Coupling Fault Diagnosis of Bearings Based on Hypergraph Neural Network

Shenglong Wang [ID], Xiaoxuan Jiao *, Bo Jing, Jinxin Pan, Xiangzhen Meng [ID], Yifeng Huang and Shaoting Pei

Aeronautics Engineering College, Air Force Engineering University, Xi'an 710038, China;
phm_wsl@outlook.com (S.W.); jingbo_sensors@outlook.com (B.J.); panjinxin_sensors@outlook.com (J.P.);
mengxz_sensor@outlook.com (X.M.); huangyiff@189.cn (Y.H.); pst_phm@outlook.com (S.P.)
* Correspondence: jiaoxx_sensor@outlook.com; Tel.: +86-02984787328

Abstract: Coupling faults that simultaneously occur during the operation of mechanical equipment are widespread. These faults encompass a diverse range of high-order coupling relationships, involving multiple base fault types. Based on the advantages of hypergraphs for higher-order relationship descriptions, two coupling fault diagnosis architectures based on the hypergraph neural network are proposed in this paper: 1. In the coupling fault diagnosis framework based on feature generation, the base faults serve as the hypergraph nodes, and each hyperedge connects the base faults. The generator, which consists of the hypergraph neural network, generates coupling faults as negative samples to enforce regularization constraints for the discriminator training. 2. In the coupling fault diagnosis framework based on feature extraction, each node represents a fault mode, and each hyperedge connects nodes with common failure modes. The multi-head attention mechanism extracts the features of base faults, and the common fault features in a hyperedge are aggregated via the hypergraph neural network. The inner product correlation is used to diagnose the fault modes. The results show that the diagnostic accuracy for coupling faults with the two frameworks reaches 88.6% and 86.76%, respectively. Both frameworks can be used for the diagnosis and analysis of high-order coupling faults.

Keywords: coupling fault diagnosis; feature generation; feature extraction; hypergraph networks

1. Introduction

Coupling faults are widespread in mechanical equipment. They are formed via the coupling of multiple types of base faults. Coupling faults are not linear superpositions of base faults, and complex fault characteristics emerge in the coupling process [1]. Therefore, simple linear decoupling cannot accurately diagnose coupling faults, and a more accurate nonlinear model needs to be established to describe the fault coupling process [2]. In this study, two kinds of data-driven methods are used to obtain the fault characteristics of multi-fault coupling, reveal the fault coupling process, and classify the coupling faults based on the obtained coupling fault characteristics.

The fault characteristics of rotating machinery can be obtained using spectral analysis [3]. Scholars have mostly used traditional frequency domain feature extraction methods [4]. Ma et al. [5] proposed adaptive demodulation analysis to extract early bearing fault features. Jiang et al. [6] used a short-time Fourier transform to obtain the bearing amplitude and used adaptive weights to extract the weighted characteristic frequency. Zhu et al. [7] determined the hyperparameters of the spectral amplitude modulation (SAM) model by taking the maximum harmonic significance as an index. They divided the frequency band and reconstructed the signal to suppress the noise component in the fault characteristics. Frequency domain features analyze all frequency bands in the spectrum, and the energy of fault features is not concentrated. Several scholars have used time–frequency domain feature extraction methods to extract the fault features of energy-concentrated frequency bands [8]. Yue et al. [9] proposed multi-scale wavelet networks

to extract multi-scale features from vibration signals and used meta-learning modules to learn the distance distribution between data to achieve model generalization. This method establishes a distance-based data space, but the definition of distance lacks explainability. Tang et al. [10] used a time–frequency domain convolutional neural network to extract the fault features of multiple energy concentrations and introduced an attention module to extract weighted representative features. Jia et al. [11] proposed a Gramian-based noise reduction strategy for noise suppression and feature extraction. Yoo et al. [12] carried out convolutional neural network (CNN) image processing based on a spectrum diagram to extract the fault features in an image. These methods extract the features of a single fault from the spectrum, but there is no effective solution to extract the features of the coupled fault. Based on the prior fault information [13], the extracted feature acquisition method decouples the fault coupling process and fits the basic fault correlation weight, which belongs to the Bayes school category.

The data-driven fault feature generation method can avoid the difficulty of spectrum extraction caused by the lack of working condition information [14]. Zhang et al. [15] generated fault features adaptively via convolutional neural networks and used support vector machine (SVM) classifiers to classify the fault types. Karnavas et al. [16] proposed using two attention mechanisms to extract global and local features, respectively, and fuse these two types of features. Sivapriya et al. [17] used multi-scale and different resolution convolution check fault information to extract and fuse it into global fault features for fault diagnosis. Yang et al. [18] reconstructed fault features via IMFs extracted from variational mode decomposition (VMD) and used SVMs for fault classification given the unbalanced and strong noise characteristics of bearing signals. Huo et al. [19] carried out the adaptive extraction of time–frequency domain features and combined harmonic detection with noise signals to continue reconstructing the features. Zhao et al. [20] reduced the dimensionality of features using depth-separable convolutional blocks and used the global feature awareness module to perform adaptive weighting processing similar to the attention mechanism for the feature signal channels to obtain the fault features. Yang et al. [21] reconstructed coupling fault features using an autoencoder to achieve feature dimension reduction and non-equilibrium data strengthening. Jang et al. [22] combined an adversarial network based on an autoencoder to generate the original data stream. Most of these generative methods reconstruct the original data, but the reconstructed data lack interpretability. Yu et al. [23] proposed a zero-sample fault diagnosis method. The digital twin method was used to generate fault data based on the health state model. The data space generated with this method was uncontrollable and could not achieve accurate fault classification. The generative feature acquisition method reconstructs the critical fault information from the existing fault information and estimates the overall probability by observing many samples; this belongs to the frequency school.

Both extraction and generative methods obtain the fault characteristics of independent bearing faults [24]. Non-Euclidean data-oriented fault diagnosis methods should be used for complex coupling faults that are widespread in bearings [25,26]. A graph neural network constructs the topological structure of the data via the nodes and the edges connecting the nodes, which can realize the correlation description of non-Euclidean data. A traditional graph neural network describes the relationship between node pairs [27]. Some scholars have used graph neural networks to extract fault features and perform fault diagnosis. Li et al. [28] established a graph convolutional network (GCN) with multiple receptive fields to conduct adaptive correlation analysis on vibration data after FFT and took data with different receptive fields as the fault characteristics. Wang et al. [29] used a bidirectional graph neural network to carry out correlation analysis of the graph data extracted from a residual network at the instance and distribution levels. This classified independent faults, but the network needs to be further optimized for correlation analysis between coupled faults. Wang et al. [30] added dynamic vertices to a graph neural network and realized coupling fault diagnosis by establishing a pair relationship topology between faults and dynamic vertices.

It is difficult for graph neural networks to achieve high-level expressions of the relationships between objects [31]. Feng et al. [32] proposed a hypergraph neural network (HGNN) to solve this problem. Bai et al. [33] introduced convolution and attention mechanisms into hypergraph neural networks, further improving the ability of representation learning in hypergraph networks. Hypergraph neural networks are mainly used in recommendation systems and multi-modal data feature extraction and have been widely used in image processing. Shi et al. [34] built unlabeled data into hypergraphs, mined the data's higher-order information, and combined an autoencoder to realize representation learning and fault classification. In this method, the hypergraph structure is taken as an autoencoder layer, and the fault features are generated from the data-driven perspective, which lacks interpretability. Su et al. [35] extracted the higher-order relationship between the same type of faults and different types of faults via the hypergraph structure. These methods are applied to independent fault diagnosis. In the field of coupled fault diagnosis, the high-order relation representation ability of hypergraph neural networks meets the requirement of complex coupled fault feature acquisition. In this paper, two fault diagnosis architectures based on hypergraph neural networks are proposed: a coupled fault diagnosis architecture based on feature generation and another based on feature extraction. The method is verified with coupling fault injection experimental data.

The main insights and contributions of this study are summarized as follows:

(a) A rotary mechanical coupling fault injection test is designed. Eight types of faults including coupling faults and independent faults are injected, and the experimental data are collected.

(b) A coupled fault diagnosis architecture based on feature extraction is proposed. A hypergraph generative adversarial network (HGGAN) is established and vibration data generation and coupling fault diagnosis are realized by applying hypergraph theory to the generative adversarial network (GAN) model.

(c) A coupled fault diagnosis architecture based on feature generation is proposed. The coupling fault characteristics are extracted via a multi-head inner product hypergraph attention network (IPHGAT), and coupling fault diagnosis and analysis are realized.

The rest of this paper is organized as follows: Section 2 introduces the preliminary knowledge, including the graph attention network, the hypergraph attention network, and the generative adversarial network. Section 3 introduces the two fault diagnosis frameworks based on a hypergraph attention network. The coupled fault injection experiment and data-preprocessing methods are introduced in Section 4. The two coupled fault diagnosis frameworks are conducted in Section 5. Finally, conclusions are drawn in Section 6.

2. Preliminary Knowledge

2.1. Graph Attention Network

In graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, $\mathcal{E}$ is the set of edges, and $\mathcal{V}$ is the set of nodes. In the GCN, one step of graph convolution is defined as follows [36]:

$$X^{l+1} = \sigma\left(\widetilde{D}^{-\frac{1}{2}} A \widetilde{D}^{-\frac{1}{2}} X^l W^l\right) \tag{1}$$

where A is the adjacency matrix, $\widetilde{D}$ is the degree matrix with a self-loop, and W is the linear transformation weight matrix. The graph attention network (GAT) adds the attention mechanism into the graph convolutional network and uses the dynamic adjacency matrix to perform the adaptive values of the edge vector according to the node data characteristics. The calculation method is as follows [37]:

$$A_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(a^T\left[h_i W \| h_j W\right]\right)\right)}{\sum_{k \in \mathcal{N}_i} \exp\left(LeakyReLU\left(a^T\left[h_i W \| h_k W\right]\right)\right)} \tag{2}$$

The GAT only describes the dynamic edges between pairwise nodes. Higher-order representation methods are needed for modeling processing for the multiple interactions in coupling faults.

2.2. Hypergraph Attention Network

The hypergraph convolutional network (HGCN) connects multiple nodes via hyperedges, which extract higher-order relationships between the nodes. The inter-layer information transmission mode is as follows:

$$X^{l+1} = D_v^{-1/2} A H D_e^{-1} A^T D_v^{-1/2} X^l W^l \tag{3}$$

where A is the adjacency matrix, D_v is the node degree matrix, D_e is the hyperedge degree matrix, H is the weight diagonal matrix of the hyperedge, and W is the linear transformation weight matrix. The hypergraph attention network (HGAT) computes the values in the adjacency matrix A, and the correlations between nodes are quantitatively described. The computation of the attention mechanism in the hypergraph structure is defined as follows:

$$A_{ij} = \frac{\exp\left(\text{LeakyReLU}(a^T [x_i W \| e_j W])\right)}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(a^T [x_i W \| e_k W]))} \tag{4}$$

where $\mathcal{N}_i$ is the neighborhood set of x_i, $[, \|,]$ is the concatenation operation, and a is the weight vector. The similarity of the two vectors can be acquired after concatenation calculation.

The attention score in the GAT describes the correlation between a pair of nodes, while the attention score in the HGAT is the correlation between nodes and hyperedges. Therefore, applying an attention mechanism in hypergraphs requires nodes and hyperedges to be in the same homogeneous domain.

2.3. Generative Adversarial Network

The adversarial generation network models the data probability distribution via the game between the generator model $G(z)$ and discriminator model $D(x)$ [38]. The input of the generator is a random vector z, and the loss function of the generator model is as follows:

$$Loss_G = \frac{1}{M_{fake}} \sum_{i=1}^{M_{fake}} \log(1 - D(G(z^{(i)}))) \tag{5}$$

where M_{fake} is the number of fake data samples. The input of the discriminator is x, and the loss function is as follows:

$$Loss_D = -\frac{1}{S} \sum_{i=1}^{M_{real}} y_{real}^i \log(D(x^i)) - \frac{1}{S} \sum_{i=1}^{M_{fake}} (1 - y_{fake}^i) \log(1 - D(G(z^i))) \tag{6}$$

where $S = M_{real} + M_{fake}$ is the total number of false samples and true samples, and M_{real} is the number of true samples. The first half of the loss function requires the discriminator to be true to the real data, and the second half requires the discriminator to be false to the false data.

3. Algorithm Flow

Coupling faults are formed by coupling multiple types of base faults [39], which interact with each other. Therefore, the fault characteristics of coupling faults are not simple linear superpositions of base faults but nonlinear complex coupling. In this study, the vibration characteristics of coupled faults are not decomposed from the complex physical mechanism, but the advantages of the hypergraph neural network in processing higher-order relationships are used to model the coupling faults.

3.1. Algorithm Flow of HGGAN

From the perspective of holism, the coupling faults of mechanical equipment comprise many kinds of faults, which show unified fault characteristics. Assuming that the fault signals of coupled faults are generated by the implicit nonlinear coupling between the fault signals of the base faults, this study uses a hypergraph neural network to model the nonlinear generation process and form the coupled fault diagnosis model, HGGAN. In this model, the graph node set $\mathcal{V}$ is composed of M, many base faults, i.e., $|\mathcal{V}| = M$. Each base fault and coupling fault constitute N, many hyperedges, i.e., $|\mathcal{E}| = N$, where $M \leq N$, which is reduced to equality if, and only if, all nodes are base faults.

A coupled fault diagnosis architecture based on feature generation is designed in this study based on the ability of the GAN to generate a probability distribution of the data. The generator consists of the HGCN, and the discriminator is composed of a multi-layer perceptron (MLP). The algorithm flow is shown in Figure 1.
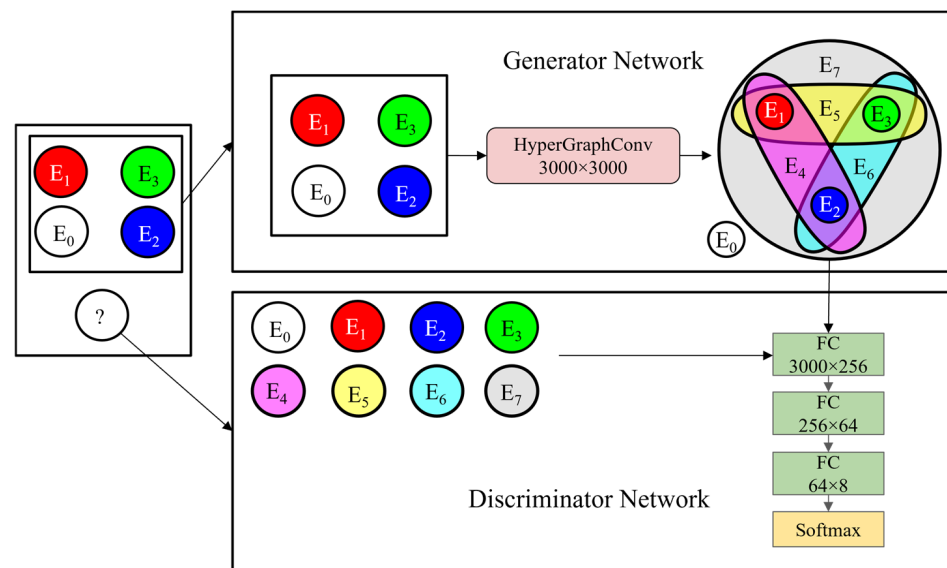
Figure 1. HGGAN structure.

In the HGGAN model proposed in this paper, the generator's loss function is rewritten:

$$Loss_G = \frac{1}{S}\sum_{i=1}^{S}\sum_{j=1}^{M}\log(1 - D(G(x_j^i))) \tag{7}$$

The loss of the generator is the classification error of the fault data coupled with the base fault, where S is the total number of samples. Compared with the traditional GAN, the input of the generator changes from a random vector z to the fault feature x of the base fault in the graph data $x_j, j \in 1, \cdots, M$, which improves the model's stability compared with random data [40]. Both true and false samples come from the same graph data sample, so the numbers of true samples and false samples are equal.

The loss function of the discriminator is rewritten as follows:

$$Loss_D = -\frac{1}{S}\sum_{i=1}^{S}y_{M+1}^i\log(D(x_{M+1}^i)) - \frac{\alpha}{S}\sum_{i=1}^{S}\sum_{j=1}^{M}(1 - y_j^i)\log(1 - D(G(x_j^i))) \tag{8}$$

The rewritten discriminator loss function is still composed of two parts. In the first part, the input x_{M+1} of the discriminator is the dynamic vertex in the graph data, that is, the fault feature of the coupling fault to be diagnosed, and its label y_{M+1} is the fault type corresponding to the coupling fault x_{M+1}. The second part is the regularization term to prevent the model from overfitting, where the hyperparameter α is set as 0.1 in this article.

The HGGAN established in this paper generates coupling fault features from the perspective of data generation. A hypergraph convolution network is used as the generator, which avoids the uncertainty caused by random vectors in a traditional GAN and provides the network structure with interpretability.

3.2. Algorithm Flow of IPHGAT

From the perspective of reductionism, the coupling faults of mechanical equipment comprise multiple base faults. The signals of coupling faults can be decomposed into those belonging to the base faults. In this framework, a coupling fault diagnosis model based on IPHGAT is established. All the fault modes, including base faults and coupling faults, are taken as graph nodes, and $|\mathcal{V}| = M$. Hyperedges are the fault characteristics of all fault modes after aggregation, and $|\mathcal{E}| = N$. The number of nodes and hyperedges are equal, i.e., $M = N$. Therefore, this study applies an attention mechanism to further analyze and process the fault information. The correlation between each base fault and each type of fault is obtained via the attention mechanism.

In the multi-head attention mechanism, the number of heads is defined as the number of independent base failures, and the multi-head attention scores are calculated as follows:

$$A_{ij}^k = \frac{\exp\left(\text{LeakyReLU}(a^{k^T}[x_i W^k \| e_j W^k])\right)}{\sum\limits_{n \in \mathcal{N}_i} \exp\left(\text{LeakyReLU}(a^{k^T}[x_i W^k \| e_n W^k])\right)} \tag{9}$$

where x is the node vector, and e is the hyperedge vector. The trainable parameter a^k is the direction of the attention calculation corresponding to the k-th head. The output of the l-th layer network is calculated according to the multi-head attention extraction features:

$$x_j^{l+1} = \sigma\left(\|_{k=1}^m \frac{1}{|\mathcal{N}_j|} \sum_{i \in \mathcal{N}_j} \alpha_{ij}^{l,k} W^l x_i^l\right) \tag{10}$$

In the hypergraph attention mechanism proposed in this paper, the updated coupled fault node vector is used to update the hyperedge vector, i.e., $e_j^{l+1} = x_j^l$.

The network structure is shown in Figure 2. The algorithm flow of the IPHGAT model is as follows.

Step 1: The constructed hypergraph data $X = [X^{(0)}, X^{(1)}] \in \mathbb{R}^{(M+1) \times Q}$ are inputted, where Q is the data feature dimension, and X includes the M base fault node data $X^{(0)} \in \mathbb{R}^{M \times Q}$ and the fault $X^{(1)} \in \mathbb{R}^Q$ to be predicted located at a dynamic vertex. The corresponding labels are $Y^{(0)} = [1, \cdots, M]$, $Y^{(1)} \in R$. The $X^{(0)}$ passes through the first layer of $m - head$ HGAT, generates m features for each hyperedge, splices the m features of each hyperedge into a whole as the features of the hypergraph, and performs feature standardization to obtain $X_1^{(0)}$.

Step 2: $X_1^{(0)}$ is inputted into the second layer of $m - head$ HGAT to obtain $X_2^{(0)}$, and then input $X_2^{(0)}$ into the fully connected network and Softmax layer to obtain the fault feature $\hat{Y}^{(0)} \in \mathbb{R}^{m \times m}$ corresponding to each hyperedge. The cross-entropy of $\hat{Y}^{(0)}$ and $Y^{(0)}$ is regarded as the loss function $\mathfrak{L}_1$.

Step 3: $X^{(1)}$ is inputted into the three-layer MLP to obtain the fault feature $\hat{Y}^{(1)} \in \mathbb{R}^m$ of the node to be predicted. The cross-entropy of $\hat{Y}^{(1)}$ and $Y^{(1)}$ is set as the loss function $\mathfrak{L}_2$.

Step 4: The inner product between the vector $\hat{Y}^{(1)}$ and $\hat{Y}^{(0)}$ corresponding to each hyperedge is calculated as $C = [c_1, \cdots, c_M]$, the label $\hat{Y}$ of the hyperedge corresponding to the vector with the largest inner product value is taken, and the model accuracy rate is obtained by judging whether $\hat{Y}$ and $\hat{Y}^{(1)}$ are equal.

The loss function of the model consists of two parts, $\mathfrak{L} = \mathfrak{L}_1 + \mathfrak{L}_2$, where $\mathfrak{L}_1 = CrossEntropy(Y^{(0)}, \hat{Y}^{(0)})$ and $\mathfrak{L}_2 = CrossEntropy(Y^{(1)}, \hat{Y}^{(1)})$. The value of the corre-

sponding fault mode's hyperedge vector tends to increase via the model training process, thus driving the inner product to increase.
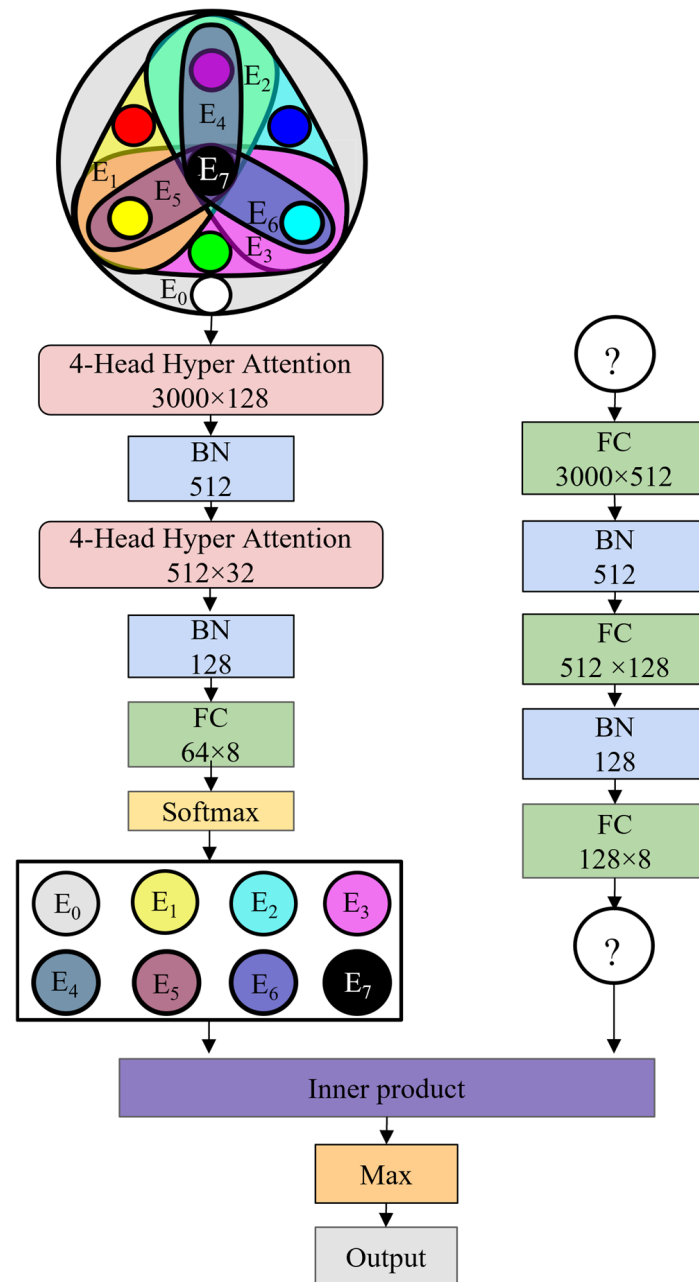

Figure 2. IPHGAT network structure.

4. Data Acquisition and Preprocessing

In the experiments in this study, 8 faults were injected into the UCPH205 bearing. The test bench is shown in Figure 3. The samples included 1 normal bearing, 4 outer-race fault bearings, 4 inner-race fault bearings, 4 ball fault bearings, 1 outer-race + inner-race fault bearing, 1 outer-race + ball fault bearing, 1 inner-race + ball fault bearing, 1 outer-race + inner-race + ball fault bearing, and 1 outer-race + inner-race + ball fault bearing, as shown in Figure 4. The fault injection experiment was carried out at 2400 rpm for each bearing. A sample rate of 20 kHz was used to collect the experimental data, and the fault size of the outer race, the inner race, and the ball was 0.3 mm.

Figure 3. Coupling fault test bench.

(**a**) Normal　　(**b**) Outer race　　(**c**) Inner race　　(**d**) Ball

(**e**) Outer race + inner race　　(**f**) Outer race + ball　　(**g**) Inner race + ball　　(**h**) Outer race + Inner race + ball
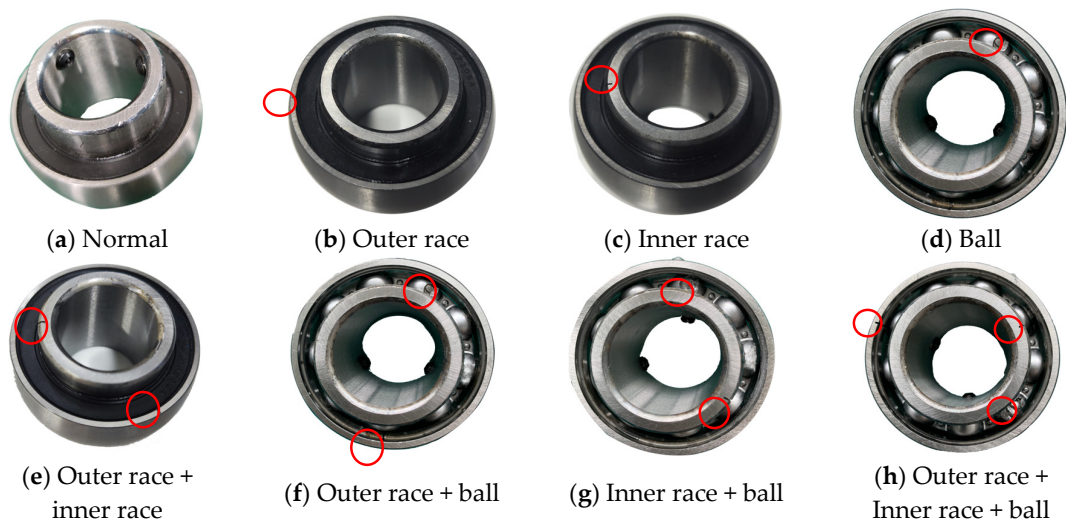
Figure 4. Coupling fault injection bearing.

Each bearing ran for 60 s at different speeds. The data for each second were taken as a sample, containing a total of 1020 samples. In this study, the *X*-axis vibration signal was selected for fault diagnosis and analysis. Firstly, the Fourier transform was performed on the data to obtain the spectra of the vibration data. The fault spectra of the base faults are shown in Figure 5.

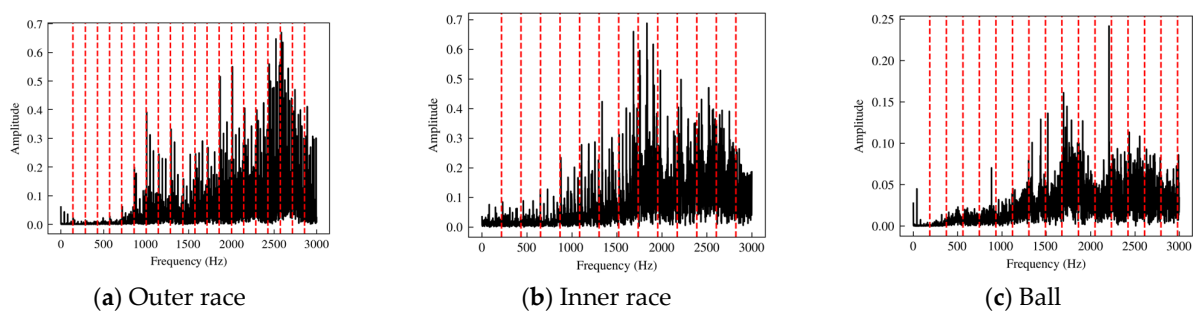

(**a**) Outer race　　(**b**) Inner race　　(**c**) Ball

Figure 5. Spectra of base fault bearings.

The red dashed line is the fundamental frequency. Its frequency—double that of the fault characteristic frequency value—was calculated theoretically based on the bearing size, in which the fundamental frequencies of the outer-race, inner-race, and ball faults were

142.93 Hz, 217.07 Hz, and 186.27 Hz, respectively. The figure shows that the fault frequency of the bearing-injected fault had a high consistency with its theoretical value.

The spectra of the coupled fault data are shown in Figure 6. The vibration spectrum of a coupled fault is not a linear superposition of the base fault spectrum but contains complex coupling relations, which requires a higher-order relationship model to obtain the coupling fault features.
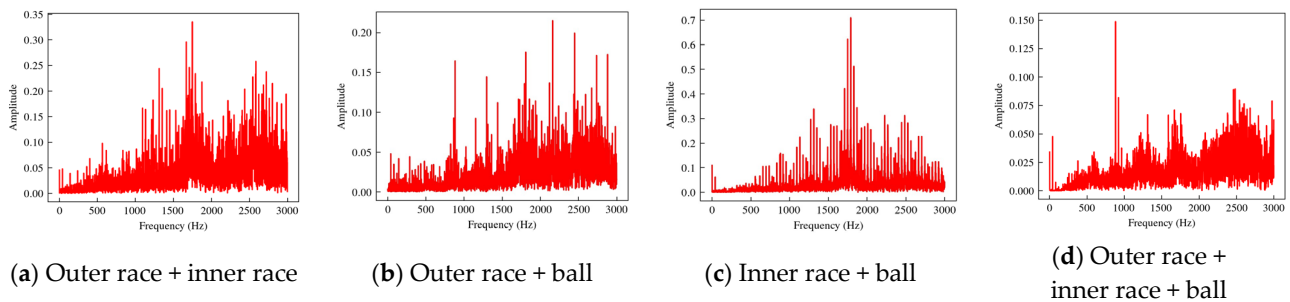

(**a**) Outer race + inner race   (**b**) Outer race + ball   (**c**) Inner race + ball   (**d**) Outer race + inner race + ball

Figure 6. Spectra of coupling fault bearings.

5. Coupling Fault Diagnosis

5.1. Coupling Fault Diagnosis Based on Feature Generation

In the HGGAN architecture, the four base faults are defined as four graph nodes, and the eight hyperedges represent eight types of coupling faults. As shown in Figure 7, the normal state constitutes an independent hyperedge; the inner-race, the outer-race, and the ball faults each constitute an independent hyperedge, in three primary colors; the pair–pair couplings constitute three hyperedges; and the three-fault coupling constitutes hyperedge E_7, comprising a total of eight hyperedges. The incidence matrix is shown in Table 1.
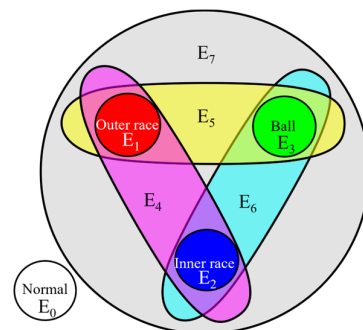
Figure 7. Coupled fault topology in HGGAN.

Table 1. Incidence matrix of HGGAN.

Fault Types	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7
Normal	1	0	0	0	0	0	0	0
Outer race	0	1	0	0	1	1	0	1
Inner race	0	0	1	0	1	0	1	1
Ball	0	0	0	1	0	1	1	1

The dataset was constructed based on the hypergraph topology, in which the training set contained 816 hypergraph samples, and the test set contained 204 hypergraph samples. The SGD optimizer was used for model training, the SGD momentum was set to 0.9, the step dynamic learning rate was adopted, the initial learning rate was 0.01, and the decay rate was 0.1. The training results are shown in Figure 8.
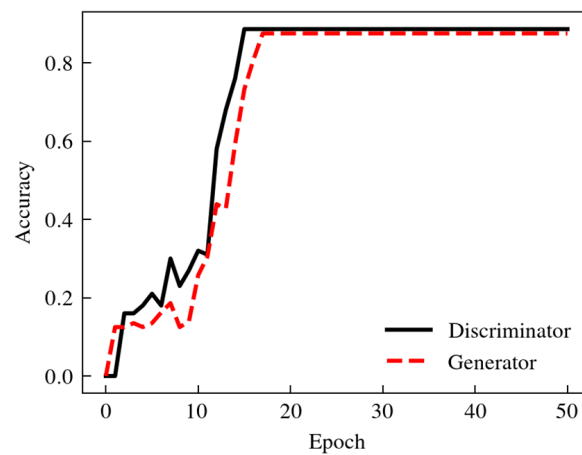
Figure 8. Accuracies of discriminator and generator during training process of HGGAN.

The accuracies of the discriminator and generator were 88.6% and 87.5%, respectively. The discriminator and generator played games with each other in the training process to jointly optimize the model's accuracy.

The generator aggregated the three types of base faults via the hypergraph network to obtain a total of $2^3 = 8$ types of faults. The experimental data and the data generated by the generator were reduced to two dimensions by the TSNE, and their distributions are shown in Figure 9. As shown in Figure 9a, the experimental data were separated after preprocessing in the feature space, but the coupling effect of the fault was not displayed in the feature space. The output data features of the generator are shown in Figure 9b, where the third-order coupling fault represented by the gray data points is approximately located in the fault center, and the second-order fault is located between the two base faults that compose it. Although there was a clear dividing line between the data generated by the generator and the original data, similar faults were approximately located in the same area in the two-dimensional plane, as shown in Figure 9c. The reason is that the generator loss function is the cross-entropy of the network output and the real label, which only has requirements for the size relationship of the output value and has low requirements for the consistency of data values so that the model has good generalization. Therefore, the data generated by the generator can realize more accurate fault diagnosis.

(**a**) Data from experiment　　　　　(**b**) Data from generator　　　　　(**c**) Coupling fault feature space
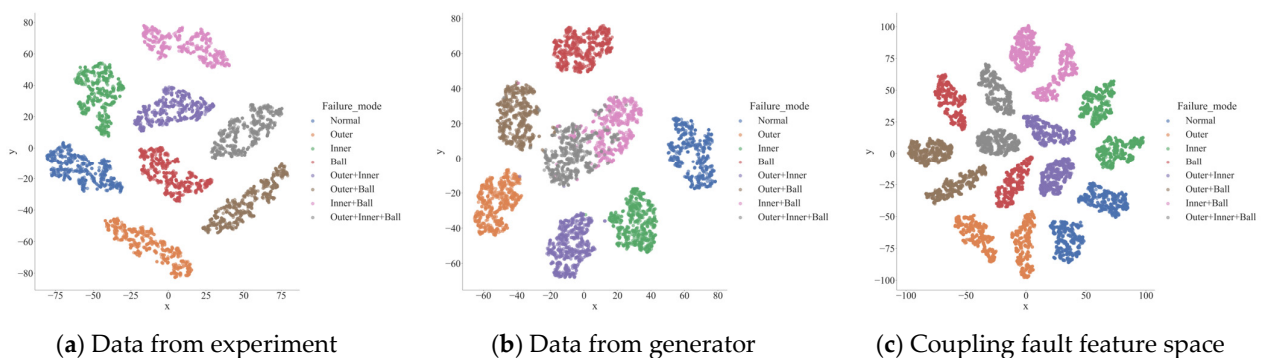
Figure 9. Visualization of data from experiment and data generated by generator.

The discriminator obtained by the HGGAN model was compared with the MLP without adding a generator and the MLP-GAN model in which the MLP was both the generator and discriminator. Moreover, it was compared with the model with a residual in the discriminator. The diagnostic accuracies are shown in Table 2. The HGGAN had advantages over the other models in terms of accuracy.

Table 2. Coupled fault diagnosis accuracies under feature generation framework.

Model	Accuracy of Generator	Accuracy of Discriminator
MLP	/	86.27%
MLP-GAN	75%	86.27%
Residual HGAN	87.5%	86.27%
HGGAN	87.5%	88.6%

Compared with the MLP, the HGGAN's discriminator had the same structure as the MLP. The HGGAN's loss function was optimized by adding the generator to HGCN to generate negative samples and optimize the MLP model parameters, so its fault diagnosis accuracy was improved. Compared with the MLP-GAN, the HGGAN used the HGCN as a generator to extract higher-order relationships more effectively and generate more accurate fault feature data. The shortcut was added to the discriminator to introduce the residual block. The residual block blurred the extracted features because of the shallow network depth, resulting in a lower discriminator accuracy. There was no improvement in the generator performance compared with the HGGAN proposed in this article.

5.2. Coupling Fault Diagnosis Based on Feature Extraction

This study considered four base faults in the IPHGAT architecture—normal, outer-race, inner-race, and ball faults—and four coupled faults—outer-race + inner-race, outer-race + ball, inner-race + ball, and outer-race + inner-race + ball faults—as the graph nodes and established eight hyperedges corresponding to all fault types. The hypergraph topology is shown in Figure 10.

Figure 10. Coupled fault topology in IPHGAT.

The correlation matrix of the hypergraph topology was obtained according to the fault coupling relationship, as shown in Table 3.

Table 3. Comparison of diagnostic accuracy for coupling faults.

Fault Type	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7
Normal	1	0	0	0	0	0	0	0
Outer race	1	1	0	0	0	0	0	0
Inner race	1	0	1	0	0	0	0	0
Ball	1	0	0	1	0	0	0	0
Outer race + inner race	1	1	1	0	1	0	0	0
Outer race + ball	1	1	0	1	0	1	0	0
Inner race + ball	1	0	1	1	0	0	1	0
Outer race + inner race + ball	1	1	1	1	1	1	1	1

The hypergraph data $X = [X^{(0)}, X^{(1)}] \in R^{9 \times 3000}$ were built for training. The SGD optimizer was used for model training, the SGD momentum was set to 0.9, a step dynamic learning rate was adopted, the initial learning rate was 0.01, and the attenuation rate was 0.1. Since the number of base faults involved in this study was four, the number of heads in the multi-head IPHGAT was set to four.

In the training process, the classification accuracy $Acc^{(0)}$ of the base fault $X^{(0)}$, the accuracy $Acc^{(1)}$ of the dynamic vertex fault $X^{(1)}$ to be predicted, and the correlation accuracy Acc of the inner product were calculated. As shown in Figure 11, $Acc^{(0)}$ and $Acc^{(1)}$ reached 99.88% and 88.60%, respectively, and the internal product correlation accuracy rate Acc reached 86.76%, so the IPHGAT can effectively classify coupling faults.
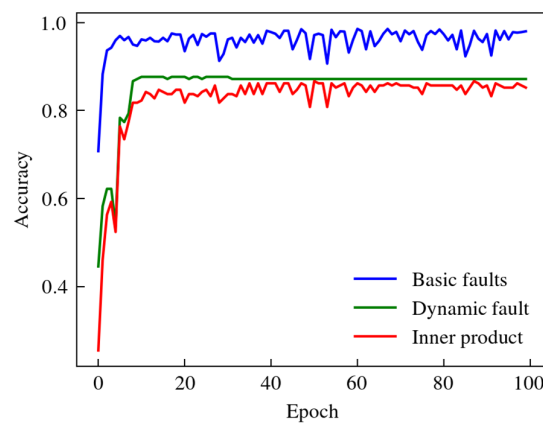
Figure 11. Diagnostic accuracy for base faults, faults in dynamic vertex, and inner product during training process of IPGAT.

Table 4 compares the coupled fault diagnostic accuracies of the three methods. The accuracy reached 78.43% by adding a residual network to the model, failing to improve the accuracy. The HGCN was used to extract coupled fault features, and the accuracy of its training set reached 88.6%, consistent with that of the IPHGAT with the MLP proposed in this paper. However, only global fault features were extracted due to its lack of local feature extraction capability. Furthermore, its generalization was poor, reaching only 83.82% accuracy on the test set.

Table 4. Coupled fault diagnostic accuracy under feature extraction framework.

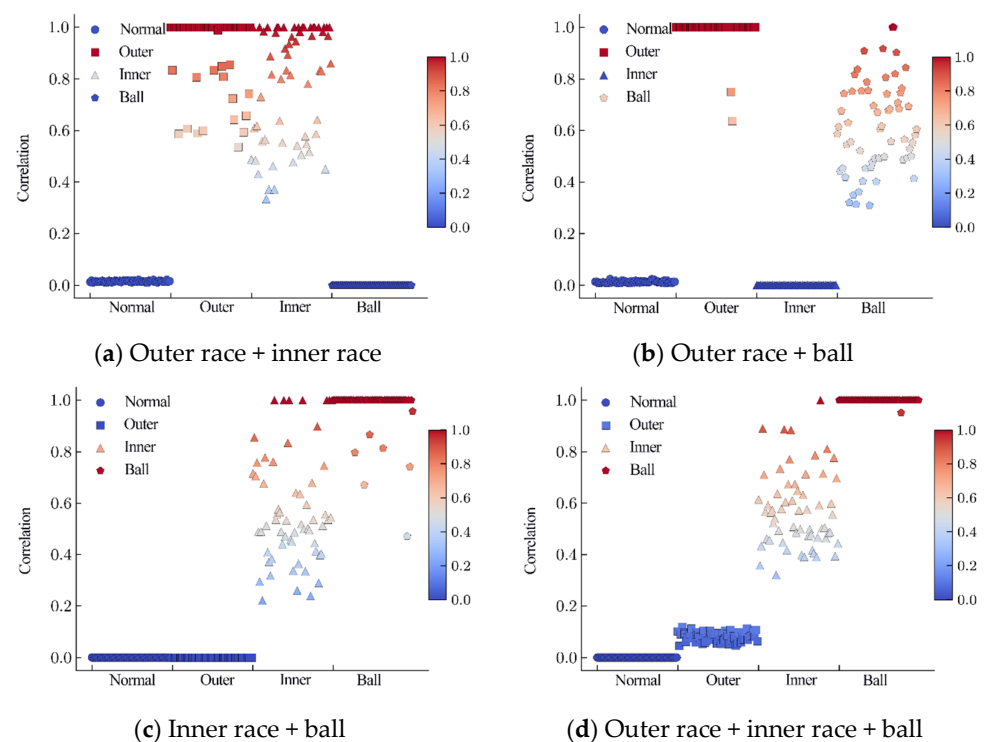| Model	Training Set Accuracy	Test Set Accuracy
IPHGAT (residual)	83.7%	78.43%
IPHGCN	88.6%	83.82%
IPHGAT (MLP)	88.6%	86.76%

The IPHGAT contained two layers of four-head hypergraph attention networks, which were used for the feature extraction of four types of basic faults. Each sample included eight types of faults, and the eight elements in the output $\hat{Y}^{(0)} \in \mathbb{R}^{8 \times 8}$ of the Softmax layer corresponded to eight types of fault modes. The first four elements $\hat{Y}_i^{(0)} = \{\hat{y}_{i,1}, \hat{y}_{i,2}, \hat{y}_{i,3}, \hat{y}_{i,4}\}$ were the correlation between the four base faults and the fault mode corresponding to the row vector x_i. The four base fault correlation elements corresponding to the eight types of faults in each sample were extracted, and the coupling mode of the faults was analyzed.

The model parameter optimization aimed to improve the fault classification accuracy during the model training. $\hat{y}_{i,i}$ is generally the maximum value in $\hat{Y}^{(0)}$ and is at a higher

order of magnitude than other elements, so this study carried out normalization processing on $\hat{y}_{i,1}, \hat{y}_{i,2}, \hat{y}_{i,3}, \hat{y}_{i,4}$ as follows:

$$\overline{y}_{i,j} = \frac{\hat{y}_{i,j} - \min(\hat{Y}_i^{(0)})}{\max(\hat{Y}_i^{(0)}) - \min(\hat{Y}_i^{(0)})} \tag{11}$$

Since the model achieved accurate feature extraction of the four types of independent base faults, a correlation analysis between the base faults and coupled faults was carried out, as shown in Figure 12. As shown in Figure 12a–c, the base faults for the second-order coupling faults were effectively identified, and the correlations between the coupling faults and the unrelated base faults were suppressed. In the base fault correlation analysis of the third-order coupling faults, all three types of base faults showed correlations with the coupled faults, and the correlation between the normal state and coupled faults approached zero, as shown in Figure 12d.

(**a**) Outer race + inner race

(**b**) Outer race + ball

(**c**) Inner race + ball

(**d**) Outer race + inner race + ball

Figure 12. Base fault correlations.

The inner product correlation between the dynamic vertex faults and eight types of fault data was $C = [c_1, \cdots, c_8] \in R^8$. The inner product correlation c_1, c_2, c_3, c_4 of the coupled fault and base fault characteristics was removed for the analysis in Figure 13. The vertical axis is the coupling fault to be predicted, and the horizontal axis is the base fault that makes up the coupling fault. The accuracy of the base fault components of the coupled faults was more than 91.67%. The coupling analysis was consistent with the fault injection model.

The IPHGAT model based on the reductionist feature extraction framework extracts the fault component representing the base fault from the coupling fault and analyzes the fault mode based on the correlation of the fault component, which enhances the model's interpretability and is significant for the coupling fault diagnosis of rotating machinery in industrial production environments.
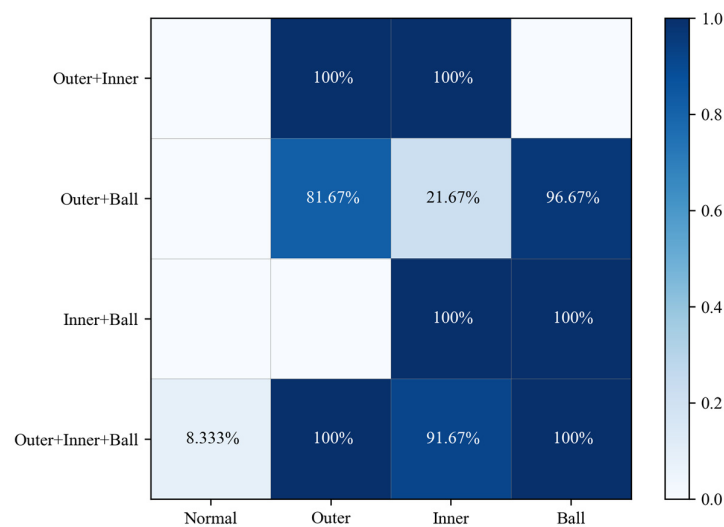
Figure 13. Coupling fault classification accuracies.

6. Discussion and Conclusions

In this study, based on the research on the coupling fault diagnosis of rotating machinery, a coupling fault diagnosis method, HGGAN, based on a fault feature generation framework, and another method, IPHGAT, based on a fault feature extraction framework, were proposed from the perspectives of holism and reductionism theory by utilizing the advantages of a hypergraph neural network for higher-order relation analysis. The HGGAN used an MLP as a discriminator and utilized an HGCN as a generator to generate coupling fault data. The generated and real data had a unified data space, and the diagnostic accuracy of the HGGAN for coupling faults reached 88.6%. IPHGAT used a multi-head hypergraph neural network to extract the fault features of coupling faults, and an MLP was used to extract the dimensionality reduction features of predictive data. The internal product correlation between the fault features and coupling faults in the dynamic vertex was calculated, and the fault diagnosis of coupling faults was realized. The diagnostic accuracy rate reached 86.76%. The two kinds of methods have their own advantages in application. The HGGAN takes the generator error as the regularization term of the discriminator, which improves its diagnostic accuracy, which is slightly higher than that of IPHGAT. IPHGAT can evaluate the correlation between coupling faults and base faults based on the local characteristics of the base faults; thus, IPHGAT has higher interpretability than the HGGAN model.

The coupling fault diagnosis framework proposed in this paper can accurately diagnose coupling faults under high-order relations and analyze the coupling characteristics of the fault characteristics, which is significant for monitoring and analyzing the health state of rotating machinery in production environments. This study qualitatively analyzed the types of base faults that constituted the coupling faults in the interpretability analysis of the model. However, it did not verify the quantitative relationship between the correlation of the base faults and coupling faults extracted by the model and the degree of faults. In the next stage of coupled fault diagnosis research, multi-degree coupled fault injection tests should be carried out to further fit the quantitative relationship of the fault correlation with several fault samples.

Author Contributions: Conceptualization, S.W. and B.J.; methodology, S.W.; validation, S.P., Y.H. and X.M.; investigation, J.P. and X.J.; writing—original draft preparation, S.W.; writing—review and editing, S.W.; supervision, B.J. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data are unavailable due to privacy.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xue, S.; Wang, C.S.; Howard, I.; Lian, P.Y.; Chen, G.G.; Wang, Y.; Yan, Y.F.; Xu, Q.; Shi, Y.; Jia, Y.; et al. The diagnostic analysis of the fault coupling effects in planet bearing. *Eng. Fail. Anal.* **2020**, *108*, 23. [CrossRef]
2. Lu, K.; Cheng, H.; Zhang, W.T.; Zhang, H.P.; Zhang, K.F.; Fu, C. Nonlinear dynamic behavior of a dual-rotor bearing system with coupling misalignment and rubbing faults. *Meas. Sci. Technol.* **2023**, *34*. [CrossRef]
3. Ying, W.M.; Zheng, J.D.; Huang, W.; Tong, J.Y.; Pan, H.Y.; Li, Y.B. Order-frequency Holo-Hilbert spectral analysis for machinery fault diagnosis under time-varying operating conditions. *ISA Trans.* **2024**, *146*, 472–483. [CrossRef]
4. Li, Y.; Cheng, G.; Ma, S.C.; Li, X. Bearing fault diagnosis method based on complete center frequency distribution feature. *Struct. Health Monit.-Int. J.* **2023**, *22*, 4100–4116. [CrossRef]
5. Ma, J.P.; Chen, G.D.; Li, C.W.; Zhan, L.W.; Zhang, G.Z. Rolling Bearing Feature Extraction Method Based on Improved Intrinsic Time-Scale Decomposition and Mathematical Morphological Analysis. *Appl. Sci.* **2021**, *11*, 2719. [CrossRef]
6. Jiang, Z.H.; Zhang, K.; Xiang, L.; Yu, G.; Xu, Y.G. A time-frequency spectral amplitude modulation method and its applications in rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **2023**, *185*, 109832. [CrossRef]
7. Zhu, D.C.; Yin, B.L.; Teng, C.X. An improved spectral amplitude modulation method for rolling element bearing fault diagnosis. *J. Braz. Soc. Mech. Sci. Eng.* **2023**, *45*, 16. [CrossRef]
8. Yang, H.; Zhang, K.; Jiang, Z.H.; Zhang, X.F.; Xu, Y.G. An adaptive time-frequency demodulation method and its applications in rolling bearing fault diagnosis. *Meas. Sci. Technol.* **2023**, *34*, 15. [CrossRef]
9. Yue, K.; Li, J.P.; Chen, J.B.; Huang, R.Y.; Li, W.H. Multiscale Wavelet Prototypical Network for Cross-Component Few-Shot Intelligent Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 11. [CrossRef]
10. Tang, J.H.; Wu, J.M.; Hu, B.B.; Qing, J.J. Towards a Fault Diagnosis Method for Rolling Bearings with Time-Frequency Region-Based Convolutional Neural Network. *Machines* **2022**, *10*, 1145. [CrossRef]
11. Jia, L.S.; Chow, T.W.S.; Yuan, Y.X. GTFE-Net: A Gramian Time Frequency Enhancement CNN for bearing fault diagnosis. *Eng. Appl. Artif. Intell.* **2023**, *119*, 105794. [CrossRef]
12. Yoo, Y.; Jeong, S. Vibration analysis process based on spectrogram using gradient class activation map with selection process of CNN model and feature layer. *Displays* **2022**, *73*, 102233. [CrossRef]
13. Ahmed, H.O.A.; Nandi, A.K. Connected Components-based Colour Image Representations of Vibrations for a Two-stage Fault Diagnosis of Roller Bearings Using Convolutional Neural Networks. *Chin. J. Mech. Eng.* **2021**, *34*, 21. [CrossRef]
14. Lin, K.Y.; Jamrus, T. Industrial data-driven modeling for imbalanced fault diagnosis. *Ind. Manag. Data Syst.* **2024**, *ahead-of-print*. [CrossRef]
15. Zhang, X.; Han, P.; Xu, L.; Zhang, F.; Wang, Y.; Gao, L. Research on Bearing Fault Diagnosis of Wind Turbine Gearbox Based on 1DCNN-PSO-SVM. *IEEE Access* **2020**, *8*, 192248–192258. [CrossRef]
16. Karnavas, Y.L.; Plakias, S.; Chasiotis, I.D. Extracting spatially global and local attentive features for rolling bearing fault diagnosis in electrical machines using attention stream networks. *IET Electr. Power Appl.* **2021**, *15*, 903–915. [CrossRef]
17. Sivapriya, A.; Kalaiarasi, N.; Verma, R.; Chokkalingam, B.; Munda, J.L. Fault Diagnosis of Cascaded Multilevel Inverter Using Multiscale Kernel Convolutional Neural Network. *IEEE Access* **2023**, *11*, 79513–79530. [CrossRef]
18. Yang, Y.; Liu, H.; Han, L.J.; Gao, P. A Feature Extraction Method Using VMD and Improved Envelope Spectrum Entropy for Rolling Bearing Fault Diagnosis. *IEEE Sens. J.* **2023**, *23*, 3848–3858. [CrossRef]
19. Huo, J.Y.; Yang, J.W.; Yao, D.C.; Sun, R.T.; Hu, Z.S.; Chen, Z.H.; Gao, C. Bearing fault diagnosis under variable speed conditions on adaptive time frequency extraction mode decomposition. *Meas. Sci. Technol.* **2024**, *35*, 15. [CrossRef]
20. Zhao, X.Q.; Guo, H.K. Rolling bearing fault diagnosis model based on DSCB-NFAM. *Meas. Sci. Technol.* **2024**, *35*, 15. [CrossRef]
21. Yang, J.; Chen, J.W.; Zhan, X.A.; Liu, C.S.; Yang, C. A feature reconstruction and SAE model based diagnosis method for multiple mixed faults. *Meas. Sci. Technol.* **2024**, *35*, 14. [CrossRef]
22. Jang, K.; Hong, S.; Kim, M.; Na, J.; Moon, I. Adversarial Autoencoder Based Feature Learning for Fault Detection in Industrial Processes. *IEEE Trans. Ind. Inform.* **2022**, *18*, 827–834. [CrossRef]
23. Yu, X.L.; Yang, Y.; Du, M.G.; He, Q.B.; Peng, Z.K. Dynamic Model-Embedded Intelligent Machine Fault Diagnosis Without Fault Data. *IEEE Trans. Ind. Inform.* **2023**, *19*, 11466–11476. [CrossRef]
24. Luczak, D. Data-Driven Machine Fault Diagnosis of Multisensor Vibration Data Using Synchrosqueezed Transform and Time-Frequency Image Recognition with Convolutional Neural Network. *Electronics* **2024**, *13*, 2411. [CrossRef]
25. Li, T.; Zhou, Z.; Li, S.; Sun, C.; Yan, R.; Chen, X. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. *Mech. Syst. Signal Process.* **2022**, *168*, 108653. [CrossRef]
26. Ning, S.H.; Ren, Y.L.; Wu, Y.K. Intelligent fault diagnosis of rolling bearings based on the visibility algorithm and graph neural networks. *J. Braz. Soc. Mech. Sci. Eng.* **2023**, *45*, 14. [CrossRef]
27. Luo, L.H.; Fang, Y.X.; Lu, M.L.; Cao, X.; Zhang, X.F.; Zhang, W.J. GSim: A Graph Neural Network Based Relevance Measure for Heterogeneous Graphs. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 12693–12707. [CrossRef]

28. Li, T.F.; Zhao, Z.B.; Sun, C.; Yan, R.Q.; Chen, X.F. Multireceptive Field Graph Convolutional Networks for Machine Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2021**, *68*, 12739–12749. [CrossRef]

29. Wang, H.; Wang, J.W.; Zhao, Y.K.; Liu, Q.; Liu, M.; Shen, W.M. Few-Shot Learning for Fault Diagnosis With a Dual Graph Neural Network. *IEEE Trans. Ind. Inform.* **2023**, *19*, 1559–1568. [CrossRef]

30. Wang, S.; Jing, B.; Pan, J.; Meng, X.; Huang, Y.; Jiao, X. Coupling Fault Diagnosis Based on Dynamic Vertex Interpretable Graph Neural Network. *Sensors* **2024**, *24*, 4356. [CrossRef]

31. Yan, X.S.; Liu, Y.; Zhang, C.A. Multiresolution Hypergraph Neural Network for Intelligent Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 10. [CrossRef]

32. Feng, Y.; You, H.; Zhang, Z.; Ji, R.; Gao, Y. Hypergraph neural networks. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; p. 437.

33. Bai, S.; Zhang, F.H.; Torr, P.H.S. Hypergraph convolution and hypergraph attention. *Pattern Recognit.* **2021**, *110*, 107637. [CrossRef]

34. Shi, M.K.; Ding, C.C.; Wang, R.; Song, Q.Y.; Shen, C.Q.; Huang, W.G.; Zhu, Z.K. Deep hypergraph autoencoder embedding: An efficient intelligent approach for rotating machinery fault diagnosis. *Knowl.-Based Syst.* **2023**, *260*, 110172. [CrossRef]

35. Su, S.Z.; Zhang, Z.P.; Zhu, Y.M.; Hou, Y.K. Fault diagnosis of rotating machinery via multi-structure fusion discriminative projection. *Meas. Sci. Technol.* **2023**, *34*, 20. [CrossRef]

36. Yuan, Z.H.; Ma, Z.Q.; Li, X.; Li, J.J. A Multichannel MN-GCN for Wheelset-Bearing System Fault Diagnosis. *Ieee Sens. J.* **2023**, *23*, 2481–2494. [CrossRef]

37. Wang, Y.P.; Zhang, S.; Cao, R.F.; Xu, D.; Fan, Y.Q. A Rolling Bearing Fault Diagnosis Method Based on the WOA-VMD and the GAT. *Entropy* **2023**, *25*, e25060889. [CrossRef]

38. Luo, W.; Yang, W.; He, J.L.; Huang, H.; Chi, H.D.; Wu, J.L.; Shen, Y. Fault Diagnosis Method Based on Two-Stage GAN for Data Imbalance. *IEEE Sens. J.* **2022**, *22*, 21961–21973. [CrossRef]

39. Liu, S.H.; Zhang, J.H.; Huang, W.D.; Lyu, F.; Wang, D.D.; Xu, B. Temporal-Spatial Attention Network: A Novel Axial Piston Pump Coupled Fault Diagnosis Method. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 15. [CrossRef]

40. Hao, C.; Du, J.; Liang, H. Imbalanced Fault Diagnosis of Rolling Bearing Using Data Synthesis Based on Multi-Resolution Fusion Generative Adversarial Networks. *Machines* **2022**, *10*, 295. [CrossRef]