Electrical Power and Energy Systems 127 (2021) 106647

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

A short-term voltage stability online prediction method based on graph convolutional networks and long short-term memory networks

Guoteng Wang [a], Zheren Zhang [a,*], Zhipeng Bian [b], Zheng Xu [a]

[a] Zhejiang University, #38 Zheda Road, Hangzhou, Zhejiang 310027, China
[b] HUAWEI Technologies CO., LTD., #410 Jianghong Road, Hangzhou, Zhejiang 310052, China

ARTICLE INFO

Keywords:
Short-term voltage stability
Online prediction
Graph convolutional networks
Long short-term memory networks

ABSTRACT

Due to complex dynamic characteristics and large scale of power systems, it is a great challenge to predict short-term voltage stability (STVS) online. To address this challenge, a STVS online prediction method based on graph convolutional networks (GCN) and long short-term memory networks (LSTM) is proposed in this paper. Firstly, we propose a novel machine learning framework, GCN-LSTM, which is a combination of GCN and LSTM. Specifically, the GCN is used to capture spatial features of power grids, the LSTM is used to capture temporal features of power grids. Secondly, a STVS online prediction method based on the GCN-LSTM model is proposed. The proposed method can capture multiplex spatial–temporal STVS evolution trends and predict STVS results. Finally, case studies are carried out in two testing systems, including a modified 39-bus system and a 68-bus system. The training and testing data is generated by Power System Simulator / Engineering (PSS/E). Simulation results illustrate the high performance of the proposed method.

1. Introduction

Voltage collapse threatens the safe and stable operation of power systems [1]. With the increase of loads, the short-term voltage stability (STVS) problem will become more and more prominent. Therefore, it is an urgent problem to evaluate the STVS online quickly and accurately. The STVS online prediction methods are mainly divided into two categories: model-driven methods and data-driven methods. Most of the studies focusing on model-driven methods are developed when power systems are less complex, and the scale of power systems are relatively small. Since the scale and dimensions of the power system has been increasing in recent years, data-driven STVS prediction methods have shown considerable promise in their ability to outperform model-driven methods.

Existing classic works have focused mainly on model-driven methods [2]. The STVS online prediction methods based stability criteria have been widely used in industry practice [3]. In order to accurately assess the STVS, scholars have proposed various stability criteria. A practical short-term voltage stability index based on voltage curves is proposed in [4]. Various indexes of voltage sag severity have been proposed to evaluate the STVS [5,6]. In addition, the Lyapunov methods can also be applied to online monitoring of the STVS [7]. In [8], Lyapunov exponents are used to monitor the STVS, the Monte Carlo simulation and the particle swarm optimization method are used to adjust embedding parameters. In addition, with the development of computing tools, it is possible to evaluate the STVS online using real-time TDS techniques [9,10]. However, the real-time TDS technique relies on computing power. Although model-driven methods have been developed maturely, model-driven methods' capability of handling high dimensional power systems is quite limited.

In recent years, data-driven STVS prediction methods have received widespread attention from scholars [11]. Various machine learning models have been used in power systems for stability prediction. Authors in [12] put forward a model and data hybrid-driven method to monitor the STVS of power systems. Generative adversarial networks are used to assess the dynamic security of power systems [13]. Convolutional neural networks (CNN) are applied to estimate the transient stability of power systems in [14,15]. The data-driven methods based on support vector machine (SVM) are also proved to be effective in stability evaluation [16,17]. The workbook-based light duty time series learning machine is adopted in [18] to evaluate the STVS. In [19], a deep learning method is proposed to evaluate the severity of voltage sags. In addition, recurrent neural networks (RNN) such as long short-term memory networks (LSTM) and gated recurrent unit networks (GRU) also have good

* Corresponding author at: #38 Zheda Road, Hangzhou, Zhejiang 310027, China.
E-mail address: 3071001296zhang@zju.edu.cn (Z. Zhang).

https://doi.org/10.1016/j.ijepes.2020.106647
Received 1 May 2020; Received in revised form 19 October 2020; Accepted 13 November 2020
Available online 17 December 2020

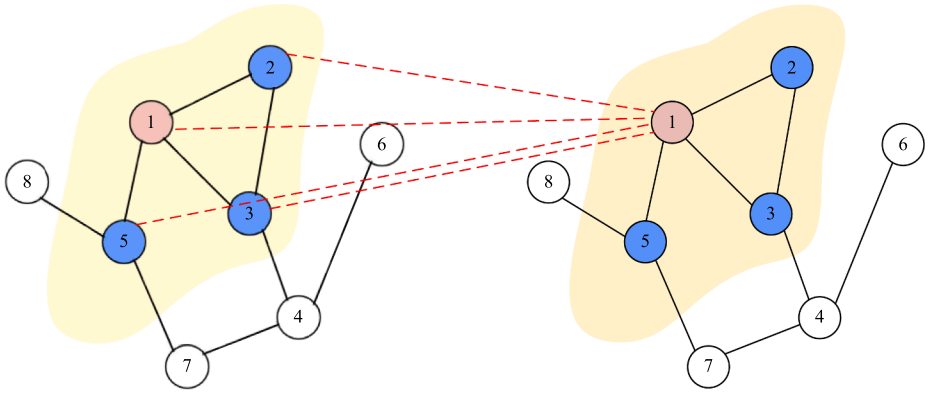

Fig. 1. Schematic diagram of graph convolution.

performance in stability evaluation [20,21]. If the STVS analysis is considered as a classification task, decision trees can be used for the STVS online prediction [22,23]. With the rapid development of data-driven methods, some machine learning models with superior performance are invented [24], and the performance of the STVS online prediction also needs to be further improved.

In this paper, we combine the GCN and the LSTM to evaluate the STVS. The contributions of this paper are summarized as follows:

(1) A novel machine learning model is proposed by combining the GCN and LSTM. Specifically, the GCN is used to capture spatial features of power grids, and the LSTM is used to capture temporal features of power grids.
(2) A STVS online prediction method based on the GCN-LSTM model is proposed. The proposed method is self-adaptive to the input time series length. By the proposed method, multiplex spatial–temporal STVS evolution trends are captured and STVS results are predicted.

The remainder of this paper is organized as follows: Section 2 briefly introduces the GCN-LSTM model. A new STVS online prediction method is proposed in Section 3. Section 4 presents the details of the model training. Two testing systems are employed in Section 5 to test the performance of the proposed method. Section 6 concludes this paper.

2. GCN-LSTM model

In this section, we first give a brief description to the GCN and LSTM, then present the GCN-LSTM model.

2.1. GCN

The power grid is a typical graph structure network, in which buses and branches are regarded as nodes and edges respectively. Therefore, a power grid can be processed into an undirected graph $\boldsymbol{g} = (\boldsymbol{\upsilon}, \boldsymbol{\varepsilon})$ with N nodes $\upsilon_{\mathrm{i}} \in \boldsymbol{\upsilon}$ and edges $(\upsilon_{\mathrm{i}}, \upsilon_{\mathrm{j}}) \in \boldsymbol{\varepsilon}$. To capture spatial features of graphs, a layer-wise linear GCN model is proposed in [24], and the layer-wise propagation rule of a multi-layer GCN is defined as (1).

$$\boldsymbol{K}^{(l+1)} = ReLU\left(\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{K}^{(l)}\boldsymbol{W}_G^{(l)}\right) \tag{1}$$

$$\boldsymbol{K}^{(0)} = \boldsymbol{X} \tag{2}$$

$$\boldsymbol{Z} = \boldsymbol{K}^{(L)} \tag{3}$$

Here, ReLU(·) is rectified linear unit, $\boldsymbol{K}^{(l)}$ is the activation matrix in the l^{th} layer ($l = 0, 1, \ldots, L$), L is the number of GCN layers, $\boldsymbol{W}_G^{(l)}$ is the layer-specific trainable weight matrix in the l^{th} layer, $\boldsymbol{X} \in \mathbb{R}^{N\times M}$ is the bus feature matrix, M is the feature number of each bus, $\boldsymbol{Z} \in \mathbb{R}^{N\times F}$ is the output matrix of the GCN, F is the number of hidden features, $\boldsymbol{A} = \boldsymbol{A} + \boldsymbol{I}_N$ is the adjacency matrix of the undirected graph $\boldsymbol{g}$ with added self-connections, $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ is the adjacency matrix, $\boldsymbol{I}_{\mathrm{N}}$ is the identity matrix, $\boldsymbol{D}$ is the degree matrix of the A, and $\boldsymbol{D}_{ii} = \sum_{j=1}^{N}\boldsymbol{A}_{ij}$.

The schematic diagram of the graph convolution is shown in Fig. 1. As shown in Fig. 1, assuming that bus 1 is the target node, the GCN model can learn the topological relationship between the target bus and its adjacent buses, and hidden features of bus 1 can be obtained considering the influence of adjacent buses.

For a power grid, the bus feature can be bus voltage magnitude, bus voltage phase, injected active power or reactive power. In a power grid, if bus i and bus j are connected directly, the elements $\boldsymbol{A}_{\mathrm{ij}}$ and $\boldsymbol{A}_{\mathrm{ji}}$ are one or weight coefficients. Otherwise, the two elements are zero. According to (1), the inputs of GCN include two matrices, one is the feature matrix $\boldsymbol{X}$ composed of features of all buses, and the other is the adjacency matrix $\boldsymbol{A}$. The output of GCN is a matrix $\boldsymbol{Z}$ composed of hidden features of all buses. In (1), $\boldsymbol{A}$ and $\boldsymbol{D}$ are obtained according to $\boldsymbol{A}$, and $\boldsymbol{W}_{\mathrm{G}}$ is the weight matrix obtained by training.

2.2. LSTM

Power systems are complex dynamic systems. The STVS of power systems is not only related to the current state, but also to the state in the past. The RNN is a class of neural networks with memory capabilities, which can effectively process time series data and make use of historical information. Due to the problem of gradient explosion or disappearance, simple RNNs can only learn short-term dependences. The transient process of power systems usually lasts for a few seconds, and a time step is only a few milliseconds in the STVS analysis. Although the transient duration of a power system is short, it involves hundreds of or even thousands of time steps. Therefore, a RNN that can learn the long-term dependences is necessary for the STVS online prediction. The LSTM is a variant of the RNN, which can effectively solve the gradient explosion or disappearance problem. The expressions of the LSTM are shown in (4)–(6). Refer to [25] for a detailed introduction of the LSTM.

$$\begin{bmatrix}\boldsymbol{c}_t\\ \boldsymbol{o}_{\mathrm{t}}\\ \boldsymbol{i}_{\mathrm{t}}\\ \boldsymbol{f}_{\mathrm{t}}\end{bmatrix} = \begin{bmatrix}tanh\\ \sigma\\ \sigma\\ \sigma\end{bmatrix}\left(\boldsymbol{W}_{\mathrm{L}}\begin{bmatrix}\boldsymbol{u}_{\mathrm{t}}\\ \boldsymbol{h}_{\mathrm{t-1}}\end{bmatrix} + \boldsymbol{B}\right) \tag{4}$$

$$\boldsymbol{c}_{\mathrm{t}} = \boldsymbol{f}_t \odot \boldsymbol{c}_{\mathrm{t-1}} + \boldsymbol{i}_{\mathrm{t}} \odot \boldsymbol{c}_{\mathrm{t}} \tag{5}$$

$$\boldsymbol{h}_{\mathrm{t}} = \boldsymbol{o}_{\mathrm{t}} \odot tanh(\boldsymbol{c}_{\mathrm{t}}) \tag{6}$$

Here, $\odot$ is the product of vector elements, $\boldsymbol{u}_{\mathrm{t}}$ is the input at moment t, $\boldsymbol{h}_{\mathrm{t}}$ is the external state at moment t, $\boldsymbol{c}_{\mathrm{t}}$ is the internal state at moment t, $\boldsymbol{c}_t$

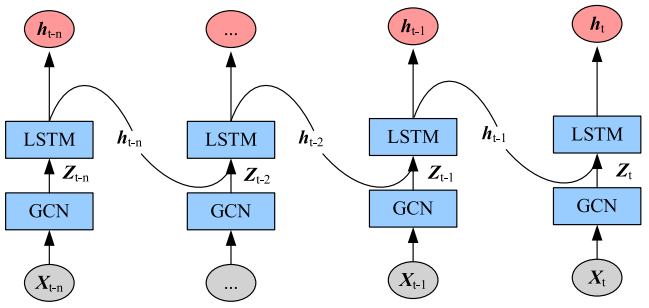

Fig. 2. Structure of the GCN-LSTM model.

is the candidate state at moment t, $\sigma(\cdot)$ is the Logistic function, $\boldsymbol{W}$ is the weight matrix, $\boldsymbol{B}$ is the bias vector, $\boldsymbol{f}_t$ is the forget gate which is used to discard part information of the internal state at the last moment $\boldsymbol{c}_{t-1}$, $\boldsymbol{i}_t$ is the input gate which is used to store part information of the candidate state at the current moment $\boldsymbol{c}_t$, $\boldsymbol{o}_t$ is the output gate which is used to output part information of the internal state at the current moment $\boldsymbol{c}_t$.

The input data and output result of the LSTM are matrix $\boldsymbol{u}_t$, and $\boldsymbol{h}_t$, repectively. In this paper, $\boldsymbol{u}_t$ is obtained from the output result of the GCN, $\boldsymbol{h}_t$ is composed of hidden features of all buses.

2.3. GCN-LSTM

The GCN and the LSTM can be combined to capture spatial–temporal features from operation data of power grids. In Fig. 2, $\boldsymbol{h}_t$ is the output of the GCN-LSTM model at time t, $\boldsymbol{X}_t \in \mathbb{R}^{N\times M}$ is the feature matrix at time t. In a power grid, the bus feature can be any electrical quantity such as voltage and injection power. As shown in Fig. 2, the bus features are first input into the GCN. The GCN captures spatial features of all buses. Then, the time series data composed of spatial features are input into the LSTM, and the hidden features carrying information about the STVS is captured. The same parameters are shared among all GCN-LSTM networks. A STVS online prediction method based on the GCN-LSTM model will be proposed in the following section.

3. STVS on-line prediction method

The STVS online prediction can be re-stated as a classification problem. In this section, we will first introduce the STVS online prediction task, and then describe how to use the GCN-LSTM model to realize the STVS online prediction task.

3.1. Formulation of the STVS on-line prediction task

Short-term voltage instability (STVI) is mainly caused by induction motors. Due to disturbance or the weakened transmission system, the induction motor may be stalled, which will lead to the STVI. When the power system faces the risk of the STVI, control measures such as load shedding need to be taken. In order to formulate control strategies, in addition to evaluating the system stability, locating the unstable buses is also necessary. In this paper, the purpose of the STVS on-line prediction is to predict and locate unstable buses based on operation data.

In general, the power system is regarded as a dynamic system, and its mathematical model is described by a set of differential algebraic equations. In this paper, the power system is described as an undirected graph with dynamic changes $\boldsymbol{g} = (\boldsymbol{v}, \boldsymbol{\varepsilon})$. The adjacency matrix $\boldsymbol{A}$ is used to represent the connection relationship among buses, $\boldsymbol{A} \in \mathbb{R}^{N\times N}$. At moment t, all bus features constitute a feature matrix $\boldsymbol{X}_t$, $\boldsymbol{X}_t \in \mathbb{R}^{N\times M}$. We regard the STVS online prediction task as a classification problem. Therefore, the process of STVS prediction can be considered as learning the mapping function f and classifying each bus, as shown in (7).

$$\boldsymbol{Y} = f(\boldsymbol{A}; (\boldsymbol{X}_{t-n}, \cdots, \boldsymbol{X}_{t-1}, \boldsymbol{X}_t)) \tag{7}$$

where $\boldsymbol{Y}$ is the label vector used to indicate whether each bus is stable, $\boldsymbol{Y} = \{y_1, y_2, \ldots, y_N\}$, y_i is an integer.

3.2. STVS on-line prediction method

The STVS online prediction task is different from offline evaluation, the start moment and duration of the disturbance are unpredictable. Therefore, a machine learning model with fixed input length is not suitable for the online prediction task. In addition, a model with a large amount of computation is also not suitable for the online prediction task. In order to meet the requirements of online application, we propose a novel STVS online prediction method, as shown in Fig. 3, where t is the current moment when monitoring. The proposed method is self-adaptive to the input time series length, and the training can be completed offline. When applying the method online, the computation amount is very small. The proposed method includes six steps, described below.

The first step is a multi-layer GCN. The input data of the GCN is obtained by the SCADA/WAMS system. The input data includes an N by N adjacency matrix $\boldsymbol{A}$ (N is the number of buses), an N by M feature matrix $\boldsymbol{X}_t$ (M is the number of features of each bus). The output of the GCN is an N by F activation matrix $\boldsymbol{Z}_t$ (F is the number of hidden features per bus). The GCN is used to capture the spatial features of buses.

The second step is a LSTM network. The spatial feature matrix $\boldsymbol{Z}_t$ at moment t, the internal state matrix $\boldsymbol{c}_{t-1}$ and the external state matrix $\boldsymbol{h}_{t-1}$ at moment t-1 are used as the input of the LSTM network. The output of

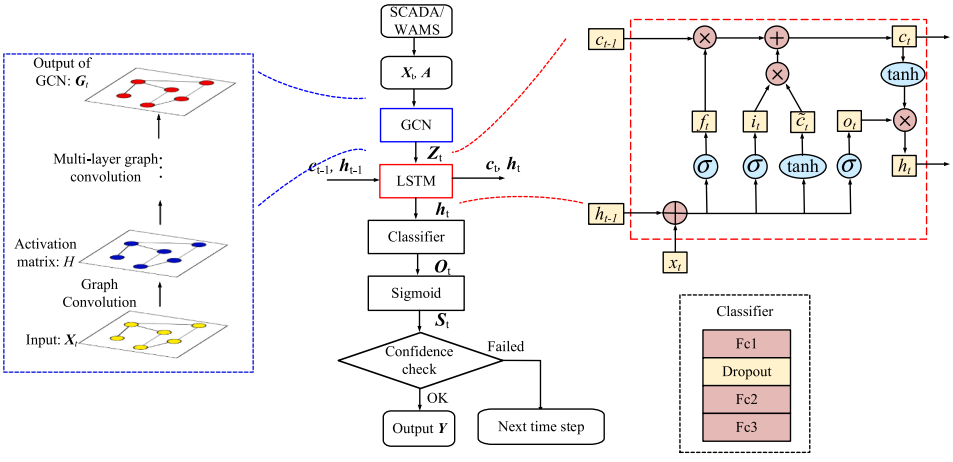

Fig. 3. Structure of STVS on-line prediction method.

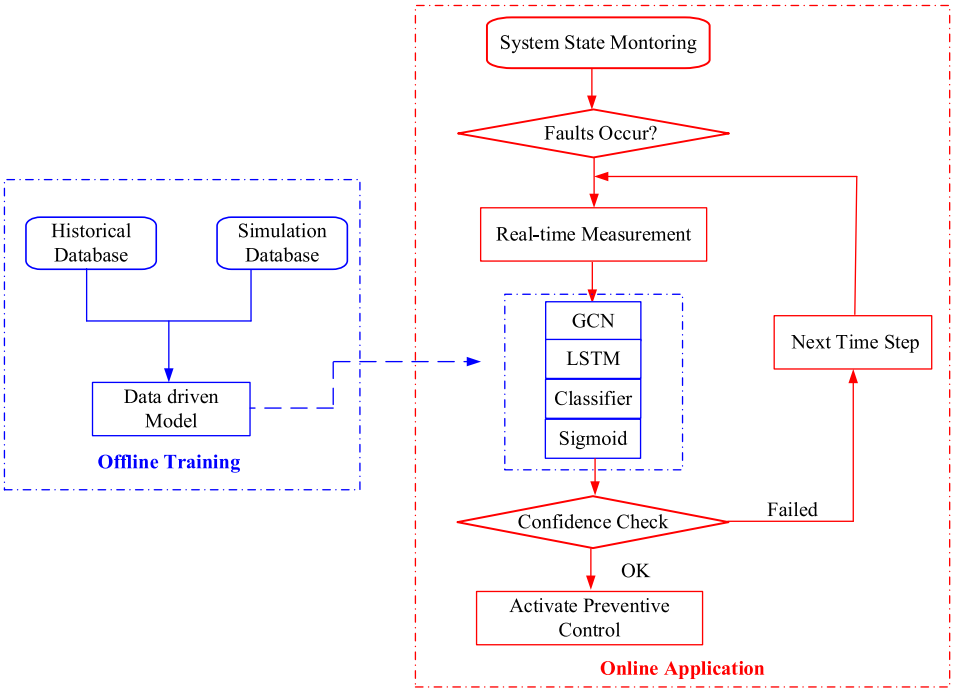

Fig. 4. Framework of model training and application.

the LSTM network includes a internal state matrix $\boldsymbol{c}_t$ and a external state matrix $\boldsymbol{h}_t$ at moment t. The LSTM is used to learn the dynamic changes of the spatial features.

The third step is a classifier. After the spatial–temporal features of each bus are extracted by the GCN-LSTM model, they are classified by a classifier. As shown in Fig. 3, full connection (FC) layers and dropout layers are adopted to form the classifier. According to [26], a three-layer FC network can approximate any continuous function at any precision. Therefore, the layer number of the FC is set as three. The expression of the FC is (8).

$$\boldsymbol{O} = \boldsymbol{\omega}^T \boldsymbol{h} + \boldsymbol{\delta} \tag{8}$$

where $\boldsymbol{h}$ and $\boldsymbol{O}$ are input and output data, $\boldsymbol{\omega}$ is the weights and $\boldsymbol{\delta}$ is the bias.

The fourth step is a sigmoid function. In this paper, the STVS online prediction is a binary classification problem which has the following two categories after classification: stable and unstable contingency (the critical stable contingency is regarded as unstable contingency in this paper). Therefore, we choose the sigmoid function as the activation function, and the expression of the sigmoid is shown as (9).

$$s(x) = \frac{1}{1 + e^{-x}} \tag{9}$$

The fifth step is a confidence check. The more information the input data carries, the more accurate the proposed method will be. After faults occur, the input data obtained from the SCADA/WAMS system will carry more information with the increase of the transient duration. However, we hope to get prediction results as soon as possible without sacrificing accuracy. Therefore, a confidence check is needed to determine when to output the results. The expression of the confidence check is (10). For a system with N buses, the output of the sigmoid is a vector $\boldsymbol{S}$ with N elements, $\boldsymbol{S} \in \mathbb{R}^N$. If (10) is satisfied, t is the moment to output the prediction results. Otherwise, the next time step $t+1$ will be taken to obtain more input data if (10) is not satisfied.

$$Tolerance = \|\boldsymbol{S}_t - \boldsymbol{S}_{t-1}\|_\infty < \varepsilon \tag{10}$$

where $\boldsymbol{S}_t$ represents the output of the sigmoid at time t, and ε is a small constant, such as 1e-04.

The sixth step is the output link. If the output of the sigmoid function passes the confidence check, the final prediction results will be output. The expression of the prediction result is (11). If the i^{th} element of the vector $\boldsymbol{S}_t$ is less than 0.5, the i^{th} bus is in STVS. On the contrary, the i^{th} bus is in STVI.

$$\boldsymbol{Y}(i) = \begin{cases} 0 & \boldsymbol{S}_t(i) < 0.5 \\ 1 & \boldsymbol{S}_t(i) \geq 0 \end{cases} \tag{11}$$

where $\boldsymbol{Y}(i)$ denotes the label of the i^{th} bus.

4. Off-line model training

In this section, we will introduce how to train the GCN-LSTM model. The GCN-LSTM model can be trained offline and applied online. The framework of the model training and application is shown in Fig. 4.

For power grids, especially transmission power grids, there is less historical data of instability. With the rapid development of power systems, the historical data cannot reflect the stability of the current system. Therefore, it is difficult to meet the requirements of training using only the historical data. Combining the historical data and the simulation data is a common means for training. The training database should include three parts: feature time series of all buses, an adjacency matrix and labels of all buses.

In terms of the feature time series, preprocessing is required when multiple features are adopted for each bus. In general, the selected feature contain the bus voltage and the injected active/reactive power. However, the magnitude difference between the voltage and the power is very large. Therefore, we use (12) to preprocess the data in this paper. Eq. (12) characterizes the deviation degree of the bus features from steady-state values.

$$x_t = \begin{cases} \dfrac{x_t - x_0}{x_0}, & x_0 \neq 0 \\ 0, & x_0 = 0 \end{cases} \tag{12}$$

where x_t represents the value of a feature at moment t, x_0 represents the value of the feature at initial steady-state point.

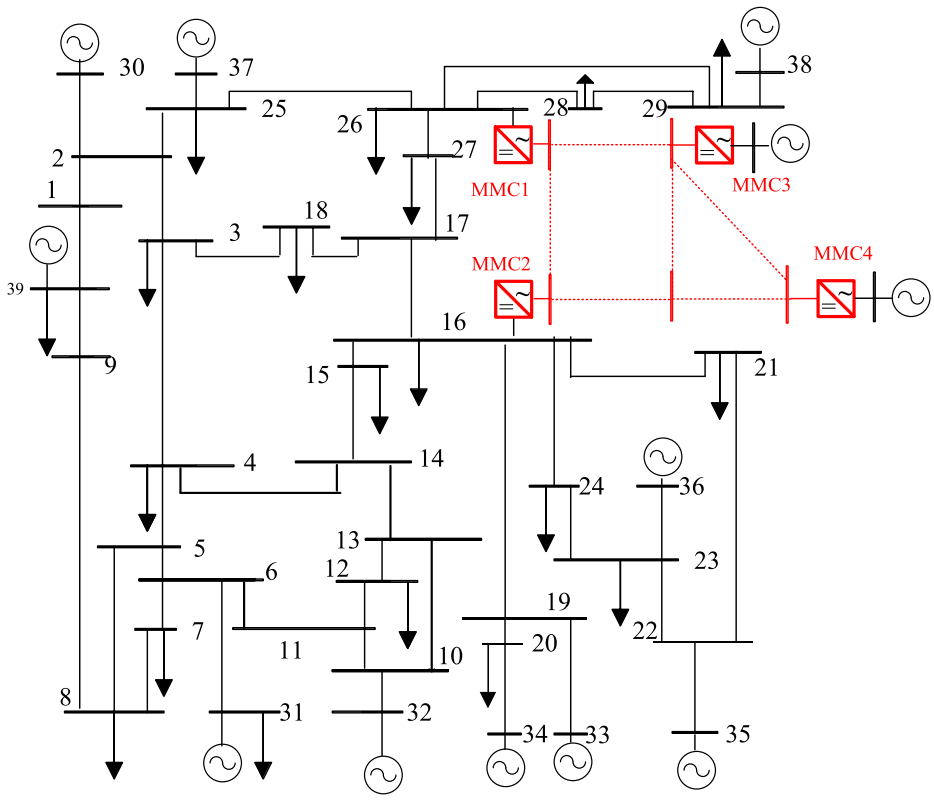

Fig. 5. Network structure of the modified New England 39 buses system.

In terms of the adjacency matrix, a binary matrix or a weighted matrix can be adopted. The binary adjacency matrix is widely used in many graph structured networks without physical meaning, such as social networks and information networks. For power grids which have clear mathematical expressions, a weighted adjacency matrix can reduce the training burden. In a power system, the nodal susceptance matrix can effectively reflect the coupling degree between buses. In this paper, the nodal susceptance matrix is used as the weighted adjacency matrix. Eq. (13) is the expression of the weighted adjacency matrix.

$$\boldsymbol{A}(i,j) = \begin{cases} \sum_{k\in i,k\neq i} b_{ik} + b_{i0}, & j = i \\ 0, & j \neq i, j \notin i \\ -b_{ij}, & j \neq i, j \in i \end{cases} \tag{13}$$

where b_{ij} is the mutual susceptance between bus i and bus j, b_{i0} is the self susceptance of bus i.

In terms of the stability label, the criteria used to determine the label in this paper are as follows:

(1) After the transient process, the bus whose voltage cannot be restored to a normal range is judged as unstable.
(2) For a load bus, if the motor load connected to the bus is stalled, the bus is judged to be unstable.
(3) For a bus with photovoltaics or wind machines connected, if the PV or wind machine is disconnected from the grid due to low AC voltage, the bus is judged to be unstable.
(4) For a bus with DC systems connected, if the DC system is blocked due to low AC voltage, the bus is judged to be unstable.

After determining the architecture of the GCN-LSTM model, the parameters can be trained by using stochastic gradient descent algorithm [27]. In addition, we choose (14) as the loss function. In (14), the first term is the cross entropy loss function used to evaluate the similarity between the prediction result $\hat{y}$ and the label annotated manually y, where T is the number of training samples, and N is the number of buses. The second term is an L2 regularization term that helps to avoid over fitting, λ is a hyper-parameter.

$$Loss = \sum_{i=1}^{T}\left(\sum_{j=1}^{N} y_{ij}\log\hat{y}_{ij}\right) + \lambda L_{reg} \tag{14}$$

5. Case study

The proposed method is implemented in Tensorflow developed by Google. In this section, a modified 39-bus system and the 68-bus system are selected respectively to generate training and testing data. The simulation software is Power System Simulator/Engineering (PSS/E).

5.1. Evaluation metrics

In this paper, three metrics are used to evaluate the performance of the proposed method, as shown in (15)–(17).

(1) Accuracy

$$AC = 1 - \frac{1}{N\times T}\sum_{i=1}^{T}\sum_{j=1}^{N}\left|y_{ij} - \hat{y}_{ij}\right| \tag{15}$$

where T is the number of testing samples, N is the number of buses, y_{ij} is the label annotated manually, $\hat{y}_{ij}$ is the prediction result. AC is a real number between 0 and 1. The closer AC is to 1, the higher the accuracy of the method is.

(2) Confidence

$$CO = 1 - \frac{1}{N\times T}\sum_{i=1}^{T}\sum_{j=1}^{N}\left|y_{ij} - s_{ij}\right| \tag{16}$$

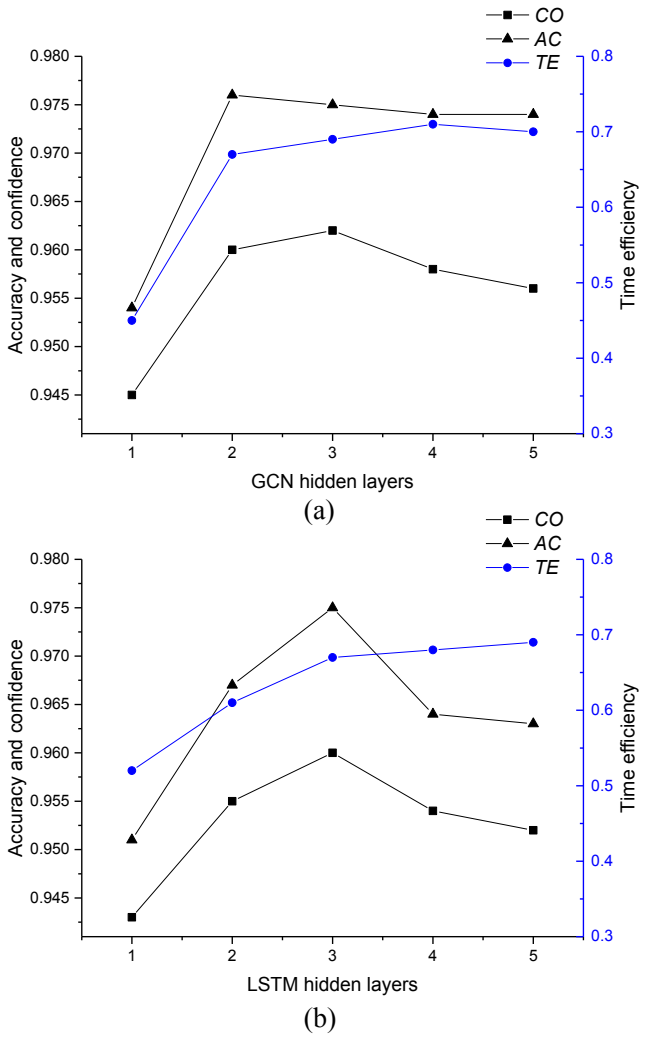

Fig. 6. Three metrics versus hyper-parameters. (a) Three metrics versus the number of the GCN hidden layers. (b) Three metrics versus the number of the LSTM hidden layers.

where s_{ij} is the output of the sigmoid function. The *CO* is used to evaluate the similarity between the sigmoid output and the manually annotated labels. The closer the sigmoid output is to 1, the higher the confidence.

(3) Time efficiency

$$TE = \frac{t_{\mathrm{STVI}} - t_{\mathrm{out}}}{t_{\mathrm{STVI}} - t_{\mathrm{f}}} \tag{17}$$

where t_{out} is the time when the GCN-LSTM model outputs the result, t_{f} is the time when faults occur, t_{STVI} is the time when the STVI occurs. If multiple buses lose stability at different moments, t_{STVI} is the earliest moment. *TE* is always less than 1, the closer the *TE* is to 1, the earlier the prediction results can be given.

5.2. Modified New England 39-bus system

The structure of the modified New England 39-bus system is shown in Fig. 5. Compared with the original New England 39-bus system, a four terminal MMC-HVDC system is added to transmit power from two wind farms. All loads are increased in equal proportion to ensure the balance of supply and demand. All generators use the 6th-order model (GENROU), with the 4th-order excitation system (IEEET1). The comprehensive distribution network load model (CLODZN) and the generic wind generator (WT1G1) are adopted. The load model CLODZN includes the

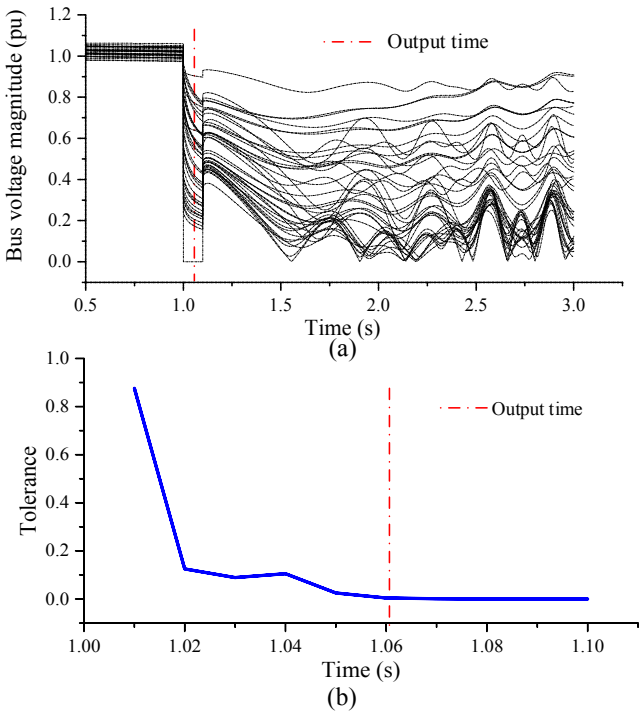

Fig. 7. Case analysis of a voltage collapse contingency. (a) Voltage curves of buses. (b) Tolerance versus time steps.

branch, the induction motor, the discharge lighting, and the constant impedance. See PSS/E manual [28] for more details of the models. In the modified New England 39 buses system, we apply various faults to perform simulations and form the data set, the types of faults include the short-circuit fault, the branch disconnection, the unit cut-off, and the converter blocked. Generated data include 5800 samples, each containing 39 labels. For all labels in the generated data, 57% are unstable and 43% are stable. In addition, 80% of the data is used as the training set and the remaining 20% is used as the testing set.

(1) Hyper-parameters

In the GCN-LSTM model, there are many hyper-parameters that need to be selected manually, including: the learning rate, the batch size, the number of training epochs, hidden units of the GCN, external states of the LSTM, and the number of hidden layers. In this experiment, we manually adjust and set the learning rate to 0.001, the batch size to 64, the training epochs to 1500, the GCN hidden units to 78 (two hidden features for one bus), and the LSTM external states to 3. Since the output of the GCN is the input of the LSTM, the number of neurons in one LSTM unit is 72. Besides, each LSTM layer contains 39 LSTM units.

The number of hidden layers is determined by experiments. Generally speaking, the more complex the system is, the more hidden layers it requires. In addition, for a given data set, excess hidden layers will result in poor performance due to over fitting. To select the most appropriate values, we experiment with different numbers of hidden layers and determine the optimal value.

Fig. 6 shows the three metrics versus the number of hidden layers. As can be seen from Fig. 6a, the accuracy is the largest when the number of the GCN hidden layers is set to 2, while the confidence is the largest when the number of the GCN hidden layers is set to 3. Considering that the improvement of the confidence and the time efficiency is slight when the number of the GCN hidden layers is larger than 2, we set the number of the GCN hidden layers as 2 for reducing the training burden. Similarly, the number of the LSTM hidden layers should be set to 3 by considering comprehensively the three metrics.

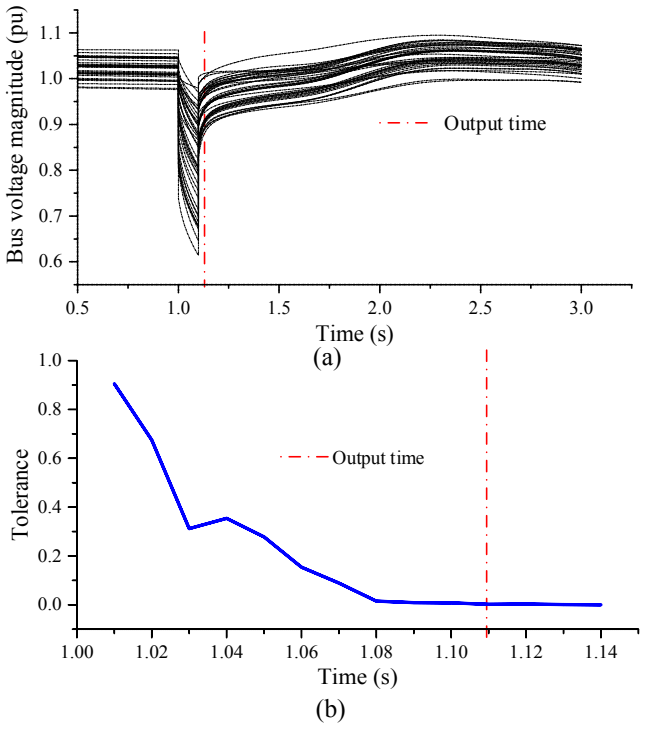

Fig. 8. Case analysis of a voltage restored contingency. (a) Voltage curves of buses. (b) Tolerance versus time steps.

Table 1
Metrics comparison between different features.

Number of case	1	2	3
Number of features	2	4	2
Bus features	U_x, and U_y	U_x, U_y, P_{in} and Q_{in}	R_{eq} and X_{eq}
AC	0.975	0.964	0.978
CO	0.96	0.953	0.963
TE	0.67	0.62	0.67

(2) Two typical contingencies

We analyze a voltage collapse contingency as shown in Fig. 7. At 1 s, a three-phase grounding short-circuit fault occurs in the 39-bus system, the voltage magnitude of bus 4 drops to zero, and the fault lasts for 0.1 s. During the transient process, a large number of induction motors are stalled, the reactive power demand of the loads is greatly increased.

Table 2
Metrics comparison between different adjacency matrices.

Type of adjacency matrix	Binary matrix	Weighted matrix
AC	0.973	0.975
CO	0.955	0.96
TE	0.67	0.67

After clearing the fault, the voltage of buses cannot be restored, indicating that the system voltage has collapsed. As shown in Fig. 7b, in the beginning, the tolerance is very large, which means that the prediction results are not yet credible. At 60 ms after the fault, the tolerance is small enough to output prediction results. If the length of the input time series is much less than 60 ms, the prediction results cannot be adopted due to poor credibility. On the contrary, an excessively long time series may cause the system to suffer greater economic losses.

Similar conclusions can be found for a voltage restored contingency, shown in Fig. 8. It can be observed that the voltage can be effectively restored after clearing the fault. But this does not mean that all buses are stable, there are still individual buses that are judged to be unstable due to stalling of connected motors. For the stable contingency, the prediction results are output at 110 ms after the fault. In summary, the proposed method is time window self-adaptive for each contingency, to make a trade-off between accuracy and time efficiency performance.

(3) Impact of different inputs

In this subsection, we first investigate the impact of different bus features on the method performance, and the corresponding testing results are illustrated in Table 1. In Table 1, U_x is the real part of the bus voltage, U_y is the imaginary part of the bus voltage, P_{in} is the injected active power of the bus, Q_{in} is the injected reactive power of the bus, R_{eq} and X_{eq} can be obtained by (18) and (19).

$$R_{eq} = \frac{U_x^2 + U_y^2}{P_{in}^2 + Q_{in}^2} \times P_{in} \tag{18}$$

$$X_{eq} = \frac{U_x^2 + U_y^2}{P_{in}^2 + Q_{in}^2} \times Q_{in} \tag{19}$$

When the equivalent resistance R_{eq} and reactance X_{eq} are adopted as bus features, the performance of the proposed method is the best, as case 3 in Table 1 shows. Otherwise, the performance of the proposed method is the worst under case 2. Considering that the bus voltage is more easily obtained online, the first case shown in Table 1 is also a good choice for our task. We also compare the loss values with epochs under different cases as shown in Fig. 9a. We found that the loss value sharply decreases with epoch increasing for case 3, while the descending rate of the loss

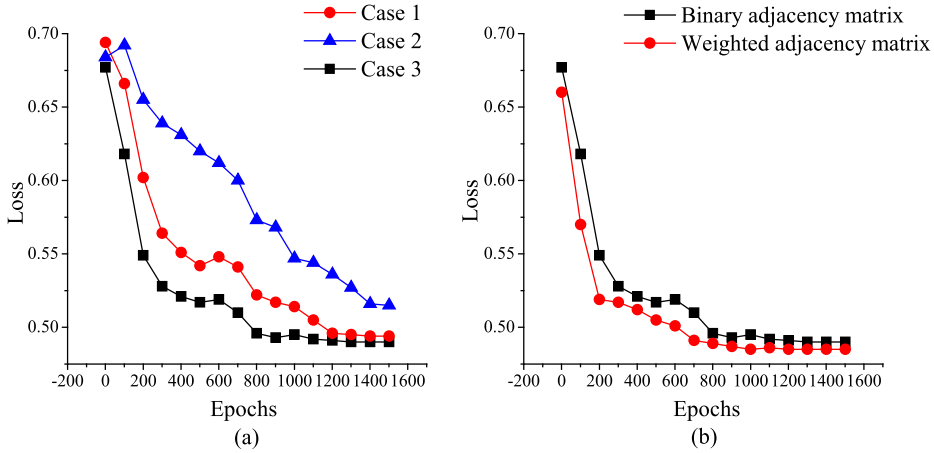

Fig. 9. Loss values versus epochs. (a) Comparison among different bus features. (b) Comparison between different adjacency matrices.

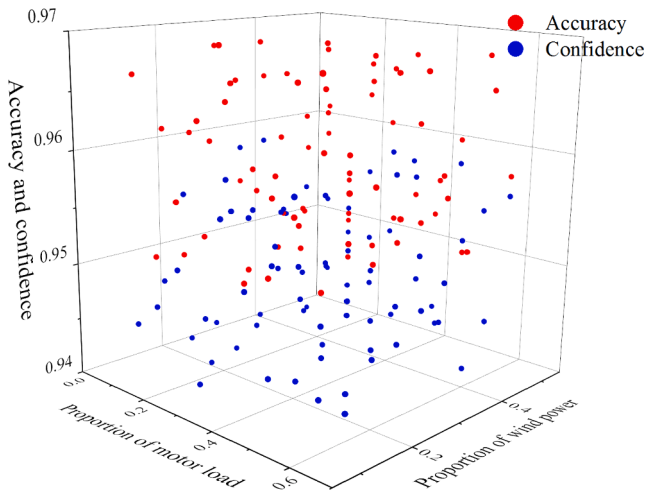

Fig. 10. Metrics under different operating conditions.

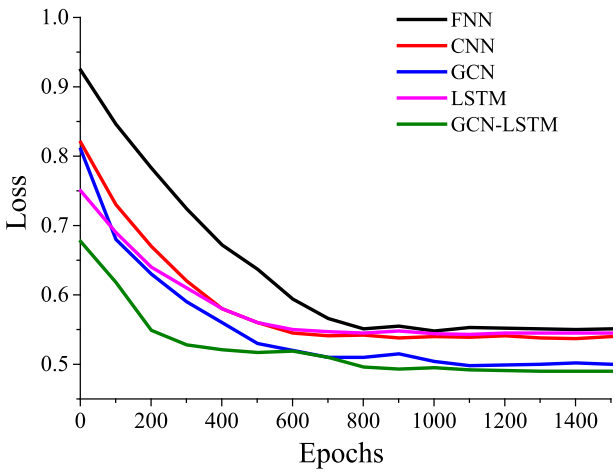

Fig. 11. Loss comparison among different models.

Table 3
Metrics comparison between different methods in New England 39 buses system.

Model	*AC*	*CO*	*TE*
FNN	0.894	0.871	0.21
CNN	0.956	0.947	0.21
GCN	0.968	0.954	0.21
LSTM	0.924	0.905	0.58
GCN + LSTM	0.975	0.96	0.67

value is the slowest for case 2.

The type of the adjacency matrix will also affect the performance of the proposed method. We obtain the test results with different adjacency matrices (Table 2) and investigate the variation of the loss value with epoch increasing during the training process (Fig. 9b). It can be concluded from Table 2 and Fig. 9b that the type of the adjacency matrix has little effect on accuracy, but the weighted matrix is beneficial to improve the training efficiency.

(4) Robustness under different operating conditions

For a power system, the proportion of the motor load has a great influence on the STVS. The higher the proportion of the motor load, the worse the STVS. Similarly, increasing the proportion of the wind power is also detrimental to the STVS. Under different operating conditions, the performance of the proposed method may be different. In order to verify the robustness of the proposed method, we obtain the corresponding metrics by changing the proportion of the motor load and the wind power, as shown in Fig. 10. The range of the motor load proportion is 0.1–0.6, and that of the wind power is 0.1–0.5. In order to ensure the balance of supply and demand, the active power of synchronous generators will be reduced in the same proportion with the wind power increasing. It should be noted that all operating conditions are considered for offline case generation to improve the applicability of the proposed method. As shown in Fig. 10, under various operating conditions, the accuracy of the proposed method is always larger than 0.95, the confidence is always larger than 0.94. Therefore, the proposed method is robust to different operating conditions.

(5) Comparison with different methods

The proposed method is compared with four existing methods: (1) FNN [29]: Feed forward neural network with three hidden layers, whose hidden unit size is 390. (2) CNN [30]: Convolutional neural network with three hidden layers, whose hidden unit size is 390. (3) GCN [24]: Graph convolutional network with three hidden layers, whose hidden unit size is 390. (4) LSTM [25]: Long short-term memory recurrent neural network with three hidden layers, the number of neurons is 140 in one LSTM unit. For FNN, CNN and GCN, the voltage time series data with time interval as 10 steps used as input, then we obtain the stability label for every bus. As for LSTM and GCN-LSTM, the samples of the input are also the voltage time series, and (11) is used to determine the input size. A summary of the results can be found in Table 3. We can found that compared with the GCN-LSTM model, the accuracy of the FNN and LSTM is poor. Although the GCN and CNN have good performance in accuracy, the size of their input data must be fixed, which makes the two models not suitable for online STVS prediction. Therefore, we can conclude from Table 3 that the GCN-LSTM model has better performance than other four existing methods.

In addition, we compare the training efficiency of the GCN-LSTM model with other models. Fig. 11 shows the validation loss curves versus the training epoch. The loss of the GCN-LSTM is the smallest among all the models. Compared with the four existing methods, the GCN-LSTM model can quickly reaches to converge with fewer epochs. Therefore, the proposed method is a more advanced method for the STVS online prediction tasks.

5.3. NETS-NYPS 68-bus system

The 68-bus system is a reduced order equivalent of the interconnected New England test system (NETS) and New York power system (NYPS). The 68 buses system contains 68 buses, 16 generators, and 35 loads. The network structure of the 68-bus system can be found in [31]. All generators use 6th-order model (GENROU), with 4th-order excitation system (IEEET1). The comprehensive load model CLODZN is adopted for all loads. Similarly, 6800 samples are generated, each sample contains 68 labels. For all labels in the generated data, 61% of which are stable, and 39% are unstable.

In order to guarantee the performance of the proposed method, the hyper-parameters need to be reset. In this experiment, the learning rate is set as 0.001, the batch size is 64, the training epoch is 1500, the number of the GCN hidden units is 136 (two hidden features for one bus), and the number of the LSTM external state is 3. Then the number of neurons in a LSTM unit can be obtained, that is 72. Besides, each LSTM layer contains 68 LSTM units. The number of hidden layers of the GCN and the LSTM are set as 2 and 3 respectively.

Similarly, we compare the proposed method with four existing methods, including FNN (680 hidden units and 3 hidden layers), CNN (680 hidden units and 3 hidden layers), GCN (680 hidden units and 3 hidden layers), and LSTM (140 neurons and 3 hidden layers). The voltage time series data is taken as the input of all the models. For FNN,

Table 4
Metrics comparison between different methods in NETS-NYPS 68 buses system.

Model	*AC*	*CO*	*TE*
FNN	0.912	0.897	0.15
CNN	0.965	0.958	0.15
GCN	0.971	0.959	0.15
LSTM	0.941	0.927	0.52
GCN + LSTM	0.981	0.963	0.59

CNN and GCN, the time series length is 10 time steps. For LSTM and GCN-LSTM, the time series length is determined by (11).The results are shown in Table 4.

Comparing Table 3 and Table 4, it can be seen that the accuracy of each model is improved. The reason is that the 68-bus system has no power electronic device, and the dynamic characteristics of the system become simpler. We also found that the proposed method performs better than benchmark methods in different systems by comparing the results of Table 4 with Table 3.

6. Conclusions and future scope

In this paper, we propose a novel STVS online prediction method based on the GCN and the LSTM. In a transient process, this method can predict and locate unstable buses, providing reference for the implementation of preventive measures. The idea is to develop a data-driven methodology to learn from real-time measurement data and obtain stability conclusions based on current available operating data. In doing so, we use a graph network to model the power system in which the nodes on the graph represent buses, the edges represent the branches, and the electrical quantity on the buses is described as the feature of the nodes. The GCN and the LSTM are coordinated to learn the graph and extract features, a classifier is applied to classify the buses, and a confidence check link is designed to determine when to output the prediction results. In short, this method provides a self-adaptive monitoring time window for each contingency, to make a trade-off between accuracy and time efficiency. The results of case studies have shown that the proposed method achieves a better prediction accuracy, compared with existing data-driven approaches. In summary, the proposed method can successfully predict and locate unstable buses with superior performance.

How to take effective measures to prevent the occurrence of instability using data-driven methods, will be the focus of future work. At present, the stability control strategies for a power system are all determined by human experience. However, the subjective judgment of operators will great effect the efficiency of the stability control strategy. Therefore, using data-driven method to achieve reasonable scheduling will be the future scope of research.

CRediT authorship contribution statement

Guoteng Wang: Conceptualization, Methodology, Software, Visualization, Writing - original draft. **Zheren Zhang:** Formal analysis, Resources, Data curation, Writing - review & editing. **Zhipeng Bian:** Validation, Investigation. **Zheng Xu:** Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Kundur P. Power system stability and control. New York (NY, USA): McGraw-Hill; 1994.
[2] Adibi MM, Hirsch PM, Jordan JA. Solution methods for transient and dynamic stability. Proc IEEE 1974;62(7):951–8.
[3] Kundur P, et al. Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions. IEEE Trans Power Syst 2004;19(3):1387–401.
[4] Zhao W, Guo Q, Sun H, Ge H, Li H. Practical short-term voltage stability index based on voltage curves: definition, verification and case studies. IET Gener Transm Distrib 2018;12(19):4292–300.
[5] Costa FB, Driesen J. Assessment of voltage sag indices based on scaling and wavelet coefficient energy analysis. IEEE Trans Power Del 2013;28(1):336–46.
[6] Shen C, Lu C. A voltage sag index considering compatibility between equipment and supply. IEEE Trans Power Del 2007;22(2):996–1002.
[7] Dasgupta S, Paramasivam M, Vaidya U, Ajjarapu V. Real-time monitoring of short-term voltage stability using PMU data. IEEE Trans Power Syst 2013;28(4):3702–11.
[8] Pinzón JD, Colomé DG. PMU-based online monitoring of short-term voltage stability using Lyapunov exponents. IEEE Lat Am Trans 2019;17(10):1578–87.
[9] Omar Faruque MD, Strasser T, Lauss G, et al. Real-time simulation technologies for power systems design, testing, and analysis. IEEE Power Energy Technol Syst J 2015;2(2):63–73.
[10] Park M, Yu I. A novel real-time simulation technique of photovoltaic generation systems using RTDS. IEEE Trans on Energy Convers 2004;19(1):164–9.
[11] Zhu L, Lu C, Kamwa I, Zeng H. Spatial–temporal feature learning in smart grids: a case study on short-term voltage stability assessment. IEEE Trans Ind Informat 2020;16(3):1470–82.
[12] Ge H, Guo Q, Sun H, et al. A model and data hybrid-driven short-term voltage stability real-time monitoring method. Int J Electr Power Energy Syst 2020;144:105373.
[13] Ren C, Xu Y. A fully data-driven method based on generative adversarial networks for power system dynamic security assessment with missing data. IEEE Trans Power Syst 2019;34(6):5044–52.
[14] Yan R, Geng G, Jiang Q, Li Y. Fast transient stability batch assessment using cascaded convolutional neural networks. IEEE Trans Power Syst 2019;34(4):2802–13.
[15] Gupta A, Gurrala G, Sastry PS. An online power system stability monitoring system using convolutional neural networks. IEEE Trans Power Syst 2019;34(2):864–72.
[16] Yang H, Zhang W, Chen J, et al. PMU-based voltage stability prediction using least square support vector machine with online learning. Electr Power Syst Res 2018;60(160):234–42.
[17] Bashiri Mosavi A, Amiri A, Hosseini H. A learning framework for size and type independent transient stability prediction of power system using twin convolutional support vector machine. IEEE Access 2018;6:69937–47.
[18] Zhu L, Lu C, Liu Y, Wu W, Hong C. Wordbook-based light-duty time series learning machine for short-term voltage stability assessment. IET Gener, Trans Distrib 2017;11(18):4492–9.
[19] Liao H, Milanović JV, Rodrigues M, Shenfield A. voltage sag estimation in sparsely monitored power systems based on deep learning and system area mapping. IEEE Trans Power Del 2018;33(6):3162–72.
[20] Yu J, Hill D, Lam A, Gu J, Li V. Intelligent time-adaptive transient stability assessment system. IEEE Trans Power Syst 2018;33(1):1049–58.
[21] Pan F, Li J, Tan B, et al. Stacked-GRU based power system transient stability assessment method. Algorithms 2018;11(8).
[22] Zhu L, Lu C, Sun Y. Time series shapelet classification based online short-term voltage stability assessment. IEEE Trans Power Syst 2016;31(2):1430–9.
[23] Khoshkhoo H, Shahrtash SM. Fast online dynamic voltage instability prediction and voltage stability classification. IET Gener, Transm Distrib 2014;8(5):957–65.
[24] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: Proc 2017 international conference on learning representations (ICLR), Toulon, France; 2017.
[25] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[26] Hornik K. Approximation capabilities of multilayer feedforward networks. Neural Netw 1991;4(2):251–7.
[27] Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge (MA, USA): MIT Press; 2016.
[28] Model Library of PSS/E 33 [Z]. America: PTI Inc; 2013.
[29] Kuschewski JG, Hui S, Zak SH. Application of feedforward neural networks to dynamical system identification and control. IEEE Trans Control Syst Technol 1993;1(1):37–49.
[30] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proc 2015 IEEE Conference On Computer Vision And Pattern Recognition (CVPR), Boston, MA; 2015.
[31] Rogers G. Power system oscillations. USA: Kluwer Academic Publishers; 2000.