Research Article

Application of Feature Fusion Using Coaxial Vibration Signal for Diagnosis of Rolling Element Bearings

Jing Jiao , Jianhai Yue, Di Pei, and Zhunqing Hu

School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Jing Jiao; jingjiao@bjtu.edu.cn

Received 22 July 2020; Revised 5 September 2020; Accepted 16 September 2020; Published 1 October 2020

Academic Editor: Baoqing Li

Copyright © 2020 Jing Jiao 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.

The research of rolling element bearings (REBs) fault diagnosis based on single sensor vibration signal analysis is very common. However, the information provided by an individual sensor is very limited, and the robustness of the system is poor. In this paper, a novel fault diagnosis method based on coaxial vibration signal feature fusion (CVSFF) is proposed to fully analyze the multisensor information of the system and build a more reliable diagnostic system. An ensemble empirical mode decomposition (EEMD) method is used to decompose the original vibration signal into a number of intrinsic mode functions (IMFs). Then the autocorrelation analysis is introduced to reduce the random noise remaining in IMFs. After that, the Rényi entropy is calculated as the feature of bearings. Finally, the features of coaxial vibration signal are fused by a multiple-kernel learning support vector machine (MKL-SVM) to classify bearing conditions. In order to verify the effectiveness of the CVSFF method in REB diagnosis, eight data sets from the Case Western Reserve University Bearing Data Center are selected. The fault classification results demonstrate that the proposed approach is a valuable tool for bearing faults detection, and the fused feature from coaxial sensors improves fault classification accuracy for REBs.

1. Introduction

The state of bearing is very important to efficient operation of mechanical equipment. Fault bearing will cause periodic impacts on vibration, which will lead to problems of other parts of mechanical system. Therefore, it is of great significance to find out bearing fault in time and replace bearing to avoid the breakdown of machine.

With the development of information technology, judging bearing state by signal analysis has become an important trend of condition-based monitoring (CBM) [1]. In recent years, many methods have been applied to fault detection of REBs, such as vibration signal, oil, temperature, and acoustic emission analysis. Among all these methods, vibration signal analysis is the most widely used and effective method owing to the great information that vibration signal contains [2, 3]. If bearing is operating with local spall, it will cause vibration impulse [4]. Through analysis of the collected vibration signals, the fault impact characteristics can be obtained, so as to

realize bearing fault diagnosis. In the research of bearing fault diagnosis based on vibration signals, many papers focus on the signal collected from a single sensor. However, for the complex mechanical system, it is uncertain to use the information of an individual sensor for fault diagnosis, which will lead to unreliable diagnosis results in many cases.

Information fusion is a technology that merges data to obtain more consistent, informative, and accurate information than the original raw data that are mostly uncertain [5]. Some scholars have made achievements in bearing fault diagnosis based on multisensor information fusion technology. These research studies are mainly divided into multisource data fusion relating to the data from the same kind of sensors and different kinds of sensors. Owing to the advantages of vibration signals, many scholars adopt the method of multiple vibration sensor signal fusion in bearing fault diagnosis. These signal fusion studies can be classified as data-level fusion, feature-level fusion, and decision-level fusion.

Yan et al. [6] proposed a concept of space-time fragments, and vibration signals captured by multiple sensors are fused at data level. For feature-level fusion, Jiang et al. [7] developed a feature-level information fusion methodology of fault diagnosis in rotating machinery. Tao et al. [8] presented a novel bearing fault diagnosis method using multivibration signals and deep belief network (DBN). Banerjee and Das [9] suggested a hybrid model for an on board fault-tolerant control system by vibration data fusion. Wang et al. [10] fused vibration data features as the health indicator of bearing status and got a good representation of bearing defect conditions. Zhou et al. [11] addressed a feature fusion approach based on NCA and coupled hidden Markov model. Hong et al. [12] introduced a preprocessing model of bearing using wavelet packet-empirical mode decomposition (WP-EMD) for feature extraction of vertical and horizontal vibrations. For decision-level fusion, Hui et al. [13] proposed an automated bearing fault diagnosis model that employs SVM and the Dempster-Shafer evidence theory in classification.

In the aspect of different kinds of signals fusion, some scholars have carried out bearing fault diagnosis based on the fusion of vibration signals and other types of signals. Safizadeh and Latifi [14] presented a new method of bearing fault diagnosis using the fusion of two primary sensors: an accelerometer and a load cell. Lu et al. [15] designed a sound-aided vibration signal adaptive stochastic resonance (SAVASR) method for bearing fault detection. All those works indicate that multisensor information fusion methods have higher classification accuracy than single sensor data analysis in bearing fault diagnosis.

Nevertheless, there are still some problems in vibration signal source of the above bearing fault diagnosis research studies. The authors of [6, 9, 10] do not specify the installation location of vibration sensors. The experiments of papers [7, 9] have too many sensors, which make the system more complex. The researchers [11, 12, 14, 15] only collect vibration signals of one or multidirections of a single bearing. In the experiment Case 2 of [11], vibration sensors are only installed on the base of the test rig. As a matter of fact, the location of vibration sensor is very important in fault diagnosis of mechanical equipment. Vibration data are composed of multiple vibration source signals and noise signals. It must pass through multiple interfaces to reach an accelerometer, which will definitely cause energy dissipation. Because a shaft and bearing inner race are rigidly connected, the shaft plays a role of vibration transmission between coaxial bearings. The fault signals of coaxial bearings are usually similar in a frequency domain. Therefore, the coaxial vibration signal feature fusion (CVSFF) algorithm proposed in this paper takes vibration sensors of coaxial bearings as data sources. Enough information about REBs can be obtained from limited multisource data. Then, the characteristics of this information are fused to realize bearing fault diagnosis.

In the research of bearing fault diagnosis based on information fusion technology, feature-level fusion is widely used. Among the existing algorithms, a support vector machine (SVM) has been widely used because of its good

classification performance. The researchers of papers [7, 9, 13] used SVM as a classifier for fault classification and have achieved good results. However, the classification performance of SVM is greatly affected by a kernel function. The determination of kernel function depends on human experience [16]. To solve this problem, multiple kernel learning (MKL) methods have been proposed. MKL learns the kernel function and classifier parameters simultaneously, which can effectively solve the problem of kernel function selection. Meanwhile, the SVM trained by MKL has more flexibility and higher classification accuracy. Many scholars have applied MKL to SVM model optimization and obtained good results [16, 17].

Based on the abovementioned analysis, a new method of fault diagnosis based on CVSFF is presented in this paper. First, an EEMD method is used to decompose the original vibration signals collected from bearings at both ends of a shaft. Then, autocorrelation is carried out to reduce the random noise in IMF components. After that, an energy ratio of each IMF component is calculated to extract probability mass function (PMF). The Rényi entropy feature matrix of coaxial sensors is obtained based on PMF. Finally, different states of REBs are classified by MKL-SVM.

This paper is organized as follows: the proposed method of CVSFF is presented in the next section. In Section 3, the algorithm is validated and analyzed by experimental data. Meanwhile, MKL-SVM classification based on an individual sensor and SVM, genetic algorithm-optimized SVM (GA-SVM), and particle swarm optimization SVM (PSO-SVM) based on coaxial features are carried out to evaluate the effectiveness of CVSFF. Section 4 presents the conclusion.

2. Methods

2.1. EEMD. There are many time-frequency data analysis methods to decompose nonlinear and nonstationary time series into a set of components, such as empirical mode decomposition (EMD) [18], ensemble empirical mode decomposition (EEMD) [19], variational mode decomposition (VMD) [20], and broadband mode decomposition (BMD) [21, 22]. Among all those methods, EEMD is the most widely used algorithm in CBM of REBs. EEMD can significantly improve the decomposition effect by reducing mode mixing. The two important parameters used in the EEMD algorithm are ratio k for standard deviation of white noise to standard deviation of signal and the total number M of EMD. Huang suggested that k is generally set as 0.2 [19]. Moreover, the amplitude of white noise should be reduced appropriately for the signal mainly composed of high-frequency components and increased for the signal mainly composed of low-frequency components. Besides, Huang found that M follows the statistical law of the following equation.

$$\varepsilon_n = \left(\frac{k}{\sqrt{M}} \right) \text{ or } \ln \varepsilon_n + \left(\frac{k}{2} \right) \ln M = 0, \quad (1)$$

where ε_n is the maximum relative error of signal decomposition. In this paper, M is taken as 100, and k is taken as 0.2.

The specific decomposition steps of EEMD are as follows:

- (1) Add random white noise $k\sigma_x n(t)$ to $x(t)$ as shown in the following equation, where $n(t)$ is the Gaussian white noise with mean value 0 and standard deviation 1 and σ_x is the standard deviation of signal $x(t)$:

$$x_m(t) = x(t) + k\sigma_x n_m(t). \quad (2)$$

- (2) EMD decomposition of signals $x_m(t)$.
- (3) Repeat Steps 1 and 2 for M times.
- (4) The average value of each IMF component is obtained by decomposing M times EMD, and the global average is obtained:

$$\bar{c}_i(t) = \frac{1}{M} \sum_{m=1}^M c_{i,m}(i = 1, 2, \dots, n), \quad (3)$$

$$\bar{r}_n(t) = \frac{1}{M} \sum_{m=1}^M r_{n,m},$$

where $\bar{c}_i(t)$ ($i = 1, \dots, n$) and $\bar{r}_n(t)$ are the i^{th} IMF and residual component, respectively.

2.2. Autocorrelation. An autocorrelation function describes the relationship of a signal at different times. The autocorrelation function of signal $x(t)$ is defined as

$$R_x(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)x(t+\tau)dt. \quad (4)$$

According to the properties of autocorrelation function, a periodic signal has the same cycle as an original signal after autocorrelation. Furthermore, the autocorrelation function of random noise attenuates quickly and tends to zero with the increase of time delay τ . If the periodic signal contains random noise, the autocorrelation function can be used to reduce noise. In this paper, autocorrelation function is applied to noise reduction of IMFs, so as to retain the useful periodic signals in IMFs and reduce random white noise.

2.3. Rényi Entropy. Self-information $I(x_i)$ refers to the amount of information contained in an event x_i of a physical system. Different events in a physical system contain diverse amounts of information. So $I(x_i)$ is a random variable that cannot be used as a measure of information about the whole system.

Shannon [23] defines the mathematical expectation of information as information entropy, that is, the average amount of information in a source:

$$S(p) = E[-\log_2(p_i)] = -\sum_{i=1}^N p_i \log_2 p_i, \quad (5)$$

where p_i is the probability mass function (PMF).

Shannon entropy is a nonparametric measure, and it is well known that it does not have significant sensitivity when dealing with noisy data. For this reason, Rényi entropy [24]

has been chosen as another potential entropy measure, which is defined as follows:

$$R(p) = \frac{1}{1-\alpha} \log_2 \sum_{i=1}^N p_i^{\alpha-1}. \quad (6)$$

The parameter α in Rényi entropy can be used to make the entropy more or less sensitive to particular segments of the probability distributions. $\alpha \rightarrow 0$ causes the Rényi entropy to become highly sensitive to changes in the tails of the distribution. And for $\alpha \rightarrow 1$, it reduces to the Shannon entropy, and hence, the Rényi entropy becomes more sensitive in the regions where the bulk of PMF is located [25].

In order to obtain the frequency distribution change of vibration signal's energy for rolling bearing, p_i is the energy ratio of IMFs to total energy:

$$p_i = \left(\frac{E_i}{E} \right), \quad (7)$$

where E_i is the energy of IMF _{i} , $E_i = \sum_{j=1}^N (\bar{c}_{i,j}(t))^2$, ($i = 1, \dots, n; j = 1, \dots, N$), n is the number of IMFs, N is the number of points of signal $\bar{c}_{i,j}(t)$, and $E = \sum_{i=1}^n E_i$.

The Rényi entropy is a parameter that characterizes the statistical properties of random variables and reflects the randomness of variables. The energy distribution of normal rolling bearing vibration signal in each frequency band is uniform. It means that the energy distribution is uncertain, so entropy is relatively large. When rolling bearing spall occurs, the energy is mainly distributed in the resonance frequency. The uncertainty of energy distribution is relatively reduced, so the entropy decreases. Therefore, the Rényi entropy is a sensitive feature for REBs classification.

2.4. MKL-SVM. SVM is a machine learning method based on statistical learning theory and structural risk minimization principle and is developed by Vapnik and his group [26]. The optimization classification function is as follows:

$$f(x) = \sum_{i=1}^n \alpha_i^* K(\mathbf{x}, \mathbf{x}_i) + b^*, \quad (8)$$

where $K(\cdot, \cdot)$ is the kernel function associated with a reproducing kernel Hilbert space (RKHS) H , $\mathbf{x}_i$ is the i^{th} training data, $\mathbf{x}$ is the data to be classified, and α_i^* and b^* are the unknown coefficients.

There are many kinds of kernel functions, such as linear kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j)$; polynomial kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j)^d$; Gaussian radial basis kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$; and sigmoid kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\beta \mathbf{x}_i^T \mathbf{x}_j + \theta)$.

The kernel function and SVM model parameters directly affect the performance of the SVM classifier. However, it is difficult to map the sample into a high-dimensional feature space by using a single kernel function for complex classification problem. In recent years, many scholars have carried out relevant research on MKL-SVM. In order to select the appropriate kernel function, MKL-SVM learns by combining different kernel functions and thus has more flexibility, better generalization ability, and stronger model interpretation ability.

In MKL framework, $K(\cdot, \cdot)$ is a convex linear combination of a set of basic kernels:

$$\begin{aligned} K(\mathbf{x}, \mathbf{x}_i) &= \sum_m d_m K_m(\mathbf{x}, \mathbf{x}_i), \\ \text{s.t. } d_m &\geq 0, \sum_m d_m = 1, \end{aligned} \quad (9)$$

where d_m is the weight of $K_m(\cdot, \cdot)$ obtained by sample learning. The decision function of MKL is as follows:

$$f(x) = \sum_m f_m(x) + b, \quad (10)$$

where each function f_m belongs to a different RKHS H_m associated with a kernel K_m . Rakotomamonjy et al. [27] proposed a simple MKL method to learn both the coefficients α_i^* and the weights d_m . They adopted the gradient method to solve the MKL problem. The optimal solution is obtained by calculating the gradient of the objective function about to d_m :

$$\begin{aligned} \min_{\{f_m\}, b, \xi, d} & \left(\frac{1}{2} \right) \sum_m \frac{1}{d_m} \|f_m\|_{H_m}^2 + C \sum_i \xi_i, \\ \text{s.t. } & y_i \sum_m f_m(x_i) + y_i b \geq 1 - \xi_i, \forall i, \\ & \xi_i \geq 0, \forall i, \\ & \sum_m d_m = 1, d_m \geq 0, \forall m. \end{aligned} \quad (11)$$

Finally, the decision functions of MKL-SVM are obtained as follows:

$$f(x) = \sum_i \alpha_i^* y_i \left[\sum_m d_m K_m(x, x_i) \right] + b^*. \quad (12)$$

2.5. Structure of CVSFF. The method process of CVSFF proposed in this paper is as follows:

- Step 1: collecting signals from coaxial vibration sensors
- Step 2: decompose vibration data through EEMD and obtain respective IMFs
- Step 3: denoise the IMFs by the autocorrelation function
- Step 4: extract the Rényi entropy of IMF components
- Step 5: randomly select the training set and test set from feature matrix
- Step 6: train MKL-SVMs
- Step 7: input the test set into MKL-SVM models, and output the classification results

The method flowchart is shown in Figure 1.

3. Experiments Analysis and Discussion

3.1. Data Set. The bearing data diagnosed in this paper were obtained from the Case Western Reserve University (CWRU) Bearing Data Center. These data sets have been

considered as a benchmark and analyzed by many researchers. As shown in Figure 2, the test rig consists of a 2 hp motor, a torque transducer, and a dynamometer.

Two coaxial accelerometers are installed at both ends of the motor to measure vibration signal of the corresponding bearing. In this paper, these two coaxial vibration signals in the drive end and fan end are used for fault diagnosis.

Smith and Randall [28] used three established bearing diagnostic techniques to provide a benchmark analysis of these widely used data sets. According to the data analysis conclusion in [28], data of eight states that are difficult to diagnose by the benchmark method were selected for analysis in this paper, as shown in Table 1.

The last column in Table 1 means categorisation of the benchmark method proposed in that paper. The explanation of diagnosis category is as follows:

Y1: data clearly diagnosable and showing classic characteristics for the given bearing fault in both the time and frequency domains

Y2: data clearly diagnosable but showing nonclassic characteristics in either or both of the time and frequency domains

P1: data probably diagnosable; e.g., the envelope spectrum shows discrete component at the expected fault frequencies, but they are not dominant in the spectrum

P2: data potentially diagnosable; e.g., the envelope spectrum shows smeared components that appear to coincide with the expected fault frequencies

N1: data not diagnosable for the specified bearing fault but with other identifiable problems (e.g., looseness)

N2: data not diagnosable and virtually indistinguishable from noise, with the possible exception of shaft harmonics in the envelope spectrum

Randall only used the benchmark method to analyze four sets of data with data set number 169, 170, 171, and 172 for inner ring fault of the drive end with 12 kHz sampling frequency. Data set 171 was selected from the four groups with poor diagnostic results. The reason for choosing data set 276 is the same as that for choosing data set 171. Except these two sets and normal bearing data, data sets in Table 1 are not diagnosable by the benchmark method in [28].

3.2. EEMD Analysis. Vibration signals of two coaxial sensors corresponding to eight states were cut into 1600 segments with ten seconds data length (120,000 points) for each state. Each state of bearing has 200 segments from two coaxial sensors (each sensor has 100 segments; each segment has 1200 points). Every segment of data is decomposed into nine IMFs and a residual component by EEMD. Drive-end raw time waveforms and part of decomposition results of eight different states are given in Figure 3. Only original signal and the first seven orders of IMFs are shown in the figure.

In the time-domain distribution of IMF, there is no fault-related impulse component. In order to view the frequency distribution of each IMF, fast Fourier transform (FFT) was

FIGURE 1: Flowchart of CVSFF.

FIGURE 2: Schematic diagram of the bearing test rig.

TABLE 1: Data set.

No.	Data set number	Bearing type	Symbol in this paper	Results of [28] DE/FE
1	99	Normal	N	—
2	171	Drive-end inner race fault	I-D	Y2/P1
3	120	Drive-end ball fault	B-D	N1/P2
4	198	Drive-end outer race fault (centred)	O6-D	N2/N2
5	276	Fan-end inner race fault	I-F	N1/Y2
6	292	Fan-end roller fault	B-F	N1/N1
7	301	Fan-end outer race fault (orthogonal)	O3-F	N1/P2
8	307	Fan-end outer race fault (opposite)	O12-F	N1/N1

performed for each component. Take the first segment of data for eight states as examples, IMF frequency-domain distribution of these data is shown in Figure 4.

It can be seen in Figure 4 that the energy of IMFs is mainly concentrated in IMF1. Due to the characteristics of EEMD, the frequency content of IMFs ranges from high to low with the increase of order. Moreover, EEMD separates the frequency bands successfully without mode mixing. It may also be noticed that the frequency distribution of normal bearing coaxial sensor data is very similar by comparing Figures 4(a) and 4(b). In Figures 4(c) to 4(p), which are fault bearing coaxial data, components except IMF1 have similar frequency distribution too. Meanwhile, it can be seen from the red box in Figure 4 that the frequency distribution of IMF components is spread over many frequency bands. The random noise in IMFs submerged part of the fault impulse, thus affecting the subsequent feature extraction effect. Therefore, the method of autocorrelation noise reduction is needed to reduce random noise in IMF components.

3.3. Autocorrelation Noise Reduction. In order to suppress the noise, the autocorrelation function of each IMF is computed in the time domain. Figure 5 is the frequency distribution of the corresponding data in Figure 4 after autocorrelation denoise.

By comparing Figures 5 and 4, it can be seen that the frequency amplitude marked by the red box in Figure 4 significantly decreases after the autocorrelation noise reduction. And the frequency band corresponding to the large amplitude of each component does not change.

The frequency distribution of IMFs is more centralized, which means random noise in the component is effectively suppressed after autocorrelation. However, Res in Figures 5(b), 5(d), and 5(f) show a same low-frequency peak distribution of 10 Hz marked in the red dotted box. We take Res of Figure 5(d) as an example to analyze this phenomenon. The time domain, autocorrelation function, and corresponding frequency-domain distribution of this component are shown in Figure 6.

FIGURE 3: Continued.

FIGURE 3: The EEMD results of eight states. (a) N. (b) I-D. (c) B-D. (d) O6-D. (e) I-F. (f) B-F. (g) O3-F. (h) O12-F.

FIGURE 4: Continued.

FIGURE 4: Frequency distribution of each IMF of the original signal. (a) N (drive end). (b) N (fan end). (c) I-D (drive end). (d) B-F (fan end). (e) B-D (drive end). (f) B-D (fan end). (g) O6-D (drive end). (h) O6-D (fan end). (i) I-F (drive end). (j) I-F (fan end). (k) B-F (drive end). (l) B-F (fan end). (m) O3-F (drive end). (n) O3-F (fan end). (o) O12-F (drive end). (p) O12-F (fan end). Res means residual component.

FIGURE 5: Frequency distribution of each IMFs after autocorrelation denoise. (a) N (drive end). (b) N (fan end). (c) I-D (drive end). (d) B-F (fan end). (e) B-D (drive end). (f) B-D (fan end). (g) O6-D (drive end). (h) O6-D (fan end). (i) I-F (drive end). (j) I-F (fan end). (k) B-F (drive end). (l) B-F (fan end). (m) O3-F (drive end). (n) O3-F (fan end). (o) O12-F (drive end). (p) O12-F (fan end). Res means residual component.

As can be seen from Figure 6, the signal becomes a monotone signal after autocorrelation. So, the frequency peak of 10 Hz in Figures 5(b), 5(d), and 5(f) correspond to the time length 0.1 s of signal. It has no actual physical meaning. Since this paper only uses IMF1-9 components to calculate the entropy value, it has no influence on the feature matrix distribution.

3.4. Entropy Feature Extraction. According to equations (6) and (7), we calculated the Rényi entropy of IMFs. The dimension of entropy multisensor feature matrix of each state is 100×2 . The row 100 of feature matrix means 100 segments of each state. The column 2 of feature matrix means 2 coaxial sensors. Table 2 shows the mean values of entropy for eight states which are calculated by taking an average of 100 Rényi entropy values for every state of bearings. The entropy values of IMFs without and with denoise are compared in Table 2.

The variation in Table 2 means the change rate of entropy mean value for the noise-reduced data comparing with original IMFs:

$$\text{variation} = \frac{(R_1 - R_2)}{R_1}, \quad (13)$$

where R_1 is the mean value of entropy for data without denoise and R_2 is mean value of entropy for data with autocorrelation denoise.

As can be seen from Table 2, the entropy of all measuring points decreases after autocorrelation. Arrays with a large reduction (greater than 50%) are highlighted in bold. The maximum reduction is 81.53%. It can also be seen from Table 2 that entropy of data from sensors located in faulty bearing end generally drops greatly. This may be because the vibration signal at the fault bearing end is more susceptible to noise interference.

Figure 7 is the Rényi entropy box plot calculated by vibration signals of two coaxial sensors. Figure 7(a) is the Rényi entropy distribution for data from the drive-end sensor. Figure 7(b) is the Rényi entropy distribution for data from the fan-end sensor. The black part in Figure 7 is the entropy value calculated from raw data, while the red part is the entropy value calculated by IMFs after autocorrelation noise reduction. It is clear that all entropy decreased after noise reduction. Furthermore, the entropy distribution for data without noise reduction from drive-end or fan-end sensor is similar to that after autocorrelation denoise. Although autocorrelation denoise hardly changes the entropy tendency of bearings in different states, it increases the differentiation of eight states. In addition, the entropy values of normal bearing in both sensors are greater than those of all faulty bearings which are consistent with the characteristics of the Rényi entropy.

In order to check the distribution of the outlier in the box diagram, we counted the number of outliers in the box plot of Figure 7. It can be clearly seen from Table 3 that the entropy distribution becomes more centralized and the number of outlier decreases after vibration noise reduction. The overall decrease of outliers was attributed to bearing

states of I-F, O3-F, and O12-F. That is, the reduction of the outlier mainly consisted of coaxial data of fan-end fault bearing.

The Rényi entropy of the vibration signal at drive end and fan end constitutes an 800×2 dimensional multisensor feature vector. To compare the distribution of feature matrix without and with noise reduction more distinctly, we draw the feature scatter plot. Their distributions are shown in Figure 8.

As can be observed, features from all conditions except normal bearing have relatively concentrated space distribution in Figure 8(a). In Figure 8(b), although conditions B-F, O3-F, and O12-F still have some similar features, the aliasing of the feature distributions of different states has been significantly improved after noise reduction. The characteristic differentiation degree of bearing in eight different states is increased.

3.5. Result Comparison and Discussion. Through all above signal preprocessing methods carried out in Section 3.2 to 3.4, we obtained the coaxial vibration feature matrix. To classify eight different states of bearing, MKL-SVM is used in this paper. Kernel functions are applied to the features of single sensor and all sensors.

Common kernel functions are divided into two types: local kernel function and global kernel function. The local kernel function has local characteristics, strong learning ability, and weak generalization ability. Linear and Gaussian kernel functions are typical local kernel function. On the contrary, the global kernel function has global characteristics, strong generalization ability, and weak learning ability. Polynomial and sigmoid kernel functions are typical global kernel functions.

In order to make the model have a good classification effect for data with different characteristics, the Gaussian kernel and polynomial kernel functions with different parameters were selected in this paper. Detailed parameters are as follows:

- (1) Five Gaussian kernel functions with different bandwidths: bandwidth σ is uniformly sampled at the interval of $[0.01, 100]$ on a logarithmic scale.
- (2) Three different levels of polynomial function: the orders d are 1, 2, and 3.

These kernel functions are applied to a single feature and all features, respectively. Therefore, there are 24 kernels in total.

3.5.1. Comparison and Discussion of Coaxial Signals with Individual Signal. In order to verify the sensitivity of coaxial signal features, we also did the same diagnosis based on single sensor. To train MKL-SVMs, 70% of feature matrix is selected as training data. The *one against all* method is adopted to construct multiclassification model. To improve computational efficiency, we jointly optimized eight binary classification problems. A total of eight MKL-SVMs and a combination of kernels are obtained for coaxial signals or

FIGURE 6: Time domain, autocorrelation function, and corresponding frequency-domain distribution of Res in Figure 5(d).

TABLE 2: Mean values of the Rényi entropy.

Sensor location	1 N	2 I-D	3 B-D	4 O6-D	5 I-F	6 B-F	7 O3-F	8 O12-F
<i>Without denoise</i>								
Drive end	2.5928	1.2033	1.0874	1.5331	1.6016	1.8180	1.7362	1.6746
Fan end	2.5244	1.9789	2.0922	1.6853	1.9023	1.3981	1.7274	1.4613
<i>Autocorrelation denoise</i>								
Drive end	1.9548	0.4336	0.2008	0.4485	1.0307	1.4850	1.2504	1.0036
Variation (%)	24.61	63.96	81.53	70.74	35.65	18.32	27.98	40.07
Fan end	1.9314	1.4479	1.6025	0.4397	1.2879	0.4887	0.5916	0.5036
Variation (%)	23.49	26.83	23.41	73.91	32.30	65.05	65.75	65.54

FIGURE 7: Box plot of the Rényi entropy distribution of (a) drive end and (b) fan end.

TABLE 3: Box plot outlier quantity of the Rényi entropy.

Signal location	1 N	2 I-D	3 B-D	4 O6-D	5 I-F	6 B-F	7 O3-F	8 O12-F	Total
<i>Without denoise</i>									
Drive end	0	2	0	1	1	0	3	4	11
Fan end	0	2	0	1	2	1	3	2	11
<i>Autocorrelation denoise</i>									
Drive end	0	3	1	2	0	0	0	0	6
Fan end	2	1	2	1	0	1	0	0	7

FIGURE 8: Rényi entropy distributions of (a) raw data and (b) after autocorrelation denoise.

individual signal. The detailed parameters of the models are shown in Table 4 and 5.

As can be observed from Table 4, weight of the kernel function acting on the data of two coaxial sensors accounts for 91.41%, which takes the most part of the MKL model. More specifically, the fusion of two-dimensional coaxial features outperformed feature of a single sensor. It can also be concluded from Table 4 that the weights obtained by MKL realize not only the selection of kernel function but also the selection of data. The kernel function parameters of model based on the data of the drive end in Table 5 are similar to those of the fan end. These selected kernel functions are sorted by weight as follows: Gaussian kernel $\sigma=0.01$, Gaussian kernel $\sigma=0.1$, and polynomial kernel $d=3$. In addition, a first-order polynomial kernel is included based on the drive-end data, but the weight is very small. According to the previous formula, the first-order polynomial kernel is a linear kernel. The classification results of these three models based on different data are shown in Table 6. The classification accuracy based on the CVSFF method is 97.50%. This is much higher than those of single signal-based classification accuracy, which are 66.67% and 60.42%, respectively. All results of three models prove that the fusion of data from two coaxial sensors improved accuracy and robustness in fault detection of bearing.

The huge difference in results is mainly due to the good differentiation of two-dimensional features based on coaxial signals, while the single-dimensional entropy distribution of the individual sensor overlaps, as shown in Figure 7. For the characteristics of drive end, I-F and O12-F are distributed with overlap in the interval $[0.6, 1.2]$; I-D and O6-D are distributed with overlap in the interval $[0.3, 0.5]$. For the characteristics of fan end, I-D and I-F are distributed with overlap in the interval $[1.1, 1.6]$; O6-D, B-F, O3-F, and O12-F are distributed with overlap in the interval $[0.3, 0.7]$.

3.5.2. Comparison and Discussion of MKL-SVM with SVM, GA-SVM, and PSO-SVM. To introduce the effectiveness of MKL-SVM, SVM, genetic algorithm-optimized SVM (GA-SVM), and particle swarm optimization SVM (PSO-SVM) based on single kernel are measured in this section. All parameters of SVM are default. That is to say, the kernel function is the Gaussian kernel; the bandwidth of the Gaussian kernel is 1.

In order to compare the results of CVSFF, we use the same feature matrix that is used in MKL-SVM to train and test the other three SVM models. The classification results are shown in Table 7. The classification accuracies are 97.50%, 95.00%, 95.83%, and 96.67% using the MKL-SVM, original SVM, GA-SVM, and PSO-SVM, respectively. The Gaussian kernel bandwidth obtained by GA and PSO optimization is 8.64 and 5.64. Although the classification accuracy of two optimized algorithms is 0.83% and 1.67% higher than that of SVM, it is still lower than MKL-SVM.

Furthermore, the running time of the models is compared. The experiments are carried out on Intel (R) Core (TM) i5-4210U CPU 2.4 GHz, 4G RAM, Win 7 and MATLAB R2016b. With the exception of the SVM model, the other three models need to optimize the parameters. Thus, the running time of these three models is greatly increased compared with that of SVM. In the first, third, and fourth models, GA-SVM has the shortest running time but the lowest accuracy. The MKL-SVM model adopted in this paper has a higher classification accuracy, although its running time is 86.17 s longer than GA-SVM. Compared with PSO-SVM, this method is not only more efficient but also more accurate in classification.

To analyze the results of these four models in detail, we draw the confusion matrix corresponding to the classification results in Table 7, as shown in Figure 9.

Apparently, the error classification is mainly concentrated in B-F, O3-F, and O12-F. And the predicted states of these

TABLE 4: Multikernel parameters for coaxial signals model.

No.	1	2	3	4	5
Kernel	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Bandwidth	0.01	0.1	1	10	100
Variable	DE + FE	DE + FE	DE + FE	DE + FE	DE + FE
Weight	0.3430	0.3724	0.1903	0	0
No.	6	7	8	9	10
Kernel	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Bandwidth	0.01	0.1	1	10	100
Variable	DE	DE	DE	DE	DE
Weight	0	0.0780	0	0	0
No.	11	12	13	14	15
Kernel	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Bandwidth	0.01	0.1	1	10	100
Variable	FE	FE	FE	FE	FE
Weight	0.0079	0	0	0	0
No.	16	17	18	19	20
Kernel	Polynomial	Polynomial	Polynomial	Polynomial	Polynomial
Order	1	2	3	1	2
Variable	DE + FE	DE + FE	DE + FE	DE	DE
Weight	0	0	0.0084	0	0
No.	21	22	23	24	
Kernel	Polynomial	Polynomial	Polynomial	Polynomial	
Order	3	1	2	3	
Variable	DE	FE	FE	FE	
Weight	0	0	0	0	

TABLE 5: Multikernel parameters for individual signal model.

No.	1	2	3	4	5
Kernel	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Bandwidth	0.01	0.1	1	10	100
Weight for individual signal of drive end	0.7169	0.2463	0	0	0
Weight for individual signal of fan end	0.7636	0.1958	0	0	0
No.	6	7	8		
Kernel	Polynomial	Polynomial	Polynomial		
Order	1	2	3		
Weight for individual signal of drive end	0.0053	0	0.0316		
Weight for individual signal of fan end	0	0	0.0405		

TABLE 6: Results of coaxial signals and individual signal.

No.	Data	Method	Label	Training data	Test data	Accuracy
1	Coaxial signals	MKL-SVM	1-8	560 * 2	240 * 2	97.50% (235/240)
2	Individual signal of drive end	MKL-SVM	1-8	560 * 1	240 * 1	66.67% (160/240)
3	Individual signal of fan end	MKL-SVM	1-8	560 * 1	240 * 1	60.42% (145/240)

TABLE 7: Results of MKL-SVM and SVM.

No.	Data	Method	Label	Training data	Testing data	Accuracy	Running time (s)
1	Coaxial signals	MKL-SVM	1-8	560 * 2	240 * 2	97.50% (235/240)	170.96
2	Coaxial signals	SVM	1-8	560 * 2	240 * 2	95.00% (228/240)	0.22
3	Coaxial signals	GA-SVM	1-8	560 * 2	240 * 2	95.83% (230/240)	84.79
4	Coaxial signals	PSO-SVM	1-8	560 * 2	240 * 2	96.67% (232/240)	367.18

FIGURE 9: Classification results of (a) MKL-SVM, (b) SVM, (c) GA-SVM, and (d) PSO-SVM.

error classifications are B-F, O3-F, and O12-F. It verifies the characteristics of the feature distribution in Figure 8(b). The lowest classification accuracy of MKL-SVM is 90.00% (O3-F, 27/30) for single state bearing, which is significantly higher than that of SVM (73.33%, B-F, 22/30). In addition, the classification accuracy of MKL-SVM for the first five states is 100%, while SVM, GA-SVM, and PSO-SVM have 1, 2, and 1 misclassification of data in I-D state, respectively.

However, it can also be seen from Figure 9 that there are three groups of misclassification for O3-F based on the method proposed in this paper, while the performance of the other three models is better. The three segments misclassified by MKL-SVM are the 79th, 88th, and 97th segments of data O3-F. GA-SVM misclassified the 71th segment of O3-F. PSO-SVM misclassified the 88th segment of O3-F which is one of the MKL-SVM misclassified groups. The 97th segment is not close to the feature of B-F, but it is still misclassified as B-F. The results that are misclassified by MKL-SVM need further study and improvement. Nevertheless, MKL-SVM outperforms SVM, GA-SVM, and PSO-SVM both in whole and single state classification accuracy.

4. Conclusion

This paper proposed a feature fusion method based on MKL-SVM using coaxial vibration signals to classify REB states. The obtained accuracy of coaxial signals is 97.50% which is much higher than the results of single sensor signal for drive end (66.67%) and fan end (60.42%). Polynomial kernel with global characteristics and Gaussian kernel with local characteristics are selected, which greatly improves the generalization ability of the model. By comparing the CVSFF, SVM, GA-SVM, and PSO-SVM-based feature fusion, 97.50%, 95.00%, 95.83%, and 96.67% accuracies are obtained, respectively. It shows that the method extracted in this paper is more effective for REB fault classification.

Data Availability

The bearing data used to support the findings of this study are obtained from the Case Western Reserve University (CWRU) Bearing Data Center (<http://csegroups.case.edu/bearingdatacenter/home>).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The research was supported by the China Railway (grant number 2017J004-H).

References

- [1] E. P. Carden and P. Fanning, "Vibration based condition monitoring: a review," *Structural Health Monitoring: An International Journal*, vol. 3, no. 4, pp. 355–377, 2004.
- [2] B. Li, M.-Y. Chow, Y. Tipsuwan, and J. C. Hung, "Neural-network-based motor rolling bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 1060–1069, 2000.
- [3] L. Guo, H. Gao, H. Huang, X. He, and S. Li, "Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring," *Shock and Vibration*, vol. 2016, Article ID 4632562, 10 pages, 2016.
- [4] H. Zhao, S. Zuo, M. Hou et al., "A novel adaptive signal processing method based on enhanced empirical wavelet Transform technology," *Sensors*, vol. 18, no. 10, p. 3323, 2018.
- [5] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," *Information Fusion*, vol. 57, pp. 115–129, 2020.
- [6] X. Yan, Z. Sun, J. Zhao, Z. Shi, and C.-a. Zhang, "Fault diagnosis of rotating machinery equipped with multiple sensors using space-time fragments," *Journal of Sound and Vibration*, vol. 456, pp. 49–64, 2019.
- [7] L.-l. Jiang, H.-k. Yin, X.-j. Li, and S.-w. Tang, "Fault Diagnosis of rotating machinery based on multisensor information fusion using SVM and time-domain features," *Shock and Vibration*, vol. 2014, Article ID 418178, 8 pages, 2014.
- [8] J. Tao, Y. Liu, and D. Yang, "Bearing fault diagnosis based on deep belief network and multisensor information fusion," *Shock and Vibration*, vol. 2016, Article ID 9306205, 9 pages, 2016.
- [9] T. P. Banerjee and S. Das, "Multi-sensor data fusion using support vector machine for motor fault detection," *Information Sciences*, vol. 217, pp. 96–107, 2012.
- [10] J. Wang, Y. Liang, Y. Zheng, R. X. Gao, and F. Zhang, "An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples," *Renewable Energy*, vol. 145, pp. 642–650, 2020.
- [11] H. Zhou, J. Chen, G. Dong, H. Wang, and H. Yuan, "Bearing fault recognition method based on neighbourhood component analysis and coupled hidden Markov model," *Mechanical Systems and Signal Processing*, vol. 66–67, pp. 568–581, 2016.
- [12] S. Hong, Z. Zhou, E. Zio, and K. Hong, "Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method," *Digital Signal Processing*, vol. 27, pp. 159–166, 2014.
- [13] K. H. Hui, M. H. Lim, M. S. Leong, and S. M. Al-Obaidi, "Dempster-Shafer evidence theory for multi-bearing faults diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 57, pp. 160–170, 2017.
- [14] M. S. Safizadeh and S. K. Latifi, "Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell," *Information Fusion*, vol. 18, pp. 1–8, 2014.
- [15] S. Lu, P. Zheng, Y. Liu, Z. Cao, H. Yang, and Q. Wang, "Sound-aided vibration weak signal enhancement for bearing fault detection by using adaptive stochastic resonance," *Journal of Sound and Vibration*, vol. 449, pp. 18–29, 2019.
- [16] V. Mygdalis, A. Tefas, and I. Pitas, "Exploiting multiplex data relationships in support vector machines," *Pattern Recognition*, vol. 85, pp. 70–77, 2019.
- [17] M. H. Zangooei and S. Jalili, "PSSP with dynamic weighted kernel fusion based on SVM-PHGS," *Knowledge-Based Systems*, vol. 27, pp. 424–442, 2012.
- [18] 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 A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [19] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [20] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [21] Y. Peng, Z. Li, K. He et al., "Broadband mode decomposition and its application to the quality evaluation of welding inverter power source signals," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 11, pp. 9734–9746, 2020.
- [22] Y. Peng, Z. Wang, K. He et al., "Modulated broadband mode decomposition for the feature extraction of double pulse metal inert gas welding," *IEEE Access*, vol. 8, pp. 134031–134041, 2020.
- [23] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [24] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, CA, USA, pp. 547–561, 1961.
- [25] P. Bošković and Đ. Juričić, "Fault detection of mechanical drives under variable operating conditions based on wavelet packet Rényi entropy signatures," *Mechanical Systems and Signal Processing*, vol. 31, pp. 369–381, 2012.
- [26] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, Berlin, Germany, 1995.
- [27] A. Rakotomamonjy, F. Bach, S. Canu et al., "Simple MKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.
- [28] 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.