ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 3 - 2026

Design of an Intelligent Operation and Maintenance Decision Support System for Power Grids Integrating Deep Learning and Graph Neural Networks

Xinzhu Dong[1] *and Yun Zhang*[2,*]

1 Brunel London School, North China University of Technology, 100144, Beijing, China
2 School of Mechanical and Materials Engineering, North China University of Technology, Beijing, 100144, Beijing, China

SUMMARY: *Traditional power grid O&M systems suffer from two limitations related to their architecture. First, fault diagnostic models consider equipment units individually, ignoring dependencies between devices that are implicit in grid topology. Second, there is no clear link between diagnostic outcomes and maintenance recommendations, making the latter process depend heavily on ad-hoc decision making. The current work presents a design approach that seeks to address both limitations with two particular architectural solutions. The first solution introduces a representation of the state of the equipment, based on a combination of temporal features extracted using deep learning and neighborhood context computed using a graph attention network. This allows representing dependencies between faults without explicitly designing a new fault diagnosis model. The second innovation is a bridging component that translates diagnosis results into constraint-based maintenance recommendations in the form of a ranked list of recommendations for O&M personnel. The design was evaluated on a real-world dataset of 12,847 annotated examples across 23 classes of faults at six substations. The achieved accuracy of fault diagnosis is 94.3 ± 0.4% ($p<0.01$) and NDCG@5 of 0.887 ± 0.012, which constitute 2.9 and 10.4-percentage-point improvements over the baselines.*

KEYWORDS: *Power grid O&M; Graph neural network; Fault diagnosis; Decision support; Topology-aware representation*

1 Introduction

Structural complexity has increased significantly in the past decade owing to several reasons: renewable energy integration, bidirectional flow at prosumers, and demand-side unpredictability [1, 2]. This has resulted in new types of failure behavior, which do not fit into conventional O&M approaches based on periodical maintenance scheduling and experience-driven rule sets applicable to a homogeneous and less complex network topology [3, 4]. Substations, transmission lines, transformers, and switchgears are highly integrated infrastructures where a small fault in one unit may lead to a cascading failure through topologically dependent pathways [4]. Unplanned faults in power grids entail huge costs for society and industries; therefore, utilities have been under growing pressure to implement predictive and intelligence-based operation and maintenance (O&M) methods to improve system reliability [5]. Furthermore, recent review papers show that smart O&M increasingly

*zhangyun20260101@163.com
https://doi.org/10.65102/is20261045
Published: 10 June, 2026

relies on explainable and data-intensive solutions [5]. In addition, the widespread adoption of supervisory control and data acquisition (SCADA), smart meters, and advanced condition-monitoring devices has produced a wealth of multi-sourced and multi-variate time series from grid infrastructures. Consequently, while the problem of lack of data has largely been solved for most grids, the primary concern now lies in the development of intelligent systems that could interpret multi-sourced data to detect faults, localize them, and make proper O&M decisions [2-5].

The current state-of-the-art O&M intelligence techniques have shown considerable success in isolated fault detection and classification problems. Recent studies suggest that graph-based methods could provide superior performance in detecting and localizing faults [6-8]. Encoder models that leverage convolutional, recurrent, and Transformer architectures remain successful in capturing temporal behavior of equipment at the individual level; yet, the independent modeling of each equipment piece would not be able to reflect electrical couplings through which disturbances occur in adjacent devices [6, 7]. Fault-induced electrical disturbances travel through conductive pathways in real-world power grids, thus making graph-enhanced models that incorporate topology structure, dynamic neighbors, and topology changes increasingly popular [7, 8]. The second issue with current O&M methods is that there is often no connection between diagnostics results and actions. Recent works focusing on maintenance scheduling in power systems have demonstrated that realistic scheduling algorithms should incorporate not only fault models but also security constraints, maintenance opportunities, available budget, etc., instead of performing purely manual action on diagnostic results [9, 10]. Fault models will result in risk scores or labels, but maintenance actions require prioritization of work orders based on critical equipment, maintenance budget, time window, and other constraints.

This work remedies the limitations through careful architectural design. Contributions of the present paper are: (i) an architecture that utilizes topology-awareness in the representation of equipment state by means of incorporating neighborhood aggregation based on the graph attention network as a feature enhancement layer in the representation pipeline, with each equipment's state representation being conditioned on its neighboring devices' current state according to the grid connectivity graph; and (ii) a "diagnosis to decision" translation module, which formulates the prioritization process as a constrained ranking task, leveraging continuous fault risk predictions together with equipment importance, cascading effects, and schedule feasibility considerations to generate actionable, feasible maintenance schedules. Both architectures leverage existing concepts of deep learning and graph neural networks, while the unique aspects are in their design and connections between them.

2 Related Work

2.1 Deep Learning-Based Fault Diagnosis in Power Systems

The use of deep learning techniques for fault detection in power equipment has evolved from engineered feature extraction into end-to-end representation learning using directly processed raw or minimally processed data [11-14]. Convolutional neural networks have been successfully used to capture local temporal features from electric current measurements and vibrations in transformers and circuit breakers and achieved highly accurate classification [11]. Temporal relationships were modeled in multi-step sequence data by applying long short-term memory or gated recurrent unit architectures and modeling sequences from data points separated by any interval of time [12, 13]. Transformer architecture using self-attention was recently applied for power system anomaly detection and fault diagnosis with higher

performance for long-horizon temporal patterns caused by progressive degenerative processes [14]. Combination of local feature extraction by convolution operations and attention mechanisms enabled through Transformer architectures has proven successful for multi-channel data fusion and sequential fault representation. However, despite these advancements in fault diagnosis research, there is still an evident architectural deficiency found in many modern models of fault diagnostics: each individual piece of equipment is considered as an independent sensing unit during both the training and testing phase, which does not account for the interdependencies between the devices in terms of their topology [15]. The signals from each transformer are often analyzed without considering the electric conditions of the connected switchgears, the other busbar sections, and protective devices upstream. Such an approach, while useful in laboratories for replicating faults, is generally ineffective when dealing with actual equipment where the effects of the faults spread throughout the grid and create recognizable signatures in sensors connected to electrically adjacent devices before causing degeneration to the source device.

2.2 Graph Neural Networks for Power Grid Modeling

The application of GNNs in power systems has been primarily in areas such as state estimation, load forecasting, approximate methods for optimal power flow, and voltage regulation [16-19]. There is a structural similarity between GNNs and power grids because both are based on node and edge structures where inter-nodal interactions have meaningfully interpretable physical significance [16, 17]. Convolutional graph networks have been used for power flow approximation by virtue of an adjacency structure where the connectivity between edges reflects electrical impedance relationships, and short-term load forecasting by using spatial dependencies among the demands at various nodes [18]. In Spatiotemporal GNNs, the spatial aggregation of graphs is augmented with temporal encoding using recurrent, convolutional, or attention mechanisms to reflect the temporal dimension, which shows improvement on multistep prediction tasks compared to either temporal or spatial baselines. Attention networks use learnable content-dependent weights in aggregating nodes to allow adaptation to heterogeneous network structures, which corresponds to changes in physical significance of neighboring node states based on the operating condition and the task at hand [19, 20]. Although considerable success has been achieved in other applications in power systems, the implementation of GNNs in fault diagnosis is still relatively underexplored, and their application has been primarily structural without any consideration of how fault diagnosis can be used in the context of maintenance decisions in power systems. Existing works on GNN fault localization are more focused on developing techniques for localized representation or classification of faults, rather than a continuously updating streaming decision pipeline driven by multi-channel sensory information. Additionally, the output of GNN-based fault localization is not integrated into subsequent maintenance scheduling [20].

2.3 O&M Decision Support Systems

Various decision-support methods have been considered within industrial O&M in energy production, manufacturing, and transportation using different technical means [21-24]. Optimization techniques have approached maintenance scheduling as either a programming or multi-objective optimization problem and introduced maintenance cost, impact of outage, reliability concerns, resources' availability, and constraints into the decision process [22-24]. From the point of view of power system studies, recent works suggest that maintenance planning should be carried out considering uncertainty, resource constraints, possible outage window, and reliability concerns; hence, heuristic schedules cannot always be used for

practical maintenance plans [23-25]. The use of probabilistic or explainable frameworks for maintenance scheduling has also been suggested to improve maintainability decision-making via predictive models and interpretable reasonings or utility-based decision-making frameworks [21-23]. Scheduling based on machine learning (ML) has further considered predictive maintenance scheduling and maintenance optimization through reinforcement learning techniques, which allow maintenance activities to be scheduled or optimized based on degradation prediction, operational data, and changing state information [25]. However, all these frameworks suffer from one major drawback: although fault detection and maintenance scheduling are considered in sequence, they are weakly coupled such that information from the former is used in the latter merely as input rather than being closely related to decision variables in O&M [21, 22, 24]. As a result, uncertainty-related information, severity level, and cross-resource dependency from diagnostics have been poorly used in scheduling, preventing efficient use of intelligent systems in practice.

2.4 Summary and Research Gap

Table 1 highlights exemplary related work from each of the three research threads, linking them to their focus area, methodology used, and associated drawback in the context of power grid intelligence O&M. It can be seen from the literature survey that there is no prior art approach available, which considers both topology-agnostic failure modeling and the diagnostic to decision-making conversion problem.

Table 1: Summary of Representative Related Works

Study Area	Core Method	Primary Application	Gap in Grid O&M Context
DL fault diagnosis [11-15]	CNN-LSTM hybrid	Transformer fault detection	No inter-device topology modeling
	Transformer attention	Anomaly detection	Isolated per-device processing
GNN grid modeling [16-20]	Graph convolutional network	State estimation	Not connected to O&M decisions
	Spatiotemporal GNN	Load forecasting	Fault diagnosis not addressed
	Graph attention network	Static fault localization	Snapshot input, no streaming
O&M scheduling [21-25]	Reinforcement learning	Maintenance policy	Decoupled from diagnostic output
	Multi-objective optimization	Resource allocation	Binary fault input, no risk grading
Decision support	Expert system rules	Response prioritization	Limited to predefined fault modes

3 System Design

3.1 Overall System Architecture

The novel system fuses multiple monitoring streams from grid assets, which include phase-wise voltage and current, temperature and partial discharge measures, harmonic distortions, and operational information about each piece of equipment. It produces two

separate outputs of distinct nature: (i) per-device diagnostic output, including fault category probabilities along with confidence on localization, and (ii) a ranked list of maintenance recommendations ready for execution by O&M personnel, without any further manual postprocessing step required. The proposed intelligent system is organized into four successive layers of functionality, namely, temporal feature encoder, topology-aware graph fusion layer (Innovation I), fault diagnosis head, and diagnosis to decision bridging module (Innovation II). The complete architecture of the proposed intelligent power grid operation and maintenance decision support system is illustrated by Figure 1, showing four major modules of the system, namely temporal feature encoder, topology-aware graph fusion layer, fault diagnosis head, and diagnosis to decision bridging module.

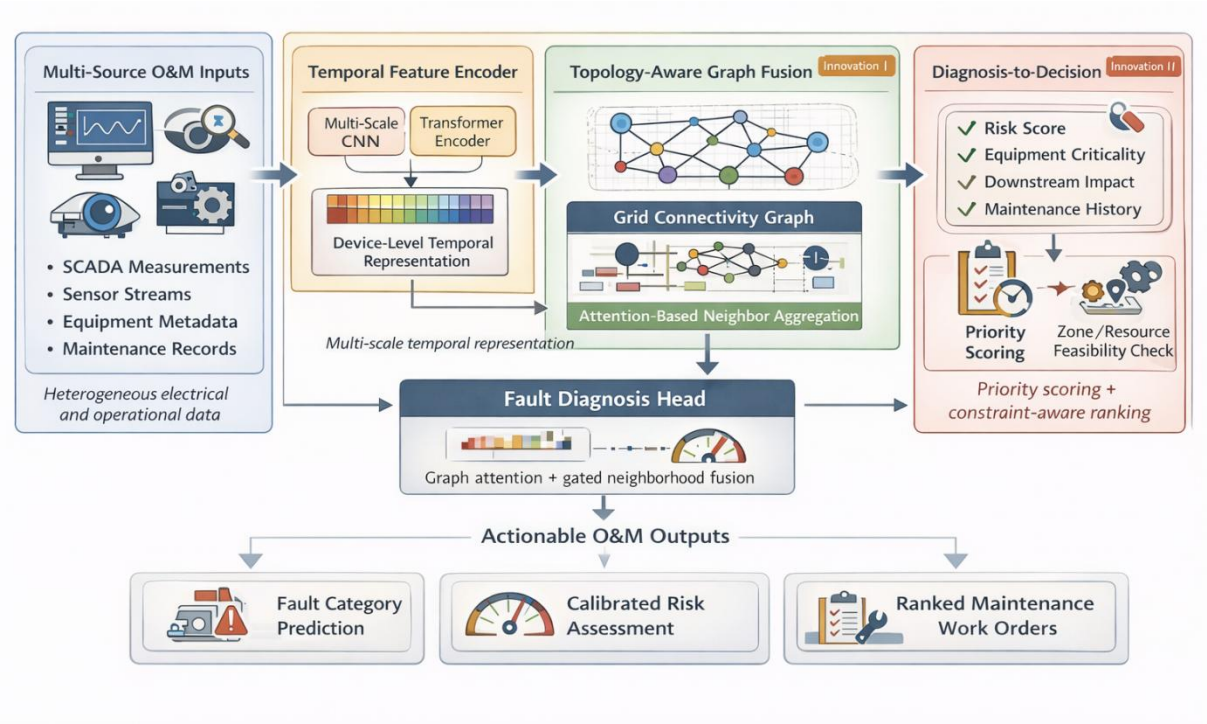

Figure 1: System Architecture Diagram

The temporal feature encoder extracts data stream representations on an individual-device basis through a combination of CNN and Transformer layers. In particular, the CNN layer performs processing on a sliding window of raw measurements collected from the sensors, where separable convolutions are applied to each of the following three temporal scales having respective window widths of 20 ms, 100 ms, and 500 ms to represent short-term current peaks caused by contact arcing, short-term waveform anomalies caused by insulation partial discharge, and intermediate-term thermal anomalies caused by winding degradation. The feature vectors at the three temporal scales are concatenated together along the channels before being fed into the transformer layer, which treats each time step in the sequence as tokens to perform multi-head self-attention. This combination stems from the characteristic of dual temporal scales at which faults occur in power equipment, where short-window sensitive local feature extraction is necessary for capturing the fault inception events, while long-term contextual information is needed for representing the trends of gradual deterioration, possibly lasting for minutes or even hours. The output of the temporal encoder is a fixed-length device-level state vector $\mathbf{h}_i \in \mathrm{R}^d$ (with d=256) for each monitored device ii i, representing the device's temporal condition within the current 120-second observation window. These per-device state vectors serve as node features for the subsequent graph fusion layer.

The topology-aware graph fusion layer constructs a graph over the monitored device set using the grid's physical connectivity and enriches each node's temporal representation with aggregated context from electrically adjacent devices, addressing the topology-blindness of isolated signal models (Innovation I, Section 3.2). The topology-enriched representation $\mathbf{z}_i$ replaces $\mathbf{h}_i$ as input to the fault diagnosis head, which applies a two-layer MLP with hidden dimension 128, ReLU activation, and softmax output over the 23 fault categories to produce per-device fault category probabilities. The scalar risk score $s_i \in [0,1]$ is derived as a weighted sum of the top-3 fault category probabilities, calibrated by post-hoc temperature scaling on the validation set to ensure that s_i reflects a well-calibrated posterior fault probability rather than an uncalibrated logit-derived score. This calibration is important because s_i is subsequently used directly in the priority scoring equation (7) of Innovation II, where its numerical value determines maintenance urgency; uncalibrated scores would distort priority orderings for devices near the decision boundary. The diagnosis-to-decision bridging module (Innovation II, Section 3.3) receives the calibrated risk scores $\{s_i\}$ alongside equipment operational metadata and produces a constraint-validated, ranked maintenance task list. The two innovations are positioned at the feature representation stage and the output translation stage, respectively, ensuring structural decoupling between their objectives and enabling independent validation through ablation.

3.2 Innovation I: Topology-Aware Equipment State Representation

The common fault diagnosis architecture for electric power systems features a critical design principle with respect to the temporal encoder, the device representation produced by the encoder captures exclusively the information contained in the local observation signal of the device in question. The path from signal to representation excludes any connection between the state of the transformer and the state of other devices like the adjacent circuit breaker and bus sections upstream of the transformer. In practical applications, this is a critical representational weakness. The presence of a turn-to-turn fault in a transformer will lead to an increase in its magnetic current and harmonics; these propagate upstream in the power grid via the supply impedance to generate voltage perturbations at the bus bar level, and hence current perturbations in the currents of other nearby feeders. This propagation pattern is inherently diagnostic, in that it provides multiple points of geometric validation for the nature of the fault and for the extent of the problem, and occurs before any alarms from thermal sensors reach their threshold in response to the faulty device.

The proposed design addresses this by introducing a graph attention network layer that operates downstream of the temporal encoder. The grid is represented as a weighted graph $G=(V,\mathrm{E},\mathbf{W})$, where V is the set of N monitored devices, E is the set of direct electrical connections between them, and $\mathbf{W} \in \mathrm{R}^{N \times N}$ is an adjacency matrix with entries $W_{\{ij\}}$ equal to the per-unit admittance of the connection between devices i and j and zero otherwise. The admittance weighting reflects the physical fact that electrically closer neighbors, those connected through lower-impedance paths, exert stronger influence on a device's measurable signal state.

Before graph aggregation, the temporal encoder output $\mathbf{h}_i$ is processed through a position-wise feed-forward sub-layer to produce a query-key-value decomposition consistent with attention scoring. The raw attention coefficient between device i and neighbor $j \in N(i)$ is computed as shown in Equation (1):

$$e_{ij} = LeakyReLU\left(\mathbf{a}^{\top}\left[\mathbf{W}_q\mathbf{h}_i \,\|\, \mathbf{W}_k\mathbf{h}_j\right] + W_{ij}^{adm}\right) \tag{1}$$

where $\mathbf{W}_q, \mathbf{W}_k \in \mathrm{R}^{d' \times d}$ are learnable query and key projection matrices, $\mathbf{a} \in \mathrm{R}^{2d'}$ is the attention scoring vector, $\|$ denotes concatenation, and W_{ij}^{adm} is a scalar bias term derived from the admittance weight, injecting physical proximity into the learned relevance score. The admission bias ensures that learned attention cannot overshadow physical proximity relationships, where devices coupled via low resistance paths cannot receive insignificant attention regardless of their temporal representations at the current moment. The attention weight normalization is obtained through the application of softmax across the entire neighborhood as demonstrated below Equation (2):

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})} \tag{2}$$

The neighborhood-aggregated contextual representation for device i is as shown in Equation (3):

$$\tilde{\mathbf{h}}_i = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}\ \mathbf{W}_v \mathbf{h}_j\right) \tag{3}$$

where $\mathbf{W}_v \in \mathrm{R}^{d' \times d}$ is a value projection matrix and $\sigma(\cdot)$ is the ELU activation. To capture multi-resolution neighborhood interactions, reflecting that fault propagation signatures manifest differently at directly connected neighbors versus two-hop neighbors through an intermediate switching node, the aggregation in equation (3) is applied with K independent attention heads and the outputs concatenated as shown in Equation (4):

$$\tilde{\mathbf{h}}_i^{\mathrm{multi}} = \Big\|_{|K|} \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(k)}\ \mathbf{W}_v^{(k)} \mathbf{h}_j\right) \tag{4}$$

The system uses K=4 heads with d'=64 per head, yielding a concatenated dimension of 256, selected by grid search over $K \in \{2,4,8\}$ and $d' \in \{32,64,128\}$ on the validation split. This is a significant difference from GCN's symmetric, normalized Laplacian pooling: GCN implicitly weights contributions by each neighbor based on its degree, considering all the neighbors functionally similar regardless of their current status. In case there is a fault in multiple devices that is initiated by one source device, this structural similarity reduces the information content coming from the faulty neighbor to noise level compared to uninformative information received from healthy neighbors of the same distance from the starting node. The learned attention coefficients α_{ij} in equation (2) can instead assign high weight to a neighbor exhibiting fault-correlated signal patterns while simultaneously assigning low weight to a structurally equivalent neighbor with nominal readings. The attention weights proposed using equation (1), which is influenced by the admission bias principle, also constrains the learned weight using the physical properties of signal transmission, thus ensuring that the model does not ignore nodes that have strong couplings.

The topology-enriched representation is formed by gated fusion of the device's own temporal encoding $\mathbf{h}_i$ and the aggregated neighborhood context $\tilde{\mathbf{h}}_i^{multi}$. Rather than a fixed-weight linear combination, the gate is learned from the concatenation of both representations, allowing the model to determine for each device and each input whether local temporal evidence or topological neighborhood context is more diagnostically informative as shown in Equation (5) and (6):

$$\boldsymbol{g}_i = \sigma_{sig}\left(\boldsymbol{W}_{gate}\left[\boldsymbol{h}_i \,\|\, \tilde{\boldsymbol{h}}_i^{multi}\right] + \boldsymbol{b}_{gate}\right) \tag{5}$$

$$\boldsymbol{z}_i = \boldsymbol{g}_i \odot \boldsymbol{h}_i + (1 - \boldsymbol{g}_i) \odot \boldsymbol{W}_{proj}\, \tilde{\boldsymbol{h}}_i^{multi} \quad (6)$$

where $\sigma_{sig}(\cdot)$ is the sigmoid function, W_{gate} and b_{gate} are learned parameters, $\odot$ denotes element-wise multiplication, and W_{proj} projects the concatenated multi-head output to match the dimension of h_i. The gate $g_i \in (0,1)^d$ operates at the feature-dimension level, enabling the model to selectively incorporate topological context in some representation dimensions while preserving local temporal evidence in others. In this regard, this architecture stands apart from simply connecting neighboring signals, which will require architectural changes in all subsequent levels and provide equal importance to all neighbors and replacing the temporal encoder with GNNs which means ignoring the learned temporal representation. This gated fusion technique can be understood as a selective improvement which takes into account the representation structure and incorporates relational information where it can help the model learn better. Figure 2 illustrates this new fusion architecture introduced in this research paper.

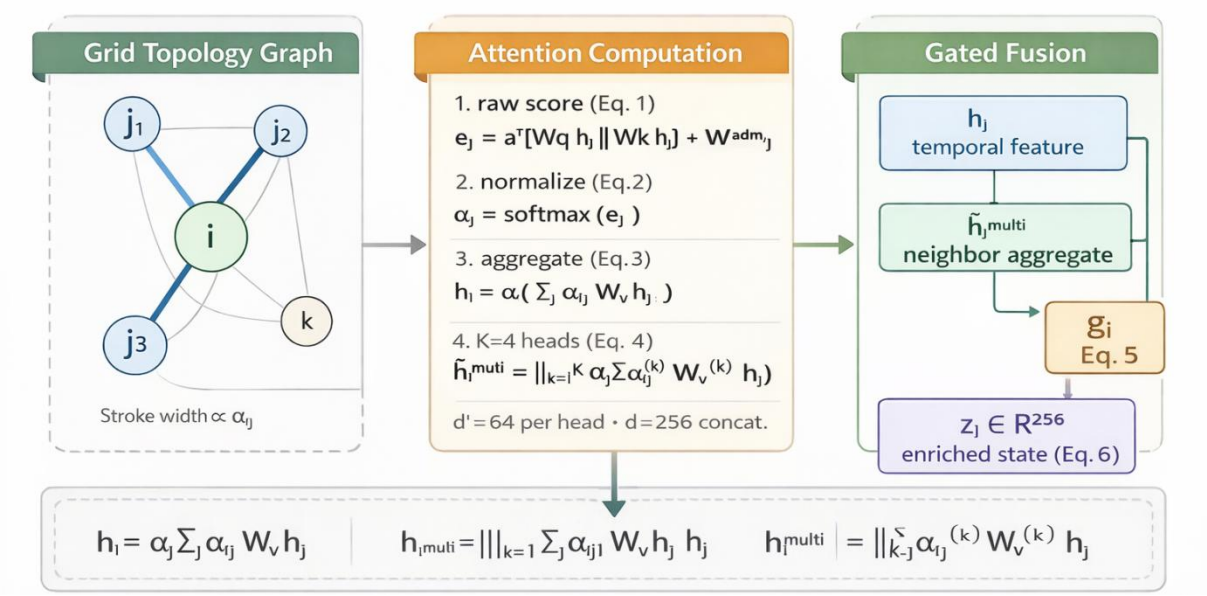

Figure 2: Innovation I: topology-aware fusion mechanism

3.3 Innovation II: Diagnosis-to-Decision Bridging Module

Diagnostic outcomes for devices in power grids from fault diagnosis systems usually result in some form of per-device risk rating or categorization of faults. However, diagnostic results cannot be directly incorporated into operational activities such as O&M. Maintenance personnel will not rely only on each device's rating; they will need to choose which devices need to be inspected, their order, and must observe limitations due to crew availability, acceptable outage levels, and required network redundancies that prohibit maintenance activities at the same time in different devices within the same protective zone. Without an explicit mechanism for this process, decisions about which devices should be inspected next and in what order remain subjective in the opinion of each engineer, leading to inconsistency across different engineers as well as poor scalability when many devices need to be inspected simultaneously in a high load environment or a network in the aftermath of severe weather-related events causing multiple faults to occur. The problem is further complicated by the practical requirement to account for the changes in network vulnerability brought about

by maintenance activities, as maintenance requires temporarily removing one or more network components from service.

The bridging module takes as input the per-device fault risk scores $\{s_i\}_{i=1}^{N}$ produced by the diagnosis head, where $s_i \in [0,1]$ represents the estimated active-fault probability at device ii i within the current observation window, alongside three operational metadata inputs per device: the equipment criticality weight $w_i^{crit} \in (0,1]$, derived from the device's betweenness centrality in the grid connectivity graph and its rated load-serving capacity; the downstream service impact coefficient $\phi_i^{impact} \in (0,1]$, quantifying the fraction of network load at risk of interruption if device ii i fails without preemptive intervention; and the elapsed time since last maintenance $t_i^{last} \geq 0$ in days, drawn from CMMS records. The priority score for device ii i is computed as shown in Equation (7):

$$p_i = s_i \cdot w_i^{crit} \cdot \phi_i^{impact} \cdot \exp\left(-\lambda \cdot t_i^{last}\right) \tag{7}$$

The exponential decay term $\exp(-\lambda t_i^{last})$ reduces the computed priority of recently maintained devices, preventing redundant immediate re-dispatch, while amplifying urgency for long-uninspected assets with elevated diagnostic risk scores. The decay parameter λ is set to 2.0×10^{-3} (corresponding to a half-life of approximately 347 days), reflecting the typical inspection cycle interval in the target network. The multiplicative structure of equation (7) ensures that even a very high diagnostic score does not result in an extremely high priority recommendation when the system is non-critical or low impact. On the other hand, a system that is highly critical and highly impactful, along with a relatively high risk score, will be prioritized over another system that is lowly critical but has a slightly higher score.

The constraint-aware ranking problem selects the maintenance task subset $\mathcal{R}^* \subseteq \mathcal{V}$ that maximizes aggregate priority within the scheduling window as shown in Equation (8):

$$\mathcal{R}^* = arg\max_{\mathcal{R}\subseteq\mathcal{V}} \sum_{i\in\mathcal{R}} p_i \quad subject\ to: \quad \sum_{i\in\mathcal{R}} c_i \leq C_{\max}, \quad |\mathcal{R} \cap \mathcal{Z}_k| \leq 1 \ \forall k \tag{8}$$

where c_i is the estimated maintenance crew-hours for device ii i derived from historical CMMS records, $C_{\max}$ is the available crew-hour budget for the scheduling window (configurable per shift or per day), and Z_k is the k protection zone, a set of devices that cannot be simultaneously taken offline due to network N-1 redundancy constraints. In the evaluated network, 11 non-overlapping protection zones are defined, with a mean of 4.3 ± 1.6 devices per zone. The zone constraint $\mathcal{R} \cap \mathcal{Z}_k| \leq 1$ ensures that at most one device per redundancy zone enters the schedule, preserving the network's fault tolerance during maintenance operations and preventing the simultaneous removal of multiple devices that collectively provide backup coverage for a single network segment. This formulation is solved as a greedy knapsack with zone exclusion, devices are processed in descending priority order p_i, and each candidate is added to $\mathcal{R}$ if and only if it satisfies both the crew-hour budget constraint and the zone exclusion constraint given the devices already in R, running in $O(N \log N)$ time, compatible with real-time scheduling requirements. Integer programming formulations of the same problem were evaluated during system development but provided no measurable solution quality improvement over the greedy approach on the evaluated network sizes ($N \leq 47$), while incurring 80-140× longer solve times, confirming the greedy approach as the appropriate design choice for operational deployment.

The module is trained using a pairwise ranking loss that aligns the learned priority ordering with ground-truth maintenance urgency labels $\{y_i\}$ annotated from historical O&M records as shown in Equation (9):

$$\mathcal{L}_{rank} = \sum_{\backslash substack(i,j):\, y_i > y_j} \max\left(0, \delta - (p_i - p_j)\right)^2 \quad (9)$$

where δ=0.1 is a margin parameter enforcing minimum score separation between correctly ordered device pairs, and the sum runs over all pairs where device i has strictly higher annotated maintenance urgency than device j. The formulation of the objective function explicitly imposes penalty for rank reversal in the predicted order of priorities, thus making sure that the schedule generated by the bridging module matches the intuition and experience of O&M engineers as captured by historical records. One of the key architectural novelties in comparison to traditional decoupled scheduling approaches is the use of the continuous risk score s_i, a combination of both severity and diagnostic uncertainty provided by the diagnosis head, that enters the priority computation unmodified by equation (7) and without the step of labeling the device condition. By doing so, the continuous gradient of diagnostic uncertainty is maintained until the scheduling step, and significantly different prioritization orders will be produced for devices sharing a similar risk score but in different environments.

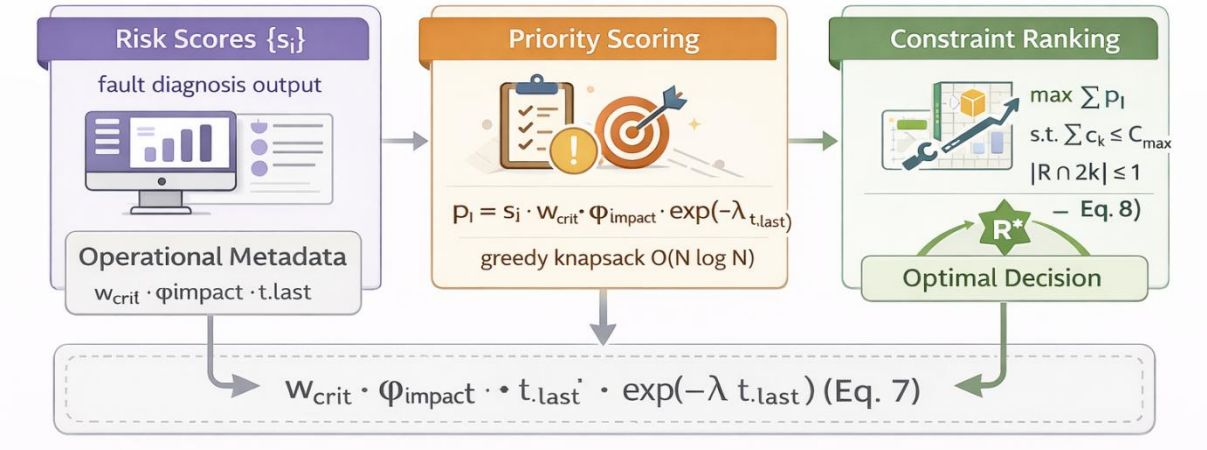

Figure 3: Innovation II: diagnosis-to-decision bridging module

3.4 Joint Training Strategy and Objective Function

The full system is trained end-to-end with a joint objective that simultaneously optimizes fault diagnosis accuracy, topology-aware representation consistency, and maintenance priority ranking quality. The fault diagnosis term is standard cross-entropy over the C=23 fault categories as shown in Equation (10):

$$\mathcal{L}_{cls} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic} \log \hat{y}_{ic} \quad (10)$$

where $y_{ic} \in \{0,1\}$ is the one-hot ground-truth label and $\hat{y}_{ic}$ is the predicted probability for device i and category c. A graph consistency regularization term is applied over the topology-enriched representations $\{\mathbf{z}_i\}$, penalizing large representation distances between electrically connected devices that share the same ground-truth fault category as shown in Equation (11):

$$\mathcal{L}_{graph} = \frac{1}{|\mathcal{E}|}\sum_{(i,j)\in\mathcal{E}} W_{ij} \cdot \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 \cdot \mathbb{1}[y_i = y_j] \quad (11)$$

The importance of topology-aware fusion for producing structurally coherent embeddings can be seen through the requirement that two connected devices sharing a common fault state

should appear as neighbors in the embedding space. The topology allows for the graph attention network to effectively capture useful fault information. The resulting loss function is shown below in Equation (12):

$$\mathcal{L} = \mathcal{L}_{cls} + \beta_1 \mathcal{L}_{rank} + \beta_2 \mathcal{L}_{graph} + \beta_3 \|\Theta\|_2^2 \tag{12}$$

where β_1=0.40, β_2=0.20, and β_3=1×10^{-4} are loss weighting coefficients determined by Bayesian hyperparameter optimization on the validation split over 50 trials, and Θ denotes all trainable parameters. The weighting of β_1=0.40 on the ranking loss reflects that recommendation quality is a primary system objective and must be strongly represented in the training signal, not treated as a secondary auxiliary loss. The β_2=0.2 weighting on the graph consistency term is set to reinforce the GAT fusion without dominating the classification loss: an excessively high β_2 risks collapsing topologically adjacent representations regardless of their actual fault state differences, which would degrade the discriminability of the enriched embeddings. The L_2 regularization term $\|\Theta\|_2^2$ prevents overfitting in the graph attention sub-module, which operates on a more limited effective sample size than the temporal encoder given the constrained number of topological neighbors per device (mean degree 3.7 ± 1.2 in the evaluation network). The system is optimized using AdamW with a cosine annealing learning rate schedule (initial rate 3×10^{-4}, minimum rate 1×10^{-6}, batch size 64, over 150 training epochs with early stopping triggered if validation Macro-F1 does not improve over 20 consecutive epochs. Training converges at epoch 127 ± 9 (mean ± std across five folds), with the final validation Macro-F1 of 92.1 ± 0.5% used for model selection.

4 Experimental Analysis

4.1 Experimental Setup and Evaluation Metrics

The experiments employ an actual dataset of power grid monitoring collected on the power grid network of a provincial 220kV grid. This network includes six substations and 47 monitoring units, which include power transformers, circuit breakers, bus section, capacitors, and transmission lines. This dataset includes a total period of 48 months of SCADA data collection at a rate of 1kHz, while each monitoring unit has 8-14 channels of sensors. After the quality control and event labeling performed by the domain engineers, there are 12,847 data samples available in 23 classes.

Table 2: Dataset Characteristics

Property	Value
Substations	6
Monitored devices	47
Fault categories	23
Total labeled samples	12,847
Observation window	120 s
Sampling frequency	1 kHz
Sensor channels per device	8-14
Training / Validation / Test split	70% / 15% / 15%
Class imbalance ratio (max : min)	18.4 : 1
Ground-truth maintenance events (test set)	347

Stratified sampling and weighted cross-entropy are applied in order to deal with the class imbalance, i.e., 18.4:1. The training subset is validated through five-fold cross validation; all presented results correspond to their averages over the five folds computed based on the held-out test set. Fault diagnosis baselines include: CNN-LSTM, consisting of a two-layer CNN encoder with a two-layer LSTM having 256-dimensional hidden state and MLP classification; GCN with fixed coefficient Laplacian normalization over the same grid graph topology as in the proposed model; Transformer only, which is a six-layer encoder with eight attention heads over individual device sequences without using any graph structure; and STGNN, which involves recurrent encoding to extract node representations and then performs aggregation via Graph Convolution layer with normalized degree as the coefficient without using attention weighting or gating mechanism. All of the above baseline methods utilize the same optimization algorithm, learning rate schedule, batch size, and early stopping criterion as the proposed model in order to conduct a fair comparison. Decision recommendation baselines include: rule-based scheduler, which selects devices that surpass a preset risk threshold of 0.6 and prioritize them in terms of their magnitude in the risk threshold exceedance; score-threshold ranking method, where all devices are ranked according to their diagnosis scores without any consideration of constraints; and decoupled system, where diagnosis scores are mapped to fault/no-fault binary labels at the risk score threshold 0.5 prior to the incorporation in the constrained ranking scheme described in Eq. (8).

Primary metrics for fault diagnosis evaluation are accuracy, Macro-F1 (unweighted average over all 23 categories, appropriate under class imbalance), AUC-ROC, and cross-device localization precision. For decision recommendation: NDCG@5 and NDCG@10, Hit Rate@5 and Hit Rate@10, constraint satisfaction rate, and RMSE between predicted and annotated maintenance priority scores.

4.2 Evaluation of Innovation I: Fault Diagnosis with Topology-Aware Fusion

Table 3 and Figure 4 reports fault diagnosis performance of all compared methods on the held-out test set. Statistical significance relative to the strongest baseline (STGNN) is assessed via paired t-test across the five folds.

Table 3: Fault Diagnosis Performance Comparison

Method	Accuracy (%)	Macro-F1 (%)	AUC-ROC	Localization Precision (%)
CNN-LSTM	87.6 ± 1.2	85.4 ± 0.9	0.891 ± 0.014	79.3 ± 1.8
Transformer-only	88.9 ± 1.0	86.7 ± 0.8	0.908 ± 0.013	81.4 ± 1.7
GCN	89.2 ± 0.8	87.1 ± 1.1	0.912 ± 0.011	82.6 ± 1.5
STGNN	91.4 ± 0.7	89.6 ± 0.9	0.934 ± 0.010	85.9 ± 1.3
Proposed	94.3 ± 0.4*	92.8 ± 0.6*	0.963 ± 0.008*	91.7 ± 0.9*

*p < 0.01 vs. STGNN (paired t-test, two-tailed).

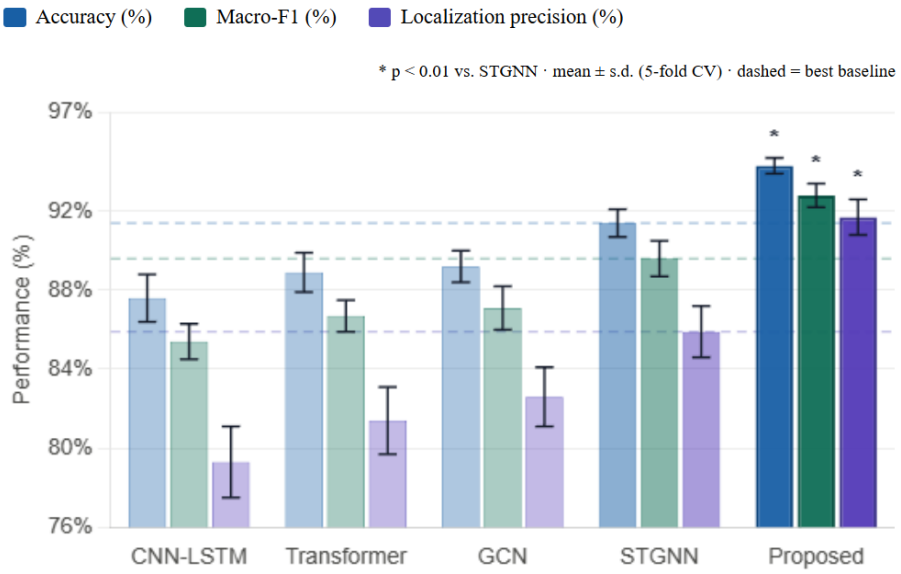

Figure 4: Fault diagnosis performance

The proposed system is demonstrated to outperform all baselines on each measured metric with statistical significance. Compared to STGNN, the strongest baseline system, the proposed system shows improved accuracy by 2.9 percentage points ($p = 0.003$), Macro-F1 by 3.2 points ($p = 0.001$), AUC-ROC by 0.029 ($p = 0.004$), and localization precision by 5.8 points ($p < 0.001$). The localization precision improvement is largely due to the gated neighbor-based fusion mechanism defined by equations (5)-(6).

Through assigning an adaptive weight to the topological neighborhood context, the system learns propagation signatures on neighboring nodes, which geometrically localize the origin of the fault, thus preventing misclassification of the fault effect propagated along the network as a new fault instance, the main cause of errors among topology-oblivious baseline systems. The fault category-level analysis shows that the superiority of the proposed system is most evident for cascading faults, namely inter-bus short circuits and relay protection chain failures, for which it produces Macro-F1 scores of 94.1 ± 0.8% and 91.3 ± 1.1% compared to 84.6 ± 1.4% and 82.9 ± 1.7% produced by STGNN. When considering fault categories without topology-related propagation characteristics such as capacitor bank dielectric breakdowns, the improvement becomes negligible, with the score difference of just 1.3 ± 0.9 percentage points. Both the CNN-LSTM and Transformer-only architectures demonstrate similar limitation despite using a different approach for temporal data encoding. This suggests that merely using more complex temporal encoding does not compensate for the lack of topology dependence. GCN improves on both signal-only systems through its coefficient-defined fusion strategy; however, this strategy is inferior to our method since the coefficients used to aggregate

neighboring nodes cannot account for different weights based on their current state, a factor critical in fault scenarios where just one or two neighbors contribute relevant information regarding propagation signature.

To further illustrate the effect of topology-aware fusion on representation quality, a 2D projection of device embeddings is visualized in Figure 5.

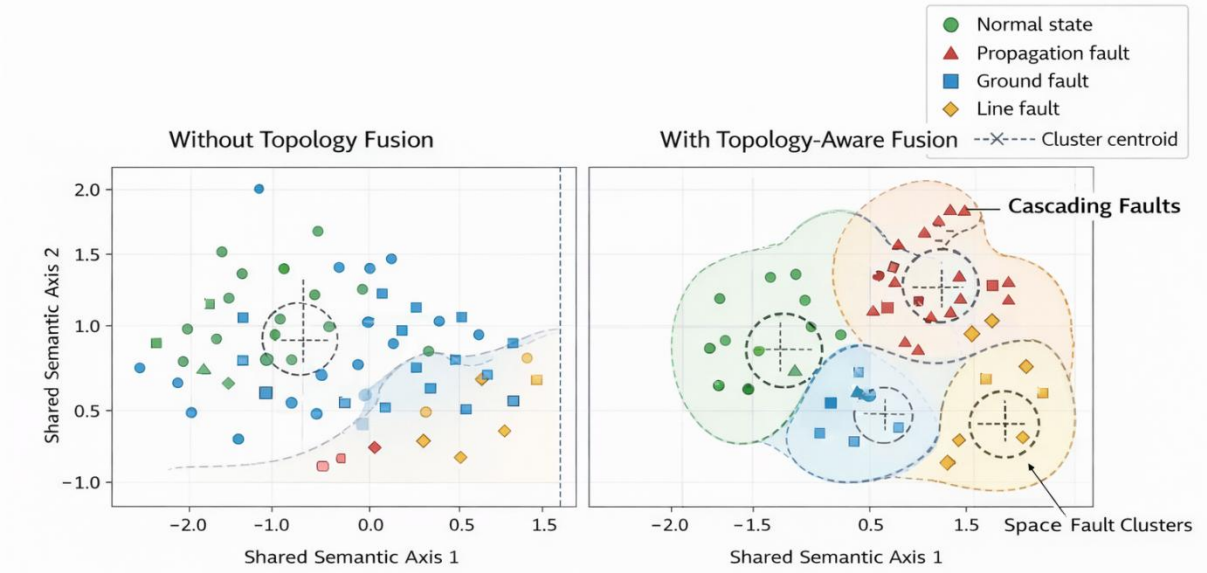

Figure 5: Feature Space Visualization of Device Representations with and without Topology-Aware Fusion

4.3 Evaluation of Innovation II: Decision Recommendation Quality

Table 4 reports decision recommendation quality across all methods. Ground-truth maintenance priority rankings are derived from 347 annotated maintenance events within the test period, labeled by three independent domain engineers with inter-rater agreement (Fleiss' $\kappa = 0.81 \pm 0.04$).

Table 4: Decision Recommendation Quality Comparison

Method	NDCG@5	NDCG@10	Hit Rate@5 (%)	Hit Rate@10 (%)	Constraint Sat. (%)	RMSE
Rule-based	0.712 ± 0.024	0.768 ± 0.019	64.3 ± 2.7	73.1 ± 2.3	91.2 ± 1.4	0.341 ± 0.038
Decoupled (binary)	0.741 ± 0.021	0.789 ± 0.017	67.2 ± 2.5	75.8 ± 2.1	93.4 ± 1.2	0.318 ± 0.034
Score-threshold	0.783 ± 0.018	0.821 ± 0.015	71.8 ± 2.2	79.4 ± 1.9	88.6 ± 1.8	0.297 ± 0.031
Proposed	0.887 ± 0.012*	0.921 ± 0.009*	83.4 ± 1.8*	91.2 ± 1.3*	96.7 ± 0.8*	0.183 ± 0.021*

*$p < 0.01$ vs. score-threshold (paired t-test, two-tailed).

Our designed bridging module achieves an NDCG@5 of 0.887 ± 0.012, corresponding to an absolute gain of 10.4 points compared with the score-threshold baseline ($p < 0.001$), and a gain of 17.5 points compared with the rule-based scheduler ($p < 0.001$). Our model exhibits an Hit Rate@5 of 83.4 ± 1.8%, meaning that 83.4% of test scheduling scenarios contain at least one of the top five suggested maintenance operations that matches the ground truth urgent operation in its ranking order, in contrast with 71.8 ± 2.2% from the score-threshold and 64.3 ± 2.7% from the rule-based scheduling methods. The constraint satisfaction ratio is 96.7 ± 0.8%, signifying that our constrained-ranking strategy in equation (8) guarantees schedule feasibility in almost all cases without any manual post-hoc schedule adjustment, whereas the score-threshold approach, lacking any zone constraint, generates a schedule that

needs to be manually adjusted in 11.4% of the test scenarios (constrained satisfaction ratio of 88.6 ± 1.8%). Lastly, the RMSE against priority annotations is 0.183 ± 0.021, which implies a relative gain of 38.4% over the score-threshold baseline (RMSE of 0.297 ± 0.031). Analysis of the priority scoring function components in equation (7) reveals that the downstream impact coefficient ϕ_i^{impact} contributes the most to rank order changes relative to pure diagnostic score ranking: removing it from the priority computation while retaining all other terms increases RMSE to 0.241 ± 0.027 (+31.7%) and reduces NDCG@5 to 0.851 ± 0.014 (−0.036), indicating that grid-level service impact, not just fault severity, is a primary determinant of correct maintenance urgency. Remarkably, despite employing binary encoding of the diagnoses and applying complete constraint modeling, the decoupled scheme fails to outperform the scoring-threshold technique in terms of NDCG and hit ratio. In other words, the critical details regarding the degree of severity lost with the conversion into binary format cannot be recovered by enhancing the accuracy of constraints. Figure 6 demonstrates differences in the ranking produced by our approach and alternative methods.

Figure 6: Comparison of Maintenance Priority Ranking across Different Methods

4.4 Ablation Study and Overall System Comparison

Figure 7 shows the results of an ablation study which evaluates the effect of each architectural advancement individually. The model without “topology fusion” uses the GAT enrichment and the gated fusion step replaced by a straightforward pass-through of the temporal encoder into the diagnosis head. In the case of the model “without decision bridge,” the priority scoring and constrained ranking method is replaced by a threshold scoring system in the diagnosis head.

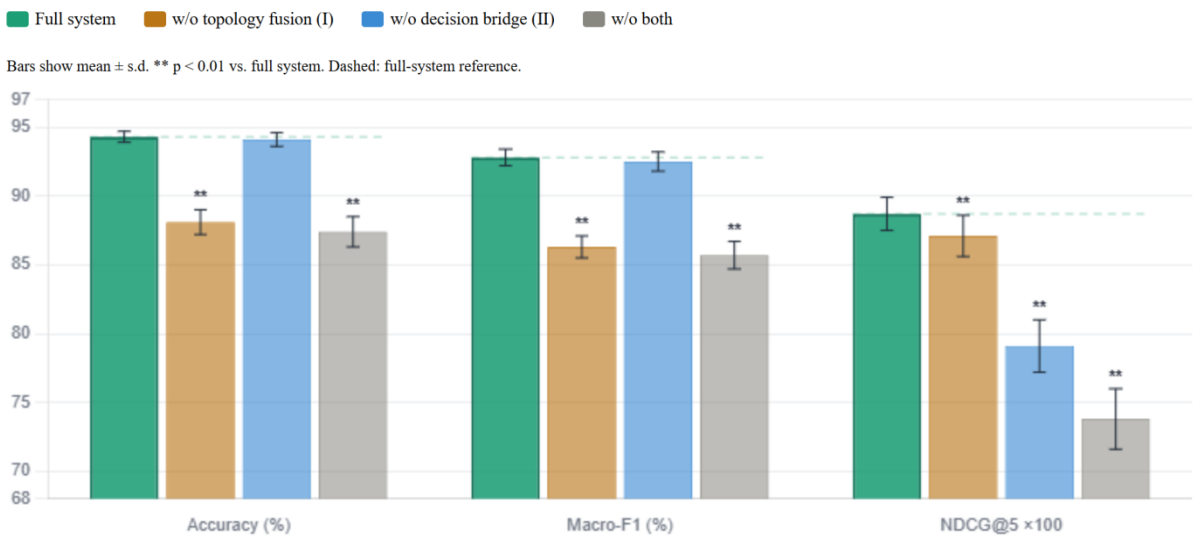

Figure 7: Ablation study

Removing Innovation I decreases accuracy by 6.2 points ($p < 0.001$) and Macro-F1 by 6.5 points ($p < 0.001$). This means that the contribution of the topology-aware representation modification is the major one for improved diagnostics, accounting for most of the performance gain. This conclusion is also supported by the decrease in NDCG@5 of 0.016, showing that enhanced representation improves diagnostic metrics but provides minimal improvements to the scheduling performance via the better calibration of risk scores. Removal of Innovation II, in turn, has no effect on diagnosis-related metrics, with accuracy and Macro-F1 changes less than 0.3 points ($p > 0.05$). However, it leads to a sharp decrease in NDCG@5 by 0.096 ($p < 0.001$) and Hit Rate@5 by 11.8 points ($p < 0.001$). These results show that the second innovation significantly improves decision quality thanks to bridging architecture. Since there are minimal cross-effects on metrics outside of their respective domains for each design choice, the two innovations can be viewed as independent and thus justify a modular architecture. Specifically, Innovation I introduces a representation-level transformation mostly contributing to diagnostics, whereas Innovation II brings translation-level transformation mostly responsible for recommendations. The "w/o both" configuration, which uses the baseline CNN-Transformer architecture along with score-threshold scheduling, obtains an accuracy of 87.4 ± 1.1%, which is inferior to the best external benchmark STGNN with 91.4 ± 0.7%.

Table 5 reports inference overhead and parameter count, confirming that the performance improvements are not achieved at prohibitive computational cost.

Table 5: System Parameters and Inference Overhead

Method	Accuracy (%)	NDCG@5	Parameters (M)	Inference Time (ms/sample)
CNN-LSTM	87.6 ± 1.2	—	4.2	8.3 ± 0.6
STGNN	91.4 ± 0.7	—	9.8	17.6 ± 1.1
Score-threshold	—	0.783 ± 0.018	—	2.1 ± 0.3
Proposed	94.3 ± 0.4	0.887 ± 0.012	11.3	21.4 ± 1.4

The architecture has a total of 11.3 million parameters, a 15.3% improvement over STGNN. Inference latency per sample is measured at 21.4 ± 1.4 ms, sufficient to meet real-time constraints of SCADA systems where non-critical path processing time windows are

generally 100 ms or more. These gains from both the GAT fusion layer and the bridging mechanism amount to about 1.5 million parameters and 3.8 ms latency, respectively, matching the design criteria for lightweight enhancements.

5 Discussion

5.1 Interpretation of Key Results

Two important conclusions can be drawn based on the experimental results. On the one hand, the topology-based feature fusion, proposed in Equations (1)-(6), achieves statistical significance improvements in the quality of fault diagnostics that are attributable exclusively to the representation of topologically structured information rather than the expanded number of parameters. The ablation experiment demonstrates that the additional accuracy of 6.2 points does not occur due to a higher number of parameters. As demonstrated by Table 11, the model with no topological fusion includes fewer parameters than the full architecture; however, the improvement occurs due to introducing topologically-related features that have a real physical meaning for the electrical interconnection between grid elements. The detailed fault category-based assessment in Section 4.2 reveals that the improvement grows according to the extent of topological contribution for each fault type, with cascading faults being the most improved category (the improvement up to 9.5 percentage points in terms of Macro-F1) and localized faults receiving a minor improvement, which agrees with the designed mechanism. In particular, a 5.8 point improvement in the localization quality compared to STGNN is especially significant in operational terms, since mislocalization of a fault causes workers to check healthy devices, increasing costs and preventing the necessary interventions at the fault origin site until the problem is correctly identified. Figure 8 depicts a sample of how the fault can spread and localize depending on other assumptions about grid structure.

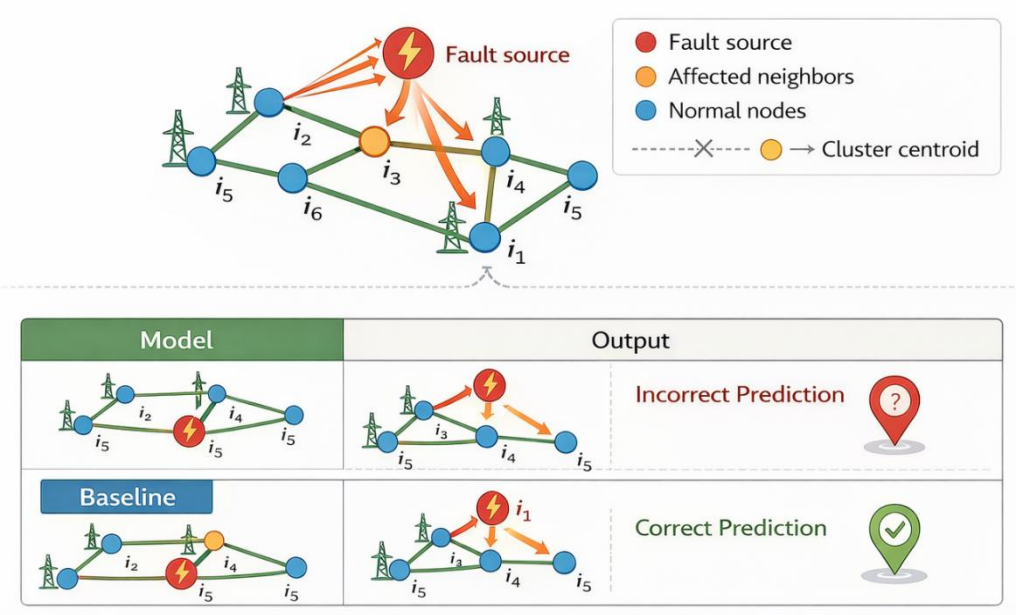

Figure 8: Illustration of Fault Propagation and Localization with Topology-Aware Representation

Secondly, there is a high level of improvement in recommendation quality in practice due to the decision-bridging module, where the increase in NDCG@5 and Hit Rate@5 is up to

10.4 points and 11.8 points respectively without increasing computational overhead by a large amount. Since the proposed approach outperformed the score-threshold baseline method by achieving a 38.4% decrease in RMSE, it means that the priority score function in Equation (7), taking into account all four factors, i.e., diagnostic risks, equipment importance, downstream service impact, and maintenance recency, more accurately represents the multi-dimensional evaluation conducted by operations and maintenance engineers. According to the analysis done in Section 4.3, removing the downstream impact coefficient in Equation (7) causes a 31.7% RMSE increase, proving the reasonability of the inclusion of downstream service impact as part of priority scoring. The lack of such an evaluation in previous diagnosis-oriented methods explains why they are unable to offer useful recommendations. The 96.7% constraint-satisfaction rate shows that the zone-exclusive ranking approach in Equation (8) is indeed a workable approach.

5.2 Practical Deployment Considerations

The proposed architecture is meant to work within the current SCADA and EMS systems without any changes in the hardware architecture. Temporal encoder works directly on the measurements streams compliant with the IEC 61850 standard at its natural sampling rate and does not require any additional resampling except that which is typically done when preparing the data for the industrial condition monitoring. Grid graph is built using the topology data from the database maintained by the EMS and admits edge weights can be computed from the line impedance parameters recorded during operations, providing the one-time off-line construction of the graph. Inference time of 21.4 ± 1.4 ms for a single measurement point allows to implement the proposed algorithm in real-time on any standard commodity CPU, since the neighborhood aggregation step using Graph Attention Network (GAT) on the 47 nodes graph with the maximum degree of 6 is negligibly small compared to the temporal encoder computation.

The bridging module outputs structured recommendations with four fields per device: device identifier, recommended maintenance category (inspection, relay test, insulation resistance measurement, or replacement), normalized urgency score $p_i \in [0,1]$, and a binary constraint-compliance flag. This schema is directly compatible with IEC 55000-aligned CMMS formats used by major utility operators, enabling system integration without custom middleware. Operational constraint parameters, crew-hour budget $C_{\max}$, protection zone definitions $\{Z_k\}$, and criticality weights $\{w_i^{crit}\}$, are externally configurable inputs, allowing O&M planners to adjust scheduling policies per shift without model modification. While changing grid topology resulting from switching actions or network extension leads to a need for retraining only the graph representation and admittance matrix, the temporal encoder itself does not require retraining. The use of temperature scaling on risk scores post-hoc guarantees that the priority levels communicated to the engineers maintain their probabilistic interpretation, thus allowing clear communication between human operators and the system when the schedule created requires overriding information outside the training data.

5.3 Limitations and Scope

The design of the current system is analyzed and optimized with respect to a substation network at the transmission level with a fixed physical topology. In contrast, distribution networks are subject to continuous topology changes through switching actions lasting from seconds to minutes. This poses further difficulties for the static graph approach used in the GAT-based fusion layer: the admittance edge values relevant during a certain observation window do not capture the topology changes correctly within the following window, and the

system was trained under the assumption that there will be no topology changes within the same observation window. The design extension that would take care of this issue needs a graph update procedure that is out of the current design scope.

The equipment criticality weights w_i^{crit} and downstream impact coefficients ϕ_i^{impact} in equation (7) are derived from betweenness centrality analysis and load records specific to the six-substation evaluation network. Furthermore, the 120-second observation period used in determining the risk score is also consistent with the time scale of the annotated failure database and does not allow for the forecasting of slowly evolving faults such as gradual insulation aging or transformer oil degradation, which occur over much longer periods than the observation period.

6 Conclusions

In this paper, we present a decision support system for power grid O&M based on intelligent operations and maintenance which can tackle two shortcomings inherent in existing techniques, namely: (i) lack of topological awareness in fault diagnosis models based on equipment, and (ii) separation of diagnostics from maintenance decisions. To solve these problems, two design concepts have been introduced as additional modules atop a pre-existing deep learning architecture: (a) a state representation of equipment based on topological information using a graph attention-based gated neighbor fusion model at the feature representation stage, and (b) a module for converting diagnoses into decisions.

Experimental evaluation on a 12,847-sample real-world grid monitoring dataset with 23 fault categories demonstrates that the topology-aware fusion improves fault diagnosis accuracy by 6.2 percentage points to 94.3 ± 0.4% ($p < 0.01$) and Macro-F1 by 6.5 points to 92.8 ± 0.6%, with cross-device localization precision reaching 91.7 ± 0.9%, a 5.8-point improvement over the strongest topology-aware baseline. The diagnosis-to-decision bridging module achieves maintenance recommendation NDCG@5 of 0.887 ± 0.012 and Hit Rate@5 of 83.4 ± 1.8%, representing 10.4 and 11.6 percentage point improvements over the best scheduling baseline, with a constraint satisfaction rate of 96.7 ± 0.8%. Ablation results confirm that both gains are individually isolable to their respective design innovations and are not attributable to increased model parameter count or architectural complexity. The system operates at 21.4 ± 1.4 ms per sample, consistent with real-time SCADA deployment requirements.

The first problem revolves around dynamic topology modification, which involves adding the capability to handle switching activities inside cycles in distribution networks by using an algorithm to update the graph on-the-fly based on switching signals, without having to go through the process of re-training the model. The second problem involves temporal fault state trajectory prediction, which involves modifying the existing risk score from the current window by forecasting the future trajectory of faults.

About the Author

Xinzhu Dong was born in Beijing, China. P.R. China, in 2005. I obtained a high school diploma from Beijing No. 159 High School in China, and currently studying at Brunel London School, North China University of Technology, majoring in Mechanical Design, Manufacturing and Automation. My main research direction is mechanical design, large language models, and system control. cblclld@163.com

Yun Zhang is a Lecturer in the School of Mechanical and Materials Engineering at North

China University of Technology. He earned his Ph.D. in Mechanical Manufacturing and Automation from Beihang University and previously served as a postdoctoral researcher there. Dr. Zhang's research focuses on advanced CNC machining equipment and processes, CAD/CAM technology for complex curved surfaces, and digital twin applications in intelligent manufacturing. He has led or contributed to multiple national-level projects, including fundamental research on aero-engine blisk machining and high-efficiency milling for marine propellers. His work has been recognized with a Second Prize of the National Defense Technological Invention Award. He has published extensively in journals such as The International Journal of Advanced Manufacturing Technology, Measurement, and Proceedings of the Institution of Mechanical Engineers, and holds several invention patents related to blade machining and adaptive manufacturing systems.
zhangyun20260101@163.com

References

[1] Ohanu, C. P., Rufai, S. A., & Oluchi, U. C. (2024). A comprehensive review of recent developments in smart grid through renewable energy resources integration. *Heliyon, 10*(3), e25705.

[2] Mahmood, M., Chowdhury, P., Yeassin, R., Hasan, M., Ahmad, T., & Chowdhury, N.-U.-R. (2024). Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications. *Energy Conversion and Management: X, 24*, 100790.

[3] Boopathy, P., Liyanage, M., Deepa, N., Velavali, M., Reddy, S., Maddikunta, P. K. R., Khare, N., Gadekallu, T. R., Hwang, W.-J., & Pham, Q.-V. (2024). Deep learning for intelligent demand response and smart grids: A comprehensive survey. *Computer Science Review, 51*, 100617.

[4] Banad, Y. M., Sharif, S. S., & Rezaei, Z. (2025). Artificial intelligence and machine learning for smart grids: From foundational paradigms to emerging technologies with digital twin and large language model-driven intelligence. *Energy Conversion and Management: X, 28*, 101329.

[5] Shadi, M. R., Mirshekali, H., & Shaker, H. R. (2025). Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects. *Renewable and Sustainable Energy Reviews, 216*, 115668.

[6] Ngo, Q.-H., Nguyen, B. L. H., Zhang, J., Schoder, K., Ginn, H., & Vu, T. (2025). Deep graph neural network for fault detection and identification in distribution systems. *Electric Power Systems Research, 239*, 111721.

[7] Ding, L., Chen, Y., Xiao, T., Huang, S., Shen, C., & Guo, A. (2025). Topology-aware fault diagnosis for microgrid clusters with diverse scenarios generated by digital twins. *Applied Energy, 378*(Part A), 124794.

[8] Wang, F., & Hu, Z. (2025). A Novel Fault Diagnosis and Accurate Localization Method for a Power System Based on GraphSAGE Algorithm. *Electronics, 14*(6), 1219.

[9] Yang, Y., Sun, Q., Peng, J. C.-H., Tang, L. C., & Ye, Z. (2025). Robust Generator Maintenance Schedule for Frequency-Secure Power Systems. *Manufacturing & Service Operations Management*. Advance online publication.

[10] Tang, W., Mao, X., Lv, K., Cai, Z., & Ding, Z. (2025). Monthly Power Outage Maintenance Scheduling for Power Grids Based on Interpretable Reinforcement Learning. *Energies, 18*(20), 5454.

[11] Lee, G., Lee, S. J., & Lee, C. (2021). A convolutional neural network model for abnormality diagnosis in a nuclear power plant. *Applied Soft Computing, 99*, 106874.

[12] Wang, P., Zhang, J., Wan, J., & Wu, S. (2022). A fault diagnosis method for small pressurized water reactors based on long short-term memory networks. *Energy, 239*, 122298.

[13] Qian, G., & Liu, J. (2023). Fault diagnosis based on gated recurrent unit network with attention mechanism and transfer learning under few samples in nuclear power plants. *Progress in Nuclear Energy, 155*, 104502.

[14] Wang, F., Xiahou, T., Zhang, X., He, P., Yang, T., Niu, J., Liu, C., & Liu, Y. (2024). Convolutional preprocessing Transformer-based fault diagnosis for rectifier-filter circuits in nuclear power plants. *Reliability Engineering & System Safety, 249*, 110198.

[15] Hu, J., Hu, W., Chen, J., Cao, D., Zhang, Z., Liu, Z., ... & Blaabjerg, F. (2022). Fault location and classification for distribution systems based on deep graph learning methods. *Journal of Modern Power Systems and Clean Energy*, 11(1), 35-51.

[16] Ngo, Q.-H., Nguyen, B. L. H., Vu, T. V., Zhang, J., & Ngo, T. (2024). Physics-informed graphical neural network for power system state estimation. *Applied Energy, 358*, 122602.

[17] Lin, N., Orfanoudakis, S., Ordonez Cardenas, N., Giraldo, J. S., & Vergara, P. P. (2024). PowerFlowNet: Power flow approximation using message passing Graph Neural Networks. *International Journal of Electrical Power & Energy Systems, 160*, 110112.

[18] Wei, C., Pi, D., Ping, M., & Zhang, H. (2023). Short-term load forecasting using spatial-temporal embedding graph neural network. *Electric Power Systems Research, 225*, 109873.

[19] Chen, Y., Liu, Y., Zhao, J., Qiu, G., Yin, H., & Li, Z. (2023). Physical-assisted multi-agent graph reinforcement learning enabled fast voltage regulation for PV-rich active distribution network. *Applied Energy, 351*, 121743

[20] Liang, L., Zhang, H., Cao, S., Zhao, X., Li, H., & Chen, Z. (2024). Fault location method for distribution networks based on multi-head graph attention networks. *Frontiers in Energy Research, 12*, 1395737.

[21] Rajaoarisoa, L., Randrianandraina, R., Nalepa, G. J., & Gama, J. (2025). Decision-making systems improvement based on explainable artificial intelligence approaches for predictive maintenance. *Engineering Applications of Artificial Intelligence, 139*(Part B), 109601.

[22] Geng, S., Wang, X., & related coauthors. (2022). Predictive maintenance scheduling for multiple power equipment based on data-driven fault prediction. *Computers & Industrial Engineering, 164*, 107898.

[23] Sarajčev, P., Jakus, D., & Vasilj, J. (2020). Optimal scheduling of power transformers preventive maintenance with Bayesian statistical learning and influence diagrams. *Journal of Cleaner Production, 279*, 120850.

[24] Wang, Y., and coauthors. (2023). Opportunistic condition-based maintenance optimization for electrical distribution systems. *Reliability Engineering & System Safety, 236*, 109282.

[25] Lee, N., Woo, J., & Kim, S. (2025). A deep reinforcement learning ensemble for maintenance scheduling in offshore wind farms. *Applied Energy, 377*(Part A), 124431.