Research Article

New State Identification Method for Rotating Machinery under Variable Load Conditions Based on Hybrid Entropy Features and Joint Distribution Adaptation

Xiaoming Xue^{1,2}, Nan Zhang², Suqun Cao², Wei Jiang², Jianzhong Zhou,³ and Liyan Liu²

¹Jiangsu Key Laboratory of Advanced Manufacturing Technology, Huaiyin Institute of Technology, Huai'an 223003, China

²Faculty of Mechanical and Material Engineering, Huaiyin Institute of Technology, Huai'an 223003, China

³College of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Correspondence should be addressed to Xiaoming Xue; taxueli@hyit.edu.cn and Suqun Cao; caosunqun@tsinghua.org.cn

Received 5 August 2019; Revised 14 April 2020; Accepted 20 May 2020; Published 10 July 2020

Academic Editor: Dan Selişteanu

Copyright © 2020 Xiaoming Xue et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fault identification under variable operating conditions is a task of great importance and challenge for equipment health management. However, when dealing with this kind of issue, traditional fault diagnosis methods based on the assumption of the distribution coherence of the training and testing set are no longer applicable. In this paper, a novel state identification method integrated by time-frequency decomposition, multi-information entropies, and joint distribution adaptation is proposed for rolling element bearings. At first, fast ensemble empirical mode decomposition was employed to decompose the vibration signals into a collection of intrinsic mode functions, aiming at obtaining the multiscale description of the original signals. Then, hybrid entropy features that can characterize the dynamic and complexity of time series in the local space, global space, and frequency domain were extracted from each intrinsic mode function. As for the training and testing set under different load conditions, all data was mapped into a reproducing space by joint distribution adaptation to reduce the distribution discrepancies between datasets, where the pseudolabels of the testing set and the final diagnostic results were obtained by the k -nearest neighbor algorithm. Finally, five cases with the training and testing set under variable load conditions were used to demonstrate the performance of the proposed method, and comparisons with some other diagnosis models combined with the same features and other dimensionality reduction methods were also discussed. The analysis results show that the proposed method can effectively recognize the multifaults of rolling element bearings under variable load conditions with higher accuracies and has sound practicability.

1. Introduction

Rolling element bearing is an important and commonly used component in rotating machines, which can effectively decrease the friction coefficient and ensure the turning accuracy of the shaft. As the running time increased, fatigue wear usually can lead to bearing parts' damage or even failures, including inner race, outer race, and rolling parts fault [1]. Meanwhile, these faults will undermine the normal operation of the machines and further cause chain fault response if without maintain services. According to the statistics, about 40%–50% of the rotating machinery faults

are directly or indirectly caused by bearings damage [2]. Combined with computer technology, communication technology, and artificial intelligence, the condition monitoring and fault diagnosis of bearings has gained extensive attention and wide application [3, 4]. Based on the collected condition data, the fault diagnosis can be seen as a typical pattern recognition problem. In the last twenty years, many works mainly focused on signal processing, feature extraction, feature dimension reduction, and classifiers have been carried out and the corresponding achievements are encouraging. It can be easily found that these application cases were just under the single operating condition and the

fault diagnosis under variable conditions gained little attention.

In practice, the operating conditions of equipment are not always constant according to different production tasks. Naturally, there is a high probability that one certain fault may occur under variable operating conditions or even this phenomenon is inevitable. Owing to complex operation states, such as different running speeds, the statistical characteristics of the signals will be changed and deeper fluctuation of the statistical distribution between data sources can be caused, especially when faults appear under new conditions [5]. As for this kind of problem, the traditional data-driven fault diagnosis methods are no longer applicable which is because they are performed based on the assumption of the distribution coherence of the training and testing set. In general, we can establish specific diagnostic models corresponding to each operating condition, but it will make the whole work more complex and inefficient. Furthermore, in many cases, the fault data under new operating conditions is scarce and difficult to obtain. So, the main purpose of this paper is to propose an effective and robust fault diagnosis method for the problem of variable load conditions. More specifically, the established model trained on historical data can be used to identify the testing samples under new operating conditions.

To eliminate the limitations of the traditional methods, some efforts with considering the influence of variable working conditions for diagnosis results have been tried. Tian et al. proposed an intelligent fault diagnosis method based on local mean decomposition, singular value decomposition, and extreme learning machine to recognize the multifaults of bearings under variable conditions [6]. Xing et al. utilized singular value features of proper rotation components and support vector machine to identify different fault types of gear under variable conditions [7]. Baraldi et al. established a diagnosis model under variable operating conditions based on the binary differential evolution algorithm and the k-nearest neighbor (KNN) classifier [8]. Obviously, in these works, the distribution discrepancies between training and testing sets have been ignored. Owing to rich historical data and better features, the diagnosis results may be relatively satisfactory in some cases. More notably, these multifault classifiers are trained by the training data obtained under different operating conditions, but they are not functional to diagnose the samples obtained from new operating conditions.

In recent years, to eliminate the limitation of distribution discrepancies for the traditional machine learning method, the transfer learning theory has been proposed and developed. The core of transfer learning focuses on knowledge migration and utilization, where the knowledge obtained in one problem can be transferred to solve a different but related problem. Owing to its crossdomain learning ability, transfer learning has been employed in many fields, such as image analysis, speech recognition, and text identification [9]. Depending on individual application requirements, Pan and Yang categorized the transfer learning approaches into

four types: instance-based transfer learning, feature-based transfer learning, parameter-based transfer learning, and relation-based transfer learning [10]. As for these four methods, some preliminary attempts in terms of fault diagnosis by using instance-based transfer learning and feature-based transfer learning have been carried out by engineers and researchers [11]. Generally, the instance-based transfer learning method only works when the distribution discrepancies between domains are relatively small. Unlike this method, the feature-based transfer learning method usually can perform well in dealing with crossdomain problems, which means it can be one potentially effective method for bearing fault diagnosis under variable operating conditions.

Among all the methods of feature-based transfer learning, the distribution adaptation method is used most frequently. The distribution adaptation means the original data matrixes can be projected into a regenerated kernel space by using the kernel function to obtain distribution coherence between domains. Aiming at reducing different probability distribution discrepancies, this method can be categorized into three types: marginal distribution adaptation (MDA) [12], conditional distribution adaptation (CDA) [13], and joint distribution adaptation (JDA) [14]. For MDA, the transformation object is to reduce the discrepancy of the marginal probability distribution between source and target domain to complete the transfer learning, while to reduce the discrepancy of conditional probability distribution is the target of CDA. By integrating the objective functions of MDA and CDA, JDA is employed to simultaneously reduce the distance of marginal probability distribution and conditional probability distribution. As a result, after transformation, the target domain data can exhibit even better between-class margin and within-class cohesion as well as display better data distinguishability. However, there are few reports about the research and application of JDA to the fault diagnosis field [15]. In this paper, the JDA method was employed to solve the bearing fault diagnosis problem under variable load conditions.

Before knowledge transfer learning, to obtain effective features which contain potential common information for signals under different operating conditions is another critical part for bearing fault diagnosis. Since the vibration signals of bearings exhibit strong nonlinearity and non-stationarity, especially under variable conditions, traditional linear features extraction methods, such as statistical analysis and Fourier transform, have limitations to characterize the signals effectively [16]. As one complexity measurement for a chaotic system, information entropy has been widely used to characterize the dynamic and randomness of time series. Its performance in terms of feature extraction in the fault diagnosis field has also been discussed and analyzed. In [17], permutation entropy and ensemble empirical mode decomposition were employed to detect and recognize bearing faults. Sample entropy of product functions obtained by local mean decomposition was proposed as the fault features to identify the fault of rolling element bearings [18]. Simultaneously, some other kinds of entropy features, such as

energy entropy [19], singular spectrum entropy [20], and power spectrum entropy [21], have also obtained attention. As for each entropy feature, it just can characterize signals unilaterally and other important information may be neglected. Jiang et al. employ three different entropies including singular spectrum entropy, power spectrum entropy, and approximate entropy to realize the fault diagnosis of rotating machinery, and the results show that the mixing entropies have higher diagnosis accuracies than each single entropy feature [22]. To obtain a comprehensive representation of the fault signals, mixing information entropies' feature can be an effective solution.

In this paper, the academic contributions are as follows. (1) The hybrid entropy feature set based on mixing domain entropy features and time-frequency analysis method was constructed to characterize the vibration signals comprehensively. (2) An advanced transfer learning method, namely, JDA, was first employed to diagnose the faults of bearings. (3). A novel diagnosis method based on hybrid entropy features and JDA was proposed to diagnose the bearing faults under variable load conditions.

Specifically, the hybrid entropy features include permutation entropy (PE), sample entropy (SE), energy entropy (EE), singular spectrum entropy (SSE), and power spectrum entropy (PSE). These features can characterize the vibration signals from the views of local space, overall space, and frequency domain. Meanwhile, the original vibration signals consist of different signal components with multiscale space and nonlinear relationships, which will easily lead to the problem of useful information loss by only extracting features from the single scale. To obtain multiscale features from the original signal, the original signal was firstly decomposed into a collection of intrinsic mode functions (IMFs) by using fast ensemble empirical mode decomposition (FEEMD) [23]. After decomposition, five entropy features mentioned above were extracted from each IMF component to construct the hybrid entropy features dataset. Next, to reduce the distribution discrepancies between source and target domains, the features dataset of training set and testing set were mapped into a reproducing space by JDA, resulting in better intraclass compactness and interclass differentiation for the features dataset. Here, the KNN algorithm was used to generate the pseudolabels of the testing set and the final diagnostic results. Finally, five cases with different load conditions for training set and testing set were constructed to verify the performance of the proposed method. The analysis results show that, compared with other methods, the proposed method can effectively recognize the multifaults of rolling element bearings under variable load conditions with higher accuracies and has sound practicability.

The rest of this paper is organized as follows. Section 2 details the theory of FEEMD and each information entropy. The crossdomain features transfer learning based on JDA is provided in Section 3. The system framework of the proposed fault diagnosis method and the specific steps are presented in Section 4. In Section 5, the proposed method is applied to the bearing fault diagnosis and some comparisons are also given and discussed. Finally, the conclusions are drawn in Section 6.

2. Fast EEMD and Five Kinds of Information Entropy

2.1. Fast EEMD. Empirical mode decomposition (EMD) is one typical time-frequency analysis method by which the original complicated signals can be decomposed into a collection of locally narrowband components, known as IMFs. For EMD, the intermittent or noise components contained in the original signal can complicate analyses and obscure physical meanings, which are defined as the mode mixing problem [24]. To solve this issue, ensemble empirical mode decomposition (EEMD) was proposed [25]. This method implements the EMD process on different polluted signals with a certain number of trials, where the polluted signals are formed by adding different Gaussian white noise series with the same amplitude to the original signal. However, the repeated implementation of EMD usually needs intensive computation for the EEMD method. The massive amount of computation cost becomes a big obstacle to the engineering applications of this method [26, 27].

Fortunately, some efforts have been done to prove that the time complexity of the EMD/EEMD is equivalent to that of the Fourier Transform, and an improved EEMD method named fast EEMD (FEEMD) with better computational performance was proposed [23]. As for the whole algorithm flow, there is no difference in fundamentals between EEMD and FEEMD, only some efficient computation methods are employed to accomplish these steps in FEEMD. The specific steps of FEEMD are given as follows:

Step 1: initialize parameters, including the number of ensemble trials M , the amplitude of the added noise, and $i = 0$.

Step 2: set $i = i + 1$, and add the white noise series $n_i(t)$ to the original signal $x(t)$ and obtain the polluted signal as follows:

$$x_i(t) = x(t) + n_i(t). \quad (1)$$

Step 3: perform EMD on $x_i(t)$ and the decomposition results are given as follows:

$$x_i(t) = \sum_{j=1}^Z \text{imf}_{ij} + r(t), \quad (2)$$

where imf_{ij} is the j th IMF components of $x_i(t)$, Z is the total number of IMFs, and $r_i(t)$ is the residue component.

In step 3, the high computation efficiency of FEEMD mainly owes to the following several points and the detailed descriptions can be found in [23].

(1) For extreme point identification, the comparison information of the consecutive points in the last step can be repeatedly employed in the next step, so some duplications of the work can be reduced.

- (2) As for spline curve fitting, the popular Thomas algorithm [28], which is the most economical way to solve tridiagonal linear equations, is employed.
- (3) A novel recursive relation with high efficiency is proposed to calculate envelopes.

Step 4: if $i < M$, return to step 2, or if $i = M$, proceed to the next step.

Step 5: after M trials, calculate the ensemble mean of the each IMF component, and the corresponding mean components are regarded as the final results of FEEMD:

$$\begin{cases} \text{imf}_j = \frac{1}{M} \sum_{i=1}^M \text{imf}_{ij}, \\ r(t) = \frac{1}{M} \sum_{i=1}^M r_i(t), \\ x(t) = \sum_{j=1}^Z \text{imf}_j + r(t), \end{cases} \quad (3)$$

where imf_j is the j th IMF component of $x(t)$.

2.2. Five Kinds of Information Entropy. To get a comprehensive representation of the vibration signals, five kinds of information entropy were employed in the feature extraction. Here, the theoretical background of each information entropy is given firstly.

2.2.1. Permutation Entropy. PE was recently proposed by Bandt and Pompe to measure the dynamic changes of the time series by comparing the neighboring values [29]. For a given signal of length N , $X = \{x(i), i = 1, 2, \dots, N\}$, and its phase space can be reconstructed as follows:

$$\begin{cases} X_1 = \{x(1), x(1 + \lambda), \dots, x(1 + (m-1)\lambda)\} \\ \vdots \\ X_k = \{x(k), x(k + \lambda), \dots, x(k + (m-1)\lambda)\} \\ \vdots \\ X_{N-(m-1)\lambda} = \{x(N - (m-1)\lambda), x(N - (m-1)\lambda + \lambda), \\ \dots, x(N)\}, \end{cases} \quad (4)$$

where m is the embedded dimension and λ is the time delay. Next, the m values of X_k can be sorted in ascending order as follows:

$$x(k + (j_1 - 1)\lambda) \leq x(k + (j_2 - 1)\lambda) \leq \dots \leq x(k + (j_m - 1)\lambda). \quad (5)$$

If the values of any two elements are equal, the permutation order is arranged according to the value of j . Thus, X_k can be projected onto a collection of symbols as $S(l) = (j_1, j_2, \dots, j_m)$, where $l = 1, 2, \dots, L$, ($L \leq m!$) and $m!$ is the largest number of different permutation modes. Let P_l

denotes the probability distribution of each symbol sequence $S(l)$ and $\sum_{l=1}^L P_l = 1$. According to Shannon's entropy of the $m!$ distinct symbol sequences, the PE value of X with m embedded dimension is defined as follows:

$$H_{\text{PE}}(m) = \sum_{l=1}^L -P_l \ln P_l. \quad (6)$$

Obviously, the upper limit of $H_{\text{PE}}(m)$ is $\ln(m!)$ when all probability distributions are equal to $1/m!$. For convenience, the value of H_{PE} can be normalized as follows:

$$\bar{H}_{\text{PE}} = \frac{1}{\ln(m!)} \sum_{l=1}^L P_l \ln P_l, \quad 0 \leq \bar{H}_{\text{PE}} \leq 1. \quad (7)$$

2.2.2. Sample Entropy. SE, as an improved version of approximate entropy [30], was originally proposed to measure the complexity of time sequences from the aspect of segment approximation [31]. The SE value of $X = \{x(i), i = 1, 2, \dots, N\}$ can be calculated as follows:

- (1) Generate a collection of m -dimension vectors $\{X_j^m: j = 1, 2, \dots, N - m + 1\}$ by constructing a phase space matrix:

$$A = \begin{bmatrix} X_1^m \\ X_2^m \\ \vdots \\ X_{N-m+1}^m \end{bmatrix} = \begin{bmatrix} x(1) & x(2) & \dots & x(m) \\ x(2) & x(3) & \dots & x(m+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-m+1) & x(N-m+2) & \dots & x(N) \end{bmatrix}. \quad (8)$$

- (2) Define the distance between any two row vectors of A as

$$d_{kl}^m = d[X_k^m, X_l^m] = \max(|x(k+n) - x(l+n)|), \quad n = 0, 1, \dots, m-1. \quad (9)$$

- (3) Given the similar tolerance γ , calculate the ratio B_k^m of the number $d_{kl}^m < \gamma$ and $N-m$:

$$B_k^m = \frac{1}{N-m} \text{num}\{d_{kl}^m < \gamma\}, \quad (10)$$

where $k, l = 1, 2, \dots, N - m + 1$ and $k \neq l$. In general, γ takes from 0.1 to 0.25 std, and std is the standard deviation of the original signal X .

- (4) The average of all B_k^m is defined as follows:

$$\bar{B}^m = \frac{1}{N-m+1} \sum_{k=1}^{N-m+1} B_k^m. \quad (11)$$

- (5) Let $m = m + 1$, and $\bar{B}^{m+1}$ can be obtained by repeating steps (1)–(4):

$$\bar{B}^{m+1} = \frac{1}{N-m} \sum_{k=1}^{N-m} B_k^{m+1}. \quad (12)$$

- (6) The sample entropy of X can be defined as follows:

$$H_{SE}(m, \gamma, N) = -\ln \frac{\bar{B}^{m+1}}{B^m}. \quad (13)$$

2.2.3. Energy Entropy. Once the working condition of bearings changes or even fault occurs, the frequency distributions of the vibration signals should be different from that of the previous condition and the signal energy distribution will also be changed accordingly. For each data point of $X = \{x(i), i = 1, 2, \dots, N\}$, its energy can be expressed as $E_i = x(i)^2$. Thus, the total energy of X can be calculated as $E = \sum_{i=1}^N E_i$. Because each data point energy is a fraction of the total energy, the ratio between E_i and E is $P_i = E_i/E$ and $\sum_{i=1}^N P_i = 1$. According to the Shannon entropy theory, the energy entropy of X can be defined as follows [32]:

$$H_{EE} = -\sum_{i=1}^N P_i \ln P_i. \quad (14)$$

2.2.4. Singular Spectrum Entropy. The singular spectrum entropy (SSE) method is proposed based on the singular spectrum analysis and the entropy theory [33]. Different from the above three entropies, SSE is used to measure the complexity and uncertainty of the signal sequence by analyzing the reconstructed phase space on the whole angle. For the given signal X , its phase space could be reconstructed in the same way as equation (8). Then, the singular value decomposition is conducted on the matrix A . The corresponding singular values are recorded as $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_m]$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$, where m represents the number of different patterns of A . The sum of all the singular values is recorded as $\lambda_{\text{sum}} = \sum_{i=1}^m \lambda_i$. By definition, we can get the singular spectrum entropy of X as follows:

$$H_{SSE} = -\sum_{i=1}^m \left(\frac{\lambda_i}{\lambda_{\text{sum}}} \right) \ln \left(\frac{\lambda_i}{\lambda_{\text{sum}}} \right). \quad (15)$$

2.2.5. Power Spectrum Entropy. The power spectrum analysis is mainly used to extract the correlation information of the original signals in the frequency domain. Combined with the entropy theory, the power spectrum entropy was developed and can be used to reflect the complexity and distribution modes of the signal energy in the frequency domain [21]. $F(f)$ is calculated by using discrete Fourier transform on the original signal $X = \{x(i), i = 1, 2, \dots, N\}$. Then, the power spectrum estimation of X is $S_x(f) = 1/N |F(f)|^2$, $f = 1, 2, \dots, N$, where N is the length of $F(f)$. According to the law of conservation of energy, the energy of X is equal to that of $F(f)$. So, $\{S_x(1), S_x(2), \dots, S_x(N)\}$ can be seen as one kind of energy partition of X in the frequency domain. The sum of all the power energy is recorded as $S_{\text{sum}} = \sum_{f=1}^N S_x(f)$. Thus, the power spectrum entropy can be defined as follows:

$$H_{PSE} = -\sum_{f=1}^N \frac{S_x(f)}{S_{\text{sum}}} \ln \frac{S_x(f)}{S_{\text{sum}}}. \quad (16)$$

3. Crossdomain Feature Learning Based on JDA

As described in Section 1, distribution adaptation is the most widely used method in transfer learning, which can complete the knowledge transferring between source and target domains aiming at reducing the distribution discrepancies. Before reviewing the JDA method, two basic terminologies about transfer learning are firstly presented below:

- (1) Domain: a domain D is the subject of the transfer learning process, which is composed of a k -dimensional feature space $X = \{x_i, i = 1, 2, \dots, n\}$ and the marginal distribution $P(x)$ of X , designated as $D = \{X, P(x)\}$, $x \in R^{1 \times k}$.
- (2) Task: given a domain D , a task of transfer learning is composed of the label set Y and a prediction function $f(x)$, i.e., $T = \{Y, f(x)\}$. In the probability theory, $f(x) = Q(y|x)$ can be interpreted as the conditional probability distribution, where $y \in Y$.

For a fault diagnosis task, the labeled training dataset is recorded as $X_s = \{x_i^s, y_i^s\}_{i=1}^{n_s}$ (source domain dataset) and the unlabeled testing dataset is $X_t = \{x_i^t\}_{i=1}^{n_t}$ (target domain dataset), where n_s and n_t represent the total number of training and testing samples, respectively. Under variable operating conditions, it can be easily concluded that the feature spaces of training and testing sets are both different in marginal and conditional distributions, i.e., $P(X_s) \neq P(X_t)$ and $Q_s(Y_s|X_s) \neq Q_t(Y_t|X_t)$. The weak form of JDA is to learn a feature transform that simultaneously minimizes the difference between these two distributions [15]. The problem formulation of JDA is established as follows:

$$\begin{aligned} \min & \|P(A^T X_s) - P(A^T X_t)\| \\ & + \min \|Q(Y_s|A^T X_s) - Q(Y_t|A^T X_t)\|, \end{aligned} \quad (17)$$

where A is the transformation matrix. In equation (17), the first minimization contributes to reducing the marginal distribution discrepancy between domains, which is the objective function of MDA, and the second part is the objective function of CDA aimed to reduce the conditional distribution difference.

Figure 1 displays the calculation purpose of different distribution adaptation methods for knowledge learning of crossdomain. As shown in Figure 1(a), it is clear that the distributions for source domain D_s and target domain D_t are different. The direct use of the trained discriminative hyperplane f may cause a certain number of misclassification results in D_t . By using different distribution adaptation methods, the knowledge of the task T in the source domain is captured by the target domain data to make a distribution coherence between domains. In Figure 1(b), the MDA contributes to improving transfer performance by aligning the distribution centers, but it still has some aliasing areas near the discriminative hyperplane. As shown in Figure 1(c), just the overall similarity of the target data can be enhanced by the CDA method and the individual clustering has been ignored. To obtain better discriminative structures, JDA

FIGURE 1: An illustration of the crossdomain learning process by using MDA, CDA, and JDA.

looks forward to minimizing the difference of both marginal and conditional distributions, as shown in Figure 1(d).

To measure the distribution discrepancies in equation (17), the Maximum Mean Discrepancy (MMD) [34] method was adopted to tackle the optimization problems in a reproducing kernel Hilbert space H . The MMD distance of marginal distribution is described as follows:

$$\text{MMD}_{\text{MD}} = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} A^T x_i - \frac{1}{n_t} \sum_{j=1}^{n_t} A^T x_j \right\|_H^2. \quad (18)$$

In order to facilitate the calculation, equation (18) can be rewritten by employing the kernel method as follows:

$$\text{MMD}_{\text{MD}} = \text{tr}(A^T X M_0 X^T A), \quad (19)$$

$$(M_0)_{ij} = \begin{cases} \frac{1}{n_s^2}, & X_i, X_j \in D_s, \\ \frac{1}{n_t^2}, & X_i, X_j \in D_t, \\ -\frac{1}{n_s n_t}, & \text{otherwise,} \end{cases} \quad (20)$$

where X is the combined data of X_s and X_t and M_0 is the MMD matrix.

Minimizing the difference between $Q_s(Y_s | X_s)$ and $Q_t(Y_t | X_t)$ is also important to obtain robust distribution adaptation. Unlike MDA, the construction of MMD distance in CDA is intractable because of none acquaintance of the label information about target domain data. Specifically, $Q_t(Y_t | X_t)$ cannot be modeled directly without label information. Because the posterior probabilities $Q_s(Y_s | X_s)$ and $Q_t(Y_t | X_t)$ are very involved, according to Bayes formula, we can employ the sufficient statistics of class-conditional distributions $Q_s(X_s | Y_s)$ and $Q_t(X_t | Y_t)$ to

approximate them [14]. Here, to establish the $Q_t(X_t | Y_t)$ model, the pseudolabel information $\hat{Y}_t$ can be predicted by using some conventional classifiers such as k-nearest neighbor (KNN) and support vector machine (SVM). Then, the MMD distance between the class-conditional distributions $Q_s(X_s | Y_s)$ and $Q_t(X_t | Y_t)$ can be represented as follows:

$$\text{MMD}_{\text{CD}} = \sum_{c=1}^C \left\| \frac{1}{n_c} \sum_{x_i \in D_s^{(c)}} A^T x_i - \frac{1}{m_c} \sum_{x_j \in D_t^{(c)}} A^T x_j \right\|_H^2, \quad (21)$$

where n_c and m_c represent the number of samples with type c in source and target domains, respectively. Equation (21) can be rewritten as

$$\text{MMD}_{\text{CD}} = \text{tr}(A^T X M_C X^T A), \quad (22)$$

$$(M_C)_{ij} = \begin{cases} \frac{1}{n_c^2}, & x_i, x_j \in D_s^{(c)}, \\ \frac{1}{m_c^2}, & x_i, x_j \in D_t^{(c)}, \\ -\frac{1}{m_c n_c}, & \begin{cases} x_i \in D_s^{(c)}, x_j \in D_t^{(c)}, \\ x_i \in D_t^{(c)}, x_j \in D_s^{(c)}, \end{cases} \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

In equation (23), $D_s^{(c)} = \{x_i: x_i \in D_s \wedge y(x_i) = c\}$ is the dataset with type c in the source domain and $y(x_i)$ is the true label of x_i . Correspondingly, $D_t^{(c)} = \{x_j: x_j \in D_t \wedge \hat{y}(x_j) = c\}$ is the dataset with type c in the target domain and $\hat{y}(x_j)$ is the pseudolabel of x_j .

By integrating equations (19) and (22), the objective function of JDA can be written as

$$\min \sum_{c=0}^C \text{tr}(A^T X M_c X^T A) + \lambda \|A\|_F^2, \quad (24)$$

where $\lambda \|A\|_F^2$ is the regularization item and λ is the regularization parameter. By adding the regularization item, the optimization model can be well-defined.

To keep the characteristics of the original dataset before and after transformation, another optimization objective with the maximization of $A^T X H X^T A$ is added to equation (23), where H is the centering matrix. Combined with the objective of equation (24), a new unified objective function is given as follows:

$$\min \frac{\sum_{c=0}^C \text{tr}(A^T X M_c X^T A) + \lambda \|A\|_F^2}{A^T X H X^T A}. \quad (25)$$

Based on the Rayleigh quotient [35], the optimization problem of equation (25) is equivalent to the problem of equation (24) with a fixed value of $A^T X H X^T A$. Finally, we can get the target function of JDA as follows:

$$\min \sum_{c=0}^C \text{tr}(A^T X M_c X^T A) + \lambda \|A\|_F^2 \text{ s.t. } A^T X H X^T A = I. \quad (26)$$

By employing the Lagrange method, the problem of equation (26) can be solved according to the following formula:

$$\left(X \sum_{c=0}^C M_c X^T + \lambda I \right) A = X H X^T A \Phi, \quad (27)$$

where $\Phi = \text{diag}(\phi_1, \dots, \phi_k)$ is the Lagrange multiplier. The solution of the final optimal adaption matrix A can be converted to solve equation (26) to obtain the smallest eigenvectors.

Besides, the initial pseudolabels in the target domain may have many mistakes. After data adaptation, the labeling quality obtained by learning the adaption data can be improved with higher accuracy. Next, the new labels are used for distribution adaptation again. As a result, we can iteratively update the pseudolabels until convergence.

4. The Proposed Fault Diagnosis Method

The purpose of this work is to establish an effective fault diagnosis model that can recognize different bearing faults under variable operating conditions. As discussed above, this situation mainly raises two thorny questions compared with that, under fixed conditions. The first one is the vibration signals could become more complex, which will directly lead to hard work in feature extraction. Specifically, it is important to construct feature spaces with little influence by variable conditions. The second one is how to reduce the distribution discrepancies between different domains. Focused on these two problems, hybrid entropy features consisted by five different information entropies are extracted to characterize the original signal and JDA is employed to reproduce new feature data spaces with well distribution coherence.

Many works indicate that the information entropy algorithm often performs well in characterizing the dynamic and mutability of the nonlinear and nonstationary vibration signals of bearings, where the typical methods include permutation entropy, sample entropy, and energy entropy [17, 36]. In this paper, instead of focusing on the performance comparisons among different entropies for feature selection, we try to obtain comprehensive information about the original signals by feature fusion. For this purpose, five entropy features (PE, SE, EE, SSE, and PSE) which have been proved to be effective in bearings fault diagnosis are integrated to construct the hybrid entropy feature. This feature can characterize the signal from the aspects of reconstruction space, integrity space, energy distribution, and frequency domain. In addition, to obtain more detailed knowledge, the original signals are first decomposed by FEEMD to obtain a collection of IMF components on different frequency bands, and then the hybrid entropy feature is calculated from each IMF component. As for the cases under variable operating conditions, after features

extraction, the direct use of training dataset to train diagnosis model often may cause many misclassifications, which is because the probability distributions between training and testing datasets are different. To eliminate the influence of distribution differences on model construction, knowledge transfer learning between training and testing data space by JDA is performed in this work.

The proposed method mainly includes four parts: signal decomposition, hybrid entropy features extraction, distribution adaptation based on JDA, and faults identification, as shown in Figure 2. The detailed steps can be expressed as follows:

Step 1: obtain the labeled training vibration signal samples $U_s = \{u_i^s, y_i^s\}_{i=1}^{n_s}$ under some known load conditions and the unlabeled testing samples $U_t = \{u_i^t\}_{i=1}^{n_t}$ under other new load conditions. For each signal sample both in training and testing domains, decompose it into a collection of IMF components $\{c_{ik}, i = 1, 2, \dots, n_s + n_t, k = 1, 2, \dots, M\}$ by using FEEMD, where M is the total number of IMF components of each signal sample.

Step 2: according to the algorithms in Section 2.2, calculate the five entropies of each IMF component for each signal sample. Then, assemble these entropies to get the hybrid entropy feature of each signal, recorded as $\{PE_i, SE_i, EE_i, SSE_i, PSE_i\}$, $i = 1, 2, \dots, n_s + n_t$, where each kind of entropy has M values.

Step 3: to complete knowledge transferring, the training set is regarded as the source domain data $X_s = \{x_i, y_i\}_{i=1}^{n_s}$ and the corresponding testing set is seen as the target domain data $X_t = \{x_i\}_{i=n_s+1}^{n_s+n_t}$, where $x_i = \{PE_i, SE_i, EE_i, SSE_i, PSE_i\}$ and y_i is the label data. In JDA, to model $Q_t(Y_t | X_t)$, the KNN method is employed to predict the pseudolabels $\hat{Y}_t$ of the testing samples. Based on JDA and KNN, after many iterations, new training and testing data with better classification quality can be obtained after crossdomain feature learning.

Step 4: the pseudolabel data of the last iteration is considered as the final diagnosis results and the corresponding KNN-based classifier can be obtained to diagnose bearing faults under variable operating conditions.

In the proposed method, KNN is selected as the base classifier since it does not need to tune crossvalidation parameters, and another reason is that it is not suitable to tune the optimal parameters when dealing with data sampled from different distributions.

5. Experimental Studies and Analysis

5.1. Bearing Test Bench and Data Description. This bearing test bench established by Bearing Data Center of Case Western Reserve University (CWRU) has been widely used in the fault diagnosis field [37], and it is composed mainly of a 2 hp Reliance Electric motor, a torque transducer, and a dynamometer, as shown in Figure 3. To simulate the actual

FIGURE 2: Flowchart of the proposed bearing fault diagnosis method.

FIGURE 3: The rolling element bearing experimental bench.

TABLE 1: Fault information of bearings with different fault locations and defect sizes.

Fault location	Rolling element (RE)	Inner race (IR)	Outer race (OR)
Defect size (mm)	0.1778	0.1778	0.1778
	0.3556	0.3556	0.3556
	0.5334	0.5334	0.5334
	0.7112	0.7112	—

failure state, the normal rolling element bearings (6205-2RS deep groove ball bearing) installed on both ends of the motor were damaged artificially by electrical discharge machining, where different single faults were separately seeded on rolling elements, the inner race, or the outer race with different defect sizes. The related fault information about different fault positions and defect sizes is listed in Table 1. Added the normal condition, twelve bearing fault conditions

will be discussed. To simulate the cases under variable operating conditions, this bearing test bench runs under four different loads (0 hp, 1 hp, 2 hp, and 3 hp) and the corresponding rotate speeds are 1797 r/min, 1772 r/min, 1750 r/min, and 1730 r/min.

To collect the vibration signals of bearings, some acceleration sensors were installed on the ends of the motor housing. Here, only the vibration signals measured from the

TABLE 2: The data constitutions of these five cases.

Case no.	Training set		Testing set	
	Loads	Sample number	Loads	Sample number
1	1 hp	480	3 hp	480
2	1 hp	480	2 hp\3 hp	960
3	1 hp	480	0 hp\2 hp\3 hp	1440
4	1 hp\2 hp	960	0 hp\3 hp	960
5	0 hp\1 hp\2 hp	1440	3 hp	480

FIGURE 4: Decomposition results of one normal signal and the FFT amplitude spectrums.

drive end are analyzed in this work and the sampling frequency is 12 kHz. For each load condition, the data length of each fault condition is 120,000 points and each signal sample is averagely divided into 40 subsignals. Hence, there are 1920 ($40 \times 4 \times 12 = 1920$) subsignal samples in total. In order to fully verify the effectiveness of the proposed method, five different fault diagnosis cases with the training and testing sets under different load conditions were constructed, and the data constitutions of these cases are listed in Table 2.

Before experiments and results analysis, we explain the meaning of variable operating conditions first. As shown in Table 2, the five cases were set with training and testing data under different load conditions. From Cases 1–3, the training data was collected under one load condition, and if the historical data is relatively rich, the model can be trained

based on the data with several load conditions, such as in Cases 4 and 5. Generally, the historical data is hard to obtain and scarce, especially with different operating conditions. So, when faults occur under new load conditions and in the cases of lacking rich training data, how to establish an effective model with the limited data is critical. As for these five cases, the points of variable conditions are mainly reflected by the variety of the load conditions between training and testing datasets. Especially from Cases 2–4, we need to identify bearing faults under several new load conditions.

5.2. Results and Analysis. By using the proposed method, at first, each original vibration signal was decomposed by FEEMD into a series of IMF components. With the

FIGURE 5: Decomposition results of one inner race fault signal and the FFT amplitude spectrums.

proposing of FEEMD, the empirical values about the number of ensemble trials, the IMF components number, and the amplitude of the added noises were given [25], and these values are also adopted in this paper. The number of ensemble trials is set to 100, the IMF components number of each signal is set to 10 and the amplitude of the added noise is 0.2 of the standard deviation of the original signal. Owing to space limitation, only the first five IMF components of one normal signal and one inner race fault signal are provided as examples displayed in Figures 4 and 5, respectively. Meanwhile, the FFT amplitude spectrums of the original signal and each IMF component are also given. It can be seen from the FFT spectrum of the original signal that the frequency components are not clear, resulting in some useful information which might be submerged by the strong information components. However, after signal decomposition, the corresponding information of each IMF component is much simpler than that of the original signal. Hence, some important weak information can be revealed by analyzing the IMF components, which is advantageous to the extraction of effective features.

After signal decomposition, by extracting the entropy features described in Section 2.2, we can obtain 10 PE values, 10 SE values, 10 EE values, 10 SSE values, and 10 PSE values of each signal sample. According to the data properties and

experiences, the parameters of PE, SE, and SSE are set as follows. (1) The embedded dimension m is set as 5; (2) The time delay λ of PE extraction is 1 and the similar tolerance γ of SE extraction is 0.2 times of the standard deviation of the corresponding IMF component. Many studies indicate that there may be some redundant and interference components in the decomposition results of FEEMD [38, 39]. As a result, the entropy values obtained from these IMF components can lead to some misclassifications about the bearing conditions. In order to determine which IMF component is good for fault identification, a graphic method to discuss the changing situations of each entropy feature on different IMF components is employed in this paper.

Figure 6 displays the PE distributions of each IMF component of different fault conditions. Here, the fault signals with 0.1778 mm defect size were considered as examples. In Figure 6(a), the PE value of each IMF component is the average value of twenty randomly selected signals. As we know, when faults occur, the fault frequency component and the corresponding multiple frequencies will appear in the collected signals, which easily make the fault signal more complex than the normal signal. Hence, the PE values of three fault signals are larger than that of normal signals, as shown in Figure 6(a). In addition, it can be seen from this figure that the PE values of the first five components of

FIGURE 6: The distribution patterns of PE on different IMF components. (a) The average PE values of each IMF component of different fault conditions. (b) 8 samples of ball fault. (c) 8 samples of inner race fault. (d) 8 samples of outer race fault.

FIGURE 7: Continued.

FIGURE 7: The distribution patterns of SE on different IMF components. (a) The average SE values of each IMF component of different fault conditions. (b) 8 samples of ball fault. (c) 8 samples of inner race fault. (d) 8 samples of outer race fault.

FIGURE 8: The distribution patterns of SSE on different IMF components. (a) The average SSE values of each IMF component of different fault conditions. (b) 8 samples of ball fault. (c) 8 samples of inner race fault. (d) 8 samples of outer race fault.

different fault conditions have obvious differences and the values of the high frequency components are larger than that of the low frequency components. Meanwhile, in Figures 6(c) and 6(d), the PE distributions of 8 randomly selected sample signals of each fault condition are given. The results indicate that the PE values of each fault condition

have high consistency. By comprehensive consideration, the PE values of the first five IMF components perform favorable discrimination ability for different fault types and high aggregation ability for the same fault.

As for SE and SSE, it can be seen from Figures 7 and 8 that the corresponding distribution patterns on different

FIGURE 9: The distribution patterns of EE on different IMF components. (a) The average EE values of each IMF component of different fault conditions. (b) 8 samples of ball fault. (c) 8 samples of inner race fault. (d) 8 samples of outer race fault.

IMF components are similar to that of PE features. The first five values of SE and SSE also show good discrimination for different fault conditions and favorable clustering ability for these samples with the same fault condition. Differed from PE, SE, and SSE, it can be seen from Figures 9(a) and 10(a) that both EE and PSE features can hold high distinguishability for diagnosing different fault conditions by the first nine entropy values. However, for the samples with same fault, the entropy values of IMF6 to IMF10 have obvious differences which can be seen from the other three graphs of Figures 9 and 10. Mostly, this difference may cause wrong diagnosis. In conclusion, for each entropy feature, the entropy values of the first five IMF components with relatively high frequency bands were selected as the final feature data to establish the diagnosis model.

Next, taken Case 4 as an example, the training dataset of hybrid entropy features obtained from different IMF numbers was employed to train three kinds of typical classifiers, including support vector machine (SVM), random forests (RF), and KNN. Then, using these classifiers to recognize all the testing samples, the corresponding diagnosis results are displayed in Figure 11. As can be seen from this figure, the diagnosis accuracies of SVM and RF can reach the maximum values when the IMF number is five. For these two classifiers, the highest diagnostic accuracies are

98.85% and 98.54%, respectively. Meanwhile, when the IMF number is larger than five, the corresponding diagnosis accuracies begin to degrade. As for KNN-based model, satisfactory diagnosis results can be obtained as the IMF number is 4 or 5, which reached nearly 95%. Through comprehensive analysis on the results of Figures 6–11, on the basis of ensuring high diagnosis accuracy and reducing features size, selecting the first five IMF components to construct features space is reasonable.

In order to illustrate the effectiveness of the proposed hybrid entropy features, the diagnosis results of different entropy features are listed in Table 3. As displayed in this table, for each single entropy feature, the diagnosis accuracies of these classifiers exceed 65%, which means each feature contains useful information about the working condition of bearings. However, just based on each single entropy feature to characterize the vibration signal is relative one-sided, some important information contained in other domains cannot be revealed. Hence, by mixing different features to improve diagnosis accuracy can be an effective way, and it has been proved by the results of different classifiers with hybrid entropy (HE) features in Table 3. As for Case 4, since the variation of load conditions between training and testing sets is relatively small, the accuracies sometimes are acceptable just by the traditional diagnosis

FIGURE 10: The distribution patterns of PSE on different IMF components. (a) The average PSE values of each IMF component of different fault conditions. (b) 8 samples of ball fault. (c) 8 samples of inner race fault. (d) 8 samples of outer race fault.

FIGURE 11: Diagnosis results of different classifiers under different IMF numbers.

methods. However, when dealing with more complex diagnosis problems, such as Cases 2 and 3, to enhance the learning ability of the target data about the source data is very important, especially for the problems under variable working

conditions. Hence, the JDA algorithm was employed to

TABLE 3: Diagnosis results of different classifiers with different entropy features.

Classifier	Entropy					
	PE (%)	SE (%)	EE (%)	SSE (%)	PSE (%)	HE (%)
KNN	79.37	75.94	79.79	90.31	65.21	94.06
SVM	77.40	81.15	79.27	76.88	65.42	98.85
RF	77.92	78.96	79.19	87.60	65.94	98.54

reduce the distribution difference between domains.

Table 4 shows the diagnosis results of Cases 1–5 by using JDA-KNN, KNN, SVM, and RF. The regularization parameter λ and the mapping dimension K of JDA are also given in Table 4. As listed in this table, the RF-based model can obtain the best diagnosis results compared with SVM and KNN for these five cases. For Cases 1–3, the diagnosis accuracies of KNN are higher than that of SVM. It indicates that this kind of weak classifier, KNN, performs better in dealing with the problems with insufficient historical data. As for Cases 4 and 5, with more adequate and comprehensive data under different load conditions to train diagnostic models, the diagnosis results of SVM are better than

TABLE 4: Diagnosis results of Cases 1–5 by using JDA-KNN, KNN, SVM, and RF.

	JDA-KNN			KNN (%)	SVM (%)	RF (%)
	λ	K	Accuracy (%)			
Case 1	110	18	99.17	88.75	84.58	95.21
Case 2	50	20	98.12	92.50	91.25	97.50
Case 3	50	20	98.06	92.78	91.25	95.63
Case 4	120	20	99.58	94.06	98.85	98.54
Case 5	50	20	98.06	88.96	92.29	97.08

FIGURE 12: Diagnosis results of Case 1 by using PCA, KPCA, LPP, and JDA under different dimension numbers.

FIGURE 14: Diagnosis results of Case 5 by using PCA, KPCA, LPP, and JDA under different dimension numbers.

FIGURE 13: Diagnosis results of Case 3 by using PCA, KPCA, LPP, and JDA under different dimension numbers.

that of KNN. By using KNN, most of the testing samples of each case can be correctly identified. However, for some cases, such as Cases 1 and 5, the diagnosis results are somewhat disappointed. By using the proposed method, the diagnosis accuracies of all these five cases can be significantly improved, where the results are more than 98%.

Finally, the performance of JDA on the variable conditions problems is compared with some dimensionality reduction techniques, including principal component analysis (PCA), kernel principal component analysis (KPCA), and locality preserving projections (LPP). Here,

just Cases 1, 3, and 5 were considered to further illustrate the effectiveness of JDA. In these experiments, the classifiers were established by using KNN. The comparison results are displayed in Figures 12–Figure 14. As can be seen from these figures, the other three methods perform better on dealing with complex nonlinear dataset than PCA. Obviously, the JDA can hold the highest accuracy when the dimension number is larger than 7 and the fluctuation of its diagnosis accuracies is also very small. Actually, by using the dimensionality reduction technique, some knowledge contained in the source domain data is also transferred to the target domain. Its transformation matrix is obtained based on the source domain data and then the target data can be projected into a new space. Like in PCA, its purpose is to get new vectors with relative large variance. However, in JDA, the transformation matrix is formed based on the source and target domain data, where the distribution discrepancies were considered during the transformation process, resulting in strong transfer learning ability. Hence, in a sense, JDA can also be an alternative and effective dimensionality reduction method, which is useful for feature extraction in the fault diagnosis field.

6. Conclusions

In the engineering practice, the faults of bearings may occur in different load conditions, which will lead to significant distribution differences between the sample sets with different load conditions. Generally, these differences can lead to a poor diagnosis model, especially dealing with samples

under new load conditions. In this paper, a novel diagnosis model based on hybrid multiscale entropy features and joint distribution adaptation is proposed to solve this kind of fault diagnosis problem. For feature extraction, five complexity measurement parameters, namely, PE, SE, EE, SSE, and PSE, were calculated from IMF components. By using different classifiers, the results show that the hybrid entropy features can obtain higher accuracies than the single entropy features do. The hybrid entropy features can effectively find the similar and potential peculiarity of signals with different load conditions. After feature extraction, the final state identification of bearings is completed by using JDA and KNN. The ability of transfer learning of JDA is compared with other nontransfer learning methods. Through extensive experiments on the five cases, the results show that the JDA with KNN outperforms the state-of-the-art approaches and the proposed method can hold high stability when dealing with cases with different load conditions. This work indicates that the JDA method shows excellent application foreground in fault diagnosis fields. However, the transfer learning-based fault diagnosis is still in the early stage. In future work, we will pursue more case studies on the real data and the parameters optimization of JDA to achieve better transfer learning results.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 51709122) and the open fund for Jiangsu Key Laboratory of Advanced Manufacturing Technology (Grant no. HGAMTL-1713). Meanwhile, the authors thank Case Western Reserve University for providing the bearing fault data freely.

References

- [1] W. Fu, J. Tan, X. Zhang, T. Chen, and K. Wang, "Blind parameter identification of mar model and mutation hybrid gwo-sca optimized svm for fault diagnosis of rotating machinery," *Complexity*, vol. 2019, Article ID 3264969, 17 pages, 2019.
- [2] X. Yu, F. Dong, E. Ding, S. Wu, and C. Fan, "Rolling bearing fault diagnosis using modified lfd and emd with sensitive feature selection," *IEEE Access*, vol. 16, pp. 3715–3730, 2017.
- [3] Y. Zhang, H. Zuo, and F. Bai, "Classification of fault location and performance degradation of a roller bearing," *Measurement*, vol. 46, no. 3, pp. 1178–1189, 2013.
- [4] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Transactions*, vol. 77, pp. 167–178, 2018.
- [5] W. Huang, J. Cheng, and Y. Yang, "Rolling bearing fault diagnosis and performance degradation assessment under variable operation conditions based on nuisance attribute projection," *Mechanical Systems and Signal Processing*, vol. 114, pp. 165–188, 2019.
- [6] Y. Tian, J. Ma, C. Lu, and Z. Wang, "Rolling bearing fault diagnosis under variable conditions using lmd-svd and extreme learning machine," *Mechanism and Machine Theory*, vol. 90, pp. 175–186, 2015.
- [7] Z. Xing, J. Qu, Y. Chai, Q. Tang, and Y. Zhou, "Gear fault diagnosis under variable conditions with intrinsic time-scale decomposition-singular value decomposition and support vector machine," *Journal of Mechanical Science and Technology*, vol. 31, no. 2, pp. 545–553, 2017.
- [8] P. Baraldi, F. Cannarile, F. Di Maio, and E. Zio, "Hierarchical k -nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions," *Engineering Applications of Artificial Intelligence*, vol. 56, pp. 1–13, 2016.
- [9] X. Li, W. Mao, and W. Jiang, "Extreme learning machine based transfer learning for data classification," *Neurocomputing*, vol. 174, pp. 203–210, 2016.
- [10] 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, 2010.
- [11] Q. Wang, G. Michau, and O. Fink, "In Domain adaptive transfer learning for fault diagnosis," in *Proceedings of the In Prognostics and System Health Management Conference*, IEEE, Paris, France, May 2019.
- [12] 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, 2011.
- [13] B. Quanz, J. Huan, and M. Mishra, "Knowledge transfer with low-quality data: a feature extraction issue," in *Proceedings of the IEEE International Conference on Data Engineering*, IEEE, Hannover, Germany, April 2011.
- [14] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, IEEE, Sydney, Australia, December 2013.
- [15] T. Han, C. Liu, W. Yang, and D. Jiang, "Deep transfer network with joint distribution adaptation: a new intelligent fault diagnosis framework for industry application," *ISA Transactions*, vol. 97, pp. 269–281, 2020.
- [16] Y. Wang, G. Xu, L. Liang, and K. Jiang, "Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 54–55, pp. 259–276, 2015.
- [17] X. Zhang, Y. Liang, J. Zhou, and Y. Zang, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, 2015.
- [18] M. Han and J. Pan, "A fault diagnosis method combined with lmd, sample entropy and energy ratio for roller bearings," *Measurement*, vol. 76, pp. 7–19, 2015.
- [19] J. Huang, X. Hu, and X. Geng, "An intelligent fault diagnosis method of high voltage circuit breaker based on improved emd energy entropy and multi-class support vector machine," *Electric Power Systems Research*, vol. 81, no. 2, pp. 400–407, 2011.
- [20] L. Yong, Y. Li, X. Han, and Z. Wang, "A sliding singular spectrum entropy method and its application to gear fault diagnosis," in *Proceedings of the 4th International Conference*

- On Intelligent Computing*, IEEE, Shanghai, China, pp. 15–18, September 2008.
- [21] C. W. Fei, G.-C. Bai, W. Z. Tang, and S. Ma, “Quantitative diagnosis of rotor vibration fault using process power spectrum entropy and support vector machine method,” *Shock and Vibration*, vol. 2014, Article ID 957531, 9 pages, 2014.
 - [22] Q. Jiang, Y. Shen, H. Li, and F. Xu, “New fault recognition method for rotary machinery based on information entropy and a probabilistic neural network,” *Sensors*, vol. 18, no. 2, pp. 337–349, 2018.
 - [23] Y. Wang, C. Yeh, H. Young, K. Hu, and M. Lo, “On the computational complexity of the empirical mode decomposition algorithm,” *Physica A Statistical Mechanics & Its Applications*, vol. 400, pp. 159–167, 2014.
 - [24] N. E. Huang, Z. Shen, S. R. Long et al., “The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
 - [25] Z. Wu and N. E. Huang, “Ensemble empirical mode decomposition: a noise-assisted data analysis method,” *Advances in Adaptive Data Analysis*, vol. 01, no. 1, pp. 1–41, 2009.
 - [26] J. Zheng, J. Cheng, and Y. Yang, “Partly ensemble empirical mode decomposition: an improved noise-assisted method for eliminating mode mixing,” *Signal Processing*, vol. 96, pp. 362–374, 2014.
 - [27] Z. Jian, R. Yan, R. X. Gao, and Z. Feng, “Performance enhancement of ensemble empirical mode decomposition,” *Mechanical Systems & Signal Processing*, vol. 24, pp. 2104–2123, 2010.
 - [28] L. H. Thomas, *Elliptic Problems in Linear Difference Equations over a Network*, Columbia University, New York, NY, USA, 1949.
 - [29] C. Bandt and B. Pompe, “Permutation entropy: a natural complexity measure for time series,” *Physical Review Letters*, vol. 88, pp. 174102–174100, 2002.
 - [30] 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. 10, pp. 2297–2301, 1991.
 - [31] J. S. Richman, D. E. Lake, and J. R. Moorman, “Sample entropy,” *Methods in Enzymology*, vol. 384, pp. 172–184, 2004.
 - [32] Y. Xiao, N. Kang, Y. Hong, and G. Zhang, “Misalignment fault diagnosis of DFWT based on IEMD energy entropy and PSO-SVM,” *Entropy*, vol. 19, no. 1, pp. 6–20, 2017.
 - [33] T. Li, X. He, J. Tang, H. Zeng, and C. Zhou, *Deep Learning for Classification of Islanding and Grid Disturbance Based on Multi-Resolution Singular Spectrum Entropy*, IOP Publishing, Bristol, UK, 2018.
 - [34] A. Gretton, K. Borgwardt, M. J. Rasch, B. Scholkopf, and A. J. Smola, “A kernel method for the two-sample problem,” *Computer Science*, vol. 1, pp. 1–10, 2008.
 - [35] J.-g. Sun, “Eigenvalues of Rayleigh quotient matrices,” *Numerische Mathematik*, vol. 59, no. 1, pp. 603–614, 1991.
 - [36] Z. Long, G. Xiong, H. Liu, H. Zou, and W. Guo, “Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference,” *Expert Systems with Applications*, vol. 37, pp. 6077–6085, 2010.
 - [37] W. A. Smith and R. B. Randall, “Rolling element bearing diagnostics using the case western reserve university data: a benchmark study,” *Mechanical Systems and Signal Processing*, vol. 64–65, pp. 100–131, 2015.
 - [38] Y. Lei, Z. He, and Y. Zi, “Application of the emd method to rotor fault diagnosis of rotating machinery,” *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1327–1338, 2009.
 - [39] X. Xue and J. Zhou, “A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery,” *ISA Transactions*, vol. 66, pp. 284–295, 2016.