Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria
 Published by Research Publishing, Singapore.
doi: 10.3850/ESREL2026061419_esrel26-p26175-cd

Toward Knowledge-Augmented Graph Neural Networks for Explainable Root Cause Analysis: A Multilevel Flow Modeling Approach

Ruixue Li

Center for Risk and Reliability, University of Maryland, College Park, MD, USA. E-mail: ruilia21@umd.edu

Katrina M. Groth

Center for Risk and Reliability, University of Maryland, College Park, MD, USA. E-mail: kgroth@umd.edu

Safety-critical process systems face challenges in fault diagnosis due to limited labeled failure data and complex non-linear coupling between components. Purely data-driven Prognostics and Health Management (PHM) methods can detect anomalies but often lack causal interpretability, limiting trust in safety-critical decision-making. This study explores a knowledge-infused framework that integrates Multilevel Flow Modeling (MFM), a functional modeling approach that encodes causal reasoning, into a Graph Neural Network (GNN) based diagnostic architecture. We construct a multi-relational graph representation where nodes represent system components and edges capture physical connections (mass/energy flow/control logic) and causal relationships (means-end relations). MFM knowledge is injected into the GNN through (1) differentiated graph structures reflecting physical and causal dependencies, and (2) causality-aware loss functions enforcing consistency with MFM propagation paths. Preliminary validation on a Minox deoxygenation system demonstrates that incorporating MFM-based functional constraints enhances Top-1 diagnostic accuracy by 10.0% compared to standard physical-topology GNNs, achieving a Top-3 accuracy of 98.5%. The framework's explainability is showcased through diagnostic heatmaps, providing a proof-of-concept for traceable decision support. This study explores the potential of bridging symbolic functional reasoning and graph-based deep learning, offering initial insights into enhancing the interpretability of safety-critical PHM.

Keywords: Prognostics and Health Management (PHM); Causal reasoning; Graph Neural Network(GNN); Knowledge-based modeling; Multilevel Flow Modelling (MFM); Artificial intelligence; Root Cause Analysis; Safety and reliability; Explainable Artificial Intelligence (XAI); Safety-Critical Systems

1. Introduction

Prognostics and Health Management (PHM) is an integrated framework combining monitoring, diagnosis, and prognosis to support maintenance decision-making in complex industrial systems (Qiu et al., 2023). In recent years, data-driven deep learning (DL) has become the cornerstone of PHM due to its powerful feature representation capabilities (Yu and Zhang, 2023). However, standard DL models like CNNs or RNNs often treat sensor data as independent streams, failing to explicitly capture the structural interdependencies and heterogeneous physical meanings of multi-sensor measurements (Moradi and Groth, 2020; Li et al., 2022). This ”black-box” nature leads to a lack of causal interpretability, which is critical for trust in safety-critical decision-making (Taqvi et al., 2021).

Graph Neural Networks (GNNs) have emerged to bridge this gap by representing industrial systems as structured graphs, where nodes symbolize sensors or components and edges capture the complex underlying dependencies or physical couplings between them (Zhou et al., 2020). The principle of GNN-based fault diagnosis is to learn the high-dimensional topological representations of a system by capturing the dependency patterns between sensor nodes (Chen et al., 2021). Recent studies have applied Graph Convolutional Networks (GCNs) to learn topology-aware features from vibration signals and process variables, improving accuracy in system-level and component-level health assessments (Ruiz-Tagle Palazuelos et al., 2020). Furthermore, research by Nguyen and Chioua (2024) has shown that embedding fault propagation paths directly into GNN architectures can enhance performance by capturing how anomalies move through chemical process topologies. By analyzing node activations within these graph-based paths, it becomes possible to verify diagnostic decisions and improve system reliability. Studies on adaptive GNNs have demonstrated the ability to dynamically construct graph structures that reflect complex sensor relationships in three-phase flow systems, effectively extracting fault features for multi-class diagnosis (Ouyang and Jin, 2024). By integrating attention mechanisms, GNNs can further isolate informative sensors and highlight suspicious fault propagation paths, offering a degree of visual interpretability (Lv et al., 2022).

Despite these advances, a significant challenge remains: most GNN-based diagnostics rely on data-driven graph construction (e.g., K-nearest neighbors or similarity matrices). Such graphs often capture statistical correlations rather than true physical or functional causality, potentially leading to misleading diagnostic results in safety-critical systems. Furthermore, purely data-driven models struggle when faced with the scarcity of labeled failure samples common in high-reliability processes. To ensure both robustness and transparency, it is essential to incorporate domain knowledge into the network architecture or loss functions (Dash et al., 2022).

In this study, we propose an exploratory knowledge-augmented GNN framework designed for explainable root cause analysis. By leveraging Multilevel Flow Modeling (MFM), a functional modeling approach that encodes causal reasoning between system goals and component functions, we define a causal topology to guide the representation learning within the GNN. Moving beyond purely data-driven pattern matching, our approach incorporates MFM-derived constraints to guide the network toward physically consistent reasoning, potentially enabling the inference of faults in components with limited sensor coverage. The primary contributions of this paper are: (1) knowledge-infused graph construction, which integrates MFM functional logic with physical mass/energy flows to build a multi-relational graph representation that captures both structural and causal dependencies; and (2) explainable root cause localization by generating node-level anomaly scores, visualized as diagnostic heatmaps, to trace failure origins. This provides a proof-of-concept for achieving transparency in safety-critical PHM and establishes a methodological foundation for integrating functional reasoning into deep learning-based diagnosis.

2. Background

2.1. *Multilevel Flow Modelling (MFM)*

MFM is ontology-based qualitative reasoning in the field of explainable AI methods (Lind, 2011). It represents the goals and functions of complex industrial processes through a hierarchical abstraction. Unlike purely physical schematics, MFM deconstructs a system into three distinct layers: objectives (goals), functional flows (mass, energy, and control information), and physical components. Objectives or goals define the desired operational states (e.g., maintaining reactor temperature). Functional Flows model the movement of mass, energy, and information using standardized functions such as Source, Sink, Transport, and Storage. The core strength of MFM lies in its means-end hierarchy, which explicitly defines the causal dependencies, linking a physical component's behavior to the fulfillment of a higher-level function, and subsequently to the system's goals. By utilizing standardized flow structures, MFM provides a robust semantic backbone for diagnosing fault propagation paths, moving beyond simple connectivity to functional causality (Song et al., 2023; Lind and Zhang, 2014). Fig. 1 shows how MFM models are constructed and applied.

2.2. *Graph Neural Network (GNN)*

GNNs are suitable for reliability engineering applications, as industrial systems are inherently structured as networks of interdependent components, forming non-Euclidean graph topologies that conventional machine learning methods struggle to represent. A Graph $\mathcal{G}$ is a mathematical structure used to model pairwise relations between objects. It is formally defined as a tuple $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where: (1) Nodes (Vertices): $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ represents the set of n entities

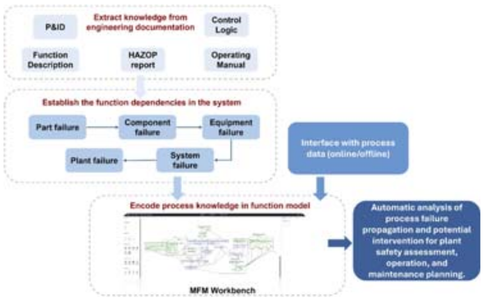

Fig. 1. MFM method for process knowledge management using MFM workbench. (Li et al., 2025)

within the system. Each node v_i is typically associated with a feature vector $\mathbf{x}_i \in \mathbb{R}^d$, which encapsulates its local attributes. The collective features are represented by the matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$. (2) Edges: $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denotes the set of links between nodes. An edge $e_{i,j} \in \mathcal{E}$ indicates a dependency or relationship between node v_i and v_j. (3) Adjacency Matrix: The topological structure is encoded in a matrix $\mathbf{A} \in \{0,1\}^{n \times n}$, where $A_{ij} = 1$ if $(v_i, v_j) \in \mathcal{E}$, and 0 otherwise.

At the core of a GNN is the message-passing mechanism (Gilmer et al., 2017), which iteratively aggregates information from a node's neighbors to update its latent representation. Formally, for a node v_i, the hidden state $h_i^{(l+1)}$ at layer $l+1$ is updated as:

$$h_i^{(l+1)} = \text{UPDATE}^{(l)}\Big(h_i^{(l)}, \text{AGGREGATE}^{(l)}\big(\{h_j^{(l)}, \forall j \in \mathcal{N}(i)\}\big)\Big) \tag{1}$$

where $\mathcal{N}(i)$ denotes the set of adjacent nodes. In a standard Graph Convolutional Network (GCN), this simplifies to a weighted sum of neighbor features followed by a non-linear transformation:

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i) \cup \{i\}} \frac{1}{\sqrt{\hat{d}_i \hat{d}_j}} \mathbf{W}^{(l)} h_j^{(l)}\right) \tag{2}$$

where $\hat{d}$ represents the node degree and $\mathbf{W}$ is a learnable weight matrix, and σ is a non-linear activation function.

3. Proposed methodology

3.1. *Multi-relational graph construction*

Industrial process systems are essentially a composite of physical entities and logical rules, modeled as a multi-relational heterogeneous graph. Rather than a single homogeneous connectivity matrix, the system topology is decoupled into several functional domains based on physical and logical principles:

- Physical flow layers ($\mathbf{A}_{phys}$): Represent the directional transport of mass and energy, governed by conservation laws.
- Control & Information layers ($\mathbf{A}_{ctrl}$): Map non-physical feedback loops and signal dependencies between sensors and actuators.
- Functional hierarchy layers ($\mathbf{A}_{mfm}$): Represent the Means-End causality, linking low-level components to functions and to high-level system goals.

The system topology is no longer defined by a single matrix, but rather by a tensor representation composed of a set of heterogeneous relationship matrices:

$$\mathcal{A} = \{\mathbf{A}_r\}_{r=1}^{R}, \quad \mathbf{A}_r \in \{0,1\}^{N \times N}$$

where A is the adjacency matrix. For each relationship $r \in \{\text{mass, energy, control, means-end}\}$, the matrix element is defined as:

$$(\mathbf{A}_r)_{i,j} = \begin{cases} 1, & \text{if a directional } r\text{-type influence} \\ & \text{exists from } i \text{ to } j \\ 0, & \text{otherwise} \end{cases}$$

The significance of this multi-relational graph construction lies in its ability to translate domain-specific engineering knowledge into a machine-readable topological prior. By decoupling the complex interdependencies of industrial processes into specialized relational domains, the framework provides a robust foundation for explainable graph reasoning, moving beyond simple correlation to causal-aware fault localization.

3.2. *Feature matrix construction*

To represent the comprehensive state of the industrial system, a multi-dimensional feature matrix $X \in \mathbb{R}^{N \times D}$ is constructed, where N is the number of nodes and D is the total feature dimension. The feature vector for each node i is formulated as a concatenation of four distinct information modalities:

$$X_i = [S_i \mid\mid T_i] \tag{3}$$

where S_i represents sensor data, T_i represents node type.

3.2.1. *Observation features: Sensor data*

The sensor data (S_i) component captures real-time physical observations from sensors (e.g., temperature, pressure, and flow rate). For each node i in the GNN-MFM graph, a fixed-length observation vector $Si \in \mathbb{R}^k$ (where k is the maximum feature slots) is constructed to maintain a uniform input tensor shape for the GNN:

$$\mathbf{S}_i = [\text{norm}(s_{i,1}), \text{norm}(s_{i,2}), \ldots, \text{norm}(s_{i,k})] \in \mathbb{R}^k$$

- Normalization: All raw readings (e.g., pressure, temperature) are scaled to a $[0, 1]$ range to eliminate unit discrepancies.
- Feature alignment: For nodes with fewer than k sensors, a zero-padding is applied.

3.2.2. *Semantic features: Node type*

To bridge the gap between raw data and physical meaning, node type (T_i) is proposed as a functional semantic mapping layer. Components are grouped into broader functional classes (e.g., flow control, energy transformation, or material separation). Physical laws that apply to a class of devices regardless of their specific location in the plant. This categorical identity is encoded as a One-hot vector:

$$\mathbf{T}_i = [t_{i,1}, t_{i,2}, \ldots, t_{i,M}] \in \{0, 1\}^M$$

where $t_{i,j} = 1$ if node i belongs to category j. This encoding serves as a structural prior, allowing the GCN layers to distinguish between different types of message passing, for instance, prioritizing energy-related features when traversing nodes of heat devices. By fusing these static functional semantics with dynamic physical states, the model achieves a deeper understanding of the heterogeneous nature of the complex engineering system.

3.3. *Multi-relational GNN architecture*

The proposed multi-relational GNN architecture is designed to perform root cause localization by integrating multi-relational physical topologies with MFM-informed causal constraints. The framework consists of six stages as illustrated in Fig. 2. The contribution of this architecture can be found in Table 1.

Multi-relational topology and feature fusion can be found in Section 3.1 and 3.2. The core of the model is an n-layer Graph Convolution Network (GCN). Each layer performs message passing where node features X are updated by aggregating information from neighbors across all relational graphs. The transformed graph is fed into a classification layer (MLP Scorer). The MLP maps the high-dimensional latent features of each node to a scalar score $S \in [0, 1]$. These scores are visualized as a diagnostic heatmap, where the node with the highest score is identified as the most probable root cause (Top-1).

3.4. *MFM-guided causal constraints*

To ensure the diagnosis aligns with physical logic, edge-level causal constraints based on MFM are implemented. This mechanism penalizes anomaly scores that violate the established means-end paths, $\mathbf{A}_{mfm}$, enabling the model to distinguish between the actual source of a fault and its downstream symptomatic neighbors. Since neural networks are probabilistic models prone to logical errors, by incorporating $pred[dst] - pred[src]$ into the loss function, we enforce the industrial principle that upstream fault scores must exceed downstream ones. The total loss represents as follows:

$$L_{total} = L_{BCE} + \lambda \cdot \frac{1}{|\mathcal{E}_{MFM}|} \sum_{(u,v) \in \mathcal{E}_{MFM}} \max(0, S_v - S_u + \gamma) \tag{4}$$

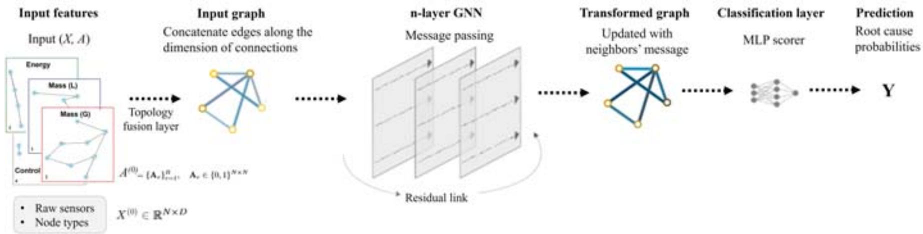

Fig. 2. Multi-relational GNN architecture.

Table 1. Comparison of GNN and proposed GNN-MFM method.

Component	Baseline GNN (Ruiz-Tagle Palazuelos et al., 2020)	GNN-MFM
Graph Structure	Single graph	Multi-relational
Number of the graph	1 (physical connection)	5 (liquid/gas/energy/control/means-end)
Domain knowledge	×	MFM-guided constraints
Output	Binary classification (healthy/unhealthy)	Heat map (anomaly score for each node)
Extra constraint method	×	Edge-level causal constraints

where $(u, v) \in \mathcal{E}_{MFM}$ represents a causal edge in the means-end diagram, where u is the means (cause) and v is the end (effect). $|\mathcal{E}_{MFM}|$ represents the total number of edges in the MFM causal graph. S_u and S_v are the anomaly scores output by the model. γ is the margin parameter.

4. Case study

4.1. *Process description*

The Minox system deoxygenates seawater for oil and gas platforms through a gas-stripping cycle, shown in Fig. 3. Seawater is mixed with stripping gas in separators (SP1/SP2) to remove oxygen; the rich gas is then heated and reacted with methanol in a deoxidizer (DeoX) to catalytically remove the captured oxygen. This exothermic process purifies the gas for continuous reuse in the stripping loop. Table 2 lists the components with related sensors.

4.2. *Feature matrix*

Each node is characterized by a 10-dimensional feature vector $\mathbf{x}_i \in \mathbb{R}^{10}$. Nodes of the GNN-MFM graph can be found in Fig. 4, including all the component nodes and the goal node. 2 types of features are included in the X matrix, the first 4 dimensions store the sensor data. Dim 4-

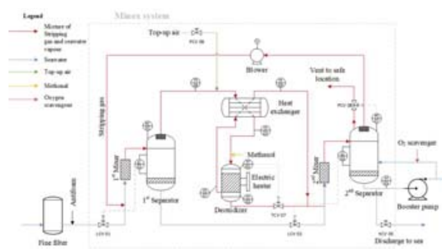

Fig. 3. The Minox system in a water injection system flow chart. (Li et al., 2026)

9 represents the node type, listed as follows. The feature matrix X matrix is shown in Eq. 5.

- Valve: [0, 6, 7]
- Mixer: [1, 8]
- Separator: [2, 9]
- Heat device: [3, 4, 5]
- Driver: [10]
- Goal: [11]

4.3. *Multi-relational graph design*

The goal of the Minox system is to decrease the oxygen content realized by the separation function and the effectiveness of the reaction. Based on the MFM knowledge and the physical layout of the

Table 2. Component of the system and related sensors

Component	ID	Related sensor	Type (units)
Separator 1	2	LC01	Level (%)
		PC01	Pressure (kPa)
		TT06	Temperature (°C)
Heat exchanger	4	TT06, TT03, TT07, TT09	Temperature (°C)
Dioxidizer	4	TT01, TT03, TT07, TT02	Temperature (°C)
Blower	10	Pp2	Power (%)
Separator 2	9	LC02	Level (%)
		PC07	Pressure (kPa)
		FT01	Flow rate (m^3/h)
		OT01	Oxygen content (ppb)

components, the relations represented in the adjacency matrix are listed in terms of mass, energy, control, and means-end edges guided by MFM, as shown in Fig. 4.

- Energy edge: (5, 4), (4, 3), (3, 4), (10, 1)
- Mass edge: Liquid: (0, 1), (1, 2), (2, 7), (7, 8); Gas: (2, 3), (3, 4), (4, 3), (3, 8), (8, 9), (9, 10), (10, 1), (1, 2), (6, 8)
- Control edge: (2, 0), (9, 7), (4, 6)
- Means-end edge: Function-goal edge: (4, 11), (5, 11), (2, 11), (9, 11); Component-function edge: (5, 4), (10, 4), (0, 2), (7, 9), (6, 4)

According to the relations listed above, the multi-relational adjacency matrix $\mathcal{A} \in \mathbb{R}^{5\times12\times10}$ is defined as:

$$\mathcal{A} = \{\mathbf{A}_{liq}, \mathbf{A}_{gas}, \mathbf{A}_{eng}, \mathbf{A}_{ctrl}, \mathbf{A}_{mfm}\}$$

where the definition of each submatrix $\mathbf{A}_r$ follows:

$$(\mathbf{A}_r)_{i,j} = \begin{cases} 1, & \text{if edge } (i,j) \text{ exists in relation } r \\ 0, & \text{otherwise} \end{cases}$$

The submatrix A_{gas} is shown as follow:

$$\mathrm{A}_{gas} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & 0 & \cdots & 0 \end{pmatrix}_{12\times10}$$

$$\mathbf{X} = \begin{bmatrix} \text{Node} & \text{Sensors (Dim 0-3)} & \text{Type (Dim 4-9)} \\ \hline \text{LCV01} & [0,0,0,0] & [1,0,0,0,0,0] \\ \text{Mixer1} & [0,0,0,0] & [0,1,0,0,0,0] \\ \text{Sep1} & [LC_{01}, PC_{01}, TT_{06}, 0] & [0,0,1,0,0,0] \\ \text{HE} & [TT_{06}, TT_{03}, TT_{07}, TT_{09}] & [0,0,0,1,0,0] \\ \text{Diox} & [TT_{01}, TT_{03}, TT_{07}, TT_{02}] & [0,0,0,1,0,0] \\ \text{Heater} & [0,0,0,0] & [0,0,0,1,0,0] \\ \text{TCV} & [0,0,0,0] & [1,0,0,0,0,0] \\ \text{LCV02} & [0,0,0,0] & [1,0,0,0,0,0] \\ \text{Mixer2} & [0,0,0,0] & [0,1,0,0,0,0] \\ \text{Sep2} & [LC_{02}, PC_{07}, FT_{01}, OT_{01}] & [0,0,1,0,0,0] \\ \text{Blower} & [PP_2, 0,0,0] & [0,0,0,0,1,0] \\ \text{Goal} & [OT_{01}, 0,0,0] & [0,0,0,0,0,1] \end{bmatrix} \tag{5}$$

4.4. *Dataset generation*

The experimental dataset is generated from a semi-physical simulation of a typical industrial Minox process, comprising 12 key functional units such as the LCV01 valve, heater, blower, and separators. We simulated three representative fault scenarios: blower failure, heater failure, and LCV01 failure. A total of 1,200 samples were generated. To replicate real-world industrial environments, Gaussian noise $\mathcal{N}(0, 0.05)$ was injected into all sensor readings to simulate measurement uncertainty. The three failures are injected based on the failure propagation paths identified by MFM to mimic the failure fingerprint.

4.5. *Experiment setup*

The proposed model employs a GCN architecture consisting of 3 layers to capture long-range fault propagation, which captures diverse physical dependencies (mass, energy, and control). To mitigate the over-smoothing problem in deep GNNs, a residual link is integrated to preserve the original sensor fingerprints from the input stage. The final classification is performed by a single-layer MLP with a Sigmoid activation, yielding a root-cause probability for each component.

4.6. *Results and discussion*

As shown in Table 3, the performance gain from Model A to Model B suggests that integrating

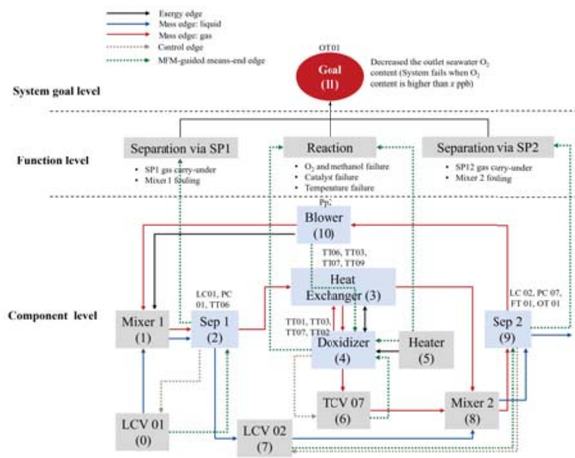

Fig. 4. GNN-MFM multi-relational graph of the Minox system

multi-domain layers (energy and control) helps resolve diagnostic ambiguities that mass-flow data alone cannot capture. The final improvement to 74.0% in Model C confirms that incorporating MFM-defined causality, specifically, Means-End logic, effectively guides the GNN from identifying mere statistical correlations toward interpretable, physics-based reasoning.

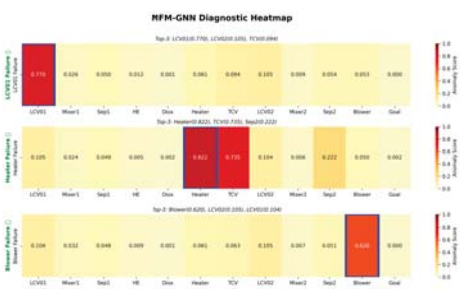

Fig. 5. Heat map for the multi-relational GNN architecture

While the MFM-GNN achieves higher overall performance, we observe confusion between TCV and Heater failures, as shown in Fig. 5. This is physically justified, as both faults manifest as temperature anomalies in the heat exchange zone, making them inherently coupled. The high score of the TCV (0.735) is physically consistent with MFM functional dependencies, as the TCV regulates the flow directly impacting the deoxidizer's performance. This demonstrates the model's ability to identify the causal chain rather than just isolated anomalies, validating the effectiveness of the knowledge-infused constraints. While finer-grained features or relations may be required to decouple these specific mechanisms to allow for a clear distinction, the 98.5% Top-3 accuracy holds significant practical value. In real-world operations, this allows maintenance teams to focus on only 3 candidate components instead of 12, effectively reducing the diagnostic search space by 75% for the vast majority of cases.

5. Conclusion

In conclusion, this study demonstrates the feasibility of the MFM-GNN framework in bridging the gap between data-driven graph networks and knowledge-based functional modeling. By embedding physical constraints and MFM causal logic into the message-passing mechanism, the model learns to identify anomalous propagation patterns rather than relying on static feature matching. This enables the framework to localize root causes that follow the underlying physical laws defined in the MFM graphs, even with limited historical data.

Although the high Top-3 accuracy confirms the

Table 3. Comparison of the experiment results of different model

Model	Graph Type	MFM-based constraint	Top-1 Acc	Top-3 Acc
A: Physical-GNN	$\mathbf{A}_{liq}$, $\mathbf{A}_{gas}$	×	64.0	93.5
B: Multi Physics-GNN	$\mathbf{A}_{liq}$, $\mathbf{A}_{gas}$, $\mathbf{A}_{ctrl}$, $\mathbf{A}_{eng}$	×	68.5	96.0
C: MFM-GNN (Proposed)	$\mathbf{A}_{liq}$, $\mathbf{A}_{gas}$, $\mathbf{A}_{ctrl}$, $\mathbf{A}_{eng}$	✓	74.0	98.5

model's ability to narrow down the diagnostic search space, the current Top-1 performance reveals challenges in definitively distinguishing between functionally coupled root causes, such as TCV versus heater failures, which exhibit nearly identical sensor signatures. To move beyond these ambiguities, future work will focus on enriching the model's knowledge base by incorporating more granular functional relationships and domain-specific approaches. This includes designing a refined architecture capable of capturing multimodal process data (like actuator feedback) to better isolate specific component signatures within functionally similar failure clusters.

Acknowledgement

The research was supported in part by the National Science Foundation under Grant No. 2045519. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the view of the National Science Foundation.

References

Chen, Z., J. Xu, C. Alippi, S. X. Ding, Y. Shardt, T. Peng, and C. Yang (2021). Graph neural network-based fault diagnosis: a review. arXiv:2111.08185 [eess].

Dash, T., S. Chitlangia, A. Ahuja, and A. Srinivasan (2022). A review of some techniques for inclusion of domain-knowledge into deep neural networks. *Scientific Reports 12*(1), 1040.

Gilmer, J., S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl (2017). Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. Pmlr.

Li, R., J. Wu, S. Mosleh, K. Groth, M. Modarres, M. Lind, O. Ravn, and X. Zhang (2025). Enhancing safety and operation of hydrogen fueling stations: A model-based method for complex failure scenario analysis. *Process Safety and Environmental Protection*, 107533.

Li, R., J. Wu, O. Ravn, and X. Zhang (2026). An integrated framework for functional model-based safety assessment of process systems using cloud-bayesian network. *Reliability Engineering & System Safety*, 112231.

Li, T., Z. Zhou, S. Li, C. Sun, R. Yan, and X. Chen (2022). The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. *Mechanical Systems and Signal Processing 168*, 108653.

Lind, M. (2011). An introduction to multilevel flow modeling. *Nuclear safety and simulation 2*(1), 22–32.

Lind, M. and X. Zhang (2014). Functional modelling for fault diagnosis and its application for NPP. *Nuclear Engineering and Technology 46*(6), 753–772.

Lv, H., J. Chen, T. Pan, T. Zhang, Y. Feng, and S. Liu (2022). Attention mechanism in intelligent fault diagnosis of machinery: A review of technique and application. *Measurement 199*, 111594.

Moradi, R. and K. M. Groth (2020). Modernizing risk assessment: A systematic integration of PRA and PHM techniques. *Reliability Engineering & System Safety 204*, 107194.

Nguyen, B. and M. Chioua (2024). A propagation path-based interpretable neural network model for fault detection and diagnosis in chemical process systems. *Control Engineering Practice 150*, 105988.

Ouyang, W. and Y. Jin (2024). Fault diagnosis of process systems based on graph neural network. In *Proceedings of the 1st International Conference on Data Mining, E-Learning, and Information Systems - DMEIS*, pp. 37–45. INSTICC: SciTePress.

Qiu, S., X. Cui, Z. Ping, N. Shan, Z. Li, X. Bao, and X. Xu (2023). Deep Learning Techniques in Intelligent Fault Diagnosis and Prognosis for Industrial Systems: A Review. *Sensors 23*(3), 1305.

Ruiz-Tagle Palazuelos, A., E. Lopez Droguett, and K. Groth (2020). A system-level prognostics and health management framework based on graph convolutional networks. In *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference (ESREL2020/PSAM15)*.

Song, M., C. Reinartz, X. Zhang, H. P.-J. Thunem, and R. McDonald (2023). MFM-based alarm root-cause analysis and ranking for nuclear power plants. *Nuclear Engineering and Technology 55*(12), 4408–4425.

Taqvi, S. A. A., H. Zabiri, L. D. Tufa, F. Uddin, S. A. Fatima, and A. S. Maulud (2021). A Review on Data-Driven Learning Approaches for Fault Detection and Diagnosis in Chemical Processes. *ChemBioEng Reviews 8*(3), 239–259.

Yu, J. and Y. Zhang (2023). Challenges and opportunities of deep learning-based process fault detection and diagnosis: a review. *Neural Computing and Applications 35*(1), 211–252.

Zhou, J., G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun (2020). Graph neural networks: A review of methods and applications. *AI open 1*, 57–81.