Received November 6, 2018, accepted November 19, 2018, date of publication November 23, 2018,
date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2883078

Bearing Fault Diagnosis Under Variable Working Conditions Based on Domain Adaptation Using Feature Transfer Learning

ZHE TONG[1], WEI LI[1], BO ZHANG[2], FAN JIANG[1], AND GONGBO ZHOU[1]

[1]School of Mechanical Engineering, China University of Mining and Technology, Xuzhou 221116, China
[2]School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China

Corresponding author: Wei Li (liwei_cmee@163.com)

This work was supported by the Outstanding Innovation Scholarship for Doctoral Candidate of Double First Rate Construction Disciplines of CUMT.

ABSTRACT Bearings, as universal components, have been widely used in the important position of rotating machinery. However, due to the distribution divergence between training data and test data caused by variable working conditions, such as different rotation speeds and load conditions, most of the fault diagnosis models built during the training stage are not applicable for the detection in the test stage. The models dramatically lead to the performance degradation for fault classification. In this paper, a novel bearing fault diagnosis method, domain adaptation by using feature transfer learning (DAFTL) under variable working conditions, is proposed to solve this performance degradation issue. The dataset of normal bearings and faulty bearings are obtained via the fast Fourier transformation of raw vibration signals, under different motor speeds and load conditions. Then, the marginal and conditional distributions are reduced simultaneously between training data and test data by refining pseudo test labels based on the maximum mean discrepancy and domain invariant clustering in a common space. Ultimately, a transferable feature representation for training data and test data is achieved. With the help of the nearest-neighbor classifier built on the transferable features, bearing faults are identified in this common space. Extensive experimental results show that the DAFTL can identify the bearing fault accurately under variable working conditions and outperforms other competitive approaches.

INDEX TERMS Fault diagnosis, vibration signal, domain adaptation, feature transfer learning.

I. INTRODUCTION

Bearings, as the most commonly used components are widely used in rotating machinery, whose health conditions, for example, the fault degree in different places under different motor speeds and loads, may have an extraordinary effect on the performance, stability and life span of the equipment [1] or even heavy casualties [2]–[4]. Hence, it is important to diagnose bearings under variable working conditions.

Bearing faults usually appear as cracks or spalls on the surfaces of the bearing especially on the roller, the outer race or inner race. The most obvious characteristic is its vibration. With the vibration signals under different conditions collected by the sensors [5], intelligent fault diagnosis methods are applied to recognize the fault types. Commonly intelligent fault diagnostic strategy covers mainly two parts: feature extraction and fault classification. The vibration signal sampled from the sensors is a raw temporal signal, which contains abundant information concerning bearings [6], including useful fault information. Therefore, it is necessary to find a way to extract the intrinsic information of bearings. Many signal processing methods are used for feature extraction, for example, time-domain statistical analysis, wavelet transformation, and Fourier spectral analysis. Then, we reduce the dimensions conducted for the sake of computational efficiency, such as principal component analysis (PCA) [7], independent component analysis (ICA) [8], and feature discriminant analysis. Finally, a classifier, such as nearest-neighbor (NN), support vector machine (SVM) or artificial neural networks (ANN), trained on the extracted features from raw vibration signal is used to verify the test data.

Recently, Huang *et al.* [9] proposed a genetic algorithm-based SVM (GA-SVM) model that can determine the optimal parameters of SVM with high accuracy and generalization ability. Amar *et al.* [10] proposed a fault diagnosis method that uses a preprocessed FFT spectrum image as input of ANN. The FFT spectrum image generated from raw vibration signal is first averaged using a 2D averaging filter and then converted to a binary image with the appropriate threshold selection. In [11], the discrete wavelet transform is used for feature extraction, and an artificial neural network is used for classification.

Many of the aforementioned works above have achieved good results under a general assumption: the training data and test data are drawn from the same distribution. In real-world applications, due to variable and complex working conditions, vibration signals sampled from sensors show large distribution differences between two domains, which lead to a dramatic drop in the performance. Therefore, there is still plenty of room for improvement under variable working conditions. Specifically speaking, we take the roller bearing fault diagnosis problem as an example; the training data for building the classifier could be sampled under certain motor speeds and load conditions, but the actual fault diagnosis application is to identify the fault types under different motor speeds and load conditions. Though the fault categories and degree are constant, the distribution differences between training data and test data are large. As a direct result, the classifier was trained with a very specific type of data, which means it may achieve high accuracy on similar data while performing poorly with another type. In other words, many diagnosis methods have poor domain adaptability. It is not uncommon to find that a classifier trained with data from one working load fails to classify samples obtained from another working load properly. Of course, we can spend a great deal of time and effort to recollect data to build a new classifier for the effective fault diagnosis on target domain. However, we cannot always replace a classifier by repetitively recollecting data. Further, it is very expensive or even impossible to rebuild the fault diagnosis model from scratch, using newly recollected training data for the actual task.

In order to avoid such recalibration effort, we might want to adapt a model built in one condition (source domain) for a new working condition (target domain) or to adapt the model trained on one rolling bearing (source domain) for a new rolling bearing (target domain). This adaptation leads to the research of domain adaptation (DA) [12] as one type of research for transfer learning, and their major purpose is to leverage knowledge learned from a source domain to use in a different but related target domain by reducing distribution differences [13]. Doing this is aimed at the purpose of minimizing the cross-domain prediction error, and the maximum mean discrepancy (MMD) [15] belonging to DA both can be applied to reduce distribution divergences.

In this paper, a novel bearing fault diagnosis method is proposed under variable working conditions based on domain adaptation using feature transfer learning (DAFTL). This method discovers a shared feature representation by jointly reducing marginal and conditional distributions simultaneously between training data and test data in a common space and preserves the important properties of training data via domain invariant clustering. Then, distribution divergence issues caused by variable working conditions are solved by the aforementioned shared feature representation without a deep neural network. First, the datasets for normal bearings and faulty bearings are obtained via the fast Fourier transformation (FFT) of raw vibration signals under variable motor speeds and load conditions. Then, in order to achieve the robust transferable feature representation for training and test domains, the pseudo outputs of the nearest-neighbor (NN) classifier in the test domain are used to refine fault diagnosis model in the training domain by using MMD, and domain invariant clustering (DIC) is embedded for preserving important properties of training data. Finally, with the help of an NN classifier built on the transferable features, bearing faults are accurately identified.

The rest of this paper is organized as follows. Section 2 reviews previous works and preliminaries, including domain adaptation and maximum mean discrepancy. Section 3 introduces the fault diagnosis method using feature transfer learning, including common feature space generation with FFT, and feature transfer learning. Section 4 presents the experimental analysis and discussion. Conclusions are provided in Section 5.

II. PREVIOUS WORKS AND PRELIMINARIES

A. DOMAIN ADAPTATION

DA, as one type of research on transfer learning, attempts to adapt a machine learning model established in a source domain for use in a different but related target domain automatically [13]. Generally, a domain is typically composed of a feature space of inputs $\mathcal{X}$ and a probability distribution of inputs $P(X)$, where $X = x_1, \cdots, x_n \in \mathcal{X}$ is a set of learning samples. Note that they have different data spaces and distributions when source domain and target domain are different, that is, $X_S \neq X_T$ and $P(X_S) \neq P(X_T)$ [13].

In this work, the goal of domain adaptation is to learn transferable features between two domains for solving the performance degradation problem caused by variable working conditions. We denote the training data from a source domain as $D_S = \{(x_{S_1}, y_{S_1}), ..., (x_{S_{n_1}}, y_{S_{n_1}})\}$, where $x_{S_i} \in \mathcal{X}$ is the input and $y_{S_i} \in \mathcal{Y}$ is the corresponding class label. Similarly, we let the test data from a target domain be $D_T = \{(x_{T_1}), ..., (x_{T_{n_2}})\}$ without labeled information, where the input $x_{T_i} \in \mathcal{X}$. We are concerned about the situation where the set of fault types remains the same in the source domain and the target domain, and the training data D_S are labeled, while the test data D_T are unlabeled.

Let $P(X_S)$ and $Q(X_T)$ be the marginal distributions of $X_S = \{x_{S_i}\}$ and $X_T = \{x_{T_i}\}$ from the training data and test data, respectively. Similarly, let $P(Y_S|X_S)$ and $Q(Y_T|X_T)$ be the conditional distributions of $X_S = \{x_{S_i}\}$ and $X_T = \{x_{T_i}\}$ from the training data and test data, respectively [23]. In our work, $P(X_S) \neq Q(X_T)$ and $P(Y_S|X_S) \neq Q(Y_T|X_T)$, which is well

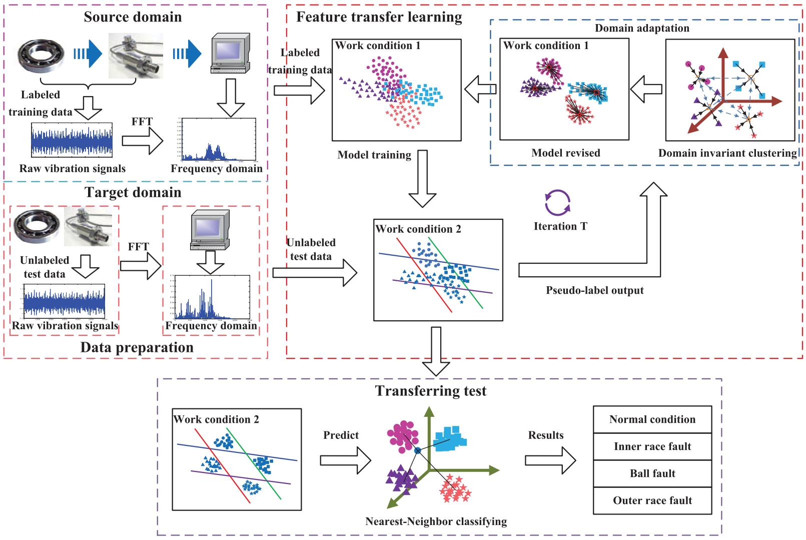


FIGURE 1. The framework of the proposed method for variable working condition fault diagnosis.

suited to fault diagnosis of bearings. We focus on that how to predict the fault types of bearing accurately in the unlabeled target domain with a different data distribution.

B. MAXIMUM MEAN DISCREPANCY

The purpose of typical domain adaptation is to reduce marginal distribution differences between two domains. In our work, the purpose of domain adaptation is to reduce both marginal and conditional distribution differences simultaneously by minimizing the empirical distance measure, which is suitable for bearing fault diagnosis under variable working conditions. In order to void expensive distribution calculation caused by the parametric criteria, a nonparametric distance measure, referred to as MMD, is widely used in the distribution adaptation. Taking data from D_S and D_T, the MMD calculates the empirical estimate of distances between two domains in the k-dimensional embedding [18]:

$$D_m(X_S, X_T) = ||\frac{1}{n_s}\sum_{i=1}^{n_s} B^T x_i - \frac{1}{n_t}\sum_{j=n_s+1}^{n_s+n_t} B^T x_j||^2 \quad (1)$$

where D_m is the distance of marginal distributions across domains, B is the adaptation matrix, and n_s and n_t denote the number of instances in the source and target domains, respectively.

III. FAULT DIAGNOSIS USING FEATURE TRANSFER LEARNING

As mentioned in Section 1, the existing large distributed differences caused by variable working conditions between training data and test data leads to performance degradation. In order to solve this issue, we need to obtain the shift between two domains and capture more robust transferable features from the training data and test data. In this section, we present a novel bearing fault diagnosis method for variable working conditions. The framework of this procedure is illustrated in Figure 1. The details of each part are elaborated in the following subsections.

A. COMMON FEATURE SPACE GENERATION WITH FFT

Raw time-series vibration signals are readily available and contain rich fault information. Due to the rotating nature of raw vibration signals from a defective bearing, the periodic impacts would appear in the obtained signals once a fault occurs. Thus, in general, these pieces of fault information can be identified in the frequency domain.

In our work, we directly catch the fixed point FFT amplitudes from the raw time-series vibration signal as samples that are from the same dimension. These samples, including labeled training data and unlabeled test data, are generated under different motor speeds and load conditions. The objective of generating a common feature space is to construct a low-dimensional robust feature representation for training data and test data that preserves the intrinsic properties of two domains after adaptation. The main steps of feature space generation are as follows:

- **Step 1**: Catch the fixed point FFT amplitudes from the raw time-series vibration signal collected under variable working condition as samples $D_{data} \in R^{n\times d}$, where n represents the number of samples and d denotes the dimensionality of the samples.
- **Step 2**: Take one of the conditions with different fault types from D_{data} as training samples $X_{tr} \in R^{n_{tr}\times d}$ with label $Y_{tr} \in R^{n_{tr}\times 1}$, and take another condition with different fault types from D_{data} as test samples $X_{te} \in R^{n_{te}\times d}$ without corresponding labels, where n_{tr} and n_{te}, respectively, represent the number of training samples and the number of test samples, and $n = n_{tr} + n_{te}$.
- **Step 3**: Project each labeled training and test sample into the common space by using PCA as follows:

$$\max_{B^T B=I} tr(B^T XHX^T B) \quad (2)$$

where $X = [X_{tr}, X_{te}] \in R^{n\times d}$, and $H = I - \frac{1}{n_{tr}+n_{te}} ll^T$, where I is considered as l as the ones vector. $B \in R^{n\times k}$ is an adaptation matrix that we want to find. Then the common feature space $Z = B^T X \in R^{k\times n}$ is generated.

B. FEATURE TRANSFER LEARNING

In the common feature space, the distribution divergence will still be very large. In order to solve the performance degradation problem caused by the distribution divergence between training data and test data under variable working conditions, transferable features that are robust for the two domains need to be extracted by explicitly minimizing the proper distance measures. MMD is applied for distance measures between x^i_{tr} and x^j_{te}:

$$||\frac{1}{n_{tr}}\sum_{i=1}^{n_{tr}} B^T x_i - \frac{1}{n_{te}}\sum_{i=1}^{n_{tr}+n_{te}} B^T x_j||^2 = tr(B^T X M_0 X^T B) \tag{3}$$

where $X = \{X_{tr}, X_{te}\}$. $M_0 = \begin{bmatrix} (M_0)_{tr,tr} & (M_0)_{tr,te} \\ (M_0)_{te,tr} & (M_0)_{te,te} \end{bmatrix}$ is the MMD matrix and is computed as follows [18], [20]

$$M_0 = \begin{cases} \frac{1}{n_{tr}n_{tr}}, & x_i, x_j \in X_{tr} \\ \frac{1}{n_{te}n_{te}}, & x_i, x_j \in X_{te} \\ \frac{-1}{n_{tr}n_{te}}, & otherwise \end{cases} \tag{4}$$

The marginal distributions between two domains are drawn close under the new representation $Z = B^T X$ by minimizing Eq.(4).

Training data and test data are from the same marginal and conditional distributions under ideal condition, while vibration signals sampled from sensors are different in practice, and reducing the differences in the marginal distributions does not guarantee that the conditional distributions between two domains can also be drawn close. In our work, the differences of conditional distribution between two domains are also reduced by mining the class-conditional distribution. MMD is modified to measure the class-conditional distributions.

$$||\frac{1}{n_{tr}}\sum_{i=1}^{n_{tr}} B^T x_i - \frac{1}{n_{te}}\sum_{i=1}^{n_{tr}+n_{te}} B^T x_j||^2 = tr(B^T X M_c X^T B) \tag{5}$$

where $M_c = \begin{bmatrix} (M_c)_{tr,tr} & (M_c)_{tr,te} \\ (M_c)_{te,tr} & (M_c)_{te,te} \end{bmatrix}$ is the MMD coefficient matrix that involves the class label c, and it can be computed with [18], [20]:

$$M_0 = \begin{cases} \frac{1}{n^c_{tr}n^c_{tr}}, & x_i, x_j \in X_{tr} \\ \frac{1}{n^c_{te}n^c_{te}}, & x_i, x_j \in X_{te} \\ \frac{-1}{n_{tr}n_{te}}, & \begin{cases} x_i \in X^c_{tr}, x_j \in X^c_{te} \\ x_j \in X^c_{tr}, x_i \in X^c_{te} \end{cases} \\ 0, & otherwise \end{cases} \tag{6}$$

The conditional distributions between two domains are drawn close under the new representation $Z = B^T X$ by minimizing Eq.(5).

Maximally preserving the local geometric structure of the data in a low-dimensional representation is beneficial for the out-of-sample problem [19]. Inspired by the linear discriminant analysis, DIC is introduced. To be specific, we minimize the within-class distance and maximize the between-class distance simultaneously based on training data. In this way, our method clusters the data with the same labels in the shared representation. We let S_w be the within-class scatter matrix and S_b be the between-class scatter matrix. We formalize the optimization problem:

$$min \frac{B^T S_w B}{B^T S_b B} \tag{7}$$

where $S_w = \sum_{\forall c\in C}\sum_{x_i\in X^c_{tr}} (x_i - \mu_c)^T(x_i - \mu_c)$ and μ_c denotes the mean of the samples in class c. $S_b = \sum N_c(\mu_c - \mu_0)^T(\mu_c - \mu_0)$, where N_c denotes the number of samples labeled class c and μ_0 denotes the mean of all samples.

In order to obtain the transferable feature representation, we aim to simultaneously minimize the differences in both the marginal and conditional distributions alongside DIC across domains by resorting the pseudo labels of test data [20], which can be obtained from a base classifier(NN classifier) trained on the labeled training data to predict the unlabeled test data. Ultimately, the optimization problem in this paper is comprised from Eq.(3), Eq.(5) and Eq.(7).

$$\min_{B^T S_b B = I} \sum_{c=0}^{C} tr(B^T X M_c X^T B + S_w) + \lambda ||B||^2_F \tag{8}$$

where I is considered as the identity matrix. λ is the regularization parameter. The goal is to find an adaptation matrix $B \in R^{n\times k}$, where the variance of data in the latent space is maximized. We derive the Lagrange function for Eq.(8) such that $\Lambda = diag(\Lambda_1, \cdots, \Lambda_k) \in R^{k\times k}$ is the Lagrange multiplier.

$$L = tr(B^T(X\sum_{c=0}^{C} McX^T + S_w + \lambda I)B) + tr((I - B^T S_b B)\Lambda) \tag{9}$$

Considering $\frac{dL}{dB} = 0$, the generalized Eigen decomposition is as follows.

$$(X\sum_{c=0}^{C} McX^T + S_w + \lambda I)B = S_b B\Lambda \tag{10}$$

Finally, the adaptation matrix B is obtained from solving Eq.(10) for the k smallest eigenvectors. The procedure of fault diagnosis based on domain adaptation using feature transfer learning(DAFTL) can be described in details as follows:

- **Step 1**: For given training data $X_{tr} \in R^{n_{tr}\times d}$ with label $Y_{tr} \in R^{n_{tr}\times 1}$ and unlabeled test data $X_{te} \in R^{n_{te}\times d}$ in the common feature space by Eq.(2).

- **Step 2**: Construct MMD matrix M_0 by Eq.(4), $Sw = \sum_{\forall c \in C} \sum_{x_i \in X_{tr}^c} (x_i - \mu_c)^T (x_i - \mu_c)$ and $S_b = \sum N_c(\mu_c - \mu_0)^T(\mu_c - \mu_0)$.
- **Step 3**: Adaptation matrix B constructed by the k smallest eigenvectors can be achieved by solving Eq.(10) through the Lagrange multiplier. Then, the robust transferable representation for the two domains is obtained as $Z = B^T X$.
- **Step 4**: Train the NN classifier on projected training data$\{B^T X_{tr}, Y_{tr}\}$, and then obtain pseudo test data labels Y_{te} that denote the conditional probability $Q(Y_{te}|X_{te})$ by using the trained NN classifier.
- **Step 5**: Update the MMD matrix $\{M_c\}_{c=1}^{C}$ by Eq.(6) according to $Q(Y_{tr}|X_{tr}) = Q(Y_{te}|X_{te})$, and then obtain the updated adaptation matrix B by solving Eq.(10) through the Lagrange multiplier. The updated transferable representation for the two domains is obtained as $Z = B^T X$, and then, we jump to **Step 4** until the end of the iteration.
- **Step 6**: Finally, the test data labels Y_{te} determined by the refined NN classifier are achieved.

IV. EXPERIMENTAL ANALYSIS AND DISCUSSION

In order to verify the effectiveness of the proposed fault diagnosis method, vast bearing vibration signals obtained from two test rigs are used. Dataset A is acquired from the bearing data center of Case Western Reserve University(CWRU) [22], and dataset B is obtained from a fault simulation testbed of the belt conveyor idler. DAFTL is compared with the baseline approaches and several successful methods.

a. Baseline1: The NN classifier with no projection and no adaptation is generated. That is, original input is directly used for diagnosis.

b. Baseline2: The NN classifier with no adaptation is created. Specifically, we use a new representation extracted from original input by PCA without domain adaptation.

c. NN SA: The NN classifier with projection and domain adaptation using subspace alignment is created [14].

d. TCA NN: Transferred component analysis(TCA) [23] is used for feature extraction; then, the NN classifier is used to classify these features.

a-b are classical methods without domain adaptation techniques, which have achieved success in many fault diagnosis applications. c is an effective subspace method proposed for solving image adaptation classification issues with domain adaptation techniques. d is one of the novel and efficient approaches in domain adaptation and has been applied successfully to fault diagnosis.

A. CASE 1: FAULT DIAGNOSIS BASED ON DATASET A

1) EXPERIMENTAL SETUP AND DATASET PREPARATION

The testbed shown in Figure 2 is composed of a driving motor, a 2 hp motor for loading, a torque sensor/encoder, a power meter, accelerometers and an electronic control unit [22], [24]. The test bearings are located in the motor shaft. They are subjected to electrosparking, and inner-race faults (IF), outer-race faults (OF) and ball faults (BF) of different sizes (0.007in, 0.014in, and 0.021in) are processed [25]. The vibration signals are sampled with the help of accelerometers attached to the rack with magnetic bases.

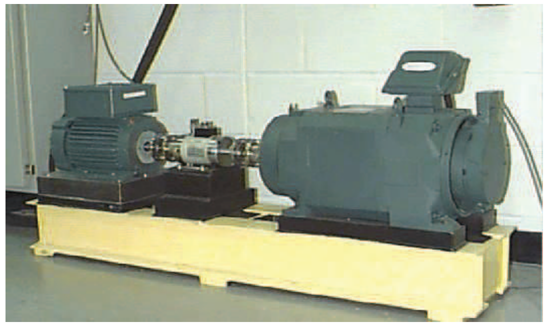


FIGURE 2. Bearing test rig of case western reserve university data center.

The working condition of the rotating machinery is variable generally in operation. In order to make the experimental results persuasive, in this experiment, dataset A is obtained under variable working conditions, which are collected from Drive End Bearing Fault Data and sampled at a frequency of 12 kHz. Dataset A includes three kinds of fault diameters (0.007in, 0.014in and 0.021in). Each fault diameter contains four fault types for the bearings: normal (NO), inner race fault (IF), outer race fault (OF) and ball fault (BF). Vibration data from each fault type are collected from four kinds of working conditions, i.e., L0 = 0 hp/1797 rpm, L1 = 1 hp/1772 rpm, L2 = 2 hp/1750 rpm and L3 = 3 hp/1730 rpm. Each sample in dataset A contains 2049 Fourier coefficients transformed from the raw vibration signal using FFT. Each domain in dataset A contains four fault types, and each fault type contains 200 samples. Due to the fact that labeled training data and unlabeled test data are sampled from different working conditions, it is impossible to find the optimal k and λ using cross-validation. Thus, empirically searching the parameter space is applied to find the optimal parameter settings, and the details are illustrated in Section 4. Finally, $k = 100$ and $\lambda = 0.1$ for diagnosis.

In order to verify the effectiveness of DAFTL, the contrast methods of a-d are also conducted simultaneously. The scenario settings of all experiments are trained on labeled training data under one single load (source domain) to diagnose the unlabeled test data under another load (target domain). In all, 48 different transferring tests are carried out on dataset A, and the detailed description of the experimental setup is illustrated in Table 1.

2) DIAGNOSIS RESULTS OF THE PROPOSED METHOD

The diagnostic results for the fault sizes of 0.007in, 0.014in and 0.021in are illustrated in Figure 3, Figure 4 and Figure 5. The average classification accuracies of the five methods are shown in Table 2.

Each figure consists of four subfigures, and the target domains in every figure are ordered clockwise from the top

TABLE 1. Description of the experimental setup in Case 1.

Target ♯ of test	Diagnose unlabeled test samples in the target domain			
	Labeled source domain	Unlabeled target domain	Fault type	Fault size
1	L0,L1,L2,L3	L0	IF,BF, OF,NO	0.007in
2	L0,L1,L2,L3	L1	IF,BF, OF,NO	0.007in
3	L0,L1,L2,L3	L2	IF,BF, OF,NO	0.007in
4	L0,L1,L2,L3	L3	IF,BF, OF,NO	0.007in
5	L0,L1,L2,L3	L0	IF,BF, OF,NO	0.014in
6	L0,L1,L2,L3	L1	IF,BF, OF,NO	0.014in
7	L0,L1,L2,L3	L2	IF,BF, OF,NO	0.014in
8	L0,L1,L2,L3	L3	IF,BF, OF,NO	0.014in
9	L0,L1,L2,L3	L0	IF,BF, OF,NO	0.021in
10	L0,L1,L2,L3	L1	IF,BF, OF,NO	0.021in
11	L0,L1,L2,L3	L2	IF,BF, OF,NO	0.021in
12	L0,L1,L2,L3	L3	IF,BF, OF,NO	0.021in

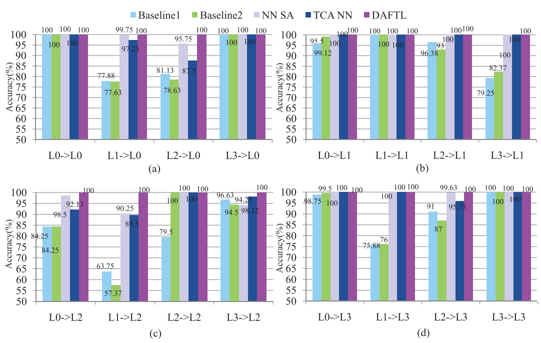


FIGURE 3. The results with a fault size of 0.007in based on a sample frequency of 12 kHz.

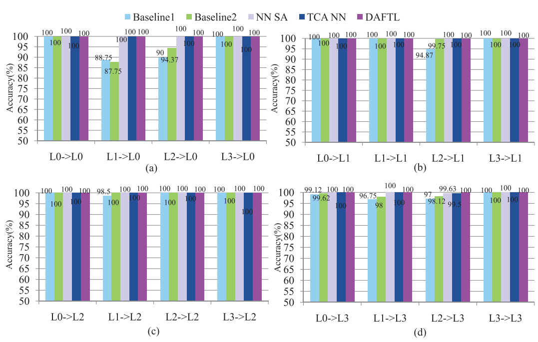


FIGURE 4. The results with a fault size of 0.014in based on a sample frequency of 12 kHz.

left: L0, L1, L2 and L3 in sequence. The left of the symbol "− >" in every part represents the source domain, and the right represents the target domain. For each set of bars

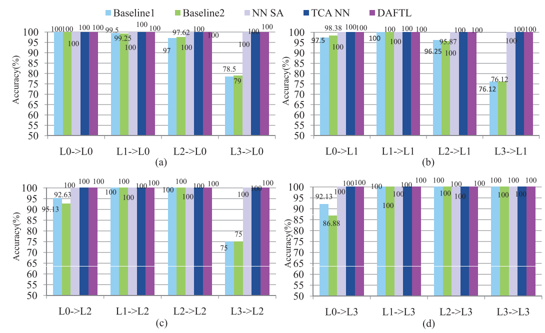


FIGURE 5. The results with a fault size of 0.021in based on a sample frequency of 12 kHz.

TABLE 2. The average classification accuracies.

AVG(%)	0.007in	0.014in	0.021in	Total
Baseline1	88.74	97.81	94.20	93.58
Baseline2	89.34	98.60	93.80	93.91
NN SA	98.63	99.98	100	99.54
TCA NN	97.52	99.97	100	99.16
DAFTL	100	100	100	100

in Figures 3, 4 and 5, the results indicate a transference from the source domain to target domain. The loads and speeds between different domains have great differences. For example, in Figure 3(a), the target domain is L0 (the load is 0 hp and the speed is 1797 rpm), and the source domains are L1 (the load is 1 hp and the speed is 1772 rpm), L2 (the load is 2 hp and the speed is 1750 rpm) and L3 (the load is 3 hp and the speed is 1730 rpm).

From the diagnostic results in Figures 3, 4 and 5, the highest accuracy rates for fault diagnosis can be obtained when the training data from one domain are the same with the test data from another domain, and this phenomenon is reasonable. It is clear that the baseline methods are very poor. For example, in Figure 3 and Figure 5, many results from the baseline methods are about 75%. Especially in Figure 3(a), several accuracies for the baseline methods cannot reach 70%, which indicates traditional methods cannot be applied to fault diagnosis under variable working conditions. NN SA is better than the baseline methods and is slightly better than TCA NN as a whole. In Figure 3(c), the inferiority of NN SA and TCA NN under variable working condition fault diagnosis is very obvious; the classification accuracy rates that transferring from L1 to L2 are only about 90%, and similar cases also appear in Figure 3(a), which can indicate that traditional transference methods also cannot be applied to solve the performance degradation problem under variable working conditions. It is exciting that DAFTL is evidently superior to the other compared methods in almost all situations, whatever the source domains and target domains are. It is worth noting that DATFL can achieve 100% accuracy. Compared to the alternate methods, the average classification accuracies of DAFTL have been obviously enhanced. It is worth

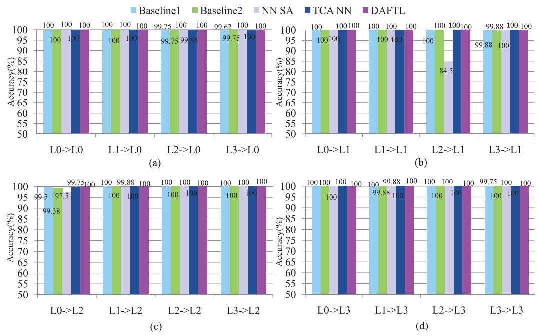


FIGURE 6. **The results with a fault size of 0.007in based on a sample frequency of 48 kHz.**

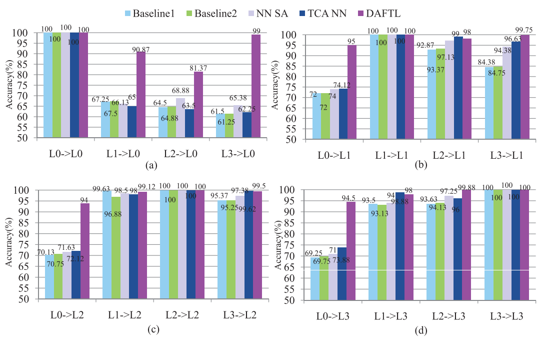


FIGURE 7. **The results with a fault size of 0.014in based on a sample frequency of 48 kHz.**

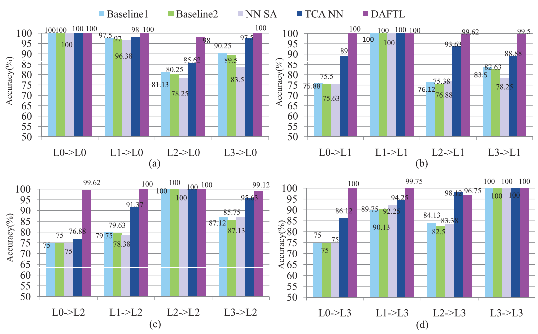


FIGURE 8. **The results with a fault size of 0.021in based on a sample frequency of 48 kHz.**

mentioning that DAFTL can be applied into the dataset from the Drive End Bearing Fault Data and sampled at a frequency of 48 kHz, which is rarely considered in other literature; the performances of the DAFTL method are clearly superior to the alternative methods shown in Figures 6, 7 and 8, when the experimental setup is the same as Case 1.

These results are all obtained from the benchmark datasets from fault diagnosis research under a relatively fair experiment condition, and therefore, we can conclude that DAFTL has very good potential for solving the accuracy-drop problem caused by variable working conditions in the field of fault classification.

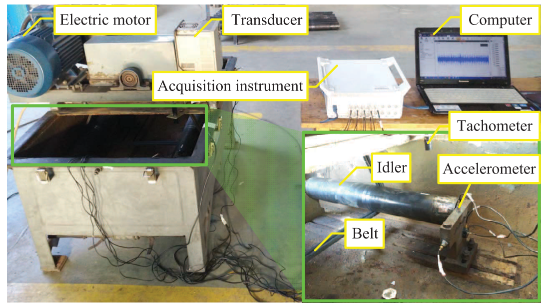


FIGURE 9. **Fault simulation testbed of the belt conveyor idler.**

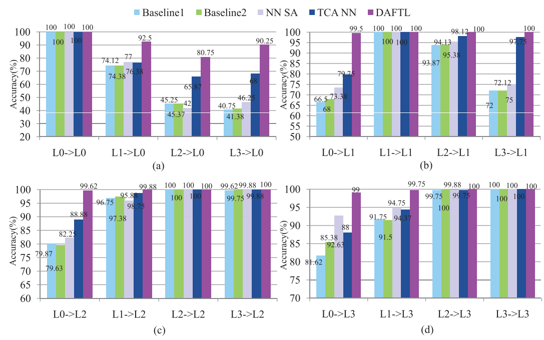


FIGURE 10. **The results on the fault simulation testbed of the belt conveyor idler.**

B. CASE 2: FAULT DIAGNOSIS BASED ON DATASET B

1) EXPERIMENTAL SETUP AND DATASET PREPARATION

Dataset B includes the raw vibration data of rolling element bearings obtained from fault simulation testbed of the belt conveyor idler illustrated in Figure 9 at a sampling frequency of 20 kHz. The testbed mainly includes an electric motor for driving, a transducer, a belt, an idler, a tachometer, accelerometers, an acquisition instrument and a computer. The driving motor is controlled by a transducer with a fixed load and synchronized with a belt, and the idler is driven through the intermediate belt. One of the bearings without defects is located in the bearing housing installed into the idler closer to the motor. The other bearing is located in the bearing housing installed into the idler farther to the motor, and it could be replaced by the test bearings. Finally the vibration signals of dataset B collected by the accelerometers at a sampling frequency of 20 kHz are used to diagnose faults.

In order to develop the proposed fault diagnosis method, four fault types for bearings with the same fault size rotor bearing are considered, i.e., normal (NO), inner race fault (IF), out race fault (OF) and ball fault (BF), and each fault type contains working conditions of L0 $=$ 300 rpm, L1 $=$ 600 rpm, L2 $=$ 900 rpm and L3 $=$ 1080 rpm, respectively. In this experiment, each sample contains 2049 data points generated from the raw vibration signal by implementing FFT in each working condition. Each domain contains four fault types, and each fault type contains 200 samples. Ultimately, $k = 100$ and $\lambda = 0.1$ for diagnosis.

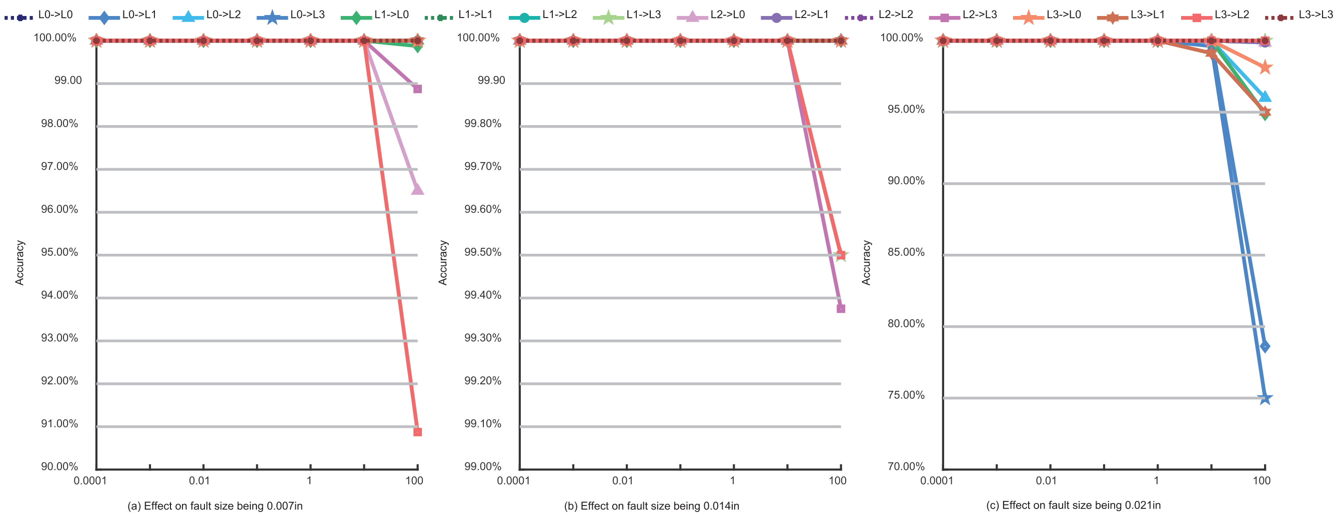


FIGURE 11. Classification accuracy (%) on different λ.

TABLE 3. Description of the experimental scenario on dataset B.

Target ♯ of test	Diagnose unlabeled test samples in target domain		
	Labeled source domain	Unlabeled target domain	Fault type
1	L0,L1,L2,L3	L0	IF,BF, OF,NO
2	L0,L1,L2,L3	L1	IF,BF, OF,NO
3	L0,L1,L2,L3	L2	IF,BF, OF,NO
4	L0,L1,L2,L3	L3	IF,BF, OF,NO

For the purpose of demonstrating the effectiveness of DAFTL, the compared methods of a-d are conducted simultaneously. The scenario settings of all experiments are trained by labeled samples collected in a certain working condition to classify the unlabeled test samples in another working condition. In all, 16 different transferring tests are carried out, and the details of the experimental scenario are described in Table 3.

2) PERFORMANCE EVALUATION UNDER VARIABLE WORKING CONDITIONS

The diagnostic results of our approach and the compared methods are illustrated in Figure 10.

Figure 10 includes three subfigures, and the target domains are ordered clockwise from the top left: L0, L1, L2 and L3 in turn. The symbol "$->$" in every part has the same meaning as Case 1. The variable working conditions are reflected in the speed differences between source domain and target domain. For example, in Figure 10(a), the target domain is L0 (the speed is 300 rpm), and the source domains are L0 (the speed is 300 rpm), L1 (the speed is 600 rpm), L2 (the speed is 900 rpm) and L3 (the speed is 1080 rpm). From these results in Figure 10, it is observed that DAFTL outperforms the compared methods obviously. The highest accuracy rates for fault diagnosis can be achieved when the training data are the same as the test data. The performance of Baseline2 is slightly better than the performance of Baseline1. However, baseline methods with no adaptation seem to fail the diagnosis under variable working conditions. For example, when transferring from L0 and L3 to L1 in Figure 10(b), the recognition rates of baseline methods can only reach about 70%. Specifically, in Figure 10(a), the classification accuracy rates of baseline methods cannot reach 50% when transferring from L2 and L3 to L0. NN SA and TCA NN are both obviously superior to the baseline methods. Unfortunately, when transferring from L2 and L3 to L0 using NN SA and TCA NN successively, the accuracies of NN SA and TCA NN are only about 45% in Figure 10(a), and similar results also are obtained in Figure 10(b). These results indicate that the contrast methods are not applicable for variable working conditions.

From Figure 10, we note that DAFTL consistently achieves favorable classification accuracies. It is worth mentioning that the classification accuracies can reach 80.75% and 90.25%, respectively, even when transferring from L2 and L3 to L0 in Figure 10(a). Overall, the results reveal that DAFTL can solve the performance degradation problem caused by the divergences between two domains under variable working conditions.

C. PARAMETER SENSITIVITY

In order to validate that DATFL can achieve optimal performance under a wide range of parameters, we investigate the influence of the parameter λ. In theory, the fact that larger values of λ can produce shrinkage regularization is more important to our work. When $\lambda \to 0$, the optimization problem is ill-defined, and when $\lambda \to \infty$, DAFTL will not be performed, that is, the transferable feature representation for training data and test data will not be constructed. Different λ values have different effects on classification accuracy. Figure 11 reports

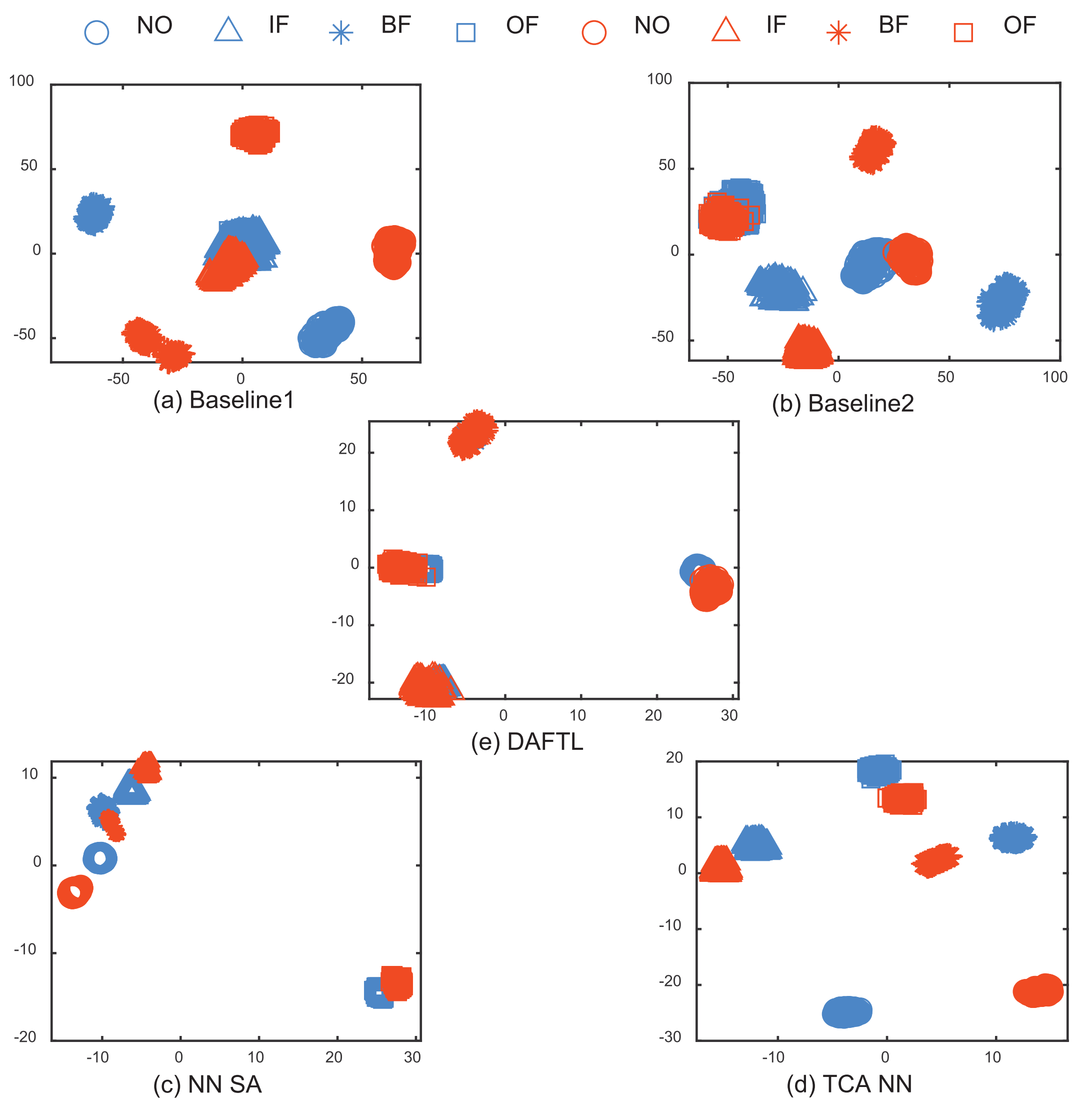


FIGURE 12. Feature visualization via t-SNE [27] over a fault diagnosis task from training domain L1(red) to test domain L2(blue) under variable working conditions. (a) Baseline1. (b) Baseline2. (c) NN SA. (d) TCA NN. (e) DAFTL.

the results. From the Figure 11, it is obvious that diagnostic results with a fault size of 0.021in are influenced to a large extent by different λ, and it has little overall effect on the performance with a fault size of 0.014in. Particularly noteworthy is the fact that the diagnosis is rarely affected by λ when the source domain and target domain are the same, and $\lambda \in [0.0001, 1]$ can be optimal parameter values, which indicates that DAFTL can achieve stable and excellent performance values under a wide range of parameters.

D. DOMAIN DISCREPANCY EFFECT OF EMPIRICAL ANALYSIS

The distribution differences between training data and test data intrinsically reflect the structure differences of data sampled from sensors under variable working conditions, and these discrepancies directly lead to performance degradation of fault classification. Thus, excavating and extracting transferable features between two domains is the key point for fault diagnosis under variable working conditions. In order to illustrate the transferability of the DAFTL features and explain why DAFTL works, we follow the t-SNE technique [27] to visualize high-dimensional features of the mentioned methods in our experiment in a two-dimensional map.

In all of the aforementioned cases, we select benchmark data (Dataset A) and take the transferring test that transfers L1 to L2 with a fault size of 0.007in as an example in Figure 12 for the analysis domain discrepancy effects under variable working conditions.

From the data in Figure 12, it is observable that DAFTL can make the distributions with the same fault types between training data and test data much closer than the compared methods. The transferable features extracted from DAFTL are of preferable divisibility and clustering than the results from the other compared methods. These results verify that

DAFTL can determine a common feature space, and the extracted features in this space are more robust for the two domains, which explains the superior adaptation performance of DAFTL under variable working conditions. The test samples can be discriminated significantly with the NN classifier built in the source domain by using DAFTL features.

E. DISCUSSION

In many actual fault diagnosis and classification scenarios, the distribution of training data is different from the test data, which leads to performance degradation for fault classification. In fact, the distribution differences between two domains reflect the differences between the data structures; thus, extracting intrinsic features from data structures is very important. DAFTL provides a way of domain adaptation using transfer learning to extract fault features and classify fault types. There are still several remarks that need to be described

(1) In this work, we present a new idea that uses domain adaptation to conduct bearing fault diagnosis under variable working conditions. Li *et al.* [25] proposed a spectrum images method, which applied two-dimensional principal component analysis (2DPCA) into the dimension reduction of the spectrum images of vibration signals, and overall highly accurate values were obtained. Unfortunately, there are still several instances that have lower accuracies. To solve this problem, we apply the domain adaptation with feature transfer learning into this field. Finally, the accuracies can all reach 100% on dataset A. In this paper, we highlight the results on data sampled at a frequency of 48 kHz. Compared with methods in case 1 and case 2, our method has absolute advantages.

(2) The results from the two diagnosis cases indicate that DAFTL is able to effectively classify mechanical health conditions under variable working conditions. Zhang *et al.* [26] proposed Deep Convolutional Neural Networks with Wide First-layer Kernel (WDCNN) and AdaBN to diagnose three datasets. Three datasets contain 10 types of health conditions (BF IF OF with fault sizes of 0.007 in, 0.014 in and 0.021 in) under three load conditions (Load1, Load2, Load3), respectively. The method in [26] obtains average accuracy, 95.9%, whereas the average accuracy of DAFTL is 100%. The main reason for this finding is that the distributions of training data and test data are very close after the use of DAFTL.

V. CONCLUSION

In this paper, a novel fault diagnosis approach for variable working conditions based on domain adaptation using feature transfer learning has been proposed. Transferable features for training data and test data were obtained by reducing the discrepancy between two domains while strengthening the recognizable information via domain invariant clustering, and it is unsupervised. The proposed method provides a novel perspective for solving the performance degradation problem of fault classification under variable working conditions. Different experimental tests under variable working conditions demonstrated the effectiveness and feasibility of the proposed method.

REFERENCES

[1] C. Bianchini, F. Immovilli, M. Cocconcelli, R. Rubini, and A. Bellini, "Fault detection of linear bearings in brushless AC linear motors by vibration analysis," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1684–1694, May 2011.

[2] K. A. Loparo, M. L. Adams, W. Lin, M. F. Abdel-Magied, and N. Afshari, "Fault detection and diagnosis of rotating machinery," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 1005–1014, Oct. 2000.

[3] X. Liu, L. Bo, X. He, and M. Veidt, "Application of correlation matching for automatic bearing fault diagnosis," *J. Sound Vib.*, vol. 331, no. 26, pp. 5838–5852, Dec. 2012.

[4] A. K. Jalan and A. R. Mohanty, "Model based fault diagnosis of a rotor–bearing system for misalignment and unbalance under steady-state condition," *J. Sound Vib.*, vol. 327, nos. 3–5, pp. 604–622, Nov. 2009.

[5] Y. Li, M. Xu, Y. Wei, and W. Huang, "A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree," *Measurement*, vol. 77, pp. 80–94, Jan. 2016.

[6] Y. Cheng, B. Zhou, C. Lu, and C. Yang, "Fault diagnosis for rolling bearings under variable conditions based on visual cognition," *Materials*, vol. 10, no. 6, p. 582, May 2017.

[7] M. Misra, H. H. Yue, S. J. Qin, and C. Ling, "Multivariate process monitoring and fault diagnosis by multi-scale PCA," *Comput. Chem. Eng.*, vol. 26, no. 9, pp. 1281–1293, Sep. 2002.

[8] A. Widodo and B.-S. Yang, "Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 241–250, 2007.

[9] J. Huang, X. Hu, and F. Yang, "Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker," *Measurement*, vol. 44, no. 6, pp. 1018–1027, Jul. 2011.

[10] 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.

[11] N. Saravanan and K. I. Ramachandran, "Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN)," *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4168–4181, Jun. 2010.

[12] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa, "Visual domain adaptation: A survey of recent advances," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 53–69, May 2015.

[13] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[14] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 2960–2967.

[15] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, Mar. 2017.

[16] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 999–1006.

[17] W. Li, L. Duan, D. Xu, and I. W. Tsang, "Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1134–1148, Jun. 2014.

[18] J. Tahmoresnezhad and S. Hashemi, "Visual domain adaptation via transfer feature learning," *Knowl. Inf. Syst.*, vol. 50, no. 2, pp. 585–605, Feb. 2017.

[19] Z. Xiao, H. Wang, and J. Zhou, "Robust dynamic process monitoring based on sparse representation preserving embedding," *J. Process Control*, vol. 40, pp. 119–133, Apr. 2016.

[20] 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. (ICCV)*, Dec. 2013, pp. 2200–2207.

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

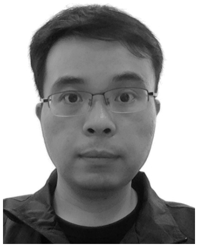
[22] *Case Western Reserve University Bearings Vibration Dataset*. Accessed: Oct. 2015. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home

[23] 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.

[24] H. Al-Bugharbee and I. Trendafilova, "A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling," *J. Sound Vib.*, vol. 369, pp. 246–265, May 2016.
[25] W. Li, M. Qiu, Z. Zhu, B. Wu, and G. Zhou, "Bearing fault diagnosis based on spectrum images of vibration signals," *Meas. Sci. Technol.*, vol. 27, no. 3, p. 035005, Feb. 2016.
[26] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017.
[27] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Feb. 2017.

ZHE TONG is currently pursuing the Ph.D. degree in mechatronic engineering with the School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou, China. His current research interests include intelligent fault diagnosis of rotating machinery and condition monitoring.

WEI LI received the Ph.D. degree from the University of Duisburg-Essen, Duisburg, Germany, in 2009. He is currently a Professor in mechatronic engineering with the School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou, China. His current research interests include fault diagnosis, remaining useful life prediction of rotating machinery.



BO ZHANG received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2014. He is currently a Lecturer with the School of Computer Science and Technology, China University of Mining and Technology. His research interests include machine learning, transfer learning, and fault diagnosis.

FAN JIANG received the Ph.D. degree from the China University of Mining and Technology, Xuzhou, China, in 2015. His current research interests include condition monitoring and fault diagnosis.

GONGBO ZHOU received the Ph.D. degree from the China University of Mining and Technology, Xuzhou, China, in 2010. His current research interests include wireless sensor networks, condition monitoring, and fault diagnosis.