IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 8, 2020, accepted October 25, 2020, date of publication October 28, 2020, date of current version November 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3034305

An Intelligent Fault Diagnosis Scheme Using Transferred Samples for Intershaft Bearings Under Variable Working Conditions

YA HE[1,2], MINGHUI HU[1,2], KUN FENG[1], AND ZHINONG JIANG[1]

[1]Key Laboratory of Engine Health Monitoring-Control and Networking, Ministry of Education, Beijing University of Chemical Technology, Beijing 100029, China

[2]Beijing Key Laboratory of High-End Mechanical Equipment Health Monitoring and Self-Recovery, Beijing University of Chemical Technology, Beijing 100029, China

Corresponding author: Kun Feng (kunfengphd@163.com)

This work was supported by the NSFC-Liaoning Joint Fund under Grant U1708257.

ABSTRACT Machine learning methods have made great development in data-driven fault diagnosis of rolling bearings. But the intelligent fault diagnosis of intershaft bearing faces the following two dilemmas: 1) the fault vibration is extremely weak, and it is difficult to extract features that can distinguish different classes; 2) due to the complex and variable working condition, the intershaft bearing does not always fail in same working conditions. That is, the labeled training data is not sufficient for every source domain. These challenges lead to the failure of traditional machine learning based fault diagnosis for intershaft bearings. Therefore, a novel intelligent fault diagnosis scheme is investigated for intershaft bearings of dual-rotor equipment under variable working conditions. The paper focuses on two key issues: 1) developing a feature extraction approach with which the fault features with excellent clustering and separation are extracted from vibration signals. This approach addresses the dilemma of weak fault feature extraction of intershaft bearing and creates a feasible precondition for subsequent feature transfer; 2) proposing a feature transfer method transforming the labeled sample features in multiple source domains into the trainable sample features in the target domain. This new transfer achieves the sharing of labeled training samples under working conditions and enriches the trainable samples in target domain. Ultimately, the faults of intershaft bearings can be diagnosed with the help of the neural network classifier trained by the transferred samples with labels. Experimental results verify that this established scheme is effective and superior to other comparable methods for the transfer diagnosis task from multiple source domains to target domain.

INDEX TERMS Intelligent fault diagnosis, feature transfer, weak fault extraction, neural networks.

I. INTRODUCTION

Intershaft bearing, as a key component, is widely used in dual-rotor equipment such as gas turbine. Its functional reliability has an extraordinary effect on the stability and safety of dual-rotor equipment. Even a small incipient defect may produce chain reaction and further lead to heavy casualties [1]. Intelligent fault diagnosis serves an essential role in pursuing the relationship between the monitoring data and the health states of bearings to prevent unpredictable failure in dual-rotor equipment [2], [3].

The framework of traditional intelligent fault diagnosis includes three main steps: 1) Data acquisition; 2) feature extraction and selection; 3) fault recognition. In the step of data acquisition, acceleration sensors are widely employed to detect the incipient faults of intershaft bearings. As both the inner and outer races of intershaft bearing rotate, it does not have fixed bearing housing. As shown in Figure 1, the intershaft bearing vibration is indirectly measured after passing through a complex transmission path [4].

For the second step, deep learning models are popular for mining the differences between different types of signals automatically, such as convolutional neural network (CNN) [5], [6], deep auto-encoder (DAE) [7] and deep belief network (DBN) [8]. However, due to the complex transfer path, the originally weak vibration of the intershaft bearing will be greatly attenuated and buried in strong vibration and interference, as the dotted line presented in Figure 1. Deep

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Xiang.

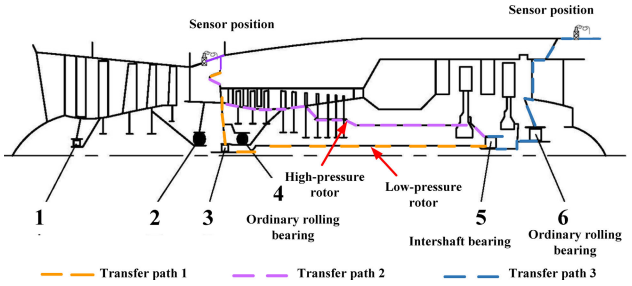

FIGURE 1. Schematic diagram of a gas turbine support structure.

learning could fail in the signal of weak fault features if the fault features are not clearly exposed. In contrast, it is more appropriate to use hand-created features based on professional knowledge [13]–[15]. To enhance weak fault features, the raw vibration signal of intershaft bearing can be preprocessed. There are many mature developments in weak feature enhancement for rolling bearings [9], such as discrete random separation (DRS), autoregression (AR), minimum entropy deconvolution (MED). The combination of AR and MED has been verified effectively to enhance the vibration generated by the faults of bearings, especially for large and high-speed machines such as gas turbines [10]–[12]. Further, mutual information (MI), as a kind of filter model, is an effective tool for feature selection after manual features creation, in order to reduce the adverse effect of redundant features in the dataset [16]–[19]. Thus, it is well worth trying to combine AR, MED and MI, named AMI, for feature extraction. AMI features could help the classifier model get better recognition results in the intelligent fault diagnosis for intershaft bearings.

In fault recognition step, the extracted features are used to train intelligent classifier, such as support vector machine (SVM) [20], [21], k-nearest neighbor (KNN) [22], artificial neural networks (ANN) [23] and so on, and the fault states can be determined by these classifiers. Because of the high self-learning ability, the ANN-based diagnosis model can automatically learn diagnosis knowledge from input data by minimizing the experience risk. It has become a popular method to classify health conditions of different machines.

The aforementioned fault diagnosis works have achieved fault classification effectively when satisfying a general condition: the training samples and testing samples are drawn from the same probability distribution, while they may fail in classifying unlabeled samples drawn from another distribution. In real-world application, due to the various working conditions, the fault samples collected from two domains show great differences in distribution, causing the decline of classification accuracy. Especially, the power performance of the dual-rotor gas turbine, the study object of this paper, depends on the speed of the high and low rotors. Any variation of the rotating speed of the two rotors will cause changes in the working conditions. This will adversely affect the intelligent fault diagnosis of the intershaft bearing from the following two aspects: 1) as mentioned above, the variation of working conditions affect easily the data distribution of intershaft bearing vibration, thus causing the failure of traditional intelligent diagnosis methods; 2) due to the complexity of working conditions, the intershaft bearing will not always fail in one certain working condition, so the training fault samples are usually collected from different working conditions. In this circumstance, the number of labeled samples in each source domain is very small. Although the classifier can be trained with all the labeled data of multiple source domains, the classifier performs badly in the target domain due to the divergence of data distribution in various source domains.

Transfer learning (TL) has provided a fine insight for tackling the first dilemma of fault recognition in the target domain through known source domain knowledge. Domain adaptive methods have been developed and received wide concern. Transferred component analysis (TCA) achieves feature transfer by searching a feature mapping to minimize the difference between source domain and target domain [24], but the shortcomings are much computation time and the poor nonlinear fitting. Joint distribution adaptation (JDA) adapts the source domain and target domain distribution by reducing the joint distribution distance [25]. However, as a supervised learning model, it will be affected by sample size and classifier performance. Differ from the TCA and JDA, Deep adaptation network (DAN) relies on the deep learning network, and it introduces three domain adaptation layers before a classifier [26]. But its limitation is enough labeled samples are necessary in training process. Besides, there are some other transfer learning methods, such as correlation comparison (Coral), adversarial domain adaptation (ADA) and so on [27], [28]. The above methods have been used widely by combining shallow learning or deep learning model in machine fault diagnosis under variable working conditions [29]–[31]. However, it is hard to get satisfactory results when applied in fault diagnosis of intershaft bearing owing to the lack of labeled data in one certain source domain, which is mainly considered by this paper. Therefore, there is still plenty of improvement room for the intelligent fault diagnosis of intershaft bearing.

In this paper, a novel intelligent fault diagnosis scheme for intershaft bearings is presented to promote the diagnosis accuracy under variable working conditions. Specifically, an AMI based feature extraction method (AMI-based FE) is developed for extracting distinguishable features from vibration signals. What's more, a feature transfer approach based on K-means and space transformation (KMST-based FT) is proposed, which transfers the labeled sample features from several source conditions to the target condition, and realizes the sharing of labeled samples across domains. After feature transfer, the labeled sample features from each source domain to the target domain are used to train the ANN classifier that can be adopted further to realize fault recognition in target condition. Even if the labeled samples number of each source domain is small, this procedure can improve the recognition results significantly. The proposed scheme is applied to analyze the experimental signals, and the results show the

effectiveness of this scheme compared with state-of-the-art methods.

The main contributions of this paper are summarized as follows.

1) This paper considers the fault signal characteristics of intershaft bearing in dual-rotor equipment, and develops the AMI-based FE method to extract weak fault features.

2) This paper makes full use of the labeled samples under various source domains to increase the number of trainable samples. The KMST-based FT approach is proposed to transfer the labeled sample features from multiple source domains to the target domain.

3) This paper has a low dependence on classifiers and can be extended from ANN to other classification models according to the practical application.

The remainder of this paper is organized as follows. Section II presents the proposed FET-NN for intershaft bearings under variable working conditions. Section III conducts the experimental results and analysis. Section IV and V draw the conclusion and discussion, respectively.

II. THE FRAMEWORK OF FET-NN

As introduced in Section I, the vibration signal characteristics of intershaft bearings that weak fault features and insufficient samples under variable working conditions lead to the accuracy degradation of traditional intelligent fault diagnosis. To overcome these obstacles, we need to extract robust features from the enhanced fault vibration signals of the intershaft bearing, and realize the transfer of labeled samples across working conditions, i.e., the sharing of the labeled samples in various working conditions. In this section, we present the proposed intelligent fault diagnosis scheme, which consists of FE based on AMI (AR, MED and MI), FT based on KMST (K-means and space transform), and fault diagnosis based on ANN. Therefore, the scheme is referred to be FET-NN, that is, feature extraction transfer and neural network. The framework of FET-NN is illustrated in Figure 2. The details of each method are elaborated in the following subsections.

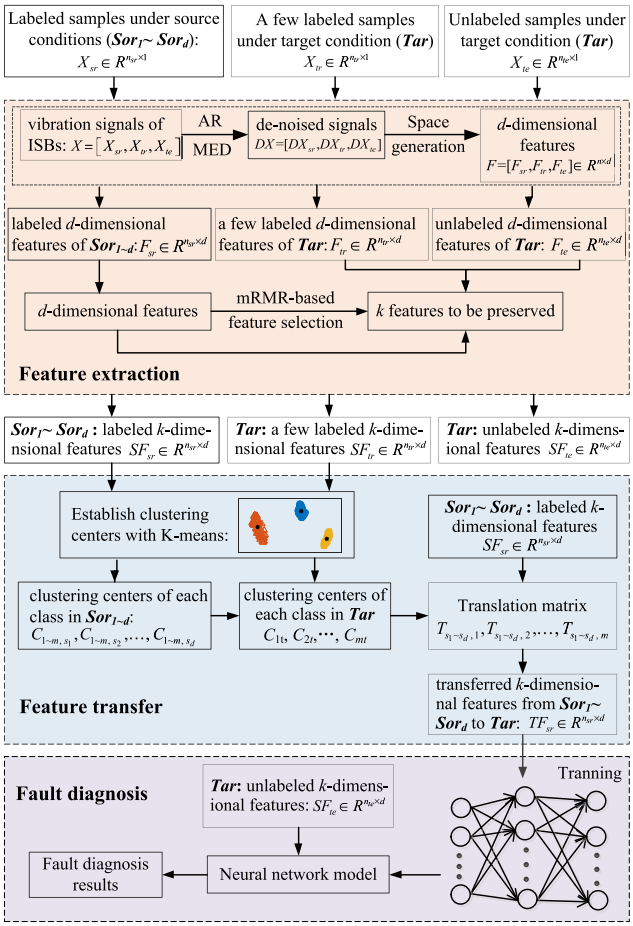

FIGURE 2. The framework of the proposed scheme FET-NN.

A. FEATURE EXTRACTION

The odious working environment and complex transmission path make the fault features of intershaft bearings in the measured signal extremely weak, and mixed with strong vibration components such as rotor and blade vibration, air flow noise and other interference. However, despite of the weak fault feature, the raw vibration signal contains rich fault information. Once the fault occurs, the periodic impact would appear in the collected signal due to the rotation of the defective bearing. Thus, these fault information can be identified in time domain and frequency domain.

In our work, the developed AMI-based FE consists of feature space generation and feature selection. In the former step, the raw vibration signals are preprocessed to expose fault impacts of intershaft bearings by AR, MED de-noising. Further, we extract multiple feature parameters from the de-noised signal in time domain and frequency domain to generate high-dimensional feature space. After the raw signal is enhanced, the generated high-dimensional feature space can represent health state of intershaft bearings. In the next step, MI based feature selection aims to optimize the high-dimensional feature space and construct a robust feature representation for training data and test data. It is capable of improving the performance of classification model in terms of recognition accuracy and reduction in training time. The main steps of AMI-based feature extraction are as follows:

Step 1: Collect the vibration data of intershaft bearings under $d+1$ working conditions as samples: $V_{data} \in R^{n\times 1}$, where n represents the number of samples, d denotes the number of source domains. Take the labeled data under condition $1\sim d$ with different fault types from V_{data} as training samples of source domain with labels: $X_{sr} \in R^{n_{sr}\times 1} Y_{sr} \in R^{n_{sr}\times 1}$, and take the labeled data of the last condition with different fault types from V_{data} as training samples of target domain with labels: $X_{tr} \in R^{n_{tr}\times 1} Y_{tr} \in R^{n_{tr}\times 1}$, and testing samples of target domain without labels $X_{te} \in R^{n_{te}\times 1}$ where $n = n_{sr} + n_{tr} + n_{te}$, and $n_{tr} \ll n_{sr}$.

Step 2: Preprocess each sample in source domain and target domain to enhance the fault feature by using AR and

TABLE 1. Feature dimension and statistical features.

No.	Statistical feature	No.	Statistical feature
1	Root mean square	9	Clearance factor
2	Square root value	10	Impulse factor
3	Peak value	11	Crest factor
4	Absolute mean value	12	Envelope negative entropy
5	Kurtosis	13	Gravity frequency center
6	Skewness	14	Mean square frequency
7	Variation factor	15	Frequency variance
8	Shape factor	16	Frequency division energy

MED de-noising as follows:

$$\begin{aligned} &\text{i.} \quad E(n) = X(n) + \sum_{k=1}^{p} a_k X(n-k) \\ &\text{ii.} \quad DX(n) = \sum_{l=1}^{L} f(l) E(n-l) \end{aligned} \tag{1}$$

where $X = [X_{sr}, X_{tr}, X_{te}]$, and a_k denotes the AR filter obtained by the Yule-Walker equation. $f(l)$ represents the inverse filter of MED, respectively. Ref. [10] provides detailed algorithm principle of MED. Then the de-noised signals $DX = [DX_{sr}, DX_{tr}, DX_{te}]$ can be obtained.

Step 3: Extract the commonly-used statistical features, displayed in Table 1, for each de-noised signal from the time domain and frequency domain to generate high-dimensional feature space $F = [F_{sr}, F_{tr}, F_{te}] \in R^{n\times d}$, $d = 25$ in this paper.

Step 4: Based on the labeled high-dimensional features and labels in source domain, feature subset sensitive to fault type can be selected by using mRMR-based MI from high-dimensional feature space:

$$\max_{f_j \in F_{sr} - S_{m-1}} \left[MI\left(f_j; y\right) - \frac{1}{S_{m-1}} \sum_{f_j \in S_{m-1}} MI\left(f_i; f_i\right) \right] \tag{2}$$

where $y \in Y_{sr}$ is the sample label and S denotes the feature subset. The mutual information $MI(X;Y)$ of discrete random variables X and Y is defined as:

$$MI(X;Y) = \sum_{x\in X} \sum_{y\in Y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)} \tag{3}$$

According to Equation (2), obtain d mRMR values and sort them, based on which k optimal features can be selected to form a new feature set. The d-dimensional feature spaces $F = [F_{sr}, F_{tr}, F_{te}] \in R^{n\times d}$ generated in Step 3 is optimized to $SF = [SF_{sr}, SF_{tr}, SF_{te}] \in R^{n\times k}$, the k-dimensional feature space, where $k < d$.

B. FEATURE TRANSFER

Although AMI-based FE can extract weak features with excellent dispersion and clustering from vibration signals of intershaft bearings under one certain working condition, it cannot solve the performance degradation problem caused by the insufficient samples of each source domain and the distribution divergence under various working conditions. Besides, the working conditions of dual-rotor equipment are complicated, so it takes too much labor and time to collect the enough labeled training data and reestablish the fault diagnosis model under each working condition.

In this section, we propose a KMST-based FT approach where the labeled features under each source domain are transferred to the target domain to realize the sharing of labeled samples across domains. This will significantly increase the labeled samples size without recollecting the labeled samples. The procedure of feature transfer can be described in details as follows.

Step 1: Give k-dimensional feature space $SF_{sr} \in R^{n_{sr}\times k}$ with labels $Y_{sr} \in R^{n_{sr}\times 1}$ in source domain ($Sor\,1 \sim Sor\,d$) and $SF_{tr} \in R^{n_{tr}\times k}$ with labels $Y_{tr} \in R^{n_{tr}\times 1}$ in target domain by feature extraction where the sample size of each condition is $s_1, s_2, \ldots, s_d$ and $n_{sr} = s_1 + s_2 + \ldots + s_d$.

Step 2: Establish the clustering centers in training feature space of source domain and target domain respectively by using K-means as follows, and more details about K-means can be seen in the Ref. [32]. Here, the number m of fault types is the number of clustering centers in the feature space.

$$\min \sum_{j=1}^{m} \sum_{i=1}^{n} d_{ji} \left\| x_i - C_j \right\|^2 \quad d_{ji} = \begin{cases} 1, & if\ x_i \in c_j \\ 0, & f\ x_i \notin c_j \end{cases} \tag{4}$$

where each subclass is expressed c_j ($j = 1, 2, \ldots, m$), and every c_j has its own clustering center C_j. The clustering centers $\left[C_{1-m,s_1}, C_{1-m,s_2}, \ldots, C_{1-m,s_d}\right] \in R^{m\times d}$ and $[C_{1t}, C_{2t}, \ldots, C_{mt}] \in R^{1\times d}$ in source domain and target domain can be established respectively.

Step 3: The translation matrix $[T_{s_1-s_d,1}, T_{1-s_d,2}, \ldots, T_{1-s_d,m}]$ of each class's cluster center from the source domain to the target domain can be calculated according to Equation (5)-(6), where $C_{jsi} \in C_{js}$, $C_{jti} \in C_{jt}$, $i = 1, 2, \ldots, d$, $j = 1, 2, \ldots, m$

$$\begin{aligned} \begin{bmatrix} C_{jt1} \\ C_{jt2} \\ \vdots \\ C_{jtd} \\ 1 \end{bmatrix} &= T \begin{bmatrix} C_{js1} \\ C_{js2} \\ \vdots \\ C_{jsd} \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 & t_1 \\ 0 & 1 & \cdots & 0 & t_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & t_d \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} C_{js1} \\ C_{js2} \\ \vdots \\ C_{jsd} \\ 1 \end{bmatrix} \\ T_j &= \begin{bmatrix} 1 & 0 & \cdots & 0 & t_1 \\ 0 & 1 & \cdots & 0 & t_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & t_d \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \end{aligned} \tag{5}$$

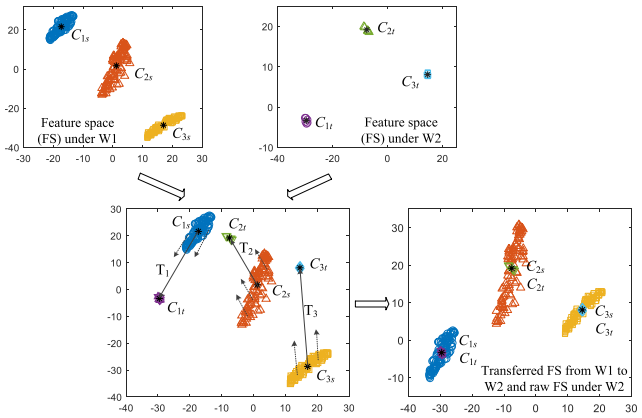

FIGURE 3. The schematic diagram of the feature transfer processing.

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 & C_{jt1} - C_{js1} \\ 0 & 1 & \cdots & 0 & C_{jt2} - C_{js2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & C_{jtd} - C_{jsd} \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{6}$$

Step 4: The labeled features can be transferred through multiplying the features in each class $c_{js} \in c_j$ of one source domain by the translation matrix corresponding to that class. The labeled features of each source domain are processed as above, and the sample features of multiple source domains are transferred. Then the transferred features $TF_{sr} \in R^{n_{sr} \times d}$ from each source domain to target domain can be achieved.

Figure 3 presents the feature transfer process with three fault types from a source domain to the target domain.

C. FAULT DIAGNOSIS

Input the transferred features $TF_{sr} \in R^{n_{sr} \times d}$ with labels $Y_{sr} \in R^{n_{sr} \times 1}$ into ANN with related parameters set for training the fault diagnosis model. It consists of input layer, one hidden layer and output layer.

Let f, l be the input vector and output vector of neural networks, respectively, where $f \in [TF_{sr}, TF_{tr}]$, $y \in [Y_{sr}, Y_{tr}]$. The commonly used objective function is the least square error (MSE), and the updating rule of parameters is as follows, where m is the total number of training samples, y_i is the target value, l_i is the network output. η is the learning rate which can be set appropriately, and ω, b are the training parameters of the neural network.

$$\begin{cases} \min L = \sum_{i=1}^{m} \frac{1}{2} (y_i - l_i)^2 \\ \omega = \omega - \eta \cdot \frac{\partial L}{\partial \omega} \quad b = b - \eta \cdot \frac{\partial L}{\partial b} \end{cases} \tag{7}$$

L gets closer and closer to the minimal value by fine-tuning these parameters for hundreds or thousands of times, thus finishing the training of neural network model.

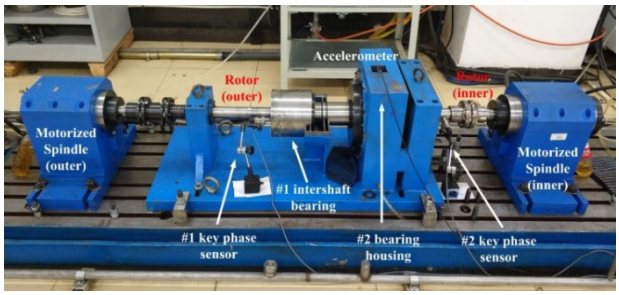

FIGURE 4. Dual-rotor vibration test rig.

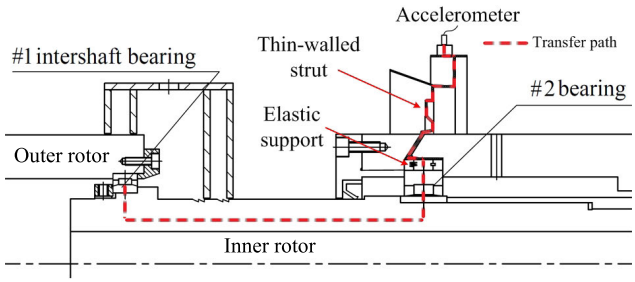

FIGURE 5. Transfer path of vibration signal of #1 intershaft bearing.

Finally, the features SF_{te} to be classified are input into the trained ANN based diagnosis model to realize the fault recognition of intershaft bearings in the target domain.

III. EXPERIMENT AND DISCUSSION

In order to verify the effectiveness and superiority of the proposed fault diagnosis scheme, the weak fault signals of bearings under variable operating conditions are collected.

A. CASE1: FAULT DIAGNOSIS BASED ON DATASET COLLECTED IN DUAL-ROTOR TEST RIG

1) EXPERIMENTAL SETUP AND DATASET PREPARATION

A dual-rotor test rig is designed and built to simulate the operating conditions and working state of the intershaft bearing on the actual dual-rotor gas turbine. As illustrated in Figure 4, it consists of two driving motors, #1 intershaft bearing, #2 bearing, #2 bearing housing and dual rotors. In particular, #2 bearings are fitted with thin-walled supports.

In actual dual-rotor equipment, since the inner and outer rings of intershaft bearings are rotating, there is no fixed bearing housing for the installation of sensors, which must require indirect measurement of vibration. The indirectly measured vibration response needs to be transmitted to the adjacent bearing via the shaft, and then to sensor position through the elastic support and complicated thin-wall paths so that the intershaft bearing's signal is buried in the strong vibration and interference.

Be similar to actual dual-rotor equipment, #1 intershaft bearing, in the dual-rotor test rig, has no directly linked bearing housing. Thus, although the #1 intershaft bearing is machined defectively, the accelerometer is installed at

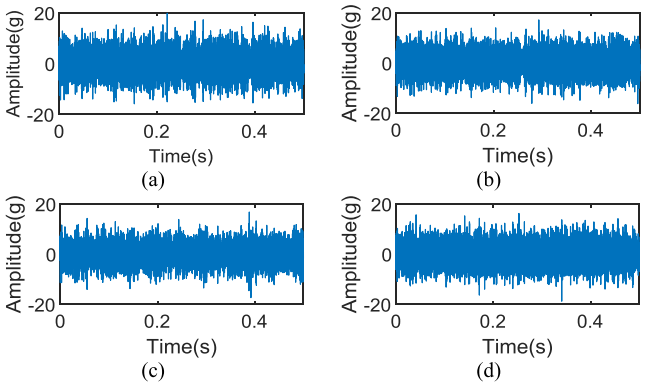

FIGURE 6. The signal waveforms of #1 bearing under condition A1. (a) normal. (b) outer race fault. (c) inner race fault. (d) rolling element fault.

#2 bearing housing. Figure 5 presents the transfer path of vibration of #1 intershaft bearing in the dual-rotor test rig. It can be seen that the test rig reproduces the running status of an actual intershaft bearing: the vibration response of #1 intershaft bearing should be transferred to adjacent bearing via the inner rotor; then, via elastic supports and thin-walled structure that comes from a real aero-engine, the response is transmitted to the #2 bearing housing. The complicated transfer path indicates that the detected bearing signal is severely attenuated during the transfer. Therefore, the fault signals of #1 intershaft bearing are similar to those in the actual dual-rotor equipment. Besides, the rotating speeds of the #1 intershaft bearing's inner and outer races are different and can vary frequently, which means the operating condition of #1 intershaft bearing in this test rig is similar to that in actual dual-rotor equipment.

Each fault type contains six kinds of working speeds, i.e. A1 = 600/0 rpm; A2 = 1500/900 rpm; B1 = 900/0 rpm; B2 = 1200/300 rpm; C1 = 1200/0 rpm; C2 = 1800/600 rpm, as shown in Table 2. Δ speed is the difference of running speed between inner rotor and outer rotor. The vibration signals are acquired by an acceleration sensor, LMS SCADAS data acquisition system [33], and collected from #2 bearing housing with a sampling frequency of 20480 Hz. Both the source domain and the target domain contain four fault types, and the original waveform of each sample contains 10240 data points. Due to the limited space, Figure 6 shows the signal waveforms of #1 bearing in condition A1 under four fault conditions. From the waveform, the amplitude of the vibration signal under the four bearing states is almost the same, and the fault vibration signal has no obvious shock.

The scenario settings of all experiments are trained on the labeled data under multiple conditions (source domain) and a few labeled data under the other condition (target domain) to diagnose the unlabeled test data under target domain. Based on the working characteristic of intershaft bearings, six different transfer tests are carried out on the dataset collected in the dual-rotor test rig (DRTR dataset), and the detailed description of transfer tests is illustrated in Table 3. In each transfer test, the part before the arrow represents all source domains, and that after the arrow refers to the target domain. Each transfer test includes five source datasets, but the sample size of each source dataset is only 20, and the sample size of target dataset is 100.

TABLE 2. Description of DRTR dataset.

Name	Fault types	Conditions	Δ speed
A1	NO,IF,OF,BF	600/0 rpm	600 rpm
A2		1500/900 rpm	
B1		900/0 rpm	900 rpm
B2		1200/300rpm	
C1		1200/0rpm	1200rpm
C2		1800/600rpm	

TABLE 3. Details of six transfer tests.

Transfer tests	Source dataset	Target dataset
A1, A2, B1, B2, C1→C2	A1,A2,B1,B2,C1: 20*5	C2: 100
A1, A2, B1, B2, C2→C1	A1,A2,B1,B2,C2: 20*5	C1: 100
A1, A2, B1, C1, C2→B2	A1,A2,B1,C1,C2: 20*5	B2: 100
A1, A2, B2, C1, C2→B1	A1,A2,B2,C1,C2: 20*5	B1: 100
A1, B1, B2, C1, C2→A2	A1,B1,B2,C1,C2: 20*5	A2: 100
A2, B1, B2,C1, C2→A1	A2,B1,B2,C1,C2: 20*5	A1: 100

TABLE 4. Parameters and optimization results of PSO.

Parameters	k_{min}	k_{max}
Range of values	1	27
Maximum evolution	50	
Maximum population	10	
Local search speed	1.5	
Global search speed	1.7	

TABLE 5. Parameters of ANN.

Parameters	Input layer	Hidden layer	Output layer
Nodes	2	10	1
Excitation function	\	tangent	linear
Iteration number		100	

2) PARAMETERS SETTING OF FET-NN

The detailed parameters for each experiment are set as follows. During the feature extraction, the optimal k in MI can be achieved by PSO (Particle Swarm Optimization) with the train samples of source domain [X2], where the main parameters' setting is presented in Table 4. The optimized k-value will be used for feature selection of both source domain and target domain samples. In the fault recognition processing, the parameter setting of ANN is shown in Table 5, with which the initial ANN classifier model can be established, and then the training samples are used to train classifier by iteration.

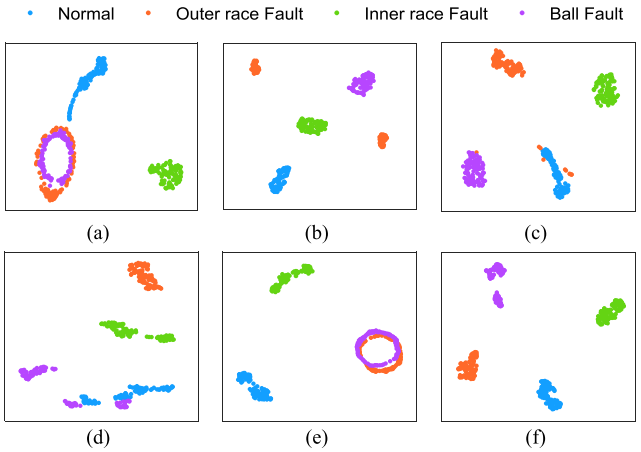

FIGURE 7. *t*-SNE visualization of features by CNN-based FSL. (a) A1. (b) A2. (c) B1. (d) B2. (e) C1. (f) C2.

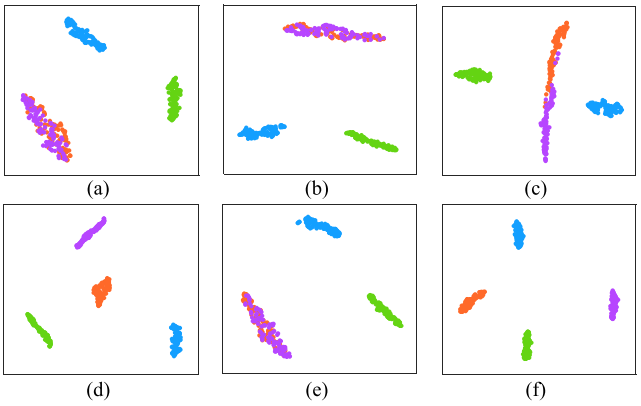

FIGURE 8. *t*-SNE visualization of features by FE without AMI. (a) A1. (b) A2. (c) B1. (d) B2. (e) C1. (f) C2.

3) FEATURE EXTRACTION EFFECT ANALYSIS

To show the superiority of the developed FE with AMI, the contrast methods of CNN-based feature self-learning (FSL) and FE without AMI are conducted respectively. Because of the weak vibration of the intershaft bearing, AR and MED are adopted to enhancing the signal features before using CNN-based FSL. We utilize the t-distributed stochastic neighbor embedding (t-SNE) technique to map the high-dimensional features into a 2-D space [34]. The results are shown in Figures 7-9.

In Figure 7, it is shown that the distribution of features obtained by FSL based on data are of poor divisibility and clustering even if the feature enhancement is performed by AR and MED before FSL. Compared with Figure 7, the features distribution under B2 and C2 conditions in Figure 8 is significantly improved. This confirms that the CNN-based FSL cannot excavate essential features for extremely weak fault features only from the vibration waveforms of intershaft bearings. It may achieve preferable results by combining frequency characteristics of signals. Figure 9 presents the features extracted by AMI. Compared to Figure 8, it can be seen that AMI addresses the dilemma that the fault features of outer race and ball cannot be distinguished under specific working conditions, i.e., A1, A2, B1 and C1.

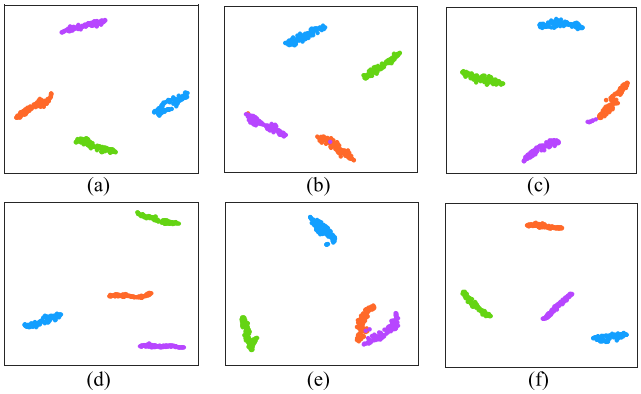

FIGURE 9. *t*-SNE visualization of features by AMI-based FE. (a) A1. (b) A2. (c) B1. (d) B2. (e) C1. (f) C2.

TABLE 6. The DB values obtained by three methods.

Methods	A1	A2	B1	B2	C1	C2
CNN-based FSL	1.95	1.64	0.55	0.85	1.22	0.42
FE without AMI	2.57	1.67	0.64	0.42	1.32	0.36
FE with AMI	0.45	0.48	0.51	0.43	0.67	0.38

Davies-Bouldin (DB) index is utilized for quantifying the feature extraction effect of the above three method. Because DB calculates the value of the distance between the samples in any two categories relative to the distance between the clustering centers of the two types, so it can evaluate well the clustering and dispersion of the extracted features [35]. DB is defined as follows.

$$DB = \frac{1}{k}\sum_{i=1}^{k}\max_{j\neq i}\left(\frac{\overline{C_i}+\overline{C_j}}{\|w_i - w_j\|_2}\right) \tag{8}$$

where k denotes the number of feature classes, w is the clustering center of a class of features, and C denotes the clustering center. The smaller the DB value, the smaller the intra class distance and the larger the inter class distance, which means the better feature extraction results. Table 6 lists the DB values of CNN-based FSL, FE without AMI and AMI-based FE. Obviously, the AMI-based FE method has an overwhelming advantage, which can well distinguish the features that make first two methods invalid. Therefore, AMI-based FE can effectively extract weak fault features of intershaft bearings, and lay a foundation for the realization of feature space transfer under different working conditions.

4) THE SUPERIORITY OF THE KMST-BASED FEATURE TRANSFER (FT)

According to Section II, the proposed KMST-based FT is a semi-supervised feature transfer approach, which needs a small number of target domain samples. Compared to training NN classifier directly with these few samples, called Direct-training, the KMST-based FT can achieve better performance.

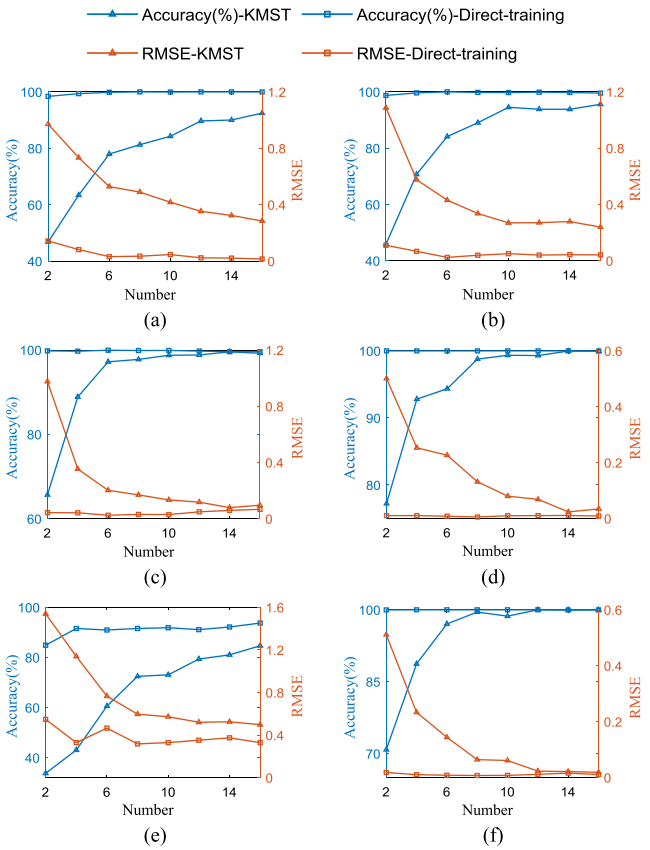

FIGURE 10. The trend of recognition accuracy with the number of labeled samples in target domain.

To validate this, we investigate the influence of the number of training samples in the target domain on the recognition accuracy of the above two methods respectively. The number of labeled training samples makes great influence on the recognition accuracy of the ANN classifier obtained by Direct-training. The performance of KMST does not depend on the number, because no matter what the number is, each class of samples must have a cluster center.

Figure 10 reports the results. From Figure 10, it is obvious that the accuracy of KMST-based FT is higher than that of Direct-training, and its RMSE is always lower than that of Direct-training at same number of samples in all six transfer tests. It is particularly noteworthy that the diagnosis based on Direct-training is evidently affected by the number of labeled samples in target domain. When there are only two labeled samples, the recognition effect of Direct-training is much poorer than KMST-based FT, and the recognition accuracy rate merely arrives at 60%. With the increasing numbers of labeled samples, the accuracy of Direct-training is significantly improved, but the convergence rate of which is lower than that of KMST-based FT. The latter is almost convergent when only four labeled samples are used. The above analysis confirms the advantage and necessity of the proposed KMST-based FT for sample feature transfer under variable working conditions.

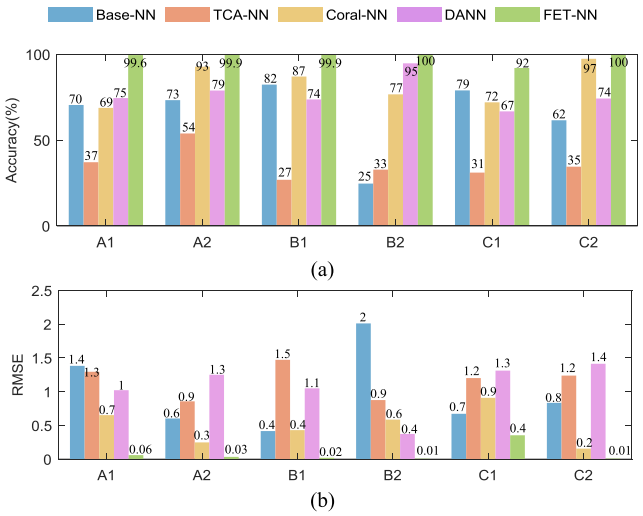

FIGURE 11. Recognition results of various methods. (a) Accuracy rate. (b) RMSE.

B. COMPARISON RESULTS

To verify the effectiveness of the scheme FET-NN, it is compared with the baseline approach and several successful fault diagnosis methods with transfer learning on the six transfer tests.

a. Base-NN: The NN classifier without feature transfer is generated. That is, the classifier is trained directly by the training samples from several source domains. Base-NN is designed to illustrate the improvement of transfer learning methods for fault diagnosis task where the samples of source domain and target domain aren't drawn from same probability distribution.

b. TCA-NN: TCA is a classical feature transfer method with MMD-regularized subspace learning [24]; then these transferred features are used to train the NN classifier.

c. Coral-NN: Coral is utilized to align their second-order statistics by linear transformation of sample features in source domain and target domain [27]; then, feature transfer is realized.

d. DANN: Deep adversarial training of neural network learns representation features that are predictive for the data samples in the source domain [26]; then it achieves fault identification.

The recognition accuracies and root mean square errors (RMSEs) are shown in Figure 11, where abscissa labels A1 $\sim$ C2 represent the working conditions corresponding to the target domain respectively. The average accuracies and RMSEs of the five methods are shown in Table 7. It can be seen that the proposed FET-NN outperforms all of the methods for comparison on six transfer tests. Specifically, through the comparison results, the following observation is obtained:

(1) Even though quite a few labeled training samples are provided by the five source domains, the Base-NN still can't obtain reliable recognition accuracy. The phenomenon is reasonable because the training data from multiple source

TABLE 7. The average recognition accuracies.

Average	Base-NN	TCA-NN	Coral-NN	DANN	FET-NN
Accuracy	65.2%	36.1%	82.5%	77.1%	98.6%
RMSE	0.987	1.157	0.497	1.071	0.081

domains are greatly different with the test data from target domain. This requires the transfer across working condition to improve the adverse situation.

(2) The performance of TCA-NN is much worse than that of Base-NN, and the recognition accuracy is generally less than 40%. This is because TCA adapts all source domain data at once, but the data distribution among different source domains is significantly distinct. It can be concluded that TCA is not suitable for the transfer work from multiple domains to one domain.

(3) Compared with TCA-NN, Coral-NN achieves better recognition accuracy because it aligns the samples of each source domain respectively. However, though the highest accuracy rate reaches 97%, it only maintains about 70% rate in other several transfer tests, which does not show strong robustness.

(4) DANN achieves stable recognition accuracy in all transfer tests, but the accuracy rate is only about 75%. This may be due to the fact that deep learning model is difficult to excavate excellent features from time waveforms, which will directly affect the subsequent transfer process.

(5) Thanks to the AMI-based FE and KMST-based FT, the average accuracies of FET-NN have been significantly enhanced, and it achieves the highest recognition accuracies and maintains excellent robustness in six transfer tests. This validates that FET transfers the training samples of multiple source domains to the target domain more effectively than TCA, Coral and DANN. Therefore, we conclude that FET-NN has very good potential for solving the classification performance drop-out problem caused by variable working conditions in the fault recognition field of intershaft bearings.

In order to illustrate the transferability of the features from FET and explain why FET works well, we also follow the t-SNE technique to visualize high-dimensional features of the aforementioned methods in a two-dimensional map. The transfer test A1, A2, B1, B2, C1→C2 is taken as an example to analyze the domain discrepancy effect. The visualization results are shown in Figure 12. Figure 12(a) plots features without transfer learning where the sample features of the each source domain are scattered in the 2-D space because of the discrepancy of feature distribution between variable working conditions. Figure 12(b)-(c) plots the features transferred by TCA and Coral respectively. It is evident that the distributions of transferred features in the latter map are closer than the former. This shows that Coral performs better than TCA in feature transfer of multi-source domains. Figure 12(d) plots the features obtained by DANN. It can be seen that since FSL cannot extract distinguishable features, the subsequent feature transfer will not obtain ideal results.

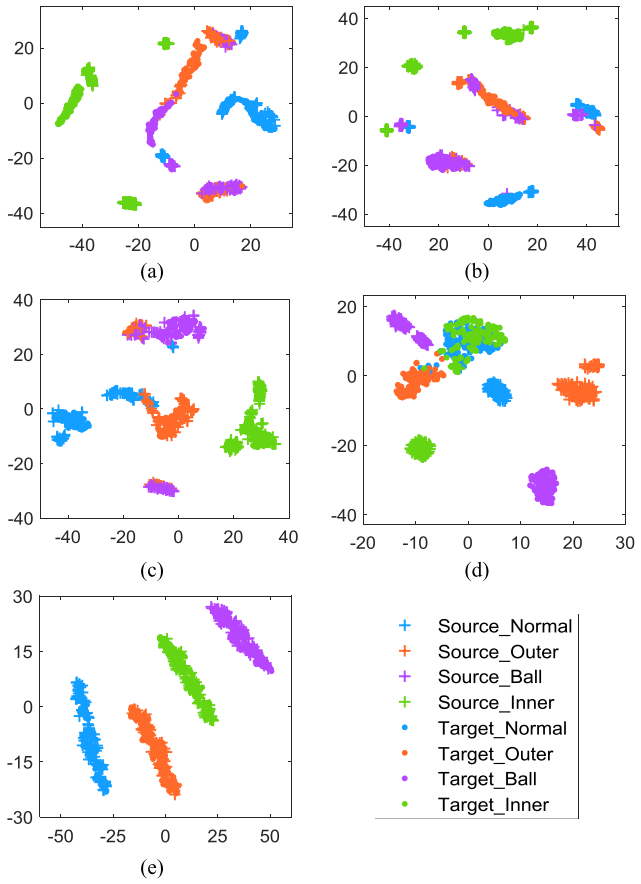

FIGURE 12. t-SNE visualization of features. (a) Baseline; (b) TCA; (c) Coral; (d) DANN; (e) FET.

Additionally, the features extracted and transferred by FET from multiple source domains can be well adapted to the distribution of target domain features. This is why the ANN classifier built with features obtained by FET can achieve high diagnostic accuracy even in the case of variable working conditions and weak fault features.

C. CASE2: SUPPLEMENTARY VERIFICATION BASED ON CWRU DATASET

In real-world application, ordinary rolling bearing is more widely used than intershaft bearing. Therefore, it is helpful to broaden the application scope of the proposed method by the validity verification of this method is also applicable to ordinary bearings. In this section, CWRU dataset is selected to the application of FET-NN to further illustrate that the proposed FET-NN has the potential to be borrowed to the fault diagnosis of rolling bearing in rotating machinery.

CWRU dataset contains the raw vibration data of roller element bearings obtained from fault simulation test rig of CWRU shown in Figure 13. The test rig is mainly composed of a driving motor, a 2 hp motor for loading, a torque sensor, a power meter, accelerometers, and an electronic control unit [36]. The test bearing are located in the drive end of motor shaft. They are subjected to electric sparking, and inner race

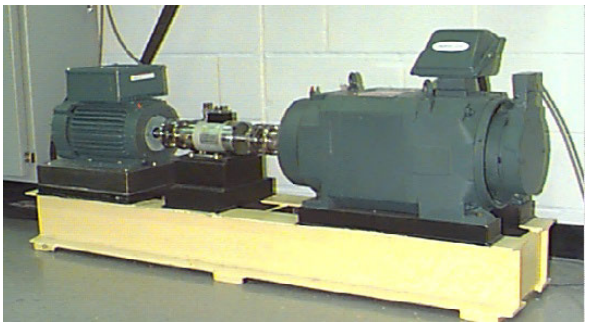

FIGURE 13. Bearing test rig of Case Western Reserve University [33].

TABLE 8. Description of CWRU dataset.

Name	Fault types	Conditions	Fault size
L0	NO,IF,OF,BF	0 hp/1797 rpm	0.021 inch
L1		1 hp/1772 rpm	
L2		2 hp/1750 rpm	
L3		3 hp/1730 rpm	
S1		0 hp/1797 rpm	0.007 inch
S2		1 hp/1772 rpm	
S3		2 hp/1750 rpm	
S4		3 hp/1730 rpm	

TABLE 9. Details of eight transfer tests.

Transfer test	Source domain	Target domain
S0, S2, S3→S1	S0, S2, S3: 40*3	S1: 100
S0, S2, S3→L1		L1: 100
S1, S2, S3→S0	S1, S2, S3: 40*3	S0: 100
S1, S2, S3→L0		L0: 100
L0, L2, L3→L1	L0, L2, L3: 40*3	L1: 100
L0, L2, L3→S1		S1: 100
L1, L2, L3→L0	L1, L2, L3: 40*3	L0: 100
L1, L2, L3→S0		S0: 100

faults (IF), outer race faults (OF) and ball faults (BF) of different sizes (0.021 inch and 0.007 inch) are processed. In this experiment, the vibration signals are collected by the accelerometers at a sampling frequency of 48 kHz.

CWRU dataset includes the following fault types: normal (NO), inner race fault (IF), outer race fault (OF), ball fault (BF) of 0.021 inch, and inner race fault (IF), outer race fault (OF), ball fault (BF) of 0.007 inch. Vibration data from each types are collected from four kind of working conditions, i.e., 0 hp/1797 rpm, 1 hp/1772 rpm, 2 hp/1750 rpm, 3 hp/1730 rpm. It is evident that the working condition parameters of the bearing test rig of CWRU are different from that of the dual-rotor vibration test rig, and rotating speed and load need to be considered. Each fault type of the source domain and the target domain contains 100 samples. The details of CWRU dataset are shown in Table 8.

For the purpose of verifying the effectiveness of FET-NN, the compared methods, i.e., Base-NN, TCA-NN, Coral-NN, DANN, are also conducted. The parameter setting of the

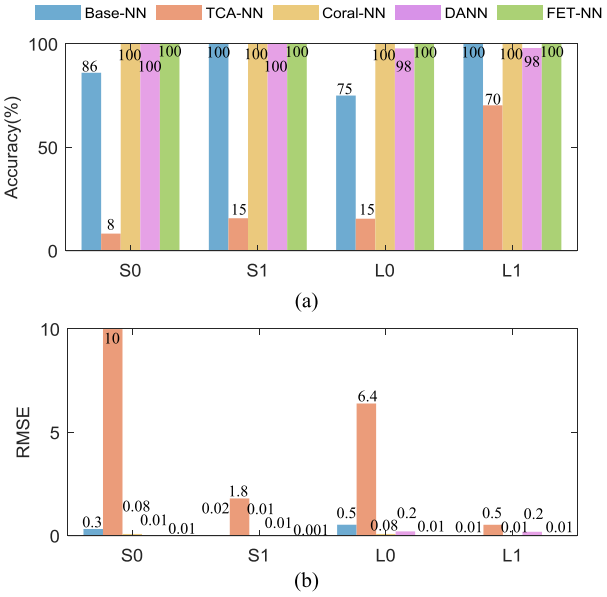

FIGURE 14. Recognition results for the across working condition transfer tasks with same fault levels.

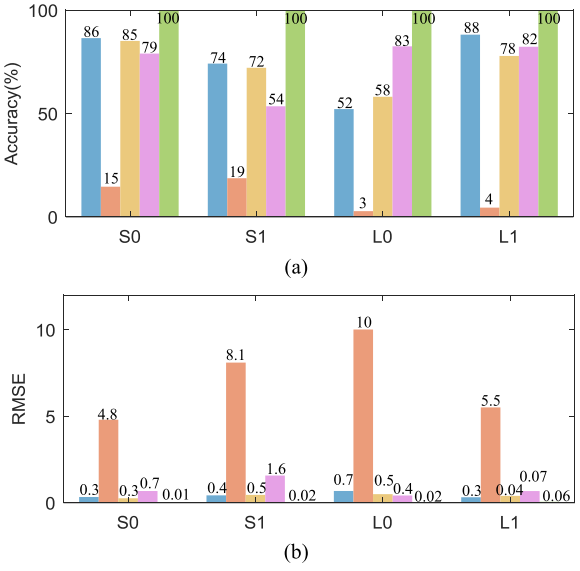

FIGURE 15. Recognition results for the across working condition transfer tasks with different fault levels.

method is the same as that of Tables 4 and 5. We consider the difference of difficulty in fault samples collection between general rotating machinery and dual-rotor equipment (more difficult for dual-rotor machinery), so eight sets of transfer tests are designed in this case, which include not only the transfer across working conditions but also the transfer across fault degrees. The detailed description of transfer tests is illustrated in Table 9. Each transfer test includes three source datasets, and the sample number of each source dataset is only 40, and the sample number of target dataset is 100.

The classified results of the proposed approach and the compared methods are presented in Figures 14 and 15. In Figure 14, the bearing vibration with the same fault degree is

collected in the source domain and the target domain, and the bearing fault degrees in the source domain and target domain are different in Figure 15. From these results, it is observed that TCA-NN is extremely inferior to other four methods. The accuracy rate of TCA-NN is as low as 4%, and its performance is even worse than Base-NN. This proves again that TCA is not suitable for transfer tasks from multiple source domains. When applied to the diagnosis task across working conditions, the effect of Base-NN, Coral-NN and DANN is better than that in the diagnosis task across working conditions and fault degrees at once. This is because there are more differences in features distribution between multiple source domains and the target domain. For example, Coral-NN and DANN reach 100% recognition accuracy in Figure 14, while in Figure 15 it can only achieve 85%, indicating that their tolerance for the distribution difference of data features is poorer than that of FET-NN.

From Figures 14-15, we can note that FET-NN achieves favorable recognition accuracies consistently and performs better than other methods obviously in all transfer tests. Not only the recognition accuracy rates are 100%, the highest RMSE value is only 0.06 that is far lower than the RMSE of other methods. It is worth mentioning that FET-NN can be applied to the intelligent fault diagnosis across operating conditions and fault levels at once, such as 0.021 inch and 0.007 inch here, which is rarely considered in other literatures. In short, FET-NN can solve the worse effect problem caused by insufficient training samples due to data distribution difference, whether the divergences originates from variable working conditions or fault development.

IV. CONCLUSION

In this paper, an intelligent fault diagnosis scheme for intershaft bearings under variable working conditions has been proposed. It consists of feature extraction (FE) based on AMI, feature transfer (FT) based on KMST, and fault diagnosis (FD) based on ANN.

(1) Assisted by the AMI-based FE, feature enhancement, multi-domain feature extraction and feature optimization are carried out respectively. Compared with CNN based FSL and FE without AMI, this method is more suitable for bearing weak fault feature extraction and has higher and adaptability under variable working conditions.

(2) As a semi-supervised transfer approach, KMST-based FT is proposed to solve the problem of insufficient labeled samples in one certain source domain. The visualization results show that it is more competent for the feature transfer task from multiple source domains than the results from the other compared methods.

(3) The ANN based FD is used to train ANN recognition model by using the labeled features transferred from multiple source domains to target domain, and then carry out fault diagnosis on the testing data of target domain. The diagnosis results demonstrate that the proposed scheme has better potential for solving the diagnosis worse accuracy problem caused by weak fault feature and shortage of labeled samples in one certain working condition for the intelligent fault diagnosis of intershaft bearing.

V. DISCUSSION

The proposed scheme can diagnose the weak fault of intershaft bearings under variable working conditions and can be applied in gas turbine or other dual-rotor equipment using intershaft bearings. The AMI-based FE in this scheme can extract robust fault features from the vibration signals mixed with strong disturbing vibration and noise. This key point contributes to the subsequent feature transfer and fault recognition. Meanwhile, the proposed KMST-based FT solves the dilemma of poor performance caused by inadequate labeled samples under variable operating conditions.

When the proposed scheme is implemented in the actual engineering equipment, it can expand the application of the following aspects: (1) the AMI-based FE method has strong robustness and adaptability for the application object. It can not only be applied to intershaft bearings, but also for the weak fault feature extraction of usual rolling bearings. However, if the bearing fault is prominent, the de-noising process based on AR or MED may be counterproductive and can be removed appropriately. (2) The KMST-based FT can achieve the sharing of labeled sample features under variable working conditions. It is also suitable for usual rolling bearings. Except the transfer between different working conditions, it can be used for the transfer across fault degree. In the next work, we will also make study on the transfer between different equipment.

However, the implementation of KMST-based FT needs to meet a critical precondition that the sample features extracted from various working conditions have great dispersion and clustering. That is, a few labeled samples in target domain can represent the clustering centers of all samples in target domain, and the labeled samples transferred from source domain to target domain can also replace a few samples in target domain for training to improve the accuracy of fault diagnosis. In this article, the AMI-based FE has provided this precondition for feature transfer obviously.

REFERENCES

[1] Z. Jiang, M. Hu, K. Feng, and H. Wang, "A SVDD and K-means based early warning method for dual-rotor equipment under time-varying operating conditions," *Shock Vib.*, vol. 2018, no. 937, pp. 1–14, 2018.
[2] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.
[3] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.
[4] Z. Jiang, M. Hu, K. Feng, Y. He, "Weak fault feature extraction scheme for intershaft bearings based on linear prediction and order tracking in the rotation speed difference domain," Applied Sciences, vol. 7, no. 9, pp. 2076–3417, 2017.
[5] W. Sun, R. Zhao, R. Yan, S. Shao, and X. Chen, "Convolutional discriminative feature learning for induction motor fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1350–1359, Jun. 2017.
[6] H. Wang, S. Li, L. Song, and L. Cui, "A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals," *Comput. Ind.*, vol. 105, pp. 182–190, Feb. 2019.

[7] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Rel. Eng. Syst. Saf.*, vol. 115, pp. 124–135, 2013.
[8] S. Haidong, J. Hongkai, L. Xingqiu, and W. Shuaipeng, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowl.-Based Syst.*, vol. 140, pp. 1–14, Jan. 2018.
[9] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," *Mech. Syst. & Signal Process.*, vol. 25, no. 2, pp. 485–520, 2011.
[10] H. Endo and R. Randall, "Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 906–919, Feb. 2007.
[11] D. Abboud, M. Elbadaoui, W. A. Smith, and R. B. Randall, "Advanced bearing diagnostics: A comparative study of two powerful approaches," *Mech. Syst. Signal Process.*, vol. 114, pp. 604–627, Jan. 2019.
[12] G. L. McDonald and Q. Zhao, "Multipoint optimal minimum entropy deconvolution and convolution fix: Application to vibration fault detection," *Mech. Syst. Signal Process.*, vol. 82, pp. 461–477, Jan. 2017.
[13] Y. Lei, M. J. Zuo, Z. He, and Y. Zi, "A multidimensional hybrid intelligent method for gear fault diagnosis," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1419–1430, Mar. 2010.
[14] X. L. Zhang, W. Chen, B. J. Wang, and X. F. Chen, "Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization," *Neurocomputing*, vol. 167, pp. 260–279, Nov. 2015.
[15] Y. Lei, Z. He, Y. Zi, and Q. Hu, "Fault diagnosis of rotating machinery based on multiple ANFIS combination with gas," *Mech. Syst. Signal Process.*, vol. 21, no. 5, pp. 2280–2294, Jul. 2007.
[16] H. C. Peng, F. H. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 1226–1238, 2005.
[17] A. Unler, A. Murat, and R. B. Chinnam, "mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification," *Inf. Syst.*, vol. 181, no. 20, pp. 4625–4641, 2011.
[18] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 483–519, 2013.
[19] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, no. 6, pp. 1157–1182, Jan. 2003.
[20] R. Liu, B. Yang, X. Zhang, S. Wang, and X. Chen, "Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis," *Mech. Syst. Signal Process.*, vol. 75, pp. 345–370, Jun. 2016.
[21] F. Chen, B. Tang, T. Song, and L. Li, "Multi-fault diagnosis study on roller bearing based on multi-kernel support vector machine with chaotic particle swarm optimization," *Measurement*, vol. 47, no. 1, pp. 576–590, 2014.
[22] D. He, R. Li, and J. Zhu, "Plastic bearing fault diagnosis based on a two-step data mining approach," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3429–3440, Aug. 2013.
[23] M. Amar, I. Gondal, and C. Wilson, "Vibration spectrum imaging: A novel bearing fault classification approach," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 494–502, Jan. 2015.
[24] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.
[25] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2200–2207.
[26] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37, 2015, pp. 97–105.
[27] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2015, pp. 1–8.
[28] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1–35, 2016.
[29] Z. Tong, W. Li, B. Zhang, F. Jiang, and G. B. Zhou, "Bearing fault diagnosis under variable working conditions based on domain adaptation using feature transfer learning," *IEEE Access*, vol. 6, pp. 76187–76197, 2018.
[30] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2018.
[31] X. Li, W. Zhang, Q. Ding, and X. Li, "Diagnosing rotating machines with weakly supervised data using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1688–1697, Jul. 2020.
[32] M. Capó, A. Pérez, and J. A. Lozano, "An efficient K-means clustering algorithm for massive data," *Data Mining Knowl. Discovery*, vol. 34, pp. 776–811, Jan. 2020.
[33] *Siemens, LMS SCADAS [EB/OL]*. Accessed: Jul. 15, 2017. [Online]. Available: https://www.plm.automation.siemens.com/zh/products/lms/testing/scadas/lab.shtml
[34] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 2605, pp. 2579–2605, Nov. 2008.
[35] S. K. Pal, S. K. Meher, and S. Dutta, "Class-dependent rough-fuzzy granular space, dispersion index and classification," *Pattern Recognit.*, vol. 45, no. 7, pp. 2690–2707, 2012.
[36] *Case Western Reserve University Bearings Vibration Dataset*. Accessed: Oct. 2015. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home

YA HE is currently pursuing the Ph.D. degree in power engineering and thermophysics with the School of Mechatronic Engineering, Beijing University of Chemical Technology, Beijing, China. Her current research interests include condition monitoring and fault diagnosis of aircraft rotor systems.

MINGHUI HU received the Ph.D. degree from the Beijing University of Chemical Technology, Beijing, China, in 2018. He is currently a Lecturer with the School of Mechatronic Engineering, Beijing University of Chemical Technology. His research interests include condition monitoring and fault diagnosis of aircraft engine.

KUN FENG received the Ph.D. degree from the Beijing University of Chemical Technology, Beijing, China, in 2013. He is currently an Associate Professor with the School of Mechatronic Engineering, Beijing University of Chemical Technology. His current research interests include fault diagnosis, machine learning, and remaining useful life prediction of rotating machinery.

ZHINONG JIANG received the Ph.D. degree in chemical process machinery from the Beijing University of Chemical Technology, Beijing, China, in 2010. He is currently a Professor in power engineering and thermophysics with the Beijing University of Chemical Technology. His research interests include condition monitoring and fault diagnosis of high-end mechanical equipment.