

Received May 5, 2021, accepted May 17, 2021, date of publication May 19, 2021, date of current version May 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3082001

A Transfer and Deep Learning-Based Method for Online Frequency Stability Assessment and Control

JIAN XIE¹, (Student Member, IEEE), AND WEI SUN, (Member, IEEE)

Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL 32826, USA

Corresponding author: Wei Sun (sun@ucf.edu)

ABSTRACT Fast and accurate prediction and control of power system dynamic frequency after disturbance is essential to enhance power system stability. Machine learning methods have great potential in harnessing data for online application with accurate predictions. This paper proposes a two-stage novel transfer and deep learning-based method to predict power system dynamic frequency after disturbance and provide optimal event-based load shedding strategy to maintain system frequency. The proposed deep learning model combines convolutional neural network (CNN) and long short-term memory (LSTM) network to harness both spatial and temporal measurements in the input data, through a four-dimensional (4-D) tensor input construction process including, 1) capture system network topology information and critical measurements from different time intervals; 2) compute a multi-dimensional electric distance matrix and reduce to a 2-D plane which can describe the system nodal distribution; 3) construct 3-D tensors based on state variables at different sample times; and 4) integrate into 4-D tensor inputs. Moreover, a transfer learning process is employed to overcome the challenge of insufficient data and operating condition changes in real power systems for new prediction tasks. Simulation results in IEEE 118-bus system verify that the CNN-LSTM method not only greatly improves the timeliness of online frequency control, but also presents good accuracy and effectiveness. Test cases in the New England 39-bus system and the South Carolina 500-bus system validate that the transfer learning process can provide accurate results even with insufficient training data.

INDEX TERMS CNN, deep learning, dynamic frequency, LSTM, spatial-temporal feature, transfer learning.

I. INTRODUCTION

Power system frequency reflects the balance between generation and load, serving as a critical indicator of power system stability [1]–[4]. Due to the increasing penetration of renewable energies, the complexity of the power system greatly increases with reduced system inertia, which causes more severe system fluctuation after disturbances. With more blackouts caused by frequency collapse, it is imperative to accurately predict system frequency online and rapidly provide reliable event-based load shedding scheme for real-time power system frequency control [3].

The methods for power system dynamic frequency analysis in literature can be classified as time-domain simulation, equivalent model, and machine learning (ML) models. First, time-domain simulation requires detailed mathematical models of power system components and network equations

to obtain system dynamic frequency responses. However, its application for real-time prediction is limited due to the high computational requirement. Next, the equivalent model-based method aggregates all the generator swing equations into an equivalent rotor model, such as the average system frequency model [5]. This method dramatically reduces the complexity of power system models and realizes the fast prediction of dynamic frequency. However, the prediction error is inevitable due to neglecting the impact of load and system topology on dynamic frequency [6]. In [7], an analytical frequency nadir prediction model is proposed to predict frequency nadir and time when it reaches. In [8], the rate of change of frequency of the center of inertia is estimated using only local frequency.

Adaptive load shedding methods are mainly studied for load shedding control in recent publications [9]–[11]. These methods first calculate power deficit or estimate the minimum frequency after disturbance, then the actual amount to be shed can be calculated or guided by a pre-defined strategy [12].

The associate editor coordinating the review of this manuscript and approving it for publication was Feng Wu.

However, this type of scheme is usually based on the equivalent system model and is limited by the accuracy of the model.

The huge amount of data and the breakthrough in learning and computing power [13] enable the applications of ML methods to improve the accuracy and efficiency of power system dynamic frequency prediction, such as decision tree (DT) [14], [15], support vector machine (SVM) [16], [17], and extreme learning machine (ELM) [6]. However, there are three major limitations, 1) they only transform the input data into one or two successive representation spaces by simple transformations, but the refined representations required by complex problems, such as dynamic frequency prediction, generally cannot be attained; 2) they cannot consider the spatial-temporal features, such as network topology, or utilize time series data, which reduce the prediction accuracy; and 3) they are unable to exploit the prior trained model for new prediction, which may result in inefficient training or insufficient training data.

Currently, enabled by more powerful computing hardware, *Deep Learning* (DL) models have demonstrated superior performance in solving high-dimensional non-linear problems [13], such as convolutional neural network (CNN) [13] and recurrent neural network (RNN) [13]. A key characteristic of CNN is that the convolution layers learn local patterns of inputs, which presents two powerful properties: 1) the patterns they learn are translation invariant, and 2) they can learn spatial hierarchies of patterns. In [18], power system stability assessment is formulated as a classification problem and solved by CNN. However, the inputs are operating conditions grouped as 1-dimensional (1-D) vector, which cannot harness the spatial feature from data. The long short-term memory (LSTM) network [13] is an effective RNN, which has been widely applied to solve sequence prediction problems, such as load forecasting and wind speed prediction.

Power system operation and structure data contain rich and useful information. Especially during the transient period, multi-channel time-series streaming data could be mined for system dynamics analysis. In this paper, a two-stage CNN-LSTM model is developed to fully exploit spatial-temporal data for online frequency prediction and load shedding control. At the first stage, the CNN-LSTM model is adopted to predict system dynamic frequency till steady state after disturbance. At the second stage, if the predicted frequency nadir is lower than thresholds, another CNN-LSTM model is used to develop a specific load shedding plan for frequency stability control. The CNN-LSTM architecture [19] integrates the unique characteristics of both CNN and LSTM, using CNN layers for feature extraction on input data and combined with LSTM networks to support dynamic frequency prediction and control.

First, in order to utilize system structure information, electric distance is used to describe the high-dimensional spatial position of power system nodes. Then the dimension reduction algorithm is applied to map system nodes into a 2-D plane. Next, critical dynamic state variables at a certain time are selected to construct a 3-D tensor. Finally, a 4-D tensor

can be generated by integrating all 3-D tensors constructed at different time intervals, which is then used as inputs for the proposed CNN-LSTM models. The output of the first CNN-LSTM model is the dynamic frequency response from the disturbance occurs and a specific load shedding strategy is the output of the second model.

In addition, due to the inability to store prior learned knowledge, DL models have to be re-trained when dealing with tasks in a new scenario, which leads to a long re-training process. Another issue is the lack of sample data for new scenarios, which can lead to poor prediction accuracy. However, *Transfer Learning* (TL) could address these challenges by harnessing the features learned to improve the performance of pre-trained DL models in new tasks [20], [21]. Aiming to accelerating and improving the accuracy of the dynamic frequency and optimal load shedding prediction process in order to meet the requirements of online prediction and control for large power grids, this paper takes advantage of TL to improve the performance of CNN-LSTM model.

In summary, the main contributions of this paper include:

- A novel two-stage CNN-LSTM-based method is proposed to predict power system dynamic frequency and optimal load shedding strategy, which can fully exploit both spatial and temporal dynamic measurements to ensure more accurate and stable predictions.
- In order to fully exploit system network information and the powerful learning ability of CNN-LSTM, power system typology is adopted to generate 4-D input tensors, as a first try in this paper.
- A transfer learning process is incorporated to address data shortage challenge and operation condition changes in real power systems, reduce the computing burden, save and apply the pre-trained models to new tasks.

The rest of the paper is structured as follows. Section II describes the basic principles of CNN-LSTM model and transfer learning. Section III presents the framework of CNN-LSTM-based power system dynamic frequency prediction and control. Simulation results are presented in section IV, and section V concludes this paper.

II. THEORETICAL FOUNDATION

A. POWER SYSTEM CENTER OF INERTIA FREQUENCY

By analyzing the measurement and simulation data, we can find that power system dynamic frequency has certain spatial-temporal characteristics. The frequency of each generator oscillates around the center of system inertia, and when the system stabilizes, the frequency of each generator will eventually approach to the center of system inertia. When the emergency controls are applied, the center of inertia (COI) frequency is usually used to represent the global state of system frequency [2], [5], [8]. In addition, most load shedding schemes use COI as the index. The definition of COI frequency f_{COI} is shown as follows:

$$f_{COI} = \frac{\sum_{i=1}^n (H_i f_i)}{\sum_{i=1}^n H_i} \quad (1)$$

where n denotes the number of generators; H_i and f_i are the inertia and rotor speed of the i^{th} generator, respectively.

When a disturbance happens to power systems, generators share the unbalanced system power based on the corresponding synchronization factor. The amount of unbalanced power that a generator undertakes is related to the initial operating state, the electrical distance to the fault point, and the size of system's unbalanced power. A generator that is closer to the fault location with respect to the electric distance will take more unbalanced power [22]. Synchronization factor SP_{ik} between generator node i and node k is defined as:

$$SP_{ik} = V_i V_k (B_{ik} \cos \delta_{ik} - G_{ik} \sin \delta_{ik}) \quad (2)$$

where V and δ are voltage amplitude and phase angle difference, respectively; B_{ik} and G_{ik} are the transfer impedance.

Let t_0 and t_f denote the moment when and after the disturbance occurs. According to (2), the active power of each generator and each load at t_0 , the voltage magnitude and angle of each node, and the unbalanced power of each generator at t_f are selected as input features. In the transient period, the power imbalance of a generator ΔP_i can be represented as [23]:

$$\begin{cases} \Delta P_i = P_{mi} - P_{ei} = 2 \frac{H_i}{f_N} \frac{df_i}{dt} \\ H_i \dot{\omega}_i = P_{m,i} - P_{e,i} - D_i (\omega_i - 1) \end{cases} \quad (3)$$

where P_{mi} and P_{ei} denote the mechanical and electrical power of i^{th} generator, respectively; and f_N denotes the nominal synchronous speed.

Considering the impact from turbine-governor [24] and taking the turbine-governor TGOV1 model as an example, the state equations of frequency after disturbance can be expressed as:

$$\begin{bmatrix} \Delta \dot{\theta} \\ \Delta \dot{\omega} \\ \Delta \dot{P}_T \\ \Delta \dot{v} \end{bmatrix} = \begin{bmatrix} 0 & K_{12} & 0 & 0 \\ K_{21} & K_{22} & K_{23} & 0 \\ 0 & K_{32} & K_{33} & K_{34} \\ 0 & K_{42} & 0 & K_{44} \end{bmatrix} \begin{bmatrix} \Delta \theta \\ \Delta \omega \\ \Delta P_T \\ \Delta v \end{bmatrix} + \begin{bmatrix} 0 \\ K_5 \\ 0 \\ 0 \end{bmatrix} \quad (4)$$

where T_1 , T_2 and T_3 are time constants in TGOV1, R is the permanent droop, and D_i is the damping coefficient in TGOV1. K_{ij} are coefficient matrices as referred in [25].

B. OPTIMAL LOAD SHEDDING

In general, the goal of event-based load shedding is to minimize the amount of shed load to maintain system frequency nadir higher than thresholds. Therefore, the minimum amount of load shedding is used as the objective function of the following linear programming optimization model:

$$\min f(P_{slj}) = \sum_{j=n+1}^m C_i P_{slj}$$

$$\text{s.t.} \begin{cases} \sum_{j=n+1}^m P_{slj} = P_s \\ 0 \leq P_{slj} \leq P_{slj \max} \\ \omega_{\min} = \omega_{set} \\ F(\theta, V, \omega) \end{cases} \quad (5)$$

where ω_{set} is the pre-defined threshold of steady state frequency; C_i is a load-shedding factor to quantitatively characterize the impacts of load importance at different load-shedding points; and $F(\theta, V, \omega)$ is the state space equations after load shedding, which can be referred in [25].

The pre-defined threshold ω_{set} will influence the optimal results. According to [26], if the system frequency cannot reach 59.5 Hz in 30 seconds for a 60-Hz system, the system will be unsafe and load shedding should be executed to improve the frequency. Thus, in this paper, the threshold is chosen as 59.5Hz.

The output in stage 2 is the optimal load shedding strategy. In the sample generation process, optimal load shedding will be solved in scenarios which require load shedding to maintain frequency stability, which will be used for training in stage 2.

C. CNN-LSTM ARCHITECTURE

To extract the rich information from the spatial-temporal dynamic measurements across different time intervals, a CNN-LSTM method is proposed with the architecture shown in Fig. 1. The proposed DL model consists of CNN and LSTM layers. In the proposed architecture, the fully connected layer in CNN is replaced by the Conv layers to learn the generic features, due to the capability of encoding spatial information, which is a key factor for the following LSTM. From the architecture, we can see that the temporal tensors are selected as inputs, and features in these inputs are first extracted by Conv layers, then pooling layers are adopted to reduce the spatial size of the representation learned by CNNs. After that, the flatten layers convert the data into a 1-D array as the input to the LSTM layer. Finally, temporal features are extracted by LSTMs, and go through the fully connected layers for regression analysis.

D. TRANSFER LEARNING PRINCIPLE

In practice, the transfer learning process is usually implemented by freezing some neuron layers in the source ML model (i.e., a well-trained model) and re-training the last one or two layers based on the data obtained from new systems. These frozen layers are usually used to extract features from inputs. For power system dynamic frequency prediction, the proposed CNN-LSTM model can be trained on one source system, and then transferred and tested on new systems where few available sample data can be obtained. The Conv and LSTM layers of trained CNN-LSTM in one system can be directly applied to other systems, which could greatly help reduce the data sampling and training cost, as shown in Fig. 2. From the mathematical view, it means only parameters in the

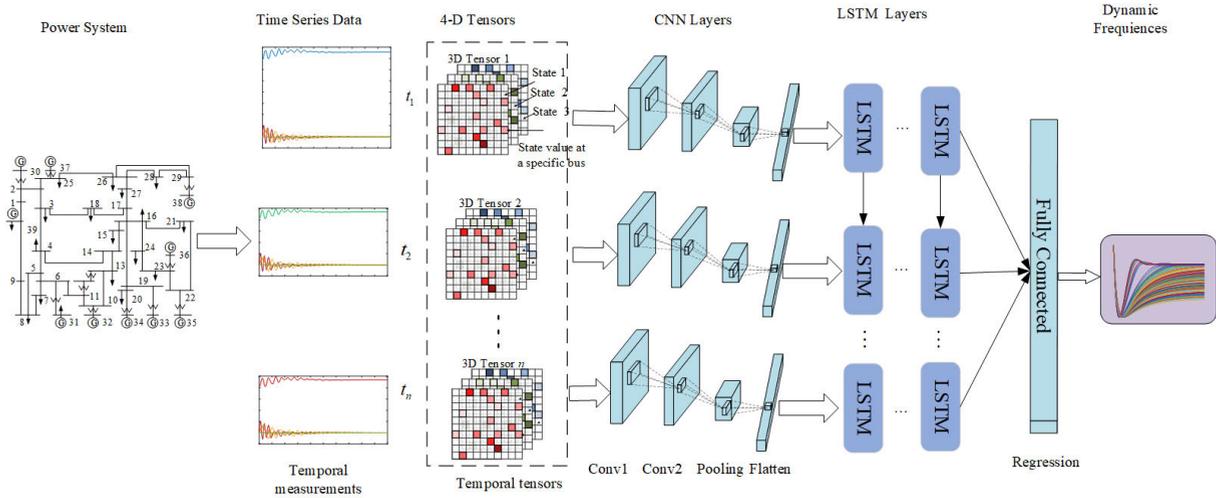


FIGURE 1. CNN-LSTM architecture.

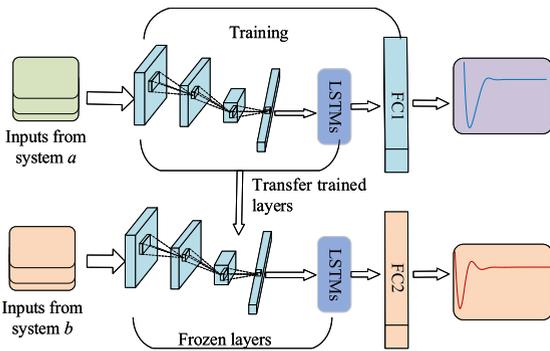


FIGURE 2. Transfer learning process.

last fully connected layer will be updated during the training process, and the equation is shown in (6).

$$\omega_{\theta} = \begin{cases} \omega_{\theta}, & \text{other} \\ \omega_{\theta} - \eta \nabla \partial L / \partial \omega_{\theta}, & \omega_{\theta} \in FC \text{ layer} \end{cases} \quad (6)$$

where ω_{θ} denotes the hyperparameters in the CNN-LSTM model.

III. TRANSFER AND CNN-LSTM-BASED FREQUENCY PREDICTION

A. INPUT FEATURE SELECTION

Adequate and high-quality data are the basis for ensuring the accuracy of ML models. Before building a CNN-LSTM model, the input data must be filtered and organized based on the characteristic of power system frequency stability. Based on the analysis in section II, 5 types of state variables, including active power P_{ei} , unbalance power of generator ΔP_i , active load power P_{li} , and bus voltage magnitude and angle V_j and V_{g_j} , are selected as the input features. In addition, the proposed CNN-LSTM method can process temporal data; therefore, state variables at different times will also

TABLE 1. Input features adopted in CNN-LSTM.

Number	Input feature
1	Active power of each generator P_{ei} from t_0 to $3f_t$
2	Unbalance power of generators ΔP_i from t_0 to $3f_t$
3	Voltage amplitude of each node V_j from t_0 to $3f_t$
4	Voltage angle of each node V_{g_j} at from t_0 to $3f_t$
5	Active power of each load P_{li} at from t_0 to $3f_t$

be captured as inputs for CNN-LSTM. In order to make a trade-off between computational time and solution accuracy, the time interval is set as $t_i \in [t_0, 3f_i]$, where t_0 represents the moment when a disturbance occurs and f_i is the sampling period. It means that critical state variables from the moment when disturbance occurs to the third sampling period are captured. The input features are summarized in Table 1.

B. INPUT TENSOR CONSTRUCTION

As stated in section I, the inputs for CNN-LSTM are 4-D tensor data, which consists of rich spatial-temporal information. In order to construct a 4-D tensor, we first build 3-D tensors that include system network data and key features at a given time $t_i \in [t_0, 3f_i]$; then a 4-D tensor can be constructed by connecting all the 3-D tensors during t_0 and $3f_i$, which is similar to videos. The following subsections introduce the process of generating a 3-D tensor.

1) ELECTRIC DISTANCE CALCULATION

There are strong electrical coupling connections between power system nodes. The electric distance is a useful index to describe the spatial distribution of nodes and measure the connection degree between each node. In addition, the electric distance determines the unbalanced power distribution of the power system at the moment of power disturbance,

which is closely related to system dynamic frequency after the disturbance. According to [27], electric distance between nodes i and j can be modeled by:

$$D_{ij} = |(Z_{ii} - Z_{ij}) - (Z_{ij} - Z_{jj})| \quad (7)$$

where Z_{ij} denotes the node impedance between nodes i and j . For an n -bus system, electric distance matrix D is an $n \times n$ square symmetric matrix.

2) DIMENSION REDUCTION OF ELECTRIC DISTANCE MATRIX

In the actual power system, the large number of nodes may cause the curse of dimension. In this paper, a 2-D matrix is built to describe the distribution of a power system in a 2-D space [28]. Therefore, a dimension reduction process is applied to the electric distance matrix D , to reduce the high dimension matrix and maintain the highly correlated nodes for each node. The dimension reduction methods in literature include Principal Component Analysis, Linear Discriminant Analysis, and t-distributed stochastic neighbor embedding (t-SNE) [13]. In order to maintain the correlation between highly related nodes in D , the non-linear method of t-SNE is selected, since it models each high-dimensional object by a two- or three-dimensional point, so that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points with high probability. The process is shown as follows:

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix} \Rightarrow Y = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ \vdots & \vdots \\ y_{n1} & y_{n2} \end{bmatrix} \quad (8)$$

where y_{i1} and y_{i2} denote the map that reflects the i^{th} variable's action set.

3) 4-D TENSOR INPUT CONSTRUCTION

The numerical number of 2-D coordinates obtained in matrix Y are between 0 and 1. In order to reflect the nodal coupling of the actual system, the elements in Y need to be enlarged to the appropriate number in terms of integers. Through the linear normalization method, the decimal node coordinates in Y are enlarged to the $[1, h]$ interval, and the normalized node coordinates are rounded to obtain the integer node coordinates y_{int} . The normalization equation is shown below:

$$y_{int} = \text{round} \left[1 + \frac{(h-1)(y - Y_{\min})}{Y_{\max} - Y_{\min}} \right] \quad (9)$$

Now we obtain the node distribution information according to the integer coordinate y_{i1} and y_{i2} . The following task is to combine state variable information and node distribution information to obtain a spatial-temporal input. The idea is to build a $h \times h$ matrix for each state variable, and then extract the corresponding node distribution from Y . Finally, each state variable measurement is assigned to the corresponding coordinate in the $h \times h$ matrix. For nodes without this input

feature, the matrix element is set with a value of 0. In this way, a second-order state feature map corresponding to an input feature at this moment is obtained. From the 4-D tensor construction process, we can see that even though the input features are different in various systems, the input for CNN-LSTM is same and it can be used for transfer learning.

Next, we take generator active power as an example to illustrate the process:

- Step 1: create a $h \times h$ null matrix M .
- Step 2: extract generator node distribution from Y :

$$G = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \\ \vdots & \vdots \\ g_{m1} & g_{m2} \end{bmatrix} \quad (10)$$

where $g_{ij} = y_{n_gi,j}$, g_{mi} denotes the corresponding coordinates related to i^{th} generator node, and n_gi is the index of the i^{th} generator.

- Step 3: assign the state variable measurement to corresponding coordinate based on the following equation:

$$M(g_{i1}, g_{i2}) = P_{ei} \quad (11)$$

- Step 4: repeat the similar process using other input features. The matrix for the k^{th} input feature at t , M_k^t , is expressed as:

$$M_k^t(Y_{i1}, Y_{i2}) = S_{ik}^t \quad (12)$$

where $i = 1, 2, \dots, n$ represents the i^{th} node; $k = 1, 2, \dots, 6$ denotes the k^{th} input feature shown in Table 1; Y_{i1} and Y_{i2} denote the coordinate elements associated with input feature; and S_{ik}^t is the measurement of k^{th} input feature at t .

- Step 5: integrate M_k^t and we can obtain a 3-D tensor at t . Then, collect measurements at another moment t_{i+1} and repeat the 3-D tensor construction process to build other tensors.
- Step 6: integrate different 3-D tensors to obtain the 4-D tensor, which are the input data containing spatial-temporal information.

C. MODEL TRAINING AND TRANSFER LEARNING

After obtaining the 4-D tensor inputs and the corresponding outputs from a specific test system, we can perform the training process to obtain a well-trained CNN-LSTM model. Then, the trained CNN-LSTM layers will be frozen, and only the last fully connected layer will be re-trained when applying to a prediction task for a new test system.

D. OVERALL PROCEDURE

The overall flow chat is shown in Fig. 3, and the pseudocode for sample generation and TL is shown in algorithm 1.

Algorithm 1 Pseudocode for the Sample Generation and TL

```

1: function IC(Simulated Data)
2:   Input: System node impedance and simulated data.
3:   Input Normalization.
4:   Compute the electric distance matrix D (12)
5:   Dimension reduction of D (13)
6:   for  $t = t_0 \rightarrow t_f$  do
7:     while  $m < 7$  do
8:       Capture the  $m^{\text{th}}$  input at  $t$ 
9:       Assign  $m^{\text{th}}$  input to the corresponding location in  $Y$ 
10:    end while
11:    Integrate the fulfilled  $Y$  into a 3-D tensor.
12:  end for
13:  Integrate the 3-D tensors into a 4-D tensor.
14: end function
15:
16: function TL(Transfer Learning)
17:  Input: Trained CNN-LSTM model, new 4-D tensors, and corresponding COIs,  $T_1$ , and  $\alpha$ .
18:  Output: A new trained CNN-LSTM model.
19:  Initialize parameters, like the number of layers and weight.
20:  for  $t = 1 \rightarrow T_1$  do
21:    Freeze the CNN and LSTM layers and train the model.
22:  end for
23: end function

```

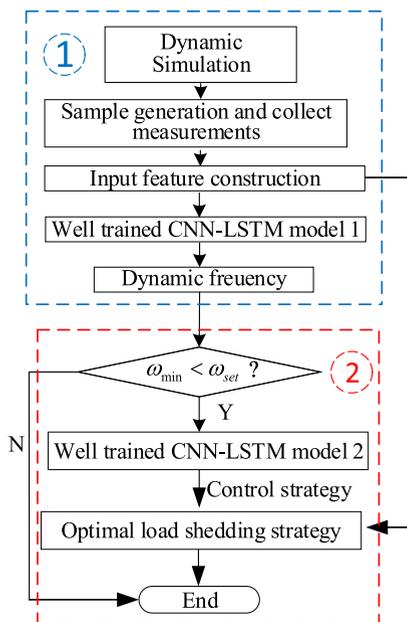


FIGURE 3. The flowchart of the proposed scheme.

IV. SIMULATION RESULTS

The proposed CNN-LSTM-based power system dynamic frequency prediction and optimal load shedding strategy are first tested on IEEE 118-bus system, and then the trained model is

transferred and applied on the New England 39-bus system and South Carolina 500-bus system ACTIVSg500 to test the performance of transfer learning.

A. SIMULATION MODEL

The IEEE 118-bus system consists of 19 generators, 9 transformers, and 91 loads.

1) SAMPLE GENERATION

In this paper, the dynamic data-sets are simulated on PSS/E, and a generator tripping disturbance is set at 0s to cause power loss and dynamic frequency. The simulation time is 40s and the sample rate is 100Hz. In order to simulate different operation status and generate enough samples, the load level is set to 120 different scenarios: 40%, 40.5%, ..., 100%. Under different operating modes, generator tripping faults are applied, and system states and dynamic frequency response after disturbances are measured and collected. A total of 2,040 sets of sample data are collected, and there are 990 samples categorized as unstable samples which require load shedding controls. For these 990 samples, a pre-optimization process is applied to obtain optimal sample outputs. In this case, 85.3% of samples are used for training, and the remaining 14.7% samples are adopted as test data.

2) BUILD ELECTRIC DISTANCE MATRIX

According to the prediction step, the electric distance matrix D is calculated based on (7), and the t-SNE method is applied on D to map the spatial distribution of nodes to a 2-D dimension-reduced matrix Y , as shown in Table 2.

TABLE 2. Two-dimensional integer node coordinates after dimension reduction and normalization.

Node	y_{11}	y_{12}
1	56	82
2	79	63
...
117	85	36
118	57	15

The node distribution of the dimension reduced matrix Y is shown in Fig. 4. It is shown that node 87 is close to node 111 but far away from other nodes, which is consistent with the electric distances between these nodes. This also verifies that the reduced matrix can represent the electric distance with embedded system network information.

3) CONSTRUCT 4-D TENSORS

As stated in section III, the elements in Y are normalized to the range between 0 and 100. Define a matrix with dimension 100×100 , then put each input feature to the corresponding node coordinate in matrix Y . Elements in Y are set to zeros for nodes without connections. After setting the dimension-reduced matrix for i^{th} input feature, we can repeat steps for another input feature.

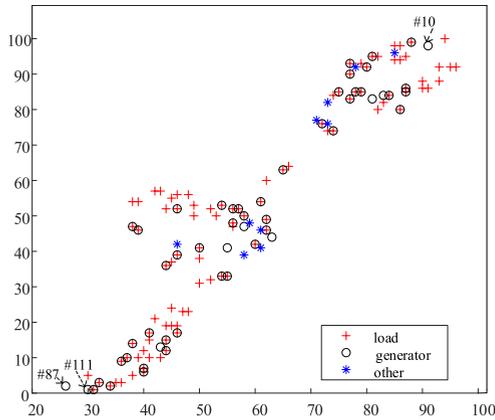


FIGURE 4. Node distribution after dimension reduction.

After completing building the matrices for six input features, we can obtain six 2-D tensors which contain state variables and system network information. Next, these six 2-D tensors will be integrated into a 3-D tensor that contains the state information at t . Similarly, by constructing and aggregating multiple 3-D tensors at different sampling moments, we can obtain the 4-D tensor input for the CNN-LSTM model.

B. FREQUENCY PREDICTION AND PERFORMANCE ANALYSIS

In this section, we adopt the CNN-LSTM model to predict the COI frequency after fault, which can provide more information than previous studies in the literature on the minimum and maximum frequency prediction. In order to test the performance of the CNN-LSTM model, among 2,040 samples, 1,740 samples are randomly selected as the training set, and the remaining 300 samples are used as the test set. Training and test process are implemented on the Keras framework.

The structure and parameters of the CNN-LSTM model need to be pre-set to make a trade-off between training time and accuracy. After several experiments, the trained CNN-LSTM model in this paper contains 4 Conv layers, one LSTM layer, and a fully connected layer to generate the output.

The proposed CNN-LSTM is compared with ANN and SVM. CNN-LSTM and ANN can be directly applied to predict dynamic frequency; while SVM is a single output model, which we need to be run multiple times to obtain a frequency curve. Inputs for SVM and ANNA are vectors containing power system state variables, i.e., generator active powers and bus voltages. All methods start at the time when a fault occurs.

In order to evaluate the prediction performance of three methods, the following index is adopted: mean absolute percentage error (MAPE) and root mean square error (RMSE) between predicted curves and actual curves, steady state frequency f_s and minimum frequency f_{min} after disturbance.

1) PREDICTION OF MINIMUM FREQUENCY

Minimum frequency is a crucial indicator of system frequency stability. This subsection presents the performance

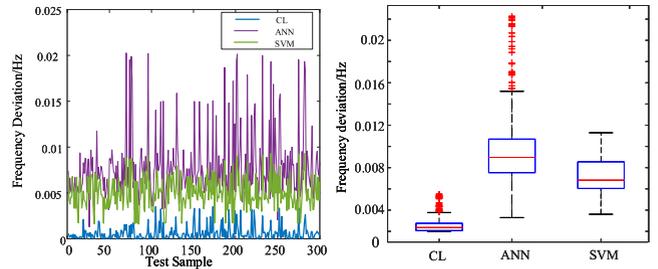


FIGURE 5. Error comparison of minimum frequency among three methods.

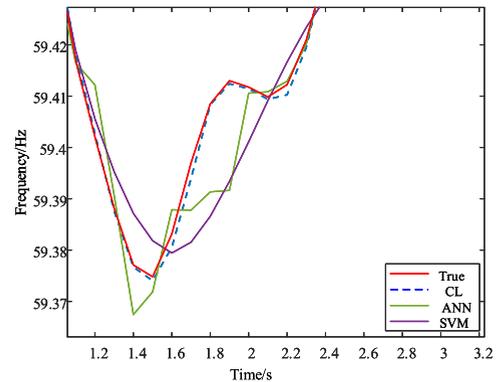


FIGURE 6. Comparison of minimum frequency prediction among three methods.

comparison of minimum frequency predictions from different methods. Fig. 5 shows the prediction error and the box plots of MAPE of minimum frequencies from three methods. Fig. 6 presents an example of the dynamic frequency prediction, in which the minimum frequency part is magnified for comparison, and CL denotes the proposed CNN-LSTM method. These comparisons show that the result from CNN-LSTM is closest to the true curve (red) in both values of X-axis and Y-axis. It means that the CNN-LSTM-based method shows the accurate minimum frequency value and time.

The powerful performance of CNN-LSTM comes from the fact that the proposed CNN-LSTM model can extract rich spatial-temporal correlation characteristics from the tensor inputs, explore the local features of the data, and combine the local features into global characteristics for prediction. On the other hand, the fully connected structure in ANN results in an inefficient training process and the poor prediction accuracy.

C. LOAD SHED AMOUNT PREDICTION AND ANALYSIS

Based on the prediction results, there are 141 cases in test samples that need to shed load in order to maintain frequency stability. Therefore, the measurements of these 141 cases are inputs to the second stage CNN-LSTM model for optimal load shedding strategy. The predicted optimal load shedding amount on load 54 is compared with the actual output value of the sample, as shown in Fig. 7. The optimal load shedding amount of each load node prediction is very close to the optimization results in test samples, with the maximum power

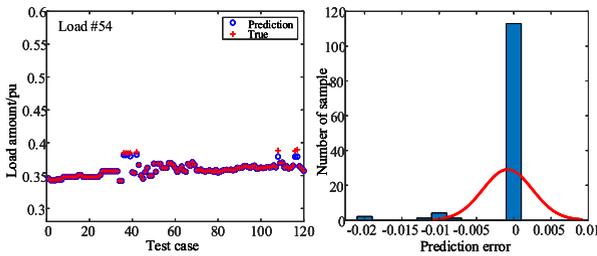


FIGURE 7. Prediction of load shedding amount.

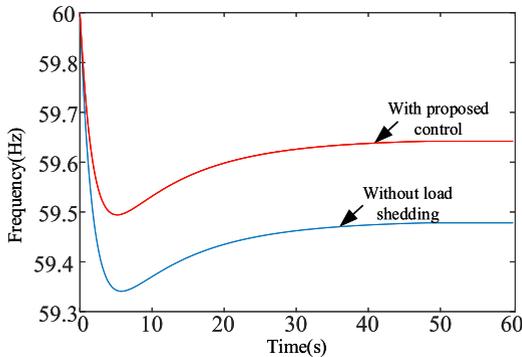


FIGURE 8. Effect of load shedding with the proposed CNN-LSTM method.

error only within 0.015 pu, which indicates that the model can provide highly accurate prediction results.

The control effect of load shedding on test case 10 is shown in Fig. 8. From this figure, it can be seen that the proposed load shedding method can ensure that the minimum frequency is maintained above 59.5Hz after a fault occurs.

D. TRANSFER LEARNING PERFORMANCE

The CNN-LSTM model has demonstrated powerful performance when both training and test data come from the same system. However, in actual power systems, the characteristics and distribution of measurements from different systems are inconsistent, resulting in the issues of taking a long time to obtain new training data and re-train the model for the application to a new system. In addition, there are insufficient measurements in some power systems, which can lead to a poor performance of DL models. Therefore, transfer learning is designed and applied to IEEE 39-bus system and South Carolina 500-bus system. There are 10 generators and 46 lines in IEEE 39-bus system, and the ACTIVSg500 system is built from a real power system [29].

In order to validate whether the features learned from IEEE 118-bus system are similar to that from other systems, we extract and compare the 3-D tensors from different systems. It is noted that there are some similarities between different inputs. For example, Fig. 9 shows the bus voltage feature map from different systems, which demonstrates the similarity exists across different systems. The vertical axis represents the voltage value in the dimension reduced electric distance matrix. The similarity is the linear increase tendency of voltage feature map. Since the status feature maps from

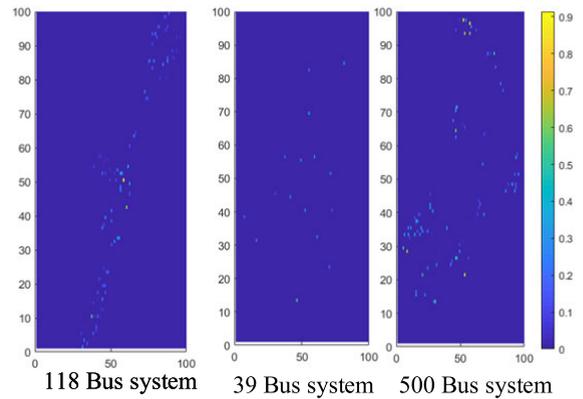


FIGURE 9. Comparison of feature maps.

different systems are similar, it is possible to transfer and apply the trained model to other systems, which can significantly reduce the training and sample generation cost.

After verifying that it is feasible to transfer the CNN-LSTM model for applications to other systems, steps to apply the trained model in the last section to another two test systems for frequency prediction are shown as follows.

1) PREPARING NEW TRAINING AND TEST DATA

Sample data are generated based on the steps introduced in the last subsection. There are 1,800 and 2,100 samples for IEEE 39-bus system and ACTIVSg500 system, respectively. In order to validate the superiority of the transferred model, non-transferred CNN-LSTM is adopted to perform the same prediction task, and only 300 samples are used for training and all the other samples are used for testing in both TL and non-TL models.

2) IEEE 39-BUS SYSTEM

Fig. 10 shows the comparison of frequency prediction for the 157th test sample. It is clear that the prediction based on transfer learning performs better. This advantage is enabled through the powerful learning ability of local patterns, while non-transferred ML models suffer from insufficient training samples with reduced prediction accuracy. In order to verify the advantages of the proposed transfer CNN-LSTM method, the proposed method and three other methods including ANN, SVM and EL are applied to the IEEE 39-bus system. There are only 300 samples for training in all four methods, and a 30 dB measurement noise is imposed on all inputs. Table 3. shows the results under different conditions. It can be observed that the proposed transfer CNN-LSTM method can always maintain a highly accurate result. Therefore, the conclusion can be drawn that the proposed transfer CNN-LSTM method has superior robustness to other existed methods.

3) ACTIVSg500 SYSTEM

Similarly, we can obtain the prediction result for ACTIVSg500 system. Tables 4 compares the RMSE and MAPE for the minimum frequency prediction in two test

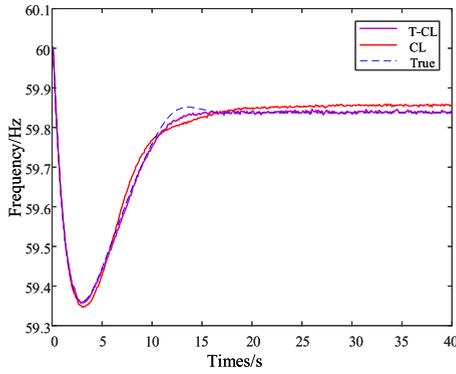


FIGURE 10. Comparison of dynamic frequency curves in IEEE 39-bus system.

TABLE 3. Test accuracy of different methods on IEEE 39-bus system.

Method	RMSE	RMSE (with Noise)
Transfer CNN-LSTM	0.0042	0.0057
SVM	0.0155	0.0181
ANN	0.0173	0.0192
EL	0.0159	0.0188

TABLE 4. Minimum frequency accuracy variation comparison.

Method	IEEE 39		ACTIVSg500	
	RMSE	MAPE	RMSE	MAPE
TL	0.0042	0.0056	3.83e-5	3.024e-5
CL	0.0084	0.0063	6.904e-4	6.078e-4

systems. It can be seen that the transferred model can improve the performance when facing the disadvantage of insufficient training data.

In addition, from the experiment, the TL model can obtain the optimal solution within 50 iteration steps, which is much faster than that of the non-TL model which takes about 230 steps to converge. This indicates the convergence rate can be significantly accelerated by transfer learning.

E. RELIABILITY IN REAL POWER SYSTEMS

As measurement noise and time delay exist in real power systems [9], it is imperative to improve the robustness and computational efficiency. In order to evaluate the robustness of CNN-LSTM method under the noise condition, different levels of white Gaussian noise are added to the corresponding inputs. The measurement noise level is described based on signal-to-noise ratios (SNRs) with the unit dB.

Fig. 11 shows the comparison of the RMSE index of dynamic frequency prediction from each method under different noise levels. The comparison results clearly show that the robustness of ANN and SVM usually decreases with the increase of prediction accuracy, and the proposed CNN-LSTM method shows significant advantages than the other two methods. This advantage comes from the fact that the CNN-LSTM’s local connection and pooling make the features more robust to noise and deformation.

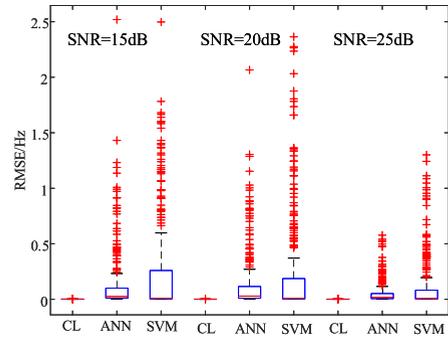


FIGURE 11. Box plot comparison for RMSE of dynamic frequency.

In [30], the authors point out that the structure of the network will greatly affect the robustness of the model. ANN is a fully connected neural network, which may be the reason why ANN is less robust. The local connection and pooling of CNN makes the features more robust to noise and deformation [30].

For measurement delays, the proposed two-stage CNN-LSTM method can quickly predict system dynamic frequency after fault, and present the optimal amount of load to be shed to recover frequency after a fault. The computational time by CNN-LSTM and load shedding algorithm is 0.024 and 0.0105s. Considering that the wide area measurement communication delay time is about 60ms and the circuit breaker operation time is 60ms, the proposed algorithm can complete dynamic frequency prediction and load shedding control within 150ms, which meets the requirement of online emergency control.

V. CONCLUSION

This paper proposed a transfer and CNN-LSTM-based method to accelerate and improve the accuracy of the dynamic frequency and optimal load shedding prediction process. The proposed method exploits system spatial-temporal information and mines the local features of inputs, which highly improves the performance compared with other machine learning methods. In addition, the proposed transfer learning process significantly enhances the generalization ability of the proposed CNN-LSTM, and overcomes the shortage of insufficient data and operation condition changes in real power systems, which provides remarkable training with sample generation savings. Simulation results on IEEE 118-bus test system, New England 39-bus test system, and ACTIVSg500 system validate the effectiveness of the proposed CNN-LSTM method. The comparisons show the proposed approach has superior accuracy because of the ability to exploit spatial-temporal information. Moreover, the proposed method is applicable to online frequency stability assessment and control to prevent frequency collapse. With more and more renewable energy integrated with modern power system, future work will be focused on investigating the impact of renewable energy generation on transient stability and extend the method to prediction of other states such as voltage.

REFERENCES

- [1] Z. Zhong, C. Xu, B. J. Billian, L. Zhang, S.-J.-S. Tsai, R. W. Conners, V. A. Centeno, A. G. Phadke, and Y. Liu, "Power system frequency monitoring network (FNET) implementation," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1914–1921, Nov. 2005.
- [2] J. Zhao, Y. Tang, and V. Terzija, "Robust online estimation of power system center of inertia frequency," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 821–825, Jan. 2019.
- [3] Q. Wang, F. Li, Y. Tang, and Y. Xu, "Integrating model-driven and data-driven methods for power system frequency stability assessment and control," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4557–4568, Nov. 2019.
- [4] L. Che, X. Liu, and Z. Shuai, "Optimal transmission overloads mitigation following disturbances in power systems," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2592–2604, May 2019.
- [5] P. M. Anderson and M. Mirheydar, "A low-order system frequency response model," *IEEE Trans. Power Syst.*, vol. 5, no. 3, pp. 720–729, Aug. 1990.
- [6] Q. Wang, C. Zhang, Y. Lü, Z. Yu, and Y. Tang, "Data inheritance-based updating method and its application in transient frequency prediction for a power system," *Int. Trans. Electr. Energy Syst.*, vol. 29, no. 6, Jun. 2019, Art. no. e12022.
- [7] L. Liu, W. Li, Y. Ba, J. Shen, C. Jin, and K. Wen, "An analytical model for frequency nadir prediction following a major disturbance," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 2527–2536, Jul. 2020.
- [8] S. Azizi, M. Sun, G. Liu, and V. Terzija, "Local frequency-based estimation of the rate of change of frequency of the center of inertia," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4948–4951, Nov. 2020.
- [9] U. Rudez and R. Mihalic, "WAMS-based underfrequency load shedding with short-term frequency prediction," *IEEE Trans. Power Del.*, vol. 31, no. 4, pp. 1912–1920, Aug. 2016.
- [10] P. He, B. Wen, and H. Wang, "Decentralized adaptive under frequency load shedding scheme based on load information," *IEEE Access*, vol. 7, pp. 52007–52014, 2019.
- [11] B. Potel, V. Debusschere, F. Cadoux, and U. Rudez, "A real-time adjustment of conventional under-frequency load shedding thresholds," *IEEE Trans. Power Del.*, vol. 34, no. 6, pp. 2272–2274, Dec. 2019.
- [12] A. K. Singh and M. Fozdar, "Event-driven frequency and voltage stability predictive assessment and unified load shedding," *IET Gener., Transmiss. Distrib.*, vol. 13, no. 19, pp. 4410–4420, Oct. 2019.
- [13] F. Chollet, *Deep Learning mit Python und Keras*. Bonn, Germany: MITP-Verlags, 2018.
- [14] T. Amraee and S. Ranjbar, "Transient instability prediction using decision tree technique," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3028–3037, Aug. 2013.
- [15] T. Behdadnia, Y. Yaslan, and I. Genc, "A new method of decision tree based transient stability assessment using hybrid simulation for real-time PMU measurements," *IET Gener., Transmiss. Distrib.*, vol. 15, no. 4, pp. 678–693, Feb. 2021.
- [16] Q. Bo, X. Wang, and K. Liu, "Minimum frequency prediction of power system after disturbance based on the V-support vector regression," in *Proc. Int. Conf. Power Syst. Technol.*, Oct. 2014, pp. 614–619.
- [17] S. Tripathi and S. De, "Dynamic prediction of powerline frequency for wide area monitoring and control," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 2837–2846, Jul. 2018.
- [18] Y. Du, F. Li, J. Li, and T. Zheng, "Achieving 100x acceleration for N-1 contingency screening with uncertain scenarios using deep convolutional neural network," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 3303–3305, Jul. 2019.
- [19] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, and W.-K. Wong, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, 2015, pp. 802–810.
- [20] M. D'Incecco, S. Squartini, and M. Zhong, "Transfer learning for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1419–1429, Mar. 2020.
- [21] C. Ren and Y. Xu, "Transfer learning-based power system online dynamic security assessment: Using one model to assess many unlearned faults," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 821–824, Jan. 2020.
- [22] P. Lagonotte, J. C. Sabonnadiere, J.-Y. Leost, and J.-P. Paul, "Structural analysis of the electrical system: Application to secondary voltage control in France," *IEEE Trans. Power Syst.*, vol. 4, no. 2, pp. 479–486, May 1989.
- [23] P. Kundur, N. J. Balu, and M. G. Lauby, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill, 1994, vol. 7.
- [24] J. H. Chow and J. J. Sanchez-Gasca, "Turbine-governor models and frequency control," in *Power System Modeling, Computation, and Control*. IEEE, 2020, pp. 327–370.
- [25] K. Liu and X. Wang, "A wide-area measurement data based method for power system dynamic frequency analysis," *Power Syst. Technol.*, vol. 37, no. 8, pp. 2201–2206, 2013.
- [26] L. Tang and J. McCalley, "Two-stage load control for severe under-frequency conditions," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 1943–1953, May 2016.
- [27] E. Bompard, R. Napoli, and F. Xue, "Analysis of structural vulnerabilities in power transmission grids," *Int. J. Crit. Infrastruct. Protection*, vol. 2, nos. 1–2, pp. 5–12, May 2009.
- [28] J. Lin, Y. Zhang, X. Wang, and Q. Chen, "Post-disturbance dynamic frequency features prediction based on convolutional neural network," 2019, *arXiv:1909.09323*. [Online]. Available: <https://arxiv.org/abs/1909.09323>
- [29] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3258–3265, Jul. 2017.
- [30] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 7533, no. 521, pp. 436–444, 2015.



JIAN XIE (Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL, USA. His research interests include power system stability and machine learning in power system resilience enhancement.



WEI SUN (Member, IEEE) received the Ph.D. degree from Iowa State University, Ames, IA, USA, in 2011. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL, USA. He is also the Director of the Siemens Digital Grid Laboratory. His research interests include power system restoration, self-healing smart grid, and cyber-physical system security and resilience.

...