International Journal of Research and Technology in Electrical Industry

journal homepage: ijrtei.sbu.ac.ir

Shahid Beheshti University

A Transfer Learning Method for Intelligent Load Shedding Using Graph Convolutional Network Considering Unknown Faults

Nazanin Pourmoradi[1] , Mohammad Taghi Ameli[1, *]

[1] *Department of Electrical Engineering, Abbaspour School of Engineering, Shahid Beheshti University, Tehran, Iran*

ARTICLE INFO

Article history:
Received: 01 December 2024
Revised: 28 December 2024
Accepted: 03 January 2025

Keywords:
Graph Convolutional Network
Unknown Fault
Transfer Learning
Event-based Load Shedding
Small Database

ABSTRACT

Event-based load shedding (ELS) is a vital emergency countermeasure against transient voltage instability in power systems. Deep learning(DL)--based ELS has recently achieved promising results. However, in power systems, faults may occur that are not in the training database, reducing the model's effective performance. In this situation, it is necessary to update the model. On the other hand, updating the model for new faults requires a large database. To address the problem of unknown faults, this paper proposes a transfer learning-based graph convolutional network (GCN) model that allows updating the model with a small database. In the first step, an ELS model is trained with a large database. Then, if a new fault occurs, the model is transferred to the new fault and updated using transfer learning and with a small database. To evaluate the performance of the proposed model, it was implemented and tested on the IEEE 39 bus system. The results show that the proposed model has high-performance accuracy and can be updated with a small database when encountering an unknown fault. According to the results, the proposed model has reduced the database size by 78.91% for optimal updating.

1. Introduction

Due to economic considerations, modern grids operate very close to their stability limits. Therefore, generator trips and line outages can cause instability in the power system. A power system's voltage stability refers to its ability to maintain a permissible voltage on all buses in the event of a fault [1,2]. Voltage instability plays a vital role in the onset of blackouts, so researchers have studied this phenomenon for the past two decades [3,4]. Corrective or preventive measures are typically used to influence the situation before and after a fault occurs to prevent voltage instability. Suppose preventive measures fail to change the system condition from critical to alarm. In that case, corrective measures such as generation switching or load shedding (LS) can be employed as a cheap but reliable way of maintaining the integrity of the network. A significant parameter in load shedding is the amount of load to be disconnected to maintain bus voltages within a stable and acceptable range. A lack of LS can lead to further failures, while a surplus of shedding can result in overdone and unnecessary curtailment. Several studies have been conducted to determine the optimal LS method using a variety of algorithms, modeling details, and objective functions.

Various heuristic methods are used for LS optimization to achieve the best results in the shortest time possible. In [5], a method based on the genetic algorithm (GA) was proposed for the steady state load shedding phenomenon in distribution networks for production shortage scenarios, which aimed to minimize the total curtailed load and system losses. According to [6], the loads were classified into fixed and random priorities based on their types, and the optimal LS was determined using a random combinations method for the random priority loads. An optimal strategy for LS under voltage was obtained by

* Corresponding author
E-mail address: m_ameli@sbu.ac.ir
https://orcid.org/0000-0002-8815-1596
http://dx.doi.org/10.48308/ijrtei.2024.237789.1066

optimizing the inertia weight and learning coefficient of the particle swarm optimization (PSO) algorithm using fuzzy rules [7]. In [8], an LS method is proposed to remove the power deficiency and restore voltage and frequency stability. Furthermore, [9] considers an under-voltage LS technique for determining an appropriate LS amount in an islanded micro grid. In [10], a UVLS optimization technique called Evolutionary Particle Swarm Optimization (EPSO) was proposed to identify the best remaining load while reducing power losses, voltage variations, and LS cost. It has also been shown that GA and artificial neural networks (ANNs) can be applied together to minimize network LS and voltage deviations [11].

The above methods are time-consuming and lack convergence. Although extensive artificial intelligence (AI) methods, including shallow neural networks [12,13] and deep learning (DL) methods [14,15], have been developed for real-time stability assessment, the approaches for stability control are very limited. Based on CNN, [16] developed a method for reducing line loads. According to [17], a deep reinforcement learning (DRL) method was developed for under-voltage LS. In [18] and [19], the extreme learning machine (ELM) algorithm was used to maintain post-fault frequency stability, and the method was further developed in [20] for fault-induced delayed voltage recovery. [21] proposes a risk-averse DL method for real-time emergency LS that trains neural networks to avoid load under-cutting events, thus reducing the cost of control failures. With an event-based model, [22] proposes intelligent LS and removing repetitive and harmful behavior to improve training and decision-making. Using DRL and data-driven strategies, [23] proposes an emergency LS technique.

LS methods based on DL have obtained promising results. However, these methods face a major challenge when implemented in real-world power systems. For LS, different models are trained for different potential faults and each model is applied to a specific fault. Assuming that the training data (i.e., the LS database) and the unknown data (i.e., the online measurement) follow the same distribution, the models usually achieve satisfactory accuracy. However, this assumption may not always hold, especially when a new fault occurs in the power system. In this situation, the accuracy of the existing methods decreases, which reduces the power system operator's confidence in data-driven models.

To solve this challenge, the authors propose updating the model for unknown faults using transfer learning (TL). For this purpose, a transfer approach based on fine-tuning the pre-trained GCN model for faults in the training database is used. In this situation, the proposed model uses the optimal parameters and weights of the pre-trained model as initial conditions to update the model for new faults. With this fine-tuning, a good initial learning start for the update is achieved, which makes it possible to transfer the model to new faults with a small database. Also, using a small database reduces model update time.

2. FOUNDATIONS OF DEEP NEURAL NETWORK MODELS

Machine learning (ML), as an important subset of AI, allows systems to automatically learn the necessary knowledge from data in a specific field. In traditional deep neural networks, excellent performance usually depends upon a large training database as well as a similar distribution between the training data (source domain) and the measured data in practical applications (target domain). However, meeting the above conditions in real-world applications is difficult for some research fields. TL is research topic in ML that focuses on storing knowledge acquired for a task and transferring that knowledge to a different but related task. The advantage of knowledge transfers or model migration is that the model with a small training database converges quickly and is trained efficiently. The original model can be trained with a large training database obtained using simulations, and knowledge transfer is used to enhance model performance with a small database. Knowledge transfer can significantly reduce the need for the target domain to have many samples for effective training.

ML-based LS builds a mapping relationship between input features (bus voltage magnitudes and angles) and outputs (optimal LS). In the LS of power systems, the offline training database covers only a limited number of faults. While in the online application of the LS model, various other faults may appear with a different data distribution from the training data. In this paper, faults that are not in the training database are called unknown faults. As a result, pre-trained ELS models may not be able to perform with high accuracy when faced with unlearned faults. The measured data from different faults in the power system have different but similar distributions. Therefore, they are suitable for applying parameter-based TL (knowledge transfer). In this situation, the source domain can be large for various faults with the database, while the target domain refers to a fault with a small database but similar characteristics. Since it is difficult and time-consuming to collect a large database when unknown faults appear, the use of TL can be beneficial for online ELS, enabling retraining for a new fault in less time and with fewer samples. The general idea of LS based on TL presented in this paper can be seen in Fig 1.

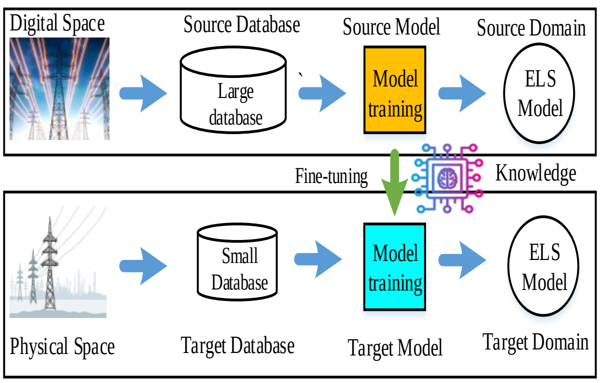

Fig. 1. Transfer learning for LS

3. GRAPH CONVOLUTIONAL NETWORK

In the GCN model, nodes and edges represent the power system buses and transmission lines. The input layer, hidden layer and output layer are constituent parts of a GCN model. Fig. 2 shows the structure of GCN. In a graph G = (N, E) nodes represent nodes, and edges represent edges between nodes [24].

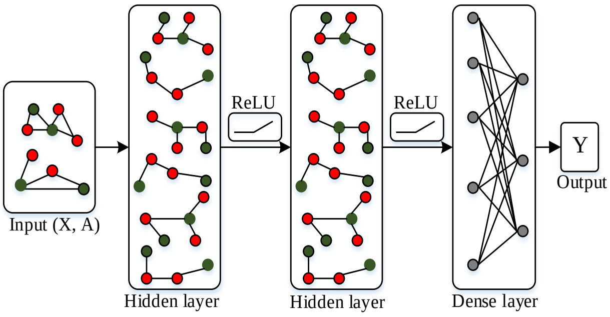

Fig. 2. GCN algorithm

- GCN Input Layer

GCN's input layer is comprised of a feature matrix and an adjacency matrix, as illustrated in equation (1). Adjacency matrices represent the relationship between graph nodes.

$$\text{Input}=(X,A) \tag{1}$$

Also, n and d respectively indicate the number of nodes or buses in the system and the number of input features. A also has $n \times n$ dimensions, which represents the adjacency matrix. A graph with an undirected adjacency matrix is expressed in the following way [30]:

$$A_{ij} = A_{ji} \begin{cases} 0, i \nrightarrow j \\ 1, i \rightarrow j \\ 1, i = j \end{cases} \tag{2}$$

As indicated in this equation, A_{ij} indicates whether node i-th connects to node j-th.

- GCN Hidden Layer

By using propagation rules, the hidden layer of the GCN can collect and send node info to the next layer. As features propagate through successive hidden layers, they become more abstract. The i-th node's layer-wise propagation rules are expressed below [24].

$$h_i^l = \sigma\left(\sum_{j=1}^{N} \bar{A}_{ij} \,.\, w^l . h_i^{l-1} + b^l\right) \tag{3}$$

$$\bar{A} = Q^{-\frac{1}{2}}.A.Q^{-\frac{1}{2}} \tag{4}$$

$$Q = \sum_{j=1}^{N} A_{ij} \tag{5}$$

Where w^l is a trainable linear transformation weight calculated by minimizing the loss function on all labeled data. b^l represents the bias variable. $\bar{A}$ is the normalized adjacency matrix. Q represents the degree matrix of the input graph. σ denotes a nonlinear activation function. h_i^l is the i-th node feature of the l-th hidden layer. Initially, $h_i^0 = X$.

- GCN Output Layer

By extracting the features from the hidden layer, the output layer produces the optimal LS value based on the fitted equation.

4. PROPOSED ELS MODELDESIGN

Fig. 3 shows the proposed model for ELS in power systems, which uses a TL-based approach to manage unknown faults. In the proposed model, a large database of various faults is first created. Subsequent to the normalization of the database, a partition was implemented, allocating 80% of the data for the training stage and reserving 20% for testing purposes. Then, GCN-based ELS models are trained. The GCN-based model integrates topology information into the learning model to exploit spatial distribution features and improve model performance. Finally, if the model encounters unknown faults, TL updates it with a small database. Below are the different parts of the proposed method.

4.1. Database Generation

In this step, a database is created to train the model. In this study, the voltage (V) and phase angle (P) of buses with generators are input features of the model. These features are directly obtained through the phasor measurement unit (PMU), making the proposed model suitable for online applications. In this case, voltage and phase angle information for different operating points are stored pre-fault, and the corresponding optimal LS is obtained with the help of genetic algorithm (By linking MATLAB and Digsilent). Finally, the database of different load levels is as follows:

$$\begin{bmatrix} V_{11} & V_{21} & V_{31} & \dots & V_{B1} & P_{11} & P_{21} & \dots & P_{B1} \\ V_{12} & V_{22} & V_{32} & \dots & V_{B2} & P_{12} & P_{22} & \dots & P_{B2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ V_{1S} & V_{2S} & V_{3S} & \dots & V_{BS} & P_{1S} & P_{2S} & \dots & P_{BS} \end{bmatrix} \begin{bmatrix} LS_1 \\ LS_2 \\ \dots \\ LS_S \end{bmatrix} \tag{6}$$

where B and S represent the number of buses and samples, respectively.

Finally, unstable points are obtained using the rotor angle index. According to this index, if the rotor angle difference between at least two generators is more than 180 degrees when the fault occurs, the system is unstable [25,26]. Optimal LS should be done to stabilize the power system. The transient stability index of the rotor angle is formulated as follows:

$$if \begin{cases} \max(\delta_{ij}) \geq \pi \rightarrow \text{unstable} \\ \max(\delta_{ij}) < \pi \rightarrow \text{stable} \end{cases} \tag{7}$$

4.2. Offline Training

Using the database created in the previous step (source domain), the ELS model is trained. This paper introduces a GCN model for ELS.

In this study, the GCN model consists of two graph convolutional layers, one dropout layers, one fully connected layer for feature interpretation, and one output layer. The graph convolution layers and the dense layers use rectified linear unit (ReLu) activation.

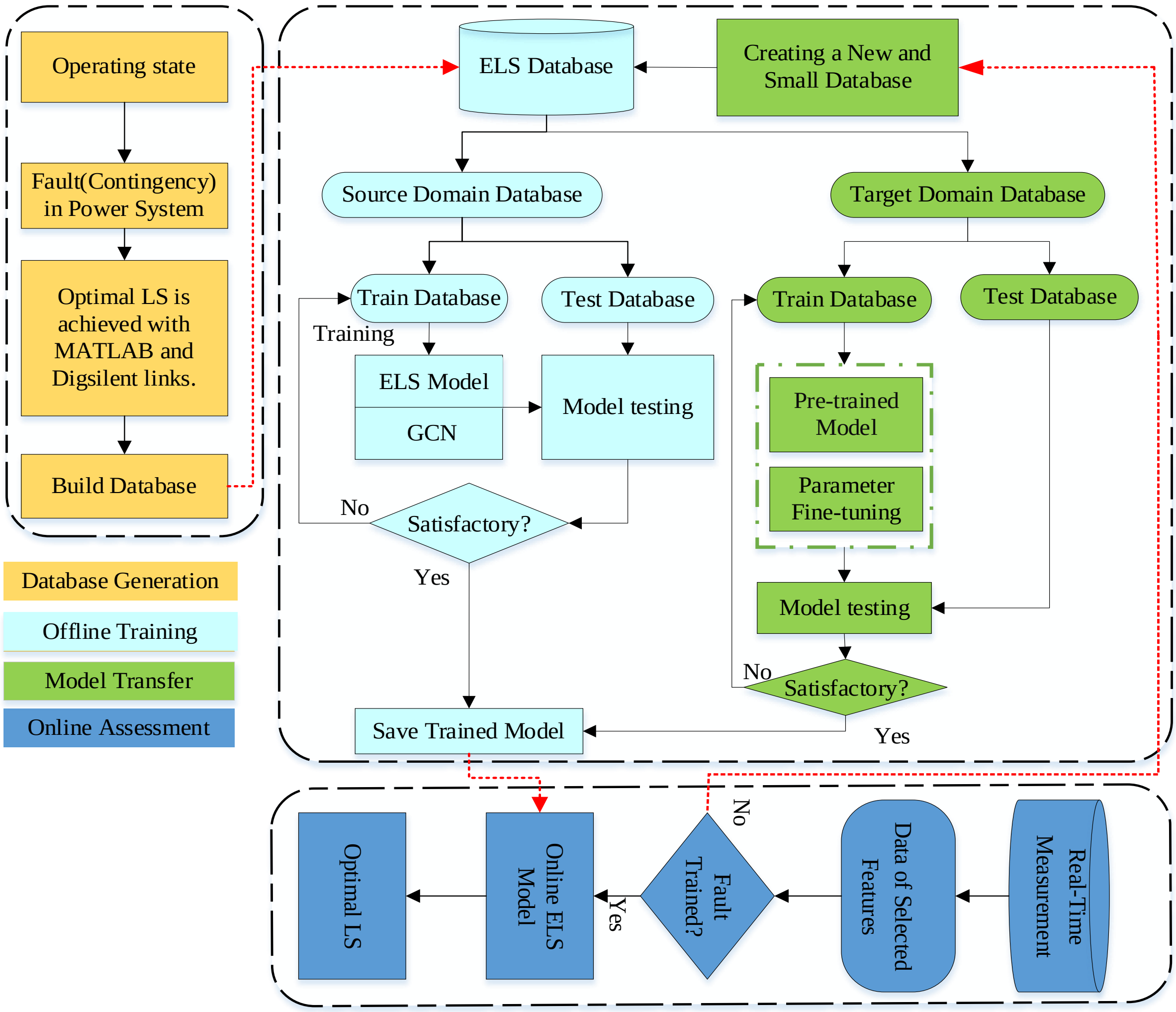

Fig. 3. Flowchart of proposed method

After the graph convolutional layer, a 0.1 probability dropout layer is applied to reduce overfitting. Finally, in the output layer, based on optimal weights, the amount of optimal LS is determined.

In offline training, the model is trained to minimize the difference between the estimated LS and the actual LS value, and learning parameters are obtained. To achieve this, a loss function and an optimization algorithm are necessary. In this situation, the estimated optimal LS is compared with the actual LS through the loss function. The optimization algorithm seeks to reduce the loss function by repeatedly updating the learning parameters. In this work, Huber is used as a loss function and Adam is used for optimization, and Adams algorithm is used for its optimization.

4.3. Model Migration Scheme in Target Domain

In case of an unknown and unlearned fault in the power system, considering that generating a large database is time-consuming and expensive [27], TL is used in this paper to enable retraining the model for the new fault with a small database. If a new fault occurs, a small database called the target domain is first created in this situation. Then, the training database of the target domain is integrated into the pre-trained LS model for fine-tune the parameters. After satisfying the model evaluation indicators, a trained and optimized model is built.

In the proposed transfer approach to retrain the model when encountering unknown faults, fine-tuning of all layers of the GCN model is performed. The structure and parameters of the pre-trained model are transferred to the new LS model and the layers are not frozen. It is used to fine-tune the layers to a small database associated with the unknown fault. In this situation, the weight matrix and bias matrix can be presented in the form of equations (٨) and (٩). The loss function is calculated and the parameters of all layers are fine-tuned to get the optimized w'' and b''.

$$w'' = initialize(w_{out}, w_l, w_{l-1}, w_{l-2}, \dots, w_1) \qquad (8)$$

$$b'' = initialize(w_{out}, w_l, b_{l-1}, b_{l-2}, \dots, b_1) \qquad (9)$$

4.4. Online Application

During the online application phase, PMUs collect real-time measurements. After collection, the measured data are entered into the LS model, whose parameters are optimized in the offline training phase. Finally, the power system's LS result is immediately determined. As shown in Fig. 3, if a new and unlearned fault occurs during the online application phase, the model is first retrained using and a small database, and then optimal LS is applied.

4.5. Evaluation Indexes

To evaluate the models' accuracy, the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) criteria have been used[28], which are presented in (10) and (11), respectively. L_Actual is the actual load, L_Forecasted is the forecasted load, and N is the number of time steps.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(L_{Forecasted} - L_{Actual})^2}{N}} \tag{10}$$

$$MAPE = \sum_{i=1}^{N}\left|\frac{L_{Actual} - L_{Forecasted}}{L_{Actual}}\right| \times \frac{100\%}{N} \tag{11}$$

5. SIMULATION RESULTS

This study implements the proposed model based on GCN on IEEE 39-bus system. Fig. 4 illustrates the system diagram, comprising 10 generators, 46 transmission lines, and 39 buses. The assumption is that PMUs are positioned on buses with generators. Fig. 4 also shows the graph structure used for the GCN-based model to consider spatial information of the power system. The Adam optimizer learning rate in the GCN-based model is 0.001 and 64 batch sizes are employed to maximize GPU utilization. Also, the number of dense units is 256. Furthermore, to avoid overfitting, a 10% dropout for graph convolutional layers is applied. During training, 200 epochs are used in the GCN-based model. The proposed model is compared against other models, including convolutional neural network (CNN), long short-term memory (LSTM), random forest (RF) and support vector machine (SVM). The proposed scheme will be implemented using the Digsilent Power Factory, Matlab and Python platforms.

5.1. Database Generation

The proposed method was evaluated using simulations on the New England 10-machine 39-bus system. Using the Monte Carlo method, operating points were generated. In order to develop the proposed model, a database containing different operating points is created as follows. To obtain the operating points, the load is randomly sampled within its practical change range. Specifically, according to the base load level for each bus, random sampling of load changes is performed. The range of load changes is considered between 0.7 and 1.25 percent of the base load. Then, using time domain simulation, each operating point is assigned a stable or unstable label depending on the set of contingency faults. A final step is to determine an optimal LS value based on the available LS resources for operating points with unstable tags. The faults in this study occur on buses 17 and 25 for 0.2 seconds. As a result of the preliminary sensitivity analysis in [29], buses 4, 8, 20 and 39 have been selected as candidates for LS. The maximum LS percentage at all candidate buses is set to 90%. GA is used as a heuristic algorithm to find the optimal emergency LS strategy for each generated sample. The size of the database is finally 4500 samples. For model training, 80% of the operating points are randomly selected, and the remaining 20% are used for model testing.

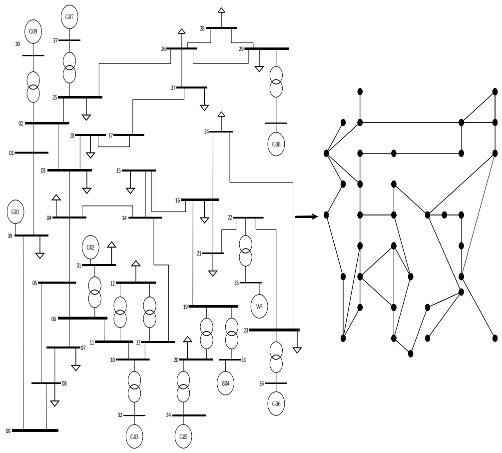

Fig. 4. The IEEE 39-bus test system

5.2. Performance Evaluation of GCN-Based Model and Comparison with Other Approach

To clarify the GCN training process, Fig. 5 shows the change in the loss function for 400 iterations. According to the Fig. 6, it is clear that, in initial iterations, the value of the loss function of the training and testing database decreases rapidly. After 200 iterations, the loss function is almost constant and does not change much. This illustrates that the GCN model has converged. In general, GCN has a stable training process and converges rapidly

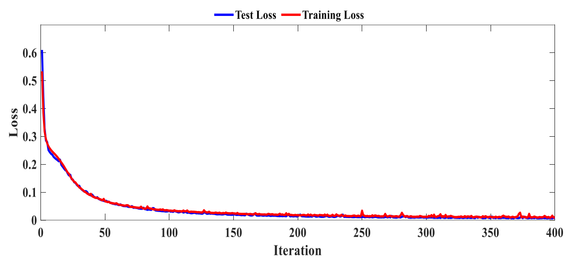

Fig. 5. Training process of GCN model

In Table I, the effectiveness of the GCN model for different numbers of convolution layers and dense layers is evaluated based on time and MAPE indicators. For better evaluation, all tests were performed 5 times with different random seeds. Model accuracy increases with increasing layers, according to the results. But it should be noted that for two layers of graph convolution and two fully connected layers, the model achieved a MAPE of 4.21%. With more layers, the model achieves almost the same MAPE, so the optimal number of layers is two graph convolution layers and two fully connected layers. Also, in this case, the model's training time is 549.52 seconds, which increases with the increase in layers. The crucial point for selecting the number of layers, in addition to model accuracy, is to pay attention to less training time. LS models may need to be updated for various reasons, and the shorter the update time, the better the model. Increasing the number of layers slows down the updating of the online LS model.

Table I. Performance of GCN model based on number of layers

Graph convolutional layer	Fully connected layer	MAPE (%)	Time(s)
1	1	7.85	315.12
2	1	6.41	442.34
3	1	5.43	571.78
1	2	6.74	394.24
2	2	4.21	495.14
3	2	4.18	604.45
1	3	5.28	452.64
2	3	4.18	588.33
3	3	4.21	657.91

Testing the proposed model on source domain data is presented in this section. In order to rationally evaluate the performance of the proposed model, comparative tests are performed with DL models such as CNN and LSTM as well as shallow ML models such as SVM and RF using evaluation indices including MAPE and RMSE. All tests were performed ۵ times with different random seeds to provide a valid comparison of other classifiers. TABLE II shows the accuracy results of different algorithms.

TABLE II. Comparison of prediction accuracy of different approaches

Approach	MAPE (%)	RMSE (%)
Proposed method	4.21	3.92
CNN	5.63	4.86
LSTM	6.15	4.95
RF	8.12	6.75
SVM	8.96	7.54

Table II shows that GCN has the highest accuracy, and CNN and LSTM have better accuracy than RF and SVM. This indicates that considering the structural information of the power system by GCN can help to achieve high-performance accuracy. One of the reasons for the superiority of GCN over other algorithms, especially CNN, is to pay attention to the topology of the power system. GCN is able to obtain the relationships between the voltage magnitude and the angle of the buses, which are the input data, according to the topology of the power system, and extract effective features for classification.

5.3. Evaluating Performance of Proposed Method for Unknown Faults

In LS studies, training databases typically encompass a limited number of faults. However, in real-world applications, other faults may emerge, characterized by distinct data distributions from those in the training database. Consequently, a pre-trained LS model may perform suboptimally for new faults that have not been previously encountered. This situation necessitates retraining the LS model for new faults. However, deep neural networks require a large training database to retrain effectively and achieve excellent performance. Collecting and labelling a large amount of data for each new fault is costly, time-consuming, and challenging. In this situation, model migration can be beneficial. The TL technique is helpful for model migration because the model can be trained effectively and quickly with a small database using this technique. In general, model migration can significantly reduce the need for a large number of samples from the target domain.

In order to test the TL capability of the proposed model, two unknown faults are considered. Table III presents the unknown fault settings for the LS model. Table IV shows the test results of the original GCN trained with source data for unknown fault scenarios. Table IV shows that the original GCN model's performance accuracy has declined significantly in unknown fault scenarios. Consequently, the original GCN trained in the source domain will need to be updated and transferred to the new unknown fault scenario. To verify the efficiency of the proposed method, three different transfer approaches are used to update the original GCN. Approach 1 is the one adopted in this paper, which is discussed in more detail in section IV, while approaches 2 and 3 are as follows:

Approach 2: The structure and parameters of the original GCN model are transferred to the new model, and only the fully connected layers are fine-tuned, and the other layers are frozen.

Approach 3: The structure of the original GCN model is transferred to the new model, and the parameters of the model are initialized randomly.

TABLE III. Contingency unknown faults

Scenario	Fault	Fault setting	Duration(s)
1	UF1	Fault bus 13, Trip 10-13	0.25
2	UF2	Fault bus 11, Trip 10-11	0.15

TABLE IV. Performance of the proposed model for unknown faults

Scenario	1	2
MAPE (%)	15.25	13.97
RMSE (%)	13.78	12.46

Fig. 6 shows GCN's performance after transferring and updating based on these three approaches. The comparison of approach 1 and approach 2 shows that only fine-tuning of fully connected layers does not provide enough learning space for new samples. In this situation, approach 2 achieved a MAPE of 7.05 % at best. That is unacceptable. Therefore, approach 2 cannot guarantee the model's successful performance when transferring to new faults. Based on a comparison of approach 1 and approach 3, sharing the structure and parameters of the original model leads to a good learning point for the new GCN. This good start to learning has caused Approach 1 to achieve a MAPE of 4.30% with 650 training samples. Therefore, updating with approach

1 only requires a database with 650 samples, which should be obtained by long simulation. Meanwhile, approach 3 with 1950 training samples reached a MAPE of 4.35%. As a result, more samples have to be generated and labeled, making updating and transferring the model more time-consuming. According to the obtained results, the proposed approach has better performance than other approaches and successfully transfers the model to new faults. Table V summarizes the performance of different approaches for model updating. Examining the results of Table V shows the superior performance of the proposed approach for updating.

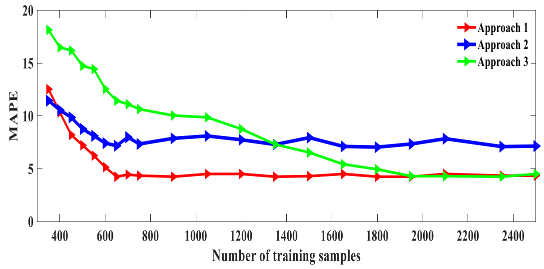

Fig. 6. Comparison of effect of different transfer approaches

Table V. Performance of different approaches to updating

	Approach 1	Approach 2	Approach 3
MAPE (%)	4.30	7.14	4.35
Database size	650	650	1950

5.4. Impact of Noisy Data

PMU s are highly accurate measuring devices, but the data sent by the PMU to the Wide-area measurement system (WAMS) centers may contain noise and be error-prone. To investigate the effect of noise data on the performance of the proposed model, two scenarios are studied in this section. In the two studied scenarios, noise is added to the database randomly so that the total vector error provided by the PMUs is below 1% [30]. In this paper, the effect of noise and measurement error of PMUs was investigated as follows:

Scenario 1: Data does not contain noise.
Scenario 2: Test data contains noise.
Scenario 3: Both training and testing data contain noise.

The test results for the described scenarios are presented in Table VI. These results demonstrate that noisy data diminishes the accuracy of the proposed models but remains within an acceptable range for LS.

TABLE VI. Accuracy of Proposed Models for Noisy Data

	Scenario 1	Scenario 2	Scenario 3
MAPE (%)	4.21	7.22	5.45
RMSE (%)	3.92	6.95	5.11

5.5. Performance Testing of Proposed Model When Changing Topology

In practical applications, an LS model's performance may degrade when the power system topology is changed. Therefore, an effective online LS model should adapt as much as possible to new topologies. To demonstrate the impact of changing the topology on the model's performance, several topologies of the IEEE 39-bus system were evaluated. In this scenario, the trained model with the original (base) topology is used in the face of unknown topologies. Table VI shows the decrease in model accuracy when facing unknown topologies. However, for T2 and T3, the decrease in performance accuracy is higher. This shows that for T2 and T3, the difference between the training data distribution and target data (T2 and T3) is greater than T1. However, the LS model maintains a satisfactory level of accuracy, as evident in the experiments in Table VI, which emphasize the model's robustness in the face of unknown topologies. Addressing the reduction of accuracy in unknown topologies is a necessary requirement that is outside the focus of this paper.

Table VI. Performance for unknown topologies

Scenario	Out of service	MAPE (%)
T1	Line 25-26	6.12
T2	Line 26-27	7.58
T3	G 06	8.64

5.6. Calculation Times

Calculation times for training and testing the proposed GCN model can be seen in Table VII. Table VII shows that the model's training time is 452.14 s, and the testing time is 21.87 s. Since the number of test data is 900 samples, the proposed model processes each sample in 0.023 s. Therefore, the proposed model is fast enough to meet online applications' data processing speed requirement, which is less than 0.033 s [31]. It is imperative to underscore that the model's training process is conducted offline. Within the domain of ML models for LS, the primary focus pivots around accuracy and test time.

Table VII. Calculation time

	Training time(s)	Testing time(s)
Proposed method	452.14	21.87

6. Conclusion

Promising advances have been made in data-driven LS methods, which have yielded promising results. However, the issue of unknown faults challenges implementing these methods in real-world power systems. This paper proposes an innovative GCN-based TL method for ELS focusing on rotor angle instability to solve the problem of unknown faults. The proposed method has been implemented on the IEEE 39 bus system, and the simulation results show its successful performance. In the proposed method, TL is used for updating; in this situation, the model does not need an extensive database for adequate updating. This reduces the generation time of target domain samples and makes the proposed method more suitable for practical online applications. The value of the proposed model for the evaluation indices MAPE and RMSE is 4.21% and 3.92%, respectively, less than other models. Also, using TL has reduced the database size for practical training by 78.1%, reducing the update time. The literature does not report similar works. TL can be an up-and-coming method for solving other problems, such as unknown topologies and operating conditions in data-driven

ELS. Since it is possible to change the topology and the occurrence of an unknown fault simultaneously in real-world power systems, the authors consider investigating unknown faults and topologies together for future work.

Reference

[1] Azad, S., Ameli, M. T., Amiri, M. M., Ameli, H., Shadi, M. R., & Strbac, G. (2024). Real-time Voltage Stability Assessment with a Novel Bus Index Considering Impact of Connection to Distribution Networks. *IEEE Access*.

[2] Ebrahimi, R., & Ahmadi, A. (2023). Real-time under-voltage load shedding method considering integer-value load model and feeder participation in demand response. Electric Power Systems Research, 217, 109115.

[3] Zhai, C., Zhang, H., Xiao, G., & Pan, T. C. (2019). A model predictive approach to protect power systems against cascading blackouts. International Journal of Electrical Power & Energy Systems, 113, 310-321.

[4] Haes Alhelou, H., Hamedani-Golshan, M. E., Njenda, T. C., & Siano, P. (2019). A survey on power system blackout and cascading events: Research motivations and challenges. Energies, 12(4), 682.

[5] Malekpour, A. R., & Seifi, A. R. (2009). An Optimal Load Shedding Approach for Distribution Networks with DGs Considering Capacity Deficiency Modelling of Bulked Power Supply. Modern Applied Science, 3(5), 143.

[6] Laghari, J. A., Mokhlis, H., Karimi, M., Bakar, A. H. A., & Mohamad, H. (2014). A new under-frequency load shedding technique based on combination of fixed and random priority of loads for smart grid applications. IEEE Transactions on Power Systems, 30(5), 2507-2515.

[7] Hong, Y. Y., & Hsiao, C. Y. (2021). Under-frequency load shedding in a standalone power system with wind-turbine generators using fuzzy PSO. IEEE Transactions on Power Delivery, 37(2), 1140-1150.

[8] Mohammadi, J., & Ajaei, F. B. (2020). Adaptive time delay strategy for reliable load shedding in the direct-current microgrid. Ieee Access, 8, 114509-114518.

[9] Rahmani, S., Rezaei-Zare, A., Rezaei-Zare, M., & Hooshyar, A. (2019). Voltage and frequency recovery in an islanded inverter-based microgrid considering load type and power factor. IEEE Transactions on Smart Grid, 10(6), 6237-6247.

[10] Usman, M., Amin, A., Azam, M. M., & Mokhlis, H. (2018, April). Optimal under voltage load shedding scheme for a distribution network using EPSO algorithm. In 2018 1st International Conference on Power, Energy and Smart Grid (ICPESG) (pp. 1-6). IEEE.

[11] Tamilselvan, V., & Jayabarathi, T. (2016). A hybrid method for optimal load shedding and improving voltage stability. Ain Shams Engineering Journal, 7(1), 223-232.

[12] Mollaiee, A., Ameli, M. T., Azad, S., Nazari-Heris, M., & Asadi, S. (2023). Data-driven power system security assessment using high content database during the COVID-19 pandemic. International Journal of Electrical Power & Energy Systems, 150, 109077.

[13] Babaali, A. H., & Ameli, M. T. (2023). Weighted ensemble learning for real‐time short‐term voltage stability assessment with phasor measurements data. IET Generation, Transmission & Distribution, 17(10), 2331-2343.

[14] Li, Y., Cao, J., Xu, Y., Zhu, L., & Dong, Z. Y. (2024). Deep learning based on Transformer architecture for power system short-term voltage stability assessment with class imbalance. Renewable and Sustainable Energy Reviews, 189, 113913.

[15] Vineeth, V. V., & Vijayalakshmi, V. J. (2024). A novel deep learning approach for estimating and classifying short-term voltage stability events in modern power systems with composite load and distributed energy resources. Electrical Engineering, 1-13.

[16] Kim, C., Kim, K., Balaprakash, P., & Anitescu, M. (2019, August). Graph convolutional neural networks for optimal load shedding under line contingency. In 2019 ieee power & energy society general meeting (pesgm) (pp. 1-5). IEEE.

[17] Huang, Q., Huang, R., Hao, W., Tan, J., Fan, R., & Huang, Z. (2019). Adaptive power system emergency control using deep reinforcement learning. IEEE Transactions on Smart Grid, 11(2), 1171-1182.

[18] Dai, Y., Xu, Y., Dong, Z. Y., Wong, K. P., & Zhuang, L. (2012). Real-time prediction of event-driven load shedding for frequency stability enhancement of power systems. IET generation, transmission & distribution, 6(9), 914-921.

[19] Wang, Q., Li, F., Tang, Y., & Xu, Y. (2019). Integrating model-driven and data-driven methods for power system frequency stability assessment and control. IEEE Transactions on Power Systems, 34(6), 4557-4568.

[20] Li, Q., Xu, Y., & Ren, C. (2020). A hierarchical data-driven method for event-based load shedding against fault-induced delayed voltage recovery in power systems. IEEE Transactions on Industrial Informatics, 17(1), 699-709.

[21] Liu, J., Zhang, Y., Meng, K., Dong, Z. Y., Xu, Y., & Han, S. (2022). Real-time emergency load shedding for power system transient stability control: A risk-averse deep learning method. Applied Energy, 307, 118221.

[22] Hu, Z., Shi, Z., Zeng, L., Yao, W., Tang, Y., & Wen, J. (2023). Knowledge-enhanced deep reinforcement learning for intelligent event-based load shedding. International Journal of Electrical Power & Energy Systems, 148, 108978.

[23] Chen, H., Zhuang, J., Zhou, G., Wang, Y., Sun, Z., & Levron, Y. (2023). Emergency load shedding strategy for high renewable energy penetrated power systems based on deep reinforcement learning. *Energy Reports*, *9*, 434-443.

[24] Tang, H., Mi, Y., Xue, F., & Cao, Y. (2020). An integration model based on graph convolutional network for text classification. IEEE Access, 8, 148865-148876.

[25] Azad, S., & Ameli, M. T. (2024). A domain adaptation-based convolutional neural network incorporating data augmentation for power system dynamic security assessment. The Journal of Engineering, 2024(7), e12400.

[26] Azad, S., & Ameli, M. T. (2025). An Imbalanced Deep Learning Framework for Pre-Fault Flexible Multi-Zone Dynamic Security Assessment Via Transfer Learning Based Graph Convolutional Network. *Results in Engineering*, 104172.

[27] Li, Y., Zhang, M., & Chen, C. (2022). A deep-learning intelligent system incorporating data augmentation for short-term voltage stability assessment of power systems. Applied Energy, 308, 118347.

[28] Shobeiry, S. M., Azad, S., & Ameli, M. T. (2023). Smart Short-Term Load Forecasting through Coordination of LSTM-Based Models and Feature Engineering Methods during the COVID-19 Pandemic. *Electric Power Components and Systems*, *51*(2), 171-187.

[29] Li, Q., Xu, Y., & Ren, C. (2020). A hierarchical data-driven method for event-based load shedding against fault-induced delayed voltage recovery in power systems. *IEEE Transactions on Industrial Informatics*, *17*(1), 699-709.

[30] Liu, S., Mao, D., Xue, T., Tang, F., Li, X., Liu, L., ... & Zhang, M. (2021). A Data-Driven Approach for Online Inter-Area Oscillatory Stability Assessment of Power Systems Based on Random Bits Forest Considering Feature Redundancy. Energies, 14(6), 1641

[31] Aghdam, T. S., Karegar, H. K., & Zeineldin, H. H. (2017). Transient stability constrained protection coordination for distribution systems with DG. *IEEE Transactions on Smart Grid*, *9*(6), 5733-5741.