

Research Article

A New Transferable Fault Diagnosis Approach of Rotating Machinery Based on Deep Autoencoder and Dominant Features Selection under Different Operating Conditions

Fei Dong ¹, Xiao Yu ^{2,3}, Xinguo Shi ⁴, Ke Liu ⁴, Zhaoli Wu ⁵, and Wanli Yu ⁶

¹School of Internet, Anhui University, Hefei 230039, China

²IOT Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221000, China

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

⁴Center of Information Technology, Zibo Mining Group Co, Ltd., Shandong, Zibo 255000, China

⁵Jiangsu Collaborative Innovation Center for Building Energy Saving and Construction Technology, Xuzhou 221116, China

⁶Institute of Electrodynamics and Microelectronics, University of Bremen, Bremen 28359, Germany

Correspondence should be addressed to Xiao Yu; yxcumt2006@163.com

Received 29 July 2021; Accepted 15 September 2021; Published 30 October 2021

Academic Editor: Jun Zhu

Copyright © 2021 Fei Dong 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.

In the actual industrial scenarios, most existing fault diagnosis approaches are faced with two challenges, insufficient labeled training data and distribution divergences between training and testing datasets. For the above issues, a new transferable fault diagnosis approach of rotating machinery based on deep autoencoder and dominant features selection is proposed in this article. First, maximal overlap discrete wavelet packet transform is applied for signals processing and mix-domains statistical feature extraction. Second, dominant features selection by importance score and differences between domains is proposed to select dominant features with high fault-discriminative ability and domain invariance. Then, selected dominant features are used for pretraining deep autoencoder (source model), which helps in enhancing the fault representative ability of deep features. The parameters of the source model are transferred to the target model, and normal state features from target domain are adopted for fine-tuning the target model. Finally, the target model is applied for fault patterns classification. Motor and bearing fault datasets are used for a series of experiments, and the results verify that the proposed methods have better cross-domain diagnosis performance than comparative models.

1. Introduction

With the prompt progress of modern industry, rotating machinery (RM) is developing towards integration and complexity [1]. RM usually operates under complex and harsh scenes, such as variable heavy loads, high temperature and speed, and strong impact [1]. Once a component fault occurs, it may further lead to the damage of other components and huge economic loss. Therefore, it is meaningful and practical to study intelligent fault diagnosis models towards real industrial scenes [1, 2].

In the last several years, intelligent fault diagnosis field has received many studies of traditional-machine-learning-(TML-) based framework, deep-learning (DL-) based

framework, and transfer-learning- (TL-) based framework, which can achieve automatically fault recognition and classification by analysing massive signals collected from mechanical equipment [1–4]. TML-based framework is often constructed by using traditional machine learning algorithms that mainly include k-nearest neighbour (KNN) [5], support vector machine (SVM) [6], artificial neural network (ANN) [7], extreme learning machine (ELM) [8], decision tree (DT) [9], and some variations of them. Generally, TML-based framework consists of three steps: signal process and features extraction, features selection or reduction, and fault classification [2, 4, 8]. Vibration signals are the most commonly used for fault diagnosis, due to the strong nonlinearity and nonstationarity. Time-frequency analysis method is widely

applied for signal process and feature extraction, such as empirical mode decomposition (EMD) [10–12], short-time Fourier transform (STFT) [13–15], wavelet packet transform (WPT) [16–19], and some variations of them. References [20–25] and [26–28] applied the variations or improvement of EMD, STFT, and WPT for fault signals processing and feature extraction. The above-mentioned time-frequency analysis method can effectively help to extract fault features, but it often leads to a high dimensional feature set which contains interference and redundancy features. Thus, feature reduction and selection are a crucial step before the fault patterns classification [3, 6, 10, 18]. In [3], the extreme gradient promotion is used for the dimensional reduction and sensitive features selection, which applies the importance of features to refine a high quality feature subset. In [6], an ant colony algorithm is applied to select features, and the selected feature subset is combined with parameter optimized SVM for enhancing the generalization of the fault diagnosis model. In [10], indexes of the cohesion and class discriminative of features are used for evaluating features, and through the combination of these two indexes, a new index, ASR (the ratio of the adjusted rand index and standard deviation), was proposed to refine the original feature set. In [18], a sensitive feature selection method and modified features dimensionality reduction method are combined to obtain a low-dimensional features subspace, which improves the diagnosis accuracy. In TML-based framework, the traditional KNN, SVM, and ANN are widely used for constructing fault classification models by researchers. For example, references [29–31], [6, 10, 18], and [7, 32, 33] applied the KNN, SVM, and ANN for fault classification, respectively. Moreover, many variations of KNN, SVM, and ANN have been studied and applied for rotating machinery fault diagnosis. In [34], an enhanced KNN (EKNN) was designed to get embedded in a dimension-reduction stage; then, by using sparse filtering, some fault-discriminative features can be extracted. In [35], in order to determine the parameter K of KNN, an improved binary particle swarm optimization was proposed to select this parameter, which construct the IBPSO-KNN for bearing fault diagnosis. In [36], due to the fact that it is difficult for the traditional squares SVM to deal with complex imbalanced data, an improved SVM, a moth-flame optimization-based LS-SVM, was proposed for bearing fault diagnosis with complex imbalanced data. In [37], SVM was optimized by intercluster distance in the feature space, which was combined with improved symplectic geometry mode decomposition to design a novel fault diagnosis scheme for rotating machinery. In [38], a perceptron multilayer ANN (MLP-ANN) is used for detecting the bearing faults. However, the main limitation of TML-based framework relies heavily on expert knowledge when the diagnosis models need to be customized for different operating states and machines [4, 39].

Deep learning algorithms have received more and more attentions because they possess a powerful hidden features automatic mining ability [4, 40]. Therefore, DL-based framework has been widely studied in the intelligent fault diagnosis field. In references [13, 19, 39, 41], convolutional neural network (CNN) [13], deep belief network (DBN) [19], deep neural networks (DNN) [39], and deep autoencoders

(DAE) [41] are used for constructing fault diagnosis models, respectively. However, some main limitations exist in most DL-based frameworks [3, 4, 40, 42, 43]. (1) Most conventional DL-based frameworks have insufficient generalization ability towards the engineering practical scenes; the reason is that an assumption of having the same distribution between the training and testing sets was widely used. In practical scenes, the real fault signals collected from machineries are inconsistent under variable operating states, which brings the distribution divergence of datasets. (2) They rely heavily on a mass of labeled data. When the labeled training data are not enough, the overfitting phenomenon may easily appear, which can lead to the reduction of diagnosis accuracy and stability. Facing at the engineering practical scenes, the sufficient labeled data are difficult to obtain due to the changeable and complex working conditions of machineries. Thus, how to enhance the fault representative ability of deep features and the stability of fault diagnosis models across different working conditions is still a challenging task.

Aiming at the limitations of DL-based frameworks mentioned above, a recent developed technique, domain adaptation under TL-based framework, intends to promote classifier learning by using labeled source domain data. At present, TL-based framework has become a study hot topic and employed for fault diagnosis of machinery [2–4, 20, 39, 42]. In the fault diagnosis, one working condition (a specific speed or load) can compose a domain. Source domain is a labeled dataset under one working condition, and the target domain is unlabeled data under another working condition. The object of domain adaptation under TL-based framework is that labeled source domain and unlabeled target domain are used to learn a cross-domain diagnosis model which can achieve desirable fault classification results of the target domain [4, 42]. Considering that deep learning methods have powerful ability to mine hidden features from original data, recently, deep TL models have been researched for cross-domain fault diagnosis of rotating machinery by many researchers. In [44], an enhanced DAE was designed by modifying the loss function, which improves the reconstruction performance of a decoder. Sufficient labeled data from source domain were employed for training the enhanced DAE model, and the corresponding parameters were transferred to target DAE model. In [45], a deep CNN with an attention mechanism was adopted for feature extraction, and a domain transformation algorithm was designed to match the distributions between source and target domains. In [46], a novel DAE model, deep transfer multiwavelet AEs, was designed for gearbox fault diagnosis by using little training samples. In this model, important features were learned by very few samples, and parameters of source model are directly migrated to target model. Although deep TL models have achieved many successful applications on cross-domain fault diagnosis of rotating machinery, how to enhance the fault representative ability of deep features and the stability of fault diagnosis models across different working conditions is still a challenging task [43]. For this issue, in this article, we propose a new transferable fault diagnosis approach of rotating machinery based on DAE and dominant features selection under different operating conditions (TFDD). In TFDD, the first step

is vibration signals processing and statistical features extraction, maximal overlap discrete wavelet packet transform (MODWPT) is used to decompose raw signals, and MODWPT is a time-frequency analysis method based on wavelet. Its advantages include two aspects: (1) it can overcome the limitation of discrete wavelet transform (DWT); that is, DWT requires the sample size to be exactly a power of 2 for the full transform because of the downsampling step; (2) MODWPT can overcome another problem that the DWT has very poor frequency resolution at low frequencies. Considering the above advantages of MODWPT, in our previous study [18, 19], MODWPT has been used for bearing fault diagnosis and compared with WPT; the performance of MODWPT is better than WPT. The second step is dominant features selection; a new feature selection method, dominant feature selection by importance score and domain differences (DSID), is proposed to evaluate the fault-discriminative ability and domain invariance of feature, which can help in enhancing the fault representative ability of deep hidden features obtained by DAE. The third step is to construct deep transfer autoencoder (DTAE) model. A DAE model (source model) is trained by feature data from source domain and the learned parameters transferred to the target model that has the same architecture as source model. Then, the normal state feature data from target domain are applied for fine-tuning the target model. The last step is that the learned DTAE model is applied to diagnose the unlabeled fault features from target domain and output fault identification accuracies. The main contributions of this article are organized as follows.

- (1) A new dominant feature selection method DSID: firstly, based on the raw feature dataset, the sufficient labeled feature data from source domain under all fault states is used to evaluate the fault-discriminative ability of features by random forest, and the importance score of features can be obtained to quantify the fault-discriminative ability. Secondly, the normal state feature data from the source and target domains are used to evaluate the domain invariance of features by computing the maximum mean discrepancy. Finally, the proposed new dominant features selection index, RIM, is constructed.
- (2) A DTAE model is learned by dominant features. For enhancing the fault representative ability of deep features and the stability of fault diagnosis models, we apply the DSID to select dominant features with high fault-discriminative ability and domain invariance to train a DTAE model, which is expected to enhance the fault representative ability of deep features.
- (3) A series of experiments are performed by using a motor and bearing faults datasets sampled from SQI-MFS test platform. The experimental results prove the availability, flexibility, and advantages of the TFDD.

The remaining contents of this article are organized as follows. Section 2 discusses the introduction of preliminary knowledge. Section 3 presents the proposed DSID and fault diagnosis framework TFDD. Experimental verification is given in Section 4. Section 5 concludes this paper.

2. Preliminaries

2.1. Deep Autoencoder (DAE). DAE is an unsupervised deep neural network [44], which is constructed by stacking several basic autoencoders (AE). Each AE has two steps: encoder and decoder; the structure of AE is presented in Figure 1. There are three layers: input layer $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$, hidden layer $\mathbf{h} = \{h_1, h_2, \dots, h_n\}$, and output layer $\mathbf{z} = \{z_1, z_2, \dots, z_m\}$.

In the step of encoder, the input data \mathbf{x} is mapped into the data of hidden layer \mathbf{h} by the activation function $\varphi_{\text{act}}(\bullet)$; the mapping process is shown as the following expression:

$$\mathbf{h} = \varphi_{\text{act}}(\mathbf{W} \cdot \mathbf{x} + \mathbf{b}), \quad (1)$$

where W and b are weight matrix and bias vector of encoder, respectively.

In the step of decoder, the data of hidden layer h are mapped into the output data z by the activation function $\varphi_{\text{act}}(\bullet)$; the mapping process is presented as the following expression:

$$\mathbf{z} = \varphi_{\text{act}}(\mathbf{W}' \cdot \mathbf{h} + \mathbf{b}'), \quad (2)$$

where \mathbf{W}' and \mathbf{b}' are weight matrix and bias vector of decoder, respectively. The hidden layer is the new feature representation and the output data are the reconstruction of the input data. The parameters of an AE model are learned by minimizing reconstruction error between the input and output layers; the reconstruction error is expressed as follows:

$$J_{\text{AE}}(\theta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \|x_i - z_i\|^2, \quad (3)$$

where N is the number of samples and $\theta = \{\mathbf{W}, \mathbf{b}, \mathbf{W}', \mathbf{b}'\}$.

The hidden layer features of an AE are used as the input of the next AE, which can stack multiple AEs to construct a DAE. The structure of a DAE is presented in Figure 2. DAE has the strong ability to mine deep features of input feature data so that it can improve accuracy of fault classification [41].

2.2. Random Forest (RF)-Based Feature Selection.

Random forest (RF), firstly proposed by Breiman [47], is one of ensemble classifiers that obtained wide attentions by researchers. RF can achieve desirable performance for classification and regression tasks with high dimensional and ill-posed feature dataset [48]. The mainly idea of RF is to construct some unbiased decision trees (DT) by using the randomly selected samples, where each tree votes for a class and the forest chooses the classification having the most votes over all the trees [47, 48].

Given a dataset $\mathbf{S} = \{(\mathbf{X}, \mathbf{Y}), \mathbf{X} = \{x_i\} \in \mathbf{R}^D, \mathbf{Y} = \{y_i\} \in \{1, 2, \dots, C\}\}_{i=1}^N$, where x_i is a feature sample with D dimension and y_i represents the class label, C and N are respectively the number of classes and training samples. RF algorithm is usually described as follows [49].

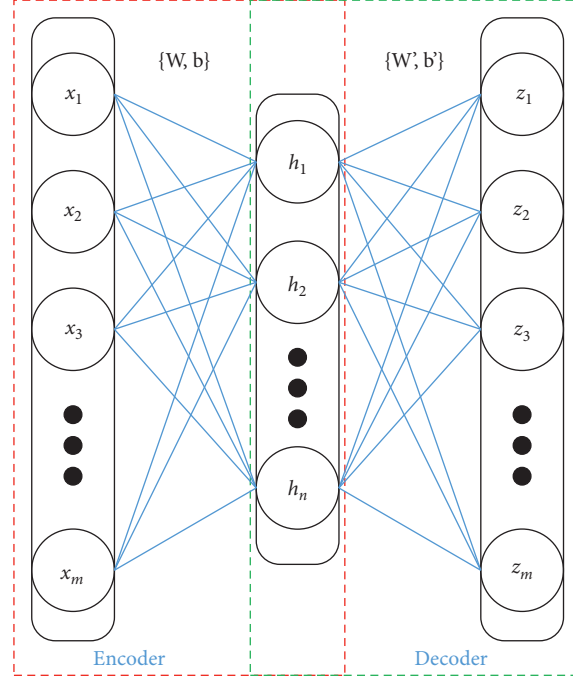


FIGURE 1: The structure of an AE.

- (1) The M (the number of decision trees) bootstrap datasets are drawn from the training sample \mathbf{S} by using bagging [50].
- (2) For each bootstrap datasets, a decision tree is constructed by employing the Classification and Regression Tree (CART) algorithm [51]. At each node of DT, a subspace including p features is sampled and split points based on this subspace are computed. Then, the best split, for example, by the maximum Gini parameters impurity decrease, is applied to segment the data and grow the tree. That is, all data are pure with regard to the class.
- (3) M DT are combined into a RF ensemble, and a majority vote manner is used to make the classification decision.

In RF algorithm, the performance and the diversity of DT can affect the performance of RF. A generalization error of RF is defined as

$$\text{error} = \frac{\bar{\rho}(1 - s^2)}{s^2}, \quad (4)$$

where $\bar{\rho}$ and s^2 represent the average correlation between DT and the average strength of DT, respectively. In addition, one of the important properties of RF is that the importance score (IS) of feature can be measured. For the high dimensional feature data, IS can be used to select relevant, compact, and discriminative features, which can help to improve the performance of classification. The Gini index (GI) is used to construct DT and determine the class in each tree [47, 49]. The GI at node t , $\text{GI}(t)$, is applied to quantify the impurity of node t ; the expression is defined as follows:

$$\text{GI}(t) = \sum_{i=1}^I \text{frac}_i (1 - \text{frac}_i), \quad (5)$$

where frac_i represents the fraction of category and i records at node t . Based on the GI, the GI information gain (GIG) of feature x_i , which is used to separate node t , is expressed as follows:

$$\text{GIG}(x_i, t) = \text{GI}(t) - (W_L \text{GI}(t^L) + W_R \text{GI}(t^R)), \quad (6)$$

where t^L and t^R are respectively the left and right child nodes of node t and W_L and W_R are respectively the corresponding fraction. Moreover, the IS of feature x_i can be obtained by calculating the following expression:

$$\text{IS}(x_i) = \frac{1}{n_{\text{DT}}} \sum_{k \in S_{x_i}} \text{GIG}(x_i, t), \quad (7)$$

where n_{DT} represents the number of DT in RF and $k \in S_{x_i}$ represents the set of split nodes. Finally, in the original high dimensional feature set, according to the IS of each feature, the features with high IS value can be selected to construct feature subset, so that many features with small IS are eliminated and improve the performance of classification.

2.3. Maximum Mean Discrepancy (MMD). Given two feature datasets $D_S = \{x_1, x_2, \dots, x_{n_S}\}$ (source domain) and $D_T = \{y_1, y_2, \dots, y_{n_T}\}$ (target domain) drawn from two different probability distributions, $P_S(D_S) \neq P_T(D_T)$, where n_S and n_T are respectively the number of D_S and D_T . For the purpose of estimating the distance between two distributions, MMD [52] was introduced by Gretton et al. for measuring distance of distributions based on reproducing

kernel Hilbert space (RKHS). The empirical distance estimate of distributions between D_S and D_T is defined as the following expression [53]:

$$\text{Dist}(D_S, D_T) = \left\| \frac{1}{n_S} \sum_{x_i \in D_S} \phi(x_i) - \frac{1}{n_T} \sum_{x_j \in D_T} \phi(x_j) \right\|_H^2, \quad (8)$$

where $\|\bullet\|_H$ represents the RKHS norm and ϕ is the kernel-induced feature map. For the issue that the inconsistent feature distribution exists in fault diagnosis across variable operating conditions, based on the above-mentioned description, the MMD can be applied to estimate the discrepancy of two distributions and align these two distributions.

3. Proposed Method and System Framework

3.1. Dominant Feature Selection by Importance Score and Domain Differences (DSID). In order to reduce redundant features in the high dimensional raw feature set (HRFS) and select dominant features (fault-discriminative but operating-condition-invariant (FDOCI) features), we suppose that features should be evaluated from two aspects: fault-discriminative ability and domain invariance. Therefore, a new feature selection approach, dominant feature selection by importance score and domain differences (DSID), is proposed in this article. In DSID, firstly, the RF is employed to quantify the fault-discriminative ability of each feature based on labeled source domain data. Secondly, the MMD is used to evaluate the domain invariance of each feature based on normal state data of source and target domain data. Finally, a new selection index, the ratio of IS and MMD (RIM), is constructed to select dominant features for enhancing the performance of fault diagnosis across different operating conditions. The specific description of the DSID is summarized as follows.

- (a) *Compute Importance Score of Features.* Given a raw feature set (RFS) $[f_1, f_2, \dots, f_p]^T$ of source domain that contains p feature samples, each sample has q features; that is, $f_i = \{f_i^1, f_i^2, \dots, f_i^q\}$, $i \in [1, p]$. Let $IS(k)$ denote the importance score of the k -th feature, according to the introduction of RF-based feature selection in Section 2.2. $IS(k)$ can be calculated by (5)–(7); thus, the sequence $IS = \{IS(1), IS(2), \dots, IS(q)\}$ of q features can be obtained. In this paper, we suppose that the fault-discriminative ability of feature is greater when the value of IS is larger.
- (b) *Evaluate the Domain Invariance of Features.* MMD is employed to estimate the distribution discrepancy of the same feature in different domains, and the value of MMD is used as the quantitative index of domain invariance of feature. Let nf_S and nf_T denote normal state feature data from source and target domains, respectively. Both nf_S and nf_T consist of p feature samples, and each sample has q features. The expressions of nf_S and nf_T are presented as follows:

$$nf_S = \begin{bmatrix} nf_S^{11} & nf_S^{12} & \cdots & nf_S^{1q} \\ nf_S^{21} & nf_S^{22} & \cdots & nf_S^{2q} \\ \vdots & \vdots & \ddots & \vdots \\ nf_S^{p1} & nf_S^{p2} & \cdots & nf_S^{pq} \end{bmatrix}, \quad (9)$$

$$nf_T = \begin{bmatrix} nf_T^{11} & nf_T^{12} & \cdots & nf_T^{1q} \\ nf_T^{21} & nf_T^{22} & \cdots & nf_T^{2q} \\ \vdots & \vdots & \ddots & \vdots \\ nf_T^{p1} & nf_T^{p2} & \cdots & nf_T^{pq} \end{bmatrix},$$

where the j -th column elements of nf_S and nf_T represent the p samples of the j -th feature from source and target domains ($j \in [1, 2, \dots, q]$), respectively. The expression of them is presented as

$$nf_S^j = [nf_S^{1n}, nf_S^{2n}, \dots, nf_S^{pn}]^T, \quad (10)$$

$$nf_T^j = [nf_T^{1n}, nf_T^{2n}, \dots, nf_T^{pn}]^T.$$

The MMD between nf_S^j and nf_T^j can be computed by (8). Therefore, a MMD sequence of q features can be further obtained, $\text{MMD} = \{\text{MMD}(1), \text{MMD}(2), \dots, \text{MMD}(q)\}$. In this article, we suppose that the domain invariance of feature is greater when the value of MMD is smaller.

- (c) *Construct the Selection Index RIM.* Based on the IS and MMD of features obtained in the previous two steps, a new selection index, RIM, is constructed for selecting dominant features from RFS. The RIM of the j -th feature is defined as follows:

$$\text{RIM}(j) = \frac{\text{IS}(j)}{\text{MMD}(j)}. \quad (11)$$

Thus, for q features, the corresponding RIM values construct a RIM sequence, $\text{RIM} = \{\text{RIM}(1), \text{RIM}(2), \dots, \text{RIM}(q)\}$. We suppose that the feature with higher value of RIM is more beneficial to cross-domain fault diagnosis, because the feature has great fault-discriminative ability and domain invariance at the same time. Finally, we can select the feature with higher value of RIM from the sorted RIM sequence that is sorted in descending mode to perform cross-domain fault diagnosis model training.

3.2. Transferable Fault Diagnosis Framework Based on DSID and DAE (TFDD)

3.2.1. The Mechanism of Deep Transfer AE Model. The structure of the deep transfer AE (DTAE) model is presented in Figure 3. In model, 1 input layer, 3 hidden layers, and 1 softmax layer are designed. The softmax layer is used to classify deep feature representations. The construction steps of a DTAE are shown in Figure 3 and stated as follows. There are 4 steps:

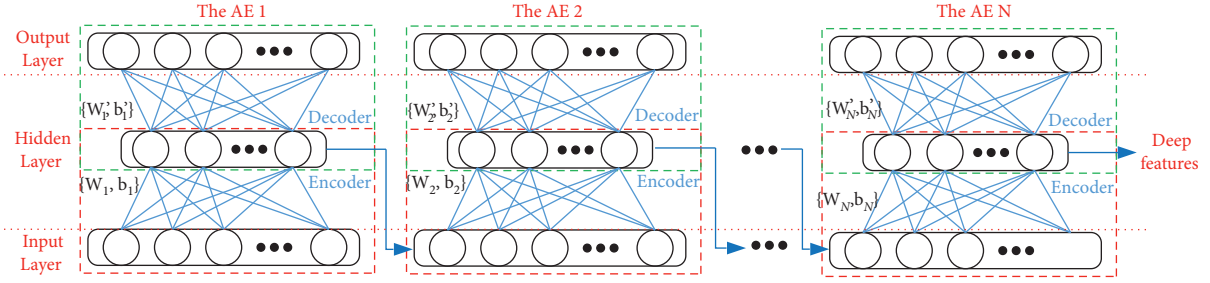


FIGURE 2: The structure of DAE.

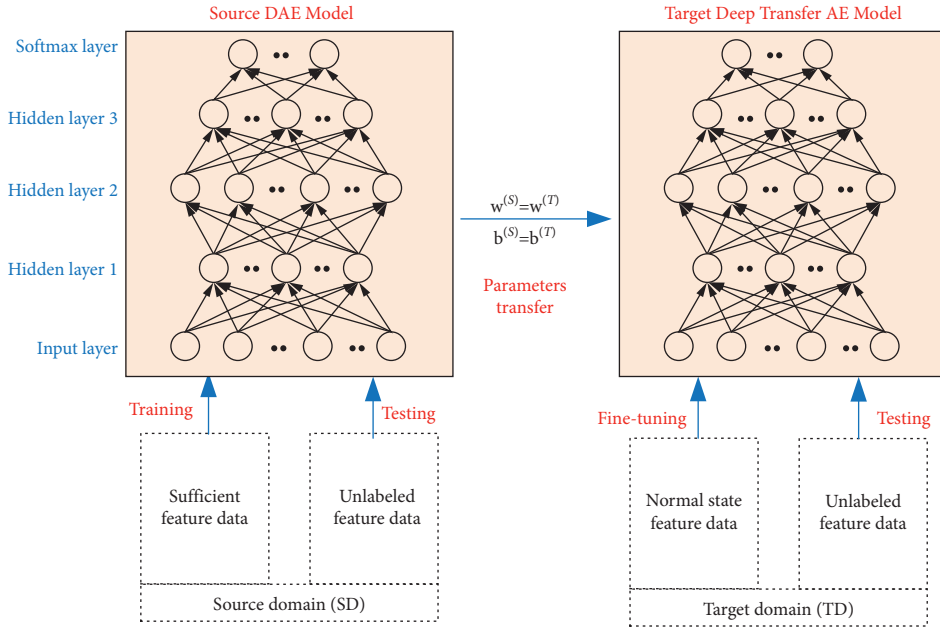


FIGURE 3: The mechanism of the DTAE model.

- (1) Train a DAE model by using sufficient feature data from SD; this is source model. The parameters weight matrix $W^{(S)}$ and bias vector $b^{(S)}$ can be obtained.
- (2) Construct another DAE model (target model), which has the same architecture as source model; that is, the number of layers and modes are the same as source model.
- (3) Parameters transfer: the parameters weight matrix $W^{(S)}$ and bias vector $b^{(S)}$ learned from the procedure of source model training are transferred to target model; that is, $W^{(S)} = W^{(T)}$ and $b^{(S)} = b^{(T)}$.
- (4) Fine-tuning target model: the normal state feature data from TD are employed to fine-tune the target DAE model. Finally, the fine-tuned target model is used to test remaining unlabeled target domain data.

3.2.2. Step Description of the Proposed Framework. In this article, TFDD, a novel transferable fault diagnosis framework based on DSID and DAE, is proposed for cross-domain fault diagnosis of rotating machinery. The framework

TFDD is given in Figure 4, and specific descriptions are organized as the following four steps.

Step 1: Signal Process and Feature Extraction. The original vibration signals sampled from rotating machinery by acceleration sensors under operating conditions 1 and 2 are respectively the source and target domains data. Then, the signal process and statistical features extraction are performed by using MODWPT and calculating statistical parameters. The mixed domains statistical characteristics are generated to construct a raw feature set (RFS).

Step 2: Dominant Features Selection. Firstly, based on the RFS obtained in Step 1, the sufficient labeled feature data from source domain under all fault states are used to evaluate the fault-discriminative ability of features by RF, and the IS of features can be obtained to quantify the fault-discriminative ability. Secondly, the normal state feature data from two domains are used to evaluate the domain invariance of features by MMD. Finally, the proposed new dominant features selection index, RIM, is constructed. The features with high value of RIM can construct feature

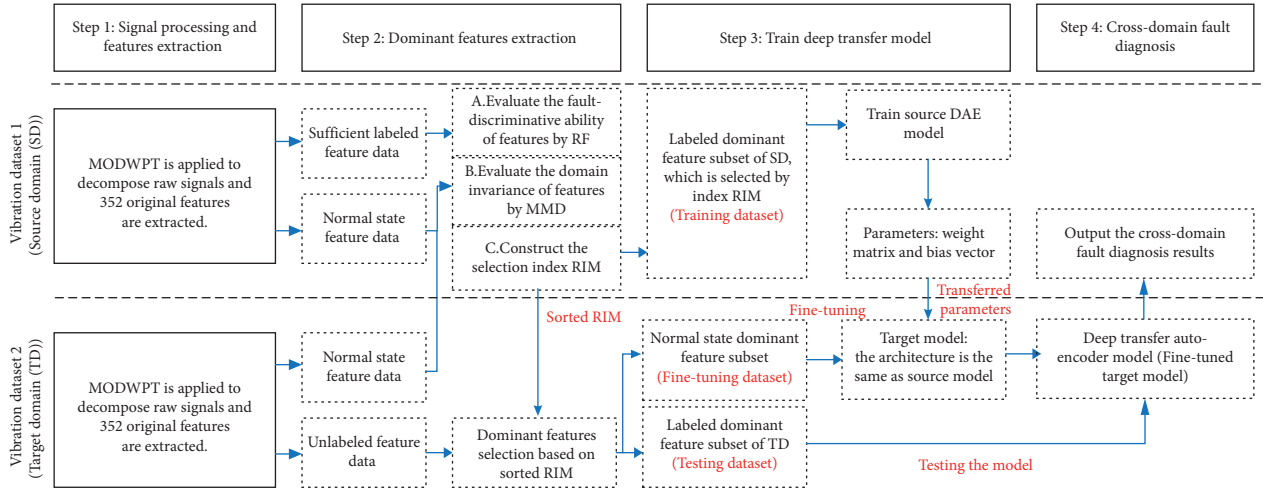


FIGURE 4: The proposed transferable fault diagnosis framework for rotating machinery across different operating conditions.

subset that is beneficial to cross-domain fault diagnosis. Thus, the sorted RIM sequence in descending mode is used for dominant features selection.

Step 3: Construct Deep Transfer Autoencoder Model. Firstly, a DAE model (source model) is trained by feature data from source domain and the parameters weight matrix $W^{(S)}$ and bias vector $b^{(S)}$ are obtained. Secondly, the parameters $W^{(S)}$ and $b^{(S)}$ are transferred to target the model that has the same architecture as the source model. Thirdly, the normal state feature data from target domain is applied to fine-tune the target model. Finally, the construction of DTAE model is completed.

Step 4: Output the Fault Diagnosis Results. Based on the learned DTAE model, the unlabeled feature data from target domain are used to test the performance of DTAE model and output diagnosis results.

4. Experimental Verification

In this article, motor and bearing fault datasets obtained from the SQI-MFS test platform [10, 18, 20, 54] are employed for experimental verification. The test platform is shown in Figure 5, and fault bearings and motors are presented in Figures 6 and 7. The vibration signals are sampled by acquisition cards and acceleration sensors installed at the drive end and fan end of the motor, and the sampling frequency is 16 kHz. Aiming at proving the availability and flexibility of the proposed transferable fault diagnosis framework across variable operating conditions, we collected faulty motor and bearing vibration data under different operating speeds, the experimental verification of two cases is carried out, and the detailed description is as follows.

4.1. Case 1: Transfer Diagnosis of Fault Motors under Different Operating Speeds

4.1.1. Introduction of Motor Dataset and Tasks. In this section, motor vibration data under two speeds of 1730 rpm

and 1750 rpm are used for experimental verification. The main parameters of motor are shown in Table 1. Four faulty motors, including broken rotor bar fault (BF), winding fault (WF), rotor bowed fault (RF), and single phase voltage unbalance fault (SF), and a normal state motor (NM) are used in experiments. Thus, there are 5 motor conditions that correspond to 5 patterns. For each pattern, 30 and 60 vibration data samples are respectively random selected as the training and testing data. Each sample contains 5000 continuous sampling points. More specific introduction of motor dataset is presented in Table 2. Based on the vibration data under speeds of 1730 rpm and 1750 rpm, we set up 2 cross-domain fault diagnosis tasks, as shown in Table 3. According to the details in Table 3, the vibration data under speeds of 1730 rpm and 1750 rpm are respectively chosen as the source datasets of tasks 1 and 2. The vibration data under speeds of 1750 rpm and 1730 rpm are respectively used as the target datasets of tasks 1 and 2. Source and target domains contain 150 and 300 samples, respectively.

4.1.2. Transfer Diagnosis Results of the Proposed TFDD Framework. According to the steps of the proposed framework TFDD, firstly, the raw vibration signals are processed by MODWPT, and statistical features are generated by calculating statistical parameters of single branch reconstruction signals of wavelet packet nodes. In this article, we apply the “dmey” as the mother wavelet in MODWPT, and the layer of wavelet decomposition is set to 4. Therefore, 16 terminal wavelet packet nodes (TWPN) are generated and the corresponding reconstruction signals (RS) are used for calculating 11 statistical parameters; thus, 176 time-domain statistical features are generated. Moreover, The Hilbert envelope spectra (HES) of 16 reconstruction signals are also used for generating 176 frequency-domain statistical features by 11 statistical parameters. These 11 statistical parameters are range, mean value, standard deviation, kurtosis, energy, energy entropy, skewness, crest factor, impulse factor, shape factor, and latitude factor, respectively [10, 18, 20, 29, 54, 55]. Therefore, 352 statistical



FIGURE 5: SQI-MFS test platform.



FIGURE 6: SER205 fault bearings.

characteristics are generated from a vibration sample to construct a raw feature set (RFS). The sampled vibration signals of 5 motor conditions under rotating speeds of 1730 rpm and 1750 rpm are shown in Figure 8, and the RS of TWPN that are obtained by decomposing normal state vibration signals are shown in Figure 9. Moreover, 352 statistical features extracted from NM and BF vibration signals under 1730 rpm and 1750 rpm are shown in Figure 10; these features have been normalized. From Figure 8, it is obvious that the distribution discrepancy existed between vibration signals from different operating speeds.

Based on the RFS obtained from signal process and features extraction, the proposed dominant feature selection method DSID is performed to evaluate the fault-discriminative ability and domain invariance of features, and the selection index RIM of each feature can be calculated by the equation (11). In this article, we suppose that the feature with higher value of RIM is more beneficial to cross-domain fault diagnosis. Then, the RIM sequence that includes RIM of 352 statistical features is obtained, the RIM sequence is sorted in descending mode, and the sorted RIM sequence is used to select dominant features for constructing feature subset. The IS, MMD, and RIM of 352 features are respectively shown in Figures 11–13.

According the sorted RIM sequence, some dominant features are chosen to construct feature subset. Then, the

DTAE model training is performed; based on the source domain, the selected dominant features are used for training source DAE model, and the learned parameters $W^{(S)}$ and $b^{(S)}$ are directly transferred to initialize the target DTAE model that has the same architecture as source DAE model. Next, the normal state feature data from target domain are used for fine-tuning the target DTAE model. Finally, the testing data (unlabeled feature data from target domain) are inputted to the DTAE model, and the softmax layer of DTAE model can achieve fault classification for testing data. In this article, some parameters used in DTAE model training are as follows: the number of hidden layers is 4 and the sizes of hidden layers are respectively 400, 100, 50, and 50. The iteration is set to 200.

The experimental results of the proposed TFDD framework are respectively given in Figure 14 and Table 4. From the details of Table 4, when dominant features number is set as 352, that is, all 352 statistical features from source domain are applied for training DATE model, the diagnosis results of tasks 1 and 2 are only 69.00% and 66.67%, respectively. However, when the proposed DSID is performed before training DTAE model, it can significantly improve diagnosis accuracy. The maximum average accuracy of tasks 1 and 2 can attain 81.67% ($dfn:101$) and 82.67% ($dfn:140$), respectively. Figure 14 presents the diagnosis accuracies of tasks 1 and 2 when the dfn is from 40 to 352. We can

conclude that the proposed TFDD framework using DSID can enhance the performance of cross-domain diagnosis when a suitable dfn is selected.

4.1.3. Comparisons with Other Models. In order to further prove the advantages of TFDD framework on cross-domain fault diagnosis, we chose some common and competitive methods for comparison. Based on these methods, some comparative models are constructed, as shown in Table 5. These comparative models can be divided into two categories. (1) The model is not combined with transfer learning method; for example, the model RFS-KNN is a common model that the RFS is directly inputted to the KNN classifier. (2) The model is combined with transfer learning method; for example, RFS-TCA is a transfer learning-based model such that RFS is directly inputted to the TCA and the SVM classifier is applied to classify the fault features. For the RFS-DSID-TCA model, it is based on the RFS-TCA model, and the proposed DSID method is employed to select dominant features from RFS for the subsequent transfer learning.

The experimental results of comparative models are given in Table 6, Figures 15 and 16. According to the details in Table 6, the diagnosis accuracies of comparative models are obviously smaller than the accuracies of TFDD. The transfer learning-based models, RFS-TCA, RFS-JDA, RFS-DSID-TCA, and RFS-DSID-TCA, can achieve better diagnosis performance than other models. When the DSID is embedded in transfer learning-based model, the diagnosis performance can be further enhanced; the diagnosis accuracies of RFS-DSID-TCA and RFS-DSID-JDA models for task 1 are 71.67% and 77.00%, which are 5.67% and 4.67% higher than RFS-TCA and RFS-JDA models, respectively. The diagnosis accuracies of RFS-DSID-TCA and RFS-DSID-JDA models for task 2 are 77% and 78.00%, which are 7.33% and 8.00% higher than RFS-TCA and RFS-JDA models, respectively. However, the models using TCA and JDA do not outperform the proposed TFDD model, and the maximum average accuracy of TFDD is higher than RFS-DSID-TCA and RFS-DSID-JDA models. These comparison results can validate the advantages of the TFDD, which includes two aspects:

- (1) For the cross-domain diagnosis tasks 1 and 2, the proposed TFDD model can effectively classify 5 motor conditions, and the maximum average accuracy can attain over 80%. The proposed dominant features selection method DSID can help in selecting features that have high fault-discriminative and domain invariance, which can significantly improve cross-domain diagnosis performance.
- (2) According to the comparison results, it reveals that the diagnosis performance of TFDD is obviously better than comparative models shown in Table 6. Moreover, the diagnosis model combined transfer learning strategy can help in enhancing diagnosis accuracy across different domains.

4.2. Case 2: Transfer Diagnosis of Fault Bearings under Different Operating Speeds

4.2.1. Introduction of Bearing Dataset and Tasks. In this section, bearing vibration data under two speeds of 1200 rpm and 1600 rpm are used to further prove the availability, flexibility, and advantages of the TFDD. Three kinds of faulty bearings (inner race fault (IRF), outer race fault (ORF), and ball fault (BF)) are manufactured by laser machining, and three kinds of fault diameters (0.05 mm, 0.1 mm, and 0.2 mm) are set for each fault type for experiments. These faulty bearings are given in Figure 6. In addition, a normal bearing is also used for experiments; thus, there are 10 bearing states that correspond to 10 patterns. For each pattern, 30 and 60 vibration data samples are respectively random chosen as the training and testing samples. Each sample contains 5000 sampling points. More details of bearings dataset are presented in Table 7. Based on the vibration data under speeds of 1200 rpm and 1600 rpm, we set up 2 cross-domain fault diagnosis tasks, as shown in Table 8.

4.2.2. Transfer Diagnosis Results of the Proposed TFDD Framework. In this section, the process of experiment is the similar to that of Section 4.1.2. The fault diagnosis results of tasks 1 and 2 obtained by TFDD framework are shown in Table 9 and Figure 17. From the experimental results, it is obvious that the TFDD can effectively diagnose bearing faults across different operating speeds, and the highest diagnosis accuracies of tasks 1 and 2 can respectively reach 90.33% (dfn : 150) and 90.00% (dfn : 152), which are 8.83% and 9% higher than the models without using DSID. From Figure 17, when a suitable dfn is selected according to the sorted RMI sequence, the performance of model can obtain an obvious enhancement. This further proves the availability of the DSID.

4.2.3. Comparisons with Other Models. For the comparative experiments, these are the same as Section 4.2.3. The models used for comparison are shown in Table 5. The experimental results are presented in Table 10, Figures 18 and 19. The diagnosis results obtained by TFDD are significantly better than other models, and the maximum accuracies of tasks 1 and 2 can respectively reach 90.33% and 90%. For task 1, the diagnosis results of RFS-SVM, RFS-KNN, RFS-DAE, RFS-DBN, RFS-CNN, RFS-TCA, RFS-JDA, RFS-DSID-TCA, and RFS-DSID-JDA are 70.83%, 65.17%, 62.83%, 75.67%, 65.00%, 56.83%, 65.50%, 69.50%, and 85.67%, respectively. For task 2, the diagnosis results of RFS-SVM, RFS-KNN, RFS-DAE, RFS-DBN, RFS-CNN, RFS-TCA, RFS-JDA, RFS-DSID-TCA, and RFS-DSID-JDA are 50.83%, 56.33%, 58.33%, 52.67%, 57.50%, 52.17%, 61.33%, 61.83%, and 83.50%, respectively. This further proves the advantages of the TFDD. According to the diagnosis results given in Figures 18 and 19, the transfer learning-based models can obtain an improvement on diagnosis accuracy by combining the DSID; thus, the availability of the DSID is also verified.



FIGURE 7: Motor fault data collection and fault motors.

TABLE 1: Main parameters of motor.

Name	Value
Poles	1
Number of stator slots	34
Number of rotor slots	24
Stator bore	82.1 mm
Rotor bore	80.5 mm
Rated power	370 W

TABLE 2: Details about the motor vibration dataset.

Motor conditions	Sampling frequency (kHz)	Rotating speeds 1730 rmp		Rotating speeds 1750 rmp		Class label
		Size of training samples	Size of testing samples	Size of training samples	Size of testing samples	
NM	16	30	60	30	60	1
BF		30	60	30	60	2
WF		30	60	30	60	3
RF		30	60	30	60	4
SF		30	60	30	60	5

TABLE 3: 2 cross-domain fault diagnosis tasks for 5 motor conditions.

	Rotating speeds (rmp)	Task 1		Task 2		
		Motor conditions	Size of samples	Rotating speeds (rmp)	Motor conditions	Size of samples
Source domain	1730	Classes 1-5	150	1750	Classes 1-5	150
Target domain	1750	Classes 1-5	300	1730	Classes 1-5	300

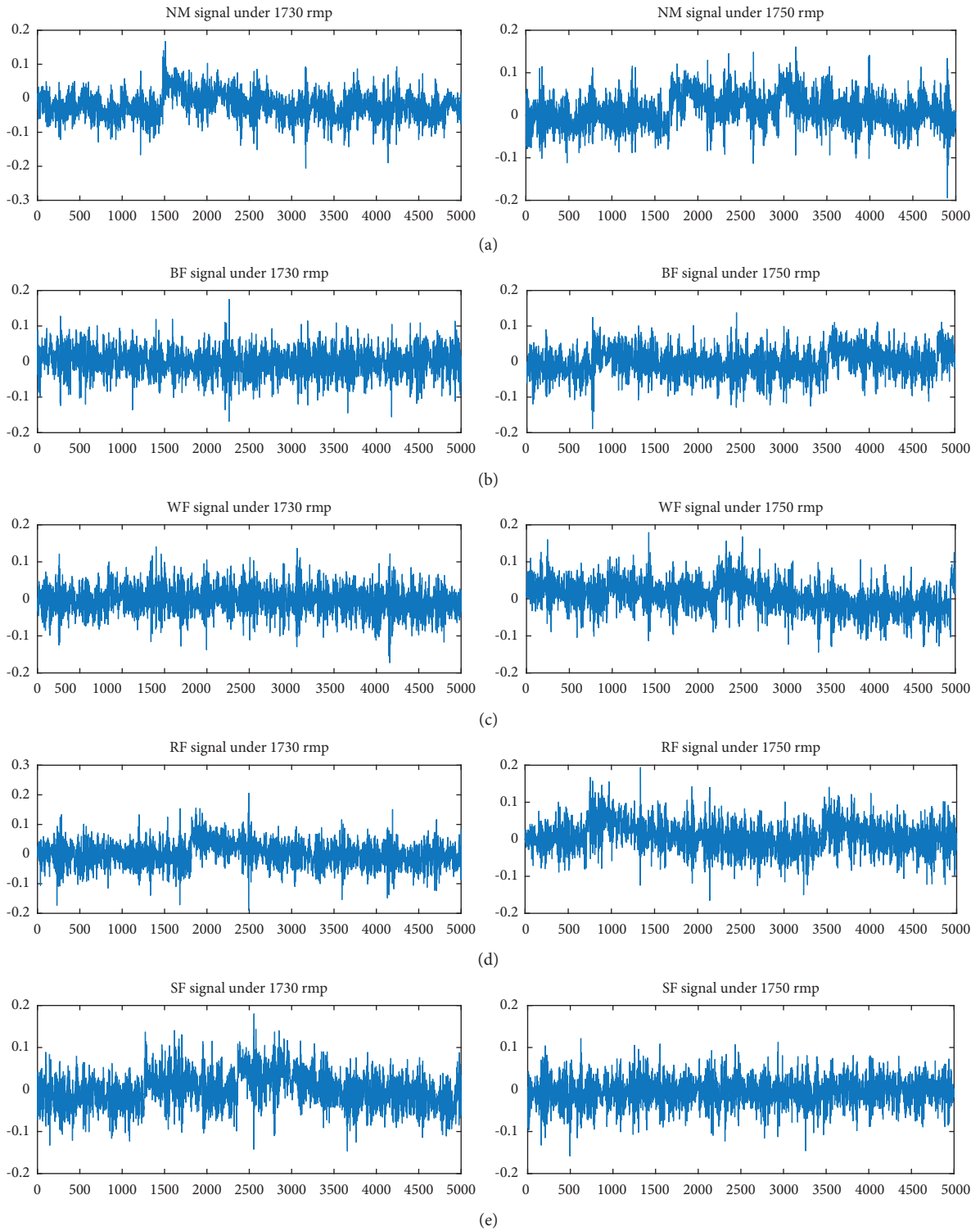


FIGURE 8: Raw signals of the 5 kinds of motor conditions. (a) Raw signal under normal state. (b) Raw signal under broken rotor bar fault. (c) Raw signal under winding fault. (d) Raw signal under rotor bowed fault. (e) Raw signal under single phase voltage unbalance fault.

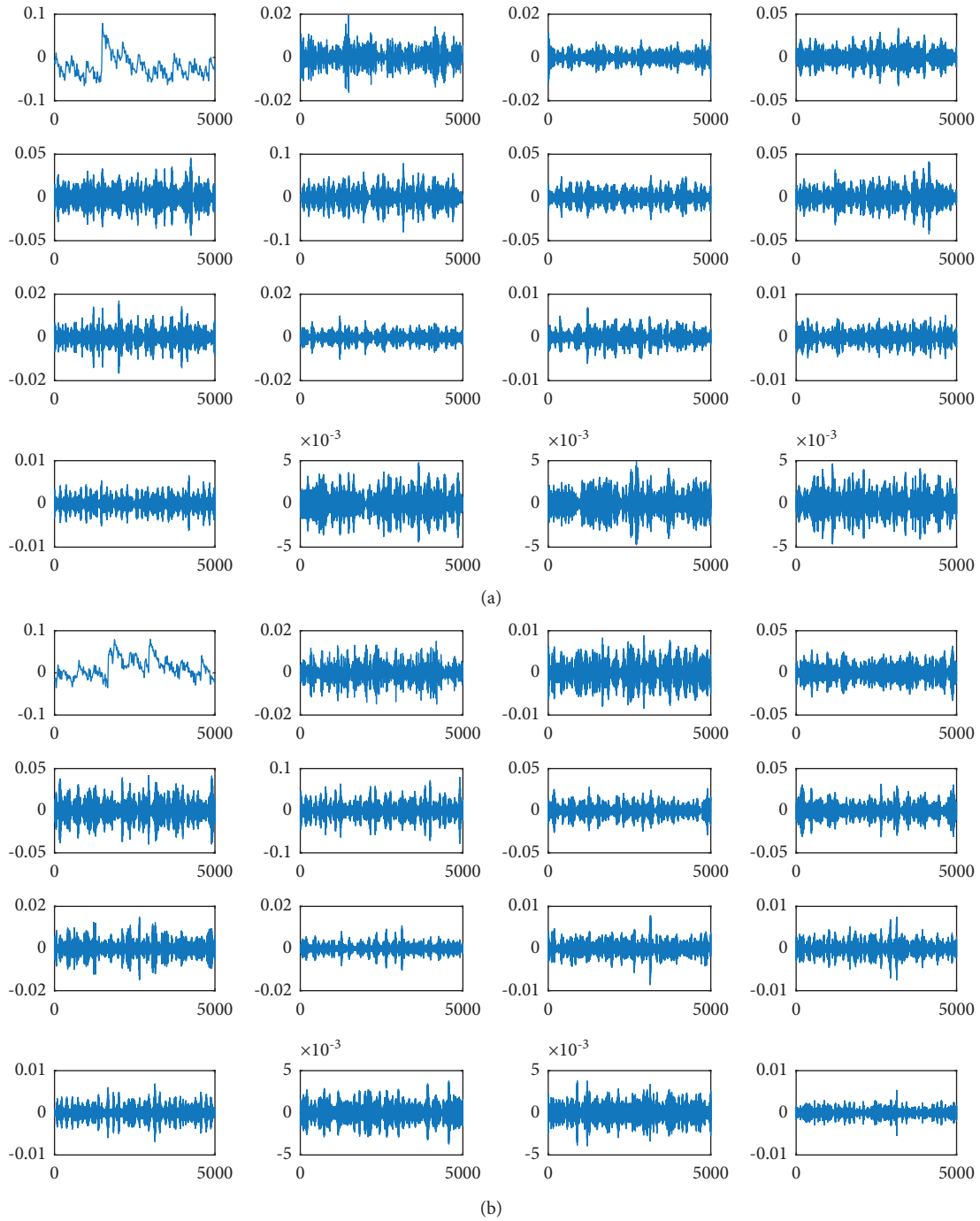


FIGURE 9: Reconstruction signals for wavelet packet nodes obtained by decomposing normal state vibration signals. (a) The reconstruction signals of NM signal under 1730 rpm by MODWPT. (b) The reconstruction signals of NM signal under 1750 rpm by MODWPT.

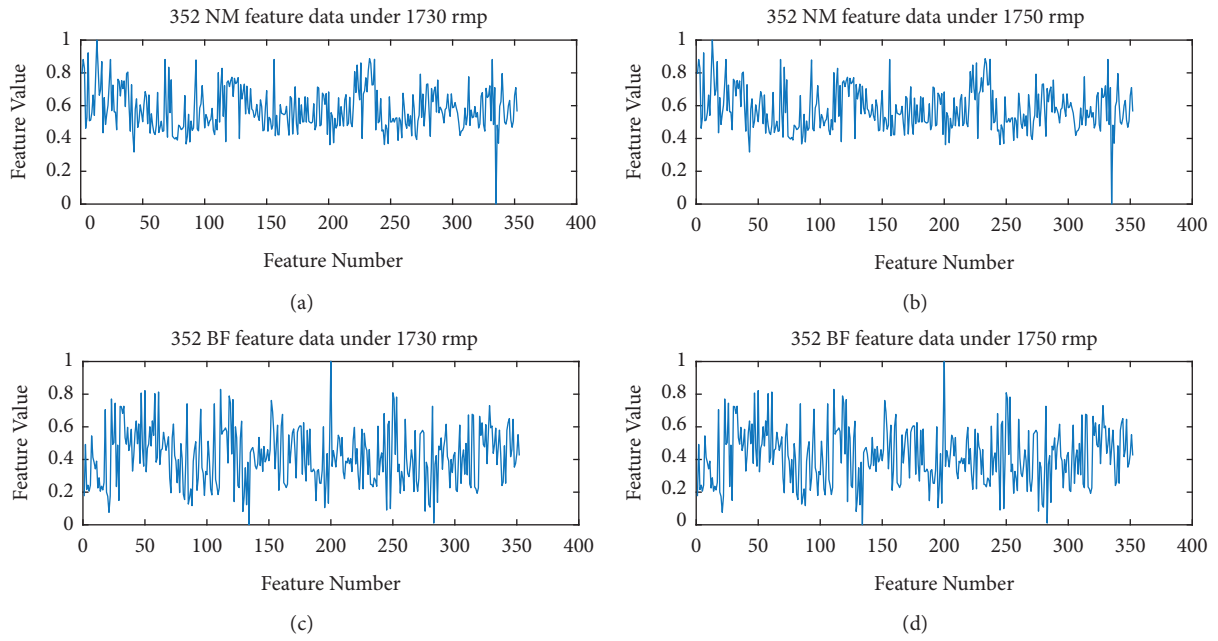


FIGURE 10: 352 statistical features extracted from NM and BF vibration signals under 1730 rpm and 1750 rpm (the feature data are normalized). (a) NM feature data under 1730 rpm. (b) NM feature data under 1750 rpm. (c) BF feature data under 1730 rpm. (d) BF feature data under 1750 rpm.

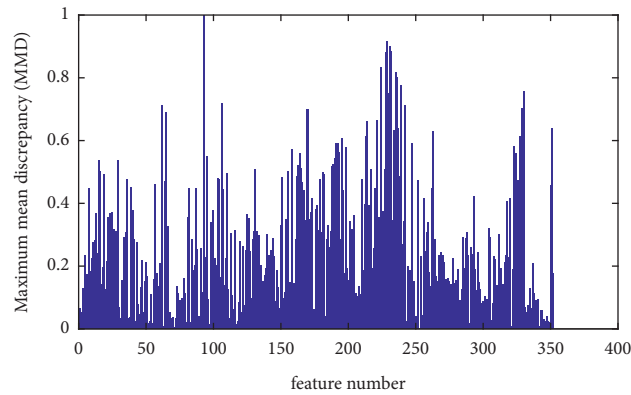


FIGURE 11: The maximum mean discrepancies of 352 features.

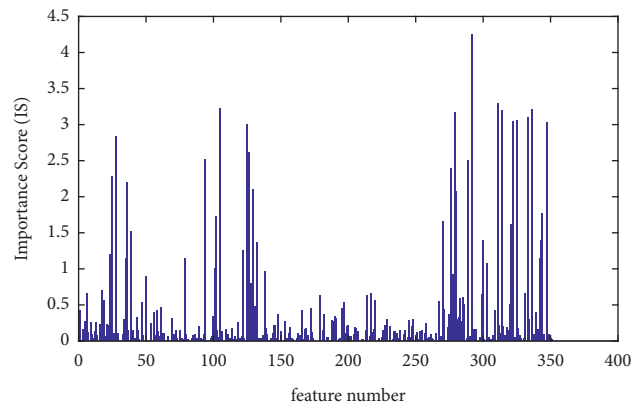


FIGURE 12: The importance scores of 352 features.

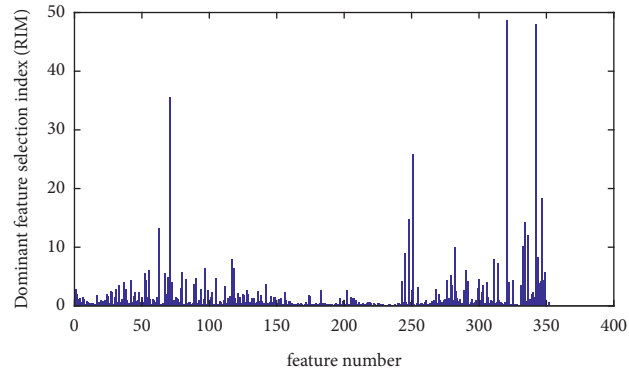


FIGURE 13: The dominant feature selection indexes of 352 features.

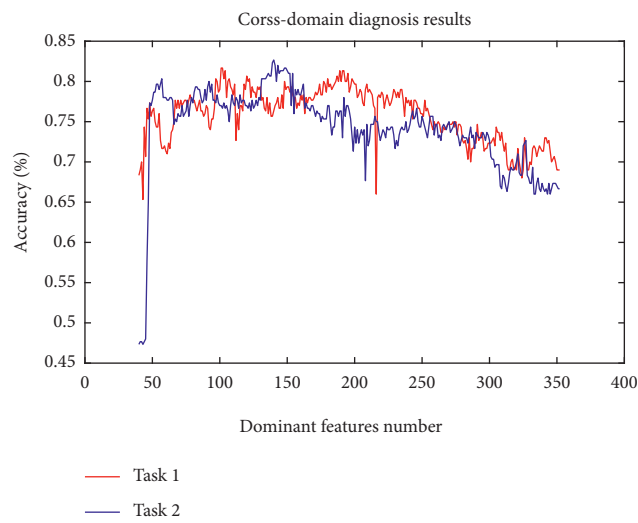


FIGURE 14: The diagram of cross-domain fault diagnosis results of TFDD framework.

TABLE 4: Cross-domain fault diagnosis results of TFDD framework.

Dominant features number (<i>dfn</i>)	Results of task 1 (%)	Results of task 2 (%)
40	68.33	47.33
50	76.33	77.67
70	76.67	75.67
90	77.00	77.67
110	78.33	77.33
130	79.33	77.67
150	79.33	81.33
170	78.00	76.33
190	80.00	76.33
210	78.67	72.00
230	78.67	71.67
250	76.00	75.67
270	74.67	75.00
290	72.33	73.67
310	72.33	68.33
330	70.00	67.33
352	69.00	66.67

TABLE 5: Comparative models based on common and competitive methods.

The model is not combined with the transfer learning method	The model is combined with the transfer learning method
RFS-SVM	RFS-TCA
RFS-KNN	RFS-JDA
RFS-DAE	RFS-DSID-TCA
RFS-DBN	RFS-DSID-JDA
RFS-CNN	TFDD

TABLE 6: Cross-domain fault diagnosis results of comparative models.

Models	Results of task 1 (%)	Results of task 2 (%)
RFS-SVM	60.33%	72.33%
RFS-KNN	48.00%	51.00%
RFS-DAE	46.33%	63.33%
RFS-DBN	36.33%	64.67%
RFS-CNN	69.67%	64.33%
RFS-TCA	66.00%	69.67%
RFS-JDA	72.33%	70.00%
RFS-DSID-TCA	71.67% (dfn:46)	77.00% (dfn:52)
RFS-DSID-JDA	77.00% (dfn:181)	78.00% (dfn:122)
TFDD	81.67% (dfn:101)	82.67% (dfn:140)

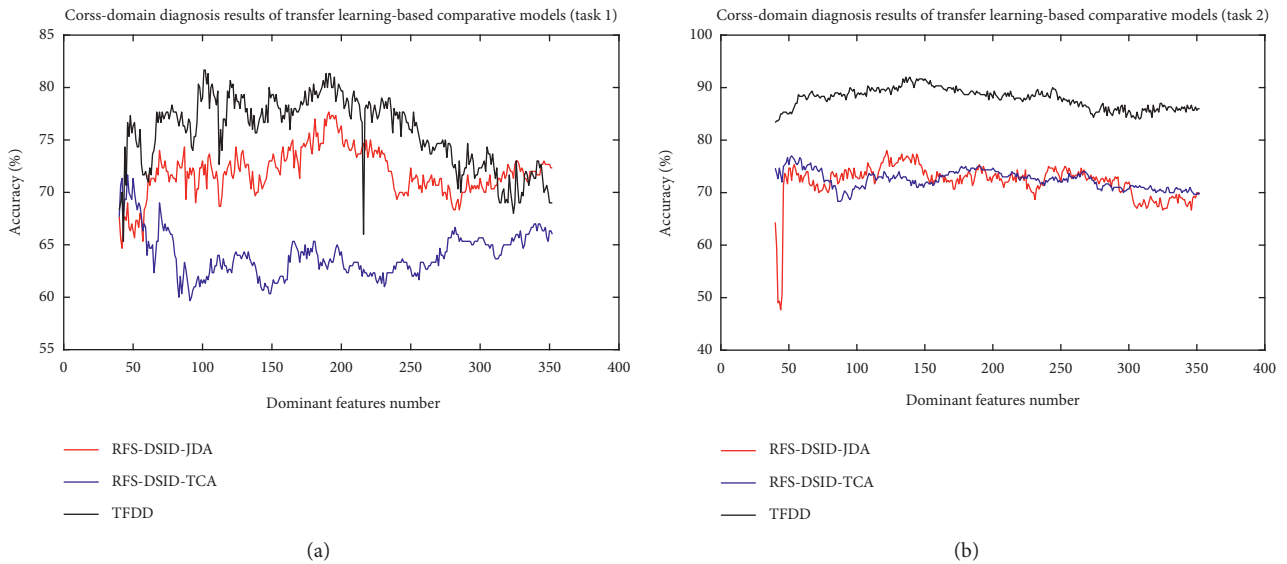


FIGURE 15: The diagram of cross-domain fault diagnosis results of transfer learning-based comparative models.

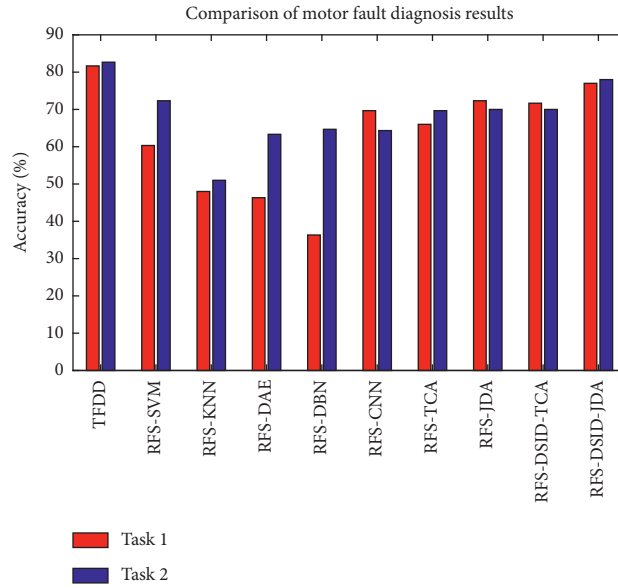


FIGURE 16: Comparison of fault diagnosis results of comparative models.

TABLE 7: Details about the bearing vibration dataset.

Bearing conditions	Fault diameters (mm)	Rotating speeds		Rotating speeds		Class label
		1200 rmp		1600 rmp		
		Size of training samples	Size of testing samples	Size of training samples	Size of testing samples	
IRF	0.05	30	60	30	60	1
	0.1	30	60	30	60	2
	0.2	30	60	30	60	3
ORF	0.05	30	60	30	60	4
	0.1	30	60	30	60	5
	0.2	30	60	30	60	6
BF	0.05	30	60	30	60	7
	0.1	30	60	30	60	8
	0.2	30	60	30	60	9
Normal	0	30	60	30	60	10

TABLE 8: 2 cross-domain fault diagnosis tasks for 10 bearing conditions.

	Rotating speeds (rmp)	Task 1		Task 2		
		Bearing conditions	Size of samples	Rotating speeds (rmp)	Bearing conditions	Size of samples
Source domain	1200	Classes 1–10	150	1600	Classes 1–10	150
Target domain	1600	Classes 1–10	300	1200	Classes 1–10	300

TABLE 9: Cross-domain fault diagnosis results of TFDD framework.

Dominant features' number (<i>dfn</i>)	Results of task 1 (%)	Results of task 2 (%)
50	85.33	83.50
70	86.33	84.00
90	85.67	83.33
110	86.83	85.17
130	86.50	84.67
150	90.33	89.33
170	87.50	86.50
190	86.00	85.83
210	82.50	81.33
230	81.67	80.83
250	81.50	80.83
270	81.50	80.83
290	81.50	80.83
310	81.67	81.00
330	81.50	80.83
352	81.50	81.00

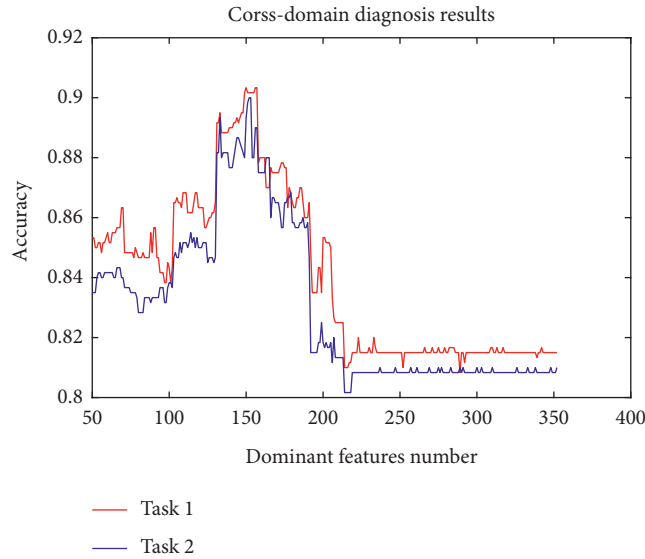


FIGURE 17: The diagram of cross-domain fault diagnosis results of TFDD framework.

TABLE 10: Cross-domain fault diagnosis results of comparative models.

Models	Results of task 1 (%)	Results of task 2 (%)
RFS-SVM	70.83	50.83
RFS-KNN	65.17	56.33
RFS-DAE	62.83	58.33
RFS-DBN	75.67	52.67
RFS-CNN	65.00	57.50
RFS-TCA	56.83	52.17
RFS-JDA	65.50	61.33
RFS-DSID-TCA	69.50% (<i>nkf</i> :142)	61.83% (<i>nkf</i> :231)
RFS-DSID-JDA	85.67% (<i>nkf</i> :175)	83.50% (<i>nkf</i> :271)
TFDD	90.33% (<i>nkf</i>:150)	90.00% (<i>nkf</i>:152)

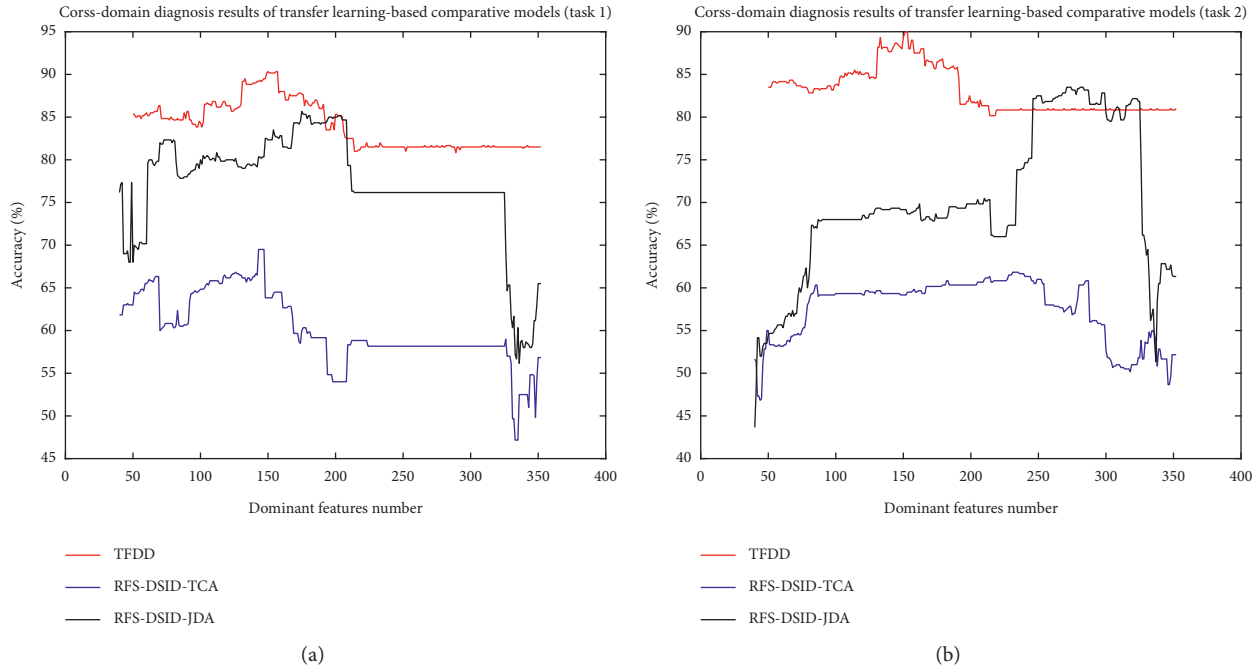


FIGURE 18: The diagram of cross-domain fault diagnosis results of transfer learning-based comparative models.

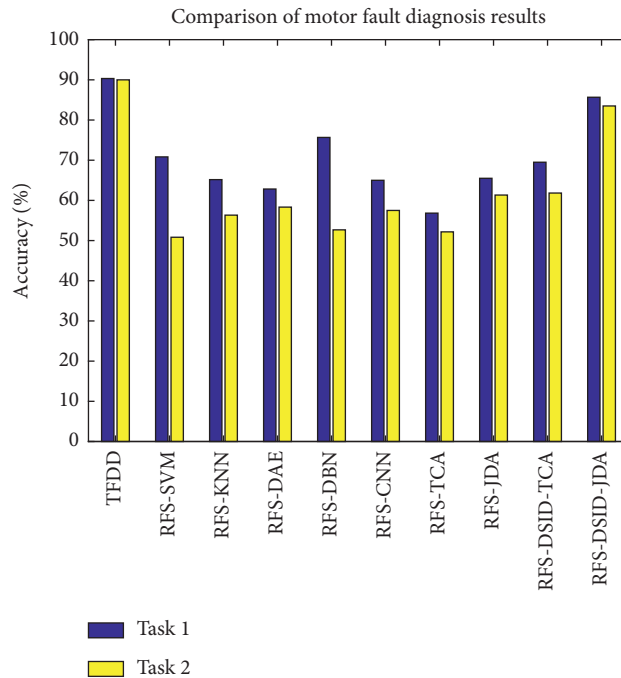


FIGURE 19: Comparison of fault diagnosis results of comparative models.

5. Conclusions

A new transferable fault diagnosis approach of rotating machinery based on deep autoencoder and dominant features selection, TFDD, is proposed. Firstly, the signal process and features extraction are performed. Then, based on the sufficient labeled source feature data and normal state target feature data, the proposed DSID is performed to evaluate the

features, the new selection index, RIM, is used to selected dominant features for training DTAE model. Next, by using labeled feature subset of source domain, a source DAE model can be learned and the corresponding parameters are transferred to the target DTAE model. Finally, this DTAE model classifies the unlabeled data from target domain.

A series of experiments are carried out by using motor and bearing fault datasets sampled from SQI-MFS test

platform. The experimental results prove the availability, flexibility, and advantages of the TFDD. The details are as the following aspects: (1) the proposed TFDD model can effectively diagnose faulty motors and bearings across different operating speeds, and the diagnosis performance significantly outperforms comparative models. (2) The proposed DSID can help to select features that have high fault-discriminative and domain invariance, when a suitable dfn is chosen, which can significantly improve cross-domain diagnosis performance.

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 are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Youth Science and Technology Fund of China University of Mining and Technology, Basic Scientific Research Business (no. 2021QN1093), the “Smart Mine” Key Technology R&D Open Fund of China University of Mining and Technology and Zibo Mining Group Co., Ltd. (no. 2019LH08), the National Key R&D Program of China (nos. 2017YFC0804400 and 2017YFC0804401), and the fund project of JiangSu Collaborative Innovation Center for Building Energy Saving and Construction Technology (no. SJXTY1603).

References

- [1] A.-S. Qin, H.-L. Mao, and Q. Hu, “Cross-domain fault diagnosis of rolling bearing using similar features-based transfer approach,” *Measurement*, vol. 172, no. 1, Article ID 108900, 2021.
- [2] D. Liu, W. Cheng, and W. Wen, “Intelligent cross-condition fault recognition of rolling bearings based on normalized resampled characteristic power and self-organizing map,” *Mechanical Systems and Signal Processing*, vol. 153, no. 11, Article ID 107462, 2021.
- [3] Z. Lei, G. Wen, S. Dong et al., “An intelligent fault diagnosis method based on domain adaptation and its application for bearings under polytropic working conditions,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2020.
- [4] D. Wei, T. Han, F. Chu, and M. J. Zuo, “Weighted domain adaptation networks for machinery fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 158, Article ID 107744, 2021.
- [5] Z. Zhou, C. Wen, and C. Yang, “Fault isolation based on k-nearest neighbor rule for industrial processes,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 4, pp. 2578–2586, 2016.
- [6] X. Zhang, W. Chen, B. Wang, and X. Chen, “Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization,” *Neurocomputing*, vol. 167, pp. 260–279, 2015.
- [7] R. S. Gunerkar, A. K. Jalan, and S. U. Belgamwar, “Fault diagnosis of rolling element bearing based on artificial neural network,” *Journal of Mechanical Science and Technology*, vol. 33, no. 2, pp. 505–511, 2019.
- [8] J. Wang, Y. Zhang, F. Zhang et al., “Accuracy-improved bearing fault diagnosis method based on AVMD theory and AWPSO-ELM model,” *Measurement*, vol. 181, Article ID 109666, 2021.
- [9] J. S. L. Senanayaka, H. Van Khang, and K. G. Robbersmyr, “Towards online bearing fault detection using envelope analysis of vibration signal and decision tree classification algorithm,” in *Proceeding of the 2017 20th International Conference on Electrical Machines and Systems (ICEMS)*, pp. 1–6, IEEE, Sydney, NSW, Australia, August 2017.
- [10] 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, 2017.
- [11] T. Guo and Z. Deng, “An improved EMD method based on the multi-objective optimization and its application to fault feature extraction of rolling bearing,” *Applied Acoustics*, vol. 127, pp. 46–62, 2017.
- [12] X. Ye, Y. Hu, J. Shen, R. Feng, and G. Zhai, “An improved empirical mode decomposition based on adaptive weighted rational quartic spline for rolling bearing fault diagnosis,” *IEEE Access*, vol. 8, Article ID 123813, 2020.
- [13] D. Zhong, W. Guo, and D. He, “An intelligent fault diagnosis method based on STFT and convolutional neural network for bearings under variable working conditions,” in *Proceedings of the 2019 Prognostics and System Health Management Conference (PHM-Qingdao)*, pp. 1–6, IEEE, Qingdao, China, October 2019.
- [14] H. Tao, P. Wang, Y. Chen, V. Stojanovic, and H. Yang, “An unsupervised fault diagnosis method for rolling bearing using STFT and generative neural networks,” *Journal of the Franklin Institute*, vol. 357, no. 11, pp. 7286–7307, 2020.
- [15] Y. Du, A. Wang, S. Wang, B. He, and G. Meng, “fault diagnosis under variable working conditions based on STFT and transfer deep residual network,” *Shock and Vibration*, vol. 2020, Article ID 1274380, 18 pages, 2020.
- [16] G. Li, C. Deng, J. Wu, Z. Chen, and X. Xu, “Rolling bearing fault diagnosis based on wavelet packet transform and convolutional neural network,” *Applied Sciences*, vol. 10, no. 3, 2020.
- [17] S. Xiong, H. Zhou, S. He, L. Zhang, and T. Shi, “Fault diagnosis of a rolling bearing based on the wavelet packet transform and a deep residual network with lightweight multi-branch structure,” *Measurement Science and Technology*, vol. 32, no. 8, Article ID 085106, 2021.
- [18] D. Fei, Y. Xiao, D. Enjie, S. Wu, C. Fan, and Y. Huang, “Rolling bearing fault diagnosis using modified neighborhood preserving embedding and maximal overlap discrete wavelet packet transform with sensitive features selection,” *Shock and Vibration*, vol. 2018, Article ID 5063527, 29 pages, 2018.
- [19] X. Yu, X. Ren, H. Wan, S. Wu, and E. Ding, “Rolling bearing fault feature extraction and diagnosis method based on MODWPT and DBN,” in *Proceedings of the 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, October 2019.
- [20] I. I. E. Amarouayache, M. N. Saadi, N. Guersi, and N. Boutasseta, “Bearing fault diagnostics using EEMD processing and convolutional neural network methods,”

- International Journal of Advanced Manufacturing Technology*, vol. 107, no. 9, pp. 4077–4095, 2020.
- [21] H. Li, T. Liu, X. Wu, and Q. Chen, "Application of EEMD and improved frequency band entropy in bearing fault feature extraction," *ISA Transactions*, vol. 88, pp. 170–185, 2019.
 - [22] J. Hou, Y. Wu, H. Gong, A. S. Ahmad, and L. Liu, "A novel intelligent method for bearing fault diagnosis based on EEMD permutation entropy and GG clustering," *Applied Sciences*, vol. 10, no. 1, Article ID 5063527, 2020.
 - [23] A. F. Aimer, A. H. Boudinar, N. Benouzza, A. Bendiabdellah, and M. E. A. Khodja, "Bearing fault diagnosis of a PWM inverter fed-induction motor using an improved short time Fourier transform," *Journal of Electrical Engineering & Technology*, vol. 14, no. 3, pp. 1201–1210, 2019.
 - [24] L. Xu, S. Chatterton, P. Pennacchi, and C. Liu, "A t order tracking method based on inverse short time fourier transform and singular value decomposition for bearing fault diagnosis," *Sensors*, vol. 20, no. 23, 2020.
 - [25] W. Zhao, Z. Wang, J. Ma, and L. Li, "Fault diagnosis of a hydraulic pump based on the CEEMD-STFT time-frequency entropy method and multiclass SVM classifier," *Shock and Vibration*, vol. 2016, Article ID 2609856, 8 pages, 2016.
 - [26] H. Shao, J. Lin, L. Zhang, and M. Wei, "Compound fault diagnosis for a rolling bearing using adaptive DTCWPT with higher order spectra," *Quality Engineering*, vol. 32, no. 3, pp. 342–353, 2020.
 - [27] H. Shao, H. Jiang, F. Wang, and Y. Wang, "Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet," *ISA Transactions*, vol. 69, pp. 187–201, 2017.
 - [28] A. Afia, C. Rahmoune, and D. Benazzouz, "New gear fault diagnosis method based on modwpt and neural network for feature extraction and classification," *Journal of Testing and Evaluation*, vol. 49, no. 2, pp. 1064–1085, 2019.
 - [29] X. Yu, W. Chen, C. Wu et al., "Rolling bearing fault diagnosis based on domain adaptation and preferred feature selection under variable working conditions," *Shock and Vibration*, vol. 2021, Article ID 8843124, 27 pages, 2021.
 - [30] R. V. Sánchez, P. Lucero, R. E. Vásquez, M. Cerrada, J. C. Macancela, and D. Cabrera, "Feature ranking for multi-fault diagnosis of rotating machinery by using random forest and KNN," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 6, pp. 3463–3473, 2018.
 - [31] J. Xiong, Q. Zhang, G. Sun, X. Zhu, M. Liu, and Z. Li, "An information fusion fault diagnosis method based on dimensionless indicators with static discounting factor and KNN," *IEEE Sensors Journal*, vol. 16, no. 7, pp. 2060–2069, 2016.
 - [32] Z. Yang, C. Kong, Y. Wang, X. Rong, and L. Wei, "Fault diagnosis of mine asynchronous motor based on MEEMD energy entropy and ANN," *Computers & Electrical Engineering*, vol. 92, Article ID 107070, 2021.
 - [33] T. A. Shifat and J.-W. Hur, "ANN assisted multi sensor information fusion for BLDC Motor fault diagnosis," *IEEE Access*, vol. 9, no. 99, pp. 9429–9441, 2021.
 - [34] J. Lu, W. Qian, S. Li, and R. Cui, "Enhanced K-nearest neighbor for intelligent fault diagnosis of rotating machinery," *Applied Sciences*, vol. 11, no. 3, 2021.
 - [35] S. Fei, "The hybrid method of VMD-PSR-SVD and improved binary PSO-KNN for fault diagnosis of bearing," *Shock and Vibration*, vol. 2019, Article ID 4954920, 7 pages, 2019.
 - [36] J. Wei, H. Huang, L. Yao, Y. Hu, Q. Fan, and D. Huang, "New imbalanced fault diagnosis framework based on Cluster-MWMOTE and MFO-optimized LS-SVM using limited and complex bearing data," *Engineering Applications of Artificial Intelligence*, vol. 96, Article ID 103966, 2020.
 - [37] X. Zhang, C. Li, X. Wang, and H. Wu, "A novel fault diagnosis procedure based on improved symplectic geometry mode decomposition and optimized SVM," *Measurement*, vol. 173, Article ID 108644, 2021.
 - [38] S. L. Souad, B. Azzedine, and S. Meradi, "Fault diagnosis of rolling element bearings using artificial neural network," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 5, 2020.
 - [39] R. Meng and J. Z. Ming, "A new strategy for rotating machinery fault diagnosis under varying speed conditions based on deep neural networks and order tracking," in *Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, December 2018.
 - [40] P. Lei, C. Shen, D. Wang, L. Chen, Z. Zhou, and Z. Zhu, "A new transferable bearing fault diagnosis method with adaptive manifold probability distribution under different working conditions," *Measurement*, vol. 173, Article ID 108565, 2021.
 - [41] X. Kong, G. Mao, Q. Wang, H. Ma, and W. Yang, "A multi-ensemble method based on deep autoencoders for fault diagnosis of rolling bearings," *Measurement*, vol. 151, Article ID 107132, 2020.
 - [42] P. Ma, H. Zhang, W. Fan, and C. Wang, "A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains," *ISA Transactions*, vol. 99, pp. 465–478, 2020.
 - [43] W. Mao, W. Feng, Y. Liu, D. Zhang, and X. Liang, "A new deep autoencoder method with fusing discriminant information for bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 150, no. 12, Article ID 107233, 2021.
 - [44] H. Zhiyi, S. Haidong, J. Lin, C. Junsheng, and Y. Yu, "Transfer fault diagnosis of bearing installed in different machines using enhanced deep autoencoder," *Measurement*, vol. 152, Article ID 107393, 2020.
 - [45] G. B. Jang and S. B. Cho, "Feature space transformation for fault diagnosis of rotating machinery under different working conditions," *Sensors*, vol. 21, no. 4, 2021.
 - [46] Z. He, H. Shao, P. Wang, J. Lin, J. Cheng, and Y. Yang, "Deep transfer multi-wavelet autoencoder for intelligent fault diagnosis of gearbox with few target training samples," *Knowledge-Based Systems*, vol. 191, Article ID 105313, 2020.
 - [47] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
 - [48] M. T. Uddin and M. A. Uddiny, "A guided random forest based feature selection approach for activity recognition," in *Proceedings of the 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, pp. 1–6, IEEE, Savar, Bangladesh, May 2015.
 - [49] Y. Wang and S. T. Xia, "A novel feature subspace selection method in random forests for high dimensional data," in *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 4383–4389, IEEE, Vancouver, BC, Canada, July 2016.
 - [50] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
 - [51] L. I. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression Trees. Wadsworth," *Biometrics*, vol. 40, no. 3, 1984.
 - [52] A. Gretton, K. Borgwardt, and M. Rasch, "A kernel method for the two-sample-problem," *Advances in Neural Information Processing Systems*, vol. 19, pp. 513–520, 2006.

- [53] 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. 199–210, 2010.
- [54] Y. Feng, J. Chen, Z. Yang et al., "Similarity-based meta-learning network with adversarial domain adaptation for cross-domain fault identification," *Knowledge-Based Systems*, vol. 217, Article ID 106829, 2021.
- [55] A. S. Minhas and S. Singh, "A new bearing fault diagnosis approach combining sensitive statistical features with improved multiscale permutation entropy method," *Knowledge-Based Systems*, vol. 218, no. 17, Article ID 106883, 2021.