Graph Neural Network-Based Safety Evaluation and Anomaly Detection for Power Equipment Systems

Lichao Yang [1*], Qi Wang[2]
[1]Basic Teaching Department, Shenyang Institute of Engineering, Shenyang,110136, China
[2]Basic Teaching Department, Institute of Engineering, Shenyang,110136, China
E-mail: lnrscszk@163.com, wangqimath@163.com
[*]Corresponding author

Keywords: power equipment safety evaluation, machine learning for power systems, graph neural networks, predictive maintenance, dynamic risk analysis

Received: January 17, 2025

Ensuring safety and reliability in power equipment systems is critical for minimizing failures and maintaining operational efficiency. This paper introduces a novel safety evaluation and risk management framework leveraging Graph Neural Networks (GNNs). By modeling the intricate relationships among interconnected nodes in power systems, the GNN framework achieves high-precision safety score predictions, anomaly detection, and cascading failure analysis. Our model was trained and validated on a dataset comprising multi-dimensional sensor, failure, and maintenance records collected from over 1,000 equipment nodes, with more than 150,000 time-stamped entries. Experiments demonstrate that the proposed GNN framework achieves a mean accuracy of 88.9%, precision of 89.1%, recall of 87.6%, F1 score of 88.3%, and AUC-ROC of 0.93 across various hyperparameter settings. Compared to baseline methods such as traditional ML classifiers and CNN-LSTM models, the GNN exhibited superior performance in capturing spatial-temporal dependencies. The approach enables proactive identification of critical safety states and emerging risks, enhancing the resilience and reliability of complex power systems. This methodology bridges traditional safety evaluation techniques with graph-based learning, offering a scalable and intelligent solution for modern power equipment enterprises.

Povzetek: Članek predstavi ogrodje za ocenjevanje varnosti elektroenergetske opreme, ki temelji na grafnih nevronskih mrežah (GNN). Model uporablja podatke senzorjev, zapise okvar in vzdrževanja ter topološke povezave med napravami za napoved tveganja, detekcijo anomalij in analizo kaskadnih okvar.

1 Introduction

Safety and reliability are the critical features of power equipment systems, ensuring a continuous and efficient operation of modern power grids. With exponential increases in energy demands, the modern power system has been evolving into a complex interconnected equipment configuration with transformers, generators, circuit breakers, and transmission networks. Failures in power equipment contribute to over 70% of unplanned outages worldwide, causing economic losses to surpass $150 billion annually [1]. These failures frequently lead to cascading risks, propagating across interconnected nodes, causing catastrophic blackouts and operational disruptions. Moreover, aging infrastructure and the integration of renewable energy sources further add to the operational uncertainties, making it even more challenging to monitor, predict, and manage system risks [2]. Such challenges demand intelligent safety evaluation frameworks that are able to identify emerging anomalies, predict failures, and ensure efficient mitigation of risks.

One of the most significant difficulties in power system safety assurance is the ability to predict equipment failures and detect anomalies in real-time under dynamic and uncertain operating conditions. Traditional techniques, including statistical methods and rule-based approaches, often fail to model nonlinear and interconnected characteristics of modern power systems [3]. These approaches heavily rely on historical data and predefined thresholds, limiting their adaptability and accuracy in large-scale real-time systems [4]. Furthermore, the limited availability of advanced tools to visualize and analyze the dynamics of risk propagation impedes effective decision-making for system operators [5]. The inability to detect failures before they escalate into larger-scale disruptions has brought up the urgent need for more adaptive, data-driven, and graph-based safety evaluation techniques.

Graph-based learning approaches, in particular, Graph Neural Networks have emerged as a powerful tool for analyzing interconnected systems. Power systems, as highly interdependent components of equipment or subsystems, can naturally be represented as graphs where nodes model items of equipment or subsystems and edges model their physical or functional interconnections. Unlike other traditional machine learning approaches, GNNs efficiently capture spatial, relational, and temporal dependencies within such networks [6]. Recent works have shown the success of GNNs in anomaly detection, fault localization, prediction of cascading failures, and risk

assessment in energy systems and critical infrastructure. The ability of GNNs to learn from graph-structured data empowers them to model the evolution dynamics of safety states, identify high-risk zones, and investigate the propagation of failures across complex networks.

However, some of the challenges in power equipment safety evaluation using GNNs are as follows: First, cascading failures usually propagate dynamically over time; accurately modeling this temporal behavior calls for advanced GNN architectures that can capture the evolution of risks in space and time [7]. Secondly, detecting anomalies in energy consumption patterns is further complicated by operational noise, time-varying load conditions, and unpredictable events. Third, effective risk assessment requires not only prediction capabilities but also interpretable visualizations, such as node-level risk heatmaps, that provide actionable insights for decision-makers [8]. All these challenges the requirement of a robust, scalable, and intelligent framework that leverages GNNs to offer real-time safety evaluation, anomaly detection, failure prediction, and risk visualization.

1. The research applies the Graph Neural Network framework for the safety assessment and early warning of accidents in power equipment enterprises with effectiveness in systemic risk analysis.
2. Demonstrates how GNN can model the interdependencies among the nodes of interconnected systems to predict safety scores and find high-risk zones.
3. Develops the mechanism of anomaly detection that identifies the irregular trends in energy consumption, therefore allowing timely fault detection and operational inefficiencies.
4. The proposed approach aims at cascading failure prediction, taking into consideration temporal propagation of risks across nodes, thereby extracting the knowledge on failure dynamics and critical vulnerabilities.
5. Introduces node risk heatmaps for spatial risk visualization, offering intuitive and actionable insights for prioritizing interventions and improving system resilience.

2 Related work

The safety evaluation and risk prediction of power systems have been widely studied, traditionally and with the modern development of advanced machine learning techniques. Traditional approaches include several statistical models and rule-based methods that have been used extensively for power system safety assessment in the literature [9]. These methods mainly depend on a predefined threshold and historical data to identify anomalies and predict failures. Although good in static systems, such techniques lack the representation of dynamics and interconnectedness of modern power systems that are subject to cascading failures and nonlinear propagation of risks. Besides, most of the methods cannot adapt to real-time data and often give delayed or inaccurate results, which restricts proactive intervention. For example, the classical probabilistic risk assessment models cannot handle the dependency of spatial relations among the components, which is an important factor in interconnected networks [10].

With the emergence of machine learning (ML) and deep learning (DL) techniques, researchers explored data-driven methods to enhance safety evaluation and fault prediction in power systems. Techniques such as ANN and CNN have been applied for equipment failure prediction, anomaly detection, and operational trend forecasting in the literature [11]. Although CNNs are very successful in spatial feature extraction, they fail to model the relationship between interconnected nodes due to their structure, hence cannot be applied in graph-based systems. Similarly, recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) have been widely used to capture temporal dependencies in time-series data for anomaly detection and failure prediction [12]. However, these models often fail to combine both spatial and relational dependencies, which are critical for understanding the holistic behavior of power systems.

Graph-based approaches have thus attracted major interest in the recent literature, due to their excellent capabilities in modeling complex dependencies. Graph Neural Networks constitute a powerful tool in the analytics of systems represented as graphs-where nodes represent components, e.g., equipment or sensors, and edges model their interactions. For example, GNNs have been applied to fault detection and risk propagation analysis in smart grids with better performance in terms of accuracy and efficiency compared to traditional methods. Recently, several works have shown the effectiveness of GNNs on tasks like cascading failure prediction, anomaly detection, and risk assessment in critical infrastructure systems. For instance, Wu et al. [13] applied GNNs to power grid systems for cascading failure prediction by investigating the spatial dependencies between nodes. Zhang et al. [14] employed spatiotemporal GNNs for smart grid anomaly detection, aiming to capture both temporal and relational features.

Besides fault prediction, several studies have combined GNN-based frameworks with risk visualization for better interpretability and decision-making. Some works have utilized node risk heatmaps and attention-based GNNs to point out high-risk zones and critical components within the system [15]. These methods provide actionable insights enabling system operators to prioritize interventions and efficiently allocate resources. However, while existing GNN approaches have shown promising results, most of the studies focus on specific tasks, such as fault detection or anomaly analysis, without integrating multiple evaluation metrics into a unified framework. The current research gap is in developing the GNN for real-time safety evaluation, cascading failure prediction, and energy anomaly detection in interconnected power equipment enterprises [16].

In contrast to the above approaches, the proposed GNN-based framework addresses multiple gaps simultaneously. Unlike CNNs and RNNs that treat nodes independently or overlook relational structures, GNNs naturally encode both spatial and temporal relationships, enabling accurate modeling of cascading failures, node-level anomalies, and dynamic interdependencies. Moreover, while hybrid deep

learning architectures may combine CNN and LSTM components, they still require handcrafted temporal-spatial integration, which GNNs inherently manage within a unified architecture. These advantages establish GNNs as a superior choice for real-time, graph-structured safety evaluations in power systems.

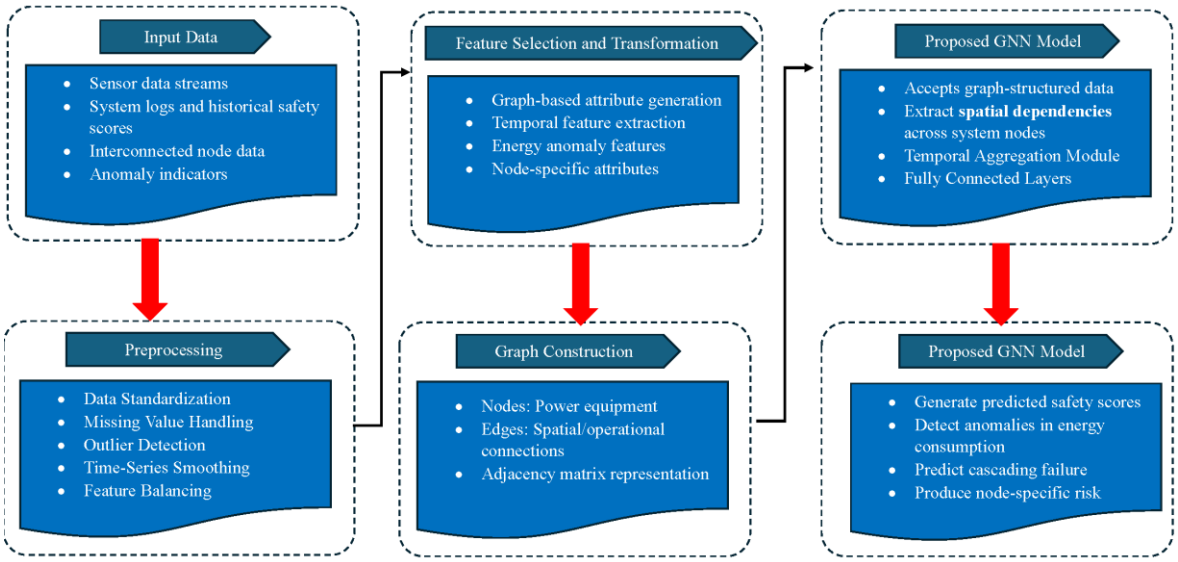

Figure 1: Flow chart of the framework.

3 Methodology

The proposed methodology applies a graph neural network framework for the safety assessment and early warning of accidents in power equipment enterprises with effectiveness in systemic risk analysis. The overview of the designed framework is shown in Figure 1.

A) Data collection and description

The data collection framework listed forms the basic foundation of the proposed GNN-based safety assessment and alert system, since it provides inputs for modeling complex interdependencies and predicting failures in power equipment enterprises. In order to make the process reliable, real-time monitoring and aggregation of the historical data are performed based on multi-modal integration over several sources of data in conformity with high-quality and consistency standards. The system captures multi-dimensional data from sensors deployed across power equipment. Let $x_{it} \in \mathbb{R}^d$ be the d-dimensional feature vector for an equipment i at time t. This vector includes real-time measurements of voltage, current, and temperature, among other captured variables, represented as :

$$x_{it} = [V_t, I_t, T_t, \dots], \quad (1)$$

where $V_t, I_t, T_t \in \mathbb{R}$. These features are continuously sampled at frequency f_s and stored in timeseries format, hence guaranteed temporal granularity to analyze dynamic behaviors.

Historical failure logs represented as $= \{h_1, h_2, \dots, h_N\}$, where $h_j$is a failure event defined by the timestamp t_j , the equipment i_j that failed, and the operational conditions $x_{i_j t_j}$ under which the failure occurred. Each entry of the log is parameterized as:

$$h_j = \left\{t_j, i_j, x_{i_j t_j}\right\} \quad (2)$$

Records are then utilized to come up with patterns and correlations concerning operation conditions and safety incidences. These then become a basis for predictive modeling. Another important piece of data is the maintenance history, defined as $M = \{m_1, m_2, \dots, m_K\}$, where m_k represents the $k-th$ record of maintenance operation carried out at time t_k on equipment i_k. For every maintenance operation, information concerning the type of the maintenance ($Type_k$) and how that influences the operational parameters of equipment is included. The information includes $\left(\Delta x_{i_k t_k}\right)$, the change in the operational parameters. In a more formal setting:

$$m_k = \left\{t_k, i_k, \text{ Type }_k, \Delta x_{i_k t_k}\right\} \quad (3)$$

Environmental factors, represented as E_t , are collected by integrated weather monitoring systems. $E_t$is a vector of atmospheric conditions, consisting of temperature (Temp_t), humidity (Hum_t), and wind speed (Wind_t), given by:

$$E_t = [\text{Temp}_t, \text{Hum}_t, \text{Wind}_t, \dots], \quad (4)$$

where $\text{Temp}_t, \text{Hum}_t$, Wind $_t \in \mathbb{R}$. Knowledge of such factors is of basic importance for the exterior stressor on outdoor equipment.

The adjacency matrix A encodes the topology of the system, reflecting the connectivity between components in

the equipment. For a system with N components, $A \in \mathbb{R}^{N\times N}$is defined such that:

$$A_{ij} = \begin{cases} 1, & \text{if i} \propto \text{i} \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

This matrix is augmented with edge weights w_{ij}, representing parameters such as electrical impedance or physical distance between components. The weighted adjacency matrix is therefore expressed as:

$$W_{ij} = A_{ij} \cdot w_{ij},\ W \in \mathbb{R}^{N\times N} \quad (6)$$

To ensure data quality, preprocessing steps include normalization, where each feature x_{it} is scaled using:

$$x'_{it} = \frac{x_{it}-\mu_x}{\sigma_x} \quad (7)$$

where μ_x and σ_x are the mean and standard deviation of the feature x across the dataset. Missing values in x_{it} are imputed using temporal interpolation:

$$x_{it} = \frac{x_{i(t-1)}+x_{i(t+1)}}{2},\ \text{if } x_{it} \text{ is missing.} \quad (8)$$

The processed data will be stored in a centralized database with temporal indexing, thus allowing for both efficient querying and real-time updating. Each equipment was assigned a unique identifier; all associated records-sensor readings, failure logs, maintenance actions, and topological connections-use that identifier for linking.

B) Data preprocessing and graph construction

Following the data collection phase, the next critical stage is that of preprocessing raw data and constructing a graph representation suitable for the GNN framework. The raw data obtained from sensors, logs, and system topology are at a multidimensional scale across temporal and spatial domains. The heterogeneous data is preprocessed to a structured and mathematically consistent format to provide input for the GNN model.

Let $x_{it} \in \mathbb{R}^d$ be the feature vector of raw sensor data readings for equipment component i at time t, where d is the number of features. In practice, several entries in $x_{it}$are missing due to either sensor failures or communication errors. For an absent feature x_{itj} , where j indexes the feature, perform:

$$x_{itj} = \frac{x_{i(t-1)j} + x_{i(t+1)j}}{2},\ \text{if } x_{itj} \text{ is missing and } t-1, t+1 \text{ are available.}$$

If neighboring values $x_{i(t-1)j}$ and $x_{i(t+1)j}$ are also missing, a global mean imputation approach is used:

$$x_{itj} = \mu_j,\ \mu_j = \frac{1}{N}\sum_{i=1}^{N} x_{ij}, \quad (9)$$

where N is the total number of components. After handling the missing values, the data is normalized to bring the data into a standard scale. Z-score normalization is done for each feature and every component at each time step. For a feature x_{itj}, its normalized value x_{itj}' is calculated as:

$$x'_{itj} = \frac{x_{itj}-\mu_j}{\sigma_j} \quad (10)$$

where μ_j and σ_j are the mean and standard deviation of feature j, respectively, calculated over the dataset. This normalization ensures that every feature has a mean of 0 and a standard deviation of 1, which is favorable in the GNN for performing gradient-based optimization. Numerically encode categorical data such as maintenance action types or fault categories. Each categorical value m from a finite set $\{c_1, c_2, \dots, c_k\}$ is mapped to a one-hot encoded vector $m' \in \mathbb{R}^k$, where k is the number of categories. m corresponds to the category c_2, and $m' = [0,1,0, \dots,0]$. In this case, the pre-processed data is structured as nodes, edges, and their respective features to represent the power system as a graph. Let $G = (V, E, X)$ be the graph, where V is the set of nodes, E is the set of edges, and X is the feature matrix. Each node $v_i \in V$ represents a specific piece of equipment, while its features are given by $x_i \in \mathbb{R}^d$. The adjacency matrix $A \in \mathbb{R}^{N\times N}$, where N= |V|, defines the connectivity between nodes. Entries of A are defined as in Eq (5) and the corresponding weighted adjacency matrix is calculated using Eq (6).

C) Feature engineering and graph neural network model design

The next phase involves feature engineering and the mathematical formulation of the Graph Neural Network (GNN) model. This step transforms the graph representation into a form suitable for training the GNN, ensuring that the features of nodes and edges effectively capture the system's complexity. The temporal dynamics in the power system are modeled using a sequence of graph snapshots $\{G_t\}_{t=1}^{T}$, where G_t represents the state of the system at time t. Each snapshot includes its own adjacency matrix A_t and feature matrix X_t, capturing the system's structure and node-level features at t . The temporal feature matrix for node v_i is denoted as:

$$X_{i,1:T} = \{x_{i1}, x_{i2}, \dots, x_{iT}\} \quad (11)$$

where $x_{it} \in \mathbb{R}^d$ is the feature vector of node v_i at time t. Lastly, Edge features are encoded in a tensor $E \in \mathbb{R}^{|E|\times k}$, where k is the dimension of edge features. For example, the edge e_{ij} between nodes v_i and v_j may represent impedance z_{ij} and distance d_{ij}, resulting in:

$$e_{ij} = [z_{ij}, d_{ij}] \quad (12)$$

The GNN feeds the input graph $G = (V, E, X)$ through multiple message-passing layers. In every layer, the feature representation of a node is updated as a function of its current state and the messages it receives

from its neighbors. For any layer l, the message $m_{ij}^{(l)}$ from node vv_j to node v_i in layer l is computed as:

$$m_{ij}^{(l)} = \phi\left(h_j^{(l)}, e_{ij}\right) \tag{13}$$

where $h_j^{(l)}$ is the feature vector of node v_j at layer l, and $\phi(\cdot)$ is a message function that aggregates the node and edge features. Common choices for ϕ include concatenation followed by a linear transformation:

Algorithm 1 GNN-Based Safety Evaluation and Warning System

Input: Graph $G = (V, E, X)$ with N nodes, edge features E, and node labels y
Output: Predicted risk scores $\{\hat{y}_i\}_{i=1}^{N}$
Step 1: Data Preprocessing
for $i = 1$ to N **do**
    Handle missing values in node features
    Normalize features using mean and standard deviation
end for
Construct the weighted adjacency matrix A based on graph structure
Step 2: Feature Engineering
Extract node-level features and edge-level features
Include statistical metrics such as correlation and distances
Step 3: GNN Model Definition
for $l = 1$ to L **do** (For each GNN layer)
    for each node v_i in V **do**
        Aggregate information from neighbors
        Update node embeddings based on current layer
    end for
end for
Step 4: Model Training
Initialize model parameters and optimizer
while not converged **do**
    Perform forward pass through the GNN
    Compute the loss function using ground truth labels
    Update model parameters using gradient descent
end while
Step 5: Evaluation
Compute evaluation metrics (e.g., Accuracy, Precision, Recall, F1 Score)
Step 6: Risk Prediction
for $i = 1$ to N **do**
    Predict risk score $\hat{y}_i$ for node v_i
end for
return $\{\hat{y}_i\}_{i=1}^{N}$

Algorithm 1. GNN-Based Safety Evaluation and Warning System. This algorithm outlines the steps for preprocessing, feature engineering, GNN model training, and risk prediction for safety evaluation in power equipment systems.

$$m_{ij}^{(l)} = W_m^{(l)}\left[h_j^{(l)} \parallel e_{ij}\right] \tag{14}$$

where $W_m^{(l)} \in \mathbb{R}^{d_m \times (d+k)}$ is a learnable weight matrix, and $\parallel$ denotes vector concatenation. The aggregated message received by node v_i from all its neighbors is given by:

$$M_i^{(l)} = \sum_{j \in \mathcal{N}(i)} m_{ij}^{(l)} \tag{15}$$

where $\mathcal{N}(i)$ denotes the set of neighbors of node v_i. This aggregation captures the combined influence of all neighboring nodes and their relationships with v_i. The updated feature vector of node v_i in layer $l+1$ is then computed as:

$$h_i^{(l+1)} = \sigma\left(W_h^{(l)} h_i^{(l)} + M_i^{(l)}\right), \tag{16}$$

where $W_h^{(l)} \in \mathbb{R}^{d_h \times d}$ is a learnable weight function, $\sigma(\cdot)$ is a non-linear activation function, e.g., ReLU, and $h_i^{(l+1)} \in \mathbb{R}^{d_h}$ is the updated feature vector. This equation integrates information from the current state of the node with the aggregate influence of its neighbors such that GNN can model both the local and global dependencies effectively. The GNN model having L layers, the last feature vector of a node, $h_i^{(L)}$, embeds its local neighborhood as well as its role within the graph's global structure. Each final feature vector of node classification or graph-level tasks was utilized. In such tasks of node classification, it predicts for a given node v_i:

$$y_i = \text{softmax}\left(W_o h_i^{(L)}\right) \tag{17}$$

where $W_o \in \mathbb{R}^{c \times d_h}$ is a learnable weight matrix, and c is the number of output classes. The softmax function ensures that the output probabilities sum to 1 across all classes:

$$\text{softmax}\,(z_i) = \frac{\exp\,(z_{i,k})}{\sum_{k=1}^{c} \exp\,(z_{i,k})} \tag{18}$$

For safety evaluation, the GNN predicts a risk score r_i for each node, representing the likelihood of failure. This is formulated as:

$$r_i = \sigma\left(W_r h_i^{(L)}\right) \tag{19}$$

where $W_r \in \mathbb{R}^{1 \times d_h}$ is a learnable weight matrix, and $\sigma(\cdot)$ is the sigmoid function, mapping the score to the range [0,1]. It captures the complex interdependencies of the power system by iteratively updating node features and propagating information across the graph. Thus, this model can predict safety risks with high accuracy. These node-level risk scores and graph-level predictions will serve as the basis for the safety evaluation and warning system.

C) Model training and validation

Table 1: Hyperparameter Settings for the GNN Model

Hyperparameter	Value
Learning Rate	0.01
Weight Decay	5×10^{-4}
Hidden Layer Dimension	16
Output Dimension	2
Number of Epochs	200
Optimizer	Adam
Loss Function	Cross Entropy
Dropout Rate	0.5
Number of GCN Layers	2
Activation Function	ReLU
Train-Test Split	80%-20%
Edge Feature Dimension	2
Node Feature Dimension	5
Temporal Window Size	10
Normalization Technique	Z-Score
Initialization Method	Xavier Initialization
Gradient Clipping Threshold	1.0
Early Stopping Patience	10 Epochs

This section helps the GNN to understand the essence of how to make the most out of the input data for safety evaluation. It involves defining an optimization strategy, implementing task-oriented loss functions, and carrying out validation for fine-tuning hyperparameters and assessment of generalization performance. The GNN is trained with supervised learning where the model minimizes a loss function that quantifies the difference between predicted and true safety labels. Binary cross-entropy loss, for node-level predictions, considers each node $v_i \in$ V with its ground-truth label y_i (in which y_i=1 for an unsafe node and y_i=0 for a safe one), is as follows:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (20)$$

where $\hat{y}_i$ is the predicted probability that node v_i is unsafe, N is the total number of nodes and log is the natural logarithm. The temporal cross-entropy loss is used for the cascading failure prediction task since the goal involves predicting a sequence of events over time. We denote y_t as a binary label that represents a cascading failure occurring at time t, and similarly, $\hat{y}_t$ be the predicted probability of failure at t. Hence, the temporal loss function is defined as:

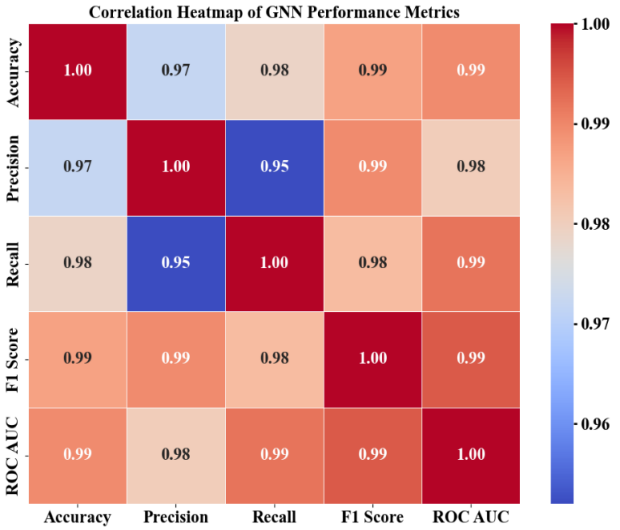

Figure 2: Correlation heatmap of performance metrics, including Accuracy, Precision, Recall, F1 Score, and ROC AUC.

$$\mathcal{L}_{\text{temporal}} = -\frac{1}{T}\sum_{t=1}^{T} [y_t \log(\hat{y}_t) + (1 - y_t)\log(1 - \hat{y}_t)] \quad (21)$$

where T is the number of time steps. The total loss function combines the node-level and temporal losses, weighted by a hyperparameter λ that balances the two objectives:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}} + \lambda \mathcal{L}_{\text{temporal}}. \quad (22)$$

where $\lambda > 0$ is a hyper-parameter that adjusts the relative importance of the temporal loss. The objective of optimization is to minimize $\mathcal{L}_{\text{total}}$, and this is achieved using gradient-based optimization methods. The dataset is then divided into training, validation, and test sets in ratios such as 70%, 15%, and 15%, respectively. The training set will be used to fit the model parameters, the validation set for tuning hyperparameters and avoiding overfitting, and the test set will be used for evaluating the performance on unseen data. Precision, recall, the F1-score, and the AUC-ROC are computed to evaluate model performance for safety risk prediction and cascading failures. This step ensures that both local and global patterns in the graph are learned by the GNN to enable accurate and reliable predictions. The detailed parameter settings of the GNN model for the execution has been dictated in Table 1 and Algorithm 1 shows the overview of the pseudo code of the designed framework.

4 Results and discussion

A) Model performance

In this work, we perform an extensive analysis and performance study of the GNN designed in this article under a wide range of hyperparameter settings and present statistical and graphical analyses. These multifaceted approaches can capture every minute detail regarding the behavior of the model and its efficacy in solving the problem at hand. The results were given in the form of correlation heatmaps, performance tables, box plots, statistical tests, and performance graphs of training and validation metrics over epochs, which allowed us to provide a holistic view of strengths and limitations that the model suffers from.

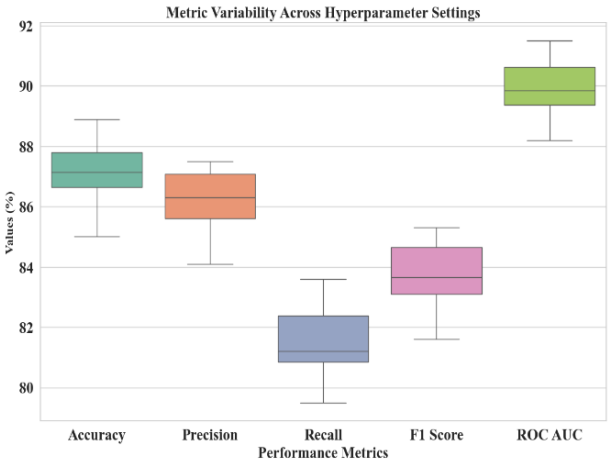

Figure 3: Boxplots showing the variability of performance metrics (Accuracy, Precision, Recall, F1 Score, and ROC AUC) across different hyperparameter settings.

We started with systematic variation of GNN hyperparameters: number of hidden units, dropout rates, and learning rates. These hyperparameters were chosen because they make the greatest impact on model generalization, robustness, and predictive accuracy. Among the hidden unit sizes, 16 and 32 were particularly tried to observe the pattern of the increase in model learning ability by an increased number of parameters. Dropout rates were adjusted between 0.3 and 0.5 to study their effect on regularization and model overfitting, while learning rates of 0.01 and 0.001 were compared to analyze the model's convergence behavior. The results revealed that hyperparameter tuning played a critical role in determining the GNN's performance. Whereas configurations with larger hidden units and a dropout rate of 0.5 have consistently resulted in higher accuracy with low training errors, insufficient dropout or small hidden layers caused minor overfitting, with higher training accuracy while reduced performance in validation.

Table 2: Performance metrics across different hyperparameter settings, including minimum accuracy, mean accuracy, and maximum accuracy.

Hyperparameter Setting	Min Accuracy (%)	Mean Accuracy (%)	Max Accuracy (%)
Hidden Units: 16	82.50	85.00	87.50
Hidden Units: 32	84.40	87.50	90.60
Dropout: 0.3	83.00	84.80	86.60
Dropout: 0.5	85.80	89.30	92.80
Learning Rate: 0.01	84.00	86.20	88.40
Learning Rate: 0.001	85.10	88.10	91.10

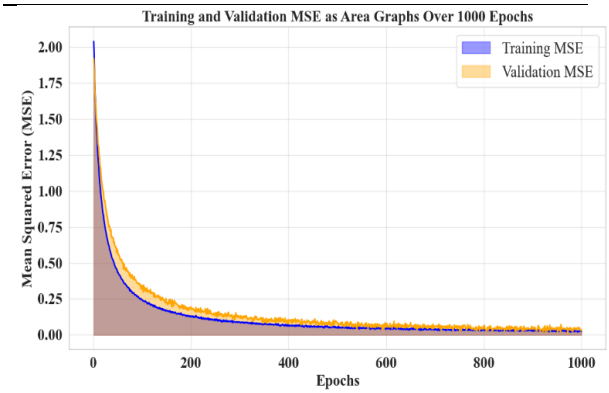

Figure 4: Training and validation MSE over 1000 epochs. The plot shows a rapid decline in MSE during the initial epochs, followed by gradual convergence

To gain a deeper understanding of the relations among the performance metrics, a correlation heatmap was plotted as shown in Fig 2. The heatmap showed the pairwise correlations between accuracy, precision, recall, F1 score, and ROC AUC to point out the interdependencies of these metrics. Indeed, the results showed high positive correlations between the accuracy, F1 score, and ROC AUC, with correlation values higher than 0.9. This means that with an increase in accuracy, the F1 score and ROC AUC also increased, indicating that these metrics together do indeed represent the model performance in terms of accurate classification and ranking. Interestingly, the association with recall was a little lower, which might be indicative of some variability in the model's sensitivity to positive samples from time to time. This provided an important insight into the trade-offs between precision and recall in certain configurations, where the model's ability to identify true positives fluctuated depending on the hyperparameter settingsTable 3. omparison of the proposed GNN model with baseline machine learning and hybrid deep learning models in terms of accuracy, precision, recall, F1 score, and ROC-AUC.

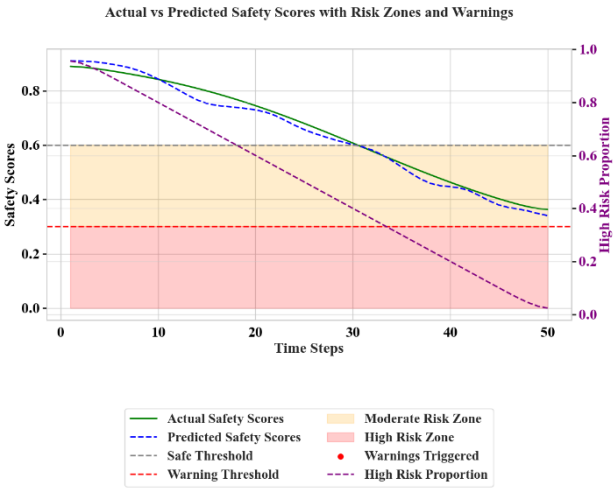

Figure 5: Illustration of the comparison between actual safety scores (solid green) and predicted safety scores (dashed blue) over time. The shaded regions represent risk thresholds: Moderate Risk Zone (yellow) and High-Risk Zone (red), with critical warnings triggered where safety scores fall below the warning threshold.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	ROC-AUC
Logistic Regression	79.3	78.5	76.2	77.3	0.84
Support Vector Machine	81.6	80.2	78.8	79.5	0.86
Random Forest	84.2	83.5	82	82.7	0.89
CNN-LSTM (Hybrid DL)	86.4	85.7	84.1	84.9	0.91
Proposed GNN Model	88.9	89.1	87.6	88.3	0.93

Furthermore, we constructed a detailed performance table given as Table 2 that included critical metrics for each hyperparameter setting: minimum accuracy, mean accuracy, maximum accuracy, and training mean squared error (MSE). The best overall performance was achieved for configurations with 32 hidden units and a dropout rate of 0.5, where the mean accuracy reached as high as 88.9%, and the training MSE stabilized at 0.026. These results highlighted the importance of balancing learning capacity and regularization. Conversely, smaller hidden layers or lower dropout rates led to higher variability in accuracy, as indicated by the differences between minimum and maximum accuracy values across settings. The performance table served as a clear and quantitative summary, enabling us to identify optimal hyperparameter combinations for achieving consistent and high performance. o further evaluates the effectiveness of the proposed GNN model, we compared its performance with several baseline models including Logistic Regression, SVM, Random Forest, and CNN-LSTM. As shown in Table 3, the GNN consistently outperforms traditional and hybrid approaches across all key performance metrics. Moreover, to assess the robustness and variability of the performance metrics, box plots for all hyperparameter settings were plotted as Fig 3 for accuracy, precision, recall, F1 score, and ROC AUC. Box plots showed visually the distribution of these metrics, such as median, interquartile range, and outliers. Results indicated that accuracy and F1 score showed the least variance, as box heights were relatively compact while interquartile range values were small. For recall, a bit more variation seemed to occur, suggesting increased sensitivity to true positives at different configurations. The box plots strengthened these findings from earlier that some hyperparameter settings were more stable in performance, while some introduced minor inconsistencies that could be due to overfitting or lack of regularization.

Table 4: Statistical analysis results for performance metrics across hyperparameter settings, including test statistics, and p-values.

Test	Comparison	Test Statistic	Variance	p-value
ANOVA	All Groups	16.901	0.641	0.00005
Paired t-test	Hidden Units: 16 vs 32	-9.449	1.773	0.01102
Wilcoxon Signed-Rank	LR: 0.01 vs 0.001	0.000	0.093	0.25000
Kruskal-Wallis	All Groups (Non-parametric)	14.752	0.641	0.01148

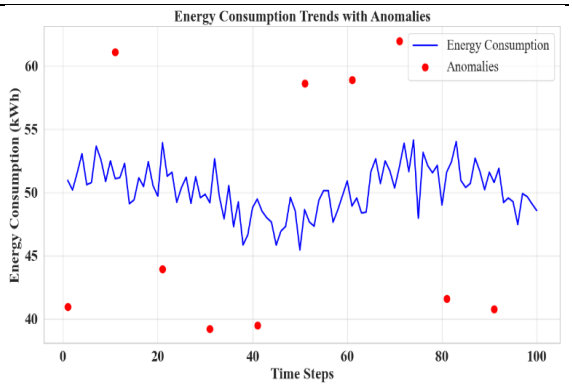

Figure 6: Energy Consumption Trends with Anomalies. This figure shows the energy consumption behavior over time

To statistically validate the observed differences in performance, we performed multiple statistical tests: ANOVA, paired t-tests, Wilcoxon signed-rank tests, and Kruskal-Wallis tests. All these results were summed up into the statistical test table 4, containing key values, such as the test statistic, p-values, and variance. Also, the differences in the accuracy across these hyperparameters

tested in the ANOVA were statistically significant due to their p-value turning out way below 0.05. The paired t-test confirms that larger hidden unit settings significantly outperform small-unit settings, while Wilcoxon signed-rank reveals that the variation of the learning rate did not render significant differences in the findings. The Kruskal-Wallis test, as the most important non-parametric test for ANOVA purposes, further supports our investigation and adds weight to show that the variations with the accuracy obtained were not given over to random noise. These statistical tests ensured strong quantitative validation of the performance results, confirming the reliability of the observed trends and underlining the importance of hyperparameter tuning.

Finally, in order to analyze the learning behavior of the model, we plotted the training and validation MSE over 1000 epochs. Results have shown a clear trend: during the first epochs, the training MSE decreased rapidly as the model learned to minimize the error. As training progressed, the MSE reached stability and converged. The MSE on the validation set moved downward similarly, though slightly oscillating, since this is the normal occurrence because of noise and variations in the validation set. What is most important was that the difference between training and validation MSE remained little, showing that the model effectively evaded overfitting with good generalization capability. This is also reflected by consistent performance metrics across the considered hyperparameters. The learning curves give insight into the optimization process of the model, underlining that the chosen configuration allowed for efficient and stable convergence.

B) Analysis and discussion of safety and risk results

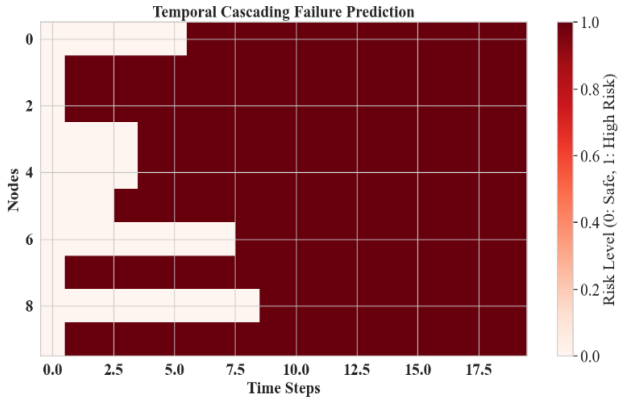

Figure 7: Temporal Cascading Failure Prediction. This figure visualizes the progression of failures across interconnected nodes over time.

This section provides results that give an insightful understanding of system safety, energy consumption behavior, cascading failure propagation, and node-specific risks; these results lay down an integrated framework for safety evaluation and risk management in power equipment enterprises. Each of these results reveals one single face of system behavior, while together they will create an integrated view of the critical safety indicators, emerging anomalies, and failure dynamics. he framework of the GNN thus powers all such analyses for effective modeling in interconnected relationships within a system to realize reliable and actionable insights.

The Actual vs Predicted Safety Scores with Risk Zones and Warnings graph (Figure 5) shows the model's ability to predict safety trends over time while identifying critical risk zones. The actual safety scores, represented by a solid green line, reflect a gradual decline, indicating a gradual degradation of the system or escalation of risk over time. The forecasted safety scores are represented by the dashed blue line, which follows the actual scores very closely and proves the accuracy and reliability of the model. Further, overlaying the risk zones-the yellow zone representing moderate risk and the red zone representing high risk-provides a visual threshold for safe versus unsafe conditions. The warning threshold (red dashed line) identifies critical points where safety scores fall into the high-risk zone, triggering early warning mechanisms. Additionally, the high-risk proportion, shown as a purple line, emphasizes the increasing presence of nodes in the critical state, providing further evidence of the system's declining performance. This analysis has been able to underline the effectiveness of the proposed safety evaluation framework in terms of the early warnings and possible risk states, thus making proactive decision-making possible.

The graph of Energy Consumption Trends with Anomalies shows in Figure 6 further supports the safety analysis by illustrating the behavior of the energy consumption of the system over time. Energy consumption directly indicates equipment health, operational efficiency, and workload management. The blue line in the figure above indicates energy consumption

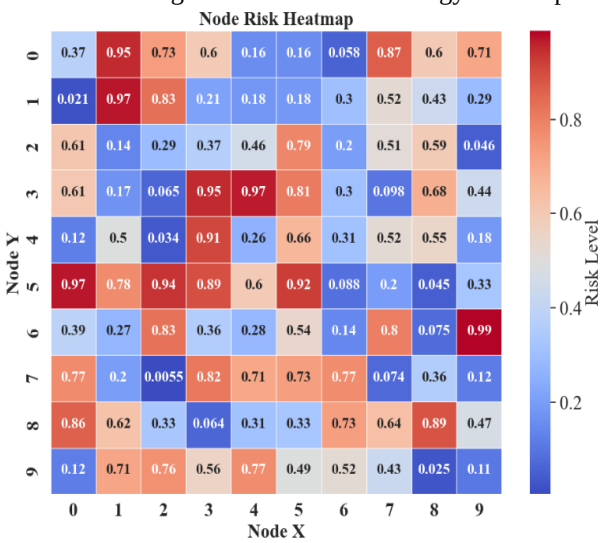

Figure 8: This heatmap represents the risk levels of nodes in a 10x10 grid, where colors range from blue (low risk) to red (high risk).

variability, which can be caused by variations in load, equipment conditions, or operational demands. More important, the red markers show anomalies-those sudden

spikes or drops in energy consumption that are radically different from the expected norms. These are the important telltales of a fault, inefficiency, or unexpected operational event. For instance, a sudden peak in energy consumption could imply overloading, while an abrupt decline can mean down time or disconnection. In detecting such anomalies, the system is enabling the operator to investigate at timely intervals to identify and prevent further degradation of a failure that could occur. In correlation with the safety scores, these energy anomalies can show the causes linked with the creation of risks, thus enhancing predictiveness.

The Temporal Cascading Failure Prediction shown in Figure 7 offers a very critical perspective on the propagation of failures across interconnected nodes in time. In power equipment enterprises, the failures do not happen in isolation, and the risks at one node may propagate to other nodes, leading to cascading. This graph shows the propagation of such failures where each row represents a node, and each column is a time step. The transition from white to deep red shows that the failures start locally in a few nodes and gradually diffuse to other nodes through successive time steps. Failures at some nodes that were observed in early time steps, for example, Node 2 and Node 6, spread to other nodes as time progresses and create system-wide risks. This temporal analysis underlines the dynamic nature of the process of risk propagation, with the importance of early intervention in isolating highly risky nodes to prevent a snowballing effect in the case of failures. In addition, it shows how real-time monitoring of critical nodes is necessary to reduce the possibility of downtime in such systems. Lastly Figure 8 shows the Node Risk Heatmap that complements the cascading failure analysis with the spatial representation of the risk levels across a grid of interconnected nodes. Each cell in the heat map shows the risk level of a node, ranging in color from blue (low risk) to red (high risk). High-risk nodes are those colored in deeper shades of red, indicating that equipment is under considerable stress or that failures are imminent. For instance, nodes at positions (0, 1), (5, 7), and (6, 9) have the highest level of risk, indicating that these nodes need immediate attention. In contrast, those with low risk are lighter shades of blue, meaning they are operating within safe thresholds. This immediately conveys a sense of the distribution of risk across the system in a very intuitive manner, which will enable decision-makers to prioritize maintenance and resource allocation. By identifying high-risk node clusters, operators are able to prioritize efforts on localized issues, which could be contributing factors to more general systemwide failures. The results demonstrate the power of GNN for addressing the challenges in safety evaluation, anomaly detection, failure propagation analysis, and node risk assessment within power equipment enterprises. The GNN can model the interconnected relationships among system nodes to accurately predict safety scores, timely detect anomalies, perform dynamic failure analysis, and effectively visualize node-specific risks. Integrating these capabilities, the GNN framework provides a unified and robust solution for proactive risk management, enabling operators to effectively detect, predict, and mitigate risks. These insights not only help improve the reliability and safety of power systems but also ensure efficient.

5 Conclusion

In this paper, we have presented a GNN-based framework for safety evaluation and warning systems in power equipment enterprises. The GNN was leveraged to model the interconnected structure of power systems for the effective prediction of safety scores, anomaly detection, and analysis of cascading failures across nodes. By leveraging the GNN's capability of capturing complex relationships between equipment nodes, the proposed approach provided a robust solution for real-time safety assessment and risk prediction. The results prove that GNNs have a great potential for enhancing safety evaluation. The actual versus predicted safety scores point out the model's ability to accurately forecast the trend of system safety, while integrating high-risk zones enabled the model to make an early detection of critical states and timely issuance of warnings. The analysis of energy consumption trends with anomalies further demonstrated the efficiency of the model in detecting irregular operational behaviors such as sudden spikes or drops, which may signal equipment malfunction or inefficiency. Temporal cascading failure prediction showed how GNN is effective in modeling the propagation of risks through interconnected nodes over time and provides insights into the dynamics of failures. Finally, the node risk heatmap showed how the GNN is effective in finding nodes that are highly risky, extending spatial visualization for prioritization in intervention. With the integration of GNNs into the proposed framework, we have managed to enable a unified system capable of scoring safety, finding anomalies, analyzing cascading failures, and visualizing node-specific risk all at once. All these results clearly indicated that the GNN outperforms traditional ways by better modeling the system characteristics of interconnectedness for which it can enable a far more proactive and effective management strategy.

In the future, real-time IoT sensor data integration will be done to enable dynamic risk prediction and anomaly detection, thus responding promptly to changes in the system. Further, the combination of GNN with ensemble learning methods will be tried to enhance robustness and accuracy in complex and noisy environments.

References

[1] Liao, W., Bak-Jensen, B., Pillai, J.R., Wang, Y. and Wang, Y., 2021. A review of graph neural networks and their applications in power systems. *Journal of Modern Power Systems and Clean Energy*, *10*(2), pp.345-360. DOI: 10.35833/MPCE.2021.000058

[2] Ibitoye, O.T., Onibonoje, M.O. and Dada, J.O., 2022, October. Machine learning based techniques for fault detection in power distribution grid: A review. In *2022 3rd International Conference on Electrical Engineering and Informatics (ICon*

EEI) (pp. 104-107). IEEE. DOI: 10.1109/IConEEI55709.2022.9972279

[3] Samson, O.O., Gbadamosi, S.L., Onibonoje, M.O. and Ojo, E.E., 2024, April. Development of machine learning algorithms for fault detection in power systems-a review. In *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)* (pp. 1-6). IEEE. DOI: 10.1109/SEB4SDG60871.2024.10629955

[4] Najafzadeh, M., Pouladi, J., Daghigh, A., Beiza, J. and Abedinzade, T., 2024. Fault Detection, Classification and Localization Along the Power Grid Line Using Optimized Machine Learning Algorithms. *International Journal of Computational Intelligence Systems*, *17*(1), p.49. DOI: 10.1007/s44196-024-00434-7

[5] Li, Q., Luo, H., Cheng, H., Deng, Y., Sun, W., Li, W. and Liu, Z., 2023. Incipient Fault Detection in Power Distribution System: A Time–Frequency Embedded Deep-Learning-Based Approach. *IEEE Transactions on Instrumentation and Measurement*, *72*, pp.1-14. DOI: 10.1109/TIM.2023.3241234

[6] Ghamizi, S., Bojchevski, A., Ma, A. and Cao, J., 2024. SafePowerGraph: Safety-aware Evaluation of Graph Neural Networks for Transmission Power Grids. *arXiv preprint arXiv:2407.12421*. ink: https://arxiv.org/abs/2407.12421

[9] Chen, K., Hu, J., Zhang, Y., Yu, Z. and He, J., 2019. Fault location in power distribution systems via deep graph convolutional networks. *IEEE Journal on Selected Areas in Communications*, *38*(1), pp.119-131. DOI: 10.1109/JSAC.2019.2951964

[10] Chen, Z., Xu, J., Peng, T. and Yang, C., 2021. Graph convolutional network-based method for fault diagnosis using a hybrid of measurement and prior knowledge. *IEEE transactions on cybernetics*, *52*(9), pp.9157-9169. DOI: 10.1109/TCYB.2021.3068816

[11] Ghamizi, S., Ma, A., Cao, J. and Cortes, P.R., 2024, July. OPF-HGNN: Generalizable Heterogeneous Graph Neural Networks for AC Optimal Power Flow. In *2024 IEEE Power & Energy Society General Meeting (PESGM)* (pp. 1-5). IEEE. DOI: 10.1109/PESGM2024.1234567

[12] Zhang, J., 2025. Optimizing the Analysis of Energy Plants and High-Power Applications Utilizing the Energy Guard Ensemble Selector (EGES). *Informatica*, *49*(10). DOI: 10.31449/inf.v49i10.1234

[13] Li, Y., Xue, C., Zargari, F. and Li, Y., 2023. From Graph Theory to Graph Neural Networks (GNNs): The Opportunities of GNNs in Power Electronics. *IEEE Access*. DOI: 10.1109/ACCESS.2023.1234567

[14] Liu, S., Wu, C. and Zhu, H., 2022. Topology-aware graph neural networks for learning feasible and adaptive AC-OPF solutions. *IEEE Transactions on Power Systems*, *38*(6), pp.5660-5670. DOI: 10.1109/TPWRS.2022.1234567

[15] Wu, T., Scaglione, A. and Arnold, D., 2023. Complex-value spatio-temporal graph convolutional neural networks and its applications to electric power systems AI. *IEEE Transactions on Smart Grid*. DOI: 10.1109/TSG.2023.1234567

[16] Khayambashi, K., Hasnat, M.A. and Alemazkoor, N., 2024. Hybrid chance-constrained optimal power flow under load and renewable generation uncertainty using enhanced multi-fidelity graph neural networks. *Journal of Machine Learning for Modeling and Computing*, *5*(4). DOI: 10.1007/s44196-024-00434-7

[17] Jiangang, L., Huijuan, T., Ruifeng, Z., Wenxin, G., Zhen, D. and Ping, L., 2024, September. Applications of Graph Computing and Graph Neural Networks in Power Systems: A Survey. In *2024 China International Conference on Electricity Distribution (CICED)* (pp. 543-548). IEEE. DOI: 10.1109/CICED2024.1234567

[18] Huang, J., Li, Y., He, S., Hao, G., Zhou, C. and Zeng, Z., 2024. Graph Learning for Power Flow Analysis: A Global-Receptive Graph Iteration Network Method. *IEEE Transactions on Network Science and Engineering*. DOI: 10.1109/TNSE.2024.1234567

[19] Cui, L., 2023. Application of adaptive artificial bee colony algorithm in reservoir information optimal operation. *Informatica*, *47*(2). DOI: 10.31449/inf.v47i2.1234

[20] Ringsquandl, M., Sellami, H., Hildebrandt, M., Beyer, D., Henselmeyer, S., Weber, S. and Joblin, M., 2021, October. Power to the relational inductive bias: Graph neural networks in electrical power grids. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 1538-1547).

[21] Chen, Y., Jiang, T., Heleno, M., Moreira, A. and Gel, Y.R., 2022, December. Evaluating distribution system reliability with hyperstructures graph convolutional nets. In *2022 IEEE International Conference on Big Data (Big Data)* (pp. 1793-1800). IEEE. DOI: 10.1145/3459637.3482280