

Detection and Classification of Transmission Line Transient Faults Based on Graph Convolutional Neural Network

Houjie Tong¹, Robert C. Qiu, *Fellow, IEEE*, Dongxia Zhang, Haosen Yang, Qi Ding, and Xin Shi

Abstract—We present a novel transient fault detection and classification approach in power transmission lines based on graph convolutional neural network. Compared with the existing techniques, the proposed approach considers explicit spatial information in sampling sequences as prior knowledge and it has stronger feature extraction ability. On this basis, a framework for transient fault detection and classification is created. Graph structure is generated to provide topology information to the task. Our approach takes the adjacency matrix of topology graph and the bus voltage signals during a sampling period after transient faults as inputs, and outputs the predicted classification results rapidly. Furthermore, the proposed approach is tested in various situations and its generalization ability is verified by experimental results. The results show that the proposed approach can detect and classify transient faults more effectively than the existing techniques, and it is practical for online transmission line protection for its rapidness, high robustness and generalization ability.

Index Terms—Graph convolutional network (GCN), power transmission line, fault detection and classification, spatio-temporal data, topology information.

I. INTRODUCTION

TRANSIENT fault detection and classification of power transmission are the basis of the analysis and treatment of power accidents, which are of great significance for improving the stability of power grid. With the growing scale of interconnection and the development of operation under stressed condition in modern power systems [1], [2], the features of transient faults become more complex, which makes fault detection and classification more urgent. Only by promptly and accurately determining the type of faults that occur in

the transmission systems can the operator take effective emergency control actions according to the classification results, which facilitates the location of the faults and reduces the time of eliminating them.

For transient fault detection and classification, the extraction of fault features is a key task. Different from the fault identification based on image data [3], the feature extraction of voltage and current data involved in this paper is more abstract. Early researches are based on transmission line fault mechanism model. In reference [4], the mechanism of fault current generation and the fault features are analyzed by establishing the expression of fault current. Reference [5] uses the fault equivalent circuit to determine the fault current and threshold to classify the faults. The above researches derive the expression of fault current or voltage through fault mechanism analysis, and finally make fault diagnosis. The above model-driven techniques may achieve good results under specific scenarios, but poor generalizability is their drawback. The key reason is that single model cannot fully depict the various mechanisms involved in electrical events, and it will become invalid in variable environments [6]. Moreover, these methods usually require many assumptions, and the modeling process involves a lot of manual calculation which is time-consuming and labor-intensive [7].

With the rise of artificial intelligence technology in the era of industrial big data, the data-driven fault detection and classification method begins to show more remarkable performance. There is an early work using support vector machine (SVM) to identify transmission line fault types under different fault working conditions, and taking wavelet singular information as characteristic parameters [8]. Reference [9] defines four multi-wavelet packet entropy to extract transmission line fault signals and uses radial basis function neural network to achieve classification results. Then some scholars adopt decision tree algorithm and K-nearest neighbor (KNN) algorithm respectively to identify transmission line fault types [10], [11]. In recent years, the end-to-end neural network (NN) featuring self-learning ability is introduced. In reference [12], sparse autoencoder (SAE) is proposed to process voltage and current signals of transmission lines for fault classification. Reference [13] regards the voltage signal matrix as a grayscale image for input, thereby using convolutional neural network (CNN) to realize fault classification. The concept of spatio-temporal matrix is mentioned in some literature such as reference [14], [15]. The authors utilize the random matrix theory (RMT) to

Manuscript received September 15, 2020; revised December 30, 2020; accepted January 19, 2021. Date of online publication April 30, 2021; date of current version May 7, 2021. This work was supported by the National Key Research and Development Program of China under Grant 2018YFF0214704.

H. J. Tong (corresponding author, email: thj_926@sjtu.edu.cn; ORCID: <https://orcid.org/0000-0002-9683-8208>), R. C. Qiu, H. S. Yang, and Q. Ding are with the Department of Electrical Engineering, Center for Big Data and Artificial Intelligence, Shanghai Jiao Tong University, Shanghai 200240, China. R. C. Qiu is also with the School of Electronic Information and Communication, Huazhong University of Science and Technology, Wuhan 430000, China.

D. X. Zhang is with China Electric Power Research Institute, Haidian District, Beijing 100192, China.

X. Shi is with the School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China.

DOI: 10.17775/CSEEJPES.2020.04970

explore the spatio-temporal correlation of abnormal data in reference [14], but the spatio-temporal matrix in it does not introduce the connection relationship between nodes. In other words, the spatial relationship of data is not provided as prior knowledge.

As pointed out above, the evolution of fault detection and classification methods is from model-driven to data-driven. With the power industry gradually becoming intelligent, the more complex characteristics of transient faults and the multi-variation of transmission system operation make data-driven fault detection and classification techniques still the leading approaches at present [1], [16]. By reviewing, we find that even though many studies investigated new approaches to improve the effect of fault detection and classification, few researchers considered the explicit spatial relations among the fault data of transmission systems. However, power system data, as a typical “industrial big data”, is indeed a kind of spatio-temporal data [17], [18]. Although the spatio-temporal correlation of fault data is vital, it is difficult to introduce it explicitly into the detection and classification task through the existing techniques. Our work aims to fill this gap. Along with the well-established research line of GCN, we come up with more new ideas as we deal with the graph structure data.

Graph NN is first introduced by Bruna *et al.* [19]. It applies convolutional layers on the graph structured data rather than just regular data such as images. Compared with the case that CNN cannot effectively process irregular data, researchers could make effective use of explicit spatial information when using GCN. Due to the universality and diversity of graph structure data in our life, the development and application of GCN are rising rapidly. It has been successfully applied in recommendation system, social network, life science and other fields [20]. In our view, the power system topology is naturally a graph and the edge information in the graph can also be extended to “electrical distance”. The topological structure of power transmission system and large amounts of measurement data provide a new opportunity for proper applications of GCN in power systems. In our work, a GCN based on power topology is used to detect and classify transient faults.

Briefly, this paper has the following contributions:

1) To the best of our knowledge, this is the first work of leveraging GCN to implement the transient fault detection and classification task.

2) A drawback of the existing techniques is pointed that the effect of explicit spatial information has not been taken into account. Therefore, we provide a novel idea of embedding the spatio-temporal relations between data into detection and classification models. To be brief, we propose to regard the transmission line topology as a graph and utilize topology parameters to construct graph elements.

3) In addition, we introduce the receiver operating characteristic (ROC) curve instead of only using accuracy to characterize the overall performance of the classifier [21]. Further, comparison studies are implemented from three aspects to analyze the generalization ability of our proposed method and the existing machine learning techniques.

The rest of our paper is organized as follows. Section II introduces graph convolutional network. Section III discusses

the problem statement and the proposed method framework. Section IV is the case study. Section V summarizes our work.

II. GRAPH CONVOLUTIONAL NEURAL NETWORK

In this section, we first introduce the graph structure and provide an example to illustrate the workflow of graph convolution.

A. Graph Structure

In our work, we treat a transmission system as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with N nodes $v_i \in \mathcal{V}$, edges $(v_i, v_j) \in \mathcal{E}$, an adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ (weighted) and a degree matrix $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$. The structure of undirected graph is depicted in Fig. 1.

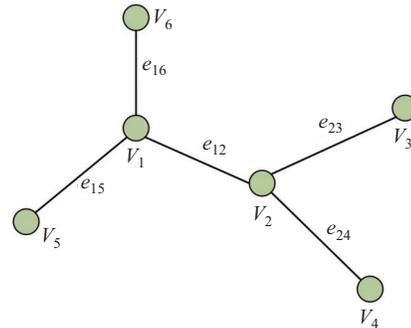


Fig. 1. Undirected graph structure with nodes v_i and edge weights e_{ij} ($i, j = 1, 2, 3, 4, 5, 6$).

The adjacency matrix \mathbf{A} represents the connection relationships of all nodes in a graph, as follows:

$$\mathbf{A} = \begin{bmatrix} 0 & e_{12} & 0 & 0 & e_{15} & e_{16} \\ e_{21} & 0 & e_{23} & e_{24} & 0 & 0 \\ 0 & e_{32} & 0 & 0 & 0 & 0 \\ 0 & e_{42} & 0 & 0 & 0 & 0 \\ e_{51} & 0 & 0 & 0 & 0 & 0 \\ e_{61} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{6 \times 6} \quad (1)$$

where e_{ij} in the matrix represents the correlation between the i th node and the j th node. If two nodes are connected by an edge, e_{ij} is equal to the weight coefficient of this edge; if not, then $e_{ij} = 0$. It is worth noting that $e_{ij} = e_{ji}$ in the undirected graph. Besides, degree matrix \mathbf{D} is diagonal and the value of the diagonal element equals the number of adjacent nodes of the corresponding node.

B. Workflow of Graph Convolution

Graph convolution was originally derived based on graph theory and convolution theorem with the purpose of applying it to graph data processing [22]. Through constant refinement and optimization of the model, the expression of GCN has become more understandable.

In practical application, we utilize the commonly used GCN for graph convolution operation, in the form of feature transfer and aggregation through self-normalized adjacency matrix. This GCN is proposed by Kipf *et al.* [23], and its one layer operation is as follows:

$$\mathbf{Z} = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X} \mathbf{W}) = \sigma(\hat{\mathbf{A}} \mathbf{X} \mathbf{W}) \quad (2)$$

where the resultant $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ is the adjacency matrix with self-loop. Self-loop can maintain the information of the target station itself in the convolution part, which is a required design strategy in GCN. And $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$ is the degree matrix of $\tilde{\mathbf{A}}$, so $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}$ represents the self-normalized adjacency matrix. \mathbf{W} is the trainable weight matrix and $\sigma(\cdot)$ is the activation function.

The workflow for graph convolution is depicted in Fig. 2. Firstly, taking the graph in Fig. 1 as an example, we assume that the feature of node v_i is $\mathbf{X}_i = [x_{i1} \ \dots \ x_{in}]^T$, so $\mathbf{X} \in R^{6 \times n}$. Secondly, the function of multiplying $\hat{\mathbf{A}}$ by \mathbf{X} is to transfer and aggregate the features of the adjacent nodes, as shown in the middle part of Fig. 2. Finally, \mathbf{Z} is the output of this GCN layer, on which all nodes contain first-order neighborhood information. It is easy to deduce that the output neurons obtained through k GCN layers can express k -order neighborhood information (spatial information). Therefore, the hidden layer data of GCN can provide more prior information for the model training, so that the trained hidden layer neurons have a deeper feature expression ability.

III. DETECTION AND CLASSIFICATION OF TRANSIENT FAULTS BASED ON GRAPH CONVOLUTIONAL NEURAL NETWORK

A. Problem Statement

Transmission line transient faults can be divided into single-phase ground fault, two-phase short circuit fault, two-phase ground fault and three-phase short circuit fault. Common causes of these faults are lightning strike, wind deviation, pollution flashover, icing, external force, bird damage and some internal faults of the system. The severity of the four main types of faults is obviously different. Three-phase short circuit fault is the most harmful fault in the power transmission system and requires the shortest clearing time. Single-phase ground fault is not as harmful as other kinds of faults, but should not be neglected due to its high occurrence frequency,

accounting for more than 90% of the total faults. When transient fault occurs in transmission line, the nodal voltage will drop to different degrees. To show the characteristics of different faults, some tests are implemented on a small transmission system with few nodes. Fig. 3 show the changes of nodal voltage waveform before and after the occurrence of various faults.

1) Normal Condition

In the normal condition of a transmission system, the nodal voltage range usually stays around 1.0 p.u. Unlike the steady state simulation data mentioned in other papers, slight voltage fluctuations may occur in the transient data under normal operating conditions.

2) Fault Condition

As shown in Fig. 3, the voltage reduction amplitude of different nodes under different fault conditions is different. Furthermore, influences on different nodes under the same fault are different due to different distances from the fault location.

The final purpose of our work is to detect the occurrence of faults and to determine which kind of fault occurs by learning deep representations of system nodal voltages.

B. Construction of Transmission System Graph Structure

In this section we elaborate on how to construct the graph structure data in transmission system. And this proposed way to construct graph can be generalized to other engineering tasks involving topology structure in power system. Here we take the IEEE 9-bus system as an example to illustrate the construction process of the graph which is shown in Fig. 4.

First of all, we define the nodes and their features. Obviously, nodes here are the bus nodes in the transmission system, whose features are bus voltages or other electrical data.

Secondly, edges are defined. The lines in the topology are the edges of the graph, which means that if there is a line between two nodes, an edge can be assumed to exist between

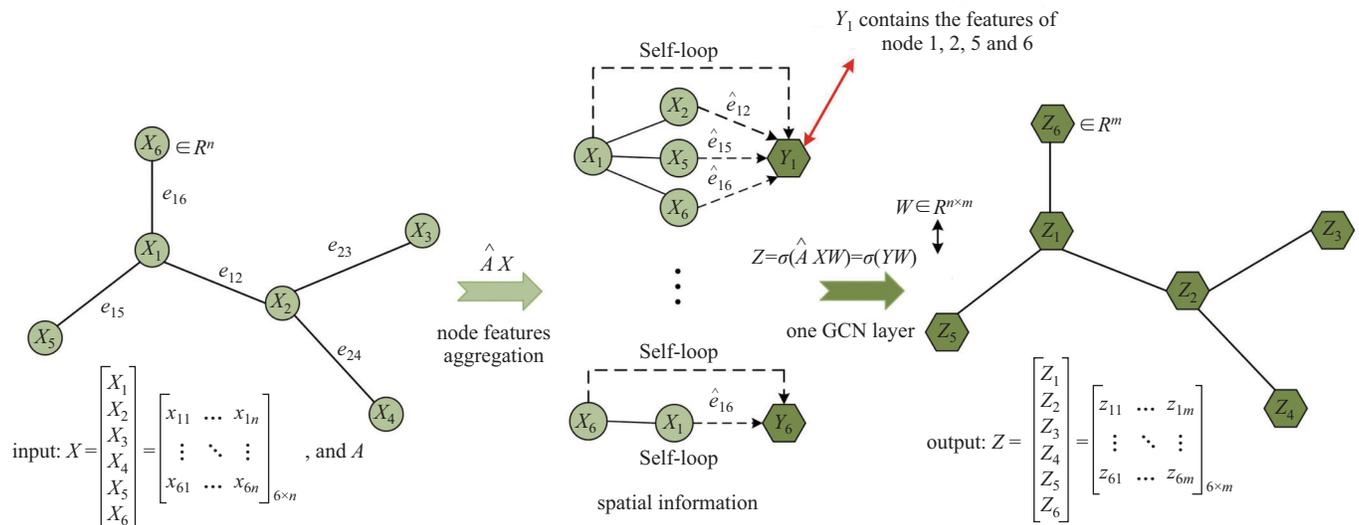


Fig. 2. Workflow of graph convolution (left part: input nodal features and edge weights; middle part: transfer and aggregate nodal features; right part: output new representations of nodal features).

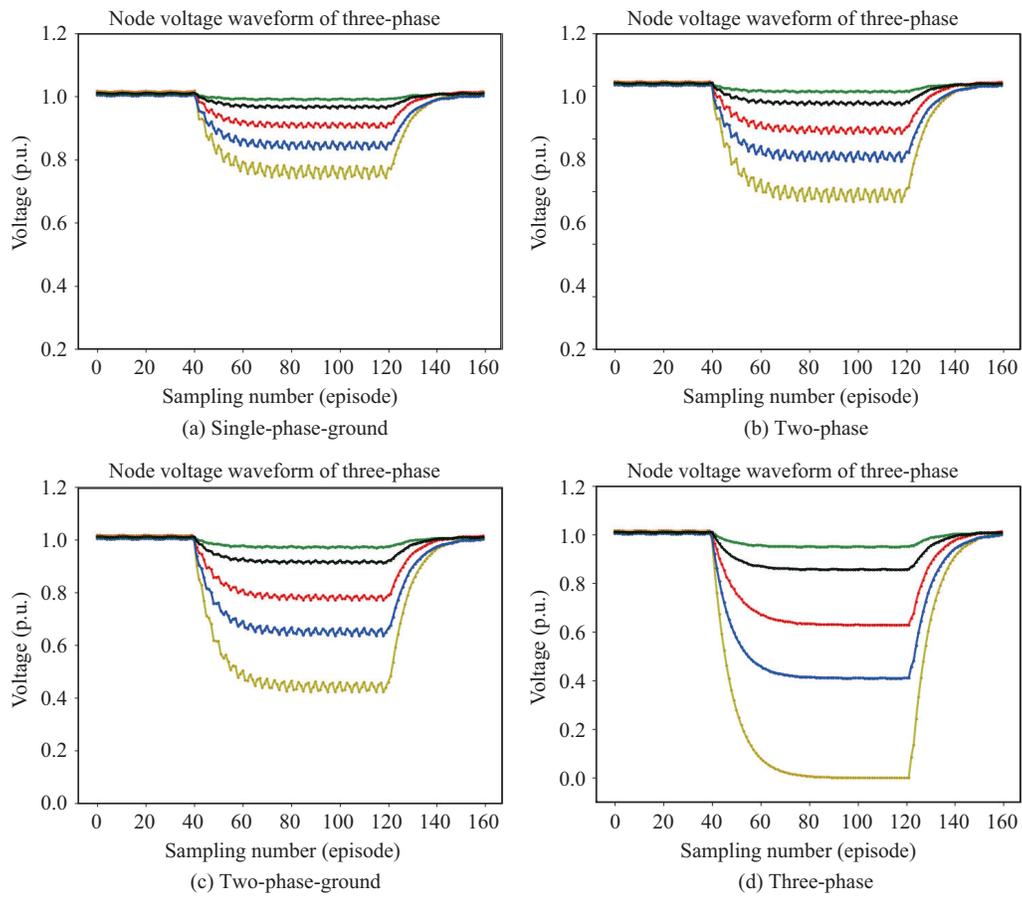


Fig. 3. Data of different fault types.

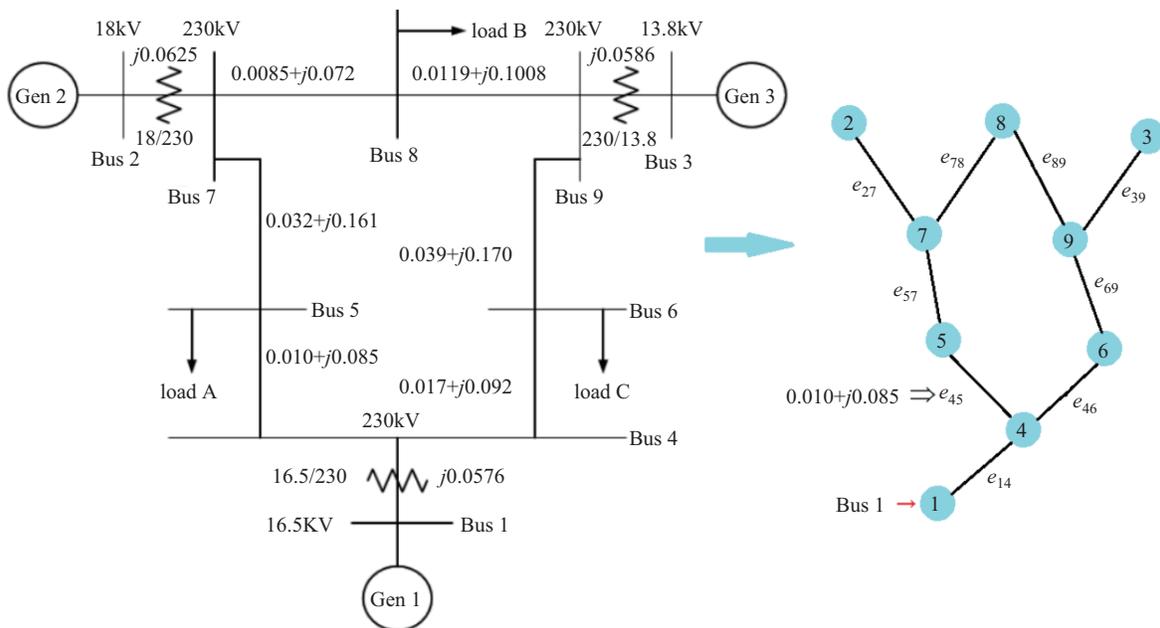


Fig. 4. Graph construction (bus \rightarrow node; line \rightarrow edge; line impedance \rightarrow edge weight).

these two nodes in the graph.

In addition, we reckon that edges should be informative. The consideration of nodal features and the existence of edges only covers the topological connections in spatial information, but

does not quantify the correlation between nodes. In reality, if the bus nodes are far apart from each other, the similarity between them may not be high even if they are connected. If the edge weights are not considered according to the actual

scene, the aggregation of neighborhood information will be less accurate in the process of graph convolution. Therefore, we create a calculation criterion to get the weight of each edge using the line parameters. The calculation formula is as follows:

$$e_{ij} = \begin{cases} \frac{1}{\sqrt{R_{ij}^2 + X_{ij}^2}} & \text{if node } i, j \text{ are connected} \\ 0 & \text{else} \end{cases} \quad (3)$$

where e_{ij} represents the edge weight coefficient between node i and node j , and R_{ij} and X_{ij} represent the line resistance and reactance parameters respectively. The significance of this equation is that a longer distance or a larger impedance indicates a smaller correlation between nodes. The idea of introducing this criterion comes from the definition of edges in relevant applications of social recommendation [24]. In the interpersonal graph, edge information is used to represent the “user-user” relationship, which is usually called “social distance”. Therefore, we come up with the idea of extending the “social distance” to the “Electrical distance” in the power system. Specifically, “social distance” is the quantification of the relationship between people, while “electrical distance” is the quantification of the spatio-temporal correlation of power data. In addition, we propose (3) to provide a reasonable idea for constructing edge weights. In reality, there can be various ways of defining edge information according to different requirements.

Finally, we construct the graph structure based on the transmission system, as shown in Fig. 4.

C. Workflow of GCN Based on Fault Detection and Classification

After building the graph structure, the next step is to build a GCN model based on fault detection and classification. Concretely, this network is expected to output a non-faulty

result when no fault occurs, and will detect the corresponding fault when a specific fault type occurs.

To illustrate the whole workflow concisely, voltages are selected as the nodal features. Then the input matrix \mathbf{X} of GCN can be written in the form of the following matrix (IEEE 9-bus system as an example):

$$\mathbf{X} = \begin{bmatrix} u_{1,t1} & u_{1,t2} & \cdots & u_{1,ts} \\ u_{2,t1} & u_{2,t2} & \cdots & u_{2,ts} \\ \vdots & \vdots & \ddots & \vdots \\ u_{9,t1} & u_{9,t2} & \cdots & u_{9,ts} \end{bmatrix}_{9 \times s} \quad (4)$$

where each row of \mathbf{X} represents the series of voltage magnitude of a particular node in IEEE 9-bus system and s means the number of voltage samplings in a sampling time. We consider using a multi-layer GCN for supervised fault classification. We first construct the self-normalized adjacency matrix $\hat{\mathbf{A}}$ according to the edge weights calculated by formula (3), and then our feedforward network takes the simple form:

$$\mathbf{H}^{(l+1)} = \sigma \left(\hat{\mathbf{A}} \mathbf{H}^{(l)} \mathbf{W}^{(l)} \right) \quad l = 0, 1, 2, \dots \quad (5)$$

Here, σ is an activation function such as RELU, Sigmoid, and $\mathbf{H}^{(0)} \in \mathbb{R}^{n \times s}$ is equal to \mathbf{X} ; if we assume that the first hidden layer has H feature maps, then $\mathbf{W}^{(0)} \in \mathbb{R}^{s \times h}$ is an input-to-hidden weight matrix. The l^{th} hidden-layer also determines the vector dimension by designing the number of feature maps. As described in Section II-B, the convolution operation process will not be repeated here. Here we assume that the final GCN layer output $\mathbf{H} \in \mathbb{R}^{n \times m}$.

Since it is a classification task, the last layer of our model needs to be a fully connected layer, as shown in Fig. 5.

The output features of the last hidden layer are stacked into a long feature vector $\mathbf{S}^{(i)}$, which is used as the input vector of a softmax classifier. The length of $\mathbf{S}^{(i)}$, n_s , is calculated as

$$n_s = n \times m \quad (6)$$

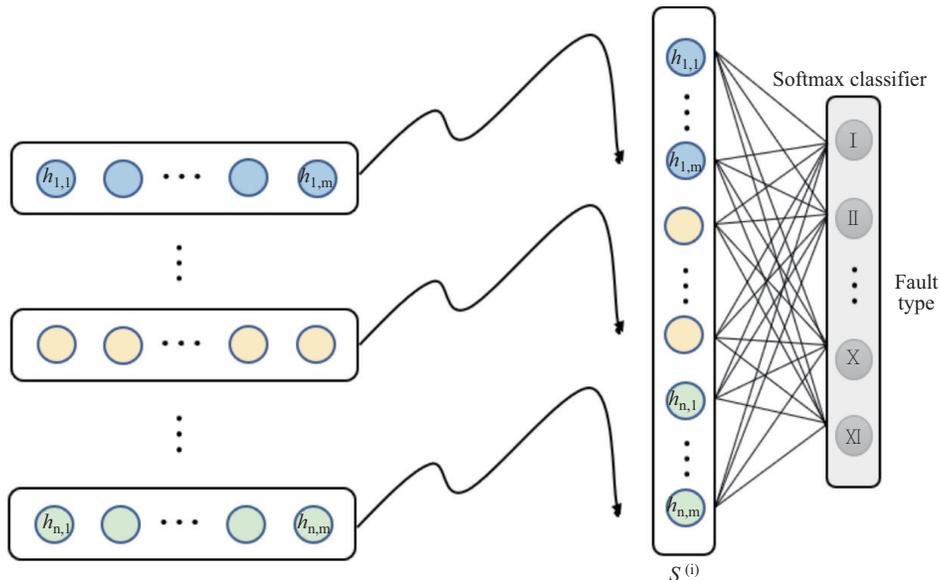


Fig. 5. Schematic diagram of softmax classifier.

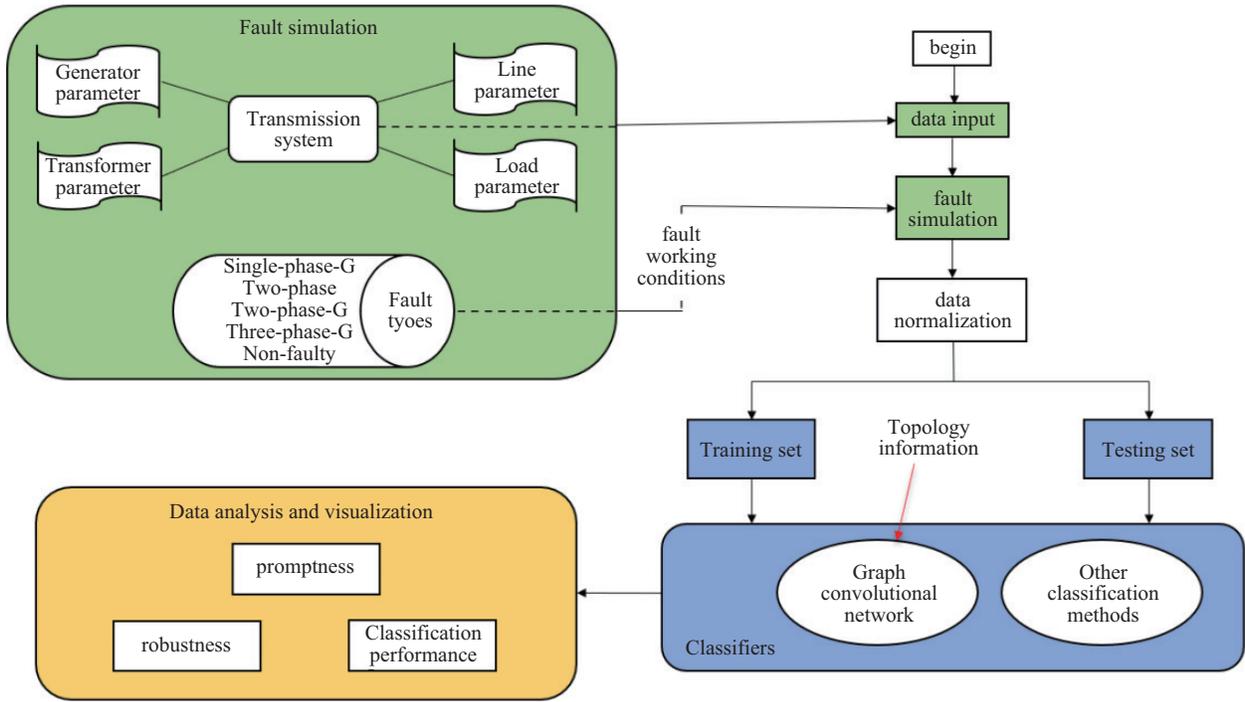


Fig. 7. Flow chart.

TABLE II
HYPERPARAMETERS OF GCN

Hyperparameter	Values or types
GCN layer	Three layers (150, 300, 150 neurons)
Loss function	Cross-entropy loss
Optimizer	Adam

choose a model with 3 hidden layers by testing and comparing the effects of different layers. The numbers of hidden neurons are determined to be 150, 300 and 150 respectively through constant tuning and optimization [27] And *Relu* is selected as the activation function of each layer. When propagating to the last hidden layer, dimension of the nodal features becomes 150, while the number of nodes remains 39. Thus the input size of softmax classifier is $39 \times 150 = 5850$. For supervised multi-classification problems, we usually choose the cross-entropy error as the cost function because it can be used to calculate the loss through a simple derivative and has a fast rate of convergence [28]. The calculation formula is as follows:

$$CE(p, q) = - \sum_{i=1}^C p_i \log(q_i) \quad (9)$$

where C represents the number of categories, p_i is the true value and q_i is the predicted value.

Adam algorithm is chosen as the optimizer owing to its fast convergence speed, high learning efficiency and small memory requirement. It is exceedingly suitable for processing large data set pairs and has great processing capacity for sparse data and data with noise samples [29]. In our test, Adam performs better than other optimizers such as Stochastic Gradient Descent (SGD) and Batch Gradient Descent (BGD). Therefore, Adam is finally selected to realize the automatic

adjustment of the learning rate.

C. Performance of the Proposed Method with Standard Data

To validate the overall performance of our proposed method, we will demonstrate the detection and classification effect of the proposed method from three aspects: classification performance in various situations, response speed, robustness. Performance comparison with common machine learning methods are indispensable. In this section, we first test the classification performance of our method with data obtained in a standard system.

For clarity, the accuracies and recall rates of the proposed method for the classification of five types are calculated and depicted in Table III. The overall classification accuracy is 98.28%, and the classification accuracy for each type is higher than 97.4%. This result shows that our proposed method is capable of classifying faults with quite high accuracies.

TABLE III
CLASSIFICATION RESULTS OF THE PROPOSED METHOD FOR DIFFERENT FAULT TYPES

Fault type	Accuracy (%)	Recall (%)
Single-phase-G	98.18	98.18
Two-phase	98.04	98.04
Two-phase-G	97.76	97.76
Three-phase-G	97.48	97.48
Non-faulty	99.93	99.93
(Average)	98.28	98.28

Further, the classification performance of the proposed method is compared with that of the common machine learning algorithms including support vector machine (SVM), decision tree (DT), K nearest neighbor algorithm (KNN), random forest (RF), linear regression (LR), naive bayes algorithm (NB), fully connected network (FCN) and convolutional neural

network (CNN). The first six methods belong to the traditional classification algorithm, while the latter two are the neural network end-to-end classification algorithm. To compare the performance of various classifiers more comprehensively, we use receiver operating characteristic (ROC) curve to measure the classification effect, which is a comprehensive index that can best reflect the overall performance of a classifier in classification problems [30]. The horizontal axis of ROC curve represents false positive rate (FPR), while the vertical axis represents true positive rate (TPR). Formulas of the two are as follows:

$$\text{FPR} = \frac{\sum_{i=1}^n \text{FP}_i}{\sum_{i=1}^n (\text{FP}_i + \text{TN}_i)} \quad (10)$$

$$\text{TPR} = \frac{\sum_{i=1}^n \text{TP}_i}{\sum_{i=1}^n (\text{TP}_i + \text{FN}_i)} \quad (11)$$

where n is the number of fault types, T/F means true or false, P/N means positive or negative, and TP_i/FP_i denotes the TP/FP of the i th type. So FPR represents the proportion of real negative samples with respect to all negative samples in positive-predicted samples. Similarly, TPR represents the proportion of real positive samples with respect to all positive samples in positive-predicted samples. By setting different thresholds for softmax output, we get different (FPR, TPR) points which constitute the ROC curve. One of the great advantages of the ROC curve is that when the distribution of positive and negative samples changes, the curve's shape remains basically unchanged. Therefore, this evaluation index can not only reduce the interference brought by different testing sets, but also measure the performance of a model more objectively. Further, the area under curve (AUC) is calculated in order to quantify classification performance. The results are shown in Fig. 8.

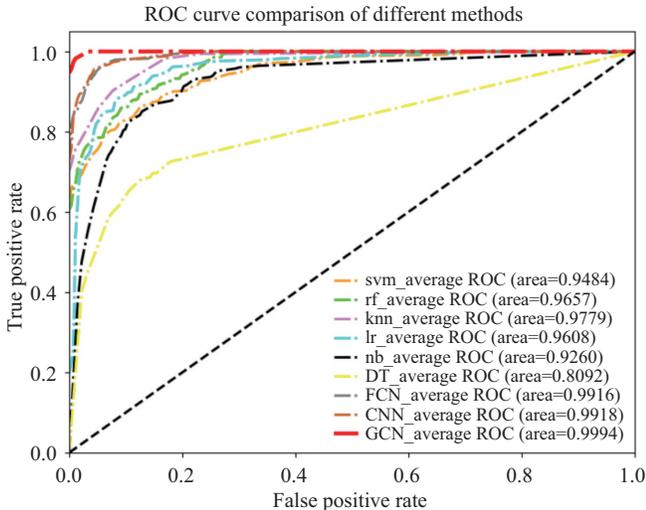


Fig. 8. Comparison of ROC curves.

According to the definitions of TPR and FPR, the ideal goal should be $\text{TPR}=1$ and $\text{FPR} = 0$. Moreover, the AUC of the ideal goal is 1.0. In other words, the closer a ROC curve is to the point (0, 1), the better the classification performance will be. We can tell from Fig. 8 that the ROC curve of GCN

is the closest to ideal classification goal ($\text{AUC} = 0.9994$). This result shows that the proposed method not only has high classification accuracy and recall rate, but also has remarkable comprehensive performance.

D. Performance of the Proposed Method With Renewable Energy Generation Integration

In order to simulate the operation of a real transmission grid, more environmental factors need to be considered. New energy power generation has increasingly become a hot spot in the industry. In view of the fact that more and more renewable energy power generations are connected to power grid, we add the renewable energy module to IEEE 39-bus system to simulate this situation.

We introduce a wind turbine into the IEEE 39-bus system to simulate the renewable energy generation integration, as shown in Fig. 9.

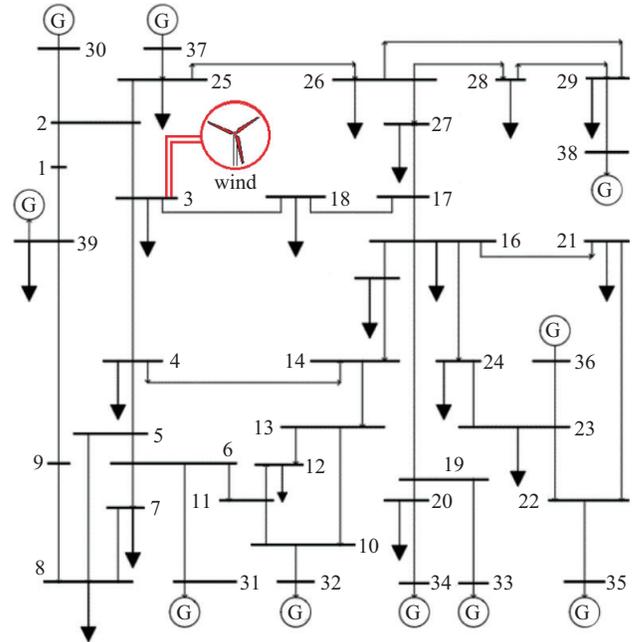


Fig. 9. Topology of IEEE39 transmission system with a wind turbine.

The selected wind turbine is a PSCAD-based calculation model [31], and we connect it to the No. 3 bus of the IEEE 39-bus system. We set the fault parameters as before, and get 3500 samples for testing the generalizability of the trained model. Fig. 10(a) and (b) represent the voltage waveforms of partial nodes under the two-phase short circuit fault. It can be seen that the characteristics of the fault data before and after the wind turbine integration are apparently different.

The classification accuracies of five types of fault data with the wind turbine are shown in Table IV below. We can see that when the characteristics of fault data become complicated due to the wind turbine integration, the well-trained model can still identify the fault with an averaged accuracy of 97.68%.

In addition, we depict the loss and accuracy curves of the training set and the testing set to verify that the model is less susceptible to over-fitting. According to the curves shown Fig. 11(a), the loss of the testing set does not increase and

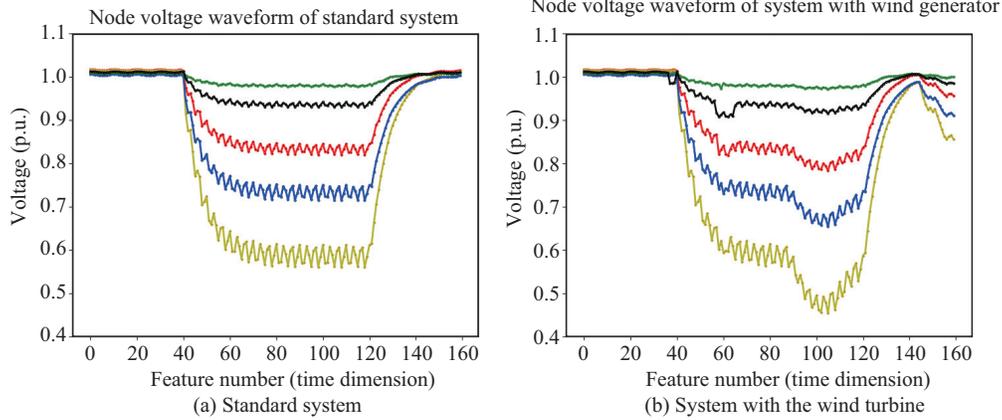


Fig. 10. Fault waveforms before and after the wind turbine integration.

TABLE IV
CLASSIFICATION ACCURACY (%) OF THE PROPOSED METHOD IN THE PRESENCE OF WIND TURBINE GENERATOR

Fault type	Standard system with wind turbine
Single-phase-G	97.48
Two-phase	97.20
Two-phase-G	97.20
Three-phase-G	96.64
Non-faulty	99.86
(Average)	97.68

remains very low in the later stage of training convergence, indicating that the model is not subjected to over-fitting [32].

E. Performance of the Proposed Method with Bad Data

Data measurement and acquisition usually brings lots of bad data in the real power grid. Therefore, we add some bad data to the standard fault data to further test the performance of the model.

Three types of bad data are considered in our paper:

1) Inaccurate measuring is simulated by multiplying standard measurements with a random number ranging from 0.75 to 1.25 and is set as 1% of the total sampling data.

2) Asynchronous sampling is simulated by selecting 5% of all PMUs and randomly moving the measurements forward or backward n sampling values. ($n \in [1, 5], n \in Z$).

3) Data loss is simulated by arbitrary discarding sampling points and is set as 1% of the total sampling data.

We add the three types of bad data to the original sample set, and get 11900 new samples which are still divided into training set and testing set at a ratio of 7:3.

Figures 12 and 13 represent the voltage waveforms of 39 nodes before and after adding bad data under the single-phase ground fault. It is obvious that the waveform of fault data after adding bad data is more complicated. As can be seen from Table V, the averaged detection and classification accuracy of fault samples with bad data is still up to 96.71%. Results of the testing set indicate that the proposed approach has good ability of bad data tolerance.

TABLE V
CLASSIFICATION ACCURACY (%) OF THE PROPOSED METHOD IN THE PRESENCE OF BAD DATA

Fault type	Standard fault samples	Fault samples with bad data
Single-phase-G	98.18	96.86
Two-phase	98.04	96.78
Two-phase-G	97.76	95.24
Three-phase-G	97.48	96.64
Non-faulty	99.93	98.04
(Average)	98.28	96.71

Similarly, the loss and accuracy curves of training and testing are depicted in Fig. 14(a) and (b). We can see that

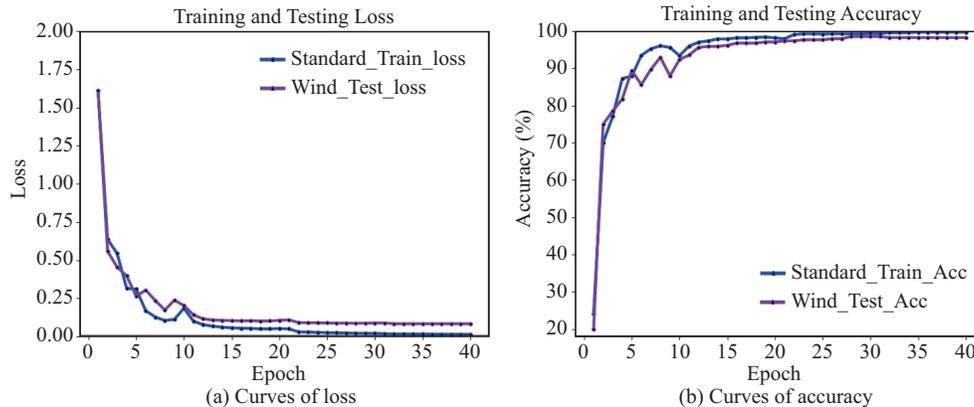


Fig. 11. Curves of training and testing loss and accuracy.

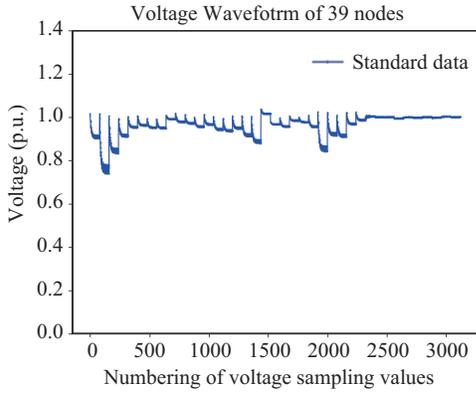


Fig. 12. Voltage waveforms of 39 nodes, “0~3120” on the x-axis represents the number of 39 (nodes) \times 80 (number of voltages).

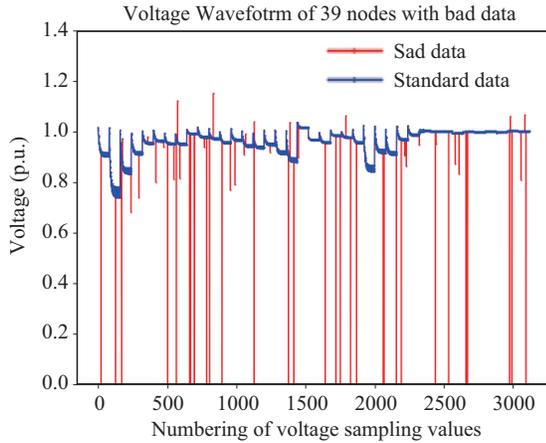


Fig. 13. Voltage waveforms of 39 nodes with bad data.

bad data do not cause the proposed method to overfit.

F. Performance of the Proposed Method with not All Buses Measured

In reality, it is not practical that all the buses are measured. Therefore, we must verify the effectiveness of the proposed method when not all the buses are measured.

Considering the real situations, we divide the cases where not all the buses are measured into three categories:

- 1) The proportion of measurable nodes is high.
- 2) The proportion of measurable nodes is low, and the measurable nodes are clustered.
- 3) The proportion of measurable nodes is low, and the measurable nodes are discretely distributed.

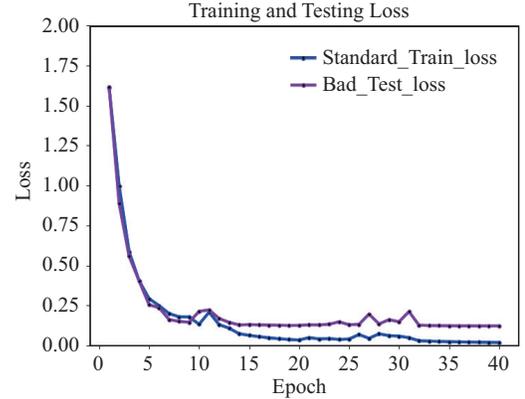
Three new datasets are generated based on the original 11900 samples, and experiments are implemented respectively. All the results are shown in Table VI.

Firstly, it can be seen that when the measurable nodes account for the majority, GCN can still achieve better performance compared with existing machine learning techniques. The adjacency matrix in this case no longer represents the neighborhood information of a few unmeasured nodes.

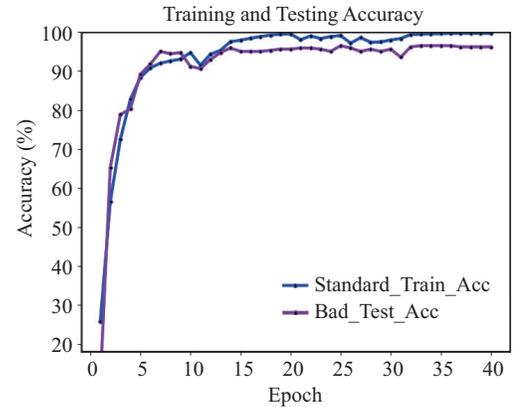
In addition, the 1st to 9th, 25th, 30th, 37th and the 39th nodes are selected as clustered measurable nodes. Obviously, the 13 measurable nodes form a local graph which is depicted in

TABLE VI
AVERAGED CLASSIFICATION ACCURACIES (%) OF THE METHODS WHEN NOT ALL THE BUSES ARE MEASURED

Name of methods	13 buses, discrete	13 buses, clustered	26 buses	39 buses
SVM	87.05	88.11	90.47	91.76
RF	93.78	94.96	95.58	96.35
KNN	86.22	85.99	92.16	91.94
FCN	86.89	85.43	93.28	95.97
CNN	95.27	95.23	96.73	97.43
GCN	95.18	96.36	97.48	98.28



(a) Curves of loss



(b) Curves of accuracy

Fig. 14. Curves of training and testing loss and accuracy.

Fig. 15. As can be seen from Table VI, the performance of GCN is also better in this situation.

At last, a small number of measurable nodes can also be discretely distributed in the topology, as shown in Fig. 16. And this situation makes the adjacency matrix a diagonal matrix, which further makes GCN unable to extract the spatial relations between the data. In this case, GCN is equivalent to an ordinary neural network, so it does not fail to detect and classify faults but cannot show its advantages. Table VI shows that the averaged accuracy of GCN is not optimal, which confirms the above inference.

G. Response Speed of the Proposed Method

Transient faults in power system often cause great damage in a short time, so the response speed of fault identification ought to be guaranteed. The previous results are obtained when we input 80 fault sample values (the fault duration

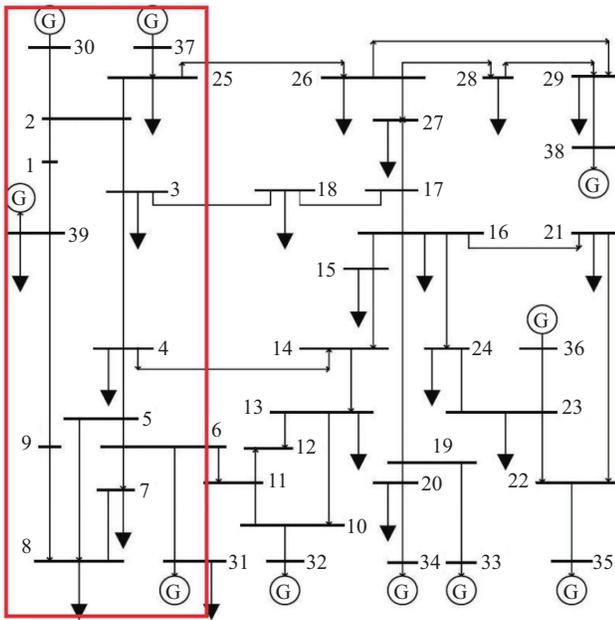


Fig. 15. Topology of IEEE39 transmission system with 13 clustered measurable buses.

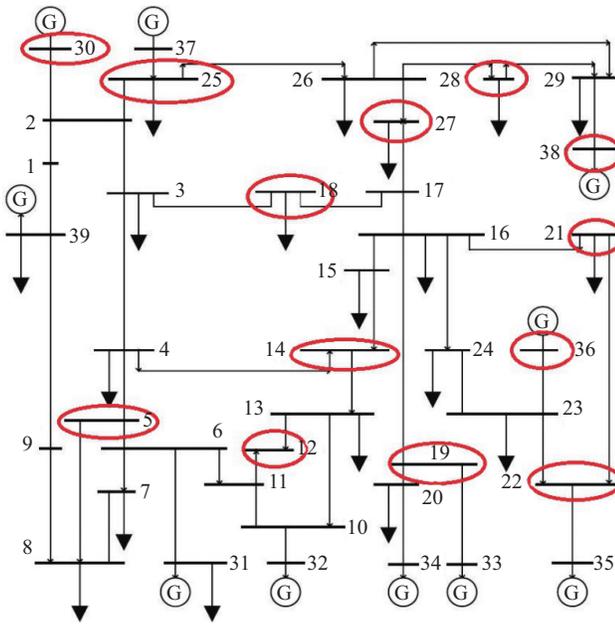


Fig. 16. Topology of IEEE39 transmission system with 13 discrete measurable buses.

within 0.1 s). For the sake of verifying the sensitivity of the proposed model, we try to identify the faults in a shorter time. Besides, under the restriction of the data acquisition equipment, the sampling frequency is not very high in practice. Thus experiments on the sensitivity of the model to faults at different sampling frequencies is supplemented. The results are depicted in Fig. 17.

First of all, the classification accuracy increases as the sampling frequency increases. This is in line with our expectation since more sampling values can be obtained by selecting a higher sampling frequency under the same sampling interval.

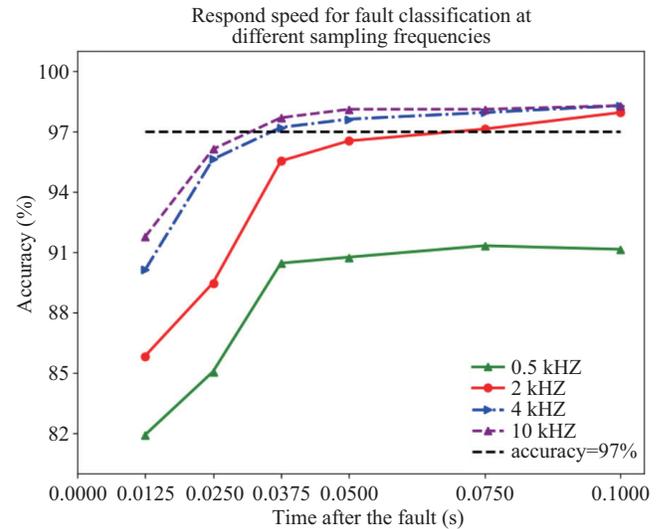


Fig. 17. Response speed for fault identification at different sampling frequencies, “0” point on the x-axis represents the moment of the fault occurrence.

Obviously, the more sample values of specific fault types can provide the richer transient characteristics of faults. In contrast, sampling at a relatively low frequency [33] will lead to insufficient data integrity.

Besides, what also satisfies us is that when sampling values within 0.0375 s after the fault occurs are taken as the model input, our method still achieves a high accuracy (above 97%), as long as the sampling frequency is not too low [33]. Considering the fact that accuracies under the common sampling frequencies are all above 97%, such response speed is quite satisfactory.

H. Robustness of the Proposed Method

It is essential to ensure that the method used to detect and classify faults can withstand noise. Noise in power transmission system refers to data fluctuation caused by load fluctuations or other uncontrollable events. We further compare the performance of the proposed method with some existing methods in the presence of noise.

Gaussian white noise is added to the fault data to test the robustness of the proposed method. The signal noise ratio (SNR) [34] of the data is 15 dB, 20 dB, 25 dB, 30 dB, 35 dB and 40 dB respectively. Other fault parameters remain the same as introduced in Table I. For comparison, we also design detection and classification networks of SVM, FCN and CNN. Radial basis function (RBF) is selected as the kernel of SVM [35], and 5-fold cross-validation and grid search are used to determine the appropriate values of the parameters γ and C . Finally, γ is set to 0.05 and C is set to 10. The structure of FCN has three hidden layers with fully connected neurons and Relu is selected as the activation function. After constant tuning and testing, the number of neurons in each hidden layer is set to 1600, 800 and 500 respectively. Besides, the CNN classifier we designed has six convolutional layers (kernel size = 3×3 , stride = 2, padding = 1), three maxpooling layers and three linear layers. LeakyRelu [36] is chosen as the

activation function. Further, dropout [37] mechanism is also used in the CNN classifier. The evaluation results are depicted in Fig. 18.

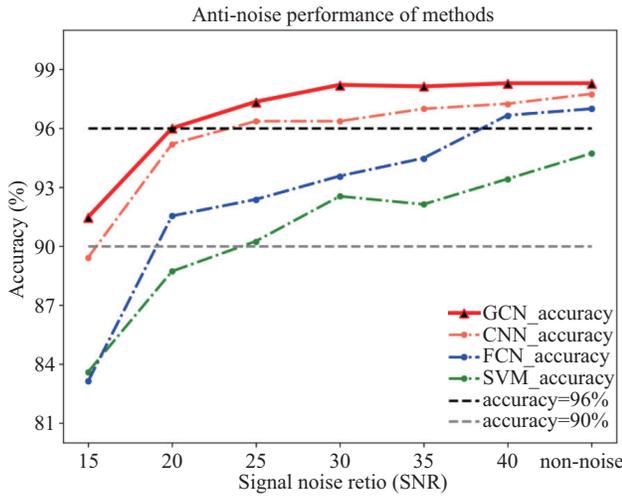


Fig. 18. Robustness for fault identification at different SNRs.

According to the results in the figure, we can see that when the SNR is above 20dB, the classification accuracy of GCN

model could still reach over 96%, which is a quite encouraging result. In order to illustrate the marvelous robustness of our method more convincingly, data waveforms under various scales of noise are shown in Fig. 19. The waveform curves of five colors in the subfigure respectively represent the data of five categories. Subfigures show that the raw data becomes very chaotic when the SNR is 25 dB, not to mention 15 dB. Moreover, this is only the voltage waveform of a single node of the transmission system. If the voltages of all nodes in the whole system are considered, the task will be much more difficult. Our model can still maintain a high classification accuracy in this situation. One explanation for this excellent anti-noise performance is the “aggregation” effect of graph convolution. As mentioned above, one of the core functions of graph convolution network is the aggregation of nodal features, which contains spatial information. The process of aggregating features offsets some effects of noise. From the above discussion, the high robustness of the model is verified.

I. Additional Experiments

1) Impedance Boundary of High Impedance Problem

The research of high impedance fault is a big challenge for power system. However, we do not discuss high impedance faults too much.

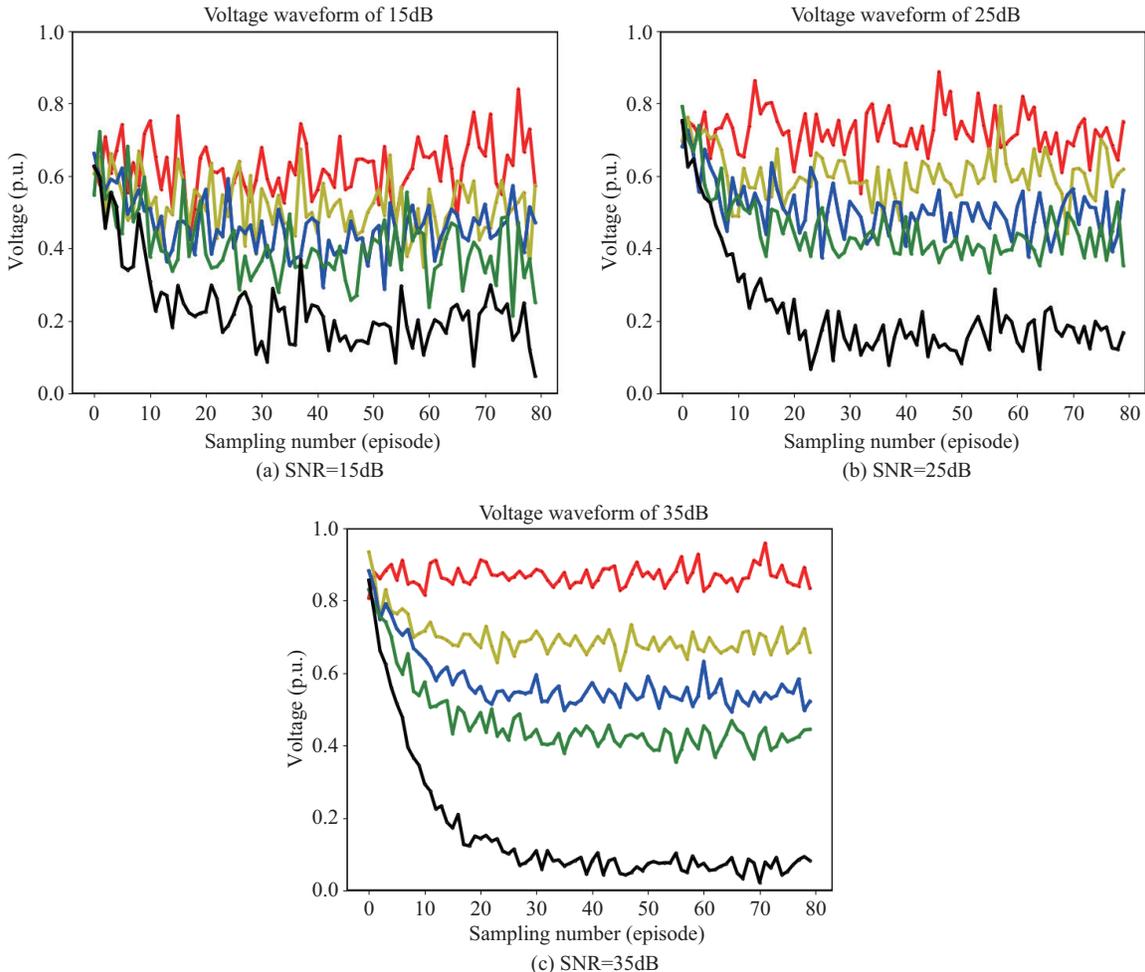


Fig. 19. Data of different SNRs.

Firstly, high impedance faults mainly occur in distribution networks (15 kV–25 kV), and power transmission systems with higher voltage level have a low probability of occurrence of high impedance fault [38]. For instance, in the case of high voltage levels, the grounding medium may be broken down when a high impedance ground fault occurs and then the high impedance fault will become a low impedance fault.

Secondly, the high impedance fault cannot be identified only through the features of nodal voltage.

As shown in Fig. 20(a) and (b), the nodal voltage waveform of single phase ground fault with resistance of 1 ohm is very similar to that of three-phase short-circuit fault with resistance of 100 ohm. And when the fault resistance is 300 ohm, the nodal voltage waveform of three phase short circuit fault even tends to the normal operating condition, as depicted in Fig. 20(c). The above conditions make it very difficult for only data-driven methods to accurately identify the fault types. In general, high impedance problems require many effective features such as the functional relationship between fault resistance and voltage variation before we can use these features to realize big data level identification.

Thirdly, the proposed approach starts from the perspective of mass data processing in the power grid. In fact, if we want to realize the detection and classification of various

fault conditions (including high impedance fault conditions and etc.), we need to combine the proposed method with the traditional protection theory to form a complete fault identification system [38]. The advantage of our method lies on that we undertake the role of data analysis when the amount of data in the power grid is huge, so that a concise and clear conclusion can be drawn from the overall analysis.

In order to determine the impedance boundary that our method can identify the five types of faults, we added extra experiments, the results of which are depicted in Fig. 21.

At total of 3500 samples are simulated and tested (700 samples for each fault resistance). It can be seen from Fig. 21 that the highest fault impedance of the sample that our detection and classification model can classify with an accuracy rate of not less than 95% is 55 ohm. In addition, our detection and classification model can identify fault samples with a fault impedance of 63 ohm under the condition that the accuracy rate is not less than 90%.

2) Validity of Adjacency Matrix

In this paper, we use the GCN-based method to detect and classify power system transient faults and its main advantage is the explicit extraction of spatio-temporal relations between data. Further, we add comparative experiments to verify the effective role of spatial information on fault detection and

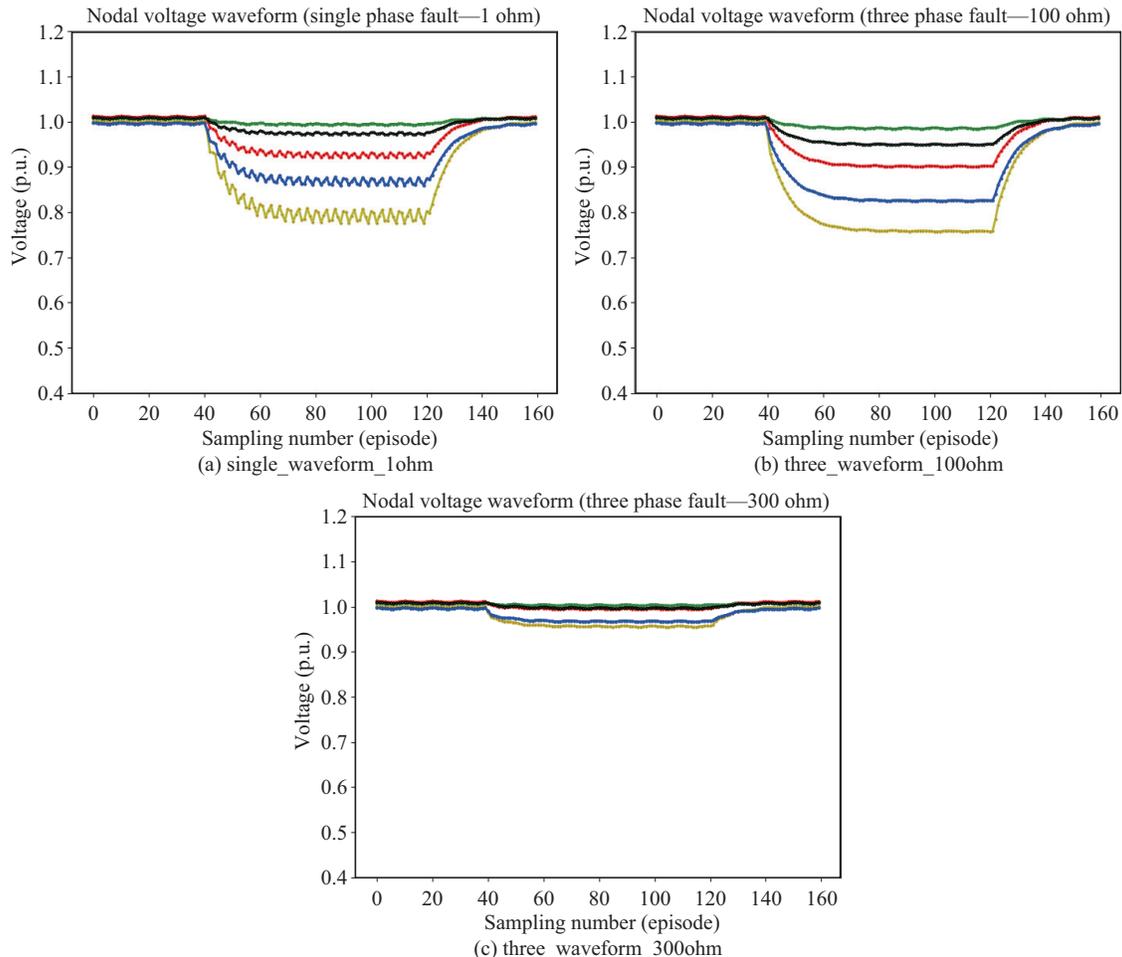


Fig. 20. Nodal voltage waveforms under different fault impedances.

TABLE VII
CLASSIFICATION ACCURACIES (%) OF GCN BASED ON DIFFERENT MATRICES

Matrix type	Only added in the first layer	Added in the first two layers	Added in the first three layers
Gaussian matrix	96.47	96.18	95.29
Uniform matrix	85.88	72.83	72.55
All-ones matrix	75.91	75.07	73.95
Identity matrix	96.76	96.76	96.76
Unweighted adjacency matrix	97.87	97.65	98.23
Weighted adjacency matrix	97.76	97.76	98.28

classification tasks.

In essence, the difference between GCN and general neural networks lies on the adjacency matrix which is used to represent topological information. Therefore, we replace the weighted adjacency matrix A with different matrices (with the same dimension as A) in the GCN framework, and retrain the model to compare the detection and classification results. The results are shown in Table VII.

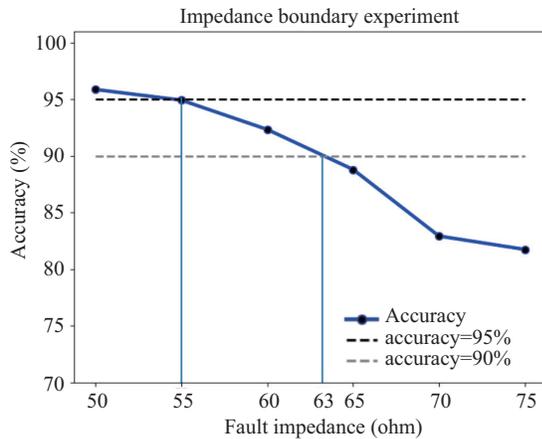


Fig. 21. Impedance boundary experiment.

The weighted adjacency matrix used in previous experiments is replaced by different matrices, including the standard Gaussian distribution matrix whose elements follow standard normal distribution [39], the standard Uniform distribution matrix [40], all-ones matrix (the matrix where all the elements are 1), Identity matrix and unweighted adjacency matrix whose elements are only “0” and “1”. It can be seen from the table that, non-adjacency matrices cannot represent the true and accurate spatial information of the transmission topology, thus the accuracy of detection and classification decreases. Adding such incorrect matrices to more GCN layers would reduce the accuracy even more. Moreover, we can see that different matrices have different negative effects on the accuracy of the model. However, adding correct adjacency matrix in a GCN layer means that the aggregation of nodal features and transform of fault information are implemented in this layer, so the accuracy is the highest compared with adding other matrices. Besides, the unweighted adjacency matrix only contains topological structure information but no parameter information [41], while GCN still achieves excellent accuracy. In theory, we reckon that edge weights can help the adjacency matrix aggregate nodal features more accurately in model training. But the results show the fault classification network is not very sensitive to edge weights.

The above experiments prove that explicitly extracting the spatio-temporal relations between nodal data helps to improve the accuracy of transient fault detection and classification, and the adjacency matrix is the key factor.

V. CONCLUSION

This paper presents a novel method for the detection and classification of power transient faults. Considering electric power data is a kind of spatio-temporal data, we regard the transmission line topology as a graph, so as to construct a graph classification model. Firstly, we propose a method for defining nodes and edge weights in the power grid topology. Secondly, we embed the topology information into the network so that the data of a single fault sample contains both temporal relationship and explicit spatial information, which provides more prior knowledge for the task and helps to improve the performance of the classifier. Experimental results on various situations show that the proposed method can distinguish several kinds of transient faults with high accuracies and strong generalizability. Further, the proposed method still shows sensitive and stable performance in the evaluation of response speed and robustness. We hold that the introduction of GCN is of great significance to the safe and stable operation of transmission system and even to the whole power system.

However, the graph convolution method introduced in this paper is a spectral convolution which has a solid theoretical foundation but poor flexibility. For example, once the adjacency matrix of a graph is determined, the structure of the graph is fixed, so dynamic grid structure cannot be dealt with. In addition, the edge weight mentioned in this paper must be more significant for fault location. We will consider using dynamic graph NN to solve the fault detection, classification and location of dynamic power grid, which ought to be a more meaningful work.

REFERENCES

- [1] D. X. Zhang, X. Miao, L. P. Liu, Y. Zhang, and K. Y. Liu, “Research on development strategy for smart grid big data,” *Proceedings of the CSEE*, vol. 35, no. 1, pp. 2–11, Jan. 2015.
- [2] X. He, Q. Ai, R. C. Qiu, W. T. Huang, L. J. Piao, and H. C. Liu, “A big data architecture design for smart grids based on random matrix theory,” *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 674–686, Mar. 2015.
- [3] Z. Ling, D. Zhang, R. C. Qiu, Z. Jin, Y. Zhang, X. He, and H. Liu, “An accurate and real-time method of self-blast glass insulator location based on faster R-CNN and U-net with aerial images,” *CSEE Journal of Power and Energy Systems*, vol. 5, no. 4, pp. 474–482, Dec. 2019.
- [4] M. K. Neyestanaki and A. M. Ranjbar, “An adaptive PMU-based wide area backup protection scheme for power transmission lines,” *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1550–1559, May 2015.

- [5] Z. X. Li, X. G. Yin, Z. Zhang, and Z. Q. He, "Wide-area protection fault identification algorithm based on multi-information fusion," *IEEE Transactions on Power Delivery*, vol. 28, no. 3, pp. 1348–1355, Jul. 2013.
- [6] J. J. Song, E. Cotilla-Sanchez, G. Ghanavati, and P. D. H. Hines, "Dynamic modeling of cascading failure in power systems," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2085–2095, May 2016.
- [7] G. Khazal and A. Zamyatin, "Feature engineering for Arabic text classification," *Journal of Engineering and Applied Sciences*, vol. 14, no. 7, pp. 2292–2301, 2019.
- [8] W. Sun, G. A. Yang, Q. Chen, A. Palazoglu, and K. Feng, "Fault diagnosis of rolling bearing based on wavelet transform and envelope spectrum correlation," *Journal of Vibration and Control*, vol. 19, no. 6, pp. 924–941, Apr. 2013.
- [9] Z. G. Liu, Z. W. Han, Y. Zhang, and Q. G. Zhang, "Multiwavelet packet entropy and its application in transmission line fault recognition and classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 11, pp. 2043–2052, Nov. 2014.
- [10] A. Jamehbozorg and S. M. Shahrtash, "A decision-tree-based method for fault classification in single-circuit transmission lines," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2190–2196, Oct. 2010.
- [11] A. Reციუი, B. Benseghier, and H. Khalfallah, "Power system fault detection, classification and location using the K-nearest neighbors," in *Proceedings of the 4th International Conference on Electrical Engineering (ICEE)*, 2015, pp. 1–6.
- [12] K. J. Chen, J. Hu, and J. L. He, "Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1748–1758, May 2018.
- [13] B. W. Guo, L. Li, and Y. Luo, "A new method for automatic seismic fault detection using convolutional neural network," presented at 2018 *SEG International Exposition and Annual Meeting*, Anaheim, California, 2018, pp. 1951–1955.
- [14] X. Shi, R. Qiu, Z. N. Ling, F. Yang, H. S. Yang, and X. He, "Spatio-temporal correlation analysis of online monitoring data for anomaly detection and location in distribution networks," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 995–1006, Mar. 2020.
- [15] H. S. Yang, R. C. Qiu, X. Shi, and X. He, "Unsupervised feature learning for online voltage stability evaluation and monitoring based on variational autoencoder," *Electric Power Systems Research*, vol. 182, pp. 106253, May 2020.
- [16] Z. Zhang, D. Zhang, and R. C. Qiu, "Deep reinforcement learning for power system applications: An overview," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 213–225, Mar. 2020.
- [17] Y. Ma, C. Huang, Y. Sun, G. Zhao, and Y. J. Lei, "Review of power spatio-temporal big data technologies for mobile computing in smart grid," *IEEE Access*, vol. 7, pp. 174612–174628, Dec. 2019.
- [18] H. S. Yang, R. C. Qiu, L. Chu, T. B. Mi, X. Shi, and C. M. Liu, "Improving power system state estimation based on matrix-level cleaning," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3529–3540, Sep. 2020.
- [19] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," arXiv:1312.6203, 2013.
- [20] K. Xu, W. H. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" arXiv:1810.00826, 2018.
- [21] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, Jul. 1997.
- [22] Z. H. Wu, S. R. Pan, F. W. Chen, G. D. Long, C. Q. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, Jan. 2021.
- [23] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv:1609.02907, 2016.
- [24] W. Q. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. L. Tang, and D. W. Yin, "Graph neural networks for social recommendation," in *Proceedings of World Wide Web Conference*, 2019, pp. 417–426.
- [25] C. Hung, J. Nieto, Z. Taylor, J. Underwood, and S. Sukkarieh, "Orchard fruit segmentation using multi-spectral feature learning," in *Proceedings of 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 5314–5320.
- [26] Q. M. Li, Z. C. Han, and X. M. Wu, "Deeper insights into graph convolutional networks for semi-supervised learning," arXiv:1801.07606, 2018.
- [27] J. Brownlee. (2018, Jul. 27). How to configure the number of layers and nodes in a neural network. [Online]. Available: <https://machinelearningmastery.com/how-to-configure-the-number-of-layers-and-nodes-in-a-neural-network/>.
- [28] Z. Y. Qin, D. Kim, and T. Gedeon, "Rethinking softmax with cross-entropy: neural network classifier as mutual information estimator," arXiv:1911.10688, 2019.
- [29] D. Kinga and J. Adam, "A method for stochastic optimization int," in *Conf on Learning Representations (ICLR)*, 2015.
- [30] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283–298, Oct. 1978.
- [31] S. K. Kim and E. S. Kim, "PSCAD/EMTDC-based modeling and analysis of a gearless variable speed wind turbine," *IEEE Transactions on Energy Conversion*, vol. 22, no. 2, pp. 421–430, Jun. 2007.
- [32] S. Yeom, I. Giacomelli, M. Fredrikson, and S. Jha, "Privacy risk in machine learning: analyzing the connection to overfitting," in *Proceedings of the IEEE 31st Computer Security Foundations Symposium*, 2017, pp. 268–282.
- [33] *Instrument Transformers— Part 9: Digital Interface for Instrument Transformers*, IEC 61869-9, 2016.
- [34] W. T. Li and M. Wang, "Identifying overlapping successive events using a shallow convolutional neural network," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4762–4772, Nov. 2019.
- [35] U. B. Parikh, B. Das, and R. Maheshwari, "Fault classification technique for series compensated transmission line using support vector machine," *International Journal of Electrical Power & Energy Systems*, vol. 32, no. 6, pp. 629–636, Jul. 2010.
- [36] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the 30th International Conference on Machine Learning*, 2013, p. 3.
- [37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.
- [38] A. Ghaderi, H. L. Ginn III, and H. A. Mohammadpour, "High impedance fault detection: a review," *Electric Power Systems Research*, vol. 143, pp. 376–388, Feb. 2017.
- [39] P. J. Forrester, "Log-gases and random matrices," *Journal of Statistical Physics*, vol. 134, no. 3, pp. 443–462, 2010.
- [40] W. Blajer, "A geometrical interpretation and uniform matrix formulation of multibody system dynamics," *ZAMM Journal of Applied Mathematics and Mechanics*, vol. 81, no. 4, pp. 247–259, Apr. 2001.
- [41] R. Seidel, "On the all-pairs-shortest-path problem in unweighted undirected graphs," *Journal of Computer and System Sciences*, vol. 51, no. 3, pp. 400–403, Dec. 1995.



Houjie Tong received the B.S. degree from Department of Electrical Engineering, North China Electric Power University in 2019. He is currently pursuing the M.S. degree in Electrical Engineering from Shanghai Jiao Tong University. His current research interests include machine learning and graph deep learning with their applications in power system.



Robert C. Qiu (F'15) received the Ph.D. degree in Electrical Engineering from New York University. He is currently the dean of the School of Telecommunications, Huazhong University of Science and Technology, and serves as a Professor in the Research Center for Big Data Engineering and Technologies, State Energy Smart Grid R&D Center, Department of Electronics and Electrical Engineering, Shanghai Jiao Tong University. He was with GTE Laboratories, Inc., Waltham and Bell Labs, Lucent Technologies. He was the Founder-

CEO and the President of Wiscom Technologies, Inc., manufacturing and marketing WCDMA chipsets. In 2008, he became a Professor at the Center for Manufacturing Research, Department of Electrical and Computer Engineering, Tennessee Technological University. He was named a fellow of IEEE in 2015 for his contributions to ultra-wideband wireless communications. His current research interests include wireless communication and networking, random matrix theory based theoretical analysis for deep learning, and smart grid technologies.



Dongxia Zhang received the M.S. degree in Electrical Engineering from the Taiyuan University of Technology, Taiyuan, Shanxi, China, in 1992 and the Ph.D. degree in Electrical Engineering from Tsinghua University, Beijing, China, in 1999. From 1992 to 1995, she was a Lecturer with Taiyuan University of Technology. Since 1999, she has been working at China Electric Power Research Institute. She is the co-author of four books, and more than 40 articles. Her research interests include power system analysis and planning, big data and AI applications in power systems. She is an Associate Editor of *Proceedings of the CSEE*.



Qi Ding is currently pursuing the master degree from Shanghai Jiao Tong University. His current research interests include model compression and machine learning application on Smart Grid.



Haosen Yang received the B.S degree from the South China University of Technology in 2017, and the M.S degree in Shanghai Jiao Tong University, 2019. He is pursuing the Ph.D. degree from the School of Electronics and Electrical Engineering, Shanghai Jiao Tong University. His research interests include voltage stability and state estimation of power grids, machine learning and data science.



Xin Shi (S'19) received the Ph.D. degree from the School of Electronics and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. He is now a Lecturer in the School of Control and Computer Engineering, North China Electric Power University, Beijing, China. His research interests include power system analysis, random matrix theory, and machine learning.