A New Bearing Fault Diagnosis Method Based on Deep Transfer Network and Supervised Joint Matching

Chengyao Liu, Fei Dong, Kunpeng Ge, and Yuanyuan Tian

***Abstract*—In practical industrial environment, variable working condition can result in shifts in data distributions, and the labeled fault data in various working conditions is difficult to collect because rotating machines often works in normal status, and the insufficient labeled fault data brings data samples imbalance and performance degradation of intelligent fault diagnosis model. To overcome these problems, by integrating the superiority of deep learning method and feature-based transfer learning method, this work proposes an innovative cross-domain fault diagnosis framework based on deep transfer convolutional neural network and supervised joint matching. First, the continue wavelet transform is used to process original bearing vibration signals and extract time-frequency images. Second, a deep transfer convolutional neural network is built by the way of fine-tuning, and the trained network is used to extract deep features from different domains. Third, a new domain adaptation approach, supervised joint matching, is developed to conduct joint feature distribution matching and instance reweighting with the consideration of maximum marginal criterion. The intelligent bearing fault diagnosis model is then trained to predict the labels of the target domain's feature data. To verify the performance of the proposed approaches, this study uses two distinct datasets pertaining to bearing defects for conducting cross-domain fault diagnosis in the presence of balanced and imbalanced data. The experimental analysis indicates that the designed methods can achieve desirable diagnostic accuracy and possess robust generalization ability.**

***Index Terms*—Deep learning, domain adaptation, fault diagnosis, time-frequency images.**

I. Introduction

ROTATING machinery plays a crucial role in industrial area, including but not limited to construction, transportation, aerospace, electricity, and other fields [1], [2]. Bearing failures within these machines can lead to operational disruptions and pose safety risks; thus, the development of robust and efficient fault diagnosis models is imperative for averting accidents and significant financial losses [3].

In recent years, artificial intelligence (AI)-based fault diagnosis techniques for rotating machinery have garnered significant attention, leading to a wealth of research conducted by numerous scholars [4]. The main challenges about the applications of AI-based methods for fault detection or prediction are data distributions shifting, class-imbalance, and insufficient labelled fault samples, etc. To overcome these challenges, researchers have investigated some popular methods for fault diagnosis, such as continual learning-based methods [5], [6], personalized diagnosis methods [7], [8], [9], simulation enlargement/augmentation and sample transfer method [10], transfer learning combined fault sample augmentation [11], conventional deep learning(DL)-based methods [12], [13], feature transfer learning(FTL)-based methods [14], [15]. In [5], aiming at that the traditional classification methods cannot effectively recognize the new fault type, the authors proposed a continual learning classification approach and applied it to fault diagnosis of equipment. In [6], a vaccine-enhanced artificial immune continual learning method (VECL) with the function of incremental learning was proposed, and the VECL can culture new types of memory cells to expand the recognition range of the model and improve the performance of bearing fault identification. In [7], a novel numerical simulation-based personalized diagnosis methodology was investigated to detect faults in a shaft, aiming at the shaft unbalance, misalignment, rub-impact and the combination of rub-impact and unbalance, this method was developed using the finite element method, wavelet packet transform and support vector machine. In [8], aiming at that faulty samples of mechanical systems are difficult to obtain, a personalized diagnosis fault method of bearings was proposed to activate the smart sensor networks using finite element method simulations. In [9], a personalized fault diagnosis of rolling bearings in trains based on digital twin and machine learning classification framework was designed. In [10], to solving the problem of insufficient labelled fault samples for gear fault detection, a new method using numerical simulation and a generative adversarial network (GAN) is proposed to enlarge fault samples for gears fault diagnosis. In [11], in order to enlarge the training samples for bearing fault diagnosis, the finite element method was adopted

Manuscript received 30 January 2024; revised 7 April 2024; accepted 16 April 2024. Date of publication 22 April 2024; date of current version 2 May 2024. This work was supported by the Joint Funds of the Zhejiang Provincial Natural Science Foundation of China under Grant LTY22E050001. *(Corresponding author: Yuanyuan Tian.)*

Chengyao Liu is with the Department of Jiaotong, Zhejiang Industry Polytechnic College, Shaoxing 312000, China (e-mail: chengyao_liu168@163.com).

Fei Dong is with the IOT Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221000, China, and also with the School of Internet, Anhui University, Hefei 230039, China (e-mail: feidong@cumt.edu.cn).

Kunpeng Ge and Yuanyuan Tian are with the IOT Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221000, China (e-mail: 370228484@qq.com; yt20231225@163.com).

Digital Object Identifier 10.1109/JPHOT.2024.3392392

to obtain sufficient and complete simulation samples of all the fault categories as the original fault samples, and the original simulation fault samples are adjusted using a generative adversarial network-based domain adaptation network to make them similar to the measured samples. The authors utilized bearing and gear fault data to conduct a series of experiments, which can validate that the proposed methods can overcome the problems of insufficient and incomplete labeled training samples. Moreover, DL methods have become one of the research hotspots due to their powerful automatic feature extraction capability, which is not possible with traditional machine learning-based approaches [12], [13]. With the help of the typical advantage of DL method, numerous intelligent fault diagnosis techniques for bearings have been developed, achieving remarkable performance. The commonly studied DL methods encompass the likes of the deep auto-encoder (DAE) [16], the deep residual network (DRN) [17], the deep belief network (DBN) [18], and the convolutional neural network (CNN) [19], among others. In [16], an enhanced DAE that employs convolutional shortcuts and a domain fusion strategy was proposed to address the challenge of limited labeled data in bearing fault diagnosis. In [17], the authors investigate a selective kernel convolution DRN, and it can leverage the channel-spatial attention mechanism and feature fusion, which helps to enhance the fault diagnosis accuracy. In [18], a novel fault diagnosis approach for bearings utilizing joint distribution adaptation and an enhanced DBN along with an improved sparrow search algorithm was developed. In [19], addressing the challenge posed by the significant training data requirements that hinder the implementation of DL-based methods, a novel bearing fault diagnosis approach utilizing a multi-channel CNN in conjunction with a multi-scale clipping fusion data augmentation technique was investigated. In [20], a novel deep network-based maximum correlated kurtosis deconvolution was developed for the problems of the filter length selection and the sensibility to random interference. In addition to the above traditional DL methods mentioned above, in recent years, the combination of deep learning and transfer learning, deep transfer learning (DTL) methods, has gradually become a research hotspot. In [21], spectrum alignment and a deep transfer convolution neural network were combined to develop a novel bearing fault diagnosis model. This model's deep transfer convolution neural network was trained to enable the extraction of robust, domain-invariant features. In [22], to address the issue of data distribution shifts, the authors constructed a deep multi-source transfer learning model. The study presented in [23] explores a deep imbalanced domain adaptation framework tailored for situations where feature distribution shifts and scarce label information co-occur across different operational states. This framework successfully mitigates class-imbalanced label shift and facilitates fine-grained matching in latent space. In [24], a new DTL-based method called classifier constrained domain adaptation network was developed for extracting transfer characteristics from simulated samples by theory model for experimental rotor fault diagnosis. In [25], to overcome the challenges of fine-grained fault feature extraction and undesirable model generalization to unseen few-shot faults in few-shot fine-grained fault diagnosis tasks, the authors investigated a new attention-based deep meta-transfer learning method. In [26], the authors proposed a new deep residual joint transfer strategy method for the cross-condition bearing fault diagnosis, in first, the deep residual network was learned in training tasks, and then three transfer strategies were applied to enhance the generalizability and adaptability of the pre-trained models to the data distribution in target tasks. The above studies have enhanced the automatic feature extraction capabilities of DL methods, leading to improved fault pattern recognition and classification performance. However, achieving optimal accuracy in DL-based fault diagnosis models typically necessitates the fulfillment of several conditions [2], [12], [14], [27]: 1) a substantial number of labeled training samples; 2) the testing samples exhibit the same distribution as the training samples. In practical engineering applications, it is often hard to acquire a sufficient quantity of labeled fault samples from real-world mechanical systems under various working conditions. Furthermore, distribution differences frequently exist between training and testing samples. Consequently, these issues present significant challenges for the deployment of DL-based fault diagnosis models in industrial settings.

FTL approaches are usually employed based on hand-craft feature extraction [14], [22]. Key methods frequently explored in FTL comprise Balanced Distribution Adaptation (BDA) [28], Joint Distribution Adaptation (JDA) [29], Transfer Component Analysis (TCA) [30], Geodesic Flow Kernel (GFK) [31], Manifold Embedded Distribution Alignment (MEDA) [32], and Joint Geometrical and Statistical Alignment (JGSA) [33]. Based on these approaches, many scholars have developed several intelligent fault diagnosis models for bearings. In [14], an improved BDA was developed by incorporating balanced distribution adaptation and balanced label propagation, and the improved BDA can achieve desirable cross-domain fault diagnosis performance. In [18], a novel approach for bearing fault diagnosis was proposed, utilizing JDA and DBN and an improved sparrow search algorithm, and the JDA component is tasked with facilitating feature transfer between different domains. In [34], the bearing fault diagnosis model was augmented with an advanced domain adaptation algorithm, specifically a weighted version of Transfer Component Analysis (TCA), to effectively operate across distinct domains. In [35], the authors present a novel domain adaptation model, designated as DAGSZ, for bearing fault diagnosis. This model incorporates GFK, feature extraction, and normalization to achieve more desirable domain adaptataion. In [36], a novel TL approach based on limit labeled data was studied, which applies bidirectional gated recurrent unit and MEDA, and the MEDA was used to synchronize sample distributions between different domains. In [37], the JGSA was applied to mitigate both distributional and geometric disparities across domains. Coupled with transfer sparse coding, this approach yielded an effectively fault diagnosis model for bearings. From the aforementioned literature review, it can be learned that the main advantages of FTL methods are low-cost, time-saving and strong ability of distribution adaptation.

Although the DL, DTL and FTL methods have been successfully used for bearing fault diagnosis and many research results have shown their enormous potential for practical application,

most intelligent diagnostic models still face challenges that will block the application in actual industrial scenarios: distribution shifting in variable working conditions, insufficient labeled fault samples and data samples imbalance in real industrial environments [14], [21], [38]. The problem of distribution shifting in variable working conditions has attracted a lot of attention, but, to the best of our knowledge, research on the issue of data samples imbalance is still quite rare. In actual industrial scenes, the collected data samples are often insufficient and unbalanced, and this problem can cause the obviously performance degradation of AI-based fault diagnosis models. Considering the above issues of bearing fault diagnosis in real-world industrial application, this work investigates a novelty frame for cross-domain fault diagnosis based on deep transfer CNN (DTCNN) extraction and supervised joint matching (SJM), termed as DTCNN-SJM, which can deal with the problem of various working conditions-caused distribution shifting under unbalanced data samples. The DTCNN-SJM model exhibits a commendable capability for cross-domain fault diagnosis with robust generalization. The main contributions of this work are described below.

1) This work utilizes continue wavelet transform (CWT) to process raw bearing vibration signals and extract the wavelet time-frequency images (WTFI). On this basis, a DTCNN, which is obtained by transferring the pre-trained CNN parameters and fine-tuning with the use of normal status data of target domain, is employed for extracting deep features from the WTFI.
2) A new FTL-based domain adaptation approach, SJM, is proposed for further processing the deep features. SJM can perform joint feature distribution matching and instance reweighting with the consideration of maximum marginal criterion, which includes four aspects: ① Introduce the manifold subspace learning for the issues of feature distortion in high-dimensional and complex feature space; ② Propose joint feature distribution matching for achieving robust distribution adaptation and overcoming the class imbalance issue; ③ Introduce instance reweighting for diminishing the distributional disparities under significant domain discrepancies; ④ Design a new maximum marginal criterion for improving the discriminability of feature data after domain adaptation. The above explorations can help to achieve more effective and robust domain adaptation while strengthening the discriminability of features.
3) To address the persistent challenges in implementing AI-based fault diagnosis approaches within real-world scenarios, we have integrated the strengths of DTL and FTL methods to develop a novel cross-domain fault diagnosis framework named DTCNN-SJM. By using two bearing datasets, it sets extensive cross-domain fault diagnosis experimental tasks that simulate the conditions of insufficient labeled fault samples and data samples imbalance in real industrial environments. The results of experimental analysis validate the outstanding and robust fault diagnosis performance.

The remaining of this article is organized as follows. In Section II, it introduces the preliminary knowledges of CNN, domain adaptation and maximum mean discrepancy. Section III presents the DTCNN-SJM framework. The details of experimental datasets and validation results are illustrated in Section IV. Section V concludes this study.

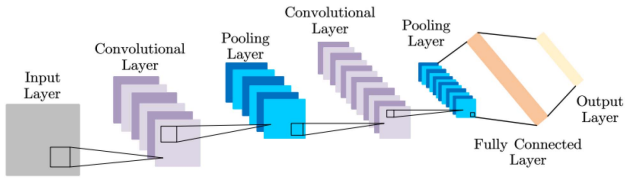


Fig. 1. Typical structure of a CNN.

II. Preliminaries

A. Convolutional Neural Network (CNN)

CNN has achieved outstanding application results in machinery fault diagnosis based on its strong deep features automatic extraction and pattern classification capabilities. This work utilizes the CNN to construct a deep features extraction model, and this section briefly describes the principle of CNN as follows.

Fig. 1 depicts a standard architecture of a CNN, in this structure, and it includes an input layer, multiple convolutional layers (CL), pooling layers (PL), a fully-connected layer (FCL), and an output layer [39].

1) Convolution Layer (CL): The CL employs multiple convolution kernels to process input data, applying the activation function (AF) to generate feature vectors. Take the l-th CL as an example, the convolution operation can be described by the subsequent equation:

$$x_j^l = f_{act}\left(\sum_i x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{1}$$

where x_j^l represents the j-th feature map generated by the l-th layer, x_i^{l-1} is the i-th input feature map from the $(l\text{-}1)$-th layer. The input feature map is a specific input data pattern, and the k_{ij}^l represents the j-th kernel connected to the i-th input feature map, $f_{act}()$ represents the AF, b_j^l is the bias vector, and $*$ represents the convolution operation [40].

2) Activation Function (AF): The AF is a crucial component in a CNN, and in this study, the Rectified Linear Unit (ReLU) is employed as the AF to alleviate issues of gradient explosion or vanishing [40]. The functional form of ReLU is denoted by:

$$f_{act}(x) = \mathrm{ReLU}(x) = \max(0,x) \tag{2}$$

3) Pooling Layer (PL): In CNN, a PL is employed to perform downsampling, which simplifies and refines the output feature map by reducing its dimensionality. Various techniques exist for this pooling process, which can decrease computational expenses and assist in extracting only the essential data.

4) Fully Connected Layer (FCL): In a CNN, the obtained feature map by several CLs and PLs is input to the FCL for the further feature extraction, and FCL combines with softmax or other classifiers to realize the fault pattern recognition and classification [41].

B. Domain Adaptation and Maximum Mean Discrepancy

Addressing the issue of distribution shift under varying operational conditions, domain adaptation (DA) emerges as a promising approach grounded in transfer learning and has been gaining increasing attention [1]. Within the realm of bearing fault diagnosis, training data (labeled fault data from a specific working condition) are designated as source domain data, while testing data (unlabeled fault data from different working conditions) constitute the target domain data. Accordingly, given source and target domains' data $D^S = \{X^S, Y^S\} = \{x_i, y_i\}_{i=1}^{n_S}$ and $D^T = \{X^T\} = \{x_i\}_{i=n_s+1}^{n_T}$, n_S andn_T represent the number of source and target domains' data, respectively. The primary objective of DA techniques is to mitigate the impact of distribution shift between the source and target domains, striving to minimize these distribution discrepancies as effectively as possible [14], [23].

Maximum mean discrepancy (MMD) is a prevalent nonparametric approach for estimating distributional differences in TL approaches [42]. The MMD between distributions of source and target domains can be given as follows:

$$\left\| \frac{1}{n_S} \sum_{x_i \in X^S} \phi(x_i) - \frac{1}{n_T} \sum_{x_j \in X^T} \phi(x_j) \right\|_H^2 \tag{3}$$

where H is the reproducing kernel Hilbert space, $\phi(\cdot)$ represents the transformation function that can transforms x_i into H. On this basis, the implement of the goal of DA method is to minimize the MMD between different domains [28].

III. DTCNN-SJM Framework

To address the persistent challenges of distribution shift and sample imbalance in the deployment of AI-based fault diagnosis methods in practical settings, this work explores a novelty cross-domain fault diagnosis framework, DTCNN-SJM. The architecture of DTCNN-SJM is depicted in Fig. 2, and it is composed of four steps: time-frequency images extraction, deep features extraction by DTCNN, domain adaptation by SJM, and fault diagnosis model training and testing.

A. Time-Frequency Images Extraction

Addressing the nonlinearity and non-stationarity inherent in bearing vibration signals, time-frequency transformation proves to be a potent method for analysis. Time-frequency analysis illustrates the signal's concurrent distribution across time domain and frequency domain. In this work, the CWT processes the raw vibration signals to yield two-dimensional time-frequency images. Subsequently, these time-frequency images are fed into a DTCNN to facilitate the extraction of deep features.

B. DTCNN Construction and Deep Features Extraction

In this work, in an effort to automatic extract deep features with better representation capability, we construct 2 CNN models shared the same network structure, in first, one CNN model is trained by the 2D time-frequency images of source domain and employed for deep feature extraction of source domain. Then another CNN model, called DTCNN, is obtained by transferring the trained CNN parameters and fine-tuning with the use of normal status data of target domain. DTCNN is employed for deep feature extraction of target domain. The detailed description of DTCNN construction and deep features extraction is as follows.

TABLE I
The Main Network Layer Parameters of the Built CNN

Network layer	Number	Size
CL 1	16	5×5
PL 1	16	2×2
CL 2	32	5×5
PL 2	32	2×2
CL 3	32	5×5
PL 3	32	2×2
FCL 1	128	—
FCL 2	10	—
Softmax layer	—	—

Firstly, it builds a CNN model that consists of three CL, three PL, two FCL, and softmax classifier layer. Each CL incorporates a batch normalization layer, and the AF is ReLU fuction. The principal parameters of the CNN are presented in Table I. Secondly, the constructed CNN model is trained by tagged source domain time-frequency images, and a trained CNN model and the corresponding network parameters (weight matrix and bias vector) are learned. Thirdly, the trained parameters of the CNN are transferred to the DTCNN to adapt to the target domain. Subsequently, time-frequency images representing the normal status from the source domain are utilized to fine-tune the DTCNN model. Finally, the CNN trained by source domain time-frequency images is used to further extract deep features, and the output of the FCL 1 is source domain deep features (SDDF). The DTCNN obtained by model parameters transfer learning and fine-tuning is applied to extract deep features, and the output of the FCL 1 is target domain deep features (TDDF).

C. Supervised Joint Matching

Although some TL-based DA approaches can achieve effective cross-domain fault diagnosis by minimizing distribution differences between different domains as much as possible, four important problems of many existing TL-based DA methods still need be considered and investigated [21], [35], [36], [37]. (1) Distribution alignment in high-dimensional and complex feature space will face the issues of feature distortion and poor domain adaptation performance. (2) The enhancement of feature discriminability is usually ignored, which may limit the cross-domain fault diagnostic accuracy after domain adaptation. (3) Underutilization of the effect of category information and neighborhood relationships of data, which may help to improve the cross-domain fault diagnostic performance and generalization ability [44], [45]. (4) In most existing TL-based domain adaptation methods, the class imbalance of data is not considered carefully during domain adaptation process. The above

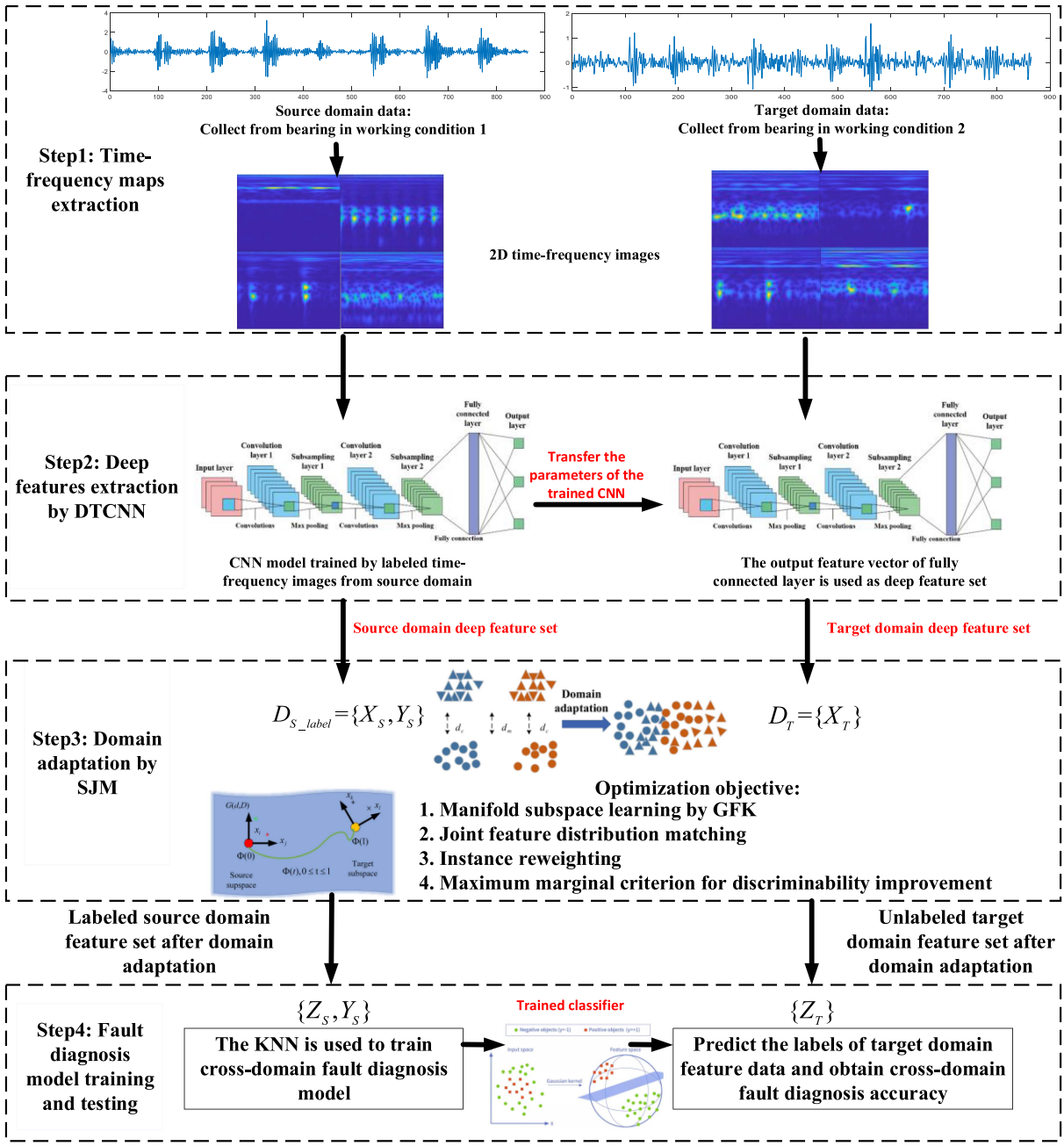


Fig. 2. The architecture of the DTCNN-SJM framework.

problems will block the achievement of desirable cross-domain fault diagnosis accuracy towards actual industrial scenes with data imbalance. Therefore, this work proposes a new TL-based domain adaptation method to overcome the above mentioned challenges, supervised joint matching (SJM). The optimization objective of SJM contains four aspects: A. Introduce the manifold subspace learning for overcoming the issues of feature distortion in high-dimensional and complex feature space; B. Propose joint feature distribution matching for achieving robust distribution alignment and overcoming the class imbalance issue; C. Introduce instance reweighting for diminishing the distributional disparities under significant domain discrepancies; D. Design a new maximum marginal criterion for improving the discriminability of feature data after domain adaptation. The details of SJM are as follows.

1) Manifold Subspace Learning (MSL): Aiming at that distribution alignment in high-dimensional and complex feature space will face the issues of feature distortion and poor domain adaptation performance, this work incorporates the classical GFK for unsupervised MSL. The objective of GFK is to extract a manifold subspace from a set of high-dimensional features, in such a way that the resultant manifold subspace is resilient to feature distortion [31], [32]. Given that the SDDF and TDDF are respectively expressed as X_S and X_T. The deep feature samples are process by GFK, which achieves that the deep features are mapped into Grassmann manifold space G(d), where d

represents the dimension of the learned manifold space. Therefore, X_S and X_T can be transformed to new feature sets Z_S and Z_T, respectively. The detailed principle of GFK can be found in reference [31].

2) Joint Feature Distribution Matching (JFDM): Due to that the class imbalance issue is not considered carefully for domain adaptation in the existing feature transfer learning methods (TCA, BDA, JDA, and etc.), they are usually difficult to attain desirable domain adaptation performance under class imbalance, which may lead to undesirable labels prediction and low cross-domain fault diagnosis accuracy. Therefore, this work proposes a new idea, JFDM, and it incorporates the joint weighted balanced distribution alignment in the optimization objective of distribution adaptation, which can adaptively change the weight of each class when performing the distribution adaptation, accordingly, it can balance the class proportion of each domain to overcome the class imbalance.

After the step of MSL, the new source domain feature set is $Z_S = \{z_i\}_{i=1}^{n_S}$ and the corresponding label set is $Y_S = \{y_i\}_{i=1}^{n_S}$, the new target domain feature set is $Z_T = \{z_j\}_{j=1}^{n_T}$. n_S and n_T represent the number of source and target domains feature data, respectively. The JFDM is employed to minimize discrepancies in their distributions, and this method encompasses three components: marginal distribution alignment (MDA), conditional distribution alignment (CDA), and joint weighted balanced distribution alignment (JWBDA), which collectively facilitate the adaptation of the data distributions across domains.

① MDA

The marginal distributions of Z_S and Z_T are $P_{MD}(Z_S)$ and $P_{MD}(Z_T)$. The MDA is conducted by minimizing the MMD between $P_{MD}(Z_S)$ and $P_{MD}(Z_T)$ [27]. The expression of MMD between $P_{MD}(Z_S)$ and $P_{MD}(Z_T)$ is defined as

$$\begin{aligned} &MMD_H^2\left(P_{MD}(Z_S), P_{MD}(Z_T)\right) \\ &= \left\| \frac{1}{n_S} \sum_{z_i \in Z_S} \phi(z_i) - \frac{1}{n_T} \sum_{z_j \in Z_T} \phi(z_j) \right\|_{\mathrm{H}}^2 \\ &= tr\left(W^T Z L_0 Z^T W\right) \end{aligned} \tag{4}$$

where H is the RKHS, $tr(W^T Z L_0 Z^T W)$ is the trace of matrix$W^T Z L_0 Z^T W$, $W$is optimal transformation matrix, and Z is constituted by Z_S and Z_T. The matrix L_0 is defined as

$$L_0 = \begin{cases} 1/n_S^2 & (z_i, z_j \in Z_S) \\ 1/n_T^2 & (z_i, z_j \notin Z_S) \\ -1/n_S n_T & (otherwise) \end{cases} \tag{5}$$

By minimizing the (4), we can derive a new representation $W^T Z$that reduces the disparity in marginal distributions between the source and target domains.

② CDA

The conditional distributions of Z_S and Z_T are $Q_{CD}(Y_S|Z_S)$ and $Q_{CD}(Y_T|Z_T)$, and because the $Y_T$is unknown, therefore, it use base-line classifier learned by $Z_S$with $Y_S$to predict the label of Z_T, and the pseudo labels $\hat{Y}_T$can replace the unknown Y_T. The expression of MMD between $Q_{CD}(Y_S|Z_S)$ and $Q_{CD}(\hat{Y}_T|Z_T)$is as follows:

$$\begin{aligned} &\sum_{c=1}^{C} MMD_H^2\left(Q_{CD}(Y_S|Z_S), Q_{CD}(\hat{Y}_T|Z_T)\right) \\ &= \sum_{c=1}^{C} \left\| \frac{1}{n_S^{(c)}} \sum_{z_i \in Z_S^{(c)}} \phi(z_i) - \frac{1}{n_T^{(c)}} \sum_{z_j \in Z_T^{(c)}} \phi(z_j) \right\|_H^2 \\ &= \sum_{c=1}^{C} tr(W^T Z L_c Z^{\mathrm{T}} W) \end{aligned} \tag{6}$$

where c is the category of feature data, $Z_S^{(c)} = \{z_i : z_i \in Z_S \wedge y(z_i) = c\}$represents the feature data pertaining to category c, and $Z_T^{(c)} = \{z_i : z_i \in Z_T \wedge \hat{y}(z_i) = c\}$ are also feature data belonging to class c. $\hat{y}(z_i)$ represents the pseudo label of z_i. $n_S^{(c)}$and $n_T^{(c)}$ represent the number of feature samples belonging to class c from Z_S and Z_T, respectively. The definition of matrix L_c is as follows:

$$L_c = \begin{cases} \frac{1}{n_S^{(c)} n_S^{(c)}} & \left(z_i, z_j \in Z_S^{(c)}\right) \\ \frac{1}{n_T^{(c)} n_T^{(c)}} & \left(z_i, z_j \in Z_T^{(c)}\right) \\ \frac{-1}{n_S^{(c)} n_T^{(c)}} & \begin{cases} z_i \in Z_S^{(c)}, z_j \in Z_T^{(c)} \\ z_i \in Z_T^{(c)}, z_j \in Z_S^{(c)} \end{cases} \\ 0 & (otherwise) \end{cases} \tag{7}$$

By minimizing the (6), we can derive a new representation $W^T Z$that reduces the disparity in conditional distributions betweenZ_S and Z_T.

③ JWBDA

Aiming at that the MDA and CDA are often treated equally in most existing TL-based domain adaptation methods, and the class imbalance is not considered carefully for domain adaptation. Therefore, this work designs JWBDA to overcome the above problems, and it uses balance factor to adaptively adjust the importance of both the MDA and CDA, and the weight of each class can be adaptively changed in the procedure of domain adaptation. The optimal expression of JWBDA is as follows:

$$\begin{aligned} &(1-\mu) MMD_H^2\left(P_{MD}(Z_S), P_{MD}(Z_T)\right) \\ &+\mu \left(\sum_{c=1}^{C} MMD_H^2\left(\hat{Q}_{CD}(Y_S|Z_S), \hat{Q}_{CD}\left(\hat{Y}_T|Z_T\right)\right)\right) \end{aligned} \tag{8}$$

whereμ is the balance factor, and the $\mu \in [0,1]$. $\sum_{c=1}^{C} MMD_H^2(\hat{Q}_{CD}(Y_S|Z_S), \hat{Q}_{CD}(\hat{Y}_T|Z_T))$ is obtained by improving the (6), the MMD matrix L_c is improved, and the new MMD

matric $\hat{L}_c$ is defined as follows:

$$\hat{L}_c = \begin{cases} \frac{P(y_S^{(c)})}{n_S^{(c)} n_S^{(c)}} & (z_i, z_j \in Z_S^{(c)}) \\ \frac{P(y_T^{(c)})}{n_T^{(c)} n_T^{(c)}} & (z_i, z_j \in Z_T^{(c)}) \\ \frac{-\sqrt{P(y_S^{(c)}) P(y_T^{(c)})}}{n_S^{(c)} n_T^{(c)}} & \begin{cases} z_i \in Z_S^{(c)}, z_j \in Z_T^{(c)} \\ z_i \in Z_T^{(c)}, z_j \in Z_S^{(c)} \end{cases} \\ 0 \quad (otherwise) \end{cases} \tag{9}$$

where the $P(y_S^{(c)})$and $P(y_T^{(c)})$ are respectively the class prior on class c in the source and target domains.

The (9) is obtained by improving the (7), the purpose of this improvement is follows: in order to overcome the issue of data class imbalance in the procedure of domain adaptation, the (9) can adaptively change the weight of each class when performing the distribution adaptation, therefore, it can balance the class proportion of each domain to overcome the class imbalance.

3) Instance Reweighting: Due to that significant domain discrepancies will inevitably result in the presence of source instances (SI) that do not have relevance to the target instances (TI), the domain adaptation performance will be degraded [46]. Therefore, to further diminish the distributional disparities, this work incorporates instance reweighting into the domain adaptation process, and it can adaptively assign greater importance to SI that is relevant to the target, while reducing the weighting of irrelevant ones. Accordingly, the $l_{2,1}$-norm structured sparsity regularizer is imposed on the mapping matrix$W$, which can introduce row-sparsity to the *W*. Because each row of *W* corresponds to a specific instance, such sparsity inherently facilitates adaptive instance reweighting [46]. The respective instance reweighting regularizer is formulated as:

$$\|W_S\|_{2,1} + \|W_T\|_F^2 \tag{10}$$

where the transformation matrices corresponding to the source and target instances are respectively denoted by $W_S$and W_T. $l_{2,1}$-norm regularizer is imposed on the SI to reweight them in the light of their relevance to the TI. By minimizing the (10), it adaptively assigns greater importance to SI that are relevant to the target, while reducing the weighting of irrelevant ones. This instance reweighting scheme enhances the robustness of domain adaptation, particularly in cases where there is a significant discrepancy due to irrelevant instances.

4) Maximum Marginal Criterion for Discriminability Improvement: Due to that the enhancement of feature discriminability is usually ignored in most existing feature transfer learning methods, only considering distribution alignment in domain adaptation will lead to poor labels prediction ability and the degradation of fault diagnosis accuracy. Therefore, this work designs a new maximum marginal criterion matrix (MMCM) for improving the discriminability of feature data, in this design, the effect of category information and neighborhood relationships of feature data are considered, and it can expand the boundaries of feature data from different classes. Inspired by the principles of local fisher discriminant analysis (LFDA) [47], the MMCM is constructed based on local within-class and between-classes scatter matrices, the expression of MMCM is as follows:

$$\text{MMCM} = I / \left(S_b^{LF} - S_w^{LF} \right) \tag{11}$$

where the $S_w^{LF}$and S_b^{LF} are within-class and between-classes scatter matrices, the I is the identity matrix. The expressions of $S_w^{LF}$and $S_b^{LF}$are as follows:

$$S_w^{LF} = \frac{1}{2} \sum_{i,j=1}^{n} \tilde{p}_{ij}^{LW} \left(z_i - z_j \right) \left(z_i - z_j \right)^T \tag{12}$$

$$S_b^{LF} = \frac{1}{2} \sum_{i,j=1}^{n} \tilde{p}_{ij}^{Lb} \left(z_i - z_j \right) \left(z_i - z_j \right)^T \tag{13}$$

where

$$\tilde{p}_{ij}^{Lb} = \begin{cases} A_{ij}(1/m - 1/m_l), & z_i = z_j = l, \\ m_l/m & z_i \neq z_j \left(j \in Nst(i) \right), \\ 1/m & else. \end{cases} \tag{14}$$

$$\tilde{p}_{ij}^{LW} = \begin{cases} A_{ij}/n_l, & z_i = z_j = l \\ 0, & z_i \neq z_j \end{cases} \tag{15}$$

where m is the count of feature samples, l represents the class label assigned to a feature data sample, and m_l signifies the quantity of samples that pertain to class l. $\tilde{p}_{ij}^{Lb}$and $\tilde{p}_{ij}^{LW}$ are designated as weight matrices. In$\tilde{p}_{ij}^{Lb}$, $Nst(i)$is the nearest neighbour of i and they pertains to different classes. $A_{ij} \in [0, 1]$ can be defined below:

$$A_{ij} = \exp\left(-\frac{\|z_i - z_j\|^2}{\gamma_i \gamma_j} \right) \tag{16}$$

where $\gamma_i = \|z_i - z_i^n\|$is the local scaling around z_i, z_i^n is the n-th nearst neighbour of z_i. When the $z_i$and $z_j$are closer, the A_{ij} is larger. By minimizing the (11), the feature samples of the same class will become increasingly close to each other, and the distance between feature samples of different classes will become increasingly distant. Furthermore, the local neighborhood relationships of the feature data are also considered by the $\tilde{p}_{ij}^{Lb}$ [47]. Therefore, the discriminability of feature data after domain adaptation will be improved.

5) Optimization Problem: In this study, we endeavor to reduce the disparity in distributions between the source and target domains by integrating the aforementioned four components, on the basis of (8), (10) and (11), we formulate the optimization problem for SJM as follows:

$$\begin{aligned} & \underset{W^T Z E Z^T W = I}{O_{SJM}} \left(Z_S, Z_T \right) \\ & = (1 - \mu) MMD_H^2 \left(P_{MD}(Z_S), P_{MD}(Z_T) \right) \\ & \quad + \mu \left(\sum_{c=1}^{C} MMD_H^2 \left(\hat{Q}_{CD} \left(Y_S | Z_S \right), \hat{Q}_{CD} \left(\hat{Y}_T | Z_T \right) \right) \right) \\ & \quad + \lambda \left(\|W_S\|_{2,1} + \|W_T\|_F^2 \right) \\ & \quad + \eta tr \left(W^T K \left(I / \left(S_b^{LF} - S_w^{LF} \right) \right) K W \right) \end{aligned} \tag{17}$$

where the $\mu \in [0,1]$and $\eta \in [0,1]$are adjustable parameters, μ modulates the balance between marginal and conditional distributions adaptation, λ represents the regularization parameter that tunes the trade-off among JFDM, instance reweighting, and discriminability improvement, $I \in \mathbb{R}^{(n_S+n_T)\times(n_S+n_T)}$and E are respectively the unit matrix and centering matrix, respectively. On the basis of the (4) and (6), it can further infer the $O_{SJM}(Z_S, Z_T)$ to

$$\begin{aligned} &\underset{W^T Z E Z^{\mathrm{T}} W=I}{O_{SJM}(Z_S, Z_T)} \\ &= \mu \sum_{c=1}^{C} tr\left(W^{\mathrm{T}} Z \hat{L}_c Z^{\mathrm{T}} W\right) + (1-\mu)\, tr\left(W^T Z L_0 Z^{\mathrm{T}} W\right) \\ &\quad + \lambda\left(\|W_S\|_{2,1} + \|W_T\|_F^2\right) \\ &\quad + \eta tr\left(W^T K\left(I/\left(S_b^{LF} - S_w^{LF}\right)\right) K W\right) \end{aligned} \tag{18}$$

According to the principles of constrained optimization, we define $\Phi = \mathrm{diag}(\phi_1, \phi_2, \cdots, \phi_k) \in \mathbb{R}^{k\times k}$ as the Lagrange multiplier and construct the corresponding function to address problem (18) as follows:

$$\begin{aligned} L = tr &\left(W^{\mathrm{T}} Z(\mu \sum_{c=1}^{C} \hat{L}_c + (1-\mu)L_0 + \eta\left(I/\left(S_w^L - S_b^L\right)\right)) Z^{\mathrm{T}} W\right) \\ &+ \lambda\left(\|W_S\|_{2,1} + \|W_T\|_F^2\right) + tr\left(\left(I - W^T Z E Z^{\mathrm{T}} W\right)\Phi\right) \end{aligned} \tag{19}$$

Setting $\partial L/\partial W = 0$, generalized eigendecomposition can be obtained as follows:

$$\begin{aligned} &\left(Z(\mu \sum_{c=1}^{C} \hat{L}_c + (1-\mu)L_0 + \eta\left(I/\left(S_w^L - S_b^L\right)\right)) Z^{\mathrm{T}} + \lambda G\right) \\ &W = Z E Z^{\mathrm{T}} W \Phi \end{aligned} \tag{20}$$

$\|W_S\|_{2,1}$is an non-smooth function at zero, therefore, it can calculate its sub-gradient as $\partial(\|W_S\|_{2,1} + \|W_T\|_F^2)/\partial W = 2GW$ [46], where G is a diagonal sub-gradient matrix as follows:

$$G_{ii} = \begin{cases} 1/2\left\|w^i\right\|, & z_i \in Z_S, w^i \neq 0 \\ 0, & z_i \in Z_S, w^i = 0 \\ 1, & z_i \in Z_T \end{cases} \tag{21}$$

where $w^i$is the ith row of W. Obtaining the optimal matrix W is reduced to solving the (20) for the h smallest eigenvectors, and new feature sets $U_S = W^T Z_S$ and $U_T = W^T Z_T$are acquired. Then, it can use labelled $U_S$to learn an pattern recognition classifier f which is then applied to derive the category labels of U_T.

D. Complete Process Steps of the DTCNN-SJM

On the basis of the cross-domain fault diagnosis framework DTCNN-SJM, the complete process steps are presented in detail below.

Step 1: The collected bearing vibration signals with class labels under a known working condition are denoted as source domainD_{source}; The collected bearing vibration signals without class labels under a unknown working condition are denoted as target domain D_{target}. The $D_{source}$and $D_{target}$are inputted in the DTCNN-SJM. Then, D_{source} and $D_{target}$are processed by the CWT to yield two-dimensional time-frequency images, and they are respectively denoted as $TFI_{source}$and TFI_{target}.

Step 2: In first, construct a CNN model for source domain, the $TFI_{source}$and the corresponding class labels are employed for train this CNN model and the network parameters of the trained CNN are obtained. Then, the obtained model parameters are transferred to a new CNN model (called DTCNN) for target domain, and the time-frequency images representing the normal status from TFI_{target} are employed to fine-tune the DTCNN. Accordingly, the CNN model for source domain is used to extract deep features from TFI_{source}, denoted as X_{source}. The fine-tuned DTCNN is utilized to extract deep features from TFI_{target}, denoted as X_{target}.

Step 3: X_{source}, the class labels $Y_{source}$of X_{source}, and X_{target} are inputted in the proposed SJM, and the parameters (manifold subspace dimension d, parametersλ,η, andμ, the dimension of feature set after domain adaptation h) involved in SJM are set. After the process of the SJM, the new source domain feature set Z_{source} and target domain feature set Z_{target} are obtained for the next step.

Step 4: $Z_{source}$and Y_{source} are used to train the KNN classifier for cross-domain fault diagnosis, and the trained KNN classifier is employed to predict the labels ofZ_{target}, finally, the cross-domain fault diagnosis results are calculated.

IV. Experimental Validation

To validate the efficacy and generalizability of the proposed methods, this study carries out cross-domain fault diagnosis experiments on two types of bearing fault datasets (Case Western Reserve University (CWRU) experimental platform [2], [14], [34], [35], [48] and SQI-MFS test platform [49], [50], [51]) under balanced and unbalanced data samples. Furthermore, to demonstrate the superiority of the proposed DTCNN-SJM, some models are built by ready-made various established techniques for comparison.

A. Case 1: Cross-Domain Fault Diagnostic Based on CWRU Test Platform

1) Introduction of Dataset and Fault Diagnosis Tasks: Fig. 3 illustrates the bearing faults experimental platform developed by (CWRU). This platform employs accelerometer sensors to capture bearing vibration signals at a sampling frequency of 12 kHz. The platform operates under four bearing load conditions: motor loads of 0 hp, 1 hp, 2 hp, and 3 hp, equivalent to rotational speeds of 1797 rmp, 1772 rmp, 1750 rmp, and 1730 rmp, respectively. The collected data includes three types of bearing defects: inner ring defect(IRD), rolling element defect(RED), and outer ring defect (ORD), with defect sizes of 0.021inch, 0.014inch and 0.007inch. Table II displays the bearing vibration

TABLE II
THE CWRU BEARING DATA FOR CROSS-DOMAIN FAULT DIAGNOSIS EXPERIMENTS

Bearing defect	Defect diameter (inches)	Number of samples(training/validating/testing)				Category labels
		0hp	1hp	2hp	3hp	
No defect	0	120/40/40	120/40/40	120/40/40	120/40/40	1
IRD	0.007	120/40/40	120/40/40	120/40/40	120/40/40	2
	0.014	120/40/40	120/40/40	120/40/40	120/40/40	3
	0.021	120/40/40	120/40/40	120/40/40	120/40/40	4
ORD	0.007	120/40/40	120/40/40	120/40/40	120/40/40	5
	0.014	120/40/40	120/40/40	120/40/40	120/40/40	6
	0.021	120/40/40	120/40/40	120/40/40	120/40/40	7
RED	0.007	120/40/40	120/40/40	120/40/40	120/40/40	8
	0.014	120/40/40	120/40/40	120/40/40	120/40/40	9
	0.021	120/40/40	120/40/40	120/40/40	120/40/40	10

TABLE III
CROSS-DOMAIN FAULT DIAGNOSIS TASKS UNDER BALANCED TRAINING SAMPLES (CASE 1)

Tasks	Training samples (SD)			Testing samples (TD)		
	Motor load	Defect types of samples	Number of samples	Motor load	Defect types of samples	Number of samples
1	0hp	1-10	1200	1hp	1-10	400
2	0hp	1-10	1200	2hp	1-10	400
3	0hp	1-10	1200	3hp	1-10	400
4	1hp	1-10	1200	0hp	1-10	400
5	1hp	1-10	1200	2hp	1-10	400
6	1hp	1-10	1200	3hp	1-10	400
7	2hp	1-10	1200	0hp	1-10	400
8	2hp	1-10	1200	1hp	1-10	400
9	2hp	1-10	1200	3hp	1-10	400
10	3hp	1-10	1200	0hp	1-10	400
11	3hp	1-10	1200	1hp	1-10	400
12	3hp	1-10	1200	2hp	1-10	400

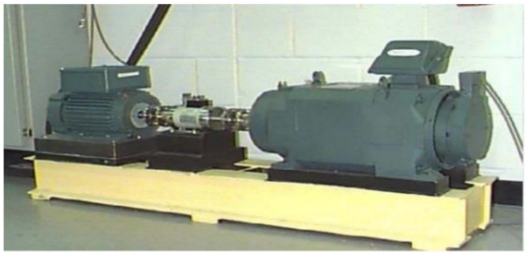


Fig. 3. CWRU test-bed.

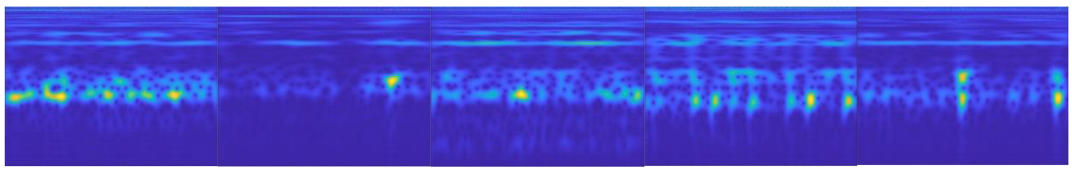
data utilized in this study. For each type of fault data, the original data is divided into 200 samples, each consisting of 864 sampling points. The training set (120 samples), validation set (40 samples), and testing set (40 samples) are divided in a 6:2:2 ratio. Consequently, we derive bearing vibration data across ten categories, designated as 1 through 10. On the basis of these bearing vibration data, 24 cross-domain fault diagnosis tasks (CFDT) under balanced and unbalanced training samples are set up, as shown in Tables III and IV. The source domain samples for both the training and validation sets are utilized to train the CNN. For the fault diagnosis tasks 1–12 under balanced data samples, the number of different types of fault samples is the same. But, the fault diagnosis under imbalanced data samples has different types of fault samples, and the number of training samples for categories 1–10 are 120, 110, 100, 90, 80, 70, 60, 50, 40, and 30, respectively. Therefore, the total number of training samples is 750.

2) The Effectiveness Verification of the DTCNN-SJM Framework: In this section, we apply the DTCNN-SJM framework to conduct 24 CFDT, as delineated in Tables III and IV. Initially, we process the raw vibration signals from both the source and target domains using CWT to generate an array of two-dimensional time-frequency images (TFI). Taking 10 types of fault vibration signal samples under 0hp motor load (1797r/min speed) as an example, the corresponding time-frequency diagrams are shown in Fig. 4. Then, a DTCNN is built to extract deep features from TFI of source and target domains, and the mainly parameters of DTCNN are shown in Table I. Accordingly, the deep features of source domain (DFSD) and the deep features of target domain (DFTD) are separately extracted. Next, the labeled DFSD and unlabelled DFTD are inputted into the SJM, and it conducts the joint feature distribution matching with the instance reweighting and maximum marginal criterion in learned Grassmann manifold subspace, which can achieve more robustness domain adaptation ability. After the procedure of SJM, the new DFSD and DFTD are obtained, accordingly, the new DFSD is used to train a classifier for cross-domain fault diagnosis. Finally, this trained classifier is applied to infer the category labels of the DFTD, thus yielding the results for cross-domain fault diagnosis.

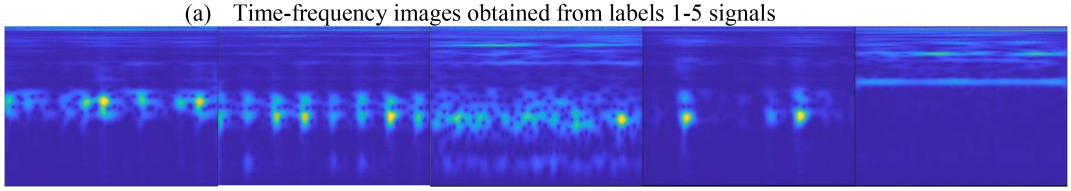
Table V presents the testing results of DTCNN-SJM framework in tasks 1–24. Tasks 1–12 are set up under balanced training

TABLE IV
CROSS-DOMAIN FAULT DIAGNOSIS TASKS UNDER UNBALANCED TRAINING SAMPLES (CASE 1)

Tasks	Training samples (SD)			Testing samples (TD)		
	Motor load	Defect types of samples	Number of samples	Motor load	Defect types of samples	Number of samples
13	0hp	1-10	750	1hp	1-10	400
14	0hp	1-10	750	2hp	1-10	400
15	0hp	1-10	750	3hp	1-10	400
16	1hp	1-10	750	0hp	1-10	400
17	1hp	1-10	750	2hp	1-10	400
18	1hp	1-10	750	3hp	1-10	400
19	2hp	1-10	750	0hp	1-10	400
20	2hp	1-10	750	1hp	1-10	400
21	2hp	1-10	750	3hp	1-10	400
22	3hp	1-10	750	0hp	1-10	400
23	3hp	1-10	750	1hp	1-10	400
24	3hp	1-10	750	2hp	1-10	400



(a) Time-frequency images obtained from labels 1-5 signals



(b) Time-frequency images obtained from labels 6-10 signals

Fig. 4. Time-Frequency Images of Various Fault Vibration Signals (Labels 1–10) under Motor Load of 3hp.

TABLE V
EXPERIMENTAL RESULTS OBTAINED BY DTCNN-SJM FRAMEWORK IN CASE 1

Tasks	Accuracies (%)	Tasks	Accuracies (%)	Tasks	Accuracies (%)	Tasks	Accuracies (%)
1	100.00	7	99.00	13	99.50	19	100.00
2	100.00	8	98.00	14	100.00	20	98.50
3	98.50	9	97.50	15	89.00	21	99.00
4	100.00	10	96.00	16	100.00	22	94.50
5	100.00	11	99.50	17	100.00	23	94.00
6	100.00	12	100.00	18	100.00	24	100.00

samples, and tasks 13–24 are set up under unbalanced training samples. For tasks 1–12, cross-domain diagnosis accuracies are respectively 100.00%, 100.00%, 98.50%, 100.00%, 100.00%, 100.00%, 99.00%, 98.00%, 97.50%, 96.00%, 99.50%, and 100.00%, the mean accuracy can attain 99.04%. For tasks 13–24, cross-domain diagnosis accuracies are respectively 99.50%, 100.00%, 89.00%, 100.00%, 100.00%, 100.00%, 100.00%, 98.50%, 99.00%, 94.50%, 94.00%, and 100.00%, the mean accuracy can attain 97.88%, which is slightly lower than mean accuracy of tasks 1–12. According to the results shown in Table V, the validity of the proposed DTCNN-SJM framework can be verified, and they also indicate that the imbalance of training samples can lead to a decrease in the fault diagnosis accuracy, nicely, the DTCNN-SJM framework can still achieve ideal cross-domain fault diagnosis results. The parameters used in SJM are as follows: the adjustable parameters $\mu$and $\eta$are 0.4 and 0.6, and the regularization parameter $\lambda$is 0.5; manifold subspace dimension d is 30, and the dimension of feature set after domain adaptation h is 30.

3) Comparative Experiments: To substantiate the superior performance of the DTCNN-SJM framework and advantages of the proposed SJM, we constructed a variety of fault diagnosis models for comparative analysis. These models utilize both conventional and widely-recognized methods such as KNN,

TABLE VI
COMPARATIVE MODELS

Specific name	Label	Specific name	Label
DTCNN-SVM	CM1	DTCNN-TJM-KNN	CM10
DTCNN-KNN	CM2	DTCNN-SJM-SVM	CM11
DTCNN-Softmax	CM3	DTCNN-SJM-KNN	CM12
DTCNN-JDA-KNN	CM4	DTCNN-CORAL-SVM	CM13
DTCNN-TCA-KNN	CM5	DTCNN-EasyTL-SVM	CM14
DTCNN-BDA-KNN	CM6	DTCNN-JPDA-SVM	CM15
DTCNN-GFK-KNN	CM7	DTCNN-MEKT-SVM	CM16
DTCNN-MEDA-KNN	CM8	DTCNN-STL-SVM	CM17
DTCNN-JGSA-KNN	CM9	DTCNN-SA-SVM	CM18

SVM, Softmax, JDA, TCA, BDA, GFK, MEDA, JGSA, TJM, CORAL [52], EasyTL [53], JPDA [54], MEKT [55], STL [56] and SA [57]. Accordingly, the Table VI lists 18 comparative models built by these methods, DTCNN and SJM. The reasons of this setup are as follows: (1) The comparative models are divided into two categories: the models without domain adaptation methods and the models integrated different classical domain adaptation methods. The models built upon SVM, KNN, and Softmax belong to the models without domain adaptation methods, and their experimental results can serve as a baseline comparison. The models built upon different domain adaptation methods (JDA, TCA, BDA, GFK, MEDA, JGSA, TJM, CORAL, EasyTL, JPDA, MEKT, STL and SA) can be compared with the models without domain adaptation methods to verify the effect of domain adaptation methods on cross-domain fault diagnosis performance. (2)The JDA, TCA, BDA, GFK, MEDA, JGSA, TJM, CORAL, EasyTL, JPDA, MEKT, STL and SA are classical feature transfer learning methods, and they have been studied by many researchers in the field of cross-domain fault diagnosis. These methods contain different mainstream optimization ideas, such as probability distributions alignment, manifold learning for domain adaptation, and joint geometrical and statistical alignment for domain adaptation. Therefore, we set comparative models constructed by them and compare the experimental results of the proposed SJM in this manuscript with theirs, which validates the superiority of the proposed SJM.

For the CM1-CM3 models, take DTCNN-SVM as an example, it means that the original vibration signals are processed by steps 1 and 2 in DTCNN-SJM framework, and deep features are extracted. Subsequently, the DFSD is used to train the machine learning classifier, and the trained classifier predicts the labels of DFTD. Consequently, the fault diagnosis results can be obtained. For the CM4-CM18 models, take DTCNN-JDA-SVM as an example, it means that the raw bearing vibration signals are processed by steps 1 and 2 in DTCNN-SJM framework, and deep features are extracted. Then, the labeled DFSD and unlabeled DFTD are inputted into JDA for domain adaptation, the next, the new DFSD after domain adaptation is used to learn the SVM classifier, and the learned SVM classifier predicts the labels of DFTD. Finally, the fault diagnosis experimental results are derived.

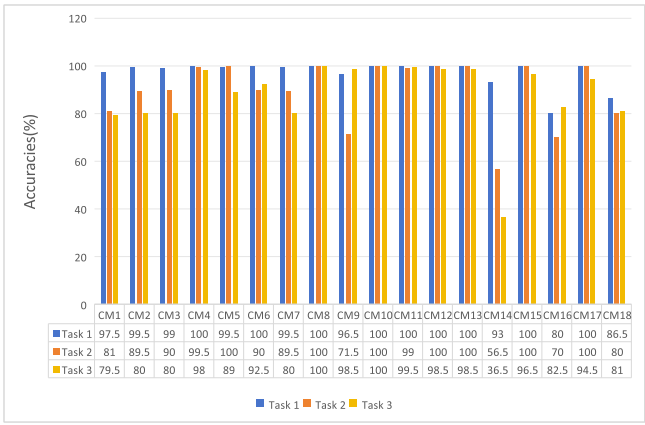


Fig. 5. Comparison of Accuracies for CM1-CM18 Models across Tasks 1–3.

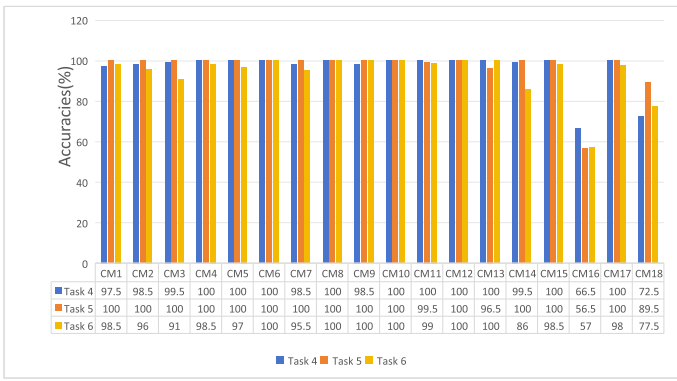


Fig. 6. Comparison of Accuracies for CM1-CM18 Models across Tasks 4–6.

Table VII shows the fault diagnosis results for the CM1-CM18 models across tasks 1–12, clearly showing that the CM11 and CM18 models, which are based on the proposed DTCNN-SJM framework, surpass the performance of the other comparative models. The mean diagnostic accuracies of CM11 and CM12 in these tasks reach 98.58% and 99.04%, respectively. By comparing the results of fault diagnosis models using different domain adaptation methods, it indicates that the performance of the models using JDA, TJM, and SJM is markedly better than models using TCA, BDA, GFK, JGSA, MEDA, CORAL, EasyTL, JPDA, MEKT, STL and SA. The performance of model using SJM is the best, the DTCNN-SJM-KNN model can attain 99.04% of mean diagnosis accuracy in tasks 1-12, which is respectively 0.63%, 3.58%, 3.50%, 8.63%, 3.21%, 9.00%, 0.54%, 5.29%, 24.79%, 4.33%, 39.79%, 7.17%, and 21.67% higher than that of models using JDA, TCA, BDA, GFK, MEDA, JGSA, TJM, CORAL, EasyTL, JPDA, MEKT, STL and SA. Figs. 5–8 present the comparison of experimental outcomes for models CM1 to CM18 across tasks 1 to 12, and the Fig. 9 gives the comparison of the average diagnostic accuracies for tasks 1 to 12 obtained by models CM1 to M18.

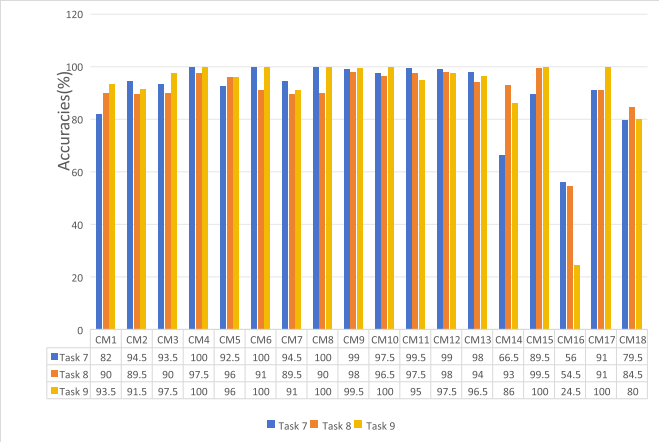
Table VIII presents the cross-domain fault diagnosis results of CM1-CM18 models in tasks 13–24, it is also apparent that the CM11 and CM12 models (based on the proposed DTCNN-SJM framework) can achieve desirable performance, which also

TABLE VII
CROSS-DOMAIN FAULT DIAGNOSIS OUTCOMES OF MODELS CM1-CM12 ACROSS TASKS 1-12

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
CM1	97.50	81.00	79.50	97.50	100.00	98.50	82.00	90.00	93.50	80.00	80.00	83.00
CM2	99.50	89.50	80.00	98.50	100.00	96.00	94.50	89.50	91.50	80.00	80.00	87.00
CM3	99.00	90.00	80.00	99.50	100.00	91.00	93.50	90.00	97.50	86.00	88.00	100.00
CM4	100.00	99.50	98.00	100.00	100.00	98.50	100.00	97.50	100.00	96.00	91.50	100.00
CM5	99.50	100.00	89.00	100.00	100.00	97.00	92.50	96.00	96.00	89.50	86.00	100.00
CM6	100.00	90.00	92.50	100.00	100.00	100.00	100.00	91.00	100.00	90.00	83.00	100.00
CM7	99.50	89.50	80.00	98.50	100.00	95.50	94.50	89.50	91.00	80.00	80.00	87.00
CM8	100.00	100.00	100.00	100.00	100.00	100.00	100.00	90.00	100.00	80.00	80.00	100.00
CM9	96.50	71.50	98.50	98.50	100.00	100.00	99.00	98.00	99.50	70.00	70.50	78.50
CM10	100.00	100.00	100.00	100.00	100.00	100.00	97.50	96.50	100.00	95.00	93.00	100.00
CM11	100.00	99.00	99.50	100.00	99.50	99.00	99.50	97.50	95.00	95.50	98.50	100.00
CM12	100.00	100.00	98.50	100.00	100.00	100.00	99.00	98.00	97.50	96.00	99.50	100.00
CM13	100.00	100.00	98.50	100.00	96.50	100.00	98.00	94.00	96.50	80.00	78.00	83.50
CM14	93.00	56.50	36.50	99.50	100.00	86.00	66.50	93.00	86.00	35.50	62.00	76.50
CM15	100.00	100.00	96.50	100.00	100.00	98.50	89.50	99.50	100.00	79.50	80.00	93.00
CM16	80.00	70.00	82.50	66.50	56.50	57.00	56.00	54.50	24.50	40.00	61.00	62.50
CM17	100.00	100.00	94.50	100.00	100.00	98.00	91.00	91.00	100.00	71.00	77.00	80.00
CM18	86.50	80.00	81.00	72.50	89.50	77.50	79.50	84.50	80.00	77.00	64.00	56.50

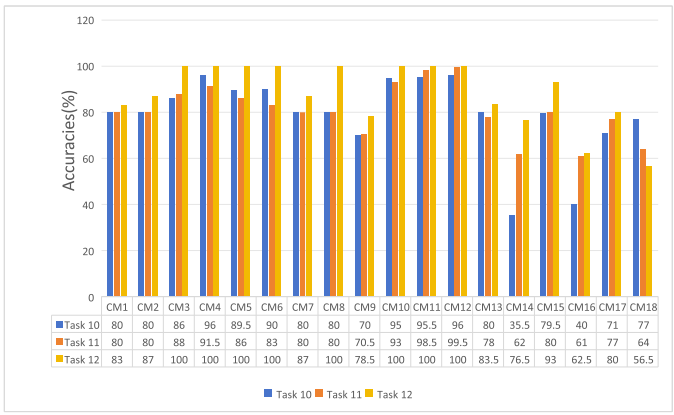
TABLE VIII
CROSS-DOMAIN FAULT DIAGNOSIS OUTCOMES OF MODELS CM1-CM12 ACROSS TASKS 13–24

Models	Accuracies (%)											
	Task 13	Task 14	Task 15	Task 16	Task 17	Task 18	Task 19	Task 20	Task 21	Task 22	Task 23	Task 24
CM1	99.50	86.50	80.00	98.00	100.00	91.50	86.00	90.00	93.50	79.00	79.50	80.50
CM2	99.50	89.50	80.00	99.00	100.00	90.50	94.00	90.00	93.00	79.50	80.00	84.50
CM3	99.00	90.00	80.00	99.00	98.50	91.00	94.50	90.00	97.00	78.50	78.50	90.50
CM4	98.50	95.50	87.50	98.50	100.00	99.50	95.00	98.50	99.50	94.50	92.00	100.00
CM5	97.00	88.00	83.00	100.00	100.00	100.00	91.50	96.00	95.50	83.50	84.00	97.50
CM6	98.50	90.00	80.00	100.00	100.00	90.00	100.00	93.00	100.00	89.50	98.50	100.00
CM7	99.50	89.50	80.00	99.00	100.00	90.50	94.00	90.00	91.50	79.50	80.00	83.50
CM8	99.00	90.00	80.00	99.50	100.00	100.00	100.00	90.00	100.00	80.00	80.00	100.00
CM9	98.50	99.50	83.50	98.50	100.00	100.00	99.50	99.00	99.00	91.50	77.50	99.00
CM10	98.50	100.00	87.50	100.00	100.00	98.50	100.00	90.00	98.00	93.00	91.00	100.00
CM11	98.00	97.50	89.50	99.50	100.00	97.50	97.50	97.50	97.00	95.00	94.50	100.00
CM12	99.50	100.00	89.00	100.00	100.00	100.00	100.00	98.50	99.00	94.50	94.00	100.00
CM13	93.00	99.00	81.50	100.00	100.00	98.00	97.00	92.50	98.50	81.00	78.00	85.50
CM14	66.50	66.00	60.50	99.50	100.00	83.50	72.00	98.00	99.00	56.00	83.50	87.00
CM15	86.50	100.00	87.50	100.00	100.00	100.00	90.00	99.50	100.00	80.00	80.00	90.50
CM16	91.00	62.00	22.50	72.50	80.50	77.00	83.00	64.00	71.50	48.50	48.00	54.50
CM17	92.00	93.00	80.00	99.50	100.00	97.00	95.00	91.50	95.00	75.00	76.50	80.00
CM18	81.00	72.50	66.50	77.50	70.00	64.00	66.50	71.00	56.50	61.00	60.50	51.50



	CM1	CM2	CM3	CM4	CM5	CM6	CM7	CM8	CM9	CM10	CM11	CM12	CM13	CM14	CM15	CM16	CM17	CM18
Task 7	82	94.5	93.5	100	92.5	100	94.5	100	99	97.5	99.5	99	98	66.5	89.5	56	91	79.5
Task 8	90	89.5	90	97.5	96	91	89.5	90	98	96.5	97.5	98	94	93	99.5	54.5	91	84.5
Task 9	93.5	91.5	97.5	100	96	100	91	100	99.5	100	95	97.5	96.5	86	100	24.5	100	80

Fig. 7. Comparison of Accuracies for CM1-CM18 Models across Tasks 7–9.



	CM1	CM2	CM3	CM4	CM5	CM6	CM7	CM8	CM9	CM10	CM11	CM12	CM13	CM14	CM15	CM16	CM17	CM18
Task 10	80	80	86	96	89.5	90	80	80	70	95	95.5	96	80	35.5	79.5	40	71	77
Task 11	80	80	88	91.5	86	83	80	80	70.5	93	98.5	99.5	78	62	80	61	77	64
Task 12	83	87	100	100	100	100	87	100	78.5	100	100	100	83.5	76.5	93	62.5	80	56.5

Fig. 8. Comparison of Accuracies for CM1-CM18 Models across Tasks 10–12.

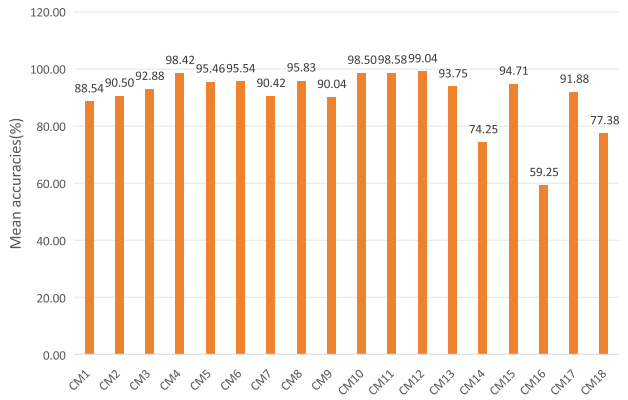
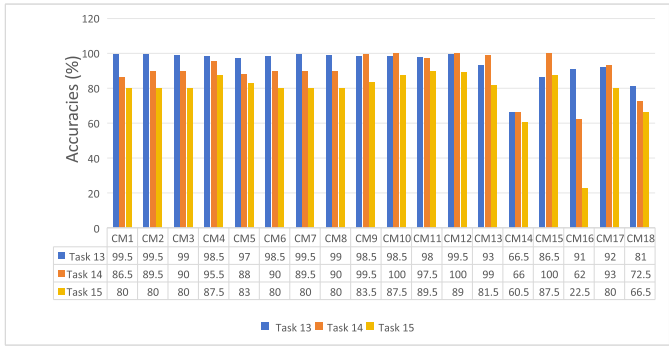


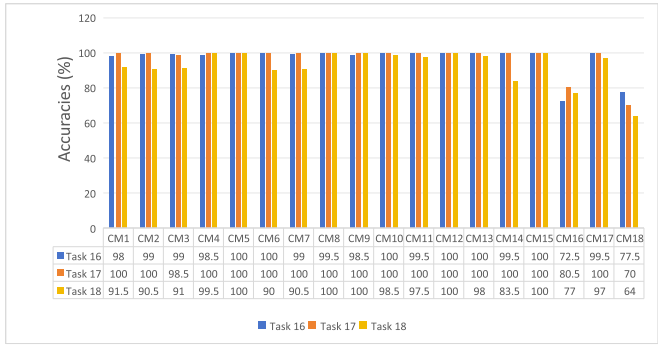
Fig. 9. The comparison of the average diagnostic accuracies for tasks 1 to 12 obtained by models CM1 to CM18.

supasses other models, and the mean diagnostic accuracies of CM11 and CM12 models in tasks 1–12 can attain 96.96% and 97.88%, respectively. From the experimental results of fault diagnosis models using different domain adaptation methods, it is evident that the performance of the models using JDA, TJM, and SJM is obviously better than models using TCA, BDA, GFK, JGSA, MEDA, CORAL, EasyTL, JPDA, MEKT, STL and SA. The DTCNN-SJM-KNN model can attain 97.88% of mean



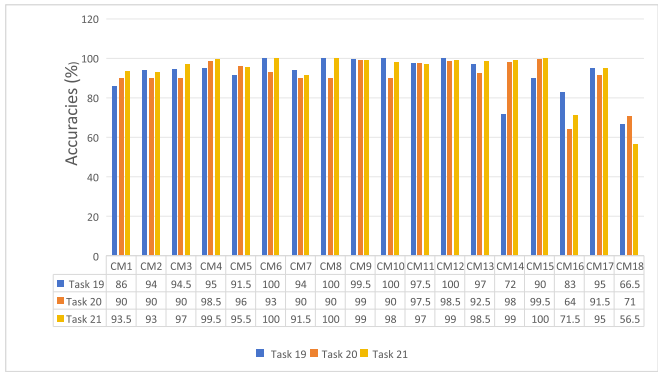
	CM1	CM2	CM3	CM4	CM5	CM6	CM7	CM8	CM9	CM10	CM11	CM12	CM13	CM14	CM15	CM16	CM17	CM18
Task 13	99.5	99.5	99	98.5	97	98.5	99.5	99	98.5	98.5	98	99.5	93	66.5	86.5	91	92	81
Task 14	86.5	89.5	90	95.5	88	90	89.5	90	99.5	100	97.5	100	99	66	100	62	93	72.5
Task 15	80	80	80	87.5	83	80	80	80	83.5	87.5	89.5	89	81.5	60.5	87.5	22.5	80	66.5

Fig. 10. Comparison of Accuracies for CM1-CM18 Models across Tasks 13–15.



	CM1	CM2	CM3	CM4	CM5	CM6	CM7	CM8	CM9	CM10	CM11	CM12	CM13	CM14	CM15	CM16	CM17	CM18
Task 16	98	99	99	98.5	100	100	99	99.5	98.5	100	99.5	100	100	99.5	100	72.5	99.5	77.5
Task 17	100	100	98.5	100	100	100	100	100	100	100	100	100	100	100	100	80.5	100	70
Task 18	91.5	90.5	91	99.5	100	90	90.5	100	100	98.5	97.5	100	98	83.5	100	77	97	64

Fig. 11. Comparison of Accuracies for CM1-CM18 Models across Tasks 16–18.



	CM1	CM2	CM3	CM4	CM5	CM6	CM7	CM8	CM9	CM10	CM11	CM12	CM13	CM14	CM15	CM16	CM17	CM18
Task 19	86	94	94.5	95	91.5	100	94	100	99.5	100	97.5	100	97	72	90	83	95	66.5
Task 20	90	90	90	98.5	96	93	90	90	99	90	97.5	98.5	92.5	98	99.5	64	91.5	71
Task 21	93.5	93	97	99.5	95.5	100	91.5	100	99	98	97	99	98.5	99	100	71.5	95	56.5

Fig. 12. Comparison of Accuracies for CM1-CM18 Models across Tasks 19–21.

diagnosis accuracy in tasks 13–24, which is respectively 1.29%, 4.88%, 2.92%, 8.13%, 4.67%, 2.42%, 1.50%, 5.88%, 16.92%, 5.05%, 33.30%, 8.34%, and 31.34% higher than that of models using JDA, TCA, BDA, GFK, MEDA, JGSA, TJM, CORAL, EasyTL, JPDA, MEKT, STL and SA. Figs. 10–13 show the comparison of experimental outcomes for models CM1 to CM18 across tasks 13 to 24, and the Fig. 14 presents the comparison of the mean diagnostic accuracies for tasks 13 to 24 obtained by models CM1 to CM18.

According to the above comparative experiments analysis, it can evidently conclude the following conclusions. (1) In the situations of balanced or unbalanced training samples, the proposed DTCNN-SJM framework can achieve desirable performance

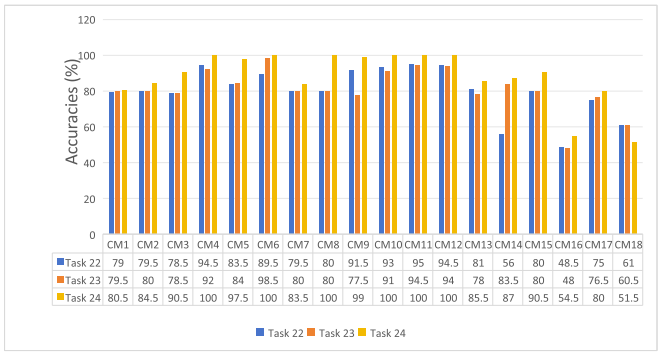


Fig. 13. Comparison of Accuracies for CM1-CM18 Models across Tasks 21–24.

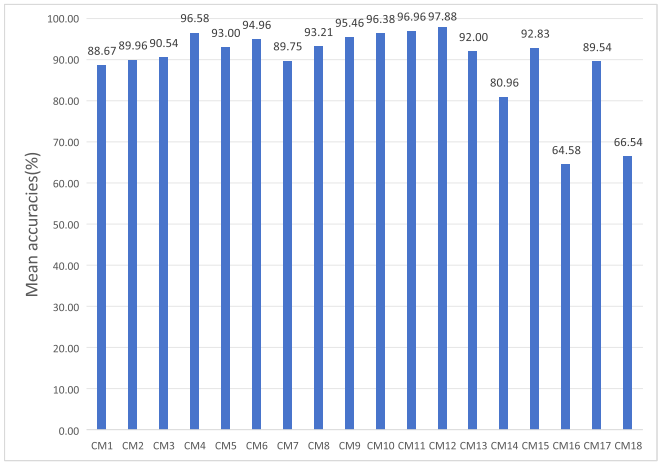


Fig. 14. The contrast of the mean diagnostic accuracies for tasks 13 to 24 obtained by models CM1 to M18.

and outperform models using other domain adaptation methods, accordingly, it indicates that the DTCNN-SJM framework is a promising approach that can be applied in practical industrial scenarios; (2) The fault diagnostic results of all models in balanced training samples (tasks 1–12) are better than that of all models in unbalanced training samples (tasks 13–24), and it evident implies that the insufficient training samples can pose challenges for achieving ideal results of fault diagnosis models and their application in practical scenarios; (3) The designed new domain adaptation method SJM possesses a desirable domain adaptation performance and obviously surpasses other traditional adaptation methods.

B. Case 2: Cross-Domain Fault Diagnostic Based on SQI-MFS Experimental Platform

1) Introduction of Dataset and Fault Diagnosis Tasks: To further demonstrate the validity and adaptability of the DTCNN-SJM framework, accordingly, cross-domain fault diagnosis experiments are conducted using a bearing vibration dataset obtained from the SQI-MFS test-bed. As depicted in Fig. 15, it utilizes the accelerometer sensors to collect the bearing vibration signals under 16 kHz sampling frequency. There are two working speeds of bearing, namely, motor speeds of 1730r/min and

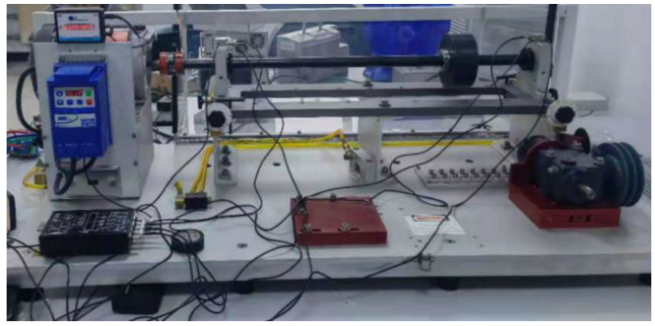


Fig. 15. The SQI-MFS test-bed.

TABLE IX
THE SQI-MFS BEARING DATA FOR CROSS-DOMAIN FAULT DIAGNOSIS EXPERIMENTS

Bearing defect	Defect diameter (mm)	Number of samples (training/validating/testing)		Category label
		1730 rmp	1750 rmp	
No defect	0	120/40/40	120/40/40	1
IRD	0.05	120/40/40	120/40/40	2
	0.1	120/40/40	120/40/40	3
	0.2	120/40/40	120/40/40	4
ORD	0.05	120/40/40	120/40/40	5
	0.1	120/40/40	120/40/40	6
	0.2	120/40/40	120/40/40	7
BD	0.05	120/40/40	120/40/40	8
	0.1	120/40/40	120/40/40	9
	0.2	120/40/40	120/40/40	10

TABLE X
CROSS-DOMAIN FAULT DIAGNOSIS TASKS BASED ON SQI-MFS BEARING DATA

Tasks	SD (Training samples)		TD (Testing samples)		Defect types of samples
	Motor speed	Samples number	Motor speed	Samples number	
25	1730 rmp	1200	1750 rmp	400	1-10
26	1750 rmp	1200	1730 rmp	400	1-10
27	1730 rmp	750	1750 rmp	400	1-10
28	1750 rmp	750	1730 rmp	400	1-10

1750r/min. The collected data includes three types of bearing defects: IRD, RED, and ORD, with defect sizes of 0.05 mm, 0.1 mm, and 0.2 mm. Table IX presents the bearing vibration data used in this experiment case. For each type of fault data, the original data is divided into 200 samples, each consisting of 864 sampling points. The dataset is partitioned into a training set of 120 samples, a validation set of 40 samples, and a testing set of 40 samples, adhering to a 6:2:2 split ratio. Consequently, the study acquires bearing vibration data spanning ten distinct categories, which are identified numerically from 1 to 10. Utilizing this data, we setup four CFDT (tasks 25–28), addressing scenarios with both balanced and unbalanced training samples, detailed in Table X. The training and validation sets of the source domain data are used to train the CNN network. For the fault diagnosis tasks 25 and 26 under balanced data samples, each fault type is represented by an equal number of samples. But, the fault diagnosis under imbalanced data samples has different types of fault samples, and the number of training samples for

TABLE XI
CROSS-DOMAIN FAULT DIAGNOSIS OUTCOMES OF MODELS CM1-CM18 ACROSS TASKS 25–28

Models	Accuracies (%)			
	Task 25	Task 26	Task 27	Task 28
CM1	80.00	81.50	78.50	79.00
CM2	82.50	84.00	81.50	80.00
CM3	80.50	81.00	79.50	77.50
CM4	87.50	89.00	86.00	85.50
CM5	82.50	80.00	81.50	83.00
CM6	84.00	84.50	85.00	86.00
CM7	79.00	78.00	76.50	78.50
CM8	83.50	80.00	82.50	81.00
CM9	82.00	83.50	81.00	80.00
CM10	88.50	86.00	87.50	85.00
CM11	93.50	92.50	87.50	86.00
CM12	94.50	94.00	91.50	89.50
CM13	72.50	77.00	62.00	56.50
CM14	80.00	82.50	75.00	71.50
CM15	82.50	78.50	79.50	72.50
CM16	78.50	73.50	64.00	54.50
CM17	77.50	78.00	66.50	61.50
CM18	72.00	65.50	68.50	65.00

categories 1-10 are 120, 110, 100, 90, 80, 70, 60, 50, 40, and 30, respectively. Therefore, the total number of training samples is 750.

2) Diagnosis Results Analysis of DTCNN-SJM Framework: To verify the efficacy and adaptability of the DTCNN-SJM framework, datasets from the SQI-MFS test rig, operating at two distinct speeds, are employed in cross-domain fault diagnosis experiments, and the comprehensive procedures are similar to that of case 1. Table XI presents the experimental results of DTCNN-SJM framework in tasks 25–28. Tasks 25 and 26 are set up under balanced training samples, and tasks 27 and 28 are set up under unbalanced training samples. For tasks 25–26, cross-domain diagnosis accuracies are respectively 94.50%, 94.00%, 91.50%, and 89.50%, and the mean accuracy can attain 92.38%. According to the results shown in Table XI, the effectiveness of the proposed DTCNN-SJM framework can be also verified, and they also indicate that an imbalance in training samples will result in reduced accuracy of fault diagnosis; however, the DTCNN-SJM framework is able to maintain optimal cross-domain fault diagnostic performance despite these challenges.

3) Comparative Analysis With Different Models: This section is also similar to that of case 1. Table XI and Fig. 16 display the comparative experimental outcomes for the CM1-CM18 models across tasks 25–28, while Fig. 17 illustrates the comparison of the average diagnostic accuracies for tasks 25 to 28 obtained by models CM1 to CM18. The experimental results underscore that the proposed DTCNN-SJM model markedly

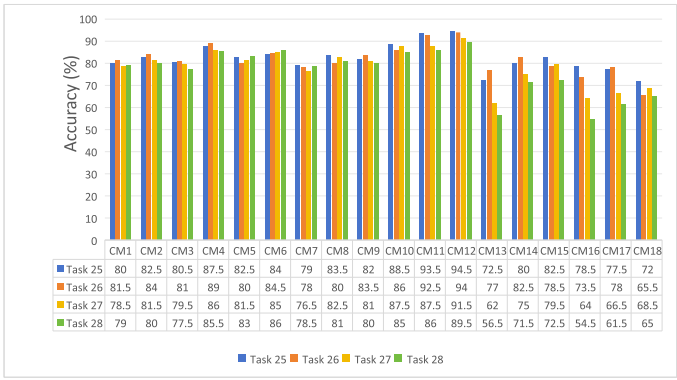


Fig. 16. Comparison of Experimental Results for Models CM1-CM18 across Tasks 25–28.

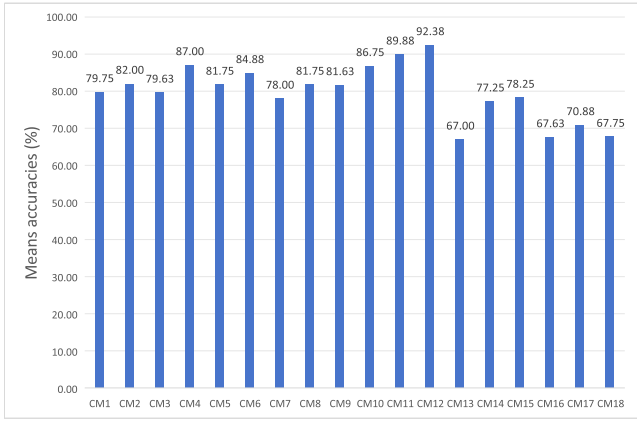


Fig. 17. The comparison of the average diagnostic accuracies for tasks 25 to 28 obtained by models CM1 to CM18.

surpasses other comparative models in cross-domain fault diagnostic performance. Specifically, the average diagnostic accuracies for the CM11 and CM12 models in tasks 25–28 achieve 89.88% and 92.38%, respectively, which are significantly superior to those of the comparative models. From the experimental results of fault diagnosis models using different domain adaptation methods, it can be also shown that the performance of the models using JDA, TJM, and SJM is obviously better than models using TCA, BDA, GFK, JGSA, MEDA, CORAL, EasyTL, JPDA, MEKT, STL and SA. The DTCNN-SJM-KNN model can attain 92.38% of mean diagnosis accuracy in tasks 13–24, which is respectively 5.38%, 10.63%, 7.50%, 14.38%, 10.63%, 10.75%, 5.63%, 25.38%, 15.13%, 14.13%, 24.76%, 21.51%, and 24.63% higher than that of models using JDA, TCA, BDA, GFK, MEDA, JGSA, TJM, CORAL, EasyTL, JPDA, MEKT, STL and SA. The aforementioned experimental analysis indicates that the proposed SJM approach significantly improves domain adaptation and augments the results of cross-domain fault diagnosis. To sum up, the findings from case 2 further verify the efficacy, advantages, and adaptability of the DTCNN-SJM framework across varying working speeds.

V. Conclusion

This study introduces an innovative fault diagnosis framework for bearings that employs DTCNN and SJM, facilitating cross-domain applications. The framework incorporates time-frequency analysis to process the raw vibration signals and extract wavelet time-frequency representation maps. Then, DTCNN construction and deep features extraction are conducted. Next, the designed SJM is performed to shrink the distribution discrepancies between source and target domains and enhance the discriminability by integrating manifold subspace learning, joint feature distribution matching, instances reweighting and maximum marginal criterion. Finally, a pattern recognition classifier is trained for cross-domain fault patterns recognition.

Extensive cross-domain fault diagnosis experiments utilizing two bearing fault datasets are conducted, encompassing scenarios with both balanced and imbalanced training samples. The experimental results indicate as follows: (1) the proposed DTCNN-SJM framework achieves superior cross-domain fault diagnosis performance in comparison to models that are built by common and traditional methods; (2) the diagnostic results of all models in balanced training samples are better than that of all models in unbalanced training samples, and it evident implies that the insufficient training samples can pose challenges for achieving ideal results of fault diagnosis models and their application in practical scenarios; (3) the designed new domain adaptation method SJM possesses a desirable domain adaptation performance and obviously surpasses other traditional domain adaptation methods. In conclusion, the proposed DTCNN-SJM framework shows significant potential for application in diverse industrial environments with fluctuating operational conditions. Future work will focus on developing domain adaptation methodologies with enhanced robustness suitable for complex fault detection scenarios across various devices.

References

[1] T. Han et al., "Fault diagnosis and health management of power machinery," *Machines*, vol. 11, no. 4, 2023, Art. no. 424.

[2] X. Chen, R. Yang, Y. Xue, M. Huang, R. Ferrero, and Z. Wang, "Deep transfer learning for bearing fault diagnosis: A systematic review since 2016," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 3508221.

[3] Z. Wang et al., "Attention-aware temporal–spatial graph neural network with multi-sensor information fusion for fault diagnosis," *Knowl.-Based Syst.*, vol. 278, 2023, Art. no. 110891.

[4] H. Te, X. Wenzhen, and P. Zhongyi, "Semi-supervised adversarial discriminative learning approach for intelligent fault diagnosis of wind turbine," *Inf. Sci.*, vol. 648, 2023, Art. no. 119496.

[5] D. Li et al., "Continual learning classification method and its application to equipment fault diagnosis," *Appl. Intell.*, vol. 52, no. 1, pp. 858–874, 2022.

[6] L. Wang, S. Liu, and H. Xiao, "Vaccine enhanced continual learning with TFE to overcome catastrophic forgetting for variable speed-bearing fault diagnosis," *IEEE Trans. Ind. Inform.*, early access, Jan. 31, 2024, doi: 10.1109/TII.2023.3345451.

[7] J. Xiang and Y. Zhong, "A novel personalized diagnosis methodology using numerical simulation and an intelligent method to detect faults in a shaft," *Appl. Sci.*, vol. 6, no. 12, 2016, Art. no. 414, doi: 10.3390/app6120414.

[8] X. Liu, H. Huang, and J. Xiang, "A personalized diagnosis method to detect faults in a bearing based on acceleration sensors and an FEM simulation driving support vector machine," *Sensors*, vol. 20, no. 2, 2020, Art. no. 420.

[9] H. Wang, J. Zheng, and J. Xiang, "Online bearing fault diagnosis using numerical simulation models and machine learning classifications," *Rel. Eng. Syst. Saf.*, vol. 234, 2023, Art. no. 109142.

[10] Y. Gao, X. Liu, and J. Xiang, "Fault detection in gears using fault samples enlarged by a combination of numerical simulation and a generative adversarial network," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 5, pp. 3798–3805, Oct. 2022, doi: 10.1109/TMECH.2021.3132459.

[11] Y. Lou, A. Kumar, and J. Xiang, "Machinery fault diagnosis based on domain adaptation to bridge the gap between simulation and measured signals," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 3514709, doi: 10.1109/TIM.2022.3180416.

[12] W. Fu et al., "Rolling bearing fault diagnosis based on 2D time-frequency images and data augmentation technique," *Meas. Sci. Technol.*, vol. 34, no. 4, 2023, Art. no. 045005.

[13] Y. Yao et al., "Uncertainty-aware deep learning for reliable health monitoring in safety-critical energy systems," *Energy*, vol. 291, 2024, Art. no. 130419.

[14] Q. Hu, X. Si, A. Qin, Y. Lv, and M. Liu, "Balanced adaptation regularization based transfer learning for unsupervised cross-domain fault diagnosis," *IEEE Sensors J.*, vol. 22, no. 12, pp. 12139–12151, Jun. 2022.

[15] J. Xiong, S. Cui, and H. Tang, "A novel intelligent bearing fault diagnosis method based on signal process and multi-kernel joint distribution adaptation," *Sci. Rep.*, vol. 13, no. 1, 2023, Art. no. 4535.

[16] S. Luo et al., "Transfer learning based on improved stacked autoencoder for bearing fault diagnosis," *Knowl.-Based Syst.*, vol. 256, 2022, Art. no. 109846.

[17] S. Zhang et al., "Selective kernel convolution deep residual network based on channel-spatial attention mechanism and feature fusion for mechanical fault diagnosis," *ISA Trans.*, vol. 133, pp. 369–383, 2023.

[18] H. Zhao et al., "Bearing fault diagnosis using transfer learning and optimized deep belief network," *Meas. Sci. Technol.*, vol. 33, no. 6, 2022, Art. no. 065009.

[19] R. Bai et al., "Rolling bearing fault diagnosis based on multi-channel convolution neural network and multi-scale clipping fusion data augmentation," *Measurement*, vol. 184, 2021, Art. no. 109885.

[20] Y. Miao et al., "Deep network-based maximum correlated kurtosis deconvolution: A novel deep deconvolution for bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 189, 2023, Art. no. 110110.

[21] F. Lu, Q. Tong, Z. Feng, and Q. Wan, "Unbalanced bearing fault diagnosis under various speeds based on spectrum alignment and deep transfer convolution neural network," *IEEE Trans. Ind. Inform.*, vol. 19, no. 7, pp. 8295–8306, Jul. 2023.

[22] M. Huang et al., "A fault diagnosis method of bearings based on deep transfer learning," *Simul. Modelling Pract. Theory*, vol. 122, 2023, Art. no. 102659.

[23] Y. Ding et al., "Deep imbalanced domain adaptation for transfer learning fault diagnosis of bearings under multiple working conditions," *Rel. Eng. Syst. Saf.*, vol. 230, 2023, Art. no. 108890.

[24] L. Xiang et al., "A novel method for rotor fault diagnosis based on deep transfer learning with simulated samples," *Measurement*, vol. 207, 2023, Art. no. 112350.

[25] C. Li et al., "Attention-based deep meta-transfer learning for few-shot fine-grained fault diagnosis," *Knowl.-Based Syst.*, vol. 264, 2023, Art. no. 110345.

[26] S. Han and Z. Feng, "Deep residual joint transfer strategy for cross-condition fault diagnosis of rolling bearings," *J. Dyn., Monit. Diagnostics*, vol. 34, pp. 51–60, 2023.

[27] T. Han et al., "Cross-machine intelligent fault diagnosis of gearbox based on deep learning and parameter transfer," *Struct. Control Health Monit.*, vol. 29, no. 3, 2022, Art. no. e2898.

[28] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfer learning," in *Proc. IEEE Int. Conf. Data Mining*, 2017, pp. 1129–1134.

[29] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2200–2207.

[30] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[31] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2066–2073.

[32] J. Wang et al., "Visual domain adaptation with manifold embedded distribution alignment," in *Proc. 26th ACM Int. Conf. Multimedia*, 2018, pp. 402–410.

[33] J. Zhang, W. Li, and P. Ogunbona, "Joint geometrical and statistical alignment for visual domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5150–5158.
[34] P. Ma et al., "A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains," *ISA Trans.*, vol. 99, pp. 465–478, 2020.
[35] Z. Zhang et al., "A novel geodesic flow kernel based domain adaptation approach for intelligent fault diagnosis under varying working condition," *Neurocomputing*, vol. 376, pp. 54–64, 2020.
[36] K. Zhao et al., "A novel transfer learning fault diagnosis method based on manifold embedded distribution alignment with a little labeled data," *J. Intell. Manuf.*, vol. 33, pp. 151–165, 2022.
[37] Y. Yu, C. Zhang, Y. Li, and Y. Li, "A new transfer learning fault diagnosis method using TSC and JGSA under variable condition," *IEEE Access*, vol. 8, pp. 177287–177295, 2020.
[38] W. Li et al., "A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges," *Mech. Syst. Signal Process.*, vol. 167, 2022, Art. no. 108487.
[39] J. -. H. Lee, J. -. H. Pack, and I.-S Lee, "Fault diagnosis of induction motor using convolutional neural network," *Appl. Sci.*, vol. 9, no. 15, 2019, Art. no. 2950.
[40] Y. Qin and X. Shi, "Fault diagnosis method for rolling bearings based on two-channel CNN under unbalanced datasets," *Appl. Sci.*, vol. 12, no. 17, 2022, Art. no. 8474.
[41] S. Gao, Z. Jiang, and S. Liu, "An approach to intelligent fault diagnosis of cryocooler using time-frequency image and CNN," *Comput. Intell. Neurosci.*, vol. 2022, 2022, Art. no. 253420.
[42] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, "Semisupervised transfer component analysis for domain adaptation in remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3550–3564, Jul. 2015.
[43] S. Yao, Q. Kang, M. Zhou, M. J. Rawa, and A. Albeshri, "Discriminative manifold distribution alignment for domain adaptation," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 53, no. 2, pp. 1183–1197, Feb. 2023.
[44] Y. Cao, M. Long, and J. Wang, "Unsupervised domain adaptation with distribution matching machines," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 1–8.
[45] X. Yu, F. Dong, B. Xia, S. Yang, E. Ding, and W. Yu, "An intelligent fault diagnosis scheme for rotating machinery based on supervised domain adaptation with manifold embedding," *IEEE Internet Things J.*, vol. 10, no. 1, pp. 953–972, Jan. 2023.
[46] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1410–1417.
[47] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, no. 5, pp. 1027–1061, 2007.
[48] Z. Lei et al., "An intelligent fault diagnosis method based on domain adaptation and its application for bearings under polytropic working conditions," *IEEE Trans. Instrum. Meas.*, vol. 70, 2020, Art. no. 3505914.
[49] F. Dong et al., "Rolling bearing fault diagnosis using modified neighborhood preserving embedding and maximal overlap discrete wavelet packet transform with sensitive features selection," *Shock Vib.*, vol. 2018, pp. 1–29, 2018.
[50] W. Ma et al., "An unsupervised domain adaptation approach with enhanced transferability and discriminability for bearing fault diagnosis under few-shot samples," *Expert Syst. With Appl.*, vol. 225, 2023, Art. no. 120084.
[51] L. Li et al., "Research on rolling bearing fault diagnosis based on variational modal decomposition parameter optimization and an improved support vector machine," *Electronics*, vol. 12, no. 6, 2023, Art. no. 1290.
[52] H. Xu et al., "Deep dynamic adaptation network: A deep transfer learning framework for rolling bearing fault diagnosis under variable working conditions," *J. Braz. Soc. Mech. Sci. Eng.*, vol. 45, no. 1, 2023, Art. no. 41.
[53] J. Wang, Y. Chen, H. Yu, M. Huang, and Q. Yang, "Easy transfer learning by exploiting intra-domain structures," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2019, pp. 1210–1215.
[54] W. Zhang and D. Wu, "Discriminative joint probability maximum mean discrepancy (DJP-MMD) for domain adaptation," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2020, pp. 1–8.
[55] W. Zhang and D. Wu, "Manifold embedded knowledge transfer for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1117–1127, May 2020.
[56] J. Wang, Y. Chen, L. Hu, X. Peng, and P. S. Yu, "Stratified transfer learning for cross-domain activity recognition," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2018, pp. 1–10.
[57] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2960–2967.