# Research On Power System Situation Awareness And Risk Assessment Based On Multimodal Perception And Human-machine Collaboration Using Graph Neural Network

Mingzhe Hou*, Zhenzi Wang, and Zhenhao Ye

*China southern power grid shenzhen power supply Co., Ltd, Shenzhen 518000, China*

*Corresponding author. E-mail: www1233332025@163.com*

*Received: Jun. 01, 2026; Accepted: Jul. 26, 2026*

To address the challenges of lagging situational awareness and insufficient risk assessment accuracy brought about by the high proportion of renewable energy grid integration and the increasingly complex grid topology, this paper proposes a power system situational awareness and risk assessment method integrating graph neural networks (GNNs), multimodal perception, and human-machine collaboration. First, a multimodal perception data system is constructed, which integrates multi-source information such as measurement data, text data, and image data. Heterogeneous data fusion is achieved through data preprocessing and feature alignment. Then, an attention-based graph neural network model (Attention-GNN) is designed. Utilizing the graph structure to adapt to the grid topology characteristics, it mines the deep correlation between multimodal features and grid situational awareness, so that can enable dynamic updates of situational awareness. Next, a human-machine collaborative risk assessment framework is established. The machine model outputs an initial risk level, which is then corrected and optimized by incorporating domain expert experience, constructing a dual-drive assessment mode of "data-driven + knowledge-guided". Finally, experimental verification is conducted based on a real power grid dataset and a simulation platform. Experimental results show that the method proposed in this paper achieves a situational awareness accuracy of 93.2% and a single-step inference time of 28.4 ms, which can meet the real-time requirements (approximately 35 FPS) for video inspections conducted by drones or fixed cameras. The accuracy of human-machine collaborative risk assessment reaches 95.6%, with an accuracy of 92.3% in extreme scenarios. Even with a 40% renewable energy penetration rate, the accuracy remains at 89.5%. In a six-month pilot project in a provincial power grid, the risk identification coverage increased from 53% to 97%, with an average early warning time of 3.2 minutes, which can prevent approximately five potential power outages and reduce direct economic losses by approximately 8 million yuan. This research provides effective technical support for the safe and stable operation of power systems.

**Keywords:** graph neural network, multimodal perception, human-machine collaboration, situational awareness, risk assessment



http://dx.doi.org/10.6180/jase.202611_34.066

## 1. Introduction

With the advancement of the "dual carbon" target, new energy sources such as wind power and photovoltaics have been connected to the grid on a large scale. In 2024, the proportion of new energy power generation in the country exceeded 30%, and the power system presented a complex form of multi-element coordination of "source, grid, load, and storage" and AC/DC hybrid connection [1]. The randomness, volatility, and intermittency of new energy sources lead to rapidly changing grid operation. For example, in 2023, a power grid in a certain region of Northwest China experienced frequency fluctuations to 49.2Hz due to a sudden drop in photovoltaic output; in 2024, a power grid in East China experienced a chain reaction caused by line tripping due to a typhoon. Both examples demonstrate that traditional situational awareness methods relying on single measurement data are insufficient to fully capture the dynamic characteristics of the system, and the risk assessment results are lagging and one-sided, easily leading to safety and stability problems [2]. Furthermore, the digital transformation of the power grid has generated massive amounts of multimodal data, and how to mine the value of this data to achieve accurate situational awareness has become a core pain point for the industry.

Situational awareness and risk assessment are the core links to ensure the safe operation of the power system. The core requirement is to grasp the system's operating status in real time and comprehensively, and to accurately identify potential risks [3]. Multimodal sensing technology can integrate multi-source data such as measurement, text, and images, breaking through the information bottleneck of single data. Graph neural networks, with their advantages in processing non-Euclidean data, are adapted to the graph structure characteristics of the power grid topology. Human-machine collaboration can integrate the data analysis capabilities of machines with the experience and knowledge of experts, which can improve the reliability of decision-making in complex scenarios [4].

Therefore, researching situational awareness and risk assessment methods based on multimodal perception and human-machine collaboration using graph neural networks is of significant theoretical and engineering importance for improving the safe and stable operation of power systems. This research uses multimodal data as a foundation, preprocessing and fusing it before inputting it into a graph neural network model. Combined with human-machine collaboration, accurate assessment is achieved. Through experimental and engineering verification and iterative optimization, a feasible technical solution is ultimately developed.

## 2. Related work

In terms of power system situational awareness, Chen et al. [5] used BiLSTM and CRF technology to extract entities and relationships from the security database to construct a knowledge graph, and combined it with network scanning and monitoring data to integrate topology, environment, and vulnerability information, thereby identifying situational elements. This method calculates the maximum possible attack path based on detected attacks and evaluates its probability of occurrence, net benefit, and coverage. Sahu et al. [6] established an initial Bayesian attack graph based on the actual power grid cyber-physical model. This method defines attack state nodes and state transition edges, sets a prior probability distribution, and uses collected data to update the posterior probability through Bayesian inference, thereby calculating the probability of security vulnerability events. However, the prior distribution of this method depends on expert experience and is difficult to update adaptively in scenarios where the power grid operation mode changes frequently, which limits its applicability in actual dynamic power grids.

In data-driven event detection research, based on PMU streaming data, Das et al. [7] established a core framework encompassing heterogeneous data processing, situational awareness, fault localization, and resilience assessment; they utilized Principal Component Analysis (PCA) to analyze bus state similarity and Local Outlier Factor (LOF) for event identification. While PCA projects high-dimensional measurement data into a low-dimensional principal component space and LOF detects anomalies by comparing the local density deviation of a sample against its k-nearest neighbors, such methods rely on extensive historical bus measurement data and labeled samples for model training. Given the low frequency of event occurrences in actual systems, a scarcity of training data can introduce model bias and compromise detection accuracy. Wei et al. [8] used Extreme Learning Machine (ELM) to achieve online detection of multiple perturbations. However, the single hidden layer structure of ELM has limited expressive power when facing complex nonlinear perturbations in the scenario of high penetration rate of new energy. Moreover, its randomly initialized input weights and biases lack stability and adaptability in the scenario of multimodal data fusion.

Regarding the application of graph neural networks in time series prediction, Zheng et al. [9] investigated the application of adversarial graph neural networks to multivariate time-series anomaly detection, employing graph attention mechanisms to capture complex inter-series dependencies. Khemani et al. [10] provided a comprehensive review of the concepts, architectures, techniques, challenges, and applications of graph neural networks, systematically outlining the evolution and applicability of variants such as GAT. Zheng et al.[11] further designed a prediction model that integrates graph attention mechanisms. By dynamically learning the correlation between sequences and adaptively allocating weights, the prediction accuracy was improved, especially suitable for multivariate electricity prediction scenarios.

In summary, existing research has three major shortcomings: First, the multimodal data fusion methods are simplistic and fail to fully utilize complementary information between modes. Second, graph neural network models have not optimized the correlation between multimodal features and power grid status. Existing models such as GAT only focus on spatial dimension attention and have not incorporated temporal evolution features into a unified modeling framework. Third, the human-machine collaboration mechanism lacks deep integration and is difficult to adapt to the risk assessment needs in complex scenarios. Therefore, this paper proposes a comprehensive approach integrating these three technologies to address the shortcomings of existing research.

## 3. Relevant theoretical basis

### 3.1. Core concepts of power system situation awareness and risk assessment

Situational awareness and risk assessment of power systems are core technologies for ensuring the safe operation of power grids. They form a progressive closed loop of "perception-assessment-decision," exhibiting new technical characteristics in scenarios with a high proportion of new energy grid connection [12]. Situational awareness follows a three-layer architecture of "data acquisition-state understanding-trend prediction." It needs to integrate measurement, text, image, and topological multimodal data. Simultaneously, it needs to accurately capture the physical state of the power grid, analyse semantic information, identify equipment visual characteristics, and adapt to dynamic topological changes. Risk assessment, based on situational results, constructs a three-level indicator system of "equipment-system-environment." Risk levels are quantified through "indicator normalization-weight allocation-comprehensive scoring," with the core being the integration of objective weights from the AHP-entropy weight method and subjective judgments from expert experience. The two achieve synergy through "data interoperability and model linkage," thereby solving the problem of disconnect between perception and assessment.

The randomness and volatility of high-proportion new energy sources present three core challenges, forming the core starting point for the design of this method [13]. First, there is the challenge of dynamic adaptation. Fluctuations in new energy output led to frequent changes in topology and power flow, requiring improvements in the model's real-time modelling capability for dynamic topology. Second, there is the challenge of uncertainty quantification. Short-term prediction errors in new energy sources result in a wide-tailed risk distribution, which traditional deterministic methods struggle to accurately characterize. Third, there is the challenge of multimodal fusion. Scenarios such as equipment failures often manifest as cross-modal correlations of measurement anomalies, image defects, and text alarms, requiring the correlation and complementary interpretation of multi-source information. These characteristics and challenges underscore the technical necessity of multimodal fusion, dynamic topology modelling, and human-machine collaboration.

### 3.2. Core theory of graph neural networks

Graph neural networks use graph $G = (V, E)$ as the data structure, where $V$ is the set of nodes corresponding to power grid devices, and $E$ is the set of edges corresponding to device connection relationships. Node embedding updates are achieved by aggregating the features of neighbouring nodes [14]. The core difference between different GNN variants lies in the feature aggregation method.

Graph Convolutional Networks (GCNs) use fixed weights to aggregate neighbor features, as shown in Eq. (1). They are computationally efficient but cannot adapt to changes in topology.

$$h_v^{(l+1)} = \sigma\left(W^{(l)} \sum_{u\in N(v)} \frac{1}{\sqrt{|N(v)||N(u)|}} h_u^{(l)} + b^{(l)}\right) \quad (1)$$

Where $h_v^{(l+1)}$ is the feature vector of node $v$ in layer $l+1$, $N(v)$ is the neighbour set of node $v$, $W^{(l)}$ and $b^{(l)}$ are the weight matrix and bias term, respectively, $\sigma$ is the activation function, and LeakyReLU is chosen in this paper. $1/\sqrt{|N(v)||N(u)|}$ is the normalization coefficient, used to balance the feature contributions of nodes with different degrees. The core assumption of GCN is that the topology of the graph is fixed and unchanging. This poses a significant limitation in scenarios where the topology of power systems changes frequently. When lines are put into service or deactivated or the network is reconfigured, GCN needs to recalculate the normalization coefficients or even retrain the model, making it difficult to meet real-time requirements.

Graph Attention Network (GAT) achieves weighted aggregation by introducing attention coefficients $\alpha_{vu}$, as shown in Eq. (2). However, focusing solely on spatial attention fails to capture the evolutionary patterns in the temporal dimension.

$$\alpha_{vu} = \frac{\exp\left(\mathrm{LeakyReLU}\left(a^T\left[Wh_v \| Wh_u\right]\right)\right)}{\sum_{u\in N(v)} \exp\left(\mathrm{LeakyReLU}\left(a^T\left[Wh_v \| Wh_u\right]\right)\right)} \quad (2)$$

Where $\alpha_{vu}$ is the spatial attention coefficient of node v to its neighbour node $u$; $\|$ is the feature concatenation operation; $W$ is the feature transformation weight matrix; $a^T$ is the attention weight vector; LeakyReLU is used to alleviate the gradient vanishing problem; the numerator $\exp()$ amplifies the attention difference, and the denominator $\exp()$ is a normalization term to ensure that the sum of the coefficients is 1 . The limitation of GAT is that its attention calculation only targets different spatial neighbors at the same time, without considering the characteristic correlation of the same node at different historical times. Therefore, it is difficult to cope with the temporal evolution caused by the fluctuation of new energy sources.

The Attention-GNN model used in this paper addresses the temporal volatility of power systems by introducing a time attention coefficient on the basis of GAT, which can construct a "space-time" dual-dimensional attention mechanism. Spatial attention quantifies the contribution of different neighboring nodes to the target node, while temporal attention quantifies the contribution of different historical moments to the current situation. These two approaches are calculated independently and then weighted and fused to achieve deep integration of topological association and temporal evolution features [15], which effectively improves the modelling accuracy in dynamic scenarios.

### 3.3. Multimodal sensing data fusion theory

Multimodal data fusion is divided into three layers: data layer, feature layer, and decision layer [16]. The technical characteristics of each layer are compared in Table 1. Data layer fusion needs to solve the problem of spatiotemporal alignment of heterogeneous data. The information utilization rate exceeds 90%, but the robustness is poor. Decision layer fusion has high computational efficiency, but serious information loss. Feature layer fusion extracts cross-modal common features through deep learning models, taking into account 70%-90% information utilization rate and moderate robustness, thereby adapting to the heterogeneous characteristics of multimodal data in power systems.

This paper employs a feature-layer fusion strategy. Its core advantage lies in its adaptability to the spatiotemporal resolution differences of multimodal data, avoiding the spatiotemporal alignment challenges of data-layer fusion. Simultaneously, it retains over 70% of the original information and allows for the mining of cross-modal associations between measured text images using deep learning. To address the modality imbalance problem, namely the degradation of fusion performance when the data quality of a certain modality is reduced or missing, this paper introduces a modality attention gating unit (see Section 4). By dynamically adjusting the weights of each modality, the remaining modality weights are automatically increased to more than 80% when a single modality is missing, which significantly enhances the fusion robustness and ensures that the total weight of the weak modality is not less than 0.4 , thus avoiding the excessive dominance of the dominant modality in the fusion results.

## 4. Construction of multimodal sensing data system for power systems

### 4.1. Multimodal data preprocessing methods

To address noise and missing values in the measurement data, an "adaptive filtering + interpolation completion" method is employed. Kalman filtering is used to remove measurement noise, and the filtering equation is as follows:

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1} \tag{3}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \left(z_k - H\hat{x}_{k|k-1}\right) \tag{4}$$

Where $\hat{x}_{k|k}$ represents the filtered state value at time $k, K_k$ represents the Kalman gain, $A$ represents the state transition matrix, $B$ represents the control input matrix, $u_{k-1}$ represents the system control input at time $k-1, z_k$ represents the original measurement observation at time $k$, and $H$ represents the measurement matrix.

For missing data, an interpolation method based on spatiotemporal correlation is used to fill in the missing data by using contemporaneous data of adjacent nodes and historical similar data. To verify the robustness of the preprocessing method, we tested the completion accuracy at different missing rates ( 10%, 20%, and 30% ), with interpolation relative errors of 2.3%, 4.1%, and 6.7%, respectively, indicating that the completion effect is acceptable when the missing rate does not exceed 30%. For extreme missing rate scenarios (>30%), the model automatically reduces the measurement modality weights through a modal attention gating unit, instead relying on text and image modal information to ensure the effectiveness of the overall evaluation.

To address the heterogeneity of text data, a "structured extraction + semantic encoding" approach is adopted. Keyword is directly extracted from structured text, and semantic encoding is performed using the BERT model for unstructured text. The text is converted into feature vectors to achieve semantic information quantification [17].

For equipment inspection images, a four-level processing flow of "image enhancement + target detection + feature extraction + state quantization" is adopted. For image enhancement, adaptive histogram equalization is used to improve the contrast of images with insufficient lighting,

**Table 1.** Hierarchy of Multimodal Data

| Fusion Layers | Core Methods | Information Utilization | Robustness | Computational Efficiency | Applicable Scenarios |
|---|---|---|---|---|---|
| Data Layer | Spatiotemporal alignment + direct concatenation | 90%+ | Poor | Low | Same-resolution data |
| Feature Layer | Deep learning + feature mapping | 70–90% | Medium | Medium | Heterogeneous multimodal data |
| Decision Layer | Weighted voting + probabilistic fusion | 50–70% | High | High | Real-time decision-making scenarios |

and Gaussian filtering is used to remove noise and improve image clarity [18]. In the context of object detection, a YOLOv8 model specialized for power equipment is employed to detect equipment regions, thereby enhancing both recall and precision. Regarding feature extraction, the ResNet50 model is utilized to extract image features, outputting 2048-dimensional feature vectors, while attention pooling is applied to highlight features associated with fault regions. During the state quantification process, an equipment state scoring system is established; state scores are calculated based on the cosine similarity between the extracted feature vectors and standard state vectors, effectively reducing quantification errors.

### 4.2. Spatiotemporal alignment and feature fusion of multimodal data

After preprocessing, multimodal data needs to undergo spatiotemporal alignment to eliminate dimensional differences between heterogeneous data, followed by feature fusion to mine cross-modal correlation information. To intuitively illustrate the entire process from preprocessing to fusion, a flowchart of multimodal data processing and fusion is presented in Fig. 1, which clearly shows the parallel processing paths and fusion logic of each modality of data.

As shown in Fig. 1, measurement, text, and image data, after feature extraction using the preprocessing methods described above, are entered into the spatiotemporal alignment stage along with the topological data. Furthermore, feature fusion generates features of a unified dimension, providing standardized input for subsequent graph neural network modelling. The entire process takes $\leqslant$ 20 ms, meeting real-time requirements.

Time alignment is based on the PMU data sampling time, and linear interpolation is used to unify other modal data to a 50Hz time scale. Spatial alignment is based on the grid topology node number, associating measurement data, text data, and image data with the corresponding topology nodes to achieve a one-to-one mapping between data and nodes.

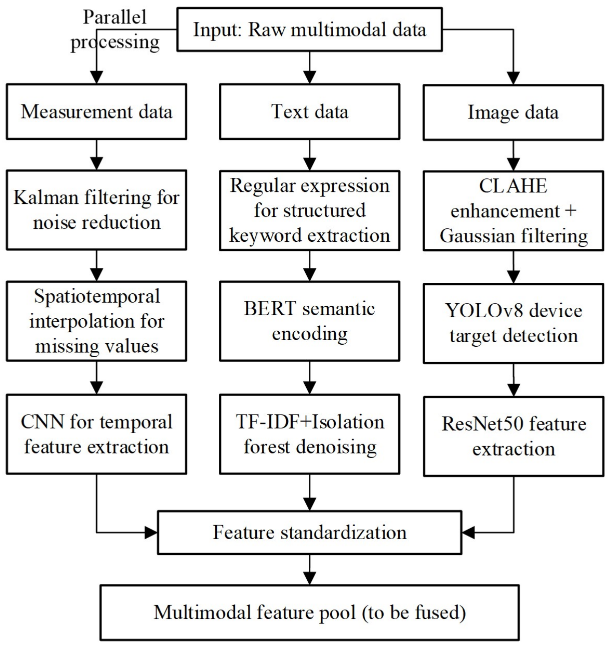


**Fig. 1.** Flowchart of multimodal data processing and fusion

In this process, measurement data focuses on "denoising, completion, and temporal feature extraction" to address real-time and completeness issues. Text data is processed through "structured extraction, semantic quantization, and denoising" to transform unstructured information. Image data is processed through "enhancement, detection, and feature extraction" to extract visual state information from the device. After preprocessing, these three types of data form standardized features, laying the foundation for subsequent fusion.

Feature fusion employs a strategy of "modality-specific features + cross-modality common features". First, specific features are extracted from the data of each modality. Then, a multimodal feature fusion layer is constructed, which concatenates the features of each modality and maps them to a unified dimension through a fully connected layer. A modal attention gating unit is embedded in the fusion layer. Its core mechanism is as follows: Let each modal feature be $h_m(m = 1, 2, 3)$. Dynamic weights $\alpha_m$ are calculated through a gating network, then normalized by Softmax and summed using weighted averages. At each time step, this gating unit dynamically adjusts the weights based on the input quality of each modality. When continuous data loss occurs, the weights of the measurement modalities are automatically reduced, while the weights of the text and image modalities are correspondingly increased, ensuring the stability of the fused features in single-modal degradation scenarios.

## 5. Design of a situational awareness and human-machine collaborative risk assessment model based on attention-gnn

### 5.1. Attention-GNN situational awareness model

To achieve deep integration of power grid topology correlation and temporal evolution characteristics, an Attention-GNN model with a hybrid architecture of "graph attention + temporal attention + bidirectional LSTM" is designed. The structural parameters of each layer are shown in Table 2. The model ingests dynamic grid graphs and historical time-series features via an input layer; following multi-layer feature aggregation, it outputs situational assessment results at both node and system levels. With a single-step inference time of 28.4 ms, it meets the real-time requirements for video inspections conducted by drones or fixed cameras.

The hierarchical structure of the Attention-GNN model is shown in Fig. 2.

As shown in Fig. 2, the model input layer incorporates the dynamic power grid graph Gt and historical $T = 10$-step multimodal features, achieving joint input of static topology and dynamic temporal features.

The graph attention layer employs a two-head attention mechanism to aggregate neighbour features, and its first-order spatial aggregation formula is:

$$h_v^{(1)} = \sigma\left(\sum_{u \in N(v)} \alpha_{vu} W_1 h_u\right) \tag{5}$$

Where $\alpha_{vu}$ is the spatial attention coefficient, $W_1$ is the aggregation weight matrix, and $\sigma$ is the LeakyReLU activation function. This layer focuses only on the contributions of neighbors at different spatial locations at the same time.

The temporal attention layer introduces a temporal attention coefficient $\beta_{vk}$ to reflect the importance of features at different historical step sizes. The modified formula is as follows:

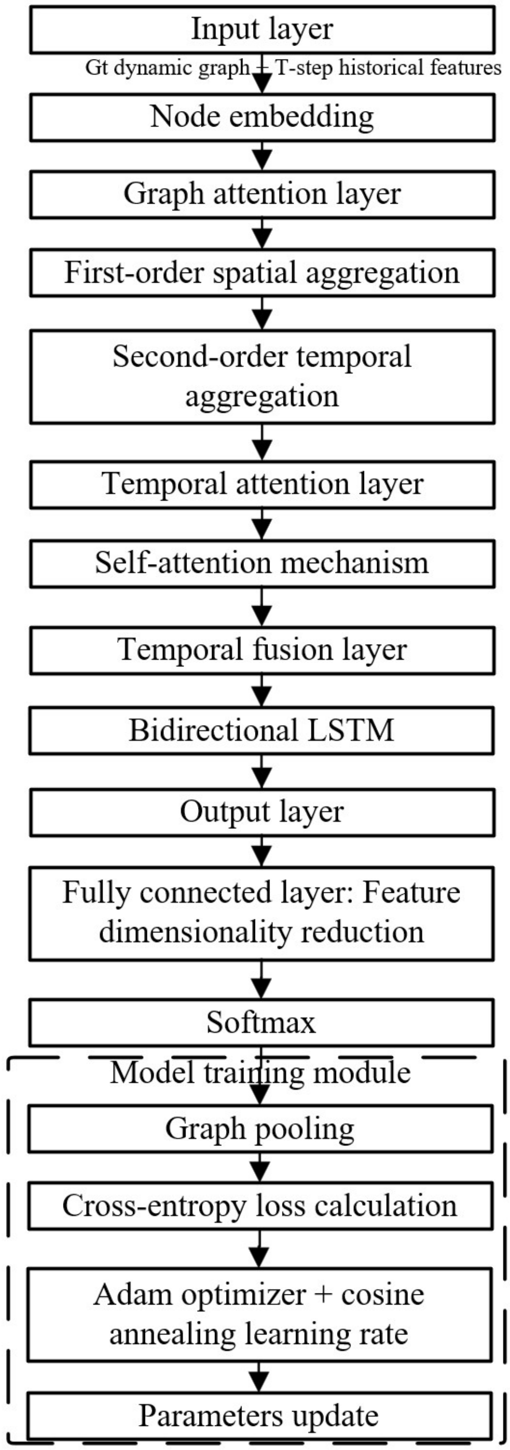


**Fig. 2.** Hierarchical structure of the Attention-GNN model

$$h_v^{(2)} = \sigma\left(\sum_{k=0}^{T-1} \beta_{vk} W_2 h_v^{(1)}(t-k)\right) \tag{6}$$

WHere, $\sum_{k=0}^{T-1} \beta_{vk} = 1$ represents the weighted fusion of features with different historical step sizes. This layer is

**Table 2.** Structural Parameters of Each Layer in the Attention-GNN Model

| Network Layer | Core Structure | Input Dimension | Output Dimension | Number of Parameters | Activation Function |
|---|---|---|---|---|---|
| Input Layer | Node embedding | 512 | 512 | 0 | – |
| Graph Attention Layer | Two-head attention | 512 | 512 | $512 \times 512 \times 2$ | LeakyReLU |
| Temporal Attention Layer | Self-attention | $512 \times 10$ | 512 | $512 \times 3$ | Sigmoid |
| Temporal Fusion Layer | Bidirectional LSTM | 512 | 512 | $(512 + 512) \times 2048$ | Tanh |
| Output Layer | Fully connected + Softmax | 512 | 3 (state classification) | $512 \times 3$ | Softmax |

decoupled from the graph attention layer. The graph attention layer outputs the spatial aggregation features, while the temporal attention layer simultaneously calculates the attention weight for each node along the time dimension. The two components are fused by element-wise multiplication followed by addition. Specifically, the spatial aggregation result $H_s \in \mathbb{R}^{N \times d}$ and the temporal aggregation result $H_t \in \mathbb{R}^{N \times d}$ are fused using a gated method:

$$H_{\text{final}} = \gamma \odot H_s + (1 - \gamma) \odot H_t, \tag{1}$$

where $\gamma$ is a learnable fusion gating parameter. This decoupling design allows the spatial and temporal contributions to be quantified separately. In the ablation experiments, either branch can be removed independently to determine the performance contribution of each branch.

The temporal fusion layer captures long-term temporal dependencies through bidirectional LSTM, enhancing the modelling of situational evolution patterns [19].

The output layer outputs node states via a fully connected layer and a Softmax function, and obtains system-level situational results after graph pooling.

Model training employs the cross-entropy loss function to quantify state classification errors and the Adam optimizer for iterative parameter updates. The learning rate is dynamically adjusted using a step decay strategy: the initial learning rate is set to 0.001 and multiplied by a decay factor of $\gamma = 0.1$ every 10 training epochs, it can ensure rapid model convergence while preventing overfitting.

### 5.2. Design of a human-machine collaboration risk assessment framework

Based on the situational awareness results, a three-level risk indicator system of "equipment-system-environment" was constructed. Among them, equipment-level indicators include equipment overload rate, voltage deviation, and temperature deviation. System-level indicators include frequency deviation, power angle stability, and power flow limit exceedance rate. Environmental-level indicators include wind speed, light intensity, and extreme weather level, etc [10].

The initial weights of each indicator were determined using the Analytic Hierarchy Process (AHP) and then corrected using the information entropy method to obtain the objective weight $w_m$.

A risk assessment sub-model is constructed by first normalizing each indicator, then calculating the risk value $R_m$ through weighted summation, and finally classifying the risk level into 5 levels (1-5), with higher levels indicating greater risk. The risk value calculation formula is as follows:

$$R = \sum_{m=1}^{M} W_m \cdot x_{m'} \tag{7}$$

Where $x_{m'}$ is the normalized indicator value, and $M$ is the total number of indicators. A closed-loop mechanism is designed: "Machine Initial Assessment-Expert Interaction Correction-Model Feedback Optimization". First, the machine outputs an initial risk level and the weights of each indicator, along with a heatmap of the contribution of each indicator. Second, experts adjust the weights based on the scenario characteristics, providing the subjective weights $(W_{m'})$. The entropy weight method is used to fuse the objective and subjective weights; the fusion formula is:

$$W_{m''} = \frac{W_m \cdot W_{m'}}{\sum_{m=1}^{M} W_m \cdot W_{m'}} \tag{8}$$

Finally, the risk level is recalculated based on the fusion weights, and the expert adjustment results are fed back to the model as labels to optimize the model parameters through transfer learning [20].

To clearly present the entire logic of "data-modelling-evaluation-optimization", the overall flowchart of the model is shown in Fig. 3. The closed-loop operation of the process ensures real-time performance and accuracy.

As shown in Fig. 3, the multimodal data, after preprocessing and fusion, drives dynamic power grid modelling

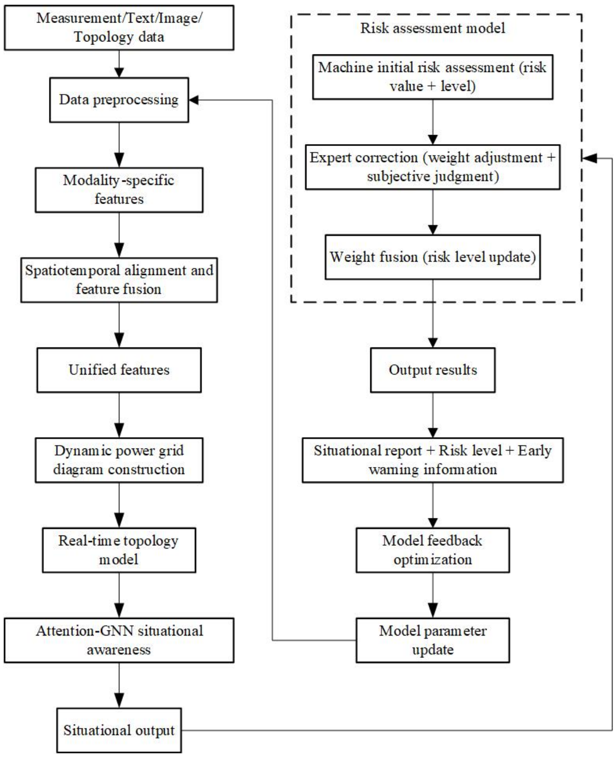


**Fig. 3.** Overall Model Flow

and Attention-GNN situational awareness. The human-machine collaboration mechanism corrects machine evaluation biases, and the feedback optimization module enables continuous model iteration. This forms a dual-drive model of "data-driven + knowledge-guided," adapting to complex dynamic power grid operation scenarios.

## 6. Experimental verification and result analysis

### 6.1. Experimental data and environment setup

The experiment combines actual operational data from a provincial power grid with PSASP simulation data. The actual data includes PMU data (100 nodes) from January to March 2024, SCADA data, dispatch logs, and 1000 equipment inspection images. The simulation data is based on a PSASP-constructed IEEE 300-node system with a 30% renewable energy penetration rate, generating data for 20 typical scenarios, including load fluctuations, renewable energy output fluctuations, and equipment failures. The dataset is divided into training, validation, and test sets in a 7:2:1 ratio.

Hardware environment: CPU: Intel Xeon Gold 6330, GPU: NVIDIA A100, RAM: 128GB. Software environment: Python 3.9, PyTorch 2.0, TensorFlow 2.10, OpenCV 4.6. To ensure fairness in model comparison, all baseline models and the Attention-GNN model presented in this paper adopted the same hyperparameter search space and tuning strategy. Specifically, all models underwent hyperparameter tuning through grid search, with the search space uniformly set as follows: number of layers $\{2, 3, 4\}$, number of hidden units $\{128, 256, 512\}$, learning rate $\{0.0005, 0.001, 0.002\}$, number of training epochs $\{100, 200, 300\}$, and optimizer {Adam, SGD}. All models used the same data partitioning (random seed 42), the same loss function, and the same evaluation metric; each hyperparameter combination was run independently three times, and the parameters corresponding to the optimal result on the validation set were used as the final settings. The final hyperparameters of each model are shown in Table 3.

Situational awareness evaluation metrics mainly include: Accuracy, Precision, Recall, and F1 score. Risk assessment evaluation metrics mainly include: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Risk Level Identification Accuracy. Computational efficiency metric: Single-step inference time.

### 6.2. Experimental results and analysis of situational awareness model

To fully verify the model performance, three types of experiments were designed [20]:

1. Benchmark comparison experiment: comparing the Attention-GNN in this paper with traditional models;
2. Ablation experiment: verifying the role of core modules such as multimodal fusion, temporal attention, and graph attention;
3. Scene adaptability experiment: testing the model performance under different new energy penetration rates and topology change frequencies.

All experiments used 5-fold cross-validation, and the results are averaged.

#### *6.2.1. Benchmark comparison experiment*

To verify the effectiveness of the model in this study, the Attention-GNN model in this paper is compared with LSTM, GCN, GAT, and GraphSAGE. The results are shown in Table 4.

As shown in Table 4, the GCN, GAT, and Attention-GNN models based on graph model theory outperform the LSTM sequence model overall, demonstrating the importance of power grid topology information. The Attention-GNN model used in this study achieves the highest accuracy, improving upon LSTM by 10.9 percentage points. The model in this paper improves accuracy by 4.3 percentage points compared to GAT, mainly because it introduces a temporal attention mechanism, thus addressing the

**Table 3.** Final Hyperparameter Settings for Each Comparison Model

| Model | Layer | Hidden Unit Number | Learning Rate | Epochs | Optimizer | Other Key Parameters |
|---|---|---|---|---|---|---|
| LSTM | 2-layer stack | 256 | 0.001 | 150 | Adam | Dropout = 0.2 |
| GCN | 3-layer | 256 | 0.001 | 200 | Adam | Dropout = 0.2 |
| GAT | 3-layer | 256 | 0.001 | 200 | Adam | Number of attention heads = 4, dropout = 0.2 |
| GraphSAGE | 3-layer | 256 | 0.001 | 200 | Adam | Number of sampled neighbors = 10, dropout = 0.2 |
| Attention-GNN | 3-layer | 512 | 0.001 | 200 | Adam | Number of attention heads = 2, $\tilde{T}$ = 10, dropout = 0.2 |

**Table 4.** Comparison Results with the Traditional Model

| Models | Accuracy (%) | Precision (%) | Recall (%) | F1 Score (%) | Single-Step Inference Time (ms) |
|---|---|---|---|---|---|
| LSTM | $82.3 \pm 1.2$ | 79.5 | 78.2 | 81.2 | 18.5 |
| GCN | $85.7 \pm 0.9$ | 83.1 | 82.5 | 84.5 | 22.3 |
| GAT | $88.9 \pm 0.7$ | 86.8 | 85.3 | 87.8 | 25.6 |
| GraphSAGE | $88.1 \pm 0.8$ | 85.2 | 84.7 | 86.3 | 26.9 |
| Attention-GNN | $93.2 \pm 0.5$ | 91.5 | 90.8 | 92.6 | 28.4 |

problem of insufficient utilization of temporal features. Although GraphSAGE can dynamically sample neighbour's, it does not incorporate multimodal data, and its accuracy is still 5.1 percentage points lower than the model in this paper. Regarding single-step inference time, although Attention-GNN is 2.8 ms slower than GAT, the 28.4 ms latency still meets the requirements for near real-time processing of video and image data (approximately 35 FPS), which can make it suitable for applications such as drone inspections and fixed-camera surveillance. It should be noted that the period of a 50 Hz electrical signal in power systems is 20 ms; the proposed method is designed for decision support in situational awareness and risk assessment, rather than for millisecond-level real-time control scenarios such as relay protection.

#### *6.2.2. Ablation test*

To verify the effectiveness of the three core modules-multimodal fusion, temporal attention, and graph attention-this study designed an ablation experiment. The "graph-free attention" variant is defined as follows: it retains the temporal attention mechanism, but the graph attention layer is replaced by a mean aggregation operation, that is, a simple average weighting of all neighbor features is performed, instead of dynamic weighting through attention coefficients. The results are shown in Table 5.

As shown in Table 5, removing the multimodal fusion module reduced the model accuracy by 7.8 percentage points, with anomaly recall decreasing by 10.2 percentage points, demonstrating that single measurement data cannot comprehensively capture device status. Removing the temporal attention module reduced accuracy by 4.5 percentage points, with an 8.3% decrease in accuracy under topology change scenarios, demonstrating the importance of temporal features for dynamic situational awareness. Replacing graph attention with mean aggregation resulted in a 3.2 percentage point decrease in accuracy, and a 12% decrease in accuracy for identifying associated faults such as device overload in topology change scenarios. This demonstrates the necessity of dynamic weighted aggregation for extracting associated features. Completely removing neighbor aggregation (only self-features) further reduced accuracy to 86.8%, proving that topological association information can provide effective gains even through simple mean aggregation. Regarding the decoupling contribution of spatial and temporal attention, in topology change scenarios, the performance decrease caused by removing temporal attention (8.3%) is greater than the decrease caused by graph attention averaging (2.7%), indicating that capturing temporal evolution patterns is more critical than refined spatial weighting when facing frequent topological changes. The synergistic effect of the three modules optimizes model performance.

**Table 5.** Results of the ablation experiment

| Models | Accuracy(%) | Recall(%) | Accuracy in topology change scenarios(%) |
|---|---|---|---|
| Attention-GNN | 93.2 | 90.8 | 90.9 |
| No multimodal fusion | 85.4 | 80.6 | 83.5 |
| No temporal attention | 88.7 | 86.3 | 82.6 |
| No graph attention (mean aggregation) | 89.0 | 87.1 | 88.2 |
| No neighbour aggregation (feature-only) | 86.8 | 84.0 | 84.7 |

#### *6.2.3. Scene adaptability experiment*

To verify the scenario adaptability of the model in this study, scenarios with different new energy penetration rates (10%, 20%, 30%, and 40%) are designed based on the IEEE 300-node system with a baseline new energy penetration rate of 30% to verify the accuracy of the model. The results are shown in Fig. 4.

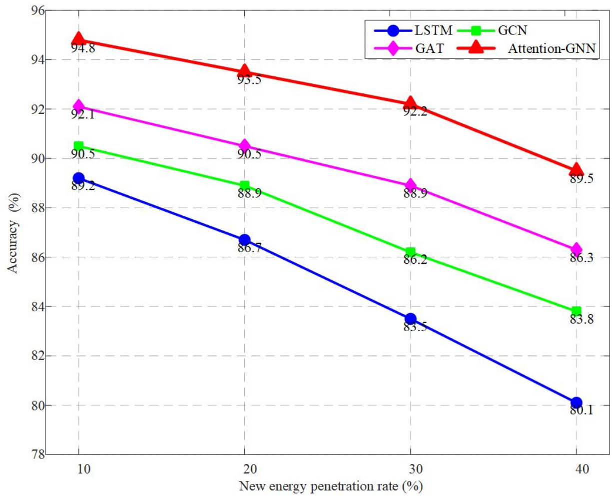


**Fig. 4.** Comparison of model accuracy under different new energy penetration rates

As shown in Fig. 4, the accuracy of all models decreased as the penetration rate of new energy sources increased. However, the Attention-GNN model in this paper showed the smallest decrease, maintaining an accuracy of 89.5% even at a penetration rate of 40%, which is 4.8 percentage points higher than GAT, demonstrating the adaptability of multimodal fusion to volatility.

To verify the model's adaptability to topological dynamics, topological change scenarios were designed using simulated switch actions. Scenarios with different topological change frequencies (1 time/hour, 5 times/hour, 10 times/hour, and 15 times/hour) are designed to verify the model's adaptability, and the results are shown in Fig. 5.

As shown in Fig. 5, when the topology change frequency is increased to 15 times/hour, the accuracy of Attention-GNN still reaches 90.1%, which is 2.3 percentage points higher than GAT's 87.8%, and the decrease is only 1.5 percentage points, while GAT's decrease is 3 percentage points, demonstrating the synergistic advantage of dynamic topology update and temporal attention mechanism.

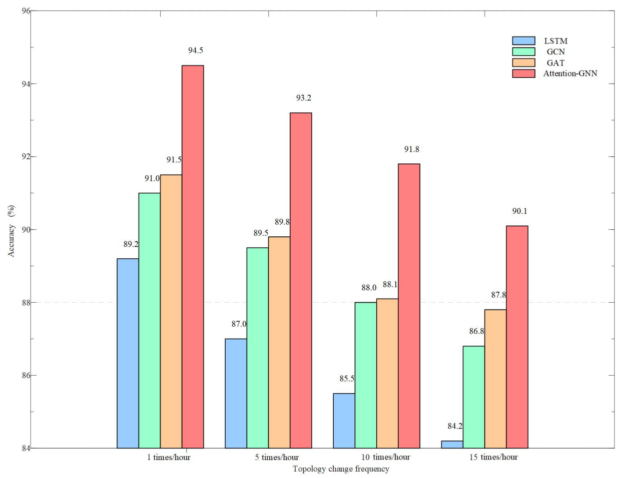


**Fig. 5.** Comparison of model accuracy under different topology change frequencies

### 6.3. Experimental results and analysis of human-machine collaborative risk assessment

#### *6.3.1. Comparison of machine risk assessment models*

To verify the risk assessment effectiveness of this model, the risk assessment sub-model in this paper is compared with the traditional AHP-entropy weight method and BP neural network model. The results are shown in Table 6.

Table 6 shows that the MAE of the machine learning model presented in this paper is reduced by 28% compared to the BP neural network, while the risk level identification accuracy is improved by 4.9 percentage points. Analysis reveals that the high-precision situational features output by the Attention-GNN provide reliable input for risk assessment, significantly improving upon the manual features used in the BP neural network. Furthermore, the risk indicator weights, after being corrected using the AHP-entropy weighting method, effectively enhance the correlation between objective weights and actual risks, better aligning with the key points of power grid safety operation.

**Table 6.** Comparison of risk assessment effects of different models

| Models | MAE | RMSE | Risk level identification accuracy (%) |
|---|---|---|---|
| AHP-entropy weight method | 0.32 | 0.45 | 80.1 |
| BP neural network | 0.25 | 0.36 | 85.3 |
| Our model | 0.18 | 0.27 | 90.2 |

#### 6.3.2. *Verification of human-machine collaboration effect*

To verify the effectiveness of human-machine collaboration in this model, three sets of comparative experiments were designed: "machine-independent evaluation", "traditional human-machine collaboration", and "human-machine collaboration in this paper". Five power grid dispatching experts were invited to participate in collaborative correction. The experimental scenarios included 10 typical risks such as equipment overload, voltage exceeding limits, frequency fluctuations, and N-1 faults. The results are shown in Table 7.

In this paper, "traditional human-machine collaboration" refers to an expert correction mode based on a fixed rule base. The system presents correction suggestions to experts according to predefined rules, such as "increasing the weight by 0.1 if the voltage exceeds the limit by more than 5%." Experts can only fine-tune the weights within the preset framework; the system does not provide visual analysis of the contribution of each indicator, nor does it support dynamic weight adjustment. When the actual operating scenario exceeds the coverage of the rule base, the expert's correction operations are limited by the rule framework, which may introduce subjective bias and lead to a decrease in evaluation accuracy. This negative result precisely reveals the limitations of the traditional collaboration mode and highlights the improvement value of the proposed "visual interactive interface + dynamic weight adjustment mechanism." In contrast, the proposed method visually displays the contribution of each indicator through heatmaps, dynamically updates weight allocation, and supports free adjustment by experts, reducing the correction time from 180 seconds to 15 seconds and improving accuracy by 8.1 percentage points.

As shown in Table 7, the human-machine collaborative assessment method presented in this paper significantly outperforms the other two methods in all indicators. The risk level identification accuracy reaches 95.6%, an improvement of 5.4 percentage points compared to machine-only assessment and 8.1 percentage points compared to traditional human-machine collaboration. In extreme scenarios, the accuracy of the proposed method reaches 92.3%, 14.6 percentage points higher than traditional human-machine collaboration. This is because the dynamic weight adjustment mechanism can quickly adapt to unknown scenarios, while traditional rule bases cannot cover extreme cases. The average expert correction time is 15 seconds, a 91.7% improvement compared to the 3 minutes of traditional human-machine collaboration. This is because the interactive interface visually displays the contribution of each indicator, assisting experts in making rapid decisions.

### 6.4. Verification of engineering practicality

A pilot application of the project was conducted at a provincial power grid dispatch centre for a period of six months (April-September 2024). The pilot covered 20 220kV and above substations, 40 power lines, and 15 renewable energy power plants, with a total of 1.2TB of multimodal data accessed. During the pilot period, the model operated stably 24 in 7-day, with an average CPU utilization of $\leqslant$ 30% and memory usage of $\leqslant$ 8 GB, meeting the operational requirements of the dispatch center.

The specific application effects are as follows:

1. Risk identification effect: A total of 32 potential risks, including equipment overload, voltage exceeding limits, frequency fluctuations, and equipment failures, are identified, 15 more than the traditional system (10 of which were potential risks). The risk identification coverage rate increases from 53% in the traditional system to 97%.

2. Early warning timeliness: The average early warning time reached 3.2 minutes, an improvement of 113% compared to the traditional system ( 1.5 minutes). The early warning time for risks related to new energy is the longest (average 4.1 minutes).

3. Decision support effect: Based on the risk assessment report output by the model, the time for dispatchers to formulate prevention and control measures is reduced from an average of 10 minutes to 3 minutes, and the accuracy rate of measure implementation increases from 82% to 95%.

4. Economic benefits: During the pilot period, based on model-based early warnings, dispatchers took timely preventative measures, which resulted in the avoidance of approximately five potential power outages

**Table 7.** Human–computer collaboration effects in different scenarios

| Evaluation method | MAE | RMSE | Risk level identification accuracy (%) | Accuracy in extreme scenarios (%) | Expert correction time (s) |
|---|---|---|---|---|---|
| Machine-independent evaluation | $0.18 \pm 0.03$ | $0.27 \pm 0.04$ | 90.2 | 78.5 | — |
| Traditional human–machine collaboration | $0.22 \pm 0.05$ | $0.31 \pm 0.06$ | 87.5 | 77.7 | $180 \pm 20$ |
| Human–machine collaboration in this paper | $0.12 \pm 0.02$ | $0.18 \pm 0.03$ | 95.6 | 92.3 | $15 \pm 5$ |

and a reduction in direct economic losses of approximately RMB 8 million (calculated based on the historical average loss per power outage and the effectiveness coefficient of dispatch measures). The program also reduced the workload of dispatchers by 30%.

Fig. 6 shows typical operational data during the pilot period. It can be seen that the model maintains a risk identification accuracy of over 94% during peak load periods of 10:00-12:00 and 18:00-20:00, and during periods of fluctuating renewable energy output of 9:00-11:00 and 14:00-16:00, so as to demonstrate good adaptability.

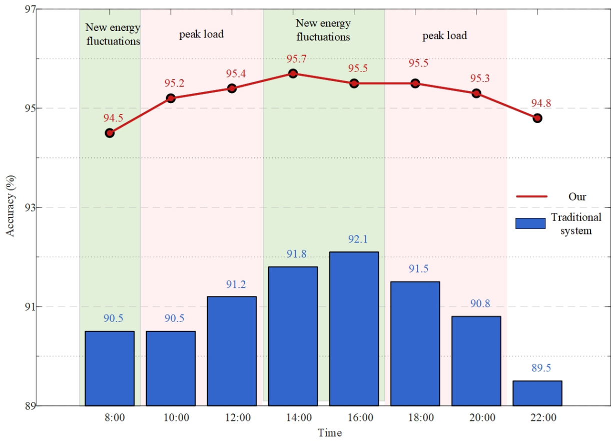


**Fig. 6.** Changes in risk identification accuracy on typical days during the pilot period

## 7. Conclusion

Combining theoretical modelling and engineering verification, this paper proposes a power system situational awareness and risk assessment method based on graph neural networks (GNNs) and human-machine collaboration. By integrating theoretical modelling, experimental verification, and engineering pilot projects, a "four-dimensional integrated" multimodal perception data system is constructed. An Attention-GNN situational awareness model is proposed, and a dynamic human-machine collaborative risk assessment framework is established. Validation results demonstrate that the model achieves a situational awareness accuracy of 93.2% and a single-step inference time of 28.4 ms (approximately 35 FPS, which can meet near-real-time processing requirements for applications such as drone inspections and video surveillance). It maintains an accuracy of 89.5% even with a 40% new energy penetration rate, balancing precision with real-time performance. The developed interactive visualization interface reduces the average expert correction time to 15 seconds-an efficiency gain of 91.7%-while achieving collaborative assessment and extreme-scenario accuracies of 95.6% and 92.3%, respectively; these figures represent improvements of 5.4 and 13.8 percentage points over standalone machine assessments. Furthermore, the model's engineering pilot project demonstrates significant effectiveness. In a six-month pilot project in a provincial power grid, the risk identification coverage rate increased from 53% to 97%, the average early warning was 3.2 minutes in advance, five power outages were avoided, direct losses were reduced by 8 million yuan, and dispatch decision-making efficiency was improved by 70%, thus verifying the engineering application value of the method.

However, this study still has shortcomings. Multimodal data annotation is costly, image and text data annotation requires the participation of professionals, the computational efficiency of the model in ultra-large-scale power grids needs to be improved, and the completion accuracy in extreme missing rate (>30%) scenarios needs further optimization. These issues need to be addressed in future work. These issues need to be addressed in future work.

## Acknowledgements

The work is supported by Research and Application of Key Technologies for Edge AI Devices Based on Brain like Chips (Grant no. SZKJXM20240027).

## References


**[1]** M. I. Abdelwanis and M. I. Elmezain, (2024) *"A comprehensive review of hybrid AC/DC networks: insights into system planning, energy management, control, and protection"* **Neural Computing and Applications** ***36***(29): 17961–17977. DOI: https://doi.org/10.1007/s00521-024-10264-5.

**[2]** D. Huang, P. Wu, X. Wang, and D. Wang. "Analysis of security and stability characteristics in new power system". In: *2024 IEEE 8th Conference on Energy Internet and Energy System Integration (EI2)*. IEEE. 2024, 3639–3646. DOI: https://doi.org/10.1109/EI264398.2024.10991342.

**[3]** D. Madurasinghe and G. K. Venayagamoorthy, (2024) *"Distributed dynamic security assessment for modern power system operational situational awareness"* **IEEE Access** ***12***: 147991–148010. DOI: https://doi.org/10.1109/ACCESS.2024.3475011.

**[4]** M. Ren, N. Chen, and H. Qiu, (2023) *"Human-machine collaborative decision-making: An evolutionary roadmap based on cognitive intelligence"* **International Journal of Social Robotics** ***15***(7): 1101–1114. DOI: https://doi.org/10.1007/s12369-023-01020-1.

**[5]** K. Chen, J. Zhu, L. Han, S. Li, and P. Gao. "A Novel Network Security Situation Awareness Model for Advanced Persistent Threat". In: *2022 7th IEEE International Conference on Data Science in Cyberspace (DSC)*. IEEE. 2022, 9–16. DOI: https://doi.org/10.1109/DSC55868.2022.00009.

**[6]** A. Sahu and K. Davis, (2023) *"Inferring adversarial behaviour in cyber-physical power systems using a Bayesian attack graph approach"* **IET Cyber-Physical Systems: Theory & Applications** *8*(2): 91–108. DOI: https://doi.org/10.1049/cps2.12047.

**[7]** S. Das and B. K. Panigrahi, (2022) *"A PMU-based data-driven approach for enhancing situational awareness in building a resilient power systems"* **IEEE Transactions on Industrial Informatics** *18*(7): 4773–4784. DOI: https://doi.org/10.1109/TII.2022.3147794.

**[8]** S. Wei, D. Wenjuan, and C. Xia, (2024) *"Power quality disturbances categorization using identity feature vector and extreme learning machine"* **Intelligent Systems with Applications** *24*: 200446. DOI: https://doi.org/10.1016/j.iswa.2024.200446.

**[9]** Z. Ye, Y. J. Kumar, G. O. Sing, F. Song, and J. Wang, (2022) *"A comprehensive survey of graph neural networks for knowledge graphs"* **Ieee Access** ***10***: 75729–75741. DOI: https://doi.org/10.1109/ACCESS.2022.3191784.

**[10]** B. Khemani, S. Patil, K. Kotecha, and S. Tanwar, (2024) *"A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions"* **Journal of Big Data** ***11***(1): 18. DOI: https://doi.org/10.1186/s40537-023-00876-4.

**[11]** B. Zheng, L. Ming, K. Zeng, M. Zhou, X. Zhang, T. Ye, B. Yang, X. Zhou, and C. S. Jensen, (2024) *"Adversarial graph neural network for multivariate time series anomaly detection"* **IEEE Transactions on Knowledge and Data Engineering** ***36***(12): 7612–7626. DOI: https://doi.org/10.1109/TKDE.2024.3419891.

**[12]** U. Pandey, A. Pathak, A. Kumar, and S. Mondal, (2023) *"Applications of artificial intelligence in power system operation, control and planning: a review"* **Clean Energy** *7*(6): 1199–1218. DOI: https://doi.org/10.1093/ce/zkad061.

**[13]** S. Liu, Y. Teng, and Z. Chen, (2025) *"Optimization framework for transient overvoltage for new energy clusters in power grids supported by integrated energy"* **Journal of Renewable and Sustainable Energy** *17*(5): DOI: https://doi.org/10.1063/5.0248715.

**[14]** R. Bhattacharya, N. K. Nagwani, and S. Tripathi, (2023) *"A community detection model using node embedding approach and graph convolutional network with clustering technique"* **Decision Analytics Journal** *9*: 100362. DOI: https://doi.org/10.1016/j.dajour.2023.100362.

**[15]** E. Mohammadi, M. Alizadeh, M. Asgarimoghaddam, X. Wang, and M. G. Simões, (2022) *"A review on application of artificial intelligence techniques in microgrids"*: DOI: https://doi.org/10.1109/JESTIE.2022.3198504.

**[16]** N. Arvinth, (2025) *"Fault Detection in Smart Grids Using Deep Learning-Based Phasor Measurement Unit Data Analysis"* **Journal of Reconfigurable Hardware Architectures and Embedded Systems** 2(2): 1–7. DOI: https://doi.org/10.17051/JRHAES/02.02.01.

**[17]** B. Azimian, R. S. Biswas, S. Moshtagh, A. Pal, L. Tong, and G. Dasarathy, (2022) *"State and topology estimation for unobservable distribution systems using deep neural networks"* **IEEE transactions on instrumentation and measurement** *71*: 1–14. DOI: https://doi.org/10.1109/TIM.2022.3167722.

**[18]** Z. Cheng, H. Chen, T. Mao, B. Zhou, W. Zhao, T. Wang, and Y. Xu, (2025) *"Data-driven topology identification and line parameter estimation for distribution network: A two-stage approach"* **Journal of Renewable and Sustainable Energy** ***17***(3): DOI: https://doi.org/10.1063/5.0267179.

**[19]** Y. Zhang, L. Guo, Y. Liu, Z. Wang, X. Li, and C. Wang, (2026) *"Physical informed data-driven topology estimation of radial distribution networks based on partial measurement"* **CSEE Journal of Power and Energy Systems**: DOI: https://doi.org/10.17775/CSEEJPES.2024.09220.

**[20]** S. Liu, Y. Teng, and Z. Chen, (2025) *"Optimization framework for transient overvoltage for new energy clusters in power grids supported by integrated energy"* **Journal of Renewable and Sustainable Energy** ***17***(5): DOI: https://doi.org/10.1063/5.0248715.