

Research Article

Bearing Fault Diagnosis Based on Multilayer Domain Adaptation

Bingru Yang,¹ Qi Li ,¹ Liang Chen ,¹ and Changqing Shen ²

¹School of Mechanical and Electric Engineering, Soochow University, Suzhou 215131, China

²School of Rail Transportation, Soochow University, Suzhou 215131, China

Correspondence should be addressed to Liang Chen; chenl@suda.edu.cn

Received 16 July 2020; Revised 13 September 2020; Accepted 15 September 2020; Published 29 September 2020

Academic Editor: Sundararajan Natarajan

Copyright © 2020 Bingru Yang et al. 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.

Bearing fault diagnosis plays a vitally important role in practical industrial scenarios. Deep learning-based fault diagnosis methods are usually performed on the hypothesis that the training set and test set obey the same probability distribution, which is hard to satisfy under the actual working conditions. This paper proposes a novel multilayer domain adaptation (MLDA) method, which can diagnose the compound fault and single fault of multiple sizes simultaneously. A special designed residual network for the fault diagnosis task is pretrained to extract domain-invariant features. The multikernel maximum mean discrepancy (MK-MMD) and pseudo-label learning are adopted in multiple layers to take both marginal distributions and conditional distributions into consideration. A total of 12 transfer tasks in the fault diagnosis problem are conducted to verify the performance of MLDA. Through the comparisons of different signal processing methods, different parameter settings, and different models, it is proved that the proposed MLDA model can effectively extract domain-invariant features and achieve satisfying results.

1. Introduction

Rolling element bearings are of great importance for mechanical equipment. They usually need to run for a long time under harsh conditions, which will inevitably cause faults. The failure of the bearings will cause a lot of economic loss and even safety problems [1]. Therefore, the study of reliable and accurate bearing fault diagnosis methods is of great importance, which can monitor and diagnose the health conditions of the bearings, so as to guarantee the normal working condition of the mechanical equipment and reduce the risk of failure. With the development of intelligent manufacturing, higher requirements have been put forward for fault diagnosis in the industrial process.

Vibration signals are high-precision indicators that can provide information for detecting the status of mechanical equipment [2]. Most traditional fault diagnosis methods based on signal processing are used to extract fault information from raw vibration signals, such as empirical mode decomposition (EMD) [3], wavelet packet transform (WPT) [4], and other time-frequency domain signal processing methods [5]. Yu et al. [6] adopted the EMD method to calculate the original statistical characteristics of the vibration

signals through the intrinsic mode functions and combined with a modified method which can reduce the feature dimension to comply bearing fault diagnosis. Liu [7] decomposed a vibration signal into sub-band signals via one-level stationary wavelet packet transform (one-level SWPT), which improved the ability to extract fault features. Signal processing methods require analysts to have certain expert knowledge to extract fault features accurately. However, the equipment status can be very complicated in the actual operation process; such methods cannot achieve sufficient accuracy. Therefore, the researchers introduced machine learning (ML) methods to make up for this deficiency. Chen et al. [8] combined rough set theory (RS) with the support vector machine (SVM) to propose a multisensor data fusion fault diagnosis method, which reduced the computing cost of the SVM but improved the effectiveness and accuracy. Yu et al. [9] applied WPT to extract fault features of the planetary gearbox; the features are discretized and regarded as the input of the flexible naive Bayesian classifier (FNBC). Fenineche et al. [10] studied the influence of the parameter selection in the artificial neural network (ANN) to obtain the best performance of fault diagnosis. However, the performance of ML models is often limited by manual feature extraction. When

the fault signals are complex, it is difficult to achieve the expected diagnostic accuracy.

With the development of deep learning (DL) [11], it can automatically extract features from nonlinear bearing signals, and the tedious signal preprocessing can be greatly reduced. Zhang et al. [12] applied the sparse autoencoder (SAE) to propose a new label generation method, which can identify samples that do not belong to known categories. Dong et al. [13] introduced the convolutional neural network (CNN) into a deep belief network (DBN) to propose a random convolutional deep belief network for the mechanical fault. By adding unsupervised components, the generalization ability of the model was improved. The novel hierarchical learning rate adaptive CNN presented by Guo et al. [14] was designed for diagnosing bearing faults and determining severity. In order to overcome the shortcomings of traditional vibration signal processing methods, Jiao et al. [15] proposed a method based on multivariate encoder information to diagnose the fault intelligently. The method presented in [16] can effectively diagnose the compound fault by combining the features automatically extracted by the model with the time-domain features designed manually. However, DL models have to satisfy the assumption that the source domain and target domain (for example, training set and test set) should obey the same distribution and feature space. Actually, in many actual industrial scenarios, the difference of distribution between source-domain samples and target-domain samples varies considerably, which degrades the diagnostic performance [17].

To tackle this challenge, the application of transfer learning (TL) to bearing fault diagnosis is an emerging research aspect in recent years. Its purpose is to fully reuse the knowledge learned from the source domain to another different but related target domain [18, 19]. Peng et al. [20] added the idea of residual learning to the model, which can effectively learn high-level and abstract features. Wen et al. [21] adopted a three-layer SAE as the feature extractor and calculated the maximum mean discrepancy (MMD) to minimize the difference between domains. The representation clustering algorithm proposed by Li et al. [22] can maximize the distance metric of interclass variations and minimize the distance metric of intraclass variations at the same time. Zhang et al. [23] improved the domain adaptation ability of the model through implementing the adaptive batch normalization method. The method presented in [24] matched the marginal distributions of the output of every convolution layer, improving the cross-domain testing performance. For the distribution discrepancy between the source domain and target domain, some studies consider that the transferability of high-level features drops significantly [25], while others believe that low-level features may be more responsible for domain shift [26]. Moreover, most existing TL methods focus on marginal distributions and ignore conditional distributions of different domains, while they both have different effects on domain adaptation.

In this paper, a novel multilayer domain adaptation (MLDA) method is proposed for TL-based intelligent bearing fault diagnosis. By calculating multikernel MMD (MK-MMD) and considering conditional distributions in multiple layers, the model can extract effective domain-invariant features, which

can clearly contribute to transfer tasks. The main contributions of this work can be summarized as the following: (1) a MLDA method is raised to diagnose the unlabeled bearing fault signals from the target domain through the shared domain features extracted from the source domain. (2) The method can diagnose the compound fault and single fault of multiple sizes at the same time. (3) A special designed residual network based on the ResNet [27] framework is adopted as the feature extractor in the fault diagnosis task to extract features automatically without complex time-frequency domain analysis. (4) In order to minimize the distribution discrepancy between the source and the target domain, MK-MMD and pseudo-label learning are adopted in multiple layers, considering both marginal distributions and conditional distributions.

The remaining parts of the paper are organized as follows. In Section 2, the domain adaptation problem is demonstrated, and MK-MMD is introduced. The proposed MLDA method for bearing fault diagnosis is raised in Section 3. The comparisons of different signal processing methods, different parameter settings, and different methods are discussed in Section 4. Finally, the conclusions are drawn in Section 5.

2. Theoretical Background

2.1. Problem Formulation. Since bearings are affected by many factors during operation, such as load and running time, the distributions of samples in the source domain are different from those of samples in the target domain. The emergence of TL provides a new idea for solving this problem. There are two important concepts in TL named domain and task. The detailed description of TL is given as follows [28–30].

The domain, abbreviated by D , contains the data space X and its marginal distribution $P(X)$, which can be described as $D = \{X, P(X)\}$.

The task, abbreviated by T , contains the label space Y and its predictive function $f(\cdot)$, which can be described as $T = \{Y, f(\cdot)\}$. $f(\cdot)$ can also be described as a conditional distribution $P(Y|X)$ from the perspective of probability.

The labeled sample space of D_s can be written as $X_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ with a relevant task T_s , and the unlabeled sample space of D_t can be written as $X_t = \{(x_j^t)\}_{j=1}^{n_t}$ with a relevant task T_t , where n_s and n_t , respectively, denote the numbers of samples of their specific domain.

TL aims to make full use of the knowledge learned from source domain D_s and source task T_s to find a target predictive function $f(\cdot)$ in target domain D_t , where $D_s \neq D_t$ or $T_s \neq T_t$. The condition $D_s \neq D_t$ indicates that $P_s(X) \neq P_t(X)$ or (and) $X_s \neq X_t$, and the condition $T_s \neq T_t$ indicates that $P_s(Y|X) \neq P_t(Y|X)$ or (and) $Y_s \neq Y_t$.

Domain adaptation (DA) can be regarded as a specific setting in TL, as shown in Figure 1, which solves the problem of $X_s \neq X_t$, but $T_s = T_t$.

2.2. Multikernel Maximum Mean Discrepancy. DA is a challenge problem when there are no (or limited) labeled data in the target domain. To address this problem, many existing methods focus on minimizing the difference

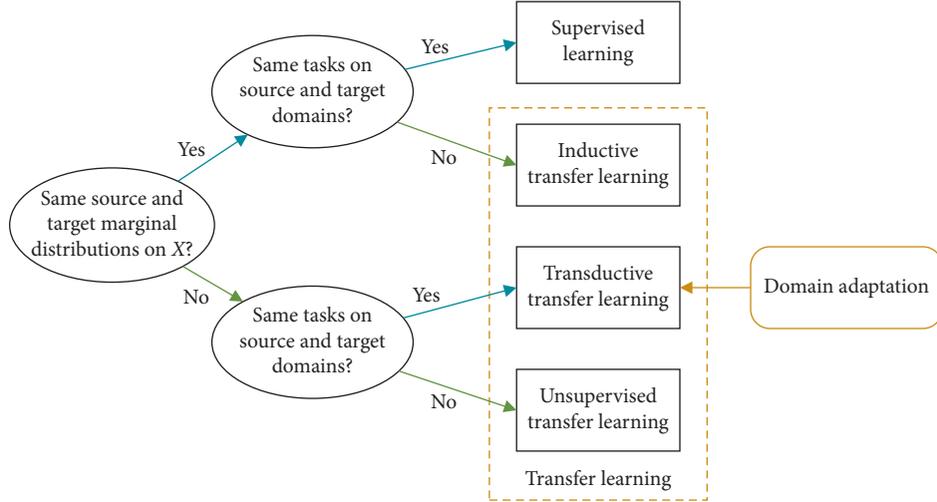


FIGURE 1: An illustration of the DA branches in a fault diagnosis tree.

TABLE 1: Detailed information of the ResNet-18 structure.

Layer name	Output size	Channels \times kernel size
Input	$3 \times 224 \times 224$	—
Conv1	$64 \times 112 \times 112$	$64 \times 7 \times 7$, stride = 2
BN, ReLu	$64 \times 112 \times 112$	—
Max pool	$64 \times 56 \times 56$	$64 \times 3 \times 3$, stride = 2
Block 1: Conv2_x	$256 \times 56 \times 56$	$\begin{bmatrix} 64 \times 3 \times 3 \\ 64 \times 3 \times 3 \end{bmatrix} \times 2$
Block 2: Conv3_x	$512 \times 28 \times 28$	$\begin{bmatrix} 128 \times 3 \times 3 \\ 128 \times 3 \times 3 \end{bmatrix} \times 2$
Block 3: Conv4_x	$1024 \times 14 \times 14$	$\begin{bmatrix} 256 \times 3 \times 3 \\ 256 \times 3 \times 3 \end{bmatrix} \times 2$
Block 4: Conv5_x	$2048 \times 7 \times 7$	$\begin{bmatrix} 512 \times 3 \times 3 \\ 512 \times 3 \times 3 \end{bmatrix} \times 2$
ReLu	$2048 \times 7 \times 7$	—
Average pool	2048×1	$2048 \times 7 \times 7$
FC, softmax	1000	—

between two domains by adopting a nonparametric distance measure called MMD, which can measure the discrepancy of marginal distributions. As stated in [31], compared with a single kernel, MK-MMD can greatly improve the efficiency of domain adaptation.

H_k denotes the reproducing kernel Hilbert space (RKHS) with a characteristic kernel k . The MK-MMD between distributions U and V is defined as the RKHS distance between the mean embeddings of U and V . The squared formulation of the MK-MMD is given as

$$\text{MMD}_k^2(U, V) = \|E_U[\phi(x^s)] - E_V[\phi(x^t)]\|_{H_k}^2. \quad (1)$$

The most important property is that $\text{MMD}_k^2(U, V) = 0$ only when $U = V$. The calculation formula of the multikernel is given by

$$k(x^s, x^t) = \sum_{i=1}^G k_{\theta_i}(x^s, x^t), \quad (2)$$

where G is the number of kernels and k_{θ_i} is the Gaussian kernel with bandwidth θ_i . Gretton et al. [32] theoretically studied that the kernel used in the mean embedding of U and V is essential to reduce the test error. The multikernel can enhance MMD test through different kernels, thus providing a method for optimal kernel selection.

3. The Proposed Method

3.1. A Special Designed Residual Network. The ResNet has been proved to have strong feature extraction capability. Considering the size of the bearing fault dataset, ResNet-18 is selected as the feature extractor of MLDA. The detailed information of ResNet-18 is shown in Table 1. Convolutional layer, batch normalization, rectified linear unit, and fully connected layer are abbreviated as Conv, BN, ReLu, and FC, respectively.

ResNet-18 contains 4 blocks, and the internal structure of the block is illustrated in Figure 2.

The block can be represented as

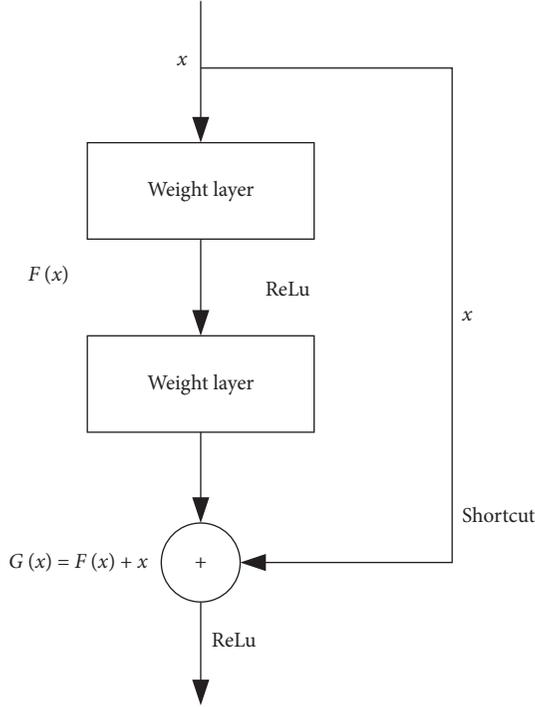


FIGURE 2: Internal structure of the block.

$$y_l = g(x_l) + F(x_l, w_l), \quad (3)$$

$$x_{l+1} = h(y_l), \quad (4)$$

where $g(x_l)$ is the input of the l th block and F is the residual function. Equation (4) represents the identity mapping, and h is the ReLU activation function. Based on equations (3) and (4), the deep features from low-level l to high-level L can be obtained:

$$x_L = x_l + \sum_{i=1}^{L-1} F(x_i, W_i). \quad (5)$$

The original ResNet-18 has achieved great success in the field of image recognition. In this paper, ResNet-18 is adopted as a feature extractor. Combined with the characteristics of the bearing signals, some modifications need to be made: (1) in order to match the input dimension of the bearing signals, the kernel size of Conv1 is changed to 3×3 . (2) To retain as much fault information as possible, the Max pool layer is removed. (3) Since ResNet-18 is only adopted as a feature extractor and its classification function is not required, the FC layer and softmax layer are removed. The modified ResNet-18 can effectively extract domain-invariant features.

3.2. Network Architecture. In order to diagnose bearing faults under variable working conditions, the architecture of MLDA is shown in Figure 3.

The data from D_s and D_t are used as the input of ResNet-18 pretrained by labeled data from the source domain. For extracting domain-invariant features effectively, the marginal

distributions are minimized through calculating MK-MMD in multiple layers. The MK-MMD loss can be obtained by

$$L_{\text{MMD}} = \sum_{l \in N^l} \sum_{k \in K} \text{MMD}_k^2(U^l, V^l), \quad (6)$$

where N^l denotes the number of blocks for calculating MK-MMD, K is the number of Gaussian kernels, U^l and V^l represent the distribution of D_s and D_t extracted from the l th block, and $\text{MMD}_k^2(U^l, V^l)$ is the MK-MMD calculated by equation (1) with kernel k .

In two domains with different working conditions, the classification categories are the same. Since the label of D_s is available, the classification loss can be minimized, and cross-entropy is used as the optimization objective:

$$L_{\text{clf}} = -\frac{1}{M} \sum_{m=1}^M \tilde{y} \ln y, \quad (7)$$

where M denotes the number of samples, y represents the true label, and \tilde{y} represents the label output by the classifier.

MK-MMD can bound the marginal distributions of the extracted features from D_s and D_t . However, unlabeled data from the target domain cannot be directly used in the training process because supervised information is not available. Pseudo-label learning [33] can be one of the solutions to this problem. The pseudo-label of the specific sample is determined by selecting the label with the maximum probability of prediction, which can be summarized into two steps: the predicted probability of labels and the conversion to the pseudo-label [34]. In MLDA, each block is followed by a matching classifier (FC layer). The predicted probability of labels given by the classifier and the softmax layer can be calculated as

$$\hat{y}_i = [P(y_i = 1 | x), \dots, P(y_i = C | x)]^T, \quad (8)$$

$$P(y_i = c | x) = \frac{e^{W_c \cdot x}}{\sum_{c=1}^C e^{W_c \cdot x}}, \quad (9)$$

where y_i is the i th sample, C is the number of categories, and W is the weight of the corresponding category. The conversion of pseudo-labels can be expressed as

$$\hat{y}_i = [\hat{y}_j, \dots, \hat{y}_j], \quad (10)$$

$$\hat{y}_j = \begin{cases} 1, & \text{if } j = \underset{j}{\operatorname{argmax}} \hat{y}_j, \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

where \hat{y}_i denotes the pseudo-label of the i th sample. The correctness of pseudo-labels will be improved during the training process so that the conditional distributions could be more similar. The pseudo-label loss of each block can be calculated by cross-entropy:

$$L_{\text{pseudo}} = -\frac{1}{N} \sum_{n=1}^N \hat{y} \ln \hat{y}. \quad (12)$$

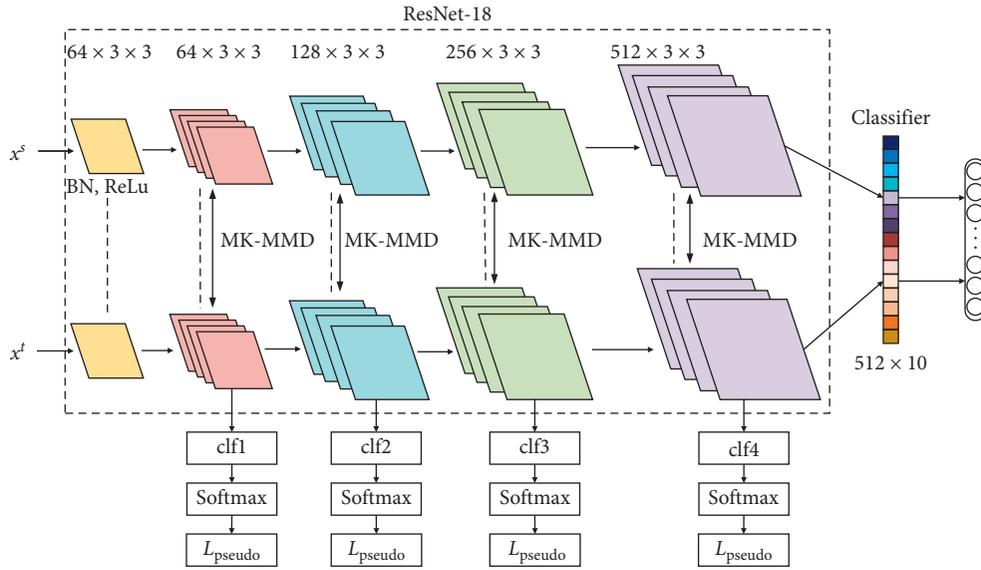


FIGURE 3: Architecture of the proposed MLDA model.

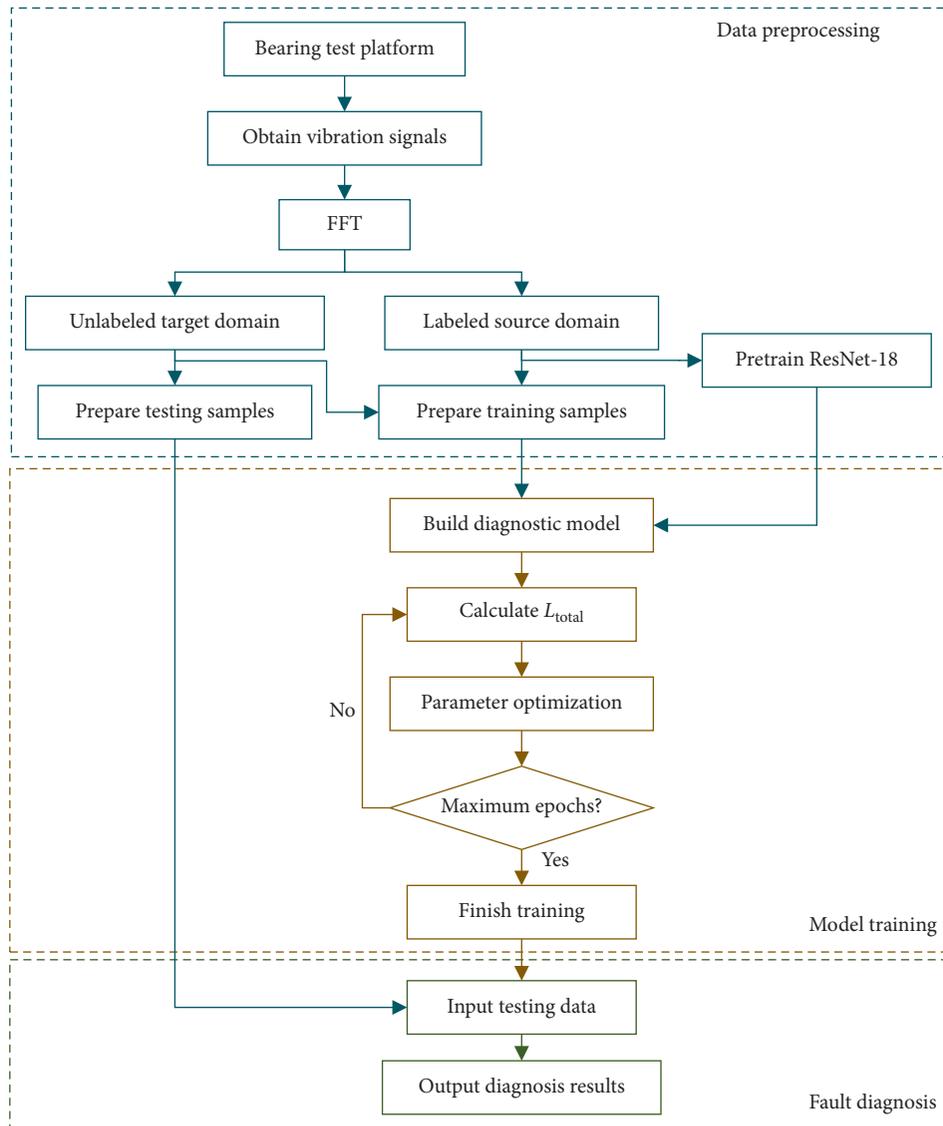


FIGURE 4: Flowchart of the proposed MLDA model.

The total pseudo-label loss can be obtained by

$$L_p = \sum_{l \in N^l} L_{\text{pseudo}}(\hat{y}^l, \tilde{y}^l). \quad (13)$$

The loss of the overall model can be expressed as

$$L_{\text{total}} = L_{\text{clf}} + \lambda_1 L_{\text{MMD}} + \lambda_2 L_p, \quad (14)$$

where λ_1 and λ_2 are tradeoff parameters.

3.3. Diagnosis Procedure. The flowchart of MLDA is developed in Figure 4.

First, the original vibration signals under different loads are collected from the bearing test platform. The frequency-domain signals are constructed via fast Fourier transform (FFT) and reshaped into 2-dimension [35, 36]. Then, the data are divided into supervised source domain and unsupervised target domain and further separated into training set and test set. Furthermore, in order to accelerate the training process, ResNet-18 is pretrained by source-domain data.

Secondly, based on the particular problem of fault diagnosis and the input dataset information, the diagnostic model built is ready for the training process. The data from the training set are fed into the pretrained ResNet-18 which can extract domain-invariant features. Both marginal distributions and conditional distributions are minimized in multiple layers. On the final layer of the network, the FC layer is adopted to identify the bearing faults with the extracted domain-shared features. The optimization objective of the model (equation (14)) is minimized through the Adam method. When the training process is over, the loss function of the overall method converges in general.

Finally, after training, the test set from the target domain is input into the model to evaluate the model capability and output the fault diagnosis results.

4. Experimental Analysis

4.1. Dataset Description. The bearing fault dataset used to evaluate the effectiveness of the raised MLDA method was collected from the bearing test platform as shown in Figure 5. The drive motor, healthy bearing, and test bearing are fixed on the same motor shaft from left to right. The data were collected by an NI PXIe-1082 data acquisition system. The adjustable loading system is settled in the radial direction of the motor shaft. An SGSF-20K dynamometer is installed in the bolt-nut system to measure the load. The sampling frequency of the PCB 352C33 accelerometer is 10 kHz, and the motor speed is 896 rpm. During the bearing operation cycle, the accelerometer continuously collects bearing signal data.

The 14 health conditions are gathered under four working conditions with different loads (0 kN, 1 kN, 2 kN, and 3 kN). There are ten health conditions for single faults, namely, normal bearing (NO), inner race fault (IF), outer race fault (OF), and ball fault (BF). Each fault condition covers three fault diameters. Furthermore, four kinds of compound faults are processed in a width of 0.2 mm: inner race and ball fault (IOB), inner race and outer race (IO), outer race and ball fault

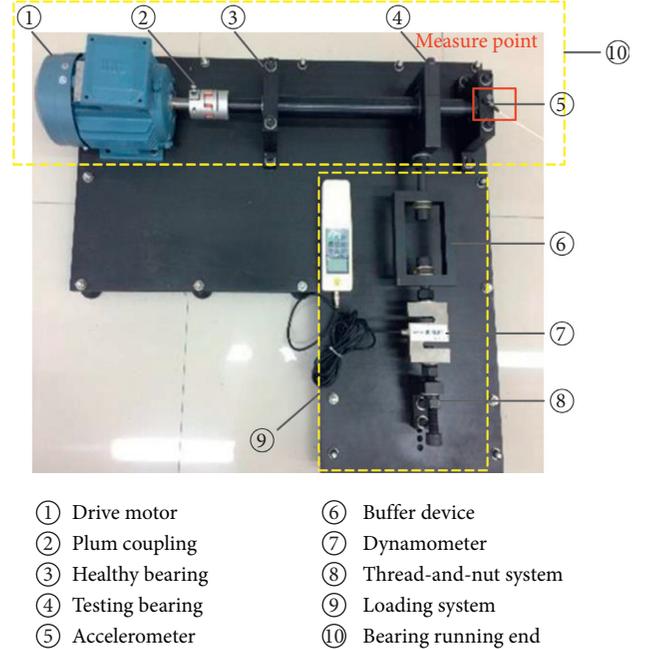


FIGURE 5: Bearing test platform.

TABLE 2: Description of 14 health conditions with both single and compound faults.

Bearing fault pattern	Diameters (mm)	Training samples	Testing samples	Label	Abbreviation
NO	—	320	160	0	NO
IF	0.2	320	160	1	IF02
	0.3	320	160	2	IF03
	0.4	320	160	3	IF04
OF	0.2	320	160	4	OF02
	0.3	320	160	5	OF03
	0.4	320	160	6	OF04
BF	0.2	320	160	7	BF02
	0.3	320	160	8	BF03
	0.4	320	160	9	BF04
IB	0.2	320	160	10	IB
IO	0.2	320	160	11	IO
OB	0.2	320	160	12	OB
IOB	0.2	320	160	13	IOB

(OB), and inner race, outer race, and ball fault (IOB). For the sake of clarity, all 14 fault patterns are summarized in Table 2. Each sample contains 2048 data points.

When the bearing rotates at a certain constant speed, different fault patterns will generate different vibration signals. The vibration signals under 0 kN load are shown in Figure 6. The vibration signal of the healthy bearing is relatively stable. For single faults, the periodicity of IF and OF can be seen obviously. However, the vibration signal of the BF has no obvious periodicity and amplitude, which is difficult to identify. Compared with the single fault, the amplitude of the compound fault increases significantly and changes greatly. The complexity of compound fault signals makes them difficult to extract features and brings challenges to academic research [37].

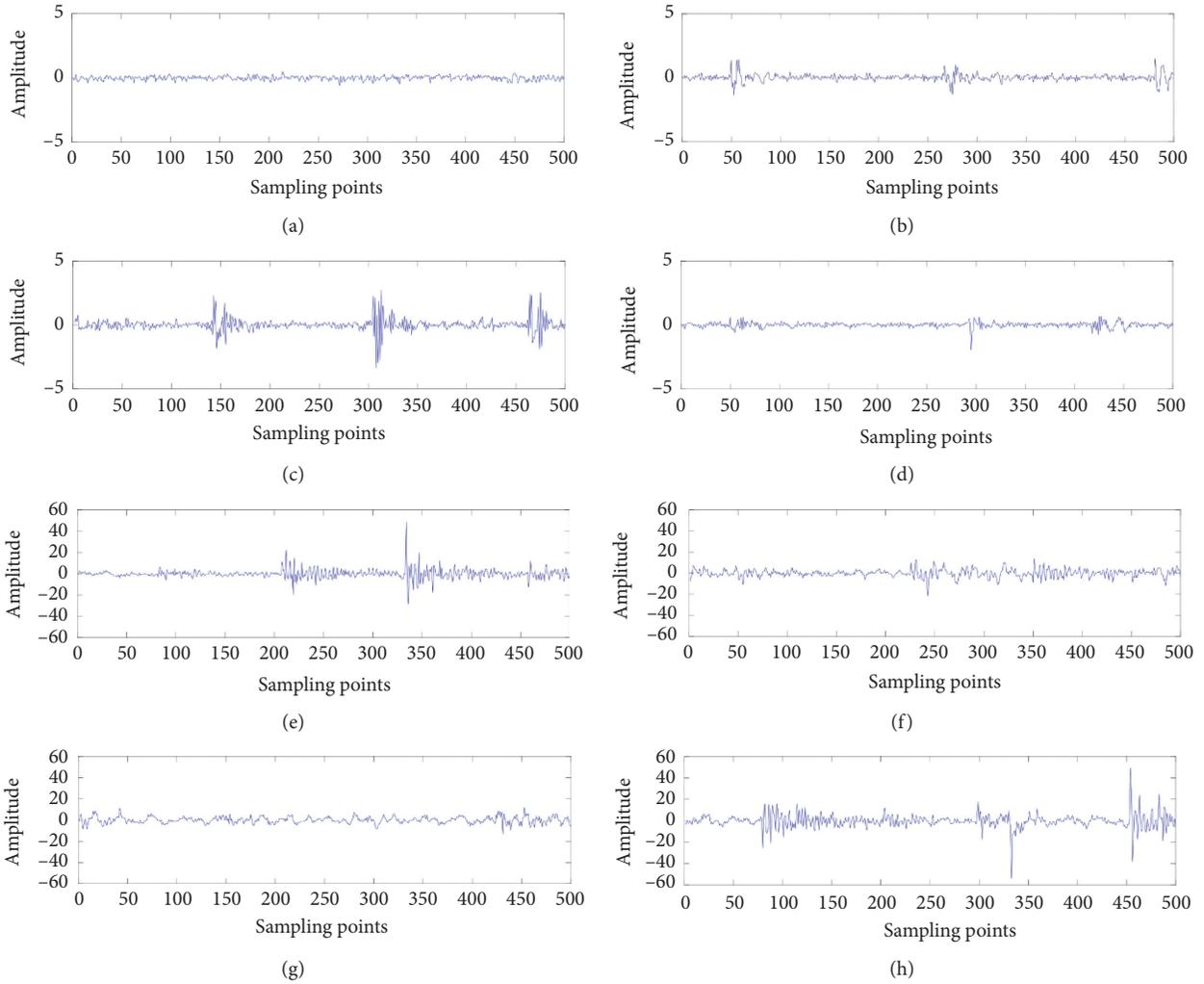


FIGURE 6: Bearing vibration signals. (a) NO. (b) IF03. (c) OF03. (d) BF03. (e) IB. (f) IO. (g) OB. (h) IOB.

TABLE 3: Experimental results of different signal processing methods.

Transfer task	Time domain + MLDA (%)	EMD + MLDA (%)	FFT + MLDA (%)
0-1	93.02	96.26	99.52
0-2	65.95	78.24	99.24
0-3	61.10	72.42	97.52
1-0	98.38	97.96	99.81
1-2	83.89	87.29	99.60
1-3	77.00	82.72	99.40
2-0	66.15	74.78	99.65
2-1	80.93	81.72	99.63
2-3	95.70	97.02	98.89
3-0	68.47	85.99	98.65
3-1	80.61	84.28	98.68
3-2	96.74	98.21	99.07

4.2. Comparison of Different Signal Processing Methods. In order to extract features from the fault-related bearing signals, traditional methods employ many tricks to process

the signal. In this part, we carried out time, frequency, and time-frequency analysis to decide the best signal pre-processing method for the MLDA model. Three experiments are carried out: (1) input time-domain signals to the MLDA model, (2) input time-frequency domain signals by empirical mode decomposition (EMD), and (3) input frequency-domain signals by FFT.

In the experiment, the learning rate is set as 0.0001, and λ_1 and λ_2 are set as 1 and 0.01, respectively. The MK-MMD adopts a mixture of 5 Gaussian kernels with bandwidths of 4, 8, 16, 32, and 64. A total of 12 transfer tasks are conducted. For example, transfer task 0-1 indicates that 0 kN is the source domain and 1 kN is the target domain, hereinafter the same.

The results are shown in Table 3. It is found FFT + MLDA can achieve the best results, which demonstrates the power of the deep neural network to model the fault-related nonlinear vibration signals. Hence, our method does not need complicated techniques on signal pre-processing. Simply transforming the signal from the time domain to the frequency domain by FFT is sufficient for the MLDA model.

TABLE 4: Experimental results of different parameter settings.

Transfer task	MLDA-1 (%)	MLDA-4 (%)	MLDA-SK (%)	MLDA (%)
0-1	90.79	95.21	92.76	99.52
0-2	84.07	80.21	84.00	99.24
0-3	74.32	86.99	85.63	97.52
1-0	94.07	96.96	94.50	99.81
1-2	91.64	95.57	94.91	99.60
1-3	87.93	92.53	91.34	99.40
2-0	80.98	89.87	84.23	99.65
2-1	88.31	98.07	94.06	99.63
2-3	94.85	98.32	86.64	98.89
3-0	64.19	76.63	75.33	98.65
3-1	82.49	83.03	89.58	98.68
3-2	79.70	94.85	88.03	99.07

4.3. *Comparison of Different Parameter Settings.* Different parameter settings will bring different effects on the experimental results. For verifying the effect of multilayer domain adaptation, two groups of experiments are designed to apply MK-MMD and pseudo-label learning only on block 1 and block 4, respectively, represented as MLDA-1 and MLDA-4. Furthermore, in order to prove the effect of MK-MMD, only one Gaussian kernel with a bandwidth of 4 is set as the third group of comparisons, represented as MLDA-SK. The experimental results are shown in Table 4.

As shown in Table 4, MLDA achieves the best accuracy in almost all tasks, with an average of 99.14%. It can be seen from MLDA-1 and MLDA-4 that both low-level features and high-level features will cause domain shift to a certain extent. Matching discrepancy of high-level features can obtain better accuracy, which indicates that the transferability of high-level features is better than that of low-level features. Moreover, when the source domain is relatively different from the target domain, such as the transfer between 0 kN and 3 kN, the advantages of applying domain adaptation in multiple layers are obvious. The comparison of MLDA-SK shows that a single kernel of MMD has a limited capability to narrow the marginal distributions. MLDA makes a great improvement in all 12 tasks, which proves the effect of mixed kernels. The radar diagram in Figure 7 shows the diagnostic results of different parameter settings. It can be seen intuitively that MLDA achieves the best results.

Figure 8 shows the classification results of the proposed method for the 0-3 transfer task. The fault characteristics of NO, IF, and OF are relatively obvious. All samples are diagnosed correctly. Misdiagnosis mostly occurs in BF04, IB, and IOB. The diagnostic results of BF04 are mainly the fault diameter recognition error, which is classified as BF03. IB is misclassified as BF04 and IOB. IOB is the category with the most misdiagnoses, and all the misclassified samples are identified as BF04. BF is easily misclassified because its fault characteristics are not obvious. Since the compound fault is a mixture of multiple fault types, the feature extraction is difficult with mixture features, especially the mixture of three fault types.

4.4. *Comparison of Different Models.* In order to further demonstrate the effect of MLDA, traditional TL methods are investigated, which include transfer component analysis

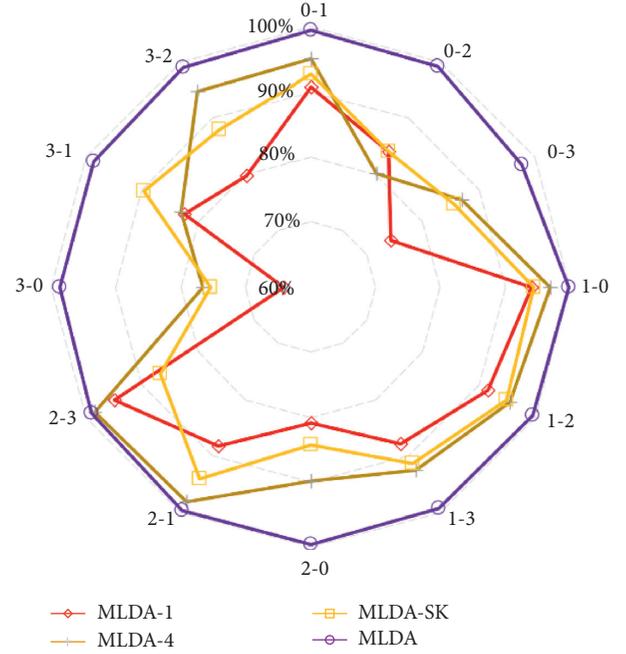


FIGURE 7: Radar diagram of the diagnostic results of different parameter settings.

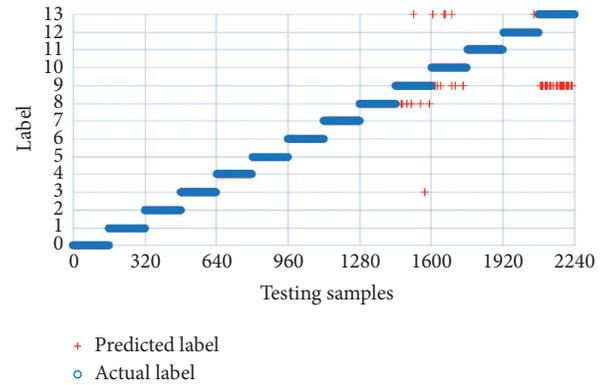


FIGURE 8: Classification results of MLDA.

TABLE 5: Experimental results of different models.

Transfer task	TCA (%)	JDA (%)	CORAL (%)	ResNet-18 (%)	MLDA (%)
0-1	98.64	99.24	97.43	91.07	99.52
0-2	83.17	92.05	83.97	77.28	99.24
0-3	75.45	84.46	80.45	72.32	97.52
1-0	99.76	99.76	99.78	96.16	99.81
1-2	91.38	98.06	91.74	86.56	99.60
1-3	86.76	90.58	89.20	86.47	99.40
2-0	90.49	92.77	90.47	77.46	99.65
2-1	98.42	99.46	92.14	89.06	99.63
2-3	95.45	97.25	94.44	87.68	98.89
3-0	84.40	85.85	85.02	71.21	98.65
3-1	85.31	85.63	85.67	84.91	98.68
3-2	94.24	95.02	91.96	86.92	99.07

(TCA) [38], joint distribution adaptation (JDA) [39], correlation alignment (CORAL) [40], and the pretraining model (ResNet-18). The comparison results are illustrated in Table 5.

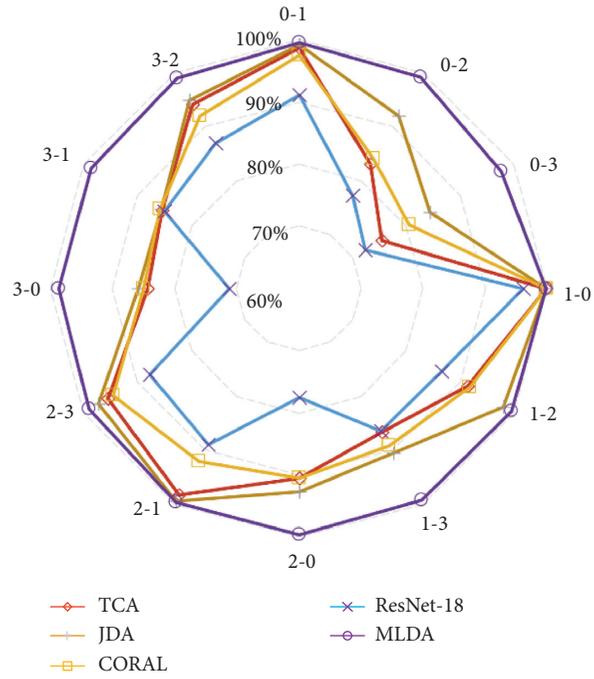


FIGURE 9: Radar diagram of the diagnostic results of different models.

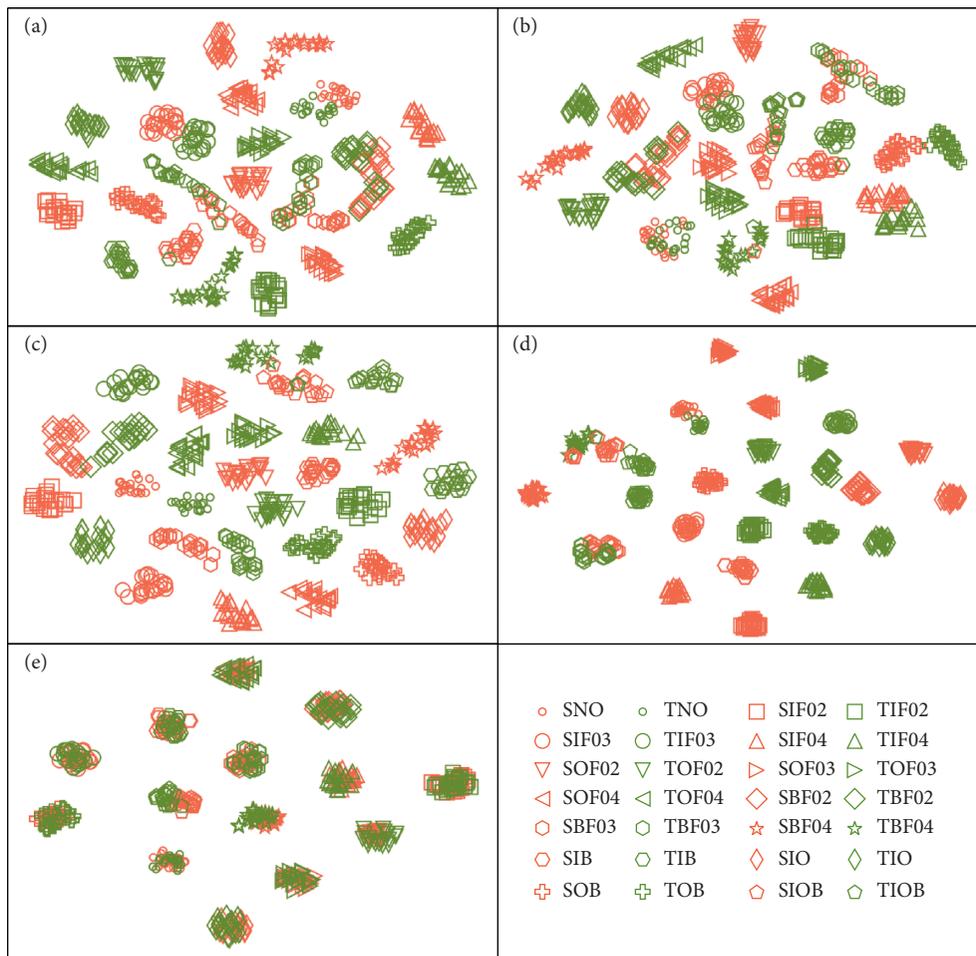


FIGURE 10: Visualization of transferable features. (a) TCA. (b) JDA. (c) CORAL. (d) ResNet-18. (e) MLDA.

From the comparison results of different methods, the conclusions can be drawn with three points: (1) the best performance is achieved by MLDA distinctly among 5 methods. Without domain adaptation, the target-domain data cannot be diagnosed effectively by the pretraining model (ResNet-18). (2) The traditional TL methods can achieve good results when the discrepancy is relatively small, such as the transfer between 0 kN and 1 kN. The conclusion can be explained by the fact that the transferability of extracted features and the degree of domain shift are affected by the degree of working conditions. (3) The unsatisfied transfer effect will occur by the traditional methods when the variety of working conditions is dramatic, which decreases the accuracy of fault diagnosis. Notably, the raised MLDA method maintains the satisfied accuracy and the generalizability.

Figure 9 illustrates diagnostic results using different models. Clearly, the best performance is achieved in various transfer tasks by the raised MLDA method, which proves its superiority.

Although MLDA achieves satisfactory results, it is still confusing whether MLDA can extract domain-invariant features. T-SNE method [41] which can reduce the dimension is introduced to visualize the features extracted by each method. The results are shown in Figure 10, which shows transfer task 3-0. 20 feature points are sampled randomly in each category.

As a pretraining model, ResNet-18 has strong feature extraction ability. However, the discrepancy between two domains can hardly be narrowed in the ResNet-18 method. The other three traditional TL methods can narrow the distribution shift between different domain features to a certain extent, but the capability is limited. The proposed MLDA method learns the feature mapping from source and target domains to the shared feature space, decreasing domain shift and effectively using the knowledge learned from the shared feature extractor to diagnose the target domain through unsupervised learning. It can clearly extract domain-invariant features with high generalizability.

5. Conclusions

In summary, this paper develops a MLDA method for bearing fault diagnosis, which can diagnose compound faults and single faults of multiple sizes simultaneously. First, modified ResNet-18 is pretrained as a feature extractor. The MK-MMD is calculated for the extracted features in multiple layers to narrow the marginal distributions. Second, the features extracted from each block are input into the matching classifier. The predicted probability is calculated through the softmax layer and converted into the pseudo-label to narrow the conditional distributions. Third, the Adam optimization method is adopted to optimize the overall model parameters and speed up the convergence of the model. Through the comparisons of different signal processing methods, different parameter settings, and different methods, the raised MLDA method classifies the fault patterns precisely and achieves better transfer performance. The proposed method is meaningful to prognostics health

management (PHM) and can provide reliable fault diagnosis results for practical industrial scenarios.

Data Availability

The data can be obtained from the Institute of Industrial Measurement, Control and Equipment Diagnostics, School of Rail Transportation, Soochow University.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (no. 51875375) and the Suzhou Science Foundation (no. SYG201802).

References

- [1] D.-T. Hoang and H.-J. Kang, "A survey on Deep Learning based bearing fault diagnosis," *Neurocomputing*, vol. 335, pp. 327–335, 2019.
- [2] D.-T. Hoang and H.-J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," *Cognitive Systems Research*, vol. 53, pp. 42–50, 2019.
- [3] S. Park, S. Kim, and J.-H. Choi, "Gear fault diagnosis using transmission error and ensemble empirical mode decomposition," *Mechanical Systems and Signal Processing*, vol. 108, pp. 262–275, 2018.
- [4] C. Wang, M. Gan, and C. a. Zhu, "A supervised sparsity-based wavelet feature for bearing fault diagnosis," *Journal of Intelligent Manufacturing*, vol. 30, no. 1, pp. 229–239, 2019.
- [5] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided VMD for fault diagnosis of rotating machines," *Journal of Sound and Vibration*, vol. 435, pp. 36–55, 2018.
- [6] X. Yu, F. Dong, E. Ding, S. Wu, and C. Fan, "Rolling bearing fault diagnosis using modified LFDA and EMD with sensitive feature selection," *IEEE Access*, vol. 6, pp. 3715–3730, 2018.
- [7] Y. Liu, "One-level stationary wavelet packet transform & hilbert transform based rolling bearing fault diagnosis," in *Proceedings of the 2018 IEEE International Conference on Information and Automation (ICIA)*, pp. 1475–1479, Fujian, China, August 2018.
- [8] G. Chen, Y. Wu, and L. Fu, "Fault diagnosis of full-hydraulic drilling rig based on RS – SVM data fusion method," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 40, no. 3, pp. 1–11, 2018.
- [9] J. Yu, M. Bai, G. Wang, and X. Shi, "Fault diagnosis of planetary gearbox using wavelet packet transform and flexible naive bayesian classifier," in *Proceedings of the 2017 36th Chinese Control Conference (CCC)*, IEEE, Dalian, China, pp. 7207–7211, July 2017.
- [10] H. Fenineche, A. Felkaoui, and A. Rezig, "A effect of input data on the neural networks performance applied in bearing fault diagnosis," in *Proceedings of the Signal Processing Applied to Rotating Machinery Diagnostics, (SIGPROMD'2017)*, pp. 34–43, Springer, Setif, Algeria, April 2017.
- [11] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

- [12] S. Zhang, M. Wang, W. Li, J. Luo, and Z. Lin, "Deep learning with emerging new labels for fault diagnosis," *IEEE Access*, vol. 7, pp. 6279–6287, 2019.
- [13] S. Dong, Z. Zhang, G. Wen et al., "Design and application of unsupervised deep belief networks for mechanical fault," in *Proceedings of the 2017 Prognostics and System Health Management Conference (PHM-Harbin)*, IEEE, Harbin, China, pp. 1–7, July 2017.
- [14] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, 2016.
- [15] J. Jiao, M. Zhao, J. Lin, and J. Zhao, "A multivariate encoder information based convolutional neural network for intelligent fault diagnosis of planetary gearboxes," *Knowledge-Based Systems*, vol. 160, pp. 237–250, 2018.
- [16] Y. Xue, D. Dou, and J. Yang, "Multi-fault diagnosis of rotating machinery based on deep convolution neural network and support vector machine," *Measurement*, vol. 156, 2020.
- [17] J. An, P. Ai, and D. Liu, "Deep domain adaptation model for bearing fault diagnosis with domain alignment and discriminative feature learning," *Shock and Vibration*, vol. 2020, Article ID 4676701, 14 pages, 2020.
- [18] C. Tan, F. Sun, T. Kong et al., "A survey on deep transfer learning," in *Proceedings of the International Conference on Artificial Neural Networks*, pp. 270–279, Munich, Germany, September 2018.
- [19] S. B. John, B. Koby, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, vol. 79, no. 1-2, pp. 151–175, 2010.
- [20] D. Peng, Z. Liu, H. Wang, Y. Qin, and L. Jia, "A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains," *IEEE Access*, vol. 7, pp. 10278–10293, 2019.
- [21] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 136–144, 2019.
- [22] X. Li, W. Zhang, and Q. Ding, "A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning," *Neurocomputing*, vol. 310, pp. 77–95, 2018.
- [23] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, 2017.
- [24] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Multi-Layer domain adaptation method for rolling bearing fault diagnosis," *Signal Processing*, vol. 157, pp. 180–197, 2019.
- [25] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proceedings of the 27th International Conference on Neural Information Processing Systems, Advances in Neural Information Processing Systems*, pp. 3320–3328, Montreal, Canada, 2014.
- [26] R. Aljundi and T. Tuytelaars, "Lightweight unsupervised domain adaptation by convolutional filter reconstruction," in *Proceedings of the European Conference on Computer Vision*, pp. 508–515, Glasgow, UK, August 2016.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Lake Tahoe, NEV, USA, July 2016.
- [28] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [29] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, 2016.
- [30] Q. Li, C. Shen, L. Chen, and Z. Zhu, "Knowledge mapping-based adversarial domain adaptation: a novel fault diagnosis method with high generalizability under variable working conditions," *Mechanical Systems and Signal Processing*, vol. 147, 2021.
- [31] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proceedings of the International Conference on Machine Learning. PMLR*, pp. 97–105, Long Beach, CA, USA, July 2015.
- [32] A. Gretton, B. Sriperumbudur, D. Sejdinovic, H. Strathmann, and M. Pontil, "Optimal kernel choice for large-scale two-sample tests," in *Proceedings of the Neural Information Processing Systems*, pp. 1205–1213, Lake Tahoe, NEV, USA, December 2012.
- [33] D. Lee, "Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks," in *Proceedings of the International Conference on Machine Learning. ICML*, pp. 2–8, Seoul, Korea, December 2013.
- [34] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mechanical Systems and Signal Processing*, vol. 122, pp. 692–706, 2019.
- [35] F. Immovilli, M. Cocconcelli, A. Bellini, and R. Rubini, "Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 11, pp. 4710–4717, 2009.
- [36] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.
- [37] X. Wang, X. Zhang, Z. Li, and J. Wu, "Ensemble extreme learning machines for compound-fault diagnosis of rotating machinery," *Knowledge-Based Systems*, vol. 188, 2019.
- [38] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 1187–1192, 2010.
- [39] 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, Cambridge, MA, USA, June 2013.
- [40] X. Wang, H. He, and L. Li, "A hierarchical deep domain adaptation approach for fault diagnosis of power plant thermal system," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5139–5148, 2019.
- [41] L. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.