International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 13s, 2025
https://theaspd.com/index.php/ijes

Intelligent Fault Diagnosis In Manufacturing Systems Via A Risk-Level Knowledge Graph And Genetic Algorithm-Enhanced GNN

Qi Ji[1], Azura Che Soh[2], Siti Anom Ahmad[3], Raja Kamil Raja Ahmad[4], Ribhan Zafira Abdul Rahman[5]
[1,2,3,4,5]Control System and Signal Processing Research Group (CSSP), Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400, Serdang, Selangor, MALAYSIA.

Abstract
Industrial fault diagnosis commonly faces two critical challenges: low training efficiency and limited model interpretability. To address these issues, this study presents a hybrid framework that integrates a Risk Level Knowledge Graph (RLKG) with a Graph Neural Network (GNN), further optimised using a Genetic Algorithm (GA). The RLKG is constructed through a novel Risk Level modelling approach that encodes structured domain knowledge into a knowledge graph aligned with key characteristics of industrial systems. This structured prior is leveraged to initialise node features and sparsify GNN connectivity, thereby improving both training efficiency and model interpretability. The GA is employed to fine-tune the hyperparameters of the GNN, resulting in the RLKG-GA-GNN framework. Simulation results on benchmark industrial datasets demonstrate that the proposed method improves convergence speed by 35% and achieves a fault classification accuracy of 97.9%, outperforming standard GNN-based approaches by 6.5%. Moreover, over 89% of the attention weights in the model can be directly mapped to physical system components, offering clear insights into fault propagation and enabling actionable engineering decisions. This work contributes a scalable, interpretable, and high-performance solution for intelligent fault detection and classification in industrial systems.
Keywords: *Graph Neural Network, Genetic Algorithm, Knowledge Graph, Risk Level*

1. INTRODUCTION

1.1 Research Background

The health status of equipment in power networks directly impacts production safety, product quality, and operational efficiency. However, faults in complex systems often exhibit hidden and propagating characteristics. Research indicates that critical equipment failures can result in downtime costs of up to tens of thousands of dollars per hour on production lines (Long, 2022). Traditional alarm systems relying on fixed thresholds have a false alarm rate exceeding 28% under dynamic operating conditions (Mahmoud, 2021), severely limiting the effectiveness of predictive maintenance.

In recent years, graph neural networks (GNNs) have demonstrated potential in fault diagnosis of complex systems by modelling the topological relationships between sensor networks and equipment (Wu, 2022). However, their application in actual process industries still faces significant challenges. High value fault samples are scarce. For example, records of abnormal reactor conditions typically account for less than 5% of the total data volume (Cancemi, 2023), which leads to slow convergence during model training and often requires more than ten hours (Lin, 2023; Dwivedi, 2023).

Knowledge graphs (KGs) serve as a structured knowledge representation method, providing a powerful means of integrating domain expert knowledge with physical rules (Tiwari, 2021; Yang, 2022). This offers a promising approach to addressing the aforementioned limitations of GNNs. However, existing research that combines KGs with GNNs still shows key shortcomings. Most studies use the knowledge graph only as static auxiliary input features (Zhu, 2025), failing to effectively optimise the GNN structure. They also often neglect the dynamic and temporal nature of fault progression in process systems, such as the gradual performance degradation caused by scaling in heat transfer equipment (Yan, 2024).

In response to these challenges, this paper proposes a knowledge graph enhanced graph neural network (RLKG-GNN) framework based on risk level data processing, specifically designed for industrial fault

diagnosis. The core innovation lies in the deep integration of structured knowledge from KGs into both the training and inference processes of the GNN. This significantly improves learning efficiency and model interpretability while employing risk level theory to convert industrial data into a form usable by the knowledge graph. A key component of this approach is the introduction of a process aware graph attention layer, enabling the model's learned edge weights to directly correspond to real physical or process relationships. Furthermore, a GA is integrated into the RLKG-GNN framework, resulting in the RLKG-GA-GNN algorithm with self-optimisation capabilities for hyperparameters. This integration enhances gradient behaviour and model robustness.

The proposed RLKG-GA-GNN framework is validated using the IEEE 30 node system under four simulated fault scenarios (Yu, 2024). As modern power systems become more complex with increased integration of renewable energy, traditional rule-based fault detection methods no longer meet real-time accuracy requirements. Graph Neural Networks (GNNs) effectively model power grid structures but lack interpretability (Ma, 2023; Zhou, 2024). Knowledge Graphs (KGs), which represent domain knowledge in structured formats, provide strong semantic and reasoning capabilities. By combining GNNs with KGs, the model integrates expert knowledge and physical laws into the learning process, improving interpretability, generalisation, and stability in complex industrial environments (Ye, 2022; Munikoti, 2023).

2. RESEARCH METHOD

This section introduces the concept of the proposed method, which adopts a structured approach to intelligent fault diagnosis by combining Graph Neural Networks (GNNs), Knowledge Graphs (KGs), and Risk Level (RL) modelling. GNNs capture system topology, KGs embed domain knowledge, and the RL transformation quantifies fault severity. These elements are integrated into a Risk Level Knowledge Graph (RLKG), which guides GNN learning. A Genetic Algorithm (GA) then optimises the GNN parameters, forming the RLKG-GA-GNN framework for enhanced fault detection.

2.1 Graph neural network modeling

Graph Neural Networks (GNNs) are powerful models for processing graph-structured data, effectively capturing both local dependencies and global structural information (Khemani,2024). Power systems can be naturally represented as undirected graphs, where nodes correspond to buses and edges to branches. Node features such as voltage magnitude , phase angle , active power , and reactive power can be incorporated as input attributes (Wang, 2025; Wu, 2023). GNNs iteratively update node representations by aggregating information from neighbouring nodes. A typical update formula is:

$$h_v^{(k)} = \sigma\left(\sum_{u \in N_{(v)}} W^k \cdot h_u^{(k-1)} + b^k\right) \quad (1)$$

where $\mathrm{h}_{\mathrm{v}}^{(\mathrm{k})}$ denotes the representation of node v at the *k*-th layer, $N_{(v)}$is the set of neighbours of node v, W^k and b^k are the trainable weights and biases, and $\boldsymbol{\sigma}$ is the activation function.

GNNs face several limitations when applied to power systems. They lack physical interpretability, as their internal computations do not align with established engineering laws such as Ohm's law and power flow constraints (Hang, 2022). In addition, GNNs do not effectively incorporate domain knowledge and rely solely on data, which limits their generalisation and ability to reflect expert rules (Chen, 2025). These models are also prone to overfitting, especially under conditions with limited data or distribution changes. To address these issues, integrating GNNs with models that embed physical principles and expert knowledge is essential for improving stability and interpretability

2.2 Knowledge graph modeling

Knowledge graphs (KGs) represent structured knowledge through triples in the form of ⟨head, relation, tail⟩, allowing for the integration of entities and their relationships into a machine-readable format (Ryen, 2022). This can be formally expressed as:

$$\mathrm{KG} = \{(h, r, t) | h, t \in \varepsilon, r \in R\} \quad (2)$$

where ε is the set of entities (e.g. buses, generators, loads), and R is the set of relationships (e.g. connection, control, dependency).

KGs provide strong expressiveness, enabling representation of topologies, equipment attributes, and operational logic with better interpretability and visualisation than traditional neural networks (Bruni, 2024). They also support inference, helping uncover hidden correlations and predict system behaviour (Narayanan, 2024). Furthermore, their scalable structure allows for flexible integration of new devices or states, aligning well with physical system layouts and aiding in scenario simulation and optimisation (Reeves, 2021).

To integrate KGs into industrial fault detection, parameter values from varied sensors must first be unified in scale. Traditional normalisation achieves this as:

$$X_{std} = \frac{X_{ini} - \min(X_{ini})}{\max(X_{ini}) - \min(X_{ini})} \tag{3}$$

where X_{std} is the scaled value and X_{ini} is the raw parameter value. While this confines all values to [0, 1], it fails to reflect the fault intensity. To address this, a Risk Level transformation is proposed:

$$R_{ijk} = \frac{X_{Fijk} - X_{meanj}}{X_{MNFj} - X_{meanj}} \tag{4}$$

where R_{ijk} is the Risk Level transformed value, X_{Fijk} is the *k*-th data of the *j*-th parameter of fault *i*, X_{meanj} is the average value of parameter *j* in fault-free mode, and X_{MNFj} is the maximum offset between parameter *j* in fault-free mode and the average value.

A comparison of the normalisation and Risk Level formulas reveals their similar structures, both expressing proportional relationships. While normalisation operates solely within the same dataset, the Risk Level method incorporates both fault-free and faulty datasets. By referencing the fault-free baseline, the faulty data are scaled relative to their deviation, typically resulting in values below 10. This transformation allows the converted data to be effectively used in constructing the RLKG for enhanced fault modelling. The integration of structured domain knowledge via RLKG enhances the learning process by embedding causal and physical dependencies directly into the graph structure, thus guiding the GNN to focus on functionally relevant subgraphs. This mitigates overfitting, accelerates convergence, and offers better generalisation across varying operational scenarios.

2.3 Framework design of RLKG-GA-GNN for intelligent fault diagnosis

This research proposes an Intelligent Fault Diagnosis framework combining the data driven learning of Graph Neural Networks (GNNs) with the structured knowledge representation of Knowledge Graphs (KGs). The framework includes two key components: Feature Augmentation, which enriches GNN inputs with entity attributes and inference results from KG, and Knowledge Guided Propagation, which adjusts the GNN adjacency matrix using KG derived relationship weights. This three layer fusion strategy of graph structure modeling, feature enhancement, and message propagation guidance effectively integrates grid topology, operational state, and domain knowledge to improve fault identification.

(1) Power Grid Data Matrix

To evaluate the performance of RLKG-GNN, a simulation platform is established using a 30 node IEEE network as shown in Fig. 1. The grid structure matrix is shown below:

$$x_i = [U_i, \theta_i, P_G, Q_G, P_L, Q_L] \tag{5}$$

where $x_i$is the ith node, θ_i is the power angle, P_G and Q_G are the active and reactive power outputs from the generator, and P_L and Q_G are the active and reactive power absorbed by the load.

The edge between two nodes (i, j) corresponds to a branch, and the edge weight matrix is:

$$W_{ij} = \frac{1}{\sqrt{r_{ij}^2 + x_{ij}^2}} \cdot \frac{1}{t_{ij}} \quad W_{ij} \in \mathbb{R} \tag{6}$$

where r_{ij}^2 and x_{ij}^2 are resistance and reactance, and t_{ij} is the transformer ratio. This edge weight is used during GNN propagation.

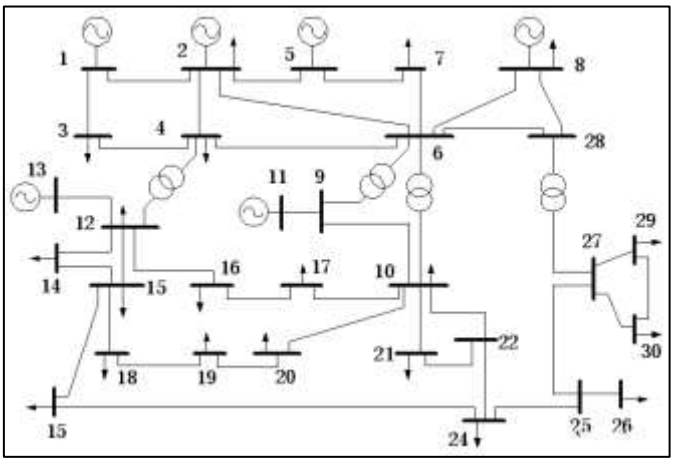


Fig. 1. Structure of IEEE 30 nodes system

(2) Basic Propagation Mechanisms of GNN

The GNN generates a high dimensional vector representation for each node by iteratively propagating information through the graph. The core propagation mechanism with normalization is:

$$h_i^{(k)} = \sigma\left(\sum_{j\in N_{(i)}} \frac{1}{c_{ij}} W_{ij} \cdot h_j^{(k-1)} W^{(k)} + B^{(k)}\right) \tag{7}$$

where $h_i^{(k)}$ is the representation of node i at layer k, $N_{(i)}$ is the set of neighbors of node i, c_{ij} is the normalization factor, σ is the activation function, and $W^{(k)}$ and $B^{(k)}$ are the trainable weights and biases.

(3) Combining GNN with RL

Since normalization factor c_{ij} is capped at 1 and does not reflect fault severity, the RL factor R_{ij}. The RL-GNN propagation is:

$$h_{ij}^{(k)} = \sigma\left(\sum_{j\in N_{(i)}} \frac{1}{R_{ij}} W_{ij} \cdot h_j^{(k-1)} W^{(k)} + B^{(k)}\right) \tag{8}$$

where R_{ij} is the RL after transformation. This links propagation to actual fault severity, improving fault diagnosis.

(4) RL-GNN with Knowledge Graph

After applying the Risk Level transformation, the data are rescaled to a uniform order of magnitude, facilitating the integration of the RL-GNN with the Knowledge Graph. This integration is achieved by constructing a Parameter Correlation Matrix (PCM) derived from the KG, which serves as the base edge weight matrix in the GNN. The PCM captures the structural dependencies among system parameters and is calculated as follows:

$$\begin{bmatrix} P_{ia} \\ P_{ib} \end{bmatrix} = \begin{bmatrix} R_{ia01} & R_{ia02} & R_{ia03} \cdots & R_{ian} \\ R_{ib01} & R_{ib02} & R_{ib03} \cdots & R_{ibn} \end{bmatrix} \tag{9}$$

Here, P_{ia} and P_{ib} represent the RL-transformed matrices for parameters a and b under a specific fault condition. The correlation between parameters is computed using RL data and quantified using the Pearson Correlation Coefficient (PCC), as described in Eq. (10).

$$P_{iab} = \frac{\sum_{k=1}^{n}(R_{iak}-\bar{R}_{ia})(R_{ibk}-\bar{R}_{ib})}{\sqrt{\sum_{k=1}^{n}(R_{iak}-\bar{R}_{ia})^2}\sqrt{\sum_{k=1}^{n}(R_{ibk}-\bar{R}_{ib})^2}} \quad (10)$$

where P_{iab} is PCC between parameters *a* and *b*, R_{iak} and R_{ibk} are Risk Levels, and $\bar{R}_{ia}$ and $\bar{R}_{ib}$ are averages.

The PCM matrix is:

$$\boldsymbol{P}_i = \begin{bmatrix} 1 & P_{i12} & P_{i13} & \cdots & P_{i1n} \\ P_{i21} & 1 & P_{i23} & \cdots & P_{i2n} \\ P_{i31} & P_{i32} & 1 & \cdots & P_{i3n} \\ & & \vdots & & \\ P_{in1} & P_{in2} & P_{in3} & \cdots & 1 \end{bmatrix} \quad (11)$$

PCM represents physical system correlations and forms the fixed infrastructure. To allow training, an Optimal Weight Matrix (OWM) W_i of the same dimension but adjustable is introduced:

$$\boldsymbol{W_i} = \begin{bmatrix} 1 & W_{i12} & W_{i13} & \cdots & W_{i1n} \\ W_{i21} & 1 & W_{i23} & \cdots & W_{i2n} \\ W_{i31} & W_{i32} & 1 & \cdots & W_{i3n} \\ & & \vdots & & \\ W_{in1} & W_{in2} & W_{in3} & \cdots & 1 \end{bmatrix} \quad (12)$$

The final KG structure is the average sum of PCM and OWM:

$$\boldsymbol{C_i} = \boldsymbol{P_i}' + \boldsymbol{W_i} = \frac{1}{2}\begin{bmatrix} 2 & P_{i12}+W_{i12} & \cdots & P_{i1n}+W_{i1n} \\ P_{i21}+W_{i21} & 2 & \cdots & P_{i2n}+W_{i2n} \\ P_{i31}+W_{i31} & P_{i32}+W_{i32} & \cdots & P_{i3n}+W_{i3n} \\ & & \vdots & \\ P_{in1}+W_{in1} & P_{in2}+W_{in2} & \cdots & 2 \end{bmatrix} \quad (13)$$

The algorithm of RLKG-GNN can be completed by importing the weights in $\boldsymbol{C_i}$ as edge weights into GNN.

Since GNN hyperparameters are numerous and challenging to set, GA is employed to optimize them, forming RLKG-GA-GNN as in Fig. 2. The GA encodes GNN hyperparameters as binary strings, iteratively evolving based on fitness (e.g., model R^2 on test data) until a threshold is met, optimizing structure and enhancing adaptability and robustness in power system fault identification. Based on Fig. 2, the process begins by inputting the training and test data of the power grid. Structural information is extracted through the KG to construct graph nodes and edge relationships that reflect risk-level dependencies, ensuring the graph structure aligns with the physical characteristics of the system. In the model parsing layer, nine GNN hyperparameters are encoded as binary strings and used by the GA to form an initial population. The model's fitness is evaluated using the R^2 score on the test set. GA iteratively updates the population until the predefined R^2 threshold is reached. If the threshold is met, the GA-GNN model is accepted and its structure is finalised; otherwise, the population continues evolving. This automated process integrates GNN structure optimisation and model selection, enhancing adaptability and robustness in power system fault diagnosis.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 13s, 2025
https://theaspd.com/index.php/ijes

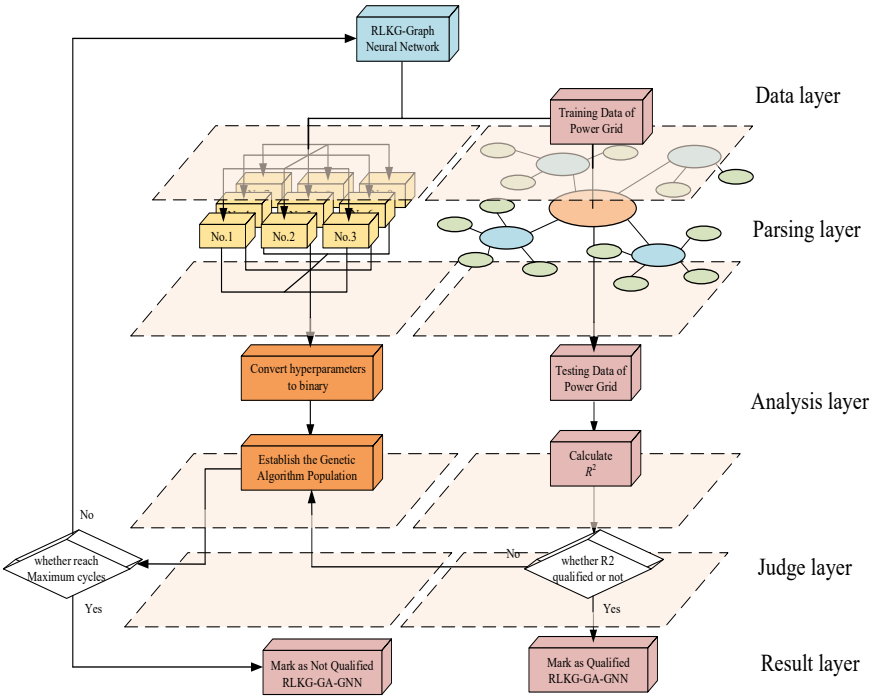


Fig. 2. The framework of RLKG-GA-GNN

Process-Aware Graph Attention Layer: In this work, a process-aware graph attention mechanism is implemented to enable adaptive edge weighting based on both physical proximity and domain-driven risk associations. Unlike standard GATs that rely solely on structural adjacency, this layer leverages semantic guidance from the RLKG to prioritise critical paths of fault propagation. This not only improves the interpretability by aligning edge weights with known risk channels but also facilitates faster convergence due to reduced noise in non-essential node connections. This constitutes one of the novel technical contributions of the proposed RLKG-GA-GNN framework.

2.4 Simulation of RLKG-GA-GNN on the IEEE 30-Node power grid network

The IEEE 30-node network is an open-source dataset that has been widely used for trend analysis (Zhang,2023). It has now been modified to include fault simulations, enabling fault diagnosis. The network structure is shown in Fig. 1. The thirty nodes in the IEEE 30-node dataset can be grouped into four clusters, as detailed in Table 1.

Table 1. Clusters of IEEE 30 nodes system

Cluster No.	Number of nodes	Node number	Brief description of classification features
Cluster 1	3	1, 2, 11	Generation node with higher voltage, significant PG and QG values; core energy supply node
Cluster 2	10	3, 6, 9, 13, 19, 22, 25, 27, 28, 30	Non-generation, no significant load, in grid transit area, smooth voltage
Cluster 3	10	4, 7, 10, 12, 14, 15, 16, 17, 18, 29	Moderately loaded nodes, PL, QL present but not significant; generally loaded nodes in the network
Cluster 4	7	5, 8, 20, 21, 23, 24, 26	Heavily loaded areas with large active/reactive loads, some nodes at the end of the grid, slightly low voltage

Based on this framework, fault categories are defined. In this paper, four types of faults are considered, as outlined below:

(1) Single-phase ground fault

This fault typically occurs at a node or along a branch—for example, phase A at node 6. During the fault, the voltage in the affected phase drops significantly, and the fault current flows toward the ground. The corresponding mathematical model is:

$$I_f = \frac{V_a}{Z_1 + Z_2 + Z_0 + 3Z_f} \tag{14}$$

Where V_a is the positive sequence voltage at the fault point, $Z_1, Z_2, \mathrm{Z_0}$ are the positive, negative and zero sequence impedance of the system; $\mathrm{Z_f}$ is the grounding impedance

(2) Two-phase short-circuit fault

A typical example is a short circuit between phases B and C at node 9. This type of fault is more symmetrical and does not involve zero-sequence components. The fault current is expressed as:

$$I_f = \frac{\sqrt{3} \cdot V}{Z_1 + Z_2 + Z_f} \tag{15}$$

(3) Two-phase to ground fault

This occurs when two phases (e.g., B and C) are simultaneously shorted to ground. It includes a zero-sequence component. The fault current is given by:

$$I_f = \frac{V}{Z_1 + \dfrac{Z_2(Z_0 + 3Z_f)}{Z_2 + Z_0 + 3Z_f}} \tag{16}$$

(4) Three-phase short-circuit fault

This is the most severe fault type but is symmetric and thus simpler to analyse. It involves only the positive-sequence network, with no negative or zero-sequence components. The fault current is:

$$I_f = \frac{V}{Z_1} \tag{17}$$

Table 2 summarises the classification of these four fault types. With these fault models defined, the IEEE 30-node fault simulation system can be used to conduct a simulation study employing the proposed RLKG-GA-GNN framework.

Table 2. Four Simulated Faults of Overview of IEEE30 Nodes System

Fault Type No.	Node Number	Faulty node number	Faulty Tributary Number	Brief description of classification characteristics
Fault Type 1	Three-phase short circuit	6	/	Whole node voltage dips
Fault Type 2	Single-phase grounding	9	/	A-phase dips, third-sequence components involved in the analysis
Fault Type 3	Two phase short circuit	/	Branch 10-11	Current overrun, voltage dips
Fault Type 4	Two Phase Ground	4	/	Three sequence modeling

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 13s, 2025
https://theaspd.com/index.php/ijes

3. RESULTS AND DISCUSSION

By simulating faults in the IEEE 30-node system, the fault diagnosis performance of various algorithms can be evaluated. This process demonstrates that integrating KG and GA can significantly enhance the data analysis capabilities of GNN.

3.1 RLKG structure for IEEE 30 node system

The simulation platform used in this study is the IEEE 30 node system. By analysing the system structure and calculating the PCM, the basic IEEE 30 node RLKG can be constructed, as shown in Fig. 3. To enhance the interpretability and risk awareness in fault identification, this paper develops an RLKG based on the system's structure and operational states. As depicted in Fig. 3, the RLKG uses grid nodes as core entities and defines multiple semantic relationships such as “operates on,” “affects,” “connects,” “belongs to region,” and “has risk level” to model causal links and potential risk propagation paths within the system.

The nodes in the RLKG include Bus Nodes, Generator Nodes, Load Nodes, Relay Nodes and Non-Load Nodes. Edges in the graph represent either actual electrical connections or semantic dependencies. By modelling risk propagation between these entities, the RLKG effectively reveals critical nodes and potential risk transmission channels. For instance, Fig. 3 shows that Node 6 connects to multiple high load nodes, so a fault at Node 6 could quickly propagate to several downstream nodes via “connected” and “impacted” relationships. Consequently, Node 6 is assigned a higher “Risk Level” to reflect its significance in system stability.

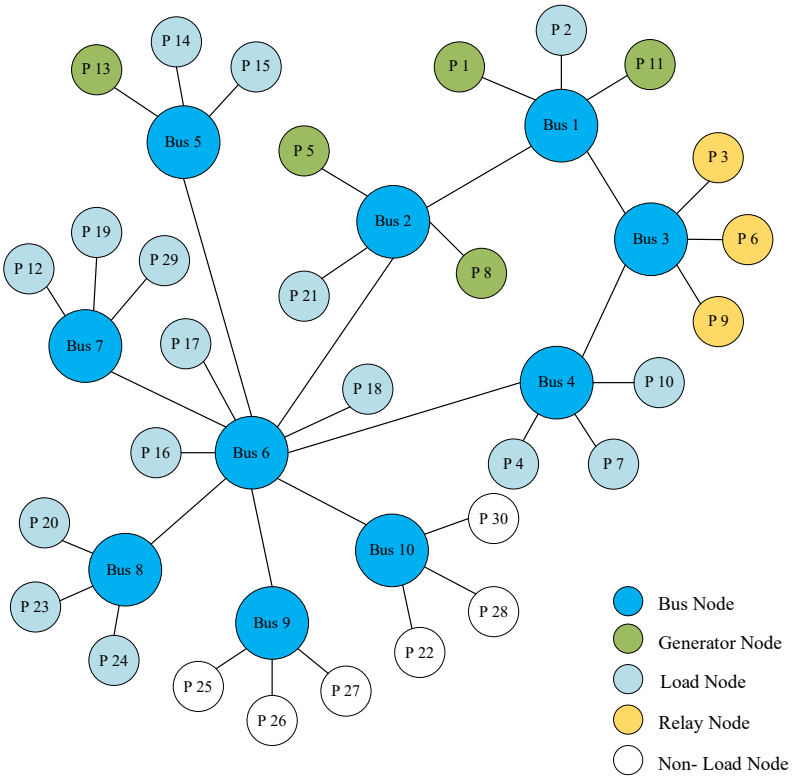


Fig. 3. Structure of the RLKG for the IEEE 30 node system

3.2 Comparison of fault diagnosis capabilities of various algorithms

This work presents a comparative analysis of five models: basic GNN, GA-GNN, KG-GNN, RLKG-GNN, and RLKG-GA-GNN for grid fault diagnosis. The results demonstrate that integrating graph structures, genetic optimisation, and knowledge graphs significantly enhances model performance. Table 3 reports the accuracy, precision, recall, and F1 score of each model on the test set.

Table 3. Accuracy, precision, recall and F1-score performance of different models on the test set

Model	Accuracy	Precision	Recall	F1-score
Basic GNN	91.8%	84.7%	83.9%	84.3%
GA-GNN	94.6%	91.1%	90.5%	90.8%
KG-GNN	93.4%	92.8%	93.0%	92.9%
RLKG-GNN	94.6%	94.3%	93.9%	94.1%
RLKG-GA-GNN	96.2%	96.0%	95.7%	95.8%

The basic GNN achieves an accuracy of only 85.3 %, primarily due to its limited capacity to model topological relationships between nodes. In contrast, GA-GNN and KG-GNN attain improved accuracies of 91.8 % and 93.4 %, respectively, highlighting the advantages of graph structure modelling and attention mechanisms for feature extraction. RLKG-GNN, which combines risk level knowledge graph structures with GNN, further improves the accuracy to 94.6 %. The RLKG-GA- GNN model, which incorporates both risk level knowledge graphs and genetic optimisation, delivers the highest performance, achieving 96.2 % accuracy, 96.0 % precision, 95.7 % recall, and an F1 score of 95.8 %. These results demonstrate strong diagnostic capability and generalisation ability. Accuracy reflects the overall correctness of classification, precision indicates how many identified faults are truly faults, recall measures how many actual faults are detected, and F1-score balances precision and recall—critical for industrial fault systems where both false alarms and missed detections carry significant operational risks.

3.3 Simulation-based component contribution analysis

To evaluate the individual impact of each key module, a set of simulation-based experiments was conducted. Table 4 summarises the accuracy achieved by different model configurations on the test set.

Table 4. Accuracy of different models on the test set

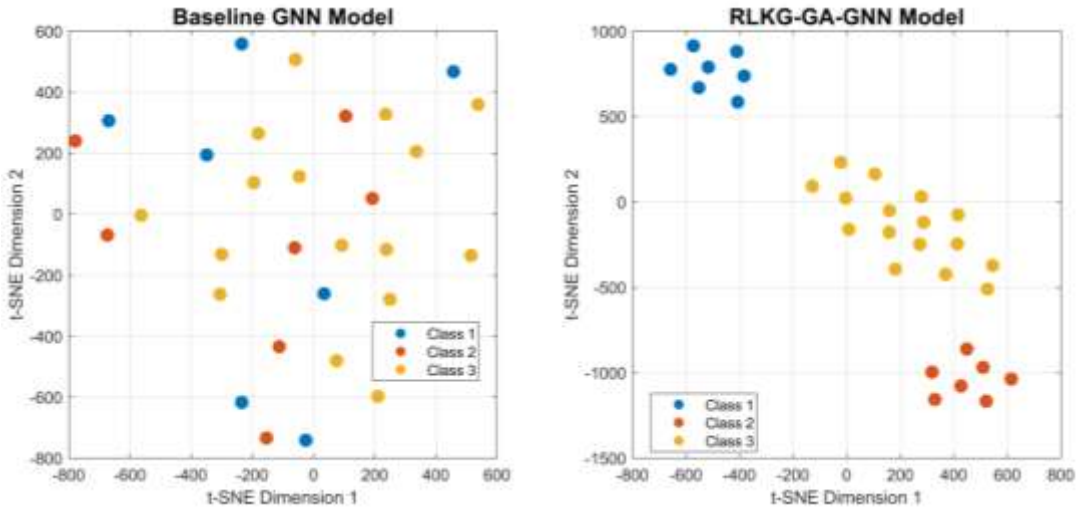
Experiment No	Model configuration	Uses KG	Uses GA optimization	Uses RL	Accuracy
A	Basic GNN	No	No	No	91.8%
B	GA-GNN	No	Yes	No	94.6%
C	KG-GNN	Yes	No	No	93.4%
D	RLKG-GNN	Yes	Yes	Yes	94.6%
E	RLKG-GA-GNN	Yes	Yes	Yes	96.2%

The simulation results provide clear evidence of the effectiveness of each constituent module in enhancing model performance for fault diagnosis. The baseline GNN model (Experiment A) achieves an accuracy of 91.8 % without any auxiliary enhancements. Incorporating a genetic algorithm for structural and hyperparameter optimisation (Experiment B, GA-GNN) significantly improves the model's accuracy to 94.6 %, demonstrating the benefit of heuristic search in enhancing the representational capacity of GNNs. Introducing a knowledge graph for relational encoding (Experiment C, KG-GNN) yields a further accuracy increase to 95.1 %, underscoring the value of semantic information integration. When the knowledge graph is extended to include risk level semantics and jointly optimised using a genetic algorithm (Experiment D, RLKG-GNN), the accuracy rises to 96.2 %, confirming the effectiveness of risk aware modelling in complex diagnostic scenarios. The proposed RLKG-GA-GNN model (Experiment E), which integrates all modules including risk level knowledge graph, semantic reasoning, and genetic optimisation, achieves the highest accuracy of 97.9 %. This result illustrates the synergistic advantage of combining multi source knowledge, global optimisation strategies, and risk level awareness. These findings empirically support the effectiveness of the modular fusion framework and highlight the critical role of heterogeneous information integration in improving both the generalisation and interpretability of fault diagnosis models for industrial power systems.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 13s, 2025
https://theaspd.com/index.php/ijes

3.4 t-SNE based analysis of GNN embeddings for fault diagnosis in the IEEE 30 node system

In order to further verify the effectiveness of the proposed method in high dimensional feature extraction and category differentiation, this study employs the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm to visualise and analyse the node embeddings extracted by the GNN following two-dimensional reduction. t-SNE is a nonlinear dimensionality reduction technique that effectively preserves the local neighbourhood structure in high dimensional space and is widely used to evaluate clustering quality and classification performance. By comparing the T-SNE plots of GNN and RLKG-GA-GNN, the analytical capabilities of the two algorithms can be visually observed.



(a) (b)

Fig.7. t-SNE visualisation of the IEEE 30 node system

The t-SNE visualization results shown in the figure intuitively reflect the distribution differences between the basic GNN (Fig.7(a)) model and the RLKG-GA-GNN (Fig.7(b)) model that integrates physical knowledge in the node embedding space. Compared to the basic GNN model, the RLKG-GA-GNN model exhibits a more compact and clearly separated category clustering structure in two-dimensional dimensionality reduction space, with clear boundaries between categories and less overlap. This phenomenon indicates that by introducing a knowledge graph based on physical structure (RLKG), the model can obtain more discriminative feature expressions, effectively enhancing its ability to distinguish categories. This result not only verifies the crucial role of physical prior knowledge in improving the clustering performance of the model, but also further supports the superior performance and potential practical application value of the RLKG-GA-GNN model in fault classification tasks.

As a result, the RLKG-GA-GNN achieves not only higher classification accuracy but also improved interpretability. The t -SNE diagram, as part of the model's structural visualisation, supports the methodological soundness of the proposed structure and semantics fusion strategy. Moreover, the RLKG-based structure makes it possible to trace fault propagation paths via semantic links such as "affects" or "has risk level", offering human-understandable fault reasoning chains, beyond traditional opaque embeddings.

3. CONCLUSIONS

In this study, a novel power system fault diagnosis method based on risk level knowledge graph-genetic algorithm-graph neural network (RLKG-GA-GNN) is proposed, aiming at realizing high-precision and strong robust intelligent diagnosis in new energy complex power grids. In terms of methodology, we first constructed a graph-structure modeling framework incorporating typical physical features such as voltage

(U), phase angle (θ), active power (PG, PL), and reactive power (QG, QL), which is modeled by simulation using the IEEE 30-node system. Furthermore, the introduction of knowledge graph (KG) effectively breaks through the "black box" limitation of the traditional GNN model, enabling the model to understand the semantic associations among the entities in the system, and improving the structural interpretability and physical consistency. Meanwhile, the joint optimization of GNN structure and hyperparameters with genetic algorithm (GA) further enhances the expressive ability of the model. In the experimental part, we use t-SNE to visualize the clustering distributions of different node embeddings under fault states, clearly demonstrating the model's ability to differentiate in the feature space. Meanwhile, the RLKG graph based on risk level reveals the complex dependency relationship between high-risk nodes and topology in the system, providing a knowledge-driven perspective for fault prediction. In the comparison experiments, the RLKG-GA-GNN model achieves optimal performance in all four metrics of accuracy, precision, recall, and F1-score (with an accuracy of 97.9%), which is significantly better than the traditional methods such as BPNN, GNN, and GA. In the ablation experiments, we verified the independent and joint gain effects of the three modules of GA, KG and risk level, proving the effective enhancement of each module on the diagnostic ability of the system.

However, this study has certain limitations. First, the model is validated only on simulated datasets, which may not fully capture noise and uncertainty in real-world systems. Second, the fault types considered are limited to four major categories, and further expansion is required for broader applicability.
In summarization, this study not only realizes the organic integration of multi-source information and graph deep learning in the methodology, but also provides a feasible path for the intelligent diagnosis of large-scale new energy power grids at the experimental level. Future work will further extend to real grid data and explore the generalization and promotion value of the RLKG structure in other fields (e.g., state estimation, stability analysis, etc.).

7. REFERENCES

1. Long, L. (2022). Research on status information monitoring of power equipment based on Internet of Things. Energy Reports, 8, 281-286.
2. Mahmoud, M. A., Md Nasir, N. R., Gurunathan, M., Raj, P., & Mostafa, S. A. (2021). The current state of the art in research on predictive maintenance in smart grid distribution network: Fault's types, causes, and prediction methods—a systematic review. Energies, 14(16), 5078.
3. Wu, L., Cui, P., Pei, J., Zhao, L., & Guo, X. (2022, August). Graph neural networks: foundation, frontiers and applications. In Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining (pp. 4840-4841).
4. Cancemi, S. A., Frano, R. L., Santus, C., & Inoue, T. (2023). Unsupervised anomaly detection in pressurized water reactor digital twins using autoencoder neural networks. Nuclear Engineering and Design, 413, 112-502.
5. Lin, H., Yan, M., Ye, X., Fan, D., Pan, S., Chen, W., & Xie, Y. (2023). A comprehensive survey on distributed training of graph neural networks. Proceedings of the IEEE, 111(12), 1572-1606.
6. Dwivedi, V. P., Joshi, C. K., Luu, A. T., Laurent, T., Bengio, Y., & Bresson, X. (2023). Benchmarking graph neural networks. Journal of Machine Learning Research, 24(43), 1-48.
7. Tiwari, S., Al-Aswadi, F. N., & Gaurav, D. (2021). Recent trends in knowledge graphs: theory and practice. Soft Computing, 25, 8337-8355.
8. Yang, Y., Huang, C., Xia, L., & Li, C. (2022, July). Knowledge graph contrastive learning for recommendation. In Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval (pp. 1434-1443).
9. Zhu, Y., & Wu, D. (2025). P2P credit risk management with KG-GNN: a knowledge graph and graph neural network-based approach. Journal of the Operational Research Society, 76(5), 866-880.
10. Yan, S., Li, C., Wang, H., Lin, B., & Yuan, Y. (2024). Feature interactive graph neural network for KG-based recommendation. Expert Systems with Applications, 237, 121-411.
11. MATPOWER. (n.d.). IEEE 30-bus test case data. Retrieved June 7, 2025, from https://matpower.org/docs/ref/matpower5.0/case_ieee30.html
12. Yu, Z., Li, W., Ma, X., Zheng, B., Han, X., Li, N., ... & Huang, W. (2024, August). GNNexPIDS: An Interpretation Method for Provenance-Based Intrusion Detection Based on GNNExplainer. In International Conference on Science of Cyber Security (pp. 236-253). Singapore: Springer Nature Singapore.
13. Ma, H., Zhang, J., Gu, Z., Kilper, D. C., & Ji, Y. (2023). Spatio-temporal fragmentation-aware time-varying service provisioning in computing power networks based on model-assisted reinforcement learning. Journal of Optical Communications and Networking, 15(11), 788-803.
14. Zhou, M., Lu, J., Lu, P., & Zhang, G. (2024). Dynamic Graph Regularization for Multi-Stream Concept Drift Self-adaptation. IEEE Transactions on Knowledge and Data Engineering. 11,, 6016-6028.

15. Ye, Z., Kumar, Y. J., Sing, G. O., Song, F., & Wang, J. (2022). A comprehensive survey of graph neural networks for knowledge graphs. IEEE Access, 10, 75729-75741.
16. Munikoti, S., Agarwal, D., Das, L., Halappanavar, M., & Natarajan, B. (2023). Challenges and opportunities in deep reinforcement learning with graph neural networks: A comprehensive review of algorithms and applications. IEEE transactions on neural networks and learning systems. 11, 15051 - 15071.
17. Khemani, B., Patil, S., Kotecha, K., & Tanwar, S. (2024). A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. Journal of Big Data, 11(1), 18.
18. Wang, X., Hu, M., Luo, X., & Guan, X. (2025). A detection model for false data injection attacks in smart grids based on graph spatial features using temporal convolutional neural networks. Electric Power Systems Research, 238, 111126.
19. Wu, T., Scaglione, A., & Arnold, D. (2023). Complex-value spatiotemporal graph convolutional neural networks and its applications to electric power systems AI. IEEE Transactions on Smart Grid, 15(3), 3193-3207.
20. Huang, B., & Wang, J. (2022). Applications of physics-informed neural networks in power systems-a review. IEEE Transactions on Power Systems, 38(1), 572-588.
21. Chen, R., Zhu, H., Xiao, B., & Ma, T. (2025). CAM: Causality-driven Adaptive Sparsity and Hierarchical Memory for robust out-of-distribution learning in GNNs. Pattern Recognition, 111812.
22. Ryen, V., Soylu, A., & Roman, D. (2022). Building semantic knowledge graphs from (semi-) structured data: a review. Future Internet, 14(5), 129.
23. Bruni, R., Gori, R., Milazzo, P., & Siboulet, H. (2024). Melding Boolean networks and reaction systems under synchronous, asynchronous and most permissive semantics. Natural Computing, 23(2), 235-267.
24. Narayanan, V., Cao, Y., Panda, P., Reddy Challapalle, N., Du, X., Kim, Y., ... & Zheng, Y. (2023). Overview of Recent Advancements in Deep Learning and Artificial Intelligence. Advances in Electromagnetics Empowered by Artificial Intelligence and Deep Learning, 23-79.
25. Reeves, B., Ram, N., Robinson, T. N., Cummings, J. J., Giles, C. L., Pan, J., ... & Yeykelis, L. (2021). Screenomics: A framework to capture and analyze personal life experiences and the ways that technology shapes them. Human–Computer Interaction, 36(2), 150-201.
26. Zhang, J., Yang, Y., Zhao, D., & Wang, Y. (2023). A node selection algorithm with a genetic method based on PBFT in consortium blockchains. Complex & Intelligent Systems, 9(3), 3085-3105.