Rotor Angle Stability Prediction Using Temporal and Topological Embedding Deep Neural Network Based on Grid-informed Adjacency Matrix

Peiyuan Sun, Long Huo, Xin Chen, and Siyuan Liang

Abstract—**Rotor angle stability (RAS) prediction is critically essential for maintaining normal operation of the interconnected synchronous machines in power systems. The wide deployment of phasor measurement units (PMUs) promotes the development of data-driven methods for RAS prediction. This paper proposes a temporal and topological embedding deep neural network (TTEDNN) model to accurately and efficiently predict RAS by extracting the temporal and topological features from the PMU data. The grid-informed adjacency matrix incorporates the structural and electrical parameter information of the power grid. Both the small-signal RAS with disturbance under initial operating conditions and the transient RAS with short circuits on transmission lines are considered. Case studies of the IEEE 39-bus and IEEE 300-bus power systems are used to test the performance, scalability, and robustness against measurement uncertainties of the TTEDNN model. Results show that the TTEDNN model performs best among existing deep learning models. Furthermore, the superior transfer learning ability from small-signal RAS conditions to transient RAS conditions has been proved.**

Index Terms—**Rotor angle stability, topological embedding, deep learning, graph convolution network.**

I. Introduction

DIGITAL transformation plays an essential role in the modernization of the power systems. The wide deployment of phasor measurement units (PMUs) enables data collection on wide-area power systems, facilitating engineers to analyze power system dynamics and predict system stability in a data-driven manner [1], [2]. The instability problem has been traditionally associated with rotor angle stability (RAS) [3], which refers to the ability of synchronous machines in a power system to remain in synchronism after being subjected to a disturbance [4]. With the increasing penetration of renewable generations in modern power systems, the RAS requires a more accurate and fast prediction due to the emerging low inertia and stochastic characteristics. Currently, the prediction methods of RAS can be classified into two categories, i.e., the model-driven methods and the data-driven methods [5].

One of the commonly used model-driven methods is the laborious time-domain simulation (TDS) based on high-dimensional nonlinear differential-algebraic equations (DAEs) that express the dynamics of power systems [6]. TDS is time-consuming since it demands the whole state trajectories to reveal the system stability. Although different approaches have been proposed to accelerate the TDS process, such as parallel computing [6] and advanced hardware [7], huge computation resources are still required to handle the increasing complexity of power systems and diverse operational scenarios. The Lyapunov function family's model-driven method is used for an analytical approach for stability assessment in power systems [8]. Unfortunately, finding a Lyapunov function to evaluate the RAS of power systems accurately has proved very difficult [9]. Therefore, some researchers find model-free methods based on maximum Lyapunov exponent [10].

Recently, the data-driven methods, especially the deep learning methods, attracted a lot of research interest in predicting RAS in power systems [1], [2], [11], [12]. Compared with model-driven methods, deep learning performance does not rely on the prior knowledge and model details of power systems. Furthermore, the strong generalization ability and the nature of offline training and online diagnosis pattern of deep learning provide great potential to meet the high accuracy and fast online requirements in practical applications [1].

Among the existing deep learning models, convolution neural network (CNN) has made significant achievements in many fields [13], including RAS prediction in power systems. For example, [2] proposed a fast power system RAS evaluation model based on CNN and the voltage phasor complex plane image. Reference [11] designed cascaded CNNs to capture data from different TDS time intervals, extract features, predict stability probability, and determine TDS termination. Besides, other deep learning models are also used for

Manuscript received: May 26, 2023; revised: June 27, 2023; accepted: July 23, 2023. Date of CrossCheck: July 23, 2023. Date of online publication: August 29, 2023.

This work was supported in part by the National Natural Science Foundation of China (No.21773182) and the HPC Platform, Xi'an Jiaotong University.

P. Sun is with the School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: spy2018@stu.xjtu.edu.cn).

L. Huo and X. Chen (corresponding author) are with the School of Electrical Engineering, and the Center of Nanomaterials for Renewable Energy, State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: eehl921105@stu.xjtu.edu.cn; xin.chen.nj@stu.xjtu.edu.cn).

S. Liang is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, 999077, Hong Kong, China (e-mail: siyuan.liang@link.cuhk.edu.hk).

DOI: 10.35833/MPCE.2023.000364

predicting RAS, such as stacked denoising autoencoder (SDAE) [14] and long short-term memory (LSTM) network [12]. The prediction results from deep learning models can be used to accomplish further operational tasks such as preventive control [15]. However, since power systems are complex dynamical networks, the architectures of the above-mentioned deep learning models need proper interpretability with the spatial correlations of power systems. Therefore, effectively using the important topological information of power network structures in deep learning remains challenging.

The graph neural network (GNN) is a promising deep learning model to extract features of the spatial correlations of power systems since GNN can naturally map the power network structure into its neural network connections. As one of the GNN family, graph convolution network (GCN) [16] combines topological structure with convolution algorithm and has been proven extremely powerful for the complex dynamical network analysis [17]. GCN demonstrates good classification and prediction capability with the graph-structured data in power systems [18]. For example, [19] developed an interpretable GCN to guide cascading failure search efficiently. Nevertheless, the GCN could be more adept at capturing the sequential characteristics, i.e., the temporal information of time series of power system dynamics. Additional techniques are needed to extract features from the time domain of power system transient dynamics. For sequence modeling [20], the convolutional technique has been developed extensively in recent works and outperformed the baseline of well-known recurrent network architectures for sequence modeling tasks [21]. As one of the convolutional technique-based recurrent architectures, temporal convolutional network, also known as TCN, has been utilized for time-series predictions in power systems, demonstrating powerful memory ability [22].

Some related methods of GNN family based RAS prediction have been proposed in recent studies. Reference [23] introduced the graph attention network (GAT) for both RAS and short-term voltage instability prediction [23]. Reference [24] proposed the multi-graph attention network with residual structure (ResGAT) for RAS assessment, which is adapted to the power system topology changes [24]. A similar GCN architecture with a residual mechanism was designed to overcome the network degradation phenomenon during model training [25]. Later, an attention-based hierarchical dynamic graph pooling network was proposed to make the deep learning model more robust against system-scale changes [26]. A multi-task recurrent graph convolutional network (RGCN) combined with LSTM was introduced for stability classification as well as critical generator identification [27]. However, to the best of our acknowledge, no existing studies considered the whole categories of RAS prediction, i.e., they focused on the scenario for either small-signal RAS [25] or transient RAS [23], [24], [26], [27]. The comprehensive prediction performance of a GNN model on both small-signal and transient RAS is still unclear. Meanwhile, few related studies discussed the model robustness against practical measurement uncertainty, i. e., the measurement noise and sampling cycle of PMUs.

In this paper, the temporal and topological embedding deep neural network (TTEDNN) model is proposed by combining GCN and TCN to capture the spatio-temporal features of transient dynamics in power systems for RAS prediction. Generally, the main contributions of this paper are as follows.

1) The TTEDNN model is proposed to predict RAS by the temporal and spatial features extracted from the post-disturbed transient dynamics. The grid-informed adjacency matrix is used to incorporate the structural and electrical parameter information of the power grid.

2) The robustness of the TTEDNN model against different levels of measurement noise and different PMU data cycles is illustrated.

In addition, the transfer learning capability of the TTEDNN model is investigated. It is found that the TTEDNN model trained with the small-signal perturbation dataset can be used as a pre-trained model for predicting the transient RAS.

The rest of this paper is organized as follows. Section II introduces the RAS of power systems. Section III proposes the architecture of the TTEDNN model. Case studies are given in Section IV. The conclusion remarks are drawn in Section V.

II. RAS of Power Systems

In this section, the concept of RAS in a power system, the RAS assessment, and disturbances imposed for the study of RAS are described.

A. Concept of RAS in a Power System

Generally, the dynamics of a power system is governed by a set of DAEs, which can be expressed in the compact form as:

$$\begin{cases} \dot{\boldsymbol{x}}=\boldsymbol{f}(\boldsymbol{x},\boldsymbol{y},t) \\ \boldsymbol{0}=\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y},t) \end{cases} \tag{1}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ denote the state and algebraic variables, respectively; $\boldsymbol{f}(\cdot)$ denotes the dynamics of synchronous machines and control systems; and $\boldsymbol{g}(\cdot)$ denotes the load flow of a power system.

Given an initial condition of $\boldsymbol{x}$ and $\boldsymbol{y}$, the solution of (1) yields time-varying trajectories of the state variables $\boldsymbol{x}$, i.e., the rotor angles and frequencies, and algebraic variables $\boldsymbol{y}$, i.e., the bus voltages and active power injections. The RAS of a power system is concerned with the ability of the interconnected synchronous machines in a power system to remain in synchronism under normal operating conditions and to regain synchronism after being subjected to a small or large disturbance [28]. According to the nature of stability problems, the RAS can be classified in terms of two subcategories: the small-signal RAS for small disturbances and the transient RAS for large disturbances. The small-signal RAS depends on the initial operating state of the system [4], i.e., the initial condition of $\boldsymbol{x}$ in (1). The transient RAS is concerned with severe disturbance such as $N-1$ contingency

[4], i.e., short circuits on transmission lines, which can be reflected by the change of $\boldsymbol{y}$ in (1).

B. RAS Assessment

The transient stability index (TSI) σ [29], a common indicator for assessing the stability status of the system, is used in this paper:

$$\sigma=\frac{360-\left|\Delta\delta_{ij}\right|_{\max}}{360+\left|\Delta\delta_{ij}\right|_{\max}} \tag{2}$$

where $\left|\Delta\delta_{ij}\right|_{\max}$ is the maximum absolute value of the difference between the rotor angles of the synchronous machines i and j. The system is stable if $\sigma>0$; otherwise, it is unstable. The traditional method to evaluate TSI requires the dynamic trajectories of rotor angles from TDS, i.e., numerically solving the DAEs in (1). For power systems with relatively large scales, the TDS becomes time-consuming, and the demand for fast on-line RAS assessment cannot be satisfied. We proposed an effective data-driven RAS prediction model called the TTEDNN model by utilizing the information embedded in PMU data, which will be introduced in Section III.

C. Disturbances

For the small-signal RAS, the disturbances of initial operating conditions in power systems have varieties of sources, including load variations, market trading, and renewable energy fluctuations. For example, energy trading happens most of the time, inducing several considerable local frequency deviations per day, even four times per hour. We mainly focus on the two most important variables, i.e., the rotor angle and rotor angular speed, for the RAS assessment and generator stability ranking [6]. The distribution of the frequency in realistic power systems demonstrates the non-Gaussian characteristics of heavy tail and skewness, which can be more accurately described by the Levy-stable distribution [30]. Additionally, the maximum fluctuations of frequency should also be set to be an appropriate value; otherwise, the disturbances will be too small to disturb the system or too large to be found in realistic power systems. Normally, the disturbances of frequency Δf are bounded to ±1% to ±4% of rated frequency (50 Hz or 60 Hz) [31]. Thus, we set the disturbance limit of angular speed as $\Delta\omega_{\max}=2\pi\Delta f=10$ rad/s. Considering a power system with N nodes, we define $m<N$ to be the number of nodes simultaneously disturbed, where $m=1$ and $m>1$ refer to the single-node disturbance case and the multiple-node disturbance case, respectively. We focus on the disturbances of small-signal RAS on rotor angle and angular speed $[\delta_i,\omega_i]$, $i=1,2,\dots,N$. For the transient RAS, power systems are subjected to more severe disturbances, i.e., the $N-1$ contingency. We consider scenarios by triggering short circuits of transmission lines and predict the transient RAS at post-fault stages. The more severe disturbances such as $N-s$ ($s\geq 2$) contingencies are rare events in power systems [32] and are not taken into account in this paper.

III. Architecture of TTEDNN Model

The TTEDNN model is proposed to predict the RAS in power systems by extracting the temporal and topological features embedded in the time-series data of PMUs.

A. Data Representation

A sample for training and testing the TTEDNN model composes of an input $\boldsymbol{X}$ and its corresponding label y. The data of transient dynamics after disturbances is collected by PMUs and used as the input of the TTEDNN model, representing the multivariate time series of state variables $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\dots,\boldsymbol{x}_F\}\in\mathbb{R}^{F\times N\times l}$, where F is the number of state variables; N is the number of nodes in power systems; l is the length of time series under fixed sampling frequency f_s of PMUs; and $\boldsymbol{x}_i$ is the time series of the i^{th} state variables, which can be expressed as:

$$\boldsymbol{x}_i=\begin{bmatrix} x_i^1(0) & x_i^2(0) & \dots & x_i^N(0) \\ x_i^1(1) & x_i^2(1) & \dots & x_i^N(1) \\ \vdots & \vdots & & \vdots \\ x_i^1(l-1) & x_i^2(l-1) & \dots & x_i^N(l-1) \end{bmatrix}^{\mathrm{T}} \tag{3}$$

where the j^{th} column of $\boldsymbol{x}_i$ is the post-fault time series of node j ($j=1,2,\dots,N$) in power systems. Four state variables are used for the input of the TTEDNN model, including the bus relative phase U_θ, the bus voltage magnitude $|U|$, the rotor angle δ, and the rotor angular speed ω, and therefore $F=4$. Although the rotor angle and rotor angular speed of synchronous machines cannot be measured directly by PMU, recent studies have shown that signals of rotor angle and rotor angular speed are available by PMU data based estimation algorithms [33], [34].

The label y is a binary, indicating the final RAS concerns the input $\boldsymbol{X}$. According to the TSI defined in (2), y is determined as:

$$y=\begin{cases} 1 & \sigma>0 \\ 0 & \sigma\leq 0 \end{cases} \tag{4}$$

where $y=1$ and $y=0$ correspond to the stable state and unstable state, respectively. The output of the TTEDNN model gives the probability p that the power system will evolve to a stable or unstable state. Numerically, we take $p>0.5$ for the stable state and $p\leq 0.5$ for the unstable state.

B. Structure of TTEDNN Model

The structure of the TTEDNN model is shown in Fig. 1. The structure has three main parts, the graph convolution (GC) modules, temporal convolution (TC) module, and multi-layer perception (MLP) prediction layer. Each GC module has five neurons, where $\vec{h}_i$ and $\vec{h}'_i$ ($i=a$ to e) represent the input states of the five neurons in each GC module. The operation of the GC module is to update the output state of the focused neurons (red) using the adjacency matrix, which is only relevant to the adjacent neurons (blue) rather than others (gray). The topological features are then extracted by the fully-connected (FC) layer and further processed by the TC module, composed of R residual blocks with dilated factors d_1 to d_R. In the end, the MLP layer generates the prediction probability function.

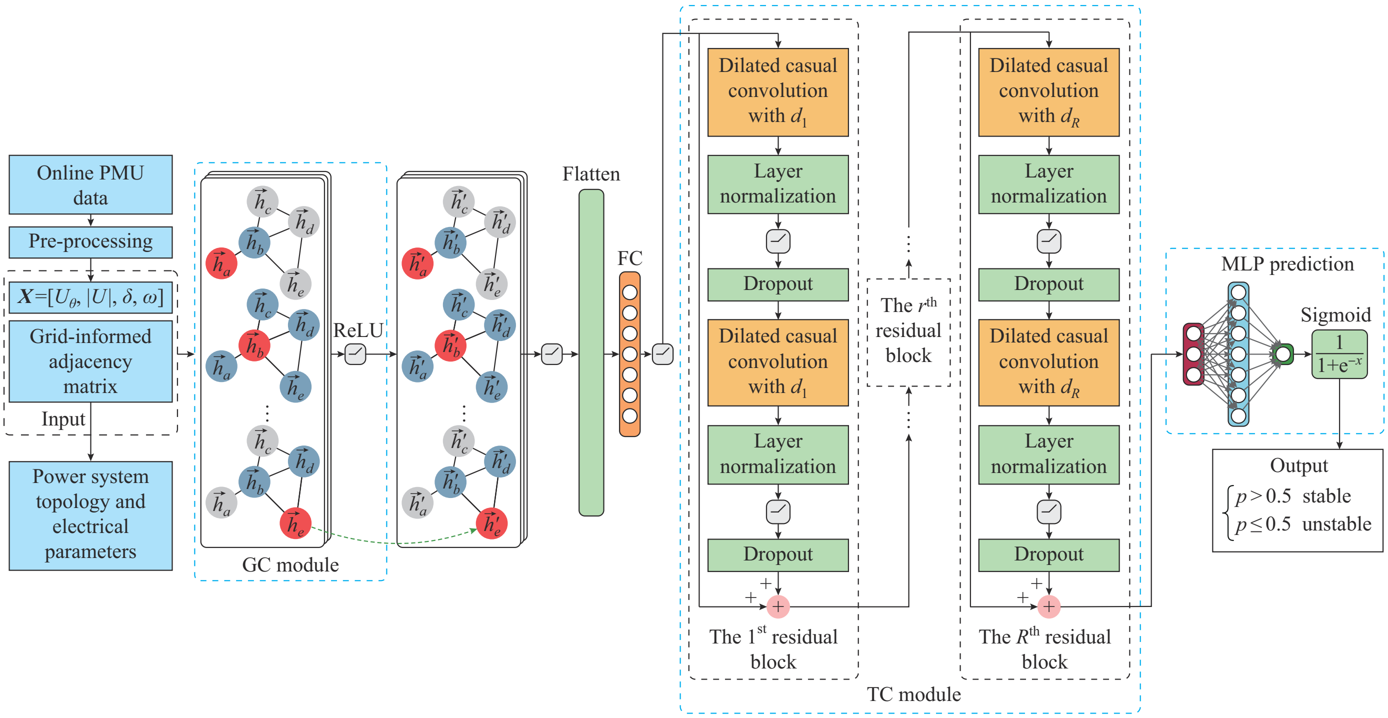

Fig. 1. Structure of TTEDNN model.

1) GC Modules

The TTEDNN model starts with n GC modules to extract topological features. Each GC module is sequentially composed of a GCN layer, a batch normalization (BN) layer, and a rectified linear unit (ReLU) activation function.

The structure of the GCN layer can be represented as an undirected graph [16] $G=(\mathcal{V},\mathcal{E},\boldsymbol{B})$, where $\mathcal{V}\in\mathbb{R}^N$ is the set of neurons; $\mathcal{E}\in\mathbb{R}^E$ is the set of links between neurons; and $\boldsymbol{B}\in\mathbb{R}^{N\times N}$ is the adjacency matrix of the graph. The re-normalized adjacency matrix $\hat{\boldsymbol{B}}'$ is often used in the GCN layer:

$$\hat{\boldsymbol{B}}'=\hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{B}}\hat{\boldsymbol{D}}^{-\frac{1}{2}} \tag{5}$$

where $\hat{\boldsymbol{B}}=\boldsymbol{B}+\boldsymbol{I}_N$ denotes the adjacency matrix with self-loop, and $\boldsymbol{I}_N$ is the identity matrix; and $\hat{\boldsymbol{D}}$ is the diagonal node degree matrix, $\hat{D}_{i,i}=\sum_j \hat{B}_{i,j}$.

The operation of the i^{th} GC module is defined as:

$$\boldsymbol{H}^{i+1}=ReLU(BN(\hat{\boldsymbol{B}}'\boldsymbol{H}^i\boldsymbol{W}^i+\boldsymbol{b}^i)) \tag{6}$$

where $ReLU(\cdot)$ denotes the activation of ReLU function; $\boldsymbol{H}^{i+1}\in\mathbb{R}^{N\times C}$ denotes the output states of the i^{th} GC module as well as the input states of the $(i+1)^{\text{th}}$ GC module; $BN(\cdot)$ is the batch normalization function; $\boldsymbol{W}^i\in\mathbb{R}^{C\times F}$ is the network weight of GCN layer; and $\boldsymbol{b}^i\in\mathbb{R}^{N\times F}$ denotes the bias. Then, the last GC module is connected to a flatten layer to reshape the output states. Following the flatten layer, a FC layer is adopted to extract the topological features to feed into the TC module.

2) TC Module

The TC module is used to further extract temporal features based on the output of the last GC module. As shown in Fig. 1, the TC module is composed of R residual blocks using the 1D fully convolutional network (1D-FCN). The 1D-FCN utilizes the dilated casual convolution consisting of the dilated convolution technique and casual convolution technique, and a residual connection. The solely 1D-FCN structure produces an output with the same length as its input, and the casual convolution technique ensures that the output emitted by a 1D-FCN layer at time step t is convolved only with elements from time step t and earlier in the previous layer. Therefore, the TC module considers the whole history information for future prediction. The disadvantage is that the casual convolution technique can only look back to the historical information with the size linear to the network depth. The casual convolution can be optimized with the dilated convolution technique by introducing the exponential receptive field. Therefore, the TC module can take all historical information into account with smaller network depth. Specifically, the dilated convolution operation $F(\cdot)$ of the r^{th} residual block $(r=1,2,\ldots,R)$ can be defined as a dilated transformation of a 1D time series data $\boldsymbol{x}$:

$$F(j)=\sum_{i=0}^{k-1} f(i)x_{j-d_r\cdot i} \tag{7}$$

where $f(i)$ is the convolution filter $f\colon 0,1,\ldots,k-1\rightarrow\mathbb{R}$; k denotes the filter size; d_r denotes the dilated factor of the r^{th} residual block, and there are R dilated factors for R residual blocks as d_1 to d_R; and $j=1,2,\ldots,n$, and n is the size of $\boldsymbol{x}$. Adjusting the dilation size can allow the top level of 1D-FCN to represent a wider range of the input as much as possible, thus expanding the receptive field extensively. Moreover, the residual connection is used to stabilize the network training to make the layers learn deep residual information as the modifications to the identity mapping when the casual convolution is dilated with considerable depth. As a result, the operation of the r^{th} residual block is defined as:

$$z^r = ReLU(\boldsymbol{x}^r + LF(F(\boldsymbol{x}^r))) \tag{8}$$

where $r = 1, 2, \ldots, R$; $\boldsymbol{x}^r$ and z^r denote the input and output of the r^{th} residual block, respectively; and $LF(\cdot)$ denotes the layer normalization technique.

3) MLP Prediction Layer

Finally, MLP with the sigmoid activation function is utilized to generate the prediction as a probability function.

Two custom-made training technics called the grid-informed adjacency matrix and class-weighted loss function are also introduced to improve the prediction performance of the TTEDNN model.

1) Grid-informed adjacency matrix

Taking the electrical and structural properties of a power system into consideration, three different grid-informed adjacency matrices $\boldsymbol{B}$ are proposed in the GC modules for the spatial feature extraction. First, according to [16], we choose the binary adjacency matrix of a power system added with self-connection (denoted as $\boldsymbol{B}^1$), whose element is expressed in (9). It considers the topology of the power system but ignores the weight of edges. Second, the active power flows are added as the weights of edges in the adjacency matrix (denoted as $\boldsymbol{B}^2$), whose element is expressed in (10). Third, we treat the maximum transmission capability of a transmission line as off-diagonal elements in the adjacency matrix together with active power injections for diagonal ones (denoted as $\boldsymbol{B}^3$), whose element is expressed in (11).

$$B_{ij}^1 = \begin{cases} 1 & (i,j) \in \mathcal{E}_p^{norm} \text{ or } i = j \\ 0 & (i,j) \notin \mathcal{E}_p^{norm} \text{ or } (i,j) \in \mathcal{E}_p^{con} \end{cases} \tag{9}$$

$$B_{ij}^2 = \begin{cases} K_{ij}\sin(\delta_i - \delta_j) & (i,j) \in \mathcal{E}_p^{norm} \\ 0 & (i,j) \notin \mathcal{E}_p^{norm} \text{ or } (i,j) \in \mathcal{E}_p^{con} \end{cases} \tag{10}$$

$$B_{ij}^3 = \begin{cases} K_{ij} & (i,j) \in \mathcal{E}_p^{norm} \\ 0 & (i,j) \notin \mathcal{E}_p^{norm} \text{ or } (i,j) \in \mathcal{E}_p^{con} \\ P_i & i = j \end{cases} \tag{11}$$

where $\mathcal{E}_p^{norm}$ and $\mathcal{E}_p^{con}$ are the normal and faulty transmission line sets under contingencies, respectively; (i,j) denotes the transmission line between node i and node j; K_{ij} is the maximum transmission capability of the transmission line (i,j); and P_i is the active power injection of node i. If $(i,j) \notin \mathcal{E}_p^{norm} \cup \mathcal{E}_p^{con}$, no transmission line exists between node i and node j. If $(i,j) \in \mathcal{E}_p^{con}$, (i,j) is a faulty transmission line during the contingency.

2) Class-weighted loss function

For training the TTEDNN model, the class-weighted binary cross entropy (BCE) is used as the loss function $Loss$ with the L_2 regularization:

$$Loss = \sum_i [\alpha_1 y_i \log_2 p_i + \alpha_0 (1 - y_i) \log_2 (1 - p_i)] + \beta \sum \frac{1}{2}\left(\| w_k \|^2 + \| b_k \|^2 \right) \tag{12}$$

where y_i and p_i denote the label and the model output of the i^{th} sample, respectively; α_0 and α_1 denote the weight factors corresponding to the stable state and unstable state, respectively; w_k and b_k are the learnable network parameters; and β is the regularization weight. Class-weighted BCE is proven significantly helpful for the training dataset with the great imbalance. In the training dataset for RAS prediction, there are fewer samples concerning unstable states. The imbalance of the dataset results from the fact that practical power systems are stable in most of the time under common disturbances (see the disturbance discussed in Section II-C).

IV. Case Study

In this section, the IEEE 39-bus and IEEE 300-bus power systems are used to test the performance, scalability, effect of PMU data cycles, and robustness against measurement noise of the TTEDNN model. Furthermore, the transfer learning ability of the TTEDNN model trained on the small-signal RAS dataset to predict the transient RAS is discussed.

A. Training Setup

The specific parameters of IEEE 39-bus and IEEE 300-bus power systems for the evaluation and scalability validation of the TTEDNN model are derived from the PST toolbox [35] and Matpower 6.0 toolbox [36]. Following the disturbance discussed in Section II-C, disturbances on the initial states of rotor angle and angular speed $[\delta_i, \omega_i]$ are considered for the small-signal RAS. The training dataset contains the set under the single-node disturbance case ($m = 1$), while the test dataset consists of the set under both the single-node and multiple-node disturbance cases ($m > 1$).

Given a power system with N nodes, procedures for generating the dataset under the single-node disturbance case are described as follows.

1) Solve the power flow and let the solution be the undisturbed initial state.

2) The undisturbed initial state for each node $i = 1, 2, \ldots, N$ is randomly disturbed $\mathcal{K}_i$ times individually according to the distribution of frequency fluctuations.

3) For each disturbed initial state, conduct TDS and use the resulting trajectories to label its TSI.

For each sample in the dataset under the multiple-node disturbance case, m ($m > 1$) different nodes are simultaneously disturbed. The corresponding data generation processes are as follows.

1) Solve the power flow and let the solution be the undisturbed initial state.

2) Randomly select M groups of nodes, and each group includes m nodes.

3) Within each group of nodes, the undisturbed initial states of m nodes are randomly disturbed $\mathcal{K}_m$ times simultaneously according to the distribution of frequency fluctuations.

4) For each group of disturbed initial states, conduct TDS and use the resulting trajectories to label its TSI.

For the single-node disturbance dataset of the IEEE 39-bus power system, given $\mathcal{K}_i = 1000$, 39000 samples in total are generated with 33004 samples of stable states and 5996 samples of unstable states. For the single-node disturbance dataset of the IEEE 300-bus system, given $\mathcal{K}_i = 441$, 52038 samples in total are generated with 48186 samples of stable states and 3852 samples of unstable states. For the multiple-node disturbance dataset of the IEEE 39-bus power system, given $m=3$, $M=60$, and $\mathcal{K}_m=200$, 12000 samples in total are generated with 7377 samples of stable states and 4623 samples of unstable states. For the multiple-node disturbance dataset of the IEEE 300-bus power system, given $m=3$, $M=60$, and $\mathcal{K}_m=200$, 12000 samples in total are generated with 11132 samples of stable states and 868 samples of unstable states. The single-node disturbance dataset is used for the training of the TTEDNN model, and 60%, 20%, and 20% of the dataset are used for training, validation, and testing, respectively. The model trained with the single-node disturbance dataset is directly used for predicting RAS under multi-node disturbance. Therefore, 100% of the multi-node disturbance dataset is used for testing. We have two test datasets, one for predicting the RAS under single-node disturbance, and the other for predicting the RAS under multi-node disturbance.

The dataset of $N-1$ contingencies for the transient RAS prediction is generated as follows.

1) Randomly change all loads from 80% to 120% at the basic load levels.

2) Solve power flow and let the solution be undisturbed initial state.

3) Conduct the TDS based on the undisturbed initial state, trigger a three-phase short-circuit fault on a randomly selected transmission line, and clear the fault after 0.1 s.

4) Label the TSI with the post-fault state.

Consequently, for the IEEE 39-bus system, 28328 samples are generated with 20986 samples of stable states and 7342 samples of unstable states. For the IEEE 300-bus system, 30850 samples are generated with 21808 samples of stable states and 9042 samples of unstable states.

The confusion matrix is helpful for the evaluation of the prediction model, which defines four values based on actual and predicted results, i.e., *TP*, *FP*, *TN*, and *FN*, where *TP* (*TN*) is the extent to which the model correctly predicts the positive (negative) class, and *FP* (*FN*) is the extent to which the model wrongly predicts the negative (positive) class. In this paper, the stable/positive and unstable/negative are interchangeable. Four metrics including accuracy *ACC*, false positive rate *FPR*, false negative rate *FNR*, and F-score F_{score} are used to measure the performance of the TTEDNN model.

$$ACC=\frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

$$FPR=\frac{FP}{FP+TN} \tag{14}$$

$$FNR=\frac{FN}{FN+TP} \tag{15}$$

$$F_{score}=(1+\gamma^2)\frac{P_{recision}R_{ecall}}{\gamma^2 P_{recision}+R_{ecall}} \tag{16}$$

where $P_{recision}=TP/(TP+FP)$ denotes the fraction of *TP* among the models classified as positive class; and $R_{ecall}=TP/(TP+FN)$ denotes the fraction of *TP* among the total number of positive samples. While *ACC*, *FPR*, and *FNR* can reveal whether the predictions are good or not, F_{score} could evaluate the prediction of the model of imbalanced samples more comprehensively for it indicates how much more important recall is than precision or vice-versa. We set $\gamma=1$ in this paper.

The TTEDNN model is based on Tensorflow 2.3.1 and deployed on a server with Intel Xeon CPU E5-2620 v3. Two groups of GC modules ($n=2F$) with the kernel sizes of 16 and 8 are used to extract topological features from PMU data input. The TC module has five RBs ($R=5$), where the exponential dilated factors $d=2^{r-1}$ for $r=1,2,\dots,R$, the kernel size k is 2, and the number of filters is 32, respectively, for each residual block. The MLP prediction layer has the dimensions of (16, 1) and (32, 1) for the input layer and the hidden layer, respectively. The learning rate and batch size for the training are set to be 10^{-3} and 128. L_2 regularization weight β is set to be 5×10^{-4}. Weight factor α_0 is set to be 1, and α_1 is calculated on each batch as:

$$\alpha_1=\begin{cases} 256\Big/\sum_{i=1}^{256}y_i-1 & \sum_{i=1}^{256}y_i\neq 0 \\ 0 & \sum_{i=1}^{256}y_i=0 \end{cases} \tag{17}$$

B. Small-signal RAS Prediction

For small-signal RAS prediction, Fig. 2 shows the validation performance of the trained TTEDNN model in terms of *ACC* and class-weighted loss at different training epochs in both the IEEE 39-bus and IEEE 300-bus power systems. It can be found that *ACC* increases sharply to 98% within 20 epochs, and the training of the TTEDNN model converges quickly and smoothly after nearly 150 epochs.

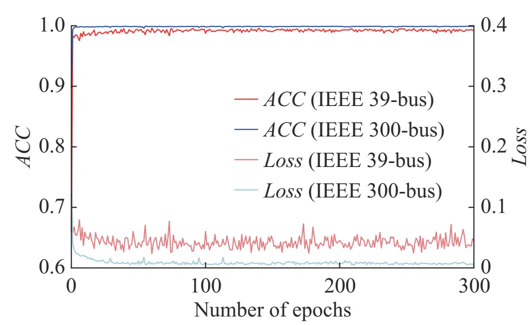

Fig. 2. Validation performance of trained TTEDNN model in terms of *ACC* and *Loss* at different training epochs for small-signal RAS prediction in both IEEE 39-bus and IEEE 300-bus power systems.

Table I shows the performance metrics for small-signal RAS prediction under the single-node disturbance dataset in the IEEE 39-bus and IEEE 300-bus power systems.

Six existing models including support vector machine (SVM), MLP, CNN [13], LSTM, GCN [19], and RGCN [27] are used to compare the performance metrics with the proposed TTEDNN model. It can be observed from Table I that the TTEDNN model outperforms the compared deep learning models under almost all performance metrics. Specifically, the TTEDNN model has the best performance in terms of *ACC* of 99.63% and F_{score} of 0.9965 for the IEEE 39-bus power system, and *ACC* of 99.88% and F_{score} of 0.9989 for the IEEE 300-bus power system.

TABLE I
PERFORMANCE METRICS FOR SMALL-SIGNAL RAS PREDICTION UNDER SINGLE-NODE DISTURBANCE DATASET IN IEEE 39-BUS AND IEEE 300-BUS POWER SYSTEMS

Model	IEEE 39-bus system				IEEE 300-bus system			
	ACC (%)	*FNR* (%)	*FPR* (%)	F_{score}	*ACC* (%)	*FNR* (%)	*FPR* (%)	F_{score}
SVM	84.27	10.36	12.17	0.8921	87.43	8.98	13.24	0.9135
MLP	98.45	0.97	7.24	0.9930	99.69	0.14	6.64	0.9845
CNN	98.36	0.86	9.33	0.9907	99.62	0.23	5.78	0.9928
LSTM	96.19	3.33	18.53	0.9558	99.25	0.18	2.15	0.9978
GCN	96.19	3.33	18.53	0.9558	99.25	0.18	2.15	0.9978
RGCN	98.15	2.20	8.91	0.9852	99.53	0.18	6.14	0.9921
Proposed	99.63	0.29	0.47	0.9965	99.88	0.17	0.00	0.9989

The correct prediction of unstable states is critically important in the practical implementation, which can be reflected by the *FPR*, the proportion of the fault prediction in all unstable samples. The TTEDNN model has the best *FPR* of only 0.47%, i.e., among all the unstable samples, only six samples are mistakenly predicted to be stable. The MLP has the best *FPR* of 0.14% under the single-node test dataset of the IEEE 300-bus power system, slightly better than that of the TTEDNN model with 0.17%.

The performance metrics for small-signal RAS prediction under the multiple-node disturbance dataset in the IEEE 39-bus and IEEE 300-bus power systems are also investigated, since the multiple-node disturbances are more likely to happen in reality and make the prediction task more complicated. As shown in Table II, the TTEDNN model has the best performance on predicting the small-signal RAS under multiple-node disturbances, i. e., *ACC* of 98.60% and F_{score} of 0.9862 for the IEEE 39-bus power system and *ACC* of 97.80% and F_{score} of 0.9785 for the IEEE 300-bus power system. It is worth noting that the compared existing models show a 3%-18% drop in terms of *ACC* and F_{score} when the condition changes from single-node disturbances to multiple-node disturbances, while the proposed TTEDNN model only has very small changes. Hence, the TTEDNN model is more robust than the existing models for the scenario where the system is subjected to multiple-node disturbances.

TABLE II
PERFORMANCE METRICS FOR SMALL-SIGNAL RAS PREDICTION UNDER MULTIPLE-NODE DISTURBANCE DATASET IN IEEE 39-BUS AND IEEE 300-BUS POWER SYSTEMS

Method	IEEE 39-bus system				IEEE 300-bus system			
	ACC (%)	*FNR* (%)	*FPR* (%)	F_{score}	*ACC* (%)	*FNR* (%)	*FPR* (%)	F_{score}
SVM	81.27	15.56	18.29	0.8130	86.96	10.92	16.43	0.8695
MLP	82.49	16.16	21.27	0.8250	90.46	2.44	20.86	0.9048
CNN	80.73	11.93	39.68	0.8072	95.29	0.41	11.57	0.9530
LSTM	82.49	16.16	21.27	0.8251	90.71	2.37	20.32	0.9072
GCN	93.21	2.31	10.45	0.9321	96.78	1.90	4.86	0.9710
RGCN	90.26	11.66	4.41	0.8999	97.36	0.52	10.34	0.9742
Proposed	98.60	0.98	2.56	0.9862	97.80	0.68	5.95	0.9785

C. Transient RAS Prediction

The TTEDNN model is also trained for transient RAS prediction, the dataset of which is generated under disturbances of $N-1$ contingencies. The validation performance of the trained TTEDNN model in terms of *ACC* and *Loss* for the transient RAS at different training epochs in both the IEEE 39-bus and IEEE 300-bus power systems are shown in Fig. 3. Both the *ACC* and *Loss* converge successfully after 200 epochs. The *ACC* increases to approximately 98% after 100 epochs.

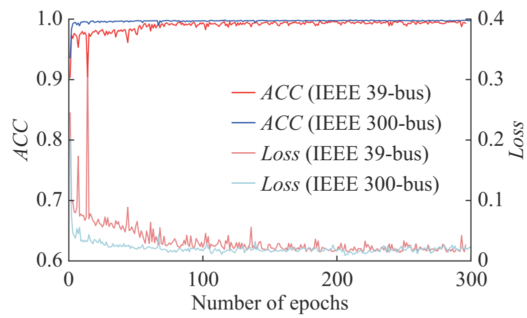

Fig. 3. Validation performance of trained TTEDNN model in terms of *ACC* and *Loss* for transient RAS at different training epochs in both IEEE 39-bus and IEEE 300-bus power systems.

The same six existing models shown in Table I are used to compare the performance metrics with the proposed TTEDNN model under disturbances of $N-1$ contingencies in the IEEE 39-bus and IEEE 300-bus power systems, as shown in Table III and Table IV, respectively.

TABLE III
PERFORMANCE METRICS FOR TRANSIENT RAS PREDICTION IN IEEE 39-BUS POWER SYSTEM

Model	*ACC* (%)	*FNR* (%)	*FPR* (%)	F_{score}
SVM	96.21	2.12	3.21	0.9624
MLP	99.15	0.41	1.32	0.9919
CNN	98.87	0.75	1.54	0.9892
LSTM	99.15	0.58	1.14	0.9919
GCN	98.20	1.19	2.46	0.9828
RGCN	99.28	0.58	0.88	0.9930
Proposed	99.63	0.34	0.40	0.9964

TABLE IV
PERFORMANCE METRICS FOR TRANSIENT RAS PREDICTION IN IEEE 300-BUS POWER SYSTEM

Model	*ACC* (%)	*FNR* (%)	*FPR* (%)	F_{score}
SVM	97.24	2.04	4.48	0.9725
MLP	98.98	0.80	1.45	0.9918
CNN	98.33	1.44	2.21	0.9835
LSTM	99.19	0.46	1.66	0.9920
GCN	98.75	1.12	1.55	0.9876
RGCN	99.42	0.34	1.16	0.9941
Proposed	99.72	0.23	0.39	0.9973

It can be observed that the proposed TTEDNN model outperforms all compared existing models for each performance metric. Specifically, the proposed TTEDNN model obtains *ACC* of 99.63% and F_{score} of 0.9964 for the IEEE 39-bus power system, and *ACC* of 99.72% and F_{score} of 0.9973 for the IEEE 300-bus power system.

Meanwhile, the advanced prediction performances in IEEE 300-bus power system under both the small-signal RAS and transient RAS also demonstrate the scalability of the proposed TTEDNN model to apply relatively large power systems. The time for predicting the RAS is also evaluated, which is important for fast online implement. Based on the Intel Xeon CPU E5-2620 v3, it takes approximately 5 ms for the trained TTEDNN model to predict both the small-signal RAS and transient RAS per batch, which is much faster than the traditional TDS.

D. Effect of PMU Data Cycles

The observation window length of post-fault PMU data affects the *ACC* and computational training time of the proposed TTEDNN model. Longer observation window length provides more information about the system dynamics that can increase the prediction performance, as shown in Table V, while longer computational training time is required. The observation window length is also called response time and is often measured by the unit of cycles [37]. With different cycles, the trade-off between *ACC* and computational training time per batch for the proposed TTEDNN model trained in the IEEE 39-bus power system is illustrated in Fig. 4. *ACC* of both the small-signal RAS and transient RAS scenarios increases to the maximum at 5 cycles while the computational training time monotonically increases. Thus, 5 cycles are chosen to be the optimal length of observation window length for the TTEDNN model, i. e., only first 5 cycles of post-fault PMU data are needed to achieve the highest *ACC* with the shortest computational training time. The existing work [38] shows the cycles of PMU data observed are longer versus the average response time around 1.5 cycles by time-adaptive methods. Nevertheless, 5 cycles are acceptable for the RAS prediction task for the following reasons. The control actions will not be executed until a waiting time of 0.15 s to 0.4 s is reached after the fault is cleared, which is still much longer than the first 5 cycles of the post-fault PMU data. Additionally, *ACC* and *FPR* indicate that the TTEDNN model has superior prediction performance and is more robust in the unstable sample prediction for the control action than the time-adaptive method.

TABLE V
TRANSIENT RAS PREDICTION WITH DIFFERENT CYCLES OF POST-FAULT PMU DATA IN IEEE 39-BUS POWER SYSTEM

Number of cycles	*ACC* (%)	*FNR* (%)	*FPR* (%)	F_{score}
2	91.58	8.38	8.46	0.9187
4	95.43	3.28	5.34	0.9548
6	99.60	0.37	0.44	0.9961
8	99.68	0.31	0.33	0.9969
10	99.63	0.34	0.40	0.9964

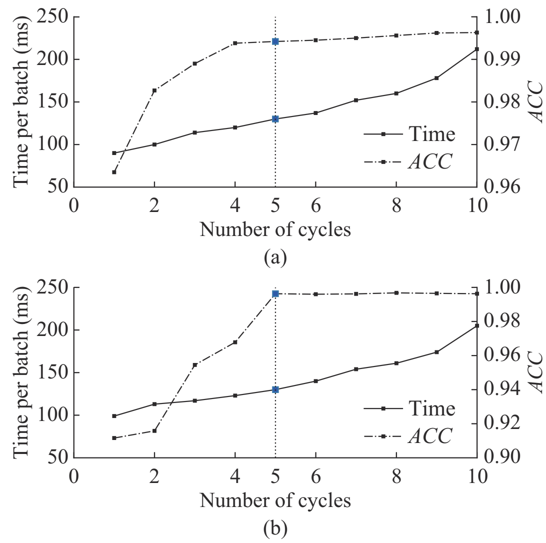

Fig. 4. *ACC* and computational training time per batch with different cycles for proposed TTEDNN model trained in IEEE 39-bus power system. (a) Small-signal RAS prediction. (b) Transient RAS prediction.

E. Effect of Grid-informed Adjacency Matrices

In (9)-(11), we introduce three grid-informed adjacency matrices incorporating the structural and electrical parameter information of the power grid. The grid-informed adjacency matrices of the IEEE 39-bus power system are visualized in Fig. 5, where the colors of each small pixel blocks represent the corresponding element value of the four adjacency matrices. The corresponding prediction performances are shown in Fig. 6. It can be observed from Fig. 6 that the adjacency matrix $\boldsymbol{B}^2$ shows the worst performance. This can be explained by the very sparse adjacency matrix visualization in Fig. 5, which indicates that although $\boldsymbol{B}^2$ contains the information of active power flow distribution, other useful information about the power system topology and electrical properties is discarded. The performance measurements of $\boldsymbol{B}^1$ and $\boldsymbol{B}^3$ are almost the same, while $\boldsymbol{B}^3$ is slightly better on true positive rate *TPR*. Hence, the grid-informed adjacency matrix $\boldsymbol{B}^3$ is used for the TTEDNN model to predict the small-signal RAS.

As for the transient RAS prediction task, when a disturbance of $N-1$ contingency happens between node i and node j, the power grid topology is changed, i.e., the transmission line (i,j) is removed during the short circuit. To this end, we revise $\boldsymbol{B}^3$ by letting B_{ij}^3 and B_{ji}^3 be zero. The revised matrix is denoted as $\hat{\boldsymbol{B}}^3$ and used for the TTEDNN model to predict the transient RAS. The performance comparison of transient RAS prediction by using different grid-informed adjacency matrices is shown in Table VI.

F. Robustness Against Noise

For RAS prediction in practical power systems, the noise in PMU data is of great concern to the performance of a prediction method [39]. The noise in PMU data has a standard deviation ranging from 0.0005 to 0.01 [40], resulting a typical signal-to-noise rate (SNR) of 45 dB. Table VII exhibits the performance on both the small-signal and transient RAS predictions in IEEE 39-bus power system under different SNR levels of PMU data.

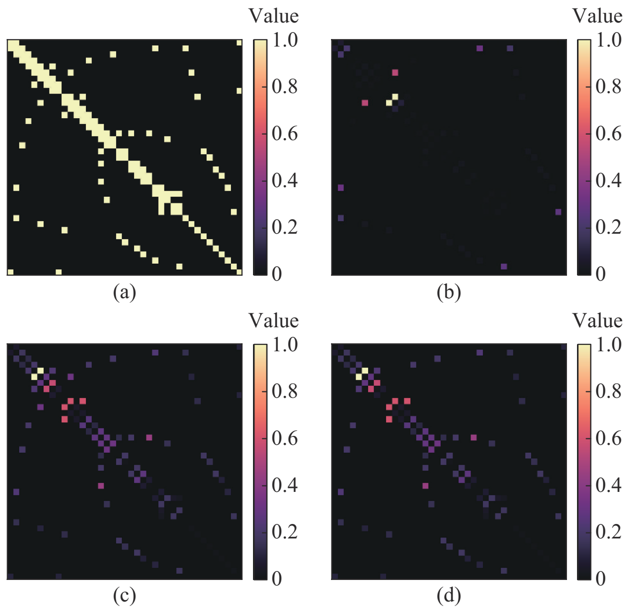

Fig. 5. Visualization of grid-informed adjacency matrices. (a) $\boldsymbol{B}^1$. (b) $\boldsymbol{B}^2$. (c) $\boldsymbol{B}^3$. (d) $\hat{\boldsymbol{B}}^3$.

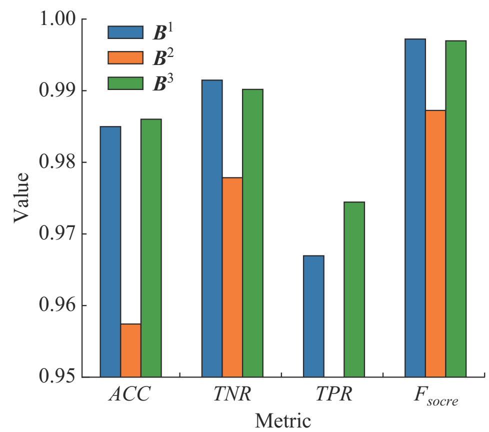

Fig. 6. Performance comparison of grid-informed adjacency matrices.

TABLE VI
PERFORMANCE COMPARISON OF TRANSIENT RAS PREDICTION BY USING DIFFERENT GRID-INFORMED ADJACENCY MATRICES

Matrix	*ACC* (%)	*FNR* (%)	*FPR* (%)	F_{score}
$\boldsymbol{B}^3$	99.59	0.31	0.51	0.9961
$\hat{\boldsymbol{B}}^3$	99.63	0.34	0.40	0.9964

The best performance is realized in the ideal environment without noise. When SNR reduces to 40 dB (lower than the typical SNR), the performance still maintains at a high level, i.e., only 0.1% and 0.25% decreases of *ACC* for the small-signal and transient RAS predictions, respectively. For strong noise levels with SNR of only 20 dB, the prediction performance degrades slightly, i.e., 0.78% and 1.20% drops of *ACC* for the small-signal and transient RAS predictions, respectively. Besides the *ACC*, other performance metrics also demonstrate only slight degrades with the decreasing SNR. Hence, the TTEDNN model is robust against the noise in PMU data.

TABLE VII
PERFORMANCE ON BOTH SMALL-SIGNAL AND TRANSIENT RAS PREDICTIONS IN IEEE 39-BUS POWER SYSTEM UNDER DIFFERENT SNR LEVELS OF PMU DATA

RAS prediction	SNR (dB)	*ACC* (%)	*FNR* (%)	*FPR* (%)	F_{score}
Small-signal	No	99.63	0.29	0.47	0.9965
	60	99.65	0.29	0.41	0.9968
	50	99.59	0.33	0.50	0.9962
	40	99.53	0.38	0.58	0.9926
	30	99.31	0.60	0.80	0.9916
	20	98.85	0.91	1.44	0.9893
Transient	No	99.63	0.34	0.40	0.9964
	60	99.61	0.34	0.44	0.9963
	50	99.58	0.37	0.51	0.9958
	40	99.38	0.54	0.70	0.9941
	30	99.10	1.02	0.77	0.9913
	20	98.43	1.42	1.72	0.9849

The performances of the proposed TTEDNN model and the RGCN model are compared under different noise levels in terms of the SNR. Six existing models shown in Table III are used to compare the performance of transient RAS prediction, and the RGCN model has the best performance among them. Figure 7 shows the comparison of *ACC* between the proposed TTEDNN model and the RGCN model under different SNR levels. It can be observed that as the SNR decreases, *ACC* of the TTEDNN model decreases slower than that of the RGCN model. Specifically, when SNR is 20 dB, *ACC* of the RGCN model decreases by 2.31%, which is twice as much as 1.20% of the TTEDNN model.

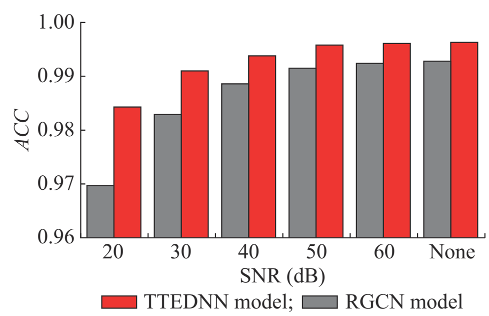

Fig. 7. Comparison of *ACC* between TTEDNN model and RGCN model under different SNR levels.

G. Transfer Learning Ability

The transfer learning ability of the proposed TTEDNN model trained on the small-signal RAS dataset to predict the transient RAS is worthful to be investigated. Usually, small disturbances happen more commonly in real power systems than serve $N-1$ contingencies. Hence, the dataset for small-signal RAS is easier to be collected. The small-signal RAS dataset can provide certain information on stable and unstable patterns for the transient RAS prediction task. Learning based on the small-signal RAS dataset can be useful for the few-shot learning of transient RAS prediction.

To investigate the transfer learning ability of pre-trained TTEDNN on the small-signal RAS dataset, three re-training tests are introduced and compared. ① Training from scratch (TFS): the whole network parameters are updated without pre-trained initialization. ② Full fine-tuning (FFT): the whole network parameters are updated with pre-trained initialization. ③ Local fine-tuning (LFT): only the layers close to the output are updated with pre-trained initialization.

The TFS test updates all the parameters of GC modules, TC modules, and MLP layer of the TTEDNN model with a random weight initializer. The FFT test updates all the parameters of the TTEDNN model with the pre-trained model with the small-signal RAS dataset. For the LFT test, part of the GC modules are frozen to keep the ability of topological feature extraction and update the parameters of the TC module and MLP layer. The performance comparison of the three re-training tests is given as follows.

Figure 8 shows the validation performance in terms of *Loss* and *ACC* during the training process of three re-training tests. The transfer learning with pre-trained initialization is proved to be effective, i.e., the *Loss* values of FFT and LFT smoothly converge to 1.5 times smaller values than those of the TFS. Moreover, it can be observed from Fig. 8(b) that the FFT and LFT enable faster early-stop with a given acceptable performance so that the training cost is reduced, i.e., only 30 to 70 epochs are needed for FFT and LFT to reach 99.5% of *ACC* while more than 200 epochs are needed for the TFS to reach the same *ACC*.

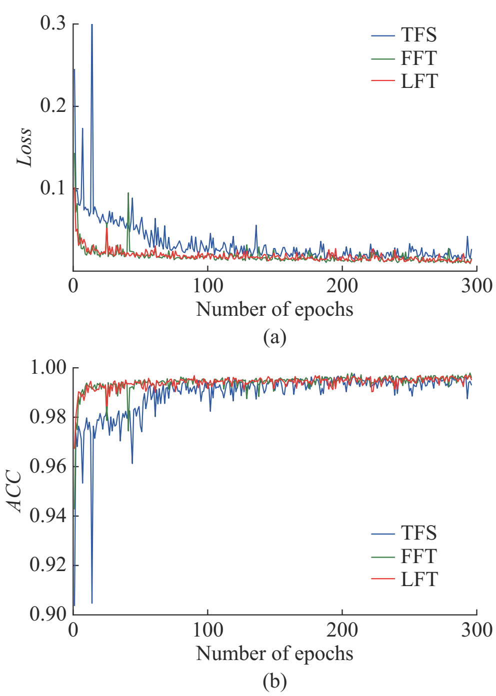

Fig. 8. Validation performance in terms of *Loss* and *ACC* for three re-training tests. (a) *Loss*. (b) *ACC*.

Table VIII shows the performance metrics for three re-training tests in the IEEE 39-bus power system. It is worth noticing that the performance metrics of transient RAS in the model pre-trained with the small-signal RAS dataset is even better than those in the model directly trained with the transient RAS dataset, which indicates that small-signal RAS dataset provides useful information for the transient RAS prediction.

TABLE VIII
PERFORMANCE METRICS FOR THREE RE-TRAINING TESTS IN IEEE 39-BUS POWER SYSTEM

Test	*ACC* (%)	*FNR* (%)	*FPR* (%)	F_{score}
TFS	99.63	0.34	0.40	0.9964
FFT	99.68	0.27	0.37	0.9969
LFT	99.77	0.20	0.26	0.9978

The time consumption of the three re-training tests in the IEEE 39-bus power system is listed in Table IX. We can notice that with the same data generation time, the FFT reduces the training time of 5042 s compared with TFS, and the LFT further reduces about 350 s compared with FFT due to part of the layers do not need to be updated.

TABLE IX
COMPARISON OF TIME CONSUMPTION OF THREE RE-TRAINING TESTS IN IEEE 39-BUS POWER SYSTEM

Test	Data generation time (s)	Training time (s)	Testing time per batch (ms)
TFS	6170	6593	4.08
FFT	6170	1551	4.08
LFT	6170	1227	4.08

To explain the mechanism of LFT for transfer learning more intuitively, the outputs of the hidden layer in the TTEDNN model are visualized with the t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction technique [41], which is shown in Fig. 9. The green dots are the samples of stable states correctly predicted, the black dots are the samples of unstable states correctly predicted, and the red dots are the samples incorrectly predicted. At epoch 0 (before training), more samples are predicted correctly with LFT. This is because for LFT, the TTEDNN model carries the information from the pre-trained dataset, while network parameters of the TTEDNN model are randomly initialized for TFS without further useful information. As the training epochs increase, fewer prediction errors and more clear classification boundaries are shown in Fig. 9(e)-(h), which demonstrates fast converge speed and better prediction ability for the LFT.

V. CONCLUSION

We proposed the TTEDNN model for small-signal and transient RAS predictions in power systems. The TTEDNN model maps the spatial information of power system topology into the GC modules as well as extracts the temporal features from the PMU data with TC modules. The TTEDNN model has the following advantages.

First, it shows the best prediction performance compared with the existing deep learning models under both small disturbances and $N-1$ contingencies.

Second, it can make a fast prediction with only the PMU data of the first five post-disturbed cycles, demonstrating its potential for online implementation.

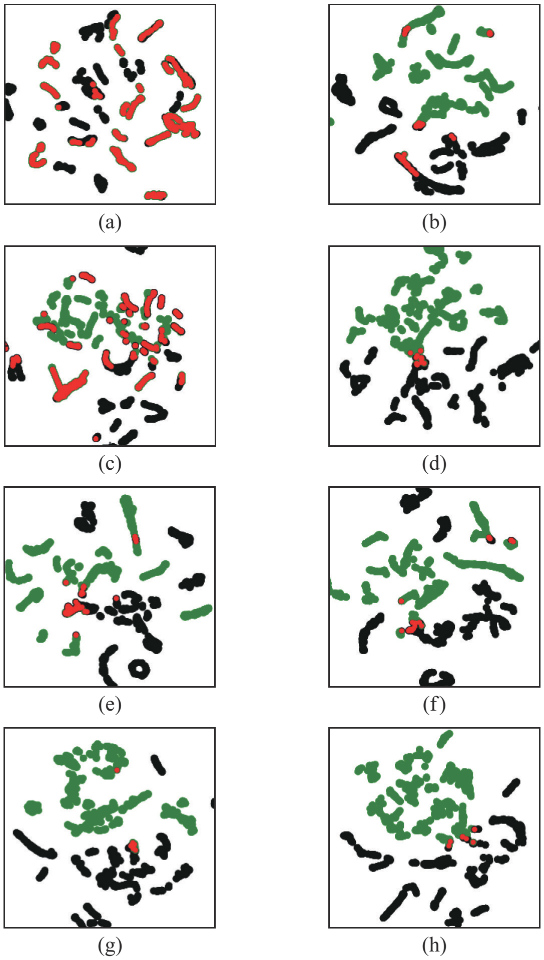

Fig. 9. Visualizations of high-dimensional activations from hidden layer in TTEDNN model. (a) Visualization of TFS at epoch 0. (b) Visualization of TFS at epoch 20. (c) Visualization of TFS at epoch 40. (d) Visualization of TFS at epoch 60. (e) Visualization of LFT at epoch 0. (f) Visualization of LFT at epoch 20. (g) Visualization of LFT at epoch 40. (h) Visualization of LFT at epoch 60.

Third, it is robust against the measurement noise of PMU data, which is necessary for practical applications.

Finally, it provides the superior transfer learning ability from small-signal RAS conditions to transient RAS conditions.

REFERENCES

[1] P. Sarajcev, A. Kunac, G. Petrovic *et al.*, "Power system transient stability assessment using stacked autoencoder and voting ensemble," *Energies*, vol. 14, no. 11, p. 3148, May 2021.

[2] J. Hou, C. Xie, T. Wang *et al.*, "Power system transient stability assessment based on voltage phasor and convolution neural network," in *Proceeding of 2018 IEEE International Conference on Energy Internet (ICEI)*, Beijing, China, May 2018, pp. 247-251.

[3] S. K. Azman, Y. J. Isbeih, M. S. E. Moursi *et al.*, "A unified online deep learning prediction model for small signal and transient stability," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4585-4598, Nov. 2020.

[4] P. Kundur, J. Paserba, V. Ajjarapu *et al.*, "Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1387-1401, Aug. 2004.

[5] S. Zhang, Z. Zhu, and Y. Li, "A critical review of data-driven transient stability assessment of power systems: principles, prospects and challenges," *Energies*, vol. 14, no. 21, p. 7238, Nov. 2021.

[6] Z. Liu, X. He, Z. Ding *et al.*, "A basin stability based metric for ranking the transient stability of generators," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1450-1459, Mar. 2019.

[7] B. Zhang, X. Jin, S. Tu *et al.*, "A new FPGA-based real-time digital solver for power system simulation," *Energies*, vol. 12, no. 24, p. 4666, Dec. 2019.

[8] T. L. Vu and K. Turitsyn, "Lyapunov functions family approach to transient stability assessment," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1269-1277, Mar. 2016.

[9] M. Anghel, F. Milano, and A. Papachristodoulou, "Algorithmic construction of Lyapunov functions for power system stability analysis," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 9, pp. 2533-2546, Sept. 2013.

[10] S. Wei, M. Yang, J. Qi *et al.*, "Model-free mle estimation for online rotor angle stability assessment with PMU data," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2463-2476, May 2018.

[11] R. Yan, G. Geng, Q. Jiang *et al.*, "Fast transient stability batch assessment using cascaded convolutional neural networks," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2802-2813, Jul. 2019.

[12] J. J. Q. Yu, D. J. Hill, A. Y. S. Lam *et al.*, "Intelligent time-adaptive transient stability assessment system," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1049-1058, Jan. 2018.

[13] M. Sahu and R. Dash, "A survey on deep learning: convolution neural network (CNN)," in *Intelligent and Cloud Computing*, Berlin: Springer, 2021, pp. 317-325.

[14] T. Su, Y. Liu, J. Zhao *et al.*, "Probabilistic stacked denoising autoencoder for power system transient stability prediction with wind farms," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3786-3789, Jul. 2021.

[15] T. Su, Y. Liu, J. Zhao *et al.*, "Deep belief network enabled surrogate modeling for fast preventive control of power system transient stability," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 315-326, Jan. 2022.

[16] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proceedings of International Conference on Learning Representations*, Toulon, France, Feb. 2017, pp. 1-9.

[17] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, Jul. 2018, pp. 3634-3640.

[18] Z. Wu, S. Pan, F. Chen *et al.*, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4-24, Jan. 2021.

[19] Y. Liu, N. Zhang, D. Wu *et al.*, "Searching for critical power system cascading failures with graph convolutional network," *IEEE Transactions on Control of Network Systems*, vol. 8, no. 3, pp. 1304-1313, Sept. 2021.

[20] S. Bai, J. Z. Kolter, and V. Koltun. (2018, Apr.). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. [Online]. Available: https://arxiv.org/abs/1803.01271

[21] A. van den Oord, S. Dieleman, H. Zen *et al.* (2016, Sept.). Wavenet: a generative model for raw audio. [Online]. Available: https://arxiv.org/abs/1609.03499

[22] X. Tang, H. Chen, W. Xiang *et al.*, "Short-term load forecasting using channel and temporal attention based temporal convolutional network," *Electric Power Systems Research*, vol. 205, p. 107761, Apr. 2022.

[23] R. Zhang, W. Yao, Z. Shi *et al.*, "A graph attention networks-based model to distinguish the transient rotor angle instability and short-term voltage instability in power systems," *International Journal of Electrical Power & Energy Systems*, vol. 137, p. 107783, May 2022.

[24] J. Huang, L. Guan, Y. Su *et al.*, "A topology adaptive high-speed transient stability assessment scheme based on multi-graph attention network with residual structure," *International Journal of Electrical Power & Energy Systems*, vol. 130, p. 106948, Sept. 2021.

[25] Y. Su, M. Guo, H. Yao *et al.*, "Power system small-signal stability assessment model based on residual graph convolutional networks," *Journal of Physics: Conference Series*, vol. 2095, p. 012011, Spet. 2021.

[26] J. Huang, L. Guan, Y. Chen *et al.*, "A deep learning scheme for transient stability assessment in power system with a hierarchical dynamic graph pooling method," *International Journal of Electrical Power & Energy Systems*, vol. 141, p. 108044, Oct. 2022.

[27] J. Huang, L. Guan, Y. Su *et al.*, "Recurrent graph convolutional network-based multi-task transient stability assessment framework in power system," *IEEE Access*, vol. 8, pp. 93283-93296, Apr. 2020.

[28] N. Hatziargyriou, J. Milanovic, C. Rahmann *et al.*, "Definition and

classification of power system stability – revisited & extended," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3271-3281, Jul. 2021.
[29] A. Gupta, G. Gurrala, and P. S. Sastry, "An online power system stability monitoring system using convolutional neural networks," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 864-872, Mar. 2019.
[30] B. Schäfer, C. Beck, K. Aihara *et al*., "Non-Gaussian power grid frequency fluctuations characterized by Lévy-stable laws and superstatistics," *Nature Energy*, vol. 3, pp. 119-126, Jan. 2018.
[31] J. Dong, J. Zuo, L. Wang *et al*., "Analysis of power system disturbances based on wide-area frequency measurements," in *Proceedings of 2007 IEEE PES General Meeting*, Tampa, USA, Jul. 2007, pp. 1-8.
[32] Q. Chen, *The Probability, Identification, and Prevention of Rare Events in Power Systems*. Ames: Iowa State University, 2004.
[33] J. Zhao, M. Netto, and L. Mili, "A robust iterated extended Kalman filter for power system dynamic state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3205-3216, Jul. 2017.
[34] E. Ghahremani and I. Kamwa, "Local and wide-area PMU-based decentralized dynamic state estimation in multi-machine power systems," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 547-562, Jan. 2016.
[35] J. Chow and K. Cheung, "A toolbox for power system dynamics and control engineering education and research," *IEEE Transactions on Power Systems*, vol. 7, no. 4, pp. 1559-1564, Nov. 1992.
[36] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12-19, Feb. 2011.
[37] R. Zhang, Y. Xu, Z. Y. Dong *et al*., "Post-disturbance transient stability assessment of power systems by a self-adaptive intelligent system," *IET Generation, Transmission & Distribution*, vol. 9, no. 3, pp. 296-305, Feb. 2015.
[38] H. Wang and S. Wu, "Transient stability assessment with time-adaptive method based on spatial distribution," *International Journal of Electrical Power & Energy Systems*, vol. 143, p. 108464, Dec. 2022.
[39] Z. Rafique, H. M. Khalid, S. Muyeen *et al*., "Bibliographic review on power system oscillations damping: an era of conventional grids and renewable energy integration," *International Journal of Electrical Power & Energy Systems*, vol. 136, p. 107556, Mar. 2022.
[40] T. Chen, H. Ren, Y. Sun *et al*., "Optimal placement of phasor measurement unit in smart grids considering multiple constraints," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 2, pp. 479-488, Mar. 2023.
[41] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579-2605, 2008.

Peiyuan Sun received his B.S. degree in electrical engineering and automation from the Xi'an Jiaotong University, Xi'an, China, in 2022. He is currently pursuing his M.S. degree in electrical engineering at the Xi'an Jiaotong University. His research interests include power system stability assessment and control and deep learning in power systems.

Long Huo received his B.Sc. degree in the College of Electrical and Information Engineering in Hunan University, Changsha, China, in 2015. He completed his M.Sc. degree in the School of Automation and Information Engineering at Xi'an University of Technology, Xi'an, China, in 2018. He is currently pursuing his Ph.D. degree in the School of Electrical Engineering at Xi'an Jiaotong University. His research interests include cascading failures and frequency synchronization stability in power grids.

Xin Chen received his B.Sc. degree from Nanjing University, Nanjing, China, in 1999, and his M.Sc. and Ph.D. degrees in chemical physics from the University of British Columbia, Vancouver, Canada, and Yale University, New Haven, USA, in 2003 and 2008, respectively. He worked as a Postdoc Fellow in the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, USA, from 2008 to 2014. He is currently working as an Associate Professor in the School of Electrical Engineering at Xi'an Jiaotong University, Xi'an, China. His research interests include artificial intelligence, feature engineering, deep reinforcement learning, stability analysis of complex dynamical systems, cascading failure in cyber-physical power systems, and power system dynamics and control.

Siyuan Liang received his B.E. degree in electrical engineering & automation from Xi'an Jiaotong University, Xi'an, China, in 2022. He is currently pursuing his Ph.D. degree in computer science and engineering at the Chinese University of Hong Kong, Hong Kong, China. His research interests include machine learning-based control solutions for power system dynamics, integrated circuit (IC) packaging technologies, and electronic design automation algorithms.