies in the fact that IndRNN can process longer sequences and construct deeper neural networks.

2) As listed in Table III, the training time of the proposed model is significantly shorter than that needed by the GRU-based model. The proposed model can be trained more efficiently since the IndRNN can be easily regulated to prevent vanishing or exploding gradients.

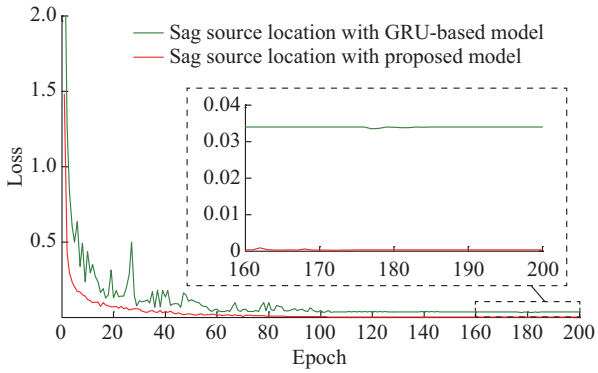

Fig. 16. Relationship between epoch and loss with different models.

TABLE III
COMPARISON BETWEEN DIFFERENT MODELS

Model	Training time for one epoch (s)	Number of training epochs with loss of 10^{-1}	Number of training epochs with loss of 10^{-3}	Number of training epochs with accuracy of 95%	Minimum loss
GRU-based model	93.1	100	Unable	30	3×10^{-2}
Model with IndRNN (proposed)	50.6	40	80	9	1×10^{-9}

Furthermore, Fig. 17 demonstrates the results of the proposed model if the phase A and phase B grounded short circuit occurs on line 20.

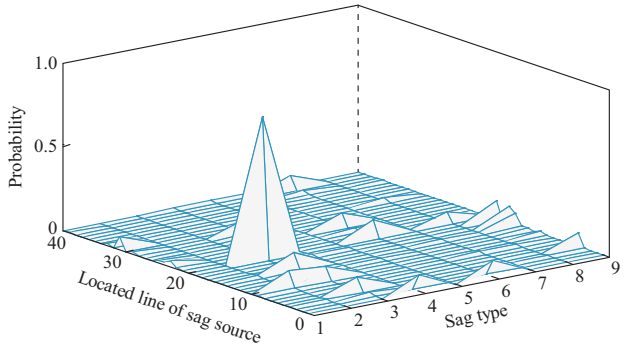

Fig. 17. Demonstrated result of proposed model.

Specially, the probability in Fig. 17 denotes the product of P_i and P_j', where P_i is the possibility that the sag source is located on line i ($i\in\{1, 2, \ldots, 37\}$); and P_j' is the possibility that the type of sag source belongs to type j ($j\in\{1, 2, \ldots, 9\}$). Here, the P_i and P_j' are respectively obtained from the proposed model as illustrated in Fig. 1 by using the monitored voltage waveforms of buses 2, 15, 21, and 25. From Fig. 16, it is observed that the maximum probability corresponds to line 20 and type 3, which shows an accurate detection. In fact, P_{20} is the maximum value among P_i, and P_3' is the maximum value among P_j'. Therefore, $P_{20}P_3'$ must be the maximum value among P_iP_j'.

F. Performance Verification for Large-scale Power Grid

We further evaluate the performance of the proposed method in a large-scale power grid. Here, the presented method is implemented on a real 134-bus system, which has been illustrated in [12]. As described in Fig. 18, the detailed procedure includes the following steps.

Step 1: meter placement can be determined to ensure that the whole system is observable by adopting the methods in [33], [35]. For the 134-bus system, 6 monitors have been installed at buses 23, 30, 63, 79, 96 and 112, respectively.

Step 2: based on the topology of power system, we can manually divide the whole system into several smaller areas or segments according to the placement of protective devices, such as reclosers, circuit breakers and sectionalizing switches. For the 134-bus system, it can be separated into 12 segments, as demonstrated in [12].

Step 3: the simulation can be implemented using MATLAB/Simulink according to Fig. 6. Meanwhile, the data for the monitors at buses 23, 30, 63, 79, 96 and 112 (as obtained from *Step 1*) are collected to generate the datasets. Similarly, 80% of the simulated data are used for training, and 20% of the simulated data are used for testing, as elaborated in the Section III-A.

Step 4: totally 12 models have been established for 12 segments (as given in *Step 2*). It should be mentioned that all these 12 models should own the specific architecture, as demonstrated in Fig. 1. Meanwhile, with the datasets generated in *Step 3*, the parameters or hyper-parameters for each of these 12 models shall be developed using the method demonstrated in Section III-B.

Step 5: each of the 12 models should be further verified as discussed in Section III-C.

Step 6: the 12 models can be used to realize sag source location and sag type recognition if their verification performances are verified as in *Step 5*. Otherwise, the parameters or hyper-parameters should be redesigned or recalculated with the method described in Section III-B.

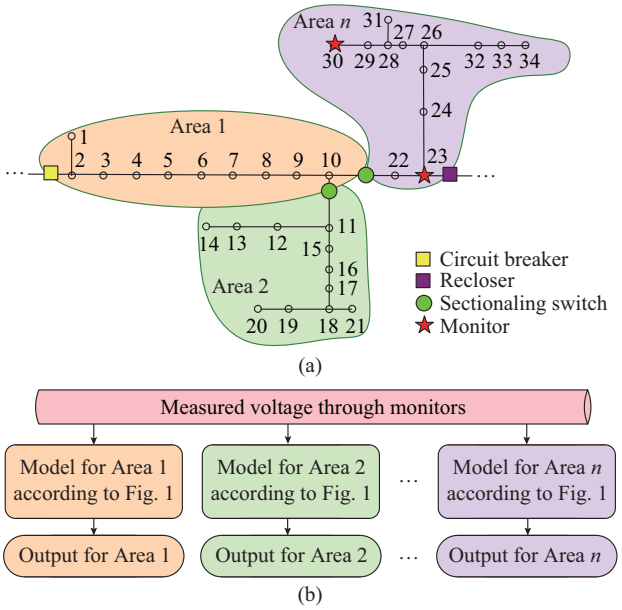

Fig. 18. Illustration of procedure for proposed model applied in practical system. (a) Real 134-bus system. (b) Procedure for proposed model.

The detailed performance for the 134-bus system is presented in Table IV and Fig. 19.

TABLE IV
ACCURACY FOR REAL 134-BUS SYSTEM

Area number	Accuracy of sag source location (%)	Accuracy of sag type recognition (%)
1	99.72	99.56
2	99.61	99.65
3	99.55	99.38
4	99.65	99.71
5	99.46	99.53
6	99.77	99.49
7	99.42	99.50
8	99.63	99.59
9	99.67	99.82
10	99.53	99.57
11	99.42	99.51
12	99.84	99.63
Average	99.61	99.58

As shown in Table IV and Fig. 19, it can be observed that the average accuracy for sag source location is 99.61% and the location accuracy for each area is higher than 99.42%. Meanwhile, the average accuracy for sag type recognition is 99.58% and the recognition accuracy for each segment is higher than 99.38%.

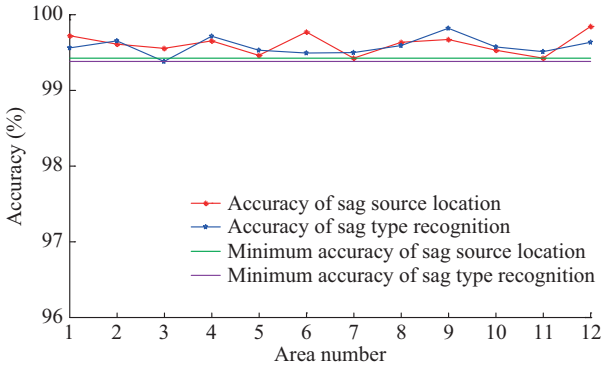

Fig. 19. Accuracy of real 134-bus system.

IV. CONCLUSION

In this paper, an attention-based IndRNN for sag source location and sag type recognition is proposed. The effectiveness of the proposed architecture is also confirmed. Using only the measured voltages with monitors in the sparsely monitored power system, the proposed method can simultaneously locate the line of sag source, recognize the type of located sag source, including motor starting, transformer energizing and short circuits, and identify the faulty phase for short circuits without knowledge of system parameters or system model. Experiments on IEEE 30-bus system and a real 134-bus system have been conducted to confirm the effectiveness and practicability of the proposed method. Moreover, a comparison between the proposed model and the GRU-based model has been performed. The comparison results show that the proposed model can be trained more efficiently and achieve higher accuracy.

The number and allocations of monitors are significant for the sag source location and sag type recognition. The optimal number of the monitors and their best allocations to ensure observability of the power system will be studied in our future work.

Since it may remain as a challenge to collect abundant high-quality data in our further studies, we will try to develop the data-driven methods requesting only a limited amount of data, or adopting an unsupervised learning so as to reduce the dependence on data or label.

REFERENCES

[1] J. Ye and H. B. Gooi, "Phase angle control based three-phase DVR with power factor correction at point of common coupling," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 1, pp. 179-186, Jan. 2020.

[2] A. P. Torres, P. Roncero-Sánchez, and V. F. Batlle, "A two degrees of freedom resonant control scheme for voltage-sag compensation in dynamic voltage restorers," *IEEE Transactions on Power Electronics*, vol. 33, no. 6, pp. 4852-4867, Jun. 2018.

[3] T. Yi, H. Jie, L. Hao *et al.*, "Method for voltage sag source location based on the internal resistance sign in a single-port network," *IET Generation, Transmission & Distribution*, vol. 10, no. 7, pp. 1720-1727, May 2016.

[4] V. Huchche, N. Patne, and A. Junghare, "Computation of energy loss in an induction motor during unsymmetrical voltage sags – a graphical method," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 5, pp. 2023-2030, May 2018.

[5] A. H. Khawaja and Q. Huang, "Estimating sag and wind-induced motion of overhead power lines with current and magnetic-flux density measurements," *IEEE Transactions on Instrumentation and Measure-*

ment, vol. 66, no. 5, pp. 897-909, May 2017.
[6] Y. Xu, Y. Wu, M. Zhang *et al*., "Sensitivity of programmable logic controllers to voltage sags," *IEEE Transactions on Power Delivery*, vol. 34, no. 1, pp. 2-10, Feb. 2019.
[7] K. Zhu, Y. Wang, P. Yin *et al*., "Voltage sag source location technology based on corresponding sequence components," *IET Generation, Transmission & Distribution*, vol. 9, no. 9, pp. 820-827, Jun. 2015.
[8] D. Zhang, X. Han, and C. Deng, "Review on the research and practice of deep learning and reinforcement learning in smart grids," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 3, pp. 362-370, Sept. 2018.
[9] Y. Xu, Z. H. Qu, R. Harvey *et al*., "Data-driven wide-area control design of power system using the passivity shortage framework," *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 830-841, Jul. 2020.
[10] G. Chang, J. Chao, H. Huang *et al*., "On tracking the source location of voltage sags and utility shunt capacitor switching transients," *IEEE Transactions on Power Delivery*, vol. 23, no. 4, pp. 2124-2131, Oct. 2008.
[11] X. Liang, S. A. Wallace, and D. Nguyen, "Rule-based data-driven analytics for wide-area fault detection using synchrophasor data," *IEEE Transactions on Industry Applications,* vol. 53, no. 3, pp. 1789-1798, May 2017.
[12] Y. Dong, C. Zheng, and M. Kezunovic, "Enhancing accuracy while reducing computation complexity for voltage-sag-based distribution fault location," *IEEE Transactions on Power Delivery*, vol. 28, no. 2, pp. 1202-1212, Apr. 2013.
[13] S. Jamali, A. Bahmanyar, and H. Borhani-Bahabadi, "A fast and accurate fault location method for distribution networks with DG using genetic algorithms," in *Proceedings of Smart Grid Conference (SGC)*, Tehran, Iran, Dec. 2015, pp. 110-114.
[14] J. C. C. L. da Silva Filho, F. Anderson Silva Borges, R. de Andrade L Rabelo *et al*., "A method for voltage sag source location using clustering algorithm and decision rule labeling," in *Proceedings of International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, Jul. 2019, pp. 1-8.
[15] F. Xu, Y. Wang, and H. Yang, "Voltage sags detection based on fundamental frequency vector of generalized S-transform modulus matrixes," in *Proceedings of 9th International Conference on Fuzzy Systems and Knowledge Discovery*, Chongqing, China, May 2012, pp. 2019-2022.
[16] J. R. Camarillo-Peñaranda and G. Ramos, "Fault classification and voltage sag parameter computation using voltage ellipses," *IEEE Transactions on Industry Applications*, vol. 55, no. 1, pp. 92-97, Jan. 2019.
[17] C. Li, J. Yang, Y. Xu *et al*., "Classification of voltage sag disturbance sources using fuzzy comprehensive evaluation method," in *Proceedings of CIRED − 24th International Conference on Electricity Distribution*, Glasgow, USA, Jun. 2017, pp. 544-548.
[18] S. Wu, L. Zheng, W. Hu *et al*., "Improved deep belief network and model interpretation method for power system transient stability assessment," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 1, pp. 27-37, Nov. 2019.
[19] G. Luo, J. Hei, C. Yao *et al*., "An end-to-end transient recognition method for VSC-HVDC based on deep belief network," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1070-1079, Dec. 2020.
[20] Y. Deng, L. Wang, H. Jia *et al*., "A sequence-to-sequence deep learning architecture based on bidirectional GRU for type recognition and time location of combined power quality disturbance," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4481-4493, Aug. 2019.
[21] H. Liao, J. V. Milanović, M. Rodrigues *et al.*, "Voltage sag estimation in sparsely monitored power systems based on deep learning and system area mapping," *IEEE Transactions on Power Delivery*, vol. 33, no. 6, pp. 3162-3172, Dec. 2018.
[22] P. Razmi, M. O. Buygi, and M. Esmalifalak, "A machine learning approach for collusion detection in electricity markets based on Nash equilibrium theory," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 1, pp. 170-180, Jan. 2021.
[23] R. Tolosana, R. Vera-Rodriguez, J. Fierrez *et al*., "Exploring recurrent neural networks for on-line handwritten signature biometrics," *IEEE Access*, vol. 6, pp. 5128-5138, Jan. 2018.
[24] H. Yang, R. Qiu, and H. Tong, "Reconstruction residuals based long-term voltage stability assessment using autoencoders," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1092-1103, Dec. 2020.
[25] C. Hu, T. Ou, Y. Zhu *et al*., "GRU-type LARC strategy for precision motion control with accurate tracking error prediction," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 1, pp. 812-820, Jan. 2021.
[26] S. Li, W. Li, C. Cook *et al*., "Independently recurrent neural network (IndRNN): building a longer and deeper RNN," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, Jun. 2018, pp. 5457-5466.
[27] K. Song, H. Yang, and Z. Yin, "Multi-scale attention deep neural network for fast accurate object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 2972-2985, Oct. 2019.
[28] Y. Zhu, C. Zhao, H. Guo *et al*., "Attention CoupleNet: fully convolutional attention coupling network for object detection," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 113-126, Jan. 2019.
[29] W. Du, Y. Wang, and Y. Qiao, "Recurrent spatial-temporal attention network for action recognition in videos," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1347-1360, Mar. 2018.
[30] B. Su, H. Chen, P. Chen *et al*., "Deep learning-based solar-cell manufacturing defect detection with complementary attention network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4084-4095, Jun. 2021.
[31] L. Huang, Y. Ma, S. Wang *et al*. (2020, Aug.). An attention-based spatiotemporal LSTM network for next POI recommendation. [Online]. *IEEE Transactions on Services Computing*. Available: http://dx.doi.org/10.1109/TSC.2019.2918310
[32] S. Wu, G. Li, L. Deng *et al*., "L1-norm batch normalization for efficient training of deep neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 7, pp. 2043-2051, Jul. 2019.
[33] H. Jiang, Y. Xu, Z. Liu *et al*., "A BPSO-based method for optimal voltage sag monitor placement considering uncertainties of transition resistance," *IEEE Access*, vol. 8, pp. 80382-80394, Apr. 2020.
[34] A. Géron. (2017, Mar.). Hands-on machine learning with scikit-learn and TensorFlow. [Online]. Available: http://www.java1234.com
[35] S. Ali, K. Wu, K. Weston *et al*., "A machine learning approach to meter placement for power quality estimation in smart grid," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1552-1561, May 2016.

Yaping Deng received the M.S. and Ph.D. degrees in electrical engineering from the Xi'an University of Technology, Xi'an, China, in 2011 and 2015, respectively. She is currently a Lecturer with the Department of Power Grid Information and Control Engineering, School of Electrical Engineering, Xi'an University of Technology. Her research interests include control and analysis of power quality, power system modeling, and applications of artificial intelligence in modern power systems.

Xinghua Liu received the B.S. degree in mathematics from Jilin University, Changchun, China, in 2009, and the Ph.D. degree in control science and engineering from the University of Science and Technology of China, Hefei, China, in 2014, respectively. From 2014 to 2015, he was a Visiting Fellow with Royal Melbourne Institute of Technology (RMIT) University, Melbourne, Australia. From 2015 to 2018, he was a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He is currently a Professor with the Department of Power Grid Information and Control Engineering, School of Electrical Engineering, Xi'an University of Technology, Xi'an, China. His research interests include state estimation and control, intelligent systems, power systems, autonomous vehicles, and cyber-physical systems.

Rong Jia received the Ph.D. degree from Xi'an University of Technology, Xi'an, China, in 1999. He is currently a Professor with the Department of Power System and Its Automation, School of Electrical Engineering, Xi'an University of Technology. He is also the head of the School of Electrical Engineering. His research interests include power system operation and control, and power equipment status monitoring.

Qi Huang received the B.S. degree in electrical engineering from Fuzhou University, Fuzhou, China, in 1996, the M.S. degree from Tsinghua University, Beijing, China, in 1999, and the Ph.D. degree from Arizona State University, Phoenix, USA, in 2003, respectively. He is currently a Professor at University of Electronic Science and Technology of China, the Executive Dean of School of Energy Science and Engineering, University of Electronic Science and Technology of China, and the Director of Sichuan State Provincial Laboratory of Power System Wide-area Measurement and Control, Chendu, China. He is a Member of IEEE since 1999. His research interests include power system instrumentation, power system monitoring and con-

trol, and power system high performance computing.

Gaoxi Xiao received the B. S. and M. S. degrees in applied mathematics from Xidian University, Xi'an, China, in 1991 and 1994, respectively, and the Ph. D. degree in computing from Hong Kong Polytechnic University, Hong Kong, China, in 1998. He was an Assistant Lecturer with Xidian University from 1994 to 1995. He was a Postdoctoral Research Fellow with Polytechnic University, Brooklyn, USA, in 1999, and a Visiting Scientist with the University of Texas at Dallas, USA, from 1999 to 2001. In 2001, he joined the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently an Associate Professor. He served as an Associate Editor or Guest Editor for the IEEE Transactions on Network Science and Engineering, Plos One, and Advances in Complex Systems, and a Technical Program Committee Member for numerous conferences including the IEEE International Conference on Communications and the IEEE Global Communications Conference. His research interests include complex systems and complex networks, communication networks, smart grids, and system resilience and risk management.

Peng Wang received the B. S. degree from Xi'an Jiaotong University, Xi'an, China, in 1978, the M.S. degree from Taiyuan University of Technology, Taiyuan, China, in 1987, and the M.S. and Ph.D. degrees from the University of Saskatchewan, Saskatoon, Canada, in 1995 and 1998, respectively, all in electrical engineering. He is currently a Professor at Nanyang Technological University, Singapore. He, as a Principle Investigator (PI) and Co-PI, has been awarded over $15 million research grant from industries and government organizations from Singapore, China, USA and Europe to work on the hybrid AC/DC micro-grids, smart grids, power system operation, planning, reliability and renewable integration. He served as an Associate Editor of the IEEE Transaction on Smart Grid and a Guest Editor of Journal of Modern Power Systems and Clean Energy for special issues on Smart Grids. He also served as an Associate Editor of IEEE Transaction on Power Delivery and Guest Editor-in-Chief of CSEE Journal of Power and Energy Systems for special issues on Hybrid AC/DC Grids for Future Power Systems. His research interests include power system planning and operation, renewable energy planning, solar/electricity conversion system, power market and power system reliability analysis.