Hindawi

Research Article

A Deep Domain-Adversarial Transfer Fault Diagnosis Method for Rolling Bearing Based on Ensemble Empirical Mode Decomposition

Xiao Yu [1,2,3] **Bing Xia,**[1,2,3] **Shuxin Yang,**[3,4] **Hongshen Yin,**[1,2,3] **Yajie Wang,**[1,2,3]
and Xiaowen Liu [3,4]

[1]*IOT Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221000, China*
[2]*School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221000, China*
[3]*The National Joint Engineering Laboratory of Internet Applied Technology of Mines, China University of Mining and Technology, Xuzhou 221008, China*
[4]*School of Electrical and Power Engineering, China University of Mining and Technology, Xuzhou 221000, China*

Correspondence should be addressed to Xiaowen Liu; xwliucumt@126.com

In recent years, the deep learning-based fault diagnosis methods for rotating mechanical equipment have attracted great concern. However, because the data feature distributions present differences in applications with varying working conditions, the deep learning models cannot provide satisfactory performance of fault prediction in such scenarios. To address this problem, this paper proposes a domain adversarial-based rolling bearing fault transfer diagnosis model EMBRNDNMD. First of all, an EEMD-based time-frequency feature graph (EEMD-TFFG) construction method is proposed, and the time-frequency information of nonlinear nonstationary vibration signal is extracted; secondly, a multi-branch ResNet (MBRN) structure is designed, which is used to extract deep features representing the bearing state from EEMD-TFFG; finally, to solve the model domain adaptation transfer problem under varying working conditions, the adversarial network module and MK-MMD distribution difference evaluation method are introduced to optimize MBRN, so as to reduce the probability distribution difference between the deep features of source domain and target domain, and to improve the accuracy of EMBRNDNMD in state diagnosis of target domain. The results of experiments carried out on two bearing fault test platforms prove that EMBRNDNMD can maintain an average accuracy above 97% in fault transfer diagnosis tasks, and this method also has high stability and strong ability of scene adaptation.

1. Introduction

The rotating mechanical equipment has broad applications in various fields, such as related industries, military, and for civil use. As an important component of rotating mechanical equipment, the rolling bearing directly affects the operating efficiency and working conditions of mechanical equipment. However, due to long-term exposure to harsh working conditions of high load, the parts of rolling bearing tend to suffer from damages. Minor damage may reduce the operating efficiency of mechanical equipment, while serious damage might lead to shutdown of equipment, which may even cause casualties. Therefore, study on the state detection and fault diagnosis of rolling bearing has important theoretical significance and engineering value for improving productivity and ensuring production safety.

In researches on fault diagnosis based on signal processing, good results have has been achieved by combining traditional feature extraction and machine learning classification [1–4]. Especially the signal processing method represented by ensemble empirical mode decomposition (EEMD), the bearing fault mechanism information can be obtained by analyzing the intrinsic mode functions (IMFs) of vibration signal. Han et al. [5] put forward a method for rolling

bearing fault diagnosis based on EEMD permutation entropy and fuzzy clustering. Wang et al. [6] proposed an improved EEMD algorithm, in which the sifting and ensemble number are self-adaptive. Yang et al. [7] and Hu et al. [8] proposed a fault detection and diagnosis method based on EEMD and support vector machine (SVM). Besides, Shao et al. [9] put forward a method called deep wavelet auto-encoder with extreme learning machine (ELM) for intelligent fault diagnosis of rolling bearing. Li et al. [10] proposed a density-based clustering method with principal component analysis (PCA) to improve the performance of variable load diagnosis in fault diagnosis. In the above researches, appropriate signal processing methods need to be selected to extract effective features according to the characteristics of data, such as EEMD, wavelet transform (WT), and PCA. However, these methods are too empirical, as a result of which, the selection of features will directly affect the diagnosis results. In order to reduce the influence of human experiences, a better method is to enable the model to automatically extract features [11, 12].

In recent years, with the rapid development of deep learning in computer vision, many scholars have applied deep learning methods to the field of fault diagnosis. Compared with machine learning, deep learning can adaptively extract deep features from signals, which has solved the difficulty to extract fault features [13]. Zhou and Yao et al. [14, 15] developed a convolutional neural network (CNN) based fault diagnosis method for rolling rearing by using the waveform of vibration as the 2-D image input of CNN. Fan et al. [16] advanced a method about the convolutional neural network and transfer learning based fault diagnosis method, aiming at the vibration image samples of rolling bearing affected by strong noise.

However, with the increase of network layers, the traditional deep learning methods suffer from the problems of gradient disappearance and gradient explosion, and as a result the weight of the model cannot be updated effectively [17, 18]. In order to solve this problem, HE et al. [19] proposed a deep residual network (ResNet) in 2015, which uses shortcuts to directly transmit the data of the front layer to the back layer of the network and completes the feature fusion through addition. Wei et al. [20] presented a novel framework that combines a residual network as a backbone and an extreme learning machine as a classifier to diagnose the faults of rotating machinery. Wang and Wen et al. [21, 22] constructed a multi-scale deep intra-class adaptation network, which uses the modified ResNet-50 to extract low-level features, and the experimental results show that the model outperforms both other deep learning models and the conventional methods. How to effectively extract features in the bearing intelligent diagnosis model is a problem worthy of study. This article proposes a method based on the EEMD and improved multi-branch ResNet to extract deep features of bearing faults from the time domain and frequency domain. However, the success of deep learning methods in fault diagnosis largely depends to sufficient and labelled training samples, but it is difficult to meet these requirements in practical works.

Transfer learning method can effectively solve the problem of data scarcity, because it is able to apply the knowledge learned in the source domain into the target domain, which can help improve the prediction accuracy of unlabelled data [23–26]. In the field of fault diagnosis, the transfer learning methods can be divided into model-based methods, such as maximum mean discrepancy (MMD), and domain distribution-based methods, such as domain adversarial training of neural networks (DANN) [27–29]. Li and Yang et al. [30, 31] presented a feature representation enhancement method based on MMD and domain confrontation training. Che et al. [32] put forward a domain-based adaptive method, which can calculate the multi-core maximum mean difference (MK-MMD) of the selected hidden layer and add it to the loss function, so as to improve the generalization ability of the deep neural network. Tang et al. [33] integrated the MK-MMD loss into the traditional fine-tuning convolutional neural network (CNN) transfer learning framework and proposed a new semi-supervised transfer learning (STL) method. Mao and Cai et al. [34, 35] advanced a novel adversarial DA method called the adversarial residual transformation network (ARTN), which directly transforms the source features into the target feature space to improve the generalization capability. Li et al. [36] came up with a novel weighted adversarial transfer network (WATN) for fault diagnosis in certain domains and achieved satisfactory performance. Both MMD and DANN have achieved good performances in fault diagnosis, but in some situations of variable working conditions, a single transfer method often performs poorly. Therefore, this paper proposes a multiple transfer fault diagnosis method combining MMD and DANN to address degradation of model diagnosis performance in scenarios with transfer of working conditions.

The deep learning technique has the advantage of adaptively extracting deep features of data, which can be used to build an end-to-end diagnosis mechanism. A lot of researches have been carried out on intelligent diagnosis models for rotating mechanical equipment based on deep learning. However, various problems are also encountered in current studies, such as vibration signals susceptible to interference of noises, insufficient samples of equipment faults, and the difference in distributions between target data and source data caused by change of equipment working conditions. Solving these problems should be the focus of deep learning-based fault diagnosis model in future studies.

In our paper, the EEMD method is adopted to preprocess the rolling bearing vibration signal, and we propose obtaining the time-frequency information of vibration signal by building EEMD-TFFG. To address the problem of degraded model diagnosis ability under transfer of working conditions, DANN and MK-MMD are introduced to optimize the multi-branch ResNet (MBRN), so as to reduce the differences in probability distributions of deep features between source domain and target domain and to improve the state diagnosis accuracy of target domain. The main contributions of this paper are as follows:

(1) We propose an EEMD-based vibration signal construction method—EEMD-TFFG, so as to achieve

time-frequency analysis and feature extraction of vibration signals

(2) A multi-branch feature extraction network MBRN based on ResNet is designed, which can extract deep features reflecting the fault state from EEMD-TFFG

(3) DANN and MK-MMD are introduced to optimize MBRN, reduce the probability distribution difference of data deep features between source domain and target domain, and improve the diagnosis ability of EMBRNDNMD under transfer of working conditions. The experimental analysis results show that EMBRNDNMD can achieve a high diagnosis accuracy for target domain states under various transfer modes, which has a strong ability to adapt to varying working conditions

Section 2 introduces the principles of various methods, including EEMD, ResNet, DANN, and MK-MMD; Section 3 presents the design idea and the structure of EMBRNDNMD model; in Section 4, experiments are carried out on the two datasets of CWRU and MFS-RDS by using EMBRNDNMD, and related analysis is conducted; Section 5 draws conclusions of our work. Furthermore, we present some acronyms in Table 1.

2. Preliminaries

2.1. Ensemble Empirical Mode Decomposition (EEMD). Hilbert-Huang transform (HHT) is one of the time-frequency analysis methods with broadest applications. First, HHT carries out empirical mode decomposition (EMD) of signal to obtain a series of intrinsic mode functions (IMFs) of different scales; then, instant frequency information with physical significance is obtained through Hilbert transform of various IMF components. However, EMD still has some problems, such as end effect and mode confusion [37], which may reduce the accuracy of fault classification. To address the mode confusion problem of EMD, Zhaohua et al. [38] proposed the ensemble empirical mode decomposition (EEMD) method based on EMD. EEMD adds white Gaussian noise on the basis of original signal, which makes the signal smooth, effectively inhibits mode confusion, and improves the precision of signal decomposition. The decomposition process of EEMD is as shown in Figure 1.

The specific decomposition procedure of EEMD consists of following steps:

(1) For the given original signal $x(t)$, initialize variable i =1, and set the mean times of EEMD set as M

(2) For the original signal, add a group of white noises $n_i(t)$, and obtain signal $x_i(t)$.

$$x_i(t) = x(t) + n_i(t), \qquad (1)$$

where $x_i(t)$ is the i-th decomposed signal of additive white noise, and $n_i(t)$ is the i-th additive white noise (i=1, 2, 3, …, M).

(3) Carry out EMD of $x_i(t)$, and obtain various IMF components and residual components as

$$x_i(t) = \sum_{j=1}^{J} c_{ij}(t) + r_{ij}(t), \qquad (2)$$

where $c_{ij}(t)$ is the j-th IMF component obtained in the i-th decomposition, and $r_{ij}(t)$ is the residual component of the i-th decomposition.

(4) Obtain the sum and average value of corresponding IMF components got in M decompositions to offset the noise, and obtain the final IMF components

$$c_j(t) = \frac{1}{M} \sum_{i=1}^{M} c_{ij}(t), \qquad (3)$$

where $c_j(t)$ (j=1,2,3, …, J) is the final j-th IMF component obtained from EEMD.

(5) Through EEMD, signal $x(t)$ is finally decomposed to:

$$x(t) = \sum_{j} c_j(t) + r(t), \qquad (4)$$

where $r(t)$ is the final residual component obtained after EEMD of signal.

2.2. Residual Network (ResNet). Convolutional neural network (CNN) is a network structure commonly used in deep learning, which has achieved broad applications in the field of fault diagnosis. CNN mainly consists of the input layer, output layer, and hidden layer, while the hidden layer can also be further divided into the convolutional layer, down-sampling layer, and fully connected layer. In the traditional CNN, the expression ability will be enhanced with the increase of depth, and more complicated features can be extracted. However, if the network layers are too deep, it may also cause gradient attenuation, gradient explosion, and other problems, leading to declined accuracy of prediction.

To solve the problem of model degradation caused by convolutional network being too deep, in 2015, Kaiming He et al. [19] from Microsoft Research proposed the residual network (ResNet). Utilizing shortcut, ResNet directly transfers data from previous layers to following layers of network and uses addition to achieve feature fusion, as shown in Figure 2.

Addition connects input x and $F(x)$ obtained after stacking weight layers cross layer, and obtains output $H(x)$ = $F(x) + x$. Here, $F(x) = H(x) - x$, which is the residual. Because the residual network has integrated such skip structure, even if the network depth is increased, the learning network is only added with the load of identity function

TABLE 1: A list of acronyms used in this paper.

| Acronym	Full form	Acronym	Full form
EMBRNDNMD	Domain adversarial-based rolling bearing fault transfer diagnosis model	TL	Transfer learning
EMD	Empirical mode decomposition	CNN	Convolutional neural networks
EEMD	Ensemble empirical mode decomposition	ResNet	Residual network
EEMD-TFFG	Time-frequency feature graph	MMD	Maximum mean discrepancy
MBRN	Multi-branch ResNet	MK-MMD	Multi-core maximum mean difference
RNB	ResNet module	GAN	Generative adversarial network
IMF	Intrinsic mode function	DANN	Domain-adversarial neural network
IMFs	Intrinsic mode functions	GRL	Gradient reversal layer
HHT	Hilbert-Huang transform	t-SNE	t-Distributed stochastic neighbor embedding
HES	Hilbert envelope Spectrum	TCA	Transfer component analysis
SVM	Support vector machine	JDA	Joint distribution adaptation
PCA	Principal component analysis	CWRU	Case Western Reserve University
ELM	Extreme learning machine	MFS-RDS	Mechanical fault diagnosis experiment platform
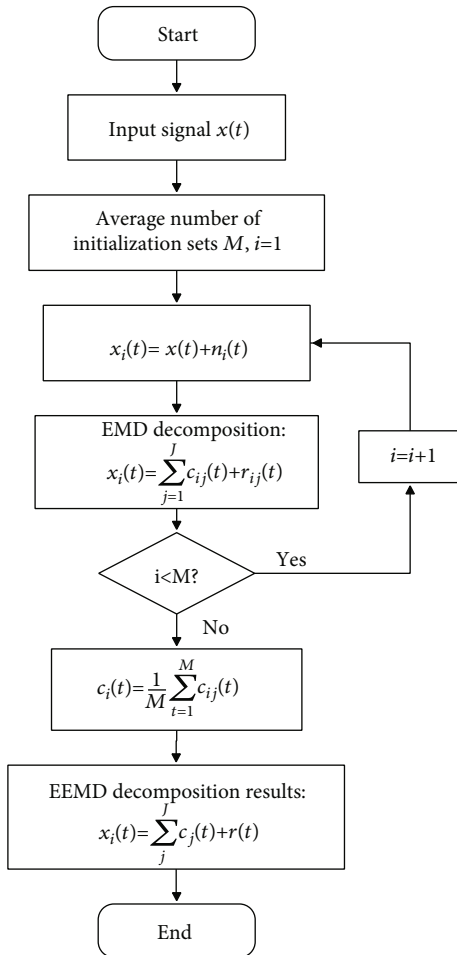
FIGURE 1: Decomposition flow chart of EEMD.

computation, and the data utilization efficiency will not be reduced. As a result, more data information can be transferred to deeper network, so as to prevent model degradation caused by convolutional network being too deep.

2.3. Domain-Adversarial Neural Network (DANN). The conventional machine learning methods not only require massive labelled data for training but also need similar probability distributions between the source domain and target domain. If the source domain and target domain present significant difference in data distribution, the generalization performance of model will degrade in the target domain. The domain adaptive transfer learning mechanism is an effective approach to solve this problem.

In recent years, with the application of generative adversarial network (GAN) in image processing, the ideal of generative adversarial has been broadly used in adaptive transfer learning applications in various fields. A classic example is the domain-adversarial neural network (DANN) proposed by Ganin Y [29] in 2016. DANN utilizes the feature extractor and domain discriminator for adversarial training, and the Nash equilibrium can be finally reached, making the domain classifier unable to determine which domain the data comes from. In this way, data from the source domain and target domain with different distributions can be mapped to the same feature space, and the classifier trained in the source domain can be used to directly classify data from target domain. The structure of DANN is as shown in Figure 3 [29].

Specifically speaking, DANN consists of the three parts of feature extractor, label predictor, and domain classifier. The feature extractor is used to (1) confuse data from the source domain and target domain to trick the domain classifier; (2) extract features required by subsequent network from the mixed data. The feature extractor and label predictor form a feed-forward neural network to achieve
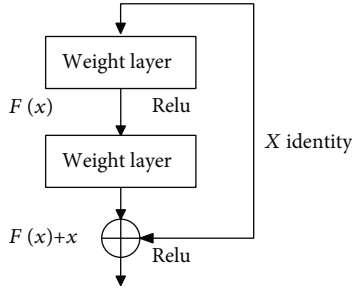
FIGURE 2: Residual module of ResNet.

adversarial training between different fields. DANN adds a domain classifier after the feature extractor, which is connected by a gradient reversal layer (GRL). With the addition of GRL, the gradient direction will be automatically flipped during back propagation in the training process of model, and identical transformation can be achieved during forward propagation.

Assuming X represents the input space, $Y = \{0, \cdots, L-1\}$ represents L labels, S is the labelled source sample and T is the unlabelled target sample. There are two different distributions of source domain D_S and target domain D_T, then the sample functions of source domain and target domain can be expressed as:

$$S = \{(x_i, y_i)\}_{i=1}^{n} \sim (D_S)^n \, ; \, T = \{(x_i)\}_{i=n+1}^{N} \sim \left(D_T^X\right)^{n'}, \quad (5)$$

where $N = n + n'$, which represents the total number of samples.

According to the source domain sample (x_i, y_i), the classification loss of label predictor can be represented by the negative logarithmic probability of correct label:

$$L_y\big(G_y\big(G_f(x_i)\big), y_i\big) = \log \frac{1}{G_y\big(G_f(x)\big)_{y_i}}, \quad (6)$$

where G_f represents feature extraction network of X; $G_y(G_f(x))$ represents the conditional probability of network mapping X to Y.

Similarly, the loss function of domain classifier can be represented as:

$$L_d\big(G_d\big(G_f(x_i)\big), d_i\big) = d_i \log \frac{1}{G_d\big(G_f(x_i)\big)} + (1 - d_i) \log \frac{1}{1 - G_d\big(G_f(x_i)\big)}, \quad (7)$$

where d_i is a binary variable representing the domain class. If $d_i = 0$, it means $X_i \sim D_S^X$; if $d_i = 1$, $X_i \sim D_T^X$. G_d represents the domain classifier.

The total loss L_{DANN} of model consists of the two parts of source label prediction loss L_y and domain classifier loss L_d:

$$L_{DANN}\big(\theta_f, \theta_y, \theta_d\big) = \frac{1}{n_S} \sum_{x_i \in D_S} L_y\big(G_y\big(G_f(x_i)\big), y_i\big) - \frac{\lambda}{n} \sum_{x_i \in (D_S \cup D_T)} L_d\big(G_d\big(G_f(x_i)\big), d_i\big). \quad (8)$$

During the training process, the feature extractor learns parameter θ_f by maximizing the loss function L_d of domain classifier, and the domain classifier adjusts its parameter θ_d by minimizing the loss function L_d.

2.4. Multi-Kernel Maximum Mean Discrepancy (MK-MMD). Maximum mean discrepancy (MMD) was proposed for double-sample test, which is used to determine the distribution difference between two types of data, and it is a common loss function in transfer learning. In MMD, the most critical step is to choose kernel parameters, and unsuitable kernel parameters will not only affect the final performance of mapping but also cause deviation in distance measurement. To prevent selecting unsuitable kernel of MMD, the MK-MMD method proposed by Gretton [39] is employed in our paper. In MK-MMD, it is assumed that the optimal kernel is obtained via linear combination of multiple kernels, which can prevent choosing unsuitable kernel parameters when only one kernel is used. Assuming the source domain dataset $X^s = \{x_i^s\}_{i=1,\cdots n}$ satisfies P distribution and the target domain dataset $X^t = \{x_j^t\}_{j=1,\cdots n}$ satisfies Q distribution, the Euclidean distance between X^s and X^t in MK-MMD is defined as:

$$d_k^2(P, Q) \triangleq \big\| E_P[\phi(X^s)] - E_Q\big[\phi(X^t)\big] \big\|_{\mathscr{H}_k}^2, \quad (9)$$

where E represents the mathematical expectation; ϕ stands for the mapping of reproducing Hilbert space; $\mathscr{H}_k$ refers to the reproducing kernel Hilbert space with feature kernel k.

The feature kernel often chooses the convex optimization combination of m kernels $\{k_u\}$ associated with features to provide effective mapping. The feature kernel is defined as follows:

$$\mathscr{K} \triangleq \left\{ k = \sum_{u=1}^{m} \partial_u k_u : \sum_{u=1}^{m} \partial_u = 1, \partial_u \geq 0, \forall_u \right\}, \quad (10)$$

where ∂_u is the weighted parameter of different kernels, and the characteristic of multi-kernel k is guaranteed via constraint of ∂_u.

3. Proposed Method and System Framework

By combining EEMD and MK-MMD, this paper proposes a deep residual adversarial transfer bearing fault analysis method EMBRNDNMD. In our paper, first, EEMD is utilized to adaptively decompose the vibration signal into empirical mode components IMFs of different scales, and IMFs and corresponding Hilbert envelope spectrum (HES) form the EEMD time-frequency feature graph (EEMD-TFFG) of time-frequency features. Then, the multi-branch ResNet structure is used to extract deep features of EEMD-TFFG, the domain adversarial mechanism is introduced to ensure consistent low-dimensional expressions of the deep features of data between source domain and target domain, and in the meantime, MK-MMD is utilized to constrain the distribution difference between them in high-dimensional
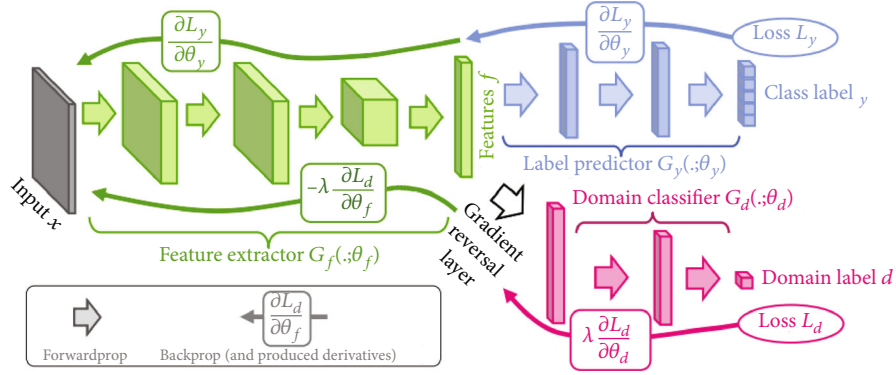
Figure 3: Schematic diagram of DANN model structure.

space. Finally, the back propagation of ResNet is optimized according to the fault state classification loss of source domain data, the discriminant loss between source domain and target domain, and the distribution difference loss of MK-MMD, so as to improve the state classification ability and domain adaptation ability of deep features, and to solve the transfer problem of state diagnosis model under different working conditions.

3.1. Construction of EEMD Time-Frequency Graph. After EEMD of vibration signal, a group of linear stable empirical mode components IMFs are obtained, and IMFs are automatically distributed from high frequency to low frequency. Considering that not every IMF can effectively represent the time-frequency characteristic or the information of original signal, Equation (11) is utilized to calculate the correlation coefficient between each IMF component and the original signal $x(t)$, so as to eliminate illusive components in IMF component.

$$\rho_{(c_i(t),x(t))} = \frac{E[(c_i(t) \cdot x(t) - E[c_i(t)] \cdot E[x(t)])]}{\sqrt{D(c_i(t))} \cdot \sqrt{D(s(t))}}, \qquad (11)$$

where $c_i(t)$ represents the i-th IMF component, $E[\cdot]$ represents the expected value of signal, and $D(\cdot)$ represents the mean square value of signal.

The bigger the correlation coefficient is, the more closely related is the IMF component to the original signal, and the richer time-frequency information it contains. Then, the Hilbert envelope spectrum (HES) used to select the IMF component is calculated. Here, the selected IMF components and its envelope spectrum are rearranged into a matrix in the order that IMF is the first and HES is the last, so as to improve the correlation of the features and obtain a group of time-frequency feature graphs, which are denoted as EEMD-TFFG, and this step aims to facilitate subsequent extraction of their deep features using 2D convolution kernel.

The construction process of EEMD-TFFG includes the following steps:

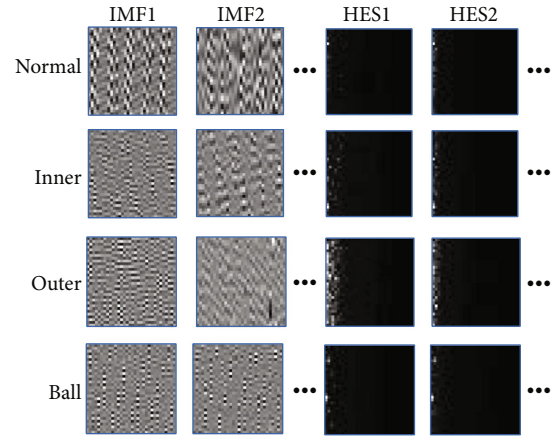(1) After EEMD of the vibration signal, a group of empirical mode components IMFs are obtained

Figure 4: EEMD-TFFG of bearing vibration signals.

(2) The correlation coefficient between each IMF component and the original signal is calculated, and the IMF components with correlation coefficients higher than the threshold value are selected for subsequent analysis

(3) Corresponding HES of IMF components selected in Step (2) is calculated

(4) The selected IMF components and HES sequences are rearranged into a matrix, and saved as a gray-scale image

3.2. Design of Network Model Structure

3.2.1. Design of Deep Feature Extraction Network. Figure 4 shows the EEMD-TFFG of a group of vibration signals under different bearing states, and each vibration signal sample includes 1024 sampling points. We can see that EEMD-TFFG has the following two characteristics:

(1) Each gray-scale image has the size of $32*32$, which is small

(2) For the same signal, the features between the gray-scale images of different IMF components are relatively independent
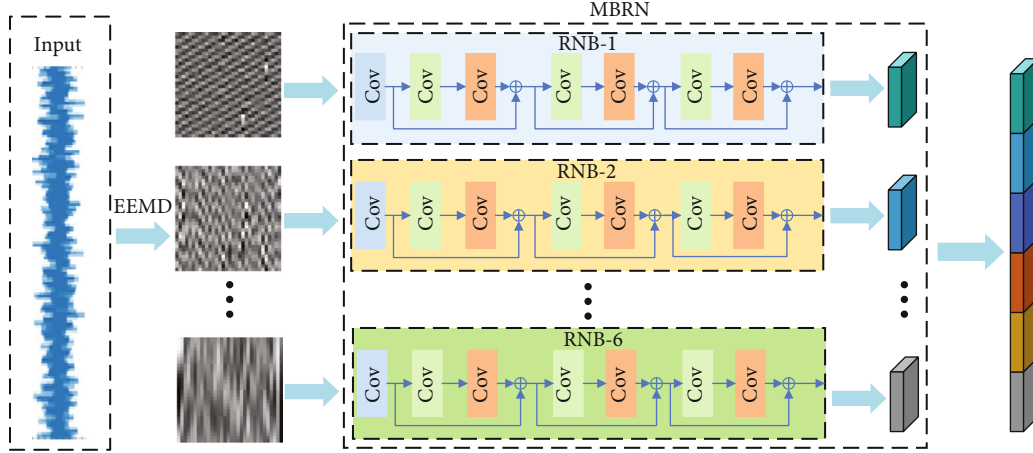
FIGURE 5: Multi-branch parallel ResNet structure.

TABLE 2: Specific parameters of RNB.

Type	Layer	Kernel/stride/channels	Input size	Output size
Input		/	(32,32,1)	(32,32,1)
	Conv2d	(3,3)/1/16	(32,32,1)	(32,32,16)
Conv1	InstanceNorm2d	/	(32,32,16)	(32,32,16)
	ReLU	/	(32,32,16)	(32,32,16)
	Conv2d	(3,3)/1/16	(32,32,16)	(32,32,16)
	InstanceNorm2d	/	(32,32,16)	(32,32,16)
Block1	ReLU	/	(32,32,16)	(32,32,16)
	Conv2d	(3,3)/1/16	(32,32,16)	(32,32,16)
	InstanceNorm2d	/	(32,32,16)	(32,32,16)
	Conv2d	(3,3)/1/32	(32,32,16)	(32,32,32)
	InstanceNorm2d	/	(32,32,32)	(32,32,32)
Block2	ReLU	/	(32,32,32)	(32,32,32)
	Conv2d	(3,3)/2/32	(32,32,32)	(16,16,32)
	InstanceNorm2d	/	(16,16,32)	(16,16,32)
	Conv2d	(3,3)/1/32	(16,16,32)	(16,16,32)
	InstanceNorm2d	/	(16,16,32)	(16,16,32)
Block3	ReLU	/	(16,16,32)	(16,16,32)
	Conv2d	(3,3)/2/32	(16,16,32)	(8,8,32)
	InstanceNorm2d	/	(8,8,32)	(8,8,32)
Pooling	AvgPool	/	(8,8,32)	(1,1,32)

Based on these two characteristics of EEMD-TFFG, we designed a multi-branch parallel ResNet structure as shown in Figure 5, which is denoted as MBRN. In Figure 5, we assume that 3 IMF components of vibration signal after EEMD and corresponding HES are selected. The parameters of each convolution layer of MBRN are shown in Table 2, in which the normalization and *relu* layers are not represented.

According to characteristic (1) of EEMD-TFFG, if the network layers are too deep, it will affect the extraction of features of small-size image, so a single ResNet module (RNB) is set with 1 convolutional layer and 3 basic residual modules, and there are 7 convolutional layers in total, and this setting can restrict the network depth. In RNB, various convolutional fea-

ture extraction layers have all used a 3∗3 convolution kernel, a small receptive field is utilized for network stacking, and the step size is set as 1. Besides, because the main redundant information has been filtered through the EEMD time-frequency figures, it will not cause information redundancy even if there is no pooling layer. So we cancel the pooling layer of the model to reduce the computational load.

According to characteristic (2) of EEMD-TFFG, a multi-branch parallel network structure MBRN is built. RNB, which has the same structure and independent parameters, is used to extract the gray-scale image features for different IMF and HES, and the deep features from the final output layers of various RNBs are combined and used as the output feature F of MBRN.

3.2.2. Loss Calculation and Back Propagation Network. The structural design of EMBRNDNMD model is as shown in Figure 6. In addition to the deep feature extraction network G_{MBRN}, the model also includes the state classification network G_y and domain discriminant network G_d. Here, G_y is a two-layer fully connected linear network, and G_d is a three-layer fully connected linear network. Three loss functions are used to optimize the network model via back propagation, and they are the bearing state classification loss L_y, the discriminant loss L_d between source domain and target domain, and the MK-MMD distribution difference loss L_{MK-MMD} between deep features of source domain and target domain, respectively. The deep feature set of source domain is denoted as $F^s = \{f_i^s\}_{i=1,..n}$, the sample label of source domain data is denoted as Y^s, and the deep feature set of target domain is denoted as $F^t = \{f_j^t\}_{j=1,..n}$.

The bearing state classification loss L_y is used to optimize G_{MBRN} and G_y. L_y is defined as:

$$L_y = \sum_{i=1}^{n} \log \frac{1}{G_y(f_i^s)_{y_i^s}}. \qquad (12)$$

L_d involves two back propagation stages, which are for optimizations of G_{MBRN} and G_d, respectively. The two stages
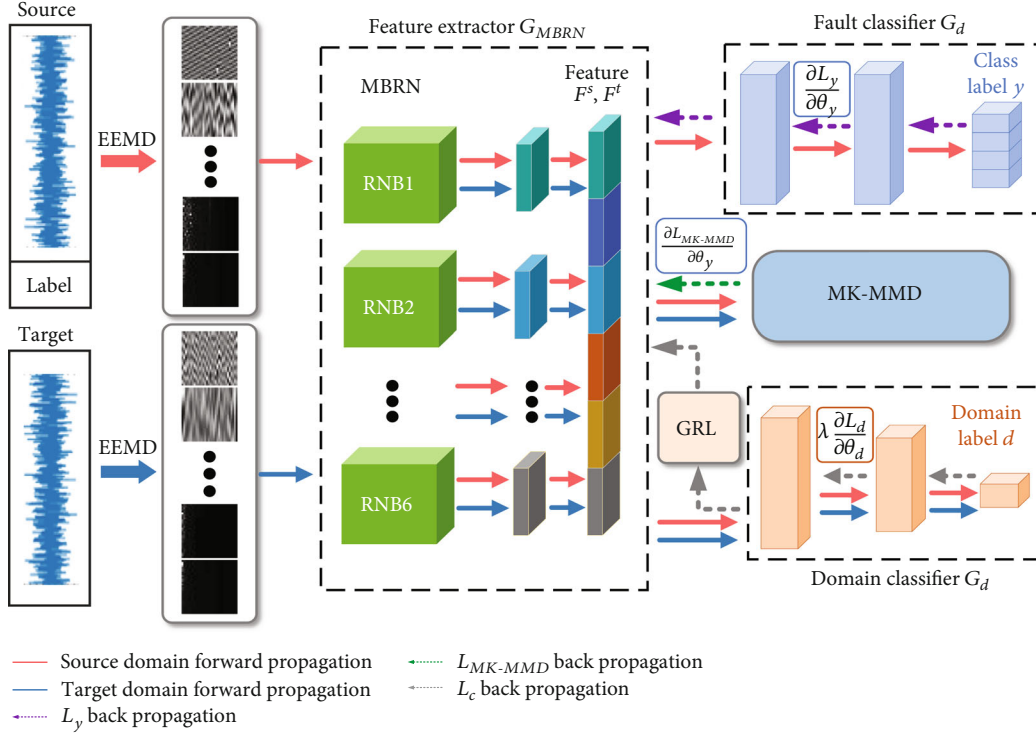
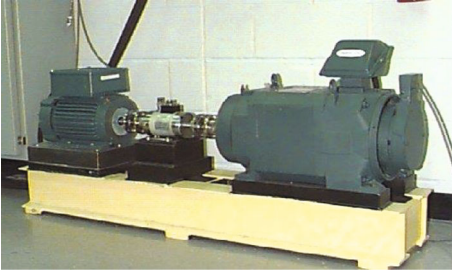FIGURE 6: Bearing fault transfer diagnosis model EMBRNDNMD.

FIGURE 7: CWRU bearing fault diagnosis and experiment platform.

of back propagation are connected by the gradient reverse layer (GRL), and the reverse mechanism of GRL is utilized to form an adversarial relation between G_{MBRN} and G_d. Back propagation optimization aims to reach Nash equilibrium between G_{MBRN} and G_d. The equation of L_d is:

$$L_d = \sum_{i=1}^{n} \log \frac{1}{G_d(f_i^s)} + \sum_{i=1}^{n} \log \frac{1}{1 - G_d(f_i^t)}. \quad (13)$$

L_{MK-MMD} represents the MK-MMD distribution difference loss, which is used to optimize G_{MBRN}. L_{MK-MMD} is defined as:

$$L_{MK-MMD} = \|E[\phi(F_s)] - E[\phi(F_t)]\|_{H_k}^2, \quad (14)$$

where E represents the mathematical expectation; ϕ stands for the mapping of reproducing Hilbert space; and H_k refers to the kernel k used by reproducing the kernel Hilbert space.

The total loss of G_{MBRN} can be expressed as:

$$L_{MBRN} = L_y - \lambda_1 L_d + \lambda_2 L_{MK-MMD}. \quad (15)$$

3.3. Procedure of Diagnosis Model. The procedure of transfer diagnosis using the EMBRNDNMD model is as follows:

(1) Collect rolling bearing vibration signals under different working conditions, assign data into the source domain or target domain, the source domain consists of labelled data, and the target domain is composed of unlabeled data

(2) Use the EEMD method to calculate the IMF and HES of vibration signal samples from the source domain and target domain, and build corresponding EEMD-TFFG

(3) Input the EEMD-TFFG of source domain and target domain into MBRN, and extract deep features F_s and F_t of EEMD-TFFG

(4) Calculate the state classification loss L_y of source domain data, and optimize G_y via back propagation

(5) Calculate the MK-MMD distribution difference between the deep features of source domain and target domain, and obtain L_{MK-MMD}

(6) Calculate the domain classifier loss L_d, and optimize G_d via back propagation

TABLE 3: Experiment datasets of bearing.

	Power	Normal		Inner race				Outer race			Rolling ball		
Damage (mm)	—	—	0.007	0.014	0.021	0.028	0.007	0.014	0.021	0.007	0.014	0.021	0.028
Label	—	0	1	2	3	4	5	6	7	8	9	10	11
Dataset A	0HP	50	50	50	50	50	50	50	50	50	50	50	50
Dataset B	1HP	50	50	50	50	50	50	50	50	50	50	50	50
Dataset C	2HP	50	50	50	50	50	50	50	50	50	50	50	50
Dataset D	3HP	50	50	50	50	50	50	50	50	50	50	50	50
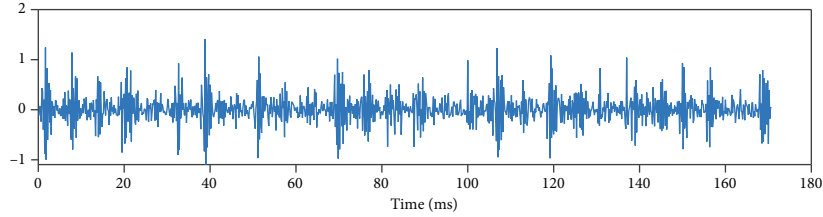
(a) Vibration signal
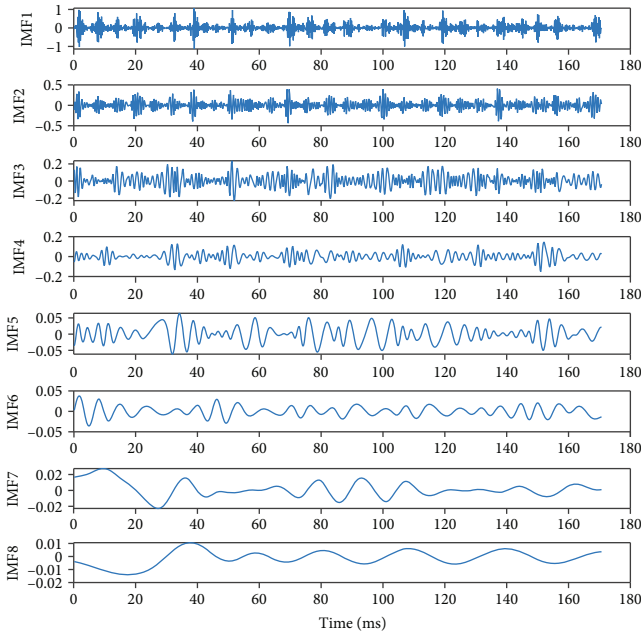
(b) IMF components
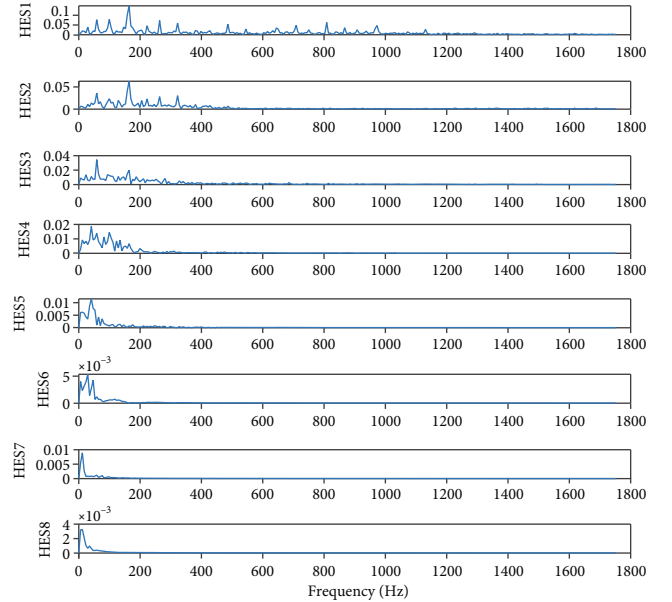
(c) Hilbert envelope spectrum

FIGURE 8: Vibration signal and its time-frequency components.

(7) Calculate the total loss L_{MBRN}, and optimize G_{MBRN} via back propagation

(8) Iterate steps (3)-(7) until L_{MBRN} is smaller than the set value or iterations have reached the target requirement, and obtain G_{MBRN} and G_y after training

(9) The trained G_{MBRN} is utilized to calculate the deep feature F_t of sample EEMD-TFFG from the source domain, and input F_t into the trained G_y to obtain the label of test samples

4. Experimental Verification

4.1. Experimental Analysis on the CWRU Bearing Dataset

4.1.1. Introduction of the CWRU Bearing Dataset. In our experiment, the CWRU bearing fault simulation and experiment platform developed by Case Western Reserve University was used, and the rolling bearing vibration signals under various states were collected to verify the performances of algorithm and model proposed in this paper. The experiment platform is presented in Figure 7, which mainly consists of the parts of motor, rolling bearing, axis of rotation, torque sensor/decoder, acceleration sensor, and signal acquisition instrument.

In the experiment, the Reliance Electric motor of 2 HP was used. Electrical discharge machining was used to create different types of faults for motor bearings, the locations included inner race, outer race, and rolling ball, and the

damage diameters were 0.007 inch, 0.014 inch, 0.021 inch, and 0.028 inch, respectively. As shown in Table 3, there are 12 types of faults and 50 samples in each type.

In this paper, the vibration signals at motor drive end with sampling frequency of 12 kHz are chosen for analysis. In the experiment, four different motor powers of 0HP, 1HP, 2HP, and 3HP were set as 4 different working conditions, and 12 transfer modes were obtained (A->B, A->C, A->D, B->A, B->C, B->D, C->A, C->B, C->D, D->A, D->B, D->C). Among them, A->B means that we set the dataset A as source domain and dataset B as target domain.

4.1.2. EEMD Analysis. First of all, we obtain the IMF components of vibration signal sample by EEMD; then, we perform Hilbert transform and spectral analysis of IMF components, and calculate the envelope spectra of IMF components. With fault at the inner race of bearing as example, the waveform of original vibration signal and the IMF components after EEMD are as shown in Figure 8.

The correlation calculation method described in Section 3.1 is used to select the IMF components. Under normal conditions and fault conditions such as inner race fault (IR), outer race fault (OR), and ball fault (BF), the correlation coefficients between the IMF components of bearing vibration signal at various order and the original signal are shown in Figure 9. According to Figure 9, with the increase of order, the correlation coefficient between the IMF component and the original signal gradually declines. The IMF components and the original signal only maintain a high correlation at the first four orders, so the IMF components of first four orders after EEMD and corresponding HES are chosen for subsequent extraction of deep features in this paper.

4.1.3. Validation of IMFs Selection. To verify that the first 4-order IMFs selected by the correlation calculation of bearing vibration signals can effectively characterize the bearing fault features, the first 3-, 5-, and 6-order IMFs (ET3, ET5, and ET6, respectively) are selected as comparative groups, and their performances are compared with the first 4-order IMFs (ET4) used in this paper in the input signal experiment. Finally, it is tested on the CWRU dataset, and the results are shown in Table 4.

It can be seen from the table that the diagnostic accuracy when using ET4 as input signal is basically higher than that of other groups, because ET3 lacks IMF4 component's fault features, resulting in incomplete expression of fault features. On the other hand, ET5 and ET6 add higher-order IMF on the basis of ET4, resulting in high redundancy in the signal, which interferes with the final results. The experimental results verify the validity of the conclusion reached in EEMD Analysis, which indicates that using the first 4-order IMF component as the input signal can effectively improve the accuracy of bearing fault diagnosis.

4.1.4. Analysis of Diagnosis Results. In this section, we test the transfer diagnosis performance of EMBRNDNMD model under four different working conditions of 0HP, 1HP, 2HP, and 3HP. To verify the theoretic analysis in Sec-
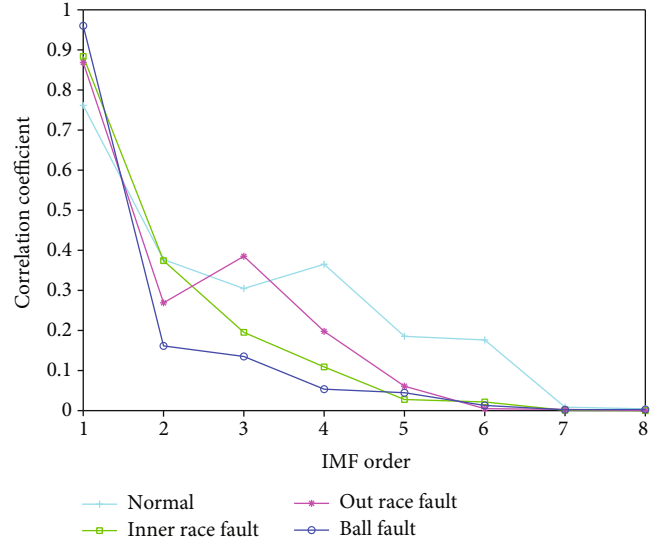

Figure 9: Correlation coefficient between the IMF and the original vibration signal.

Table 4: Test results of different IMFs input signals.

Working condition	IMF1-IMF3	IMF1-IMF5	IMF1-IMF6	IMF1-IMF4
A->B	94.17%	96.67%	97.83%	97.83%
A->C	93.50%	95.33%	96.50%	97.33%
A->D	94.00%	96.83%	97.33%	96.50%
B->A	92.83%	94.83%	95.83%	96.67%
B->C	95.83%	98.33%	99.33%	100%
B->D	94.67%	97.83%	97.50%	97.17%
C->A	92.33%	94.33%	94.83%	96.50%
C->B	92.83%	98.83%	97.33%	97.33%
C->D	94.83%	97.17%	98.67%	99.50%
D->A	93.17%	95.83%	96.33%	97.17%
D->B	94.50%	96.17%	98.00%	97.33%
D->C	95.17%	96.67%	98.83%	99.83%
Average	93.99%	96.18%	97.36%	98.17%

tion 3 and to evaluate the performance of EMBRNDNMD model, we designed some models for comparative analysis, and the specific designs include:

(1) EMBRN model: Compared to the EMBRNDNMD model, this model also uses MBRN to extract deep features of EEMD-TFFG and inputs deep features into the state classification network, but it does not involve the MK-MMD loss or the domain adversarial network

(2) EMBRNDN model: On the basis of the EMBRN model, it integrates a domain adversarial network to optimize MBRN

(3) EMBRNMD network: On the basis of the EMBRN model, it combines the MK-MMD loss to optimize MBRN via back propagation
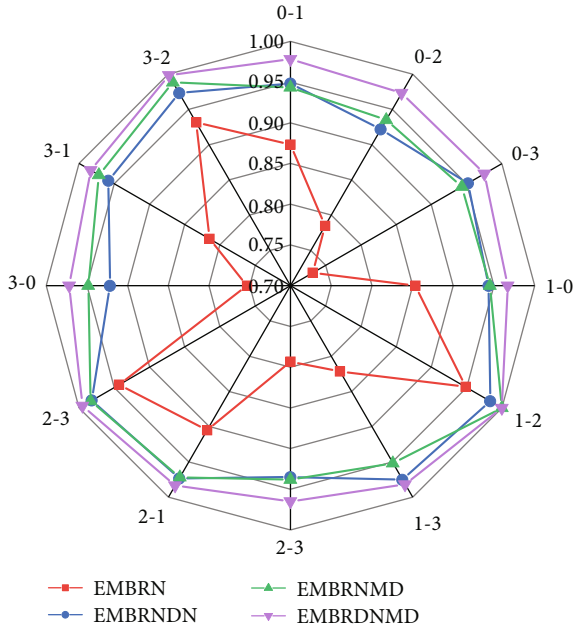
Figure 10: Accuracy distributions of various models.

Table 5 lists the state identification accuracies of every diagnosis model, and Figure 10 shows the radar comparison maps of the identification accuracies of these models. According to Table 5 and Figure 10, we can draw the following conclusions:

(1) The diagnosis accuracy of EMBRN is significantly lower than that of the other three models, which indicates that the deep features of data present distribution differences under different working conditions, and the domain adversarial network and MK-MMD domain adaptation method can well solve this problem

(2) EMBRNDNMD has higher diagnosis accuracy than EMBRNDN and EMBRNMD, which is consistent with the theoretic analysis in Section 3.2. The reason is that the EMBRNDNMD model has not only considered the consistency of deep feature distribution in high-dimensional kernel space (MK-MMD loss) but also increased the distribution similarity in low-dimensional space (domain classification loss).

(3) EMBRNDN and EMBRNMD have poor performances under partial transfer modes, but EMBRNDNMD can maintain a high accuracy under all transfer modes, and it also has better stability than the other models for comparison, which proves the effectiveness and reliability of EMBRNDNMD model

Figure 11 shows how the diagnosis accuracies of various models change with iterations on various transfer modes. According to Figure 11, in all transfer modes, every model

Table 5: Fault classification results using various network models.

Working condition	EMBRN	EMBRNDN	EMBRNMD	EMBRNDNMD
A->B	87.33%	94.83%	94.33%	97.83%
A->C	78.50%	92.17%	93.50%	97.33%
A->D	73.17%	95.17%	94.33%	97.50%
B->A	85.33%	94.33%	94.50%	96.67%
B->C	94.83%	98.33%	100%	100%
B->D	82.17%	97.50%	95.17%	98.17%
C->A	79.33%	93.50%	93.83%	96.50%
C->B	90.50%	97.33%	97.17%	98.33%
C->D	94.33%	98.17%	98.33%	99.50%
D->A	75.32%	92.17%	94.83%	97.17%
D->B	81.50%	95.83%	97.17%	98.33%
D->C	93.17%	97.33%	98.83%	99.83%
Average	84.50%	95.67%	95.83%	98.17%

can converge after 2000 iterations and become stable after 1000 iterations. Compared to the other three models, EMBRNDNMD has the fastest convergence speed, and its accuracy curve is the most stable. The analysis results show that under various transfer modes, EMBRNDNMD can not only provide high diagnosis accuracies but also has higher stability.

Figure 12 shows the t-SNE diagrams of deep features under the transfer mode of A->B by using different models, and the high-dimensional features are mapped to the two-dimensional space. According to Figure 12, compared to EMBRN, by integrating the domain transfer method, the deep features of models EMBRNMD and EMBRNDN have a bigger between-class distance and a smaller within-class distance, and the confusion problem among features under various states is significantly alleviated. By combining the MK-MMD loss and DANN, the separability of deep features of the EMBRNDNMD model is further improved, and the between-class confusion is also further reduced. The t-SNE analysis proves that compared to the other three models, the deep features extracted using EMBRNDNMD have better cross-domain invariance, and it also has stronger adaptation ability to working condition transfer.

4.1.5. The Influence of Hyperparameters on the Model. Four kinds of optimizers—Ada Delta, RMS Prop, SGD, and Adam—are selected for the test. The learning rates range from 0.001 to 0.2, and the results are listed in Table 6. It can be seen that when the learning rates are less than 0.1, the accuracy remains at a higher level. However, when the learning rates are higher than 0.1, it will make the network difficult to converge and obtain satisfactory training results. The Adam optimizer has the highest accuracy when the learning rate is 0.001, reaching 99.79%, so we ultimately choose the Adam optimizer to optimize the network parameters.

4.1.6. Comparison with Other Diagnosis Methods. To verify the effectiveness of the EMBRNDNMD model proposed in this paper under transfer of working conditions, we choose
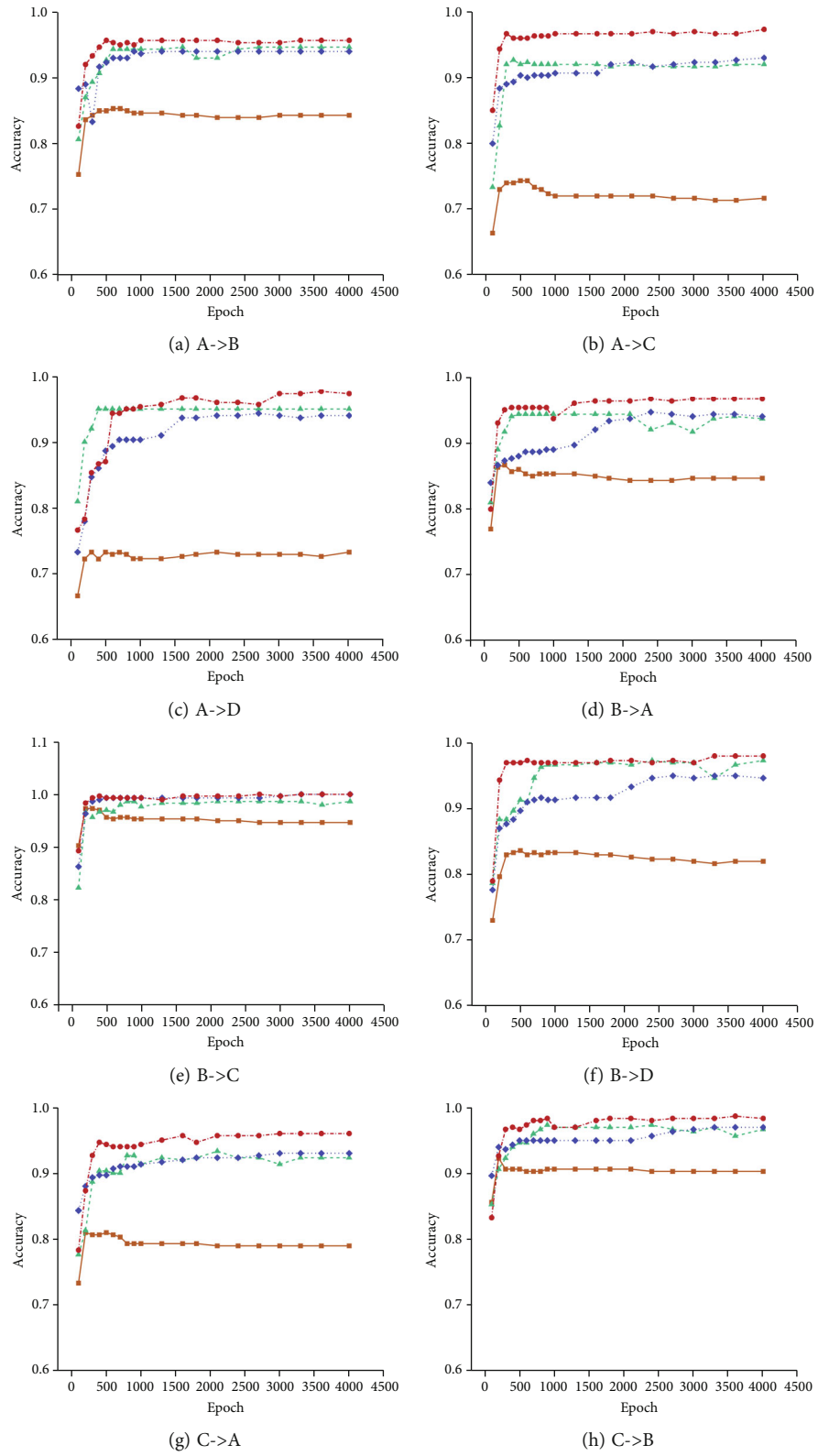
(a) A->B

(b) A->C

(c) A->D

(d) B->A

(e) B->C

(f) B->D

(g) C->A

(h) C->B

FIGURE 11: Continued.

(i) C->D
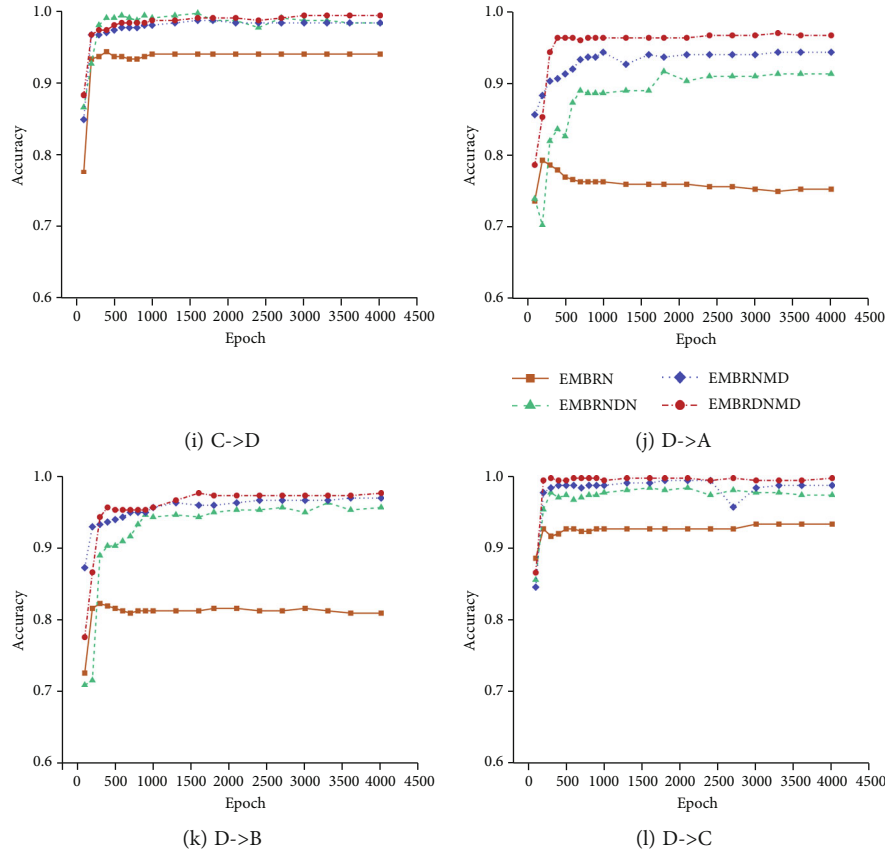
(j) D->A

(k) D->B

(l) D->C

FIGURE 11: Comparison of model accuracy-iterations.

TABLE 6: Hyperparameters experimental results.

Learning rates	Ada delta	RMS prop	SGD	Adam
0.001	99.25%	98.29%	99.46%	99.79%
0.05	99.17%	98.12%	99.21%	99.42%
0.01	98.67%	97.95%	99.16%	99.29%
0.1	94.21%	92.86%	95.37%	96.25%
0.2	93.63%	90.28%	93.71%	93.33%

some classic diagnosis models based on machine learning and deep learning to test on the CWRU dataset, including SVM, CNN, TCA, and JDA, and their diagnosis accuracies under 12 transfer modes are obtained. The results are listed in Table 7. According to comparison and analysis results, we can find:

(1) Under varying working conditions, EMBRNDNMD can provide higher diagnosis accuracies than the methods of SVM, CNN, TCA, and JDA

(2) JDA has the closest diagnosis accuracies to EMBRNDNMD, and its accuracies are even higher than 90% under some transfer modes. However, it also has poor performance under some other transfer modes, and its overall performances are not as stable as EMBRNDNMD

(3) Compared with the conventional models, EMBRNDNMD is more advantageous in solving the problem of working condition transfer, which also proves the effectiveness of the design of EMBRNDNMD model

4.2. Tests on the MFS-RDS Experiment Platform and Related Analysis

4.2.1. Introduction of MFS-RDS Experiment Platform. To verify the generalization ability of the proposed EMBRNDNMD model, the mechanical fault diagnosis experiment platform (MFS-RDS) was used to further evaluate the model performance. The MFS-RDS platform mainly consists of a three-phase motor, AC variable frequency drive (VFD) and tachometer. The sound and vibration data recorder WebDAQ-504 (MCC, US) was used for data collection. The vibration acceleration sensor was installed above the bearing seat. The experiment platform is as shown in Figure 13. In the experiments, bearings under the four states of normal condition, damage of 0.1 mm inner ball, damage of 0.1 mm outer ball, and damage of 0.1 mm rolling ball were used.

In the experiment, the vibration signals with sampling frequency of 8 kHz under the three speeds of 900 r/min, 1200 r/min, and 1800 r/min were collected, corresponding to the three working conditions of E, F, and G. With the vibration signal of 1024 continuous sampling points as a
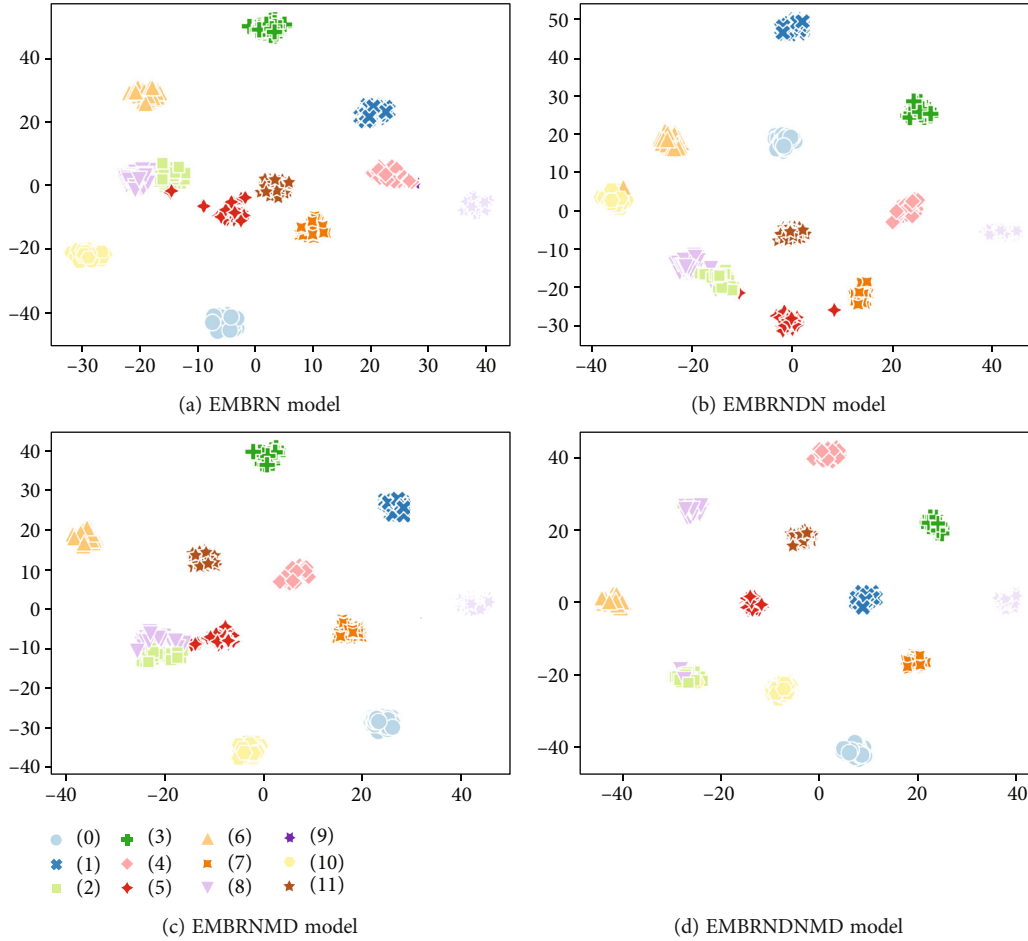
(a) EMBRN model

(b) EMBRNDN model

- (0)
- (1)
- (2)
- (3)
- (4)
- (5)
- (6)
- (7)
- (8)
- (9)
- (10)
- (11)

(c) EMBRNMD model

(d) EMBRNDNMD model

Figure 12: t-SNE diagrams of all models under the transfer learning task A->B.

Table 7: Comparison of different models in diagnosis accuracy.

Working condition	SVM	CNN	TCA	JDA	EMBRNDNMD
A->B	67.25%	78.33%	80.38%	91.75%	97.83%
A->C	67.42%	79.33%	75.67%	84.58%	97.33%
A->D	67.37%	78.17%	76.25%	85.31%	97.50%
B->A	75.13%	80.53%	91.75%	93.28%	96.67%
B->C	62.38%	78.87%	74.63%	91.27%	100%
B->D	67.56%	80.05%	82.18%	93.83%	98.17%
C->A	65.37%	71.08%	84.37%	91.08%	96.50%
C->B	61.22%	69.50%	84.53%	87.58%	98.33%
C->D	67.51%	79.33%	82.75%	91.32%	99.50%
D->A	62.34%	78.00%	80.52%	92.64%	97.17%
D->B	64%	78.83%	86.78%	91.16%	98.33%
D->C	67.36%	74.17%	86.43%	90.93%	99.83%
Average	66.24%	77.14%	82.34%	90.39%	98.17%

sample, 120 vibration signal samples were collected under each bearing state, as listed in Table 8. Three working conditions correspond to 6 transfer modes (E->F, E->G, F->E, F->G, G->E, G->F).

4.2.2. Experimental Results and Analysis. On the MFS-RDS bearing dataset, the diagnosis results of different methods under various transfer modes are presented in Table 9 and Figure 14. According to analysis of the experimental results, we can come up to the following conclusions:

(1) Under various transfer modes, the average diagnosis accuracy of EMBRN model reaches 85.39%; after introducing the domain adversarial module, the average accuracy of EMBRNDN is 93.51%; after introducing the MK-MMD loss, the average accuracy of EMBRNMD is 96.01%. This further proves that the domain adaptation mechanism can effectively improve the fault diagnosis accuracy under varying working conditions of bearing

(2) EMBRNDNMD maintains a high accuracy under all transfer modes and also shows great stability. Its average accuracy reaches 98.54%, which proves that the distribution consistency between deep features
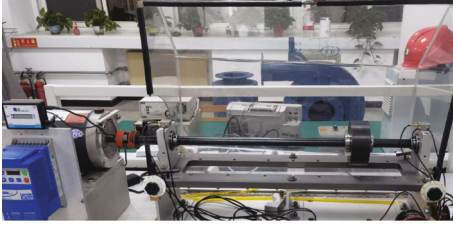
FIGURE 13: Mechanical fault diagnosis and rotor dynamic simulation and experiment platform.

TABLE 8: Introduction of MFS-RDS bearing dataset.

	Speed (r/min)	Normal	Inner race	Outer race	Rolling ball
Damage (mm)	—	—	0.1	0.1	0.1
Label	—	0	1	2	3
Dataset E	900	120	120	120	120
Dataset F	1200	120	120	120	120
Dataset G	1800	120	120	120	120

FIGURE 14: Transfer results statistics on the MFS-RDS bearing dataset.

TABLE 9: Results of various network models on the MFS-RDS bearing dataset.

	EMBRN	EMBRNDN	EMBRNMD	EMBRNDNMD
E->F	86.04%	94.38%	97.71%	99.79%
E->G	85.63%	92.92%	96.04%	98.96%
F->E	83.96%	92.08%	94.37%	98.33%
F->G	85.21%	93.95%	97.79%	97.51%
G->E	86.46%	94.79%	96.86%	98.12%
G->F	85.02%	92.92%	92.63%	98.54%
Average	85.39%	93.51%	96.01%	98.54%

TABLE 10: Comparison of different models in diagnosis accuracy.

Working condition	SVM	CNN	TCA	JDA	EMBRNDNMD
E->F	72.92%	81.67%	88.33%	94.17%	99.79%
E->G	69.17%	84.17%	86.67%	92.71%	98.96%
F->E	68.33%	80.62%	81.04%	91.04%	98.33%
F->G	66.46%	79.79%	79.79%	87.29%	97.51%
G->E	69.38%	74.79%	83.33%	93.33%	98.12%
G->F	69.58%	80.83%	86.67%	92.29%	98.54%
Average	69.31%	80.31%	84.31%	91.81%	98.54%

from source domain and target domain can be effectively improved by combining the MK-MMD loss and domain adversarial module

To further prove the above conclusions, we use the confusion matrix of the test dataset sample labels and prediction labels of various models to analyze the diagnosis precision, and use the t-SNE diagrams to carry out visual analysis of the deep features extracted by every model. Figure 15 shows the confusion matrices and t-SNE diagrams of every model under transfer mode G->E. According to the confusion matrices, we can see that EMBRNDNMD designed in this paper has the best performance. With the introduction of MK-MMD loss and domain adversarial module, the types and number of false classifications by both EMBRNDN and EMBRNMD show remarkable decline. Moreover, by combing the MK-MMD loss and domain adversarial mechanism, the number of false classifications by EMBRNDNMD is further reduced. In the meantime, the t-SNE diagrams show that compared to other models, the deep features extracted by EMBRNDNMD present bet-
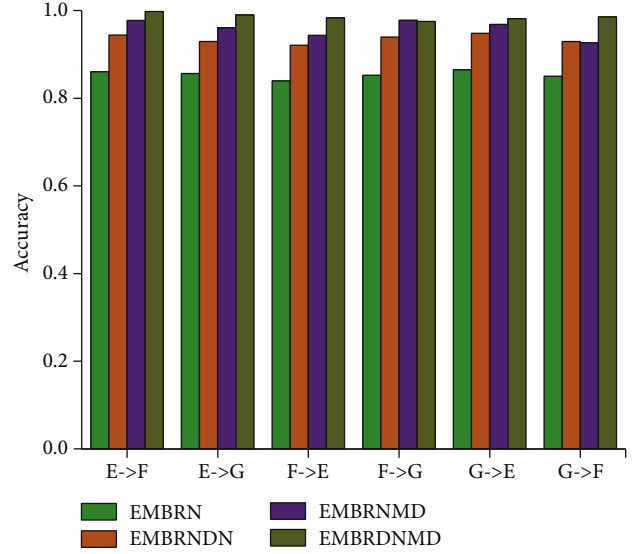
ter class separability, which proves that EMBRNDNMD has better adaptability to various scenes.

4.2.3. Comparison with Other Diagnosis Methods. To verify the generalization ability of the EMBRNDNMD model, the same comparative experiment as in Section 4.1.6 is set and carried out on the MFS-RDS dataset. The experimental results are shown in Table 10.

From the table, it can be seen that the experimental results of EMBRNDNMD on MFS-RDS datasets are basically consistent with the results in Section 4.1. The diagnostic accuracies of EMBRNDNMD under varying working conditions are higher than the other groups, and the accuracies are all above 97%. This shows that the EMBRNDNMD model still performs well in cross-platform device diagnostics. It also has excellent stability when running on the MFS-RDS datasets, which can effectively improve the fault diagnosis accuracy under varying working conditions of bearing.
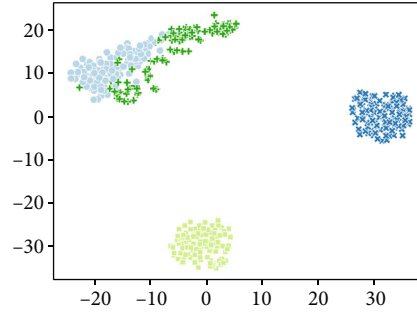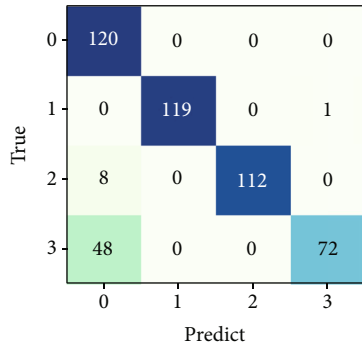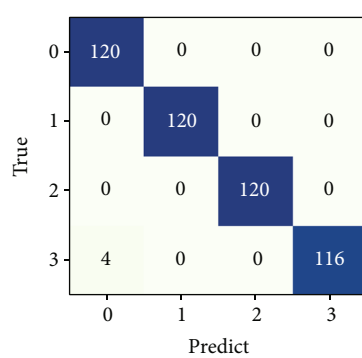
(a) EMBRN
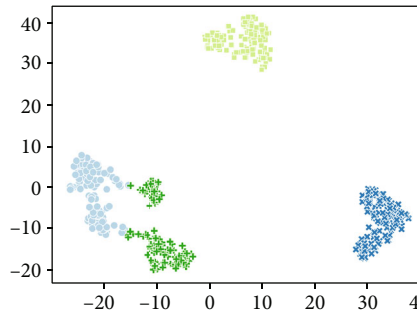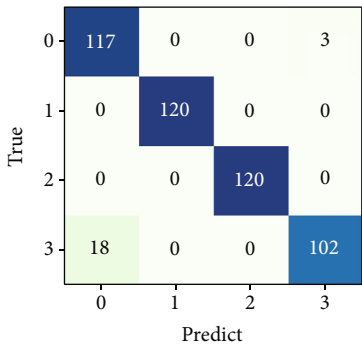
(b) EMBRNDN
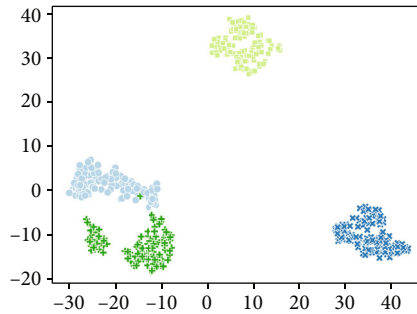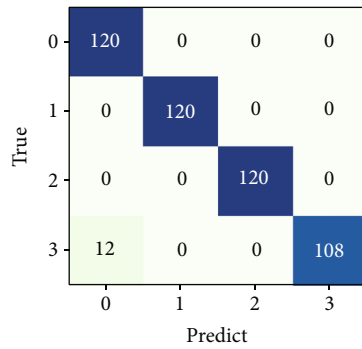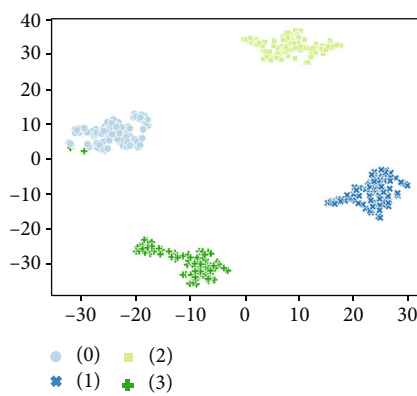

(c) EMBRNMD

(d) EMBRNDNMD

FIGURE 15: Compare the confusion matrices and t-SNE diagrams of different models.

5. Conclusions

This paper proposes a transfer diagnosis method EMBRNDNMD for rolling bearing faults. In this method, the EEMD method is used to extract the time-frequency information of the vibration signal, and the time-frequency feature graph EEMD-TFFG is constructed; then, the feature extraction network MBRN is designed according to the characteristics of EEMD-TFFG to extract deep features of EEMD-TFFG fault status; finally, the MBRN is optimized by combining DANN and MK-MMD, which improve the diagnosis ability of EMBRNDNMD under transfer of working conditions. According to theoretical derivation and experimental verification, we can draw the following conclusions:

(1) Using the EEMD method to perform time-frequency analysis of vibration signals, a construction method of EEMD-TFFG is proposed, which can provide time-frequency feature information reflecting the state of rolling bearings for subsequent deep learning networks

(2) MBRN is designed according to the characteristics of EEMD-TFFG. The multi-branch network structure and residual stacking mechanism can solve various problems of EEMD-TFFG, such as small size, scattered features, and independent time-frequency features of different scale information

(3) A joint domain transfer mechanism is designed based on DANN and MK-MMD, which can effectively improve the consistency of data deep features between the source domain and target domain, and reduce the distribution differences of deep features in high-dimensional kernel space between the source domain and target domain. It can effectively improve the diagnosis ability of EMBRNDNMD under transfer of working conditions. The results of experiments carried out on two bearing fault test platforms show that the EMBRNDNMD model can achieve high diagnostic accuracy in various working condition transfer modes

Data Availability

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

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] W. Fu, K. Shao, J. Tan, and K. Wang, "Fault diagnosis for rolling bearings based on composite multiscale fine-sorted dispersion entropy and SVM with hybrid mutation SCA-HHO algorithm optimization," *Access*, vol. 8, pp. 13086–13104, 2020.

[2] Z. Wang, L. Yao, Y. Cai, and J. Zhang, "Mahalanobis semisupervised mapping and beetle antennae search based support vector machine for wind turbine rolling bearings fault diagnosis," *Renewable Energy*, vol. 155, pp. 1312–1327, 2020.

[3] W. Gong, H. Chen, Z. Zhang et al., "A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion," *Sensors*, vol. 19, no. 7, p. 1693, 2019.

[4] Z. Huo, Y. Zhang, L. Shu, and M. Gallimore, "A new bearing fault diagnosis method based on fine-to-coarse multiscale permutation entropy, Laplacian score and SVM," *IEEE Access*, vol. 7, pp. 17050–17066, 2019.

[5] L. Han, C. Li, L. Zhan, and X. Li, "Rolling bearing fault diagnosis method based on EEMD permutation entropy and fuzzy clustering," in *Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC)*, Qinhuangdao, China, 2016.

[6] C. S. Wang, C. Y. Sha, M. Su, and Y. K. Hu, "An algorithm to remove noise from locomotive bearing vibration signal based on self-adaptive EEMD filter," *Journal of Central South University*, vol. 24, no. 2, pp. 478–488, 2017.

[7] S. Yang and J. Yue, "Fault diagnosis of EMU rolling bearing based on EEMD and SVM," in *International Conference on Computer-aided Design*, Marrakech, Morocco, May 2018.

[8] C. Z. Hu, M. Y. Huang, Q. Yang, and W. J. Yan, "On the use of EEMD and SVM based approach for bearing fault diagnosis of wind turbine gearbox," in *2016 Chinese Control and Decision Conference (CCDC)*, Yinchuan, China, 2016.

[9] S. Haidong, J. Hongkai, L. Xingqiu, and W. Shuaipeng, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowledge-Based Systems*, vol. 140, pp. 1–14, 2018.

[10] G. Li and Y. Hu, "Improved sensor fault detection, diagnosis and estimation for screw chillers using density-based clustering and principal component analysis," *Energy & Buildings*, vol. 173, pp. 502–515, 2018.

[11] Q. Qian, Y. Qin, Y. Wang, and F. Liu, "A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis," *Measurement*, vol. 178, article 109352, 2021.

[12] S. Guo, T. Yang, W. Gao, C. Zhang, and Y. Zhang, "An intelligent fault diagnosis method for bearings with variable rotating speed based on pythagorean spatial pyramid pooling CNN," *Sensors*, vol. 18, no. 11, p. 3857, 2018.

[13] S. Tang, S. Yuan, and Y. Zhu, "Data preprocessing techniques in convolutional neural network based on fault diagnosis towards rotating machinery," *Access*, vol. 8, pp. 149487–149496, 2020.

[14] F. Zhou, W. Zhou, D. Chen, and C. Wen, "Rolling bearing real time fault diagnosis using convolutional neural network," in

2019 Chinese Control And Decision Conference (CCDC), vol. 12, Nanchang, China, 2019.

[15] Y. Yao, S. Zhang, S. Yang, and G. Gui, "Learning attention representation with a multi-scale CNN for gear fault diagnosis under different working conditions," *Sensors*, vol. 20, no. 4, p. 1233, 2020.

[16] H. Fan, C. Xue, X. Zhang, X. Cao, S. Gao, and S. Shao, "Vibration images-driven fault diagnosis based on CNN and transfer learning of rolling bearing under strong noise," *Shock and Vibration*, vol. 2021, no. 2, pp. 1–16, 2021.

[17] K. Zhang, B. Tang, L. Deng, and X. Liu, "A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox," *Measurement*, vol. 179, no. 10, article 109491, 2021.

[18] C. Wang, Y. Xie, and D. Zhang, "Deep learning for bearing fault diagnosis under different working loads and non-fault location point," *Journal of Low Frequency Noise Vibration and Active Control*, vol. 40, no. 1, 2019.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, USA, 2016.

[20] W. Hao, Q. Zhang, M. Shang, and Y. Gu, "Extreme learning Machine-based classifier for fault diagnosis of rotating Machinery using a residual network and continuous wavelet transform," *Measurement*, vol. 183, article 109864, 2021.

[21] X. Wang, C. Shen, M. Xia, D. Wang, J. Zhu, and Z. Zhu, "Multi-scale deep intra-class transfer learning for bearing fault diagnosis," *Reliability Engineering, System Safety*, vol. 202, p. 107050, 2020.

[22] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Computing and Applications*, vol. 32, no. 10, pp. 6111–6124, 2020.

[23] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: a survey," *Knowledge-Based Systems*, vol. 80, pp. 14–23, 2015.

[24] F. Zhuang, Z. Qi, K. Duan et al., "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[25] A. Mohammed and B. Carlos, "Domain adversarial for acoustic emotion recognition," *IEEE/ACM Transactions on Audio Speech & Language Processing*, vol. 26, no. 12, pp. 2423–2435, 2018.

[26] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, "Deep learning-based adversarial multi-classifier optimization for cross-domain machinery fault diagnostics," *Journal of Manufacturing Systems*, vol. 55, pp. 334–347, 2020.

[27] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2296–2305, 2017.

[28] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.

[29] Y. Ganin, E. Ustinova, H. Ajakan et al., "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 1, p. 2096, 2017.

[30] Y. Li, Y. Song, L. Jia, S. Gao, Q. Li, and M. Qiu, "Intelligent fault diagnosis by fusing domain adversarial training and maximum mean discrepancy via ensemble learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2833–2841, 2021.

[31] B. Yang, Y. Lei, F. Jia, N. Li, and Z. Du, "A polynomial kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 11, pp. 9747–9757, 2020.

[32] C. Che, H. Wang, X. Ni, and Q. Fu, "Domain adaptive deep belief network for rolling bearing fault diagnosis," *Computers & Industrial Engineering*, vol. 143, p. 106427, 2020.

[33] Y. Tang, B. Wu, L. Peng, and C. Liu, "Semi-supervised transfer learning for convolutional neural network based Chinese character recognition," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, 2018.

[34] W. Mao, Y. Liu, L. Ding, A. Safian, and X. Liang, "A new structured domain adversarial neural network for transfer fault diagnosis of rolling bearings under different working conditions," *IEEE Transactions on Instrumentation and Measurement.*, vol. 70, pp. 1–3, 2020.

[35] G. Cai, Y. Wang, L. He, and M. Zhou, "Unsupervised domain adaptation with adversarial residual transform networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 8, pp. 3073–3086, 2019.

[36] W. Li, Z. Chen, and G. He, "A novel weighted adversarial transfer network for partial domain fault diagnosis of machinery," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1753–1762, 2020.

[37] J. Dybala and R. Zimroz, "Rolling bearing diagnosing method based on Empirical Mode Decomposition of machine vibration signal," *Applied Acoustics*, vol. 77, pp. 195–203, 2014.

[38] W. U. Zhaohua and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.

[39] A. Gretton, D. Sejdinovic, H. Strathmann et al., "Optimal kernel choice for large-scale two-sample tests," in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, California, USA, 2012.