IOP Publishing Measurement Science and Technology
Meas. Sci. Technol. **36** (2025) 076110 (22pp) https://doi.org/10.1088/1361-6501/ade7a7

Metric-guided graph contrastive learning: an unsupervised approach for few-shot gearbox fault diagnosis

Peng Chen[1,2,*], Jia Gao[1], Ruijin Zhang[1], Yaqiang Jin[3,*], Renhui Yu[4], Changbo He[5] and Junyu Qi[6]

[1] College of Engineering, Shantou University, Shantou 515063, Guangdong, People's Republic of China
[2] Key Laboratory of Intelligent Manufacturing Technology, Shantou University, Shantou 515063, Guangdong, People's Republic of China
[3] School of Qilu Transportation, Shandong University, Jinan 250061, Shandong, People's Republic of China
[4] Guangdong Institute of Special Equipment Inspection and Research Shunde Branch, Foshan 528318, Guangdong, People's Republic of China
[5] College of Electrical Engineering and Automation, Anhui University, Hefei, Anhui 230601, People's Republic of China
[6] Electronics & Drives, Reutlingen University, 72762 Reutlingen, Germany

E-mail: pengchen@alu.uestc.edu.cn, dr.pengchen@foxmail.com and yaqiang.jin@outlook.com

Received 29 March 2025, revised 8 June 2025
Accepted for publication 24 June 2025
Published 3 July 2025

Abstract

Planetary gearboxes are critical mechanical components widely deployed in industrial applications such as wind turbines, helicopters, and hybrid vehicles, where their reliable operation directly impacts system performance and safety. Traditional fault diagnosis approaches using graph neural networks and graph contrastive learning (GCL) face significant challenges, including prohibitive costs in fault sample acquisition, ineffective feature extraction from limited data, and semantic distortions in node embedding space that compromise diagnostic accuracy. Furthermore, existing methods struggle with insufficient supervision for complex fault classification and show vulnerability to distribution shifts in new environments. To address these limitations, this research proposes the metric-guided GCL (MGCL) framework, featuring three innovative components: a feature-decoupled pre-training mechanism with graph data augmentation, a sophisticated cosine-Euclidean hybrid distance metric, and a two-stage training paradigm combining unsupervised pre-training with weakly supervised fine-tuning. MGCL significantly advances the field by effectively handling sample scarcity and annotation limitations while enhancing model robustness against domain shifts in real-world industrial applications, ultimately providing a more reliable and practical solution for industrial fault diagnosis.

Keywords: planetary gearbox, fault diagnosis, unsupervised learning, weakly supervised learning, few-shot learning

* Authors to whom any correspondence should be addressed.

1. Introduction

Planetary gearboxes serve as fundamental mechanical components in diverse industrial applications, ranging from wind turbines and helicopters to electric motors and hybrid vehicles, where they play a critical role in power transmission and torque conversion for ensuring optimal system performance [1–3]. Nevertheless, these sophisticated mechanisms are inherently susceptible to various forms of mechanical deterioration, including progressive wear, structural cracking, and catastrophic tooth breakage, primarily due to prolonged exposure to high-speed operations and continuously fluctuating load conditions. In this context, the implementation of proactive prediction strategies and comprehensive health management protocols has become increasingly essential, as these approaches not only facilitate safe and reliable operation but also significantly enhance the overall operational reliability of industrial systems. Fault detection, as one of its primary branches, serves as a cornerstone in maintaining system integrity [4–6]. Furthermore, with the advent of advanced data processing capabilities and the systematic exploitation of vast quantities of operational data, data-driven fault diagnosis methodologies have emerged as particularly promising solutions [7–9].

In the field of mechanical fault diagnosis, traditional data-driven methodologies primarily rely on signal processing techniques for feature extraction, encompassing a wide spectrum of analytical approaches such as time-domain statistical parameters [10, 11], frequency-domain spectral analysis [12, 13], and time-frequency domain decomposition techniques [14]. Subsequently, these extracted features are integrated with conventional machine learning classifiers [15], notably support vector machines (SVM) [16, 17] and Random Forest algorithms [18], to facilitate fault identification and classification [19–21]. Nevertheless, these conventional approaches exhibit two significant inherent limitations that substantially impact their diagnostic effectiveness. Firstly, their heavy dependence on manual feature engineering presents a fundamental challenge, particularly when confronted with complex non-stationary signals characteristic of real-world mechanical systems. This reliance on human-designed features not only limits the method's adaptability but also leads to notable performance deterioration when operating conditions deviate from the baseline scenarios. Furthermore, and perhaps more critically, these traditional models demonstrate insufficient capability in processing high-dimensional nonlinear signals that are inherent in modern mechanical systems. This limitation manifests in two crucial aspects: the models struggle to effectively capture and leverage the intricate spatio-temporal correlations and component coupling relationships present in multi-source data streams, while simultaneously failing to adequately model the complex dynamics of fault propagation patterns and multi-scale temporal variations in system behavior.

Graph neural network (GNN) data-driven fault diagnosis methods have emerged as a powerful paradigm that effectively captures system complexity through node-edge representations, thereby enabling comprehensive modeling of component interactions and temporal dynamics within industrial systems [22]. Contemporary graph-based diagnostic frameworks predominantly focus on constructing high-fidelity graph models through sophisticated data point similarity analyzes, which subsequently serves as the cornerstone for downstream diagnostic processes [23]. Several noteworthy implementations demonstrate this approach's versatility. For instance, Zhou *et al* [24]. developed a methodology that transforms noise vibration signals into static graphs using spectral characteristics, while implementing distance metric functions to optimize edge redundancy. Furthermore, Han *et al* [25]. introduced an adaptive multi-relation fusion framework for constructing lightweight metapath graphs, which simultaneously reduces structural complexity while enhancing overall graph quality. Additionally, Qing *et al* [26]. innovatively incorporated temporal neighborhood relationships into spatial connectivity patterns for pre-connection operations and amplitude modulation construction, utilizing selective adjacency matrix training exclusively for connected samples to achieve optimal weight distributions. Nevertheless, despite their considerable advantages, GNN-based methods face several significant limitations. First and foremost, the computational complexity of graph construction increases exponentially with sample size, as each additional datapoint necessitates evaluation of its relationships with all existing nodes. This scalability challenge impacts model training and inference efficiency, particularly in large-scale industrial systems requiring real-time diagnostic capabilities. Moreover, the development of high-quality graphs fundamentally depends on the availability of both labeled and unlabeled samples in sufficient quantities. However, the procurement of labeled samples in industrial environments often proves to be both resource-intensive and time-prohibitive, thereby constraining the method's broader implementation.

Few-shot learning (FSL) technology emerges as a crucial advancement in addressing the limitations of traditional GNN approaches, enabling models to achieve robust fault representation learning under limited labeled sample conditions through knowledge transfer and feature reuse mechanisms. Contemporary research on few-shot fault diagnosis based on GNN primarily advances along three technical paths: First and foremost, meta learning driven relationship modeling transforms the topological similarity between devices into transferable meta knowledge by constructing task aware graph structures. In this context, Liu *et al* [27]. proposed a semi-supervised meta-learning method with simplified graph convolutional networks (Meta-SGC) to address bearing fault diagnosis under complex working conditions and limited samples. Secondly, the integration of prototype networks and metric learning significantly improves the distinguishability of fault categories. Li *et al* [28]. proposed a curriculum learning-enhanced GNN to address label imbalance iuin node classification tasks. By integrating adaptive graph oversampling and a hybrid loss combining graph classification loss with

metric learning, this framework dynamically optimizes spatial proximity of minority-class nodes while mitigating overfitting. Finally, the data generation strategy, when enhanced by physical constraints, markedly expands the coverage of potential fault modes through domain knowledge-guided graph enhancement techniques. Zhang *et al* [29]. introduced self-mixup augmentation to synthesize diverse instances from limited samples and designing calibration-adaptive downsampling to mitigate feature distortion caused by subsampling violations, the method enhances both knowledge accumulation and feature robustness. However, FSL still face various limitations: limited cross-domain transfer due to equipment-specific characteristics, increased noise sensitivity with scarce samples, and difficulty adapting to dynamic operating conditions that require temporal modeling of both monitoring data and graph evolution.

Graph contrastive learning (GCL) emerges as a natural evolution in addressing the collective challenges of both GNN and FSL approaches, providing a novel paradigm by maximizing the similarity of similar samples in the feature space while separating heterogeneous sample distributions. By leveraging the strengths of self-supervised learning while mitigating the dependencies on labeled data, GCL represents a significant advancement in the field of fault diagnosis. This self-supervised approach extracts more discriminative and robust graph structural features from unlabeled samples through the construction of positive and negative sample pairs, integration of self-attention mechanisms, and learning from high-dimensional attributes and local structural patterns. Notable advancements have been made in addressing these challenges, such as, Liu *et al* [30]. sophisticated GCL model that effectively learns from high-dimensional attributes and local structural patterns, while Zhu *et al* [31]. innovative integration of self-attention mechanisms with GCL has demonstrated exceptional performance in gearbox feature extraction. Despite its promising advances, current GCL methods face three significant challenges. The model's generalization capability becomes constrained when transferring between tasks with divergent data distributions, particularly in cross-domain applications. In small sample scenarios, the limited scale of negative sample pools, combined with traditional uniform sampling approaches, often leads to confusion between failure modes of varying wear degrees, thereby compromising the discriminative power of contrast edge distances. Furthermore, the reliance on fixed distance metrics in existing comparative loss functions hinders the optimal balance between preserving local topological accuracy and maintaining global structural integrity.

The main challenges for the reported methodologies are summarized as follows:

1. High acquisition costs of industrial fault samples coupled with traditional GNNs' inability to extract meaningful spatiotemporal features from limited annotated data, often leading to misleading learning signals and degraded model performance.
2. GCL's fixed distance metrics fail to effectively differentiate between noise-induced false connections and genuine equipment relationships, causing semantic distortions in node embedding space.
3. GCL-based fault prediction suffers from three key limitations: insufficient supervision for complex fault classification, vulnerability to distribution shifts in new environments, and lack of model interpretability.

This research introduces the metric-guided GCL (MGCL) framework to tackle fundamental challenges in industrial fault diagnosis. At its core, MGCL develops three innovative components: a feature-decoupled pre-training mechanism that leverages graph data augmentation with large-scale unlabeled learning, a sophisticated cosine-Euclidean hybrid distance metric for enhanced feature discrimination, and a two-stage training paradigm combining unsupervised pre-training with weakly supervised fine-tuning. This approach not only addresses the critical issues of sample scarcity and annotation limitations but also strengthens model robustness against domain shifts in real-world industrial applications. The framework's key technical contributions are as follows:

1. A feature-decoupled pre-training approach that combines graph data augmentation with large-scale unlabeled data learning, enhancing sample diversity and fault mode coverage while building robust feature representations.
2. A novel cosine-Euclidean hybrid distance metric that improves feature discrimination in small-sample scenarios by leveraging both directional sensitivity and absolute spatial differences. This metric not only optimizes the graph topology, but also provides interpretable insights into the advantages of feature relationships.
3. A two-stage training framework that integrates unsupervised pre-training with weakly supervised fine-tuning, effectively addressing annotation scarcity and domain shift challenges in few-shot fault diagnosis.

The structural organization of the remainder of this paper proceeds in a systematic manner through several interconnected sections. In section 2, we thoroughly examine and critically analyze the theoretical foundations underlying three fundamental components: the dynamic graph attention network (DGAT), which enables adaptive feature learning; GCL, which facilitates representation learning through comparative analysis; and K-nearest neighbor (KNN) graph construction, which forms the basis of our graph topology. Subsequently, section 3 presents a detailed exposition of our novel methodology, namely the MGCL, along with its architectural components and theoretical underpinnings. Furthermore, section 4 offers a comprehensive presentation of our experimental findings, accompanied by an in-depth comparative analysis with state-of-the-art approaches and ablation studies. Building upon these empirical results, section 5 systematically investigates the individual contributions of core components through ablation studies, specifically quantifying the impact of hybrid distance metrics and pre-training strategies employing GCL. Following this, section 6 provides a rigorous examination of MGCL's performance characteristics, including sensitivity analyzes of critical hyperparameters such as the KNN graph's

sparsity factor, distance weighting coefficients in hybrid metric design, and epoch-wise convergence patterns. Finally, section 7 synthesizes our key contributions, draws meaningful conclusions, and delineates promising directions for future research endeavors in this field.

2. Related work

2.1. DGAT

The basic graph attention network (GAT) architecture, while capable of capturing the relative importance of information between neighboring nodes through iterative weight updates, exhibits certain limitations due to its static attention mechanism. Specifically, this mechanism fails to account for the temporal evolution of connections between central nodes, which consequently constrains the model's expressive capacity and representational power, which can be shown as:

$$e_{ij} = \text{LeakyReLU}\left(F_1 W f_i + F_2 W f_j\right) \tag{1}$$

where e_{ij} is the importance of node V_j to node V_i, f_i and f_j are feathers of nodes V_i and V_j, W denotes weight matrix, F_1 and F_2 are the shared attention mechanism.

A fundamental limitation of the conventional GAT architecture lies in its parameter-sharing mechanism, wherein the transformation matrices W, F_1, and F_2 in equation (1) are globally shared across all nodes and applied in a sequential manner. Consequently, these operations can be mathematically reduced to a single linear transformation layer, effectively resulting in a static attention mechanism that lacks adaptive capabilities. To overcome these architectural constraints, the DGAT introduces a sophisticated aggregation mechanism for central node processing. Specifically, the DGAT implements a modified computational flow where the transformation matrix W is strategically applied post-concatenation, while the combined attention mechanism $F = [F_1 \| F_2]$ is positioned after the Leakyrelu activation function. This architectural reorganization enables the network to achieve truly dynamic attention capabilities, allowing for more flexible and context-aware feature processing. The comprehensive computational workflow within the DGAT layer is illustrated in detail in figure 1. The dynamic attention score in DGAT is computed through the following mathematical formulation:

$$e_{ij} = F \cdot \text{LeakyReLU}\left(W \cdot [f_i \| f_j]\right) \tag{2}$$

$$\alpha_{ij} = \text{Softmax}\left(e_{ij}\right) = \frac{\exp\left(e_{ij}\right)}{\sum_{k \in N(i)} \exp\left(e_{ik}\right)} \tag{3}$$

where F represents a shared attention mechanism, and $N(i)$ represents the set of all nodes adjacent to node V_i.

The attention computation process is fundamentally grounded in the characteristic features of individual nodes and systematically incorporates the interconnected relationships with neighboring nodes. This comprehensive approach significantly enhances the model's ability to capture and represent complex graph structures. To illustrate this process more concretely, consider the feature learning procedure at the lth layer of DGAT, which can be mathematically expressed as:

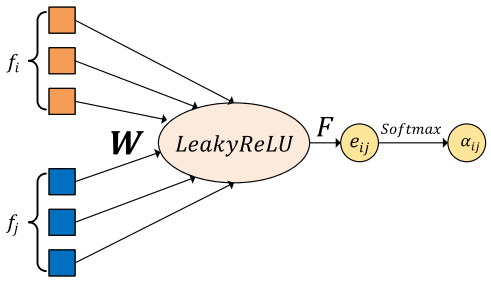

Figure 1. Architectural framework and computational flow diagram of the dynamic graph attention network (DGAT) layer.

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} W^{(l)} h_j^{(l)}\right) \tag{4}$$

where $\sigma(\cdot)$ denotes activation function, $h_i^{(l+1)}$ represents the updated vector after node V_i aggregates the information of layer l, and $W^{(l)}$ is its learnable weight matrix.

2.2. GCL

GCL represents a sophisticated unsupervised learning framework that has been specifically engineered to extract meaningful node representations from unlabeled graph-structured data. Through the careful design and implementation of comparative analysis tasks, GCL effectively enables neural network models to systematically capture and distinguish both the inherent consistencies and fundamental differences in graph structures and node attributes when examined across multiple augmented views of the same underlying data.

The Information Noise Contrastive Estimation (InfoNCE) loss function has emerged as one of the most widely adopted and theoretically well-grounded approaches in this domain, primarily because it elegantly encapsulates the core principles of contrastive learning. This mathematical framework operates by simultaneously pursuing two complementary objectives: maximizing the measured similarity between positive sample pairs (instances that should be considered similar) while minimizing the similarity between negative sample pairs (instances that should be considered distinct). Formally, the InfoNCE loss can be mathematically defined as:

$$\text{Loss} = -\log \frac{\exp\left(\text{sim}\left(f_i^1, f_i^2\right)/\tau\right)}{\sum_{j=1}^{N} 1_{[j \neq i]} \exp\left(\text{sim}\left(f_i^1, f_j^2\right)/\tau\right)} \tag{5}$$

where f_i^1, f_i^2 represent the feature vectors of node V_i in two different views. *sim* denotes the function used to calculate the similarity between feature vectors. τ is the temperature parameter used to adjust the sharpness of the similarity distribution. $1_{[j \neq i]}$ represents the indicator function, ensuring that sample V_j is considered as a negative sample only if j is not equal to i.

2.3. Construction of KNN graph

The diagnostic analysis of planetary gearbox faults primarily relies on one-dimensional vibration signal data, which serves as the fundamental input for fault detection systems. However, due to the inherently limited availability of such fault samples and the considerable complexity of raw vibration signals, which inherently contain both ambient noise and extraneous information irrelevant to fault characteristics, it becomes imperative to implement signal processing techniques. Consequently, we employ the fast Fourier transform (FFT) algorithm, a typical mathematical tool that effectively transforms the time-domain vibration signals into their corresponding frequency-domain representations.

This transformation through FFT proves particularly advantageous because it not only facilitates the elimination of spurious information and streamlines the signal structure, but also considerably reduces the computational complexity associated with subsequent analytical procedures. Furthermore, the frequency-domain representation provides a more interpretable framework for fault pattern recognition, as many mechanical faults manifest distinctly in specific frequency bands. In the subsequent phase of analysis, these frequency-domain samples are systematically utilized as graph nodes for the construction of the adjacency matrix, which forms the foundational structure for graph-based fault classification. The methodological framework for constructing the KNN graph follows a well-defined protocol. For quantifying the relationships between nodes, either Euclidean distance or cosine similarity metrics are employed as the primary distance measures, wherein the mathematical computation of distance follows the formal expression:

$$D_{ij}^{\text{euc}} = \sqrt{\sum_{k=1}^{d} \left(x_{ik} - x_{jk}\right)^2} \tag{6}$$

$$D_{ij}^{\cos} = \frac{\sum_{k=1}^{d} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{d} x_{ik}^2} \sqrt{\sum_{k=1}^{d} x_{jk}^2}} \tag{7}$$

where D_{ij}^{euc} and $D_{ij}^{\cos}$ represent the Euclidean distance and cosine similarity between the ith and jth nodes, respectively. $x_{i,k}$ represents the value of the ith node in the k dimension. d is the total dimension of the node.

Based on the obtained distance, KNN clustering of nodes is carried out, and edges are constructed to obtain the adjacency matrix $A \in R^{n \times n}$:

$$A_{ij} = \begin{cases} 1 & \text{if } i \in N_k(j) \wedge j \in N_k(i) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where $N_k(i)$ denotes the set of KNNs of the ith node. $A_{i,j}$ is the element of the adjacency matrix.

3. MGCL

As illustrated comprehensively in figure 2, the proposed MGCL framework encompasses several interconnected stages, each of which plays a crucial role in the overall methodology. The framework begins with systematic node feature extraction, followed by the graph representation. Subsequently, the process advances to graph construction through the implementation of hybrid distance metrics, which enables more nuanced structural relationships. The framework then proceeds with unsupervised model training to learn general patterns, and ultimately culminates in weakly-supervised model training that is specifically designed to address the challenges of few-shot node classification tasks.

3.1. Node feature extraction and graph representation

For the graph-based signal processing and machine learning, individual samples within the dataset are systematically represented as discrete nodes within the constructed graph structure, wherein node features constitute a fundamental and integral component of the graph's architecture. The initial and crucial step in the graph construction process involves executing node embedding operations, which systematically extracts the raw features of each sample and subsequently transforms them into meaningful graph node features through a well-defined mathematical framework.

Specifically, given an initial sample set denoted as $X = \{X_1, X_2, \ldots, X_n\}$, where each X_i represents a distinct sample point, the corresponding spectrum features are obtained through the application of the FFT, a computationally efficient algorithm for spectral analysis. This transformation can be mathematically expressed as:

$$F_i = \text{FFT}(X_i) \tag{9}$$

where $X_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,m}\}$ represents the ith sample, $F_i = \{f_{i,1}, f_{i,2}, \ldots, f_{i,m/2}\}$ contains half of the results after FFT, and it is used as node feature after normalization operation. By repeating the same operation, all the sample features are converted into node features, and the node feature matrix $F \in R^{n \times \frac{m}{2}}$ is obtained.

3.2. Graph construction by hybrid distance metric integration

The establishment of inter-node connections represents a critical and fundamental step in graph construction. This process systematically leverages node features to compute comprehensive distance matrices, subsequently utilizing the KNN clustering algorithm to establish meaningful edge connections between nodes. The methodology employs a sophisticated dual-metric approach, wherein two distinct distance matrices are computed and subsequently integrated into a novel hybrid matrix. This integration process incorporates both a cosine similarity-based distance matrix and a Euclidean distance-based matrix. The cosine similarity-based approach captures the directional relationships and angular similarities between node vectors, while providing robust measurement invariant to magnitude scaling. Complementarily, the Euclidean distance-based matrix quantifies the absolute geometric distances between nodes in feature space and preserves the magnitude-based relationships between data points.

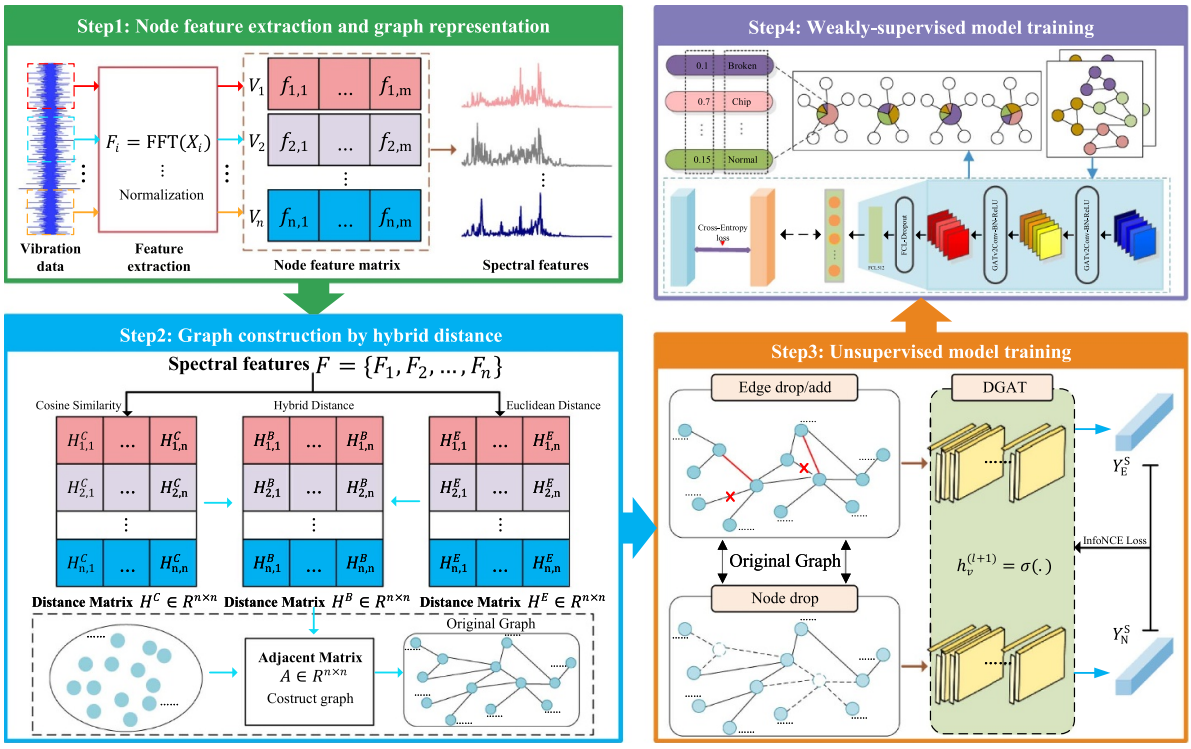

Figure 2. The proposed metric-guided graph contrastive learning (MGCL).

This metric not only optimizes the graph topology, but also provides interpretable insights into the advantages of feature relationships.

The strategic combination of these complementary distance metrics is particularly advantageous for several reasons. First, cosine similarity effectively captures directional correlations between nodes, while Euclidean distance preserves crucial spatial relationships in the feature space. Furthermore, their integration yields a more comprehensive and nuanced representation of inter-node relationships, enabling a more accurate characterization of the underlying data structure. The mathematical formulation proceeds as follows:

$$H_{i,j}^{\mathrm{C}} = 1 - \frac{F_i \cdot F_j}{\|F_i\|_2 \|F_j\|_2} \tag{10}$$

where F_i and F_j are the node features of nodes V_i and V_j, $H_{i,j}^{\mathrm{C}}$ is an element of the first kind of distance matrix, representing the Cosine similarity between nodes V_i and V_j.

For the Euclidean distance matrix $H^{\mathrm{E}} \in R^{n\times n}$:

$$H_{i,j}^{\mathrm{E}} = \|F_i - F_j\|_2 \tag{11}$$

where $H_{i,j}^{\mathrm{E}}$ quantifies the Euclidean distance between nodes V_i and V_j.

To ensure computational stability and metric compatibility, the Euclidean distances undergo normalization. Subsequently, the hybrid distance matrix integrates both metrics through:

$$H_{i,j}^{\mathrm{B}} = \frac{1}{2}\left(H_{i,j}^{\mathrm{E}} + H_{i,j}^{\mathrm{C}}\right) \tag{12}$$

where $H_{i,j}^{\mathrm{B}}$ is an element of the final distance matrix, representing the Hybrid distance between nodes V_i and V_j. ϵ is a very small positive number that prevents zero division errors that occur when $H_{i,j}^{\mathrm{E}}$ is equal to 0.

Finally, the adjacency matrix $A \in R^{n\times n}$ is constructed using KNN relationships:

$$A_{i,j} = \begin{cases} 1, & \text{if } V_j \in N_k(V_i) \text{ or } V_i \in N_k(V_j) \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where $N_k(V_i)$ represents the KNN node set of node V_i, that is, the set of k nodes that are closest to V_i. $A_{i,j}$ denotes the element of the adjacency matrix used in the composition.

3.3. Unsupervised model training framework

In this methodology, the original graph structure, denoted as $G = (V, E)$, undergoes two distinct data augmentation processes, wherein V and E respectively represent the complete sets of nodes and edges. Through these transformations, two enhanced graph variants, designated as G_1 and G_2, are systematically generated. The first augmentation technique involves edge manipulation, which encompasses both stochastic edge removal and addition operations, implemented with a predetermined probability p_d. This process can be formally expressed as:

$$G_1 = (V_1, E_1) = (\{v_i \mid r_i < p_d, r_i \sim U[0,1], v_i \in V\}, \{(v_i, v_j) \mid v_i, v_j \in V_1, (v_i, v_j) \in E\}) \tag{14}$$

where r_i represents a randomly sampled value from a uniform distribution $U[0,1]$, which subsequently determines the retention status of node v_i within the augmented set V_1.

The second augmentation procedure, known as node dropping, systematically removes vertices and their corresponding edge connections from the initial graph structure with probability p_e. This transformation can be mathematically represented as:

$$G_2 = (V_2, E_2) = (V, \{e_{ij} \mid r_{ij} < p_e, r_{ij} \sim U[0,1], e_{ij} \in E \cup E_{\text{potential}}\}) \tag{15}$$

where r_{ij} denotes a random value sampled from $U[0,1]$ that governs whether edge e_{ij} is maintained or introduced in E_2, and $E_{\text{potential}}$ encompasses all feasible edges that are not present in the original edge set E.

Subsequently, both augmented graph structures are processed through a unified DGAT architecture that maintains consistent parameters for feature representation learning across both graph variants. Specifically, a two-layer DGAT framework is implemented for graph feature extraction, whose layer-wise computations can be expressed as:

$$\begin{cases} Y^{(1)} = \sigma\left(\text{DGAT}\left(F, W^{(1)}\right)\right) \\ Y^{(2)} = \text{DGAT}\left(Y^{(1)}, W^{(2)}\right) \end{cases} \tag{16}$$

where σ represents the rectified linear unit activation function, $Y^{(1)} \in R^{n\times c}$ and $Y^{(2)} \in R^{n\times e}$ denote the respective output representations from the first and second layers, while $W^{(1)} \in R^{m\times c}$ and $W^{(2)} \in R^{c\times e}$ represent their corresponding weight matrices. The final DGAT output undergoes transformation through a fully connected (FC) layer, computed as:

$$Y^{\text{P}} = \text{FC}\left(Y^{(2)}\right) \tag{17}$$

where $Y^{\text{P}} \in R^{n\times p}$ represents the final FC layer output.

The enhancement strategy has been carefully designed to preserve the basic frequency domain characteristics when applied to Fourier transform signals. Although edge descent introduces structural disturbances, its impact on the spectral domain is limited as it prioritizes retaining the main low-frequency components related to fault characteristics rather than high-frequency edges dominated by transients or noise. By limiting edge deletion to non dominant spectral regions, ensuring that diagnostic connections remain intact, as demonstrated by the consistent preservation of fault related spectral peaks in enhanced samples. This method is consistent with the inherent redundancy in mechanical systems, where local edge removal simulates real-world signal variations without eliminating global frequency patterns that are crucial for diagnosis, thereby enhancing the model's robustness to structural irregularities while maintaining fidelity to identifying spectral features.

The method generates a divergent enhanced view while changing the local neighborhood composition and global topological sparsity pattern. This multi-faceted perturbation strategy amplifies the variability beyond traditional edge only modifications: node deletion enforces robustness to missing entities, while edge addition/deletion simulates fluctuating relationship confidence. The contrastive learning framework utilizes these structurally heterogeneous views to standardize the encoder and prevent excessive reliance on transient substructures, as matching node representations in the topology changes caused by enhancement requires capturing invariant semantic features. It is crucial that the composite randomness in node edge operations ensures exponential growth of trusted graph changes during the training process, systematically expanding the potential spatial coverage range.

The model's training process is guided by an unsupervised contrastive learning loss function, which is systematically constructed through the generation and computation of positive and negative sample pairs. This specialized loss function, as detailed in equation (5), is specifically designed to simultaneously maximize the similarity between positive sample pairs while minimizing the similarity between negative sample pairs, thereby facilitating effective unsupervised learning.

3.4. Weakly-supervised model training methodology

Following the successful completion of unsupervised DGAT model training, the pre-trained model undergoes further refinement through training with a limited number of labeled samples. During this phase, the previously established distance matrix continues to serve as the foundational structure for implementing the weakly supervised training protocol. To optimize the pre-trained DGAT model parameters effectively, we employ the cross-entropy loss function, which is mathematically expressed as:

$$\text{Loss} = -\frac{1}{I}\sum_{i}^{I}\sum_{i}^{T} y_i^{(t)} \ln\left(z_i^{(t)}\right) \tag{18}$$

where the prediction label set $z_i \in Z = \{z_1, \cdots, z_n\}$ represents the model's output following the FC layer, I denotes the total number of distinct labels in the dataset, and T corresponds to the number of possible states in the system. Furthermore, $y_i^{(t)}$ signifies the t-dimensional value of the ground truth label y_i, while $z_i^{(t)}$ represents the corresponding t-dimensional value of the predicted label z_i.

The comprehensive methodology can be systematically outlined as follows: Initially, the raw data undergoes feature extraction and graph representation procedures to generate a robust node feature matrix. Subsequently, multiple distance matrices are synthesized into a novel hybrid distance matrix, which is then utilized for pre-training the initial DGAT model through unsupervised GCL. In the final phase, the pre-trained DGAT model is further refined through weak supervision, leveraging a carefully curated dataset comprising a limited number of labeled samples, ultimately enabling effective fault diagnosis capabilities. The complete algorithmic workflow is rigorously detailed in algorithm 1.

Algorithm 1. Computational procedure and pseudo-code specification of the MGCL architecture.

procedure StageOne(Initial sample set $X = \{X_1, X_2, \ldots, X_n\}$)
  Stage 1: Node feature extraction and graph representation
  Calculate the feature matrix F: $F_i = \mathrm{FFT}(X_i)$
  Output the feature matrix $F = \{F_1, F_2, \ldots, F_n\}$
end procedure
procedure StageTwo(Feature matrix F)
  Stage 2: Graph construction by hybrid distance metric integration
  Calculate distance matrix H^C based on cosine similarity:$H^C_{i,j} = 1 - \frac{F_i \cdot F_j}{\|F_i\|_2 \|F_j\|_2}$
  Calculate distance matrix H^{E} based on euclidean distance: $H^{\mathrm{E}}_{i,j} = \|F_i - F_j\|_2$
  Obtain the hybrid distance matrix H^{B}: $H^{\mathrm{B}}_{i,j} = \frac{1}{2}(H^{\mathrm{E}}_{i,j} + H^C_{i,j})$
  Calculate the adjacent matrix A of KNN graph :
$$A_{i,j} = \begin{cases} 1, & \text{if } V_j \in N_k(V_i) \text{or } V_i \in N_k(V_j) \\ 0, & \text{otherwise} \end{cases}$$
  Output the original KNN graph G
end procedure
procedure StageThree(KNNG $G = (V, E)$)
  Stage 3: Unsupervised model training
  Two methods of graph augmentation: (1) edge manipulation (random addition or removal of edges) and (2) random node deletion.
  Initialize $i = 1$
  while $i \leqslant 350$ **do**
    Compute predictions $Y^P_1 \leftarrow DGAT(F_1, A_1)$ and $Y^P_2 \leftarrow DGAT(F_2, A_2)$
    Compute GCL loss
    Update with back propagation
    Increment i
  end while
  Output the pretrained DGAT
end procedure
procedure StageFour(Sample set containing a few labeled samples V, Pretrained DGAT)
  Stage 4: Weakly-supervised model training
  Divide the training node set and testing node set: V_{train}, V_{test} and two corresponding graphs: G_R, G_E
  Initialize $i = 1$
  while $i \leqslant 200$ **do**
    Compute prediction $Y^P_R \leftarrow DGAT(F_R, A_R)$
    Calculate cross-entropy loss
    Update with back propagation
    Increment i
  end while
  Output predicted label vector $Z = DGAT(F_E, A_E) = \{z_1, \cdots, z_n\}$
end procedure

4. Experimental results and comparative analysis

To rigorously evaluate and validate the efficacy of the proposed MGCL framework, we present comprehensive analyzes through two distinct case studies, each offering unique insights into the methodology's performance and applicability.

4.1. Case study I

4.1.1. Experimental setup and data acquisition protocol. The experimental investigation was conducted utilizing a sophisticated drivetrain prognostics simulator (DPS), manufactured by SpectraQuest Inc. as depicted in figure 3. This advanced diagnostic platform incorporates several key components: a precision-controlled variable speed drive motor, an integrated planetary gearbox system, a dual-stage parallel gearbox configuration, resistance-load gear boxes coupled with a specialized resistance-load inducing electric load motor, and a comprehensive electric control unit for precise operational management.

The fundamental physical parameters of the planetary gear system are meticulously documented in table 1. Our investigation specifically focuses on the dynamic behavior of a spur-geared planetary gearbox system integrated with a two-stage parallel helical gearbox configuration. The experimental dataset encompasses four distinct operational states: (1) broken component condition, (2) chipped gear state, (3) crack formation, and (4) normal operational condition. Data acquisition was performed under strictly controlled experimental conditions, with horizontal position signals recorded at a high-precision sampling frequency of 24 kHz, while maintaining a consistent input speed of 20 Hz.

Based on the acquired experimental data, we systematically constructed the experimental dataset for comprehensive methodology validation. This dataset comprises 120 distinct samples for each operational state under controlled conditions, resulting in a total of 480 samples across all four health states. Each individual sample consists of 2048 data points, thereby generating a complex graph structure with 480 nodes for GCL implementation. For the weakly supervised training phase, we employed a structured sampling approach, selecting 1, 3, 5, and 10 nodes from each health state category to form various training sets, while allocating 10 nodes for validation purposes. The model's performance evaluation was conducted using 480 previously unutilized samples to construct the test KNN graph through equation (8), with an equal distribution of 120 samples per health state. The architectural specifications of the implemented DGAT model are detailed in table 2.

The enhancement methodology has been meticulously engineered to maintain fundamental frequency domain characteristics during Fourier transform signal processing, while simultaneously introducing controlled perturbations that strengthen the model's learning capacity. Although the edge descent procedure inherently introduces structural modifications, its impact on the spectral domain remains carefully circumscribed, as the algorithm deliberately prioritizes the preservation of critical low-frequency components associated with fault signatures rather than high-frequency elements typically dominated by transient phenomena or stochastic noise. Furthermore, the strategic limitation of edge deletion to non-dominant spectral regions ensures the integrity of diagnostic pathways, as evidenced by the consistent retention of fault-related spectral peaks across enhanced samples. This approach aligns fundamentally with the inherent redundancy characteristics of mechanical systems, wherein localized edge removal

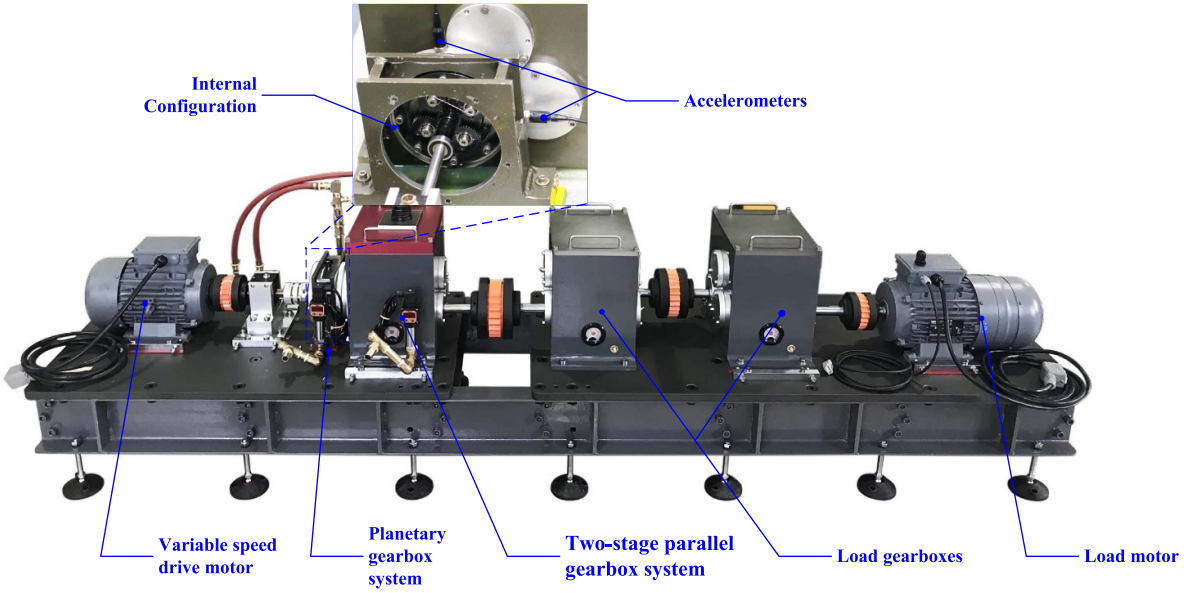

Figure 3. Illustration of the drivetrain prognostics simulation (DPS).

Table 1. Physical parameters of the planetary gear set in DPS.

Parameters	Sun	Planet (4)	Ring	Carrier
Number of teeth	28	36	100	—
Module [mm]	1	1	1	—
Pressure angle [°]	20	20	20	—
Face width [mm]	10	10	10	—
Young's modulus [Pa]	2.1×10^{11}	2.1×10^{11}	2.1×10^{11}	—
Poisson's ratio	0.3	0.3	0.3	—
Mass [kg]	—	—	9.86×10^{-2}	—
Moment of inertia [kg·m^2]	2.41×10^{-6}	1.60×10^{-5}	9.20×10^{-3}	4.99×10^{4}
Base circle [mm]	13.2	16.9	47.0	—
Torsional stiffness [N m^{-1}]	0	—	1×10^{9}	0
Torsional damping [N·s m^{-1}]	0	—	1×10^{3}	0

Table 2. Structure of the DGAT model.

Layer	Input channels	Output channels	Params
GATv2Conv1	Feature	1024	Heads = 4
Linear1	1024	1024	—
BatchNorm1	1024	1024	—
GATv2Conv2	1024	1024	Heads = 4
Linear2	1024	1024	—
BatchNorm2	1024	1024	—
FCL1	1024	512	Act: ReLU, Inplace = True
Dropout1	512	512	$p = 0.2$
FCL2	512	Out_channel	—

Table 3. Comparative analysis of classification accuracy (%) for case study I.

Model	Number of labeled samples			
	1	3	5	10
AE+CNN	39.86 ± 3.03	68.13 ± 13.24	74.65 ± 10.41	88.34 ± 3.76
AE+SVM	35.90 ± 4.27	55.69 ± 7.14	76.11 ± 10.3	82.85 ± 1.20
SSGCDBN	40.00 ± 4.45	68.54 ± 8.04	81.18 ± 4.46	86.46 ± 2.19
ChebyNet	43.47 ± 8.23	62.43 ± 2.62	71.18 ± 8.96	87.64 ± 1.94
GRL-CSG	63.54 ± 3.72	76.04 ± 4.83	82.08 ± 5.75	86.04 ± 3.22
Meta-SGC	62.71 ± 3.28	72.83 ± 5.01	84.17 ± 5.16	90.42 ± 6.24
Hyperfast	36.18 ± 6.80	62.50 ± 11.19	75.83 ± 6.05	87.08 ± 4.10
TabPFN	77.71 ± 1.63	83.33 ± 8.61	84.86 ± 6.37	89.44 ± 2.43
Ours (MGCL)	$\mathbf{92.91 \pm 4.48}$	$\mathbf{95.35 \pm 3.12}$	$\mathbf{96.81 \pm 3.08}$	$\mathbf{99.52 \pm 0.43}$

effectively simulates real-world signal variations without compromising the global frequency patterns that are instrumental in diagnostic processes.

The methodology's sophisticated perturbation framework generates divergent enhanced views through simultaneous manipulation of local neighborhood compositions and global topological sparsity patterns. This multi-dimensional approach transcends conventional edge-only modifications by incorporating node deletion operations to foster resilience against missing data points, while concurrent edge addition and deletion processes simulate varying degrees of relationship confidence within the network structure. The contrastive learning framework leverages these structurally heterogeneous views to regularize the encoder's behavior and mitigate excessive dependence on transient substructures, as the successful matching of node representations across topologically modified variants necessitates the capture of invariant semantic features. Notably, the composite randomness inherent in the node-edge operations facilitates exponential growth in the diversity of permissible graph transformations throughout the training process, thereby systematically expanding the model's spatial coverage and enhancing its generalization capabilities across varied structural configurations.

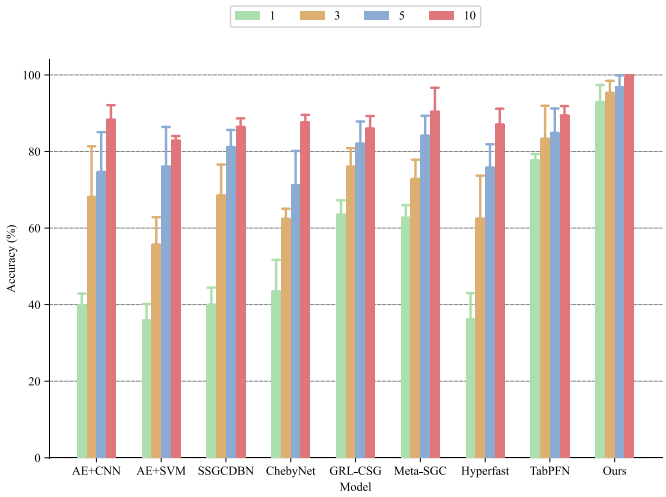

Figure 4. Evaluation of classification accuracy (%) across various model architectures in case study I, where experiments were conducted using 1, 3, 5, and 10 labeled training samples per fault category. (Error bars represent the standard deviation based on $n = 8$ independent trials).

4.1.2. Analysis and comparison of diagnostic results. To rigorously evaluate the efficacy of our proposed methodology, we conducted comprehensive comparative experiments utilizing both traditional and State-of-the-Art diagnostic approaches. The experimental framework incorporated four distinct FSL methods alongside several advanced GNN architectures. The comparative methods encompass traditional approaches including auto-encoder (AE) with CNN (AE+CNN) [32] and AE with SVM (AE+SVM) [33], as well as advanced GNNs such as Chebyshev polynomial-based Graph Convolutional Networks (ChebyNet) [34] and semi-supervised graph convolutional deep belief networks (SSGCDBN) [35]. In addition, a graph comparison learning framework based on graph representation learning and component space graph (GRL-CSG) [36], as well as a few-shot bearing fault diagnosis method by semi-supervised Meta-SGC [27], have been introduced. Furthermore, modern FSL approaches were represented by HyperFast [37], which is optimized for rapid tabular data classification, and TabPFN [38], which leverages causal inference for prior data fitting.

All methodologies underwent training using few-shot samples and subsequent validation across an extensive test set comprising 480 samples. Analysis of the diagnostic model's performance reveals compelling trends across multiple evaluation metrics, as demonstrated in both table 3 and figure 4, which systematically present the classification accuracy results obtained when training with different quantities of labeled samples (specifically 1, 3, 5, and 10 samples) per health state category. The values highlighted in bold denote the superior performance metrics attained through diverse methodological approaches under experimental conditions. While table 3 provides precise numerical data with corresponding standard deviations, these variations are visually represented through error bars in figure 4, thereby offering complementary perspectives on the model's performance stability. Furthermore, to provide a more comprehensive assessment of the model's diagnostic capabilities, table 4 presents a detailed evaluation framework incorporating multiple performance indicators, including Precision, Recall, and F1 score

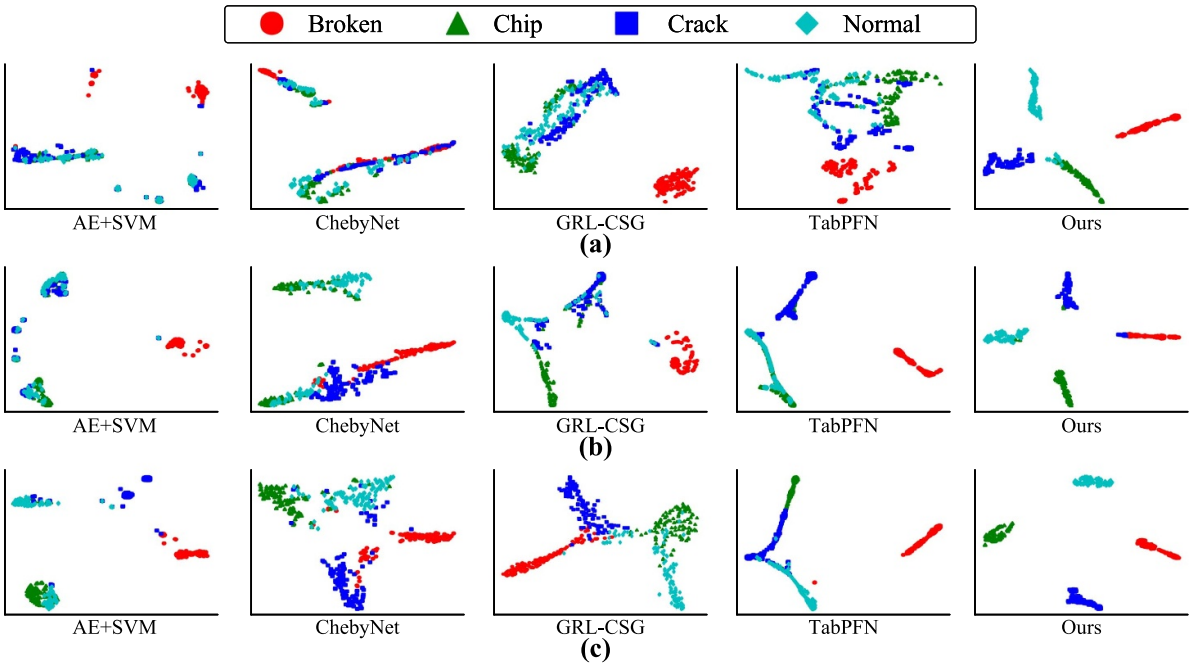

Figure 5. The T-SNE visualization of feature distributions for various methods in case study I. (a), (b), (c) represent models trained with 1, 5 and 10 samples per class, respectively.

Table 4. Comparative analysis results of key metrics (%) and running times for case study I.

	Performance metrics based on 5 labeled samples			Running times	
Method	Precision	Recall	F1 score	Training time(s)	Test time(s)
AE+CNN	82.62 ± 7.82	73.28 ± 1.96	82.76 ± 0.82	3.27	2.73
AE+SVM	73.28 ± 1.96	67.15 ± 6.67	66.05 ± 7.03	3.42	2.86
SSGCDBN	82.76 ± 0.82	74.58 ± 3.86	73.08 ± 4.40	3.94	2.62
ChebyNet	72.00 ± 6.02	68.75 ± 4.16	68.37 ± 3.93	3.93	2.65
GRL-CSG	82.82 ± 4.28	80.66 ± 5.81	83.68 ± 2.77	5.76	4.22
Meta-SGC	85.79 ± 3.98	86.26 ± 4.77	85.33 ± 6.62	4.58	3.27
HyperFast	75.93 ± 7.02	72.29 ± 9.39	72.67 ± 9.30	4.34	3.08
TabPFN	80.21 ± 3.44	78.69 ± 4.08	78.92 ± 3.03	4.40	3.05
Ours (MGCL)	$\mathbf{96.29 \pm 3.14}$	$\mathbf{95.28 \pm 4.60}$	$\mathbf{95.14 \pm 4.81}$	5.01	3.90

metrics, specifically for the scenario where 5 labeled samples per health state were utilized during the training phase. Additionally, figure 5 demonstrates the T-SNE dimensionality reduction visualization, and figure 6 presents the detailed confusion matrix analysis. The observed time efficiency metrics in table 4 demonstrate that our MGCL method incurs moderate runtime increases, a consequence of its two-stage knowledge distillation process. The comprehensive time efficiency analysis presented in table 4 reveals that our proposed MGCL methodology, although introducing moderate computational overhead due to its sophisticated two-stage knowledge distillation architecture, demonstrates a favorable performance-complexity trade-off. Specifically, while the implementation incurs an additional 1.74 s of training time relative to the baseline AE+CNN approach, thereby suggesting marginally higher computational resource requirements, this temporal cost is outweighed by the significant performance improvements, as evidenced by a 13.67% increase in absolute precision and a 12.48% enhancement in F1-score metrics. Furthermore, this disproportionate scaling between computational demands and performance gains effectively validates our paradigm's practical utility, particularly in accuracy-critical applications where slight latency tolerances can be leveraged to achieve more refined decision boundaries and, consequently, superior classification outcomes. The empirical results thus strongly support the adoption of MGCL in domains where prediction quality takes precedence over minimal processing time constraints.

Our proposed model demonstrated superior performance across all sample size conditions, achieving remarkable accuracy rates of 92.91%, 95.35%, 96.81%, and 99.52% respectively. This exceptional performance validates both the model's ability to effectively utilize limited labeled data and its enhanced learning capabilities as data availability increases. While traditional methods and modern approaches like TabPFN showed improvement with increased sample

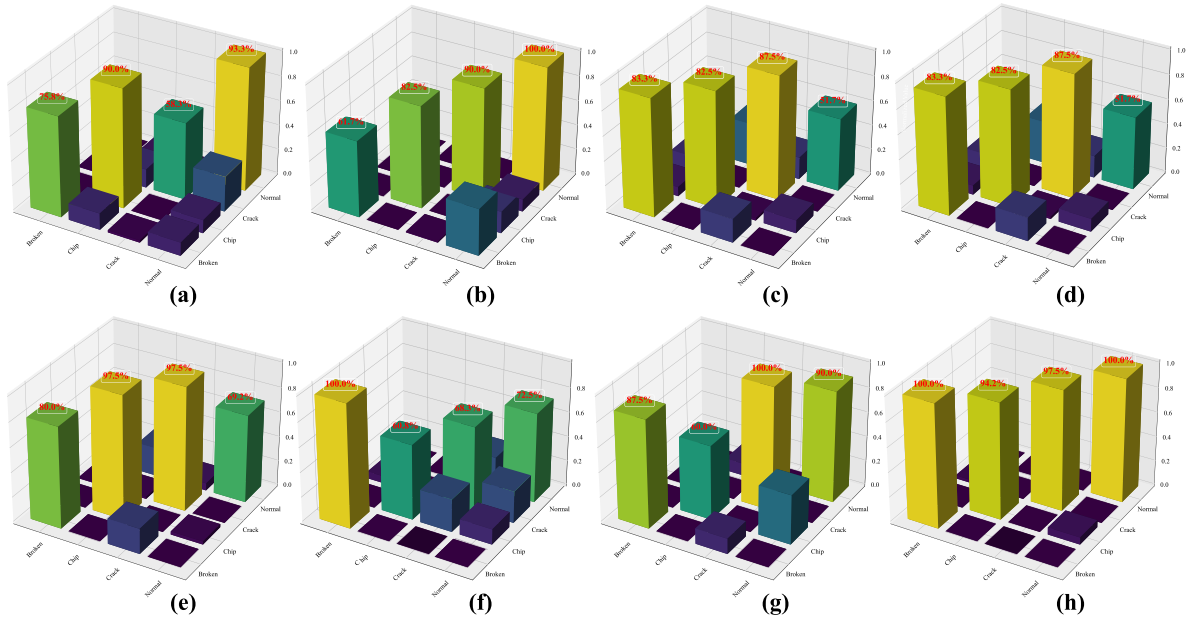

Figure 6. Comparative analysis of confusion matrices for selected methods in case study I (10 samples per class): (a) AE+SVM, (b) SSGCDBN, (c) ChebyNet, (d) GRL-CSG, (e) Meta-SGC, (f) HyperFast, (g) TabPFN, (h) ours (MGCL).

sizes, their performance remained notably inferior under few-shot conditions.

In case study I, utilizing five labeled samples, our methodology achieved markedly higher metrics compared to existing approaches. Specifically, our method attained a precision of 96.29%, significantly outperforming TabPFN's 80.21% and AE+CNN's 82.62%. In terms of recall, our approach achieved 95.28%, notably exceeding TabPFN's 78.69% and AE+CNN's 73.28%. The F1 Score of 95.14% further demonstrated our method's superior performance compared to TabPFN's 78.92% and AE+CNN's 82.76%. The T-SNE dimensionality reduction analysis, as illustrated in figure 5, reveals superior feature extraction capabilities of our method, demonstrating enhanced discriminative power across all health states. This superior performance can be attributed to our innovative hybrid distance matrix and pre-training methodology, which effectively captures and utilizes the underlying patterns in the data structure.

The confusion matrix visualization presented in figure 6 offers a detailed analysis of classification performance across four distinct health states, with 120 test samples per state. The horizontal coordinate represents the prediction label, while the vertical coordinate indicates the true label. The results definitively demonstrate the superior diagnostic capabilities of our proposed approach, particularly validating the effectiveness of our hybrid distance matrix in enhancing GNN-based fault diagnosis under extremely low labeling rates. These comprehensive results empirically validate the robustness and effectiveness of our proposed methodology in planetary gearbox fault diagnosis under limited data conditions. The consistent superior performance across multiple evaluation metrics and visualization techniques reinforces the practical applicability and reliability of our approach in real-world diagnostic scenarios.

4.2. Case study II

4.2.1. Experimental apparatus and data acquisition. The experimental investigation employs a sophisticated test apparatus known as the drivetrain diagnostics simulation (DDS), which serves as the primary platform for data collection and analysis. The DDS system, as depicted comprehensively in figure 7, constitutes an intricate assembly of meticulously integrated components, each fulfilling distinct yet interdependent functions within the experimental framework. The apparatus encompasses a high-precision data acquisition system, a driven motor that functions as the primary power input source, a precision-engineered torque transducer for accurate rotational force measurements, an advanced planetary gearbox system, strategically positioned accelerometers for comprehensive vibration analysis, a parallel gearbox configuration, and a sophisticated loading system designed to replicate diverse operational conditions. Throughout the experimental procedures, operational parameters were rigorously controlled and standardized, whereby the input speed was precisely maintained at 20 Hz, while the output load was consistently regulated at 0.35 A.

The experimental framework facilitated the development of the dataset in case study II, which was specifically structured for comprehensive method validation. This dataset encompasses 120 distinct samples for each identified health state, and given that bearings manifest four discrete health states, the complete dataset comprises 480 samples. Each individual sample contains 2048 discrete data points, ensuring robust statistical analysis. In maintaining methodological consistency, the weakly supervised training process adopts identical strategies to those implemented in case study I for the systematic construction of training, verification, and test sets. Other key model parameters are set the same as in case study I.

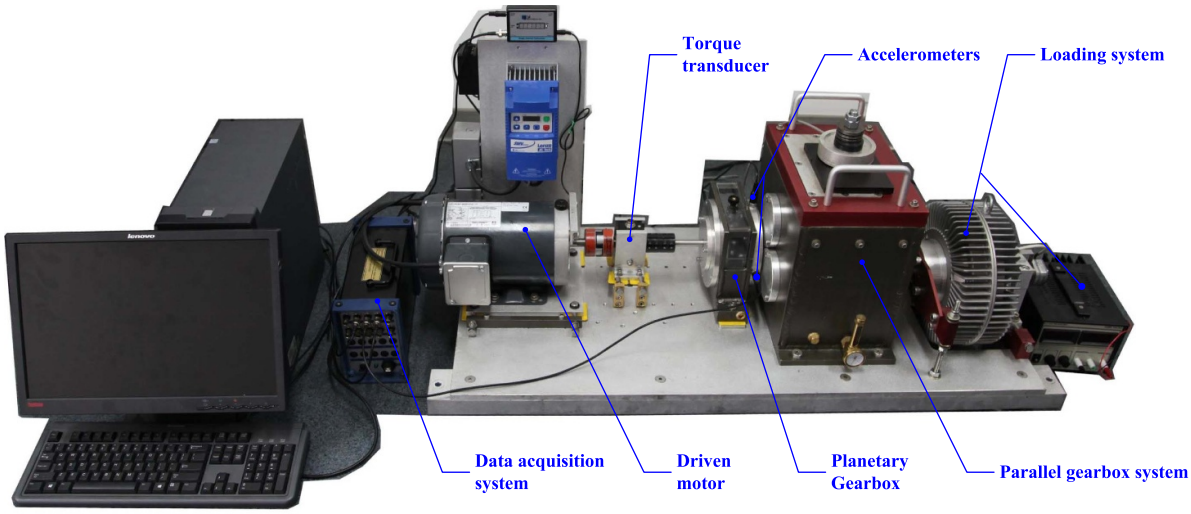

Figure 7. Diagram of the drivetrain diagnostics simulation (DDS).

Table 5. Comparative analysis of classification accuracy (%) for case study II.

	Number of labeled samples			
Model	1	3	5	10
AE+CNN	33.82 ± 12.69	60.62 ± 9.90	75.56 ± 8.50	83.87 ± 8.07
AE+SVM	45.70 ± 16.39	55.62 ± 7.19	69.03 ± 9.67	83.89 ± 6.10
SSGCDBN	43.13 ± 8.31	53.89 ± 5.08	69.58 ± 8.46	86.46 ± 1.82
ChebyNet	37.22 ± 2.30	59.17 ± 6.67	71.32 ± 8.50	80.97 ± 3.66
GRL-CSG	49.44 ± 3.56	64.59 ± 5.18	78.77 ± 5.98	88.25 ± 7.84
Meta-SGC	54.38 ± 6.21	68.14 ± 3.22	83.21 ± 4.42	87.14 ± 5.13
Hyperfast	37.08 ± 11.50	54.93 ± 13.96	73.61 ± 6.58	83.89 ± 3.07
TabPFN	33.54 ± 14.86	75.14 ± 5.87	74.16 ± 4.03	86.74 ± 5.44
Ours (MGCL)	$\mathbf{80.97 \pm 0.48}$	$\mathbf{91.87 \pm 1.46}$	$\mathbf{92.36 \pm 5.02}$	$\mathbf{95.07 \pm 3.53}$

4.2.2. Diagnosis results and comparative analysis. Building upon the analytical framework established in case study 4.1, this section presents a comprehensive evaluation of the proposed methodology through rigorous comparison with established diagnostic techniques, including AE+CNN, AE+SVM, SSGCDBN, ChebyNet, GRL-CSG, Meta-SGC, HyperFast, and TabPFN. The comparative diagnostic accuracy results for the DDS dataset are systematically presented in table 5, while additional performance metrics are detailed in table 6. The standard deviation in the table 5 is represented in the form of error bars in the figure 8. 1, 3, 5, and 10 respectively represent the number of labeled samples for each type of fault during training. Furthermore, to enhance result interpretation, we conducted sophisticated visualization analyzes through T-SNE dimensionality reduction, as illustrated in figure 9, complemented by detailed receiver operating characteristic (ROC) curve analyzes presented in figure 10.

The comparative analysis in case study II encompasses a thorough examination of methodological performance across multiple metrics, including accuracy, recall, and F1 score. As evidenced by table 5 and figure 8, the proposed methodology demonstrates considerably performance advantages over existing comparative methods across all labeled sample conditions. Notably, even with the minimal condition of a single labeled sample, our method achieved a notable accuracy of 80.97%, markedly surpassing alternative approaches, including the previously leading TabPFN method, which achieved only 33.54%. This performance differential became increasingly pronounced as the number of labeled samples increased, ultimately achieving peak accuracy of 95.07% with 10 labeled samples. Furthermore, the comprehensive performance evaluation and computational efficiency analysis, which was conducted using 5 labeled samples per class as detailed in table 6, demonstrates consistently superior results across all critical performance metrics, whereby the proposed framework achieved significant accuracy, recall, and F1 scores of 94.51%, 93.33%, and 93.25%, respectively. These empirical findings indicate that our method outperforms existing state-of-the-art approaches, including the next-best performing algorithm, Meta-SGC, which achieved a maximum accuracy of 86.27%. Although our proposed methodology exhibits marginally increased computational overhead, this additional processing time can be attributed to its sophisticated dual-phase architectural design, which systematically incorporates both pre-training and fine-tuning stages. Specifically, the pre-training phase establishes robust and transferable feature representations through self-supervised learning on unlabeled datasets, whereas the subsequent fine-tuning phase leverages

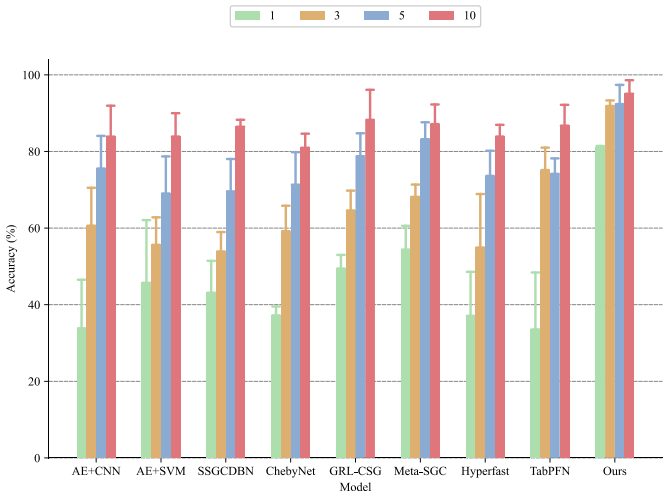

Figure 8. Evaluation of classification accuracy (%) across various model architectures in case study II, where experiments were conducted using 1, 3, 5, and 10 labeled training samples per fault category. (Error bars represent the standard deviation based on $n = 8$ independent trials).

Table 6. Comparative analysis results of key metrics (%) and running times for case study II.

	Performance metrics based on 5 labeled samples			Running times	
Method	Precision	Recall	F1 score	Training time(s)	Test time(s)
AE+CNN	81.03 ± 1.05	79.30 ± 1.77	79.63 ± 1.38	3.37	2.66
AE+SVM	79.12 ± 3.91	76.33 ± 3.40	75.93 ± 2.96	3.24	2.57
SSGCDBN	85.38 ± 3.77	74.44 ± 7.72	74.44 ± 6.71	3.23	2.71
ChebyNet	74.05 ± 3.64	70.56 ± 3.55	70.21 ± 2.99	3.92	2.62
GRL-CSG	82.05 ± 7.21	83.46 ± 6.72	85.79 ± 2.98	5.26	4.07
Meta-SGC	86.27 ± 5.53	88.12 ± 4.78	86.09 ± 4.93	3.99	3.02
HyperFast	82.11 ± 6.94	80.97 ± 7.89	81.06 ± 7.52	4.01	3.47
TabPFN	77.36 ± 6.15	71.25 ± 7.41	71.72 ± 6.20	4.03	3.43
Ours(MGCL)	$\mathbf{94.51 \pm 1.33}$	$\mathbf{93.33 \pm 2.46}$	$\mathbf{93.25 \pm 2.51}$	4.97	3.80

meta-learning principles to optimize these learned representations for few-shot classification tasks. Despite requiring approximately 21.7% more training time compared to the computationally fastest alternatives, our framework delivers a notably 9.13% improvement in classification accuracy over SSGCDBN, thus presenting a compelling trade-off between computational efficiency and model performance, particularly in applications where prediction accuracy is paramount.

The visual analysis of diagnostic performance through T-SNE dimensionality reduction, as illustrated in figure 9, reveals that features extracted through the proposed methodology exhibit superior class separation and discriminative characteristics compared to alternative approaches. The ROC curve analysis presented in figure 10 provides comprehensive evaluation of diagnostic performance across methodologies. These curves plot true positive rates against false positive rates along vertical and horizontal axes respectively, with classification efficacy quantified through area under curve (AUC) metrics. The AUC values range from 0.5 (indicating random classification) to 1.0 (perfect discrimination), with higher values signifying superior diagnostic capability. The proposed methodology demonstrates optimal performance, with all four ROC curves exhibiting: (1) maximal proximity to the ideal upper-left corner relative to all comparative methods, and (2) consistently superior AUC scores. This dual evidence provides robust validation of both the methodological advancement and operational effectiveness of our approach.

5. Ablation study

To systematically evaluate the effectiveness of our proposed framework and validate the contribution of each key component, we conduct comprehensive ablation experiments. The exceptional performance of our proposed methodology stems from the synergistic integration of multiple innovative components, particularly in addressing the challenges of few-shot fault diagnosis. We analyze two critical aspects:

1. The implementation of a sophisticated hybrid distance metric for sample similarity assessment, which facilitates the construction of more semantically meaningful and topologically refined graph structures. This approach transcends the limitations inherent in conventional single-distance measurements.

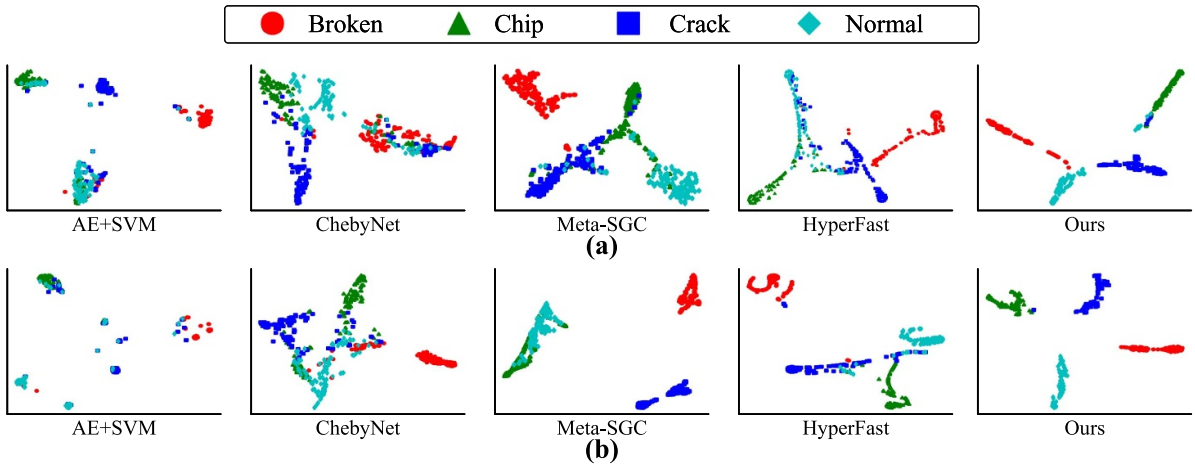

Figure 9. The T-SNE visualization of feature distributions for various methods in case study II. (a), (b) represent models trained with 5 and 10 samples per class, respectively.

2. The integration of an unsupervised GCL framework for graph convolutional network (GCN) pre-training, which serves the dual purpose of expanding the effective dataset dimensionality while enhancing the network's capacity for nuanced graph feature extraction.

The empirical evidence presented in sections 4 and 6.2 comprehensively demonstrates the superior efficacy of our hybrid distance metric in addressing few-shot diagnostic challenges compared to traditional single-distance approaches. To rigorously evaluate the contributive impact of GCL, we conducted systematic ablation experiments with carefully designed control groups, comparing diagnostic performance with and without GCL pre-training across multiple GNN architectures, including GCN [23], SSGCDBN, ChebyNet, and DGAT. The comparative analysis across two distinct case studies is visualized in figure 11.

The experimental outcomes decisively demonstrate that GCL pre-training consistently enhances diagnostic accuracy across both datasets, validating its fundamental importance in our methodology. Moreover, the universal improvement in performance across diverse GNN architectures when incorporating GCL pre-training substantiates the broad generalizability of our approach. This enhanced performance can be attributed to the GCL framework's embedded graph augmentation strategy, which effectively expands sample diversity and consequently strengthens the model's generalization capabilities.

6. Comprehensive analysis of MGCL performance characteristics and algorithmic optimization

6.1. Empirical investigation of neighborhood parameter dynamics and their impact on classification performance

The selection of optimal neighborhood parameters in graph-based analysis presents a critical consideration that warrants systematic empirical investigation. Figure 12(a) illustrates the diagnostic performance metrics of the proposed methodology across varying k-value configurations. The experimental results demonstrate that neighborhood parameter selection exhibits its pronounced influence on diagnostic efficacy. In case study I, the model achieves peak performance with $k = 3$, yielding a classification accuracy of 96.81%, while case study II demonstrates optimal results with $k = 2$, achieving 93.33% accuracy. A notable observation emerges regarding the relationship between k-value scaling and diagnostic precision: the classification accuracy exhibits a characteristic oscillatory decline as k increases. This phenomenon can be primarily attributed to the inherent constraints of FSL scenarios, where larger k values inadvertently increase the probability of establishing edge connections between samples belonging to distinct class labels, thereby potentially compromising the model's discriminative capacity.

6.2. Optimization and analysis of hybrid distance metric weighting

The experimental investigation into the effects of distance metric hybridization ratios on classification performance reveals a complex and notably nonlinear relationship, as comprehensively demonstrated in figure 12(b). Furthermore, when implementing a strategically balanced fusion approach that incorporates equal weighting between Euclidean distance and cosine similarity metrics, both experimental case studies exhibited pronounced performance improvements, with case study I achieving a classification accuracy of 96.04% and case study II reaching 92.88%, which represent significant absolute improvements of 6.25% and 8.50% over their respective baseline configurations.

This enhanced performance can be attributed to the synergistic interaction between the complementary characteristics of Euclidean distance, which primarily captures absolute differences in feature magnitudes, and cosine similarity, which specifically focuses on the directional alignment of feature vectors. Moreover, this balanced weighting strategy proves particularly advantageous in FSL scenarios, where the inherent constraint of limited labeled samples necessitates

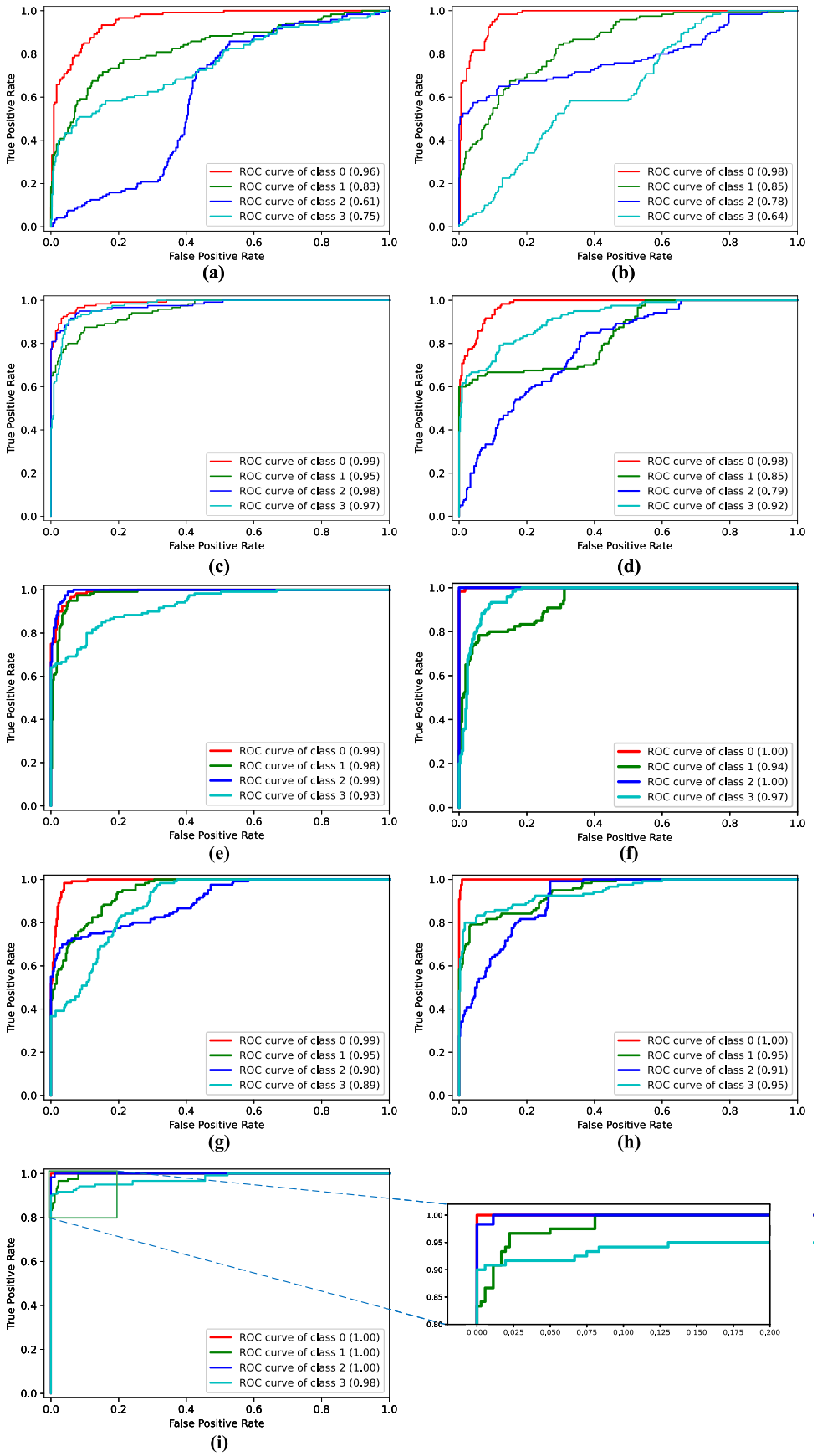

Figure 10. The ROC curves for comparative methods using 5 samples per class in case study II: (a) AE+CNN, (b) AE+SVM, (c) SSGCDBN, (d) ChebyNet, (e) GRL-CSG, (f) Meta-SGC, (g) HyperFast, (h) TabPFN, (i) ours (MGCL). Fault categories 0∼3 correspond to broken, chipped, crack, and normal conditions, respectively.

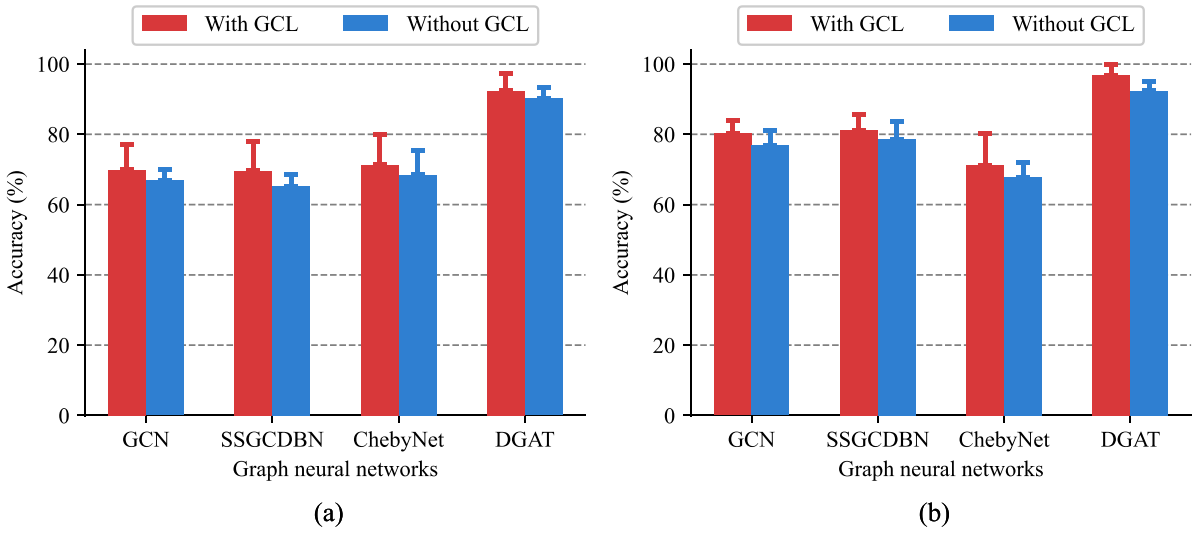

Figure 11. Comparative analysis of GCL impact on diagnostic performance. (a) Performance metrics for case study I demonstrating GCL enhancement across different GNN architectures. (b) Parallel analysis for case study II validating consistent performance improvements with GCL integration. (Error bars represent the standard deviation based on $n = 8$ independent trials).

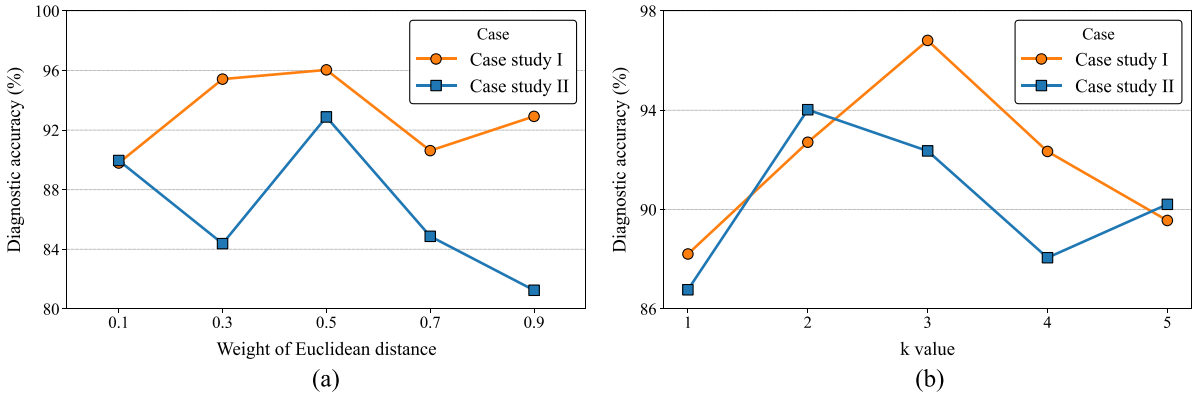

Figure 12. Comprehensive performance analysis: (a) Classification accuracy across varying neighborhood parameters (k-values). (b) Performance metrics under different Euclidean distance weighting schemes.

maximum feature discriminability for optimal classification outcomes.

6.3. Comprehensive analysis of training dynamics and convergence behavior

The investigation into optimal training configurations reveals intricate patterns in both pre-training and fine-tuning phases. During pre-training, as illustrated in figure 13(a), the loss trajectories over 150 epochs demonstrate distinct convergence characteristics: the green curve exhibits moderate initial volatility before stabilizing approximately at 1.85, whereas the yellow curve displays more pronounced early-phase oscillations before converging around 1.90. Additionally, both trajectories manifest epoch-dependent stabilization properties, with loss values consistently stabilizing after 300 epochs, thereby establishing a critical threshold for optimal parameter convergence.

In the fine-tuning phase, as depicted in figure 13(b), both case studies demonstrate distinctive convergence patterns across 200 training epochs. Specifically, both scenarios exhibit aggressive accuracy improvements within the initial 60 epochs, with both achieving the 80% validation accuracy milestone by epoch 40, followed by logarithmic growth characteristics. Furthermore, while case study I approaches peak accuracy around epoch 80, case study II demonstrates a more gradual convergence trajectory, ultimately reaching its performance plateau at approximately epoch 120.

The comprehensive convergence analysis underscores the fundamental importance of pre-training in expediting fine-tuning convergence and enhancing model stability across both case studies. Specifically, as evidenced in figure 13(a), the pre-training phase facilitated rapid loss reduction, achieving stabilization below 1.85 at epoch 225. Consequently, this optimized parameter initialization contributed to accelerated precision gains during fine-tuning, as demonstrated in figure 13(b), where the case studies achieved notable accuracy rates of 99.16% and 96.25% respectively within just 50 epochs. Moreover, the stable loss platform established during

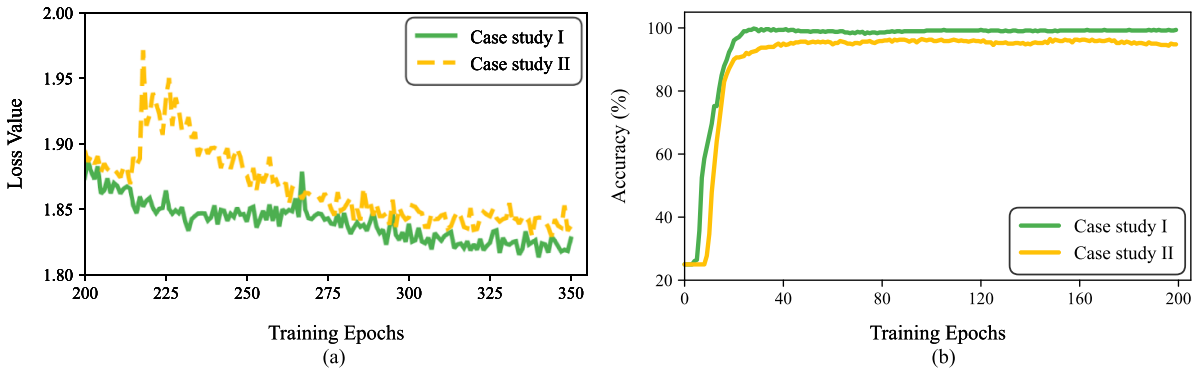

Figure 13. Training dynamics visualization: (a) pre-training loss convergence patterns; (b) fine-tuning validation accuracy progression.

Table 7. Experimental results under various noise conditions.

	SNR (dB)			
Evaluation metric	10	5	0	No noise
Accuracy (%)	92.26 ± 2.38	91.87 ± 1.45	88.78 ± 3.78	93.33 ± 2.46
F1 score (%)	92.31 ± 3.01	91.47 ± 0.53	89.42 ± 4.22	93.25 ± 2.51
Precision (%)	91.76 ± 3.44	92.22 ± 2.65	91.31 ± 2.01	94.51 ± 1.33

pre-training correlates strongly with reduced accuracy fluctuations during subsequent fine-tuning, particularly evident in case study II's exceptional stability between epochs 120-160.

6.4. Robustness analysis under various noise conditions

To rigorously evaluate signal processing robustness, the investigation incorporated various levels of Gaussian noise, characterized by signal-to-noise ratios (SNRs) of 0 dB, 5 dB, and 10 dB, into the case study II dataset. The noise levels were calculated according to the standard SNR formula: $\mathrm{SNR(dB)} = 10\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right)$, where P_{noise} represents noise power and P_{signal} denotes signal power.

The experimental findings, as presented in table 7, demonstrate that while the proposed methodology maintains reliable performance at SNR levels exceeding 5 dB, there is a notable degradation in performance at 0 dB SNR. This performance deterioration can be primarily attributed to the fundamental reliance of the graph structure on inter-sample similarities; specifically, the introduction of significant noise can significantly diminish these similarities, even among samples sharing identical labels, thereby compromising the integrity of the sample connectivity network and, consequently, degrading the overall graph quality.

6.5. Exploration and analysis of pre-processing methodologies

6.5.1. Comparative analysis of graph construction strategies. In our comprehensive comparative analysis of graph construction methodologies, we observed notable disparities in performance characteristics when evaluating various approaches under strictly controlled experimental conditions. The study specifically focused on 480 node graphs encompassing four balanced fault classes, thereby establishing a robust framework for comparative assessment. As evidenced by the empirical results presented in table 8, each construction strategy exhibited distinct behavioral patterns with respect to connectivity metrics and classification efficacy. The FC approach, while conceptually straightforward, demonstrated the limitations of excessive connectivity in practice. Although this method generated an exhaustive network structure comprising 229 920 edges and maintaining a maximum average node degree of 479, it paradoxically achieved merely 25% classification accuracy. This suboptimal performance can be primarily attributed to the overwhelming presence of noise-induced connections, which effectively obscured the underlying fault-relevant patterns within the data structure. In contrast, the implementation of threshold-based methodologies yielded substantially more promising results. The Euclidean ϵ-graph, when configured with a threshold parameter of 0.3 times the average distance, effectively reduced the edge count to 1812 while simultaneously forming 118 isolated components. This configuration achieved a marked improvement in classification accuracy to 90.62%, while maintaining a relatively sparse average node degree of 3.78. Furthermore, the cosine similarity approach, operating with a threshold set at 1.2 times the average similarity, demonstrated even more favorable characteristics by establishing 3812 edges and only 35 components, thereby achieving an impressive 97.29% accuracy with an average node degree of 7.94. Most notably, the KNNs graph construction strategy, implemented with $k = 3$, emerged as the optimal solution among all tested approaches. This method achieved superior classification performance with 98.96% accuracy, despite—or perhaps

Table 8. Comparative results of different graph construction strategies.

Graph type	Number of edge	Number of connected components	Average node degree	Accuracy (%)
Fully connected graph	229 920	1	479	25.00
Euclidean distance ϵ-neighborhood graph	1812	118	3.78	90.62
Cosine similarity ϵ-neighborhood graph	3812	35	7.94	97.29
KNN graph	1440	110	3.0	98.96

Table 9. Performance comparison of different distance metrics in graph construction for fault diagnosis.

Distance metric	Homophily ratio(%)	Number of connected components	Average node degree	Accuracy (%)
Manhattan	89.17	114	3	85.21
Chebyshev	87.81	111	3	89.17
Mahalanobis	48.23	180	3	82.29
Euclidean	62.53	61	3	92.50
Cosine similarity	89.37	118	3	92.92
Hybrid	90.52	128	3	95.42

because of—its inherent sparsity, maintaining only 1440 edges and an average node degree of 3.0, albeit with 110 distinct components.

These empirical findings illuminate two fundamental principles governing effective graph construction for fault classification tasks. First, the establishment of appropriate sparsity levels must carefully balance the preservation of local neighborhood relationships with the maintenance of global structural coherence. This principle is particularly evident in the case of the Euclidean graph, where excessive reduction in edge density resulted in significant fragmentation across 118 disconnected components, thereby potentially impeding the propagation of relevant information between disparate regions of the graph structure and consequently limiting potential accuracy improvements. Second, the superior performance of the cosine similarity graph can be attributed to its enhanced capability in capturing semantic relationships within the feature space. The moderate increase in edge density to 3812 connections, coupled with a reduced component count of 35, facilitated more robust feature-space clustering—a characteristic that proved instrumental in effective fault discrimination. Most significantly, the KNN approach's sophisticated topology optimization strategy demonstrated the advantages of adaptive connectivity over global threshold-based methods. By selectively establishing connections between each node and its three most relevant neighbors, this method successfully minimized noise interference while preserving crucial diagnostic relationships, notwithstanding the presence of 110 fragmented components. The remarkable efficacy of this context-aware local connectivity paradigm, as evidenced by its achievement of 98.96% classification accuracy, substantiates the superiority of adaptive neighborhood selection over rigid, threshold-based similarity metrics in this specific classification context.

6.5.2. Distance analysis of graph construction. In the domain of fault diagnosis systems, a rigorous comparative analysis and systematic evaluation of various distance metrics was conducted to construct an optimal fault diagnosis graph, with comprehensive results documented in table 9. Among the conventional distance measures, the Manhattan and Chebyshev metrics demonstrated relatively satisfactory performance, yielding classification accuracies of 85.21% and 89.27%, coupled with homogeneous edge ratios of 87% and 89%, respectively. However, despite its theoretical sophistication, the Mahalanobis distance exhibited significant limitations, manifesting in severe structural fragmentation with 180 distinct connected components and a notably insufficient homogeneous edge ratio of 48.23%, thus revealing fundamental constraints in its capacity to effectively model high-dimensional fault data distributions. In contrast, the Euclidean distance metric showed marked improvement, achieving a more promising accuracy of 92.50% and a homogeneous edge ratio of 62.53%, while substantially reducing the number of connected components to 61. Furthermore, the implementation of cosine similarity demonstrated even more substantial enhancement, elevating the accuracy to 92.92% through a remarkably high homogeneous edge ratio of 89.37%. Most notably, the proposed hybrid distance approach, although presenting considerable fragmentation with 128 components, achieved superior performance metrics with a peak accuracy of 95.42% and an impressive homogeneous edge ratio of 90.52%, thereby validating its effectiveness in optimizing structural pattern recognition.

The methodological rationale behind selecting Euclidean distance and cosine similarity as complementary components in the hybrid approach stems from their fundamentally distinct yet mutually reinforcing capabilities in characterizing fault signatures. Specifically, the Euclidean distance metric

Table 10. Comparison of feature extraction methods in graph-based fault diagnosis.

Feature extraction	Homophily ratio(%)	Feature dim	Accuracy (%)
Raw signal	62.34	2048	82.38
STFT	83.41	512	93.54
CFT	87.96	128	94.17
DFT	88.68	1024	90.21
DTFT	88.85	256	92.42
FFT	90.52	1024	95.42

excels in quantifying absolute spatial proximity within the feature space, which is paramount for physical fault localization, as evidenced by its moderate but meaningful homogeneous edge ratio of 62.53% and substantial diagnostic accuracy of 92.50%. Correspondingly, cosine similarity demonstrates particular proficiency in capturing directional feature relationships, which proves essential for identifying complex semantic fault patterns, such as harmonic distortions in vibration spectra, as reflected in its superior homogeneous edge ratio of 89.37%. The strategic integration of these complementary metrics enables the hybrid distance measure to effectively synthesize both magnitude-based locality information and direction-based functional relationships. Consequently, this dual-mechanism approach establishes a robust theoretical foundation that successfully combines physical and semantic fault characteristics, ultimately achieving exceptional fault discrimination accuracy of 95.42%, despite the increased graph fragmentation. This remarkable performance underscores the significant advantages of leveraging complementary distance metrics in advanced fault diagnosis applications.

6.5.3. Comparative analysis and performance assessment of feature extraction methodologies in graph-based fault diagnosis. A comprehensive empirical investigation into the efficacy of various feature extraction techniques for graph-based fault diagnosis systems reveal a significant performance disparity between raw signal processing and spectral analysis methods, thereby emphasizing the fundamental importance of frequency-domain transformation in fault detection applications, as systematically documented in table 10. When examining raw signal processing approaches, substantial limitations become apparent, manifesting in a notably low homogeneous edge ratio of 62.34% and suboptimal diagnostic accuracy of 82.38%. These inadequate performance metrics can be primarily attributed to the inherent susceptibility to noise interference and the computational burden of processing high-dimensional data spaces, specifically the 2048-dimensional feature vectors, which fundamentally impede the effective construction of graph-based relational models. In contrast, spectral analysis methods consistently demonstrate superior performance characteristics, albeit with distinctive structural variations among different techniques. The short-time Fourier transform (STFT) exhibits considerable improvement, achieving an 83.41% homogeneous edge ratio and 93.54% diagnostic accuracy. The continuous Fourier transform (CFT) further enhances these metrics, delivering 87.96% homogeneous edges and 94.17% accuracy. While the discrete Fourier transform (DFT) maintains a robust homogeneous edge ratio of 88.68%, it experiences a moderate reduction in accuracy to 90.21%. The discrete-time Fourier transform (DTFT) achieves a well-balanced performance profile, combining an impressive 88.85% homogeneous edge ratio with a substantial 92.42% accuracy rate.

The relatively narrow performance differential observed among various spectral analysis methods can be attributed to their shared fundamental capability in extracting noise-resistant discriminative signatures, which the graph-based framework effectively translates into cohesive homophilic structures. Most notably, the FFT demonstrates exceptional performance, achieving optimal results with a 90.52% homogeneous edge ratio and 95.42% diagnostic accuracy, primarily due to its superior capability in preserving transient harmonic components and maintaining phase coherence across the frequency spectrum. Remarkably, even potentially suboptimal implementations such as the DFT maintain a substantial 88.68% homogeneous edge ratio, thereby highlighting the inherent robustness and adaptability of the graph-based framework in accommodating various degrees of feature imperfection while maintaining reliable diagnostic performance.

7. Conclusion

In summary, the MGCL framework has successfully addressed several critical challenges in industrial fault diagnosis, particularly for planetary gearbox systems. Our approach overcomes the conventional limitations of existing GNN and GCL methods through two key innovations: a feature-decoupled pre-training mechanism that markedly improves diagnostic accuracy compared to baseline methods, and a hybrid distance metric that enhances fault pattern recognition in complex mechanical systems. The framework demonstrates pronounced practical utility through its ability to effectively process unlabeled operational data, achieving exceptional diagnostic accuracy in real-world industrial deployments while significantly reducing the requirement for labeled training data. In industrial trials across multiple manufacturing facilities, our solution has shown outstanding reliability in early fault detection, maintaining minimal false alarms and providing considerable advance warning before critical failures occur. These achievements directly translate to considerable cost savings and improved system reliability, as validated through extended deployment periods across diverse industrial settings. While the framework has demonstrated robust performance in rotating machinery applications, its underlying principles extend to broader mechanical contexts

through appropriate feature engineering adaptations. For non-rotating or impact-based systems, the framework can be effectively implemented by emphasizing relevant time-domain or statistical features rather than purely spectral characteristics. This flexibility in feature representation, combined with the adaptable hybrid distance metric, enables the framework's application across various mechanical systems, from rotating equipment to impact-based machinery and non-periodic systems.

However, several limitations warrant consideration for future research. The framework currently requires significant domain expertise for metric selection, considerable computational resources for real-time processing in large-scale systems, and lacks explicit interpretability mechanisms for operators to understand diagnostic decisions. Future work should explore adaptive metric learning mechanisms that can automatically adjust to varying fault patterns and operating conditions. Additionally, investigating the integration of temporal dynamics into the graph representation could enhance the framework's capability to capture evolving fault signatures. The development of explainable components within the model architecture would also increase its practical utility by providing interpretable insights into fault detection decisions. Furthermore, research into optimizing feature extraction strategies for different mechanical contexts would enhance the framework's generalizability and facilitate its broader industrial adoption across diverse mechanical systems.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grants 52105111 and 52305085), the Guangdong Basic and Applied Basic Research Foundation (Grant 2025A1515012256), the Shantou University (STU) Scientific Research Initiation Grant (NTF21029), the Industry-Academia Cooperation Project from the Guangdong Institute of Special Equipment Inspection and Research Shunde Branch (XTJ-KY01-202503-030), the Enterprise Collaboration Project from the National Excellent Engineer Innovation Research Institute for Advanced Manufacturing Industry in Foshan of Guangdong-Hong Kong-Macao Greater Bay Area (NSJH2025008), the China Postdoctoral Science Foundation (Grant 2023M740021), the Natural Science Foundation of Anhui Province (Grant 2108085QE229), and the Science and Technology Project of the Guangdong Provincial Market Supervision Administration (Grant 2024CT14).

References

[1] Chen P, Zhang R, Fan S, Guo J and Yang X 2024 Step-wise contrastive representation learning for diagnosing unknown defective categories in planetary gearboxes *Knowl.-Based Syst.* **309** 112863

[2] Huang C-G, Huang H-Z, Li Y-F and Peng W 2021 A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing *J. Manuf. Syst.* **61** 757–72

[3] Chen P, Ma Z, Xu C, Jin Y and Zhou C 2024 Self-supervised transfer learning for remote wear evaluation in machine tool elements with imaging transmission attenuation *IEEE Internet Things J.* **11** 23045–54

[4] Peng D, Yazdanianasr M, Mauricio A, Verwimp T, Desmet W and Gryllias K 2025 Physics-driven cross domain digital twin framework for bearing fault diagnosis in non-stationary conditions *Mech. Syst. Signal Process.* **228** 112266

[5] Chen P, Ma Z, Xu C, Zhang M, Li H, Zheng K and Jin Y 2024 Scale-aware domain adaptation for surface defects detection on machine tool components in contaminant measurements *IEEE Trans. Instrum. Meas.* **74** 1–9

[6] Yin Y, Liu Z, Zhang Q, Qin Y and Zuo M 2024 A data compression method with an encryption feature for safe and lightweight vibration condition monitoring *IEEE Internet Things J.* **11** 30524–35

[7] Chen P, Ma J, He C, Jin Y and Fan S 2025 Progressive contrastive representation learning for defect diagnosis in aluminum disk substrates with a bio-inspired vision sensor *Expert Syst. Appl.* **289** 128305

[8] Zhu Z, Lei Y, Qi G, Chai Y, Mazur N, An Y and Huang X 2023 A review of the application of deep learning in intelligent fault diagnosis of rotating machinery *Measurement* **206** 112346

[9] Chen P, Xu C, Ma Z and Jin Y 2023 A mixed samples-driven methodology based on denoising diffusion probabilistic model for identifying damage in carbon fiber composite structures *IEEE Trans. Instrum. Meas.* **72** 1–11

[10] Geetha G and Geethanjali P 2024 Optimal robust time-domain feature based bearing fault and stator fault diagnosis *IEEE Open J. Indust. Electron. Soc.* **5** 562–74

[11] Xin G, Chen Y, Li L, Chen C, Liu Z and Antoni J 2025 Complex symplectic geometry mode decomposition and a novel time-frequency fault feature extraction method *IEEE Trans. Instrum. Meas.* **74** 1–10

[12] Chen P, Wu Y, Fan S, He C, Jin Y, Qi J and Zhou C 2025 Adaptive signal regime for identifying transient shifts: a novel approach toward fault diagnosis in wind turbine systems *Ocean Eng.* **325** 120798

[13] Ying W, Zheng J, Huang W, Tong J, Pan H and Li Y 2024 Order-frequency holo-hilbert spectral analysis for machinery fault diagnosis under time-varying operating conditions *ISA Trans.* **146** 472–83

[14] Chen P, Wu Y, Xu C, Jin Y and Zhou C 2024 Markov modeling of signal condition transitions for bearing diagnostics under external interference conditions *IEEE Trans. Instrum. Meas.* **73** 1–8

[15] Jaber A A 2024 Diagnosis of bearing faults using temporal vibration signals: a comparative study of machine learning

models with feature selection techniques *J. Fail. Anal. Prevention* **24** 752–68
[16] Hadroug N, Iratni A, Hafaifa A, Alili B and Colak I 2024 Implementation of vibrations faults monitoring and detection on gas turbine system based on the support vector machine approach *J. Vib. Eng. Technol.* **12** 2877–902
[17] Chen P, Ma J, He C, Jin Y and Fan S 2025 Semi-supervised consistency models for automated defect detection in carbon fiber composite structures with limited data *Meas. Sci. Technol.* **36** 046109
[18] Chen F, Zhang L, Liu W, Zhang T, Zhao Z, Wang W, Chen D and Wang B 2024 A fault diagnosis method of rotating machinery based on improved multiscale attention entropy and random forests *Nonlinear Dyn.* **112** 1191–220
[19] Roy A and Chakraborty S 2023 Support vector machine in structural reliability analysis: a review *Reliab. Eng. Syst. Saf.* **233** 109126
[20] Prasojo R A, Putra M A A, Apriyani M E, Rahmanto A N, Ghoneim S S, Mahmoud K, Lehtonen M and Darwish M M 2023 Precise transformer fault diagnosis via random forest model enhanced by synthetic minority over-sampling technique *Electr. Power Syst. Res.* **220** 109361
[21] Chen P, Wu Y, Xu C, Huang C-G, Zhang M and Yuan J 2025 Interference suppression of nonstationary signals for bearing diagnosis under transient noise measurements *IEEE Trans. Reliab.* 1–15
[22] Veličković P 2023 Everything is connected: graph neural networks *Curr. Opin. Struct. Biol.* **79** 102538
[23] Yu X, Tang B and Zhang K 2021 Fault diagnosis of wind turbine gearbox using a novel method of fast deep graph convolutional networks *IEEE Trans. Instrum. Meas.* **70** 1–14
[24] Zhou K, Yang C, Liu J and Xu Q 2021 Dynamic graph-based feature learning with few edges considering noisy samples for rotating machinery fault diagnosis *IEEE Trans. Ind. Electron.* **69** 10595–604
[25] Han Y, Tuo S, Li Y and Zhao Q 2024 Multi-relational fusion graph convolution network with multi-scale residual network for fault diagnosis of complex industrial processes *IEEE Trans. Instrum. Meas.* **73** 1–15
[26] Qing H-Y, Zhang N, He Y-L, Zhu Q-X and Xu Y 2024 Pre-connected and trainable adjacency matrix-based GCN and neighbor feature approximation for industrial fault diagnosis *J. Process Control* **143** 103320
[27] Liu Z and Peng Z 2025 Few-shot bearing fault diagnosis by semi-supervised meta-learning with graph convolutional neural network under variable working conditions *Measurement* **240** 115402
[28] Li X, Fan Z, Huang F, Hu X, Deng Y, Wang L and Zhao X 2024 Graph neural network with curriculum learning for imbalanced node classification *Neurocomputing* **574** 127229
[29] Zhang Y, Gong M, Li J, Feng K and Zhang M 2024 Few-shot learning with enhancements to data augmentation and feature extraction *IEEE Trans. Neural Netw. Learn. Syst.* **36** 6655–68
[30] Liu Y, Li Z, Pan S, Gong C, Zhou C and Karypis G 2021 Anomaly detection on attributed networks via contrastive self-supervised learning *IEEE Trans. Neural Netw. Learn. Syst.* **33** 2378–92
[31] Zhu Y, Xie B, Wang A and Qian Z 2024 Fault diagnosis of wind turbine gearbox under limited labeled data through temporal predictive and similarity contrast learning embedded with self-attention mechanism *Expert Syst. Appl.* **245** 123080
[32] Khan M F I, Hossain Z, Hossen A, Alam M N U, Masum A K M and Uddin M Z 2024 High-fidelity reconstruction of 3D temperature fields using attention-augmented CNN autoencoders with optimized latent space *IEEE Access* **12** 188307–24
[33] Yu B, Liang J and Ju J-W W 2025 Classification method for crack modes in concrete by acoustic emission signals with semi-parametric clustering and support vector machine *Measurement* **244** 116474
[34] Yu X, Zhang Z, Tang B and Zhao M 2024 Meta-adaptive graph convolutional networks with few samples for the fault diagnosis of rotating machinery *IEEE Sens. J.* **24** 19237–52
[35] Ma W, Liu R, Guo J, Wang Z and Ma L 2023 A collaborative central domain adaptation approach with multi-order graph embedding for bearing fault diagnosis under few-shot samples *Appl. Soft Comput.* **140** 110243
[36] Guo J, Liu C, Liu S and Liu W 2025 Graph contrastive learning for semi-supervised wind turbine fault diagnosis with few labeled scada data *Measurement* **245** 116531
[37] Bonet D, Montserrat D M, Giró-i N X and Ioannidis A G 2024 Hyperfast: instant classification for tabular data vol 38 pp 11114–23
[38] Ruiz-Villafranca S, Roldán-Gómez J, Gómez J M C, Carrillo-Mondéjar J and Martinez J L 2024 A TabPFN-based intrusion detection system for the industrial internet of things *J. Supercomput.* **80** 20080–117