Expert Systems With Applications 306 (2026) 130962

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Environment-Aware graph relational reasoning for interpretable and generalizable mechanical transmission system distributed fault diagnosis

Chao Zhao [a,b,*], Weiming Shen [c], Enrico Zio [d,e], Hui Ma [a,b,*]

[a] School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China
[b] Key Laboratory of Vibration and Control of Aero-Propulsion System Ministry of Education, Northeastern University, Shenyang 110819, China
[c] Key Laboratory of Future Intelligent Manufacturing Technologies for High-end Equipment (Fuyao University of Science and Technology), Ministry of Education, China
[d] MINES Paris PSL University, CRC, Sophia Antipolis, France
[e] Energy Department, Politecnico di Milano, Milan, Italy

ARTICLE INFO

Keywords:
Mechanical transmission system
Domain generalization
Graph neural network
Distributed fault diagnosis
Multiple sensors

ABSTRACT

In recent years, numerous fault diagnosis models have been developed to monitor the health status of mechanical transmission systems under dynamic environments. However, these models generally perform point-to-point monitoring of individual components, focusing on a local perspective while overlooking the coupling relationships among multiple components. Faults originating in a single component can propagate to adjacent components through vibration transmission, which may ultimately lead to misdiagnosis or misinterpretation. To address this issue, this paper proposes an environment-aware graph relational reasoning framework based on a discover-evaluate-refine paradigm, aiming to achieve comprehensive system-level health monitoring of mechanical transmission systems. The framework constructs stable relational subgraphs by identifying the significance of sensors in the diagnostic decision process and capturing collaborative signal variations among sensors. Samples from different source domains are then used to perturb each other, simulating variations in working conditions and evaluating the robustness of subgraph structures. This assessment provides feedback that guides the refinement of subgraph discovery, ensuring the model effectively captures environment-invariant correlations. Extensive experiments conducted on a self-built experimental platform, a high-speed train, and a metro train bogie demonstrate the superiority of the proposed method. Visualization of the relational subgraphs provides interpretability support for the diagnostic results of the model. Our code and dataset are publicly available at: https://github.com/CHAOZHAO-1/EAGRR.

1. Introduction

Mechanical transmission systems are widely used in critical equipment such as high-speed trains, industrial conveyor systems, and wind turbine generator units (Chen & Jiang, 2020). These complex systems operate under harsh working conditions and are therefore prone to failure. When a failure occurs, it can result in significant economic losses and even casualties (Fan et al., 2025). Therefore, comprehensive monitoring is essential to maintain the health of mechanical transmission systems (Wang et al., 2025; Xiao et al., 2025; Yu et al., 2025).

With the rapid advancement of computing power and sensor networks, data-driven approaches have become a key tool for monitoring the health status of mechanical transmission systems (Lei et al., 2020). Generally, these systems comprise several critical components, such as shafts, gearboxes, pulleys, and bearings, each of which plays an indispensable role in overall system reliability (Shang et al., 2024). Therefore, various methods have been proposed to monitor these critical components. For example, Wang et al. (Jinhai Wang et al., 2022) proposed a time–frequency representation-based convolutional neural network for diagnosing gearbox of railway vehicle. Zhang et al. (Zhang et al., 2023) developed an information stream fusion approach for intelligent fault diagnosis of offshore wind turbine bearings. Liu et al. (Liu et al., 2021) utilized statistical methods with sound and thermal infrared signals for idler fault diagnosis. Algburi et al. (Algburi et al., 2025) combined a hierarchical hyper-Laplacian prior and singular spectrum analysis for industrial robot diagnosis.

However, most methods typically focus on individual components in isolation, constructing independent diagnostic models that result in

* Corresponding authors.
E-mail addresses: zhaoc@me.neu.edu.cn (C. Zhao), wshen@ieee.org (W. Shen), enrico.zio@polimi.it (E. Zio), huima@me.neu.edu.cn (H. Ma).

https://doi.org/10.1016/j.eswa.2025.130962
Received 25 September 2025; Received in revised form 22 December 2025; Accepted 23 December 2025
Available online 24 December 2025

fragmented analyses and fail to capture system-level fault propagation mechanisms (Fan et al., 2023; Wang et al., 2022). In practical mechanical transmission systems, faults often propagate along the power transmission chain, where an anomaly in one component can trigger abnormal behavior in adjacent components. To address this problem, it is essential to consider the transmission system as an integrated whole and implement distributed monitoring through multiple sensors. For example, Wang et al. (Jinxin Wang et al., 2022) constructed a bond graph of dynamical model and two multivariate statistic measures for diesel engine lubrication system fault detection. Yang et al. (Yang et al., 2024) analyzed data from vibration acceleration sensor, hall induction coil and the motor encoder for evaluating the state of harmonic drive. Further discussion about system-level fault diagnosis can be found in the related work section.

The above-discussed methods have considered the component relationships from a global perspective but still exhibit poor diagnostic performance under dynamic environments (Raouf et al., 2024; Wu et al., 2024). The highly variable working conditions in industrial site pose significant challenges for diagnostic models to adapt to unseen environments (Zio, 2022). Most existing methods are trained on limited and specific data distributions and thus struggle to cope with environmental shifts, which limits their generalization capacity under dynamic conditions. Fortunately, advanced domain generalization techniques can improve cross-environment generalization without requiring access to target domain data (Zhao et al., 2024). Moreover, most existing models suffer from limited interpretability, making it difficult to provide clear and trustworthy decision support for frontline maintenance engineers, which severely restricts practical deployment in industrial applications.

To overcome the aforementioned two drawbacks, this paper proposes an ***E***nvironment-***A***ware ***G***raph ***R***elational ***R***easoning (EAGRR) framework for interpretable and generalizable mechanical transmission system distributed fault diagnosis. The key motivation of methodology design is that collaborative signal variations among sensors are discriminative for diagnosis tasks, and such collaborative relationships remain stable under environmental changes. To effectively capture the collaborative relationships among sensors, the subgraph discoverer constructs relational structures, while the representation learner evaluates their stability across varying environments. The evaluation outcomes are subsequently fed back to refine the subgraph structure learning, establishing a synergistic paradigm that unites relational reasoning with environment-aware validation.

The main contributions of this study, encompassing the research scenario, dataset, and methodology, are summarized as follows:

(1) **A research scenario**. This study introduces the concept of system-level panoramic perception under dynamic environments for mechanical transmission systems, leveraging spatially distributed multi-sensor parallel observations. The goal is to monitor the overall health of multiple interacting components under varying operating conditions, overcoming the limitations of traditional methods, which typically focus on individual components under single operating conditions. This scenario aligns with the practical requirements of industrial transmission systems.
(2) **A novel dataset**. To explore the above scenario, a experimental platform for mechanical transmission systems was designed and constructed. Time-series data from multiple sensors under diverse operating conditions were collected to form a new benchmark dataset, which will be made publicly available to facilitate further research in this field.
(3) **A novel methodology**. To tackle the challenge of system-level perception, an environment-aware graph relational reasoning framework is proposed. Comprehensive experiments conducted on three datasets demonstrate the effectiveness of the proposed method.

The remainder of this paper is organized as follows. Section 2 reviews some related work. Section 3 details the proposed method. Section 4 presents experimental results and analyses. Finally, Section 5 draws some conclusions and discusses our future work.

2. Related work

2.1. Multiple-Sensor-based intelligent fault diagnosis

Multi-sensor arrangements improve mechanical system observability by covering multiple subsystems, spatial locations, and measurement dimensions. Accordingly, researchers have developed various advanced data fusion methods, which can be broadly categorized into homogeneous sensor-based and heterogeneous sensor-based approaches, depending on the nature and compatibility of the input data sources.

By deploying a single-type sensor at multiple locations, it is possible to capture spatially distributed responses of a mechanical system. Zhang et al. (Zhang et al., 2025) used multi-sensor vibration signals collected from an axial flow pump as input for a multi-scale deep feature memory and recovery network to address channel missing scenarios. Wang et al. (Wang et al., 2025) focused on the system temporal evolution of subway train bogies and developed a selective spatio-temporal graph neural network.

Multi-modal signals from heterogeneous sensors provide complementary information that enhances fault characterization, especially for complex or subtle failure modes. Peng et al. (Peng et al., 2024) constructed a multimodal knowledge graph based on bearing vibration signals and text description. Sun et al. (Sun et al., 2023) extracted structural and non-structural damage information of the gearbox from infrared thermal images and acoustic data. Similarly, Zhang et al. (Zhang et al., 2024) developed an innovative approach that combines vibration signals with infrared images to achieve accurate gearbox fault diagnosis. Mao et al. (Mao et al., 2025) proposed a federated relation self-perception graph network for aero-engine rotor, considering issues of heterogeneous data islands and multi-sensor information fusion.

2.2. Graph-based System-Level mechanical fault diagnosis

Previous studies have primarily focused on monitoring the health conditions of individual components, typically treating each component in isolation. However, multiple components often function collaboratively, and their interactions and interdependencies are critical to the reliability of the overall system. Therefore, it is crucial to consider the collaborative dynamics between components to develop more robust and holistic monitoring approaches. Graph-based representations offer a powerful means to model such complex dependencies and to enable the effective fusion of multi-sensor data. For train bogie, Ding et al. (Ding et al., 2024) proposed a component spatial relationship-based graph neural diagnostic network, while Yan et al. (Yan et al., 2025) proposed a framework that integrates physical knowledge and statistical learning. Xiao et al. (Xiao et al., 2025) designed multilayer graphs by analyzing correlation between different sensors of pumped storage units.

2.3. Domain generalization-based intelligent fault diagnosis

Generalizability is crucial for intelligent diagnostic models in dynamic environments (Zhao et al., 2024). To address domain shift problem, Zhao and Shen (Zhao & Shen, 2022) explored domain invariance and retained domain specificity simultaneously. Furthermore, He and Shen (He & Shen, 2024) developed a federated domain generalization method for machine-level motor fault diagnosis. Zhu et al. (Zhu et al., 2024) utilized fault causality and physical prior knowledge to enhance performance of the diagnostic models. Qian et al. (Qian et al., 2025) proposed a new DG-Softmax loss for cross-machine fault diagnosis.

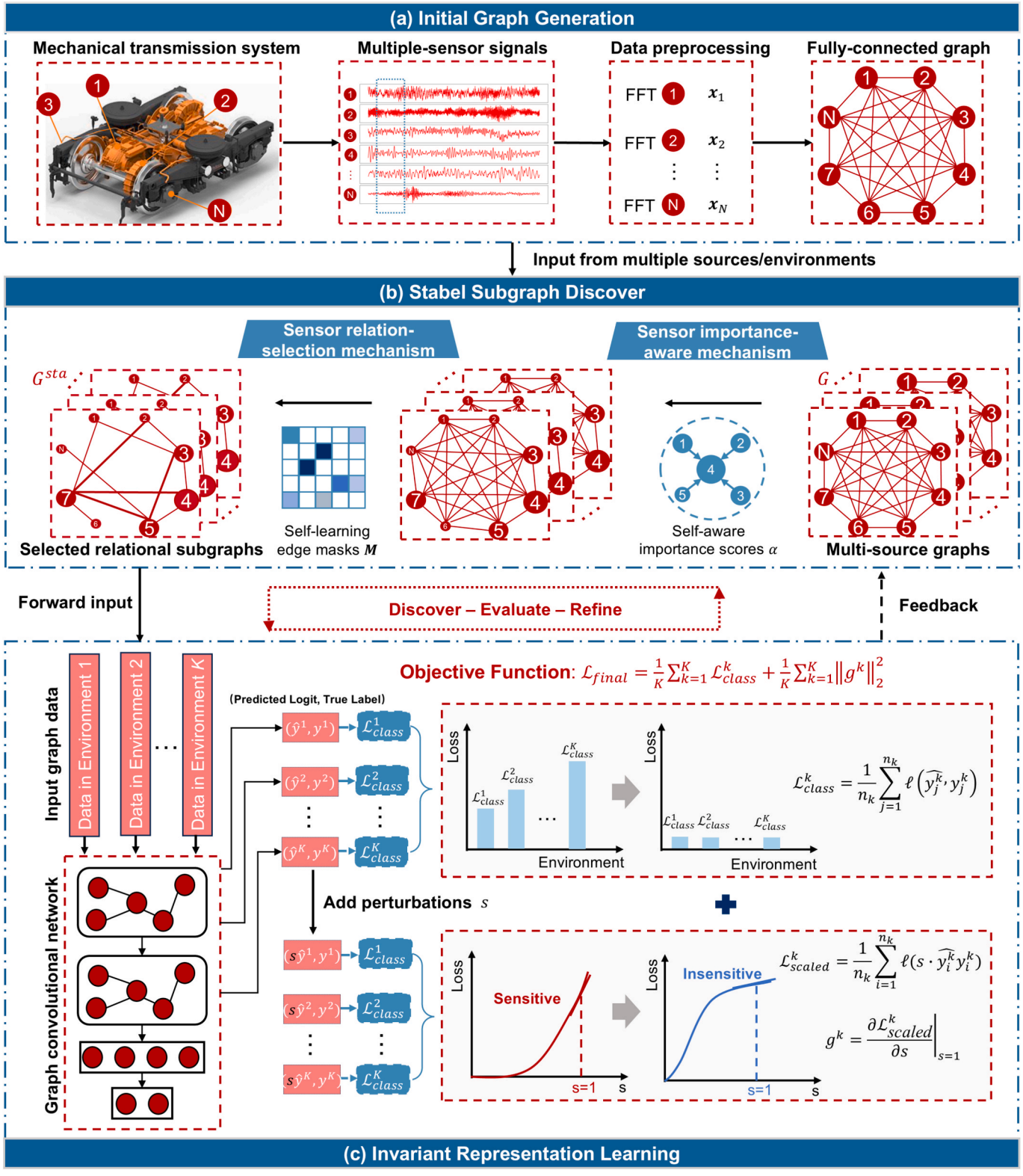

Fig. 1. Overview of the proposed EAGRR.

3. Proposed method

Mechanical transmission systems generally operate under dynamic environments, resulting in significant differences in the signal distributions measured by different sensors across these conditions. Traditional data-driven methods tend to capture spurious correlations specific to certain conditions. Moreover, the complex physical couplings between transmission components make it challenging to model them a priori or to precisely reflect the true sensor dependencies. To address these two problems, an environment-aware graph relational reasoning framework (see Fig. 1) is proposed. Table 1 lists the details of the proposed architecture.

Table 1
Network architecture of the proposed EAGRR.

Module name		Layer name	Parameters	Operation	Activation
Stable subgraph discoverer $S(\bullet)$	Importance aware	Linear @1	Input: 1024 → Output: 1024	/	/
		Self attention @1	Input: 256 → Output: N, 4 heads	Softmax normalization	LeakyReLU
	Relation selection	Linear @2	Input: 2048 → Output: 256	/	ReLU
		Linear @3	Input: 256 → Output: 1	Top-k edge selection	/
Invariant representation learner $R(\bullet)$		Graph conv @1	Input: 1024 → Output: 512	Batch normalization	ReLU
		Graph conv @2	Input: 512 → Output: 128	Batch normalization	ReLU
		Graph pooling @1	Size: 128	Mean pooling over nodes	/
		Linear @4	Input: 128 → Output: class number	/	/

This method automatically discovers stable subgraph structures among different sensors to transform the implicit correlations between multi-sensor measurements into explicit graph topologies, while simultaneously learning environment-invariant fault representations, thereby mitigating condition-specific spurious correlations and enhancing diagnostic robustness under unseen operating conditions.

3.1. Problem definition

The diagnostic definition of multiple sensor-based mechanical transmission system is given. Let $\boldsymbol{X}$ denote an input space and $\mathscr{Y}$ denote a label space. During the training stage, given K source training domains $\mathscr{D}_{train} = \left\{\mathscr{D}^k \middle| k = 1, \cdots, K\right\}$, where $\mathscr{D}^k = \left\{\boldsymbol{x}_j^k, y_j^k\right\}_{j=1}^{n_k}$ denotes the k-th domain with n_k data samples. Each data sample $\boldsymbol{x}_j^k = \left\{\left(\boldsymbol{x}_j^k\right)_n \middle| n = 1, \cdots, N\right\} \in \boldsymbol{X}$ is comprised of N different subsamples collected by different sensors, and $y_j^k \in \mathscr{Y}$ denotes the corresponding health state label. The distributions among multiple domains are different: $P\left(\boldsymbol{X}^1\right) \neq P\left(\boldsymbol{X}^2\right) \neq \cdots \neq P\left(\boldsymbol{X}^K\right)$. The goal is to construct a generalized multi-sensor feature extractor $\boldsymbol{z} = F(\boldsymbol{x})$ and a robust state classifier $y = C(\boldsymbol{z})$. The constructed model can minimize classification risk $\mathbb{E}_{(\boldsymbol{x},y)\ \mathscr{D}_{test}}[C(F(\boldsymbol{x})) \neq y]$ on an unseen target testing domain $\mathscr{D}_{test}$, where data distributions differ from those of the training domains $\left\{P\left(\boldsymbol{X}^{test}\right) \neq P\left(\boldsymbol{X}^{train}\right)\right\}$. Note, the main mathematical symbols used throughout the paper are summarized in **Appendix A1**.

3.2. Interpretable stable subgraph discovery

In mechanical transmission systems, component couplings vary dynamically due to complex power transmission paths and structural interactions. Graph-based models provide an effective way to capture such nonlinear, time-varying dependencies in multi-sensor data. However, predefined adjacency structures may neglect critical relationships or embed incorrect assumptions, as sensor couplings are often unknown. To address this issue, a sensor importance-aware mechanism and a sensor relation-selection mechanism are developed to achieve interpretable stable subgraph discovery, enabling the autonomous learning of subgraphs that can adapt to diverse systems and can generalize across different working conditions.

Interpretable stable subgraph discovery proceeds in three steps: initial graph construction, sensor importance assessment, and sensor relation selection. During the initial graph construction phase, fully connected graphs are generated based on synchronized multi-sensor time series to provide a comprehensive dependency space for the subsequent learning processes. Each node in the graph corresponds to a sensor, with its node features derived from the signals acquired by that sensor. Let G denote a fully connected graph, defined as follows,

$$G = (\boldsymbol{V}, \boldsymbol{A}) \tag{1}$$

where $\boldsymbol{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_N\}$ is the set of nodes corresponding to N sensors, and $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix. The feature node $\boldsymbol{v}_n$ is the time-series signal collected by sensor n. Given a raw multi-sensor time-series sample $\boldsymbol{x}_j^k = \left\{\boldsymbol{x}_{j,1}^k, \boldsymbol{x}_{j,2}^k, \cdots, \boldsymbol{x}_{j,N}^k\right\}$, it is mapped to an initial graph:

$$G_j^k = \left(A_{ini}, \boldsymbol{x}_j^k\right) = \left((1)^{N \times N}, \left(\boldsymbol{x}_{j,n}^k\right)_{n=1}^{N}\right) \tag{2}$$

where $\boldsymbol{A}_{ini} = 1^{N \times N}$ denotes a fully connected adjacency matrix initialized with ones, and $\boldsymbol{x}_{j,n}^k$ is the input signal of sensor n in the j-th sample of domain k.

It is intuitively reasonable that the importance of different sensors in diagnosis decision-making varies across distinct system health states. For example, when a fault occurs in the motor, sensors located in close proximity to the motor naturally become more informative and critical for accurate fault diagnosis. This fault-scenario-dependent variability in sensor significance necessitates that the model be capable of adaptively adjusting the contribution weights of sensor inputs, thereby accurately capturing their relevance under the current health condition. To this end, a sensor importance-aware mechanism is proposed to enable such dynamic adjustment, which in turn facilitates more precise and robust health state identification.

The importance coefficients quantify the direct influence strength of neighboring nodes on the state of target node in the current task, thereby capturing the dynamic correlations among different nodes. Specifically, the initial node feature $\boldsymbol{x}_j^k$ is linearly projected into a new feature space as follows:

$$\boldsymbol{h}_j^k = W\boldsymbol{x}_j^k \tag{3}$$

where W is a learnable parameter matrix, and $\boldsymbol{h}_j^k$ is the transformed node feature. Next, the importance score between a target node n and one of its neighbor nodes m is computed. This process begins by concatenating their transformed features:

$$\boldsymbol{h}_{j,nm}^k = \left[\boldsymbol{h}_{j,n}^k \middle\| \boldsymbol{h}_{j,m}^k\right] \tag{4}$$

where $\|$ denotes feature concatenation. The concatenated feature $\boldsymbol{h}_{j,nm}^k$ is then passed through a learnable importance vector $\boldsymbol{a}$, followed by a LeakyReLU activation function σ, to produce the unnormalized importance score $\boldsymbol{o}_{j,nm}^k$,

$$\boldsymbol{o}_{j,nm}^k = \sigma\left(\boldsymbol{a}^\top \boldsymbol{h}_{j,nm}^k\right) \tag{5}$$

After obtaining $\boldsymbol{o}_{j,nm}^k$, the importance scores for all neighboring nodes $m' \in \mathscr{N}(n)$ are normalized using the softmax function to obtain the importance coefficients $\alpha_{j,nm}^k$,

$$\alpha_{j,nm}^k = \frac{e^{\boldsymbol{o}_{j,nm}^k}}{\sum_{m' \in \mathscr{N}(n)} e^{\boldsymbol{o}_{j,nm'}^k}} \tag{6}$$

where $\mathscr{N}(n)$ denotes the set of neighbor nodes of node n. The initial graph is fully connected, where every node is initially connected to all other nodes, so the neighbors of a node n include all other nodes in the graph except n itself. $\mathscr{N}(n) = \{m \in \{1, 2, ..., N\} | m \neq n\}$. In the case of multi-head importance with T independent importance heads, the model performs the above steps T times in parallel. Each head computes

its own set of importance coefficients and transformed neighbor features. The outputs of all heads are concatenated to form the importance-aware representation of node n:

$$\boldsymbol{h}_{j,n}^{k,imp} = \|_{t=1}^{T} \sum_{m \in \mathcal{N}(n)} \alpha_{j,nm}^{k,t} \boldsymbol{h}_{j,m}^{k,t} \tag{7}$$

Sensor importance-aware mechanism automatically identifies which sensors contribute most to the recognition of the current abnormal signals. This mechanism lays the foundation for subsequent relation subgraph selection, as the selected edges must be based on such dynamic interactions rather than isolated node features.

Due to strong dependencies between sensors may actually arise from co-variations under the same operating conditions. Part of the statistical correlations reflects the essential characteristics of the system state, while others are merely by-products of the working condition. The model is still susceptible to interference from such condition-induced correlations, which weakens its generalization capability across different working scenarios. Therefore, it is necessary to further filter out stable relationships from coupled correlations.

Sensor relation-selection mechanism is designed not only to enhance the robustness of the graph structure but also improves the interpretability of the relation subgraph, clarifying which sensor interactions play a critical role in fault detection. Specifically, for the j-th sample, the edge feature between node n and node m is constructed by concatenating their node embeddings:

$$\boldsymbol{e}_{j,nm}^{k} = \left[\boldsymbol{h}_{j,n}^{k} \middle\| \boldsymbol{h}_{j,m}^{k}\right] \tag{8}$$

where $\boldsymbol{h}_{j,n}^{k}$ is the embedding of n-th node in the j-th sample from k-th source domain. A shared linear layer is applied to each edge feature to compute the edge score:

$$\boldsymbol{s}_{j,nm}^{k} = f\left(\boldsymbol{e}_{j,nm}^{k}\right) \tag{9}$$

where $\boldsymbol{s}_{j,nm}^{k}$ denotes the scalar attention score for the edge connecting node n and node m in the j-th sample from domain k. $f(\bullet)$ denotes a learnable transformation implemented as a two-layer multilayer perceptron in our implementation. During training, the parameters of $f(\bullet)$ are optimized jointly with the rest of the model using the back-propagation algorithm. These scores are then assembled into a symmetric edge score matrix:

$$\boldsymbol{S}_{j}^{k} = \left[\boldsymbol{s}_{j,nm}^{k}\right] \in \mathscr{R}^{N \times N} \tag{10}$$

Score-ranking based edge selection extracts the top-q most informative edges. A mask ratio $\rho \in (0,1)$ is defined. The number of selected edges is determined as follows:

$$q = \lfloor \rho \bullet N^2 \rfloor \tag{11}$$

Let $\boldsymbol{I}_{j}^{k} = \text{top\_}q\text{\_indices}\left(\boldsymbol{S}_{j}^{k}, q\right)$ denotes the indices of the top-q edges. Then, a binary edge selection mask $\boldsymbol{M}_{j}^{k} \in \{0,1\}^{N \times N}$ is constructed as,

$$\left(\boldsymbol{M}_{j}^{k}\right)_{nm} = \begin{cases} 1, if\ i(n,m) \in \boldsymbol{I}_{j}^{k} \\ 0, if\ i(n,m) \notin \boldsymbol{I}_{j}^{k} \end{cases} \tag{12}$$

where $i(n,m)$ denotes the index of the edge connecting the n-th and m-th nodes within the set of all node pairs. This mask retains the most important q edges, which are assumed to represent candidate stable interactions among the sensors. The resulting stable adjacency matrix is then defined as:

$$\boldsymbol{A}_{j}^{k,sta} = \boldsymbol{M}_{j}^{k} \tag{13}$$

Accordingly, the refined subgraph learned by stable subgraph discoverer $S(\cdot)$ is given by:

$$G_{j}^{k,sta} = S\left(G_{j}^{k}\right) = \left(\boldsymbol{A}_{j}^{k,sta}, \boldsymbol{h}_{j,n}^{k,imp}\right) \tag{14}$$

3.3. Environment-Invariant representation learning

Considering that different working conditions introduce variations in environment variables, it is essential to ensure that the diagnostic model possesses generalization ability across environments. Therefore, environment-invariant representation learning is developed to learn invariant features that remain robust despite changes in working conditions.

An invariant representation learner $R(\bullet)$ is applied to extract high-level representations $\boldsymbol{z}_{j}^{k}$ from subgraph $G_{j}^{k,sta}$. In this study, $R(\bullet)$ is implemented as a two-layer graph convolutional network as shown in Table 1. This step performs feature extraction over the selected graph structure, generating fused node embeddings under the sensor relations for downstream fault state recognition.

$$\boldsymbol{z}_{j}^{k} = R\left(G_{j}^{k,sta}\right) \tag{15}$$

In the diagnosis representation learning process, each working condition is treated as a distinct environment k. Within each environment, the diagnostic model must first achieve accurate fault state recognition. Therefore, for each environment k, the classification loss is computed individually:

$$\mathscr{L}_{class}^{k} = \frac{1}{n_k} \sum_{j=1}^{n_k} \ell\left(\widehat{y}_{j}^{k}, y_{j}^{k}\right) \tag{16}$$

where n_k denotes the number of samples in environment k, $\widehat{y}_{j}^{k}$ is the model output logits, y_{j}^{k} is the corresponding ground truth label, and $\ell(\bullet)$ represents the cross-entropy loss function.

However, optimizing solely for environment-specific correlation may lead the diagnostic model to overfit the statistical patterns of particular working conditions, rather than capturing the inherently stable relationships underlying faults. Motivated by previous research (Cha et al., 2021), which demonstrates that solutions with less sensitive loss landscapes tend to generalize better, a perturbation-based strategy is applied during training to simulate variations in working conditions. This approach ensures that the learned fault representations remain predictive under distribution shifts.

The central concept involves the incorporation of a perturbation-sensitivity measure. A controlled perturbation is applied to the model predictions, and the derivative of the loss with respect to the perturbation magnitude is calculated to quantify the vulnerability of the model to variations in environmental conditions. This metric is subsequently employed as a regularization term to promote the acquisition of more robust representations. Specifically, a scaling factor s is applied to the prediction logits to assess the sensitivity of the model loss under such perturbations. For each environment k, the perturbed loss is computed as:

$$\mathscr{L}_{scaled}^{k} = \frac{1}{n_k} \sum_{i=1}^{n_k} \ell\left(s \bullet \widehat{y}_{i}^{k}, y_{i}^{k}\right) \tag{17}$$

Next, the gradient of the perturbed loss with respect to the scaling factor s is computed at $s = 1$, yielding the environment-specific sensitivity:

$$g^{k} = \left.\frac{\partial \mathscr{L}_{\text{scaled}}^{k}}{\partial s}\right|_{s=1} \tag{18}$$

The squared ℓ_2-norm of this gradient is used to quantify the sensitivity score:

$$U^{k} = \|g^{k}\|_{2}^{2} \tag{19}$$

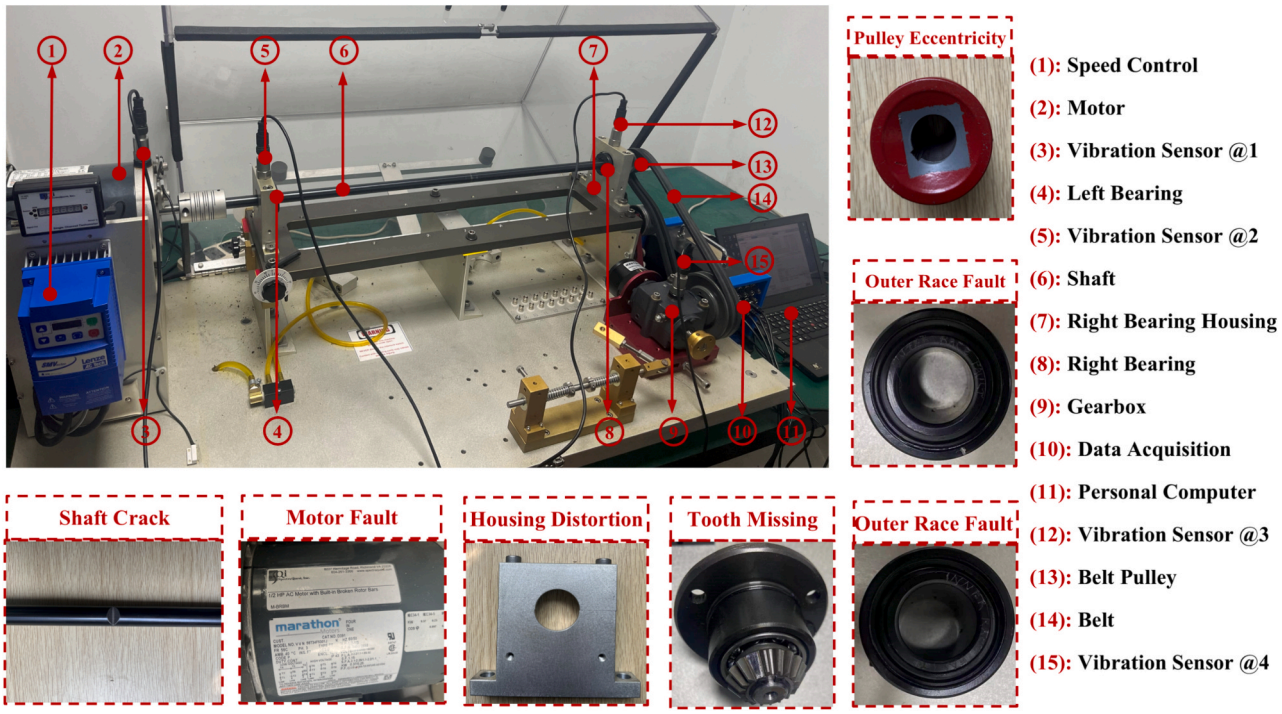

Fig. 2. HUST transmission system test bench and associated fault components.

If the model has successfully learned truly environment-invariant features, the sensitivity scores U^k should remain small and stable across different environments, indicating that the model predictions are insensitive to global scaling perturbations. Conversely, if the model overfits to environment-specific noise, the sensitivity score in that environment will be significantly larger, reflecting the model reliance on spurious correlations.

The overall training objective consists of two components: (1) the average empirical risk $\mathscr{L}_{\text{class}}$ across all environments to ensure predictive performance, and (2) a sensitivity regularization term U to reduce reliance on environment-specific correlations. Formally, they are expressed as:

$$\mathscr{L}_{class} = \frac{1}{K}\sum_{k=1}^{K}\mathscr{L}_{class}^{k} \tag{20}$$

$$U = \frac{1}{K}\sum_{k=1}^{K} U^k \tag{21}$$

By jointly optimizing these objectives, the model is encouraged to minimize overfitting to spurious environment-specific patterns and accurately recognize transmission system states across diverse working conditions.

3.4. Environment-Aware graph relational reasoning

In the proposed EAGRR framework, interpretable stable subgraph discovery and environment-invariant representation learning operate in a synergistic paradigm (discover-evaluate-refine). The subgraph discovery module autonomously identifies critical sensor interactions, constructing relational structures that capture stable dependencies in multi-sensor signals while filtering out environment-specific correlations. The representation learning module evaluates the robustness of subgraph structures under varying operating conditions via perturbation-based sensitivity analysis. This assessment quantifies the reliance of model predictions on environment-specific correlations, providing feedback that guides the refinement of edge selection and node aggregation in the subgraph discovery process. This closed-loop interaction enables relational reasoning and environment-aware validation to mutually reinforce each other.

As a result, EAGRR integrates structural interpretability with environment-adaptive learning, yielding meaningful sensor relationships and fault representations that are robust across different working conditions. EAGRR mitigates overfitting to spurious correlations and ensures reliable distinction of fault states in complex transmission systems.

Algorithm 1: EAGRR.

 # Training stage

Input: Multi-source multi-sensor dataset $\left\{\{\boldsymbol{X}_n^k\}_{n=1}^{N}\right\}_{k=1}^{K}$.

Model: A subgraph discoverer $S(\bullet)=\{\boldsymbol{W},\boldsymbol{a},f(\bullet)\}$ and a representation learner $R(\bullet)$

1: for *epoch* = 1 to *epochs* do

2: Randomly sample training source data from $\left\{\{\boldsymbol{X}_n^k\}_{n=1}^{N}\right\}_{k=1}^{K}$.

3: Generate initial graphs based on **Eq. (2)**.

4: Discover stable subgraphs via **Eq. (14)**.

5: Learn robust representations by **Eq. (20)** and **Eq. (21)**.

6: Forward propagation to calculate the total loss using **Eq. (22)**.

7: Backward propagation to update subgraph discoverer $S(\bullet)=\{\boldsymbol{W},\boldsymbol{a},f(\bullet)\}$ and representation learner $R(\bullet)$ using Adam optimizer using **Eq. (23)**.

8: end for

Return: The optimal subgraph discoverer $S(\bullet)$ and representation learner $R(\bullet)$.

#Inferece stage

Input: Unknown target testing dataset.

Model: Trained EAGRR.

Output: Final diagnostic decision and interpretable insights.

The overall optimization objective of the proposed EAGRR is formulated as follows:

$$\mathscr{L}_{final} = \mathscr{L}_{class} + \lambda U \tag{22}$$

where λ is a tradeoff parameter. Network parameters are updated in an end-to-end manner using the Adam optimizer, with the optimization procedure at each training epoch formulated as follows:

Table 2
Task descriptions. Tasks 1–10 correspond to the HUST Transmission System Datasets. Tasks 11–15 are based on the BJTU-RAO Bogie Datasets. Tasks 16–18 involve the High-Speed Train Bogie Fault Datasets. The rest means all other operating conditions.

Task index	Source domains (Speed, Hz)	Target domain (Speed, Hz)	Task index	Source domains	Target domain
Task 1	20, 30, 40, 50, 60, 70	0–70-0	Task 11	40 Hz/0kN, 40 Hz/+10 kN, 40 Hz/-10 kN	The rest
Task 2	20, 30, 40, 50, 70	0–70-0	Task 12	60 Hz/0kN, 60 Hz/+10 kN, 60 Hz/-10 kN	The rest
Task 3	20, 40, 50, 70	0–70-0	Task 13	40 Hz/0kN, 20 Hz/+10 kN, 20 Hz/-10 kN	The rest
Task 4	20, 50, 70	0–70-0	Task 14	60 Hz/0kN, 40 Hz/+10 kN, 40 Hz/-10 kN	The rest
Task 5	20, 30, 40	50, 60, 70	Task 15	20 Hz/0kN, 60 Hz/0 kN, 60 Hz/-10 kN	The rest
Task 6	50, 60, 70	20, 30, 40	Task index	Source domains (Speed, Hz)	Target domain
Task 7	30, 50, 70	20, 40, 60			
Task 8	20, 40, 60	30, 50, 70	Task 16	30, 40	50
Task 9	20, 50, 60	30, 40, 70	Task 17	30, 50	40
Task 10	30, 40, 70	20, 50, 60	Task 18	40, 50	30

$$\begin{aligned} \boldsymbol{W}^{q+1} &\leftarrow \boldsymbol{W}^{q} - \mu \frac{\partial(\mathscr{L}_{class} + \lambda U)}{\partial \boldsymbol{W}^{q}} \\ \boldsymbol{a}^{q+1} &\leftarrow \boldsymbol{a}^{q} - \mu \frac{\partial(\mathscr{L}_{class} + \lambda U)}{\partial \boldsymbol{a}^{q}} \\ f(\bullet)^{q+1} &\leftarrow f(\bullet)^{q} - \mu \frac{\partial(\mathscr{L}_{class} + \lambda U)}{\partial f(\bullet)^{q}} \\ R(\bullet)^{q+1} &\leftarrow R(\bullet)^{q} - \mu \frac{\partial(\mathscr{L}_{class} + \lambda U)}{\partial R(\bullet)^{q}} \end{aligned} \tag{23}$$

where μ denotes learning rate, and q represents q-th epoch update. In summary, EAGRR can be reported as: **Algorithm 1**.

4. Case study

In this section, comprehensive experiments on one self-collected mechanical transmission dataset and two open-source train bogie datasets are conducted to validate the effectiveness of the proposed method.

4.1. Dataset description and experimental setting

(1) HUST Transmission system Dataset. This dataset was collected from a machinery fault simulator, equipped with four vibration sensors, as illustrated in Fig. 2. The transmission system operates on six constant motor speeds (20 Hz, 30 Hz, 40 Hz, 50 Hz, 60 Hz, and 70 Hz) and one time-varying motor speed (0–70-0 Hz). In totally, there are 14 health states, including one heathy state, seven single fault, and six compound faults. The sample rate is 25.6 kHz. The power transmission chain consists of the motor, shaft, left bearing, right bearing, pulley, belt, and gearbox. Further details of this dataset can be found in **Appendix A2**.

(2) BJTU-RAO Bogie Dataset (Ding et al., 2024). This dataset was collected from a scaled-down experimental platform of a real subway train bogie, educed and simplified at a 1:2 scale from the original design. It includes six vibration sensors and covers nine working conditions (motor speed/transverse load): 20 Hz/0kN; 40 Hz/0kN; 60 Hz/0kN; 20 Hz/+10kN; 40 Hz/+10kN; 60HZ/+10kN; 20 Hz/-10kN; 40 Hz/-10kN; 60 Hz/-10kN. The sampling rate is 64 kHz, and there are eight health states in total. Further details of this dataset can be found in **Appendix A3**.

(3) High-speed Train Bogie Fault Dataset (Li et al., 2025). Similarly, this dataset was collected from a bogie system based on a 1:2 scaled design of an actual high-speed train. The platform primarily consists of a bogie frame, traction motors, axle boxes, gearboxes, rail axles, wheels, and loading devices. It includes three working conditions defined by different combinations of vertical load, transverse load, and motor speed: 1300 kg/0 kg/30 Hz, 1300 kg/0 kg/40 Hz, and 1300 kg/0 kg/50 Hz. There are nine health states in total and five vibration sensors. The sampling rate is 25.6 kHz. Further details of this dataset can be found in **Appendix A4**.

Table 3
Details of comparison methods. ✓ means the module is equipped, while × means the module is not equipped.

Method	Description
CCDG	A domain generalization-based fault diagnosis method using multi-channel CNN and contrastive loss.
CNN-C	A domain generalization-based method for fault diagnosis utilizing multi-channel CNN and center loss.
GSAT	A graph domain generalization method employing fully connected graphs and explicit attention mechanism.
LECI	A graph domain generalization method using multi-channel CNN and label-environment causal independence.

Method	Sensor relation-selection mechanism	Sensor importance-aware mechanism	Environment-invariant representation learning
G1	×	×	×
G2	✓	×	×
G3	✓	✓	×
EAGRR	✓	✓	✓

For all three datasets, the sample length is set to 2048, and each class contains 300 samples. In total, there are 18 fault diagnosis tasks, as summarized in Table 2.

4.2. Experimental settings and comparison methods

To verify the superiority of the proposed method, four state-of-the-art domain generalization methods including convolution neural

Table 4
Diagnosis results (%) of HUST Transmission System diagnosis tasks.

Methods	CCDG	CNN-C	GSAT	LECI	Proposed
Task 1	77.55 ± 0.92	82.45 ± 1.44	81.95 ± 1.09	81.76 ± 1.69	**86.14 ± 0.25**
Task 2	77.83 ± 0.83	81.74 ± 0.17	80.64 ± 1.83	79.60 ± 1.18	**85.50 ± 1.47**
Task 3	76.50 ± 0.75	80.67 ± 2.14	79.10 ± 2.39	81.33 ± 0.66	**85.79 ± 1.19**
Task 4	74.19 ± 1.32	79.26 ± 0.23	76.76 ± 1.16	80.31 ± 1.95	**87.52 ± 1.67**
Task 5	71.10 ± 0.12	77.60 ± 2.95	71.26 ± 1.48	76.10 ± 0.37	**94.95 ± 1.59**
Task 6	50.67 ± 1.1	59.52 ± 2.47	37.69 ± 1.59	57.02 ± 0.56	**76.60 ± 1.97**
Task 7	86.93 ± 2.8	89.98 ± 1.7	81.17 ± 0.56	85.86 ± 0.44	**94.36 ± 0.30**
Task 8	87.69 ± 1.53	93.88 ± 0.68	90.90 ± 2.41	96.48 ± 0.84	**99.38 ± 0.22**
Task 9	88.71 ± 1.79	92.38 ± 1.62	87.52 ± 1.18	92.93 ± 3.09	**96.83 ± 1.50**
Task 10	86.60 ± 1.45	95.83 ± 0.23	81.38 ± 0.90	87.98 ± 2.74	**96.69 ± 1.75**
Average	77.78	83.33	76.84	81.94	**90.38**

Table 5
Diagnosis results (%) of BJTU-RAO Bogie diagnosis tasks.

Methods	CCDG	CNN-C	GSAT	LECI	Proposed
Task 11	97.69 ± 0.70	**98.84 ± 0.54**	90.57 ± 1.99	96.10 ± 2.80	95.88 ± 0.52
Task 12	83.58 ± 1.86	80.56 ± 1.82	87.99 ± 1.66	83.58 ± 1.98	**88.72 ± 0.22**
Task 13	**98.90 ± 0.50**	96.80 ± 0.78	93.34 ± 1.24	98.65 ± 1.39	96.72 ± 0.72
Task 14	**98.35 ± 0.23**	98.10 ± 0.02	93.09 ± 1.38	97.40 ± 1.33	95.13 ± 0.98
Task 15	89.78 ± 0.84	93.48 ± 0.78	93.43 ± 1.11	94.84 ± 1.53	**95.33 ± 0.25**
Average	93.66	93.56	91.68	94.11	**94.35**

Table 6
Diagnosis results (%) of High-speed Train Bogie diagnosis tasks.

Methods	CCDG	CNN-C	GSAT	LECI	Proposed
Task 16	78.48 ± 2.43	**83.67 ± 1.87**	65.56 ± 1.85	68.81 ± 1.24	78.55 ± 0.57
Task 17	92.70 ± 2.88	95.74 ± 0.58	86.70 ± 1.67	98.70 ± 0.12	**99.51 ± 1.08**
Task 18	59.00 ± 3.09	62.15 ± 0.26	72.63 ± 2.26	84.74 ± 0.79	**85.60 ± 0.15**
Average	76.73	80.52	74.96	84.09	**87.89**

network (CNN) based (CCDG(Ragab et al., 2022) and CNN-C(Y. Yang et al., 2020)) and graph neural network based (GSAT(Miao et al., 2022), LECI(Gui et al., 2023)) are selected for performance comparison. In addition, a rigorously ablation study (G1, G2, and G3) is conducted to evaluate the effectiveness of the proposed module. The detail of comparison methods is listed in Table 3.

The experiments are conducted using the PyTorch framework with an NVIDIA 4070 GPU for accelerated computing. The average accuracy and standard deviation are calculated over ten repeated trials. The batch size is set to 256, and the models are trained for 100 epochs. For the proposed method, the hyperparameters are set as $\lambda = 1$ and $\rho = 0.6$. The hyperparameter settings for the comparison methods are adopted from existing literature or selected based on their optimal performance.

4.3. Results Analyses

(1) Main Results

The diagnostic results of the HUST Transmission System diagnosis tasks are listed in Table 4. It can be seen that the proposed method achieves the highest accuracy across all 10 tasks, outperforming the best comparison method by an average accuracy margin of 6.32 %. This demonstrates the superiority of the proposed method in mechanical transmission system monitoring. CNN-C and CCDG exhibit relatively weak generalization performance. Although feature space constraints improve generalization to some extent, multi-channel CNNs struggle to effectively learn the interdependencies between sensors located at different positions, limiting their ability to capture complex spatial relationships. GSAT models attention only on edges within the graph, neglecting the importance of nodes. This incomplete modeling of graph structures hinders its generalization performance. LECI perform well on several tasks and exhibits certain generalization capability. However, its performance is still inferior to the proposed method, indicating that the proposed method offers stronger robustness across varying operating conditions.

The diagnostic results for the BJTU-RAO bogie diagnosis tasks are summarized in Table 5. It can be observed that the proposed method achieves the highest average accuracy, further validating its generalizability. Due to the relative simplicity of these five tasks, all five methods demonstrate comparably good performance. In addition, the diagnostic results for the high-speed train bogie diagnosis tasks are also presented in Table 6. It is evident that the proposed method surpasses the best comparison method by 3.80 % in accuracy. These two train bogie-related diagnostic tasks collectively demonstrate the practical feasibility and effectiveness of the proposed method.

To provide a more intuitive illustration of the uncertainty in the performance of the proposed method, Fig. 3 presents the 95 % confidence intervals. As shown in Fig. 3, the confidence intervals of the proposed method are generally narrow, which demonstrates the stability and reliability of the method.

In addition, a statistical significance test was conducted to assess whether the proposed method achieves a significant improvement over the other methods. A paired t-test with a $\alpha = 0.05$ significance level was carried out, assuming the null hypothesis $H_0 : Acc_A = Acc_B$, where A and B denote the proposed and the compared methods, respectively. The results of the paired t-test on the three datasets are summarized in Table 7. As all P-values are below the significance threshold of 0.05, the null hypothesis is rejected, indicating a statistically significant difference between the proposed and compared methods. Specifically, since the proposed method achieves higher accuracy than the compared methods, it can be concluded that the proposed method significantly outperforms the compared methods.

(2) Ablation Study

A rigorous ablation study was conducted by progressively removing individual modules from the proposed method, with the diagnostic results presented in Fig. 4. It is evident that as the modules are gradually eliminated, the model accuracy consistently decreases across diagnostic tasks in three different scenarios. This observation confirms the positive contribution of each module to the model generalization performance. A more detailed analysis reveals that, although the sensor importance-aware meacnism and environment-invariant representation learning contribute less significantly to performance improvement compared to

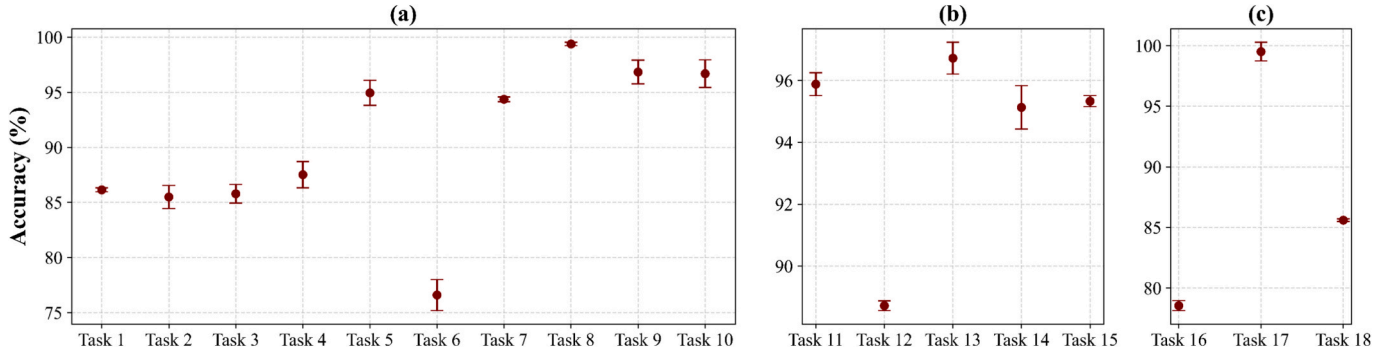

Fig. 3. Performance of the proposed method with 95% confidence intervals.

Table 7
Results of the paired *t*-test between the proposed method and compared methods.

Method	HUST Transmission System		BJTU-RAO Bogie		High-speed Train Bogie	
	T-value	P-value	T-value	P-value	T-value	P-value
CCDG	19.671	5.612×10^{-36}	2.140	4.092×10^{-2}	5.272	1.191×10^{-5}
CNN-C	12.562	3.281×10^{-22}	2.092	4.529×10^{-2}	3.420	1.885×10^{-3}
GSAT	13.197	1.518×10^{-23}	8.469	3.702×10^{-11}	41.178	2.788×10^{-27}
LECT	14.006	3.216×10^{-25}	2.068	4.772×10^{-2}	4.433	1.224×10^{-4}

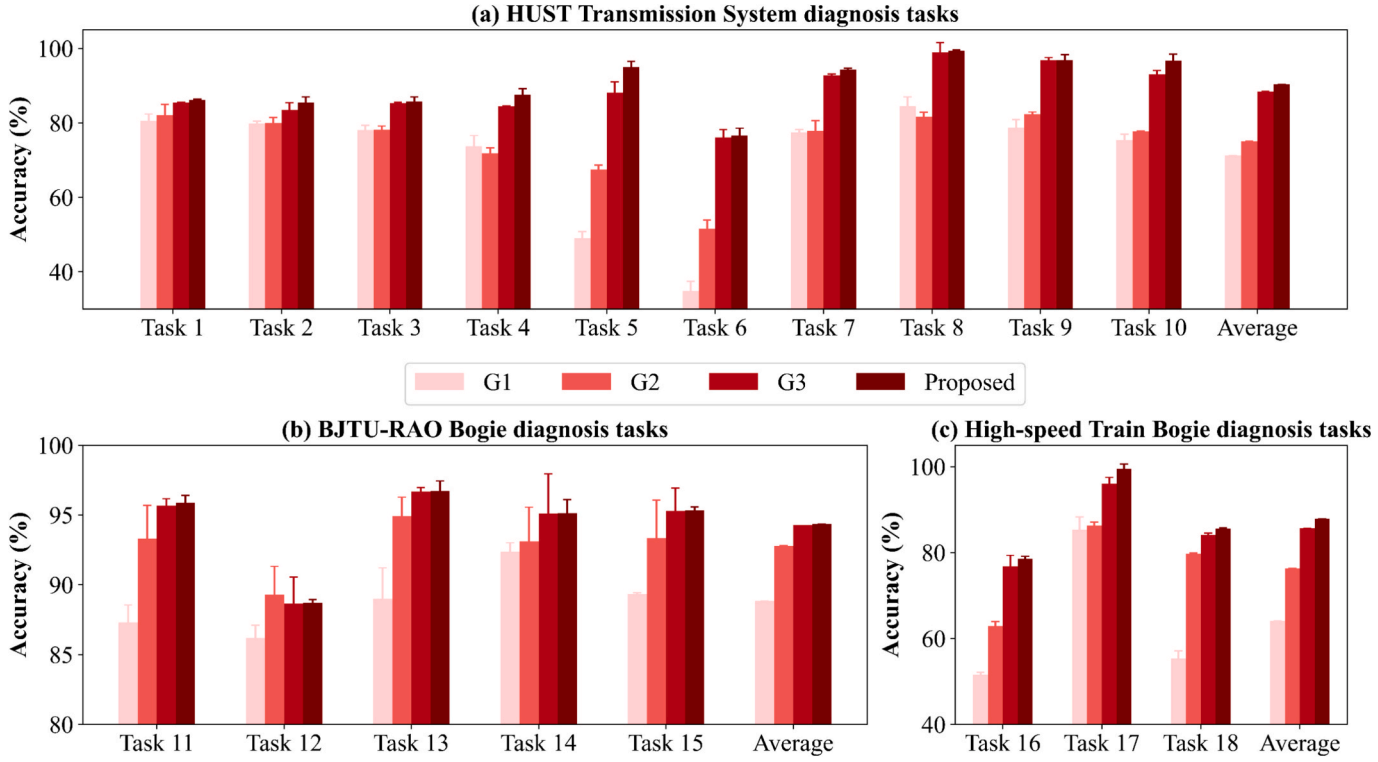

Fig. 4. Diagnosis results for ablation study.

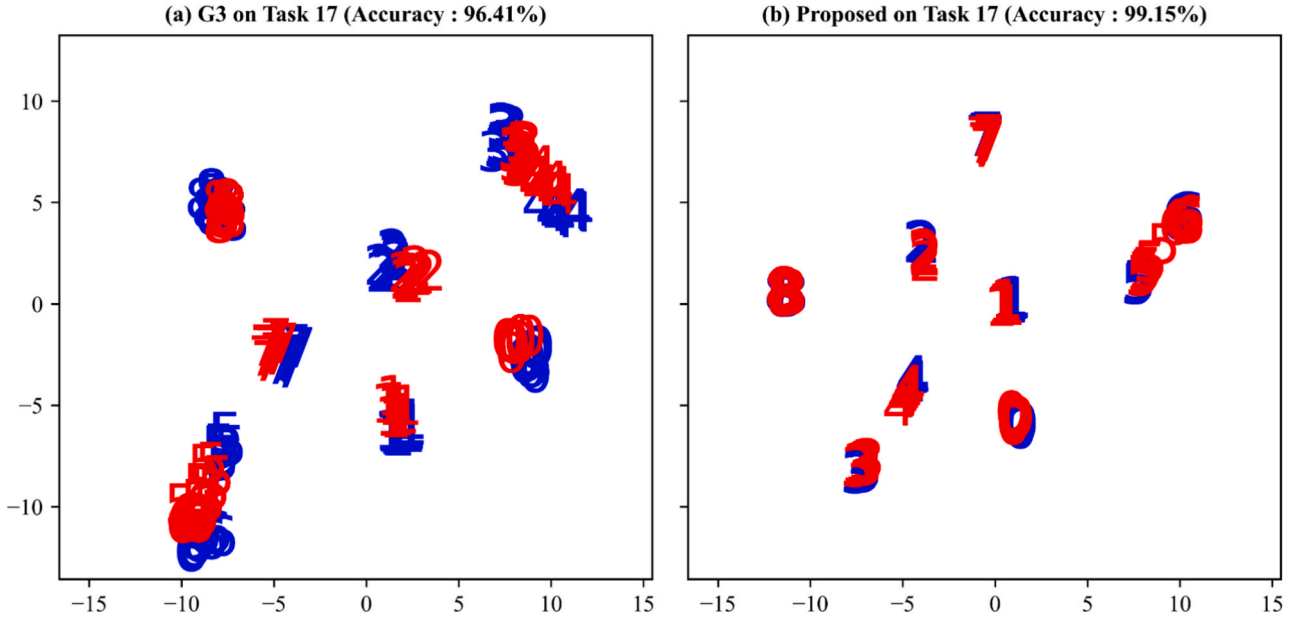

Fig. 5. Visualization of learned features via t-SNE on task 17.

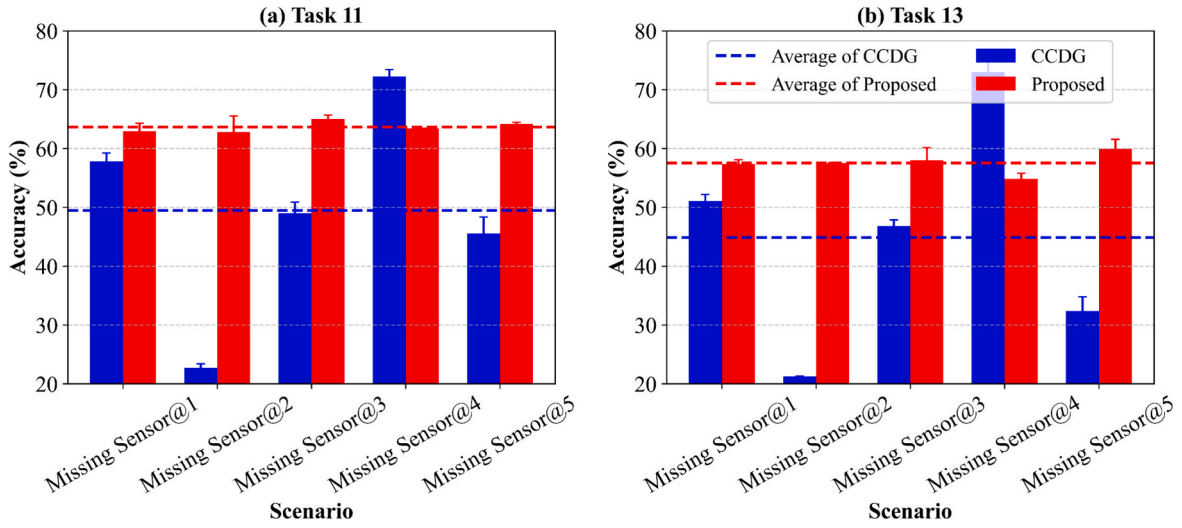

Fig. 6. Diagnosis results of two methods facing sensor signal missing scenarios on Tasks 11 and 13.

the sensor-relation selection meacnism, they still provide measurable enhancements. These findings indicate that filtering stable relational graphs to uncover dynamic relationships is crucial for improving diagnostic capability in mechanical transmission systems.

To intuitively demonstrate the effectiveness of the environment-invariant representation learning, t-SNE is employed to visualize the high-dimensional features learned by two methods (G3 and EAGRR). As shown in Fig. 5 (a), the features learned by G3 exhibit partial alignment between the source and target domains, and some classes show decent clustering. However, certain overlap remains among different classes, indicating limited class separability. In contrast, Fig. 5 (b) illustrates the results after incorporating the environment-invariant representation learning module. The class boundaries become much clearer, and samples from the same class but different domains are better aligned. These observations confirm the effectiveness of environment-invariant representation learning in enhancing cross-environment generalization and in combating unknown distribution shifts.

(3) Stable Performance Analyses

The real-time monitoring of health status in mechanical transmission systems is commonly achieved through the deployment of numerous sensors. As equipment operates over long periods, some sensors inevitably suffer from failures or signal loss. Therefore, ensuring the robustness of diagnostic models under sensor anomalies is of critical importance. In this subsection, models are trained using complete multi-sensor data, while during the testing phase, the input from each sensor is sequentially set to zero. This setup is designed to simulate real-world scenarios of sensor malfunction or data dropout. CCDG, a representative method based on a CNN architecture, is selected for comparative performance evaluation.

As shown in Fig. 6, experimental results on Task 11 and Task 13 demonstrate that under sensor anomaly conditions, the proposed method significantly outperforms CCDG, achieving an average accuracy improvement of over 10 %. Moreover, as summarized in Table 5, CCDG exhibits slightly better performance than the proposed method when the sensor data are complete and undisturbed. However, when the critical second sensor fails, the performance of CCDG degrades dramatically, indicating a lack of robustness and stability. This phenomenon is attributed to the inability of CNN-based models to capture structural relationships across sensor channels. When a specific sensor provides abnormal or zeroed signals, CNNs still treat them as valid inputs and often leads to the generation of misleading representations, ultimately degrading the discriminative capability of the diagnostic model. In contrast, the proposed method, which leverages a graph-based architecture, maintains stable accuracy across different sensor failure scenarios. Each node not only relies on its own features but also aggregates equipment information from neighboring nodes via the graph structure. Even when a node is zeroed out, it can still receive information from others, thus mitigating the impact of missing or corrupted data and preventing catastrophic performance degradation.

(4) Interpretability Fault Analyses

To intuitively demonstrate the interpretability of the stable relational subgraph for mechanical system fault localization, Fig. 7 visualizes the subgraphs extracted by the proposed method from Tasks 8, 11, and 17.

Fig. 7(a) illustrates the subgraph when the mechanical system is under a motor fault condition. It can be observed that the node corresponding to Sensor @1, which is directly connected to the motor, exhibits the highest importance score, while the other three nodes show significantly lower intensity. Edge analysis indicates that, under motor failure, the model tends to direct information flow from surrounding nodes toward the motor node, highlighting an information-gathering behavior. Fig. 7(b) depicts the case of a compound fault, comprising motor fault, left bearing fault, and gear fault. The importance scores of Sensors @1, @2, and @4 are significantly enhanced, whereas Sensor @3 shows a relatively weak node intensity.

Fig. 7(c) also corresponds to a motor fault scenario. In this case, both Sensor @1 and Sensor @2, which are installed on the motor, show prominently high node strengths, reinforcing the physically consistent importance mechanism of the model. Fig. 6(d) reflects a situation involving simultaneous motor and gear faults, where all four nodes (Sensor @1, @2, @3, and @4) demonstrate high importance scores.

Fig. 7(e) corresponds to a bearing fault in the axle box. Here, Sensor @1 exhibits the highest node strength, and edge patterns suggest that most of the information is funneled toward this node, highlighting its critical role under this failure mode. Fig. 7(f) visualizes the condition of a gear fault. Since the gearbox is most closely associated with Sensor @3 and Sensor @4, these nodes exhibit stronger importance intensities accordingly.

In summary, the stable subgraph structure provides a valuable and structured interpretability framework for the diagnostic model. Different fault types exhibit distinct node importance distributions and edge connectivity patterns within the graph. Additionally, the contribution of sensor signals varies depending on the fault location, and the cooperative behavior among nodes constitutes key information flow

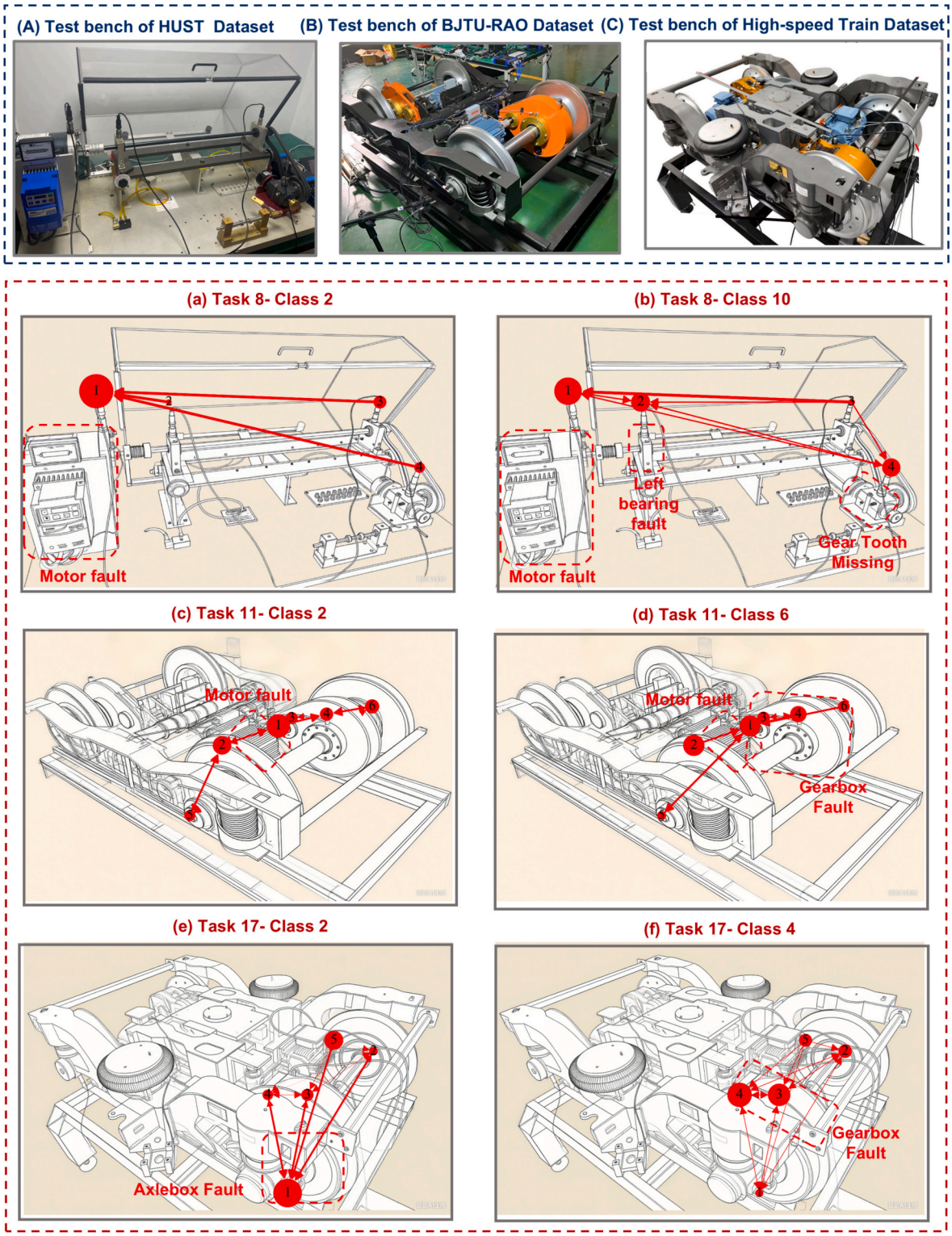

Fig. 7. Stable subgraph-interpretable visualization. (a) Task 8. (b) Task 11. (c) Task 17. Node size reflects the importance of corresponding node, while edge thickness indicates the strength of inter-node associations.

pathways that underpin specific fault classifications.

(5) Subgraph Consistency Analyses

To evaluate the consistency of subgraph structures across domains for the same class, both quantitative and qualitative analyses were conducted. Specifically, cosine similarity was employed to assess Node Importance Consistency (NIC), while Frobenius distance was used to evaluate Edge Structure Consistency (ESC). NIC and ESC for class c are defined as:

$$NIC(c) = \frac{1}{K}\sum_{k=1}^{K} \frac{\overline{\alpha}_c^k}{\|\overline{\alpha}_c^k\|} \bullet \frac{\overline{\alpha}_c^t}{\|\overline{\alpha}_c^t\|}$$

Table 8
Subgraph consistency cross domains in Task 8. Note that values of node importance consistency and edge structure consistency closer to 1 suggest stronger structural alignment.

Class	Node Importance Consistency (NIC)	Edge Structure Consistency (ESC)
1	0.9872	0.9766
2	0.9911	0.9749
3	0.9337	0.9751
4	0.9815	0.9733
5	0.9866	0.9772
6	0.9807	0.9456
7	0.9884	0.9707
8	0.9876	0.9795
9	0.9980	0.9782
10	0.9926	0.9598
11	0.9840	0.9125
12	0.9926	0.9601
13	0.9876	0.9610
14	0.9898	0.9775

$$ESC(c) = 1 - \frac{1}{K}\sum_{k=1}^{K} \|\overline{\boldsymbol{A}}_c^k - \overline{\boldsymbol{A}}_c^t\|_F \tag{23}$$

where $\overline{\boldsymbol{\alpha}}_c^k$ denotes the average node importance vector of class c in the k-th source domain, and $\overline{\boldsymbol{\alpha}}_c^t$ represents the corresponding average node importance in the target domain. $\overline{\boldsymbol{A}}_c^k$ denotes the average edge adjacency matrix for class c in the k-th source domain, and. $\overline{\boldsymbol{A}}_c^t$ is the corresponding matrix in the target domain.

Table 8 reports the NIC and ESC of the proposed method for different fault classes in Task 8. As shown in the Table 8, the NIC for most classes is approximately 0.98, while the ESC for the majority of classes hovers around 0.97. This high level of consistency confirms that the proposed method is capable of learning stable and domain-invariant fault relational subgraphs under varying operating conditions.

Furthermore, Fig. 8 presents the visualizations of subgraphs corresponding to Classes 2, 3, and 5 across different domains in Task 8. It can be observed that the subgraph structures for the same fault category exhibit remarkable consistency across domains, further validating the effectiveness of the proposed method.

(6) Hyperparameter Sensitivity Analyses

The two key hyperparameters in the proposed method, the edge retention ratio ρ and the regularization coefficient λ, were analyzed. Fig. 9 shows the diagnostic accuracy under different combinations of ρ and λ. The model performance improves significantly as ρ increases from 0.1 to the range of 0.4–0.6, indicating that appropriately retaining more edges when constructing a robust relational graph can lead to more robust representation learning. However, when ρ exceeds 0.6, accuracy declines due to the inclusion of too many task-irrelevant edges. When λ is small, the regularization term has little impact on learning, and the model tends to focus on empirical risk minimization. As λ increases, the model gradually enforces stronger constraints for environment invariance, which helps improve generalization within a certain range. However, when λ becomes too large, the dominance of the regularization term in the loss suppresses the ability of model to recognize health

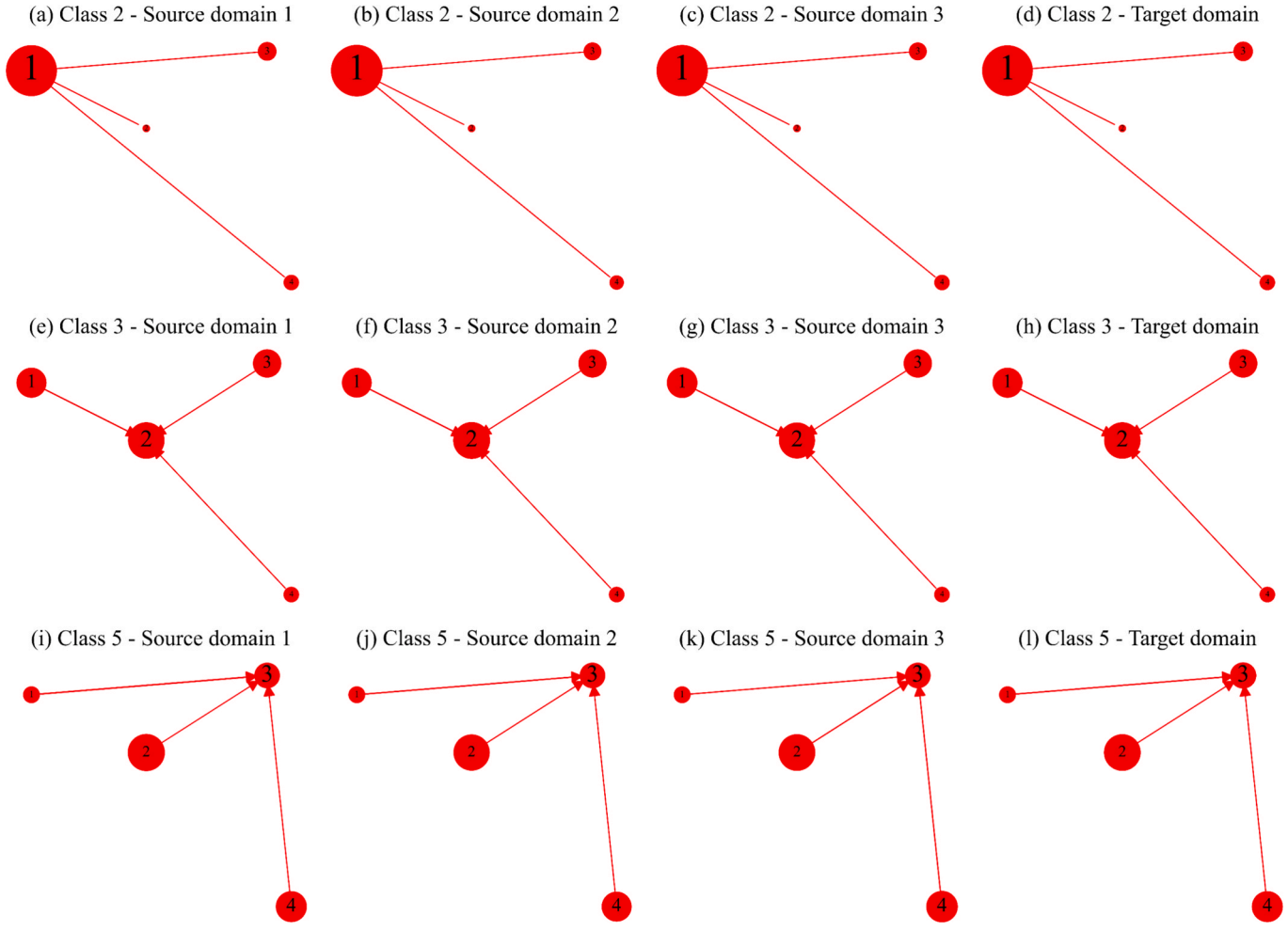

Fig. 8. Stable subgraph visualization on Task 8 with Classes 2, 3, 5.

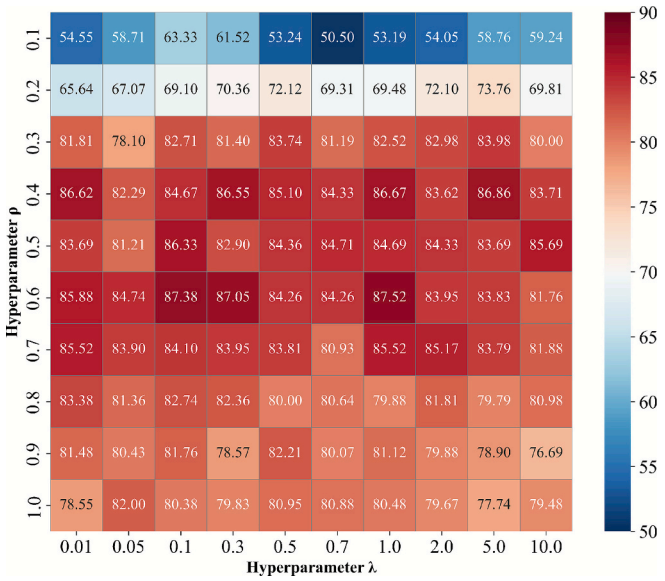

Fig. 9. Diagnosis results (%) of the grid search on the validation Task 4.

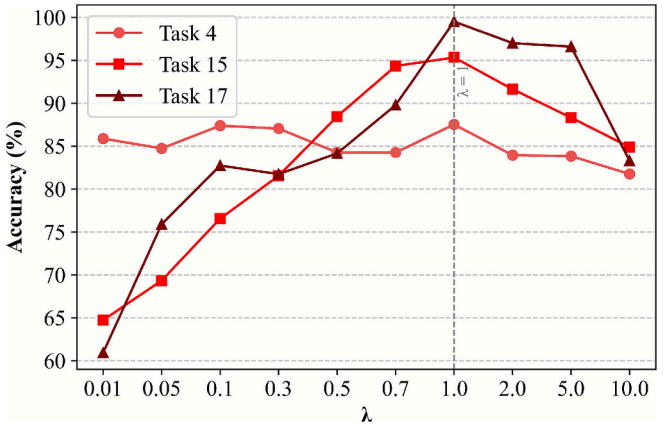

Fig. 10. Sensitivity analysis of λ.

states, leading to a performance drop. In experiment, the model achieved the highest accuracy of 87.52 % when $\rho = 0.6$ and $\lambda = 1$.

To verify the generalizability of $\lambda = 1$ across different tasks, Tasks 4, 15, and 17 were randomly selected from three cases. Fig. 10 presents the performance of the proposed method under different values of λ. It can be seen that the method performs well when λ ranges from 0.7 to 2, with the best performance observed at $\lambda = 1$.

(7) Model Performance Analysis

Fig. 11 illustrates the convergence performance of the proposed method on three different tasks. It can be observed that both the average empirical risk and the sensitivity regularization term gradually stabilize and converge as the number of training epochs increases.

Table 9 presents the number of parameters for each method, along with their corresponding training and testing times. Note that the testing time refers to the time required to make predictions for all test samples. Due to the simple architecture of the convolutional networks, CCDG and CNN-C have relatively few parameters, and consequently their training and testing times are short. In contrast, the three graph-based methods (GSAT, LECI, and the proposed method) require a larger number of parameters to be trained, resulting in longer training and testing times. Since these methods follow an offline training and online testing paradigm, the additional computational cost is considered acceptable. Notably, the testing times of the proposed method are 0.45 s, 1.79 s, and 0.31 s for the respective tasks, demonstrating relatively good real-time performance.

5. Conclusion

This study proposes an environment-aware graph relational reasoning framework for health monitoring of mechanical transmission systems. The method leverages parallel observations from spatially distributed multi-sensor data and offers strong interpretability and generalization capabilities. In the proposed discover-evaluate-refine paradigm, stable relational subgraphs are mined based on both the contribution of each sensor to the diagnostic decision-making process and the collaborative signal variations among sensors. Subsequently, samples from different source domains are employed to mutually perturb one another, simulating variations in operating conditions and thereby evaluating the robustness of the subgraph structures. Finally, sensitivity regularization is further applied to assist in refining the robust fault subgraph structures. Extensive quantitative experiments on three complex mechanical transmission systems validate the effectiveness of the proposed method. Rigorous ablation studies further demonstrate the contribution of each module. Additionally, a visual analysis of the stable relational subgraphs provides interpretability support for the model outputs. However, this study only considers single-modal signals in system monitoring. In the future, multi-modal heterogeneous signals such as acoustic signals, thermal imaging, and others will be investigated. Moreover, for complex mechanical systems, exploring a minimal yet reliable sensor deployment strategy is highly meaningful for enhancing monitoring system reliability and reducing costs. This direction is also planned to be investigated in future work.

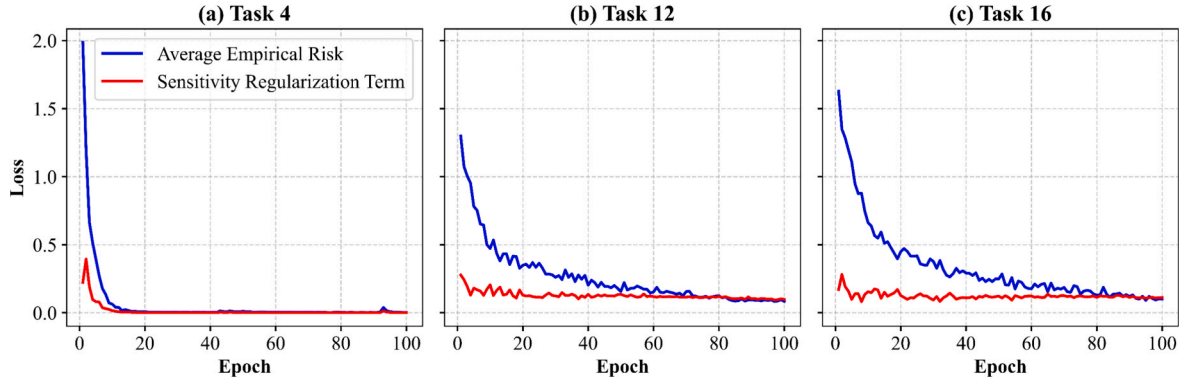

Fig. 11. Convergence performance of the proposed method.

Table 9
Number of model parameters and computation time of different methods.

Method	HUST Transmission System			BJTU-RAO Bogie			High-speed Train Bogie		
	Parameter number	Training time (s)	Testing time (s)	Parameter number	Training time (s)	Testing time (s)	Parameter number	Training time (s)	Testing time (s)
CCDG	47,134	24.57	0.04	45,592	36.93	0.15	46,873	32.74	0.03
CNN-C	47,134	109.17	0.04	45,592	182.10	0.16	46,873	192.98	0.03
GSAT	13,250,319	71.23	0.15	13,248,777	107.61	0.72	13,249,034	88.31	0.11
LECT	129,139	767.70	3.21	127,684	775.18	11.02	127,573	2882.24	1.96
Proposed	2,168,975	507.3	0.45	2,168,201	517.17	1.79	2,168,330	509.84	0.31

CRediT authorship contribution statement

Chao Zhao: Conceptualization, Methodology, Data curation, Writing – original draft. **Weiming Shen:** Supervision, Funding acquisition. **Enrico Zio:** Supervision, Writing – review & editing. **Hui Ma:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2025.130962.

Data availability

Data will be made available on request.

References

Algburi, R. N. A., Aljibori, H. S. S., Al-Huda, Z., Gu, Y. H., & Al-antari, M. A. (2025). Advanced fault diagnosis in industrial robots through hierarchical hyper-laplacian priors and singular spectrum analysis. *Complex and Intelligent Systems, 11*(6), 1–26. https://doi.org/10.1007/s40747-025-01915-8

Cha, J., Chun, S., Lee, K., Cho, H.-C., Park, S., Lee, Y., & Park, S. (2021). *SWAD: Domain Generalization by Seeking Flat Minima*. 1–22. http://arxiv.org/abs/2102.08604.

Chen, H., & Jiang, B. (2020). A review of fault detection and diagnosis for the traction system in high-speed trains. *IEEE Transactions on Intelligent Transportation Systems, 21*(2), 450–465. https://doi.org/10.1109/TITS.2019.2897583

Ding, A., Qin, Y., Wang, B., Guo, L., Jia, L., & Cheng, X. (2024). Evolvable graph neural network for system-level incremental fault diagnosis of train transmission systems. *Mechanical Systems and Signal Processing., 210*(December 2023), Article 111175. https://doi.org/10.1016/j.ymssp.2024.111175

Fan, C., Wang, P., Zhang, Y., Ma, H., Li, X., & Wang, Q. (2025). Digital twin assisted degradation assessment of bearing cage performance. *IEEE Transactions on Industrial Informatics, PP, 1–11*. https://doi.org/10.1109/TII.2025.3552655

Fan, L., Chen, X., Chai, Y., & Lin, W. (2023). Attribute fusion transfer for zero-shot fault diagnosis. *Advanced Engineering Informatics, 58*, Article 102204. https://doi.org/10.1016/j.aei.2023.102204

Gui, S., Liu, M., Li, X., Luo, Y., & Ji, S. (2023). Joint learning of label and environment causal independence for graph out-of-distribution generalization. *Advances in Neural Information Processing Systems, 36*(1), 1–34.

He, Y., & Shen, W. (2024). FedITA: A cloud–edge collaboration framework for domain generalization-based federated fault diagnosis of machine-level industrial motors. *Advanced Engineering Informatics, 62*(PC), Article 102853. https://doi.org/10.1016/j.aei.2024.102853

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing, 138*, Article 106587. https://doi.org/10.1016/j.ymssp.2019.106587

Li, Z., Zhang, K., Zheng, Q., Ding, G., Hao, W., Zhang, H., & Zhang, W. (2025). Unsupervised fault detection with multi-source anomaly sensitivity enhancing convolutional autoencoder for high-speed train bogie bearings. *Expert Systems with Applications,, 281*(August 2024), Article 127570. https://doi.org/10.1016/j.eswa.2025.127570

Liu, Y., Miao, C., Li, X., Ji, J., & Meng, D. (2021). Research on the fault analysis method of belt conveyor idlers based on sound and thermal infrared image features. *Measurement, 186*(September), Article 110177. https://doi.org/10.1016/j.measurement.2021.110177

Mao, G., Li, H., Xue, L., Li, Y., Cai, Z., & Noman, K. (2025). FedPM-SGN: A federated graph network for aviation equipment fault diagnosis by multi-sensor fusion in decentralized and heterogeneous setting. *Information Fusion, 117*(May 2024), 102876. Doi: 10.1016/j.inffus.2024.102876.

Miao, S., Liu, M., & Li, P. (2022). Interpretable and generalizable graph learning via stochastic attention mechanism. *Proceedings of Machine Learning Research, 162*, 15524–15543.

Peng, C., Sheng, Y., Gui, W., Tang, Z., & Li, C. (2024). A rolling bearing fault diagnosis method based on multimodal knowledge graph. *IEEE Transactions on Industrial Informatics, 20*(11), 1–11. https://doi.org/10.1109/tii.2024.3431074

Qian, Q., Wen, Q., Tang, R., & Qin, Y. (2025). DG-Softmax: A new domain generalization intelligent fault diagnosis method for planetary gearboxes. *Reliability Engineering and System Safety*, *260*(November 2024), 111057. https://doi.org/10.1016/j.ress.2025.111057.

Ragab, M., Chen, Z., Zhang, W., Eldele, E., Wu, M., Kwoh, C. K., & Li, X. (2022). Conditional contrastive domain generalization for fault diagnosis. *IEEE Transactions on Instrumentation and Measurement, 71*(1–12). https://doi.org/10.1109/TIM.2022.3154000

Raouf, I., Kumar, P., & Kim, H. S. (2024). Deep learning-based fault diagnosis of servo motor bearing using the attention-guided feature aggregation network. *Expert Systems with Applications, 258*(August), Article 125137. https://doi.org/10.1016/j.eswa.2024.125137

Shang, J., Xu, D., Qiu, H., Gao, L., Jiang, C., & Yi, P. (2024). A novel data augmentation framework for remaining useful life estimation with dense convolutional regression network. *Journal of Manufacturing Systems, 74*(February), 30–40. https://doi.org/10.1016/j.jmsy.2024.02.011

Sun, D., Li, Y., Jia, S., Feng, K., & Liu, Z. (2023). Non-contact diagnosis for gearbox based on the fusion of multi-sensor heterogeneous data. *Information Fusion*, *94*(September 2022) pp112–125 Doi: 10.1016/j.inffus.2023.01.020.

Wang, C., Wang, Y., Zheng, F., Li, M., & Xia, R. (2025). SSTG: An interpretable spatio-temporal Selective State-Space Model for multi-sensor data fusion in intelligent diagnosis. *Knowledge-Based Systems, 316*(February), Article 113278. https://doi.org/10.1016/j.knosys.2025.113278

Wang, J., Yang, J., Wang, Y., Bai, Y., Zhang, T., & Yao, D. (2022). Ensemble decision approach with dislocated time–frequency representation and pre-trained CNN for fault diagnosis of railway vehicle gearboxes under variable conditions. *International Journal of Rail Transportation, 10*(5), 655–673. https://doi.org/10.1080/23248378.2021.2000897

Wang, Jinxin, Sun, X., Zhang, C., & Ma, X. (2022). An integrated methodology for system-level early fault detection and isolation. *Expert Systems with Applications., 201*(September 2021), Article 117080. https://doi.org/10.1016/j.eswa.2022.117080

Wang, R., Jiang, H., Zhu, K., Wang, Y., & Liu, C. (2022). A deep feature enhanced reinforcement learning method for rolling bearing fault diagnosis. *Advanced Engineering Informatics, 54*(August), Article 101750. https://doi.org/10.1016/j.aei.2022.101750

Wang, S., Shuai, H., Hu, J., Zhang, J., Liu, S., Yuan, X., & Liang, P. (2025). Few-shot fault diagnosis of axial piston pump based on prior knowledge-embedded meta learning vision transformer under variable operating conditions. *Expert Systems with Applications., 269*(December 2024), Article 126452. https://doi.org/10.1016/j.eswa.2025.126452

Wu, Y., Sicard, B., & Gadsden, S. A. (2024). Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring. *Expert Systems with Applications, 124678*. https://doi.org/10.1016/j.eswa.2024.124678

Xiao, X., Li, C., He, H., Huang, J., & Yu, T. (2025a). Rotating machinery fault diagnosis method based on multi-level fusion framework of multi-sensor information. *Information Fusion., 113*(August 2024), Article 102621. https://doi.org/10.1016/j.inffus.2024.102621

Xiao, Y., Shao, H., Wang, J., Cai, B., & Liu, B. (2025b). Domain-augmented meta ensemble learning for mechanical fault diagnosis from heterogeneous source domains to unseen target domains. *Expert Systems with Applications., 259*(November 2023), Article 125345. https://doi.org/10.1016/j.eswa.2024.125345

Yan, B., Sun, Q., Shen, L., & Ma, X. (2025). A Physical-statistical framework on complex mechanical system fault isolation. *IEEE Transactions on Reliability, PP*, 1–15. https://doi.org/10.1109/TR.2025.3549216

Yang, G., Tao, H., Wu, K., Du, R., & Zhong, Y. (2024). Fault diagnosis of harmonic drives using multimodal collaborative meta network with severely missing modality. *IEEE Transactions on Industrial Informatics, PP*, 1–9. https://doi.org/10.1109/TII.2024.3396339

Yang, Y., Yin, J., Zheng, H., Li, Y., Xu, M., & Chen, Y. (2020). Learn generalization feature via convolutional neural network: A fault diagnosis scheme toward unseen

operating conditions. *IEEE Access, 8*, 91103–91115. https://doi.org/10.1109/ACCESS.2020.2994310

Yu, S., Pang, S., Ning, J., Wang, M., & Song, L. (2025). ANC-Net: A novel multi-scale active noise cancellation network for rotating machinery fault diagnosis based on discrete wavelet transform,. *Expert Systems with Applications., 265*(July 2024), Article 125937. https://doi.org/10.1016/j.eswa.2024.125937

Zhang, T., Jiang, L., Liu, J., Zhang, X., & Zhang, Q. (2025). A multi-scale deep feature memory and recovery network for multi-sensor fault diagnosis in the channel missing scenario. *Engineering Applications of Artificial Intelligence, 145*(February), Article 110228. https://doi.org/10.1016/j.engappai.2025.110228

Zhang, Y., Ding, J., Li, Y., Ren, Z., & Feng, K. (2024). Multi-modal data cross-domain fusion network for gearbox fault diagnosis under variable operating conditions. *Engineering Applications of Artificial Intelligence, 133*(PC), Article 108236. https://doi.org/10.1016/j.engappai.2024.108236

Zhang, Y., Yu, K., Lei, Z., Ge, J., Xu, Y., Li, Z., Ren, Z., & Feng, K. (2023). Integrated intelligent fault diagnosis approach of offshore wind turbine bearing based on information stream fusion and semi-supervised learning. *Expert Systems with Applications., 232*(September 2022), Article 120854. https://doi.org/10.1016/j.eswa.2023.120854

Zhao, C., & Shen, W. (2022). A domain generalization network combing invariance and specificity towards real-time intelligent fault diagnosis. *Mechanical Systems and Signal Processing, 173*. https://doi.org/10.1016/j.ymssp.2022.108990

Zhao, C., Zio, E., & Shen, W. (2024). Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study. *Reliability Engineering and System Safety, 245*(November 202), Article 109964. https://doi.org/10.1016/j.ress.2024.109964

Zhu, Y., Zi, Y., Li, J., & Xu, J. (2024). PhysiCausalNet: a causal- and physics-driven domain generalization network for cross-machine fault diagnosis of unseen domain. *IEEE Transactions on Industrial Informatics, 20*(6), 8488–8498. https://doi.org/10.1109/TII.2024.3369240

Zio, E. (2022). Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. *Reliability Engineering and System Safety, 218*, Article 108119. https://doi.org/10.1016/j.ress.2021.108119