

Dual-branch spatio-temporal graph network for bearing fault diagnosis

Received: 21 October 2025

Accepted: 26 February 2026

Published online: 11 March 2026

Cite this article as: Wang Y., Li Y., Li C. *et al.* Dual-branch spatio-temporal graph network for bearing fault diagnosis. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-42504-0>

Yajun Wang, Yang Li, Chenggang Li & Qi Dai

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

Funding Declaration

This study supported by CHN Shenhua Energy Co.Ltd. Shendong Coal Branch No: E210100625

Author Contribution

LI Yang conducted literature research;
WANG Yajun and LI Yang provided the subject source;
WANG Yajun provided financial support;
WANG Yajun LI Yang designed the algorithm;
DAI Qi were responsible for the experimental design;
LI Chenggang implement the algorithm in code;
DAI Qi have sorted out the data;
LI Yang wrote the main manuscript text;
All authors reviewed the manuscript.

WANG Yajun (Co-corresponding author)

Shendong Technology Research Institute, CHN Energy Group, Yulin 719315, China
State Energy Group Shendong Coal Group Co., Ltd., Yulin 719315, China
yajun.wang.h@chnenergy.com.cn

LI Yang (Corresponding author)

Branch Institute of Emergency Science, Chinese Institute of Coal Science, Beijing 100013, China
State Key Laboratory of Disaster Prevention and Ecology Protection in Open-Pit Coal Mines, Chinese
Institute of Coal Science, Beijing 100013, China
drliyang@ieee.org

LI Chenggang

College of science, North China University of Science and Technology, Tangshan, 063210, China
lichenggang@stu.ncst.edu.cn

DAI Qi

College of science, North China University of Science and Technology, Tangshan, 063210, China
daiq_cup@126.com

Dual-branch spatio-temporal graph network for bearing fault diagnosis

WANG Yajun^{1 2} (Co-corresponding author), LI Yang^{3 4} (Corresponding author), LI Chenggang⁵, DAI Qi⁵

1. Shendong Technology Research Institute, CHN Energy Group, Yulin 719315, China

2. State Energy Group Shendong Coal Group Co., Ltd., Yulin 719315, China

3. Branch Institute of Emergency Science, Chinese Institute of Coal Science, Beijing 100013, China

4. State Key Laboratory of Disaster Prevention and Ecology Protection in Open-Pit Coal Mines, Chinese Institute of Coal Science, Beijing 100013, China

5. College of science, North China University of Science and Technology, Tangshan, 063210, China

Abstract: Bearings are one of the critical components in rotating machinery. Bearing failures can lead to equipment damage, reduced performance, and even major safety accidents. Therefore, improving the ability to diagnose bearing faults can help improve the availability, reliability, and safety of rotating machinery. However, the original vibration signals in rotating machinery often contain noise and irregularities, making it difficult for traditional vibration analysis to extract effective high-dimensional features. Inspired by the construction of spatio-temporal graphs and dual-branch graph networks, a bearing fault diagnosis method based on dual-branch spatio-temporal graph networks (DBSGN) is proposed. Firstly, the vibration signal is modeled based on spectrum theory and the spectrum analysis method to construct a spatio-temporal graph. Secondly, Laplace-based spectral decomposition is used to extract the feature vectors of samples in the spatio-temporal graph. Finally, we designed a dual-branch fusion network to train and verify the bearing data and adjusted the model's learning of the bearing data through a dynamic attention mechanism. The experimental results on three benchmark datasets indicate that DBSGN outperforms traditional models in terms of stability and accuracy.

Keywords: rotating machinery, fault diagnosis, graph neural network, dual-branch network

1. Introduction

With the rapid development of society, rotating machinery has been widely used in automation and modern industry and plays a vital role. Rolling bearings are one of the most important components in modern industrial rotating machinery. Due to long-term operation and harsh working conditions, rotating machinery such as bearings often suffers unexpected failures and structural damage. This

irreversible damage may lead to economic losses such as equipment damage and reduced production capacity. In severe cases, it may even cause major safety accidents. Therefore, accurate diagnosis of rolling bearing faults has become an important research topic in the engineering field [1][2].

At present, rolling bearing fault diagnosis technology has gone through three stages: manual diagnosis, signal processing, and artificial intelligence. Early manual diagnosis mainly relied on the sensory experience of engineering personnel, such as hearing and touch, and simple vibration measurement tools. Although intuitive, it had low diagnostic efficiency and high subjectivity, making it difficult to cope with complex working conditions and identify early subtle faults. With the advancement of signal processing technology, diagnostic methods have gradually shifted to time domain, frequency domain, and time-frequency domain analysis of vibration and acoustic emission signals. For example, technologies such as short-time Fourier transform (STFT) [3][4], empirical wavelet transform (EWT) [5][6], empirical mode decomposition (EMD) [7][8], and Hilbert-Huang transform (HTT) [9] are used to extract fault features from raw noise signals, thereby effectively identifying specific fault types. However, after discovering fault-related features, traditional signal processing methods still rely on manual or rule-based algorithms to make the final decision. In recent years, with the rise of artificial intelligence technologies such as machine learning and deep learning, rotating machinery fault diagnosis has begun to develop towards intelligence and automation. This method can learn the characteristics of different fault categories from massive monitoring data, achieve high-precision, end-to-end fault classification and status prediction, and greatly improve the accuracy and robustness of fault diagnosis.

Currently, fault diagnosis methods based on deep learning have become a research hotspot, such as convolutional neural networks (CNNs) [10][11], recurrent neural networks (RNNs) [12][13], graph neural networks (GNNs) [14], and their variant models. These models autonomously extract fault features directly from input data without the need for prior knowledge, thereby performing accurate classification. When processing bearing vibration data, existing deep learning methods usually treat the data as independent time series samples or two-dimensional image data. This processing method makes it difficult to fully mine the key fault-related correlation information in the data. On the one hand, predicting continuous vibration signals as time series samples will lose the inherent correlation of bearing fault signals. On the other hand, when converting data into two-dimensional images to adapt to convolutional neural networks, although spatial features can be extracted, the fine time domain details in the original signal may be lost, and the frequency domain signal cannot be effectively modeled. These limitations result in the existing methods being insufficient and inadequate in fault feature extraction and robustness under complex operating conditions such as variable speed and strong noise, which in turn affects the stability and accuracy of the diagnosis results.

To solve the above problems, we propose a rolling bearing fault diagnosis

method based on the graph neural network, namely, bearing fault diagnosis based on dual-branch spatio-temporal graph network (DBSGN). Our proposed method first models the vibration signal based on spectral graph theory and spectral analysis methods to construct a spatio-temporal graph. Secondly, Laplace-based spectral decomposition is used to extract the feature vectors of samples in the spatio-temporal graph. Finally, we designed a dual-branch fusion network to train and validate the bearing data and adjusted the model's learning of the bearing data through a dynamic attention mechanism. The main contributions of our approach can be summarized as follows.

(1) We propose a spatio-temporal graph construction method, which converts bearing vibration signals into graph-structured data, more effectively captures the complex relationships in the graph, and aggregates complex information in neighbor nodes through a message passing mechanism.

(2) We propose a bearing fault detection model based on a dual-branch fusion network, which enhances the accuracy and robustness of the model and stabilizes the detection results.

(3) This study conducts experimental verification on three widely used public datasets, and the experimental results demonstrate the effectiveness of our method in distinguishing abnormal signals from normal signals and achieving more efficient and accurate fault identification.

The rest of this paper is structured as follows. Section 2 introduces graph neural networks and a review of related literature. Section 3 gives the experimental preparation and the theoretical basis of graph convolution. Section 4 describes the proposed fault detection model in detail. In Section 5, we conduct experiments and conduct an in-depth analysis of the experimental results to verify the effectiveness of the model. In Section 6 we conclude our work and discuss future directions.

2. Related Work

2.1 Machine Learning

With the rapid development of machine learning, more and more researchers have begun to use machine learning to predict bearing fault signals. Common machine learning methods include random forest (RF), principal component analysis (PCA), K nearest neighbor (KNN), support vector machine (SVM), etc. Applying these methods requires deep exploratory data analysis of the dataset and feature engineering to extract features from vibration signals. Finally, it is passed to the ML algorithm for prediction. Chen et al. [15] proposed an imbalanced sample diagnosis framework based on clustering and density methods, which improved performance and interpretability by adjusting sample distribution and fusing multi-domain features. Transfer learning methods have also been applied across different working conditions and machines. Wan et al. [16] proposed a

cross-domain adaptation network that combines layered decoding and an attention mechanism, enabling cross-machine diagnosis without labels. Cui et al. [17] combined unsupervised learning and sensitivity analysis in their research on wind turbines and proposed a three-stage learning framework to effectively extract features under diverse working conditions. Wang et al. [18] proposed an application method that combines numerical simulation models with machine learning classification. They used a mature dynamic model to build a simulation model, modified the fault influencing factors through the Pearson correlation coefficient (PCC), obtained simulation data, and used the machine learning model for prediction. Habbouche et al. [19] proposed an early diagnosis method for bearing faults based on variational mode decomposition (VMD) and a machine learning method, which uses VMD for feature extraction for fault detection and then triggers the use of a one-dimensional convolutional neural network for multi-scale feature extraction for classification and diagnosis. Brusamarello et al. [20] proposed a support vector machine (SVM) classifier to identify the fault severity level by selecting the four highest peaks in the spectrum and PCA techniques for feature extraction, and used a supervised SVM classifier for prediction.

2.2 Deep Learning

Although machine learning methods have achieved good performance in bearing fault detection, there are still some challenges, such as the need for extensive domain expertise, complex feature engineering, and data explosion. In order to solve the above problems, more and more researchers have adopted deep learning algorithms to perform fault diagnosis on bearing vibration signals and have achieved good model performance and experimental results. Common deep learning algorithms include convolutional neural networks (CNN), autoencoders, long short-term memory networks (LSTM), generative adversarial networks (GAN), etc. Compared with classic ML algorithms, these DL algorithms have the advantages of automatic feature extraction and good transferability. Hoang D et al. [21] systematically reviewed the application of deep learning methods such as autoencoders, restricted Boltzmann machines and convolutional neural networks in bearing diagnosis, pointing out their advantages in automatic feature learning. Neupane and Seok [22] specifically summarized the current status of deep learning research based on the CWRU dataset, explaining that this dataset has become an important benchmark for validating diagnostic methods. In terms of specific methods, Hatipoğlu et al. [23] proposed a framework that combines multi-domain features, LSTM and attention mechanism, and achieved high-precision diagnosis on both CWRU and HUST datasets. Rezazadeh et al. [24] proposed an unsupervised domain adaptation framework called WaveCORAL-DCCA, which innovatively combines discrete wavelet transform for robust time-frequency feature extraction with an enhanced deep canonical correlation analysis (DCCA) network and incorporates a correlation alignment (CORAL) loss function to achieve superior domain-invariant representation learning. Du et al. [25] proposed VMD decomposition combined with GASF-CNN and BiLSTM parallel network to

improve the performance under complex working conditions. To address the sample imbalance problem, Li et al. [26] proposed a diagnostic method combining CNN-KAN with diffusion network enhancement, which significantly improved the accuracy under few-sample conditions. Ma and Guo [27] combined time-frequency joint features with a hybrid deep network to achieve robust diagnosis in variable speed and strong noise environments. Xie et al. [28] combined the enhanced VAE-WGAN with the WDCNN dual-channel structure, showing superior performance under class imbalance conditions. In addition, scholars at home and abroad have proposed a variety of improvement strategies, such as ACGAN-SN based on spectral normalization [29], capsule GAN combined with wavelet features [30], deep feature enhancement GAN [31], etc., which have made significant progress in the diversity and reliability of generated samples.

2.3 Graph Neural Networks

Although rolling bearing data contains rich relational structures, practical engineering often faces the problem of limited sample size, which makes the model unable to fully explore the features in the bearing data. To address this problem, graph neural networks have become an effective method in the field of rolling bearing fault diagnosis in recent years due to their ability to model structured data. Yang et al. [32] proposed a super graph method to convert the original vibration signal into graph structured data, and then used a graph convolutional network for final diagnosis. Yu et al. [33] proposed a new fault diagnosis framework based on graph neural network, which converts the features extracted from the original vibration signal into a graph structure and applies GNN to the graph data for fault diagnosis research. Zhang et al. [34] constructed an adjacency matrix through feature extraction and causal test quantization weights and input it into GNN classification. This method effectively alleviated the impact of noise on vibration signals and enhanced the interpretability of the model. Jiang et al. [35] proposed a multi-head graph attention network, which aggregates the discriminative features of graph data into new feature representations through a multi-head attention mechanism. Sun et al. [36] introduced a novel intelligent fault diagnosis method based on frequency-time sequence graph (FTSG) and graph generative classification adversarial network. Certain frequencies in the spectrum are used as node features, connected by edges in chronological order, and the generator is used to expand the training samples, effectively alleviating the impact of the scarcity of fault samples. Xiao et al. [37] simplified the generated weighted graph structure data by defining an edge failure threshold to improve data quality and improved the graph convolutional network (GCN) to reduce computational complexity. Rezazadeh et al. [38] proposed a novel structural health monitoring (SHM) framework for composite materials under temperature-induced variations. This method leverages the powerful feature extraction capabilities of GAT and advanced domain adaptation (DA) techniques, achieving scalable comparison of temperature domain feature distributions by combining maximum mean discrepancy (MMD) and correlation alignment (CORAL) losses with a domain

discriminative adversarial model.

3. Preliminaries

3.1 Problem Definition

A graph G is represented as $G=(V,E,X,A)$, where $V=\{v_i|i=1,2,\dots,n\}$ represents the set of nodes in the graph G , $E=\{e_i|i=1,2,\dots,m\}\subseteq V\times V$ represents the set of connected edges in the graph G , $X\in\mathbb{R}^{n\times d}$ represents the node features, and $A\in\{0,1\}^{n\times n}$ represents the adjacency matrix, $|V|=n$ and d are the number of nodes and the feature dimension of each node, respectively. $A=[a_{ij}]$ represents the relationship between nodes v_i and v_j . If there is an edge between v_i and v_j (i.e., $(v_i,v_j)\in E$), then $a_{ij}=1$, otherwise $a_{ij}=0$. In addition, the labels of fault classes are denoted as $C=\{c_i|i=1,2,\dots,C\}$, where c_i represents the label belonging to a specific class and C represents the total number of classes.

3.2 Theoretical Basis

Graph convolutional networks (GCNs) [39] represent a class of neural networks specifically designed to process graph-structured data. They can be roughly divided into spectral methods rooted in graph signal processing and spatial methods that directly define convolutions on graph topology. The GCNs used in this work belong to the spectral domain category. However, traditional spectral graph convolution has two limitations. First, the graph convolution kernel is global and has a large number of parameters. The size of the convolution kernel is the same as the input signal, and the number of parameters is the same as the number of graph nodes. Second, the graph convolution operation is highly complex and relies on the inverse operation of the adjacency matrix. The computational complexity is $\mathcal{O}(N^2)$. When N is large, the amount of calculation is too large, which is not conducive to deep network training. In order to solve the above problems, Chebyshev proposed to use polynomial expansion to approximate the calculation of graph convolution [40], that is, to perform polynomial approximation on the parameterized frequency response function:

$$g_{\theta}(\Lambda)\approx\sum_{k=0}^{K-1}\theta_k\Lambda^k(1)$$

Where k represents the highest order of the polynomial. The number of parameters of the convolution kernel is reduced from n to k , but the complexity of the entire graph convolution operation is still $\mathcal{O}(N^2)$.

To further reduce computational complexity, Defferrard et al. [41] used the k -th order Chebyshev polynomial $T_k(x)$ to better approximate traditional graph convolution and proved that the complexity can be reduced to $\mathcal{O}(N|E|)$. The equation definition is as follows:

Where $\hat{g}_{\theta}(\bullet)$ represents the filter function, θ_k is the trainable coefficient of the Chebyshev coefficient, and $T_k(x)$ is the recursive calculation of the Chebyshev

polynomial, as shown in Eq. 3. $\tilde{\Lambda}$ is the normalized version of Λ , and the calculation formula is shown in Eq. 4:

$$\begin{cases} T_0(x)=1, T_1(x)=x \\ T_k(x)=2xT_{k-1}(x)-T_{k-2}(x), k \geq 2 \end{cases} (3)$$

$$\tilde{\Lambda} = \frac{2\Lambda}{\lambda_{max}} - I_N (4)$$

Where λ_{max} represents the maximum eigenvalue of the diagonal matrix Λ , and the eigenvalues in $\tilde{\Lambda}$ range from $[-1, 1]$.

By multiplying the Chebyshev graph convolution g_θ with the feature matrix X of the input signal, the aggregation of multi-order neighbor features is achieved, and the feature matrix after dimensionality reduction is finally output. The calculation formula for Chebyshev graph convolution feature aggregation is shown in Eq. 5:

$$M = \sum_{k=0}^{K-1} \theta_k U T_k(\tilde{\Lambda}) U^T X (5)$$

The graph convolution operation can be viewed as filtering the signal without changing the feature dimension. After the graph signal is multiplied by the trainable weight matrix, the dimension of the feature can be changed, as shown in Eq. 6:

$$X^i = \text{Cheb}(M, W^i) = M W^i (6)$$

Where W^i is the trainable weight matrix of the i -th graph convolution layer, $\text{Cheb}(\cdot, \cdot)$ represents the function of Chebyshev convolution, and X^i is the final output of the i -th graph convolution layer.

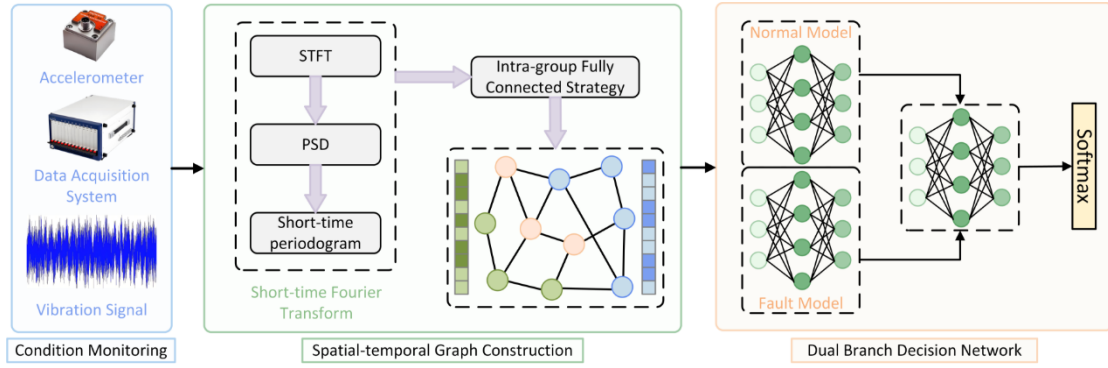
Referring to previous work [42][43], this experiment selected five Chebyshev convolutional layers as the prediction model. The definition of its prediction model is shown in Eq. 7:

$$Y = \text{softmax}(\text{Cheb}(\sigma(\text{Cheb}(X^i, W^i)))) (7)$$

Where $\text{softmax}(\cdot)$ is the normalized exponential function, and $\sigma(\cdot)$ represents the activation function rectified linear unit (RELU).

4. The dual-branch space-time graph network

This paper introduces in detail the proposed new rolling bearing fault diagnosis model based on dual-branch spatio-temporal graphs network (DBSGN) and derives the calculation process of the model. The overall architecture of DBSGN is shown in Fig. 1. The model is mainly divided into three parts: spatio-temporal graph generation, dual-branch decision network and decision fusion. In Section 4.1, the process of generating a spatio-temporal graphs from the rolling bearing dataset is described in detail. In Section 4.2, a dual-branch Chebyshev network is used to learn the characteristic patterns of different data in the bearing fault dataset. Finally, the final decision is generated through an adaptive decision mechanism to improve the robustness of the final classification decision, as shown in Section 4.3.



4.1 Spatial-temporal Graph Generation

In bearing fault diagnosis, traditional deep learning models perform bearing fault diagnosis by constructing time-frequency domain features into time series samples or two-dimensional image data. These methods are difficult to simultaneously capture the temporal dynamics, spatial correlation, and frequency distribution characteristics of the fault, resulting in weak prediction accuracy and robustness of the model. To address this issue, we propose a spatio-temporal graph generation framework to convert unstructured bearing vibration signals into graph-structured data, thereby capturing the spatial correlation and temporal dynamics between features and providing an adaptive input format for graph convolutional networks.

The original bearing vibration signal contains a lot of noise, and it is impossible to extract effective fault features. In order to preserve the local time domain characteristics of the fault, this paper adopts a non-overlapping sliding window strategy to segment the input vibration signal and provide standardized input for subsequent time-frequency analysis.

Since the generated normal and faulty samples have high similarity in time domain waveforms, it is difficult to accurately distinguish normal samples from faulty samples only through the time domain waveforms. In order to further improve the recognition of fault features, this paper uses short-time Fourier transform (STFT) to map time domain samples to the time-frequency domain, present the fault signal features in three-dimensional form, and improve the visibility of fault features. The STFT calculation formula in the continuous time-frequency domain is shown in Eq. 8:

$$S(\tau, f) = \int_{-\infty}^{+\infty} x(t) \cdot \omega(t-\tau) \cdot e^{-j2\pi ft} dt \quad (8)$$

Where $S(\tau, f)$ is a complex time-frequency matrix, f is the frequency parameter, τ is the time frame parameter, $x(t)$ represents the time domain signal of the short-time sample, t represents the range of the sliding window length, and $\omega(t-\tau)$ represents the selected window function. The choice of window function directly affects the time-frequency analysis effect and needs to meet the requirements of high sidelobe attenuation rate and moderate mainlobe width. This study uses Hanning window as the sliding window, and its formula is expressed as:

$$\omega(t) = 0.5 \left[1 - \cos\left(\frac{2\pi t}{T}\right) \right] \quad (9)$$

Where T_s represents the window length of the STFT.

Through short-time Fourier transform, the transformation of time domain signals into time-frequency features is achieved, preserving the temporal dynamics and frequency distribution characteristics of the fault. However, the amplitude of the time-frequency matrix output by the short-time Fourier transform can only reflect the signal intensity and cannot directly quantify the energy distribution. To solve the above problems, the power spectral density (PSD) is used to convert the amplitude of the time-frequency matrix into an energy index with clear physical meaning, amplifying the energy difference between the fault characteristics and the normal characteristics, making the energy peak at the fault characteristic frequency more significant. The calculation formula for PSD is shown in Eq. 10:

$$PSD(s, f, t) = \frac{|S_{s, f, t}|^2}{T} \quad (10)$$

Where $PSD(s, f, t)$ represents the power spectral density value, $S_{s, f, t}$ represents the short-time Fourier transform assignment matrix, and T represents the Hanning window length.

Traditional deep learning models are usually used to process image data or time series data. Traditional deep learning models are typically used to process image data or time series data. In contrast, GNN models use the adjacency matrix A and feature matrix X in the graph data G as input, fully utilizing the hidden structural and topological information in the graph to improve the accuracy of model predictions. However, the process of how to convert one-dimensional vibration signal data into Euclidean data deserves careful consideration. To this end, this paper proposes a spatio-temporal graph construction method based on the Laplace matrix to solve the above problems. Compared with traditional feature extraction methods, spatio-temporal graphs establish a mapping relationship between time-frequency features and graph structure, so that the graph not only includes time-frequency domain information, but also contains the structural information and topological relationship of the graph, further improving the discrimination of the extracted features and the prediction accuracy of the GNN model. The specific construction process of the spatial-temporal graph G is as follows:

After processing the vibration signals in the bearing dataset using STFT, we treat all the resulting fault samples as nodes in graph G . Subsequently, we group all the nodes V and construct the graph. Due to the spatio-temporal-frequency global coupling characteristics of bearing vibration signals, fault features often exist in the form of discrete frequency bands or their harmonic clusters, and the coupling relationships are complex. A fully connected topology within each group can maximize the preservation of potential correlation information within the group and prevent the loss of weak fault features. Therefore, we use a fully connected strategy within each group to construct the spatio-temporal graph G

for all nodes in different groups.

In order to obtain the features of each node in the graph, we first define a weight matrix W . This paper uses Euclidean distance and Gaussian kernel to calculate the power spectral density matrix of each node to measure the similarity value between different frequencies of each node. The greater the feature similarity between different frequencies, the higher the weight value in the adjacency weight matrix, and the better the effect of subsequent feature extraction. The calculation formula of the weight matrix W is shown in Eq. 11:

$$W_{ij} = \exp\left(-\frac{1}{\sigma^2} \sum_{t=t_0}^{t_m} \text{DisPSD}(F_i, t), \text{PSD}(F_j, t)\right) \quad (11)$$

Where W_{ij} represents the elements of the undirected adjacency weight matrix, with a value range of $(0,1]$. The weight of nodes with exactly the same features is 1, and the weight of nodes with very different features tends to 0. i and j are node indices of different frequencies. $\text{PSD}(F_i, t)$ represents the power spectral density value of the F_i -th frequency at a certain time point t . $\text{Dis}\{\bullet, \bullet\}$ represents the Euclidean distance between frequencies F_i and F_j at a certain time point t , which is a commonly used distance function. σ is the Gaussian kernel parameter used to control the degree of influence of distance on weight.

In spectral graph analysis, the symmetric normalized Laplacian matrix is the core input of graph convolution, which updates the current node features through the features of neighboring nodes to avoid the interference of isolated nodes on feature learning. Therefore, we use the symmetric normalized Laplace matrix to process the weight matrix. First, calculate the degree matrix D of the weight matrix. The calculation formula of the Laplace matrix L is shown in Eq. 12-13:

$$L = D - W \quad (12)$$

$$d_{ii} = \sum_{n=1}^f \sum_{j=1, j \neq i}^f W_{ij} \quad (13)$$

Where D represents the diagonal matrix, the diagonal elements are the sum of the weights of the corresponding nodes, L represents the Laplace matrix, and d_{ii} represents the degree of the i -th node, that is, the sum of the weights of all its neighboring nodes. Then calculate the symmetric normalized Laplace matrix L_{sym} , and the calculation formula is shown in Eq. 14:

$$L_{sym} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \quad (14)$$

Where L_{sym} represents the symmetric normalized Laplace matrix, I is the identity matrix, and $D^{-\frac{1}{2}}$ is the inverse square root of the degree matrix, which is used to normalize node weights to prevent nodes with large degrees from excessively influencing their neighbors.

Subsequently, we use orthogonal decomposition to calculate the normalized Laplace matrix, and the calculated eigenvalues are used as node features. The calculation formula for orthogonal decomposition is shown in Eq. 15:

$$L_{sym} = U \Lambda U^T \quad (15)$$

where $\Lambda = \text{diag}([\lambda_0, \lambda_1, \dots, \lambda_{d-1}])$ represents the eigenvalue diagonal matrix of the

Laplacian matrix after orthogonal decomposition, and $U=[u_0, u_1, \dots, u_{d-1}]$ represents the eigenvector matrix of the Laplacian matrix after orthogonal decomposition. Each node feature in the spatio-temporal graph can be represented by a one-dimensional feature vector x , which consists of eigenvalues and is expressed as $x=[\lambda_0, \lambda_1, \dots, \lambda_{d-1}]$.

The final generated spatio-temporal graph is $G=(X,A)$, where each node in V represents a bearing fault sample, and $X=\{x_1, x_2, \dots, x_n\} \in \mathbb{R}^{n \times d}$ represents the feature matrix of the bearing fault sample.

4.2 Dual-branch Decision Network

Single-branch Chebyshev graph convolutional networks have obvious adaptability issues. First, the characteristic distributions of normal and abnormal bearing signals are extremely different, which causes the gradient of fault samples to be dominated by normal samples, resulting in insufficient fault feature learning. Secondly, the characteristics of minor fault samples are masked by major fault samples, causing the model to misclassify minor fault samples as major fault samples. To address these issues, we designed a parameter-independent dual-branch Chebyshev graph convolutional network. By differentially training all features and fault features, we continue to learn the characteristic patterns of fault samples based on the features of all samples, thereby improving the specificity and accuracy of feature extraction.

To ensure that the two branches focus on learning all features and fault features respectively and avoid gradient interference between the two models, we designed a differentiated training mechanism to ensure stable network training. For the normal model, we use all samples in the training set for training, perform gradient updates on the parameters of the normal model through the loss function, and learn the characteristic patterns of all signals. The loss function L_{nor} of the normal model is defined as shown in Eq. 19:

$$L_{nor} = E_{j \in V_{train}^{nor}} [H(y_{v_i, c}, \hat{y}_{v_i, c})] \quad (16)$$

where V_{train}^{nor} is all the training data used by the normal model, $H(\cdot, \cdot)$ is the cross entropy loss function, $y_{v_i, c}$ represents the true label of node v_i belonging to category c , and $\hat{y}_{v_i, c}$ represents the probability predicted by the model that node v_i belongs to category c .

For the fault model, we use the fault samples in the training set for training, perform gradient updates on the parameters of the fault model through the loss function, and learn the characteristic patterns of fault signals of different categories. The loss function L_{fal} of the fault model is defined as shown in Eq. 17:

$$L_{fal} = E_{j \in V_{train}^{fal}} [H(y_{v_i, c}, \hat{y}_{v_i, c})] \quad (17)$$

Where V_{train}^{fal} is the fault training data used by the fault model, $H(\cdot, \cdot)$ is the cross-entropy loss function, $y_{v_i, c}$ represents the true label of node v_i belonging to category c , and $\hat{y}_{v_i, c}$ represents the probability predicted by the model that node v_i belongs to category c .

4.3 Decision Fusion

The dual-branch Chebyshev graph convolutional network has different emphasis on the data in the training set, resulting in significant differences in the training process of the dual-branch model. Specifically, when the proportion of normal samples in the data set is higher than that of fault samples, the feature learning of fault samples by the normal model is easily dominated by normal samples, resulting in high recognition accuracy for normal samples, but insufficient feature capture ability for minor fault samples. Since the fault model focuses on training fault samples, the recognition accuracy of fault samples is significantly improved, but the features of normal samples are not sufficiently fitted, and normal samples are easily misjudged as minor faults. However, traditional fixed-weight fusion cannot take into account both the recognition accuracy of normal samples and the sensitivity of fault samples, resulting in the accuracy after fusion being lower than the optimal value of a single branch.

To address the aforementioned problems, we propose an adaptive attention-based decision fusion mechanism. Its core adjustment rule is an adaptive allocation mechanism based on sample feature differences, guided by multi-objective loss, and constrained by weight normalization. This mechanism dynamically learns attention weights through a multi-layer perceptron (MLP), differentially weighting the model's output. This ensures that normal samples are prioritized by the normal model, faulty samples by the fault model, and slightly faulty samples balance the outputs of both models, thus solving the performance bias problem of a single model.

In order to enable MLP to simultaneously utilize the feature information output by the dual-branch network, the output features of the normal model and the fault model need to be spliced in the feature dimension to form a fusion input matrix. The feature splicing formula is shown in Eq. 18:

$$C=[O_{nor}, O_{fal}](18)$$

Among them, $O_{nor} \in \mathbb{R}^{N \times C}$ is the output feature matrix of the normal model, $O_{fal} \in \mathbb{R}^{N \times C}$ is the output feature matrix of the fault model, and $C \in \mathbb{R}^{N \times 2C}$ is the concatenated fusion input matrix. Each row of the fusion input matrix C corresponds to the normal branch feature and fault branch feature of a node, and contains the complete feature basis for determining the label category to which the node belongs.

In order to dynamically allocate attention weights, we use MLP to map the 2 C -dimensional fusion input matrix into a 2-dimensional attention weight vector, and normalize the attention weights through the softmax function to obtain the final attention weight matrix. The calculation of the attention weight matrix ω_{att} is shown in Eq. 19:

$$X^l = f(W^{l-1} X^{l-1} + b^{l-1})(19)$$

Where $X^0 = C$ represents the fused input matrix of layer 0, $f(\cdot)$ represents the activation function, such as ReLU, Sigmoid, etc., W^{l-1} is the weight matrix of layer $l-1$, b^{l-1} is the bias of layer $l-1$, and $X^l \in \mathbb{R}^{N \times 2}$ is the unnormalized attention

weight. Then, X^l is normalized to the interval $[0,1]$ through the Softmax function to obtain the attention weight matrix ω_{att} .

Based on the attention weight matrix ω_{att} , the output features of the two-branch Chebyshev network are weighted and summed to obtain the final fusion feature matrix O_{fus} . The calculation formula for weighted feature fusion is shown in Eq. 20:

$$O_{fus} = \omega_{att}^{nor} \cdot O_{nor} + \omega_{att}^{fal} \cdot O_{fal} \quad (20)$$

Where $O_{fus} \in \mathbb{R}^{N \times C}$ is the final fusion feature matrix, ω_{att}^{nor} and ω_{att}^{fal} represent the weights of the normal model and fault model output, respectively. The obtained fusion feature vector O_{fus} is converted into a category probability distribution through the log_softmax activation function, and the fusion loss function is calculated. The fusion loss function L_{fus} is defined as shown in Eq. 21:

$$L_{fus} = E_{i \in V_{train}} [H(y_{v_i, c}, \hat{y}_{v_i, c})] \quad (21)$$

Where V_{train} is the training data used by the MLP model, $H(\cdot, \cdot)$ is the cross-entropy loss function, $y_{v_i, c}$ represents the true label of node v_i belonging to category c , and $\hat{y}_{v_i, c}$ represents the probability that node v_i belongs to category c after fusion.

In order to simultaneously optimize the feature fitting capabilities of the multi-layer perceptron, normal model, and fault model, and avoid the degradation of model performance caused by single-objective optimization, we designed a multi-objective loss function to optimize the model. The calculation formula of the multi-objective loss function is shown in Eq. 22:

$$L_{total} = L_{fus} + \alpha \cdot L_{nor} + (1 - \alpha) \cdot L_{fal} \quad (22)$$

Where α is the weight coefficient.

4.4 Computational Complexity Analysis

We performed a complexity analysis of the DBSGN in this study. The DBSGN mainly consists of three parts: spatio-temporal graph generation, a dual-branch decision network, and decision fusion. Therefore, the computational cost also comes from these three parts. In the spatio-temporal graph generation stage, the complexity of the short-time Fourier transform is $O(LT \log T)$, the calculation of the adjacency weight matrix is $O(N^2 d)$, and the orthogonal eigenvalue decomposition of the Laplacian matrix is $O(N^3)$; the dual-branch decision network contains K convolutional layers, the complexity of a single branch is $O(KNd^2)$, and the total complexity of the dual-branch is $O(KNd^2)$; the complexity of feature concatenation, weight learning, and fusion calculation in the decision fusion part is low and can be ignored. The overall time complexity is $O(N^3 + KNd^2)$. In sparse graphs or practical scenarios, when $N \gg d$ and $N \gg K$, the dominant term is $O(N^3)$. The spatial complexity is mainly dominated by the storage of the $N \times N$ adjacency matrix, which is $O(N^2)$, while the storage cost of the feature matrix and network parameters is negligible. The pseudo code of the model is shown in Algorithm 1.

Algorithm 1: Dual-branch Space-time Graph Network □DBSGN □

Input: the raw signal X ;

Output: the fault category \hat{Y} ;

Perform short-time Fourier transform on X to calculate the short-time periodogram using Eq.8-10

Construct the adjacency matrix of spatial-temporal graph G based on the grouped fully connected strategy for all nodes

For each node, calculate the frequency weight matrix using Eq.11

Calculate the Laplacian matrix using Eq. 12-13

Extract the eigenvectors of each node to construct the feature matrix of the spatio-temporal graph using Eq. 14-15

Obtain the G

Construct the dual branch decision network and adaptive attention fusion network

for epoch in epochs **do**

Calculate the output of a dual branch decision network using Eq.7

Calculate the fused output using Eq. 18-20

Calculate loss function using Eq. 16-17, 21-22

Update with back propagation;

end for

5. Experiment

5.1 Experimental Setup

Dataset. To evaluate the proposed model, we conduct extensive experiments on data provided by Case Western Reserve University (CWRU) [44], University of Paderborn (PU) [45], and University of Ottawa (uOttawa) [46]. The details of the fault dataset are shown below.

CWRU bearing dataset [44]: The CRWU dataset is provided by Case Western Reserve University and is a public dataset widely used in the field of rolling bearing fault diagnosis. This dataset is bearing test data obtained from an experimental test bench, which consists of a 2HP induction motor, a torque sensor, a dynamometer, and a deep groove ball bearing. Vibration data is captured using accelerometers installed at different positions on the motor housing.

The experiment collected data under four different motor load conditions: 0HP, 1HP, 2HP, and 3HP, and the vibration signal was sampled at a frequency of 12 kHz. The 12k drive-end bearing failure data includes four failure diameters: 0.007 inches, 0.014 inches, 0.021 inches, and 0.028 inches. The 0.028 inches uses an NTN-6205-2RS equivalent bearing, and the others all use SKF-6205-2RS bearings. Each fault diameter includes three fault types: inner race (IR), outer race (OR), and ball (B), with the outer race fault data missing for 0.028 inches. Based on the above, we constructed a comprehensive dataset containing 12 different bearing health conditions: 1 normal condition and 11 fault conditions. Table 1 summarizes

the details of the dataset composition.

Table 1 Composition of the CWRU rolling bearing experimental dataset.

No.	Fault Diameter	Fault Location	Fault Class	Label	Samples size
1	0	Normal	0	0	200
2		Inner raceway	IR007	1	200
3	0.007	Ball	B007	2	200
4		Outer raceway	OR007	3	200
5		Inner raceway	IR014	4	200
6	0.014	Ball	B014	5	200
7		Outer raceway	OR014	6	200
8		Inner raceway	IR021	7	200
9	0.021	Ball	B021	8	200
10		Outer raceway	OR021	9	200
11		Inner raceway	IR028	10	200
12	0.028	Ball	B028	11	200

PU Bearing Dataset [45]: The PU dataset is published by Paderborn University. Its core goal is to provide support for the research and application of data-driven classification methods. It is mainly used to monitor and diagnose bearing damage in electromechanical drive systems through motor current signals. This dataset contains bearing test data obtained from an experimental test bench. The test bench consists of a drive motor, a torque measurement shaft, a rolling bearing experimental module, a flywheel, and a load motor. The test bench obtains signal data under different working conditions by applying different loads and speeds to the experimental bearings.

The sampling frequency of the vibration signal in this dataset is 64 kHz. Each bearing fault data contains four operating conditions. The detailed information of different operating conditions is shown in Table 2. The dataset for each working condition contains three fault types: healthy, inner loop fault, and outer loop fault. Based on the above, we constructed a dataset containing four working conditions, each of which includes three fault types.

Table 2 Working condition information of PU rolling bearing experimental dataset.

No.	Rotational Speed [rpm]	Load Torque [Nm]	Radial force [N]	Name of Setting	Samples size
1	1500	0.7	1000	N15_M07_F10	300

2	900	0.7	1000	N09_M07_F10	300
3	1500	0.1	1000	N15_M01_F10	300
4	1500	0.7	400	N15_M07_F04	300

uOttawa bearing dataset[46]: The uOttawa dataset is provided by the University of Ottawa. The experiments were conducted on a SpectraQuest Mechanical Fault Simulator (MFS-PK5M), which consists of an AC drive, a motor, two ER16K ball bearing support shafts, and bearings. Vibration data was collected by placing an ICP accelerometer on the housing of the experimental bearing. In addition, an EPC incremental encoder was installed to measure shaft speed.

The data contains vibration signals collected from bearings in different health conditions under time-varying speed conditions. The vibration signals are sampled at a frequency of 200kHz. Each dataset contains two experimental settings: bearing health condition and speed change condition. The health status of a bearing includes health, inner ring defect failure, outer ring defect failure, ball defect failure, and combined defect failure of the inner ring, outer ring and ball. The operating speed conditions are increasing speed, decreasing speed, increasing speed and then decreasing speed, and decreasing speed and then increasing speed. Based on the above, we constructed a comprehensive dataset containing five different bearing health states: one normal condition and four fault conditions. Table 3 summarizes the details of the dataset composition.

Table 3 Composition of the uOttawa rolling bearing experimental dataset.

Bearing Health Conditions	Speed Varying Conditions				Samples size
	Increasing Speed	Decreasing Speed	Increasing then decreasing speed	Decreasing then increasing speed	
Healthy	H-A-1	H-B-1	H-C-1	H-D-1	600
	H-A-2	H-B-2	H-C-2	H-D-2	
	H-A-3	H-B-3	H-C-3	H-D-3	
Inner Race Defect	I-A-1	I-B-1	I-C-1	I-D-1	600
	I-A-2	I-B-2	I-C-2	I-D-2	
	I-A-3	I-B-3	I-C-3	I-D-3	
Outer Race Defect	O-A-1	O-B-1	O-C-1	O-D-1	600
	O-A-2	O-B-2	O-C-2	O-D-2	
	O-A-3	O-B-3	O-C-3	O-D-3	
Ball Defect	B-A-1	H-B-1	B-C-1	B-D-1	600
	B-A-2	B-B-2	B-C-2	B-D-2	
	B-A-3	B-B-3	B-C-3	B-D-3	
Combined Defects on the Inner Race, the Outer Race and a Ball	C-A-1	C-B-1	C-C-1	C-D-1	600
	C-A-2	C-B-2	C-C-2	C-D-2	
	C-A-3	C-B-3	C-C-3	C-D-3	

Baseline. To demonstrate the advantages of our model, this study selected

some representative and popular methods in the field of fault diagnosis as baseline models for comparison with our proposed method, as shown below:

Convolutional Neural Network (CNN) [47]: CNN is a benchmark method in deep learning. In this experiment, bearing signals are predicted by converting them into image information.

SuperGraph [32]: SuperGraph converts raw data into a hypergraph through short-time Fourier transform (STFT) and uses graph neural networks to predict nodes in the hypergraph.

WPT-FFT [48]: This method decomposes the vibration signal through wavelet packet transform (WPT) and Fourier transform (FFT), extracts high-resolution features from the short-term vibration signal, and uses the Bayesian optimized random forest algorithm for classification.

multi-scale CNN-Transformer[49]: This method uses the whale optimization algorithm to determine the optimal parameters for continuous variational mode decomposition. After reconstructing the signal using intrinsic mode functions, a multi-scale convolutional neural network and a Transformer are used to extract local and global features, thereby achieving fault type classification.

MultiPatchTST[50]: This method utilizes multiple PatchTST models based on the Transformer architecture to generate different intermediate results, and then uses an MLP layer to generate the final prediction.

HMDN[51]: HMDN employs a tree-structured labeling scheme to annotate multi-granularity faults and improves intra-class compactness and reduces inter-class hierarchical similarity through a hierarchical multi-granularity diagnostic network and a multi-granularity fault loss function.

Evaluation indicators. To evaluate the effectiveness of our model, this experiment selected accuracy, precision, recall, F1-score, AUC curve, and Matthews correlation coefficient (MCC) as evaluation indicators to measure the performance of the model in the bearing fault multi-classification task. Among them, accuracy is the basic indicator that reflects the overall correct recognition ratio of the model for all samples. Precision, recall, and F1 score judge the performance of the model from the perspective of category balance and overall sample recognition effect. The AUC curve and MCC can more robustly evaluate the classification stability of the model in data imbalance scenarios. By considering these six evaluation indicators, the evaluation bias caused by category shift of a single indicator can be avoided.

Detailed Experimental Setup. For all experiments in this study, we refer to some previous experimental setups [32][42][48]. Specifically, for the CWRU, PU, and uOttawa datasets, we did not add any noise, meaning the signal-to-noise ratio level is 0 dB. Then, we used stratified sampling to divide the data into training, validation, and test sets, with proportions of 60%, 20%, and 20%, respectively. This method ensures that the category distribution of each dataset is consistent with the original data, thus avoiding class imbalance, and also guarantees that the test samples are completely independent of the training samples, preventing the

risk of temporal data leakage. Subsequent spatio-temporal graph construction and model training were performed based on the training set for parameter learning. After model training, the trained model was tested using the test set, ensuring no cross-set information leakage throughout the entire process. In addition, all experiments were repeated 10 times independently, and the model parameters were reinitialized in each experiment. The final result was the mean of the 10 experiments to avoid the influence of random errors in a single experiment on the conclusion. The training and evaluation of all experiments were performed on a computer with an Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz, an NVIDIA GTX 1650 GPU, a Kingston 32GB memory, and Windows 10 Professional. **The model in this study is based on Python 3.9 (1) and is developed using open-source libraries such as PyTorch 2.2 (2), scikit-learn 1.5.2 (3), and NumPy 1.26.3 (4).**

(1) <https://www.python.org/downloads/release/python-390/>

(2) <https://pytorch.org/blog/pytorch2-2/>

(3) https://scikit-learn.org/1.5/whats_new/v1.5.html

(4) <https://numpy.org/doc/2.1/release/1.26.3-notes.html>

5.2 Comparison with Other Advanced Methods

To evaluate the performance of our proposed DBSGN on three bearing fault datasets, we selected six bearing fault diagnosis methods for comparative experiments with our method. We also calculated the performance of different bearing fault diagnosis methods under six evaluation indicators. The experimental results are shown in Table 4.

Table 4. Performance of different models under six evaluation metrics in three datasets. Bold and underlined lines indicate the best and second best results, respectively.

Dataset	Model	Evaluation Metric					
		Accuracy	Precision	Recall	F1-Score	AUC	MCC
CWRU	CNN	<u>99.72</u>	99.72	<u>99.72</u>	<u>99.72</u>	99.99	99.58
	SuperGraph	98.54	98.43	98.51	98.44	99.90	98.42
	WPT-FFT	99.66	99.59	99.59	99.59	<u>100.0</u>	99.62
	multi-scale	99.66	99.59	99.59	99.59	<u>100.0</u>	99.62
	CNN-Transformer	99.24	99.16	99.16	99.16	99.98	99.25
	MultiPatchTS	99.42	<u>99.73</u>	99.54	99.63	99.99	<u>99.6</u>
	T	99.42	<u>99.73</u>	99.54	99.63	99.99	<u>99.6</u>
HMDN	96.67	96.94	96.67	96.72	97.99	96.32	

				99.8		100.	99.7
	DBSGN	99.79	99.80	5	99.82	00	7
							74.0
	CNN	76.10	77.31	76.05	75.86	94.78	7
							95.9
	SuperGraph	97.22	97.44	96.67	96.91	99.88	0
							93.8
	WPT-FFT multi-scale	95.83	95.87	95.83	95.80	99.53	0
PU	CNN- Transformer	<u>99.68</u>	<u>99.73</u>	<u>99.74</u>	<u>99.73</u>	<u>99.99</u>	<u>99.8</u>
	MultiPatchTS						98.7
	T	98.32	98.63	98.89	98.75	99.94	4
							96.1
	HMDN	96.22	96.25	96.22	96.22	97.83	0
			100.0	100.		100.	100.
	DBSGN	100.00	0	00	100.00	00	00
							89.2
	CNN	91.06	92.17	91.06	90.80	99.04	0
							96.9
	SuperGraph	97.50	97.59	97.58	97.49	99.68	3
							98.9
	WPT-FFT multi-scale	99.17	99.17	99.17	99.16	99.99	6
uOttawa	CNN- Transformer	<u>99.58</u>	<u>99.59</u>	<u>99.58</u>	<u>99.58</u>	<u>99.99</u>	<u>99.5</u>
	MultiPatchTS						7
	T	98.67	98.67	98.67	98.67	99.78	2
							96.1
	HMDN	96.22	96.25	96.22	96.22	98.78	0
				99.8		100.	99.7
	DBSGN	99.83	99.83	4	99.84	00	9

Experimental results show that compared to bearing fault diagnosis methods such as CNN, SuperGraph, WPT-FFT, multi-scale CNN-Transformer, and MultiPatchTST, the proposed DBSGN model significantly improves fault diagnosis performance on these three datasets. On the CWRU dataset, compared to the baseline models, the accuracy of the DBSGN model increased by approximately 0.07%, and the precision, recall, F1-score, and MCC also increased by 0.07%, 0.13%, 0.1%, and 0.15%, respectively. On the PU dataset, all evaluation metrics of the DBSGN model reached 100.00%, showing a significant improvement compared to other baseline models. This indicates that our model has efficient recognition capabilities for this bearing fault dataset, even achieving zero-error recognition of fault samples on this dataset. Furthermore, on the uOttawa dataset, compared to the best baseline model multi-scale CNN-Transformer, the accuracy

of the DBSGN model increased by approximately 0.66%, and the precision, recall, F1-score, AUC, and MCC also increased by 0.66%, 0.67%, 0.67%, 0.01%, and 0.83%, respectively.

5.3 Further Analysis

To further validate the performance of our proposed DBSGN model in bearing fault detection, this paper conducts a detailed analysis of the model's class discrimination ability and probability prediction reliability from three perspectives: confusion matrix, ROC curve, and convergence curve. The experimental results are shown in Figures 2-4.

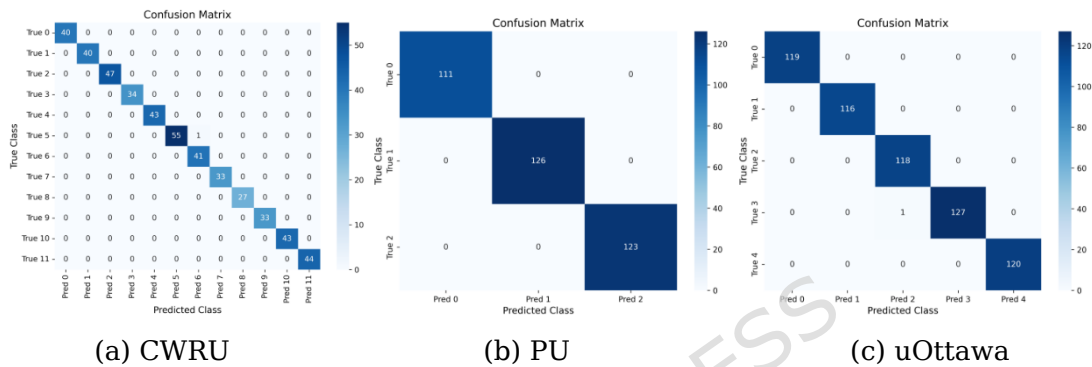


Fig. 2 Confusion matrix of three datasets

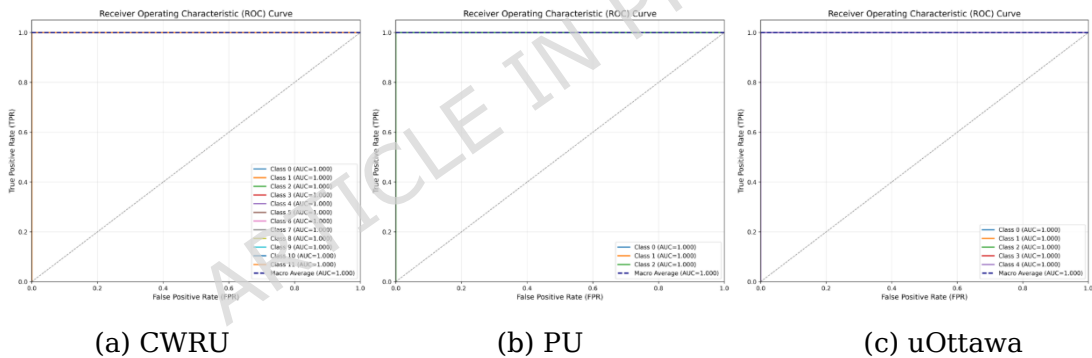


Fig. 3 ROC curves of three datasets

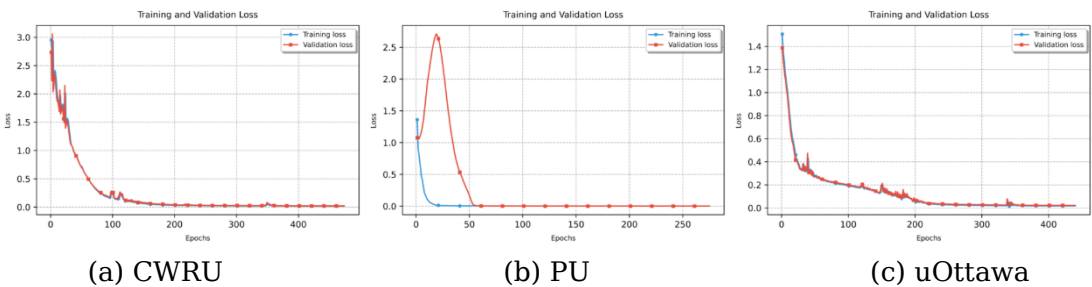


Fig. 4 Loss function convergence curves of three datasets

In the confusion matrix of Fig. 2, each row corresponds to the true category label, each column corresponds to the predicted category label, and the diagonal elements represent the number of correctly classified samples for each type of bearing condition. In the confusion matrix of the CWRU dataset, the number of samples on the main diagonal accounts for a very high proportion, while the number of samples on the off-diagonal accounts for a very low proportion,

indicating that the DBSGN model has high recognition accuracy for various types of faults in the CWRU dataset, and the vast majority of samples can be correctly classified into the corresponding categories. In the confusion matrix of the PU bearing dataset, all samples are completely concentrated on the main diagonal, indicating that the model can achieve zero-error classification performance for the faulty samples in this dataset. In the confusion matrix of the uOttawa dataset, only a very small number of samples have cross-category prediction bias, verifying that the model still maintains high classification accuracy in multi-category, fine-grained fault scenarios.

From the ROC curve results in Fig. 3, it can be seen that the ROC curves of the three bearing data sets are all close to the upper left corner of the coordinate system, and the AUC values of each category and macro average are all 1. This shows that the model's ability to distinguish between positive and negative categories for each type of fault has reached the theoretical optimal level, and the discriminatory degree of probability prediction is extremely strong. It can achieve perfect distinction between faults and non-faults, as well as between different faults, in data sets with different numbers of categories and different fault degrees.

Figure 4 shows the training and validation loss convergence curves of the proposed model on the CWRU, PU, and uOttawa bearing datasets. In the CWRU dataset's loss convergence curve, both training and validation losses decrease rapidly with the number of iterations and then stabilize, with minimal fluctuations and slight numerical differences. This indicates that the model's fitting process is stable on this dataset, without significant overfitting or underfitting issues. In the uOttawa dataset's loss convergence curve, both training and validation losses decrease rapidly initially and then remain at a low level. Although the validation loss shows slight fluctuations, the difference between it and the training loss remains small and the trends are consistent, verifying the model's stable convergence performance in multi-class, fine-grained fault scenarios. In the PU bearing dataset's loss convergence curve, the training loss converges rapidly to a stable level in the initial iterations, and the validation loss fluctuates briefly before decreasing to near zero and remaining stable. The two show highly consistent trends in the later stages of training, reflecting the model's fast convergence speed and strong generalization ability.

Based on the analysis of confusion matrix and ROC curve, the DBSGN model proposed in this paper demonstrates high-precision and high-robustness classification performance in bearing fault diagnosis tasks, and can provide reliable technical support for bearing health monitoring in industrial scenarios.

5.4 Ablation Study

In this section, we conduct detailed ablation experiments to evaluate the effectiveness of the proposed two-branch decision network. The rest of the experimental settings are consistent with those used in our previous comparative experiments.

In Sections 4.2 and 4.3, we proposed a dual-branch decision network, fused

the dual-branch output features through an adaptive attention decision fusion mechanism, and performed fault prediction based on the fusion results. To verify the effectiveness of the algorithm, a single Chibyshev graph convolutional network is used as a comparison method. The network removes the dual-branch structure and adaptive fusion mechanism and directly performs single-branch graph convolution prediction on graph data. Then, the fault prediction performance of the dual-branch decision network and the single network for vibration signals is evaluated on three bearing fault datasets: CWRU, PU and uOttawa. Fig. 5 demonstrates the importance of the dual-branch decision network. Experimental results show that on the CWRU dataset, our proposed DBSGN improves accuracy, precision, and recall by approximately 2% compared to GCN. On the PU dataset, DBSGN's accuracy, precision and other indicators all reached 100%, significantly outperforming the performance of GCN. On the uOttawa dataset, DBSGN achieves significant improvements in core indicators compared to GCN, with an average improvement of about 1.2%. This fully proves that the dual-branch decision network combined with the adaptive fusion mechanism is effective and can significantly enhance the recognition ability of bearing fault characteristics through differentiated feature learning and dynamic weight allocation.

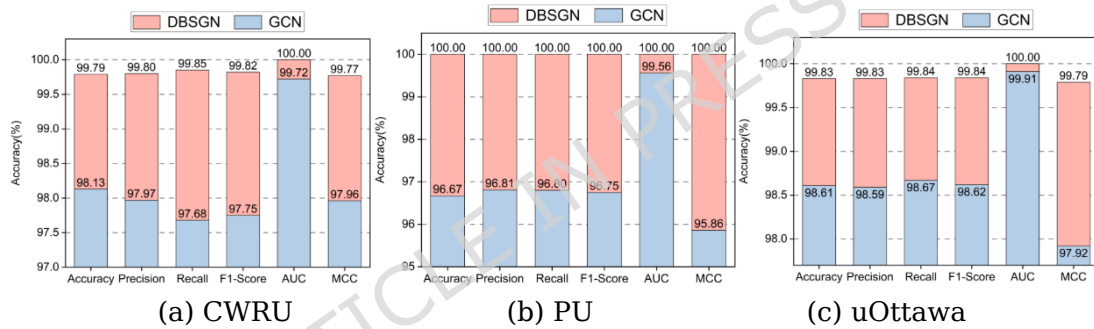


Fig. 5 Ablation study

We further analyzed the differences in computational complexity and efficiency of the model before and after ablation, and the experimental results are shown in Table 5. Table 5 compares the FLOPs, number of parameters, training time per sample, and time per iteration before and after ablation on the CWRU, PU, and uOttawa bearing fault datasets. The results show that the number of parameters, FLOPs, and training time per sample of DBSGN are all increased compared to GCN, and the increase varies with the heterogeneity of the dataset. The increase is most significant in the PU dataset, and the iteration time only increases in the PU dataset, remaining stable in CWRU and uOttawa. The increased computational cost brings significant gains in model performance. The core metrics on the CWRU dataset improved by approximately 2%, the accuracy on the PU dataset reached 100%, and the core metrics on the uOttawa dataset improved by an average of approximately 1.2%. This confirms the reasonable trade-off between performance and cost of the dual-branch structure and adaptive fusion mechanism. Further optimization of the lightweight design of the branches and batch processing strategies can be explored to balance model performance and computational efficiency, considering the graph structure characteristics of

the datasets.

Table 5. Computational Complexity and Efficiency of the Model Before and After Ablation

Datasets	Model	FLOPs/M	Sample Time/s	Iteration Time/s	Total Parameters
CWRU	GCN	41.330	0.00002	0.020	6,187
	DBSGN	133.710	0.00657	0.020	21,642
PU	GCN	1.900	0.00003	0.007	5,503
	DBSGN	382.260	0.01122	0.034	44,792
uOttawa	GCN	48.400	0.00332	0.020	5,655
	DBSGN	157.510	0.00642	0.020	20,060

5.5 Parameter Sensitivity Analysis

In order to study the effect of the hyperparameter α in Eq. 22 on the performance of DBSGN for bearing fault diagnosis, we set the variation range of α to 0-1 with an increment of 0.2, and the experimental results are shown in Fig. 6. Note that the other experimental settings remain the same as those used in the previous experiments.

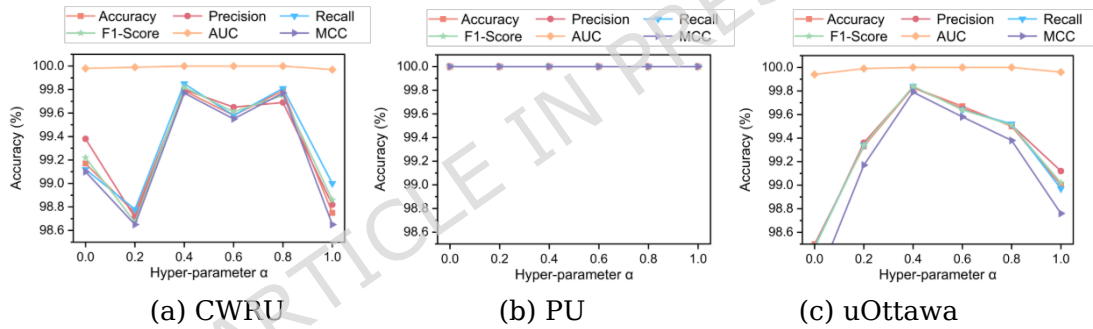


Fig. 6 Parameter sensitivity analysis

Experimental results show that on the CWRU and uOttawa datasets, when α is 0.4, the model's accuracy, precision, recall and other indicators reach the best. After deviating from this value, the indicators show a downward trend. However, on the PU dataset, when α varies from 0 to 1, the results of each evaluation indicator are all 100%. This phenomenon shows that in the CWRU and uOttawa datasets, the performance of the DBSGN model is more sensitive to α and needs to be fine-tuned to match the data distribution characteristics. In the PU dataset, however, changes in the hyperparameter α have no effect on the model. This may be because the data characteristics such as fewer fault samples and strong fault mode discrimination make the model less sensitive to α . Based on the experimental results of the above parameter sensitivity analysis, this paper finally sets the hyperparameter α in Eq. 22 to 0.4.

5.6 Visualize Feature Distribution

To intuitively verify the capabilities of the dual-branch decision network and

the dynamic attention-based decision fusion mechanism in feature learning, we used the t-SNE visualization method to analyze the feature distributions of the CWRU, PU, and uOttawa datasets. The experimental results are shown in Figures 7-9.

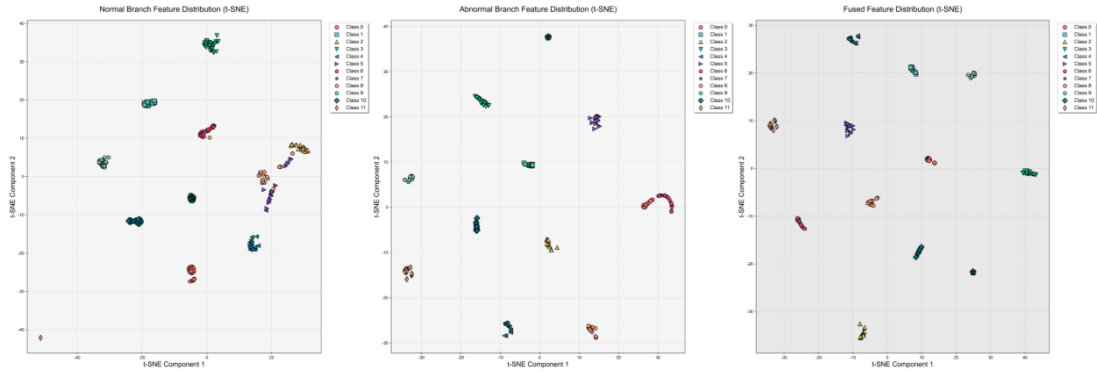


Figure 7. T-SNE feature visualization of the CWRU dataset

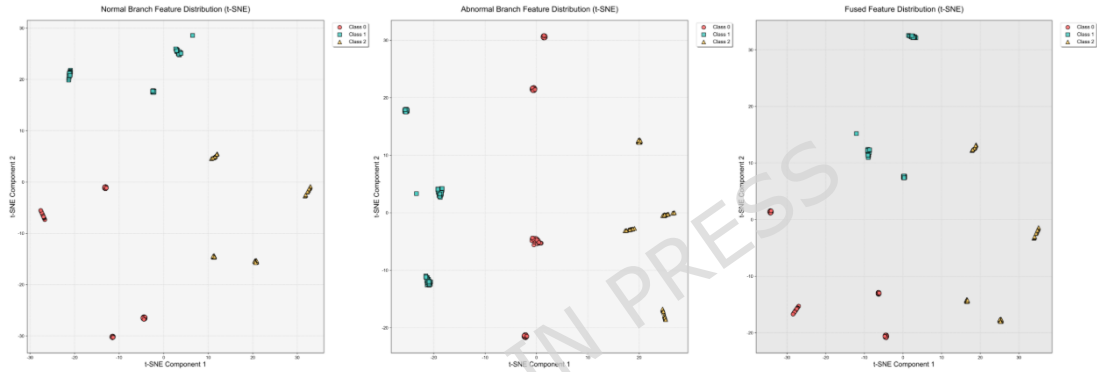


Figure 8. T-SNE feature visualization of the PU dataset

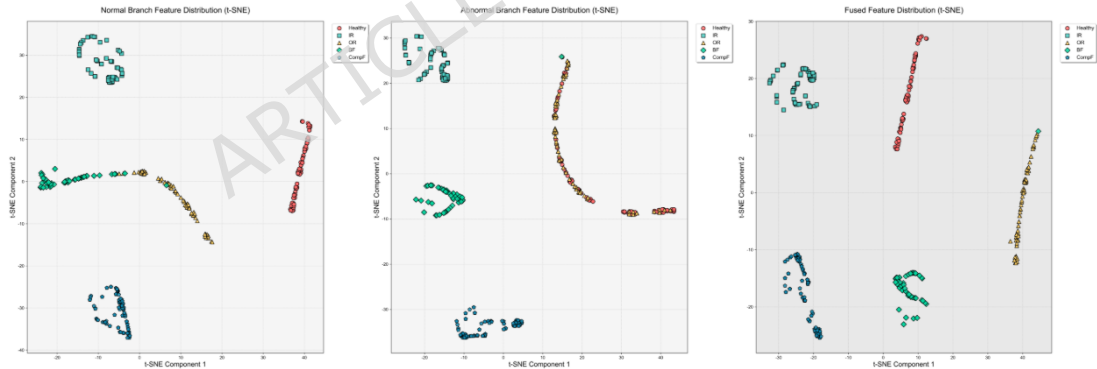


Figure 9. T-SNE feature visualization of the uOttawa dataset

Experimental results show that in the CWRU dataset, the normal branch can only achieve preliminary class separation for 12 types of samples, with significant overlap between feature clusters corresponding to similar faults; although the abnormal branch can strengthen the clustering of fault samples, it causes the normal samples to be scattered. By using an adaptive attention-based decision fusion mechanism to fuse the output features of the dual branches, the feature clusters of each category are completely separated, and the inter-class boundaries are clearly distinguishable. For the PU dataset, the results from the normal branch show marginal overlap in the feature clusters of a few fault categories, while the abnormal branch effectively improves the feature separability of all categories.

After fusion, the feature clusters of each category are perfectly separated, and this feature distribution characteristic is highly consistent with the 100% classification accuracy achieved by the model. In the uOttawa dataset, the results from the normal branch show feature cluster crossover among fault categories. Although the abnormal branch eliminates this crossover problem, it leads to a wider distribution range of the healthy sample cluster. After fusion, the feature clusters of each category achieve completely independent distribution, and the feature separability reaches the optimal level. In summary, the normal branch excels at learning and representing global features, while the abnormal branch focuses on the precise extraction of fault features. The decision fusion mechanism proposed in this study effectively integrates the learning advantages of both branches for normal and fault features, resulting in fused features that possess optimal intra-class clustering and inter-class separability. This fully validates the rationality and effectiveness of the DBSGN model structure design.

6. Conclusion and Future Work

This study proposes a dual-branch graph fusion network based on spatio-temporal graphs, which effectively solves the problems of insufficient stability and accuracy in the case of small sample data. The model first models the vibration signal based on spectral graph theory and spectral analysis method to construct a spatio-temporal graph. Secondly, Laplace-based spectral decomposition is used to extract the feature vectors of samples in the spatio-temporal graph. Based on the constructed spatio-temporal graph, we designed a dual-branch fusion network to train and verify the bearing data, and adjusted the model's learning of the bearing data through a dynamic attention mechanism. Extensive experiments show that the proposed model outperforms traditional models in terms of stability and accuracy compared with other models, verifying the effectiveness of the proposed method in fault diagnosis.

In this paper, our research mainly focuses on how to construct the vibration signal of the bearing into a graph data structure, but the construction of the graph data structure depends on manually preset thresholds. In addition, Laplace spectral decomposition has limited ability to characterize vibration signals. Therefore, in future work, new frequency domain feature extraction methods and graph data structure construction methods can be proposed to improve the model's representation of vibration signals.

Data Availability Statement

The datasets used in this study are openly available as follows: The CWRU dataset is available from the Case Western Reserve University Bearing Data Center website (<http://csegroups.case.edu/bearingdatacenter>). The PU dataset is provided by the University of Paderborn Bearing Data Center website

(<https://mb.uni-paderborn.de/kat/datacenter>). The uOttawa dataset is bearing vibration data collected in Ottawa under time-varying speed conditions and is available from Mendeley Data (<https://data.mendeley.com/datasets/v43hmbwaxpm/1>).

- [1] Singh, V., Gangsar, P., Porwal, R., & Atulkar, A. (2023). Artificial intelligence application in fault diagnostics of rotating industrial machines: A state-of-the-art review. *Journal of Intelligent Manufacturing*, 34(3), 931-960.
- [2] Nath, A. G., Udmale, S. S., & Singh, S. K. (2021). Role of artificial intelligence in rotor fault diagnosis: A comprehensive review. *Artificial Intelligence Review*, 54(4), 2609-2668.
- [3] Zhang, Q., & Deng, L. (2023). An intelligent fault diagnosis method of rolling bearings based on short-time Fourier transform and convolutional neural network. *Journal of Failure Analysis and Prevention*, 23(2), 795-811.
- [4] Tao, H., Wang, P., Chen, Y., Stojanovic, V., & Yang, H. (2020). An unsupervised fault diagnosis method for rolling bearing using STFT and generative neural networks. *Journal of the Franklin Institute*, 357(11), 7286-7307.
- [5] Chegini, S. N., Bagheri, A., & Najafi, F. (2019). Application of a new EWT-based denoising technique in bearing fault diagnosis. *Measurement*, 144, 275-297.
- [6] Li, L., Guo, A., & Chen, H. (2021). Feature extraction based on EWT with scale space threshold and improved MCKD for fault diagnosis. *IEEE Access*, 9, 45407-45417.
- [7] Sun, Y., Li, S., & Wang, X. (2021). Bearing fault diagnosis based on EMD and improved Chebyshev distance in SDP image. *Measurement*, 176, 109100.
- [8] Shah, A. K., Yadav, A., & Malik, H. (2018). EMD and ANN based intelligent model for bearing fault diagnosis. *Journal of Intelligent & Fuzzy Systems*, 35(5), 5391-5402.
- [9] Pang, B., Tang, G., Tian, T., & Zhou, C. (2018). Rolling bearing fault diagnosis based on an improved HTT transform. *Sensors*, 18(4), 1203.
- [10] Chen, X., Zhang, B., & Gao, D. (2021). Bearing fault diagnosis base on multi-scale CNN and LSTM model. *Journal of Intelligent Manufacturing*, 32(4), 971-987.
- [11] Sinitsin, V., Ibryaeva, O., Sakovskaya, V., & Eremeeva, V. (2022). Intelligent bearing fault diagnosis method combining mixed input and hybrid CNN-MLP model. *Mechanical Systems and Signal Processing*, 180, 109454.
- [12] Liu, H., Zhou, J., Zheng, Y., Jiang, W., & Zhang, Y. (2018). Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA transactions*, 77, 167-178.
- [13] Joungh, B. G., Nath, C., Li, Z., & Sutherland, J. W. (2024). Bearing anomaly detection in an air compressor using an LSTM and RNN-based machine learning model. *The International Journal of Advanced Manufacturing Technology*, 134(7), 3519-3530.

- [14]Li, Y., Liu, X., Hu, J., Liang, P., Wang, B., Yuan, X., & Zhang, L. (2025). Graph optimization algorithm enhanced by dual-scale spectral features with contrastive learning for robust bearing fault diagnosis. *Knowledge-Based Systems, 315*, 113275.
- [15]Chen, Y., Feng, G., Chen, H., Gou, L., & Wang, S. (2025). A machine learning-based interpretable fault diagnosis method for rolling bearings under unbalanced samples. *Engineering Applications of Artificial Intelligence, 158*, 111505.
- [16]Wan, S., Liu, J., Li, X., Zhang, Y., Yan, K., & Hong, J. (2022). Transfer-learning-based bearing fault diagnosis between different machines: A multi-level adaptation network based on layered decoding and attention mechanism. *Measurement, 203*, 111996.
- [17]Cui, B., Weng, Y., & Zhang, N. (2022). A feature extraction and machine learning framework for bearing fault diagnosis. *Renewable Energy, 191*, 987-997.
- [18]Wang, H., Zheng, J., & Xiang, J. (2023). Online bearing fault diagnosis using numerical simulation models and machine learning classifications. *Reliability Engineering & System Safety, 234*, 109142.
- [19]Habbouche, H., Amirat, Y., Benkedjoh, T., & Benbouzid, M. (2021). Bearing fault event-triggered diagnosis using a variational mode decomposition-based machine learning approach. *IEEE Transactions on Energy Conversion, 37*(1), 466-474.
- [20]Brusamarello, B., Da Silva, J. C. C., de Moraes Sousa, K., & Guarneri, G. A. (2022). Bearing fault detection in three-phase induction motors using support vector machine and fiber Bragg grating. *IEEE Sensors Journal, 23*(5), 4413-4421.
- [21]Hoang, D. T., & Kang, H. J. (2019). A survey on deep learning based bearing fault diagnosis. *Neurocomputing, 335*, 327-335.
- [22]Neupane, D., & Seok, J. (2020). Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *Ieee Access, 8*, 93155-93178.
- [23]Hatipoğlu, A., Süpürtülü, M., & Yılmaz, E. (2025). Enhanced fault classification in bearings: A multi-domain feature extraction approach with LSTM-attention and LASSO. *Arabian Journal for Science and Engineering, 50*(14), 10795-10812.
- [24]Rezazadeh, N., De Oliveira, M., Lamanna, G., Perfetto, D., & De Luca, A. (2025). WaveCORAL-DCCA: A Scalable Solution for Rotor Fault Diagnosis Across Operational Variabilities. *Electronics, 14*(15), 3146.
- [25]Du, Y., Cao, Y., Wang, H., & Li, G. (2024). A Rolling Bearing Fault Diagnosis Method Combining MSSSA-VMD with the Parallel Network of GASF-CNN and BiLSTM. *Lubricants, 12*(12), 452.
- [26]Li, Y., Gu, X., & Wei, Y. (2024). A Deep Learning-Based Method for Bearing Fault Diagnosis with Few-Shot Learning. *Sensors, 24*(23), 7516.

- [27]Ma, Z., & Guo, H. (2024). Fault diagnosis of rolling bearing under complex working conditions based on time-frequency joint feature extraction-deep learning. *Journal of Vibroengineering*, 26(7), 1635-1652.
- [28]Xie, Y., Zhang, L., Liu, H., & Hu, C. (2025). Imbalanced few-shot bearing fault diagnosis via hybrid enhanced VAE-WGAN and attention-guided WDCNN. *Measurement Science and Technology*, 36(8), 086134.
- [29]Tong, Q., Lu, F., Feng, Z., Wan, Q., An, G., Cao, J., & Guo, T. (2022). A novel method for fault diagnosis of bearings with small and imbalanced data based on generative adversarial networks. *Applied Sciences*, 12(14), 7346.
- [30]Liu, Y., Jiang, H., Liu, C., Yang, W., & Sun, W. (2022). Data-augmented wavelet capsule generative adversarial network for rolling bearing fault diagnosis. *Knowledge-Based Systems*, 252, 109439.
- [31]Liu, S., Jiang, H., Wu, Z., & Li, X. (2022). Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis. *Mechanical Systems and Signal Processing*, 163, 108139.
- [32]Yang, C., Zhou, K., & Liu, J. (2021). SuperGraph: Spatial-temporal graph-based feature extraction for rotating machinery diagnosis. *IEEE Transactions on Industrial Electronics*, 69(4), 4167-4176.
- [33]Yu, Z., Zhang, C., & Deng, C. (2023). An improved GNN using dynamic graph embedding mechanism: A novel end-to-end framework for rolling bearing fault diagnosis under variable working conditions. *Mechanical Systems and Signal Processing*, 200, 110534.
- [34]Zhang, Z., & Wu, L. (2024). Graph neural network-based bearing fault diagnosis using Granger causality test. *Expert Systems with Applications*, 242, 122827.
- [35]Jiang, L., Li, X., Wu, L., & Li, Y. (2022). Bearing fault diagnosis method based on a multi-head graph attention network. *Measurement Science and Technology*, 33(7), 075012.
- [36]Sun, S., Xia, X., & Zhou, H. (2025). Bearing fault diagnosis under time-varying speeds with limited samples using frequency temporal series graph and graph generative classified adversarial networks. *Neurocomputing*, 130613.
- [37]Xiao, X., Li, C., Huang, J., Yu, T., & Wong, P. K. (2023). An improved graph convolutional networks for fault diagnosis of rolling bearing with limited labeled data. *Measurement Science and Technology*, 34(12), 125109.
- [38]Rezazadeh, N., De Luca, A., Perfetto, D., Lamanna, G., Annaz, F., & De Oliveira, M. (2025). Domain-Adaptive Graph Attention Semi-Supervised Network for Temperature-Resilient SHM of Composite Plates. *Sensors*, 25(22), 6847.
- [39]Kipf, T. N. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- [40]Hammond, D. K., Vandergheynst, P., & Gribonval, R. (2011). Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2), 129-150.

- [41]Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29.
- [42]Liao, J., Deng, Y., Xie, X., & Zhang, Z. (2024). Research on Chebyshev graph convolutional neural network modeling method for rotating equipment fault diagnosis under variable working conditions. *Applied Sciences*, 14(20), 9208.
- [43]Wu, Z., Long, Z., Luo, C., Wang, S., & Ma, X. (2025). DRL-GCNet: A Deep Reinforcement learning and Graph Convolutional Network for Harmonic Drive Fault Diagnosis. *IEEE Transactions on Instrumentation and Measurement*.
- [44]CWRU, Case Western Reserve University Bearing Data Center Website, Cleveland, OH, USA. Available online: <http://csegroups.case.edu/bearingdatacenter/home>.
- [45]Lessmeier, C., Kimotho, J. K., Zimmer, D., & Sextro, W. (2016, July). Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In *PHM society European conference* (Vol. 3, No. 1).
- [46]Huang, H., & Baddour, N. (2018). Bearing vibration data collected under time-varying rotational speed conditions. *Data in brief*, 21, 1745-1749.
- [47]Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12), 6999-7019.
- [48]Aburakhia, S. A., Myers, R., & Shami, A. (2022). A hybrid method for condition monitoring and fault diagnosis of rolling bearings with low system delay. *IEEE Transactions on Instrumentation and Measurement*, 71, 1-13.
- [49]Xu, Z., Tao, J., Hu, Y., Feng, H., & Ma, L. (2025). A WOA-SVMD and multi-scale CNN-transformer method for fault diagnosis of motor bearing. *Measurement and Control*, 00202940241312665.
- [50]Ko, S., & Lee, S. (2025). Multi-Patch Time Series Transformer for Robust Bearing Fault Detection with Varying Noise. *Applied Sciences*, 15(3), 1257.
- [51]Yan, H., Tan, J., Luo, Y., Wang, S., & Wan, J. (2024). Multi-Condition Intelligent Fault Diagnosis Based on Tree-Structured Labels and Hierarchical Multi-Granularity Diagnostic Network. *Machines*, 12(12), 891.