Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P3345-cd

Causal Intervention-Based GNNs for OOD Generalization in Fault Diagnosis of Wind Turbines

Xinming Li

the School of Mechanical-Electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, China. E-mail: lxm18134462961@163.com

Yanxue Wang*

the School of Mechanical-Electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, China. Corresponding author. E-mail: yan.xue.wang@gmail.com

Wenhan Lyu*

the School of Mechanical-Electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, China. Corresponding author. E-mail: lvwenhan@bucea.edu.cn

Jiachi Yao

the School of Mechanical-Electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, China. E-mail: yaojc.2020@tsinghua.org.cn
the Guangxi Key Laboratory of Manufacturing System Advanced Manufacturing Technology, Guilin University of Electronic Technology, China.
the Key Laboratory of Marine Power Engineering and Technology of Ministry of Transport, Wuhan University of Technology, China.

Meng Li

the School of Mechanical-Electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, China. E-mail: limeng23451@163.com

The generalization of machine learning models in out-of-distribution (OOD) scenarios remains a significant challenge, particularly in the context of high-end equipment diagnostics, where dynamic operating environments introduce complex distribution shifts. This study proposes a novel intelligent diagnostic framework based on graph causal intervention, designed to improve model adaptability and robustness under heterogeneous conditions. The framework leverages causal inference principles to infer pseudo-environment labels, enabling the removal of environmental confounding effects without requiring explicit environmental annotations. By integrating causal intervention mechanisms into graph-structured data, the proposed method effectively learns stable causal relationships across diverse environments, enhancing its generalization capabilities. The proposed framework demonstrates reliable performance in addressing OOD generalization challenges, significantly surpassing conventional methods. By dynamically regulating the propagation branch count, it achieves optimal recognition accuracy while reducing redundant computations and feature noise. This study offers a robust and scalable solution for OOD generalization in intelligent diagnostics, providing a foundation for practical applications in high-end industrial systems.

Keywords: Graph causal intervention, out-of-distribution generalization, pseudo-environment estimation, dynamic environmental adaptation, intelligent diagnostics.

1. Introduction

High-end equipment is a critical component of modern industry and operates in a complex, dynamic environment influenced by factors such as load and temperature. These external factors frequently cause mechanical component failures, which severely affect operational efficiency and significantly increase maintenance costs. Developing fault diagnosis systems with high real-time performance and accuracy is crucial to ensuring reliability and stability. In recent years, deep learning has made significant advances in intelligent diagnosis, gradually becoming a core technology for equipment condition monitoring and fault prediction. By analyzing historical operational data, deep learning models automatically extract key features from complex conditions, enabling high-precision fault prediction. These models demonstrate exceptional capabilities, particularly in handling high-dimensional, multivariable, and complex interactive data Amiri et al. (2025).

Specifically, Convolutional Neural Networks (CNNs) perform excellently in fault identification using vibration signals Jiang et al. (2024). CNNs use convolution operations to effectively extract local time-frequency features from the signals, enabling accurate classification of various fault modes. Recurrent Neural Networks (RNNs) and their variants excel at modeling long-term dependencies in time-series signals Vo et al. (2024). Multi-scale feature extraction and deep fusion methods significantly enhance the model's adaptability in fault diagnosis, particularly in cases of scarce data or complex environments. This improves diagnostic accuracy and robustness under varying operating conditions and fault modes. However, traditional deep learning models face limitations when dealing with complex topologies and multi-dimensional signal interactions, especially in high-complexity equipment systems, where they struggle to capture internal structures and dynamic changes.

Recently, Graph Neural Networks (GNNs) have shown great potential in handling complex topologies and non-Euclidean data, and have been successfully applied in fields like autonomous driving Ji et al. (2024), and intelligent diagnosis Li et al. (2025). Specifically, in intelligent diagnosis, a study proposed a synergistic similarity graph construction method, allowing GNNs to better capture component relationships, significantly improving fault diagnosis accuracy and robustness Wang et al. (2024). This novel method offers new perspectives and technical approaches for the intelligent diagnosis of complex systems. However, most existing GNN models and deep learning methods assume that training and test data come from the same distribution. This assumption often fails in practical applications, especially in high-end equipment environments. Environmental factors such as changes and sensor drift can cause distribution shifts, severely affecting model performance. Existing deep learning models, particularly when faced with out-of-distribution (OOD) data, often exhibit overconfidence, leading to inaccurate predictions and compromising the reliability of critical diagnostic tasks Li et al. (2024).

To address the challenges of OOD data in high-end equipment operations, this paper proposes an innovative intelligent diagnosis framework based on graph causal intervention. The framework leverages causal inference principles and introduces a novel learning objective. It removes environmental interference from the data by leveraging inferred pseudo-environment information, eliminating the need for environmental labels and addressing environmental confounding biases. This method enables the model to learn stable causal relationships without relying on specific environments, thereby improving its adaptability across environments. Unlike traditional methods that rely on explicit environmental labels, the proposed framework enhances the model's generalization ability in complex environments through causal intervention mechanisms, enabling better handling of dynamic changes and unforeseen disturbances in complex systems like mechanical equipment. Notable innovations include:

(a) The framework introduces a pseudo-environment label estimator, which adaptively estimates environment-related information based on data characteristics. This eliminates the reliance on predefined environmental labels and addresses confounding biases effectively.

(b) A novel learning objective is designed to integrate causal intervention into graph-structured data, enabling the model to identify causal features while reducing environmental interference, thereby improving generalization to OOD scenarios.

(c) The framework incorporates a mechanism to control the propagation branch count, balancing feature extraction efficiency and environmental complexity. This results in optimal recognition performance under heterogeneous conditions.

This paper is organized as follows: Section 2 explores the causal effects of environmental variables on GNN performance. Section 3 introduces the proposed graph causal intervention framework, while Section 4 presents the experimental setup and analyzes the results. Finally, Section 5 concludes the paper and highlights future research directions.

2. Causal Insights into the Effects of Environmental Variables on GNNs

In the current GNN design framework, node representation learning relies primarily on aggregating information from neighboring nodes. Through iterative updates, GNNs integrate node and neighbor features during network propagation to gradually generate final node embeddings, which serve as the foundation for prediction tasks Lin et al. (2023). Specifically, at the l-th layer, the embedding $\mathbf{h}_v^{(l)}$ of node v is updated as follows:

$$\mathbf{h}_v^{(l+1)} = \delta\left(\Psi^{(l)}\left(\{\mathbf{h}_u^{(l)} : u \in \Gamma_v \cup \{v\}\},\right)\right), \tag{1}$$

where $\Psi^{(l)}$ is the graph operation function at layer l, and Γ_v is the set of neighbors of node v. Notably, the training objective of GNNs is based on the ego-graph $\mathcal{G}_v$, which is centered on node v. The prediction y_v for node v is expressed as $f_\theta(\mathcal{G}_v)$, where f_θ represents the parameterized function of the GNN model.

During the model training process, the Maximum Likelihood Estimation (MLE) is a widely used optimization criterion. The goal is to adjust the model parameters so that the predictive distribution $q_\theta\left(Y \mid G\right)$ approximates the true data distribution as closely as possible. For node prediction tasks, the optimization objective is defined by the cross-entropy loss function:

$$\theta^* = \arg\min_\theta -\frac{1}{|\mathcal{V}_{tr}|}\sum_{v\in\mathcal{V}_{\mathrm{tr}}} y_v^{\mathrm{T}} \cdot \log\left(f_\theta(\mathcal{G}_v)\right). \tag{2}$$

In practice, the relationship between the ego-graph $\mathcal{G}_v$ and its label Y is influenced by both direct interactions and latent environmental factors $\boldsymbol{M}$. These variables may include data distribution biases, external interference, and other complex factors that constrain model predictions via intricate causal mechanisms.

To better capture this dependency, the model's learning objective can be reformulated as an expectation minimization problem that accounts for the impact of environmental factors $\boldsymbol{M}$:

$$\theta^* = \arg\min_\theta \mathbb{E}_{m\sim p_{\mathrm{tr}}(M),\,(G_v,y_v)\sim p(G,Y|M=m)} \left[-y_v^\top \log f_\theta(\mathcal{G}_v)\right]. \tag{3}$$

where $\mathbb{E}_{m\sim p_{\mathrm{tr}}(M)}$ denotes the expectation over the joint distribution of environmental variables $\boldsymbol{M}$ and graph structure $\mathcal{G}_v$. Therefore, the model's parameter optimization is constrained by the distribution characteristics of $\boldsymbol{M}$, which impact the predictive labels y due to environmental complexity.

3. Proposed Methods

3.1. *Hierarchical Graph Network Modeling*

In diagnosing and identifying the states of complex industrial systems, data collected under multiple working conditions often exhibit significant nonlinearity and high dimensionality. To address this challenge, we propose a hierarchical graph-based modeling approach utilizing GNNs. This approach transforms time-series data into a structured graph format, capturing both local fine-grained features and global behavioral patterns. Assume there are n distinct working states (e.g., varying rotational speeds or damage sizes). Time-series data collected under each working condition are denoted as $\mathcal{T}_n^{(k)}$, where $k = 1, 2, \cdots, K$ represents the k-th condition of state n, and each time-series segment is labeled as $y^{(n)} \in \{1, 2, \cdots, K\}$. To effectively capture sequential features, the original time series $\mathcal{T}_n^{(k)}$ is partitioned into non-overlapping subsequences using a time window of length Δt. Each subsequence is mapped to a node $v_{i,n}^{(k)}$ in the graph. Simultaneously, a feature vector $h_{i,n}^{(k)}$ is extracted for each node to represent dynamic behaviors within the time window. Within each working state, intra-layer connections are constructed to quantify the similarity between nodes. Specifically, the Euclidean distance is used to measure the feature similarity between any two nodes $v_{i,n}^{(k)}$ and $v_{j,n}^{(k)}$. Using the nearest-neighbor strategy, each node is connected to its most similar neighbors, forming a sparse local graph structure $\mathcal{G}_n^{(k)} = \left(\mathcal{V}_n^{(k)}, \mathcal{E}_n^{(k)}\right)$. This strategy preserves strong node correlations, enabling precise capture of local dynamic features under single working conditions.

To characterize global behavioral patterns across multiple working conditions, we propose a cross-layer connection mechanism. This mechanism identifies behavioral similarities among different states. For nodes $v_{i,n}^{(k)}$ and $v_{j,b}^{(c)}$ from different layers, the cross-domain similarity is computed using cosine similarity. When the similarity between two nodes exceeds a predefined threshold ς, a cross-layer edge connection is established between nodes $v_{i,n}^{(k)}$ and $v_{j,b}^{(c)}$. This connection mechanism integrates features across multiple working conditions and models multi-level dependencies, effectively capturing complex behavioral patterns across the system.

3.2. *Enhancing Model Stability and Generalization with Causal Inference*

To enhance the OOD generalization capability of GNNs for node attribute prediction tasks, this paper introduces a robustness enhancement strategy based on causal inference. This strategy guides the model to discover stable causal relationships in the data that remain invariant to environmental shifts. Specifically, we apply a do-operation to intervene on the environmental variable $\boldsymbol{M}$ Pearl et al. (2016). The do-operation removes the confounding effects of $\boldsymbol{M}$ on graph features $\boldsymbol{\mathcal{G}}_v$. This allows the model to capture the stable causal relationship $q_\theta\left(Y|\, do\left(G\right)\right)$ between the graph structure $\boldsymbol{\mathcal{G}}_v$ and the label Y, while avoiding the influence of noise and unstable factors. While directly computing $q_\theta\left(Y|\, do\left(G\right)\right)$ is ideal, it is often infeasible in real-world applications due to high experimental costs and resource limitations. To address this issue, we adopt the backdoor adjustment strategy using observational data to approximate the causal intervention. This adjustment effectively removes the interference of environmental factors on predictions.

To enhance the model's robustness to environmental shifts, we propose an approximation method based on pseudo-environment labels. This method introduces latent variables to decouple pseudo-environment labels from graph features $\boldsymbol{\mathcal{G}}_v$. The process includes designing a pseudo-environment estimator $p_\Omega\left(M|\, G\right)$ to infer the environmental variable $\boldsymbol{M}$ based on the ego-graph features $\boldsymbol{\mathcal{G}}_v$. The inferred $\boldsymbol{M}$ and ego-graph features $\boldsymbol{\mathcal{G}}_v$ are fed into the GNN prediction model for joint optimization, expressed as follows:

$$\begin{aligned}\log q_\theta(Y \mid do(G)) \geq \mathbb{E}_{p_\Omega(M|G)}\left[\log q_\theta(Y \mid G, M)\right] \\ - KL(p_\Omega(M \mid G) \parallel q_0(M)).\end{aligned} \tag{4}$$

In this formulation, the first term represents the supervised loss $\mathcal{L}_{\text{sup}}$ which improves the model's prediction accuracy, while the second term is the regularization loss $\mathcal{L}_{reg}$, ensuring the independence of $\boldsymbol{M}$ from ego-graph features $\boldsymbol{\mathcal{G}}_v$. This independence decouples environmental information from graph structural dependencies, allowing the model to identify causal patterns robust to environmental changes. Consequently, the model's generalization to OOD scenarios is enhanced, addressing challenges caused by data distribution shifts.

3.3. *Dynamic Generalization Enhancement through Pseudo-Environment Mechanisms*

In complex scenarios, data often lack explicit environmental labels, which makes determining true environmental states using prior knowledge challenging. Moreover, complex node connections often cause distributional shifts, which degrade the model's generalization performance. To address these challenges, we propose a method called pseudo-environment representation modeling and adaptive expert ensemble.

To capture the latent environmental effects, we introduce a pseudo-environment estimator $p_\Omega\left(M|\, G\right)$. This estimator infers the environmental variable $\boldsymbol{M}$ during the multi-layer feature aggregation process of the graph neural network. Specifically, the pseudo-environment variable $m_v^{(l)}$, considered as a latent variable, is inferred from the node feature vector$h_v^{(l)}$ and represented as a categorical distribution. To resolve the non-differentiability issue of categorical distribution sampling, we adopt the Gumbel-Softmax technique, which ensures stable gradient propagation through continuous approximation. The node embedding, incorporating the pseudo-environment representation, is updated as follows:

$$\Theta_v^{(l)} = softmax\left(W^{(l)} h_v^{(l)}\right), \tag{5}$$

where $W^{(l)}$ represents the learnable weight matrix at layer l. Using this method, pseudo-environment labels are dynamically inferred during training, ensuring effective modeling of node-level categorical distributions.

In OOD generalisation tasks, models must possess the capability to dynamically adapt to variations across diverse environments. To enhance the model's adaptability to changing environmental conditions, we propose the hierarchical adaptive expert ensemble GNN. This method employs multiple expert branches to dynamically update node embeddings. Within this framework, each expert branch aggregates node features dynamically based on the inferred pseudo-environment labels and the ego-graph structure $\boldsymbol{\mathcal{G}}_v$. The node feature update is formulated as:

$$\mathbf{h}_v^{(l+1)} = \varphi\left(\sum_{z=1}^{Z} m_{v,z}^{(l)} \left(\sum_{u\in\mathcal{N}(v)} \frac{1}{\sqrt{\delta_v \delta_u}} \overline{W}^{(l,z)} \mathbf{h}_u^{(l)}\right) + W^{(l,z)} \mathbf{h}_v^{(l)}\right), \tag{6}$$

where δ_v and δ_u represents the degrees of nodes v and u, used to normalize neighboring node features. $\widetilde{W}^{(l,z)}$ and $W^{(l,z)}$ are transformation matrices for the current node and its neighbors in expert branch $\boldsymbol{Z}$, respectively. The activation function $\varphi()$ captures nonlinear interactions among node features. This framework can be interpreted as a causal representation of GCNs, where the dynamically inferred $m_{v,z}^{(l)}$ guides the selection of $\boldsymbol{Z}$ convolutional filters, enabling adaptive propagation.

3.4. *The Overall Framework*

To tackle the challenges of OOD tasks in complex industrial environments, this paper presents an intelligent diagnostic framework integrating causal inference, a hierarchical adaptive expert mechanism, and GNNs. The framework combines causal modeling with adaptive feature learning to dynamically adjust to varying working conditions, enhancing robustness and generalization performance for complex, non-Euclidean data. The specific design includes the following three core steps:

(a) Vibration signals and node relationships are first processed and transformed into a multi-level graph representation. Using graph-based modeling, the framework captures both the topological structure and multi-scale feature interactions among nodes.

(b) To mitigate distributional shifts caused by changing environmental conditions, the framework incorporates pseudo-environment variables as latent intermediate variables. A pseudo-environment label estimator is designed based on causal inference theory.

(c) A hierarchical adaptive expert mechanism is introduced during the node feature learning process. At each layer, node representations are updated using a multi-branch expert network. Expert branches dynamically select optimal convolution kernels based on the inferred pseudo-environment labels, enabling efficient node feature aggregation and interaction modeling.

The proposed intelligent diagnostic framework integrates the robustness of causal inference with the flexibility of the adaptive expert mechanism. It achieves accurate node attribute prediction and enhanced generalization under varying working conditions, offering an efficient and reliable solution for intelligent diagnostic tasks.

4. Experimental Setup and Results Analysis

4.1. *Experimental Platform and Dataset Description*

This study uses the BUCEA_Bearing dataset, collected from a self-constructed doubly-fed wind turbine test platform. The dataset contains six types of bearing fault conditions, divided into two categories: (a) Single fault states: inner ring fault (IF), outer ring fault (OF), rolling element fault

(BF), and normal state (NA); (b) Compound fault states: outer ring + rolling element fault (OBF) and inner ring + rolling element fault (IBF). To evaluate the model's adaptability and generalization under various fault conditions, faults with different groove sizes were simulated: 0.4×1 mm, 2×2 mm, 2.8×3 mm, 3.4×4 mm, and 4×4 mm. During data acquisition, five rotational speeds were set to simulate operating conditions under varying wind speeds: 500 r/min, 700 r/min, 900 r/min, 1000 r/min, and 1100 r/min. A sampling frequency of 25.6 kHz was used to ensure high-resolution acquisition of vibration signals, enabling detailed capture of subtle fault characteristics and dynamic variations. The experimental design comprises two schemes: (a) BUCEA-R experiment: The fault size was fixed at 2×2 mm, and rotational speeds were varied to generate ID and OOD data. This experiment evaluated the model's adaptability to distribution shifts across different rotational speeds. (b) BUCEA-S Experiment: At a fixed rotational speed of 900 r/min, the impact of varying fault sizes on model performance was analyzed. This experiment aimed to assess the model's generalization performance and robustness under varying fault scales.

4.2. *Experimental Implementation Details*

During data preprocessing, vibration signals were segmented using the sliding window technique with a window length of 1024. Each vibration signal, under a specific fault condition, had a total length of 102400 and was divided into 100 nodes after sliding window processing. These nodes were then used to construct and train graph-structured data. To maintain consistency and rigor, the dataset was split into 60% for training, 10% for validation, and 30% for testing. All models were implemented under a unified hardware and software environment to ensure fairness and result reproducibility. PyTorch served as the implementation framework, with a learning rate of 0.001 and 300 training epochs.

4.3. *Results and Analysis*

4.3.1. *Impact of Rotational Speed and Fault Size on Generalization*

This section examines the model's generalization performance in ID and OOD scenarios using the BUCEA-R and BUCEA-S datasets. The analysis focuses on how rotational speed variations and fault size differences influence the model's performance and explores the potential reasons behind these effects.

Using the BUCEA-R dataset, experiments were conducted with a fixed fault size of 2×2 mm to analyze the model's performance under various rotational speeds. The results are shown in Fig.1. The model exhibited better generalization when trained on high-speed ID data and tested on low-speed OOD data. Conversely, training on low-speed ID data and testing on high-speed OOD data led to a significant performance decline. This suggests that high-speed vibration signals, characterized by stronger amplitudes and high-frequency features, enable the model to learn clearer and more stable fault patterns during training. As a result, the model effectively captured core fault features and maintained high recognition accuracy, even when tested on lower-speed data. In contrast, low-speed samples exhibited weaker signals with fault patterns lacking diversity and generalizability. This limitation restricted the model's ability to fully learn the complexity of fault distributions during training.

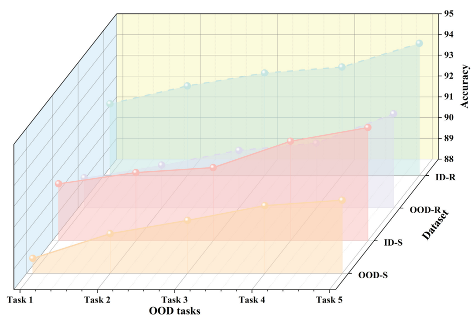

Fig. 1. OOD recognition results under different rotational speed and fault size.

Further analysis with the BUCEA-S dataset investigated the impact of fault size variations on model performance. Training on large fault sizes led to significantly improved recognition performance on small-size OOD data. This improvement is attributed to the distinct feature patterns of large-size fault signals, which allow the model to learn clearer fault representations during training and demonstrate strong feature transfer during testing. Conversely, training on small-size fault samples, characterized by weak and indistinct signals, limited the model's ability to extract critical features effectively. As a result, the model struggled to adapt to feature distribution changes when tested on large-size OOD data, resulting in a significant decline in recognition accuracy.

4.3.2. *Comparison of OOD Generalization Methods*

To evaluate the model's performance in OOD tasks, this study employs GCN as the base encoder and systematically compares several classical and state-of-the-art OOD generalization methods. The experimental results are presented in Table 3. The comparison methods specially include: ERM, IRMArjovsky et al. (2019) and SRGNNZhu et al. (2021). The results as shown in Fig.2 indicate that the proposed method achieved superior performance across all experimental datasets, particularly on the highly heterogeneous BUCEA-R and BUCEA-S datasets, where it demonstrated significant performance advantages. Compared to methods such as IRM and SRGNN, the proposed approach achieved substantial accuracy improvements. This suggests that the high heterogeneity and substantial distribution shifts in the BUCEA-R and BUCEA-S datasets present severe challenges to traditional generalization methods. By adaptively estimating pseudo-environment labels, the proposed approach effectively captured core feature patterns under varying conditions, improving modeling precision and generalization capability for distribution-shifted data.

4.3.3. *Impact of Propagation Branch Count on Model Performance*

To better understand the impact of key hyperparameters on model performance, this study

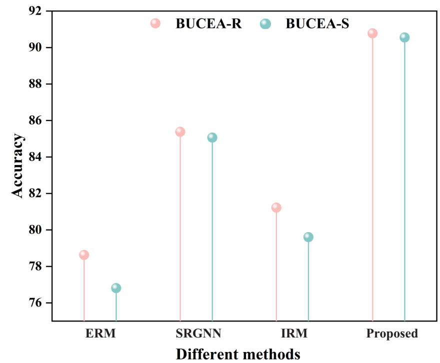

Fig. 2. Comparison of results across different baselines.

systematically analyzed the propagation branch count Z, assessing its role and limitations in OOD generalization tasks. The experimental results are shown in Fig.3. The results indicate a non-monotonic relationship between model performance and Z. When Z is small, the limited number of propagation branches fails to capture the multi-dimensional features of complex environments, restricting the model's ability to adapt to environmental variations. This leads to suboptimal detection performance and reduced generalization capability.

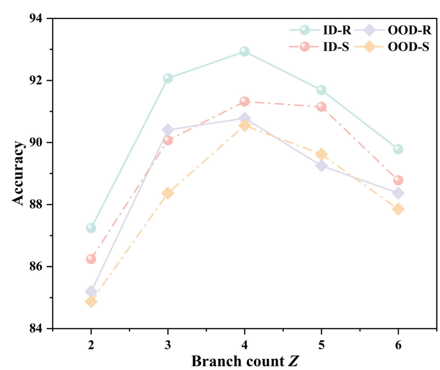

Fig. 3. Impact of branch count Z on model performance.

As Z increases, model performance improves, suggesting that the pseudo-environment estimator effectively captures diverse feature patterns across environments and facilitates efficient propagation path regulation. At Z=4, the model achieves its optimal state by balancing branch count and environmental feature complexity. This balance ensures the efficient extraction and propagation of critical information, resulting in optimal OOD recognition performance on heterogeneous data. However, as Z continues to increase, model performance begins to degrade. This degradation is mainly caused by excessive branches introducing redundant computations and feature noise, which weaken the model's focus on critical features and ultimately reduce its generalization capability. This phenomenon is especially pronounced in the highly complex dataset, highlighting that an excessive branch count can compromise the model's adaptability to diverse environments.

5. Conclusion

This study introduces a novel diagnostic framework based on graph causal intervention to tackle the challenges of out-of-distribution (OOD) generalization in intelligent diagnostics. By integrating causal inference mechanisms, the framework infers pseudo-environment labels, effectively mitigating the influence of environmental confounders and enhancing adaptability across dynamic and heterogeneous conditions. The framework demonstrates superior performance compared to existing methods, showcasing its ability to capture stable causal relationships in graph-structured data and significantly reduce the impact of distribution shifts. This robust and scalable framework is broadly applicable to fault detection tasks in high-end equipment diagnostics, addressing the complexities of dynamic environments and unforeseen disturbances. Future research will focus on extending this approach to other complex systems, exploring its potential for broader generalization and improved causal reasoning.

Acknowledgement

This study was supported by the National Natural Science Foundation of China (52275079, 12204031, 52405146), the Youth Beijing Scholars program, the Guangxi Science and Technology Major Project (AA23062031), the Young Teachers Research Capacity Im-provement Program (X23003), the Guangxi Key Laboratory of Manufacturing System & Advanced Manufacturing Technology (24354S014), the Key Laboratory of Marine Power Engineering and Technology of Ministry of Transport (KLMPET2023-03) and the BUCEA Post Graduate Innovation Project.

References

Amiri, M. H., N. M. Hashjin, M. K. Najafabadi, A. Beheshti, and N. Khodadadi (2025). An innovative data-driven ai approach for detecting and isolating faults in gas turbines at power plants. *Expert Systems with Applications 263*, 125497.

Arjovsky, M., L. Bottou, I. Gulrajani, and D. Lopez-Paz (2019). Invariant risk minimization. *arXiv preprint arXiv:1907.02893*.

Ji, Q., X. Wen, J. Jin, Y. Zhu, and Y. Lv (2024). Urban traffic control meets decision recommendation system: A survey and perspective. *IEEE/CAA Journal of Automatica Sinica 11*(10), 2043–2058.

Jiang, L., C. Shi, H. Sheng, X. Li, and T. Yang (2024). Lightweight cnn architecture design for rolling bearing fault diagnosis. *Measurement Science and Technology 35*(12), 126142.

Li, X., M. Li, J. Gu, Y. Wang, J. Yao, and J. Feng (2024). Energy-propagation graph neural networks for enhanced out-of-distribution fault analysis in intelligent construction machinery systems. *IEEE Internet of Things Journal*.

Li, X., Y. Wang, S. Zhao, J. Yao, and M. Li (2025). Adaptive convergent visibility graph network: An interpretable method for intelligent rolling bearing diagnosis. *Mechanical Systems and Signal Processing 222*, 111761.

Lin, H., M. Yan, X. Ye, D. Fan, S. Pan, W. Chen, and Y. Xie (2023). A comprehensive survey on distributed training of graph neural networks. *Proceedings of the IEEE*.

Pearl, J., M. Glymour, and N. P. Jewell (2016). Causal inference in statistics: A primer. 2016. *Internet resource*.

Vo, T.-T., M.-K. Liu, and M.-Q. Tran (2024). Harnessing attention mechanisms in a comprehensive deep learning approach for induction motor fault diagnosis using raw electrical signals. *Engineering Applications of Artificial Intelligence 129*, 107643.

Wang, X., X. Zhuang, J. Ge, J. Xiang, and D. Zhou (2024). Ssgc-gat: Synergistic similarity graph construction strategy combined with gat network for wind turbine anomaly identification using scada data. *IEEE Transactions on Instrumentation and Measurement*.

Zhu, Q., N. Ponomareva, J. Han, and B. Perozzi (2021). Shift-robust gnns: Overcoming the limitations of localized graph training data. *Advances in Neural Information Processing Systems 34*, 27965–27977.