A multi-rate sampling PMU-based event classification in active distribution grids with spectral graph neural network

Mohammad MansourLakouraj [*], Mukesh Gautam, Hanif Livani, Mohammed Benidris

Department of Electrical and Biomedical Engineering, University of Nevada, Reno, Reno, NV 89557, United States of America

A B S T R A C T

Phasor measurement units (PMUs) are time-synchronized measurement devices that have been proliferated in transmission networks during the last two decades. Recently, there have been efforts to bring this technology to distribution grids for different applications such as three-phase state estimation, fault and event analyses, and phase identification. Streamed time-synchronized voltage and current phasor data can be used for events classification and region identification along distribution feeders to determine the type and location of events, which are important features of any fault and event detection, location, and isolation software. In this paper, the spectral theory-based graph convolution is used for event classification and region identification. The proposed model uses modified graph convolution filters to aggregate the regional multi-rate samples of PMU data, i.e., voltage magnitude and angles from several nodes. Besides these temporal data of the measured nodes, the physical configuration of the network containing edge features are given to the graph convolution network (GCN) to not only classify the event type, but also identify the affected region and location. The proposed graph-based method is tested on a standard test system with capacitor and distributed energy resources-related events, malfunction of voltage regulator, sudden load changes, and different types of faults. The results are compared with baseline methods, Chebyshev graph neural network (GNN), decision tree, logistic regression and K-nearest neighbor using the accuracy, recall, precision and F-1 score metrics. Furthermore, performance sensitivity analysis is carried out with respect to the number of installed PMUs, measurement noise level, size of available historical data, availability of network edge features, and different designs of GNN.

1. Introduction

1.1. Motivation

Phasor measurement units (PMUs) are time-synchronized sensor technologies that have been successfully deployed in transmission networks for variety of applications [1], from state estimation to faults and events detection, classification, and localization [2]. On the other hand, these measurement devices have not been implemented on distribution grids at large-scale yet. Historically, it has not been cost-effective to equip passive distribution feeders with such advanced sensors such as PMUs [3]. However, due to the increasing penetration of distributed energy resources (DERs) and flexible loads, distribution grids are becoming more active compared to the past [4]. The increasing penetration level of DERs brings new challenges for developing more effective and efficient monitoring, protection, and control methods for the evolving grids. Therefore, there is an increasing need for equipping distribution networks with more advanced sensors, such as PMUs, to better provide network and asset situational awareness, such as events analysis [5].

Events in distribution grids include short-circuit faults [6], capacitor bank (CB) switching [7], voltage regulator (VR) malfunctioning, sudden load changes [8], and distributed energy resource (DER) outages. PMUs capture voltage and current phasor data before, during, and after events which can be then used for events classification and region identification. This is a necessary tool to develop more accurate network and assets situational awareness in advanced distribution management system (ADMS) platforms. PMUs installed on different nodes may have different sampling rates from 30 to 120 samples-per-second (SPS), and therefore multi-rate data fusion is needed to simultaneously process all captured data while maintaining the spatio-temporal features of the streamed data. To the best of the authors' knowledge, multi-rate data fusion has recently been proposed for power system applications [9], but not in the context of event classification and location in active distribution grids.

Additionally, having a prior knowledge about the physical connectivity of the grid and distances between installed PMUs, provides

* Corresponding author.
E-mail addresses: mansour@nevada.unr.edu (M. MansourLakouraj), mukesh.gautam@nevada.unr.edu (M. Gautam), hlivani@unr.edu (H. Livani), mbenidris@unr.edu (M. Benidris).

spatial information of the data stream that can be used for more precise event classification and region identification. The physical distance can be represented as a graph matrix containing the edges' (grid lines') information between the nodes. Traditional classification algorithms usually do not incorporate the spatial characteristics of the sensor location and may only rely on the captured temporal features of PMU data stream. The above challenges have motivated us to focus on a graph-based data-driven event classification model considering multi-rate data stream provided by PMUs. An efficient event classification and region identification assists system operators to have better situational awareness and take proper corrective actions in case of severe events or faults.

1.2. Literature review

Active distribution grids are susceptible to experience different events which might cause damage to the assets. Hence, it is crucial to analyze the events, including identifying their type and location. A prompt event analysis framework can mitigate the adverse consequence of an event, leading to improvement in grid reliability and resilience. There have been many studies about fault and events analysis in power distribution networks using different methods. Recently, machine learning-based (ML) methods have been widely used for events and fault analysis in distribution networks. In [10], measured voltages are used as inputs to a support vector machine (SVM) classifier to specify the fault types. Authors of [11] propose the use of SVM for fault classification and deep neural network for fault location. Ref. [12] classifies faults and cyber attacks, and specifies the location of the occurred incident in distribution grids using deep neural network (DNN). In [13], a fault classification method is proposed via a latent structure learning under a multi-task learning algorithm to perform an accurate classification with limited labeled data. Authors of [14] propose a pattern recognition technique for fault detection in transmission grids by selecting the primitive patterns in recorded waveform signals. In this study, the attribute grammar is deployed for feature extraction as it can handle semantic and syntactic information [15]

Several studies have also focused on event classification, considering a wide variety of events such as short-circuit faults and physical failure of installed equipment in distribution grids. In [16], a PMU-based data-driven model is developed with autoencoders to recognize normal and disruptive events from each other. This study focuses on capacitor banks (CBs) and voltage regulators (VRs) outages, sudden load changes and switching of feeders. A multi-class SVM classifier is proposed for classification of CBs switching, distribution grid oscillation, and non-specific events [17]. However, regional faults and other events such as DER outages are not classified individually. A recent study in [18] uses k-Shape Stream method to cluster voltage sag and transformer tap changes as distribution grid events [18]. Authors of [19] leverage convolution neural network (CNN) models for event classification in power grids using PMU data. In this study, line faults and frequency events are considered as contingencies of the grid.

The majority of previously proposed methods for event and fault classification have focused on temporal aspects of the recorded data stream with less attention to the spatial relation among the measurement devices located on different nodes. Therefore, conventional ML algorithms may not be able to incorporate the topology of the grid containing the location of the PMUs as an additional set of information for enhancing the accuracy of the classification. However, authors of [20] have developed a CNN-based approach in the context of spectral graph theory, extracting the local and stationary information of the network via graph convolution layers. There are other studies that have incorporated Graph Neural Network (GNN) to include the spatial information of the installed PMUs and the connectivity configuration of the distribution network. The study conducted in [21] uses a graph convolution network (GCN) for fault location in distribution networks without considering the penetration of DERs. This study does not use

the temporal data and the events are limited to different types of faults. In [22], the fault location study is performed on active distribution systems using approximation in GCN filters. The study indicates the fast convergence and desirable accuracy of the proposed model.

The use of GCN is proposed for other power system problems as well. In [23], GCN is used for voltage stability assessment of power grids using spatial–temporal data features of captured data. GCN model is also used for transient fault classification and detection on transmission grids in [24]. Online short-term voltage stability prediction of power networks is performed using GCN and long-short memory networks (LSTM) in [25]. Another application of GCN is studied for cascading failure assessment and root cause analysis in [26]. Authors of [27] propose the use of GCN for residential load forecasting using geographical and temporal features of captured data from the field. A GNN model is proposed for classifying the events in transmission systems without knowing the physical configuration of the grid [28]. However, event localization is not studied using multi-rate sampling PMU data in [28]. To the best of our knowledge, there are very limited studies for the use of GNNs in distribution network studies with the penetration of DERs, and particularly for fault and events analysis, including region/location identification and classification.

1.3. Contributions

With respect to the existing literature, this paper proposes an applicable GNN algorithm for event classification and region identification in DER-penetrated distribution grids with limited recorded phasor data by PMUs. Therefore, the main contributions of this paper are as follows:

- Graph neural network (GNN) is proposed to capture spatial–temporal features of PMU data for event classification and region identification at sub-circuit level in distribution feeders. Potential events such as capacitor banks (CBs) switching, DER outages, sudden load changes, malfunctioning of tap changes in VRs, and all nodal short-circuit faults are considered and studied in a standard distribution test system. The proposed GNN-based method assists system operators to effectively determine events type and region leading to enhancement in network and assets situational awareness.

- The proposed method is developed to process multi-rate PMU data captured at different locations. Measurement devices with different sampling rates bring new challenges in simultaneous processing of data toward each specific application such as event classification. A multi-rate data fusion is appropriately adopted, making the proposed method applicable to real-world conditions where sensors with different sampling rates might be available. Moreover, the physical distance between PMUs and the configuration of the network is defined as a weighted adjacency matrix of a graph representing the distribution grid, and used as a prior knowledge in the classification problem.

- We use an efficient layer-wise linear filter of GCN, which is obtained by the first order approximation of spectral graph convolution [29]. Moreover, this model is not limited to the explicit parameterization given by Chebyshev polynomial, and is more efficient in terms of computation time and convergence rate. Also, the proposed linear GNN (L-GNN) gives a desired classification accuracy compared to Chebyshev-GNN (C-GNN), constructed with Chebyshev filter, and other conventional machine learning (ML) methods.

The rest of the paper is presented as follows: Section 2 presents and discusses L-GNN model. Section 3 shows different case studies and considerations. Sections 4 and 5 scrutinize the numerical evaluations and discussions, respectively. Finally, Section 6 presents a brief conclusion.

2. Graph neural network

A distribution network and installed PMUs on its nodes can be described as a graph with nodal and edge information. Using GNN, the feature signal of each node and edge characteristics are processed in graph convolution layers (GCNs). The GCN layers are the most important part of the proposed model. Two GCN types are defined such as spatial and spectral convolutions. Spatial models are stemmed from convolution networks by aggregating the information of neighboring nodes in the graph. Nevertheless, the traditional convolution network cannot be applied on signal of spatial-based features and it has lack of the theoretical basis. Also, the input data for convolution networks must be in two-dimension images or a regular one-dimensional data stream. However, spectral graph convolution is applicable in this situation due to its strong theoretical basis in graph signal processing domain [20]. In order to provide the background to the proposed solutions, the basics of graph theory and the mathematical formulations of spectral convolution considering re-normalization and approximation techniques are presented in the following sections.

2.1. Graph theory

A distribution network can be treated as an un-directed graph $G = (N, K, A)$, where N denotes the set of nodes, K indicates the set of edges, and A is the adjacency matrix of G. Assuming two nodes of the graph as n and m, a typical element of A is a_{nm}, as defined in (1) with weighted and zero amounts [30].

$$a_{nm} = \begin{cases} k_{nm} & \text{When the connection exists between } n, m \\ 0 & \text{When the connection does not exist between } n, m \end{cases} \quad (1)$$

The weighted adjacency matrix A inherently shows the spatial correlation between the sensors installed on different nodes. If the nodes have a strong correlation, the k_{nm} should have a large value. As an example, in distribution grids, the sensors installed on nodes with shortest distance would have a higher correlation for detecting an event or fault on their adjacent nodes. Therefore, we define a metric based on the distance between nodes given in (2).

$$k_{nm} = 1/d_{nm} \quad (2)$$

where d_{ij} indicates the physical distance between two nodes of the distribution grid. Due to the inverse relation of edge weights with the distance, close distances between nodes correspond to high correlation among the nodes, which is technically meaningful in power system event and fault studies.

The correlation among nodes can be defined based on the impedance of the lines, which is also indirectly relevant to the distance between the nodes. In this regard, different definitions for the weighted values of the adjacency matrix have been discussed in [24]. Additionally, the diagonal degree matrix of the graph is D, in which the diagonal values indicate the number of connected nodes to the corresponding node.

2.2. Spectral graph convolution network

Multiplying signal v, which is the node phasor voltage, with a parameterized filter h_θ is the spectral convolution on the graph in the Fourier domain and defined as follows:

$$h_\theta * v = U h_\theta U^T v \quad (3)$$

where U is the eigenvectors matrix of the normalized Laplacian graph represented in (4). The eigen decomposition of matrix L is represented in (5), where $\Lambda = diag\ (\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n)$ is the diagonal matrix with ordered positive eigenvalues.

$$L = I - D^{-0.5} A D^{-0.5} \quad (4)$$

$$L = U \Lambda U^T \quad (5)$$

In (3), the multiplication of the eigenvector matrix with the filter is computationally intensive. To tackle this challenge, h_θ, defined as the function of eigen values $h_\theta(\Lambda)$ considering the Fourier transforms, can be equated with the truncated Chebyshev polynomial terms as (6) up to Gth order [31]. This functional approximation stabilizes the training in filters [20]. Besides, vector θ' contains the Chebyshev coefficients. Equation (7) shows the re-scaled representation of the eigenvalues, where λ_{max} is the maximum eigenvalue of the Laplacian matrix.

$$h_{\theta'}(\Lambda) \simeq \sum_{g=0}^{G} \theta' T_g(\hat{\Lambda}) \quad (6)$$

$$\hat{\Lambda} = 2/\lambda_{max} \times \Lambda - I \quad (7)$$

The Chebyshev polynomials operating recursively are written as $T_g(y) = 2y T_{g-1}(y) - T_{g-2}(y)$. With respect to the filters' definition, the convolution of the signal and filter is modified as (8). This representation is driven by $(U \Lambda U^T)^g = U\ \Lambda^g\ U^T$ property. The scaled Laplacian matrix is demonstrated in (9).

$$h_{\theta'} * v \simeq U \sum_{g=0}^{G} \theta' T_g(\hat{\Lambda}) U^T v = \sum_{g=0}^{G} \theta' T_g(\hat{L}) v \quad (8)$$

$$\hat{L} = 2/\lambda_{max} \times L - I \quad (9)$$

Equation (8) has Gth order in Laplacian, showing that it is G-localized. In other words, it depends on maximum G neighboring nodes of the central node of the graph [20].

2.3. Approximation in graph convolutions

In (8), the G is set at 1 for making a linear approximation. The first two terms of the truncated Chebyshev polynomial expansion are defined as $T_0(y) = 1$ and $T_1(y) = y$. These terms are the linear function of Laplacian graph L. According to [29], this approximation reduces the chance of overfitting in training, and allows us to build a deeper model with less computational burden than conventional methods. This consideration is in contrast with the model proposed in [20], and motivates us to take advantage of this linear model for the event classification problem. Additionally, the linear representation of the convolution filter of the graph is simplified with $\lambda_{max} = 2$ and the approximation of (8) can be denoted as follows:

$$h_{\hat{\theta}} * v = \hat{\theta}(I + D^{-0.5} A D^{-0.5})v \quad (10)$$

The Chebyshev coefficients for the first and second terms are approximated with a single parameter $\hat{\theta}$. The range of the eigenvalues in $I + D^{-0.5} A D^{-0.5}$ is [0, 2], leading to instability and vanishing gradients in deep neural network. The work presented in [29] proposes a re-normalization technique to compensate the downside of (10). Equation (11) indicates the final expression obtained for the convolution in GNN.

$$h_\theta * v = \theta(\hat{D}^{-0.5} \hat{A} \hat{D}^{-0.5})v \quad (11)$$

where $\hat{A} = A + I$, and $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$.

Equation (11) can be extended with matrix $V \in R^{N \times M}$ containing time-series voltage information as (12). The temporal data have M features captured by PMUs (before, during and after the event). Matrix $\Omega \in R^{M \times Q}$ defines the parameter matrix of the filter with Q feature maps or filters. Finally, F is the matrix of convoluted signals.

$$F = (\hat{D}^{-0.5} \hat{A} \hat{D}^{-0.5})V\Omega \quad (12)$$

2.4. GNN structure

More specifically, the forward model inside two GCN layers is interpreted as (13). In (13), $W^{(0)} \in R^{M \times P}$ is the input for hidden weight matrix with P feature maps. Besides, $W^{(1)} \in R^{P \times Q}$ indicates the weight matrix of the hidden layer to the output.

$$F = \hat{A} ReLU(\hat{A} V W^{(0)}) W^{(1)} \tag{13}$$

GCN layers capture the weighted average of the information at adjacent nodes on the graph. The output of GCN layers is given to a layer for reshaping the dimension. Then the output of this layer σ is given to the softmax classifier as (14). ω and c are the parameters of the layer before the softmax classifier. GNN determines the label vector Y through (14), but the actual labels are predefined in set Z.

$$Y = softmax(\sigma(\omega F + c)) \tag{14}$$

In addition, the cross-entropy energy function is used as the loss function, represented in (15) [29]. In (15), i and j represent the size of the training set and the number of locational/regional events, respectively. The parameters in the entire GNN are optimized by the backpropagation algorithm in an iterative manner. The loss function is also minimized by updating the parameters considering the predicted Y and actual Z labels [32].

$$Loss = -\sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij} Ln(Z_{ij}) \tag{15}$$

2.5. L-GNN for event classification and region identification

The spectral GNN is developed for determining the event type and location based on the proposed framework in Fig. 1. The hybrid input data consist of multi-rate sampling PMUs phasor voltages and the physical configuration of the grid. More specifically, the configuration of the network is shown with the weighted adjacency matrix of a graph, known as the prior knowledge in this classification problem [32]. The simulation environment is developed to create different loading and PV generation scenarios, and then to simulate different types of events. After simulating each event under different loading and generation scenarios, PMUs are assumed to record the voltage phasor data on the nodes at which they are installed. The recorded stream of data for each scenario, created with a specific loading, PV generation, and event, are ordered in a PMU matrix V, as shown on the left side of Fig. 1 The stored data in each row of V is the time-series phasor data stream. The data stream contains pre-, during- and post-events voltage phasors recorded by PMUs on different nodes. As PMU devices might have different sampling rates, the stream of data in each row of V contains different number of samples. In order to place them in the measurement matrix V, the largest sampling rate determines the number of columns or features. For instance, an interval of 300 msec of pre-, during-, and post-events leads to 18 samples by PMU-A with 60 sample-per-second (SPS) rate, and 9 samples by PMU-B with 30 SPS. In order to simultaneously process the samples from PMU-A and PMU-B as the same measurement matrix, we put zero in time slots, in which the information are not available by the PMU-B due to the slower rate of sampling. Fig. 2 illustrates the schematic of data fusion for both PMUs in the measurement matrix V. As shown in Fig. 2, although PMU-A and PMU-B have different time instances, they can be ordered in the measurement matrix V according to their corresponding time frame. This method can be generalized for further multi-rate sampling PMU devices and even other sensors technologies. It should be noted that the zeroed out samples from the slower rate PMUs can be substituted using proper interpolation or data recovery algorithms [33]. Thereby, we build the matrix of data as $V \in R^{N \times M}$, where N indicates the node number of the distribution grid and M shows the temporal PMU raw data. Among all nodes N, P of them contains PMU information, indicating the number of installed PMUs. M is set with respect to the largest

sampling rate during the selected time period for the event analysis. To be more specific, the recorded features in each node and each time slot contain three-phase voltage phasor information, i.e., magnitude and angle, as $V_i^p = (Vm_1, Vm_2, Vm_3, Va_1, Va_2, Va_3)_i^p \in R^6$. Therefore, the PMU sample in each time slot contains 6 voltage magnitudes and angles. So, if the sampling rate of the PMU is defined as 18 per event time, we will have $18 \times 6 = 108$ features in the corresponding row of V.

Additionally, as it was discussed in Section 2.1, matrix A includes the spatial information of PMUs and the network features. In other words, matrix A contains the connectivity status and the distance metric of lines to consider the intensity of the correlation among the nodes. Combining these prior physical information with measured PMU data assists the algorithm in aggregating the most important information of neighboring nodes by spectral graph convolution layers [29]. Accordingly, input data, including the temporal measured voltages and the spatial correlation of nodes and multi-rate sampling PMUs, are given to the GCN layers, as indicated in Fig. 1. The GCN layers are followed by non-linear ReLU function. The mean pooling layer is also set after these layers to improve the classification accuracy. Also, the dropout layer is considered to enhance the performance of the model and to prevent overfitting during the training [34]. The last GCN layer gives the output in a flatten shape to the classifier layer determining the event region and type. It is to be mentioned that the Pytorch Geometric library is implemented for preparing the proposed model [35]. In this library, graph convolution filters for L-GNN and C-GNN are developed and known as GCNConv and ChebConv, respectively.

3. Case study and considerations

The proposed linear spectral GNN model is tested on the modified IEEE 34-bus system to identify the event region and type. This test grid has 35 buses including the source bus and is modified with PV generation units and CBs, as shown in Fig. 3. The aggregated power capacity of PVs on buses 810, 822 and 838 is 900 kW, which is equally distributed between them. The CBs located in 848 and 840 can generate 200 kVar reactive power individually, but the CB located near to bus 824 can provide 110 kVar power. The large-scale loads in 844 and 890 consume 405 and 450 kW, respectively. Two types of PMUs with fast and slow sampling rates are considered, indicated as PMU-A and PMU-B, respectively. The sampling rate of PMU-A is 60 SPS (i.e., 18 samples for 300 msec), and the slower rate PMU-B captures 9 samples for the same duration.

The required voltage phasor data sets for training the GNN algorithm are generated using OpenDSS software and Python under different events. A variety of single and multi-phase faults are created in different load buses. A portion of PV generations is tripped off during fault and the remaining units are assumed to have fault-ride through capability and remains connected during the fault. The temporary and permanent outages of PVs are simulated, and the series of data is recorded within the simulation time. Moreover, the switching of CBs and large-scale loads are simulated. These events can also last for a sudden short period (during event time slots) or they can remain as permanent incidents even in post-event period. The malfunction of VRs is considered with undesired sudden tap changes. The stream of the generated data includes before, during and post event periods. The voltage angles and magnitudes are stored by 8 PMUs with different sampling rates. To generate realistic data sets under all probable operational conditions, different load levels changing from 0.4 to 1.2 times of the base load and PV generations varying between 0 to 100% are taken into account during the simulation. Therefore, a more robust model can be proposed for real-world implementation by training it under various load and PV uncertainties.
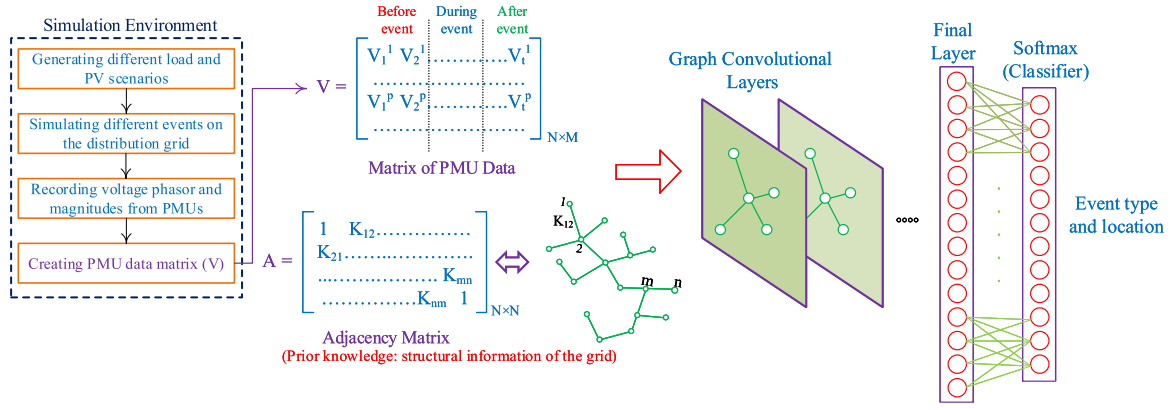
Fig. 1. The schematic of the proposed GNN model for event classification along with simulation flowchart of the training data problem.
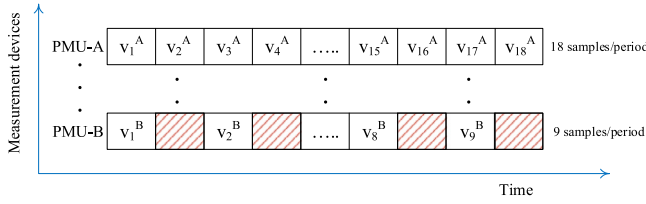

Fig. 2. Data fusion of PMUs with different multi-rate samplings.

4. Numerical analysis

For the sake of training and testing the proposed L-GNN, 5435 different event scenarios are generated. The scenarios are devides as follows: 70% of the scenarios are used to train the model; 15% of the scenarios are kept for validation; and 15% of scenarios is assigned for testing the model after completing the training process and tuning the hyperparameters using validation accuracy criteria. Different testing size can be considered, but the typical range is between 10%–30% [36]. The impact of testing size is also evaluated in the following sections. The constructed L-GNN model has two linear spectral graph convolution layers with 512 and 512 channels, followed by a layer with 256 hidden nodes. The activation function is a non-linear ReLU function and the dropout rate is adjusted to 0.5. The learning rate is set to be 0.0001 for the Adam optimizer. The mini-batch that groups the graph samples is set to 32. These parameters are adjusted with the parameter tuning in training process and observing the validation accuracy. The desirable classification accuracy is obtained using the adjusted parameters in an acceptable time. Comprehensive sensitivity analysis are presented in this regard to justify why these value are selected. The softmax classifier located in the output of L-GNN framework separates the events in 12 different classes. Each class represents an event, its location, and the region where it happens. For faults, the classes are defined based on the regions due to the closeness of the faulty nodes in each area. Three regions are shown with dotted lines in Fig. 3. These regions are divided based on the physical distance of the nodes. Table 1 specifies the label assigned to each event with respect to their locations and regions.

Furthermore, the performance of L-GNN is tested under different case studies. The proposed model is compared with other baseline models using statistical classification metrics. The optimal structure of base models are represented in the following section using the hyper-parameter tuning to demonstrate a fair comparison with the proposed model.

Table 1
Event labels, types and locations.

Labels	Event type	Nodal location	Region
1	CB switching	848 and 840	3
2	CB switching	824	2
3	Sudden load changes	844	3
4	Sudden load changes	890	2
5	PV outages	822	2
6	PV outages	838	3
7	PV outages	810	1
8	VR malfunction	814–850	1
9	VR malfunction	832–852	2
10	Fault	Load nodes in region 1	1
11	Fault	Load nodes in region 2	2
12	Fault	Load nodes in region 3	3

4.1. Hyperparameters tuning

To verify the competence of the proposed framework, the following efficient classifiers are opted for comparison. All the hyperparameters are adjusted using the same set of training data with several experiments.

1- Chebyshev GNN (C-GNN): This method implements CNN in the context of spectral graph theory, taking the advantage of fast localized convolution filters on irregular grids. The representation of the convolution operation on graph is given in (6) with Chebyshev polynomial of order G. Detailed descriptions of C-GNN is discussed profoundly in [20]. In this study, two GCN layers are considered, which is a similar structure to L-GNN to be a comparable graph-based method. In C-GNN, the order of Chebyshev polynomials G is set on 2 for the graph convolution layers. Also, other approximation techniques and re-normalization used for the proposed L-GNN are ignored. The Adam optimizer and learning rates are set similar to the L-GNN model. The softmax classifier is also added to the last layer of this model.

2- Decision tree (DT): This is a non-parametric learning algorithm, which can easily handle the multi-output problems [37]. Implementing this method, a decision tree is created using the branches as conjunctions of the features and leaves as the labels of the classes. The maximum depth of tree is evaluated and finally set on 10. This value is selected because the greater value than 10 for the tree depth does not improve the accuracy of the model in hyperparameter tuning process, as shown in Fig. 4.

3- K-nearest neighbors (KNN): It calculates the distances between the unknown measured samples and the training samples based on the K-closet neighbors [38]. The optimal K is analyzed by setting different K, and is tuned on 7, by which the misclassification error reaches to minimum, as shown in Fig. 5.

4- Logistic regression (LR): This algorithm is employed for multi-class classification based on a logistic function [39]. The $lbfgs$ optimizer and $L2$ penalty are selected by tuning the parameters. Also, the
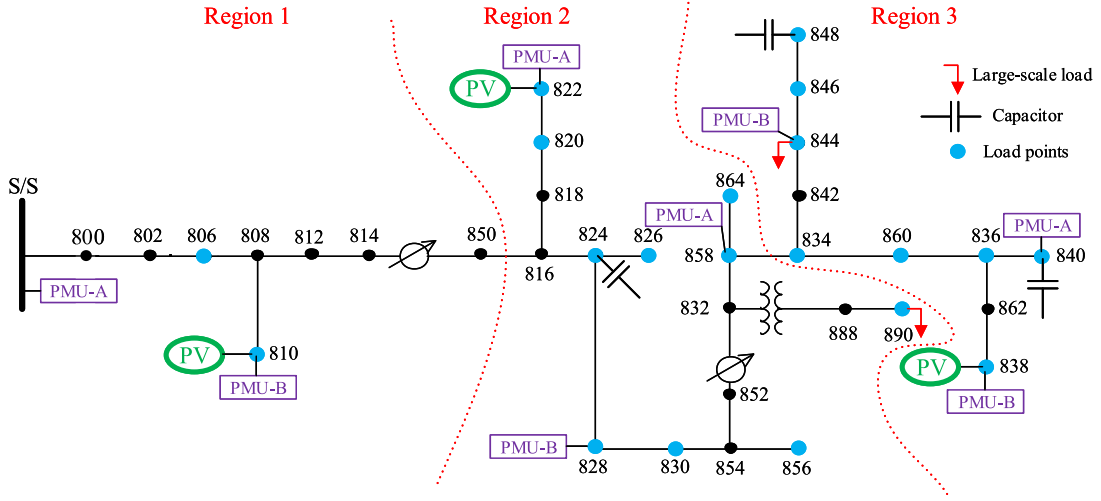
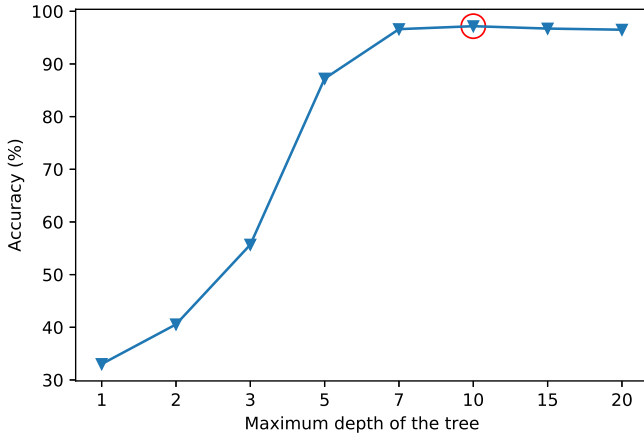**Fig. 3.** Schematic of modified IEEE 34-bus system.

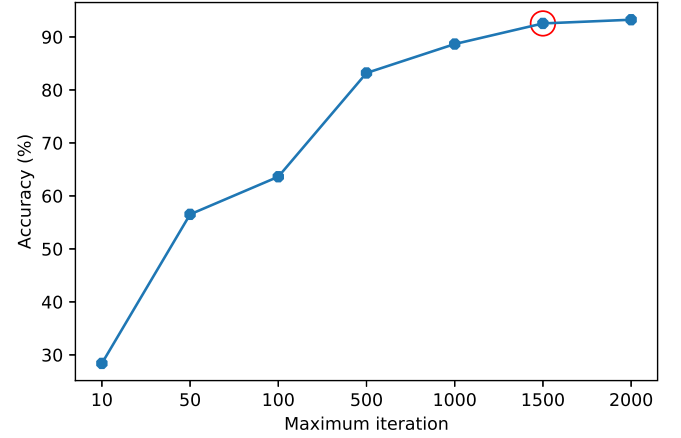**Fig. 4.** The process of hyperparameter tuning for DT.

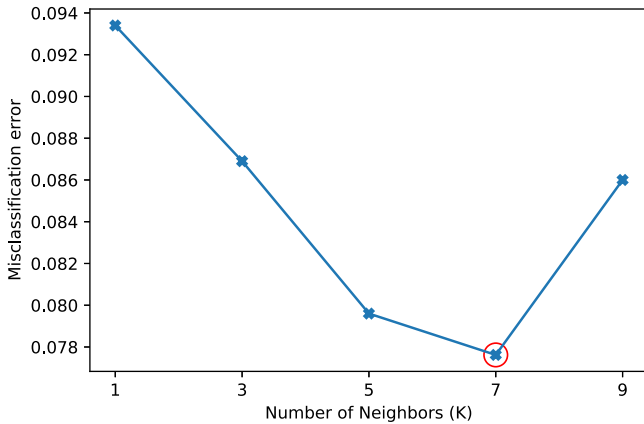**Fig. 5.** The process of hyperparameter tuning for kNN.

Fig. 6. The process of hyperparameter tuning for LR.

4.2. Performance of models on event classification

The performance of event classification using L-GNN is compared with different models during the testing process, as tabulated in Table 2. The testing accuracy of L-GNN is 98.34%, which is higher than the other methods. This desirable accuracy ensures that it can predict the event types and location more efficiently. To better evaluate the classification performance of the proposed model, three classification metrics such as precision, recall, and F-1 score are used based on the macro-average definition.

- Precision implies the classifier's ability to predict actual positives (APs) for the class l among all APs as well as false positives (FPs) predicted in the corresponding class. Then, the macro average of whole classes is calculated using (16).

$$Macro - Precision = \frac{\sum_{l=1}^{L} Precision_l}{L} = \frac{\sum_{l=1}^{L} AP_l/(AP_l + FP_l)}{L} \tag{16}$$

- Recall shows the efficacy of the classifier to predict APs in class l among whole APs plus false negatives (FNs) related to that class. The macro average of this metric for multiple classes is defined as (17).

$$Macro - Recall = \frac{\sum_{l=1}^{L} Recall_l}{L} = \frac{\sum_{l=1}^{L} AP_l/(AP_l + FN_l)}{L} \tag{17}$$

maximum iteration is adjusted to 1500. Fig. 6 shows that increasing maximum iteration does not significantly change the accuracy after 1500 iterations while it remarkably increases the computation time.

Table 2
The comparison of results.

Approach	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
$L - GNN$	**98.34**	**99.26**	**97.92**	**98.46**
$C - GNN$	98.28	98.59	97.59	98.36
DT	97.17	97.11	96.66	96.83
LR	92.55	93.65	89.52	90.74
kNN	92.24	91.52	90.41	90.88

Table 3
The comparison of results with adding noise.

Approach	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
$L - GNN$	**95.37**	95.35	**93.73**	**94.05**
$C - GNN$	95.15	**95.48**	93.08	93.99
DT	94.51	93.74	92.67	93.03
LR	91.72	92.74	88.90	89.58
kNN	90.49	90.69	88.56	89.18

- F_1 score is calculated by the weighted harmonic mean of recall and precision for each class. The macro-F_1, calculating the average of F_1 score of each class over the whole considered classes, is represented in (18).

$$Macro - F_1 = \frac{\sum_{l=1}^{L} Precision_l * Recall_l / (Precision_l + Recall_l)}{L}$$

(18)

Considering the proposed classification metrics, it is concluded that L-GNN gives better scores compered to baseline approaches. After that, the C-GNN and DT show desirable performances in classification of events. The presented scores shown in Table 2 are calculated by taking the average amount of four individual training and testing process results. To show more details regarding the classification accuracy for each event, the confusion matrices of L-GNN and KNN are compared and represented as Fig. 7. The confusion matrix illustrates the accurate and inaccurate event classification for each region and location. The y-axis shows the actual classes or events, which are shown in Table 1, and x-axis presents the predicted events with their corresponding regions. The L-GNN model works satisfactorily and identifies the location and types of 10 regional incidents and equipment malfunctions with 100% accuracy. However, KNN could just classify 3 events with 100% accuracy, and has difficulty in differentiating the incident location of specific events such as capacitor switching and load changes. For instance, KNN fails to predict 20.51% of the location of capacitor outages in region 1, and labels them as region 3. This algorithm also misclassifies 35.90% of load changes in region 3 as load events of region 2. This misclassification happens as KNN cannot capture the spatial correlation of the nodes to determine the region or location of the events. Moreover, regarding the identification of faulty regions, L-GNN fails to predict 22.95% of the faults occurring in region 1, and it classifies them as the faults of region 2. Additionally, an almost comparable misclassification is done by KNN for fault events, but still L-GNN works slightly better in classifying all possible regional faults. The failure in detecting faulty region usually occurs in distribution networks since the nodes are geographically close to each other, and the impact of short-circuit faults on one node can easily spread to adjacent regions.

4.3. Convergence rate and computation time

One of our main goals in this study is to demonstrate the efficiency of L-GNN compared to C-GNN in terms of convergence speed and computation time of training. As shown in Fig. 8, L-GNN has a faster convergence rate and reaches to the stable validation accuracy in 500 epochs. However, C-GNN is slower in attaining the desired accuracy. Additionally, because of the approximation considered in the filter of L-GNN, the computation time of training is remarkably lower, preparing the algorithm for real-time application in a shorter time. More explicitly, the computation time of L-GNN and C-GNN are 8.6 and 11.13 min for the first 50 epochs, respectively. This shows 29.41% increment in the computation time when using C-GNN model. The calculated computation time is related to the offline training process. However, during online testing, the results are obtained in a few milliseconds. Therefor, L-GNN has a higher convergence rate, less computation time in training process and as we discussed in previous section, it has a better classification performance than the other methods.

4.4. Performance of model with measurement noise

To examine the robustness of the proposed L-GNN under real-world situation, Gaussian noise with zero mean and a standard deviation as $10^{-SNR/20}$ is added to the measured data, with signal-to-noise rate (SNR) of 45 dB. Table 3 shows the accuracy and classification metrics for the proposed method and the baseline approaches under noisy situations. L-GNN demonstrates more robustness against the noisy data and performs better in other metrics in overall. C-GNN also shows a comparable robustness and satisfactory accuracy in the event classification. DT is also vulnerable with noisy data and experiences 2.66% reduction in accuracy compared to the first case without considering noises. LR and kNN detect the event type and region with very close and lower accuracy. The rest of the study in following sections is performed considering the noisy data.

4.5. Testing size impact

Initially the performance of L-GNN is shown by testing the trained model on 15% of the data. This means that 85% of the data is used in the training and validation processes. However, the massive number of recorded PMU data may indicate limited information regarding the type of the events. So, the training may need to be performed with limited number of accessible data, while the testing can be performed with larger data sets. To this end, different size of testing set is used as 40%, 70% and 90% of the event database. Fig. 9 shows better performance of both GNN-based methods compared to other methods in terms of accuracy for testing size of 15%, 40%, and 70%. However, when the test size is 90%, meaning that the size of available training data set is reduced significantly, LR performs better. Therefore, LR has more robustness when limited historical data set is available for training. In this study, the classification accuracy of kNN and C-GNN is more susceptible to the smallest size of training data set, calculated as 78.96% and 77.85%, respectively. To solve this problem, a data augmentation technique is a viable solution to expand the size of historical training data when limited number of events are recorded in the past [40]. This strategy, which is out of the scope of this study, enhances the performance of ML models with small training sets.

4.6. Different configuration of PMUs

In this part, the configuration of PMUs is changed to study the efficiency of the proposed GNN model under limited number of recorded PMU data. Three new sets are considered, indicating different configurations of the installed multi-rate PMUs. The location of PMUs is specified as A: [S/S, 844], B: [S/S, 828, 840, 844], C: [S/S, 810, 822, 828, 840, 844]. These configurations are heuristically selected to make sure that different regions of the network can be observed by the operators. Also, each set contains equal numbers of fast PMU-A and slow PMU-B devices. Optimal placement of PMUs is beyond the scope of this paper.

The average test accuracy of L-GNN is tabulated in Table 4 for different sets of PMUs under four individual experiments. The accuracy of the classification is 87.07% and 93.75% with just two and four multi-rate sampling PMUs, respectively. Hence, L-GNN performs satisfactorily when limited spatial information is captured in the grid.
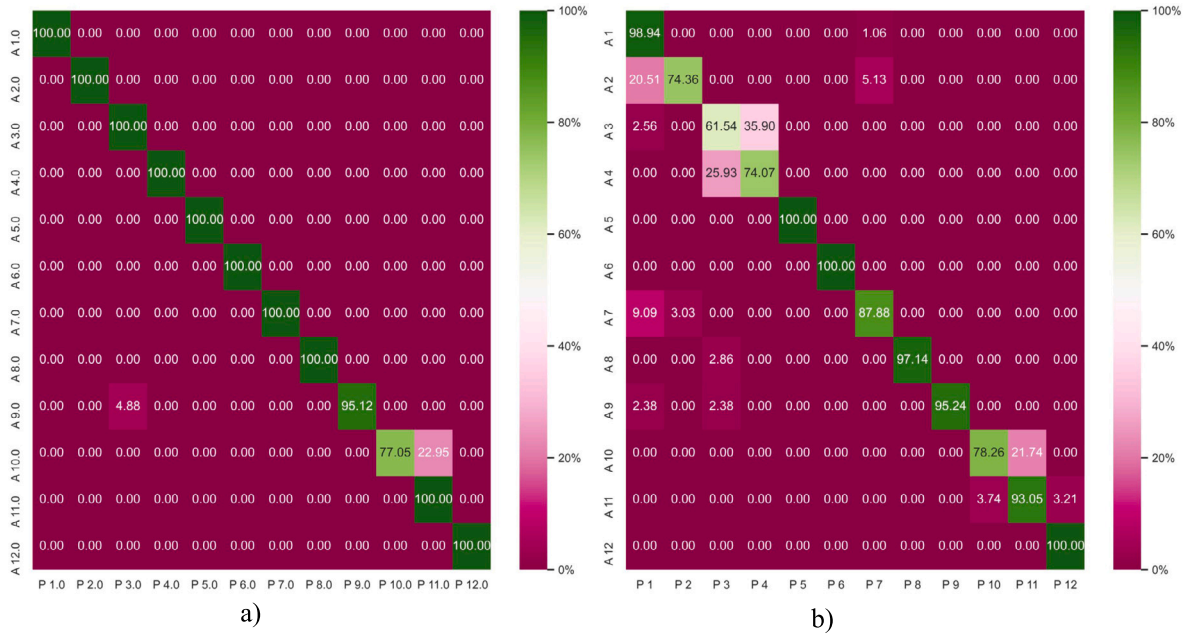
Fig. 7. Confusion matrix of (a) L-GNN and (b) KNN (A: Actual label, P: Predicted label).
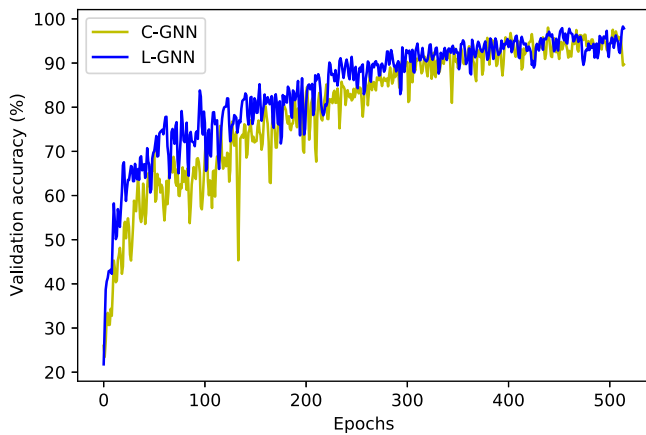

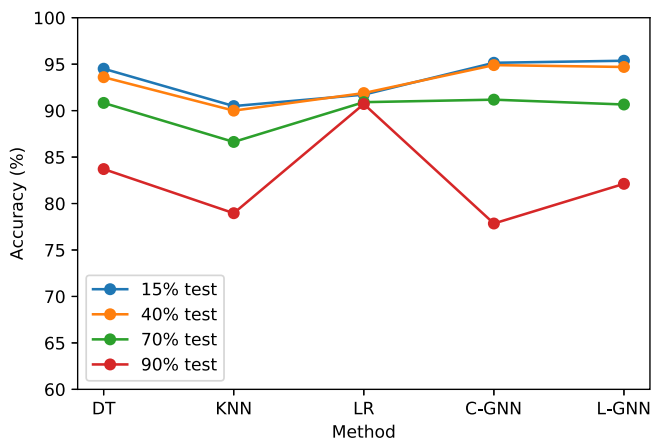**Fig. 8.** Validation accuracy versus epochs.

Fig. 9. Test accuracy with different size of testing sets.

Table 4

The accuracy of event classification with various configuration of PMUs.

Classes	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
A	87.07	86.62	84.03	84.40
B	93.75	93.45	91.96	92.28
C	95.40	95.66	94.14	94.67

Other statistical metrics such as precision, recall, and F1 scores also indicate an acceptable performance of the model, as shown in Table 4. Moreover, having six installed PMUs in whole network as class C, gives a comparable and close test accuracy and statistical metrics compared with the base case with eight PMUs. It means that the proposed method performs well if limited number of sensors are available. This analysis is presented as PMUs are expensive devices and they are usually installed on a few number of nodes at the distribution level.

4.7. Temporal missing data

In this section, we consider the missing information in the time-series data from the PMUs. We randomly remove 20% 50% and 80% of the data stream and shrink the feature dimension of the input data of the classifier. Table 5 shows the average accuracy of the event classification and feature dimension for these three scenarios. Missing 80% of PMU data stream, as the worst scenario, leads to 92.93% accuracy, while losing less temporal data as 20% does not considerably affect the overall performance of the classification, and the accuracy reaches to 95.22%. It needs to be noted that PMUs record limited number of samples during events, specially for transient events, and missing data in these periods can reduce the classification accuracy. This problem motivates us to explore the application of waveform measurement units (WMU) with higher sampling rate and more recorded information during the event in our future studies [41]. Also, [42] is suggested as a new reference for the application of WMUs and PMUs for event classification.

4.8. Edge feature

Incorporating the adjacency matrix of the graph (distribution grid topology) in GNN models makes substantial differences compared to

Table 5
The accuracy of event classification considering temporal missing data.

Temporal missed data (%)	Test accuracy (%)	Feature dimension
20	95.22	87
50	94.36	55
80	92.93	23

Table 6
Model evaluation without edge features.

Adjacency matrix	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
No edge features	94.79 (−0.40)	94.96 (−0.39)	93.30 (−0.43)	93.77 (−0.28)

other traditional methods. This matrix containing topological and edge information enables GCN to aggregate the spatial–temporal nodal features of the graph. In this regard, the effect of weighted and unweighted adjacency matrices are compared in the event classification problem to show the impact of the edge information (features) on the performance. The unweighted adjacency matrix just contains the connectivity status of the grid with binary values. Using this matrix, the accuracy of L-GNN reduces to 94.79%, which is 0.4% less than the L-GNN model in which weighted adjacency matrix with the distance metric is considered. Other metrics also drop compared to the base L-GNN case as indicated in the parenthesis in Table 6. Even with unweighted adjacency matrix used in the proposed GNN-based method, the obtained accuracy is higher than the ones obtained by DT, KNN and LR. Therefore, considering the topological structure of the graph even without edge features is beneficial in better regional and locational event classification.

4.9. Design of GNN

There are several elements that can lead to enhancement of the performance of L-GNN, which are explained below.

Number of layers: GCN layers are substantial parts of the proposed GNN model. The number of these layers specifies the receptive field of graph nodes. This means that increasing the stacked layers leads to a deeper model and higher-order neighborhoods. According to our analysis, the accuracy slightly improves to 95.74% and 95.71% with three and four GCN layers, respectively, while the computation time is also increased during the training process. It is worth mentioning that increasing the number of stacked layers, extending the receptive fields of nodes does not always enhance the classification accuracy due to the complexity of the model [29]. As shown in Fig. 10, the accuracy remains almost constant for more than one GCN layer. Also, the accuracy drops with just one GCN layer as constructing a single layer is not sufficient to take advantage of spatial–temporal features of signals for the classification.

Learning rate: Another important parameter that significantly changes the performance of GNN is the learning rate. The learning rate determines the step size during the training process, and how fast the GNN is adapted to the event classification problem. For example, small learning rate causes slow training process with tiny updates of the weights in layers. To show the functionality of GNN with different learning rates and step sizes, we adjust it to 0.001 and 0.00001. The problem runs for 600 epochs, and the best obtained accuracy is 78.67% and 87.86% for 0.001 and 0.00001, respectively; however, these results are not desirable and this is why the optimal learning rate is set on 0.001 in the previous experiments. Fig. 11 demonstrates the validation accuracy for several epochs, proving that higher and lower step size does not improve the performance of L-GNN and a trade-off is required for tuning this parameter.

Batch size: Furthermore, the training scenarios are not given one-by-one to GCN. The adjacency matrices of the scenarios are stacked in a sparse diagonal form, forming one big graph representing the entire generated scenarios. Their node and target labels are also concatenated
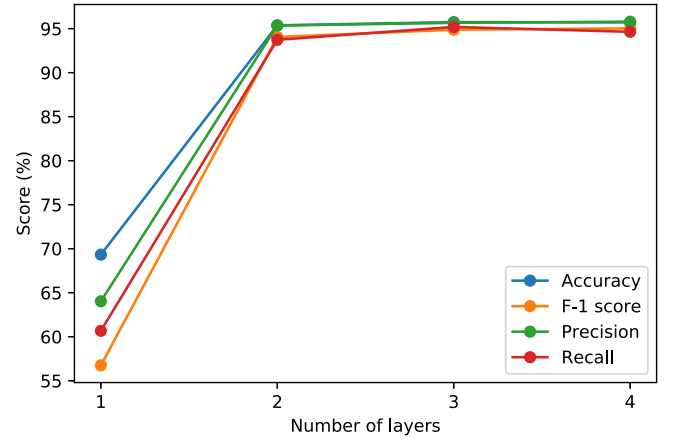
Fig. 10. Scores vs. number of GCN layers.
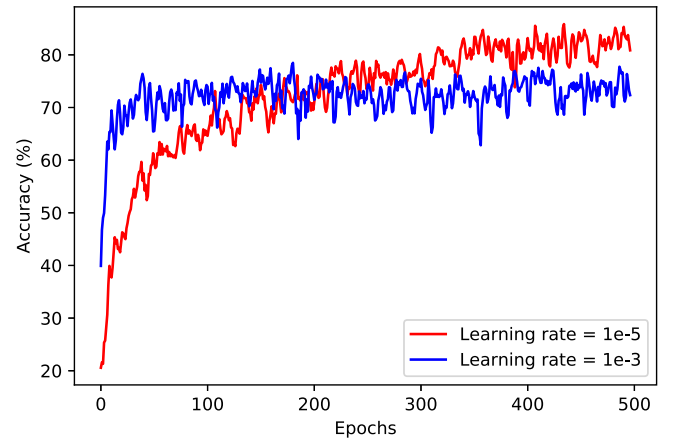
Fig. 11. Accuracy plots for different learning rates.

with respect to the nodal dimension. It should be mentioned here that different grouped graphs (different batches) are isolated, so their features cannot be transmitted to the other grouped information and this demonstrates how the parallel computation for the scenarios is performed. The batch size is defined to group a number of sample graphs, making a large individual samples. We set batch size to 64 and 128, and the obtained accuracy is 93.56% and 94.42%, respectively. F1 score is calculated as 91.69% and 92.79% considering 64 and 128 batches, respectively. Consequently, less number of sample graphs in each group leads to a better classification accuracy.

4.10. Comparison with the state-of-art

Table 7 shows the comparison between the proposed classifier and the state-of-art methods in [16,17] and [21]. The proposed model in [16] classifies the events related to CB, OLTC, switching and load changes considering similar sampling rate for the PMUs and ignoring the penetration of DGs. The method discussed in [17] focuses on event classification using classical ML-based methods that do not incorporate the network-related correlation among the sensors using graph. Moreover, several types of events are grouped in the same class for event classification, and are not classified individually. However, the study in [17] is performed using real-world data sets recorded by PMUs with similar sampling rates. In [21], fault location using graph-based method namely GNN in passive distribution grid is proposed, and other types of event are not included. Our previous study in [22] extended [21] by considering PV-penetrated distribution grids and different graph filter

Table 7
Comparison with the state-of-art methods.

	Methodology	Event type	Multi-rate sensors	Active grid
[16]	Autoencoder	CB, OLTC, Load changes, Network switching	No	No
[17]	SVM	CB, Oscillation Load changes, DER, Transformer	No	Yes
[21]	GNN	Fault	No	No
Proposed study	GNN	CB, OLTC, fault, DER, Load changes	Yes	Yes

with mathematical approximations. However, the multi-rate sampling sensors and various types of events are not addressed. The proposed method of this study addresses various types of events, such as faults, CB outages, VR malfunction, load changes and DER outages. Additionally, this study is formulated to include multi-rate sampling data for having a more realistic event classification approach. Regarding the approach, two GNN models with different filters are studied to incorporate the connectivity information of the grid and the location of sensors into the training process to enhance the event classification and network situational awareness. The performance of the proposed physics-informed GNN models is also compared with conventional ML-based approaches, indicting a better classification accuracy.

5. Discussion

The proposed L-GNN can be a viable solution to improve event classification performance in DER-penetrated distribution grids. Specific advantages of the proposed model are discussed and future directions are explained as follows:

- *Situational awareness*: The L-GNN assists system operators for better distribution grid monitoring and enhances situational awareness under different types of temporary and permanent contingencies. Utilities can use the trained L-GNN model when events occur to identify the type and the corresponding zone. This enables them to instantly mitigate the side effects of events, specifically permanent ones, by dispatching the repair crew to the affected locations, optimally managing the remaining active and reactive power resources, and isolating the faulted region.

- *Correlation of events*: Identifying the CBs switching and tap-changer failure events as reactive/voltage related resources as well as the DER and load incidents as the main active power-based units are complicated problems, as they can lead so similar patterns on voltage phasors. Also, due to the high ratio of R/X in distribution grids, even active power fluctuations can impact the voltage profile like reactive-based resources events and it even makes the classification more challenging. Most of the existing studies focus on a limited number of events or they categorize the highly correlated events as one class, leading to a more limited event classification. However, the proposed algorithm can successfully classify a variety of individual events with an accuracy of 95.37% and 98.34% with and without adding noise, respectively. Additionally, it can identify events' region and even location using the physics-aware features of PMUs phasor data.

- *Modified PMU data*: In real-world operation, utilities may have limited budget to install PMU technologies in their grids and installed PMUs might experience outages. Thus, robustness of the proposed model has been evaluated under limited set of available PMUs. The classification accuracy reaches to 87.07% even with two PMUs. Additionally, the recorded time-series data may contain missing information due to sudden failure of sensors in

several locations. In the worst case scenario with the missing rate of 80%, L-GNN classifies 92.93% of the events accurately. Also, the sensors' data might be corrupted with noise. It is shown that the L-GNN is more robust compared to other methods for event classification with 95.37% accuracy, considering measurement noise of 45 dB SNR.

- *Data fusion*: Proliferation of DERs necessitates the use of more sensors in distribution networks in the future. Thus, the number of measurement devices with different sampling rates can increase and data fusion of the recorded data is a challenging task for many applications including event classification studies. Although we have combined two groups of PMUs with different sampling rates in the measurement matrix during a given time interval, other types of data fusion can be studied, which is one of our main future directions.

In our future studies, a real-time digital simulator (RTDS) platform integrated with inverters and monitoring devices will be used to create a near-to-real world data sets under various events. Also, working with real-world events data with grid topological characteristic are among our future research directions to better evaluate the performance of the proposed event classification model. Another direction will be the application of other efficient GCN filters in the GNN framework, for enhancing the classification performance.

6. Conclusions

This paper has proposed a graph neural network model for event classification and region identification in distribution networks. Temporary and permanent events such as capacitor banks (CBs) and DER outages, malfunction of transformer tap-changer, sudden load changes, and faults are studied in this paper. The physical characteristics of the DER-penetrated distribution grid and phasor measurements from a limited number of phasor measurement devices (PMUs) installed on distribution feeders' nodes are considered as the spatial–temporal information during the learning of the L-GNN model with linear spectral convolution filters. Moreover, multi-rate sampling PMUs are considered in this study to resemble real-world situation. The performance of the proposed L-GNN classifier is validated with adding noise, shrinking data size, changing PMUs configuration, and removing a portion of PMUs data. The simulation results on the modified IEEE 34-bus system shows that the L-GNN outperforms GNN with Chebyshev filter in terms of classification accuracy, convergence rate and computation time. The superior performance of the L-GNN is validated with respect to other conventional machine learning algorithms, based on the calculated statistical metrics, such as precision, recall, and F-1 score.

CRediT authorship contribution statement

Mohammad MansourLakouraj: Conceptualization, Methodology, Data curation, Writing – original draft, Visualization, Software, Investigation, Validation, Formal analysis, Writing – review & editing. **Mukesh Gautam:** Writing – original draft, Visualization, Data curation, Formal analysis, Writing – review & editing. **Hanif Livani:** Conceptualization, Methodology, Formal analysis, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Mohammed Benidris:** Conceptualization, Methodology, Supervision, Writing – review & editing.

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.

Acknowledgments

References

[1] Y. Yuan, Y. Guo, K. Dehghanpour, Z. Wang, Y. Wang, Learning-based real-time event identification using rich real PMU data, IEEE Trans. Power Syst. 36 (6) (2021) 5044–5055.

[2] S. Pandey, A.K. Srivastava, B.G. Amidan, A real time event detection, classification and localization using synchrophasor data, IEEE Trans. Power Syst. 35 (6) (2020) 4421–4431.

[3] A.L. Liao, E.M. Stewart, E.C. Kara, Micro-synchrophasor data for diagnosis of transmission and distribution level events, in: 2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), IEEE, 2016, pp. 1–5.

[4] S. Fahmy, R. Gupta, M. Paolone, Grid-aware distributed control of electric vehicle charging stations in active distribution grids, Electr. Power Syst. Res. 189 (2020) 106697.

[5] N. Duan, E.M. Stewart, Frequency event categorization in power distribution systems using micro PMU measurements, IEEE Trans. Smart Grid 11 (4) (2020) 3043–3053.

[6] L.G. de Oliveira, M.d.L. Filomeno, H.V. Poor, M.V. Ribeiro, Fault detection and location in power distribution systems: The usefulness of the HS-OFDM scheme for time-domain reflectometry, Electr. Power Syst. Res. 203 (2022) 107600.

[7] M. Izadi, H. Mohsenian-Rad, Characterizing synchronized Lissajous curves to scrutinize power distribution synchro-waveform measurements, IEEE Trans. Power Syst. (2021).

[8] A. Shahsavari, M. Farajollahi, H. Mohsenian-Rad, Individual load model parameter estimation in distribution systems using load switching events, IEEE Trans. Power Syst. 34 (6) (2019) 4652–4664.

[9] M. Ghosal, V. Rao, Fusion of multirate measurements for nonlinear dynamic state estimation of the power systems, IEEE Trans. Smart Grid 10 (1) (2017) 216–226.

[10] H. Livani, C.Y. Evrenosoglu, V.A. Centeno, A machine learning-based faulty line identification for smart distribution network, in: 2013 North American Power Symposium, NAPS, IEEE, 2013, pp. 1–5.

[11] Z.S. Hosseini, M. Mahoor, A. Khodaei, AMI-enabled distribution network line outage identification via multi-label SVM, IEEE Trans. Smart Grid 9 (5) (2018) 5470–5472.

[12] M. Ganjkhani, M. Gilanifar, J. Giraldo, M. Parvania, Integrated cyber and physical anomaly location and classification in power distribution systems, IEEE Trans. Ind. Inf. (2021).

[13] M. Gilanifar, H. Wang, J. Cordova, E.E. Ozguven, T.I. Strasser, R. Arghandeh, Fault classification in power distribution systems based on limited labeled data using multi-task latent structure learning, Sustainable Cities Soc. (2021) 103094.

[14] C. Pavlatos, V. Vita, A.C. Dimopoulos, L. Ekonomou, Transmission lines' fault detection using syntactic pattern recognition, Energy Syst. 10 (2) (2019) 299–320.

[15] C. Pavlatos, V. Vita, Linguistic representation of power system signals, in: Electricity Distribution, Springer, 2016, pp. 285–295.

[16] I. Niazazari, H. Livani, A PMU-data-driven disruptive event classification in distribution systems, Electr. Power Syst. Res. 157 (2018) 251–260.

[17] A. Shahsavari, M. Farajollahi, E.M. Stewart, E. Cortez, H. Mohsenian-Rad, Situational awareness in distribution grid using micro-PMU data: A machine learning approach, IEEE Trans. Smart Grid 10 (6) (2019) 6167–6177.

[18] M. Bariya, A. von Meier, J. Paparrizos, M.J. Franklin, K-ShapeStream: Probabilistic streaming clustering for electric grid events, in: 2021 IEEE Madrid PowerTech, IEEE, 2021, pp. 1–6.

[19] M. Pavlovski, M. Alqudah, T. Dokic, A.A. Hai, M. Kezunovic, Z. Obradovic, Hierarchical convolutional neural networks for event classification on PMU measurements, IEEE Trans. Instrum. Meas. (2021).

[20] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, Adv. Neural Inf. Process. Syst. 29 (2016) 3844–3852.

[21] K. Chen, J. Hu, Y. Zhang, Z. Yu, J. He, Fault location in power distribution systems via deep graph convolutional networks, IEEE J. Sel. Areas Commun. 38 (1) (2019) 119–131.

[22] M. MansourLakouraj, R. Hossain, H. Livani, M. Ben-Idris, Application of graph neural network for fault location in PV penetrated distribution grids, in: 2021 North American Power Symposium, NAPS, IEEE, 2021, pp. 01–06.

[23] Y. Luo, C. Lu, L. Zhu, J. Song, Data-driven short-term voltage stability assessment based on spatial-temporal graph convolutional network, Int. J. Electr. Power Energy Syst. 130 (2021) 106753.

[24] H. Tong, R.C. Qiu, D. Zhang, H. Yang, Q. Ding, X. Shi, Detection and classification of transmission line transient faults based on graph convolutional neural network, CSEE J. Power Energy Syst. 7 (3) (2021) 456–471.

[25] G. Wang, Z. Zhang, Z. Bian, Z. Xu, A short-term voltage stability online prediction method based on graph convolutional networks and long short-term memory networks, Int. J. Electr. Power Energy Syst. 127 (2021) 106647.

[26] Y. Liu, N. Zhang, D. Wu, A. Botterud, R. Yao, C. Kang, Searching for critical power system cascading failures with graph convolutional network, IEEE Trans. Control Netw. Syst. (2021).

[27] W. Lin, D. Wu, B. Boulet, Spatial-temporal residential short-term load forecasting via graph neural networks, IEEE Trans. Smart Grid (2021).

[28] Y. Yuan, Z. Wang, Y. Wang, Learning latent interactions for event identification via graph neural networks and PMU data, IEEE Trans. Power Syst. (2022).

[29] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2016, arXiv preprint arXiv:1609.02907.

[30] R. Levie, F. Monti, X. Bresson, M.M. Bronstein, Cayleynets: Graph convolutional neural networks with complex rational spectral filters, IEEE Trans. Signal Process. 67 (1) (2018) 97–109.

[31] D.K. Hammond, P. Vandergheynst, R. Gribonval, Wavelets on graphs via spectral graph theory, Appl. Comput. Harmon. Anal. 30 (2) (2011) 129–150.

[32] Z. Chen, J. Xu, T. Peng, C. Yang, Graph convolutional network-based method for fault diagnosis using a hybrid of measurement and prior knowledge, IEEE Trans. Cybern. (2021).

[33] A. Ghasemkhani, I. Niazazari, Y. Liu, H. Livani, V.A. Centeno, L. Yang, A regularized tensor completion approach for PMU data recovery, IEEE Trans. Smart Grid 12 (2) (2020) 1519–1528.

[34] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.

[35] PyTorch Geometric, URL https://pytorch-geometric.readthedocs.io/en/latest/.

[36] M. Izadi, H. Mohsenian-Rad, A synchronized lissajous-based method to detect and classify events in synchro-waveform measurements in power distribution networks, IEEE Trans. Smart Grid (2022).

[37] W.D. Oliveira, J.P. Vieira, U.H. Bezerra, D.A. Martins, B.d.G. Rodrigues, Power system security assessment for multiple contingencies using multiway decision tree, Electr. Power Syst. Res. 148 (2017) 264–272.

[38] N. García-Pedrajas, J.A.R. Del Castillo, G. Cerruela-García, A proposal for local *k* values for *k*-nearest neighbor rule, IEEE Trans. Neural Netw. Learn. Syst. 28 (2) (2015) 470–475.

[39] M. Coutinho, L.L. Novo, M. de Melo, L. de Medeiros, D. Barbosa, M. Alves, V. Tarragô, R. dos Santos, H.L. Neto, P. Gama, Machine learning-based system for fault detection on anchor rods of cable-stayed power transmission towers, Electr. Power Syst. Res. 194 (2021) 107106.

[40] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, H. Xu, Time series data augmentation for deep learning: A survey, 2020, arXiv preprint arXiv:2002.12478.

[41] I. Niazazari, H. Livani, A. Ghasemkhani, Y. Liu, L. Yang, Event cause analysis in distribution networks using synchro waveform measurements, in: 2020 52nd North American Power Symposium, NAPS, IEEE, 2021, pp. 1–5.

[42] M. Izadi, H. Mohsenian-Rad, Synchronous waveform measurements to locate transient events and incipient faults in power distribution networks, IEEE Trans. Smart Grid 12 (5) (2021) 4295–4307.