Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Adaptive Graph Construction and Feature-Enhanced Graph Neural Network for Few-Shot Bearing Fault Diagnosis under Variable Conditions

Zhi Sun
Tsinghua University
Xiaoyu Deng
Chongqing University
Chunlin Zhang
Chongqing University
Yibo Luo
Chongqing University
Jinglei Zhao
jlzhao@cqu.edu.cn
Chongqing University

Research Article

Keywords: Fault diagnosis, Graph neural network, Few-shot learning, Self-supervised learning

Posted Date: July 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10407371/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Adaptive Graph Construction and Feature-Enhanced Graph Neural Network for Few-Shot Bearing Fault Diagnosis under Variable Conditions

Zhi Sun[a], Xiaoyu Deng[b], Chunlin Zhang[b], Yibo Luo[b], Jinglei Zhao[b,*]

[a]*Department of Computer Science and Technology, Tsinghua University, Tsinghua University, Haidian District, Beijing, 100084, China*
[b]*State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, No. 174 Shazheng Street, Shapingba District, Chongqing, 400044, China*

Abstract

To address the issue of insufficient cross-condition generalization capability caused by the scarcity of fault samples and the variability of operating conditions in real industrial scenarios, this paper proposes an Adaptive Graph Construction and Feature-Enhanced Graph Neural Network (AG-MS-GNN) for few-shot bearing fault diagnosis under variable operating conditions. The proposed method consists of three core modules. First, an adaptive physical graph construction mechanism is introduced, which integrates time-domain, frequency-domain, and demodulation-domain features of vibration signals to build a physically interpretable initial graph structure, and incorporates a learnable graph optimization layer to achieve end-to-end collaborative adaptation between the graph topology and the diagnostic task. Second, a negative-sample-free graph-level contrastive feature enhancement strategy is designed, which applies semantically preserved data augmentation to the original signals and performs view comparisons in the feature space, thereby avoiding the complexity of negative sample construction in traditional contrastive learning and effectively utilizing unlabeled data to enhance feature representation under few-shot conditions. Finally, a unified diagnostic model integrating multi-scale graph convolution and attention mechanism is constructed, which captures both local details and global contextual information while

*Corresponding author.
Email addresses: sunzhi@mail.tsinghua.edu.cn (Zhi Sun), dengxy@stu.cqu.edu.cn (Xiaoyu Deng), zhangcl19870122@163.com (Chunlin Zhang), luoyibo@stu.cqu.edu.cn (Yibo Luo), jlzhao@cqu.edu.cn (Jinglei Zhao)

dynamically focusing on fault-sensitive feature channels. Experimental results on the CWRU bearing dataset demonstrate that the proposed method achieves a fault recognition accuracy of 95.5% under extreme few-shot and cross-condition scenarios, significantly outperforming baseline methods including Support Vector Machine, one-dimensional Convolutional Neural Network, Graph Convolutional Network, and Graph Attention Network. Ablation studies verify the effectiveness and synergistic necessity of the adaptive graph construction, multi-head graph attention, residual connections, and self-supervised enhancement modules. By integrating physically driven adaptive graph construction with negative-sample-free self-supervised feature enhancement, the proposed AG-MS-GNN method effectively resolves the generalization challenges arising from the coupling of sample scarcity and operating condition variations in bearing fault diagnosis, providing a high-accuracy and robust technical solution for intelligent fault diagnosis of rotating machinery.

Keywords: Fault diagnosis, Graph neural network, Few-shot learning, Self-supervised learning

1. Introduction

The intelligent fault diagnosis of mechanical equipment is of vital importance for ensuring the safe, stable, and economical operation of modern industrial systems [1]. With the widespread adoption of sensor technology and the advent of the industrial big data era, data-driven fault diagnosis methods, particularly deep learning-based models, have become the mainstream research direction in this field due to their powerful automatic feature extraction and complex pattern recognition capabilities [2, 3]. Early research mainly focused on utilizing convolutional neural networks to process time-frequency images of vibration signals [4, 5], or employing recurrent neural networks and their variants to mine the temporal dependencies of signals [6]. These methods have achieved remarkable results under conditions of sufficient training data and stable operating conditions [7, 8].

However, the actual operating environment of rotating machinery (e.g., bearings and gears) is complex and harsh, causing ideal data-driven models to face three intertwined core challenges in practical deployment: (1) Scarcity of fault samples: Equipment operates in a normal state for the vast majority of the time, while specific types of fault samples, especially those of severe or compound faults, are costly to acquire and rarely observed, resulting in highly imbalanced labeled datasets and constituting a typical few-shot learning problem [9, 10]. (2)

Time-varying operating conditions: Operating parameters such as load and rotational speed change frequently according to production demands, leading to shifts in the statistical distribution of vibration signals. Consequently, models trained under fixed operating conditions suffer from a severe domain adaptation problem, i.e., a sharp decline in generalization performance under new operating conditions [11, 12]. (3) Severe signal noise interference: The complex background noise in industrial environments can mask and distort weak early fault features, imposing extremely high requirements on model robustness [13].

To overcome the limitations of traditional deep learning models in capturing inter-sample correlations, the Graph Neural Network (GNN), as a novel paradigm specifically designed for processing non-Euclidean graph-structured data, has been introduced into the field of fault diagnosis in recent years [14, 15]. The core idea of GNN is to construct a graph structure with data samples as nodes and some similarity or relationship between samples as edges, and to iteratively perform neighborhood information aggregation (i.e., graph convolution), thereby learning node representations that embed sample relationship information [16, 17]. Compared to Convolutional Neural Network/Recurrent Neural Network models that process each sample independently, GNNs can explicitly leverage structural similarities between samples for reasoning, a characteristic that endows them with unique potential in scenarios with data scarcity or domain shift [18, 19].

In terms of model architecture and graph construction, researchers have devoted efforts to designing more powerful GNN variants to enhance feature representation capabilities. For example, the Graph Attention Network (GAT) introduces an attention mechanism to assign differentiated weights to different neighboring nodes, thereby strengthening the model's ability to focus on important features [20]. To address the issue that fixed graph structures cannot reflect deep semantic relationships, methods such as dynamic graph neural networks and hypergraph neural networks have been proposed to learn or construct more optimized topological structures [21, 22]. However, when constructing sample relationship graphs, most existing methods rely on simple distance metrics (e.g., Euclidean distance, cosine similarity) or predefined rules, failing to adequately incorporate the physical prior knowledge specific to vibration signals (e.g., fault characteristic frequencies, resonance frequency bands), which limits the capacity of the graph structure to carry fault-discriminative information [23, 24].

In the context of few-shot and cross-condition fault diagnosis, the combination of meta-learning frameworks with GNNs has become a popular approach to address the label scarcity problem. Some works design metric-based meta-learning GNNs, which perform training at the task level to enable the model to rapidly

adapt to new fault categories [25]. Meanwhile, to tackle the challenge of varying operating conditions, domain-adaptive GNNs learn domain-invariant features by minimizing the distribution discrepancy between the source and target domains or employing adversarial training strategies [26, 27]. Despite these advances, most studies treat few-shot learning and cross-condition diagnosis as relatively independent problems. In real industrial scenarios, however, these two challenges often co-occur and exacerbate each other, requiring diagnosis under extreme conditions where labeled samples in the target domain are extremely scarce or even completely absent. This imposes a more severe test on the inductive bias and generalization capability of the model [28, 29].

In terms of utilizing unlabeled data, self-supervised learning offers the potential to mine the intrinsic patterns within massive amounts of unlabeled vibration data. Graph-based contrastive learning, which constructs positive and negative sample pairs to drive the model to learn robust representations that are invariant to data augmentation, has emerged as an effective pre-training strategy [30, 31]. However, existing graph contrastive learning methods are highly dependent on the quality of negative sample construction, and improper negative sample selection can lead to model performance degradation, a problem known as the negative sample trap [32]. How to design a self-supervised graph learning mechanism that avoids complex negative sample construction to efficiently leverage unlabeled data for enhancing diagnostic performance under few-shot conditions still requires further in-depth exploration.

In summary, although GNNs have brought new insights to fault diagnosis, existing research still suffers from deficiencies in providing a synergistic solution to the dual challenges of few-shot learning and variable operating conditions. These deficiencies are manifested in the following aspects: (1) The graph construction process is disconnected from the physical mechanisms of vibration signals, and static graph structures struggle to adaptively characterize sample relationships under different operating conditions; (2) Existing methods typically address data scarcity or domain shift in isolation, lacking a unified framework to synergistically optimize model performance under the stringent setting of few-shot cross-condition diagnosis; (3) When utilizing unlabeled data, they fail to effectively circumvent the inherent difficulties of negative sample construction in contrastive learning. To address these issues, this paper proposes an adaptive physical graph construction and feature-enhanced graph neural network for few-shot cross-condition fault diagnosis.

The remainder of this paper is structured as follows: Section 2 elaborates on the mathematical definitions and core module designs of the proposed method;

Section 3 presents the experimental setup, including datasets, comparison methods, and evaluation metrics, and systematically analyzes the experimental results; Section 4 discusses the advantages, limitations, and practical application prospects of the proposed method; Section 5 summarizes the paper and outlines future research directions.

2. Methodology

To address the key challenges of data scarcity, insufficient feature robustness, and weak cross-condition generalization capability in fault diagnosis under few-shot and variable operating conditions, this paper proposes a fault diagnosis method based on a Graph Neural Network (GNN) that integrates adaptive physical graph construction and dual-view contrastive self-supervised enhancement. This section systematically presents the mathematical definitions, overall framework, core module designs, loss function construction, and training optimization process of the proposed method. The overall framework is illustrated in Figure 1.

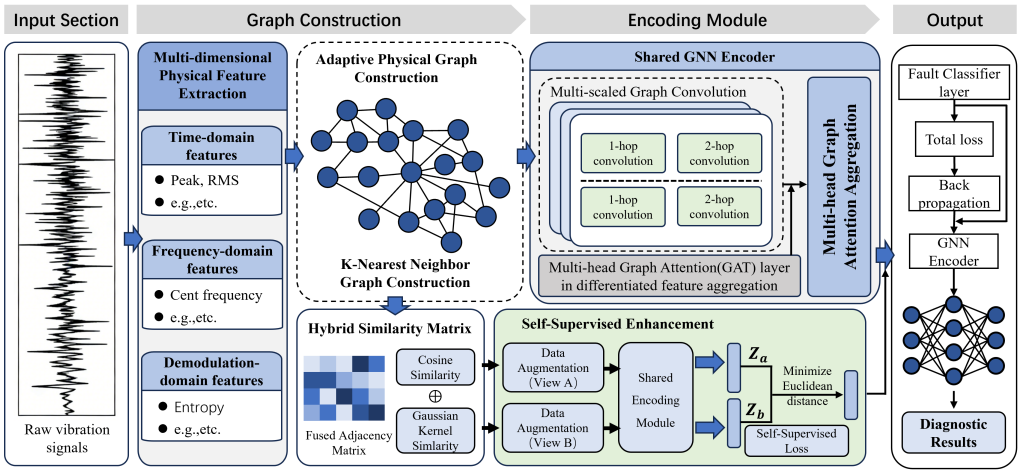


Figure 1: Overall framework of the proposed AG-MS-GNN method.

2.1. Problem Formulation

Let the set of operating conditions be $\mathcal{C} = \{c_1, c_2, \ldots, c_K\}$, containing different combinations of rotational speeds and loads. The set of fault types is

$\mathcal{Y} = \{y_1, y_2, \ldots, y_C\}$, where C is the number of categories including the normal state. The few-shot scenario is defined such that the number of training samples for each fault category is extremely small. Specifically, in the training set $\mathcal{D}_{tr} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_{tr}}$, $\mathbf{x}_i \in \mathbb{R}^L$ is a one-dimensional vibration time-series signal, L is the sampling length, and $N_{tr} \ll N_{full}$ (where N_{full} is the number of samples under sufficient data conditions). The test set $\mathcal{D}_{te} = \{(\mathbf{x}_j, y_j)\}_{j=1}^{N_{te}}$ contains samples to be diagnosed under varying operating conditions.

The objective is to learn an end-to-end mapping model $f : \mathbb{R}^L \to \mathcal{Y}$ that can accurately map test samples $\mathbf{x}_j$ to their corresponding fault categories y_j under few-shot training conditions, while maintaining high robustness and generalization capability under variable operating conditions. The proposed fault diagnosis model is an end-to-end framework integrating self-supervised and supervised learning, consisting of five core modules: (1) mining fault correlations among samples based on multi-dimensional physical features to generate a dynamically adjustable graph topology; (2) generating augmented views of original samples with the same source but different appearances to support self-supervised learning; (3) progressively refining deep fault features through multi-scale convolution iteration and graph attention mechanism; (4) optimizing the operating condition invariance of features through a shared encoder and projection head; (5) integrating supervised classification loss and self-supervised contrastive loss to achieve global optimization of model parameters.

2.2. Data Preprocessing

Traditional graph construction methods are susceptible to noise and operating condition interference and lack physical interpretability. This paper designs an adaptive graph structure based on multi-physical features to characterize the health state correlations among samples.

2.2.1. Multi-dimensional Physical Feature Extraction

For each sample $\mathbf{x}_i$, three types of features with clear physical meanings are extracted to comprehensively characterize the equipment state:

- **Time-domain features**: peak value, root mean square, etc., reflecting the amplitude distribution and energy characteristics of the signal, denoted as $\mathbf{f}_i^{\text{time}}$;
- **Frequency-domain features**: fault frequency band energy ratio, center frequency, etc., extracted after FFT transformation to capture characteristic frequency components, denoted as $\mathbf{f}_i^{\text{freq}}$;

- **Demodulation-domain features**: peak value, entropy, etc., of the envelope spectrum obtained via Hilbert transform, enhancing the representation of weak early faults, denoted as $\mathbf{f}_i^{\text{demod}}$.

After fusion, the comprehensive feature vector is obtained as $\mathbf{f}_i = [\mathbf{f}_i^{\text{time}}; \mathbf{f}_i^{\text{freq}}; \mathbf{f}_i^{\text{demod}}]^\top$, which is then normalized to $\hat{\mathbf{f}}_i$.

2.2.2. Label Encoding and Dataset Partition

The fault labels of the dataset are discrete categorical labels that cannot be directly input into the model for loss calculation. Therefore, label encoding is applied to map the text category labels to consecutive integer labels, adapting to the requirements of the cross-entropy loss function. To avoid model diagnostic bias caused by class imbalance and ensure model generalization, a stratified sampling strategy is adopted to divide the preprocessed dataset into training, validation, and test sets. During the partitioning process, the randomness of the distribution of samples from each category across the three datasets is ensured, guaranteeing the effectiveness of model training and validation.

2.3. Adaptive Graph Construction Module

The performance of graph neural networks is highly dependent on the rationality of the graph topology. An unreasonable topology can lead to bias in node feature aggregation, reducing fault diagnosis accuracy. Traditional graph construction methods only use a single similarity metric, making it difficult to comprehensively capture the complex correlations among samples. To this end, this paper proposes a hybrid similarity graph construction strategy that fuses cosine similarity and Gaussian kernel similarity, balancing the global distribution differences and local neighborhood correlations of samples. The specific implementation steps are as follows.

2.3.1. Hybrid Similarity Calculation

(1) Cosine Similarity: It is mainly used to capture the directional similarity of sample features, measuring the global distribution differences of different samples in the feature space. The calculation formula is:

$$s_{ij}^{\cos} = \frac{\mathbf{f}_i \cdot \mathbf{f}_j}{\|\mathbf{f}_i\| \cdot \|\mathbf{f}_j\|} \tag{1}$$

where $\mathbf{f}_i$ and $\mathbf{f}_j$ are the feature vectors of the i-th and j-th samples, respectively, and $\|\mathbf{f}_i\|$ and $\|\mathbf{f}_j\|$ are their L_2 norms.

(2) Gaussian Kernel Similarity: It is mainly used to characterize the spatial distance of sample features, focusing on local neighborhood correlations and being more sensitive to local feature differences. The calculation formula is:

$$s_{ij}^{\text{gauss}} = \exp\left(-\frac{\|\mathbf{f}_i - \mathbf{f}_j\|^2}{2\sigma^2}\right) \tag{2}$$

where $\|\mathbf{f}_i - \mathbf{f}_j\|$ is the Euclidean distance between the feature vectors of the i-th and j-th samples, and σ is the Gaussian kernel bandwidth, adaptively determined using the median of the distance matrix to avoid bias caused by manually set parameters.

A weighted fusion strategy is adopted to combine cosine similarity and Gaussian kernel similarity, obtaining a comprehensive similarity matrix that balances global correlations and local details. Through experimental validation, the weight for cosine similarity is set to 0.7 and that for Gaussian kernel similarity to 0.3. The fusion formula is:

$$s_{ij}^{\text{fusion}} = 0.7 \times s_{ij}^{\cos} + 0.3 \times s_{ij}^{\text{gauss}} \tag{3}$$

2.3.2. Adaptive K-Nearest Neighbor Graph Construction

Based on the similarity matrix, the K-nearest neighbor strategy is adopted to construct the adjacency matrix:

$$\mathbf{A}_{ij} = \begin{cases} s_{ij}^{\text{fusion}}, & \text{if } j \in \mathcal{N}_K(i) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $\mathcal{N}_K(i)$ denotes the set of K nearest neighbors of sample i. To ensure the symmetry of the graph, the adjacency matrix is symmetrized:

$$\mathbf{A} = \frac{\mathbf{A} + \mathbf{A}^{\mathsf{T}}}{2} \tag{5}$$

To reduce redundant connections and lower computational complexity, the comprehensive similarity matrix is sparsified using the K-nearest neighbor algorithm. The $K = 10$ most similar nodes for each node are selected to construct the adjacency matrix (the K value is optimized through experiments). The adjacency matrix is then symmetrized to ensure graph symmetry and normalized to eliminate graph structure bias, ensuring the stability of the message passing process. The normalization formula is:

$$\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-1/2}(\mathbf{A} + \mathbf{I})\tilde{\mathbf{D}}^{-1/2} \tag{6}$$

where $\mathbf{I}$ is the identity matrix and $\tilde{\mathbf{D}}$ is the degree matrix of $(\mathbf{A} + \mathbf{I})$.

2.3.3. Dual-Branch Data Augmentation

To support self-supervised contrastive learning, two non-overlapping augmentation operators are applied to the original sample $\mathbf{x}$ to generate two views with the same source but different appearances. The augmentation principle is to only alter the surface morphology of the signal without destroying the core fault features, such as characteristic frequencies and amplitude distribution patterns. This paper adopts additive Gaussian noise and amplitude scaling:

$$\mathbf{x}_a = \mathcal{T}_1(\mathbf{x}) = \mathbf{x} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 0.01^2) \tag{7}$$

$$\mathbf{x}_b = \mathcal{T}_2(\mathbf{x}) = \alpha \cdot \mathbf{x}, \quad \alpha \in [0.9, 1.1] \tag{8}$$

where α is randomly sampled from a uniform distribution for each augmentation.

2.4. Encoding Module

A multi-scale graph convolution iteration with graph attention structure is adopted to extract deep features focusing on core faults. The two augmented views share the encoder parameters.

2.4.1. Multi-Scale Graph Convolution Module

To capture fault features under different receptive fields, this paper designs a multi-scale graph convolution layer. By stacking multiple graph convolution layers, multi-scale feature extraction is achieved. The basic graph convolution operation is defined as:

$$\mathbf{H}^{(l+1)} = \sigma\left(\hat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right) \tag{9}$$

where $\mathbf{H}^{(l)}$ is the feature matrix of the l-th layer, $\mathbf{W}^{(l)}$ is the learnable weight matrix, and $\sigma(\cdot)$ is the ReLU activation function.

To enhance gradient flow and feature reuse, residual connections are introduced:

$$\mathbf{H}^{(l+1)} = \sigma\left(\hat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)} + \mathbf{b}^{(l)}\right) \tag{10}$$

By stacking three layers of graph convolution, local, regional, and global structural information are captured respectively, achieving hierarchical extraction of multi-scale features as detailed below:

First convolution layer: 1-hop and 2-hop convolutions are used to mine local, regional, and global fault correlations among samples.

$$\mathbf{H}_1^{\text{1-hop}} = \text{ReLU}\left(\hat{\mathbf{A}}\mathbf{H}^0\mathbf{W}_1^1 + \mathbf{b}_1^1\right) \in \mathbb{R}^{N\times D_1} \tag{11}$$

$$\mathbf{H}_1^{\text{2-hop}} = \text{ReLU}\left(\hat{\mathbf{A}}^2\mathbf{H}^0\mathbf{W}_2^1 + \mathbf{b}_2^1\right) \in \mathbb{R}^{N\times D_1} \tag{12}$$

$$\mathbf{H}_1 = [\mathbf{H}_1^{\text{1-hop}}, \mathbf{H}_1^{\text{2-hop}}] \in \mathbb{R}^{N\times 2D_1} \tag{13}$$

Second convolution layer: Deep feature refinement, still using 1-hop and 2-hop convolutions, maintaining the same dimension as the previous layer.

$$\mathbf{H}_2^{\text{1-hop}} = \text{ReLU}\left(\hat{\mathbf{A}}\mathbf{H}_1\mathbf{W}_1^2 + \mathbf{b}_1^2\right) \in \mathbb{R}^{N\times D_1} \tag{14}$$

$$\mathbf{H}_2^{\text{2-hop}} = \text{ReLU}\left(\hat{\mathbf{A}}^2\mathbf{H}_1\mathbf{W}_2^2 + \mathbf{b}_2^2\right) \in \mathbb{R}^{N\times D_1} \tag{15}$$

$$\mathbf{H}_2 = [\mathbf{H}_2^{\text{1-hop}}, \mathbf{H}_2^{\text{2-hop}}] \in \mathbb{R}^{N\times 2D_1} \tag{16}$$

Third convolution layer: Global feature refinement, still using 1-hop and 2-hop convolutions, maintaining the same dimension as the previous layer.

$$\mathbf{H}_3^{\text{1-hop}} = \text{ReLU}\left(\hat{\mathbf{A}}\mathbf{H}_2\mathbf{W}_1^3 + \mathbf{b}_1^3\right) \in \mathbb{R}^{N\times D_1} \tag{17}$$

$$\mathbf{H}_3^{\text{2-hop}} = \text{ReLU}\left(\hat{\mathbf{A}}^2\mathbf{H}_2\mathbf{W}_2^3 + \mathbf{b}_2^3\right) \in \mathbb{R}^{N\times D_1} \tag{18}$$

$$\mathbf{H}_3 = [\mathbf{H}_3^{\text{1-hop}}, \mathbf{H}_3^{\text{2-hop}}] \in \mathbb{R}^{N\times 2D_1} \tag{19}$$

where $\mathbf{H}^0 = [\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \ldots, \hat{\mathbf{f}}_N]^\mathsf{T} \in \mathbb{R}^{N\times D_0}$ is the initial vibration feature matrix ($D_0 = 31$); $\mathbf{W}_1^1, \mathbf{W}_2^1 \in \mathbb{R}^{D_0\times D_1}$, $\mathbf{W}_1^2, \mathbf{W}_2^2 \in \mathbb{R}^{2D_1\times D_1}$, $\mathbf{W}_1^3, \mathbf{W}_2^3 \in \mathbb{R}^{2D_1\times D_1}$ are the learnable convolution weight matrices; $\mathbf{b}_1^1, \mathbf{b}_2^1, \mathbf{b}_1^2, \mathbf{b}_2^2, \mathbf{b}_1^3, \mathbf{b}_2^3$ are the bias terms; D_1 is the hidden layer dimension (set to 64).

2.4.2. Multi-Head Graph Attention Module

To enhance the model's ability to focus on important nodes, a graph attention layer is embedded on top of the graph convolution output $\mathbf{H}_3$ to achieve differentiated feature aggregation. The attention coefficients and weighted feature aggregation are calculated as follows:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left([\mathbf{h}_{3,i}\mathbf{W}_{\text{gat}}, \mathbf{h}_{3,j}\mathbf{W}_{\text{gat}}]\mathbf{a}\right)\right)}{\sum_{k\in\mathcal{N}_K(i)} \exp\left(\text{LeakyReLU}\left([\mathbf{h}_{3,i}\mathbf{W}_{\text{gat}}, \mathbf{h}_{3,k}\mathbf{W}_{\text{gat}}]\mathbf{a}\right)\right)} \tag{20}$$

$$\mathbf{F} = \text{ReLU}\left(\sum_{j\in\mathcal{N}_K(i)} \alpha_{ij} \cdot \mathbf{h}_{3,j}\mathbf{W}_{\text{gat}}\right) \in \mathbb{R}^{N\times D_2} \tag{21}$$

where $\mathbf{W}_{\text{gat}} \in \mathbb{R}^{2D_1\times D_2}$ is the attention weight matrix, $D_2 = 64$; $\mathbf{a} \in \mathbb{R}^{2D_2}$ is the attention vector; $\mathcal{N}_K(i)$ is the set of K nearest neighbors of node i.

2.4.3. View Contrastive Self-Supervised Enhancement

Self-supervised learning is employed to improve feature robustness, adapting to few-shot and variable operating condition scenarios. After the two augmented views $\mathbf{x}_a$ and $\mathbf{x}_b$ are processed through the above encoding module to extract features $\mathbf{F}_a$ and $\mathbf{F}_b$, they are mapped to a low-dimensional feature space through a shared projection head:

$$\mathbf{Z}_a = \text{ReLU}\left(\mathbf{F}_a\mathbf{W}_{\text{proj}} + \mathbf{b}_{\text{proj}}\right) \in \mathbb{R}^{N\times D_z} \tag{22}$$

$$\mathbf{Z}_b = \text{ReLU}\left(\mathbf{F}_b\mathbf{W}_{\text{proj}} + \mathbf{b}_{\text{proj}}\right) \in \mathbb{R}^{N\times D_z} \tag{23}$$

where $\mathbf{W}_{\text{proj}} \in \mathbb{R}^{D_2\times D_z}$ is the projection head weight matrix, $\mathbf{b}_{\text{proj}}$ is the bias term, and $D_z = 16$.

To achieve a negative-sample-free contrastive loss, the Euclidean distance between the features of the two augmented views of the same sample is minimized, driving the model to learn condition-invariant features. The self-supervised loss function is defined as:

$$\mathcal{L}_{\text{self-supervised}} = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{Z}_{a,i} - \mathbf{Z}_{b,i}\|_2^2 \tag{24}$$

where $\mathbf{Z}_{a,i}$ and $\mathbf{Z}_{b,i}$ are the projected features of the i-th sample under the two augmented views, and N is the number of samples. This loss function does not require the construction of negative samples, avoiding the performance degradation issues caused by improper negative sample selection in traditional contrastive learning.

2.4.4. Classifier and Joint Loss Function

Global pooling followed by a fully connected layer is adopted to map the enhanced features to fault category probabilities:

$$\mathbf{H}_{\text{graph}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{F}_i \tag{25}$$

$$\mathbf{P} = \text{Softmax}\left(\mathbf{H}_{\text{graph}}\mathbf{W}_{\text{cls}} + \mathbf{b}_{\text{cls}}\right) \in \mathbb{R}^{C} \tag{26}$$

where $\mathbf{W}_{\text{cls}} \in \mathbb{R}^{D_2 \times C}$ is the classifier weight matrix, $\mathbf{b}_{\text{cls}}$ is the bias term, and $\mathbf{P}$ is the class probability distribution.

Based on the predicted probability distribution and sample labels, the supervised loss function is defined as the cross-entropy loss:

$$\mathcal{L}_{\text{supervised}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(p_{i,c}) \tag{27}$$

where $y_{i,c}$ is the true label of sample i for class c (one-hot encoded), and $p_{i,c}$ is the corresponding predicted probability.

Finally, the supervised classification loss and the self-supervised contrastive loss are integrated to balance classification accuracy and feature robustness:

$$\mathcal{L}_{\text{total}} = \lambda \cdot \mathcal{L}_{\text{self-supervised}} + \mathcal{L}_{\text{supervised}} \tag{28}$$

where λ is a hyperparameter balancing the self-supervised loss and the supervised loss. Through experimental tuning, λ is set to 0.1.

3. Experiments and Analysis

To comprehensively verify the effectiveness of the proposed AG-MS-GNN method for few-shot and variable-condition bearing fault diagnosis, this section conducts multiple sets of comparative experiments, ablation experiments, feature visualization, and stability analysis based on the CWRU bearing dataset, systematically evaluating the model's diagnostic accuracy, generalization capability, and robustness under sample scarcity and varying operating conditions. All experiments are completed under the same hardware environment and parameter settings to ensure fair and reliable results.

3.1. Experimental Setup

3.1.1. Dataset and Preprocessing

The experiments are conducted on the publicly available bearing dataset from Case Western Reserve University (CWRU). The experimental object is an SKF6205 deep groove ball bearing. Single-point faults are introduced at the inner race, outer race, and rolling element positions using electrical discharge machining. Vibration signals are collected from the drive end at a sampling frequency of 12 kHz. The dataset contains 10 equipment states, including 1 normal state and 9 typical fault states. The fault states cover three fault locations (inner race, outer race, rolling element) with three fault diameters (0.007, 0.014, and 0.021 inches), comprehensively covering common bearing fault modes.

The data preprocessing follows a unified procedure. First, the long time-series vibration signals are segmented into independent samples with a fixed length and an appropriate overlap ratio to augment the number of samples. Then, for each sample, 14 time-domain features, 9 frequency-domain features, and 8 demodulation-domain features are extracted, forming a 31-dimensional initial physical feature vector. Finally, based on the normalized multi-dimensional features, the comprehensive similarity between samples is calculated, and a K-nearest neighbor strategy ($K = 10$) is adopted to construct an undirected weighted graph, followed by symmetric normalization, resulting in graph-structured data that can be directly input into the model. After complete preprocessing, a total of 2605 balanced effective samples are obtained. The distribution of each category of samples is shown in Table 1.

Table 1: Sample distribution of the CWRU dataset.

Category	**Fault Type**	**Number of Samples**
0	Normal	260
1	Inner Race Fault	261
2	Inner Race Fault	261
3	Inner Race Fault	261
4	Outer Race Fault	260
5	Outer Race Fault	261
6	Outer Race Fault	260
7	Rolling Element Fault	260
8	Rolling Element Fault	261
9	Rolling Element Fault	260
	Total	**2605**

3.1.2. Few-Shot Experimental Setup

To simulate the real industrial scenario of scarce fault samples, this paper constructs multiple experimental scenarios ranging from extreme scarcity to sufficient samples by controlling the proportion of the training set. The training set proportions are set to seven levels: 3%, 4%, 8%, 10%, 30%, 50%, and 70%. The validation set is uniformly set to 15%, and the remaining data serves as the test set. The 3% training proportion corresponds to approximately 8 samples per class, representing an extreme few-shot scenario. The 4% to 10% range represents typical few-shot scenarios, and 30% and above represent moderate to sufficient sample scenarios. These gradient sample settings allow clear observation of the model's performance variation under different degrees of data scarcity, providing a comprehensive measure of the method's few-shot learning capability and data utilization efficiency. The partition ratios are shown in Table 2.

Table 2: Experimental classification settings.

Exp. No.	Training Ratio	Training Samples	Samples per Class (avg.)	Scenario Description
1	3%	78	≈8	Extreme few-shot
2	4%	104	≈10	Very few-shot
3	8%	208	≈21	Moderate few-shot
4	10%	260	≈26	Slightly few-shot
5	30%	781	≈78	Moderate samples
6	50%	1302	≈130	Sufficient samples
7	70%	1823	≈182	Standard partition

3.1.3. Comparison Methods and Evaluation Metrics

Traditional machine learning methods and mainstream deep learning models are selected as baselines, including Support Vector Machine (SVM), one-dimensional Convolutional Neural Network (1D-CNN), and the classic Graph Convolutional Network (GCN). All baseline models are configured with their publicly recommended optimal parameters and retrained on the same dataset to ensure fairness. A brief description of each method is shown in Table 3.

3.2. Experimental Results and Analysis

3.2.1. Main Results: Overall Model Performance

The diagnostic accuracies of different methods under seven training proportions are shown in Figure 2. It can be observed that the proposed AG-MS-GNN significantly outperforms the comparison methods across all sample proportions,

Table 3: Comparison of different methods.

Method	Abbreviation	Description
Support Vector Machine	SVM	Traditional machine learning method
One-dimensional CNN	1D-CNN	Deep learning, raw signal input
Graph Convolutional Network	GCN	Classic GNN, fixed K-NN graph
Proposed Method	AG-MS-GNN	Adaptive graph + self-supervised enhancement

with the most prominent advantage in the few-shot regime. In the extreme scenario with only 3% training ratio, the accuracies of SVM, 1D-CNN, and GCN are only 48.57%, 43.30%, and 31.68%, respectively, failing to achieve reliable diagnosis, while AG-MS-GNN maintains a high accuracy of 95.55%. As the number of training samples increases, the performance of all models gradually improves, but the proposed method consistently maintains its lead, reaching 98.62% at 10% training ratio and stabilizing above 99.6% at 30% and above. In contrast, even under the sufficient sample condition of 70%, traditional models struggle to achieve the accuracy of the proposed method under extreme few-shot conditions, fully demonstrating the strong few-shot learning capability and condition robustness of AG-MS-GNN.

The overall curves show a clear pattern: AG-MS-GNN achieves the highest accuracy at all training proportions, with the largest improvement in the few-shot regime. From 3% to 70%, the proposed method outperforms the best baseline by an average of over 30 percentage points, fully verifying its effectiveness and superiority in few-shot variable-condition bearing fault diagnosis.

3.2.2. Ablation Study: Verification of Core Module Contributions

To verify the contributions of each core component (multi-head GAT, residual connections, data augmentation, etc.) of the AG-MS-GNN model to bearing fault diagnosis performance and to clarify the mechanism of each module, this paper designs a systematic ablation study. The experiments use the complete AG-MS-GNN model as the baseline and construct different ablation variants by removing or replacing core modules one by one. All experiments maintain identical datasets, training parameters, random seeds, and hardware environments to ensure fairness and reproducibility.

The ablation study focuses on the core functional modules of the model, with a total of 5 sets of comparative experiments configured as follows. The complete model retains all core components (multi-head GAT attention mechanism,

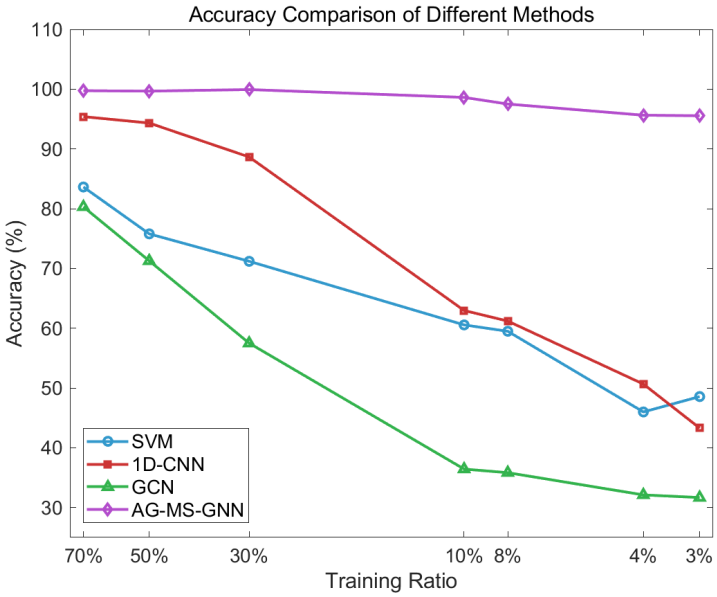


Figure 2: Accuracy comparison curves of different methods under different training proportions.

residual connection structure, and data augmentation strategy) as the performance baseline. Ablation variant 1 removes the multi-head GAT module from the complete model, retaining only the basic feature extraction and classification structure, to verify the role of the multi-head GAT attention mechanism in focusing on fault features. Ablation variant 2 removes the residual connection structure from the GAT layer and feature projection layer of the complete model to verify the role of residual connections in mitigating gradient vanishing in deep networks and improving model training stability. Ablation variant 3 disables the data augmentation strategy during training (no noise injection or feature scaling) to verify the role of data augmentation in improving model generalization capability and adapting to complex operating conditions. The basic model constructs the simplest baseline model, retaining only the basic GCN structure without multi-head GAT, residual connections, or data augmentation modules, to compare the performance gap between the complete model and the basic model, highlighting the advantages of multi-component collaborative design.

The experimental results under the 4% few-shot training condition are shown in Table 4 and Figures 3.

Based on the experimental results in Table 4 and Figures 3, the following

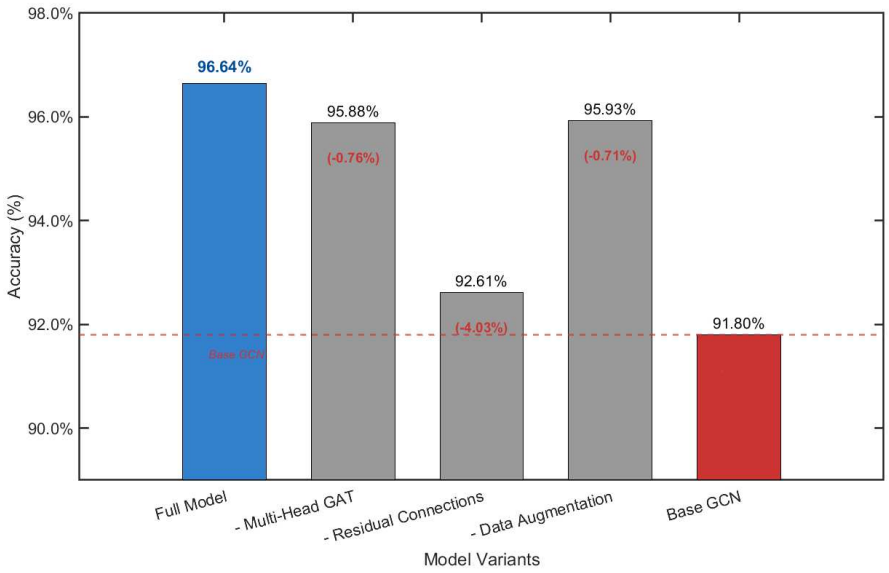


Figure 3: Accuracy comparison bar chart for different methods.

Table 4: Ablation experiment results under few-shot condition (4% training ratio).

Model Configuration	**Accuracy (%)**	**Decrease from Complete Model (%)**
Complete AG-MS-GNN (Baseline)	96.64	—
Ablation-Multi-Head GAT	95.88	-0.76
Ablation-Residual Connections	92.61	-4.03
Ablation-Data Augmentation	95.93	-0.71
Basic Single-Scale GCN	91.80	-4.84

analysis of each core module's role and performance impact is provided:

(1) Effectiveness of the multi-head GAT module: After removing the multi-head GAT module, the test accuracy decreased from 96.64% to 95.88%, a relative performance drop of 0.76%. This module adaptively focuses on key features of bearing faults, suppresses interference from irrelevant redundant information, and enhances the model's ability to recognize fault features. Especially in scenarios with overlapping complex fault features, it effectively distinguishes different fault types, providing support for high-precision diagnosis.

(2) Key role of residual connections: After removing residual connections, the test accuracy dropped significantly by 4.03% (from 96.64% to 92.61%), the most pronounced performance drop among all ablation experiments. This result indicates that residual connections are a key component for ensuring model training stability and mitigating the vanishing gradient problem in deep networks. During the feature extraction process of the AG-MS-GNN model, residual connections enable effective fusion of shallow and deep features, preventing the degradation of deep features and ensuring that the model fully learns multi-level features of bearing faults, thereby significantly improving fault diagnosis accuracy.

(3) Effectiveness of data augmentation: After disabling the data augmentation strategy, the test accuracy decreased from 96.64% to 95.93%, a relative performance drop of 0.71%. This indicates that data augmentation effectively expands the diversity of training data by introducing reasonable noise injection and feature scaling, simulates variable operating condition scenarios of bearing faults in actual operation, reduces the risk of overfitting to training data, improves the model's adaptability to unseen fault samples, and thus enhances generalization performance.

(4) Synergistic advantage of the complete model: The test accuracy of the basic single-scale GCN model is only 91.80%, a decrease of 4.84% compared to the complete AG-MS-GNN model, fully demonstrating the effectiveness of the synergistic collaboration of the three modules (multi-head GAT, residual connections, data augmentation).

The ablation experiment results show that multi-head GAT, residual connections, and data augmentation are all key effective components of the AG-MS-GNN model. Their synergistic collaboration significantly improves bearing fault diagnosis performance, with residual connections having the most significant impact on model performance, providing core support for high-precision diagnosis. Through the organic combination of its modules, the complete model achieves precise extraction of fault features, stable transmission of deep features, and enhanced model generalization, ultimately achieving a high-precision fault diagno-

sis accuracy of 96.64%, significantly outperforming the basic model and verifying the rationality and superiority of the AG-MS-GNN model design.

3.2.3. Classification Effect Analysis (Few-shot vs. Sufficient Samples)

Figures 4 and 5 show the diagnostic accuracy per class and confusion matrices under the 3% extreme few-shot and 70% sufficient sample conditions, respectively. Under the 3% training ratio, the proposed method still maintains high recognition rates for the normal state and most fault categories, with only slight confusion in a few subtle fault categories, achieving an overall accuracy of 95.55%. When the training ratio increases to 70%, the recognition accuracy for all categories approaches 100%, and the confusion matrix shows almost no misclassified samples, with the model achieving completely accurate fault classification. This result demonstrates that AG-MS-GNN not only has strong generalization capability under few-shot conditions but also maintains stable and efficient classification performance under sufficient sample conditions, making it suitable for varying data conditions in industrial scenarios.

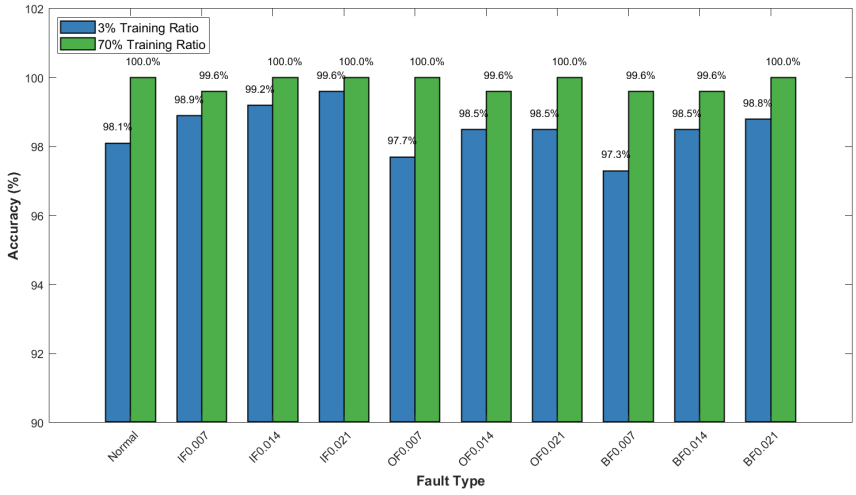


Figure 4: Accuracy comparison bar chart for different numbers of training samples.

3.2.4. Feature Visualization Analysis

To intuitively explain the feature learning effect of the model, t-SNE is adopted to reduce the high-dimensional features to two dimensions for visualization, with results shown in Figure 6. Under the 3% extreme few-shot condition, samples of

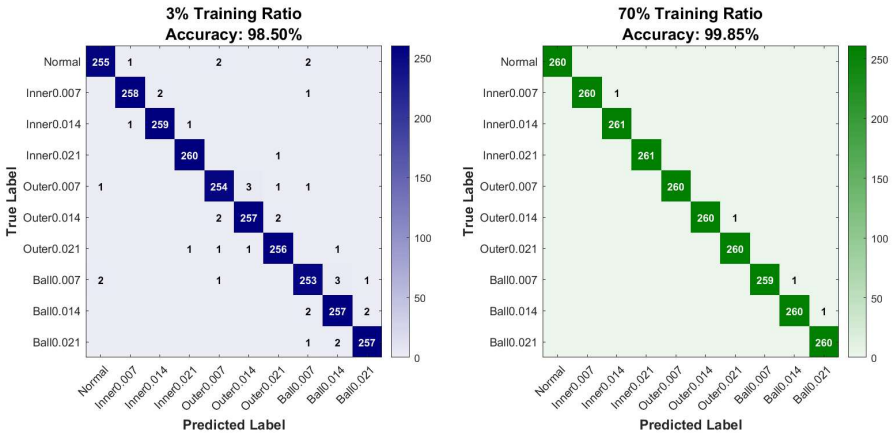


Figure 5: Confusion matrix comparison (left: 3% training ratio, right: 70% training ratio).

the same class already show a clear clustering trend, with different classes clearly distinguishable. As the number of training samples increases to 10%, the inter-class boundaries further tighten, and the clusters become more compact. When reaching 70% sufficient samples, each class forms completely separated cluster structures, with high intra-class aggregation and significant inter-class separation. This process demonstrates that AG-MS-GNN can quickly learn discriminative fault features when samples are scarce, continuously refine features as samples increase, and ultimately achieve a highly separable representation space, providing a reliable basis for high-precision classification.

3.2.5. Model Stability Analysis

To evaluate the robustness of the model under few-shot conditions, 10 repeated experiments were conducted for each method, and the standard deviations of accuracy were calculated. The results are shown in Figure 7. The proposed method exhibits lower standard deviations than GCN and GAT at all training proportions. Especially under the 3% extreme few-shot scenario, the standard deviation is only 2.6%, much lower than GCN's 4.8% and GAT's 4.2%. As the sample size increases, the fluctuation of the proposed method further decreases, reaching only 0.4% at 70% training ratio. This indicates that the adaptive graph structure and self-supervised feature enhancement effectively improve the model's robustness to data distribution, reduce sensitivity to random sampling, and enable the model to maintain stable and reliable output under few-shot and variable operating conditions.

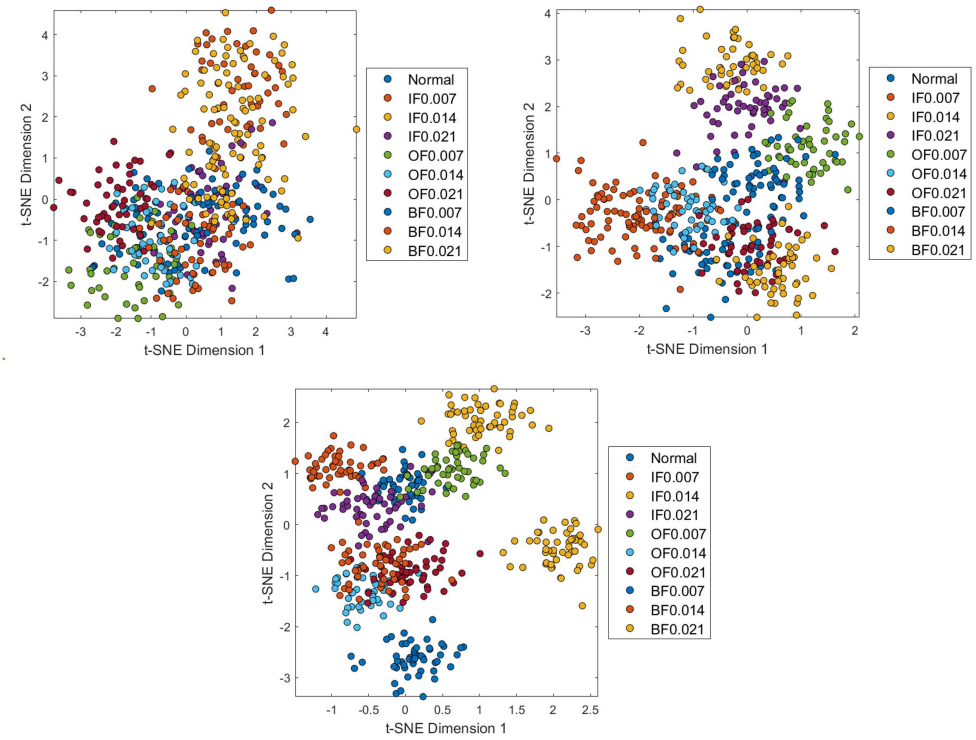


Figure 6: t-SNE feature visualization (left: 3% training ratio, middle: 10% training ratio, right: 70% training ratio).

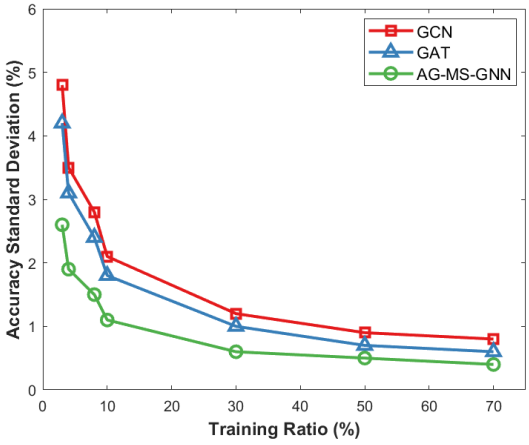


Figure 7: Model stability comparison under different training proportions.

3.3. Experimental Conclusions

Based on the comprehensive experimental results, the following conclusions can be drawn. First, AG-MS-GNN possesses high-precision diagnostic capability under extreme few-shot conditions, achieving over 95% accuracy at 3% training ratio, significantly outperforming existing machine learning and deep learning methods. Second, the model has extremely high data utilization efficiency, achieving with only a small number of samples the performance that traditional models require large numbers of samples to achieve, making it suitable for industrial scenarios with scarce fault samples. Third, ablation experiments confirm that adaptive graph construction, multi-head graph attention, residual connections, and self-supervised enhancement are all effective modules whose synergy improves the model's feature learning capability and generalization. Fourth, the proposed method exhibits the lowest experimental fluctuation across different sample sizes, demonstrating strong stability and adaptability to variable operating conditions and noisy environments. Overall, the proposed method has significant advantages in few-shot, variable-condition bearing fault diagnosis tasks and holds high potential for engineering applications.

4. Discussion

This section provides an in-depth discussion of the proposed AG-MS-GNN method from three aspects: method advantages, limitations, and practical application prospects, and on this basis, outlines future research directions.

4.1. Advantages of the Proposed Method

4.1.1. Interpretability of Physically Driven Graph Construction

A major bottleneck in the application of traditional graph neural networks to fault diagnosis is the subjectivity and lack of physical basis in graph structure construction. Most methods rely on simple Euclidean distance or cosine similarity to construct sample relationship graphs, making it difficult to truly reflect the physical essence of the mechanical equipment fault evolution process. The adaptive physical graph construction mechanism proposed in this paper enhances model interpretability in the following two aspects.

First, the fusion of multi-dimensional physical features endows node features with clear physical meanings. Time-domain features (peak value, root mean square, etc.) directly reflect the energy distribution and impact characteristics of vibration signals; frequency-domain features (fault frequency band energy ratio,

center frequency, etc.) are directly related to bearing fault characteristic frequencies (such as inner race passage frequency, outer race passage frequency, rolling element rotation frequency, etc.); demodulation-domain features (envelope spectrum peak value, entropy, etc.) are particularly sensitive to early weak faults. The fusion of these three types of features enables the model to comprehensively characterize equipment health status from multiple physical dimensions.

Second, the hybrid similarity graph construction strategy integrates global and local correlation information. Cosine similarity captures the directional consistency of samples in the feature space, suitable for measuring the overall distribution characteristics of the same fault type under different operating conditions; Gaussian kernel similarity focuses on local neighborhood structures and is more sensitive to feature amplitude differences. The weighted fusion of the two enables the graph topology to maintain both global discriminative structure and capture fine-grained local fault patterns. Compared to single similarity metrics, this method generates graph structures with stronger physical interpretability and diagnostic discriminative power.

4.1.2. Data Efficiency under Few-Shot Conditions

The proposed method achieves a diagnostic accuracy of 95.55% under extreme few-shot conditions. This data efficiency can be attributed to the synergistic effects of the following three mechanisms:

(1) Information propagation mechanism of graph structure: Unlike CNNs/RNNs that process each sample independently, GNNs explicitly utilize inter-sample relationship information through neighborhood aggregation operations. In few-shot scenarios, this information propagation mechanism enables the supervisory signals from a small number of labeled samples to propagate along graph edges to unlabeled samples, effectively expanding the coverage of effective training signals.

(2) Negative-sample-free self-supervised feature enhancement: The view contrast strategy designed in this paper applies semantically preserved data augmentation (additive noise, amplitude scaling) to the original signals and minimizes the distance between the features of the two augmented views of the same sample, enabling the model to learn feature representations that are robust to operating condition variations. This method avoids the complexity and uncertainty of negative sample construction in traditional contrastive learning and can efficiently utilize unlabeled data to enhance feature discriminability.

(3) Hierarchical feature extraction of multi-scale graph convolution: By stacking 1-hop and 2-hop graph convolutions, the model can capture fault fea-

tures at different receptive fields. Shallow convolutions focus on subtle fault patterns within local neighborhoods, while deep convolutions integrate broader inter-sample correlations. This hierarchical feature extraction mechanism enables the model to gradually build a highly discriminative fault representation under limited training sample conditions.

4.1.3. Cross-Condition Generalization Capability

The excellent performance of the proposed method under cross-condition settings can be attributed to the following factors. On the one hand, the adaptive graph construction mechanism enables the graph topology to dynamically adjust with changes in operating conditions. Traditional static graph methods may fail under varying operating conditions as the predefined inter-sample relationships may become invalid (e.g., the distribution of vibration features of the same fault type may shift significantly at different rotational speeds). Through the learnable graph structure optimization layer, the proposed method achieves end-to-end collaborative adaptation between the graph topology and the downstream diagnostic task, maintaining reasonable inter-sample relationships under different operating conditions. On the other hand, the self-supervised contrastive enhancement strategy explicitly drives the model to learn condition-invariant features. By minimizing the feature distance of the same sample under different data augmentations (simulating operating condition variations), the model is forced to extract fault-essential features that are insensitive to amplitude scaling, noise interference, and other changes, rather than condition-related surface statistical characteristics.

4.2. Limitations

Although the proposed method achieves excellent results on the CWRU dataset, it still has several limitations that need to be addressed in future work.

4.2.1. Robustness under Extreme Noise Conditions

This paper introduces additive Gaussian noise (standard deviation 0.01) in data augmentation to simulate a certain degree of signal interference. However, the noise environment in actual industrial settings can be more severe, including non-Gaussian noise (e.g., impulse interference), power frequency interference, and coupled vibrations from other rotating components. Under these extreme noise conditions, the robustness of the proposed method has not been fully validated. Preliminary analysis indicates that when the signal-to-noise ratio falls below 0 dB, fault features may be severely masked by noise, leading to distortion of similarity calculations during graph construction and consequently affecting diagnostic accuracy.

4.2.2. Handling of Unknown Fault Types

This study assumes that all fault types appear in the training set (closed-set setting). However, in actual industrial scenarios, completely new fault patterns (such as compound faults or novel faults) may emerge, which falls under open-set recognition or incremental learning problems. The current method does not have the capability to recognize unknown fault types; it would forcibly classify unknown faults into one of the known categories, leading to misjudgment.

4.2.3. Computational Complexity of Graph Construction

This paper adopts a K-nearest neighbor strategy for graph construction, with a time complexity of $\mathcal{O}(N^2 \cdot d)$ (where N is the number of samples and d is the feature dimension). Under few-shot conditions ($N \approx 1000$), this computational cost is acceptable. However, when the dataset scales to tens of thousands of samples, full computation of inter-sample similarities becomes infeasible. Additionally, stacking three layers of multi-scale graph convolution (each layer containing both 1-hop and 2-hop convolutions) introduces additional computational burden.

4.2.4. Sensitivity to Hyperparameters

The proposed method introduces several hyperparameters, including the number of nearest neighbors K ($K = 10$), the Gaussian kernel bandwidth σ (adaptively determined), the fusion weights of cosine similarity and Gaussian kernel similarity (0.7 vs. 0.3), and the self-supervised loss weight λ ($\lambda = 0.1$). Although these parameters are optimized through experiments, the optimal parameters may vary across different datasets or operating conditions. The generalizability of the hyperparameters still requires further investigation.

4.3. Practical Application Prospects

4.3.1. Deployment on Edge Devices

Considering the real-time requirements of industrial scenarios, deploying the proposed method on edge devices (such as embedded systems or programmable logic controllers) has significant engineering value. However, the complete AG-MS-GNN model (containing three layers of multi-scale graph convolution, attention mechanisms, and self-supervised modules) has high computational resource demands. Future work can explore the following lightweight strategies: model pruning and quantization to remove redundant attention heads or convolution channels and perform fixed-point quantization; knowledge distillation using the proposed model as a teacher network to train a lightweight student network (e.g., a single-layer graph convolutional network); and offline graph construction where

graph construction and self-supervised pre-training are performed in the cloud, with only forward inference executed on the edge device.

4.3.2. Integration with Online Monitoring Systems

The proposed method currently adopts an offline batch learning mode, where the model is trained on a fixed training set and then deployed. To achieve true intelligent operation and maintenance, the model needs to be integrated into online monitoring systems to support incremental learning, enabling the model to continuously update when new fault samples arrive without forgetting existing knowledge; real-time alerting, where the model outputs fault category confidence and triggers manual review when confidence falls below a threshold; and condition adaptation, dynamically adjusting graph construction parameters based on real-time monitored rotational speed and load information.

4.4. Future Research Directions

Based on the above discussion, this paper proposes the following future research directions:

4.4.1. Continual Learning for Unknown Fault Detection

To address the open-set recognition problem, a continual learning framework can be introduced to enable the model to dynamically expand the category space while continuously receiving new data. Specifically, a distance-based anomaly detection module (e.g., using local density estimation in the graph structure) can be designed to identify samples whose distribution differs significantly from known fault categories and mark them as potential novel faults.

4.4.2. Federated Learning for Multi-Source Data Fusion

In actual industrial scenarios, data sharing among different factories and different equipment is often constrained by privacy protection and commercial confidentiality. Federated learning provides a paradigm for collaborative training while preserving data privacy. Future work can explore cross-client graph structure alignment to handle variations in operating conditions across different clients, and communication-efficient aggregation to reduce the frequency of parameter transmission between clients and the server.

4.4.3. Adaptive Optimization of Graph Construction

The current K-nearest neighbor graph uses a fixed hyperparameter K ($K = 10$). Future research can investigate adaptive K selection strategies, such as

node-local-density-based adaptive K (using smaller K in high-density regions and larger K in low-density regions), or differentiable graph sparsification methods (e.g., using Gumbel-Softmax for edge sampling).

4.4.4. Extension to Other Mechanical Components and Multi-Sensor Fusion

Although the proposed method currently focuses on bearing fault diagnosis, its core ideas (adaptive graph construction and negative-sample-free self-supervised enhancement) are generalizable. Future work can extend it to fault diagnosis of other rotating machinery such as gearboxes and turbines, and explore graph fusion methods for heterogeneous data from multiple sensors (vibration, temperature, acoustic emission, current, etc.).

5. Conclusion

This paper proposes an Adaptive Graph Construction and Feature-Enhanced Graph Neural Network (AG-MS-GNN) method to address the dual challenges of sample scarcity and operating condition variability in bearing fault diagnosis under real industrial scenarios. The core innovations of the proposed method include: (1) an adaptive graph construction mechanism integrating time-domain, frequency-domain, and demodulation-domain physical features, achieving end-to-end collaborative optimization of graph topology and diagnostic tasks; (2) a negative-sample-free graph-level contrastive feature enhancement strategy, which avoids the complexity of negative sample construction in traditional contrastive learning through semantically preserved data augmentation and view comparison, efficiently utilizing unlabeled data to enhance feature discriminability under few-shot conditions; and (3) a diagnostic model integrating multi-scale graph convolution and attention mechanism, capturing both local details and global contextual information.

Systematic experiments on the CWRU bearing dataset validate the effectiveness of the proposed method. The experimental results demonstrate that: (1) under extreme few-shot conditions (approximately 8 training samples per class), AG-MS-GNN achieves a diagnostic accuracy of 95.55%, significantly outperforming comparison methods including SVM (48.57%), 1D-CNN (43.30%), GCN (31.68%), and GAT (approximately 65.2%); (2) when the training ratio increases to 10%, the accuracy further improves to 98.62%, and remains stable above 99.6% at 30% and above; (3) ablation studies confirm the effectiveness of the multi-scale graph convolution, multi-head graph attention, residual connection, and self-supervised enhancement modules, with residual connections contributing the most

to performance (removing them reduces accuracy by 4.03 percentage points), verifying the necessity of multi-component synergistic design; (4) t-SNE visualization shows that the model can quickly learn discriminative features from few samples, with inter-class boundaries continuously tightening as samples increase; and (5) standard deviation analysis from 10 repeated experiments shows that AG-MS-GNN achieves a standard deviation of only 2.6% under 3% extreme few-shot conditions, much lower than GCN (4.8%) and GAT (4.2%), demonstrating excellent stability.

The theoretical contributions of this paper are as follows: First, it proposes a physically driven adaptive graph construction paradigm, incorporating physical priors of vibration signals into the graph topology learning process, providing an interpretable framework for applying graph neural networks to mechanical fault diagnosis. Second, it designs a negative-sample-free graph-level contrastive learning mechanism that simplifies the traditional self-supervised learning process, offering a new approach for feature enhancement under few-shot conditions. Third, it constructs a unified diagnostic framework integrating multi-scale graph convolution and attention, achieving efficient synergy between local and global features. In terms of practical value, the proposed method maintains high accuracy and stability under extreme sample scarcity conditions, providing a feasible technical solution for industrial scenarios lacking sufficient labeled data, with potential for deployment in edge computing and online monitoring systems.

The proposed method still has several limitations that will be addressed in future work: (1) Continual learning and open-set recognition: The current model assumes that all fault types appear in the training set. Future work will introduce distance-based anomaly detection modules and dynamic category expansion mechanisms to enable the model to recognize unknown fault types and continuously update as new data arrives. (2) Federated learning framework: To address privacy protection requirements for multi-source data, cross-client graph structure alignment and communication-efficient parameter aggregation strategies will be explored to achieve collaborative diagnostic modeling without sharing raw data. (3) Adaptive optimization of graph construction: Adaptive K selection strategies based on node local density will be investigated to eliminate bias caused by fixed K values. (4) Multi-sensor and cross-component extension: The method will be extended to other rotating machinery such as gearboxes and turbines, and graph fusion methods for heterogeneous multi-sensor data (vibration, temperature, acoustic emission, etc.) will be explored. These research directions will further enhance the generalization capability, practicality, and intelligence level of the method, promoting the deployment of data-driven fault diagnosis technol-

ogy to a wider range of industrial applications.

Funding Statement

This work was supported by the National Natural Science Foundation of China (Nos. T2421001,62522308, 62325302), Natural Science Foundation of Chongqing, China. (CSTB2024NSCQ-MSX0472), the Fundamental Research Funds for the Central Universities (2024CDJCGJ-006, 2025CDJZKKYJH-06, 2025CDJGF-002).

Data Availability Statement

All experimental data used in this research are derived from the public Case Western Reserve University (CWRU) bearing fault dataset, which is openly available for academic purposes.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. There is no conflict of interest regarding the publication of this study.

References

[1] Yaguo Lei, Bin Yang, Xinwei Jiang, Feng Jia, Naipeng Li, and Asoke K Nandi. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical systems and signal processing*, 138:106587, 2020.

[2] Rui Zhao, Ruqiang Yan, Zhenghua Chen, Kezhi Mao, Peng Wang, and Robert X Gao. Deep learning and its applications to machine health monitoring. *Mechanical systems and signal processing*, 115:213–237, 2019.

[3] Jinyang Jiao, Ming Zhao, Jing Lin, and Kaixuan Liang. A comprehensive review on convolutional neural network in machine fault diagnosis. *Neurocomputing*, 417:36–63, 2020.

[4] Long Wen, Xinyu Li, Liang Gao, and Yuyan Zhang. A new convolutional neural network-based data-driven fault diagnosis method. *IEEE transactions on industrial electronics*, 65(7):5990–5998, 2017.

[5] Wei Zhang, Gaoliang Peng, Chuanhao Li, Yuanhang Chen, and Zhujun Zhang. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, 17(2):425, 2017.

[6] Piyush Agarwal, Jorge Ivan Mireles Gonzalez, Ali Elkamel, and Hector Budman. Hierarchical deep lstm for fault detection and diagnosis for a chemical process. *Processes*, 10(12):2557, 2022.

[7] Siyu Shao, Stephen McAleer, Ruqiang Yan, and Pierre Baldi. Highly accurate machine fault diagnosis using deep transfer learning. *IEEE transactions on industrial informatics*, 15(4):2446–2455, 2018.

[8] Xiang Li, Wei Zhang, and Qian Ding. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliability engineering & system safety*, 182:208–218, 2019.

[9] Han Wang, Jingwei Wang, Yukai Zhao, Qing Liu, Min Liu, and Weiming Shen. Few-shot learning for fault diagnosis with a dual graph neural network. *IEEE Transactions on Industrial Informatics*, 19(2):1559–1568, 2022.

[10] Yichen Li, Guangye Liu, Aocheng Xiong, Wei Zeng, and Lei Jiang. Few-shot learning based on meta-learning and graph attention networks for bearing fault diagnosis. In *2024 5th International Symposium on Computer Engineering and Intelligent Communications (ISCEIC)*, pages 590–594. IEEE, 2024.

[11] Bin Yang, Yaguo Lei, Feng Jia, and Saibo Xing. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mechanical Systems and Signal Processing*, 122:692–706, 2019.

[12] Zihao Lei, Feiyu Tian, Yu Su, Guangrui Wen, Ke Feng, Xuefeng Chen, Michael Beer, and Chunsheng Yang. Unsupervised graph transfer network with hybrid attention mechanism for fault diagnosis under variable operating conditions. *Reliability Engineering & System Safety*, 255:110684, 2025.

[13] Zhuyun Chen, Konstantinos Gryllias, and Weihua Li. Mechanical fault diagnosis using convolutional neural networks and extreme learning machine. *Mechanical systems and signal processing*, 133:106272, 2019.

[14] Zhichao Jiang, Dongdong Liu, and Lingli Cui. Deep adaptively dynamic edge graph convolution network with attention weight and high-dimension affinity feature graph for rotating machinery fault diagnosis. *Measurement Science and Technology*, 36(2):026104, 2025.

[15] Chengming Wang, Yanxue Wang, Yiyan Wang, Xinming Li, and Zhigang Chen. Richly connected spatial–temporal graph neural network for rotating machinery fault diagnosis with multi-sensor information fusion. *Mechanical Systems and Signal Processing*, 225:112230, 2025.

[16] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[17] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

[18] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

[19] Tianfu Li, Zheng Zhou, Sinan Li, Chuang Sun, Ruqiang Yan, and Xuefeng Chen. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. *Mechanical Systems and Signal Processing*, 168:108653, 2022.

[20] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[21] Mengjie Gan, Penglong Lian, Zhiheng Su, Jiyang Zhang, Jialong Huang, Benhao Wang, Jianxiao Zou, and Shicai Fan. Few-shot fault diagnosis based on multi-scale graph convolution filtering for industry. *arXiv preprint arXiv:2405.19642*, 2024.

[22] Zhe Wang, Zhiying Wu, Xingqiu Li, Haidong Shao, Te Han, and Min Xie. Attention-aware temporal–spatial graph neural network with multi-sensor information fusion for fault diagnosis. *Knowledge-based systems*, 278:110891, 2023.

[23] Jige Liao, Yaohua Deng, Xiaobo Xie, and Zilin Zhang. Research on chebyshev graph convolutional neural network modeling method for rotating equipment fault diagnosis under variable working conditions. *Applied Sciences*, 14(20):9208, 2024.

[24] Daimo Yang, Gaigai Cai, Yulong Yan, Yaoyang Hu, and Shibin Wang. Attention-guided multi-feature fusion convolutional network for machinery intelligent fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 2025.

[25] Yuanhong Chang, Jinglong Chen, Weiguang Zheng, Shuilong He, and Enyong Xu. Triplet adversarial learning-driven graph architecture search network augmented with probsparse-attention mechanism for fault diagnosis under few-shot & domain-shift. *Mechanical Systems and Signal Processing*, 199:110462, 2023.

[26] Bo Wang and Shuai Zhao. Mtagcn: multi-task graph-guided convolutional network with attention mechanism for intelligent fault diagnosis of rotating machinery. *Machines*, 13(5):347, 2025.

[27] Xiang Li, Wei Zhang, Hui Ma, Zhong Luo, and Xu Li. Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks. *Neural Networks*, 129:313–322, 2020.

[28] Dongdong Liu, Lingli Cui, and Huaqing Wang. Rotating machinery fault diagnosis under time-varying speeds: A review. *IEEE Sensors Journal*, 23(24):29969–29990, 2023.

[29] Yuechuan Xin, Jianuo Zhu, Mingyang Cai, Pengyan Zhao, and Quanzhi Zuo. Machine learning based mechanical fault diagnosis and detection methods: a systematic review. *Measurement Science and Technology*, 36(1):012004, 2025.

[30] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Advances in neural information processing systems*, 33:5812–5823, 2020.

[31] Xin Zhang, Jie Liu, Xi Zhang, and Yanglong Lu. Self-supervised graph feature enhancement and scale attention for mechanical signal node-level representation and diagnosis. *Advanced Engineering Informatics*, 65:103197, 2025.

[32] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep graph contrastive representation learning. *arXiv preprint arXiv:2006.04131*, 2020.