An Intelligent Fault Diagnosis Method Based on Domain Adaptation and Its Application for Bearings under Polytropic Working Conditions

Zihao Lei, Guangrui Wen, Shuzhi Dong, Xin Huang, Haoxuan Zhou, Zhifen Zhang, Xuefeng Chen, *Senior Member, IEEE*

***Abstract*—In engineering practice, mechanical equipment is usually in polytropic working conditions, where the data distribution of training set and test set is inconsistent, resulting in insufficient generalization ability of the intelligent diagnosis model. Simultaneously, different tasks often need to be modeled separately. Domain adaptation, as one of the research contents of transfer learning, has certain advantages in solving the problem of inconsistent feature distribution. This paper designs and establishes a Domain Adaptation Framework based on Multi-scale Mixed Domain Feature (DA-MMDF) for cross-domain intelligent fault diagnosis of rolling bearings under polytropic working conditions. The proposed method first uses the multi-scale mixed domain feature extractor to obtain features from the collected data, which constructs a complete feature space through Variational Mode Decomposition (VMD) and mixed domain feature extraction to fully mine the state information and intrinsic attributes of the vibration signal. Secondly, the dimensionality reduction and optimization of features are achieved through extreme gradient promotion, and meaningful and sensitive features are selected according to the importance of features to eliminate redundant information. The optimized important features are combined with the manifold embedded distribution alignment method to realize the distribution alignment of data in different fields and cross-domain diagnosis. In order to verify the effectiveness of the proposed approach, the rolling bearing data sets gathered from the laboratories are employed and analyzed. The analysis result confirms that DA-MMDF is able to achieve effective transfer diagnosis between polytropic working conditions. Compared with traditional intelligent fault diagnosis methods and domain adaptation methods, the method proposed in this article achieved state-of-the-art performances.**

***Index Terms*—Multi-scale mixed domain feature, XGBoost, domain adaptation, rolling bearings, intelligent fault diagnosis.**

I. Introduction

In recent years, with the continuous improvement of the industrial level, high-end equipment such as aero engines, wind turbines, and CNC machine tools are developing in complexity and large-scale. As a key component of rotating machinery, bearings are being widely used in ships, automobiles, aerospace and other fields[1]. On account of bearings constantly working under harsh and changeable conditions like high temperature, high pressure, and heavy load, their reliability is greatly reduced. If measures are not taken in time, serious accidents are likely to take place. Even worse, huge economic losses and casualties will occur[2]. In order to avoid major economic losses and catastrophic failures, it is of great significance to study advanced intelligent fault diagnosis algorithms[3].

Due to the abundant information contained in the vibration signal, the current fault diagnosis methods based on the vibration signal have become the focus of research. Generally speaking, the fault diagnosis of the bearing is realized in three different ways: methods based on signal processing, methods based on feature extraction combined with pattern recognition, and fault diagnosis methods based on deep learning. Relevant research progresses are described as follows:

(1) The method based on signal processing mainly relies on rules and experience, and the filter-based analysis algorithm aims to find the optimal center frequency and bandwidth[4-6]. The algorithm based on parametric time-frequency analysis mainly includes Ensemble Empirical Mode Decomposition (EMD) and Synchronous Wavelet Transform (SWT)[7-9]. Algorithms based on wavelet analysis aim to find wavelet basis functions matching the target pulse[10-12]. The algorithm based on sparse representation aims to design a suitable sparse dictionary and mine low-dimensional prior knowledge[13-15]. And mathematical morphological filtering provides a nonlinear filtering approach. Li et al. proposed an improved time-varying morphological filtering (TMF), which can extract fault features of defective rolling element bearings with high computational efficiency and has been well supported by experiments[16]. Variational Mode Decomposition (VMD), as a recently proposed adaptive decomposition method[17], has been playing an important role in the signal processing and feature extraction of rolling bearings. On this basis, relevant scholars have improved and optimized it, and applied it to the field of fault diagnosis[18, 19]. Although the above methods have sound physical explanations, they often face the following problems when they are being used alone. On the one

This research was supported in part by The National Key Research and Development Program of China (No. 2020YFB1710002), National Natural Science Foundation of China (No. 51775409), and Equipment Pre-research Fund of China (No. 61420030301).

The authors are with School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China. Guangrui Wen is also with the Key Laboratory of Education Ministry for Modern Design & Rotor-Bearing System, Xi'an, Shaanxi 710049, China. Corresponding author: Guangrui Wen, e-mail: grwen@mail.xjtu.edu.cn.

hand, it is difficult to obtain a feature set that can fully characterize the signal. On the other hand, this kind of method is only used for fault diagnosis of a small number of bearings under some specific conditions, which is subject to poor adaptability and low automation level.

(2) The method based on feature extraction combined with pattern recognition mainly includes three steps, data acquisition, feature construction and pattern recognition. The key step is the construction of feature engineering, since the quality of feature set will directly determine the accuracy of fault diagnosis. At present, relevant scholars have carried out research on the extraction of statistical parameters in time domain, frequency domain and time-frequency domain on the basis of signal processing methods, and achieved certain results[20, 21]. However, when the mechanical system has coupling components and is exposed in a complex environment of strong background noise, the system often appears in a nonlinear behavior. At this point, the statistical feature parameter extraction based on signal processing is unlikely to accurately identify the fault features. While as a complexity measure, entropy index has been widely used in time series analysis[22], and the application of entropy index in fault diagnosis has been widely studied[23-25]. According to the literature being reviewed, few studies unify the advanced signal processing methods with the extraction of statistical parameters or entropy index in time domain, frequency domain and time-frequency domain, so as to construct a complete multi-scale mixed domain feature set to represent the internal attributes of the system.

(3) Due to the rapid development of artificial intelligence, deep learning algorithm is becoming more and more popular and is gradually applied in the field of fault diagnosis[26-28]. Although deep learning can learn and extract features adaptively, there are still application paradigms of Data Driven & Result Oriented, and most of the methods are to adjust parameters to improve the accuracy. In addition, the research on deep learning itself is still in the Black Box State, and the physical meaning of the deep network itself and the interpretability of deep features still need to be further advanced.

With the development of fault diagnosis technology, this process often involves more feature dimensions. It is worth noting that after obtaining high-dimensional feature sets, How to select features that are useful for diagnosis will become very significant. Because some irrelevant features not only consume computing resources but also have negative effects on fault diagnosis[29]. Some scholars have explored many feature selection techniques, including Principal Component Analysis (PCA), Euclidean Distance (ED) and Isometric Mapping (ISOMAP) methods and so on[30-32]. Nevertheless, these methods have some limitations, where they cannot refine the importance of each original feature or select the feature subset sensitive to diagnosis. XGBoost, as a new machine learning algorithm has been widely used in engineering field. The algorithm has the advantages of high precision, fast running speed and strong generalization ability. In order to improve the modeling efficiency in the training process, the algorithm will evaluate the effect and give a score to each feature according to different gain in each iteration[33].

Although the intelligent fault diagnosis methods have achieved great success, these successes are based on a common assumption, where there are enough labeled data to train the intelligent diagnosis model[34]. However, in engineering practice, the above assumptions are often not satisfying. Transfer learning provides an idea to solve the above problems by applying the knowledge learned from one domain to another[35]. The related methods are TCA[36], JDA[37], JGSA[38] and MEDA[39]. With the development of deep learning, deep models are used to automatically learn transferable features from different domains. Some achievements have been made in the field of computer vision, such as Dynamic Adversarial Adaptation Network (DAAN) [40] and Deep Subdomain Adaptation Network (DSAN)[41]. In the field of fault diagnosis and transfer learning, some scholars have carried out relevant research[42, 43]. Dong et al.[44] used wavelet packet decomposition to construct feature space, and then JGSA was used for cross-domain feature mapping and to reduce distribution differences adaptively. Wen et al. used Sparse AutoEncoder to extract spectrum features of bearings under different working conditions, and then MMD was used to learn the transferable feature distribution[45]. Lei et al. introduced the domain adaptive module into deep learning, which can effectively learn transferable features[46]. However, compared with the deep transfer model, the shallow transfer model is easier to understand the process of knowledge-based migration from the original feature space to the target domain, reducing the number of training samples, and avoiding the training and optimization of deep network. More importantly, the shallow transfer model can select the feature set to be migrated after sorting the feature importance according to the above method, so as to obtain the fault sensitive subset and avoid *Negative Transfer* to a certain extent.

In summary, considering that the latest existing methods cannot solve the completeness, interpretability and transferability of features at the same time, a Domain Adaptation method based on Multi-scale Mixed Domain Feature (DA-MMDF) is proposed in this paper, which is innovative in theories based on knowledge transfer and cross-domain applications. This method overcomes the non-completeness of the previous methods, which either extract features from one domain or a few domains or extract features from a single scale. The proposed method first designs a multi-scale mixed domain feature extractor. Compared with the existing methods, the innovation is that it can achieve complete feature extraction, especially for nonlinear systems with excellent feature extraction capabilities. Secondly, although existing deep learning methods can also extract features adaptively, they are always a black box model and lack of interpretability of features. However, the GS_XGBoost module of the proposed method can obtain the importance of features and sort them, and it can obtain fault-sensitive features easily, at the same time, it has high computational efficiency. Finally, the domain

adaptive module can realize cross-domain fault diagnosis through feature mapping and reduce distribution differences adaptively. Compared with the latest existing traditional transfer learning methods, it can effectively avoid negative transfer and realize knowledge transfer. The contribution of this paper can be summarized as follows:

1) A Multi-scale Mixed Domain Feature extractor is proposed for the first time. The feature extractor integrates an advanced signal processing method called VMD, and the characteristic parameters such as time domain, frequency domain, time-frequency domain and entropy index. Compared with previous methods, it can characterize the signal more completely and exert strong nonlinear extraction ability in complex environment.

2) A feature optimizer based on GS_XGBoost is proposed, which not only reduces the dimension of features and eliminates redundant information, but also gives the importance ranking of features to understand the contribution of different features for diagnosis and recognition, and then selects sensitive feature subsets to diagnosis and gives its physical significance.

3) A domain adaptive intelligent fault diagnosis framework (DA-MMDF) is proposed based on the Manifold Embedding Distributed Alignment. The completeness, interpretability and transferability of features are considered simultaneously to solve the problem of cross-domain diagnosis and knowledge-based transfer under different working conditions.

4) Experiments under variable speed and variable load are designed and verified by the proposed framework. Compared with traditional intelligent diagnosis methods and transfer learning methods, the effectiveness and superiority of the framework are proved.

The rest of this paper is arranged as follows. Section II briefly reviews the theoretical background of the proposed method, and section III introduces the proposed framework in detail. In Section IV, the proposed method is used to verify and analyze three cases. Finally, in Section V, the conclusion of this paper and the future work are presented.

II. Theoretical Foundation

A. Variational Mode Decomposition(VMD)

As an adaptive and quasi-orthogonal signal decomposition method, VMD aims to decompose the signal into sparse Intrinsic Mode Function (IMF), and its essence is equivalent to generalizing the Wiener filter as an adaptive frequency band with narrow-band characteristics. It overcomes the problems of end effect and mode component aliasing in EMD method, and has a more solid mathematical theoretical foundation. It can reduce the non-stationarity of time series with high complexity and strong nonlinearity, and decompose to obtain relatively stable subsequences with different frequency scales. Its adaptability is manifested in determining the number of mode decompositions of a given sequence according to the actual situation. During the subsequent search and solution process, it can adaptively match the best center frequency and limited bandwidth of each mode. Therefore, signal decomposition based on VMD can be regarded as a constrained variational problem, and the adaptive decomposition of the signal can be realized by searching for the optimal solution of the constrained variational model. By searching for the optimal solution of the variational mode and updating the mode functions and central frequencies, several mode functions IMFs with certain broadband are obtained. The specific process is shown in Eq. (1) - Eq. (5). The constrained variational model is as follows:

$$\begin{cases} \min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} \sum_k u_k = f \end{cases} \tag{1}$$

where u_k is the decomposed IMF, and ω_k is the center frequency of the IMF. In order to solve the constrained variational problem effectively, the quadratic penalty parameter and the Lagrangian multiplication operator are introduced. Therefore, the constrained variational problem can be rearranged in the form of Eq. (2):

$$\begin{aligned} L(\{u_k\},\{\omega_k\},\lambda) = \; & \alpha \sum_k \| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right)^* u_k(t) \right] e^{-j\omega_k t} \|_2^2 \\ & + \| f(t) - \sum_k u_k(t) \|_2^2 \\ & + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \tag{2}$$

where $f(t)$ is the original signal. Eq. (2) can be solved by an alternate direction multiplier algorithm, which converts the original problem into an alternate update of the IMF and the corresponding center frequency.

$$\begin{cases} \hat{u}_k^{n+1}(\omega) = \dfrac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \dfrac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \\ \omega_k^{n+1} = \dfrac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \end{cases} \tag{3}$$

When the IMF and the center frequency are updated, the Lagrangian multiplication operator is also updated.

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n-1}(\omega) \right) \tag{4}$$

The updating process is repeated until the convergence criterion of "(5)" is met. The value of ε is usually set 10^{-6}.

$$\sum_k \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2 / \left\| \hat{u}_k^n \right\|_2^2 < \varepsilon \tag{5}$$

B. Feature Extraction

When the equipment fails, relevant indexes will change simultaneously including the amplitude and probability distribution in time domain, the energy of different frequencies in frequency domain and the position of the main energy spectrum front, the structural distribution of the time-frequency energy, the internal distribution characteristics and complexity of the signal, and the disorder degree of the system. In this section, a feature extractor for multi-scale mixed domains is proposed. In order to fully

TABLE I
THE EXPRESSION OF TIME-DOMAIN FEATURE PARAMETERS.

Feature expression	Feature expression	Feature expression
$F_1=\frac{1}{N}\sum_{i=1}^{N}x_i$	$F_6=\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert$	$F_{11}=F_8/F_7$
$F_2=\frac{1}{N}\sum_{i=1}^{N}x_i^2$	$F_7=\sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}$	$F_{12}=F_8/F_6$
$F_3=\frac{1}{N}\sum_{i=1}^{N}(x_i-F_1)^2$	$F_8=\max\{\lvert x_i\rvert\}$	$F_{13}=F_8/F_5$
$F_4=\frac{1}{N}\sum_{i=1}^{N}(x_i-F_1)^2$	$F_9=\lvert\max(x_i)-\min(x_i)\rvert$	$F_{14}=E\left[\left(\frac{x_i-\mu}{\sigma}\right)^3\right]$
$F_5=\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x_i\rvert}\right)^2$	$F_{10}=F_7/F_6$	$F_{15}=\frac{1}{N}\sum_{i=1}^{N}\frac{(x_i-\mu)^4}{\sigma^4}$

characterize the vibration signal of the bearing, we mainly extract features from time domain, frequency domain, time-frequency domain, complexity domain, combining with VMD to further extract information-measured features and the above-mentioned mixed domain features.

We extracted 15-time domain features by statistical method, whose expressions are shown in Table I, where $x_i, i=1,2,\cdots,N$, and N is the number of data points of the signal. As shown in Table I, 9 dimensional statistical parameters called F1~F9 respectively represent mean value, mean square value, variance, standard deviation, root amplitude, mean amplitude, mean square amplitude, peak value, peak-to-peak value. 6 dimensionless statistical parameters called F10~F15 are separately represent waveform indicator, peak indicator, impulse indicator, margin indicator, skewness and kurtosis.

In the frequency domain, 12 frequency domain characteristic parameters[47] and 7 power spectral density parameters[48] are extracted, which respectively reflect the magnitude of frequency domain vibration energy and the dispersion or concentration degree of frequency spectrum.

In addition, when the power of each frequency component of vibration signal changes, the center of gravity position and energy distribution of power spectrum will also change. Among them, barycenter frequency and mean square frequency represent the change of barycenter position of power spectrum, and frequency variance represents the dispersion degree of power spectrum energy distribution. Therefore, the barycenter frequency(FC), the mean square frequency(MSF) and the frequency variance(VF) are extracted respectively, and their calculation formulas are as follows:

$$FC=\frac{\int_0^{\pi}\omega * S(\omega)d\omega}{\int_0^{\pi}S(\omega)d\omega} \tag{6}$$

$$\mathrm{MSF}=\frac{\int_0^{\pi}f^2S(\omega)d\omega}{\int_0^{\pi}S(\omega)d\omega} \tag{7}$$

$$VF=MSF-FC^2 \tag{8}$$

where $S(\omega)$ is the power spectrum of the signal and is the angular frequency of the signal.

In the time-frequency domain, 8 continuous wavelet energy spectrum features are extracted, which are expressed as follows:

$$E_l=E_l/\sum_{i=1}^{n}E_i,\quad l\in[1,n] \tag{9}$$

where E_l is the wavelet energy spectrum of the original signal on *n* scales. When the scale is *a*, the wavelet energy spectrum is calculated as follows:

$$E(a)=\int_{-\infty}^{\infty}|W_f(a,b)|^2db \tag{10}$$

where $W_f(a,b)$ is the amplitude of wavelet transform.

Complexity measures mainly analyze the distribution characteristics and complexity of signals. Entropy, as a complexity measure, has been widely used in time series analysis. Usually, when a mechanical system fails, the system has nonlinear characteristics, which increases the complexity of system response. Entropy measure is just suitable for quantifying this dynamic change, so as to distinguish different system states. On such basis, this paper further extracts four complexity measures, namely permutation entropy[49], approximate entropy[50], sample entropy[51] and fuzzy entropy[52].

C. Extreme Gradient Boosting(XGBoost)

XGBoost algorithm is an improved version of the traditional Gradient Boosting Decision Tree algorithm. Its basic idea is to combine multiple models with lower accuracy to build a model with relatively higher accuracy. The model can be calculated iteratively, and each iteration generates a new model to fit the error of the previous model. This process is called gradient boosting. In short, the boosting algorithm first uses simple model to obtain a relatively general result, and then constantly adds simple models to the original model, thereby improving the accuracy of the model.

The XGBoost model is integrated by multiple CART regression trees, and the algorithm principle is derived as follows:

Assuming that the model has *t*-1 decision trees, the *t*-th decision tree can be expressed as Eq. (11)

$$\hat{y}_i^{(t)}=\sum_{k=1}^{t}f_k(x_i)=\hat{y}_i^{(t-1)}+f_t(x_i) \tag{11}$$

where t is the number of trees, f is a specific CART tree, $\hat{y}_i$ is the predicted value and (x_i) is the input sample. The objective optimization function is shown in Eq. (12).

$$\begin{aligned}Obj^{(t)}&=\sum_{i=1}^{n}l\left(y_i,\hat{y}_i^{(t)}\right)+\sum_{k=1}^{t}\Omega(f_k)=\\&\sum_{i=1}^{n}l\left(y_i,\hat{y}_i^{(t-1)}+f_t(x_i)\right)+\\&\Omega(f_t)+constant\end{aligned} \tag{12}$$

where l is the defined loss function, and $\Omega(f_t)$ is the complexity of the model, that is, the regular term. In general, if the Taylor series of the loss function is extended to order 2 and the constant term is removed, the objective function of the *t*-th step is shown as in Eq. (13).

$$Obj^{(t)}=\sum_{i=1}^{n}\left[g_i f_t(x_i)+\frac{1}{2}h_i f_t^2(x_i)\right]+\Omega(f_t) \tag{13}$$

$$g_i=\partial_{\hat{y}_i^{(t-1)}}l\left(y_i,\hat{y}_i^{(t-1)}\right) \tag{14}$$

$$h_i=\partial^2_{\hat{y}_i^{(t-1)}}l\left(y_i,\hat{y}_i^{(t-1)}\right) \tag{15}$$

where g_j and h_j are the first and second derivatives respectively, and the calculation formulas are shown in Eq. (14) and Eq. (15) separately.

For a decision tree, its complexity can be defined as Eq. (16).

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|\omega\|^2 \tag{16}$$

where T represents the number of leaf nodes, and ω is the score on the leaf nodes.

As a result, for the *t*-th decision tree, when a tree structure is determined the objective function can be calculated as follows:

$$\omega_j^* = -\frac{\sum_{i\in I_j} g_i}{\sum_{i\in I_j} h_i + \lambda} \tag{17}$$

$$Obj^* = -\frac{1}{2}\sum_{j=1}^{T}\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i + \lambda} + \gamma T \tag{18}$$

where I_j is the sample set on the *j*-th leaf node. Therefore, to complete the establishment of the *t*-th decision tree model, we only need to find the tree structure that minimizes the objective function.

XGBoost uses the information-gain algorithm similar to the decision tree to determine the tree structure. In another word, the information gain of all the segmentation nodes of all features is calculated according to Eq. (19) and bifurcates at the segmentation boundary point with the maximum information gain. This process can be calculated in parallel.

$$\text{Gain} = \frac{1}{2}\left[\frac{(\sum_{i\in I_L} g_i)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{(\sum_{i\in I_R} g_i)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{(\sum_{i\in I} g_i)^2}{\sum_{i\in I} h_i + \lambda}\right] - \gamma \tag{19}$$

where *I* is the set before splitting, I_L and I_R are the sets of left and right nodes after splitting. During the training process, the model will automatically calculate the total number of times used by each feature and take it as the importance index of the feature. In the general use process, the *total gain* is often used to sort the importance of features.

D. Manifold Embedded Distribution Alignment (MEDA)

Manifold Embedded Distribution Alignment is an adaptive distribution adaptation method, which solves the problem of how to quantitatively evaluate the importance of marginal(P) distribution and conditional distribution(Q) in transfer learning. Firstly, the features of the original feature space are mapped to the Grassmann manifold space by manifold feature transformation to reduce the data drift between domains. Secondly, the relative importance of marginal distribution and conditional distribution is dynamically measured by introducing adaptive factors. Finally, the classifier is learned by Structural Risk Minimization (SRM) principle.

Assuming $g(\cdot)$ is a function of manifold feature transformation, then can be represented as

$$\begin{aligned} f = \arg\min_{f\in\sum_{i=1}^{n}\mathcal{H}_K} \quad & \ell\left(f\left(g\left(\mathbf{x}_i\right)\right), y_i\right) + \eta \| f \|_K^2 \\ & + \lambda \overline{D_f}\left(\mathcal{D}_s, \mathcal{D}_t\right) + \rho R_f\left(\mathcal{D}_s, \mathcal{D}_t\right) \end{aligned} \tag{20}$$

where $\| f \|_K^2$ is the squared norm of f, and $\overline{D_f}$ is the dynamic distribution alignment.

In addition, R_f is regarded as Laplacian regularization to take advantage of the similar geometric property of the nearest points in the manifold. η, λ and ρ are regularization parameters respectively.

When mapping manifold features, the subspace is first embedded into the manifold G, then each *d*-dimensional subspace of the original feature can be regarded as a point on G. Let $\mathcal{S}_s = \Phi(0)$ and $\mathcal{S}_t = \Phi(1)$, then find a geodesic from $\Phi(0)$ to $\Phi(1)$ to map the original feature space into the manifold feature space[53].

$$\left\langle \mathbf{z}_i, \mathbf{z}_j \right\rangle = \int_0^1 \left(\Phi(t)^T \mathbf{x}_i\right)^T \left(\Phi(t)^T \mathbf{x}_j\right) dt = \mathbf{x}_i^T \mathbf{G} \mathbf{x}_j \tag{21}$$

Then the original feature space can be transformed into Grassmann manifold with $z = g(\mathrm{x}) = \sqrt{G}x$.

Next, an adaptive factor is introduced to measure the importance of marginal distribution and conditional distribution.

$$\overline{D_f}\left(\mathcal{D}_s, \mathcal{D}_t\right) = (1-\mu) D_f\left(P_s, P_t\right) + \mu \sum_{c=1}^{C} D_f^{(c)}\left(Q_s, Q_t\right) \tag{22}$$

where μ is the adaptive factor, $c \in \{1, \cdots, C\}$ is the category. $D_f\left(P_s, P_t\right)$ is the marginal distribution adaptation, and $D_f^{(c)}\left(Q_s, Q_t\right)$ is the conditional distribution adaptation for class C.

With the help of Maximum Mean Difference (MMD) to calculate the distribution difference, adaptive distribution adaptation can be expressed as follows:

$$\begin{aligned} \overline{D_f}\left(\mathcal{D}_s, \mathcal{D}_t\right) = \quad & (1-\mu) \| \mathbb{E}\left[f\left(\mathbf{z}_s\right)\right] - \mathbb{E}\left[f\left(\mathbf{z}_t\right)\right] \|_{\mathcal{H}_K}^2 \\ & + \mu \sum_{c=1}^{C} \| \mathbb{E}\left[f\left(\mathbf{z}_s^{(c)}\right)\right] - \mathbb{E}\left[f\left(\mathbf{z}_t^{(c)}\right)\right] \|_{\mathcal{H}_K}^2 \end{aligned} \tag{23}$$

The *A*-distance is used to estimate the distance of different distributions. For the marginal distribution difference, the *A*-distance between the source domain and the target domain is calculated and recorded as A_M. For the conditional distribution difference, first cluster the target domain into category C, then calculate the difference between the source domain and the target domain from the same category of data A-distance, taking its mean value as A_c. Therefore, the adaptive factor can be estimated as follows:

$$\hat{\mu} \approx A_M / \left(A_M + \sum_{c=1}^{C} A_c\right) \tag{24}$$

Finally, the classifier on the feature space is learnt after the manifold transformation based on the principle of minimizing institutional risk.

III. The Proposed Architecture

In order to improve the fault diagnosis accuracy of rolling bearing under variable working conditions, a machine fault diagnosis pipeline (DA-MMDF) is proposed in this section. Its flow chart is shown in Fig.2, which is mainly divided into three steps (Multi-scale Mixed Domain Feature Extraction, Feature Optimization and Domain Adaptation). The specific details are as follows.

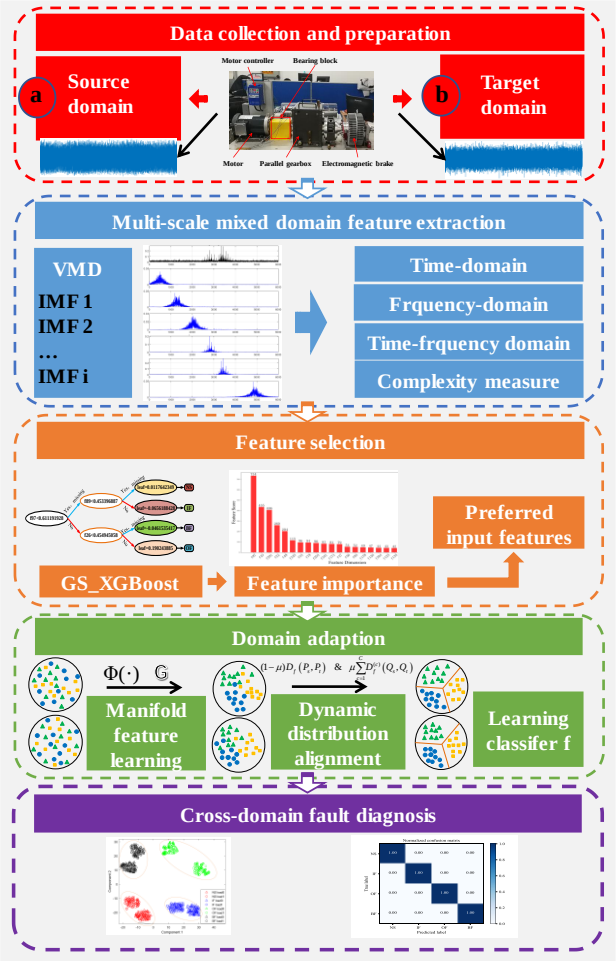


Fig. 1. Architecture of the proposed DA-MMDF model.

Step 1: The acceleration sensor is used to collect the original vibration signals of the rolling bearing under different working conditions.

Step 2: According to Section II, the signal is firstly decomposed by VMD, the original signal and IMFs of each order are extracted in time domain, frequency domain, time-frequency domain and complexity measure respectively to construct a complete Multi-scale Mixed Domain feature set that can represent the signal state information and internal attributes, then the feature set is normalized by maximum-minimum normalization, where the relevant parameters of VMD are set according to experience in Table II.

Step 3: A XGBoost model is built and XGBoost parameters are optimized using GridSearchCV algorithm, where the optimal parameter combination is selected after cross-validation as shown in Table II. Then *total gain* is selected as the feature importance index in 5 common importance assessment indexes,

TABLE II
Parameters Settings For VMD And Optimized XGBoost Model By GridSearchCV.

VMD parameters	Value	XGBoost parameters	Value
alpha	2891	learning_rate	0.1
tau	0	max_depth	3
K	6	min_child_weight	1
DC	0	gamma	0.1
init	1	subsample	0.6
tol	1e-7	colsample_bytree	0.6
/	/	reg_alpha	1e-5
/	/	reg_lambda	1

TABLE III
MEDA Model Parameters Setting.

MEDA parameters	Setting
d	8
T	10
lambda	10
eta	0.1
rho	1.0
base	1NN

which represents the *total gain* brought by features in each split node. Then the importance of the above features is prioritized and optimized.

Step 4: Manifold feature transformation for the features optimized in Step 3 is performed so as to reduce data drift between different domains. Then, the marginal distribution and conditional distribution of the source domain and target domain are adapted dynamically. Finally, the classifier f is learned through the principle of SRM to realize the cross-domain diagnosis of rolling bearing under different working conditions, The parameters setting of MEDA are shown in Table III, where d represents the dimensionality after manifold feature transformation. T is the number of iterations, lambda, eta, and rho are regularization parameters, and the base classifier is 1NN.

IV. Experimental Verification

In this section, we first verify the effectiveness of the proposed framework in fault diagnosis under variable operating conditions through the motor bearing data set of the Case Western Reserve University (CWRU) laboratory. Secondly, the data set of planetary gear box in the laboratory of Xi'an Jiaotong University (XJTU) is further tested and verified. Compared with traditional intelligent diagnosis methods and traditional transfer learning algorithms, the superiority of the proposed framework is highlighted.

A. Variable Load Datasets from CWRU Lab

1) Data description

The motor bearing fault data set from the CWRU Lab is used as a verification case. The experimental platform is shown in Fig.3, which mainly consists of an induction motor, a torque transducer and a load motor. During testing each bearing was operated under four loads (0,1,2,3HP). Single-point defects of different degrees such as 0.007, 0.014, 0.021, 0.028 inches are set by using spark machining to simulate inner race fault, outer race fault and ball fault. The accelerometer is placed near the

drive end of the motor to measure the vibration signal. In the experiment, the sampling frequency is set to 12kHz.

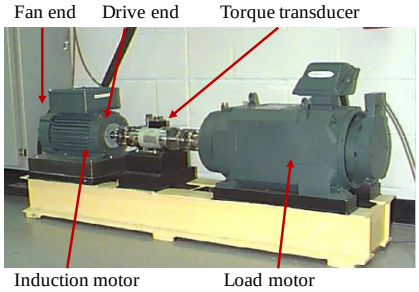


Fig. 2. Rolling bearing fault simulation test platform in CWRU.

In this section, the collected vibration signals under 4 different loads are used as analysis data sets A, B, C, and D. Each data set includes normal state(NS), inner race fault (IF), outer race fault (OF) and ball fault(BF). And there are a total of 400 samples in every dataset, of which 100 samples of each type, and each sample has 2400 sampling points. The single-point defect of the faulty bearing here is 0.028 inches, and the data of the 6 o'clock direction is used for the outer race fault. Therefore, 12 transfer learning tasks are set up, namely A→B, A→C, A→D, B→A, B→C, B→D, C→A, C→B, C→D , D→A, D→B and D→C. Data sets A, B, C, and D are respectively regarded as the source domain and target domain.

2) Experimental Results and Analysis

According to the steps in section III, the proposed framework is applied to the data set in Variable Load Datasets from CWRU Lab. Firstly, the mixed domain feature sets including time domain, frequency domain, time-frequency domain and complexity measures are extracted from the original vibration signal. Then, VMD was used to decompose the signal, and the signal after VMD decomposition is shown in Fig. 3. Taking a group of rolling bearing inner race fault data from CWRU Lab as an example, and VMD is carried out. The bearing fault size is 0.007 inches, the working speed is 1750 r/min, and the load is 3 HP, the sampling frequency is set to 12kHz, and the sampling number is 12000 points. The waveform of the original signal and the 6 order IMFs decomposed by VMD are shown in the left figure, and the spectrum of the original signal and the 6 order IMFs decomposed by VMD are shown in the right figure.

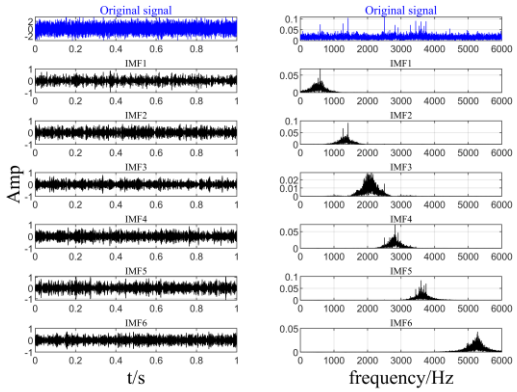
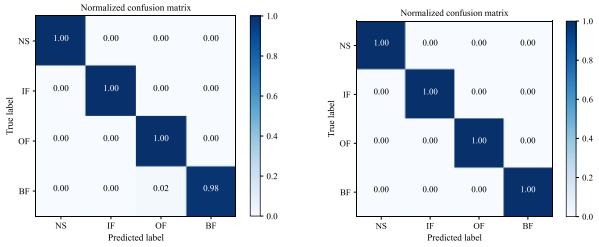


Fig. 3. Waveform and spectrum of original signal and IMFs for CWRU Lab's rolling bearing by VMD.

Among them, the original signal is blue, and each order of IMF is black. Next, the first 6 IMFs that contain useful information for research are selected, and the mixed domain feature sets are extracted from these 6 IMFs to form a multi-scale mixed domain feature set, and then the feature sets are normalized.

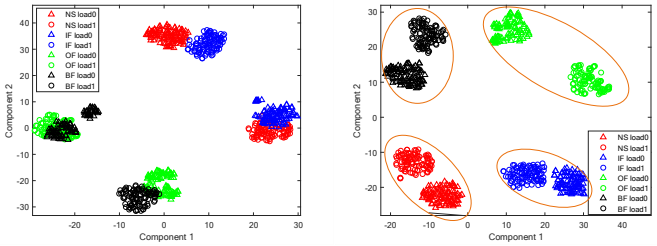
Moreover, the XGBoost model is built, and the candidate values of each parameter are set in advance, then the Grid Search algorithm is used to optimize the XGBoost model, namely GS_XGBoost.

According to the established GS_XGBoost model, total gain is selected as the evaluation index of feature importance. After training with XGBoost, the features will be sorted according to their importance and the preferred features will be obtained.



(a) Before domain adaptation (b) After domain adaptation

Fig. 4. Confusion Matrix of Proposed Method on CWRU dataset B.



(a) Before domain adaptation (b) After domain adaptation

Fig. 5. Visualization result of cross process domain adaptation on dataset A(source domain) and dataset B(target domain).

TABLE IV
ACCURACY(%) OF PROPOSED METHOD ON CWRU'S DATASETS.

Task	Dataset under different working loads			
	A	B	C	D
A	/	100	99.75	100
B	100	/	100	100
C	100	99.50	/	100
D	100	100	100	/

Then the preferred features are combined with MEDA to realize manifold feature transformation and dynamic distributed alignment, and finally the fault diagnosis across operating conditions is realized

Taking task A→B as an example, the accuracy of DA-MMDF is 100%. What's more, according to the confusion matrix shown in Fig.4, the proposed DA-MMDF can correctly classify NS, IF, OF and BF. In order to understand the process of transfer learning more intuitively. As can be seen from the confusion matrix in Fig.4(a) and Fig.4(b), domain adaptation can effectively improve the accuracy of cross-domain diagnosis. the t-Distribution Stochastic Neighbour Embedding(*t-sne*) algorithm[54] is introduced. This algorithm can reduce the dimension of features and directly draw the distribution of features. In addition, Fig.5(a) is the feature description before

domain adaptation. It can be found that although different types of fault samples can be well distinguished under the same working condition, the same type of fault can not be effectively distinguished under different working conditions. After domain adaptation, the distribution of the same type of fault samples under different working conditions is getting closer and closer, as shown in Fig.5(b). The proposed method is further verified under 12 variable load conditions in the CWRU data set, and the relevant results are shown in Table IV. The above results prove the effectiveness of the proposed method.

B. Variable Load Datasets from XJTU Lab

In order to further explore the generalization ability and superiority of the proposed method, we used the wind turbine drive train fault diagnosis simulator platform of Xi'an Jiaotong University (XJTU) to collect bearing vibration data under different loads and different speeds. Compared with traditional intelligent diagnosis methods and traditional transfer learning algorithms, the proposed method achieves state-of-the-art performances.

1) Data description

The fault simulation experiment was carried out on the Xi'an Jiaotong University (XJTU) experimental platform. The experimental device is shown in Fig.6, which mainly consists of a motor, a motor controller, a bearing block, a parallel gear box, and an electromagnetic brake. Fig.7 shows the three fault modes of bearings, including IF, OF and BF. The IF and OF are damaged by laser burns. The damage diameter is about 2mm, and

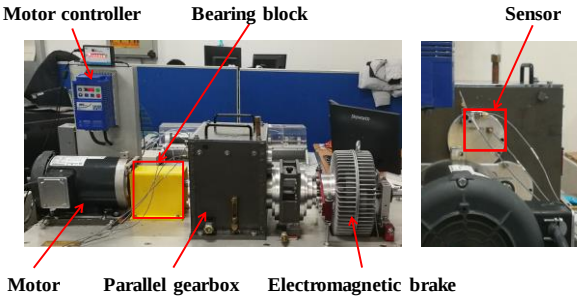


Fig. 6. Wind turbine drive train fault diagnosis simulator.

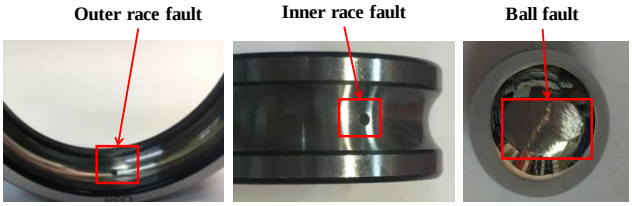


Fig. 7. Three fault conditions of the experimental bearing.

the damage depth is about 0.2mm. Due to the small volume of the rolling elements, the use of laser burns will cause secondary heat treatment of the rolling elements, so the rolling elements are ground with a grinder to preset the fault.

Because the change of load can be controlled by changing the current, the current and corresponding load are set as follows: 0A (load0 = 0HP), 0.4A (load1 = 1HP), 0.8A (load2 = 2HP) and 1.2A (load3 = 3HP). The speed of the DC motor is set to four speed conditions of 500r/min, 1000r/min, 1500r/min, and 2000r/min. The vibration signal is collected by Bruel & Kjaer

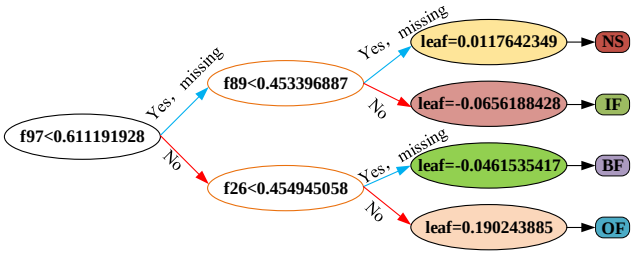


Fig. 8. Decision tree for 4-class problem using multi-scale mixed features.

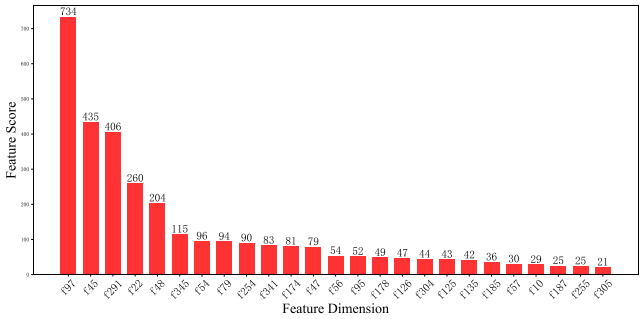


Fig. 9. Sorted histogram according to the importance of features(Top 25 dimensions).

Deltatron accelerometer Type 4508B. The characteristics and metrological characterization of the sensors are listed as follows: the amplitude frequency range is from 0.3 Hz to 8 kHz, the phase frequency range is from 2 Hz to 5 kHz, the sensitivity is 10.23 mV/g, the mounted resonance frequency is 25 kHz, the measuring range is ± 70 g peak, the bias voltage is 13 V ± 1 V, the power supply requirements is from 2 to 20 mA, the inherent noise is less than 350 mg, the maximum shock level is 5000 g, the environmental temperature range is from -54 ℃ to 121 ℃, the electrical connector is 10-32 UNF and the weight is 4.8 gram. The signal acquisition card DT9837 is used to acquire vibration signals, the sampling frequency is 20480Hz, and the sampling number is 2400 points.

In this section, the collected vibration signals under 4 different loads are used as analysis data sets A, B, C, and D. Each data set include normal state(NS), inner race fault (IF), outer race fault (OF) and ball fault(BF). And there are a total of 400 samples in every dataset, of which 100 samples of each type, and each sample has 2400 sampling points. The related setting schemes can be seen in Table 7. According to Table 7, we also set up 12 transfer learning tasks, namely E→F, E→G, E→H, F→E, F→G, F→H, G→E, G→F, G→H , H→E, H→F and H→G. Data sets E, F, F, and H are respectively regarded as the source domain and target domain.

2) Experimental Results and Analysis

In this section, the XJTU's variable load datasets will be used to verify the effectiveness and superiority of the proposed method. Similar to Case A, firstly, the multi-scale mixed domain feature extractor is used to extract the feature set, then normalized, and then GS_XGBoost is used to train the model.

Fig.8 is one of the decision trees in the training process. With reference to the decision tree for 4-class problem in Fig. 8, the most important feature from GS_XGBoost is "f97", and its corresponding feature is the sample entropy of the first-order IMF decomposed by VMD, reflecting the complexity of the time series signal. The larger the value, the more complex the sample

sequence. When the value is <0.611191928，the tree splits to "f89", the corresponding feature is the third wavelet packet energy spectrum feature of the first-order IMF decomposed by VMD, which reflects the energy distribution of the signal in the time-frequency domain. If the value "f89" <0.453396887, the input is categorized as NS, or it is IF. When the value of "f97" ≥611191928, the tree splits to "f26", the corresponding feature is the frequency domain characteristic parameter of the original signal, which reflects the degree of dispersion or concentration of the spectrum. If the value "f26" <0.454945058, the input is categorized as BF, or it is OF

The feature importance is sorted according to the index total gain in XGBoost. Due to space limitations, the 25-dimensional features with the highest importance ranking are used here to draw a bar graph, as shown in Fig.9. In order to save space, the remaining feature importance is omitted. The features of the first 25 dimensions and their physical meanings are shown in Table V. The most important feature is the sample entropy of the first-order IMF and the second place is the wavelet scale energy of the input signal. It can be seen that the extracted features of the designed multi-scale mixed domain feature extractors have different sensitivity for diagnosis, and their contribution to fault diagnosis is also different.

Moreover, the features with high contribution are mainly concentrated on the original signal and the IMFs of the first 3 orders. In particular, due to the outstanding non-linear feature extraction ability and good anti-noise ability, approximate entropy, sample entropy and permutation entropy have become the main sensitive features. In addition, the mean square value, root amplitude, waveform indicator of time domain, the spectral characteristic parameters and the power spectral density characteristic parameters of frequency domain, and the wavelet scale energy of time-frequency domain all contribute to varying degrees. This shows that the multi-scale mixed domain feature extractor proposed in the article can extract a complete feature set.

In order to determine the feature dimension to be input after XGBoost optimization, the feature dimension is gradually increased according to the importance of the feature and input into SVM for identification. The 12 data sets of Case B are tested on, and the average of each accuracy rate is thus taken. With the increase of the input feature dimension, the average recognition rate gradually increases. When the first 37-dimensional features are input, the average recognition rate reaches 75.81%, and then as the input dimension continues to increase, the recognition rate decreases to some extent, but it also stabilizes at about 74.13%.The reason may be that there is redundancy and interference in the feature set after adding features of low importance, resulting in negative transfer. Therefore, the best feature subset is the first 37 dimensional features.

In order to illustrate the superiority of the first 37-dimensional features optimized by XGBoost, the multi-scale mixed domain feature set (MMF) is compared with the first 37-dimensional features (XGB37) selected in this paper using the class separability index. The class separability index *J* is defined as the ratio of the trace of the intra-class divergence to the trace of the inter-class divergence. The smaller its value is, the better the separability between classes. The calculation formula of J is defined as follows:

$$J = tr(\mathrm{S_w}) / \mathrm{tr}(\mathrm{S_b}) \tag{25}$$

where S_w is the intra-class divergence, S_b is the inter-class divergence, and tr(·) represents the trace of the matrix.

$$S_w = \sum_{j=1}^{c}\sum_{i=1}^{N_j}\left(x_i^j - \mu_j\right)\left(x_i^j - \mu_j\right)' \tag{26}$$

$$S_b = \sum_{j=1}^{c}\left(\mu_j - \mu\right)\left(\mu_j - \mu\right)' \tag{27}$$

TABLE V
THE NOTABLE FEATURES OF THE TOP 25 DIMENSIONS AND THEIR PHYSICAL MEANING.

Order	Feature Dimension	Physical Meaning	Feature Score
1	97	1st IMF's sample entropy	734
2	45	Input signal's 8th wavelet scale energy	435
3	291	5th IMF's approximate entropy	406
4	22	Input signal's 7th characteristic parameter of spectral	260
5	48	Input signal's sample entropy	204
6	345	6th IMF's mean square value	115
7	54	1st IMF's root amplitude	96
8	79	1st IMF's frequency variance	94
9	254	5th IMF's peak-to-peak value	90
10	341	6th IMF's approximate entropy	83
11	174	3rd IMF's12th characteristic parameter of spectral	81
12	47	Input signal's approximate entropy	79
13	56	1st IMF's mean square amplitude	54
14	95	1st IMF's permutation entropy	52
15	178	3rd IMF's 1st characteristic parameters of power spectral density	49
16	126	2nd IMF's center frequency	47
17	304	6th IMF's waveform indicator	44
18	125	2nd IMF's 12th characteristic parameter of spectral	43
19	135	2nd IMF's 7th characteristic parameters of power spectral density	42
20	185	3rd IMF's 1st wavelet scale energy	36
21	57	1st IMF's peak value	30
22	10	Input signal's waveform indicator	29
23	187	3rd IMF's 3rd wavelet scale energy	25
24	255	5th IMF's waveform indicator	25
25	305	6th IMF's peak indicator	21

TABLE VI
COMPARISON OF FEATURE SET INDICATORS.

Feature set	tr(S_w)	Tr(S_b)	J	Average accuracy using SVM
MMF	5.8206	6.1847	0.9411	74.13%
XGB37	0.7294	0.8310	0.8777	75.81%

where x_i^j is the i-th sample in the j-th category, μ_j is the mean of the j-th sample, c is the number of categories of the sample, N_j is the total number of samples in the j-th category, μ is the mean of all samples.

According to Table VI, the J value of the optimized feature subset XGB37 is smaller and the average accuracy rate is higher. Therefore, it can be explained that the optimized feature subset has more advantages than the original features extracted by the multi-scale mixed domain feature extractor.

Next, the MEDA method is then used for cross-domain fault diagnosis for XJTU Lab's variable load datasets. For each sample, the preferred feature (XGB37) proposed by the above method is obtained, and a manifold feature transformation (10-dimensional) is performed on it in order to reduce the domain difference between the source domain and the target domain. Finally, cross-domain fault diagnosis can be realized through adaptive distribution alignment and the learning of simple classifier 1NN.

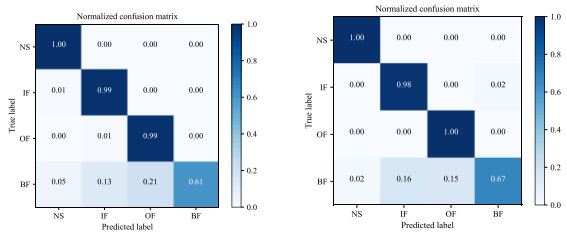
Taking task F→E as an example, the accuracy of DA-MMDF is 91.50%. What's more, according to the confusion matrix shown in Fig.10, the proposed DA-MMDF can correctly classify NS, IF and OF, but misidentify the category RF for the unlabeled target domain samples. From the confusion matrix of Fig. 10 (a) and Fig. 10 (b), it can be seen that domain adaptation can effectively improve the accuracy of cross-domain diagnosis. The result of *t-sne* visualization is shown in Fig.11. It can be found that the proposed method can reduce the data drift between different domains.

The method in this paper is further verified under 12 variable load conditions in the XJTU data set, then compared with the traditional intelligent diagnosis method and transfer learning method. The results are shown in Table VII. It can be seen that the DA-MMDF method achieves the highest accuracy rate under 7 variable load conditions, and at the same time, the average accuracy rate of the proposed method is also the highest.

C. Variable Speed Datasets from XJTU Lab

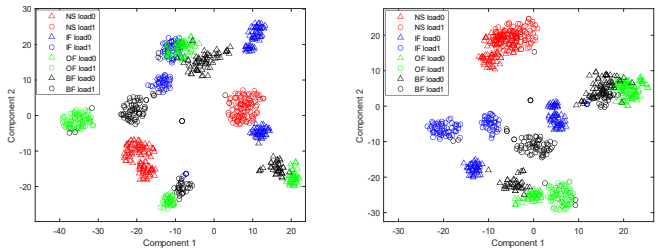
1) Data description

This case uses the same XJTU experimental platform as Case B to collect data, for details of related experimental devices in Case B. In this section, the collected vibration signals under 4 different speeds are used as analysis data sets I, J, K, and L. The speed of the DC motor is set to four speed conditions of 500r/min, 1000r/min, 1500r/min, and 2000r/min, so the load is set as 0.4A (load1 = 1HP). And there are a total of 400 samples in every dataset , of which 100 samples of each type, and each sample has 3200 sampling points. What's more, we also set up 12 transfer learning tasks, namely I→J, I→K, I→L, J→I, J→K, J→L, K→I, K→J, K→L , L→I, L→J and L→K. Data sets I, J, K, and L are respectively regarded as the source domain and target domain.



(a) Before domain adaptation (b) After domain adaptation

Fig. 10. Confusion Matrix of Proposed Method on XJTU dataset E.



(a) Before domain adaptation (b) After domain adaptation

Fig. 11. Visualization result of cross process domain adaptation on dataset F(source domain) and dataset E(target domain).

TABLE VII
ACCURACY(%) OF PROPOSED METHOD ON XJTU'S VARIABLE LOAD DATASETS.

ID	Task	1NN	SVM	TCA	GFK[55]	BDA[56]	JDA	CORAL [57]	TJM [58]	JGSA	EasyTL (g)[59]	EasyTL (c)	**DA-MMDF**
1	E→F	60.25	71.75	71.25	61.25	65.75	78.00	72.00	82.50	**95.25**	54.50	54.50	82.00
2	E→G	63.25	71.00	64.75	65.50	63.50	68.25	70.50	**74.75**	72.25	61.50	57.75	69.75
3	E→H	71.75	79.75	67.25	72.50	52.00	70.75	**82.00**	74.50	81.25	71.50	60.00	79.00
4	F→E	61.50	67.75	71.75	65.00	70.25	80.00	76.50	73.75	89.25	64.50	63.75	**91.50**
5	F→G	88.75	92.75	94.50	90.75	96.50	96.50	92.25	95.50	82.25	82.75	80.75	**96.50**
6	F→H	72.75	89.00	91.75	85.25	91.75	93.50	85.75	96.50	87.00	86.75	82.5	**96.75**
7	G→E	66.25	54.50	66.00	58.25	53.75	75.25	55.00	73.00	85.50	66.50	74.25	**88.50**
8	G→F	93.00	94.00	94.75	95.00	96.25	94.25	93.25	96.50	96.50	96.00	95.00	**96.75**
9	G→H	93.25	94.00	94.00	92.75	94.50	93.50	94.00	94.75	88.00	85.75	86.25	**97.25**
10	H→E	30.75	33.75	52.50	35.75	47.75	55.75	33.00	54.50	54.75	42.75	**63.25**	62.75
11	H→F	64.25	67.00	85.00	73.50	83.75	86.50	73.00	89.50	95.50	92.00	84.00	**94.00**
12	H→G	90.25	94.5.	91.50	91.75	92.50	91.25	93.25	**94.00**	83.50	80.75	83.50	91.00
-	AVG	71.33	75.81	78.75	73.94	75.69	81.96	76.71	83.31	84.25	73.77	73.79	**87.15**
-	STD	0.1737	0.1801	0.1412	0.1651	0.1806	0.1125	0.1745	0.1278	0.1108	0.1493	0.1228	**0.1098**
-	Average rank	12	7	5	9	8	4	6	3	2	11	10	**1**

TABLE VIII
ACCURACY(%) OF PROPOSED METHOD ON XJTU'S VARIABLE SPEED DATASETS.

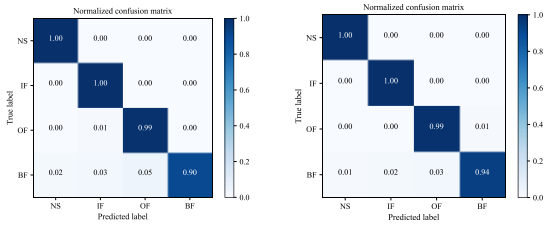
ID	Task	1NN	SVM	TCA	GFK	BDA	JDA	CORAL	TJM	JGSA	EasyTL (g)	EasyTL (c)	**DA-MMDF**
1	I→J	63.00	73.00	56.50	58.75	44.75	64.00	70.25	57.50	82.50	58.25	62.75	**83.75**
2	I→K	42.25	67.25	51.00	45.50	61.00	55.25	68.00	61.25	**92.00**	75.50	79.25	85.50
3	I→L	40.00	63.00	45.25	44.50	47.75	45.00	55.00	48.50	61.00	48.75	53.50	**79.00**
4	J→I	52.00	59.50	49.75	52.50	40.75	53.25	61.00	46.75	59.75	43.00	52.50	**61.75**
5	J→K	71.75	94.25	90.25	78.75	87.00	92.75	92.50	91.75	97.75	84.75	87.50	**98.25**
6	J→L	49.75	62.00	52.75	50.75	46.75	50.25	68.50	61.75	69.25	40.25	32.25	**79.75**
7	K→I	44.50	46.25	49.50	44.75	44.75	42.00	46.50	52.50	**58.25**	42.25	49.50	48.50
8	K→J	54.50	65.50	80.50	54.50	78.50	89.00	70.00	91.25	94.75	58.75	54.50	**96.25**
9	K→L	47.25	55.00	56.25	47.50	47.00	54.00	57.00	63.00	**71.75**	52.50	53.00	56.00
10	L→I	27.50	32.75	33.75	28.50	32.00	40.50	32.75	33.50	38.75	32.50	40.50	**45.25**
11	L→J	51.50	56.75	48.25	49.50	37.75	52.50	**58.00**	48.75	41.75	42.75	44.50	41.50
12	L→K	50.25	83.75	73.25	58.00	49.75	85.50	79.75	53.25	66.75	55.50	50.50	**94.75**
-	AVG	49.52	63.25	57.25	51.12	51.48	60.33	63.27	59.15	69.52	52.90	55.02	**72.52**
-	STD	**0.1069**	0.1539	0.1538	0.1128	0.1559	0.1771	0.1478	0.1642	0.1856	0.1444	0.1474	0.1996
-	Average rank	12	4	7	11	10	5	3	6	2	9	8	**1**

2) Experimental Results and Analysis

For the datasets under variable speed conditions, the same processing method as Case A and Case B is adopted. Taking task J→K as an example, the accuracy of DA-MMDF is 98.25%.

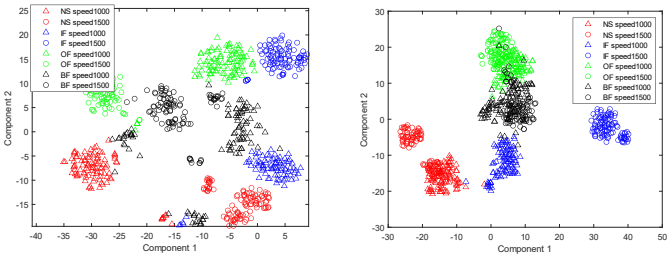
Moreover, the confusion matrix and the result of *t-sne* visualization are shown in Fig.12 and Fig.13 respectively. Similarly, it can be found from Fig. 13 (a) and Fig. 13 (b) that domain adaptation plays a certain role in improving the accuracy of cross-domain diagnosis. The method in this paper is further verified under 12 variable speed conditions in the XJTU datasets, and compared with the traditional intelligent diagnosis method and transfer learning method. The results are shown in Table VIII. It can be seen that the DA-MMDF method achieves the highest accuracy rate under 8 variable speed conditions, and at the same time, the average accuracy rate of the proposed method is also the highest.

The average accuracy of transfer diagnosis based on the variable load and variable speed datasets of XJTU laboratory is shown in Fig.14. It can be found that the average accuracy of the proposed method is higher than other methods. In addition,



(a) Before domain adaptation (b) After domain adaptation

Fig. 12. Confusion Matrix of Proposed Method on XJTU dataset K.



(a) Before domain adaptation (b) After domain adaptation

Fig. 13. Visualization result of cross process domain adaptation on dataset J(source domain) and dataset K(target domain).

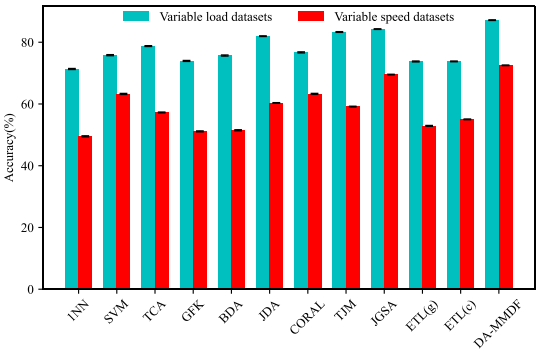


Fig. 14. Average accuracy of different methods on variable load and variable speed datasets.

the average accuracy of the variable load datasets is generally higher than the average accuracy of the variable speed datasets. Additionally, it is found that the accuracy of transfer diagnosis is not high when the data set with a rotation speed of 500r/min is used as the source domain or the target domain. It can be understood that when the speed is too low, the fault frequency cannot be effectively stimulated, so the fault characteristics are also difficult to effectively extract.

V. CONCLUSION

In this paper, a domain adaptive intelligent fault diagnosis algorithm (DA-MMDF) is proposed. This method is based on Multi-scale Mixed Domain Feature extractor to achieve complete feature extraction, and further uses GS_XGBoost to sort the importance of features and select sensitive feature set that is sensitive to fault. Furthermore, its superiority is proved. Finally, the intelligent fault diagnosis of rolling bearing under polytropic working conditions is realized by domain adaptive method. The effectiveness of DA-MMDF is proved by three experimental cases. Compared with the traditional intelligent diagnosis method and transfer learning method, the generalization ability and superiority of the proposed method is further demonstrated.

In the future work, we will continue to carry out the following research on DA-MMDF method.

1) The proposed method can also be applied to the fault diagnosis of gearbox fault, rotor fault, motor fault and hydraulic equipment fault.

2) With the proposed method, the key parameters of VMD algorithm are determined based on experience, and the optimal parameters can be solved by swarm intelligence optimization algorithm to further improve the accuracy of cross-domain fault diagnosis.

3) On the basis of this paper, we will explore the transfer diagnosis experiments under different fault levels and different types of components, and try to analyze the influencing factors.

REFERENCES

[1] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," Mechanical Systems and Signal Processing, vol. 122, pp. 692-706, 2019.

[2] Y. Lei, Z. He, and Y. Zi, "A new approach to intelligent fault diagnosis of rotating machinery," Expert Systems With Applications, vol. 35, no. 4, pp. 1593-1600, 2008.

[3] G. Chen, F. Liu, and W. Huang, "Sparse discriminant manifold projections for bearing fault diagnosis," Journal of Sound and Vibration, vol. 399, pp. 330-344, 2017.

[4] J. Antoni and R. B. Randall, "The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines," Mechanical Systems And Signal Processing, vol. 20, no. 2, pp. 308-331, Feb 2006.

[5] J. Antoni, "The spectral kurtosis: a useful tool for characterising non-stationary signals," Mechanical Systems And Signal Processing, vol. 20, no. 2, pp. 282-307, Feb 2006.

[6] G. L. McDonald and Q. Zhao, "Multipoint Optimal Minimum Entropy Deconvolution and Convolution Fix: Application to vibration fault detection," Mechanical Systems And Signal Processing, vol. 82, pp. 461-477, Jan 1 2017.

[7] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," Mechanical Systems And Signal Processing, vol. 35, no. 1-2, pp. 108-126, Feb 2013.

[8] S. Wang, X. Chen, C. Tong, and Z. Zhao, "Matching Synchrosqueezing Wavelet Transform and Application to Aeroengine Vibration Monitoring," IEEE Transactions on Instrumentation And Measurement, vol. 66, no. 2, pp. 360-372, Feb 2017.

[9] Y. Yang, Z. Peng, W. Zhang, and G. Meng, "Parameterised time-frequency analysis methods and their engineering applications: A review of recent advances," Mechanical Systems And Signal Processing, vol. 119, pp. 182-221, Mar 15 2019.

[10] W.-B. Shangguan, G.-F. Zheng, S. Rakheja, and Z. Yin, "A method for editing multi-axis load spectrums based on the wavelet transforms," Measurement, vol. 162, Oct 1 2020, Art. no. 107903.

[11] A. R. Bastami and S. Vahid, "Estimating the size of naturally generated defects in the outer ring and roller of a tapered roller bearing based on autoregressive model combined with envelope analysis and discrete wavelet transform," Measurement, vol. 159, Jul 15 2020, Art. no. 107767.

[12] Y. Hu, X. Tu, and F. Li, "High-order synchrosqueezing wavelet transform and application to planetary gearbox fault diagnosis," Mechanical Systems And Signal Processing, vol. 131, pp. 126-151, Sep 15 2019.

[13] Z. Zhao, S. Wang, C. Sun, R. Yan, and X. Chen, "Sparse Multiperiod Group Lasso for Bearing Multifault Diagnosis," IEEE Transactions on Instrumentation And Measurement, vol. 69, no. 2, pp. 419-431, Feb 2020.

[14] Z. Zhao, S. Wu, B. Qiao, S. Wang, and X. Chen, "Enhanced Sparse Period-Group Lasso for Bearing Fault Diagnosis," IEEE Transactions on Industrial Electronics, vol. 66, no. 3, pp. 2143-2153, Mar 2019.

[15] S. Wang, I. Selesnick, G. Cai, Y. Feng, X. Sui, and X. Chen, "Nonconvex Sparse Regularization and Convex Optimization for Bearing Fault Diagnosis," IEEE Transactions on Industrial Electronics, vol. 65, no. 9, pp. 7332-7342, Sep 2018.

[16] Y. Li, X. Liang, and M. J. Zuo, "A new strategy of using a time-varying structure element for mathematical morphological filtering," Measurement, vol. 106, pp. 53-65, Aug 2017.

[17] K. Dragomiretskiy and D. Zosso, "Variational Mode Decomposition," IEEE Transactions on Signal Processing, vol. 62, no. 3, pp. 531-544, Feb 2014.

[18] Y. Miao, M. Zhao, Y. Yi, and J. Lin, "Application of sparsity-oriented VMD for gearbox fault diagnosis based on built-in encoder information," ISA Transactions, vol. 99, pp. 496-504, Apr 2020.

[19] X. Li, Z. Ma, D. Kang, and X. Li, "Fault diagnosis for rolling bearing based on VMD-FRFT," Measurement, vol. 155, Apr 2020, Art. no. 107554.

[20] M. Seera, M. L. D. Wong, and A. K. Nandi, "Classification of ball bearing faults using a hybrid intelligent model," Applied Soft Computing, vol. 57, pp. 427-435, Aug 2017.

[21] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," Neurocomputing, vol. 313, pp. 47-64, Nov 3 2018.

[22] Z. Huo, M. Martinez-Garcia, Y. Zhang, R. Yan, and L. Shu, "Entropy Measures in Machine Fault Diagnosis: Insights and Applications," IEEE Transactions on Instrumentation And Measurement, vol. 69, no. 6, pp. 2607-2620, Jun 2020.

[23] R. Q. Yan and R. X. Gao, "Complexity as a measure for machine health evaluation," IEEE Transactions on Instrumentation And Measurement, vol. 53, no. 4, pp. 1327-1334, Aug 2004.

[24] R. Yan and R. X. Gao, "Approximate Entropy as a diagnostic tool for machine health monitoring," Mechanical Systems And Signal Processing, vol. 21, no. 2, pp. 824-839, Feb 2007.

[25] R. Yan, Y. Liu, and R. X. Gao, "Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines," Mechanical Systems And Signal Processing, vol. 29, pp. 474-484, May 2012.

[26] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," Mechanical Systems And Signal Processing, vol. 72-73, pp. 303-315, May 2016.

[27] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," Mechanical Systems And Signal Processing, vol. 115, pp. 213-237, Jan 15 2019.

[28] Z. Xiang, X. Zhang, W. Zhang, and X. Xia, "Fault diagnosis of rolling bearing under fluctuating speed and variable load based on TCO Spectrum and Stacking Auto-encoder," Measurement, vol. 138, pp. 162-174, May 2019.

[29] V. Sugumaran and K. I. Ramachandran, "Effect of number of features on classification of roller bearing faults using SVM and PSVM," Expert Systems With Applications, vol. 38, no. 4, pp. 4088-4096, Apr 2011.

[30] X. Gao and J. Hou, "An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process," Neurocomputing, vol. 174, pp. 906-911, Jan 22 2016.

[31] S. P. Patel and S. H. Upadhyay, "Euclidean distance based feature ranking and subset selection for bearing fault diagnosis," Expert Systems With Applications, vol. 154, Sep 15 2020, Art. no. 113400.

[32] M. Balasubramanian and E. L. Schwartz, "The Isomap algorithm and topological stability," Science, vol. 295, no. 5552, Jan 4 2002.

[33] T. Chen, C. Guestrin, and M. Assoc Comp, XGBoost: A Scalable Tree Boosting System (Kdd'16: Proceedings Of the 22nd Acm Sigkdd International Conference on Knowledge Discovery And Data Mining). 2016, pp. 785-794.

[34] 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," Mechanical Systems And Signal Processing, vol. 138, Apr 2020, Art. no. 106587.

[35] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," IEEE Transactions on Knowledge And Data Engineering, vol. 22, no. 10, pp. 1345-1359, Oct 2010.

[36] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain Adaptation via Transfer Component Analysis," IEEE Transactions on Neural Networks, vol. 22, no. 2, pp. 199-210, Feb 2011.

[37] M. Long, J. Wang, G. Ding, J. Sun, P. S. Yu, and IEEE, "Transfer Feature Learning with Joint Distribution Adaptation," in 2013 IEEE International Conference on Computer Vision(IEEE International Conference on Computer Vision, 2013, pp. 2200-2207.

[38] J. Zhang, W. Li, P. Ogunbona, and IEEE, "Joint Geometrical and Statistical Alignment for Visual Domain Adaptation," in 30th IEEE Conference on Computer Vision And Pattern Recognition(IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5150-5158.

[39] J. Wang et al., Visual Domain Adaptation with Manifold Embedded Distribution Alignment (Proceedings Of the 2018 ACM Multimedia Conference). 2018, pp. 402-410.

[40] Y. Chaohui, W. Jindong, C. Yiqiang, and H. Meiyu, "Transfer Learning with Dynamic Adversarial Adaptation Network," 2019 IEEE International Conference on Data Mining (ICDM), pp. 778-86, 2019 2019.

[41] Y. Zhu et al., "Deep Subdomain Adaptation Network for Image Classification," IEEE transactions on neural networks and learning systems, vol. PP, 2020-May-04 2020.

[42] J. Jiao, M. Zhao, J. Lin, and C. Ding, "Classifier Inconsistency-Based Domain Adaptation Network for Partial Transfer Intelligent Diagnosis," IEEE Transactions on Industrial Informatics, vol. 16, no. 9, pp. 5965-5974, Sep 2020.
[43] T. Han, C. Liu, W. Yang, and D. Jiang, "Learning transferable features in deep convolutional neural networks for diagnosing unseen machine conditions," ISA Transactions, vol. 93, pp. 341-353, Oct 2019.
[44] D. Shuzhi, W. Guangrui, and Z. Zhifen, Bearing fault diagnosis under different operating conditions based on cross domain feature projection and domain adaptation (2019 IEEE International Instrumentation and Measurement Technology Conference). 2019, pp. 6 pp.-6 pp.
[45] L. Wen, L. Gao, and X. Li, "A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis," IEEE Transactions on Systems Man Cybernetics-Systems, vol. 49, no. 1, pp. 136-144, Jan 2019.
[46] 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 Transactions on Industrial Electronics, vol. 66, no. 9, pp. 7316-7325, Sep 2019.
[47] Y. Lei, Z. He, and Y. Zi, "Fault diagnosis based on novel hybrid intelligent model," Chinese Journal of Mechanical Engineering, vol. 44, no. 7, pp. 112-17, July 2008.
[48] P. Chen, M. Taniguchi, T. Toyota, and Z. J. He, "Fault diagnosis method for machinery in unsteady operating condition by instantaneous power spectrum and genetic programming," Mechanical Systems And Signal Processing, vol. 19, no. 1, pp. 175-194, Jan 2005.
[49] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," Physical Review Letters, vol. 88, no. 17, Apr 29 2002, Art. no. 174102.
[50] S. M. Pincus, "APPROXIMATE ENTROPY AS A MEASURE OF SYSTEM-COMPLEXITY," Proceedings Of the National Academy Of Sciences Of the United States Of America, vol. 88, no. 6, pp. 2297-2301, Mar 1991.
[51] D. Labate, F. La Foresta, G. Morabito, I. Palamara, and F. C. Morabito, "Entropic Measures of EEG Complexity in Alzheimer's Disease Through a Multivariate Multiscale Approach," IEEE Sensors Journal, vol. 13, no. 9, pp. 3284-3292, Sep 2013.
[52] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," IEEE Transactions on Neural Systems And Rehabilitation Engineering, vol. 15, no. 2, pp. 266-272, Jun 2007.
[53] B. Gong, Y. Shi, F. Sha, K. Grauman, and IEEE, "Geodesic Flow Kernel for Unsupervised Domain Adaptation," in 2012 IEEE Conference on Computer Vision And Pattern Recognition(IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2066-2073.
[54] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," Journal Of Machine Learning Research, vol. 9, pp. 2579-2605, Nov 2008.
[55] B. Gong, Y. Shi, F. Sha, K. Grauman, "Geodesic Flow Kernel for Unsupervised Domain Adaptation," in 2012 IEEE Conference on Computer Vision And Pattern Recognition(IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2066-2073.
[56] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced Distribution Adaptation for Transfer Learning," in 2017 17th IEEE International Conference on Data Mining, V. Raghavan, S. Aluru, G. Karypis, L. Miele, and X. Wu, Eds. (IEEE International Conference on Data Mining, 2017, pp. 1129-1134.
[57] B. Sun, J. Feng, K. Saenko, and Aaai, Return of Frustratingly Easy Domain Adaptation (Thirtieth Aaai Conference on Artificial Intelligence). 2016, pp. 2058-2065.
[58] M. Long, J. Wang, G. Ding, J. Sun, P. S. Yu, and IEEE, "Transfer Joint Matching for Unsupervised Domain Adaptation," in 2014 IEEE Conference on Computer Vision And Pattern Recognition(IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1410-1417.
[59] J. Wang, Y. Chen, H. Yu, M. Huang, Q. Yang, "EASY TRANSFER LEARNING BY EXPLOITING INTRA-DOMAIN STRUCTURES," in 2019 IEEE International Conference on Multimedia And Expo(IEEE International Conference on Multimedia and Expo, 2019, pp. 1210-1215.

Zihao Lei received the B.S. degree in mechanical engineering from Southwest Jiaotong University, Chengdu, China, in 2018. He is current working toward the Ph.D. degree in mechanical engineering in the Department of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China. His current research is focused on transfer learning, deep learning, machinery condition monitoring, intelligent fault diagnosis and prognosis.

Guangrui Wen received his B.S. degree, M.S. degree and Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 1998, 2001 and 2006 respectively. He was a postdoctor fellow at Xi'an Shaangu Power Co., LTD. Xi'an, China from 2008 to 2010. He is a member of Chinese Mechanical Engineering Society (CIME) and Chinese Society for Vibration Engineering (CSVE). He has published 2 books and more than 80 articles and held more than 20 items of patents. His research interests include Mechanical System Fault Diagnosis & Prognosis, Mechanical Equipment Life Cycle Health Monitoring & Intelligent maintenance.

Shuzhi Dong received his B.S. degree from Southwest Jiaotong University, Chengdu, China, in 2015. He is currently working towards the Ph.D. degree in mechanical engineering in the Department of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China. His research interests are Fault Diagnosis and health monitoring of Mechanical Equipment, and mainly related to the deep learning and other artificial intelligence methods.

Xin Huang received his B.S. and M.S. degrees in mechanical engineering from Xinjiang University, Urumqi, China, in 2013 and 2016, respectively. He is currently working towards the Ph.D. degree in mechanical engineering in the Department of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China. His research interests include mechanical signal processing, mechanical system fault diagnosis, and prognosis.

Haoxuan Zhou received his B.S. degree from Southwest Petroleum University in 2016 and M.S. degree from Kunming University of Science and Technology in 2019. He is currently working towards the Ph.D. degree in mechanical engineering in the Department of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China. His main research direction is mechanical signal processing, rotating machinery fault diagnosis, and mechanical system health monitoring based on industrial big data.

Zhifen Zhang received the B.S., M.S. and Ph. D. degrees in Materials Processing Engineering from the Lanzhou University of Technology, GS in 2007, 2010, and Shanghai Jiao Tong University, Shanghai in 2015. She was a lecturer from 2015 to 2019 with the

Department of Mechanical Engineering, Xi'an Jiao Tong University, Xi'an, China. Since Oct.2019, she has been an associate professor at the same department. She has published more than 10 magazine papers including published in IEEE Trans.Industrial Informatics, Mechanical Systems and Signal Processing, Journal of Materials Processing Technology and so on. Her research interests include intelligent manufacturing of laser shocking peening and information fusion.

Xuefeng Chen received the Ph.D. degree from Xian Jiaotong University, Xian, China, in 2004. He is currently a Professor of Mechanical Engineering with Xian Jiaotong University. His current research interests include finite-element method, mechanical system and signal processing, diagnosis and prognosis for complicated industrial systems, smart structures, aero-engine fault diagnosis, and wind turbine system monitoring. Dr. Chen was a recipient of the National Excellent Doctoral Dissertation of China in 2007, the Second Award of Technology Invention of China in 2009, the National Science Fund for Distinguished Young Scholars in 2012, and a Chief Scientist of the National Key Basic Research Program of China (973 Program) in 2015.