

Research Article

A New Framework Based on Supervised Joint Distribution Adaptation for Bearing Fault Diagnosis across Diverse Working Conditions

Chengyao Liu ¹ and Fei Dong ^{2,3}

¹Department of Jiaotong, Zhejiang Industry Polytechnic College, Shaoxing 312000, China

²School of Internet, Anhui University, Hefei 230039, China

³School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221000, China

Correspondence should be addressed to Fei Dong; feidong@cumt.edu.cn

Received 4 September 2023; Revised 20 November 2023; Accepted 14 December 2023; Published 3 January 2024

Academic Editor: Zhipeng Zhao

Copyright © 2024 Chengyao Liu and Fei Dong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To address the degradation of diagnostic performance due to data distribution differences and the scarcity of labeled fault data, this study has focused on transfer learning-based cross-domain fault diagnosis, which attracts considerable attention. However, deep transfer learning-based methods often present a challenge due to their time-consuming and costly nature, particularly in tuning hyperparameters. For this issue, on the basis of classical features-based transfer learning method, this study introduces a new framework for bearing fault diagnosis based on supervised joint distribution adaptation and feature refinement. It first utilizes ensemble empirical mode decomposition to process raw signals, and statistical features extraction is implemented. Then, a new feature refinement module is designed to refine domain adaptation features from high-dimensional feature set by evaluating the fault distinguishability and working-condition invariance of feature data. Next, it proposes a supervised joint distribution adaptation method to conduct improved joint distribution alignment that preserves neighborhood relationships within a manifold subspace. Finally, an adaptive classifier is trained to predict fault labels of feature data across varying working conditions. To prove the cross-domain fault diagnosis performance and superiority of the proposed methods, two bearing datasets are applied for experiments, and the experimental results verify that the model built by the proposed framework can achieve desirable diagnosis performance under different working conditions and that it apparently outperforms comparative models.

1. Introduction

In the last several years, with the speedy and sustained advancement of modern industrial equipment, rotating machinery plays a major role in various production scenarios, such as transportation, mining, logistics, electricity, and manufacturing [1]. Due to that, the bearing is one of the most important units of industrial machinery, and the malfunction of the bearing may cause serious accidents and economic losses. Moreover, bearing typically operates under complicated operating circumstances, which may cause it easy to malfunction. Importantly and challengingly, it is mostly difficult to collect fault samples of real-world mechanical facilities under variable operating conditions [2].

Therefore, when facing real-world industrial scenes, the most existing artificial intelligence-based fault diagnosis techniques of rolling bearing still suffer from some challenges, such as data distribution differences and inadequate fault samples [3, 4].

Artificial intelligence technologies applied to fault diagnosis of bearings are mainly divided into three classes: classical machine learning-based method (CMLM), deep learning-based method (DLM), and transfer learning-based method (TLM) [5, 6]. Commonly, CMLM that has been widely studied since many years ago include the support vector machine (SVM) [7], artificial neural network (ANN) [8], k-nearest neighbor (KNN) [9], extreme learning machine (ELM) [10], and random forest (RF) [11]. These

methods possess some major drawbacks, including heavy reliance on expert knowledge under variable working conditions and a default assumption that the samples share the same probability distribution [3, 6]. At present, DLM has attracted widespread attention and research with the help of their powerful ability to automatically extract deep features with better representation performance. Commonly, studied approaches include deep auto-encoder (DAE) [12], deep residual network [13], deep belief network (DBN) [14], and convolutional neural network (CNN) [15]. Nevertheless, several shortcomings of DLM are still prominent [1, 3]. Particularly, the fault diagnosis of rotating machinery based on traditional DLM adheres to the hypothesis that the data under diverse working conditions follow the identical distribution, which is adversarial to data distribution deviation under actual operating status. Furthermore, a bearing fault diagnosis model based on DLM requires sufficient training samples to achieve ideal fault diagnosis performance, which contradicts the insufficient fault data under actual industrial scenes. Furthermore, DLM usually involves a high-cost and high-time-consuming procedure to tune numerous hyperparameters [3].

To date, TLM has made increasing attention and research in cross-domain fault diagnosis (CFD) due to their distribution adaptation ability that is hope to tackle the above challenges of CMLM and DLM. TLM intends to learn the related domain knowledges from source domain (SD) and utilize them to target domain (TD). In the bearing fault diagnosis field, a fault dataset under one working state can constitute a domain. Transfer learning methods can be mainly divided into two classes: classical manual feature extraction-based transfer learning (TL) approaches and deep transfer learning (DTL) approaches [3, 16]. Although DTL methods have attracted increasing attentions in bearing fault diagnosis towards different working conditions, they still have some drawbacks. A common and important one is that a desirable DTL-based fault diagnosis model requires a high-cost and time-consuming procedure because of the adjustment of numerous hyperparameters. Accordingly, in this article, we focus on the typical feature-based TL approach to achieve the desirable CFD of rolling bearing in real-world industrial scenarios. Commonly studied feature-based TL methods mainly include the balanced distribution adaption (BDA) [17], joint distribution adaption (JDA) [18], transfer component analysis (TCA) [19], geodesic flow kernel (GFK) [20], and joint geometrical and statistical alignment (JGSA) [21]. Based on these methods, some intelligent models for cross-domain diagnosis have been investigated. In [22], a transfer deep learning network was proposed to resolve the drawbacks of existing rolling bearing fault algorithms on the basis of deep learning. In this network, the feature transfer using TCA and a pretrained convolutional neural network is performed. In [23], a source domain multisample JDA (SM-JDA) approach was used for the bearing fault diagnosis under variable operating conditions. In [24], the BDA was introduced to facilitate the domain adaptation on bearing cross-domain fault diagnosis. In [25], aiming at the domain shift (distribution discrepancy) issue in the field of bearing fault diagnosis, the multikernel joint distribution adaptation

(MKJDA) with dynamic distribution alignment is proposed for bearing fault diagnosis. In [3], based on BDA, a new balanced adaptation regularization was designed to solve the problem of sample distribution discrepancy-caused degradation of CFD performance. In [26], an adaptive manifold probability distribution was studied for CFD; in this method, the GFK was implemented for distribution adaptation, and a domain adaptive classifier was further trained to diagnose the target domain under different working conditions. In [27], transfer sparse coding and JGSA were combined to construct a novel fault diagnosis approach for bearing under different operating status. Although the above-mentioned methods have successfully realized CFD of bearings, three issues are still blocking the application of these methods in actual industrial scenarios. (1) The implementation of distribution adaptation in most studies is based on the probability distributions alignment in the primitive characteristic space, which makes it difficult to tackle the issue of feature distortion and may lead to the poor domain adaptation (DA) performance [28]. (2) The goals of mostly distribution adaptation of TLM merely concentrate on decreasing probability distribution differences and enhancing the transferability of features, and the class distinguishability of features is usually neglected, which may lead to the poor classification performance [29, 30]. (3) In the process of distribution adaptation, the impact of class information and neighborhood relationships of feature data on distribution adaptation has not been effectively considered, which may restrict the CFD accuracy and generalization ability of the model [28, 31].

Considering three issues of the above-mentioned TLM approaches, we investigate a new DA idea, that is, joint distribution alignment with neighborhood relationship preserving in manifold subspace. Moreover, for improving the DA capability, we consider the impact of fault discriminability and working-condition invariance (WCI) of features in the procedure of DA. Therefore, we designed a feature refinement module to refine features with the better domain adaptability from the original high-dimensional feature set (OHFS). In view of the above discussion, this study proposes a new CFD framework for bearing on the basis of feature refinement and supervised JDA. There are four modules in this framework: signal processing and feature extraction module, feature refinement module, DA module, and classifier module for CFD. In the signal processing and feature extraction module, it uses ensemble empirical mode decomposition (EEMD) to decompose the raw signals collected from bearing and conducts feature extraction. For the feature refinement module, a domain adaptation feature refinement based on classification accuracy and distribution discrepancies (DFCD) is investigated to estimate the fault distinguishability and WCI of feature. In DA module, a new DA method, termed improved JDA with manifold subspace learning and neighborhood relationship preserving (IDAMN), is proposed. Finally, in the cross-domain classifier module, the classical machine learning classifier, the KNN, is trained by labeled data of SD, and the trained classifier predicts the labels of data of TD. The main contributions are summarized as follows:

- (1) A feature refinement module is designed, named domain adaptation feature refinement based on classification accuracy and distribution discrepancies (DFCD). The classical classifier KNN is utilized to estimate the fault distinguishability of features, and the maximum mean discrepancy (MMD) and Kullback–Leibler divergence (KLD) are employed to quantify the WCI of features. Accordingly, it constructs a new feature estimation index to refine DA features with better fault distinguishability and WCI from the primitive characteristic set.
- (2) It proposes a new DA method, improved JDA with manifold subspace learning, and neighborhood relationship preserving (IDAMN). IDAMN performs improved JDA of different domains in a learned manifold subspace with the consideration of neighborhood relationship preserving and category information, which help to shrink distribution differences while overcoming feature distortion and enhancing the discriminant performance of features.
- (3) Aiming at the key challenges still exist in applying artificial intelligence-based fault diagnosis approaches to actual application scenes, a new fault diagnosis framework constructed by the DFCD and IDAMN is designed, termed as DFCD-IDAMN. This framework can prominently strengthen CFD performance. Two bearing datasets are utilized to set up a series of CFD tasks in experimental verification. The outcome shows that DFCD-IDAMN significantly outperforms other comparative models that use common baseline methods.

The rest of the contents are arranged as follows. In Section 2, the preliminary knowledges of ensemble empirical mode decomposition, domain adaptation, MMD, and local fisher discriminant analysis are introduced, respectively. Section 3 describes the DFCD-IDAMN framework. In Section 4, the experimental validation is given to illustrate the performance of the proposed methods. The conclusions of this work are presented in Section 5.

2. Preliminaries

2.1. Ensemble Empirical Mode Decomposition (EEMD). EEMD was proposed to overcome the mode confusion problem of empirical mode decomposition (EMD), its basic principle is that Gaussian white noise is added into raw signals, and signals can be automatically distributed to the appropriate reference scale. Therefore, EEMD can achieve the better time-frequency analysis of nonstationary vibration signals from bearings [32, 33]. The procedure of EEMD is illustrated in Figure 1, and the specific implementation process of EEMD is as follows [34]:

- (1) Given an original signal $s(t)$, set up the variable i as 1, and set up the average times of EEMD as N .
- (2) Add the Gaussian white noise (GWN) $n_i(t)$ to $s(t)$, and the signal $s_i(t)$ can be obtained. The expression of $s_i(t)$ is as follows:

$$s_i(t) = s(t) + n_i(t). \quad (1)$$

- (3) Apply EMD to process $s_i(t)$, and various intrinsic mode functions (IMF) and the corresponding residual components can be obtained; the expression of $s_i(t)$ can be presented as follows:

$$s_i(t) = \sum_{j=1}^J \text{IMF}_{ij}(t) + r_{ij}(t), \quad (2)$$

where $\text{IMF}_{ij}(t)$ represents the j -th IMF component obtained by EMD, J is the numbers of IMF, and $r_{ij}(t)$ represents the residual components.

- (4) Add different GWN to $s(t)$ and repeat steps (2) and (3), obtain the sum and average of the IMF components reached in N decompositions to offset the GWN, and the final IMF components can be obtained as follows:

$$\text{IMF}_j(t) = \frac{1}{N} \sum_{i=1}^N \text{IMF}_{ij}(t) + r_{ij}(t). \quad (3)$$

- (5) Through the above steps, $s(t)$ is finally decomposed to

$$s(t) = \sum_j \text{IMF}_j + r(t). \quad (4)$$

2.2. Domain Adaptation and MMD. Domain adaptation (DA) is a bright transfer learning-based approach in the situation that traditional pattern recognition and classification models do not achieve ideal results due to the distribution discrepancies between the training and testing samples [1, 5]. Given a SD $D^S = \{X^S, Y^S\} = \{x_i, y_i\}_{i=1}^{n_S}$ and a TD $D^T = \{X^T, Y^T\} = \{x_i, y_i\}_{i=n_S+1}^{n_T}$, n_S and n_T are, respectively, the number of samples from D_S and D_T , respectively. $X = \{x_i\}_{i=1}^n \in \chi$ represents data of SD and TD; $Y = \{y_i\}_{i=1}^n \in \gamma$ represents the corresponding label set of X . D^T and D^S are drawn from two different probability distributions, and the optimization goal of DA is to shrink the distribution discrepancies between SD and TD [35].

MMD [36], a widely used nonparametric distance estimation in TL, was proposed by Gretton et al. for estimating the distance of distributions based on reproducing kernel Hilbert space (RKHS). The MMD between distributions of D_S and D_T can be expressed as

$$\text{MMD}(D^S, D^T) = \left\| \frac{1}{n_S} \sum_{x_i \in X^S} \phi(x_i) - \frac{1}{n_T} \sum_{x_j \in X^T} \phi(x_j) \right\|_H^2, \quad (5)$$

where $\|\bullet\|_H$ represents the RKHS norm and $\phi(\cdot)$ is the transformation function that transforms data to a RKHS. Aiming at the challenge of that inconsistent feature distribution is existed in CFD, the MMD has been widely utilized to estimate distribution discrepancies between domains and align data distributions.

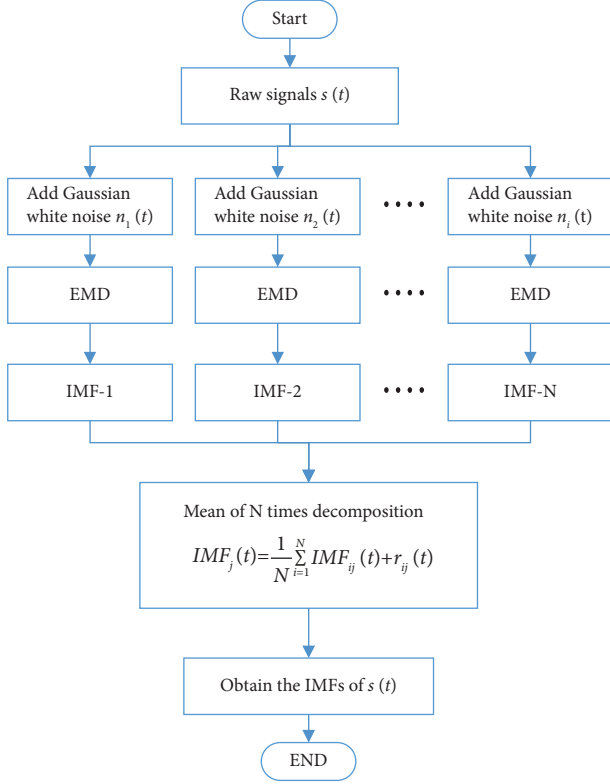


FIGURE 1: The procedure of EEMD.

2.3. Local Fisher Discriminant Analysis (LFDA). LFDA was proposed by improving local fisher analysis (LFA) by Sugiyama [37], and it is a classical supervised dimensionality reduction approach. Let $f_i \in R^d$, $i = 1, 2, \dots, n$ be d -dimensional data and $y_i \in [1, c]$ be the corresponding category labels, where n and c are, respectively, the number of f_i and the class number of data. According to the literature [37, 38], the objective of LDA is to maximize the proportion of the between-class scatter matrix (BSM) S_b to the within-class scatter matrix (WSM) S_w :

$$J(A) = \max_A \frac{|A^T S_b A|}{|A^T S_w A|}, \quad (6)$$

where A is a mapping matrix, and the definitions of S_b and S_w are as follows:

$$S_b = \frac{1}{2} \sum_{i,j=1}^n p_{ij}^b (f_i - f_j)(f_i - f_j)^T, \quad (7)$$

$$S_w = \frac{1}{2} \sum_{i,j=1}^n p_{ij}^w (f_i - f_j)(f_i - f_j)^T, \quad (8)$$

where

$$p_{ij}^b = \begin{cases} \frac{1}{n} - \frac{1}{n_l}, & y_i = y_j = l, \\ \frac{1}{n}, & y_i \neq y_j, \end{cases} \quad (9)$$

$$p_{ij}^w = \begin{cases} \frac{1}{n_l}, & y_i = y_j = l, \\ 0, & y_i \neq y_j, \end{cases} \quad (10)$$

where n_l is the number of samples in class l . Compared to LDA, the higher objective of LFDA is that the between-categories divisibility is maximized and the within-category local manifold structure is preserved simultaneously in a new feature space with reduced dimension. Based on the above S_b and S_w , the local relationship of feature data can be incorporated into the definition of weight. Accordingly, the new BSM \tilde{S}_b and WSM \tilde{S}_w have been substituted for S_b and S_w , respectively. The expressions of \tilde{S}_b and \tilde{S}_w are presented as follows [37]:

$$\tilde{S}_b = \frac{1}{2} \sum_{i,j=1}^n \tilde{p}_{ij}^b (f_i - f_j)(f_i - f_j)^T, \quad (11)$$

$$\tilde{S}_w = \frac{1}{2} \sum_{i,j=1}^n \tilde{p}_{ij}^w (f_i - f_j)(f_i - f_j)^T, \quad (12)$$

where

$$\tilde{p}_{ij}^b = \begin{cases} A_{ij} \left(\frac{1}{n} - \frac{1}{n_l} \right), & y_i = y_j = l, \\ \frac{1}{n}, & y_i \neq y_j, \end{cases} \quad (13)$$

$$\tilde{p}_{ij}^w = \begin{cases} \frac{A_{ij}}{n_l}, & y_i = y_j = l, \\ 0, & y_i \neq y_j, \end{cases} \quad (14)$$

where the definition of A_{ij} is shown as follows:

$$A_{ij} = \exp\left(-\frac{\|f_i - f_j\|^2}{\gamma_i \gamma_j}\right), \quad (15)$$

where γ_i and γ_j are the local scaling around f_i and f_j .

3. DFCD-IDAMN Framework

To achieve a desirable CFD of bearing, this work designs a new DFCD-IDAMN framework based on the domain adaptation feature refinement method DFCD and supervised joint distribution adaptation IDAMN. The whole

structure is presented in Figure 2. DFCD-IDAMN framework is constituted by four modules: signal processing and feature extraction module, feature refinement module, domain adaptation module, and adaptive classifier module. The specific introduction is presented as follows.

3.1. Signal Processing and Feature Extraction Module. Due to that, the original bearing vibration signals usually possess severe nonlinearity and nonstationarity; in an effort to tackle this issue and extract features that can effectively reflect fault states and help pattern recognition and classification, first of all, EEMD is applied to decompose the collected vibration signals into several different IMFs, and these IMFs are utilized to calculate the Hilbert envelope spectrum (HES) and Hilbert marginal spectrum (HMS). By using raw vibration signals, the decomposed IMFs, and the corresponding HES, this module calculates the statistical parameters of them to obtain the corresponding statistical features. The procedure of this module is drawn in Figure 3.

3.2. Feature Refinement Module. In order to strengthen the performance of domain adaptation procedure, this work designs the feature refinement module to refine domain adaptation features with more satisfying fault discriminability and WCI from the original high-dimensionality feature set. This module, domain adaptation feature refinement based on classification accuracy and distribution discrepancies (DFCD), is built by using the classical classifier KNN, maximum mean discrepancy (MMD), and Kullback–Leibler divergence (KLD). The structure of DFCD is drawn in Figure 4. The feature datasets extracted from vibration signals in a certain operating status and other operating states are used as SD and TD, respectively. In order to adapt to conditions as close as possible to actual industrial scenarios, the DFCD runs on this input: tagged feature data in fault states and normal state from SD, untagged feature data in fault states from TD, and feature data in normal state from TD. The reason for setting such input is that in actual industrial scenes, it is unknown that which category the newly collected samples belong to and samples in all fault states under one specific working condition are usually easy to prepare and obtain; therefore, the inputted feature data from TD is untagged. However, for any mechanical equipment, the samples in their normal state under all working conditions are easily accessible. Accordingly, the labeled feature data from SD are used to evaluate the fault discriminability due to its known label, and only feature data in the normal state from TD are used to measure the WCI of feature.

According to the structure shown in Figure 4, the labeled feature data (contains multiple fault categories) in a certain operating condition and normal status feature data in other operating conditions are used for feature evaluation. Firstly, it randomly divides the labeled feature data into the training and testing data, and it trains a KNN classifier to predict the class labels of the testing data. Accordingly, the classification accuracy of each feature can be used to measure the fault discriminability. Then, the normal state feature data in two

working conditions is implemented to calculate the MMD and KLD of features, which accomplishes the quantification of the WCI of feature. Finally, a novel evaluation index for domain adaptation features refinement, the domain adaptability index (DAI), is built. In this study, we presume that the feature with higher DAI is more advantageous to domain adaptation and fault classification. The detailed description of DFCD is as follows.

3.2.1. Evaluate Fault Discriminability of Feature Based on Classification Accuracy. Given a high-dimensional original feature set (OFS) that includes P feature samples containing K class data, that is, $\text{OFS} = [f_1, f_2, \dots, f_P]^T$. For each sample, it is constructed by Q features, that is, $f_i = \{f_i^1, f_i^2, \dots, f_i^Q\}$, and $i \in [1, P]$, where f_i^q represents the q -th feature of the i -th sample. Accordingly, the OFS can be presented as follows:

$$\text{OFS} = \begin{bmatrix} f_1^1 & f_2^1 & \dots & f_P^1 \\ f_1^2 & f_2^2 & \dots & f_P^2 \\ \vdots & \vdots & \ddots & \vdots \\ f_1^Q & f_2^Q & \dots & f_P^Q \end{bmatrix}. \quad (16)$$

The row of OFS, the first feature data $[f_1^1, f_2^1, \dots, f_P^1]$, is used to obtain classification accuracy by KNN classifier. The labeled q -th feature data from source domain are randomly divided into the training dataset $\{D_{\text{train}}, Y_{\text{train}}\}$ and the testing dataset $\{D_{\text{test}}, Y_{\text{test}}\}$. The D_{train} and D_{test} present the training and testing data samples, respectively. The Y_{train} and Y_{test} are the corresponding labels of D_{train} and D_{test} . On this basis, the $\{D_{\text{train}}, Y_{\text{train}}\}$ is employed to train the KNN classification, and the trained KNN predicts the labels of D_{test} . Accordingly, the predicted labels of D_{test} , termed as $Y_{\text{test}}^{\text{predict}}$, can be obtained. By comparing Y_{test} and $Y_{\text{test}}^{\text{predict}}$, the number of correctly predicted samples for labels $Y_{\text{test}}^{\text{correct}}$ is obtained. Based on the $Y_{\text{test}}^{\text{correct}}$ and Y_{test} , the expression of accuracy (q) is presented as follows:

$$\text{accuracy}(q) = \frac{Y_{\text{test}}^{\text{correct}}}{Y_{\text{test}}} \times 100\%. \quad (17)$$

The remaining features are also handled in the same way. Let accuracy (q) denote the classification accuracy of the q -th feature. Therefore, it can obtain the classification accuracy sequence, $\{\text{accuracy}(1), \text{accuracy}(2), \dots, \text{accuracy}(Q)\}$. In this study, we presume that the higher value of classification accuracy indicates the better fault discriminability.

3.2.2. Measure WCI of Feature Based on MMD and KLD. For a more comprehensive WCI evaluation of features, MMD and KLD are employed to evaluate the distribution difference between feature samples from SD and TD. The basic principle of MMD is introduced in Section 2.2. The details of KLD are described as follows [39].

KLD is an effective metric tool to estimate the distribution differences [40], and it is often applied in the fields of statistical learning, information technique, signal

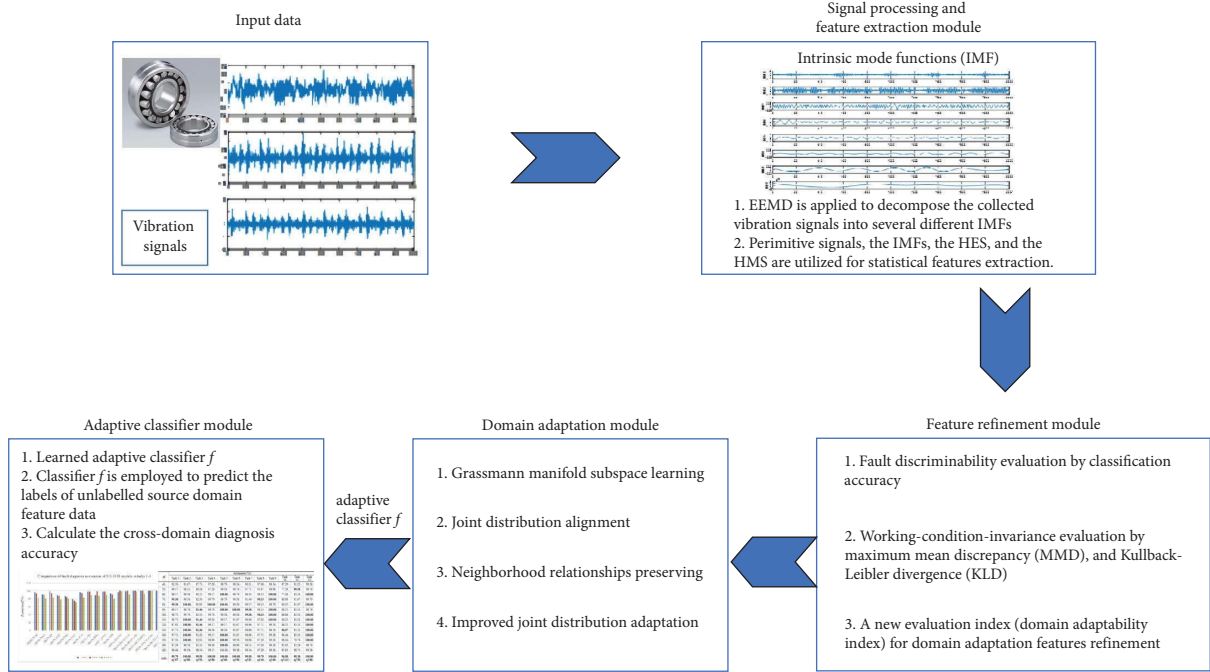


FIGURE 2: The architecture of the DFCD-IDAMN framework.

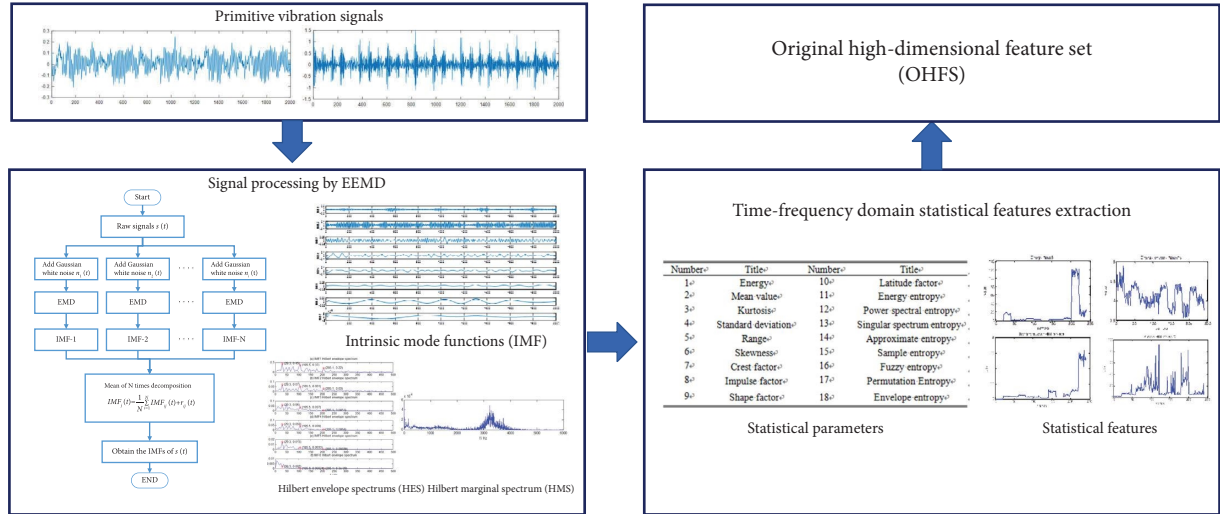


FIGURE 3: The architecture of signal processing and feature extraction module.

processing, etc. Given two probability density functions of two different variables as $\text{pro_}d_1$ and $\text{pro_}d_2$, the KLD is represented on the basis of the definition of information entropy.

$$I(\text{pro_}d_1 \parallel \text{pro_}d_2) = \int \text{pro_}d_1(x) \log \frac{\text{pro_}d_1(x)}{\text{pro_}d_2(x)} dx, \quad (18)$$

where the function $I(\bullet)$ has no symmetry, that is, $I(\text{pro_}d_1 \parallel \text{pro_}d_2) \neq I(\text{pro_}d_2 \parallel \text{pro_}d_1)$.

According to the references [39–41], the expression of KLD in symmetric form can be denoted as

$$K(\text{pro_}d_1, \text{pro_}d_2) = I(\text{pro_}d_1 \parallel \text{pro_}d_2) + I(\text{pro_}d_2 \parallel \text{pro_}d_1). \quad (19)$$

Based on the basic principles of MMD and KLD, given normal state feature sets $\text{OFS}_s^{\text{normal}}$ and $\text{OFS}_T^{\text{normal}}$ from source and target domains, respectively, $\text{OFS}_s^{\text{normal}}$ and $\text{OFS}_T^{\text{normal}}$ are expressed as follows:

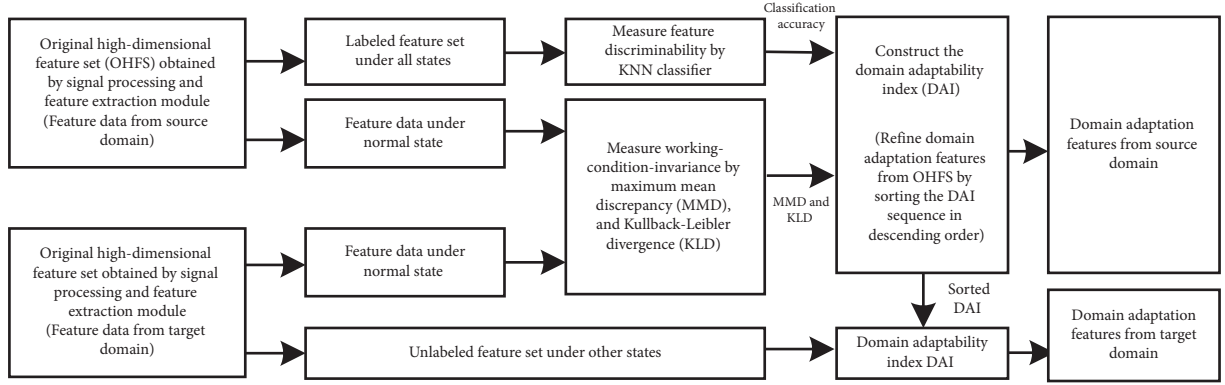


FIGURE 4: The process of DFCD.

$$\text{OFS}_s^{\text{normal}} = \begin{bmatrix} f_{s1}^1 & f_{s2}^1 & \cdots & f_{sM}^1 \\ f_{s1}^2 & f_{s2}^2 & \cdots & f_{sM}^2 \\ \vdots & \vdots & \ddots & \vdots \\ f_{s1}^Q & f_{s2}^Q & \cdots & f_{sM}^Q \end{bmatrix}, \quad (20)$$

$$\text{OFS}_T^{\text{normal}} = \begin{bmatrix} f_{T1}^1 & f_{T2}^1 & \cdots & f_{TM}^1 \\ f_{T1}^2 & f_{T2}^2 & \cdots & f_{TM}^2 \\ \vdots & \vdots & \ddots & \vdots \\ f_{T1}^Q & f_{T2}^Q & \cdots & f_{TM}^Q \end{bmatrix},$$

where f_{sM}^Q represents the q -th feature of the M -th sample from SD, f_{TM}^Q represents the q -th feature of the M -th sample from TD, and M is the number of normal state feature sample. The first row of $\text{OFS}_s^{\text{normal}}$ and $\text{OFS}_T^{\text{normal}}$ and the first feature data $[f_{s1}^1, f_{s2}^1, \dots, f_{sM}^1]$ and $[f_{T1}^1, f_{T2}^1, \dots, f_{TM}^1]$ are used to calculate the MMD and KLD, which can obtain the MMD and KLD of the first feature data between SD and TD. The remaining features data are also handled in the same way. Let $\text{mmd}(q)$ and $\text{kld}(q)$ denote the MMD and KLD of the q -th feature, respectively. Therefore, it can obtain the MMD sequence $\{\text{mmd}(1), \text{mmd}(2), \dots, \text{mmd}(Q)\}$ and MMD sequence $\{\text{kld}(1), \text{kld}(2), \dots, \text{kld}(Q)\}$. In this study, we presume that the WCI of feature is better when the sum of MMD and KLD is smaller.

3.2.3. Build the Domain Adaptability Index. According to the estimation of fault discriminability and WCI of features, based on the classification accuracy, MMD, and KLD, a new domain adaptability index, DAI, is proposed to assist refine domain adaptation features. For the n -th feature, the definition of DAI is presented as follows:

$$\text{DAI}(q) = \frac{\text{accuracy}(q)}{\theta \cdot \text{mmd}(q) + (1 - \theta) \cdot \text{kld}(q)}, \quad (21)$$

where θ is a trade-off parameter. Then, we can obtain the DAI sequence of Q features, $\text{DAI} = \{\text{DAI}(1), \text{DAI}(2), \dots, \text{DAI}(Q)\}$. In this work, it is supposed that the domain adaptability of feature is stronger when the corresponding value of DAI is higher. Accordingly, we can refine domain

adaptation features from OHFS by sorting the DAI sequence in descending order, and the features with high DAI values are used to form feature subset for domain adaptation.

3.3. Improved JDA with Manifold Subspace Learning and Neighborhood Relationship Preserving (IDAMN). Aiming at three significant issues of many existing DA approaches based on feature-based TL: (1) the implementation of distribution adaptation in most studies is based on the probability distributions alignment in the original complex and high-dimensional feature space, which is difficult to tackle the issue of feature distortion and may lead to the poor domain adaptation performance [28]. (2) The optimization goals of numerous ready-made DAs of TLM merely concentrate on decreasing the distribution differences and enhancing the transferability of features, and the class distinguishability of feature is usually neglected, which may lead to the poor classification performance [29, 30]. (3) In the process of distribution adaptation, the impact of class information and neighborhood relationships of feature data on distribution adaptation has not been effectively considered, which may degrade the CFD performance and generalization ability of the model [28, 30]. Therefore, in this section, on the basis of the idea that is joint distribution alignment with neighborhood relationship preserving in manifold subspace, a novel domain adaptation method, IDAMN, is designed. There are four steps of IDAMN. (1) Grassmann manifold subspace learning; (2) joint distribution alignment; (3) neighborhood relationships preserving; and (4) improved joint distribution adaptation. The details of IDAMN are presented as follows.

3.3.1. Grassmann Manifold Subspace Learning. This work applies the classical unsupervised manifold learning approach of the geodesic flow kernel (GFK) to learn low-dimensional manifold structure of feature set in original high-dimensional space [42]. Accordingly, some features with certain geometrical structures in the manifold subspace can be obtained, which can overcome the problem of feature distortions in the raw feature space [28, 43]. Given that the labeled feature dataset of SD and TD are, respectively,

expressed as X_S and X_T , then, the GFK is implemented to map original feature data X_S and X_T into Grassmann manifold (GM) space $G(d)$ by $Z = g(X) = \sqrt{GX}$ [20, 42], and the Z_S and Z_T can be obtained, respectively. The detailed introduction of GFK can be referred to [20, 42].

In particular, the prevailing subspace dimension of GFK must be set to less than half of the input feature space dimension. Therefore, aiming at the scenario that the input feature dimension is less than twice the set dimension of manifold subspace, before executing unsupervised manifold learning of GFK, it will conduct dimension size comparison and automatic adjustment. Specifically, if the feature dimension is less than twice the dimension of the set manifold subspace, the dimension of the manifold subspace will be set as the half of the feature dimension. Conversely, if the feature dimension is greater than twice the dimension of the set manifold subspace, GFK will be implemented under the set manifold subspace dimension.

3.3.2. Joint Distribution Alignment. In order to further shrink the distribution divergences between SD and TD, joint distribution alignment is introduced. It includes two parts: marginal distribution alignment (MDA) and conditional distribution alignment (CDA).

(1) *MDA.* Let Z_S and Z_T denote the representations of SD and TD data on the GM space, respectively. The corresponding marginal distributions of them are $P(Z_S)$ and $P(Z_T)$. The marginal distribution alignment is conducted by minimizing the MMD between $P(Z_S)$ and $P(Z_T)$ [17]. The expression of MMD between $P(Z_S)$ and $P(Z_T)$ is shown as follows:

$$\begin{aligned} \text{MMD}_H^2(P_S, P_T) &= \left\| \frac{1}{n_S} \sum_{z_i \in Z_S} \phi(z_i) - \frac{1}{n_T} \sum_{z_j \in Z_T} \phi(z_j) \right\|_H^2 \\ &= \text{tr}(W^T Z L_0 Z^T W), \end{aligned} \quad (22)$$

where H represents the RKHS. $\text{tr}(W^T Z L_0 Z^T W)$ represents the trace of $W^T Z L_0 Z^T W$, W is optimal transformation matrix, and Z denotes the input feature data matrix composed of Z_S and Z_T . The definition of matrix L_0 is presented as follows:

$$L_0 = \begin{cases} \frac{1}{n_S^2}, & (z_i, z_j \in Z_S), \\ \frac{1}{n_T^2}, & (z_i, z_j \notin Z_S), \\ \frac{-1}{n_S n_T}, & (\text{otherwise}), \end{cases} \quad (23)$$

where n_S and n_T are the number of Z_S and Z_T , respectively. By minimizing equation (22), a new representation $W^T Z$

can be obtained to achieve that the marginal distribution discrepancies between the SD and TD are narrowed.

(2) *CDA.* The CDA is conducted by minimizing the MMD between conditional distributions $Q_S(Y_S | Z_S)$ and $Q_T(Y_T | Z_T)$ [18]. Aiming at the lack of Y_T , it utilizes base classifier f trained on the Z_S with Y_S , the pseudo labels \hat{Y}_T of the TD data Z_T can be easily predicted by f [18]. Due to that, the $Q_S(Y_S | Z_S)$ and $Q_T(Y_T | Z_T)$ are posterior probabilities and quite involved, it can explore the sufficient statistics of $Q_S(Z_S | Y_S = c)$ and $Q_T(Z_T | Y_T = c)$ instead [44]. c is the category in the label set and $c \in [1, 2, \dots, C]$ (C is the total number of categories) [18]. Therefore, the MMD between the $Q_S(Z_S | Y_S = c)$ and $Q_T(Z_T | Y_T = c)$ can be expressed as follows:

$$\begin{aligned} \sum_{c=1}^C \text{MMD}_H^2(Q_S^{(c)}, Q_T^{(c)}) &= \sum_{c=1}^C \left\| \frac{1}{n_S^{(c)}} \sum_{z_i \in Z_S^{(c)}} \phi(z_i) - \frac{1}{n_T^{(c)}} \sum_{z_j \in Z_T^{(c)}} \phi(z_j) \right\|_H^2 \\ &= \sum_{c=1}^C \text{tr}(W^T Z L_c Z^T W), \end{aligned} \quad (24)$$

where $Z_S^{(c)} = \{z_i: z_i \in Z_S \wedge y(z_i) = c\}$ and $Z_T^{(c)} = \{z_i: z_i \in Z_T \wedge \hat{y}(z_i) = c\}$ are, respectively, feature sets pertaining to class c . $\hat{y}(z_i)$ is the pseudo tag of the TD data z_i . $n_S^{(c)}$ and $n_T^{(c)}$ are the number of samples pertaining to class c , respectively. Accordingly, the MMD matrix L_c can be obtained by the following equation:

$$L_c = \begin{cases} \frac{1}{n_S^{(c)} n_S^{(c)}}, & (z_i, z_j \in Z_S^{(c)}), \\ \frac{1}{n_T^{(c)} n_T^{(c)}}, & (z_i, z_j \in Z_T^{(c)}), \\ \frac{-1}{n_S^{(c)} n_T^{(c)}}, & \begin{cases} z_i \in Z_S^{(c)}, z_j \in Z_T^{(c)}, \\ z_i \in Z_T^{(c)}, z_j \in Z_S^{(c)}, \end{cases} \\ 0, & (\text{otherwise}). \end{cases} \quad (25)$$

When the minimum of equation (24) is achieved, a new representation $W^T Z$ can be obtained to achieve that the conditional distribution discrepancies between $Q_S(Y_S | Z_S)$ and $Q_T(Y_T | Z_T)$ are narrowed.

3.3.3. Neighborhood Relationships Preserving. In order to consider the impact of class information and neighborhood relationships of feature data in the process of distribution adaptation, inspired by the principles of LDA [45] and LFDA [37], a new local minimum margin criterion matrix (LMMCM) is designed to utilize the label information while preserving the local neighborhood geometry of the feature data. The expression of LMMCM is presented as follows:

$$\text{LMMCM} = S_w^L - S_b^L, \quad (26)$$

where S_w^L and S_b^L are local WSM and local BSM. The S_w^L and S_b^L are expressed as follows:

$$S_w^L = \frac{1}{2} \sum_{i,j=1}^n \tilde{p}_{ij}^{LW} (z_i - z_j)(z_i - z_j)^T, \quad (27)$$

$$S_b^L = \frac{1}{2} \sum_{i,j=1}^n \tilde{p}_{ij}^{Lb} (z_i - z_j)(z_i - z_j)^T, \quad (28)$$

where

$$\tilde{p}_{ij}^{Lb} = \begin{cases} A_{ij} \left(\frac{1}{n} - \frac{1}{n_l} \right), & z_i = z_j = l, \\ \frac{n_l}{n}, & z_i \neq z_j (j \in \text{Nst}(i)), \\ \frac{1}{n}, & \text{else,} \end{cases} \quad (29)$$

$$\tilde{p}_{ij}^{LW} = \begin{cases} \frac{A_{ij}}{n_l}, & z_i = z_j = l, \\ 0, & z_i \neq z_j, \end{cases} \quad (30)$$

where n , l , and n_l are, respectively, the number of feature sample, class label of feature sample, and the number of feature samples that belongs to the class l . \tilde{p}_{ij}^{Lb} and \tilde{p}_{ij}^{LW} constitute weight matrices. In \tilde{p}_{ij}^{Lb} , the meaning of $z_i \neq z_j (j \in \text{Nst}(i))$ is that j is the nearest neighbor of i and they pertain to different classes. $A_{ij} \in [0, 1]$ is defined as follows:

$$A_{ij} = \exp \left(- \frac{\|z_i - z_j\|^2}{\gamma_i \gamma_j} \right), \quad (31)$$

where $\gamma_i = \|z_i - z_i^m\|$ represents the local scaling around z_i and z_i^m is the m -th nearest neighbor of z_i . When z_i and z_j are closer, the A_{ij} is larger, if not, the A_{ij} is smaller. By introducing the LMMCM, the local neighborhood geometry of the feature data, including the neighborhood relationships between data of the same category and the neighborhood relationships between data of different classes, can be considered. Furthermore, the class label information is effectively introduced, and it can improve the discriminability of feature data by minimizing the LMMCM.

3.3.4. Improved Joint Distribution Adaptation. On the basis of the above three contents, we design an improved joint distribution adaptation, $D_{\text{IJDA}}(Z_S, Z_T)$; it is defined as follows:

$$D_{\text{IJDA}}(Z_S, Z_T) = \beta \sum_{c=1}^C \text{MMD}_H^2(Q_S^{(c)}, Q_T^{(c)}) + (1 - \beta) \text{MMD}_H^2(P_S, P_T) + \eta \text{tr}(W^T K(S_w^L - S_b^L)KW), \quad (32)$$

where $\beta \in [0, 1]$ and $\eta \in [0, 1]$ are adjustable parameters and β tunes the proportion of the marginal and conditional distributions adaptation. According to equations (22) and (24), the $D_{\text{IJDA}}(Z_S, Z_T)$ can be further expressed as follows:

$$D_{\text{IJDA}}(Z_S, Z_T) = \beta \sum_{c=1}^C \text{tr}(W^T ZL_c Z^T W) + (1 - \beta) \text{tr}(W^T ZL_0 Z^T W) + \eta \text{tr}(W^T Z(S_w^L - S_b^L)Z^T W). \quad (33)$$

According to the optimization objective of JDA and equation (26), the optimization goal of IDAMN can be defined as

$$\begin{aligned} \min_W \beta, & \sum_{c=1}^C \text{tr}(W^T ZL_c Z^T W) \\ & + (1 - \beta) \text{tr}(W^T ZL_0 Z^T W) \\ & + \eta \text{tr}(W^T Z(S_w^L - S_b^L)Z^T W) + \lambda \|W\|_F^2, \end{aligned} \quad (34)$$

s.t., $W^T ZEZ^T W = I,$

where λ is the regularization parameter with $\|\cdot\|_F^2$ the Frobenius norm and $\lambda \|W\|_F^2$ is used to ensure the optimization problem to be well-defined. $I \in \mathbb{R}^{(n_S+n_T) \times (n_S+n_T)}$ and E represent the unit matrix and centering matrix, respectively. $E = \mathbf{I} - (1/(n_S + n_T))\mathbf{1}$, and $\mathbf{1}$ is the $(n_S + n_T) \times (n_S + n_T)$ matrix of ones. For the solution of equation (34), based on the constrained optimization theory, set Lagrange multipliers $\Phi = \text{diag}(\phi_1, \phi_2, \dots, \phi_k) \in \mathbb{R}^{k \times k}$; accordingly, the Lagrange function for solving equation (34) is as follows:

$$L = \text{tr} \left(W^T Z \left(\beta \sum_{c=1}^C L_c + (1 - \beta)L_0 + \eta(S_w^L - S_b^L) \right) Z^T W \right) + \lambda \|W\|_F^2 + \text{tr}((I - W^T ZEZ^T W)\Phi). \quad (35)$$

By setting derivative $\partial L / \partial W = 0$, the solution of equation (34) can be derived as a generalized eigendecomposition problem as follows:

$$\left(Z \left(\beta \sum_{c=1}^C L_c + (1 - \beta)L_0 + \eta(S_w^L - S_b^L) \right) Z^T + \lambda I \right) W = ZEZ^T W \Phi. \quad (36)$$

According to equation (36), finally, the optimal adaptation matrix W is built by using the k smallest eigenvectors, and new feature representations $U_S = W^T Z_S$ and $U_T = W^T Z_T$ are obtained. Then, it can use labeled U_S to learn an adaptive classifier f , and the learned adaptive classifier f is employed to predict the label of unlabeled U_T .

In summary, the overall complete procedures of IDAMN are presented as follows:

- (1) Input: Source and target domains feature set X_S and X_T , true labels Y_S of X_S , manifold subspace dimension d , regularization parameters λ , β , and η , and dimension of output source and target domains

feature space k . The iteration is i . The dimension of X_S or X_T is d_f . When the $d_f < 2 \times d$, the manifold subspace dimension d will be set as $0.5 \times d_f$.

- (2) By equation (28), learn the Grassmann manifold transformation kernel G to transform the original feature data (X_S and X_T) into $G(d)$ with $Z = g(X) = \sqrt{G}X$. Accordingly, the new source domain Z_S and new target domain Z_T are obtained.
- (3) Learn a base classifier on Z_S and conduct prediction on Z_T to obtain its pseudo labels \hat{Y}_T .
- (4) Constitute $Z = [Z_S, Z_T]$; compute L_0 and L_c by equations (23) and (25). Compute S_w^L and S_b^L by equations (27) and (28).
- (5) Solve the eigendecomposition problem in equation (36) and use k smallest eigenvectors to form adaptation matrix W , and $U = [U_S, U_T] = W^T Z = [W^T Z_S, W^T Z_T]$.
- (6) Train an adaptive classifier f on $\{W^T Z_S, Y_S\}$ and update the pseudo labels \hat{Y}_T of target domain data, $\hat{Y}_T = f(W^T Z_T)$.
- (7) Construct the MMD matrices $\{L_c\}_{c=1}^C$ by equation (25).
- (8) Repeat the step (4) until the iteration i .
- (9) Output the learned adaptive classifier f .

3.4. Complete Process of the Cross-Domain Fault Diagnosis Based on the DFCD-IDAMN. Based on the DFCD-IDAMN framework and cross-domain fault diagnosis tasks, the complete process is described in detail as follows:

- (1) Input collected fault vibration signals under a specific working condition and unknown working condition, and denote S_{source} and S_{target} , respectively. S_{source} and S_{target} represent source and target domains data, respectively. The class information of S_{source} is known, but the class information of S_{target} is unknown.
- (2) S_{source} and S_{target} are decomposed into several different IMFs by EEMD, respectively. Then, these IMFs are utilized to calculate HES and HMS. On the basis of raw vibration signals S_{source} and S_{target} , the IMFs, the corresponding HES and HMS, it calculates statistical parameters of them to obtain the corresponding statistical features, and the high-dimensional statistical feature sets $\text{OFS}_{\text{source}}$ and $\text{OFS}_{\text{target}}$ of source and target domains are built.
- (3) $\text{OFS}_{\text{source}}$ and $\text{OFS}_{\text{target}}$ are inputted in the DFCD module, and set parameter in this step, the fault discriminability evaluation of features, and the WCI measurement of features are conducted, which can obtain the domain adaptability index of features for refining $\text{OFS}_{\text{source}}$ and $\text{OFS}_{\text{target}}$. Therefore, the new feature sets of source and target domains X_S and X_T are obtained for the subsequent step.

- (4) Input: Source and target domains feature set X_S and X_T , true labels Y_S of X_S , manifold subspace dimension d , regularization parameters λ, β , and η , and dimension of output source and target domains feature space k . The iteration is i . On this basis, the proposed IDAMN is performed; accordingly, new feature sets U_S, U_T and adaptive classifier f are obtained. Finally, the cross-domain fault diagnosis accuracy is calculated.

4. Experimental Verification

In this work, for validating the performance and superiority of the proposed methods, two bearing fault datasets, obtained from the Case Western Reserve University (CWRU) test platform [3, 6, 29, 31, 46–48] and the SQI-MFS test platform [29, 31, 44, 46] are employed for a set of case studies. To clearly illustrate the superiority of the proposed methods (DFCD and IDAMN), some comparative models are built by ready-made common methods: KNN, SVM, DAE, CNN, DBN, JDA, TCA, JGSA, BDA, and GFK.

4.1. Case 1: Fault Diagnostic of Bearing Dataset 1 across Different Working Loads

4.1.1. Description of Bearing Dataset and Fault Diagnosis Tasks. In case 1, it utilizes the bearing vibration dataset gained from the CWRU test-bed to conduct CFD experiments. The test platform is presented in Figure 5. This bearing vibration signals are sampled through acceleration sensors under 12 kHz sampling frequency. Table 1 lists the description of bearing vibration dataset. There are three categories of bearing defect: inner raceway defect (IRD), ball defect (BD), and outer raceway defect (ORD). The defect parameters include 0.028 inch, 0.021 inch, 0.014 inch, and 0.007 inch. Moreover, vibration data for bearings without defects is also used. In order to set CFD tasks, bearing data under motor loads of 0 hp, 1 hp, 2 hp, and 3 hp are chosen for experiments. Therefore, it can obtain bearing vibration data of 12 classes, labeled 1–12. For each class, 60 samples are used to build a training set and a testing set, and it randomly divides 20 and 40 samples as training and testing samples. Each sample is composed of 2000 continuous data points from original vibration signals. Based on the bearing data presented in Table 1, 12 CFD tasks are arranged, as listed in Table 2.

4.1.2. Diagnosis Results of the DFCD-IDAMN Framework. In this section, according to the overall procedure of the DFCD-IDAMN framework shown in Figure 2, it first conducts signal processing and feature extraction, and primitive signals are decomposed into several IMFs by EEMD. Although the obtained IMFs are distributed from high frequency to low frequency by default, it is not that each IMF can represent the time-frequency feature of a fault signal with effect [49]. For this issue, according to related previous research work [49–51], correlation coefficient between each IMF and raw vibration signal is

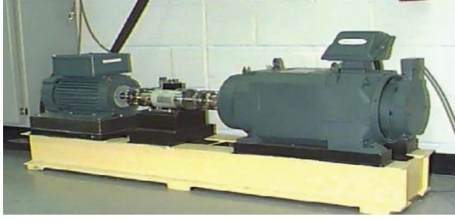


FIGURE 5: CWRU test platform.

utilized to reduce redundant IMFs. The IMF is more closely related to the original vibration signal and has richer time-frequency information when the value of the correlation coefficient is higher. Therefore, in this work, we refer to the literature [49]; the first four IMFs are used for feature extraction; furthermore, four Hilbert envelope spectrums (HES) of four IMFs and one Hilbert marginal spectrum (HMS) calculated from four IMFs are also used to generate statistical features. Accordingly, it can obtain 4 IMFs, 4 HES, and 1 HMS from a vibration signal, then calculate 18 statistical parameters [29, 31, 44, 52–55] of them listed in Table 3, from which 162 statistical features can be extracted to form the original high-dimensional feature set. Vibration signal samples of no defect bearing and inner raceway defect under motor loads of 0 hp, 1 hp, 2 hp, and 3 hp are presented in Figure 6. The corresponding IMFs from these samples are presented in Figures 7 and 8.

Secondly, it carries out the feature refinement module. The proposed DFCD evaluates the fault distinguishability and WCI of 162 statistical features, which obtains the DAI of them and helps to refine features with better domain adaptability from high-dimensional original feature set. Take the no defect vibration data under motor load of 0 hp as an example. Figure 9 presents the DAI of 162 statistical features. From the figure, it can be seen that different features have different DAI values, and it indicates the different domain adaptability quantification results of different features. For the 39th and 42nd features, their DAI values are significantly higher than other features, and it shows that their domain adaptability is more prominent. Therefore, in this study, we assume that the higher DAI value indicates the greater domain adaptability. Therefore, the DFCD can help to refine some features (they are more advantageous to domain adaptation) by manually select a threshold of the DAI value, and these refined features are processed by the subsequent domain adaptation module.

Next, the refined features obtained by performing the feature refinement module constitute a cross-domain adaptation feature set (CDAF), and the labeled CDAF of the SD and the unlabeled CDAF of the TD are inputted into the proposed IDAMN domain adaptation method, achieving the joint distribution alignment with neighborhood relationship preserving is performed in Grassmann manifold subspace, and learning an adaptive classifier f for CFD. Finally, the learned classifier f is learned and it can predict the labels of the target domain feature set; therefore, the CFD result can be calculated.

After performing the above steps, the experimental results of 12 CFD tasks are listed in Table 4. It shows the mean diagnosis accuracies of 12 bearing defect types under different numbers of domain adaptation features (nf). According to the diagnosis accuracies of these 12 CFD tasks, it can easily conclude the following analysis. Firstly, the proposed DFCD-IDAMN framework for CFD of bearings can achieve ideal fault diagnosis result. The diagnosis accuracies of tasks 2, 4, 5, 6, 9, and 12 can reach 100% with the suitable nf . Tasks 1, 3, and 7 can attain over 99.5% diagnosis accuracy. Accordingly, the effectiveness of the DFCD-IDAMN framework can be validated. Secondly, it is evident that the use of the proposed DFCD has an apparent effect on the fault diagnosis accuracy. Without using DFCD, all of 162 features are utilized for the subsequent IDAMN domain adaptation method and CFD, the diagnosis result is not ideal. The diagnosis accuracies of tasks 1–12 are 96.46%, 99.58%, 83.33%, 99.17%, 100.00%, 89.58%, 98.54%, 97.29%, 99.38%, 95.83%, 82.29%, and 99.58%, respectively. When the DFCD is applied and the refined CDAF is employed for the subsequent procedure, it can attain desirable CFD accuracies that are apparently higher than that of diagnosis without using DFCD. The maximum accuracies (mda) of 12 CFD tasks are 99.79%, 100.00%, 99.58%, 100.00%, 100.00%, 100.00%, 99.58%, 98.54%, 100.00%, 96.88%, 91.67%, and 100.00%, respectively. Therefore, the effectiveness of the DFCD with a suitable nf for improving fault diagnosis accuracy can be verified. The above CFD experiment involves some parameters of DFCD and IDAMN that need to be manually chosen. For the basis for setting hyperparameters of the proposed methods, the specific values of these parameters are set based on experimental experience. Therefore, we directly present the relevant parameter values in this manuscript. For the DFCD, the corresponding parameters set in DFCD include trade-off parameter $\theta = 0.5$. The parameters set in IDAMN include manifold subspace dimension $d = 50$, regularization parameters $\lambda = 0.1$, $\beta = 0.3$, and $\eta = 0.5$, dimension of output source and target domains feature space $k = 20$. Iterations $i = 10$. In particular, although the manifold subspace dimension is set as 50, when the feature dimension after the proposed feature refinement (that is nf) is less than twice of the set manifold subspace dimension, the manifold subspace dimension will be automatic adjusted as the half of nf . In Table 4, when the nf is 40, 50, 60, 70, 80, and 90, the manifold subspace dimension will be automatic adjusted as 20, 25, 30, 35, 40, and 45. On the contrary, when nf is not less than twice of the set manifold subspace dimension (when the nf is 100 to 162), the GFK is implemented under the set manifold subspace dimension 50.

4.1.3. Comparative Analysis with Other Fault Diagnosis Models. In an effort to further validate the advantages of the DFCD-IDAMN framework for CFD, some common and competitive approaches are used to conduct a series of comparison experiments, these methods include KNN, SVM, DAE, CNN, DBN, JDA, TCA, JGSA, BDA, and GFK. The reason of this set up is as follows: (1) it choses three

TABLE 1: The CWRU bearing data for experiments.

| Category of bearing defect | Defect diameter (inches) | Number of training/testing samples | | | Class label |
|----------------------------|--------------------------|------------------------------------|------------------|------------------|-------------|
| | | Motor loads 0 hp | Motor loads 1 hp | Motor loads 2 hp | |
| No defect | 0 | 20/40 | 20/40 | 20/40 | 1 |
| Inner raceway defect (IRD) | 0.007 | 20/40 | 20/40 | 20/40 | 2 |
| | 0.014 | 20/40 | 20/40 | 20/40 | 3 |
| | 0.021 | 20/40 | 20/40 | 20/40 | 4 |
| | 0.028 | 20/40 | 20/40 | 20/40 | 5 |
| Outer raceway defect (BD) | 0.007 | 20/40 | 20/40 | 20/40 | 6 |
| | 0.014 | 20/40 | 20/40 | 20/40 | 7 |
| | 0.021 | 20/40 | 20/40 | 20/40 | 8 |
| Ball defect (ORD) | 0.007 | 20/40 | 20/40 | 20/40 | 9 |
| | 0.014 | 20/40 | 20/40 | 20/40 | 10 |
| | 0.021 | 20/40 | 20/40 | 20/40 | 11 |
| | 0.028 | 20/40 | 20/40 | 20/40 | 12 |

TABLE 2: The CFD tasks for case 1.

| Tasks | Training samples (SD) | | | Testing samples (TD) | | |
|-------|-----------------------|-------------------------|-------------------|----------------------|-------------------------|-------------------|
| | Motor load (hp) | Defect types of samples | Number of samples | Motor load (hp) | Defect types of samples | Number of samples |
| 1 | 0 | Classes 1–12 | 240 | 1 | Classes 1–12 | 480 |
| 2 | 0 | Classes 1–12 | 240 | 2 | Classes 1–12 | 480 |
| 3 | 0 | Classes 1–12 | 240 | 3 | Classes 1–12 | 480 |
| 4 | 1 | Classes 1–12 | 240 | 0 | Classes 1–12 | 480 |
| 5 | 1 | Classes 1–12 | 240 | 2 | Classes 1–12 | 480 |
| 6 | 1 | Classes 1–12 | 240 | 3 | Classes 1–12 | 480 |
| 7 | 2 | Classes 1–12 | 240 | 0 | Classes 1–12 | 480 |
| 8 | 2 | Classes 1–12 | 240 | 1 | Classes 1–12 | 480 |
| 9 | 2 | Classes 1–12 | 240 | 3 | Classes 1–12 | 480 |
| 10 | 3 | Classes 1–12 | 240 | 0 | Classes 1–12 | 480 |
| 11 | 3 | Classes 1–12 | 240 | 1 | Classes 1–12 | 480 |
| 12 | 3 | Classes 1–12 | 240 | 2 | Classes 1–12 | 480 |

TABLE 3: 18 statistical parameters.

| Number | Title |
|--------|---------------------------|
| 1 | Energy |
| 2 | Mean value |
| 3 | Kurtosis |
| 4 | Standard deviation |
| 5 | Range |
| 6 | Skewness |
| 7 | Crest factor |
| 8 | Impulse factor |
| 9 | Shape factor |
| 10 | Latitude factor |
| 11 | Energy entropy |
| 12 | Power spectral entropy |
| 13 | Singular spectrum entropy |
| 14 | Approximate entropy |
| 15 | Sample entropy |
| 16 | Fuzzy entropy |
| 17 | Permutation entropy |
| 18 | Envelope entropy |

categories methods: classical machine learning methods, classical deep learning methods, and classical transfer learning methods, which are used to compare the effectiveness differences between them. (2) KNN and SVM are classic classifiers that have been widely used and are very representative. (3) DAE, CNN, and DBN are widely developed and studied classical deep learning approaches. (4) JDA, TCA, JGSA, BDA, and GFK are representative transfer learning methods that have gradually received attention and study from many researchers in recent years.

Table 5 presents comparative models built by these methods, DFCD and IDAMN. These comparative models are labeled as M1–M18 and can be divided into three types. (1) The models are not combined with domain adaptation methods, and they only utilize the original high-dimensional feature set (OHFS) and classical classifiers. Take M1 as an example; it is a classical classifier-based model, and the OHFS is directly inputted in the SVM classifier for cross-domain fault diagnosis. (2) The models are combined with domain adaptation methods, and they use the OHFS, domain adaptation methods, and base

classifier. Take M7 as an example, it is a domain adaptation-based model. The OHFS is firstly processed by TCA, and the output features are inputted in the KNN classifier. (3) The models are combined with DFCD and domain adaptation methods, and they use the OHFS, DFCD, domain adaptation methods, and base classifier. Take M13 as an example, the OHFS is firstly refined by the proposed DFCD, then, the refined features are processed by TCA, and finally the output features are inputted in the KNN classifier.

The fault diagnosis results of M1–M18 models are shown in Table 6 and Figures 10–14. It is obvious that the M18 model obtained by the proposed DFCD-IDAMN framework can achieve the better CFD performance than other comparative models. The detailed comparative analysis can be easily drawn as follows. (1) Compared with the M1–M6 (base classifier-based models), the fault diagnosis accuracies of tasks 1–12 of DFCD-IDAMN model are remarkably higher than that of M1–M6 models. In Figure 14, the mean fault diagnosis accuracy of 12 tasks of DFCD-IDAMN model can reach 99.57%, which is respectively 8.11%, 8.69%, 7.36%, 12.32%, 13.37%, and 18.01% higher than M1–M6 models. (2) Comparing OHFS-IDAMN (M12) model with M7–M11 models (domain adaptation-based models), the accuracies of 12 tasks are noticeably higher than M7–M11 models. Accordingly, the DA ability of the IDAMN outperforms traditional TCA, JDA, BDA, JGSA, and GFK. (3) Comparing M7–M12 (domain adaptation-based models without DFCD) with M13–M18 (domain adaptation-based models with DFCD), it is easily found that the use of the DFCD has a significant enhancement on the fault diagnosis accuracy of domain adaptation-based model, take OHFS-TCA (M7) and OHFS-DFCD-TCA (M13) as examples, the diagnosis accuracies of tasks 1–12 of M13 model are, respectively, 98.75%, 99.79%, 90.63%, 97.92%, 100.00%, 97.92%, 97.50%, 96.67%, 99.38%, 89.79%, 97.92%, and 100.00%, which surpasses that of the M7 model. Therefore, it implies that the DFCD can help to refine features with strong domain adaptability, which can effectively strengthen DA performance and increase fault diagnosis accuracy.

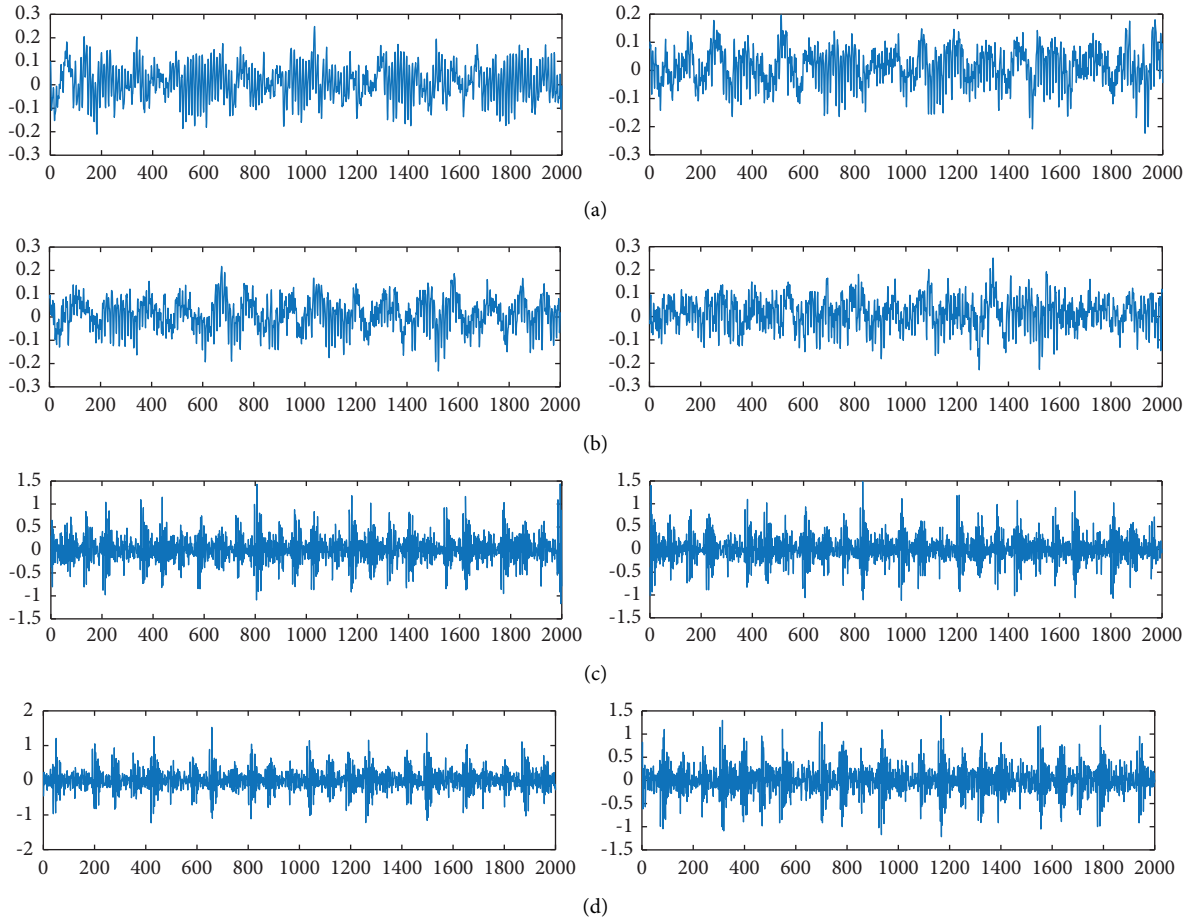


FIGURE 6: Vibration signal samples of no defect bearing and inner raceway defect bearing. (a) A vibration signal sample of no defect bearing under motor loads of 0 hp and 1 hp. (b) A vibration signal sample of no defect bearing under motor loads of 2 hp and 3 hp. (c) A vibration signal sample of inner raceway defect bearing under motor loads of 0 hp and 1 hp. (d) A vibration signal sample of inner raceway defect bearing under motor loads of 2 hp and 3 hp.

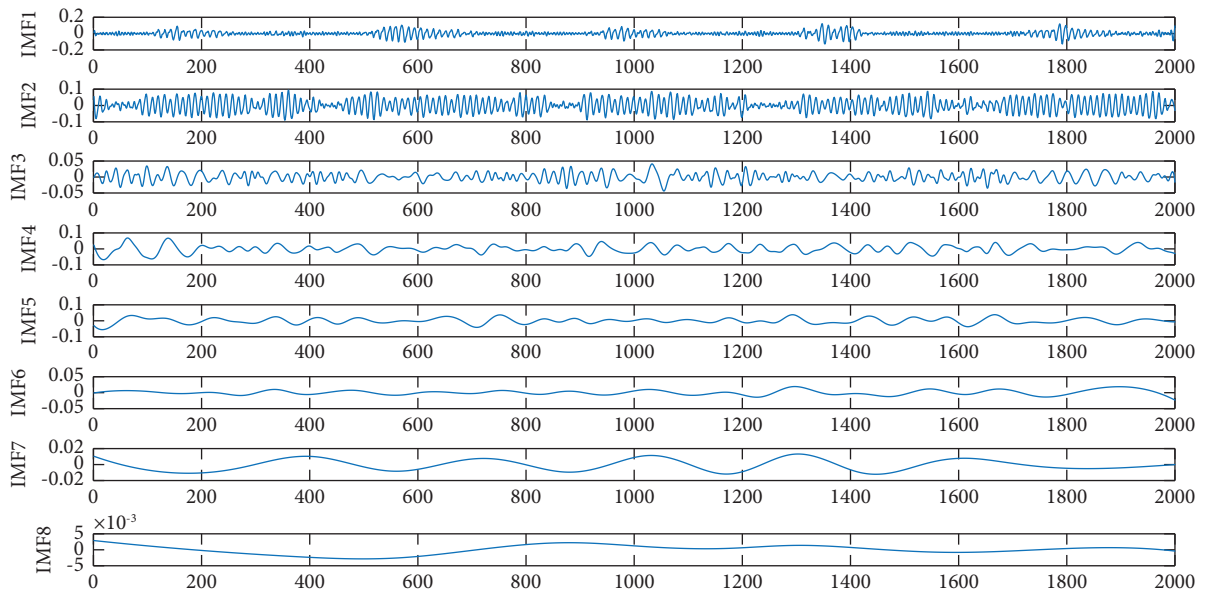


FIGURE 7: IMF1-IMF8 of no defect bearing vibration signal under motor load of 0 hp.

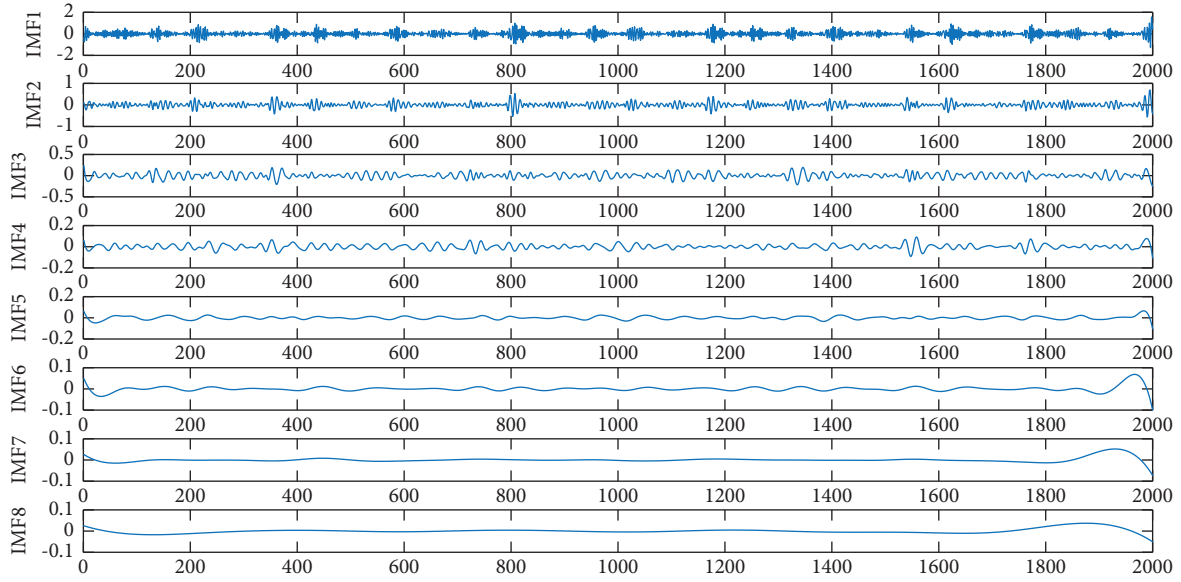


FIGURE 8: IMF1-IMF8 of inner raceway defect bearing vibration signal under motor load of 0 hp.

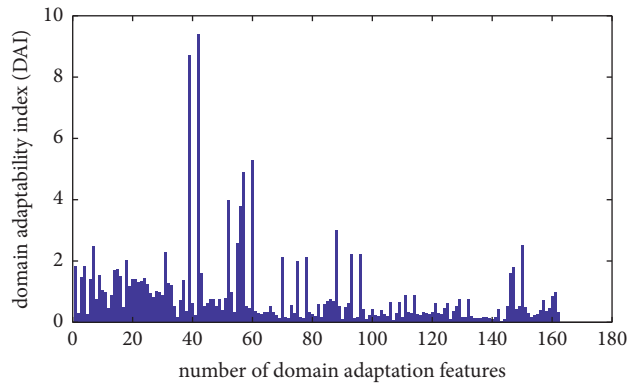


FIGURE 9: DAI of 162 statistical features that extracted from no defect bearing vibration data under motor load of 0 hp (trade-off parameter θ is 0.5).

TABLE 4: CFD results obtained by DFCD-IDAMN framework in case 1.

| nf | Accuracies (%) | | | | | | | | | | | |
|-------|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|---------------|---------------|
| | Task 1 | Task 2 | Task 3 | Task 4 | Task 5 | Task 6 | Task 7 | Task 8 | Task 9 | Task 10 | Task 11 | Task 12 |
| 40 | 92.50 | 91.67 | 87.71 | 97.29 | 99.79 | 98.54 | 90.21 | 97.08 | 98.54 | 87.29 | 91.25 | 99.38 |
| 50 | 99.17 | 98.33 | 89.58 | 97.29 | 99.58 | 99.79 | 97.71 | 91.67 | 98.96 | 77.29 | 99.38 | 99.79 |
| 60 | 99.17 | 99.38 | 90.21 | 99.17 | 100.00 | 99.79 | 98.33 | 98.13 | 100.00 | 77.08 | 83.33 | 100.00 |
| 70 | 99.38 | 99.58 | 82.50 | 99.79 | 99.79 | 99.58 | 91.46 | 98.33 | 100.00 | 86.88 | 91.67 | 99.79 |
| 80 | 99.38 | 100.00 | 90.83 | 100.00 | 100.00 | 99.58 | 99.17 | 98.13 | 99.79 | 86.25 | 91.67 | 100.00 |
| 90 | 99.17 | 99.79 | 91.46 | 99.79 | 100.00 | 100.00 | 99.38 | 98.13 | 100.00 | 86.25 | 83.33 | 99.79 |
| 100 | 98.75 | 99.79 | 83.33 | 99.79 | 99.58 | 89.58 | 99.38 | 98.33 | 100.00 | 86.88 | 83.33 | 100.00 |
| 110 | 98.75 | 100.00 | 91.46 | 99.38 | 99.17 | 91.67 | 98.96 | 97.92 | 100.00 | 86.25 | 83.33 | 100.00 |
| 120 | 97.92 | 100.00 | 91.46 | 99.17 | 99.17 | 91.67 | 98.96 | 97.71 | 99.58 | 86.25 | 83.33 | 100.00 |
| 130 | 97.71 | 100.00 | 91.46 | 99.38 | 99.38 | 91.67 | 98.96 | 97.71 | 99.58 | 96.67 | 83.33 | 100.00 |
| 140 | 97.71 | 100.00 | 91.25 | 99.17 | 100.00 | 91.67 | 98.96 | 97.71 | 99.38 | 96.46 | 83.33 | 100.00 |
| 150 | 97.50 | 100.00 | 82.92 | 98.96 | 100.00 | 99.58 | 98.96 | 97.29 | 99.38 | 96.04 | 79.79 | 100.00 |
| 160 | 97.29 | 99.58 | 83.33 | 99.38 | 100.00 | 88.96 | 98.13 | 97.08 | 99.38 | 95.83 | 82.29 | 99.79 |
| 162 | 96.46 | 99.58 | 98.54 | 99.17 | 100.00 | 99.38 | 98.54 | 97.29 | 99.38 | 95.83 | 98.75 | 99.58 |
| mda | 99.79 | 100.00 | 99.58 | 100.00 | 100.00 | 100.00 | 99.58 | 99.79 | 100.00 | 96.88 | 99.38 | 100.00 |
| | nf: 67 | nf: 80 | nf: 92 | nf: 80 | nf: 60 | nf: 90 | nf: 81 | nf: 46 | nf: 60 | nf: 131 | nf: 50 | nf: 60 |

The bold values highlight that the experimental results are desirable.

TABLE 5: Comparative models.

| The model without domain adaptation method | Label | The model with domain adaptation method | Label | The model with DFCD and domain adaptation method | Label |
|--|-------|---|-------|--|-------|
| OHFS-SVM | M1 | OHFS-TCA | M7 | OHFS-DFCD-TCA | M13 |
| OHFS-KNN | M2 | OHFS-JDA | M8 | OHFS-DFCD-JDA | M14 |
| OHFS-RF | M3 | OHFS-BDA | M9 | OHFS-DFCD-BDA | M15 |
| OHFS-DAE | M4 | OHFS-JGSA | M10 | OHFS-DFCD-JGSA | M16 |
| OHFS-DBN | M5 | OHFS-GFK | M11 | OHFS-DFCD-GFK | M17 |
| OHFS-CNN | M6 | OHFS-IDAMN | M12 | DFCD-IDAMN | M18 |

TABLE 6: CFD results of M1–M18 comparative models in case 1.

| Models | Accuracies (%) | | | | | | | | | | | |
|--------|----------------|---------------|---------------|----------------|----------------|----------------|----------------|---------------|----------------|----------------|----------------|---------------|
| | Task 1 | Task 2 | Task 3 | Task 4 | Task 5 | Task 6 | Task 7 | Task 8 | Task 9 | Task 10 | Task 11 | Task 12 |
| M1 | 96.46 | 93.13 | 81.25 | 98.33 | 99.79 | 88.75 | 91.46 | 94.17 | 96.88 | 83.75 | 83.33 | 94.79 |
| M2 | 91.25 | 89.79 | 81.04 | 95.63 | 99.79 | 87.50 | 93.54 | 92.71 | 99.17 | 84.38 | 83.13 | 92.92 |
| M3 | 98.75 | 94.38 | 82.71 | 97.08 | 99.58 | 96.88 | 91.25 | 92.08 | 99.38 | 83.75 | 84.38 | 92.29 |
| M4 | 89.38 | 87.29 | 78.33 | 86.25 | 93.96 | 91.67 | 88.54 | 90.00 | 92.50 | 79.79 | 80.83 | 90.42 |
| M5 | 87.29 | 84.58 | 80.00 | 88.54 | 91.04 | 89.17 | 88.33 | 90.63 | 88.75 | 76.25 | 79.79 | 91.04 |
| M6 | 78.96 | 74.79 | 71.46 | 79.17 | 82.50 | 80.42 | 87.08 | 91.25 | 89.38 | 78.13 | 78.96 | 84.17 |
| M7 | 96.04 | 94.58 | 82.50 | 97.08 | 98.96 | 87.29 | 96.88 | 94.58 | 94.79 | 84.79 | 83.33 | 91.04 |
| M8 | 98.13 | 98.33 | 89.17 | 95.42 | 99.79 | 98.96 | 96.67 | 95.83 | 99.17 | 90.83 | 83.33 | 98.54 |
| M9 | 89.17 | 98.75 | 87.08 | 94.58 | 99.79 | 98.75 | 97.08 | 91.67 | 98.75 | 88.75 | 92.29 | 92.50 |
| M10 | 97.71 | 97.92 | 90.42 | 95.83 | 99.79 | 98.33 | 96.25 | 94.58 | 98.96 | 91.88 | 94.38 | 98.75 |
| M11 | 93.54 | 91.25 | 80.21 | 96.67 | 99.38 | 84.17 | 94.38 | 93.13 | 98.33 | 88.75 | 82.92 | 95.42 |
| M12 | 96.46 | 99.58 | 98.54 | 99.17 | 100.00 | 99.38 | 98.54 | 97.29 | 99.38 | 95.83 | 98.75 | 99.58 |
| M13 | 98.75 | 99.79 | 90.63 | 97.92 | 100.00 | 97.92 | 97.50 | 96.67 | 99.38 | 89.79 | 97.92 | 100.00 |
| | nf: 40 | nf: 40 | nf: 40 | nf: 127 | nf: 122 | nf: 56 | nf: 130 | nf: 61 | nf: 80 | nf: 72 | nf: 71 | nf: 71 |
| M14 | 99.38 | 100.00 | 97.92 | 99.58 | 100.00 | 99.17 | 98.96 | 97.92 | 99.79 | 95.21 | 98.96 | 100.00 |
| | nf: 122 | nf: 40 | nf: 80 | nf: 55 | nf: 40 | nf: 67 | nf: 56 | nf: 45 | nf: 98 | nf: 134 | nf: 134 | nf: 41 |
| M15 | 100.00 | 100.00 | 91.46 | 98.96 | 100.00 | 99.38 | 98.96 | 96.88 | 99.58 | 93.75 | 98.33 | 100.00 |
| | nf: 73 | nf: 45 | nf: 82 | nf: 99 | nf: 40 | nf: 99 | nf: 140 | nf: 43 | nf: 95 | nf: 140 | nf: 95 | nf: 57 |
| M16 | 99.58 | 100.00 | 99.79 | 98.33 | 100.00 | 99.38 | 99.17 | 98.33 | 99.79 | 91.17 | 99.17 | 100.00 |
| | nf: 62 | nf: 40 | nf: 70 | nf: 73 | nf: 40 | nf: 100 | nf: 74 | nf: 58 | nf: 95 | nf: 74 | nf: 95 | nf: 45 |
| M17 | 99.38 | 99.79 | 89.79 | 100.00 | 100.00 | 98.75 | 96.04 | 97.29 | 99.58 | 90.63 | 91.25 | 99.58 |
| | nf: 86 | nf: 62 | nf: 47 | nf: 56 | nf: 45 | nf: 44 | nf: 115 | nf: 46 | nf: 101 | nf: 40 | nf: 44 | nf: 46 |
| M18 | 99.79 | 100.00 | 99.58 | 100.00 | 100.00 | 100.00 | 99.58 | 99.79 | 100.00 | 96.88 | 99.38 | 100.00 |
| | nf: 67 | nf: 80 | nf: 92 | nf: 80 | nf: 60 | nf: 90 | nf: 81 | nf: 46 | nf: 60 | nf: 131 | nf: 50 | nf: 60 |

The bold values highlight that the experimental results are desirable.

Moreover, we select some other literature that used similar DA methods for cross-domain fault diagnosis experiments that are similar to ours and compare our experimental results with them. Table 7 presents the comparison results. It is obviously true that our proposed fault diagnosis method outperforms other methods proposed in the corresponding literatures. To sum up, extensive comparative experiments are conducted, and the results prove the validity and advantages of the DFCD-IDAMN framework under diverse working loads.

4.2. Case 2: Fault Diagnostic of Bearing Dataset 2 across Different Working Speeds

4.2.1. Description of Bearing Dataset and Fault Diagnosis Tasks. To further prove the validity and flexibility of the

DFCD-IDAMN framework for CFD, in this case, it utilizes bearing vibration dataset sampled from the SQI-MFS test platform to conduct fault diagnosis experiments. The test platform is presented in Figure 15. This bearing vibration signals are sampled through acceleration sensors under 16 kHz sampling frequency. Table 7 lists the description of bearing vibration dataset. There are three categories of bearing defect: inner raceway defect (IRD), ball defect (BD), and outer raceway defect (ORD). The defect parameters include 0.05 mm, 0.1 mm, and 0.2 mm. Moreover, vibration data for bearings without defects is also used. To set CFD tasks, it utilizes the bearing vibration data under different motor speeds for implementing experiments. Therefore, it can obtain bearing vibration data of 10 classes, labeled 1–10. For each class, 90 samples are used to build a training set and a testing set, and it randomly divides 30 and 60 samples as training and testing samples. Each sample is

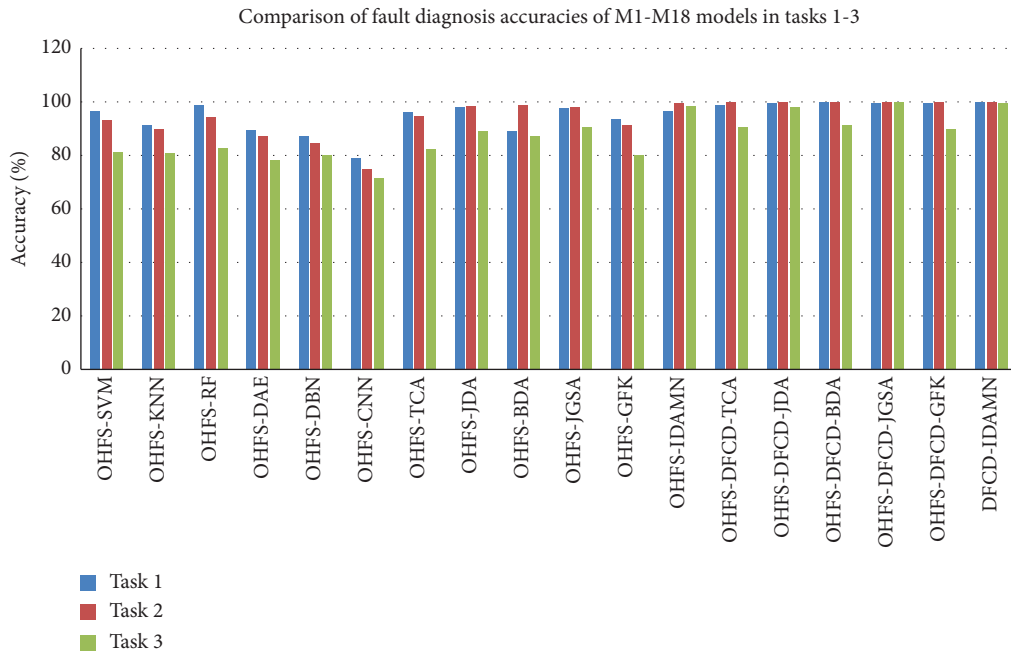


FIGURE 10: Comparison of CFD results of M1–M18 models (tasks 1–3).

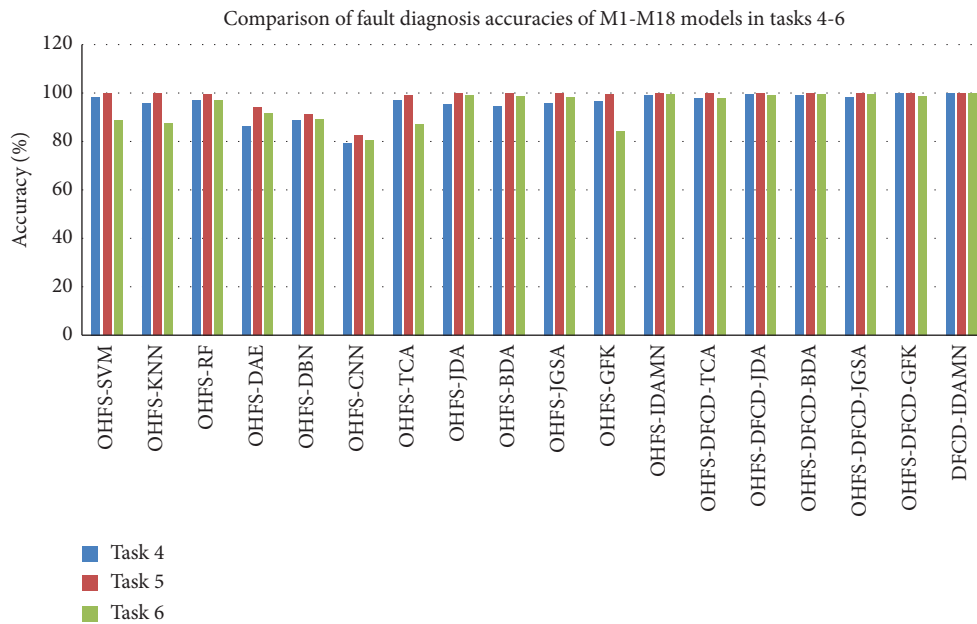


FIGURE 11: Comparison of CFD results of M1–M18 models (tasks 4–6).

composed of 5000 continuous data points from original vibration signals. On the basis of the bearing vibration data listed in Table 8, it sets 2 CFD tasks for experiments, and the details are shown in Table 9.

4.2.2. *Diagnosis Results of the Proposed DFCD-IDAMN Framework.* To further demonstrate the performance and advantages of the DFCD-IDAMN framework, bearing datasets from the SQI-MFS test-bed under diverse working speeds are employed for CFD experiments, and the contents

are similar to that of case 1. Take the no defect vibration data under a motor speed of 1730 rpm as an example. Figure 16 presents the DAI of 162 statistical features. From the figure, it can be seen that different features have different DAI values, and it indicates the different domain adaptability quantification results of different features. For the 3rd, 6th, 16th, 21st, and 24th features, their DAI values are significantly higher than other features, and it shows that their domain adaptability is more significant. Due to that, this work assumes that the higher DAI value indicates the greater domain adaptability; therefore, the DFCD can help to refine

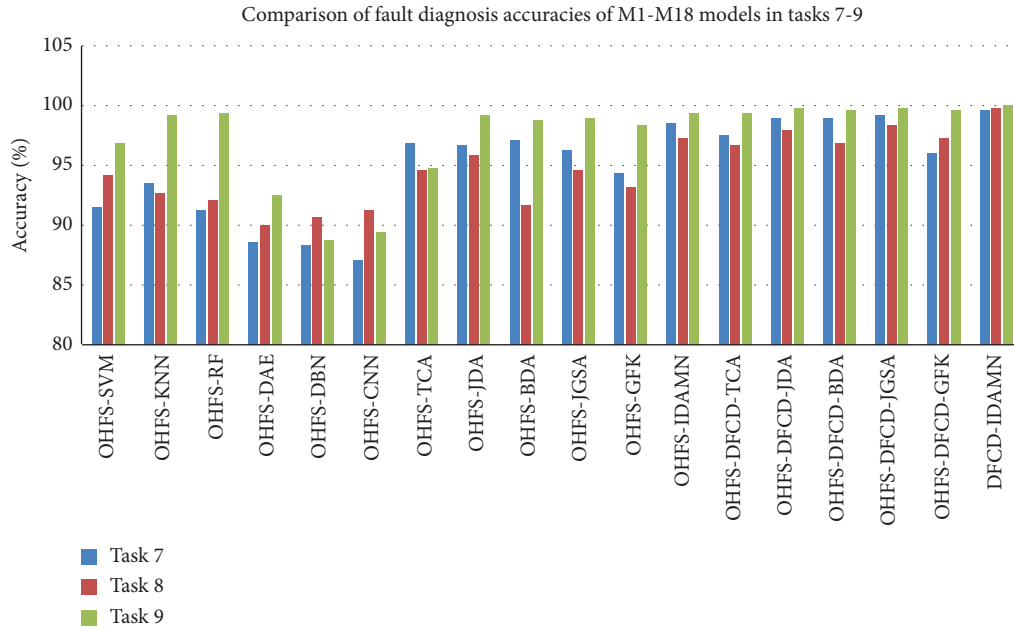


FIGURE 12: Comparison of CFD results of M1–M18 models (tasks 7–9).

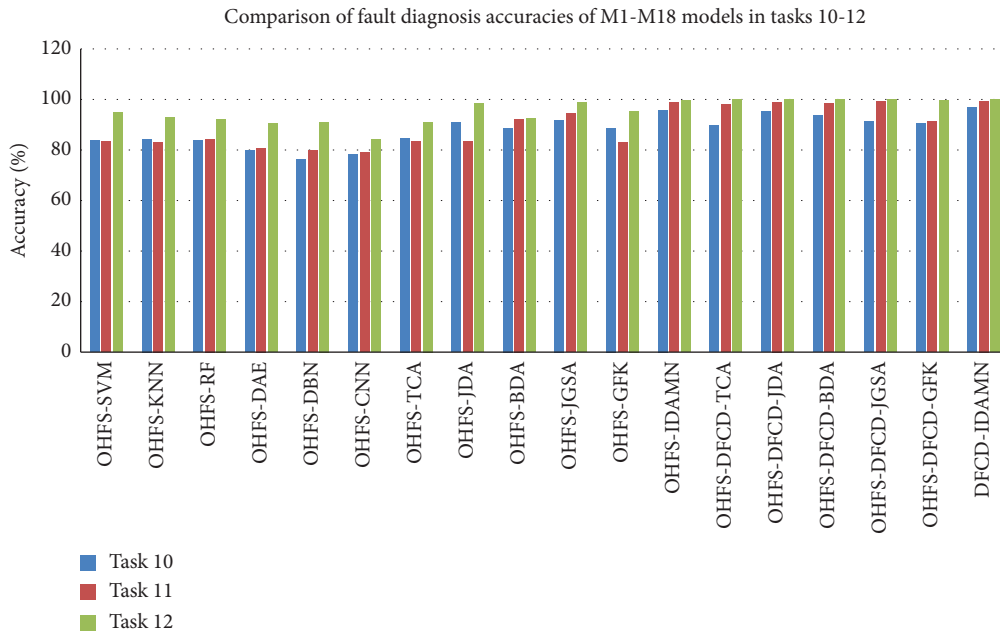


FIGURE 13: Comparison of CFD results of M1–M18 models (tasks 10–12).

some features (they are more advantageous to domain adaptation) by manually select a threshold of the DAI value, and these refined features are processed by the subsequent domain adaptation module. Table 10 lists the diagnosis results of 2 CFD tasks under different nf , it is easy to draw conclusions similar to the experimental analysis for case 1. Firstly, the model built by the DFCD-IDAMN framework attains an ideal result, the maximum diagnosis accuracies of tasks 1 and 2 are 91.83% and 95.17%, respectively. Secondly, the significant enhancement effect of the use of DFCD on CFD performance is further proven. When the DFCD is not

applied, all of 162 features are employed for the subsequent IDAMN domain adaptation method and fault classification, the diagnosis result (task 1: 86.17%, task 2: 86.83%) is not ideal. When the DFCD is used and the refined CDAF is employed for the subsequent procedure, it can attain obviously improved CFD accuracies. Therefore, the effectiveness of DFCD-IDAMN framework is validated again. The above CFD experiment involves some parameters of DFCD and IDAMN that should be manually set. For the basis for setting hyperparameters of the proposed methods, the specific values of these parameters are set based on

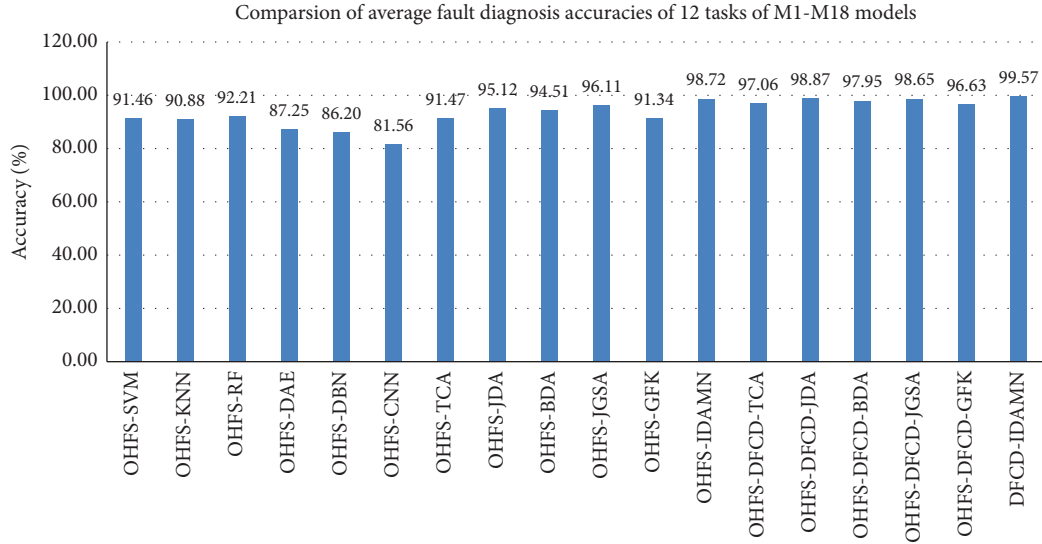


FIGURE 14: Contrast of mean diagnosis accuracies of 12 tasks obtained by M1–M18 models.

TABLE 7: comparison of experimental results between DFCD-IDAMN and relevant methods from other literatures.

| Abbreviation of methods | DA method | Literature | Experimental data | Cross-domain tasks | Maximum mean accuracy (%) |
|-------------------------|-----------|---|---|--|---------------------------|
| DFCD-IDAMN | IDAMN | This article | Bearing data from CWRU and our own test-bed | 12 tasks under 4 working conditions | 99.57 |
| MTSDE [56] | MMD | A novel cross-domain intelligent fault diagnosis method based on entropy features and transfer learning | Bearing from PHM2009, CWRU, and MFPT | 6 diagnosis tasks under 2 working conditions | 97.10 |
| BARTL [3] | BDA | Balanced adaptation regularization based transfer learning for unsupervised cross-domain fault diagnosis | Bearing data from Jiangnan university and Politecnico di Torino | 6 diagnosis tasks under 2 working speeds | 98.73 |
| FT-IDJ [57] | JDA | An intelligent fault diagnosis method for rolling bearings based on feature transfer with improved DenseNet and joint distribution adaptation | Bearing data from CWRU | 12 diagnosis tasks under 4 working speeds | 98.50 |
| TCA-based [58] | TCA | Transfer learning based data feature transfer for fault diagnosis | Bearing data from CWRU | 6 diagnosis tasks under 2 working speeds | 91.40 |
| AMPD [26] | GFK | A new transferable bearing fault diagnosis method with adaptive manifold probability distribution under different working conditions | Bearing data from own test rig | 12 diagnosis tasks under 4 working speeds | 98.85 |
| JGSA-FTFE [59] | JGSA | Time frequency feature analysis of rolling bearing fault based on deep transfer learning | Bearing data from CWRU and own test-bed | 2 diagnosis tasks under 2 working conditions | 95.55 |

The bold values highlight that the experimental results are desirable.

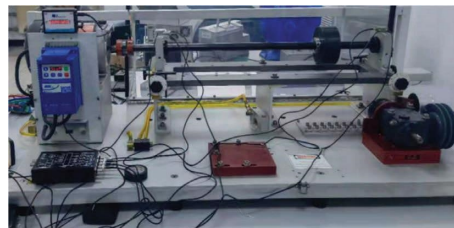


FIGURE 15: The SQI-MFS test-bed.

TABLE 8: The bearing data from SQI-MFS test-bed.

| Category of bearing defect | Defect diameter (mm) | Number of training/testing samples | | Class label |
|----------------------------|----------------------|------------------------------------|-----------------------|-------------|
| | | Motor speeds 1730 rpm | Motor speeds 1750 rpm | |
| No defect | 0 | 30/60 | 30/60 | Classes 1 |
| IRD | 0.05 | 30/60 | 30/60 | Classes 2 |
| | 0.1 | 30/60 | 30/60 | Classes 3 |
| | 0.2 | 30/60 | 30/60 | Classes 4 |
| ORD | 0.05 | 30/60 | 30/60 | Classes 5 |
| | 0.1 | 30/60 | 30/60 | Classes 6 |
| | 0.2 | 30/60 | 30/60 | Classes 7 |
| BD | 0.05 | 30/60 | 30/60 | Classes 8 |
| | 0.1 | 30/60 | 30/60 | Classes 9 |
| | 0.2 | 30/60 | 30/60 | Classes 10 |

TABLE 9: The CFD tasks for case 2.

| Tasks | SD (training samples) | | | TD (testing samples) | | |
|-------|-----------------------|-------------------------|-------------------|----------------------|-------------------------|-------------------|
| | Motor speed (rpm) | Defect types of samples | Number of samples | Motor speed (rpm) | Defect types of samples | Number of samples |
| 1 | 1730 | Classes 1–10 | 300 | 1750 | Classes 1–10 | 600 |
| 2 | 1750 | Classes 1–10 | 300 | 1730 | Classes 1–10 | 600 |

experimental experience. Therefore, we directly present the relevant parameter values in this manuscript. For the DFCD, trade-off parameter $\theta = 0.5$. The parameters manual set in IDAMN include: manifold subspace dimension $d = 40$, regularization parameters $\lambda = 0.1$, $\beta = 0.3$, and $\eta = 0.5$, dimension of output source and target domains feature space $k = 20$. Iterations $i = 10$. In particular, although the manifold subspace dimension is set as 40, when the feature dimension after the proposed feature refinement (that is nf) is less than twice of the set manifold subspace dimension, the manifold subspace dimension will be automatic adjusted as the half of nf . In Table 10, when the nf is 40, 50, 60, and 70, the manifold subspace dimension will be automatic adjusted as 20, 25, 30, and 35. On the contrary, when nf is not less than twice of the set manifold subspace dimension (when the nf is 80 to 162), the GFK is implemented under the set manifold subspace dimension 40.

4.2.3. Comparative Analysis with Other Fault Diagnosis Models. The comparative models used in this section are also shown in Table 6, and the experimental contents are the same as case 1. The corresponding cross-domain fault diagnosis results are listed in Table 11 and Figure 17. It is also obviously concluded that the performance of the model built by the DFCD-IDAMN framework significantly surpasses that of the other models. The detailed comparative analysis is illustrated as follows. (1) Comparing the DFCD-IDAMN model with M1–M6 (base classifier-based models), the diagnosis accuracies of tasks 1 and 2 of DFCD-IDAMN model are remarkably higher than that of the M1–M6 models. Moreover, the OHFS-IDAMN model can achieve the higher diagnosis accuracies in tasks 1 and 2 than M1–M6 models. (2) Comparing the OHFS-IDAMN (M12) model with the M7–M11 models (domain adaptation-based models), the

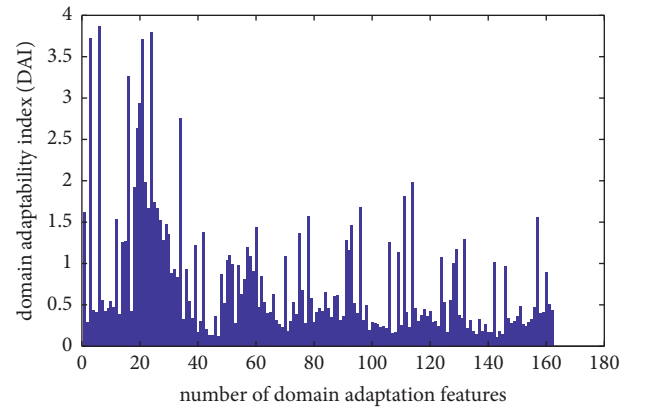


FIGURE 16: DAI of 162 statistical features that extracted from no defect bearing vibration data under motor speed of 1730 rpm (trade-off parameter θ is 0.5).

diagnosis accuracies of tasks 1 and 2 are noticeably higher than M7–M11 models. The accuracy of the M12 model in task 1 can attain 86.17%, which is, respectively, 10.17%, 3.67%, 20.00%, 6.00%, and 11.67% higher than the M7–M11 models. Accordingly, for domain adaptation ability, it is evident that the proposed IDAMN outperforms traditional JDA, BDA, TCA, JGSA, and GFK, which can effectively increase the CFD accuracy. (3) Comparing M7–M12 (domain adaptation-based models without DFCD) with M13–M18 (domain adaptation-based models with DFCD), it is easily found that the utilization of the DFCD has a remarkable improvement on the diagnosis accuracy of domain adaptation-based model, take OHFS-JDA (M8) and OHFS-DFCD-JDA (M14) as examples, the accuracies of tasks 1 and 2 of the M14 model are, respectively, 89.00% and 83.83%; nevertheless, the M8 model only attains

TABLE 10: CFD results of DFCD-IDAMN in case 2.

| <i>nf</i> | Accuracies of task 1 (%) | Accuracies of task 2 (%) |
|------------|-----------------------------|-----------------------------|
| 40 | 77.67 | 74.17 |
| 50 | 87.67 | 72.17 |
| 60 | 87.00 | 91.00 |
| 70 | 88.67 | 92.50 |
| 80 | 90.00 | 89.50 |
| 90 | 89.83 | 93.17 |
| 100 | 88.83 | 92.33 |
| 110 | 89.00 | 91.50 |
| 120 | 91.67 | 87.67 |
| 130 | 91.50 | 87.83 |
| 140 | 91.17 | 90.67 |
| 150 | 90.00 | 90.50 |
| 160 | 81.50 | 88.00 |
| 162 | 86.17 | 86.83 |
| <i>mda</i> | 91.83 | 95.17 |
| | nf: 121 | nf: 65 |

The bold values highlight that the experimental results are desirable.

TABLE 11: CFD results of M1–M18 models in case 2.

| Model | Accuracies of task 1 (%) | Accuracies of task 2 (%) |
|-------|-----------------------------|-----------------------------|
| M1 | 81.17 | 77.83 |
| M2 | 76.33 | 75.00 |
| M3 | 75.17 | 80.83 |
| M4 | 71.83 | 73.00 |
| M5 | 69.33 | 67.67 |
| M6 | 66.17 | 59.33 |
| M7 | 76.00 | 69.83 |
| M8 | 82.50 | 72.00 |
| M9 | 66.17 | 70.83 |
| M10 | 80.17 | 74.00 |
| M11 | 74.50 | 75.50 |
| M12 | 86.17 | 86.83 |
| M13 | 77.67 | 79.17 |
| | nf: 136 | nf: 75 |
| M14 | 89.00 | 83.83 |
| | nf: 116 | nf: 99 |
| M15 | 74.33 | 76.83 |
| | nf: 112 | nf: 88 |
| M16 | 84.67 | 86.83 |
| | nf: 118 | nf: 56 |
| M17 | 81.67 | 78.83 |
| | nf: 95 | nf: 110 |
| M18 | 91.83 | 95.17 |
| | nf: 121 | nf: 65 |

The bold values highlight that the experimental results are desirable.

82.50% and 72.00% accuracies, respectively, which is obvious inferior than the M14 model. Accordingly, the above experimental analysis once again shows that the DFCD can help to refine features with strong domain adaptability, which can effectively enhance domain

adaptation performance and increase CFD accuracy. To sum up, extensive experiments are carried out, and the results further validate the validity, adaptability, and superiority of the DFCD-IDAMN framework under diverse working speeds.

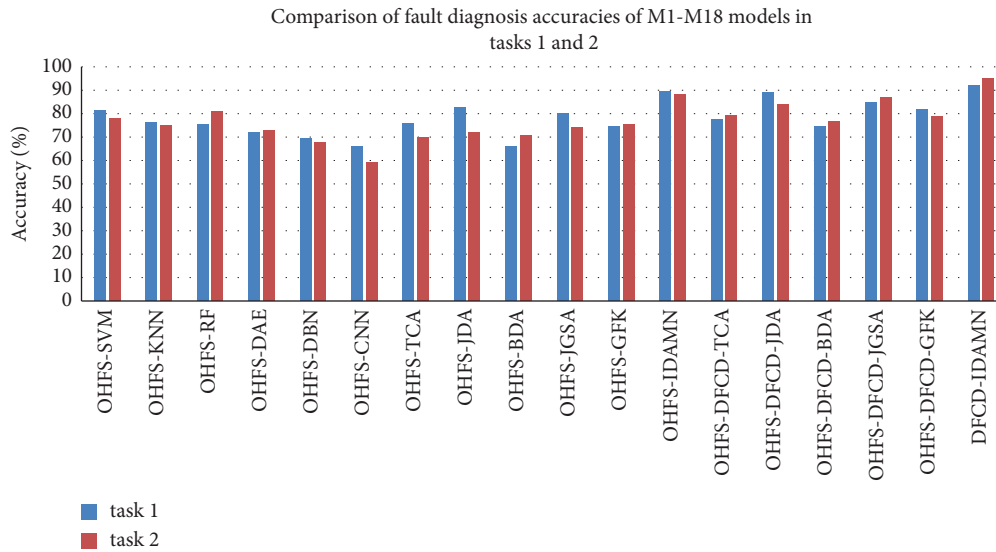


FIGURE 17: Contrast of diagnosis results of M1–M18 models in tasks 1–2.

5. Conclusions

This work designs a new framework based on the proposed DFCD and IDAMN for rolling bearing across diverse operating conditions. In this framework, the EEMD is first applied for signals processing and statistics-based features extraction. Then, the DFCD is employed to refine the features by evaluating the fault distinguishability and WCI. Next, the IDAMN is performed to maps the feature data into a GM subspace and further achieves improved JDA with neighborhood relationship preserving. Finally, an adaptive classifier is trained for fault diagnostic.

By utilizing bearing data collected from two experimental platforms, extensive fault diagnosis experiments are conducted. These experimental results show the following: (1) the DFCD can effectively refine features with the better domain adaptability; accordingly, the utilization of the DFCD has a significant enhancement on the diagnosis accuracy of domain adaptation-based models. (2) IDAMN possesses more robust domain adaptation ability than JDA, TCA, BDA, JGSA, and GFK. (3) The model built by the DFCD and IDAMN can attain a desirable cross-domain fault diagnosis accuracy with a suitable nf , which presents a promising capability for employing it in practical industrial scenarios with variable working conditions. In future, we are planning to develop stronger domain adaptation-based approaches for more complicated fault detection scenes and conduct research on adaptive optimization methods for related parameters used in the proposed methods.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the Innovation and Entrepreneurship Training Program for College Students of China under Grant no. 202210357121 and the Joint Funds of the Zhejiang Provincial Natural Science Foundation of China under Grant no. LTY22E050001.

References

- [1] Y. Xia, C. Shen, D. Wang, Y. Shen, W. Huang, and Z. Zhu, "Moment matching-based intraclass multisource domain adaptation network for bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 168, Article ID 108697, 2022.
- [2] T. Han, Y. F. Li, and M. Qian, "A hybrid generalization network for intelligent fault diagnosis of rotating machinery under unseen working conditions," *Institute of Electrical and Electronics Engineers Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.
- [3] Q. Hu, X. Si, A. Qin, Y. Lv, and M. Liu, "Balanced adaptation regularization based transfer learning for unsupervised cross-domain fault diagnosis," *Institute of Electrical and Electronics Engineers Sensors Journal*, vol. 22, no. 12, pp. 12139–12151, 2022.
- [4] T. Zhou, T. Han, and E. L. Droguett, "Towards trustworthy machine fault diagnosis: a probabilistic Bayesian deep learning framework," *Reliability Engineering & System Safety*, vol. 224, Article ID 108525, 2022.
- [5] D. Wei, T. Han, F. Chu, and M. J. Zuo, "Weighted domain adaptation networks for machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 158, Article ID 107744, 2021.

- [6] T. Han, W. Xie, and Z. Pei, "Semi-supervised adversarial discriminative learning approach for intelligent fault diagnosis of wind turbine," *Information Sciences*, vol. 648, Article ID 119496, 2023.
- [7] X. L. Zhang, W. Chen, B. J. Wang, and X. F. Chen, "Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization," *Neurocomputing*, vol. 167, pp. 260–279, 2015.
- [8] R. S. Gunerkar, A. K. Jalan, and S. U. Belgamwar, "Fault diagnosis of rolling element bearing based on artificial neural network," *Journal of Mechanical Science and Technology*, vol. 33, no. 2, pp. 505–511, 2019.
- [9] Z. Zhou, C. Wen, and C. Yang, "Fault isolation based on k-nearest neighbor rule for industrial processes," *Institute of Electrical and Electronics Engineers Transactions on Industrial Electronics*, vol. 63, no. 4, pp. 2578–2586, 2016.
- [10] H. Wei, Q. Zhang, M. Shang, and Y. Gu, "Extreme learning Machine-based classifier for fault diagnosis of rotating Machinery using a residual network and continuous wavelet transform," *Measurement*, vol. 183, Article ID 109864, 2021.
- [11] J. Ma and F. Liu, "Bearing fault diagnosis with variable speed based on fractional hierarchical range entropy and hunter-prey optimization algorithm—optimized random forest," *Machines*, vol. 10, no. 9, p. 763, 2022.
- [12] S. Luo, X. Huang, Y. Wang, R. Luo, and Q. Zhou, "Transfer learning based on improved stacked autoencoder for bearing fault diagnosis," *Knowledge-Based Systems*, vol. 256, Article ID 109846, 2022.
- [13] S. Zhang, Z. Liu, Y. Chen, Y. Jin, and G. Bai, "Selective kernel convolution deep residual network based on channel-spatial attention mechanism and feature fusion for mechanical fault diagnosis," *International Society of Automation Transactions*, vol. 133, pp. 369–383, 2023.
- [14] H. Zhao, X. Yang, B. Chen, H. Chen, and W. Deng, "Bearing fault diagnosis using transfer learning and optimized deep belief network," *Measurement Science and Technology*, vol. 33, no. 6, Article ID 65009, 2022.
- [15] R. Bai, Q. Xu, Z. Meng, L. Cao, K. Xing, and F. Fan, "Rolling bearing fault diagnosis based on multi-channel convolution neural network and multi-scale clipping fusion data augmentation," *Measurement*, vol. 184, Article ID 109885, 2021.
- [16] Y. Zhang, Z. Ren, K. Feng, K. Yu, M. Beer, and Z. Liu, "Universal source-free domain adaptation method for cross-domain fault diagnosis of machines," *Mechanical Systems and Signal Processing*, vol. 191, Article ID 110159, 2023.
- [17] J. Wang, Y. Chen, H. Shuji, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfer learning," in *Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM)*, pp. 1129–1134, IEEE, Beijing, China, July 2017.
- [18] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proceedings of the IEEE international conference on computer vision*, pp. 2200–2207, Sydney, NSW, Australia, December 2013.
- [19] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *Institute of Electrical and Electronics Engineers Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.
- [20] B. Gong, Y. Shi, S. Fei, and G. Kristen, "Geodesic flow kernel for unsupervised domain adaptation," in *Proceedings of the 2012 IEEE conference on computer vision and pattern recognition*, pp. 2066–2073, IEEE, Providence, RI, USA, June 2012.
- [21] J. Zhang, W. Li, and O. Philip, "Joint geometrical and statistical alignment for visual domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1859–1867, Wollongong, Australia, May 2017.
- [22] Z. Cheng, X. Li, G. Peng, Y. Deng, Z. Xie, and L. Liu, "Transfer deep learning network for rolling bearing fault diagnosis of wind turbines," *Journal of Physics: Conference Series*, vol. 2503, no. 1, Article ID 12095, 2023.
- [23] X. Li and T. Zhang, "Bearing fault diagnosis under different operating conditions based on source domain multi sample joint distribution adaptation," in *Proceedings of the 2022 34th Chinese Control and Decision Conference (CCDC)*, pp. 1827–1832, IEEE, Hefei, China, August 2022.
- [24] Z. Zhong, H. Liu, W. Mao, X. Xie, and Y. Cui, "Rolling bearing fault diagnosis across operating conditions based on unsupervised domain adaptation," *Lubricants*, vol. 11, no. 9, p. 383, 2023.
- [25] J. Xiong, S. Cui, and H. Tang, "A novel intelligent bearing fault diagnosis method based on signal process and multi-kernel joint distribution adaptation," *Scientific Reports*, vol. 13, no. 1, p. 4535, 2023.
- [26] P. Lei, C. Shen, D. Wang, L. Chen, Z. Zhou, and Z. Zhu, "A new transferable bearing fault diagnosis method with adaptive manifold probability distribution under different working conditions," *Measurement*, vol. 173, Article ID 108565, 2021.
- [27] Y. Yu, C. Zhang, Y. Li, and Y. Li, "A new transfer learning fault diagnosis method using TSC and JGSA under variable condition," *Institute of Electrical and Electronics Engineers Access*, vol. 8, pp. 177287–177295, 2020.
- [28] J. Wang, W. Feng, Y. Chen, Y. Han, M. Huang, and P. S. Yu, "Visual domain adaptation with manifold embedded distribution alignment," in *Proceedings of the 26th ACM International Conference on Multimedia*, pp. 402–410, New York, NY, USA, October 2018.
- [29] W. Ma, Y. Zhang, L. Ma, R. Liu, and S. Yan, "An unsupervised domain adaptation approach with enhanced transferability and discriminability for bearing fault diagnosis under few-shot samples," *Expert Systems with Applications*, vol. 225, Article ID 120084, 2023.
- [30] Y. Cao, M. Long, and J. Wang, "Unsupervised domain adaptation with distribution matching machines," *Proceedings of the Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
- [31] X. Yu, W. Chen, C. Wu et al., "Rolling bearing fault diagnosis based on domain adaptation and preferred feature selection under variable working conditions," *Shock and Vibration*, vol. 2021, no. 99, Article ID 8843124, 27 pages, 2021.
- [32] X. Lu, Z. Lu, Q. Wu et al., "Soft Fault diagnosis of analog circuit based on EEMD and improved MF-DFA," *Electronics*, vol. 12, no. 1, p. 114, 2022.
- [33] Y. Zhao, Y. Fan, H. Li, and X. Gao, "Rolling bearing composite fault diagnosis method based on EEMD fusion feature," *Journal of Mechanical Science and Technology*, vol. 36, no. 9, pp. 4563–4570, 2022.
- [34] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 01, no. 01, pp. 1–41, 2009.
- [35] T. Han, R. Liu, Z. Zhao, and P. Kundu, "Fault diagnosis and health management of power machinery," *Machines*, vol. 11, no. 4, p. 424, 2023.
- [36] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, "Semisupervised transfer component analysis for domain

- adaptation in remote sensing image classification,” *Institute of Electrical and Electronics Engineers Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3550–3564, 2015.
- [37] M. Sugiyama, “Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis,” *Journal of Machine Learning Research*, vol. 8, no. 5, pp. 1027–1061, 2007.
- [38] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science (New York, NY)*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [39] Y. Wu, X. Liu, Y.-L. Wang, Q. Li, Z. Guo, and Y. Jiang, “Improved deep PCA and Kullback–Leibler divergence based incipient fault detection and isolation of high-speed railway traction devices,” *Sustainable Energy Technologies and Assessments*, vol. 57, Article ID 103208, 2023.
- [40] Q. Zhang, K. Liu, D. Han, G. Su, and Y. Xia, “Design of stealthy deception attacks with partial system knowledge,” *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 68, no. 2, pp. 1069–1076, 2023.
- [41] A. K. Seghouane and S. I. Amari, “The AIC criterion and symmetrizing the Kullback–Leibler divergence,” *Institute of Electrical and Electronics Engineers Transactions on Neural Networks*, vol. 18, no. 1, pp. 97–106, 2007.
- [42] J. Hamm and D. D. Lee, “Grassmann discriminant analysis: a unifying view on subspace-based learning,” in *Proceedings of the 25th International Conference on Machine Learning*, pp. 376–383, Philadelphia, PA, USA, May 2008.
- [43] Z. Zhang, H. Chen, S. Li, Z. An, and J. Wang, “A novel geodesic flow kernel based domain adaptation approach for intelligent fault diagnosis under varying working condition,” *Neurocomputing*, vol. 376, pp. 54–64, 2020.
- [44] F. Dong, X. Yu, X. Shi, K. Liu, Z. Wu, and W. Yu, “A new transferable fault diagnosis approach of rotating machinery based on deep autoencoder and dominant features selection under different operating conditions,” *Shock and Vibration*, vol. 2021, Article ID 7383255, 21 pages, 2021.
- [45] R. Fisher, “The use of multiple measurement sin taxonomic problems,” *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [46] F. Dong, X. Yu, E. Ding, S. Wu, C. Fan, and Y. Huang, “Rolling bearing fault diagnosis using modified neighborhood preserving embedding and maximal overlap discrete wavelet packet transform with sensitive features selection,” *Shock and Vibration*, vol. 2018, Article ID 5063527, 29 pages, 2018.
- [47] Z. Lei, G. Wen, S. Dong et al., “An intelligent fault diagnosis method based on domain adaptation and its application for bearings under polytropic working conditions,” *Institute of Electrical and Electronics Engineers Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.
- [48] G.-B. Jang and S.-B. Cho, “Feature space transformation for fault diagnosis of rotating machinery under different working conditions,” *Sensors*, vol. 21, no. 4, p. 1417, 2021.
- [49] X. Yu, B. Xia, S. Yang, H. Yin, Y. Wang, and X. Liu, “A deep domain-adversarial transfer fault diagnosis method for rolling bearing based on ensemble empirical mode decomposition,” *Journal of Sensors*, vol. 2022, Article ID 8959185, 18 pages, 2022.
- [50] W.-L. Qin, W.-J. Zhang, and C. Lu, “Rolling bearing fault diagnosis based on ensemble empirical mode decomposition, information entropy and random forests,” *Vibroengineering Procedia*, vol. 5, pp. 211–216, 2015.
- [51] C. Zhong, J.-S. Wang, and W.-Z. Sun, “Fault diagnosis method of rotating bearing based on improved ensemble empirical mode decomposition and deep belief network,” *Measurement Science and Technology*, vol. 33, no. 8, Article ID 85109, 2022.
- [52] X. Yu, F. Dong, E. Ding, S. Wu, and C. Fan, “Rolling bearing fault diagnosis using modified LFDA and EMD with sensitive feature selection,” *Institute of Electrical and Electronics Engineers Access*, vol. 6, pp. 3715–3730, 2018.
- [53] S. Rajabi, M. Saman Azari, S. Santini, and F. Flammini, “Fault diagnosis in industrial rotating equipment based on permutation entropy, signal processing and multi-output neuro-fuzzy classifier,” *Expert Systems with Applications*, vol. 206, Article ID 117754, 2022.
- [54] C. Ma, Y. Li, X. Wang, and Z. Cai, “Early fault diagnosis of rotating machinery based on composite zoom permutation entropy,” *Reliability Engineering & System Safety*, vol. 230, Article ID 108967, 2023.
- [55] A. S. Minhas, S. Singh, and S. Singh, “A new bearing fault diagnosis approach combining sensitive statistical features with improved multiscale permutation entropy method,” *Knowledge-Based Systems*, vol. 218, no. 17, Article ID 106883, 2021.
- [56] Y. Li, Y. Ren, H. Zheng, Z. Deng, and S. Wang, “A novel cross-domain intelligent fault diagnosis method based on entropy features and transfer learning,” *Institute of Electrical and Electronics Engineers Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.
- [57] C. Qian, Q. Jiang, Y. Shen, C. Huo, and Q. Zhang, “An intelligent fault diagnosis method for rolling bearings based on feature transfer with improved DenseNet and joint distribution adaptation,” *Measurement Science and Technology*, vol. 33, no. 2, Article ID 25101, 2021.
- [58] W. Xu, Y. Wan, T. Y. Zuo, and X. M. Sha, “Transfer learning based data feature transfer for fault diagnosis,” *Institute of Electrical and Electronics Engineers Access*, vol. 8, pp. 76120–76129, 2020.
- [59] J. Cui, R. Zhou, and Z. Wang, “Time frequency feature analysis of rolling bearing fault based on deep transfer learning,” in *Proceedings of the 2021 4th IEEE international conference on industrial cyber-physical systems (ICPS)*, pp. 679–685, IEEE, Victoria, BC, Canada, May 2021.