

Article citation info:

Gao Y, Li X, Fault Diagnosis of Centrifugal fan Bearings Based on I-CNN and JMMD in the Context of Sample Imbalance, *Eksploracja i Niezawodność – Maintenance and Reliability* 2024; 26(4) <http://doi.org/10.17531/ein/191459>

Fault Diagnosis of Centrifugal fan Bearings Based on I-CNN and JMMD in the Context of Sample Imbalance

Indexed by:



Yang Gao^a, Xueyi Li^{b,c,*}

^a College of Mathematical Sciences, Daqing Normal University, China

^b Department of Mechanical Engineering, Tsinghua University, China

^c College of Mechanical and Electrical Engineering, Northeast Forestry University, China

Highlights

- This paper conducts Fast Fourier transform on the signals to enhance sample features.
- Parallel CNNs are employed to capture bearing fault information at various scales.
- Maximize domain adaptation through joint mean discrepancy.
- Introduces concentrated loss (C-Loss), prioritizing minority samples.
- Integrates lead weight factors to enhance focus on easily confused samples.

Abstract

Bearing fault diagnosis is an effective technical means to improve the reliability of centrifugal fan bearings. In this paper, a transfer learning-based fault diagnosis method for Centrifugal fan bearings is proposed, utilizing the improved CNN (I-CNN) and Joint Maximum Mean Discrepancy (JMMD) algorithms. The raw vibration signals of bearings are enhanced through fast Fourier transform for feature representation. The enhanced signals are then processed by parallel multi-scale CNNs with an embedded Squeeze-and-Excitation (SE) attention mechanism to extract and focus on key features. Furthermore, the JMMD is introduced as a metric for quantifying the disparity between the source and target domains, thereby mitigating domain shift. In the loss function, weight factors and scaling factors are introduced to increase attention on minority samples and easily confused samples within the imbalanced dataset. The proposed method is validated on the Centrifugal fan bearing dataset from Jiangnan University and the CWRU dataset.

Keywords

centrifugal fan bearings, sample imbalance, transfer learning, fault diagnosis

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

Centrifugal fans find extensive application across diverse industrial sectors including manufacturing, chemical engineering, and energy production[1-4]. They play an indispensable role in ventilation, cooling, dust removal, and exhaust gas emission, among other applications[5]. As crucial components of Centrifugal fan transmission systems, Centrifugal fan bearings operate at high speeds for extended periods in complex and variable environments, often experiencing faults due to fluctuating loads and mechanical wear. Studies indicate that approximately 30% of failures in

rotating machinery are attributed to bearings[6]. Faults in Centrifugal fan bearings can lead to sudden shutdowns or severe vibrations, posing safety risks to personnel and equipment. Timely diagnosis and maintenance can reduce the probability of accidents and enhance workplace safety. Therefore, achieving rapid and accurate fault diagnosis of rolling bearings is of paramount importance.

Traditional fault diagnosis methods typically involve processing and analyzing signals such as vibration, sound, and temperature from bearings[7]. These signals can be acquired

(*) Corresponding author
E-mail addresses:

Y. Gao (ORCID: 0000-0003-1952-3955) gy19790607@163.com, X. Li (ORCID: 0000-0003-0335-0594) lixueyinefu@163.com,

through sensors, monitoring devices, etc., and then analyzed and diagnosed using signal processing techniques[8]. Common signal processing methods include wavelet transform (WT), Fourier transform (FT), power spectral analysis (PSA), autocorrelation function (AF), variational mode decomposition (VMD)[9-12]. Although traditional fault diagnosis methods have good results, their feature extraction relies on manual experience, and they often face challenges such as slow processing speed when dealing with large amounts of data, leading to many limitations in the field of fault diagnosis[13].

Recently, with a significant boost in computational power, deep learning has emerged and rapidly found application in bearing fault diagnosis. Researchers have leveraged the powerful feature extraction capabilities of deep learning to diagnose faults in critical components of rotating machinery such as bearings, ensuring the smooth operation of machines[14]. Li et al.[15] proposed a method based on a combination optimization algorithm, using the ResNet18 network for classifying and diagnosing bearing faults. Tang et al.[16] proposed a new deep confidence network embedded with a Kalman filter, which utilizes multi-sensor information to achieve bearing fault diagnosis under noisy conditions.

Machine learning-based fault diagnosis of Centrifugal fan bearings has drawn significant attention from researchers. Xie et al.[17] introduced a fault diagnosis technique for fan bearings, employing continuous wavelet transform and autocorrelation analysis. This method offers a novel approach to diagnose and predict faults in cooling fans used in electronic equipment. Cui et al.[18] introduced a method that converts one-dimensional vibration signals into SDP images and utilizes convolutional neural networks (CNN) for fault identification in mine fan bearings. He et al.[19] introduced a vibration-based health monitoring approach for cooling fans, employing wavelet filters to enable early detection and severity assessment of fan bearing faults.

Traditional fault diagnosis typically involves training and diagnosing networks under the same operating conditions, which can effectively handle bearing fault diagnosis under specific conditions[20]. However, the operating conditions of rotating machinery are often variable. Diagnosis under various conditions using traditional approaches requires collecting large amounts of labeled data for each condition. To address this issue,

researchers have considered cross-domain (CD) fault diagnosis of bearings, where source domain (SD) data is used to train models to diagnose fault data in the target domain. Zhao et al.[21] proposed a rolling bearing fault diagnosis method based on twin-domain adversarial transfer learning, improving the convolutional and pooling layers of the transfer learning feature extraction using twin neural networks. This approach reduces differences in fault sample distributions under different operating conditions, enhances model generalization, and achieves CD fault diagnosis. Cao et al.[22] introduced an unsupervised shared-domain CNN for effective fault transfer diagnosis from stable to time-varying speeds, achieving cross-domain diagnosis of bearings. Xiao et al.[23] simulated SD bearing fault signals using simulation techniques to train neural networks, and then applied transfer learning techniques to target domain (TD) data, realizing a data-physics coupled fault diagnosis approach.

Furthermore, traditional data-driven bearing fault diagnosis methods often use simulated data with an equal number of samples per class[18]. However, in practical working conditions, once a problem occurs with Centrifugal fan bearings, the turbine needs to be shut down for inspection and repair, making it difficult to collect fault data. Moreover, due to the long accumulation period, time consumption, and incomplete fault data obtained during the collection of Centrifugal fan bearing fault data, healthy data is inevitably much more abundant than fault data. Especially as we enter the big data era, the density of data collection has grown exponentially, leading to even more healthy data and exacerbating data imbalance. Therefore, bearing fault diagnosis inevitably faces the challenge of dealing with data imbalance. Mao et al.[24] proposed an unbalanced fault diagnosis method based on Generative Adversarial Networks (GANs) and conducted detailed comparative studies. Hang et al.[25] proposed a two-step clustering algorithm to enhance the imbalanced data classification of the original synthetic minority oversampling technique algorithm. Lu et al.[26] proposed an improved active learning intelligent fault diagnosis method for unbalanced sample rolling bearings, which obtains the distribution representation of samples by constructing a Gaussian mixture model.

This paper conducts research based on the background of transfer learning and sample imbalance. By leveraging CNN

networks and the JMMD algorithm, an unsupervised fault diagnosis method for Centrifugal fan bearings under imbalanced data is proposed, termed I-CNN and JMMD. The primary contributions of this paper include:

1. Prior to utilizing neural networks to process bearing fault signals, this paper conducts Fast Fourier transform (FFT) on the signals to enhance sample features. Subsequently, parallel CNNs with different kernel sizes are employed to capture bearing fault information at various scales.

2. This paper adopts a transfer learning approach to address the time-consuming and labor-intensive signal acquisition problem in real-world operating conditions. By considering the joint maximum mean discrepancy (JMMD) between the SD and the TD as a crucial term in the loss function, domain shift is minimized, achieving domain adaptation.

3. In response to the common challenge of imbalanced data in reality, this paper innovatively introduces concentrate loss (C-Loss), lead weight factors and scaling factors into the loss function, enhancing the focus on minority samples and easily confused samples.

The remaining sections of the paper are structured as follows: Section 2 presents an overview of related work, while Section 3 elaborates on the proposed method for diagnosing bearing faults; Section 4 outlines the experimental details on the CWRU dataset; Section 5 covers the experimental details on the JNU dataset; Lastly, Section 6 provides the conclusion of the paper.

2. Related Work

2.1. Fast Fourier transform

The FFT is an efficient algorithm for computing the Fourier transform[27, 28]. It reduces the computation time of calculating the Fourier transform of a discrete sequence from $O(n^2)$ to $O(n \log n)$, where n is the length of the sequence. FFT finds extensive applications in signal processing, image processing, and various other fields.

Suppose have a complex sequence x_0, x_1, \dots, x_{N-1} , of length N . Its Discrete Fourier Transform (DFT) is defined as:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N}, \quad k = 0, 1, \dots, N-1 \quad (1)$$

Where X_k is the transformed sequence, x_n is the element of the original sequence, and k is the frequency index.

The FFT algorithm is based on the divide-and-conquer strategy, decomposing a DFT of length N into two DFTs of

length $\frac{N}{2}$. Specifically, for even N , we can decompose X_k into two parts: E_k containing elements with even indices and O_k containing elements with odd indices:

$$E_k = \sum_{n=0}^{N/2-1} x_{2n} e^{-i2\pi kn/N} \quad (2)$$

$$O_k = \sum_{n=0}^{N/2-1} x_{2n+1} e^{-i2\pi kn/N} \quad (3)$$

According to Euler's formula, $e^{ix} = \cos(x) + i \sin(x)$, E_k and O_k can be rewritten as:

$$E_k = \sum_{n=0}^{N/2-1} x_{2n} \left(\cos\left(\frac{2\pi kn}{N}\right) - i \sin\left(\frac{2\pi kn}{N}\right) \right) \quad (4)$$

$$O_k = \sum_{n=0}^{N/2-1} x_{2n+1} \left(\cos\left(\frac{2\pi kn}{N}\right) - i \sin\left(\frac{2\pi kn}{N}\right) \right) \quad (5)$$

Then, by utilizing the relationship $X_k = E_k + e^{-i2\pi k/N} O_k$, we can recursively calculate E_k and O_k to obtain X_k .

The FFT algorithm recursively halves the length of the sequence and exploits the symmetry of frequencies, significantly improving computational efficiency. Its time complexity is $O(N \log N)$.

2.2. Unsupervised Transfer Learning

Transfer learning involves leveraging knowledge gained from a SD to tackle issues in a target domain. Utilizing unsupervised transfer learning for fault diagnosis in bearings under various operating conditions can decrease the reliance on labeled fault data specifically for rolling bearings[29]. The dataset containing fault labels under specific operating conditions is denoted as the SD, while the dataset lacking fault labels under different operating conditions is referred to as the target domain. The distribution of bearing fault data differs across various operating conditions. To implement unsupervised transfer learning for rolling bearings across diverse operating conditions, a common strategy involves integrating neural network models with domain adaptation techniques. Throughout network training, domain adaptation is conducted on both the SD and TD to mitigate distribution disparities, consequently enhancing the predictive capabilities of the network model for the TD. As shown in Figure 1, let the SD data be D_s and the TD data be D_t . Through training on the SD data, the relationship between samples and true labels, $X_s \rightarrow Y_s$, is established. Based on the fundamental assumption of transfer learning, there exists a certain correlation between the SD data and the TD data. After feature mapping of the SD data and the TD data, there exists domain bias between the SD and TD, but there is still a certain

degree of feature similarity, indicating a shared feature space. Transfer learning algorithms can mitigate the domain bias

between the SD and TD. This enables accurate classification of TD data using a model trained on the SD data.

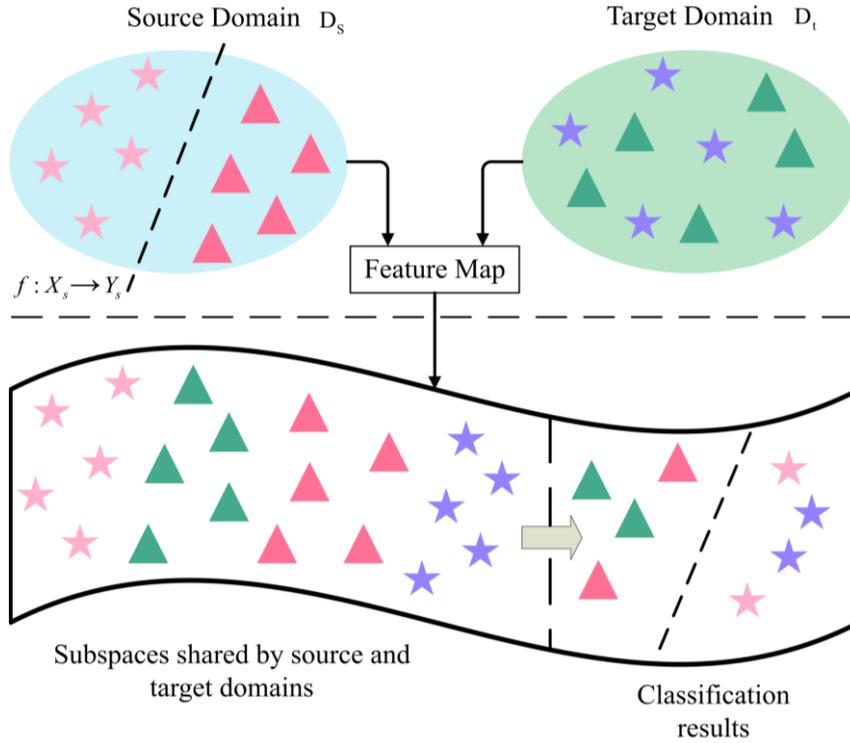


Fig. 1. The workflow diagram of transfer learning.

2.3. Maximum Mean Discrepancy

The Maximum Mean Discrepancy (MMD) is a commonly used metric in transfer learning to measure the distribution discrepancy between the SD and TD[30]. The MMD function calculates the mean discrepancy after mapping the source and target domains to the reproducing kernel Hilbert space. A smaller MMD value indicates a greater similarity between the distributions of the SD and TD. The expression for calculating MMD is:

$$\begin{aligned}
 MMD(X^s, X^t)^2 &= \frac{1}{m_s} \sum_{i=1}^{m_s} f(x_i^s)^2_{\mathcal{N}} - \frac{1}{m_t} \sum_{i=1}^{m_t} f(x_i^t)^2_{\mathcal{N}} \\
 &= \frac{1}{m_s^2} \sum_{i=1}^{m_s} \sum_{j=1}^{m_s} k(x_i^s, x_j^s) - \frac{2}{m_s m_t} \sum_{i=1}^{m_s} \sum_{j=1}^{m_t} k(x_i^s, x_j^t) \\
 &\quad + \frac{1}{m_t^2} \sum_{i=1}^{m_t} \sum_{j=1}^{m_t} k(x_i^t, x_j^t) \quad (6)
 \end{aligned}$$

where, x_i^s is the i -th sample vector from the SD, x_j^t is the j -th sample vector from the TD; m_s is the number of samples in the SD; m_t is the number of samples in the TD; \mathcal{N} is the reproducing kernel Hilbert space; $f(\bullet)$ is the nonlinear mapping function that maps the SD and TD data to the Hilbert space. In this paper, a Gaussian kernel function $K(\bullet)$ is used as the mapping function, expressed as:

$$f(\bullet) = k(a, a') = e^{-\frac{a-a'}{2\sigma^2}} \quad (7)$$

where, a can represent the i -th sample vector from SD x_i^s or the i -th sample vector from the TD x_i^t ; a' can be represented as the transpose of a ; σ is the bandwidth, which influences the local effect range of $K(\bullet)$. In transfer learning, MMD algorithm can be utilized to reduce the discrepancy between the SD and TD, thereby enhancing the accuracy of fault diagnosis.

3. Proposed Method

The proposed method in this paper for Centrifugal fan bearing fault diagnosis under imbalance-sample based on I-CNN and JMMD transfer learning is outlined as shown in Figure 2.

Step 1: Collect raw vibration signals of Centrifugal fans under multiple operating conditions, divide the collected data samples into SD training set, SD validation set, and TD validation set, and perform fast Fourier transform on the samples.

Step 2: Input the SD samples into a multiscale parallel neural network embedded with SE attention mechanism for network training and validation. Utilize multiscale convolutional windows to capture sample features at different granularities and focus on key features using Squeeze-and-Excitation

attention mechanism. Additionally, introduce concentrate loss(C-Loss) with weighting factors and scaling factors to address the issue of sample imbalance and the presence of easily confused samples.

Step 3: Input the target domain samples into the network trained on the SD training set and validated on the SD validation set. Utilize the trained network to diagnose TD data, and

introduce maximum mean discrepancy calculation to narrow the domain bias between the SD and TD, further optimizing the fault diagnosis performance of transfer learning.

Step 4: Apply publicly available datasets and Centrifugal fan datasets to validate the proposed method, analyze experimental results, and demonstrate the effectiveness of the method. Details are discussed in Sections 4 and 5.

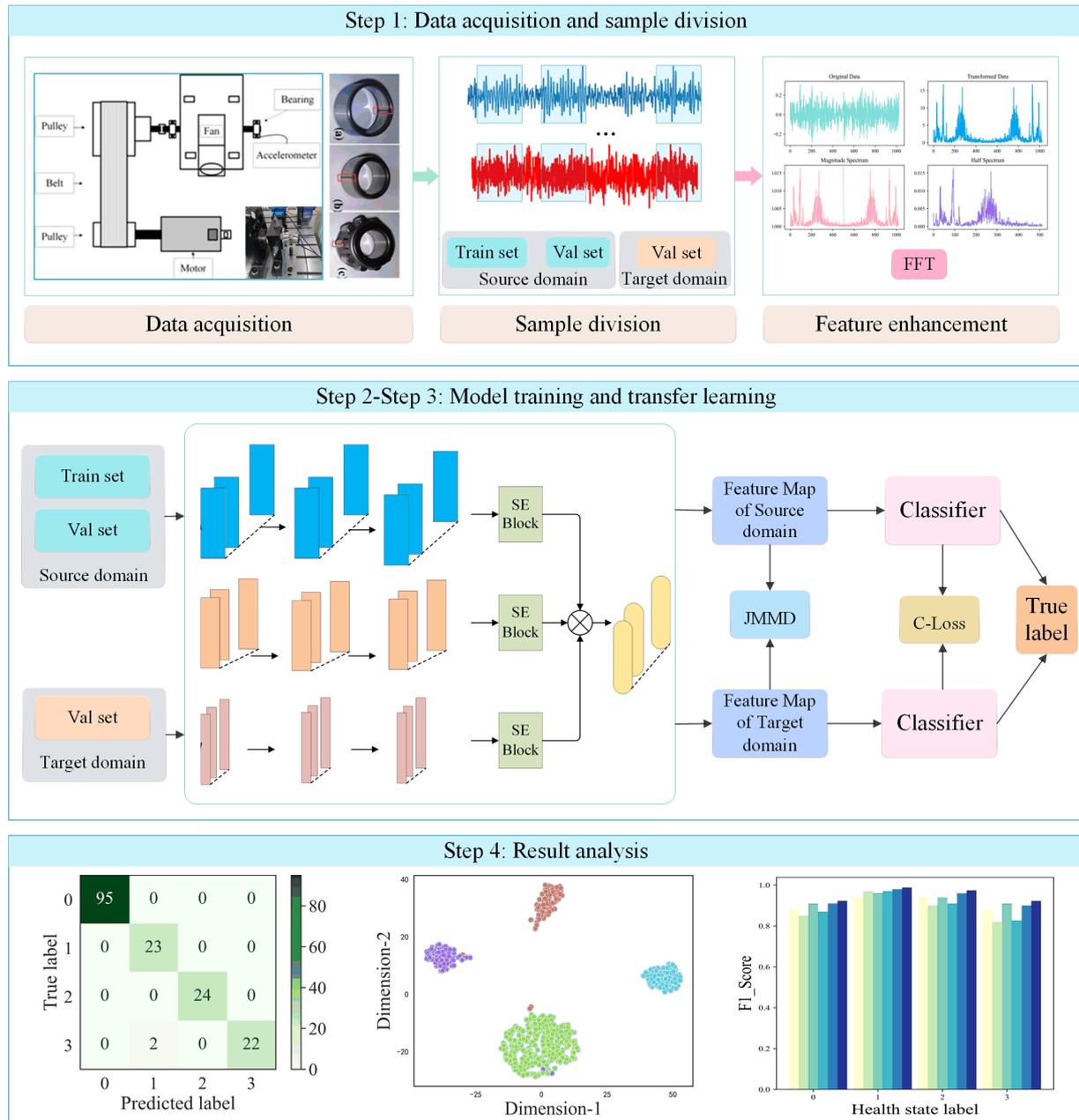


Fig. 2. The flowchart of this study.

3.1. Fast Fourier transform of Fault Signals

Performing Fourier Transform on bearing fault signals transforms the time domain signal $x(t)$ into the frequency domain signal $X(f)$, where t represents time and f represents frequency.

The mathematical expression for Fourier Transform is as follows:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft} dt \quad (8)$$

However, for practical digital signal processing, we use Discrete Fourier Transform (DFT), The mathematical

expression as follows:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-i2\pi kn/N} \quad (9)$$

Where $x(n)$ is the discrete sample of the time-domain signal, $X(k)$ is the discrete sample of the frequency-domain signal, N is the number of samples in the time domain signal, and k is the frequency index.

By employing the efficient algorithm for computing DFT known as FFT, the computational complexity is reduced from $O(n^2)$ to $O(N \log N)$, expediting the process of spectrum analysis. As shown in Figure 3, this article performs FFT transformation on the original signal, normalizes it, and then selects only the first half based on Nyquist's theorem.

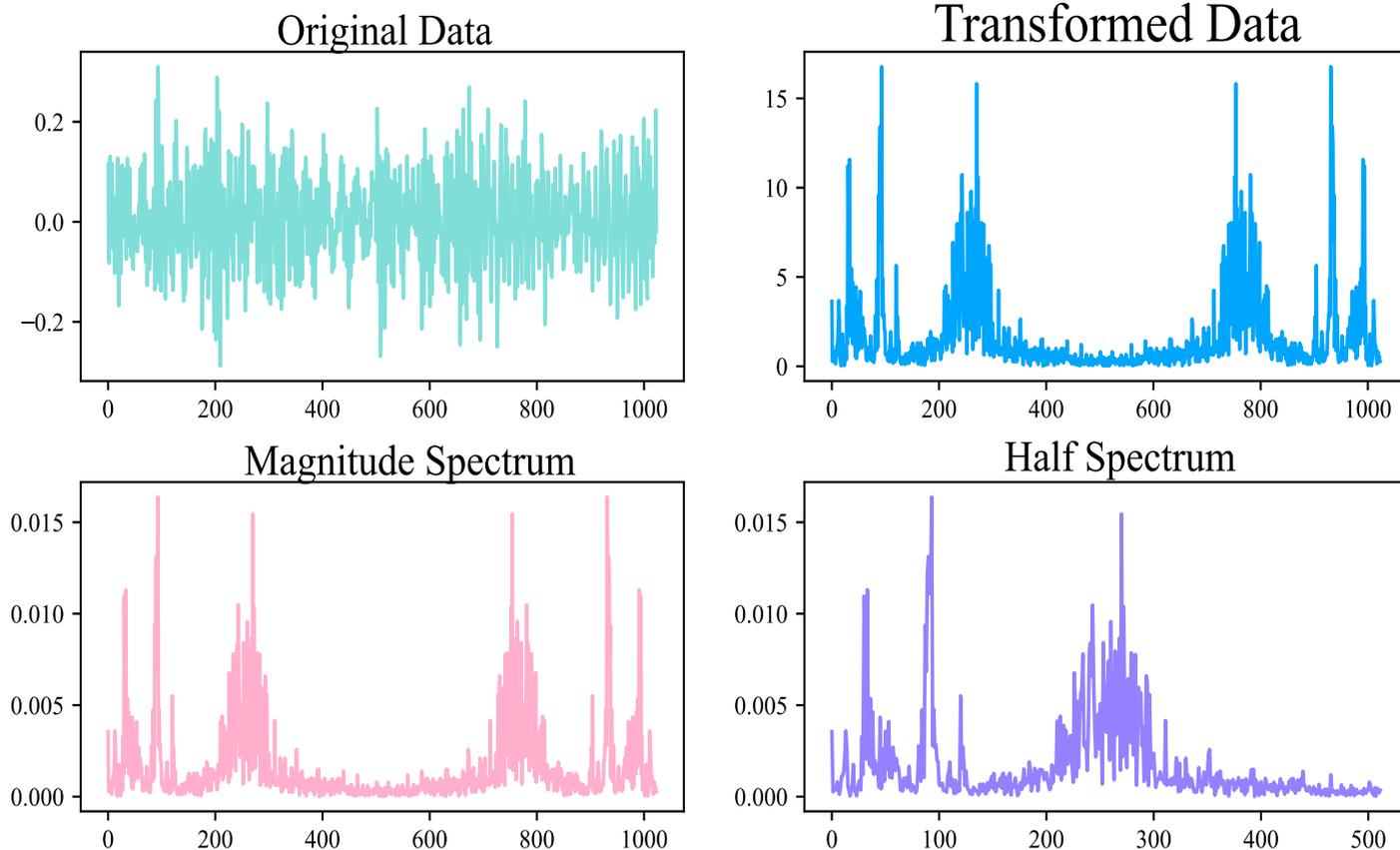


Fig. 3. The FFT processing diagram of fault signals.

3.2. Multiscale Neural Network

The ability to accurately and effectively extract key features that reflect differences between signals is crucial for accurate bearing fault diagnosis. Traditional single-scale neural networks can only cover specific periods of signals, and their feature extraction process is often mechanical, lacking adaptability to changing and complex operating conditions and environments. In contrast, multi-scale neural networks use convolutional units with different sizes of convolutional kernels, allowing the multi-scale feature extraction network to perceive the input signal's field of view with different kernel sizes. This not only reduces the empirical requirements for selecting convolutional kernel sizes but also enables the extraction of robust multi-scale features.

Compared to single-scale features, multi-scale features better capture the description of different fault data. As shown in Figure 4, the proposed method constructs parallel channels of the same shape in the multi-scale network, utilizing convolutional kernels of different sizes paired with varying numbers of filters to extract multi-scale features from samples. Smaller convolutional kernels focus more on local connections within the data, emphasizing the localization of key information in the signal, while larger convolutional kernels are conducive to extracting global features of the signal. In order to enrich the scale of feature perspectives, the convolutional kernel sizes should cover a certain range. Choosing odd-sized kernels can match the center point of the data, reducing the likelihood of feature shifting. Therefore, the convolutional kernel sizes for different parallel channels are set as 3, 11, and 17.

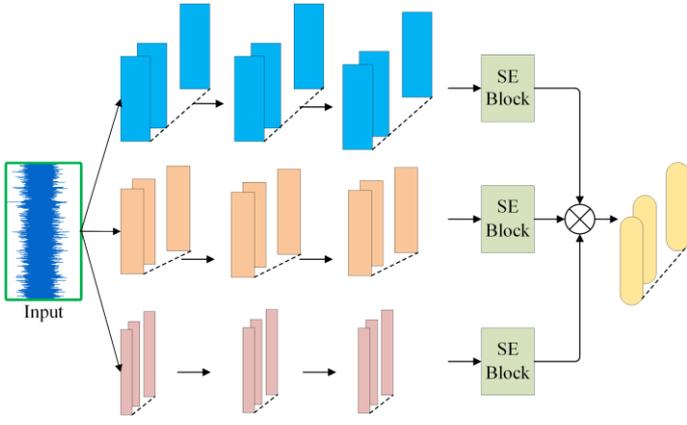


Fig. 4. The diagram of multi-scale neural network.

3.3. Squeeze-and-Excitation Attention Mechanism

The multi-scale neural network can capture fault information from different granularities of the original vibration signals. In order to further extract important features from the signals, this paper integrates an attention mechanism into the network architecture, as shown in Figure 4. The attention mechanism can learn the importance of different features in bearing fault diagnosis, thereby weighting the features. Consequently, the model can focus more on the features relevant to fault diagnosis, reduce reliance on irrelevant features, and improve diagnostic accuracy.

The SE attention mechanism dynamically adjusts the responses of different channels in the feature map by learning the importance of each channel, thereby enhancing the network's representational capacity. Assuming the input feature map is $X \in \mathbb{R}^{H \times W \times C}$, where H and W denote the height and width of the feature map, respectively, and C represents the number of channels. The operations of the SE network can be divided into two steps: Squeeze and Excitation.

Squeeze. In the Squeeze step, global pooling operation is applied to the feature map of each channel, compressing it into a single value. Global average pooling operation is used. For each channel C, its compressed representation z_c can be computed as:

$$z_c = AvgPool(X_c) \quad (10)$$

Excitation. In the Excitation step, each channel's compressed representation is mapped to a new representation space through fully connected layers and an activation function. This process can be represented by a subnetwork, which captures relationships between channels by learning weights for each channel. Assuming the parameters of the Excitation

subnetwork are W_{exc} and b_{exc} , and ReLU activation function is used, the excitation value s_c for each channel c can be computed as:

$$s_c = \sigma(\text{ReLU}(W_{exc} \cdot z_c + b_{exc})) \quad (11)$$

Finally, by multiplying each channel's excitation value s_c with the original feature map, we obtain the weighted feature map:

$$Y = X \otimes s \quad (12)$$

Where \otimes represents element-wise multiplication operation.

3.4. Joint Maximum Mean Discrepancy

JMMD is an extension of MMD that introduces an embedding function to enhance the performance of distribution comparison. The SD dataset and TD dataset are denoted as X and Y respectively, where X contains n vibration signal samples and Y contains m vibration signal samples.

For the sample set X , compute the mean μ_X and covariance matrix ΣX :

$$\mu_X = \frac{1}{n} \sum_{i=1}^n f(x_i) \quad (13)$$

$$\Sigma X = \frac{1}{n} \sum_{i=1}^n (f(x_i) - \mu_X)(f(x_i) - \mu_X)^T \quad (14)$$

For the sample set Y , compute the mean μ_Y and covariance matrix ΣY :

$$\mu_Y = \frac{1}{m} \sum_{j=1}^m g(y_j) \quad (15)$$

$$\Sigma Y = \frac{1}{m} \sum_{j=1}^m (g(y_j) - \mu_Y)(g(y_j) - \mu_Y)^T \quad (16)$$

Measure the similarity between two sets of vibration signals by calculating the mean discrepancy between embedded samples. Specifically, you can compute the square of JMMD as:

$$L_{JMMD} = JMMD^2(X, Y) = \|\mu_X - \mu_Y\|_F^2 \quad (17)$$

During the construction of the transfer learning network, JMMD is combined with the cross-entropy loss function. This integration is intended to enhance the similarity between the predicted data distribution of the model and the actual data distribution, simultaneously mitigating the distribution gap between the SD and TD.

3.5. Concentrate Loss (C-Loss)

The traditional cross-entropy loss function computes loss by assessing the disparity between the predicted probability

distribution and the true labels:

$$\text{Cross-Entropy Loss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(\hat{y}_{ij}) \quad (18)$$

Where N represents the number of samples, while C denotes the number of classes, y_{ij} is the true label of the i -th sample for the j -th class, which is 1 if the i -th sample belongs to the j -th class, and 0 otherwise; \hat{y}_{ij} is the predicted probability by the model for the i -th sample for the j -th class.

For traditional balanced fault diagnosis, the cross-entropy loss function treats the classification cost for each class equally, the total loss is the summation of losses across all samples. However, in the face of data imbalance, the loss function also affects the performance of fault diagnosis. This is mainly manifested in:

$$L = L_1 + L_2 = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C \omega_j (1 - \hat{y}_{ij})^{\alpha} y_{ij} \log(\hat{y}_{ij}) + L_{JMMD} \quad (19)$$

Improvements over traditional loss functions include:

- 1) Introducing domain adaptation loss using the JMMD algorithm to minimize the difference between the two domains.
- 2) Introducing weighting factor ω_j to assign different weights to samples of different quantities, to balance the difference in quantity between healthy samples and fault samples.
- 3) Incorporating a scaling factor to adjust the weighting of losses, diminishing the impact of easily classified samples and augmenting the significance of challenging samples. This adjustment directs the neural network's focus towards the more challenging samples during training.

4. Case Study 1: Case Western Reserve University Dataset

4.1. Introduction to the Case Western Reserve University Experimental Platform

The CWRU dataset is derived from the Case Western Reserve University's bearing fault simulation test rig, as shown in Figure 5. The test rig consists of a 1.5 kW motor, a torque sensor, drive-end bearings, and fan-end bearings. Accelerometers, affixed to the casing through magnetic bases, are employed for collecting vibration data. The accelerometers are situated at the 12 o'clock positions on both the drive end and fan end of the motor casing.

This paper utilizes fan-end bearing fault data, with the bearing model being SKF6203. The fault is created through

electrical discharge machining, with a fault diameter of 0.1778 mm and a depth of 0.2794 mm. The sampling frequency is set at 12 kHz.

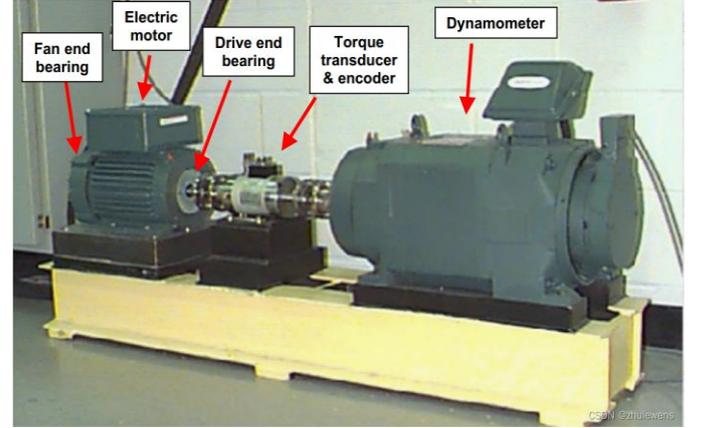


Fig. 5. The test rig of CWRU fault experimental.

4.2. Construction of Imbalanced Dataset

To validate the result of the proposed method, the CWRU public dataset is utilized for validation. This paper selects data from four health conditions: normal, inner race fault, outer race fault and rolling element fault condition for the fan-end bearing. The health conditions and corresponding labels are shown in Table 1 (Bold labels represent imbalanced data). Additionally, experiments on transfer learning are conducted considering data at different operating speeds: 1772, 1750, and 1730 r/min. Six transfer tasks are established, as outlined in Table 2.

Table 1. Labels and health conditions.

Health conditions	Label
Normal	0
0.007_InnerRace	1
0.007_Ball	2
0.007_OutRace6	3

Table 2. Transfer tasks and results

Task	Transfer Task	Accuracy(r/min)	Loss
α	1772rpm→1750rpm	98.80%	0.0495
β	1772rpm→1730rpm	99.40%	0.0537
γ	1750rpm→1772rpm	99.39%	0.0276
δ	1750rpm→1730rpm	99.80%	0.0151
ϵ	1730rpm→1772rpm	98.79%	0.0916
ζ	1730rpm→1750rpm	98.80%	0.0542

In this study, data from the fan-end bearing at three different speeds are selected. The sampling length is set to 1024, with a total of 827 samples for each transfer task, including 827 SD

samples and 166 TD samples. The dataset is divided into SD training set, SD validation set, and TD validation set. To simulate the imbalance in the dataset, out of the 827 SD samples, there are 573 healthy samples and a total of 254 fault samples across various categories. The SD data is divided into SD training set and SD validation set in an 8:2 ratio. The TD data consists of 95 healthy samples and 71 fault samples out of 166 samples. Through the aforementioned data processing, an imbalanced dataset with a majority of healthy samples is constructed.

4.3. Experimental Results and Analysis

The solver settings include an initial learning rate of 1×10^{-4} and a batch size of 16. The model is trained for 300 epochs. The experimental findings are presented in Table 2, where each task is repeated 10 times, and the average accuracy and average loss over 10 repetitions are taken as the experimental results. To further illustrate the effectiveness of the proposed method, confusion matrices for the 6 transfer tasks are plotted in Figure 6. It can be observed from the confusion matrices that the

proposed transfer learning method for bearing fault diagnosis performs well on all 6 tasks. To showcase the effectiveness of the proposed method, it is benchmarked against five other methods: CNN_1d[22], Resnet18[15], S-CNN[31], MK-CNN[32], and Swin Transformer[33]. t-SNE plots are drawn in Figure 7. It can be concluded that the proposed method achieves better clustering results and can perform better fault classification compared to the other five methods. This article also includes accuracy and loss as evaluation indicators. The experimental results are shown in Table 3, Table 4 and Figure 8. From the table and figure, it can be seen that the method proposed in this article has a high accuracy in each migration task, with relatively small loss values, and the area under the ROC curve for each migration task is also the largest.

To demonstrate the effectiveness of the proposed method, we conducted ablation experiments, including E-CNN, E-CNN+JMMD, E-CNN+SE, and E-CNN +JMMD+SE. The results are shown in Figure 9 (ROC curves). The curve of our method is closest to the upper-left corner and has the largest area underneath it, demonstrating the effectiveness of our approach.

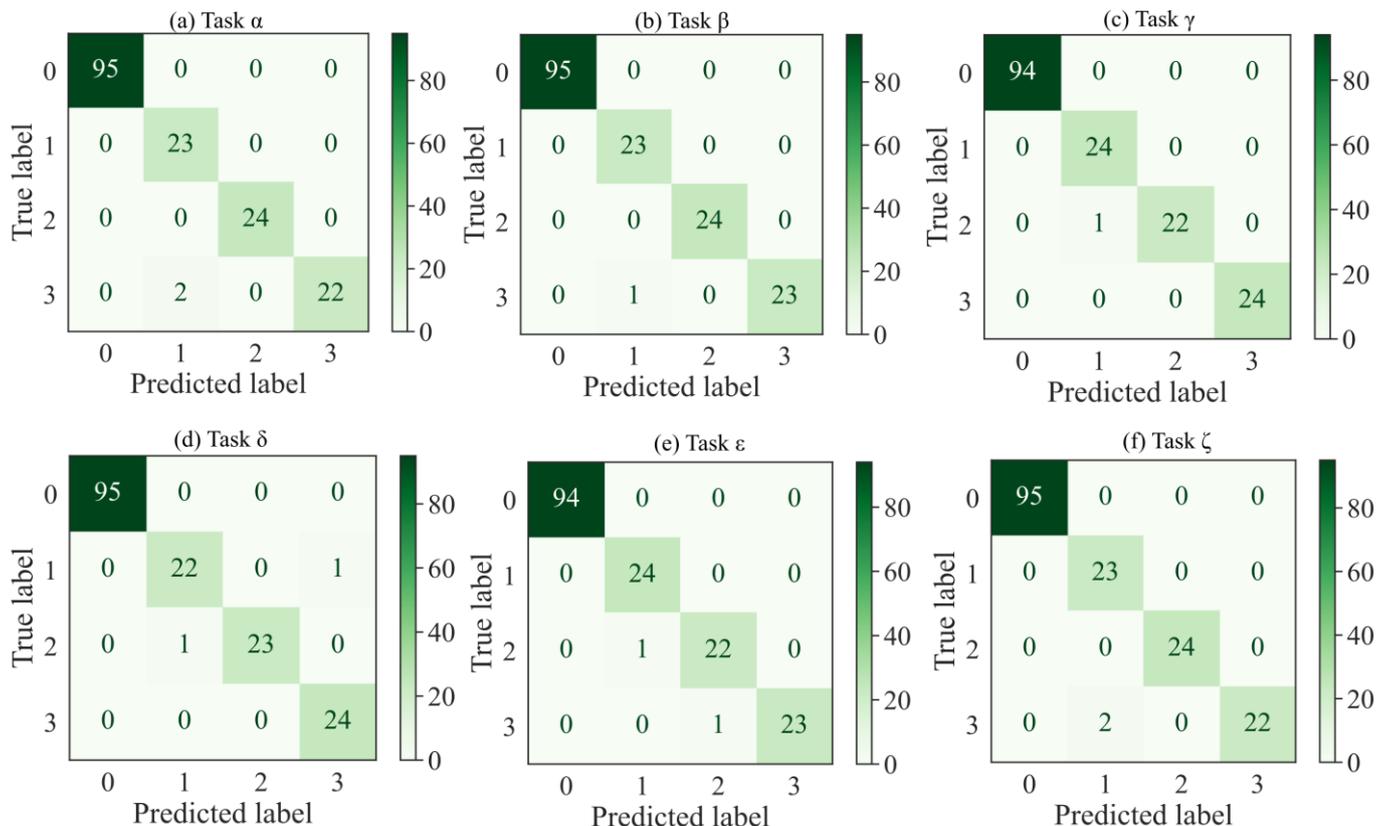


Fig. 6. The confusion matrix on CWRU dataset.

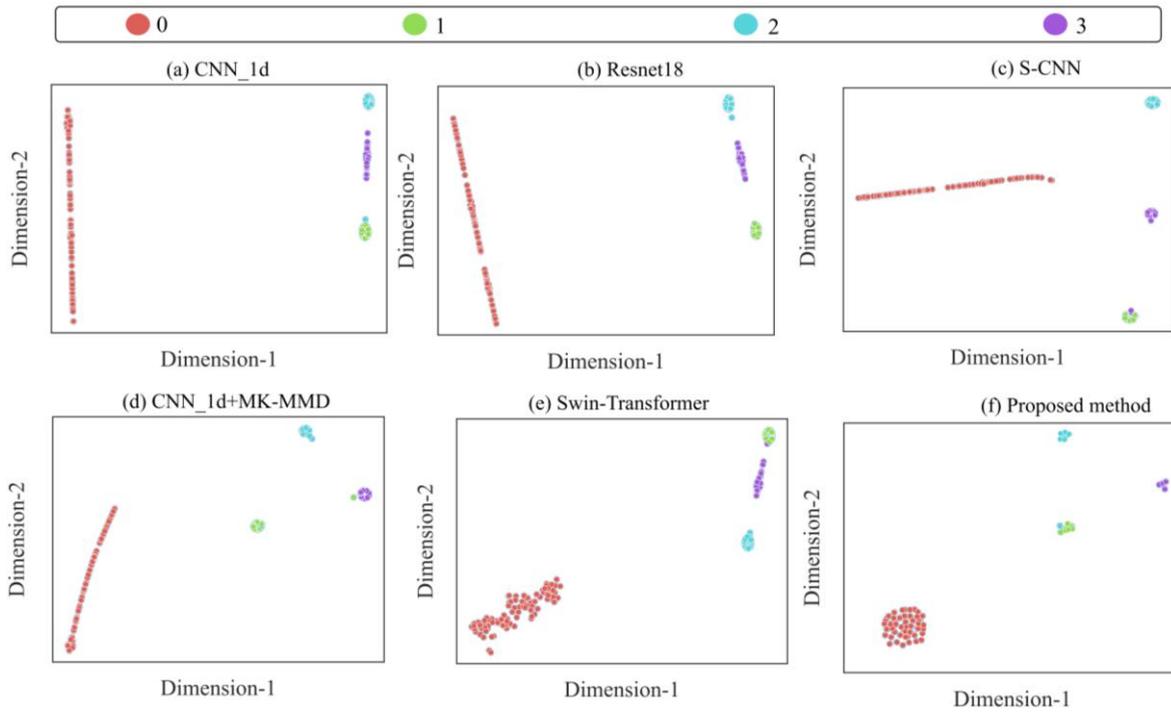


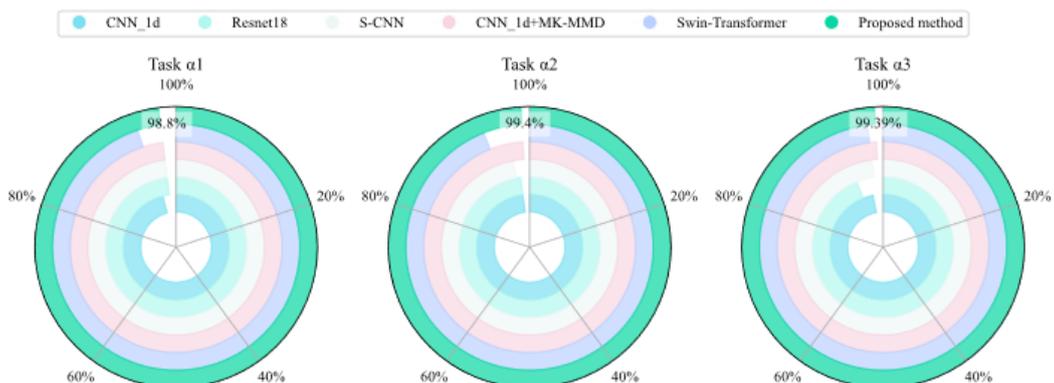
Fig. 7. t-SNE clustering plot.

Table 3. The accuracy of comparison method in case1.

Task	CNN 1d	Resnet18	S-CNN	CNN 1d+MK-MMD	Swin-transformer	Proposed method
α	96.04%	98.34%	98.20%	98.64%	95.88%	98.80%
β	98.31%	98.73%	97.78%	99.16%	94.33%	99.40%
γ	97.51%	94.74%	98.10%	99.06%	98.39%	99.39%
δ	98.68%	99.12%	99.27%	98.54%	97.37%	99.80%
ε	94.74%	96.49%	98.60%	98.89%	95.91%	98.79%
ζ	94.44%	96.58%	97.86%	97.32%	95.30%	98.80%

Table 4. The loss of comparison method in case1.

Task	CNN 1d	Resnet18	S-CNN	CNN 1d+MK-MMD	Swin-transformer	Proposed method
α	0.6910	0.1123	0.5719	0.0833	0.1056	0.0495
β	0.3476	0.2364	0.7004	0.0804	0.6700	0.0537
γ	0.4573	0.5081	0.2056	0.0304	0.0916	0.0276
δ	0.1020	0.0163	0.0603	0.0384	0.0390	0.0151
ε	1.0034	1.0987	0.0890	0.0459	0.0565	0.0916
ζ	1.1008	0.9816	0.0450	0.0987	0.0961	0.0542



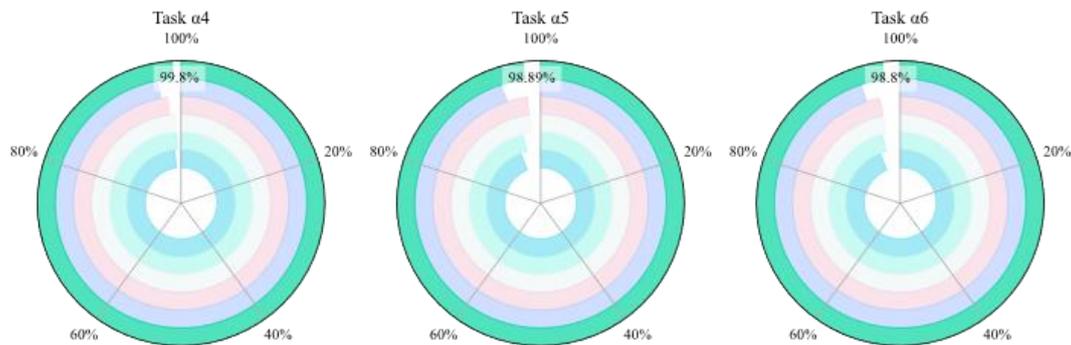


Fig. 8. Accuracy ring chart of case 1.

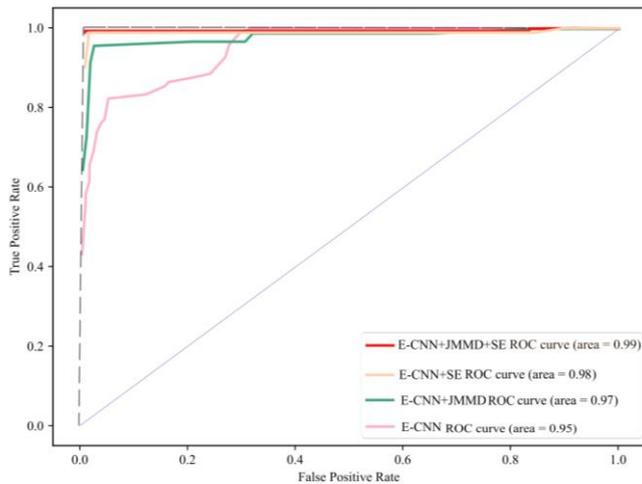


Fig. 9. The ROC curve on CWRU dataset.

5. Case Study 2: Jiangnan University Fan Bearing Dataset

The purpose of this paper is to diagnose faults in fan bearings.

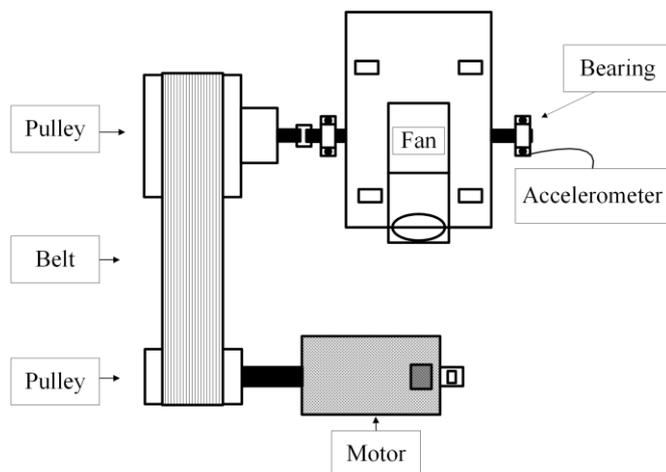


Fig. 10. The test rig of JNU dataset[34].

The Jiangnan University fan bearing data includes four health conditions: inner race fault, rolling element fault, outer race fault, and normal. The correspondence between fault types and labels is shown in Table 5 (Bold labels represent imbalanced data). As shown in Figure 11, the bearing faults are generated

In addition to the commonly used CWRU dataset for fault diagnosis, we also select the fan bearing dataset to validate the proposed method for fan bearing fault diagnosis-related research.

5.1. Introduction to Jiangnan University Experimental Setup

The Jiangnan University bearing dataset is collected from the Jiangnan University fan test rig (as shown in Figure 10)[34]. The centrifugal fan test rig at Jiangnan University consists of a motor, transmission device, coupling, fan bearing under test, accelerometer, fan, fan casing, etc. The electric fan is mainly a type commonly used in industry. Therefore, this experimental data is adopted in this paper to validate the effectiveness of the proposed method.

by wire cutting. Bearing data for each health condition is tested at speeds of 600, 800, and 1000 r/min, respectively, to simulate fan bearing fault data under different operating conditions. Likewise, six distinct transfer tasks are established to validate the effectiveness of our method, as outlined in Table 6.

Table 5. Labels and health conditions

Health conditions	Label
inner race fault	0
normal	1
out race	2
roller fault	3

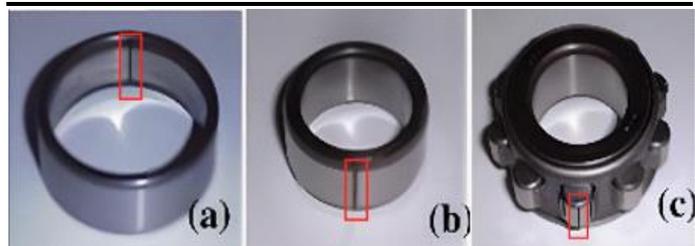


Fig. 11. Schematic Diagram of Bearing Faults. (a) Outer Race Fault, (b) Inner Race Fault, (c) Rolling Element Fault.

Table 6. Transfer tasks and results.

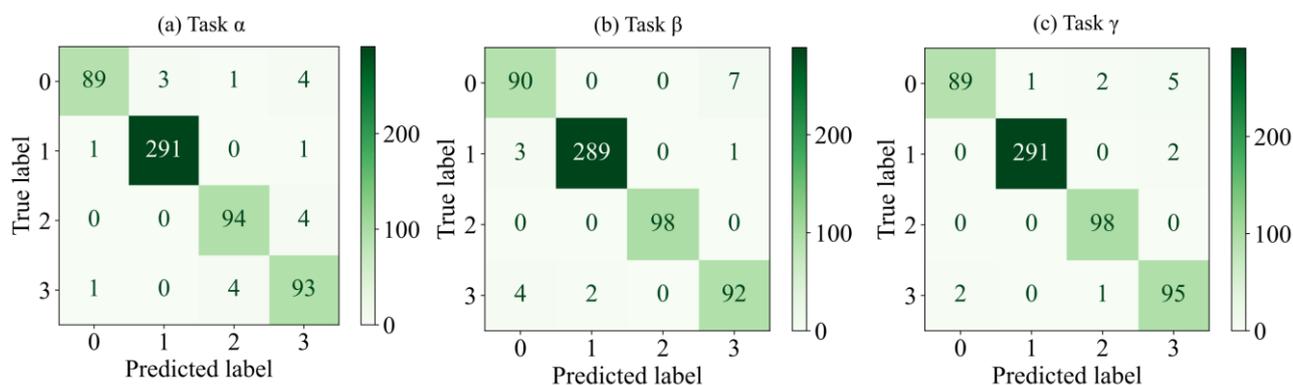
Task	Transfer Task	Accuracy(r/min)	Loss
α	600rpm→800rpm	95.90%	0.1964
β	800rpm→600rpm	97.09%	0.1632
γ	800rpm→1000rpm	98.78%	0.1827
δ	1000rpm→800rpm	96.80%	0.2972
ε	600rpm→1000rpm	95.09%	0.3016
ζ	1000rpm→600rpm	96.01%	0.4037

Similar to the data processing for the CWRU dataset, the dataset is divided into source domain training set, source domain validation set, and target domain validation set. To construct an imbalanced dataset, each source domain training set for the transfer tasks consists of 2432 samples, while each source domain validation set consists of 586 samples, and each target domain validation set consists of 586 samples. Among these, there are 1172 healthy samples in the source domain data and 293 healthy samples in the target domain data, with a total of 293 fault samples across various fault types in the target domain. Through data selection and partitioning, an imbalanced dataset is established.

5.3. Experimental Results

The solver settings and operations on the Jiangnan University dataset are essentially the same as those on the CWRU dataset. The accuracy and loss statistics are also summarized in Table 6. Similarly, confusion matrices for the six transfer tasks are plotted, as shown in the Figure12. It can be observed that the proposed method can effectively identify bearing faults in an imbalanced dataset. To illustrate the superiority of our method, we compared it with five other methods on the Jiangnan University Centrifugal fan bearing dataset and plotted t-SNE clustering graphs, as shown in Figure 13. It is evident that our method outperforms the others in terms of clustering effectiveness. In order to better compare with other methods, we included the accuracy and loss under 6 migration tasks of our method and five comparison methods in the evaluation indicators. The results are shown in Figure 14, Table 7 and Table 8, indicating that our method has a higher accuracy and a lower loss than other methods.

Additionally, to demonstrate the effectiveness of our method, we conducted ablation studies on our approach and plotted ROC curves, as shown in Figure 15. It is evident that our method improves the accuracy of bearing fault diagnosis. Furthermore, to better demonstrate the proposed method's effectiveness and evaluate the classification performance of each method for each type of fan bearing fault, F1 score plots are generated as shown in Figure 16. From the F1 score plots, it is evident that the proposed method performs well in each fault category, generally outperforming the other methods. Although there may be instances where the proposed method's score for certain categories is lower than that of other methods in individual tasks, it still achieves high scores for these categories and can achieve high-accuracy fault diagnosis.



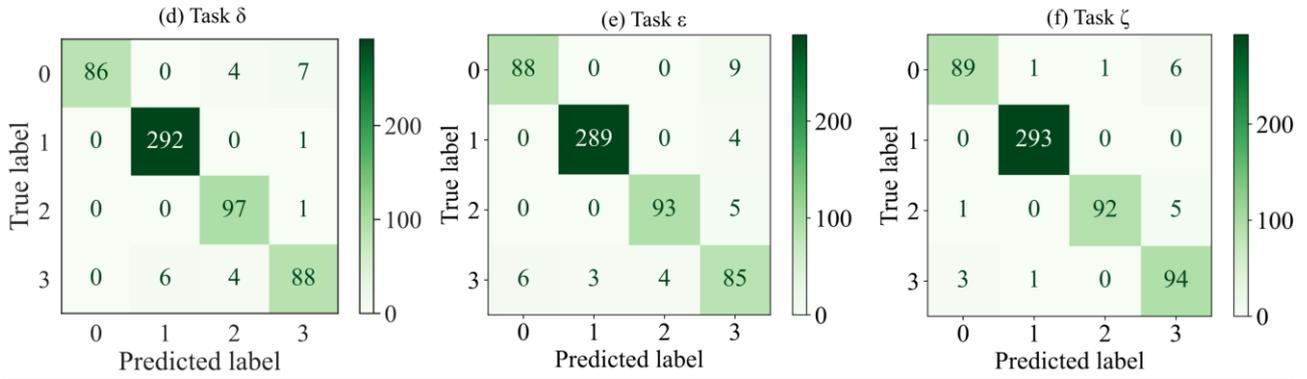


Fig. 12. The confusion matrix on JNU dataset.

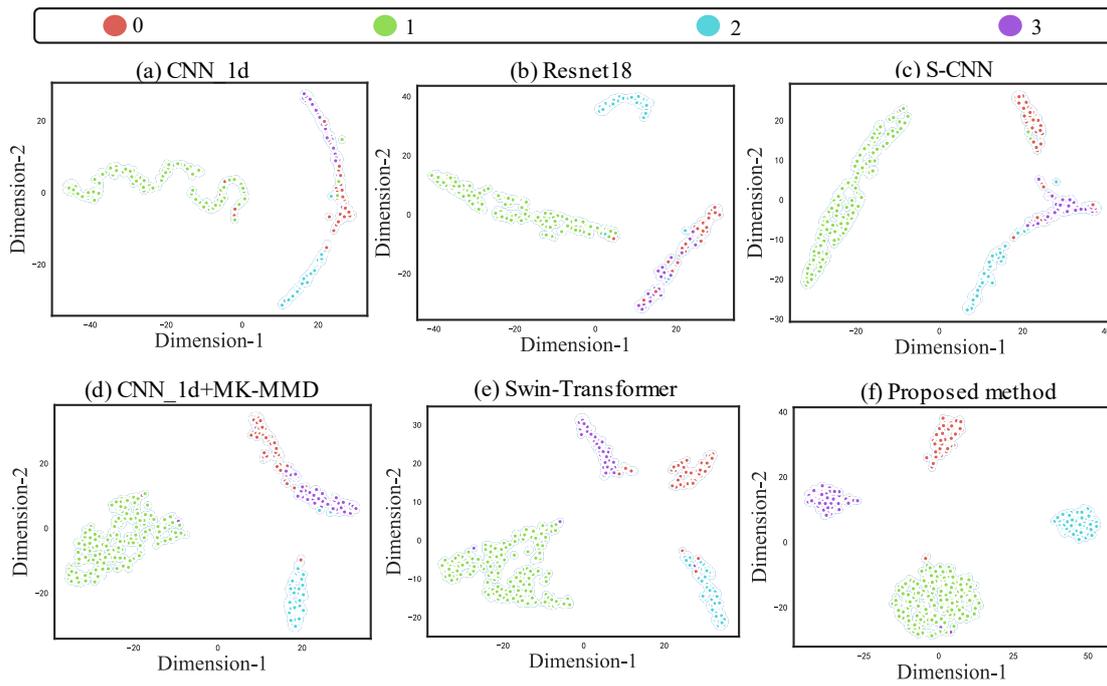


Fig. 13. The t-SNE clustering plot on JNU dataset.

Table 7. The accuracy of comparison method in case 2.

Task	CNN_1d	Resnet18	S-CNN	CNN_1d+MK-MMD	Swin-transformer	Proposed method
α	87.72%	90.88%	92.01%	92.94%	94.69%	95.90%
β	92.22%	93.13%	92.30%	96.72%	95.33%	97.09%
γ	97.51%	94.24%	94.73%	98.39%	96.05%	98.78%
δ	89.96%	91.76%	94.30%	96.62%	93.50%	96.80%
ϵ	90.34%	93.28%	93.51%	94.71%	94.23%	95.09%
ζ	92.71%	95.50%	94.00%	95.03%	95.09%	96.01%

Table 8. The loss of comparison method in case2.

Task	CNN_1d	Resnet18	S-CNN	CNN_1d+MK-MMD	Swin-transformer	Proposed method
α	5.6027	1.4490	1.0122	0.9907	0.6000	0.1964
β	2.0933	0.9006	0.9990	0.3074	0.4501	0.1632
γ	1.2704	0.7041	0.5038	0.2106	0.3340	0.1827
δ	3.9000	1.4098	0.4602	0.3003	0.8976	0.2972
ϵ	2.5092	0.9081	0.9973	0.7034	0.6043	0.3016
ζ	2.3078	0.3901	0.4072	0.4156	0.4889	0.4037

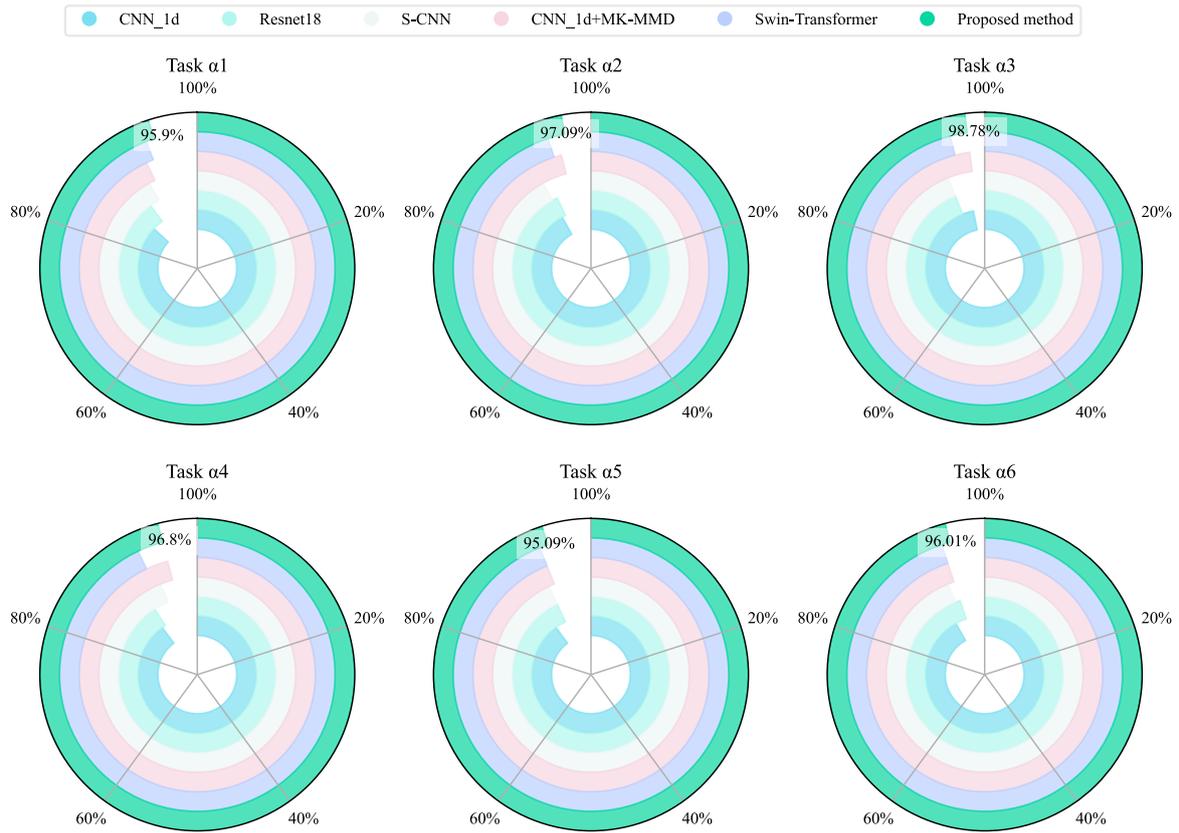


Fig. 14. Accuracy ring chart of case 2.

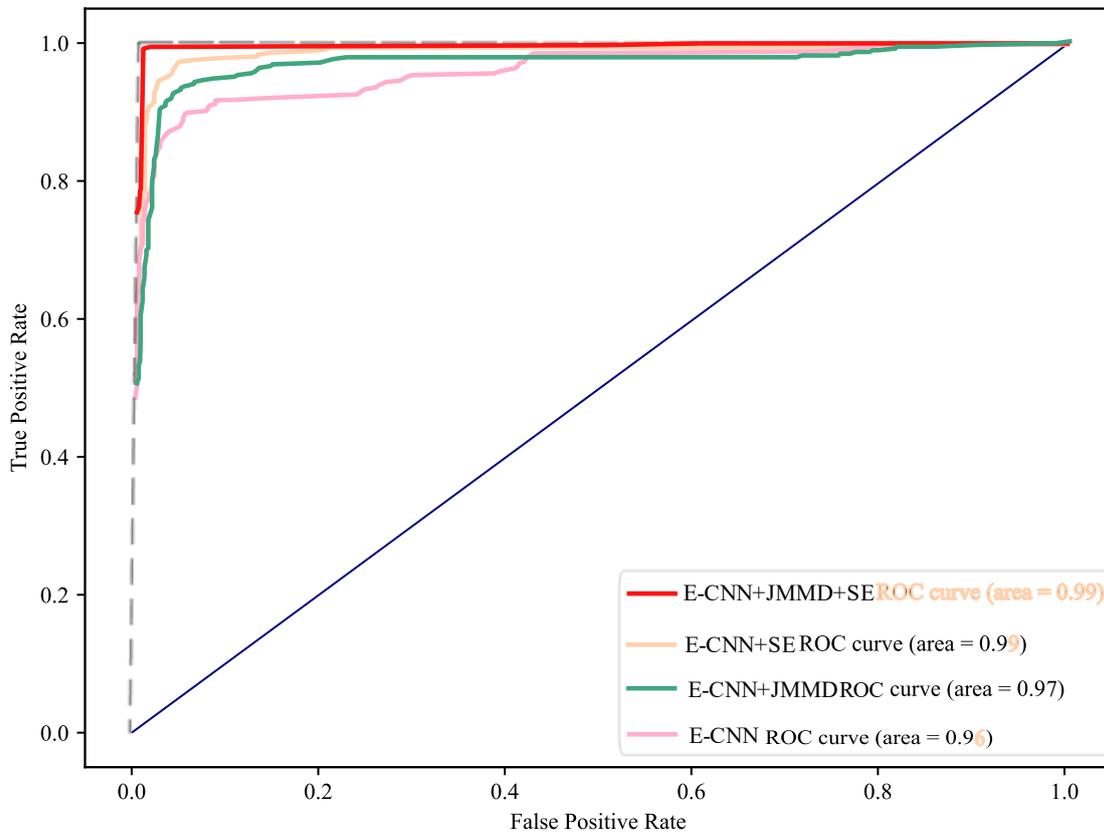


Fig. 15. The ROC curve plot on JNU dataset.

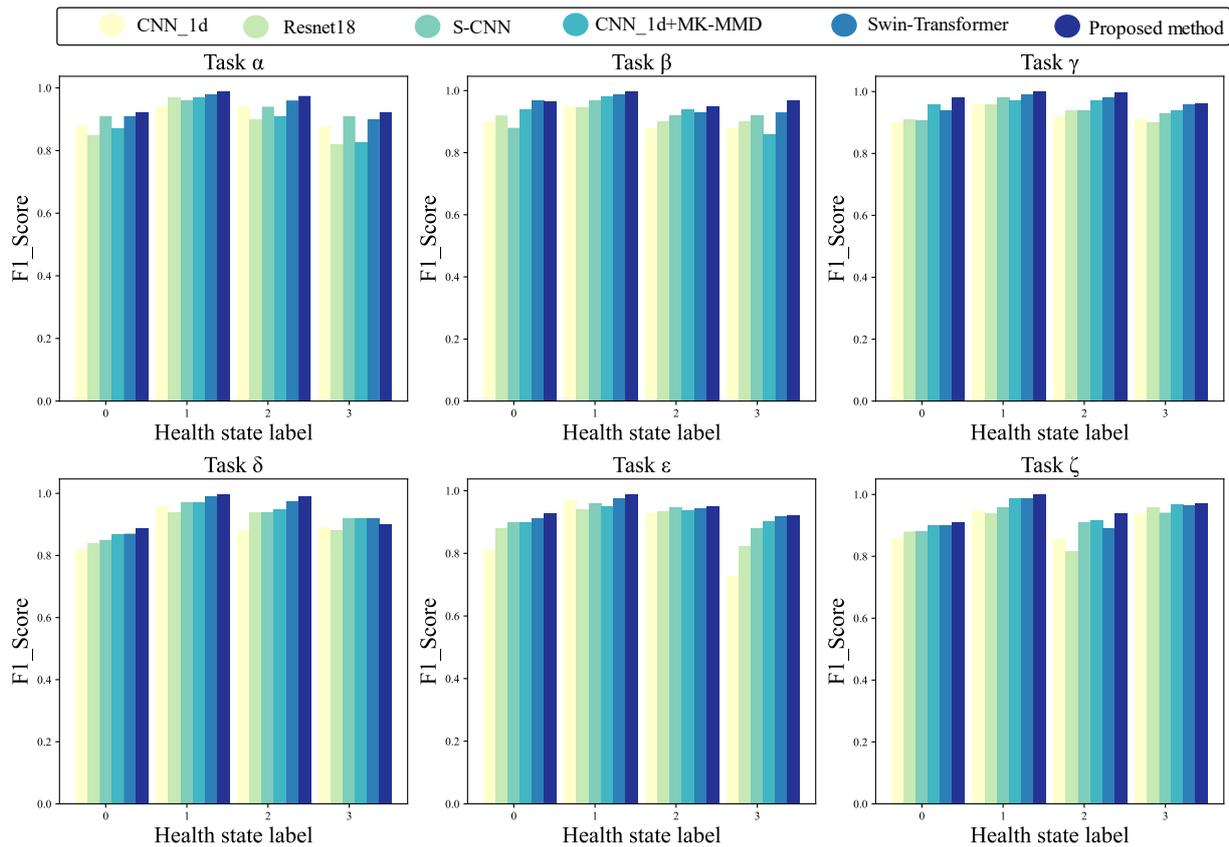


Fig. 16. The F1 score on JNU dataset.

In order to compare common signal processing methods such as wavelet transform, Fourier transform, power spectrum analysis, autocorrelation function, variational mode decomposition, etc., we conducted experiments on the wind turbine bearing dataset of Jiangnan University using Python 1.10.2 software; CPU is i5-12400F; The GPU is RTX3050. We selected the data under the working condition of 600 rpm for analysis, with 4 data files, each containing 500000 single column data points. We selected the model running time as the evaluation index to evaluate the processing speed of the model, and the results are shown in Table 2. Although the amount of data we selected is relatively small compared to the large amount of data for big data analysis, as shown in Table 9, the Fourier transform takes the shortest time and has the fastest processing speed. Therefore, in this article, the Fourier transform is used for preliminary data processing. These common signal processing methods cannot directly classify faults. The main reason for the time-consuming processing of large-scale data is to obtain relevant features through these analyses and then rely on manual classification or commonly used classification methods for fault classification and

recognition.

Table 9. The comparison in common signal processing methods

Method	WT	FT	PSA	AF	VMD
Time	0.4427	0.1410	0.3160	0.3420	64.2839

In addition, we also made corresponding comparisons with commonly used fault signal classification methods. We also selected data at a speed of 600 and compared it with SVM and random forest methods. We also conducted one training session, and the time and accuracy are shown in Table 10. From the table, it can be seen that the CNN model in this article is more suitable for processing large-scale data. The amount of data in this article is relatively small compared to the large industrial data. As the amount of data increases, the powerful feature extraction ability of convolutional neural networks can play a greater role.

Table 10. Comparison of common signal classification methods

Method	SVM	Random forest	CNN
Time(s)	1.4989	10.1697	1.4972
Accuracy	0.4659	0.5935	0.7362

We added Gaussian noise with standard deviations of 0.5 and 1 to the original signals in the Jiangnan University dataset to simulate the noise situation in industrial practice and verify the robustness of our method to noise. Then, a new experimental

study was conducted on the migration task, and the results are shown in Table 11. It can be seen that although the accuracy of our method fluctuates after adding noise, it still remains above 90%, indicating that the proposed model has a certain degree of robustness when facing noise.

Table 11. Accuracy under different levels of noise on JNU

Task	Transfer Task	Accuracy	
		Std 0.5	Std 1.0
α	600rpm→800rpm	92.02%	90.90%
β	800rpm→600rpm	96.42%	94.60%
γ	800rpm→1000rpm	96.30%	95.09%
δ	1000rpm→800rpm	93.42%	91.04%
ε	600rpm→1000rpm	94.44%	91.46%
ζ	1000rpm→600rpm	95.69%	93.99%

6. Conclusion

In the context of imbalanced samples, a transfer learning method for fan bearing fault diagnosis based on I-CNN and JMMD is proposed. This method addresses the issue of sample

imbalance in fault diagnosis while also considering the challenges of data collection in the target domain and insufficient training data in the target domain by applying transfer learning algorithms. Effective features are extracted by performing FFT transformation on the data before processing them in the neural network. Furthermore, the SE attention mechanism is embedded in a parallel multi-scale neural network to extract key information from the signals. The JMMD algorithm is introduced within the transfer learning framework to calculate the maximum mean difference between the SD and TD, thereby minimizing losses while reducing domain shift between the SD and TD. Additionally, to address the issue of sample imbalance, a loss function based on weight factors and scaling factors is proposed, which focuses more on small samples and easily confused samples in imbalanced samples, thereby improving fault diagnosis performance in the context of sample imbalance.

Acknowledgement

This work is supported in part by the Natural Science Foundation of Heilongjiang Province (LH2020A017), in part by the Key Laboratory of Vibration and Control of Aero-Propulsion System, Ministry of Education, Northeastern University (VCAME202209) and in part by the Harbin science and technology innovation talent project (2023HBRCCG004)

Reference

1. A. Fortini, A. Suman, and N. Zanini, "An experimental and numerical study of the solid particle erosion damage in an industrial cement large-sized fan," *Engineering failure analysis*, vol. 146, p. <https://doi.org/10.1016/j.engfailanal.2023.107058>
2. N. Ouerghemmi, S. B. Salem, K. Bacha, and A. Chaari, "V-belt Fault Detection using Motor Current Signature Analysis in Centrifugal Fan System," in *2021 IEEE 2nd International Conference on Signal, Control and Communication (SCC)*, 2021: IEEE, pp. 1-6. <https://doi.org/10.1109/SCC53769.2021.9768345>
3. C. Medrea, D. Papageorgiou, H. Bravos, and I. Chicinaş, "Failure analysis of a fan blade holding assemblment, installed on the cooling tower of a power plant," *Engineering Failure Analysis*, vol. 127, p. <https://doi.org/10.1016/j.engfailanal.2021.105505>
4. M. Kumar, T. Maity, and M. K. Kirar, "Energy savings through VOD (Ventilation-on-Demand) analysis in Indian underground coal mine," *IEEE Access*, vol. 10, pp. <https://doi.org/10.1109/ACCESS.2022.3203710>
5. Y. Yang *et al.*, "Purification technology of oil mist in industrial buildings: A review," *Building and Environment*, p. <https://doi.org/10.1016/j.buildenv.2023.110229>.
6. J. Liu and Y. Shao, "Overview of dynamic modelling and analysis of rolling element bearings with localized and distributed faults," *Nonlinear Dynamics*, vol. 93, no. 4, 2018. <https://doi.org/10.1007/s11071-018-4314-y>
7. Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, p. <https://doi.org/10.1016/j.ymssp.2019.106587>
8. A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Systems with Applications*, vol. 166, p. 114060, 2021. <https://doi.org/10.1016/j.eswa.2020.114060>
9. R. Yan *et al.*, "Wavelet transform for rotary machine fault diagnosis: 10 years revisited," *Mechanical Systems and Signal Processing*, vol. 200, p. 110545, 2023. <https://doi.org/10.1016/j.ymssp.2023.110545>
10. K. O. M. Touati, M. Boudiaf, L. Mazouz, and L. Cherroun, "Efficient hybrid strategy based on FFT and fuzzy logic techniques applied to

- fault diagnosis in power transmission line," *Soft Computing*, pp. 1-20, 2023. <https://doi.org/10.1007/s00500-023-09089-6>
11. Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," *Mechanical Systems and Signal Processing*, vol. 140, p. 106683, 2020. <https://doi.org/10.1016/j.ymssp.2020.106683>
 12. X. Jiang *et al.*, "An adaptive and efficient variational mode decomposition and its application for bearing fault diagnosis," *Structural Health Monitoring*, vol. 20, no. 5, pp. 2708-2725, 2021. <https://doi.org/10.1177/1475921720970856>
 13. A. Abid, M. T. Khan, and J. Iqbal, "A review on fault detection and diagnosis techniques: basics and beyond," *Artificial Intelligence Review*, vol. 54, no. 5, pp. 3639-3664, 2021. <https://doi.org/10.1007/s10462-020-09934-2>
 14. J. Jiao, M. Zhao, J. Lin, and K. Liang, "A comprehensive review on convolutional neural network in machine fault diagnosis," *Neurocomputing*, vol. 417, pp. 36-63, 2020. <https://doi.org/10.1016/j.neucom.2020.07.088>
 15. X. Li, K. Su, D. Li, Q. He, Z. Xie, and X. Kong, "Transfer learning for bearing fault diagnosis: adaptive batch normalization and combined optimization method," *Measurement Science and Technology*, vol. 35, no. 4, p. 046106, 2024. <https://doi.org/10.1088/1361-6501/ad19c2>
 16. H. Tang *et al.*, "Feature extraction of multi-sensors for early bearing fault diagnosis using deep learning based on minimum unscented kalman filter," *Engineering Applications of Artificial Intelligence*, vol. 127, p. 107138, 2024. <https://doi.org/10.1016/j.engappai.2023.107138>
 17. L. Xie, Q. Miao, Y. Chen, W. Liang, and M. Pecht, "Fan bearing fault diagnosis based on continuous wavelet transform and autocorrelation," in *Proceedings of the IEEE 2012 Prognostics and System Health Management Conference (PHM-2012 Beijing)*, 2012: IEEE, pp. 1-6. <https://doi.org/10.1109/PHM.2012.6228837>
 18. W. Cui *et al.*, "Fault Diagnosis of Rolling Bearings in Primary Mine Fans under Sample Imbalance Conditions," *Entropy*, vol. 25, no. 8, p. 1233, 2023. <https://doi.org/10.3390/e25081233>
 19. W. He, Q. Miao, M. Azarian, and M. Pecht, "Health monitoring of cooling fan bearings based on wavelet filter," *Mechanical Systems and Signal Processing*, vol. 64, pp. 149-161, 2015. <https://doi.org/10.1016/j.ymssp.2015.04.002>
 20. C. Qian, J. Zhu, Y. Shen, Q. Jiang, and Q. Zhang, "Deep transfer learning in mechanical intelligent fault diagnosis: application and challenge," *Neural Processing Letters*, vol. 54, no. 3, pp. 2509-2531, 2022. <https://doi.org/10.1007/s11063-021-10719-z>
 21. X. Zhao, M. Ma, and F. Shao, "Bearing fault diagnosis method based on improved Siamese neural network with small sample," *Journal of Cloud Computing*, vol. 11, no. 1, p. 79, 2022. <https://doi.org/10.1186/s13677-022-00350-1>
 22. H. Cao, H. Shao, X. Zhong, Q. Deng, X. Yang, and J. Xuan, "Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds," *Journal of Manufacturing Systems*, vol. 62, pp. 186-198, 2022. <https://doi.org/10.1016/j.jmsy.2021.11.016>
 23. Y. Xiao, H. Shao, Z. Min, H. Cao, X. Chen, and J. J. Lin, "Multiscale dilated convolutional subdomain adaptation network with attention for unsupervised fault diagnosis of rotating machinery cross operating conditions," *Measurement*, vol. 204, p. 112146, 2022. <https://doi.org/10.1016/j.measurement.2022.112146>
 24. W. Mao, Y. Liu, L. Ding, and Y. Li, "Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study," *Ieee Access*, vol. 7, pp. 9515-9530, 2019. <https://doi.org/10.1109/ACCESS.2018.2890693>
 25. Q. Hang, J. Yang, and L. Xing, "Diagnosis of rolling bearing based on classification for high dimensional unbalanced data," *IEEE Access*, vol. 7, pp. 79159-79172, 2019. <https://doi.org/10.1109/ACCESS.2019.2919406>
 26. J. Lu, W. Wu, X. Huang, Q. Yin, K. Yang, and S. Li, "A modified active learning intelligent fault diagnosis method for rolling bearings with unbalanced samples," *Advanced Engineering Informatics*, vol. 60, p. 102397, 2024. <https://doi.org/10.1016/j.aei.2024.102397>
 27. V. Rai and A. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform," *Mechanical systems and signal processing*, vol. 21, no. 6, pp. 2607-2615, 2007. <https://doi.org/10.1016/j.ymssp.2006.12.004>
 28. G. Mesnil *et al.*, "Unsupervised and transfer learning challenge: a deep learning approach," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 2012: JMLR Workshop and Conference Proceedings, pp. 97-110.
 29. Z. Zhao *et al.*, "Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-28, 2021. <https://doi.org/10.1109/TIM.2021.3116309>
 30. R. K. Sanodiya and L. Yao, "Unsupervised transfer learning via relative distance comparisons," *IEEE Access*, vol. 8, pp. 110290-110305,

2020. <https://doi.org/10.1109/ACCESS.2020.3002666>

31. X. Li, P. Yuan, X. Wang, D. Li, Z. Xie, and X. Kong, "An unsupervised transfer learning bearing fault diagnosis method based on depthwise separable convolution," *Measurement Science and Technology*, vol. 34, no. 9, p. 095401, 2023. <https://doi.org/10.1088/1361-6501/acda55>
32. C. Che, H. Wang, Q. Fu, and X. Ni, "Deep transfer learning for rolling bearing fault diagnosis under variable operating conditions," *Advances in Mechanical Engineering*, vol. 11, no. 12, p. 1687814019897212, 2019. <https://doi.org/10.1177/1687814019897212>
33. Z. Liu *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10012-10022. <https://doi.org/10.1109/ICCV48922.2021.00986>
34. K. Li, L. Su, J. Wu, H. Wang, and P. Chen, "A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine," *Applied Sciences*, vol. 7, no. 10, p. 1004, 2017. <https://doi.org/10.3390/app7101004>