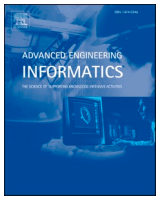
Advanced Engineering Informatics 60 (2024) 102536

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei



Towards multi-scene learning: A novel cross-domain adaptation model based on sparse filter for traction motor bearing fault diagnosis in high-speed EMU

Feiyu Lu [a,*], Qingbin Tong [a,*], Jianjun Xu [a], Ziwei Feng [a], Xin Wang [a], Jingyi Huo [a], Qingzhu Wan [b]

[a] School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China
[b] School of Electrical and Control Engineering, North China University of Technology, Beijing 100144, China

ARTICLE INFO

Keywords:
Bearing fault diagnosis
Sparse filter
Cross-domain adaptation

ABSTRACT

Fault diagnosis of traction motor bearing is of great significance to improve the reliability and safety of high-speed electric multiple units (EMU). While the fault diagnosis method based on cross-domain adaptation has been successful in scenarios involving speed or load fluctuations, existing methods ignore the independence and diversity of features, resulting in unsatisfactory diagnostic results under multi-scene learning, thereby reducing the generalization ability. Moreover, the development of complex models is time-consuming, and their computational efficiency is low. To address these issues, this study proposes a novel cross-domain adaptation model based on sparse filtering (SFCDA), which consists of only two fully connected (FC) layers. Firstly, pre-training is conducted to utilize the soft reconstruction penalties to constrain the weights of sparse filtering and improve the independence of features. The weights of unsupervised training are used to initialize the parameters of the first FC layer of the SFCDA model. Secondly, a multiple sparse regularization (MSR) algorithm is proposed and used to constrain the SFCDA. Then, fine-tuning is conducted, in which the structural alignment function is used to measure the distribution distance between the source and target domain data. Minimizing the kernel norm can improve the diversity of features and enhance the robustness. Finally, the effectiveness of SFCDA in multi-scene learning is proved theoretically. It is validated in three fault diagnosis scenes in four different bearing datasets. The results show that the suggested approach is more straightforward and has a better fault diagnosis effect than the state-of-the-art domain adaptive approaches.

1. Introduction

The safety and reliability of high-speed electric multiple units (EMU) bogies are crucial for ensuring trains' operational quality and safety. When a high-speed EMU is in operation, the traction motor utilizes the interaction forces between the steel rails and the train wheels. Additionally, the traction motor frequently operates under various conditions such as startup, braking, and load variation. As the power source of the EMU bogie, the traction motor relies heavily on the rolling bearing as a core and fundamental component. Bearing failure can pose significant risks to high-speed EMU, which stresses the need for timely safety warnings. Fault diagnosis, therefore, plays a critical role in ensuring the proper functioning of the machine.

The vibration signal is highly sensitive to the pulse shock signal, which makes it an important source of fault information. As such, many scholars consider vibration signal-based fault diagnosis as their primary choice [1–3]. Traditional fault diagnosis methods require the extraction of numerous eigenvalues to describe the vibration signal, which depends on the theoretical knowledge of professional technicians and can be time-consuming, particularly in complex diagnosis tasks [4]. In recent years, with the development and application of machine learning and deep learning technologies, such as support vector machine [5], sparse filter (SF) [6], autoencoders [7], and convolutional neural networks (CNN)[8], fault diagnosis technology has gradually moved towards intelligence and automation [9,10].

The intelligent fault diagnosis model of end-to-end learning can

* Corresponding authors.
E-mail addresses: 21117039@bjtu.edu.cn (F. Lu), qbtong@bjtu.edu.cn (Q. Tong), jjxu@bjtu.edu.cn (J. Xu), 22110480@bjtu.edu.cn (Z. Feng), jyhuo@bjtu.edu.cn (J. Huo), wanqz@ncut.edu.cn (Q. Wan).

https://doi.org/10.1016/j.aei.2024.102536
Received 14 December 2023; Received in revised form 29 February 2024; Accepted 8 April 2024

traverse the error of the objective function to each layer of the model through the back-propagation algorithm, making full use of the existing label information [11–13]. Based on end-to-end learning, a lot of intelligent diagnosis methods have been proposed [14]. The fault classification model of vibration signal under multi-sensor is built by using CNN [15]. Li et al. [16] considered the possible influence of non-Euclidean space between vibration signals, adjusted the receptive file of existing Graph Convolutional Networks (GCNs), proposed a multi receptive field GCNs model, and applied it to gear fault diagnosis. Besides, Wei et al. [17] adopted the end-to-end model based on CNN, realised diagnosis in cross domain by migrating the same features, and have been verified in multiple data sets.

On the other hand, the models based on deep transfer learning are widely used in fault diagnosis [18–20]. Considering that the training and testing set are not the same distribution under some cases, Ma et al. [21] proposed an improved domain adaptation algorithm-weighted transfer component analysis under the framework of domain adaptive transfer learning, and the model was applied in five fault datasets of two bearings. Similarly, through adversarial learning and distance metric learning, [22,23] reduce the distance between the source domain and target domain in the feature space, so that the unsupervised fault diagnosis can be realized by an ordinary classifier. The feature representation alignment networks (FRAN) were developed [24]. This model maximizes the mutual information between different domains to improve knowledge transferability.

In addition to domain shift in fault diagnosis, numerous scholars have taken into consideration the scarcity and imbalance of fault samples due to varying probabilities of fault occurrences in practical engineering applications [25]. To address this, several small or imbalanced fault diagnostic models have been proposed [26]. Meng et al. [27] introduced an adaptive feature fusion-assisted generative adversarial network to mitigate the issue of sample imbalance. This approach employs features to guide the training process of the generative model, expediting model convergence, and has been validated on two bearing fault datasets to demonstrate its efficacy. Furthermore, Wang et al. [28] proposed a generative adversarial minority enlargement method to rectify the learning bias in imbalanced data, substantiated through feasibility verification on 28 datasets. Addressing the challenge of small sample datasets, a novel feature-level consistency regularization module [29] has been devised. This model aims to reduce the Earth Mover's Distance between real and generated data, thereby augmenting the sample data.

Recently, some scholars have gradually paid attention to the issue of fault diagnosis of traction motor bearings [30]. Considering the problem of model performance degradation caused by variable operating conditions in practical engineering, Zhang et al. [31] proposed a frequency-domain data analysis method that overlays multi-frequency band information. They combined it with a fault frequency adaptive decision model, achieving the bearing fault diagnosis. And in addressing the fault diagnosis of subway traction motor bearings, He et al. [32] proposed an intelligent fault diagnosis method based on the fusion of multiple signals from vibration and sound sources. They validated the method using a dataset of subway traction motor bearing faults. However, this method, employing Markov transition fields and deep residual networks, led to suboptimal efficiency in model construction. Addressing the challenges of poor feature extraction capability and low diagnostic accuracy in the diagnosis of train traction motor bearing faults, [33] proposed an improved deep residual network that embeds squeezing excitation into the residual network. They conducted validation on their self-built traction motor test bench. Additionally, [34] proposed the multi-sensor data fusion and dual-scale residual network. This method implemented strategies for multi-information fusion at both the data and feature levels. The validity of the suggested method was analyzed in bearing datasets and publicly available bearing datasets.

While the above methods have demonstrated good performance in bearing fault diagnosis, they still have several shortcomings: 1) Single scene: most existing models are designed to diagnose faults under single-scene conditions, such as speed shift or unbalanced samples. 2) Computational time: most diagnostic models trade off accuracy for time, without considering the time cost required for the model. This approach consumes a significant amount of computation resources. For high-speed EMU traction motor fault diagnosis, there are still the following pain points and difficulties:

(1) In actual industrial scenarios, due to the different failure probability of each part of the bearing, the fault data is small and unbalanced.
(2) For high-speed EMU, the high operating speed, heavy load, and variable speed of the bearings make the diagnosis of traction motor bearing faults in high-speed trains extremely difficult.
(3) Currently, most fault diagnosis model structures are very complex, occupying too much memory and consuming a significant amount of training time. This complexity is not conducive to the transition of the model from theory to application, limiting its application on high-speed EMU. In practical engineering, building a model for each scenario increases both system resources and manpower costs.
(4) Current fault diagnosis models are unable to simultaneously handle fault transfer tasks in multiple scenarios, such as: Case1: One source domain and one target domain (1S1T). Case2: One source domain and multiple target domains (1SmT). Case3: Small and imbalanced (S&I) fault diagnosis. Constructing fault diagnosis models for multiple scenarios can effectively enhance the generalization capability of the high-speed traction motor bearing fault diagnosis model and promote the practical application of the model.

Therefore, it is crucial to mine the data commonality of multiple scenarios and build fault diagnosis models, particularly for fault diagnosis problems in multiple scenarios, such as speed shift, small sample data, and unbalanced data.

SF is an unsupervised feature extraction model. Unlike most unsupervised learning algorithms, sparse filtering learns the distribution of features rather than the distribution of data, which relaxes the limitation of SF on input data. And the characteristics of the bearing vibration signal have not changed significantly after the speed/load shift. Therefore, SF is a promising tool for fault diagnosis in multiple scenarios. Recently, many SF based methods have been presented. For instance, Lei et al. [35] proposed an unsupervised fault feature extraction framework based on SF, and realized the classification of bearing and gear signals. Zhang et al. [36] added $\ell_{3/2-2}$−norm regularization terms to the loss function, which was verified in the single domain fault diagnosis with noise. Zhang et al. [37] proposed a reconstruction sparse filtering (RSF). RSF adopts the idea of orthogonal constraints and adds constraints to the weight matrix in SF, so that the output eigenvectors have orthogonality. Cheng et al. [38] established a deep neural network based on multiple SF modules, and took the frequency domain data of vibration signals as input to realize the single domain fault identification of bearings. Zhang et al. [39] used SF and maximum mean discrepancy (MMD) to extract fault features from the time–frequency diagram of vibration signals, and realized fault diagnosis under domain shift.

Although the SF-based models mentioned above have achieved good results, there are still two main defects in current research: 1) the potential of SF in the field of cross domain adaptation has not been brought into full play. The existing adaptive algorithm based on SF domain is phased [39], and this non-end-to-end architecture cannot find the global optimal solution of the model, which limits the application of SF in multiple scenarios, especially cross domain adaptation. 2) The SF based diagnostic model usually ignores the inconsistent data distribution between the training set and the test set and the S&I data issue. Single domain models based on SF cannot handle the domain shift question, and the S&I data also reduces the performance of the diagnosis model in

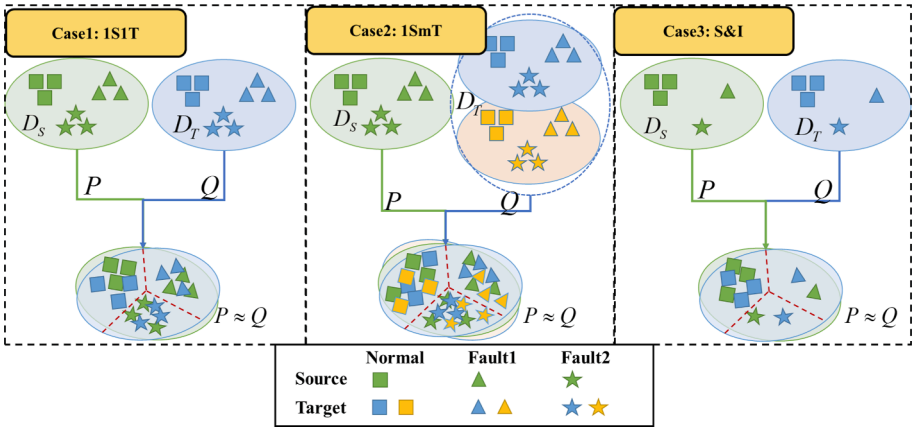


Fig. 1. The three cases of cross-domain adaptation.

the scenario where it is difficult to obtain fault samples [26].

Based on the above analysis, the current fault diagnosis model still has the problem of complex structure and poor performance in multi-task scenarios. Therefore, for traction motor bearing fault diagnosis in high-speed EMU under in multiple scenarios, we propose a cross-domain adaptive fault diagnosis model based on sparse filter, named SFCDA. This model combines the unsupervised feature extraction method SF with the pre-trained and fine-tuning framework to achieve cross domain adaptive fault diagnosis even under S&I data. The contributions of this paper are as follows:

(1) A pre-training method based on the soft reconstruction penalty of SF is designed. This method allocates pre-trained weights to a single fully connected neural network, initializing the weights of the first network layer, thereby significantly improving the model's feature aggregation capability. Experimental discussions indicate that this pre-training method can be combined with various deep learning architectures, enhancing the model's performance.
(2) A multiple sparse regularization (MSR) algorithm is proposed, which can adapt to domain shift data under different speeds and has strong regularization constraints.
(3) A joint domain adaptation loss function is constructed for the fine-tuning stage of the model. This loss function, from the perspective of the marginal distribution, promotes the clustering of source and target domain features while increasing the diversity of the model's output features.
(4) An end-to-end domain adaptation architecture based on a two-layer fully connected neural network, named SFCDA, is proposed. To our information, SFCDA is the first end-to-end domain adaptation model based on SF. The structure of SFCDA is simple and practical, and compared to recently proposed fault diagnosis models, SFCDA achieves higher accuracy in fault diagnosis tasks. As SFCDA can simultaneously address domain adaptation tasks in three different scenarios, reducing the model's development cycle and improving computational efficiency, it is more suitable for handling practical engineering application problems.
(5) Detailed theoretical analysis is conducted, elucidating how SFCDA promotes the fundamental principles of multi-scene learning by improving feature independence and diversity.

The remainder of this paper is organized as follows. Section 2 provides the definition of the problem and presents the SF theory. In Section 3, we describe the proposed SFCDA model. Section 4 gives the data preparation and experimental configuration process. Section 5 presents the experimental results of the proposed method and compares them with other advanced methods. In Section 6, we conduct an in-depth discussion of the model. Finally, the conclusion is presented in Section 7.

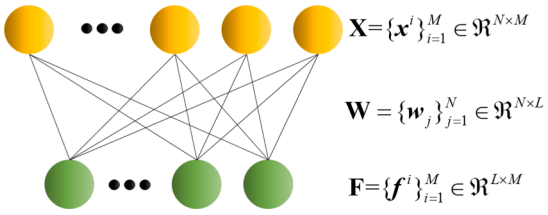


Fig. 2. The SF structure diagram.

2. Preliminary

2.1. Problem formulation

This paper defines fault diagnosis problems toward multi-scene learning as follows. Datasets D^m at different speeds/loads, m is the change of speeds/loads. The source domain $D_S^{m\_s} = \{(\boldsymbol{x}_s^{(i)}, \boldsymbol{y}_s^{(i)})\}_{i=1}^N$ and the target domain $D_T^{m\_t} = \{(\boldsymbol{x}_t^{(i)})\}_{i=1}^M$ correspond to the bearing dataset under different speeds, and $D^m \supseteq \{D_S^{m\_s} \cup D_T^{m\_t}\}$, N,M are the number of samples in the source domain and the target domain, m_s and m_t are the corresponding speed/load in the source domain and the target domain. Besides, the label $\{(\boldsymbol{y}_s^{(i)})\}_{i=1}^N$ of $D_S^{m\_s} = \{(\boldsymbol{x}_s^{(i)})\}_{i=1}^N$ is known, the label of $D_T^{m\_t} = \{(\boldsymbol{x}_t^{(i)})\}_{i=1}^M$ is unknown. Let P, Q be the feature discrepancy of $D_S^{m\_s}$ and $D_T^{m\_t}$. Usually, due to domain shift, $P \neq Q$. The goal of domain adaptation is $P \approx Q$. Given the above analysis, for simplicity, a normal label and two faulty labels are used as illustrations. Fig. 1 presents the three domain adaptation problems to be solved in this paper. Case1: One source domain and one target domain (1S1T). Case2: One source domain and multiple target domains (1SmT). Case3: S&I fault diagnosis.

2.2. Sparse filter (SF)

SF can be understood as a double-layer neural network structure with bias = 0, as shown in Fig. 2. Given a set of input signal $X = \{\boldsymbol{x}^i\}_{i=1}^M \in \Re^{N\times M}$, where N is the length of the $\boldsymbol{x}^i$, M is the number of samples. The feature vector corresponding to the sample is $F = \{f^i\}_{i=1}^M \in \Re^{L\times M}$. L is the length of the feature vector after SF. The forward propagation process of SF is shown in (1).

$$\boldsymbol{F} = \sigma(\boldsymbol{W}^T\boldsymbol{X}) = \sigma\left(\left(\boldsymbol{w}^i\right)^T\boldsymbol{x}^i\right) \tag{1}$$

where $\sigma(t) = \sqrt{t^2 + 10^{-8}}$ is the soft-absolute function, and $\boldsymbol{W} =$

$\{\boldsymbol{w}_j\}_{j=1}^{N} \in \Re^{N\times L}$.

As described in [6], SF obtains the weight matrix of unsupervised learning by learning the sparse distribution of $\boldsymbol{W}$, including population sparsity, lifetime sparsity and high dispersal. The specific operations are as follows:

First, regularize each line of $\boldsymbol{f}_j$ through ℓ_2-normal, as follows:

$$\widetilde{\boldsymbol{f}}_j = \boldsymbol{f}_j \Big/ \left\|\boldsymbol{f}_j\right\|_2 \tag{2}$$

Then, use ℓ_2-normal to regularize the column vector of $\widetilde{\boldsymbol{f}}_j$, so that the features lie on the unit ℓ_2-ball.

$$\widehat{\boldsymbol{f}}^i = \widetilde{\boldsymbol{f}}^i \Big/ \left\|\widetilde{\boldsymbol{f}}^i\right\|_2 \tag{3}$$

Lase, do ℓ_1 penalty for $\widehat{\boldsymbol{f}}^i$. We can obtain the loss function in the back propagation process.

$$\underset{\boldsymbol{W}}{\text{minimize}} \sum_{i-1}^{M} \left\|\widehat{\boldsymbol{f}}^i\right\|_1 = \sum_{i-1}^{M} \left\| \frac{\widetilde{\boldsymbol{f}}^i}{\left\|\widetilde{\boldsymbol{f}}^i\right\|_2} \right\|_1 \tag{4}$$

2.3. Structure alignment loss

Structure alignment loss MMD [40] is used to describe the distance between two random distributions. It makes the two distributions similar by reducing the distribution distance between the source domain distribution P and the target domain distribution Q in a reproducing kernel Hilbert space (RKHS) $\mathscr{F}$. This distance is as shown in Eq. (11).

$$\mathrm{MMD}[\mathscr{F}, P, Q] = \sup_{f\in\mathscr{F}} (\boldsymbol{E}_{\boldsymbol{x}_s}[f(\boldsymbol{x}_s)] - \boldsymbol{E}_{\boldsymbol{x}_t}[f(\boldsymbol{x}_t)]) \tag{5}$$

According to the statistics theory [41], a biased[2] empirical estimate is utilized to simply Eq. (11). And the MMD loss function can be obtained.

$$L_{MMD}[\mathscr{F}, P, Q] = \sup_{f\in\mathscr{F}} \left(\frac{1}{N}\sum_{i=1}^{N} f\left(\boldsymbol{x}_s^{(i)}\right) - \frac{1}{M}\sum_{i=1}^{M} f\left(\boldsymbol{x}_t^{(i)}\right) \right) \tag{6}$$

where, $\mathscr{F}$ is a class of functions, sup (*) is the supremum, $\boldsymbol{E}_*$ is the expectations of the domain distribution, and other parameters are defined as above.

2.4. Batch nuclear norm maximization

High discrimination corresponds to low uncertainty of prediction results, so Shannon entropy [42] is used to measure uncertainty, which is denoted as follows:

$$H(\boldsymbol{Y}) = -\frac{1}{B}\sum_{i=1}^{B}\sum_{j=1}^{C} \boldsymbol{Y}_{i,j}\log\left(\boldsymbol{Y}_{i,j}\right) \tag{7}$$

where, $\sum_{j=1}^{C}\boldsymbol{Y}_{i,j} = 1 \forall i \in 1\ldots B$, $\boldsymbol{Y} \in \Re^{B\times C}$ is the prediction result matrix, B is batch size of model input. Let C be the number of labels. Frobenius-norm of output matrix $\boldsymbol{Y}$ is as follows:

$$\begin{aligned} \|\boldsymbol{Y}\|_F &= \sqrt{\sum_{i=1}^{B}\sum_{j=1}^{C}\left|\boldsymbol{Y}_{i,j}\right|^2} \\ &\leqslant \sqrt{\sum_{i=1}^{B}\left(\sum_{j=1}^{C}\boldsymbol{Y}_{i,j}\right)\cdot\left(\sum_{j=1}^{C}\boldsymbol{Y}_{i,j}\right)} = \sqrt{\sum_{i=1}^{B} 1\cdot 1} = \sqrt{B} \end{aligned} \tag{8}$$

According to [43], the monotonicity of $\|\boldsymbol{Y}\|_F$ and $H(\boldsymbol{Y})$ is opposite. Therefore, maximizing $\|\boldsymbol{Y}\|_F$ is equivalent to minimizing $H(\boldsymbol{Y})$. Thus, we can maximize $\|\boldsymbol{Y}\|_F$ to enhance the discriminability. In [44,45], the Frobenius-norm $\|\boldsymbol{Y}\|_F$ and the nuclear-norm $\|\boldsymbol{Y}\|_\odot$ have the following relationship:

$$\frac{1}{\sqrt{D}}\|\boldsymbol{Y}\|_\odot \leqslant \|\boldsymbol{Y}\|_F \leqslant \|\boldsymbol{Y}\|_\odot \leqslant \sqrt{D}\cdot\|\boldsymbol{Y}\|_F \tag{9}$$

where, $D = \min(B, C)$. And Eq. (9) shows that $\|\boldsymbol{Y}\|_\odot$ tends to be larger with the increase of $\|\boldsymbol{Y}\|_F$. So we can use the nuclear-norm $\|\boldsymbol{Y}\|_\odot$ to enhance prediction discriminability. The nuclear norm can express the diversity of the prediction results of the model [43].

Based on the above analysis, the batch nuclear norm maximization (BNM) loss function is denoted as follows:

$$L_{BNM} = -\frac{1}{B}\|\boldsymbol{Y}\|_\odot \tag{10}$$

where, $\|\boldsymbol{Y}\|_\odot$ is the nuclear-norm of the model prediction matrix.

3. The proposed SFCDA

This paper explores an intelligent fault diagnosis model with a simple structure and high performance, which can deal with the domain shift problem caused by different speed/load conditions, and has high discrimination and multi scenario application ability. The block diagram of bearing fault diagnosis based on SFCDA is shown in Fig. 4, and it is composed of three blocks: data acquisition, pre-training, and fine-tuning.

Step 1: Use the signal collection system to acquire the original vibration signals from the equipment installed with bearings under variable working conditions They are divided into source domain data with fault labels and target domain data without labels.
Step 2: Perform unsupervised training on the target domain data by RSF, and initialize the parameters of the first full connection layer (FC1) with the obtained $\boldsymbol{W}_{tr}$.
Step 3: Build two fully connected neural networks with MSR, named SFCDA. The forward propagation process is as follows: each sample is equally divided into k segments, which are input into FC1 respectively to obtain K one-dimensional feature vectors, which are processed by average and MSR, and finally input into the second full connection layer (FC2) to output the model prediction results.
Step 4: Combine the joint loss function with the labeled data in the source domain to make fine-tuning for SFCDA.
Step 5: Input the test set in target domain into SFCDA to get the diagnostic results.

Therefore, the pseudocode of SFCDA can be summarized as follows:

Algorithm 1 The fault diagnosis process for the proposed SFCDA

Input:

— *The source domain* $D_S^{h-s} = \{(\boldsymbol{x}_s^{(i)}, \boldsymbol{y}_s^{(i)})\}_{i=1}^{N}$ *and the target domain* $D_T^{h-t} = \{(\boldsymbol{x}_t^{(i)})\}_{i=1}^{M}$
— *Initialize lr, Bs, max_epoch,* α, β.
1: ***Pre-training***: *Use RSF to train target domain data and initialize FC*1 *parameters with the obtained weight* $\boldsymbol{W}_{tr}$.
2: ***Fine-tuning***: *Utilize the joint loss function* L_{All} *to make fine-tuning for SFCDA.*
3: ***While*** *epoch* $\leq$ *max_epoch* ***do***
4: ***For*** *epoch* ***do***
5: ***For*** *batch* ***do***
6: *Calculate the cross-entropy loss using* L_C;
7: *Calculate the structure alignment loss by* L_{MMD};
8: *Calculate the discriminability and diversity loss with* L_{BNM};
9: *Obtain the overall objective with* L_{All};
10: *Train and update the parameters of FC*1 *and FC*2;
11: ***end for***
12: ***end for***
Output: *The SFCDA model with trained parameters.*
Test: *Input the test data into SFCDA to get the diagnostic results.*

3.1. Reconstruction sparse filter (RSF) for pre-training

The effectiveness of intelligent fault diagnosis model based on sparse filtering has been proved [35,46]. However, in the cross-domain

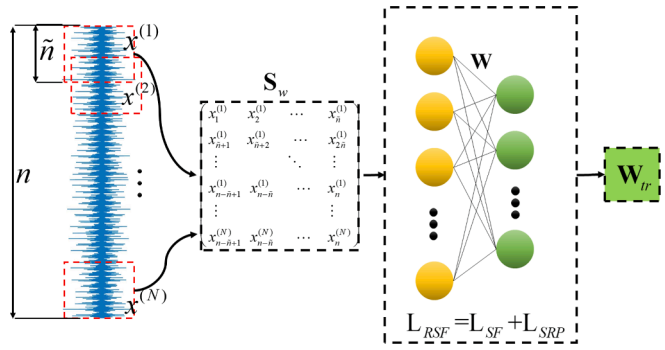


Fig. 3. Pre-training for the FC1.

adaptation field, the performance of sparse filtering has not been fully developed. Therefore, based on RSF [37], this paper adopts RSF to pre-training the end-to-end model SFCDA towards the goal of simplification and unsupervised of the model, and the steps are as follows:

Suppose $\boldsymbol{X}=\{\boldsymbol{x}^{(i)}\}_{i=1}^{N}$ is the fault signal set contains N samples, and the length of each sample is n, sub sample length is $\tilde{n}$, stride is $\triangle n$, overlap rate is $\eta=(\tilde{n}-\triangle n)/\tilde{n}$. Segment each sample to obtain the final sub sample set $\boldsymbol{S}\in\Re^{\tilde{n}\times M_{tr}}$.

$$\boldsymbol{S}=\begin{pmatrix} x_1^{(1)}x_2^{(1)}\cdots x_{\tilde{n}}^{(1)} \\ x_{\tilde{n}+1}^{(1)}x_{n+2}^{(1)}\cdots x_{2\tilde{n}}^{(1)} \\ \vdots\ddots\vdots \\ x_{n-\tilde{n}+1}^{(1)}x_{n-\tilde{n}}^{(1)}\cdots x_n^{(1)} \\ \vdots\vdots \\ x_{n-\tilde{n}+1}^{(N)}x_{n-\tilde{n}}^{(N)}\cdots x_n^{(N)} \end{pmatrix} \tag{11}$$

M_{tr} is the number of sub samples.

$$M_{tr}=N\times\left(\frac{n-\tilde{n}}{\triangle n}+1\right) \tag{12}$$

Then, whitening $\boldsymbol{S}$ [37] to enhance feature learning ability for the pre-training.

$$\boldsymbol{S}_w=\boldsymbol{E}\boldsymbol{D}^{-1/2}\boldsymbol{E}^T\boldsymbol{S} \tag{13}$$

where, $\boldsymbol{D}$ is the eigenvalue of the covariance matrix of $Scov(\boldsymbol{S})$, and $cov(\boldsymbol{S})=\boldsymbol{E}\boldsymbol{D}\boldsymbol{E}^T$, $\boldsymbol{E}$ is an orthogonal matrix.

$\boldsymbol{S}_w$ is the input matrix $\boldsymbol{X}$ of SF. Calculate the eigenvector matrix $\boldsymbol{F}$ corresponding to $\boldsymbol{S}_w$ through (1–3). Based on the loss function (4) of SF, do soft reconstruction penalty (SRP) [47] on the weight matrix $\boldsymbol{W}$ to obtain the RSF loss function [37]

$$L_{RSF}=L_{SF}+L_{SRP}=\sum_{i-1}^{M}\left\|\widehat{\boldsymbol{f}}^i\right\|_1+\frac{\lambda}{4}\sum_{i-1}^{M}\left\|\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{x}^i-\boldsymbol{x}^i\right\|_2^2 \tag{14}$$

In fact, (14) is an unconstrained optimization problem. We can use the limited memory Broyden Fletcher Goldfarb Shanno (L-BFGS) algorithm to solve it. The trained parameter matrix $\boldsymbol{W}_{tr}$ is obtained and used to initialize the FC1 layer parameters of SFCDA. The whole process is shown in Fig. 3.

Next, we will explain why RSF is effective for multi scenario from the perspective of the output features independence.

For the second item of Eq. (14), Eq. (15) can be considered:

$$\begin{aligned} &\left\|\boldsymbol{W}\boldsymbol{W}^T-\boldsymbol{I}_d\right\|_F^2 \\ &=\mathrm{tr}\left(\left(\boldsymbol{W}\boldsymbol{W}^T-\boldsymbol{I}_d\right)^T\left(\boldsymbol{W}\boldsymbol{W}^T-\boldsymbol{I}_d\right)\right) \\ &=\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{W}^T\right)-2\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{W}^T\right)+\mathrm{tr}\left(\boldsymbol{I}_d\right) \\ &=\mathrm{tr}\left(\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W}\right)-2\mathrm{tr}\left(\boldsymbol{W}^T\boldsymbol{W}\right)+\mathrm{tr}\left(\boldsymbol{I}_D\right)+d-D \\ &=\mathrm{tr}\left(\left(\boldsymbol{W}^T\boldsymbol{W}-\boldsymbol{I}_D\right)^T\left(\boldsymbol{W}^T\boldsymbol{W}-\boldsymbol{I}_D\right)\right)+d-D \\ &=\left\|\boldsymbol{W}^T\boldsymbol{W}-\boldsymbol{I}_D\right\|_F^2+d-D \end{aligned} \tag{15}$$

Where D and d are the input and output size of SF, respectively. tr(•) is the sum of diagonal elements. $\mathbf{I}$ is the identity matrix. Discuss the case of Eq. (15) in $D\leqslant d$ and $D>d$.

When $D\leqslant d$, under the constraint of Eq. (14), the second term tended to zero, namely $\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{x}\approx\boldsymbol{x},\boldsymbol{W}^T\boldsymbol{W}\approx\mathbf{I}_D$. Thus $\left\|\boldsymbol{W}^T\boldsymbol{W}-\boldsymbol{I}_d\right\|_F^2$ tended to be

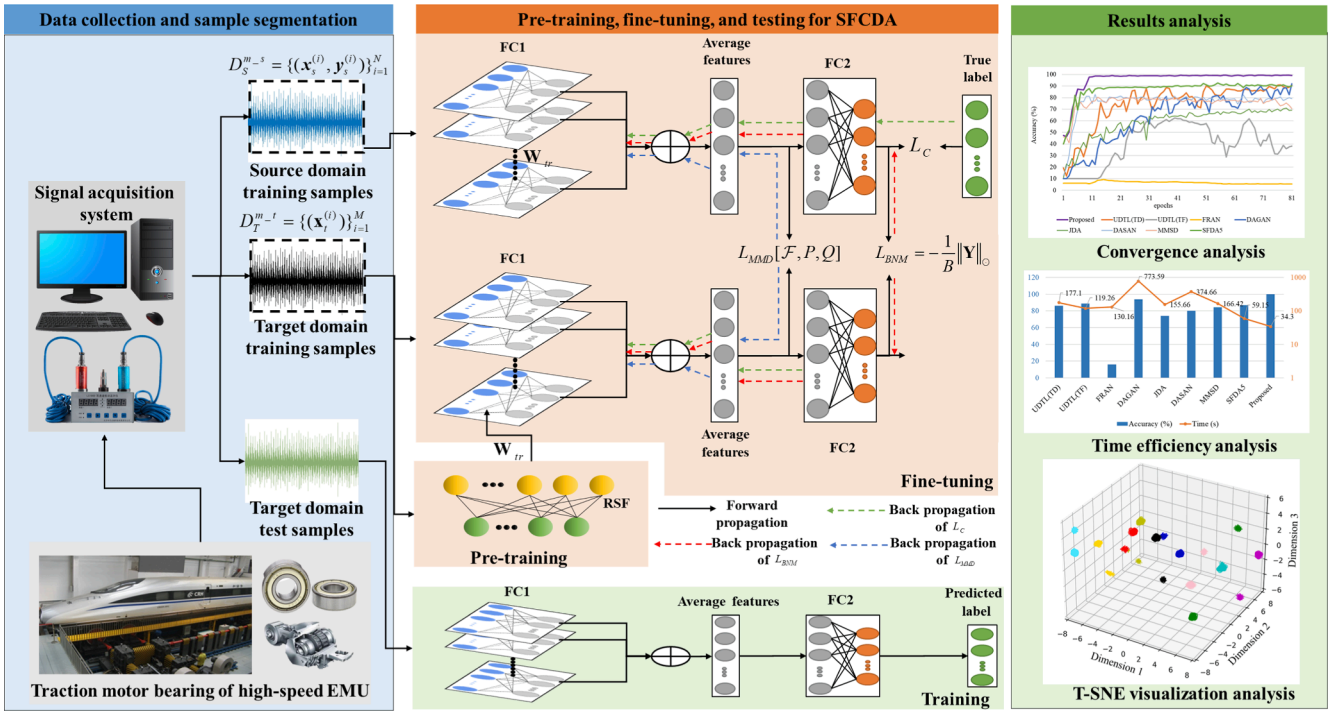


Fig. 4. Structure of the proposed fault diagnosis method.

zero. It is shown that **W** is a linearly independent matrix, and the output features $\boldsymbol{W}^T\boldsymbol{X}$ of SF are independent.

When $D > d$, the data dimension is increased. $\boldsymbol{W}^T\boldsymbol{W} \approx \mathbf{I}_D$ and $\left\|\boldsymbol{W}^T\boldsymbol{W} - \boldsymbol{I}_d\right\|_F^2 < 0$, it is still possible to make **W** linearly independent while Eq. (14) converges. Unless $D \ll d$, obviously, this situation does not hold.

As can be seen from the above analysis, RSF is an unsupervised learning algorithm that does not require additional label information for data. Therefore, no matter how the source and target domain data changes, as iterative algorithms run, there must be $L_{RSF}(\boldsymbol{W}^{t+1}) \leqslant L_{RSF}(\boldsymbol{W}^t)$ and $L_{RSF}(\boldsymbol{W}) \geqslant 0$, t is the number of iterations. It is shown that Eq. (14) can iteratively calculate the minimum value of $L_{RSF}(\boldsymbol{W})$ using the L-BFGS algorithm. Therefore, there is a minimum value that makes **W** linearly independent, which means that the output features of FC1 are as uncorrelated as possible. It is proved that the pre-training has a positive effect on fault diagnosis problems in multiple scenarios.

3.2. Multiple sparse regularization (MSR)

In solving the problem of training and testing data distribution bias arising from speed fluctuations, a study is conducted from the perspective of data sparsity, aiming to increase the number of zero-valued features with the expectation of mitigating distribution bias. This approach is implemented within the framework of transfer learning to tackle cross-device tasks. For the input data matrix **X**, after undergoing MSR processing, a feature matrix **Y** is obtained. The mathematical model for MSR is as follows:

$$Y_1 = X \Big/ \left(f_{msr}(f_{msr}(X))\right) = X_{ij} \Big/ \left(f_{msr}\left(\sqrt{\sum_{k=1}^{m} X_{kj}^2 + \varepsilon}\right)\right) \tag{16}$$

$$Y = Y_1 \Big/ f_{msr}(Y_1) = Y_{ij} \Big/ \sqrt{\sum_{k=1}^{m} Y_{ik}^2 + \varepsilon} \tag{17}$$

where $f_{msr} = \sqrt{\sum_{k=1}^{m} X_{kj}^2 + \varepsilon}$ is the soft absolute value function, with ε taking the value of 1e-8. It is evident that the role of the activation function in Eq. (16) and Eq. (17) respectively performs row and column regularization. According to the explanation in reference [6], applying regularization to each row of data can ensure that the transformation process from data to features for each sample is activated by fewer features, a concept referred to as population sparsity. Regularizing each column of the data matrix enforces high dispersal, thereby preventing the same features from remaining continuously activated. This helps avoid difficulties in feature learning under the distribution shift of time-varying speed data. The feature transformations through equations Eq. (16) and Eq. (17) alleviate the domain shift problem.

3.3. Fine-tuning using the joint loss function

In the fine-tuning stage, the optimization goal of the model consists of three parts, as shown in Fig. 3. including cross-entropy loss, structure alignment loss, and discriminability and diversity loss.

The cross-entropy loss function usually follows the FC. The functional mapping between the softmax value of the output result of the FC and the real label, such as Eq. (18).

$$L_C = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C} y_c^{(n)} \log \frac{\exp\left(\widehat{y}_c^{(n)}\right)}{\sum_{\tilde{c}=1}^{C}\exp\left(\widehat{y}_{\tilde{c}}^{(n)}\right)} \tag{18}$$

where, N equals the sample size, C is the category corresponding to the sample. $y_c^{(n)}$ is a symbolic function. If the true category of sample n is c, take 1, otherwise take 0. $\widehat{y}_c^{(n)}$ is the characteristic value of the n-th sample in FC at the c label.

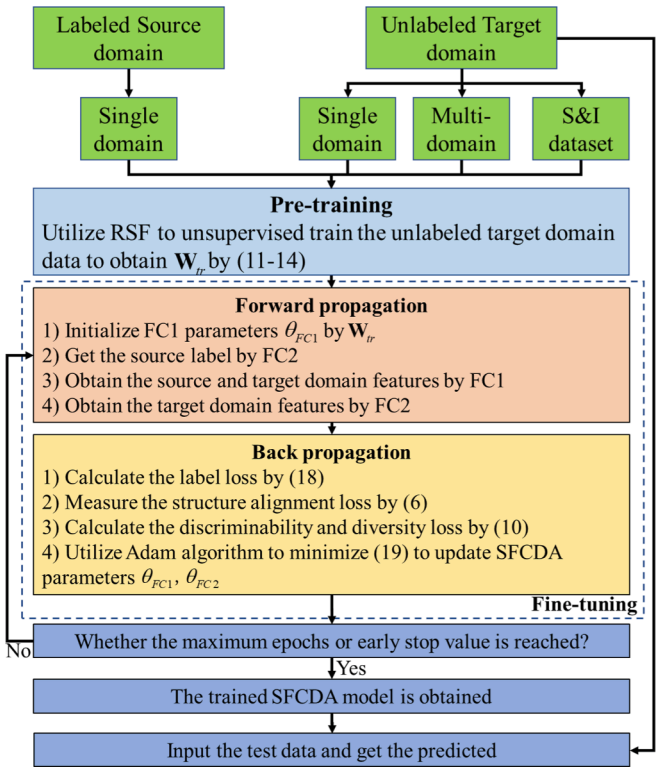


Fig. 5. Flowchart of SFCDA training process with fine-tuning.

Reference [43] proved that discriminant and diversity are equivalent to maximizing the kernel norm of the matrix, and proposed the BNM loss function. Adding BNM to semi-supervised learning and unsupervised domain adaptation can significantly enhance the performance. Inspired by it, we introduce BNM into the bearing domain shift problem, to enhance the discriminant and accuracy of the model.

According to the cross-entropy loss, structure alignment loss, and discriminant and diversity loss, we can obtain the following joint loss function for fine-tuning:

$$L_{All} = L_C + \alpha L_{MMD} + \beta L_{BNM} \tag{19}$$

where, α, β are the weight factors.

As depicted by Eq. (19), the cross-entropy loss leverages the label information. The imposition of kernel norm on the FC2 output outcomes of the target domain data curtails the ambiguity associated with the model's forecasting. Furthermore, the application of Eqs. (8–9) manifests that the kernel norm also regulates the diversity of the FC1 output features of the target domain data, thereby enhancing the model's aptitude to learn from multiple scenarios. Minimizing the loss function L_{All} can maximize the generalization ability.

The optimization goal of SFCDA is to minimize a non-convex loss L_{All} through the utilization of the stochastic gradient descent algorithm, yielding:

$$\theta_{t+1} = \theta_t - \eta \nabla L_{ALL}(\theta_t) \tag{20}$$

where, θ_t represents the model parameters, η signifies the learning rate, and ∇ denotes the gradient operator.

In accordance with deep learning theory, the feasibility of SFCDA can be theoretically substantiated. For non-convex optimization problems, random initialization of FC1 weights may easily lead the model into local minima, a concern exacerbated in the context of domain-adaptive fault diagnosis tasks. Consequently, employing an initialization method based on RSF weights enables the model to identify a more favourable starting point within the parameter space, thereby expediting the

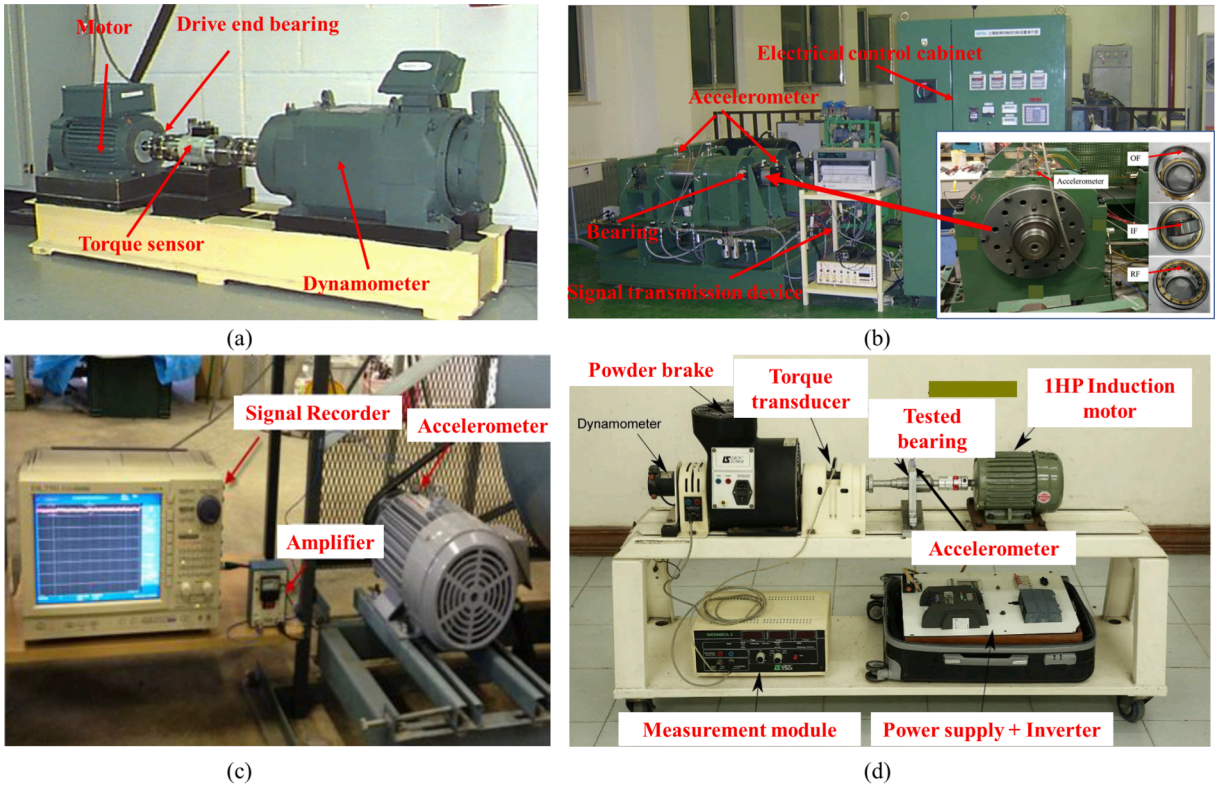


Fig. 6. (a) CWRU bearing test bench, (b) The high-speed EMU traction motor test bench and the faulty bearing.

convergence process. Furthermore, since the initialized weights ensure the model possesses relatively stable weight values from the onset of training, this serves to alleviate issues associated with gradient vanishing and exploding. Lastly, the proposed pre-training method in this paper represents a form of sparse parameter initialization, effectively reducing the likelihood of model overfitting and thereby enhancing the model's adaptability to transfer learning across diverse fault scenarios.

3.4. Model training with fine-tuning

The model is pre trained by RSF, and the SFCDA model is fine-tuned by minimizing the objective loss function Eq. (19) and back propagation algorithm. Adam algorithm [48] is used to update the parameters θ_{FC1}, θ_{FC2} of FC. Until the model meets the maximum epochs condition or the early stop value, the model stops training, as shown in Fig. 5. Consequently, we can obtain the trained cross-domain adaptation model for bearing fault diagnosis.

4. Data preparation and experimental configuration

4.1. Dataset descriptions

Our research is conducted under the auspices of the Power Traction Ministry of Education Engineering Research Center affiliated with the School of Electrical Engineering at Beijing Jiaotong University. This center is mainly engaged in research in the fields of rail transit traction drive systems, bearings and lubrication, fault diagnosis and health management. In collaboration with the Japanese company NTN (https://www.ntn.co.jp/japan/index.html), they have constructed a specialized test rig for high-speed EMU traction motor bearings with world-class capabilities. This test rig can be used for experimental research on the traction motor bearings under different operating conditions, including data and information collection and analysis of the dynamic train drive system, experimental verification of ground platforms, and research work throughout the life cycle, enabling health prediction and status repair of high-speed train traction motor bearings. The private dataset is from the high-speed EMU traction motor bearing test bench of Beijing Jiaotong University (BJTU), as shown in Fig. 6 (a). The testing platform consists of an electrical control cabinet, accelerometer, multiple test bearings, and signal transmission device. The vibration signals of the bearings are collected through sensors and signal transmission systems. The experimental bearings are deep groove ball bearings (6311) and cylindrical roller bearings (NU214) produced by NTN. The dataset consists of four condition types: normal condition (NC), outer race fault (OF), inner race fault (IF) and roller fault (RF). The width of the fault point is 0.1 mm and the depth is 0.15 mm, as shown in Fig. 6 (a). Bearings in different health conditions are installed respectively and driven by the motor, while vibration data is collected by the acceleration sensor under high-speed conditions (2766 rpm, 3480 rpm, 4400 rpm) to simulate the speed condition of the high-speed EMU locomotive. The sensor has a voltage sensitivity of 101.5 mV/g.

To verify the generality of our proposed method, we also utilized three open dataset (Dataset A, Dataset B, and Dataset C) for evaluation. The public dataset is from Case Western Reserve University (CWRU) [49], as shown in Fig. 6 (b). The testing platform consists of a dynamometer, a drive motor, multiple drive end test bearings, and torque sensors. The vibration signals of the bearings are collected through sensors and signal transmission systems. Dataset A is collected at 12 kHz under four loads/speeds, OF, IF and RF contain three fault sizes: 0.18 mm, 0.36 mm and 0.53 mm. There are a total of 10 labels in dataset A. Dataset B is collected at 100 kHz under three loads/speeds, with a total of 4 labels.

Dataset C [50] records the fault diagnosis test data of the centrifugal fan system of Jiangnan University (JNU), as shown in Fig. 6 (c). The test used a Mitsubishi SB-JR induction motor to simulate vibration signals under different operating conditions by changing voltage and applying torque loads, including 600 rpm, 800 rpm, and 1000 rpm. Data were collected using accelerometers and signal recorders to detect different fault states of rolling element bearings, including OF, IF, and RF. By artificially manufacturing faulty bearings and testing them with

Table 1
Dataset description.

Dataset	Condition type	Load	Speed (rpm)	Sampling frequency
CWRU (A)	NC, OF, IF, RF	0hp, 1hp, 2hp, 3hp	1797, 1772, 1750, 1730	12 kHz
BJTU (B)	NC, OF, IF, RF	2800 N, 2600 N 2400 N	2766, 3480, 4400	100 kHz
JNU (C)	NC, OF, IF, RF	/	600, 800, 1000	50kHZ
HUST (D)	NC, OF, IF, RF	0 W, 200 W, 400 W	/	51.2kHZ

different bearing models, a diversified data reference was provided for fault diagnosis.

Dataset D [51] comes from the Hanoi University of Science and Technology (HUST) bearing test platform, as shown in Fig. 6 (d), which includes a multi-stage shaft driven by a 750 W induction motor and a Leroy Somer powder brake to simulate load conditions. Torque sensors and force gauges were used to monitor the motor's load and speed, and accelerometers were installed at three different loads (0 W, 200 W, 400 W) to measure vibration. Faulty bearings were installed on different types of bearing seats, allowing for flexible replacement on multi-stage shafts. The dataset includes four health states, including NC, OF, IF, and RF.

The number of samples of datasets A, B, C, and D under each load/speed is 1000, and the sample length is 1200 points, with corresponding sampling times of 0.012 s and 0.1 s, respectively. The details are described in Table 1. Due to space limitations of the article, only the signals of datasets A and B are visualized here. Figs. 7–10 show the vibration signal time-domain and frequency-domain waveforms of dataset A at 1797 rpm and dataset B at 4400 rpm under four health conditions, respectively.

4.2. Compared approaches

For comparative analysis, several methods similar to our proposed model are employed.

1) **SF**: SF [37] provides a baseline model without fine tuning, which is non end-to-end learning. And to reveal the impact of loss function on the accuracy and the performance of SFCDA, different combinations of loss functions are tested, including SFDA1, SFDA2, SFDA3, SFDA4.
2) **SFDA1:** Based on the proposed SFCDA model, using the domain-adversarial [52] and label loss function for the fine tuning.
3) **SFDA2:** The feature loss [24] and label loss are embedded in RSF with fine tuning.
4) **SFDA3:** The MMD [40] and label loss function are considered in this algorithm.
5) **SFDA4:** The BNM [43] and label loss function are utilized in back propagation

4.3. Diagnosis tasks and implementation details

We set three cases, which are single domain to single domain, single domain to mutil-domain, and imbalanced data.

Specifically, for Dataset A, Case 1: cross validation between different speed conditions, A1-A3 represents the transfer of source domain data under 0hp to target domain data under 1/2/3hp, and so on. Case 2: A13-A15: transfer task of source domain data under 0hp to target domain data under 1, 2hp/1, 3hp/2, 3hp, A16-A24 have the same principle. Case 3: Test the impact of S&I data on the model under Case 2, as shown in Table 2.

For dataset B, Case 1: Cross validation between different speeds, B1-

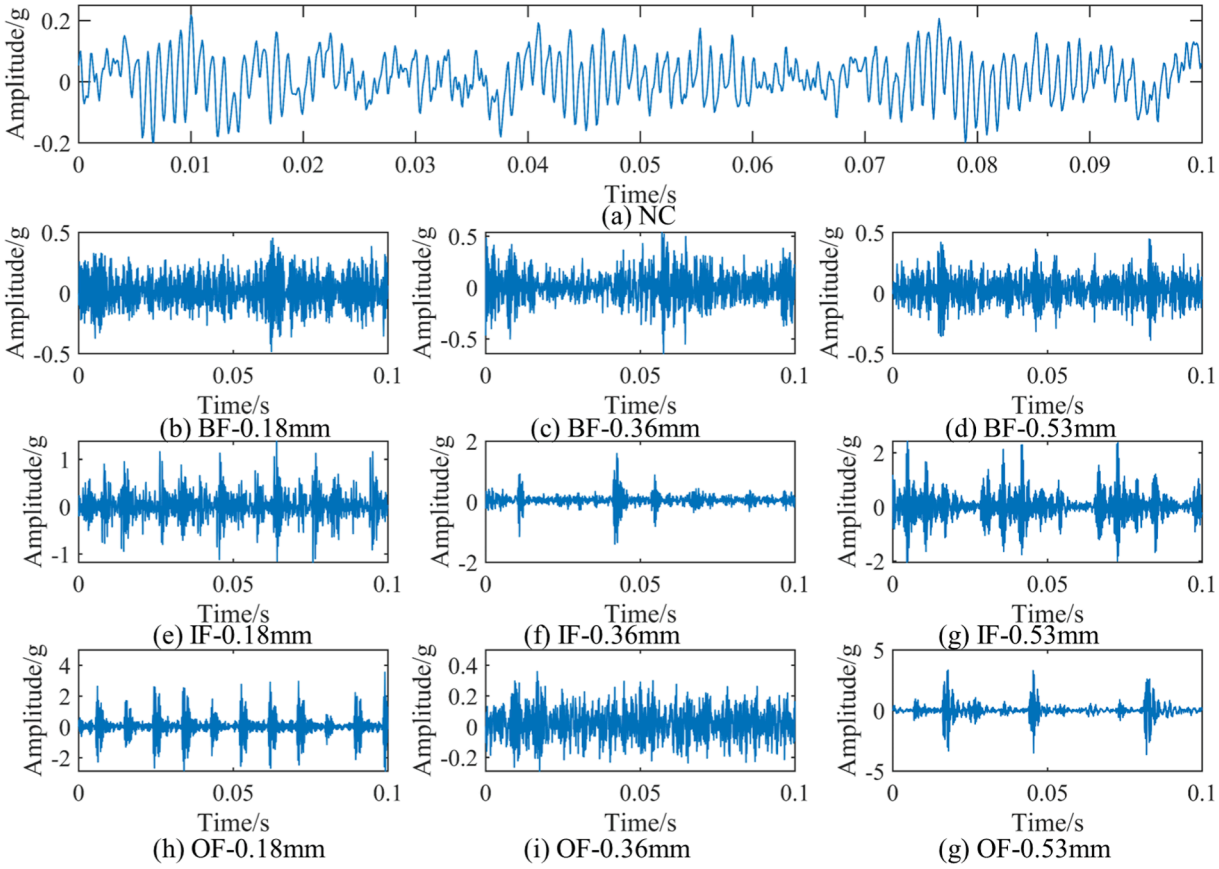


Fig. 7. Time domain graphical representation of CWRU data.

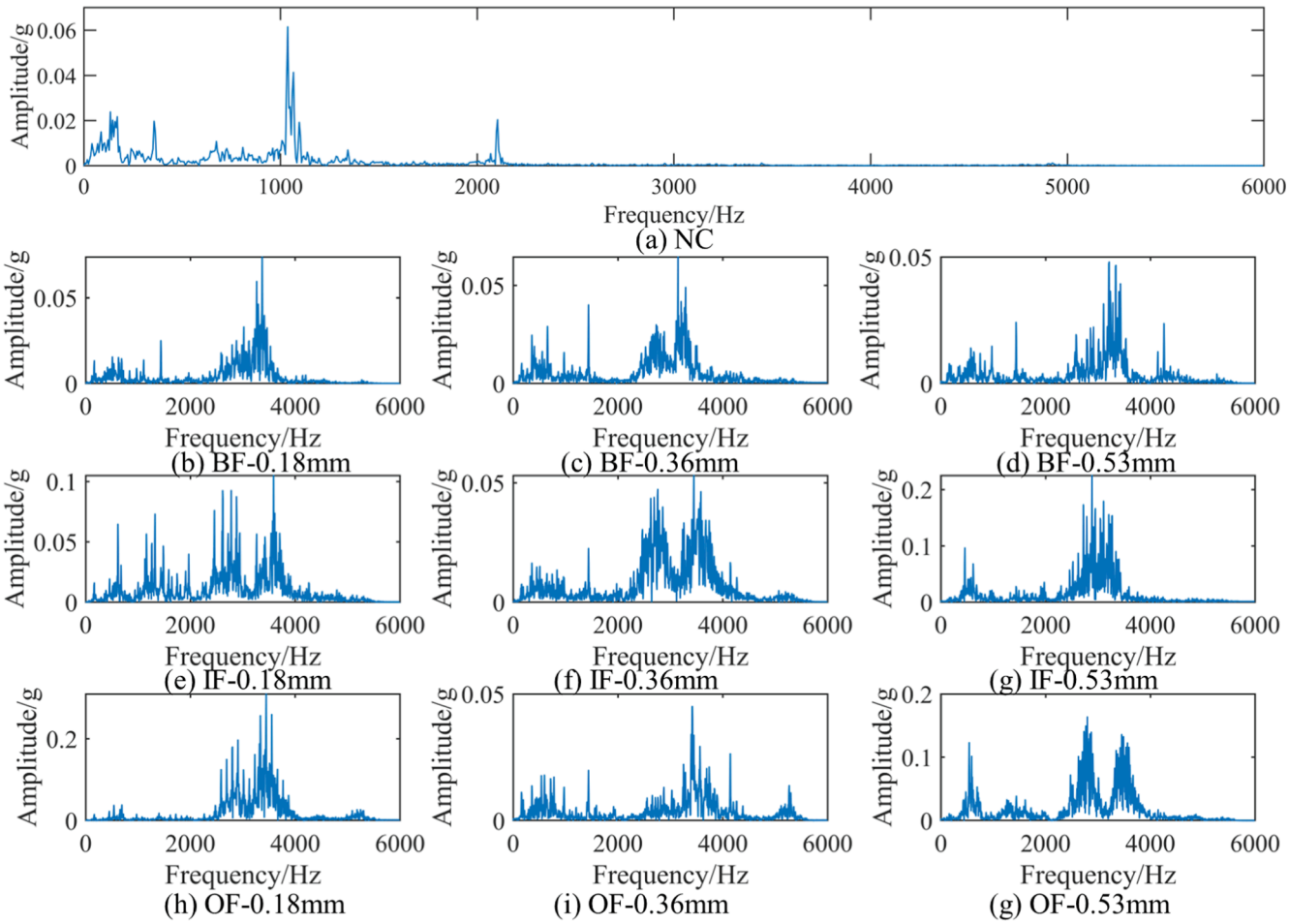


Fig. 8. Frequency-domain illustration for samples from CWRU dataset.

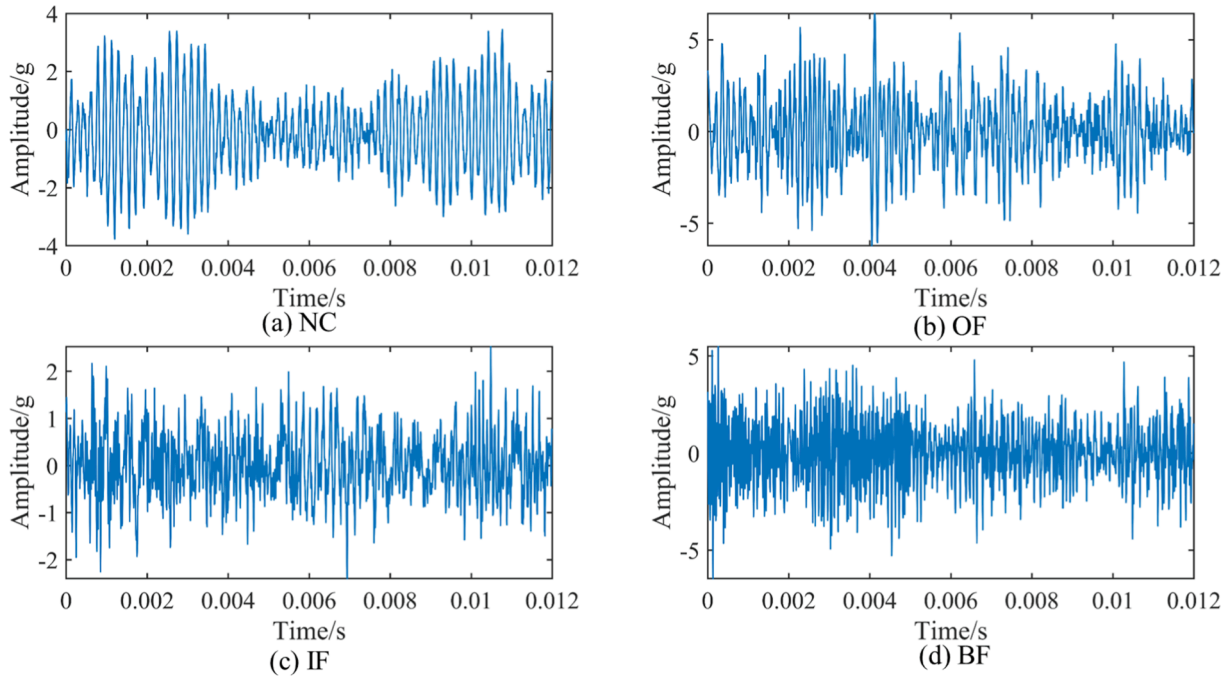


Fig. 9. Time domain graphical representation of BJTU data.

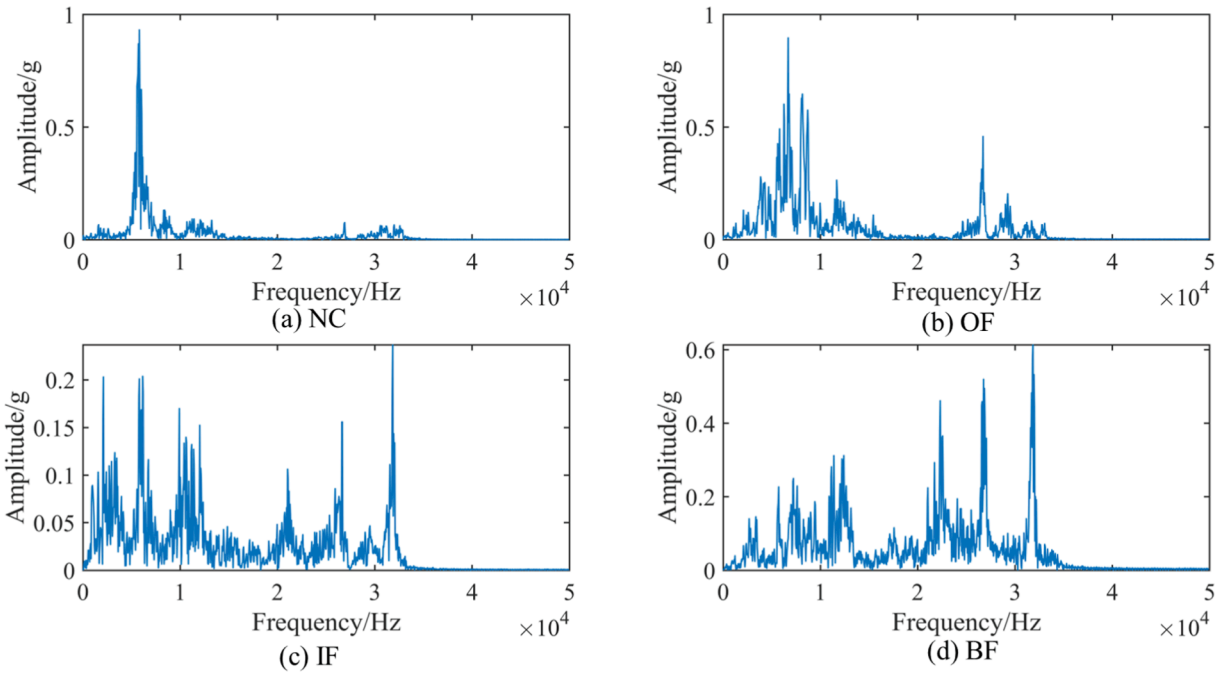


Fig. 10. Frequency-domain illustration for samples from BJTU dataset.

Table 2
Experimental configuration for CWRU.

Case 1				Case 2				Case 3	
Tasks	Source → target domain	Tasks	Source → target domain	Tasks	Source → target domain	Tasks	Source → target domain	Tasks	Source → target domain
A1	0 hp → 1 hp	A7	2 hp → 0 hp	A13	0 hp → 1 hp, 2 hp	A19	2 hp → 0 hp, 1 hp	Rate1	All Case 2 tasks
A2	0 hp → 2 hp	A8	2 hp → 1 hp	A14	0 hp → 2 hp, 3 hp	A20	2 hp → 1 hp, 2 hp	Rate2	All Case 2 tasks
A3	0 hp → 3 hp	A9	2 hp → 3 hp	A15	0 hp → 3 hp, 1 hp	A21	2 hp → 3 hp, 1 hp	Rate3	All Case 2 tasks
A4	1 hp → 0 hp	A10	3 hp → 0 hp	A16	1 hp → 0 hp, 2 hp	A22	3 hp → 0 hp, 1 hp	Rate4	All Case 2 tasks
A5	1 hp → 2 hp	A11	3 hp → 1 hp	A17	1 hp → 2 hp, 3 hp	A23	3 hp → 1 hp, 2 hp	Rate5	All Case 2 tasks
A6	1 hp → 3 hp	A12	3 hp → 2 hp	A18	1 hp → 3 hp, 1 hp	A24	3 hp → 2 hp, 3 hp		

Table 3
Experimental configuration for BJTU.

Case 1		Case 2		Case 3	
Tasks	Source → target domain	Tasks	Source → target domain	Tasks	Source → target domain
B1	2800 N → 2600 N	B7	2800 N → 2600 N, 2400 N	Rate1	All Case 1 tasks
B2	2800 N → 2400 N	B8	2600 N → 2800 N, 2400 N	Rate2	All Case 1 tasks
B3	2600 N → 2800 N	B9	2400 N → 2800 N, 2600 N	Rate3	All Case 1 tasks
B4	2600 N → 2400 N			Rate4	All Case 1 tasks
B5	2400 N → 2800 N			Rate5	All Case 1 tasks
B6	2400 N → 2600 N				

Table 4
Experimental configuration for JNU.

Case 1		Case 2		Case 3	
Tasks	Source → target domain	Tasks	Source → target domain	Tasks	Source → target domain
C1	600 rpm → 800 rpm	C7	600 rpm → 800 rpm, 1000 rpm	Rate1	All Case 1 tasks
C2	600 rpm → 1000 rpm	C8	800 rpm → 600 rpm, 1000 rpm	Rate2	All Case 1 tasks
C3	800 rpm → 600 rpm	C9	1000 rpm → 600 rpm, 800 rpm	Rate3	All Case 1 tasks
C4	800 rpm → 1000 rpm			Rate4	All Case 1 tasks
C5	1000 rpm → 600 rpm			Rate5	All Case 1 tasks
C6	1000 rpm → 800 rpm				

Table 5
Experimental configuration for HUST.

Case 1		Case 2		Case 3	
Tasks	Source → target domain	Tasks	Source → target domain	Tasks	Source → target domain
D1	0 W → 200 W	D7	0 W → 200 W, 400 W	Rate1	All Case 1 tasks
D2	0 W → 400 W	D8	200 W → 0 W, 400 W	Rate2	All Case 1 tasks
D3	200 W → 0 W	D9	400 W → 0 W, 200 W	Rate3	All Case 1 tasks
D4	200 W → 400 W			Rate4	All Case 1 tasks
D5	400 W → 0 W			Rate5	All Case 1 tasks
D6	400 W → 200 W				

Table 6
SFCDA structure settings.

Component	Layers	Input / Output Size
Pre-training	RSF	100/100
Fine-tuning	FC1	100/100
	Average	100/100
	FC2	100/num-class

Table 7
The sample ratio setting of training set and test set.

Training/Testing rate	Case 1	Case 2	Case 3
CWRU	20 %/80 %	20 %/80 %	Rate1-5
BJTU	10 %/90 %	30 %/70 %	Rate1-5
JNU	20 %/80 %	40 %/60 %	Rate1-5
HUST	20 %/80 %	40 %/60 %	Rate1-5

Table 8
The values of the three hyperparameters.

$[\lambda,\alpha,\beta,lr,Bs]$	Case 1	Case 2	Case 3
CWRU	[0.6, 1, 1, 0.6, 64]	[0.6, 1, 1, 0.03, 32]	[0.6, 1, 1, 0.15, 20]
BJTU	[0.6, 1, 0.5, 0.01, 32]	[0.6, 1, 0.1, 0.0023, 30]	[0.6, 1, 0.5, 0.08, 80]
JNU	[0.6, 0.1, 0.1, 0.1, 64]	[0.6, 1, 0.1, 0.1, 64]	[0.6, 0.1, 0.1, 0.05, 32]
HUST	[0.6, 1, 1, 0.1, 64]	[0.6, 1, 0.1, 0.1, 64]	[0.6, 0.1, 0.1, 0.1, 32]

B2 represents the transfer of source domain data under 2800 N to target domain data of 2600 N/2400 N, and so on. Case 2: B7: transfer task of source domain data under 2800 N to 2600 N and 2400 N target domain data, B8-B9 have the same principle. Case 3: Test the impact of S&I data on the model under Case 1, as shown in Table 3. For dataset C and D, three different fault diagnosis scenarios are created by imitating dataset B, as shown in Tables 4 and 5, respectively.

The parameter setting of SFCDA model is shown in Table 6. The input and output size of SRF in the pre-training stage is 100. According to the literature [35], the overlap rate $\eta = 0.8$. In the fine-tuning stage, the input and output size of FC1 is 100, and the output size of FC1 is the number of fault labels, corresponding to num-class.

The hyperparameters of the model are λ, α, β,learning rate lr, and batch size Bs. Where, λ is set to 0.6 based on experience [37], and the early stopping epoch is 50. The optimizer is Adam. The sample ratio settings of the training set and test set are shown in Table 7. The values of the three hyperparameters β, lr, Bs are shown in Table 8.

Table 9
Summary of fault diagnosis accuracy.

Tasks	SF	SFDA1	SFDA2	SFDA3	SFDA4	Proposed
A1	88.50	94.50	94.63	98.75	99.88	100.00
A2	84.00	94.38	96.88	97.88	100.00	100.00
A3	67.50	82.88	82.00	96.38	100.00	100.00
A4	91.00	99.00	98.25	99.88	100.00	100.00
A5	95.50	99.88	99.88	100.00	100.00	100.00
A6	80.50	93.14	89.13	98.50	100.00	100.00
A7	88.00	98.38	98.00	98.38	93.13	100.00
A8	99.50	98.00	99.00	99.38	99.00	99.89
A9	90.00	99.50	99.75	100.00	98.00	100.00
A10	72.00	86.00	86.13	99.38	99.13	99.88
A11	96.50	93.50	91.00	98.88	99.63	99.91
A12	99.50	100.00	100.00	100.00	100.00	100.00
Avg (%)	87.71	94.93	94.55	98.95	99.06	99.97

5. Experimental validation

5.1. Fault diagnosis of CWRU

5.1.1. Analysis of 1S1T

Table 9 and Fig. 11 show the fault diagnosis accuracy of the proposed comparison method under 12 domain adaptive tasks. We find that the accuracy of SF is only 87.71 %, because the non-end-to-end model cannot take into account the global optimal solution. The accuracy of different end-to-end models SFDA1, SFDA2, SFDA3, SFDA4 is more than 94 %, which shows that the end-to-end model has superior fault diagnosis performance. The accuracy of SFDA3 with MMD and SFDA4 with BNM is 98.95 % and 99.06 % respectively, which shows that the accuracy of cross domain fault diagnosis can be improved by introducing domain adaptive and discriminant loss function, and the average diagnosis accuracy can be improved to 99.97 % by combining MMD and BNM.

The analysis of classification performance based on three-dimensional (3D) t-SNE plots provides a method for visualizing and quantitatively evaluating the classification ability of models in high-dimensional data spaces. Taking task A8 as an example, we use t-SNE technology to visualize the input and output characteristics of each layer of neural network, as shown in Fig. 12. The results highlight that after the original data passes through FC1 layer, different fault data are well separated, which proves the feature extraction ability of SF model. After FC2 and joint loss function processing, the data of the same label in the source and target domain are more compact, and there are differences between different label data. It can be seen that after processing by the proposed SFCDA model, the deviation between the source domain and target domain data has been greatly reduced.

5.1.2. Analysis of 1SmT

The performance results of the proposed and comparison method in task A13-A24 are summarized in the 3-D column diagram, as shown in Fig. 13. We can see that the diagnosis accuracy of the comparison method is unstable in Case2, and the diagnosis accuracy is significantly lower than that of Case1. The diagnostic accuracy of the proposed method in task A22 is 93.73 %. One possible reason is that: the domain transfer task from low speed 1730 rpm to high speed 1797 rpm and 1772 rpm is greatly affected by speed fluctuations. On the whole, the comprehensive performance of the proposed method is the best.

5.1.3. Analysis of imbalanced fault diagnosis

The unbalanced ratio is defined as Rate = normal data/ fault data. There are five different ratios, which are Rate1-Rate5. For example, Rate1 (100/80*9) indicates that there are 100 samples of normal data, a total of 9 kinds of fault data, and 80 of each kind of fault data. Based on

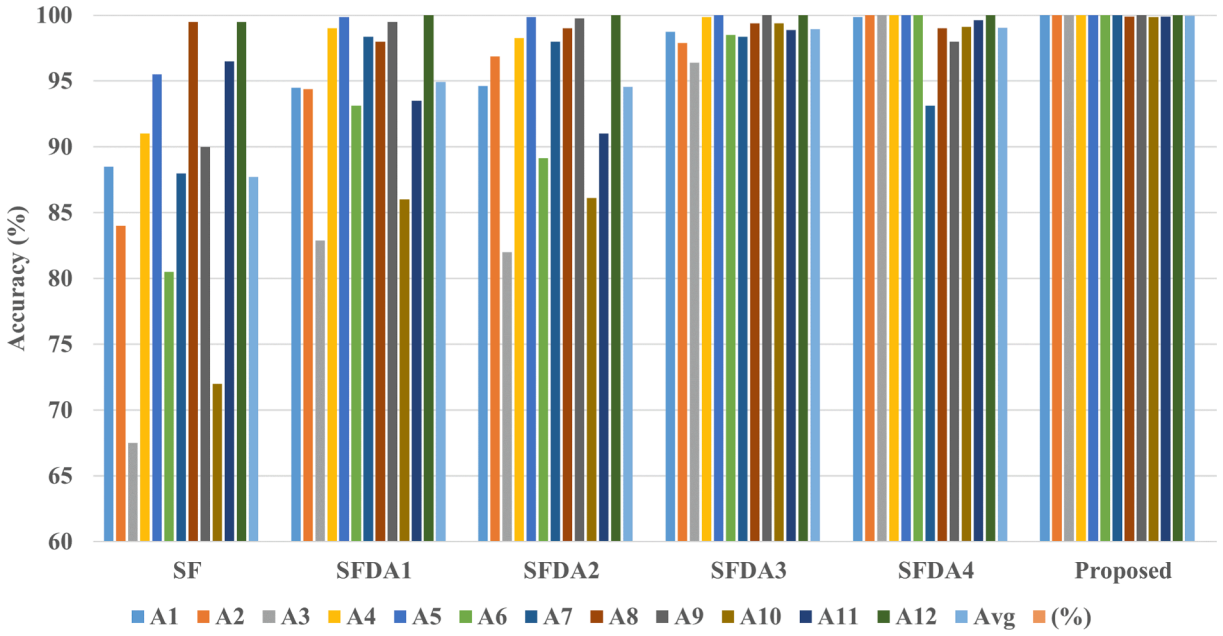


Fig. 11. The histogram of results on Case 1 for CWRU.

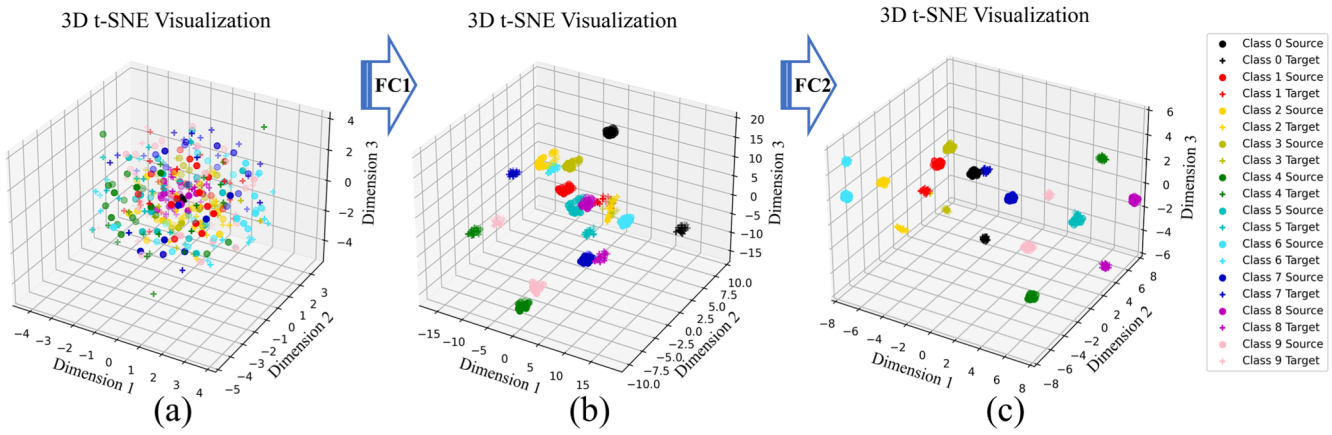


Fig. 12. Visual clustering process of features by t-SNE on task A8. (a) is the raw data in the .
Source domain and target domain. (b) is the feature distribution after FC1. (c) is the feature distribution of source domain and target domain data after FC2

Case2, the fault diagnosis accuracy of different models under different proportions is shown in Table 10. It can be seen that the accuracy of all models decreases with the increase of unbalance. This reflects the interference of category unbalanced data on the model. The SFCDA has the best performance in Rate1-Rate4.

At various imbalance ratios (Rate1 to Rate5), our proposed model demonstrates overall better or competitive fault diagnosis accuracy compared to other methods, including SF and SFDA1 to SFDA4. Specifically, as the imbalance level of the data gradually increases from Rate1 to Rate5, i.e., from lower to higher imbalance, the accuracy of our proposed model remains above 97 % for the initial four ratios (Rate1 to Rate4), indicating excellent stability and robustness of the model, especially in the face of imbalanced data challenges. However, when the ratio is Rate5, the accuracy of the proposed model slightly decreases to 90.95 %. Although it still maintains relatively high performance, it is slightly lower compared to the SFDA4 method at 91.94 %. We believe this might be due to the following possible factors:

(1) Extremely high imbalance level: When the ratio of normal samples to fault samples reaches 10:1, the model may face more challenges in identifying sparse fault samples. When the data imbalance level is too high, the information of minority class samples may be neglected by the model because the model may overly favor the features of the majority class. In such cases, the model may perform poorly in identifying minority classes
(2) Characteristics of the SFCDA model: Although MMD is an effective domain adaptation strategy that can reduce differences between different distributions, when the categories in the source domain are extremely imbalanced, MMD may lead to insufficient feature representation of minority classes in the target domain. This is because MMD focuses on making the overall feature distributions of the two domains as similar as possible, without specifically considering the weights of each individual class in this process.
(3) Dataset characteristics: In extremely imbalanced data situations, even small changes in data or feature distributions may have a significant impact on the model's performance

5.1.4. *Comparison of relate works*

To fully verify the performance of SFCDA, it is compared with the

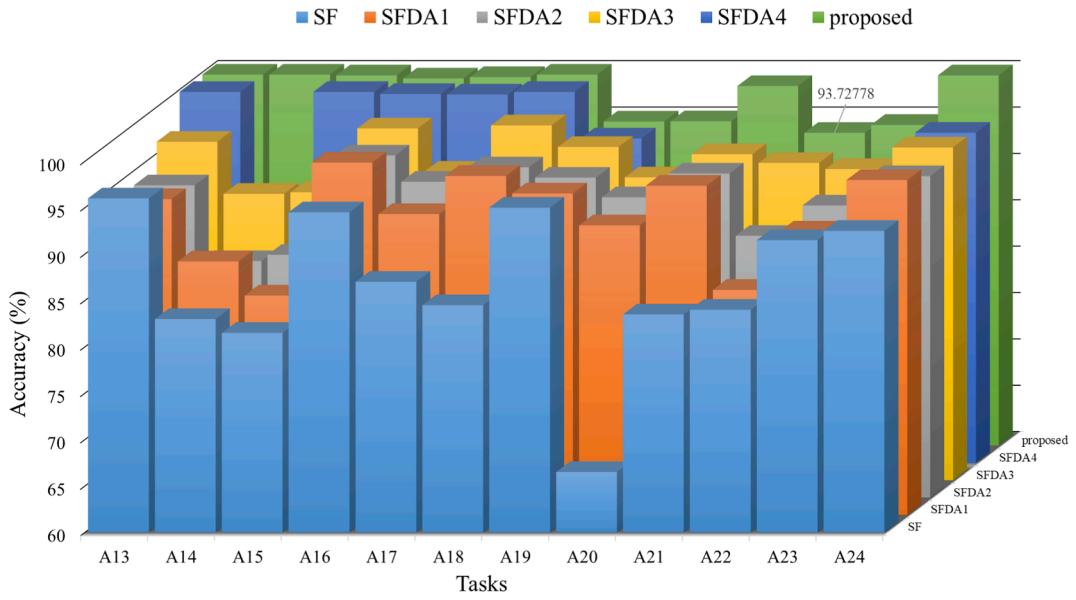


Fig. 13. 3-D column diagram of accuracy corresponding to different methods.

Table 10
Accuracy of S&I fault diagnosis.

Methods	Number of samples (Normal/fault)				
	Rate1	Rate2	Rate3	Rate4	Rate5
	100/80*9	100/60*9	100/40*9	100/20*9	100/10*9
SF	86.77	87.85	86.11	86.58	85.09
SFDA1	92.98	91.45	91.81	91.21	89.78
SFDA2	91.93	91.53	91.77	90.71	90.10
SFDA3	96.71	94.83	94.74	92.93	88.88
SFDA4	96.31	95.88	96.42	93.77	**91.94**
Proposed	**97.73**	**97.93**	**97.33**	**94.74**	90.95

Table 11
Comparison with related work.

Tasks	Models	Training ratio (%) (S/T)	Accuracy ± STDs	Time(s)
Case1	UDTL(TD)	20/20	86.13 ± 6.32	177.10
	UDTL(TF)	20/20	88.84 ± 9.30	119.26
	FRAN	20/20	16.12 ± 9.85	130.16
	DAGCN	20/20	93.97 ± 5.47	773.59
	JDA	20/20	73.91 ± 5.30	155.66
	DASAN	20/20	79.96 ± 6.64	374.66
	MMSD	20/20	84.19 ± 4.55	166.42
	SFDA5	20/20	86.81 ± 28.87	59.15
	Proposed	**20/20**	**99.97 ± 0.05**	**34.30**
Case2	UDTL(TD)	20/20	82.95 ± 2.81	90.65
	UDTL(TF)	20/20	85.01 ± 5.24	62.69
	FRAN	20/20	9.85 ± 2.24	150.86
	DAGCN	20/20	94.17 ± 3.63	406.80
	JDA	20/20	75.62 ± 4.81	197.68
	DASAN	20/20	77.38 ± 4.13	427.07
	MMSD	20/20	80.69 ± 4.52	218.73
	SFDA5	20/20	87.38 ± 5.30	97.40
	Proposed	**20/20**	**96.49 ± 3.17**	**95.98**
Case3	UDTL(TD)	Rate5	63.08 ± 6.50	97.32
	UDTL(TF)	Rate5	67.02 ± 4.55	98.68
	FRAN	Rate5	8.13 ± 1.73	51.32
	DAGCN	Rate5	75.92 ± 5.53	479.44
	JDA	Rate5	70.33 ± 4.49	165.49
	DASAN	Rate5	77.78 ± 2.07	396.25
	MMSD	Rate5	80.02 ± 4.46	178.03
	SFDA5	Rate5	84.92 ± 6.64	19.63
	Proposed	**Rate5**	**90.95 ± 5.36**	**18.30**

Table 12
Summary of fault diagnosis accuracy.

Tasks	SF	SFDA1	SFDA2	SFDA3	SFDA4	Proposed
B1	84.00	79.00	81.75	100.00	100.00	100.00
B2	52.00	57.13	49.00	99.13	72.88	100.00
B3	95.50	82.63	80.75	100.00	100.00	100.00
B4	85.50	81.38	83.38	99.88	100.00	100.00
B5	25.00	51.00	57.50	100.00	61.63	100.00
B6	75.50	79.50	74.50	100.00	100.00	100.00
Avg (%)	69.58	71.77	71.15	99.83	89.08	100.00

advanced cross domain fault diagnosis models in recent years: including UDTL (TD) and UDTL (TF) both adopt the best model in literature [53]: domain adaptive model based on JMMD loss function; FRAN model based on feature alignment [24]; DAGCN model based on graph neural network [23]; JDA [54] is a domain adaptation model that jointly considers marginal distribution and conditional distribution; DASAN [55] is a feature alignment model based on adversarial domain adaptation; MMSD [56] is a novel domain adaptation model that considers mean and variance; SFDA5 is a variant of the SFCDA, where $\lambda = 0.10$ experiments were performed under each test task. The sample size (source domain/target domain), average overall accuracy, standard deviation and average test time of each domain adaptive method are summarized in Table 11.

The performance of FRAN model is the worst, because it has a demand for the amount of data in the training set, and the training samples set in this paper are few. The diagnostic accuracy of the proposed method in Case1 and Case2 is better than that of the comparison method. The diagnostic accuracy of the proposed method in Case3 is 90.95 %, and the average training time of the proposed method is only 18.30 s.

5.2. Fault diagnosis of BJTU

5.2.1. Analysis of 1S1T

Table 12 and Fig. 14 summarize the diagnostic accuracy of Case1 in BJTU data. The proposed method achieves 100 % diagnostic accuracy, and the diagnostic accuracy of B1-B6 is higher than that of the comparison method. Fig. 15 illustrates the t-SNE feature visualization results, which demonstrate that the proposed SFCDA model effectively

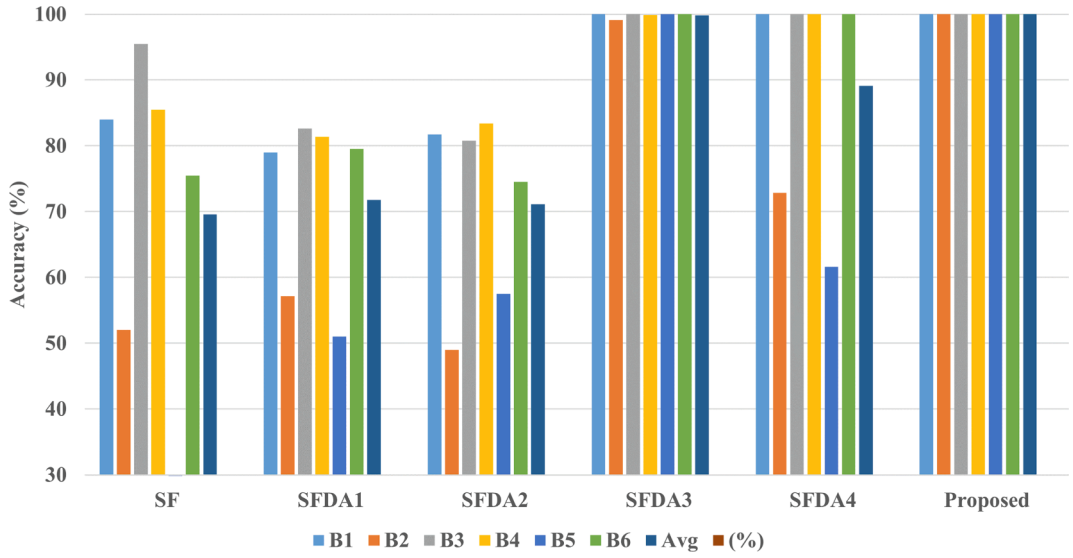


Fig. 14. The histogram of results on Case 1 for BJTU.

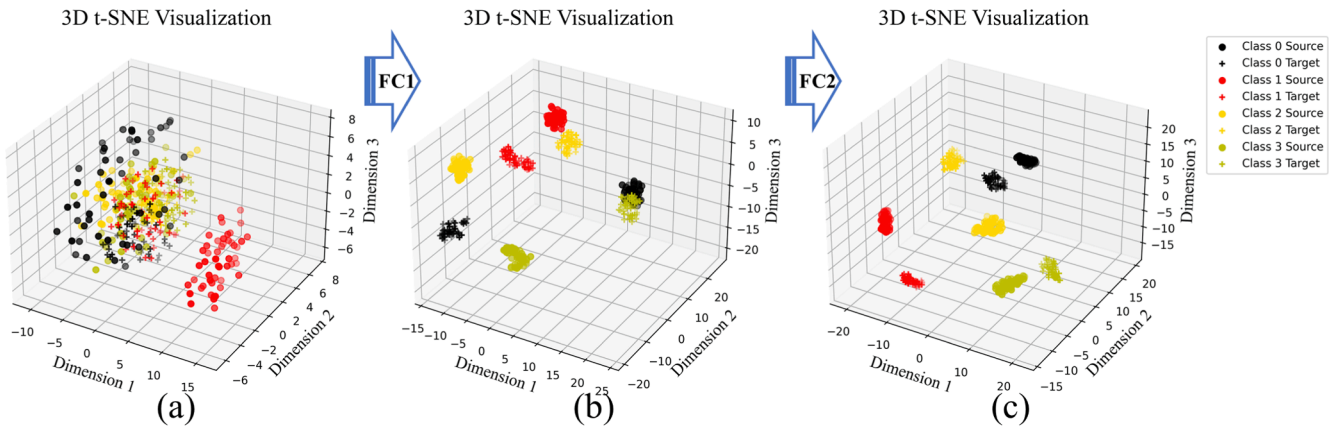


Fig. 15. Visual clustering process of features by t-SNE on task B2. (a) is the raw data in the . Source domain and target domain. (b) is the feature distribution after FC1. (c) is the feature distribution of source domain and target domain data after FC2

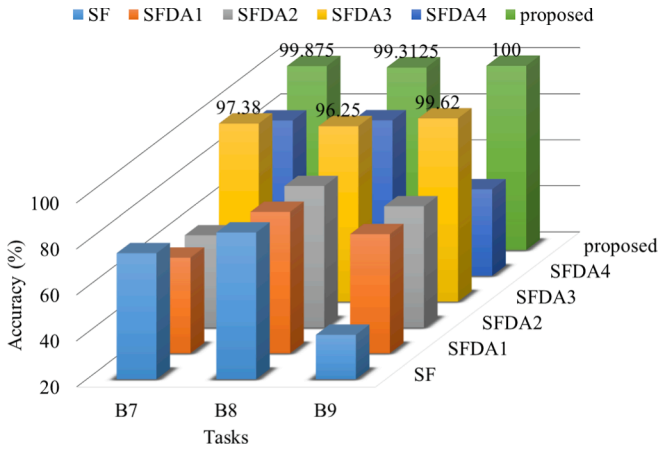


Fig. 16. 3-D column diagram of accuracy corresponding to different methods.

clusters data with similar labels in both the source and target domain data via FC1 and FC2, while effectively separating data with different labels. Comparing Fig. 12 and Fig. 15, the visualization effect of Dataset B is superior to Dataset A, with Dataset B exhibiting a more uniform intra-class distribution and more obvious within-class compactness. Generally, this is because different datasets have different numbers of label categories, and the fewer the number of categories, the better discriminability of model.

Table 13
Accuracy of S&I fault diagnosis.

Methods	Number of samples (Normal/fault)				
	Rate1	Rate2	Rate3	Rate4	Rate5
	125/125*3	125/100*3	125/75*3	125/50*3	125/25*3
SF	74.60	77.10	75.63	71.07	73.03
SFDA1	74.63	74.31	70.00	75.76	68.00
SFDA2	74.60	77.10	75.63	71.07	73.03
SFDA3	99.73	99.63	94.77	76.13	72.63
SFDA4	84.10	59.50	79.83	50.53	73.80
Proposed	**99.93**	**100.00**	**95.60**	**80.23**	**73.90**

5.2.2. *Analysis of 1SmT*

Fig. 16 shows the 3-D column diagram of the diagnostic accuracy of each method in Case2 task. The classification accuracy of the proposed method in the B7-B9 task is 99.88 %, 99.31 % and 100 % respectively. In addition to SFDA3, the classification performance of other methods is

Table 14
Comparison with related work.

Tasks	Models	Training ratio (%) (S/T)	Accuracy (%) ± STDs	Time (s)
Case1	UDTL (TD)	20/20	89.48 ± 17.04	64.69
	UDTL(TF)	20/20	54.27 ± 31.37	16.84
	FRAN	20/20	17.92 ± 0.79	69.90
	DAGCN	20/20	90.98 ± 12.93	276.90
	JDA	20/20	56.92 ± 18.12	73.03
	DASAN	20/20	62.02 ± 13.61	192.48
	MMSD	20/20	62.48 ± 8.63	82.72
	SFDA5	20/20	83.79 ± 24.62	31.93
	Proposed	**10/10**	**99.90 ± 0.17**	**26.12**
Case2	UDTL (TD)	40/40	68.00 ± 8.92	50.54
	UDTL(TF)	40/40	41.35 ± 11.37	14.09
	FRAN	40/40	20.47 ± 2.24	67.80
	DAGCN	40/40	75.96 ± 9.44	192.75
	JDA	40/40	52.19 ± 17.22	60.48
	DASAN	40/40	56.54 ± 15.70	193.21
	MMSD	40/40	56.04 ± 14.15	74.63
	SFDA5	40/40	77.68 ± 2.70	30.47
	Proposed	**30/30**	**90.34 ± 5.13**	**31.33**
Case3	UDTL (TD)	Rate5	71.00 ± 5.17	55.55
	UDTL(FD)	Rate5	54.17 ± 18.90	20.23
	FRAN	Rate5	18.72 ± 1.37	68.28
	DAGCN	Rate5	63.27 ± 12.54	200.77
	JDA	Rate5	47.93 ± 6.35	56.65
	DASAN	Rate5	57.53 ± 22.80	186.37
	MMSD	Rate5	47.53 ± 7.75	73.60
	SFDA5	**Rate5**	**80.65 ± 2.16**	**19.83**
	Proposed	Rate5	73.90 ± 3.14	19.70

Table 15
Summary of fault diagnosis accuracy.

Tasks	SF	SFDA1	SFDA2	SFDA3	SFDA4	Proposed
C1	69.00	90.38	91.38	93.75	97.63	**96.63**
C2	63.50	85.75	87.38	89.25	89.63	**93.00**
C3	63.50	72.88	74.88	72.38	74.13	**77.13**
C4	71.00	87.25	85.50	84.50	86.75	**88.88**
C5	62.00	69.50	**71.88**	71.25	60.63	69.88
C6	78.00	85.63	86.63	84.38	**91.88**	90.38
Avg (%)	67.83	81.90	82.94	82.58	83.44	**85.98**

low. The diagnostic accuracy of SFDA3 was 97.38 %, 96.25 % and 99.62 % respectively. In a word, the performance of the proposed method is the best.

5.2.3. Analysis of imbalanced fault diagnosis

Based on the results obtained in Case 1, the diagnostic accuracy of different methods under different unbalanced ratios is shown in Table 13. For Rate1-Rate5, the diagnosis accuracy of SFCDA is the highest. It is worth mentioning that in the proportion of Rate1-Rate3, the diagnosis accuracy of the proposed method is more than 95.6 %, and the performance of other comparison methods is insufficient.

5.2.4. Comparison of relate works

The comparison method is consistent with the CWRU experiment, and the results are displayed in Table 14. To highlight the advantages of the proposed method, in Case1 and Case2, the volume of data of the proposed method is smaller than that of the comparison method. The results show that the performance of FRAN is also the worst, which is closely related to the small training set data. The diagnosis accuracy of the proposed method is higher than that of the comparison method in the case of insufficient data. In Case3, the average accuracy of the proposed method is 73.9 % and SFDA5 is 80.65 %, which indicates that the proposed method has room for parameter adjustment in unbalanced

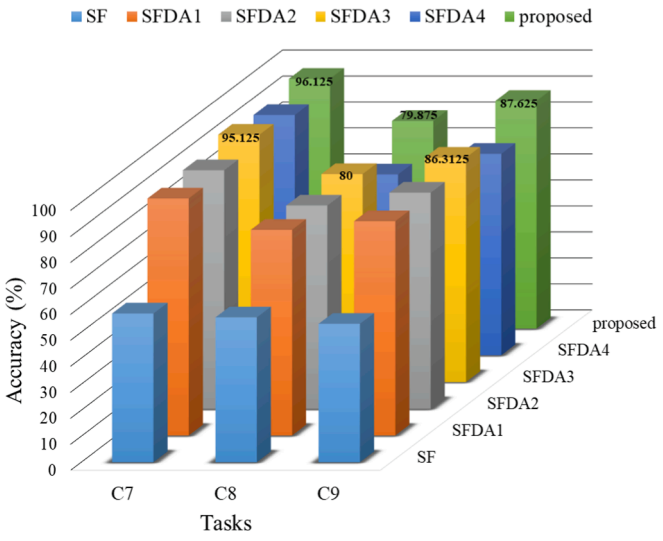


Fig. 17. 3-D column diagram of accuracy corresponding to different methods.

Table 16
Accuracy of S&I fault diagnosis.

Methods	Number of samples (Normal/fault)				
	Rate1	Rate2	Rate3	Rate4	Rate5
	125/125*3	125/100*3	125/75*3	125/50*3	125/25*3
SF	64.97	64.94	63.48	66.85	70.75
SFDA1	89.70	89.07	86.03	86.07	82.23
SFDA2	88.13	88.70	87.00	85.23	80.37
SFDA3	91.03	89.47	86.63	81.70	74.77
SFDA4	91.27	87.03	80.20	73.57	74.13
Proposed	**94.87**	**93.23**	**91.17**	**87.87**	**83.00**

Table 17
Comparison with related work.

Tasks	Models	Training ratio (%) (S/T)	Accuracy (%) ± STDs	Time (s)
Case1	UDTL (TD)	20/20	80.81 ± 10.51	62.70
	UDTL(TF)	20/20	82.10 ± 6.54	60.94
	FRAN	20/20	16.70 ± 0.69	68.80
	DAGCN	20/20	73.83 ± 11.28	310.50
	JDA	20/20	49.23 ± 4.91	65.58
	DASAN	20/20	56.44 ± 11.83	189.86
	MMSD	20/20	51.00 ± 8.35	79.70
	SFDA5	20/20	74.79 ± 9.63	18.96
	Proposed	**20/20**	**85.98 ± 9.38**	**18.33**
Case2	UDTL (TD)	40/40	82.67 ± 10.83	38.83
	UDTL(TF)	40/40	82.04 ± 9.96	41.33
	FRAN	40/40	17.54 ± 0.48	67.19
	DAGCN	40/40	77.81 ± 8.89	216.16
	JDA	40/40	50.27 ± 5.80	61.68
	DASAN	40/40	61.65 ± 6.74	185.28
	MMSD	40/40	57.33 ± 1.06	72.69
	SFDA5	40/40	72.08 ± 9.50	19.38
	Proposed	**40/40**	**87.88 ± 6.64**	**19.88**
Case3	UDTL (TD)	Rate5	79.30 ± 6.95	51.35
	UDTL(FD)	Rate5	81.87 ± 6.27	49.49
	FRAN	Rate5	17.26 ± 0.80	69.25
	DAGCN	Rate5	63.80 ± 8.08	252.69
	JDA	Rate5	47.53 ± 5.09	60.74
	DASAN	Rate5	50.50 ± 8.41	182.88
	MMSD	Rate5	50.57 ± 6.47	70.96
	SFDA5	Rate5	69.90 ± 9.57	29.54
	Proposed	**Rate5**	**83.00 ± 9.38**	**27.49**

Table 18
Summary of fault diagnosis accuracy.

Tasks	SF	SFDA1	SFDA2	SFDA3	SFDA4	Proposed
D1	99.00	92.50	99.50	100.00	100.00	100.00
D2	79.50	83.25	98.25	99.88	100.00	100.00
D3	99.00	88.75	100.00	100.00	100.00	100.00
D4	99.50	99.50	100.00	100.00	100.00	100.00
D5	77.00	77.13	100.00	100.00	100.00	100.00
D6	91.50	97.25	100.00	100.00	100.00	100.00
Avg (%)	90.92	89.73	99.63	99.98	100.00	100.00

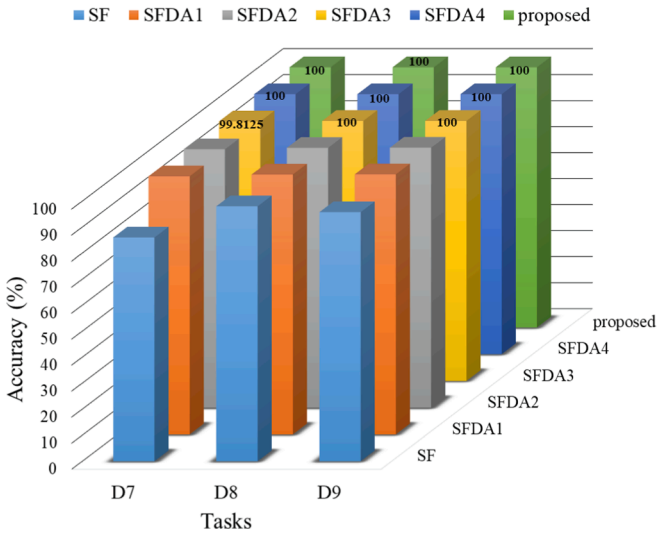


Fig. 18. 3-D column diagram of accuracy corresponding to different methods.

tasks. In a word, the SFCDA has the best performance and the least training time among the three cases.

5.3. Fault diagnosis of JNU

5.3.1. Diagnostic performance of different models in multiple scenarios

To save space, we present the multi-scenario fault diagnosis tasks under the JNU dataset in one section. Table 15, Fig. 17, and Table 16 respectively summarize the diagnostic accuracy of the proposed method and the comparison methods in Case 1, Case 2, and Case 3.

A thorough comparison and discussion of the experimental results are conducted. In Case 1, we observed that under different experimental tasks (C1 to C6), the proposed method achieved an average diagnostic rate of 85.98 %. In the fault diagnosis scenarios of Case 2, we found variations in the performance of each method under different diagnostic tasks (C7 to C9). Overall, the proposed method still demonstrated excellent accuracy. In Case 3, we observed the recognition accuracy of different methods under different fault rates. In a word, considering all imbalanced sample ratios, the proposed method exhibited strong robustness and reliability for imbalanced sample tasks.

5.3.2. Comparison of relate works

The summary table of diagnostic effectiveness of domain adaptive fault diagnosis methods in three fault diagnosis scenarios is shown in Table 17. Through analysis of the experimental results, the following conclusions can be drawn:

In the first scenario, comparing the performance of each model at a 20 % training rate, it can be observed that the proposed method significantly outperforms others in terms of accuracy and training time, with an accuracy of 85.98 %±9.38 and a training time of only 18.33 s. In Case2 scenario, the proposed method still maintains a high accuracy of 87.88 %±6.64, with a relatively short training time of 19.88 s. Additionally, significant performance differences among various advanced models were observed, with the DAGCN model having the longest training time. In Case3 scenario, we focused on performance comparison under specific imbalanced data ratios (Rate5), and the results showed that the proposed method still exhibited the highest accuracy (83.00 % ±9.38) and was also the most efficient in terms of training time (27.49 s).

In conclusion, the proposed method demonstrates excellent performance and efficient training speed in multiple fault diagnosis scenarios, indicating its great potential for practical applications.

Table 19
Accuracy of S&I fault diagnosis.

Methods	Number of samples (Normal/fault)				
	Rate1	Rate2	Rate3	Rate4	Rate5
	125/125*3	125/100*3	125/75*3	125/50*3	125/25*3
SF	93.10	94.51	92.95	93.70	93.83
SFDA1	99.73	99.93	99.90	99.87	99.37
SFDA2	99.90	99.90	99.80	99.80	99.77
SFDA3	100.00	99.93	99.93	99.97	98.77
SFDA4	100.00	100.00	100.00	100.00	99.83
Proposed	**100.00**	**100.00**	**100.00**	**100.00**	**99.97**

Table 20
Comparison with related work.

Tasks	Models	Training ratio (%) (S/T)	Accuracy (%) ± STDs	Time (s)
Case1	UDTL (TD)	20/20	99.81 ± 0.42	54.82
	UDTL(TF)	20/20	99.52 ± 0.68	50.88
	FRAN	20/20	23.63 ± 1.88	67.26
	DAGCN	20/20	99.65 ± 0.63	263.54
	JDA	20/20	91.23 ± 4.33	76.17
	DASAN	20/20	96.71 ± 4.64	186.73
	MMSD	20/20	97.08 ± 1.58	77.59
	SFDA5	20/20	89.69 ± 8.49	17.41
	Proposed	**20/20**	**100.00 ± 0.00**	**18.44**
Case2	UDTL (TD)	40/40	99.96 ± 0.06	37.20
	UDTL(TF)	40/40	100.00 ± 0.00	36.96
	FRAN	40/40	27.39 ± 1.27	66.08
	DAGCN	40/40	99.94 ± 0.00	221.45
	JDA	40/40	96.92 ± 2.33	62.80
	DASAN	40/40	99.98 ± 0.03	189.68
	MMSD	40/40	98.58 ± 0.72	70.55
	SFDA5	40/40	94.73 ± 2.30	18.89
	Proposed	**40/40**	**100.00 ± 0.00**	**14.15**
Case3	UDTL (TD)	Rate5	99.90 ± 0.15	51.32
	UDTL(FD)	Rate5	99.90 ± 0.15	47.83
	FRAN	Rate5	24.69 ± 1.08	68.16
	DAGCN	Rate5	90.77 ± 6.25	242.21
	JDA	Rate5	86.63 ± 7.69	71.42
	DASAN	Rate5	94.10 ± 2.92	177.72
	MMSD	Rate5	95.13 ± 3.39	70.60
	SFDA5	Rate5	86.73 ± 7.76	29.94
	Proposed	**Rate5**	**99.97 ± 0.07**	**22.34**

5.4. Fault diagnosis of HUST

5.4.1. Diagnostic performance of different models in multiple scenarios

Table 18, Fig. 18, and Table 19 respectively summarize the diagnostic accuracy of the proposed method and comparison methods in Case 1, Case 2, and Case 3. This study conducted a detailed comparison of the performance of the proposed method and comparison methods in multiple fault diagnosis scenarios. The results show that the proposed method achieved outstanding performance in tasks D1 to D6, with 100 % accuracy in all tasks. In comparison, SF, SFDA1, SFDA2, and SFDA3

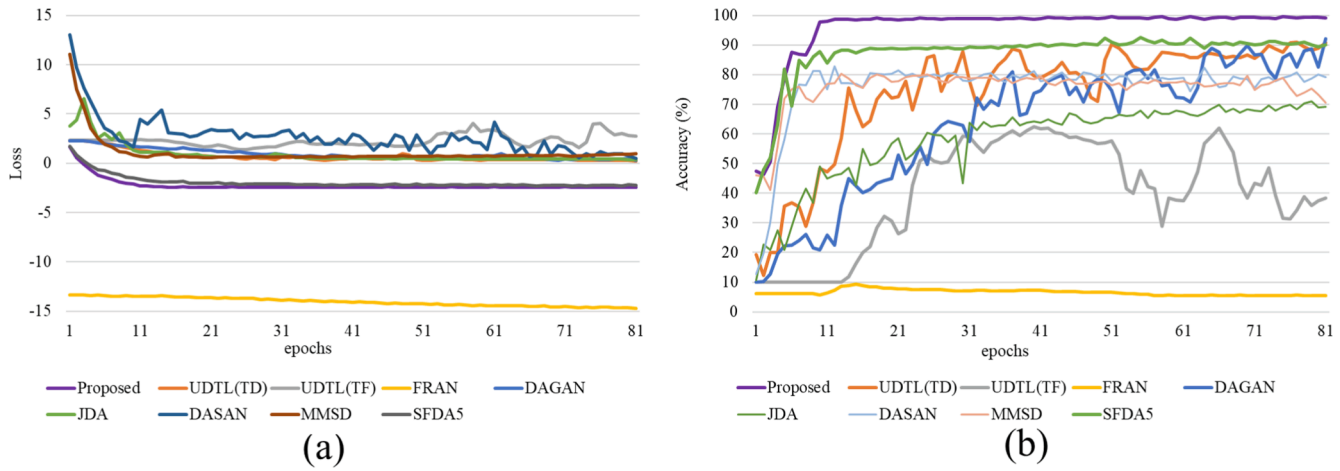


Fig. 19. Convergence curve of task A10. (a) Loss curve. (b) Accuracy curve.

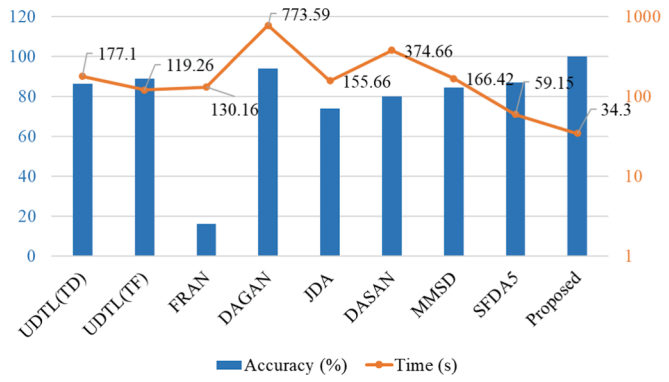


Fig. 20. The accuracy and time of different methods under task A10.

were slightly inferior in specific tasks. Further observation of tasks D7 to D9 revealed that all methods achieved close to or reached 100 % accuracy, with the proposed method being notably significant. Additionally, under various sample imbalance ratios, the proposed method maintained high accuracy, demonstrating its robustness under different conditions. In summary, the proposed method exhibits superior performance in fault diagnosis tasks across multiple scenarios.

5.4.2. *Comparison of relate works*

In this section, we compared the diagnostic performance of advanced models in multi-scenario fault diagnosis tasks. The results are summarized in Table 20.

Across the three different scenarios, we observed that the proposed method consistently achieved the best performance in all cases. Firstly, in Case 1, the proposed method outperformed other models with an accuracy of 100.00 %±0.00 and relatively short training time (18.44 s). In Case 2, the proposed method maintained a perfect accuracy of 100.00 %±0.00 and exhibited relatively high efficiency in training time (14.15 s), further confirming its robustness and reliability across different fault scenarios. In Case 3, even under the imbalanced data ratio of Rate5, the proposed method maintained extremely high accuracy (99.97 %±0.07) and the lowest training time (22.34 s), which further demonstrates the robustness and reliability of the method.

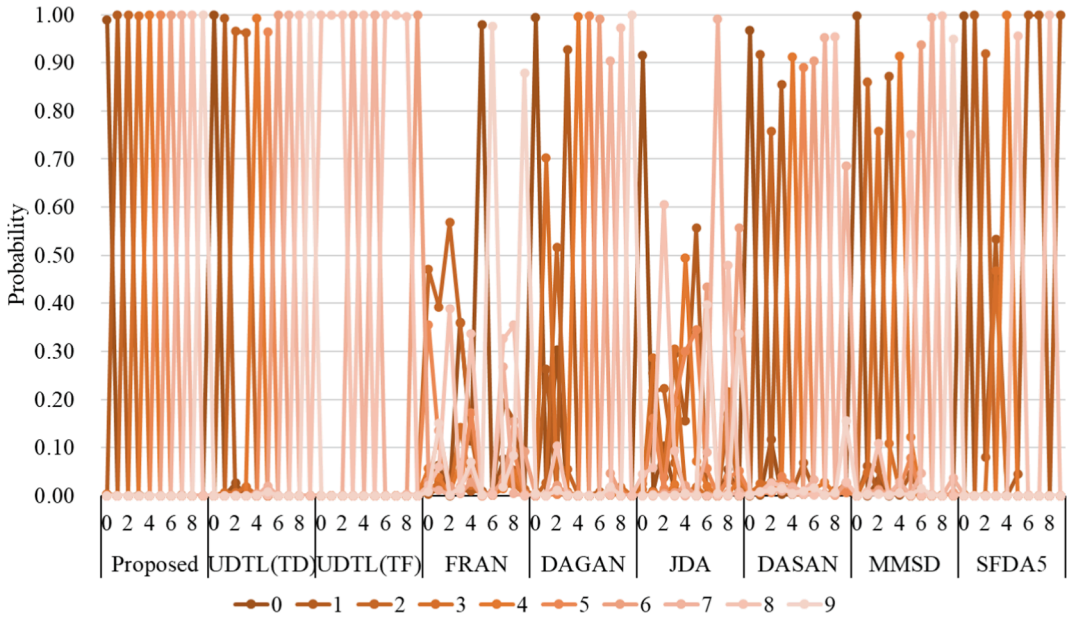


Fig. 21. The softmax values of different methods under task A10.

Table 21
Algorithm complexity.

Methods	TC(FLOPs)	SC(Bytes)
UDTL(TD)	682,232,832	232,906
UDTL(TF)	333,773,824	232,906
FRAN	788,099,584	19,221,618
DAGCN	676,575,744	315,532
JDA	17,322,770,432	1,997,018
DASAN	227,658,300	1,491,655
MMSD	60,906,240	266,970
SFDA5	11,552,000	11,011
Proposed	11,552,000	11,011

6. Model discussion

6.1. Convergence analysis and time efficiency

Illustrated by Case1 in the CWRU dataset, Fig. 19 presents the convergence curve of accuracy and loss of various methods under the A10 task. It is evident that in the process of model training, the loss of the proposed method decreased rapidly, and the accuracy of 100 % was reached at the 11th epoch. Additionally, Fig. 20 shows the comparison of diagnostic accuracy and calculation time of different methods. These results shown that SFCDA has a faster convergence rate, higher accuracy, and reduces the calculated time in a limited epoch.

6.2. Discrimination analysis

The discrimination of model prediction results is an index to evaluate whether a diagnostic model is firm or not, and to some extent reflects the prediction performance and robustness of the model. Taking Case1 in the CWRU data as an example, Fig. 21 shows the softmax probability values output by the proposed method and the comparison method on the test set for the data with labels 0–9. It can be seen that the prediction probability of the proposed method for each label is higher, close to 1.0, which also proves that the loss function L_{BNM} plays a positive role.

6.3. Complexity analysis of SFCDA

In substantiating the simplicity of our proposed methodology, we draw upon the insights presented in [23] and conduct a comprehensive analysis employing temporal complexity (TC) and spatial complexity (SC). The utilization of Big-O notation allows us to articulate the asymptotic upper bounds governing the order of magnitude of the function. The formulas for TC and SC calculations in a single-layer neural network are detailed as follows:

$$TC_{FC} = O(D_{out} \cdot w \cdot h \cdot D_{in}) \tag{19}$$

$$SC_{SF} = O(D_{out} \cdot D_{in}) \tag{20}$$

here, D_{out} and D_{in} signifie the characteristic dimension,w and h denote the width and height of the input feature. Consequently, the TC and SC computation expressions for our proposed methodology are succinctly stated as:

$$TC_{Proposed} = TC_{RSF} + TC_{FC1} + TC_{Average} + TC_{FC2} \tag{21}$$

$$SC_{Proposed} = SC_{RSF} + SC_{FC1} + SC_{Average} + SC_{FC2} \tag{22}$$

As per the aforementioned formulations, the TC and SC values for SFCDA are computed as 11,552,000 FLOPs and 11,011 Bytes, respectively. The TC and SC values for each method are systematically derived using Big-O notation and are presented in Table 21. Remarkably, SFCDA consistently demonstrates the most minimized TC and SC values. Compared with the most complex FRAN, TC and SC are reduced by about 68 times and 1745 times respectively. It is fully proved that the proposed model is the simplest and most efficient.

6.4. Hyperparameter sensitivity analysis

In this section, the hyperparameter sensitivity analysis on the

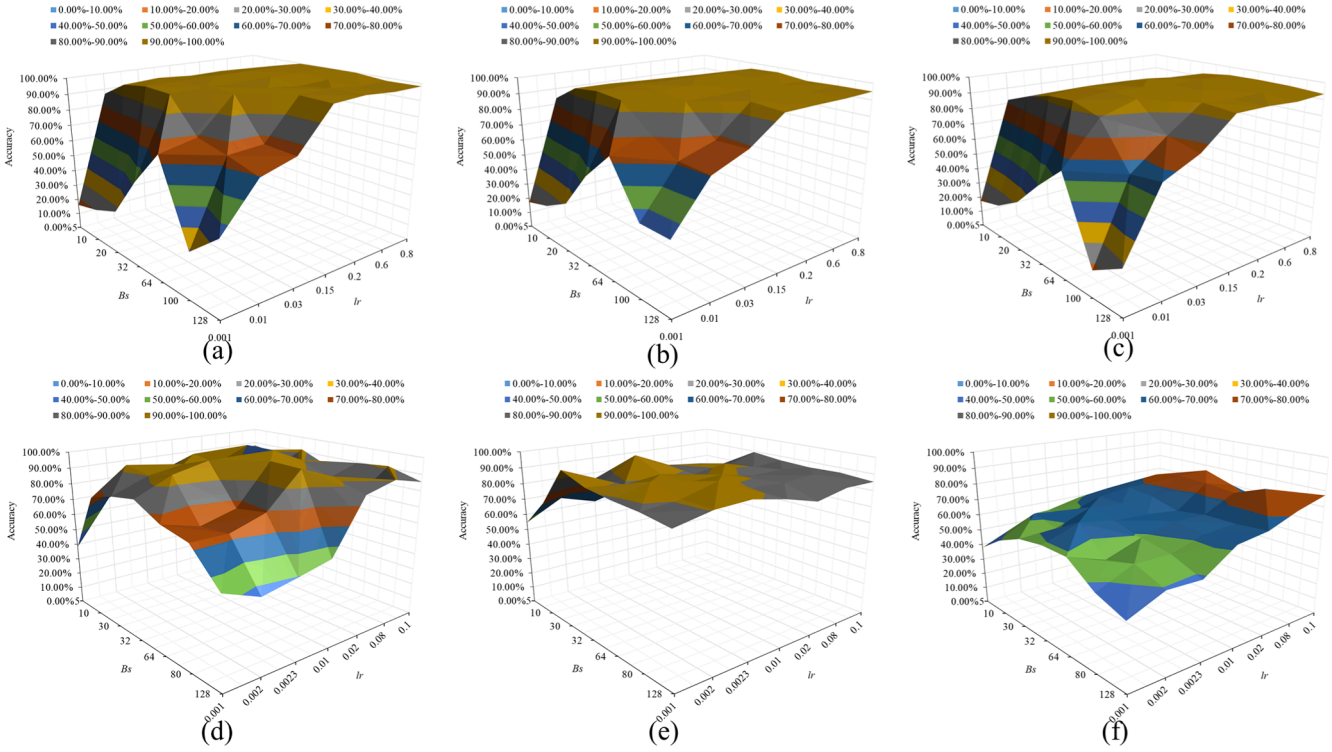


Fig. 22. The results of hyperparametric sensitivity analysis for *lr* and *Bs*. (a) CWRU dataset in Case1. (b) CWRU dataset in Case2. (c) CWRU dataset in Case3. (d) BJTU dataset in Case1. (e) BJTU dataset in Case2. (f) BJTU dataset in Case3.

Table 22
The loss function combination.

Combo	α	β	Combo	α	β	Combo	α	β
1	0.0	0.0	**13**	0.5	0.0	**25**	2.0	0.0
2	0.0	0.1	**14**	0.5	0.1	**26**	2.0	0.1
3	0.0	0.5	**15**	0.5	0.5	**27**	2.0	0.5
4	0.0	1.0	**16**	0.5	1.0	**28**	2.0	1.0
5	0.0	2.0	**17**	0.5	2.0	**29**	2.0	2.0
6	0.0	5.0	**18**	0.5	5.0	**30**	2.0	5.0
7	0.1	0.0	**19**	1.0	0.0	**31**	5.0	0.0
8	0.1	0.1	**20**	1.0	0.1	**32**	5.0	0.1
9	0.1	0.5	**21**	1.0	0.5	**33**	5.0	0.5
10	0.1	1.0	**22**	1.0	1.0	**34**	5.0	1.0
11	0.1	2.0	**23**	1.0	2.0	**35**	5.0	2.0
12	0.1	5.0	**24**	1.0	5.0	**36**	5.0	5.0

parameters *lr* and *Bs* is executed, and the results are depicted in Fig. 22. Specifically, Fig. 22 (a)-(f) show the three-dimensional diagram of diagnostic accuracy changes of the CWRU and BJTU datasets in Case1, Case2, and Case3 scenarios. Overall, a small *lr* and large *Bs* have a negative impact on improving the accuracy of the model, while a large learning rate and an appropriate *Bs* can yield better diagnostic results. Consequently, based on the trend of precision changes, we identify the best corresponding parameter value for precision, as presented in Table 8.

6.5. Ablation experiment

To investigate the weight on the loss function's performance on the model α, β, we conduct a sensitivity analysis and obtain various loss function ablation experiments. First, we construct possible combinations of the joint loss function proposed in this paper, resulting in a total of 36 function combinations, as shown in Table 22. Then we conduct experimental tests under three cases using the CWRU and BJTU datasets. Six experiments are performed to test the diagnostic accuracy of 36 joint loss functions on the model. The experimental results are presented in Fig. 23. As can be seen from the figure, when the weight is greater than 1, the diagnostic performance of the model is relatively poor. The best parameter value can be selected from the experimental results corresponding to each combination, where $\alpha = 1$. And refer to Table 8 for other parameter values.

6.6. Statistical test

In this paper, fault diagnosis experiments under three scenarios are conducted in four datasets, encompassing a total of 14 fault diagnosis methods and 12 fault diagnosis tasks. The critical difference diagram is utilized to analyze the performance differences between each method, as illustrated in Fig. 24. A lower CD value indicates the superior performance of the corresponding method. Notably, our proposed method outperforms the others and exhibits the best performance.

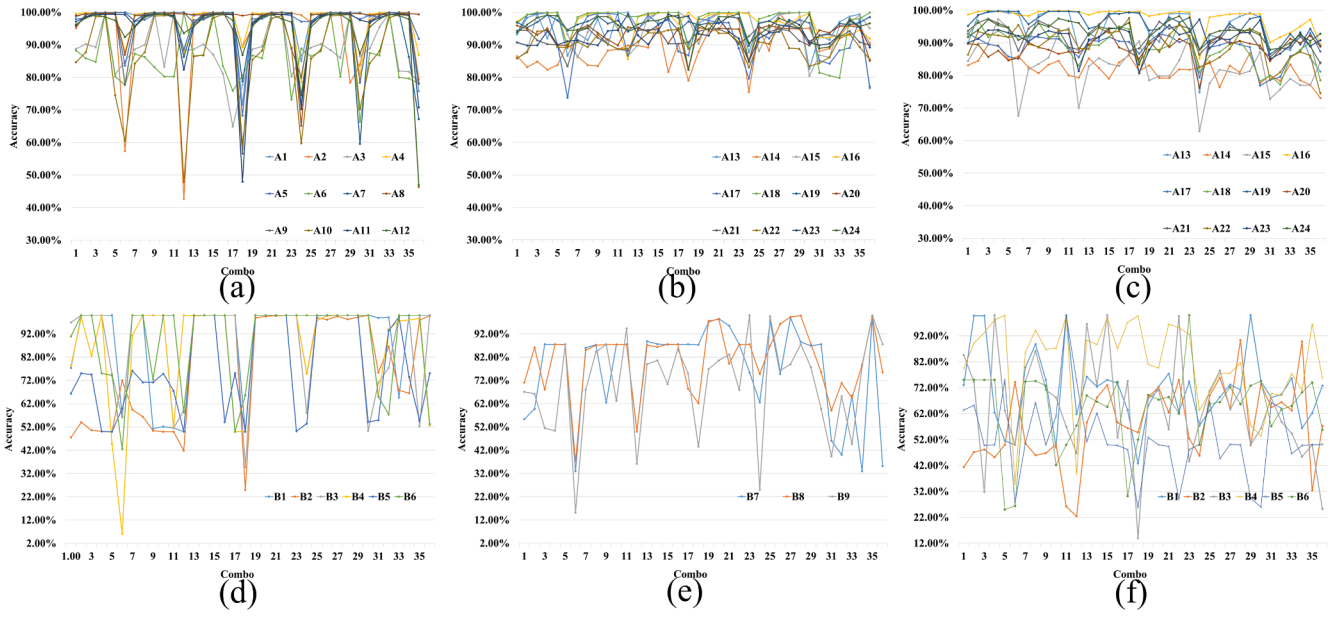


Fig. 23. The results of ablation study for α and β. (a) CWRU dataset in Case1. (b) CWRU dataset in Case2. (c) CWRU dataset in Case3. (d) BJTU dataset in Case1. (e) BJTU dataset in Case2. (f) BJTU dataset in Case3.

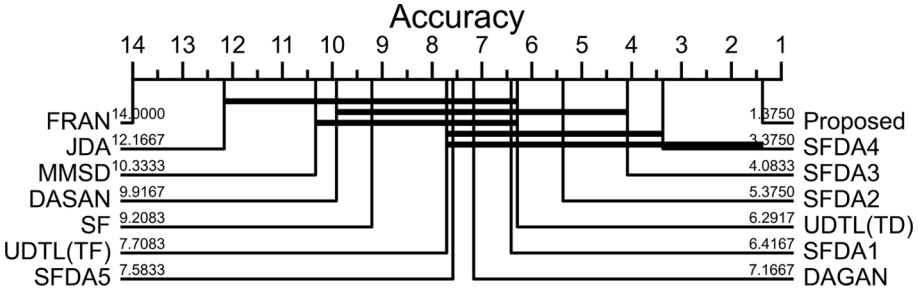


Fig. 24. The CD diagram.

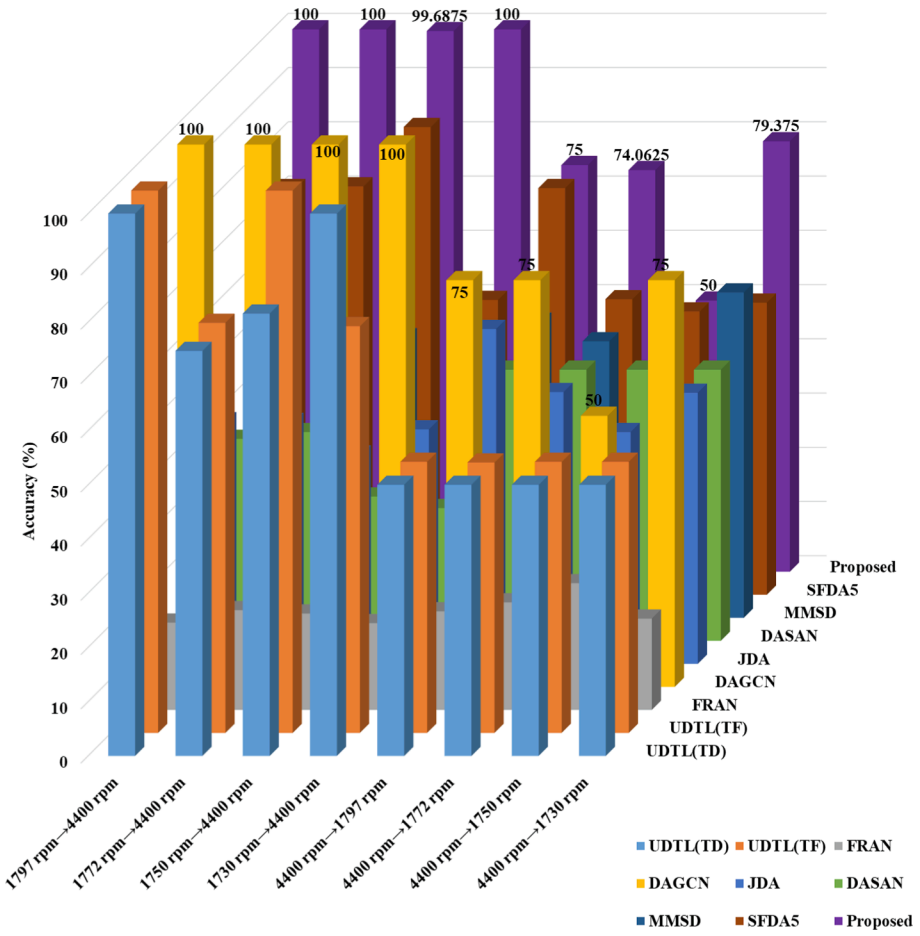


Fig. 25. 3-D column diagram of accuracy corresponding to different methods.

6.7. Further experimental Research: Cross-device fault diagnosis

To further validate the domain adaptation capability of the proposed SFCDA model in this study, we devise a cross-device fault diagnosis task. Specifically, we execute a cross-device transfer task from the publicly available dataset of high-speed EMU traction motor bearing fault data, transitioning from the CWRU dataset to the BJTU dataset. We utilize CWRU data under four distinct loads, corresponding to 1797 rpm, 1772 rpm, 1750 rpm, and 1730 rpm, for the transfer task to high rotational speed 4400 rpm BJTU data. The performance of SFCDA in this cross-device scenario is compared with eight contrasting methods outlined. This is done to assess SFCDA's adaptability in handling cross-device scenarios. The results are illustrated in Fig. 25. It is evident that the diagnostic effectiveness of the proposed method is superior. The lowest diagnostic accuracy is observed in the FRAN model, whereas the DAGCN model exhibited the highest diagnostic accuracy. The incorporation of a domain discrimination loss function within DAGCN confers a notable advantage in addressing domain transfer tasks under cross-device conditions. However, it is noteworthy that DAGCN incurs a considerable computational time overhead. Thus, considering the compromise between performance and computational time, the SFCDA model proposed in this paper remains a commendable domain adaptation model.

6.8. Comparative analysis of different backbone models

The SFCDA model proposed in this article consists of two layers of fully connected neural networks, and its structure is very simple and effective. To explore the reasons behind this, we conduct research on different backbone models. Specifically, for the adaptive problem in Case1 domain of traction motor bearing fault diagnosis in high-speed EMU, the pre-training method proposed in this article is used to operate on various deep neural network models, including 1-layer CNN, 2-layer CNN, 3-layer CNN, 4-layer CNN, 5-layer CNN, 6-layer CNN, long short term memory (LSTM) neural network, multi-layer perceptron (MLP), and transformer classification model. The detailed structure of each model is summarized in Table 23.

In the experimental design phase, to extend the pre-trained weights to other deep learning models, we conduct three different experiments for each model: Strategy1: Deep learning models without pre-trained weights; Strategy2: Deep learning models with pre-trained weights in FC1 participating only as classifier, with the MMD loss function behind FC1 layer; Strategy3: Parameters of FC1 with pre-trained weights, but the MMD loss function in the second-to-last layer of the deep learning model. The code for all comparative models can be found at https://github.com/John-520/Models-for-SFCDA. The final experimental accuracy and computation time are summarized in Fig. 26.

From the results chart, it can be observed that different deep learning models exhibit varying patterns of diagnostic accuracy under different pre-training strategies. Overall, under the influence of pre-training weights, the performance of all deep learning models has improved. However, the accuracy improvement is most pronounced for the MLP based on the fully connected neural network, possibly because of the stronger connectivity between FC1 and MLP. In contrast, models based on CNN exhibit the lowest performance. This may be attributed to the general nature of the CNN model constructed in this study, which did not consider the impact of convolutional kernel size on fault diagnosis tasks. We believe this comparative setting better highlights the advantages of the proposed pre-training strategies.

Table 23
Structural details of different models.

Models	Details	Strategy
CNN_1 CNN_2 CNN_3 CNN_4 CNN_5 CNN_6	Kernel size: 4 × 2, BatchNormal, activation: ReLU; Kernel size: 8 × 2, BatchNormal, activation: ReLU; MaxPool1d: 2 × 2; Kernel size: 8 × 2, BatchNormal, activation: ReLU; Kernel size: 16 × 2, BatchNormal, activation: ReLU; Kernel size: 16 × 2, BatchNormal, activation: ReLU; Kernel size: 32 × 2, BatchNormal, activation: ReLU; AdaptiveMaxPool1d: 4, Linear: 4, activation: ReLU; Dropout.	Strategy1 Strategy2 Strategy3
LSTM	input_size: 1, hidden_size: 64, output_size: 4	Strategy1 Strategy2 Strategy3
MLP	Linear: 1024, BatchNormal, activation: ReLU; Linear: 600, BatchNormal, activation: ReLU; Linear: 500, BatchNormal, activation: ReLU; Linear: 500, BatchNormal, activation: ReLU; Linear: 128, BatchNormal, activation: ReLU; Linear: 4, BatchNormal, activation: ReLU;	Strategy1 Strategy2 Strategy3
Transformer	Number of encoders: 4, Sentence length: 10, Word embedding dimension: 10 Number of attention heads: 10	Strategy1 Strategy2 Strategy3

Although the diagnostic accuracy of the above models is not high, this is because model parameters have not been fine-tuned. This paper only analyzes the weight strategy based on pre-training, without considering the issue of model and data adaptation. Readers can apply this strategy to their models to improve the performance.

6.9. The relationship between model layers and accuracy

To determine the relationship between the number of proposed SFCDA layers and diagnostic accuracy, experimental investigations under different layer numbers are performed. Similarly, research is conducted based on the Case1 task in high-speed traction motor bearings in Section 5.9. We use 2–6 layers of fully connected neural networks for experimental verification. The results of diagnosis accuracy and calculation time are summarized in Fig. 27. We can know that as the number of layers increases, the training time consumption of the model becomes larger and the diagnostic performance fluctuates. The two-layer fully connected layer structure has the best performance and the shortest computing time.

6.10. The interpretability of the SFCDA model

To better explain the transfer learning process and results of SFCDA, we conducted research from two perspectives: 1) Intrinsic interpretability, focusing on the sparsity of FC1 weights. 2) Extrinsic interpretability, exploring the interpretable relationship between model prediction results and original data.

Firstly, we plot two-dimensional color maps of the FC1 weight $\boldsymbol{W}$ with and without pre-training, as shown in Fig. 28. It can be observed that the weight matrix based on pre-trained weights is sparser, indicating that the model can have stronger generalization ability during the testing process.

Secondly, the SHAP algorithm [57] is employed to explain the relationship between the prediction results of the SFCDA model and the samples. SHAP value analysis is an additive interpretable method based on Shapley values, and it can be used to explain any deep learning model. Fig. 29 presents the interpretation of classification results for high-speed traction motor bearings under NC, OF, IF, and RF conditions at 3480 rpm. It can be observed that different sample points have different weights, where the redder the color of the weight value, the greater the contribution of that point to the model's prediction results. Through this interpretable method, we can provide sample-based explanations for the prediction results of different samples, supporting the interpretability of the model's decisions.

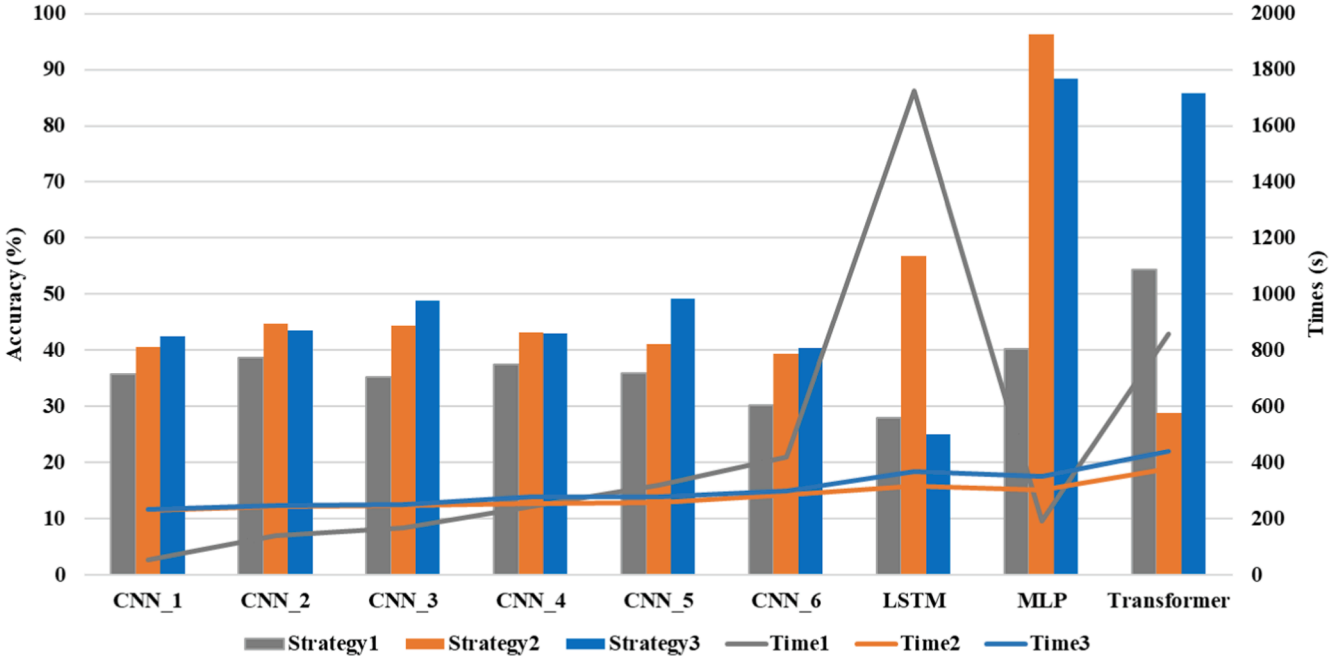


Fig. 26. Performance comparison results of different deep learning models under pre-training support.

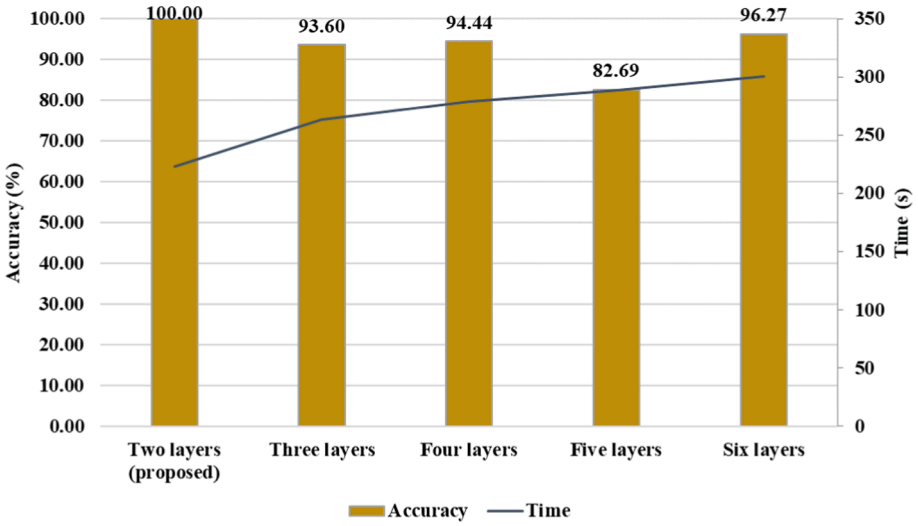


Fig. 27. The relationship diagram between model layers and accuracy.

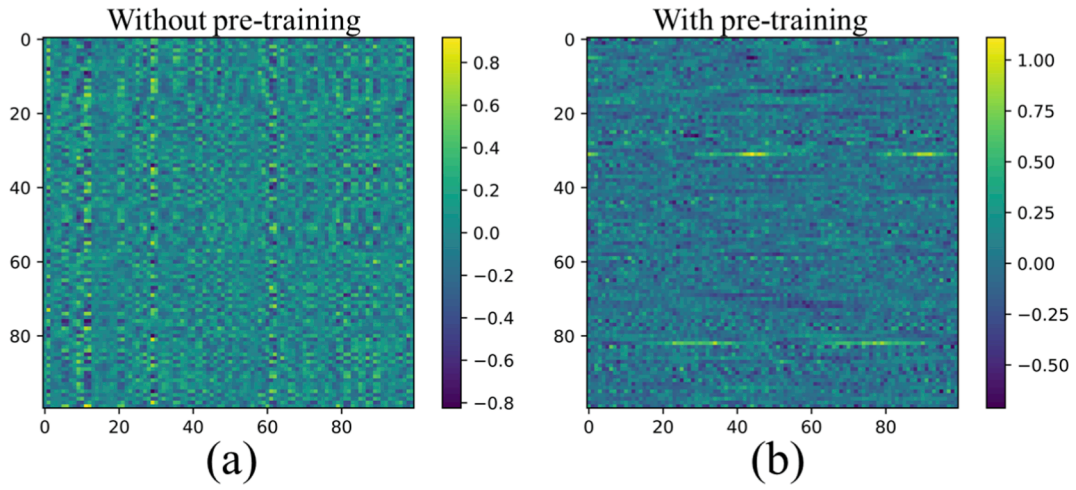


Fig. 28. The comparison of weight matrices without and with pre-training.

6.11. *Advantages and disadvantages*

The proposed SFCDA model is effectively validated in the traction motor bearings of high-speed EMU and publicly available motor bearings. The demonstrated fault diagnosis method has the following advantages:

(1) The newly proposed RSF pre-training strategy for FC layers can enhance feature extraction and aggregation capabilities. Experimental results combining various deep learning models confirm the universality of this strategy, improving the performance of fundamental backbone models.
(2) The constructed joint loss function enhances the model's discriminative ability from the perspective of output result diversity, thereby improving diagnostic performance. The principles of this loss function can be applied to other domain adaptive models, leading to potential performance improvements. Experimental verification suggests that researchers should use the proposed RSF pre-training module by directly adding different deep learning models behind the first FC layer to enhance the backbone model's performance.
(3) The fault diagnosis model based on SFCDA is characterized by its simple structure and short training time, which is crucial for practical high-speed EMU bearing fault diagnosis. This implies that the proposed model lays the groundwork for lightweight fault diagnosis systems in the future, significantly reducing the time and financial costs of the diagnosis recognition process.
(4) The proposed fault diagnosis method can simultaneously address three different fault transfer learning scenarios, providing new insights into exploring the generality of fault diagnosis models.

Of course, the proposed model still has limitations. Although SFCDA is unsupervised learning for the target domain data, data is still required. How to extend SFCDA to scenarios where target domain data is inaccessible is a future research direction. In addition, this paper has not explored fault diagnosis scenarios under time-varying speed conditions. In future research, we consider combining the proposed method with signal resampling techniques to develop a more robust fault diagnosis method for high-speed EMU traction motor bearings under time-varying speed conditions.

7. Conclusion

In this paper, a new cross-domain adaptive model based on SF is proposed, and the traction motor bearing fault diagnosis in high-speed EMU in three cases is tested. Experimental results affirm that the

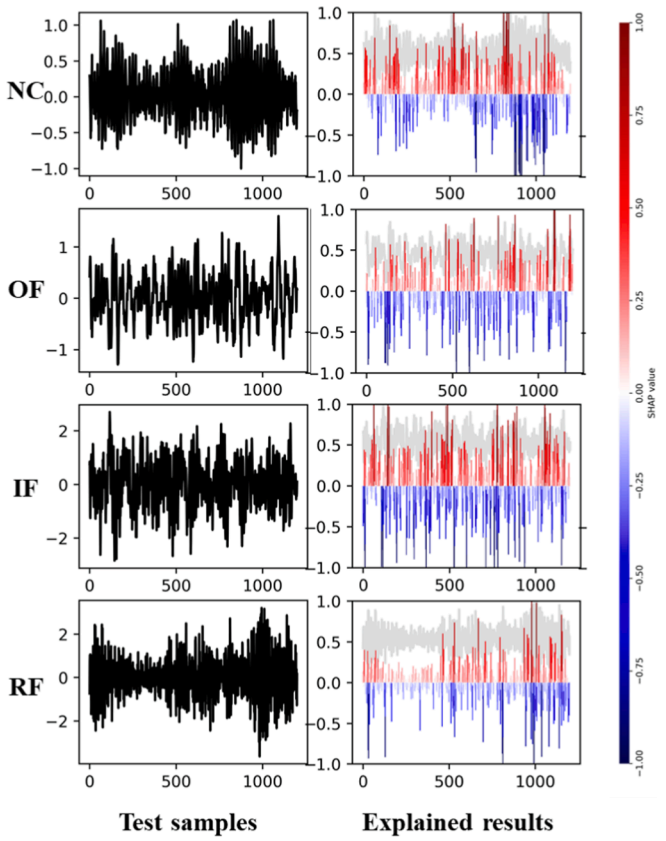


Fig. 29. SHAP-based sample interpretable results under different health states.

SFCDA model achieves the highest diagnostic accuracy and exhibits lower computational costs in multi-scene learning compared to similar methods and state-of-the-art domain adaptive approaches. The primary conclusions derived from this study are outlined as follows:

1) By introducing pre-training based on RSF, the feature extraction ability of FC can be effectively improved.
2) The SFCDA model, consisting of a double-layer FC, is simple and efficient. Compared to the complex RFAN model, it reduces time and space complexity by approximately 68 times and 1745 times, respectively.
3) Experimental results unequivocally demonstrate the superior performance of the SFCDA model in addressing three distinct bearing fault diagnosis problems, coupled with a reduced demand for training set data.

CRediT authorship contribution statement

Feiyu Lu: Writing – original draft, Software, Methodology, Data curation. **Qingbin Tong:** Investigation, Conceptualization. **Jianjun Xu:** Writing – review & editing. **Ziwei Feng:** Visualization, Supervision. **Xin Wang:** Formal analysis. **Jingyi Huo:** Software. **Qingzhu Wan:** Validation, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the Beijing Natural Science Foundation (Grant no. L211010) and the Fundamental Research Funds for the Central Universities (2023JBZY039). The authors wish to extend their sincere thanks for the support from the Beijing Municipal Science & Technology Commission of China.

References

[1] C. Liu, C. Qin, X. Shi, Z. Wang, G. Zhang, Y. Han, TScatNet: an interpretable cross-domain intelligent diagnosis model with antinoise and few-shot Learning capability, IEEE Trans. Instrum. Meas. 70 (2021) 1–10.
[2] R.B. Randall, J. Antoni, Rolling element bearing diagnostics-a tutorial, Mech. Syst. Sig. Process. 25 (2011) 485–520.
[3] D. Zhao, S. Liu, H. Du, L. Wang, Z. Miao, Deep branch attention network and extreme multi-scale entropy based single vibration signal-driven variable speed fault diagnosis scheme for rolling bearing, Adv. Eng. Inf. 55 (2023) 101844.
[4] D. Liu, L. Cui, W. Cheng, Flexible generalized demodulation for intelligent Bearing fault diagnosis under Nonstationary conditions, IEEE Trans. Ind. Inf. 1–12 (2022).
[5] J.-Y. Chen, Y.-W. Feng, D. Teng, C. Lu, C.-W. Fei, Support vector machine-based similarity selection method for structural transient reliability analysis, Reliab. Eng. Syst. Safety 223 (2022) 108513.
[6] J. Ngiam, Z. Chen, S. Bhaskar, P. Koh, A. Ng, Sparse Filtering, in: J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, K.Q. Weinberger (Eds.) Advances in Neural Information Processing Systems, Curran Associates, Inc., 2011.
[7] Y. Zhang, L. Gao, X. Wen, H. Wang, Intelligent fault diagnosis of machine under noisy environment using ensemble orthogonal contractive auto-encoder, Expert Syst. Appl. 203 (2022).
[8] Z. Xie, J. Chen, Y. Feng, K. Zhang, Z. Zhou, End to end multi-task learning with attention for multi-objective fault diagnosis under small sample, J. Manuf. Syst. 62 (2022) 301–316.
[9] Y.G. Lei, B. Yang, X.W. Jiang, F. Jia, N.P. Li, A.K. Nandi, Applications of machine learning to machine fault diagnosis: a review and roadmap, Mech. Syst. Sig. Process. 138 (2020).
[10] D. Ruan, J. Wang, J. Yan, C. Gühmann, CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis, Adv. Eng. Inf. 55 (2023) 101877.
[11] G. Jiang, H. He, J. Yan, P. Xie, Multiscale convolutional neural networks for fault diagnosis of wind turbine Gearbox, IEEE Trans. Ind. Electron. 66 (2019) 3196–3207.
[12] C. He, H. Shi, J. Si, J. Li, Physics-informed interpretable wavelet weight initialization and balanced dynamic adaptive threshold for intelligent fault diagnosis of rolling bearings, J. Manuf. Syst. 70 (2023) 579–592.
[13] P. Liang, Z. Yu, B. Wang, X. Xu, J. Tian, Fault transfer diagnosis of rolling bearings across multiple working conditions via subdomain adaptation and improved vision transformer network, Adv. Eng. Inf. 57 (2023) 102075.
[14] M. Deng, A. Deng, Y. Shi, Y. Liu, M. Xu, A novel sub-label learning mechanism for enhanced cross-domain fault diagnosis of rotating machinery, Reliab. Eng. Syst. Safety 225 (2022).
[15] M. Xia, T. Li, L. Xu, L. Liu, C.W. de Silva, Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks, IEEE/ASME Trans. Mechatron. 23 (2018) 101–110.
[16] T. Li, Z. Zhao, C. Sun, R. Yan, X. Chen, Multireceptive field graph convolutional networks for machine fault diagnosis, IEEE Trans. Ind. Electron. 68 (2021) 12739–12749.
[17] D. Wei, T. Han, F. Chu, M.J. Zuo, Weighted domain adaptation networks for machinery fault diagnosis, Mech. Syst. Sig. Process. 158 (2021).
[18] W. Li, R. Huang, J. Li, Y. Liao, Z. Chen, G. He, R. Yan, K. Gryllias, A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: theories, applications and challenges, Mech. Syst. Sig. Process. 167 (2022).
[19] J. Luo, H. Shao, H. Cao, X. Chen, B. Cai, B. Liu, Modified DSAN for unsupervised cross-domain fault diagnosis of bearing under speed fluctuation, J. Manuf. Syst. 65 (2022) 180–191.
[20] M. Xia, H. Shao, D. Williams, S. Lu, L. Shu, C.W. de Silva, Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning, Reliab. Eng. System Safety 215 (2021) 107938.
[21] P. Ma, H. Zhang, W. Fan, C. Wang, A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains, ISA Trans., 99 (2020) 465-478-478.
[22] Z. Zhao, T. Li, J. Wu, C. Sun, S. Wang, R. Yan, X. Chen, Deep learning algorithms for rotating machinery intelligent diagnosis: an open source benchmark study, ISA Trans. 107 (2020) 224–255.
[23] T. Li, Z. Zhao, C. Sun, R. Yan, X. Chen, Domain Adversarial graph convolutional network for fault diagnosis under Variable working conditions, IEEE Trans. Instrum. Meas. 70 (2021) 1–10.
[24] J. Chen, J. Wang, J. Zhu, T.H. Lee, C.W. de Silva, Unsupervised cross-domain fault diagnosis using feature representation alignment networks for rotating machinery, IEEE/ASME Trans. Mechatron. 26 (2021) 2770–2781.
[25] F. Lu, Q. Tong, Z. Feng, Q. Wan, Unbalanced Bearing fault diagnosis under Various speeds based on Spectrum alignment and deep transfer convolution neural network, IEEE Trans. Ind. Inf. 19 (2023) 8295–8306.

[26] T. Zhang, J. Chen, F. Li, K. Zhang, H. Lv, S. He, E. Xu, Intelligent fault diagnosis of machines with small & imbalanced data: a state-of-the-art review and possible extensions, ISA Trans 119 (2022) 152–171.
[27] Z. Meng, H. He, W. Cao, J. Li, L. Cao, J. Fan, M. Zhu, F. Fan, A novel generation network using feature fusion and guided adversarial learning for fault diagnosis of rotating machinery, Expert Syst. Appl. 234 (2023) 121058.
[28] K. Wang, T. Zhou, M. Luo, X. Li, Z. Cai, Generative adversarial minority enlargement—A local linear over-sampling synthetic method, Expert Syst. Appl. 237 (2024) 121696.
[29] T. Zhang, C. Li, J. Chen, S. He, Z. Zhou, Feature-level consistency regularized semi-supervised scheme with data augmentation for intelligent fault diagnosis under small samples, Mech. Syst. Sig. Process. 203 (2023) 110747.
[30] C. He, H. Shi, J. Li, IDSN: a one-stage interpretable and differentiable STFT domain adaptation network for traction motor of high-speed trains cross-machine diagnosis, Mech. Syst. Sig. Process. 205 (2023) 110846.
[31] X. Zhang, Y. Li, Y. Liu, D. Li, X. Chen, Adaptive fault diagnosis and Decision-making method based on multi-Spectrum evaluation and fusion for Traction motor Bearings, IEEE Trans. Instrum. Meas. 72 (2023) 1–19.
[32] K. He, Y. Xu, Y. Wang, J. Wang, T. Xie, Intelligent diagnosis of rolling Bearings fault based on multisignal fusion and MTF-ResNet, Sensors, MDPI AG (2023) 6281.
[33] H. Sun, D. He, Z. Lao, Z. Jin, C. Liu, S. Shan, Fault diagnosis of train traction motor bearing based on improved deep residual network, Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci. (2023).
[34] D. He, Z. Lao, Z. Jin, C. He, S. Shan, J. Miao, Train bearing fault diagnosis based on multi-sensor data fusion and dual-scale residual network, Nonlinear Dyn. 111 (2023) 14901–14924.
[35] Y. Lei, F. Jia, J. Lin, S. Xing, S.X. Ding, An intelligent fault diagnosis method using unsupervised feature Learning Towards mechanical big data, IEEE Trans. Ind. Electron. 63 (2016) 3137–3147.
[36] Z. Zhang, S. Li, J. Wang, Y. Xin, Z. An, X. Jiang, Enhanced sparse filtering with strong noise adaptability and its application on rotating machinery fault diagnosis, Neurocomputing 398 (2020) 31–44.
[37] Z. Zhang, Q. Yang, Unsupervised feature learning with reconstruction sparse filtering for intelligent fault diagnosis of rotating machinery, Appl. Soft Comput. 115 (2022).
[38] C. Cheng, W. Liu, W. Wang, M. Pecht, A novel deep neural network based on an unsupervised feature learning method for rotating machinery fault diagnosis, Measure. Sci. Technol. 32 (2021).
[39] Z. Zhang, H. Chen, S. Li, Z. An, Sparse filtering based domain adaptation for mechanical fault diagnosis, Neurocomputing 393 (2020) 101–111.
[40] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, J. Mach. Learn. Res. 13 (2012) 723–773.
[41] A. Müller, Integral probability metrics and their generating classes of functions, Adv. Appl. Probab. 29 (1997) 429–443.
[42] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423.
[43] S. Cui, S. Wang, J. Zhuo, L. Li, Q. Huang, Q. Tian, Towards discriminability and diversity: batch Nuclear-norm maximization under label insufficient situations. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[44] M. Fazel, T.K. Pong, D. Sun, P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, SIAM J. Matrix Anal. Appl. 34 (2013) 946–977.
[45] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of Linear matrix equations via Nuclear norm minimization, SIAM Review 52 (2010) 471–501.
[46] Z. Zhang, Q. Yang, Z. Wu, Sparse filtering with Adaptive basis weighting: a novel representation Learning method for intelligent fault diagnosis, IEEE Trans. Syst., Man, Cybern.: Syst. 52 (2022) 1019–1025.
[47] Q.V. Le, A. Karpenko, J. Ngiam, A. Ng, ICA with reconstruction cost for efficient overcomplete feature Learning, NIPS (2011).
[48] D. Kingma, J. Ba, Adam: a method for stochastic optimization, Computer Science (2014).
[49] Case Western Reserve University Bearing Data Center Website [Online] Available: http://csegroups.case.edu/bearingdatacenter/home.
[50] K. Li, X. Ping, H. Wang, P. Chen, Y. Cao, Sequential fuzzy diagnosis method for motor roller Bearing in Variable operating conditions based on vibration analysis, Sensors, MDPI AG (2013) 8013–8041.
[51] N.D. Thuan, H.S. Hong, HUST bearing: a practical dataset for ball bearing fault diagnosis, BMC Res. Notes 16 (2023) 138.
[52] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, J. Mach. Learn. Res. 17 (2016) 2096.
[53] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, X. Chen, Applications of unsupervised deep transfer Learning to intelligent fault diagnosis: a survey and Comparative study, IEEE Trans. Instrum. Meas. 70 (2021) 1–28.
[54] T. Han, C. Liu, W. Yang, D. Jiang, Deep transfer network with joint distribution adaptation: a new intelligent fault diagnosis framework for industry application, ISA Trans. 97 (2020) 269–281.
[55] Y. Liu, Y. Wang, T.W.S. Chow, B. Li, Deep Adversarial subdomain Adaptation network for intelligent fault diagnosis, IEEE Trans. Ind. Inf. 18 (2022) 6038–6046.
[56] Q. Qian, Y. Qin, J. Luo, Y. Wang, F. Wu, Deep discriminative transfer learning network for cross-machine fault diagnosis, Mech. Syst. Sig. Process. 186 (2023) 109884.
[57] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inform. Process. Syst. 30 (2017).