

## Research Article

# A Fault Diagnosis Method of Rolling Bearing Integrated with Cooperative Energy Feature Extraction and Improved Least-Squares Support Vector Machine

Zhang Xu, Darong Huang , Tang Min, and Yunhui Ou

Chongqing Jiaotong University, School of Information Science and Engineering, Chongqing 400074, China

Correspondence should be addressed to Darong Huang; drhuang@cqjtu.edu.cn

Received 5 November 2020; Revised 23 November 2020; Accepted 10 December 2020; Published 24 December 2020

Academic Editor: Yong Chen

Copyright © 2020 Zhang Xu 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.

To solve the problem that the bearing fault of variable working conditions is challenging to identify and classify in the industrial field, this paper proposes a new method based on optimization of multidimension fault energy characteristics and integrates with an improved least-squares support vector machine (LSSVM). First, because the traditional wavelet energy feature is difficult to effectively reflect the characteristics of rolling bearing under different working conditions, based on analyzing the wavelet energy feature extraction in detail, a collaborative method of multidimension fault energy feature extraction combined with the method of Transfer Component Analysis (TCA) is constructed, which improves the discrimination between the different features and the compactness between the same features of rolling bearing faults. Then, for solving the problem of the local optimal of particle swarm optimization (PSO) in fault diagnosis and recognition of rolling bearing, an improved LSSVM based on particle swarm optimization and wavelet mutation optimization is established to realize the collaborative optimization and adjustment of LSSVM dynamic parameters. Based on the improved LSSVM and optimization of multidimensional energy characteristics, a new method for fault diagnosis of rolling bearing is designed. Finally, the simulation and analysis of the proposed algorithm are verified by the experimental data of different working conditions. The experimental results show that this method can effectively extract the multidimensional fault characteristics under variable working conditions and has a high fault recognition rate.

## 1. Introduction

In industrial equipment, the rolling bearing is an essential part of high-speed rating machinery. During actual operation, once the rolling bearings have failures (such as internal cracking, abrasion, external cracking, et al.), the safety and reliability of the entire system will be directly affected. In this situation, the condition monitoring and fault diagnosis of rolling bearings have become a hot research topic in prognostic and health management (PHM) of industrial systems.

Over the past decade, the accretion data, which reflect the running performance of rolling bearings, are usually used to analyze and test the fault characteristics of the rolling bearings. However, because the working operation of rolling bearings is influenced by various kinds of dynamic factors, the fault characteristics of the acceleration data may be quickly submerged by the ambient noise. Thus, these all

cause a huge difficulty in diagnosing the fault of rolling bearings. And fortunately, when the fault characteristic of rolling bearings is located in the blind areas, the energy waveform of the original signal shows the characteristics of high efficiency and low amplitude. So, some researchers have tried to extract the energy feature from the acceleration data to accomplish the target of fault diagnosis for the rolling bearings. Unfortunately, because the fault data signal of rolling is nonlinear and nonstationary vibration response in the industrial environment, the energy characteristics extracted by based-wavelet packet and improved methods cannot effectively distinguish the differences of different features and the tightness of the similar characteristics for the complex industrial environment.

To extract useful information of energy features, Wavelet Theory is a usual method to analyze the vibration data of rolling bearing in previous work [1]. The key reason is that

the wavelet packets can adaptively be selected according to the characteristics of the signal and may divide a frequency band into multiple frequency bands. Based on this, some scholars had presented some improved extracted method (such as wavelet packet transform (WPT), the fuzzy mutual information of wavelet packet transform (FMIWPT), dual-tree complex wavelet packet transform, support vector machine based- WPT, complex wavelet packet energy moment entropy and maximal overlap wavelet packet transform, et al.) of energy feature in [2–7]. These methods may not only implement the initial enhancement of the fault feature but also extract multiple permutation entropy features in the real application.

To address this problem, some scholars have tried to decompose the fault data signal into different frequency bands by using the wavelet packet and reconstruct the nodes in the frequency band in [8]. The advantage of the method is that characteristic frequency points may be located as quickly as possible in an industrial environment. Meanwhile, to treat the irregular vibration signal, the fault features may be extracted in time domain by using wavelet transform (see [9]). Besides, the optimization of structure of the energy characteristics has been also discussed briefly using Transfer Component Analysis (TCA) in [10]. By using the algorithm, the data properties may be preserved and the data distributions in different domains may converge to a stable scale. However, the running state of rolling bearings is affected by the endogenous factors; the different decomposition depths of the energy features are a very key problem in our working condition. Due to the diversity and variability of the actual fault diagnosis distribution, some methods (such as optimized transfer learning (TL) algorithm and regularization terms of multilayer) are aimed at solving the domain adaptation and reducing the distribution discrepancy and the among-class distance of the learned transferable features in [11, 12]. These methods can optimize the structure of feature sets better. At the same time, for getting better effectiveness in fault diagnosis under variable working conditions, some improved methods based on transfer learning (such as high-order Kullback–Leibler, parameter transfer, improved joint distribution adaptation, et al.) were also presented in [13–16]. So, it is essential to find out a new method to further optimize the structure of energy characteristics of rolling bearings in the real application. This problem is the first key core of this paper.

Additionally, on the one hand, the purpose of extracting the energy features is to implement the accurate diagnosis of fault state in industrial scenes. For this reason, the diagnosis method needs to be also simultaneously concerned while the multifeatures of fault signal are extracted. In the light of this, the support vector machine (SVM), which has the preferable ability of the classification, is usually used to implement the classification and recognition of fault in running processing of rolling bearing. However, the algorithm does not suit the situation of large amounts of data. Thus, some researchers presented improved algorithms such as binary SVM or based-HV SVM to identify the multifault types of the rolling bearing [17]. Further, the least-squares support vector machine (LSSVM) was constructed to reduce the difficulty of

calculation and improve the recognition speed. The algorithm and model have solved the inequality constraint in SVM. But how does equality constraint substitute the inequality constraint is very difficult in practice. To overcome this problem, some optimized methods (such as multiclass LSSVM, trend analysis based-LSSVM, et al.) have been presented to diagnose the fault state in [18–22]. These models can better identify the fault state in a complex industrial scenario. Meanwhile, to get the better classification performance of fault states, some integrated intelligent diagnosis methods and models (such as based-SVM neural network, based-LSSVM neural network, et al.) were also established in [23]. The result showed that the improved algorithms have the classification performance for the rolling bearing in industrial systems.

However, although these improved models and algorithms may achieve the desired goal in fault diagnosis of rolling bearing, there are two crucial parameters of LSSVM worth noting, i.e., the penalty factor and the kernel function parameter. Because the penalty factor trades off between misclassification samples and interface simplicity and the kernel function defines the size of the impact of the single training sample, the accuracy of fault diagnosis is decided by them to a great extent. At present, the optimization of the two parameters has not yet been resolved. From the point of practical engineering application, there are few methods to synergistically adjust the structure of the feature to make it better for practical fault diagnosis. To overcome the problem, some optimized algorithms (such as multimode PSO, the PSO based on the Mahalanobis distance (MD), implements mutation based on PSO, et al.) was proposed to adjust significant parameters in [24–26]. So, the second area that we were focusing on in this paper is considering the interactive impact between the optimization selecting of the energy feature and the accuracy of fault diagnosis [27–31].

According to the statement, the method in this paper uses three effective methods to construct a bearing fault diagnosis model. First, wavelet transform and energy features are used to represent the characteristics of the bearing signal, while the eight-dimension energy feature set cannot distinguish the difference between the five bearing states. Then, TCA is introduced to optimize the distribution of the energy feature set. Because the TCA can both reduce the distribution between different bearing states and increase the distance of the learned transferable features, the optimized feature set is beneficial to the improvement of diagnostic accuracy. At last, the improved PSO aims to find the optimal parameters of LSSVM.

According to these two points, a new fault diagnosis method of rolling bearing was presented by integration with cooperative energy feature extraction and improved LSSVM to extract the multidimensional feature set and enhance the accuracy of fault diagnosis. The rest of the paper is arranged as follows. In Section 2, the cooperative energy feature extraction rule has been discussed in detail combined with TCA and WP. In Section 3, we have established an improved LSSVM algorithm with dynamic parameter adjustment-based Particle Swarm Optimization (PSO) and Wavelet Mutation Optimization (WMO). In Section 4, the fault data

coming from the laboratory of the Guangdong Institute of Petrochemical Technology was used to verify the effectiveness of the model algorithm. Finally, some promising applications of the model have been discussed in detail in Section 5.

## 2. Cooperative Energy Feature Extraction Model and Algorithm for Vibration Signal of Rolling Bearings

In general, the extraction of the reasonable feature from original signal data is a universal method in an industrial scenario. But, as we all know, the original signal data of the bearing is large and complicated in the real industrial scene. In addition, the original data set is disturbed by complex noises. Therefore, in this situation, how to extract the energy features of the original signal data is very important to exactly represent the running state of the rolling bearing. Based on this, the Wavelet Theory and Transfer Component Analysis are introduced to construct the cooperative energy feature extraction model. The advantage of this processing method has the following two points: the first point is that the primary signal components with different frequency bands may be in detail depicted by wavelet packet because the wavelet packet may provide satisfactory localization properties in both time and frequency domains; the second point is that the structure of energy feature can be optimized by the TCA. Also, this cooperative processing method can get the internal form of the energy feature. Next, the energy feature extraction model based on wavelet packet shall first be expounded.

*2.1. Energy Feature Extraction Model Based on Wavelet Packet.* To address the above first problem, the vibration data may be divided into multiple frequency bands by wavelet packet. In a real application, the internal characteristics of the signal can be adaptively selected. To better understand the idea, the detailed algorithm shall be in detail described by using a wavelet packet in the next step.

To further analyze the data resource, let  $L^2(R) = \oplus_{j \in Z} W_j$  indicate the fact that multiresolution analysis is based on different scale factors of  $j$ . In the multifrequency analysis,  $L^2(R)$  is decomposed into a series of subspaces of the orthogonal sum of  $W_j$  ( $j \in Z$ ), where  $W_j$  is the subspace of the wavelet function. In our work, the wavelet space of  $W_j$  is refined in binary mode to achieve the goal of increasing the frequency resolution. To ensure the mapped performance between the scale-space  $V_j$  and wavelet subspace of  $W_j$  in a new subspace  $U_j^n$ , the iteration formula was defined as follows:

$$\begin{cases} U_j^0 = V_j, \\ U_j^1 = W_j, \\ j = Z, \end{cases} \quad (1)$$

where the subspace of  $U_j^n$  is the closure space of the function  $\omega_n(t)$  and  $U_j^{2^n}$  is the closure space of function  $\omega_{2^n}(t)$ ; the following two-scale equations should be also satisfied:

$$\begin{cases} \omega_{2^n}(t) = \sqrt{2} \sum h(k) \omega_n(2t - k), \\ \omega_{2^{n+1}}(t) = \sqrt{2} \sum g(k) \omega_n(2t - k), \end{cases} \quad (2)$$

where  $g(k) = (-1)^k h(1 - k)$ ; the sequence of  $\{\omega_n(t), n \in Z\}$  is the basis function. And then the sequence constructed is determined by the basis function  $\omega_0(t) = \phi(t)$  and is called the orthogonal wavelet packet;  $\omega_0(t)$  and  $\omega_1(t)$  are the scaling function of  $\phi(t)$  and the wavelet basis function  $\psi(t)$ , respectively.

In addition, the normalized orthogonal basis of  $L^2(R)$  is composed of  $\omega_n(t - k), \omega_n(t - l) = \delta_{kl}$  and  $\{\omega_n(t), n \in Z\}$ . The wavelet packet series of  $h(k)$  is described as  $\{\omega_n(t), n \in Z\}$ .

Further, for an arbitrary  $c_j^n(t) \in U_j^n$ ,  $c_j^n(t)$  can be expressed as follows:

$$c_j^n(t) = \sum_t d_i^{k,n} \omega_n(2^j t - l), \quad (3)$$

where  $U_j^{2^n} \perp U_j^{2^{n+1}}, U_{j+1}^n = U_j^{2^n} \oplus U_j^{2^{n+1}}$ . And then,  $c_{j+1}^n(t)$  can be decomposed into  $c_j^{2^n}(t)$  and  $c_j^{2^{n+1}}(t)$  by using wavelet decomposition.

In addition,  $\{d_l^{i+1,n}\}$  is used to obtain the equations for  $\{d_l^{j,2^n}\}$  and  $\{d_l^{j,2^{n+1}}\}$  according to the following formula:

$$\begin{cases} d_l^{j,2^n} = \sum_k h_{k-2l} d_k^{j+1,n}, \\ d_l^{j,2^{n+1}} = \sum_k g_{k-2l} d_k^{j+1,n}. \end{cases} \quad (4)$$

In conventional approaches, the three-layer wavelet packet decomposition structure is shown in Figure 1.

From Figure 1, for an arbitrary signal  $S$  at which the frequency range is in  $[0, f]$ , it may be decomposed into a high-frequency part  $D_1$  and a low-frequency part  $A_1$ . After the first layer in the multiresolution analysis framework, the frequency range of the high-frequency part is  $[f/2, f]$ , and the frequency range of the low-frequency part signal is  $[0, f/2]$ . Once the first layer was ended, the decomposition in the second layer starts to perform; i.e., the low-frequency part  $AA_2$  and the high-frequency part  $DA_2$  are obtained from decomposing the low-frequency part  $A_1$ . The high-frequency part  $D_1$  is also decomposed to obtain the low-frequency component  $AD_2$  and the high-frequency component  $DD_2$ . This means that the four frequency ranges may be indicated as  $[0, f/4], [f/4, f/2], [f/2, 3f/4],$  and  $[3f/4, f]$ . Analogously, the signal data set may be implemented to decompose layer by layer. The decomposition relationship for signal  $S$  may be formulated as follows:

$$S = AAA_3 + DAA_3 + ADA_3 + DDA_3 + AAD_3 + DAD_3 + ADD_3 + DDD_3. \quad (5)$$

Through the above algorithm, the different orthogonal wavelet spaces of  $U_j^n$  have different time–frequency resolution spaces, and all  $U_j^n$  can cover the entire bandwidth of signal  $S$ . Obviously, the time–frequency domain analysis can adaptively project the spectral components of the signal onto the

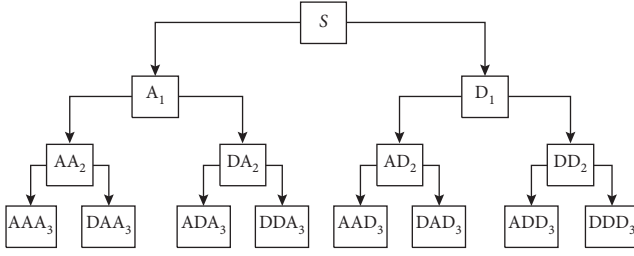


FIGURE 1: Schematic diagram of the wavelet packet decomposition structure.

orthogonal wavelet packet space of the corresponding frequency band. In engineering of energy feature extraction of rolling bearing, because the components of the original signal at each decomposition level represent the signal information in the corresponding local time–frequency area, the information of the component signal may be always intact. Of course, the energy of the signal distribution has been calculated at a certain decomposition level, and the energy in the orthogonal wavelet packet space at a certain decomposition level can be calculated. Then, the frequency indices of energy wavelet packets are arranged to form the eigenvectors of the original signals.

To better characterize the energy feature, suppose that the calculation formula of the wavelet packet energy is as follows:

$$E(j, n) = \sum_{k \in Z} [\omega_n(2t - k)]^2, \quad (6)$$

where  $\omega_n(2t - k)$  is the wavelet packet transform coefficient.

To further understand and analyze the distribution of energy features, the statistical distribution of the energy is calculated according to the decomposed signals at different frequency bands.

Unfortunately, when the energy feature with different working conditions is input into the classifier, the result of training accuracy is 97.5%, and the test accuracy is only 87.2%. The energy characteristics cannot fully depict the differences among different states of the bearing, which results in low accuracy for bearing fault diagnosis. Thus, to find a method to make up for the shortage of wavelet packet is necessary. To reduce the data dimension and optimize the data distribution, the TCA theory was used to further optimize the feature sets in our research. Next, an improved cooperative energy feature extraction shall be established to solve the problem of combining with wavelet packet and TCA.

### 3. An Improved Cooperative Energy Feature Extraction Method Based on the Transfer Component Analysis Algorithm

In the real operation of extracted energy feature for signal data of rolling bearing, how to accurately distinguish the differences among the states in variable working conditions is very crucial. To ensure that the energy feature set of rolling bearing has the characteristics such as stronger class and

compact inner class, the TCA was introduced to reduce the distribution discrepancy and among-class distance of the learned transferable features. The main role of TCA is to optimize the structure of energy characteristics gotten in the above section. To accommodate more flexible modeling, based on introducing the basic concept and approach of transfer feature, this section would design and implement a cooperative energy feature extraction method by using the TCA algorithm.

**3.1. Basic Concepts of Transfer Feature.** Notice that the TCA algorithm can adjust the edge distribution probability of the data set, and the edge distribution probability represents the probability distribution of the data set. The distribution of the bearing feature set is insufficient to meet the accuracy requirements of fault diagnosis. To reduce the distance in the same feature set and expand the gap of different feature sets, the method can reduce the distribution between the source domain and target domain data. The transfer feature mapping process is designed in Figure 2.

The circle and triangle represent source domain and target domain, respectively. A and B mean different data. Before the common mapping process is implemented, the edge probability distribution between the feature set of the source domain differs from the feature set of the target domain. The mapping relationship from the source domain to the target domain should be depicted.

For simplicity of further analysis, assume that the source domain is  $D_S = \{X_S, X_T\}$ , the target domain is  $D_T = \{X_T\}$ ,  $X_S$  is the feature set of the source domain,  $Y_S$  is the label set, and  $X_T$  is the feature set of the target domain. And then,  $P(X_S) \neq P(X_T)$ .

In fact, after feature mapping by using the TCA algorithm, the edge probabilities of  $M(X_S)$  and  $M(X_T)$  are as similar as possible, and the following relationship should be satisfied:

$$P[M(X_S)] \approx P[M(X_T)]. \quad (7)$$

Once the above formula is correct, the source domain feature sample set and the target domain feature sample set are mapped to the shared subspace, and the knowledge of the feature sample transfer process can be fully utilized to improve the cross-domain learning ability.

## 4. Energy Feature Extraction Method Based on the Transfer Component Analysis Algorithm

To ensure that the difference between the source domain and the target domain should be reduced by finding the common points, the distance between the transfer method and retaining the original features of the two data sets is defined as follows:

$$D_S = \{X_S, X_T\} = \{x_{S_i}, y_{S_i}\}_{i=1}^{n_s}, \quad (8)$$

where  $n_s$  is the number of labeled source domain training samples data and  $n_t$  is the number of unlabeled data in the target domain for  $X_T = \{x_{T_i}\}_{j=1}^{n_t}$ .



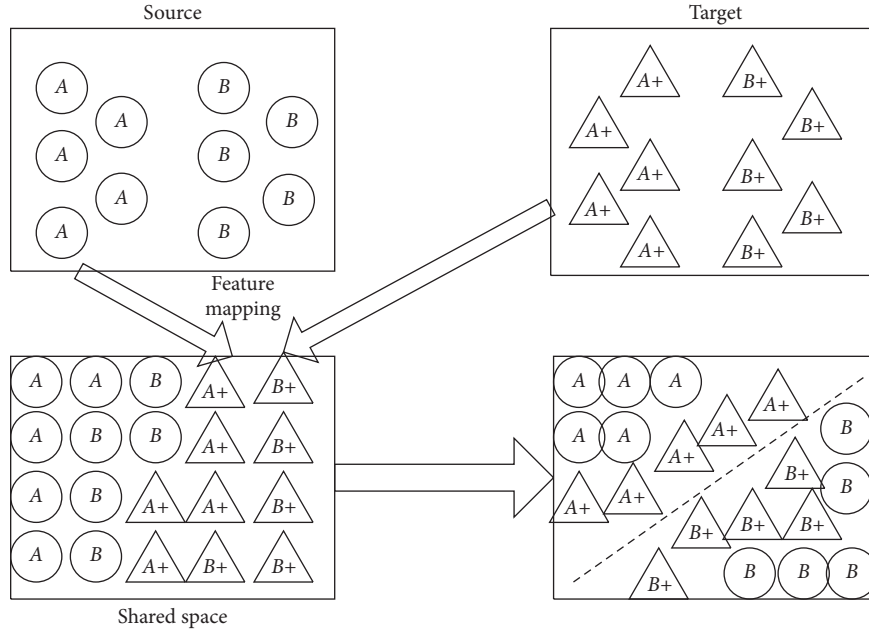


FIGURE 2: Transfer learning feature mapping diagram.

In this situation, the goal is to predict the sample label of  $y_{T_i}$ . At the same time, the data mapping function  $\phi$  between the source domain and the target domain is defined as follows:

$$\begin{aligned} X_S &\longrightarrow \phi(X_S) = X_S^*, \\ X_T &\longrightarrow \phi(X_T) = X_T^*. \end{aligned} \quad (9)$$

The objective of this process is to reduce the difference between the edge probability distributions  $P(X_S)$  and  $P(X_T)$  so that  $P(X_S^*) \approx P(X_T^*)$ .

Similarly, for a given source domain data set  $X_S$  and associate target domain data set  $X_T$ , the distance function MMD between the two data sets can be expressed as follows:

$$\text{MMD}(X_S, X_T) = \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(x_{S_i}) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(x_{T_j})^2, \quad (10)$$

where  $\phi(x_{T_j})^2_H$  is the squared standard operation performed in the regenerative kernel Hilbert space. The source and target domain data are mapped into a shared low-dimensional potential space through the nonlinear mapping, and then the kernel functions can be solved as follows:

$$K = \begin{pmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{pmatrix} \in \mathbb{R}^{(n_s+n_t) \times (n_s+n_t)}. \quad (11)$$

In equation (11),  $K_{S,S}$ ,  $K_{S,T}$ ,  $K_{T,S}$ , and  $K_{T,T}$  are the corresponding kernel functions obtained from the source domain, the target domain, and the hybrid domain, respectively.

Further, the formula can be rewritten as

$$\text{MMD}(X_S, X_T) = \text{Tr}(KL), \quad (12)$$

where  $\text{Tr}$  represents the trace of a matrix

For simplicity,  $L_{i,j}$  maybe expressed as follows:

$$L_{i,j} = \begin{cases} \frac{1}{n_s^2}, & x_i, x_j \in X_S, \\ \frac{1}{n_t^2}, & x_i, x_j \in X_T, \\ -\frac{1}{(n_s n_t)}, & \text{others.} \end{cases} \quad (13)$$

Through the above analysis, the distribution among different data may be reduced, and shared feature representation of the two domains is realized. The representation may also maintain the data feature attributes of the two domains. Also, the method may achieve this goal and extract the data components for the transfer of data from different but related fields. The main purpose of the algorithm is twofold. First, the distance between  $\phi(X_S)$  and  $\phi(X_T)$  is minimized; second, the main feature attributes of the raw data sets  $X_S$  and  $X_T$  are preserved.

For the whole mapped samples, we can find an embedded matrix  $W \in \mathbb{R}^{(n_s+n_t) \times m}$  ( $m \ll n_s + n_t$ ), s.t.

$$\tilde{K} = (KK^{-1/2}\tilde{K})KWW^T K. \quad (14)$$

Based on equation (14), equation (11) may be rewritten as follows:

$$\text{MMD}(X_S^*, X_T^*) = \text{Tr}((KWW^T K)L) = \text{Tr}(W^T K L K W). \quad (15)$$

Once the covariance matrix  $W$  may be found, the largest variance of the energy feature can be maintained into the

newly created subspace. The concrete kernel matrix formula can be indicated as  $\tilde{\Sigma}$ , i.e.,

$$\tilde{\Sigma} = W^T K H K W, \quad (16)$$

where  $H = I - (1/(n_s + n_t)1^T) \in R^{(n_s+n_t) \times (n_s+n_t)}$  indicates the center matrix.

Therefore, the problem may be transformed into the optimal problem of  $\tilde{\Sigma} = I_m$ , and  $I_m \in R^{m \times m}$  is a unit matrix. The final core learning problem can be established as follows:

$$\begin{aligned} \min_W & Tr(W^T K L K W) + \mu Tr(W^T W), \\ \text{s.t.} & W^T K H K W = I_m, \end{aligned} \quad (17)$$

where  $\mu$  is the trade-off parameters and  $\mu > 0$ .

Next, the optimization problem can be transformed into a maximum mapping matrix  $W$ , which can be obtained by matrix decomposition. First, we need to calculate the matrix  $(K L K + \mu I)^{-1} K H K$  to obtain  $W$ .

So far in this discussion, the core energy features can be selected by integration with the above model. On the other hand, because the distance between the same-state features becomes increasingly similar, the separability of energy feature becomes increasingly clear for different states. All in all, the compactness of features has been greatly improved after integrating with TCA. It is convenient to use classifiers to improve the fault diagnosis accuracy of bearings.

**4.1. Design and Analysis of a Cooperative Energy Feature Extraction Algorithm.** According to the above theoretical analysis, an improved cooperative energy feature extraction algorithm may be designed as follows.

- (i) *Step 1.* Original signals in different working conditions of the bearing are input to the wavelet packets for three-layer decomposition.
- (ii) *Step 2.* According to the signal component, the energy of every component is calculated, and the bearing feature set is constructed.
- (iii) *Step 3.* The training sample set of the source domain is built based on the energy characteristics with explicit working conditions. Moreover, feature sets under unknown working conditions are constructed to collect test samples in the target field.
- (iv) *Step 4.* The source domain feature set and the target domain feature set are mapped into the kernel space together. The maximum mean distance between the source domain feature samples and the target feature samples is measured in space. The calculated maximum mean distance is used as a criterion for judging the source domain data.
- (v) *Step 5.* The data are input into the optimized LSSVM, training is performed with the source domain data, and the target domain data are used to test the training result. Finally, the classification results are obtained and accuracy is assessed. The detailed flowchart is shown in Figure 3.

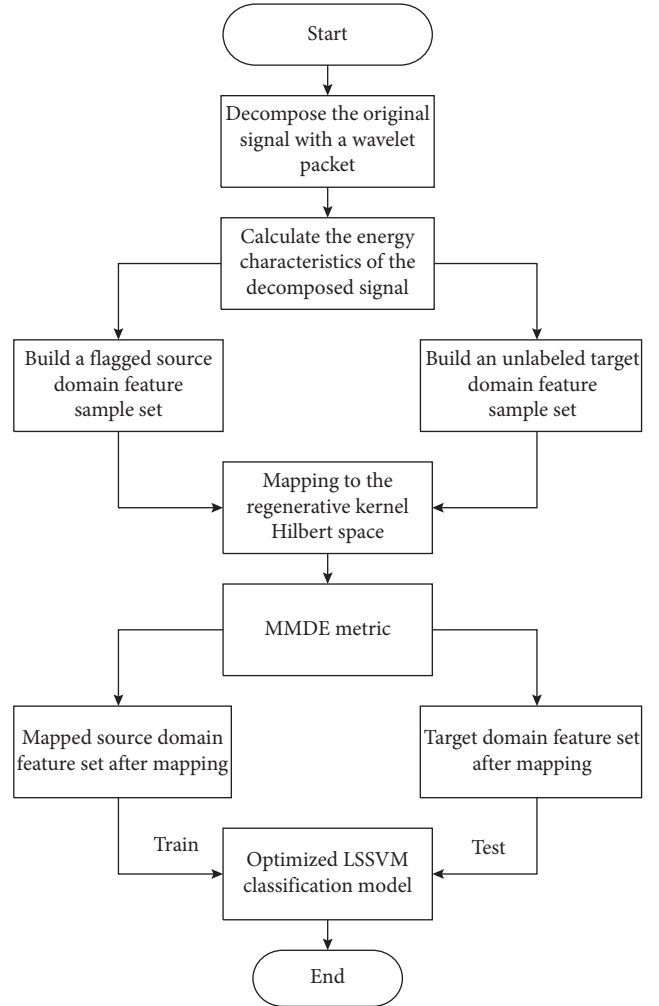


FIGURE 3: Flowchart of the bearing fault diagnosis method.

Whether the extraction mechanism of the energy feature is improved, the final goal of the energy feature is to improve the accuracy of the fault recognition of rolling bearing. Of course, good differentiation among different states and a high correlation among the same states will bring some gains in diagnostic accuracy. That is to say, it is also convenient to use classifiers to improve the fault diagnosis accuracy of rolling bearings. In the next step, the fault diagnosis method of rolling bearing shall be established to solve the goal.

## 5. Classification Process of Improved LSSVM with Dynamic Parameter Adjustment

According to the above analysis, an improved fault diagnosis method combining with improved LSSVM with dynamic parameter adjustment is listed in Figure 4.

- (i) *Step 1.* Input the extracted data features into the improved LSSVM model and train the two parameters that need to be optimized.
- (ii) *Step 2.* Initialize the parameter in particle swarm, such as evolutionary algebra, the learning factor, the

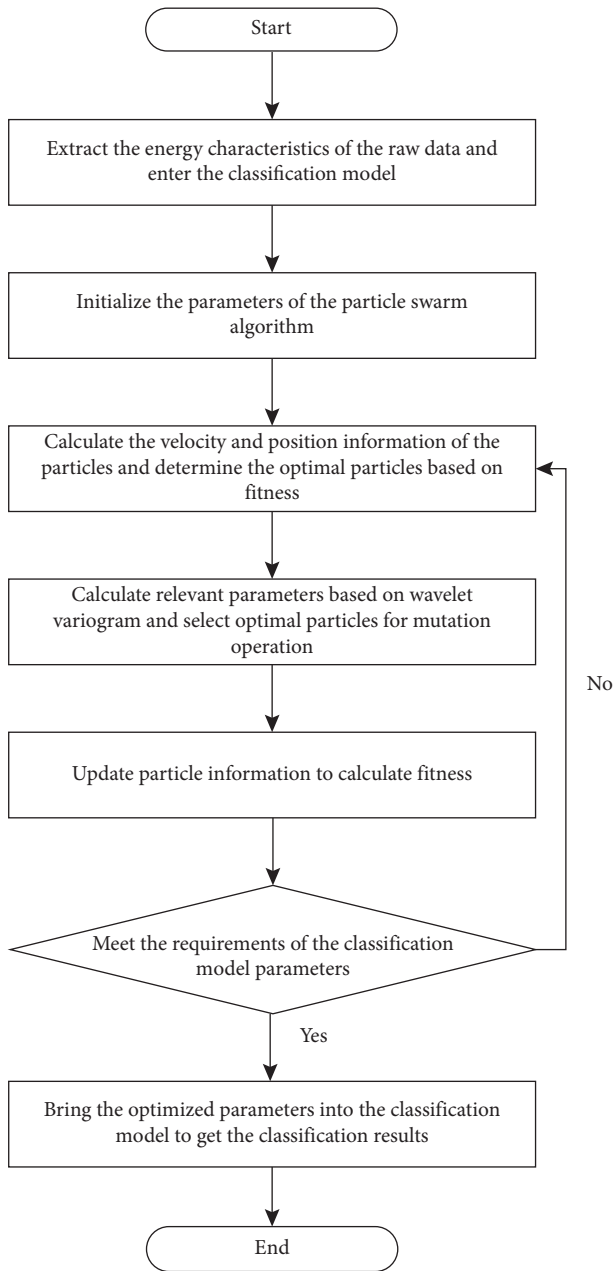


FIGURE 4: Schematic diagram of the optimized LSSVM classification model.

initial position  $x_{id}$  of each particle, the initial velocity  $v_{id}$ , et al.

- (iii) *Step 3.* The best position is set as the initial position of each particle. The optimal fitness equals the best position of each particle. The speed and position of each particle are calculated according to the formula.
- (iv) *Step 4.* Calculate the scale parameter and wavelet function value from the wavelet variogram. The mutation operation is performed on the current optimal particle according to the wavelet function formula.

- (v) *Step 5.* Update  $p_{best}$  and  $g_{best}$  according to the fitness value of the particle. Then, update the velocity and position information of the particle at the same time.

- (vi) *Step 6.* Determine whether the results of the algorithm reach the optimal condition. The training classification accuracy of the classification model is defined as the fitness degree of the PSO. If the fitness value calculated in the current cycle is the best, the current particle is saved as the best particle. If the fitness is not the best, the optimal parameters from the end of the previous cycle are used. The optimal particle search continues until the end of the cycle. The punishment coefficient  $C$  and Gaussian radial kernel function  $R$  are saved to construct the LSSVM classification model.

Through this algorithm, the cooperative energy feature extraction model and algorithm for the vibration signal of a rolling bearing are used to build a multidimensional feature set. And the fault diagnosis may also be implemented. The special flowchart of fault diagnosis is designed in Figure 4.

## 6. Experiments and Discussions

To verify the effectiveness of the proposed fault diagnosis method, the experimental acceleration data of bearings are used for fault diagnosis. The experimental data were obtained from the multifault diagnosis equipment of the rotary unit in the State Key Laboratory of Bearings, Guangdong University of Petrochemical Technology. Figure 5 shows the single-stage centrifugal fan fault diagnosis unit. Figure 6 shows the schematic diagram of the inner and outer cracks of bearings.

With this experimental platform, the data for each fault can be acquired under five states: normal, external cracking, internal cracking, missing bearings, and wearing bearings.

*6.1. Signal Processing and Feature Set Construction.* In our testing rig, the acceleration signals of five different conditions in the normal, external cracking, internal cracking, wearing, and missing states of the bearing during operation are used as the original signals for fault diagnosis.

The five different working conditions are in *A* of length 1150 mm, speed 2870 r/min; in *B* of length 1730 mm, speed 2980 r/min; in *C* of length 1800 mm, speed 2970 r/min; in *D* of length 2450 mm, speed 2980 r/min; and in *E* of length 2200 mm, speed 4800 r/min. The 300 characteristics of the first three working conditions are used as the source domain set; the 200 characteristics of the latter two working conditions are used as the target domain set. The original signals for the original states of the bearing are shown in Figure 7.

To verify the effectiveness of our algorithm, a sample data set containing 10240 sampling points in a sample period is used to extract the energy feature. To facilitate signal processing and extraction, the original signals are divided into  $1024 \times 10$  groups. Figure 7 shows the acceleration signal of the bearing under normal conditions. From Figure 7, the original data set has been divided into  $1024 \times 10$  groups. And



FIGURE 5: Single-stage centrifugal fan fault diagnosis unit.

the signal in each group is decomposed into 8 frequency bands by wavelet packet algorithms as shown in Figure 8.

The signal has a large volume due to high sampling frequency, and it is difficult to distinguish faults from these signals. First, the original signal is decomposed into three layers of wavelet packets to obtain signal components with eight different frequency bands. Figure 8 is a diagram showing the signal components obtained by decomposing the original signals for the original states of the bearing. The frequency band of the original signal is divided into multiple bands. According to the characteristics of the signal, the corresponding frequency band is adaptively selected to match the signal frequency, thereby improving the resolution of the signal frequency. Figure 7 only shows that the signal is decomposed into signal components of different frequency bands, and it does not reflect obvious fault characteristics. Therefore, the next step is to further extract the energy characteristics of the signal components.

After the original signal in five different states is decomposed by the wavelet packet, the characteristic histogram is obtained by calculating the energy of the node in the component signal. The energies described the multidimensional feature set of the bearing and the energy features extracted from one group of the normal signal. From the using point of view, constructing the feature set of signals is reasonable by using them in a sample period. As shown in Figure 9, the distribution of energy is different under different bearing states. The energy characteristics can initially show the difference, and then the energy feature values are extracted to construct the energy feature table. So, these features may be used to constitute a complete feature set for structure processing and fault diagnosis. Based on this, the energy feature values extracted from the  $1024 \times 10$  group of the normal signal are listed in Table 1.

In our experiment, the data sets for the five bearing states include  $10 \times 10240$  groups, and each group of signals is divided into ten groups for signal decomposition. We can obtain 8 different frequency bands from the original signal. The energy characteristics of the nodes are used to construct a multidimensional energy feature set for the bearing. Table 1 shows the multidimensional energy feature data sets for the original states; it is obvious that different bearing state has different energy features. Then, the table of energy features is input into the classifier.

The energy features extracted from the bearing fault vibration signal constitute a feature set that has been normalized. The labeled source domain data sample set and the

unlabeled target domain data sample set are mapped to the regenerative Hilbert kernel space. Between the source domain and target domain, the difference in the total maximum mean value reflects the difference in the distribution. The smaller the maximum mean difference is between the source domain and target domain, the stronger the source domain to target domain mobility. It is beneficial to select source domain data with high similarity to the target domain data.

Unfortunately, when the energy feature with different working conditions is input into the classifier, the result of training accuracy is 97.5%, and the test accuracy is only 87.2%. The energy characteristics cannot fully depict the differences among different states of the bearing, which results in low accuracy for bearing fault diagnosis. Thus, to find a method to make up for the shortage of wavelet packet is necessary. To reduce the data dimension and optimize the data distribution, the TCA theory was used to further optimize the feature sets in our research. Next, an improved cooperative energy feature extraction shall be established to solve the problem of combining with wavelet packet and TCA.

Energy feature is recalculated from each component by the improved cooperative energy feature extraction algorithm. In our simulation experiment, the feature set in A is inputted into TCA which is used to optimize the distribution of the feature set. In this hidden subspace, a classifier can be trained using the tagged samples from the mapped source domain, and the classifier is used to test the target domain data in the hidden space. The simulation results are shown in Figures 10 and 11.

Figure 10 shows the original energy distributions of the bearing. The five state characteristics of the bearing (normal, outer crack, inner crack, wear, and missing steel ball in bearing) are not distinct, and a poor energy distribution leads to low classification accuracy. Obviously, after inputting the energy characteristics into the TCA algorithm, the energy distribution of the bearing is shown in Figure 11. For indeed, the distance between same-state features becomes increasingly similar and the energy features possessed the advantage of the time-space concentricity. That has shown that our model and algorithm are effective.

Whether the extraction mechanism of the energy feature is improved, the final goal of the energy feature is to improve the accuracy of the fault recognition of rolling bearing. Of course, good differentiation among different states and a high correlation among the same states will bring some gains in diagnostic accuracy. That is to say, it is also convenient to use classifiers to improve the fault diagnosis accuracy of rolling bearings. In the next step, the fault diagnosis method of rolling bearing shall be established to solve the goal.

**6.2. Comparative Experimental Analysis.** As a classifier, the optimized LSSVM is used for random cross-validation experiments. The data set of the source domain is used as a training set, and the data set of the target domain is used as a test set.



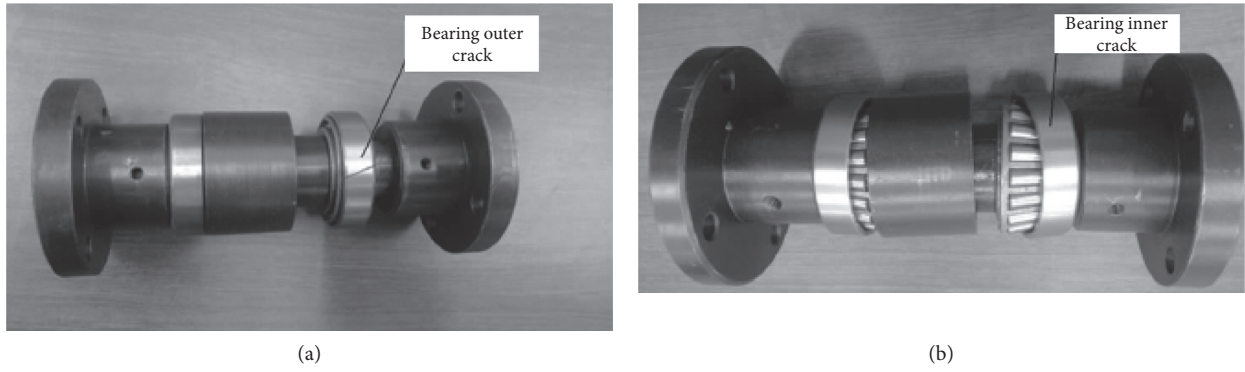


FIGURE 6: (a) Bearing outer crack. (b) Bearing inner crack.

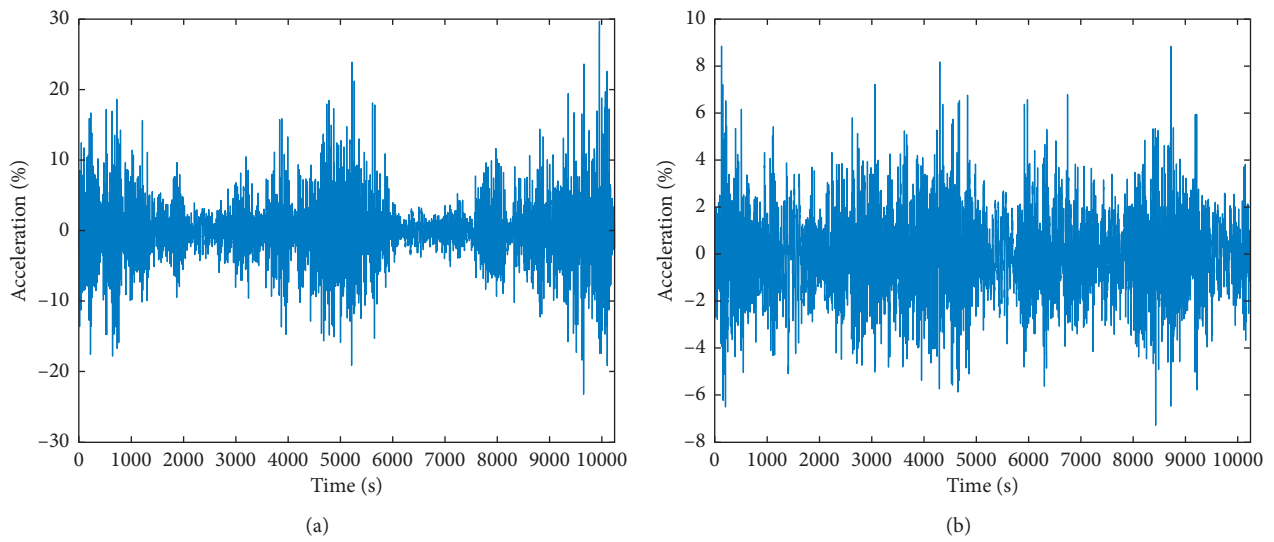


FIGURE 7: (a) Normal original signal diagram. (b) Outer crack original signal diagram.

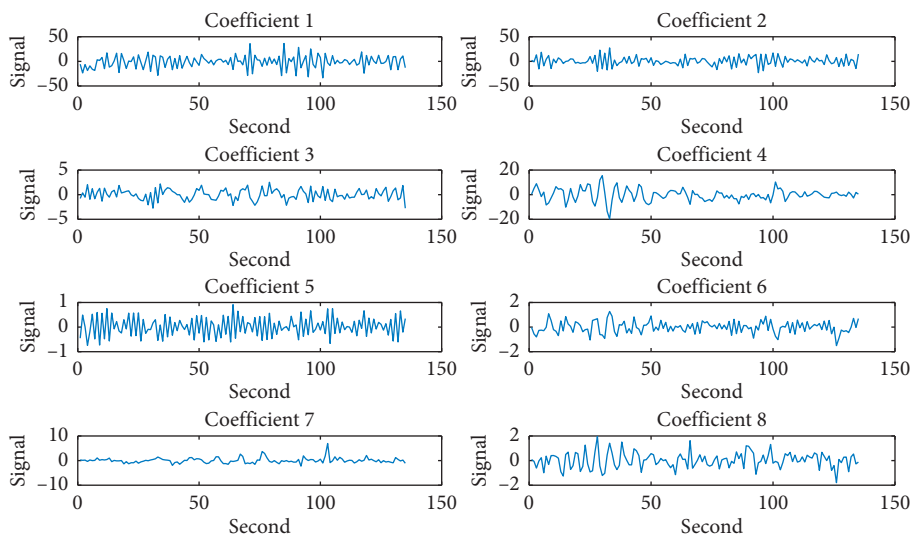


FIGURE 8: Normal signal component.

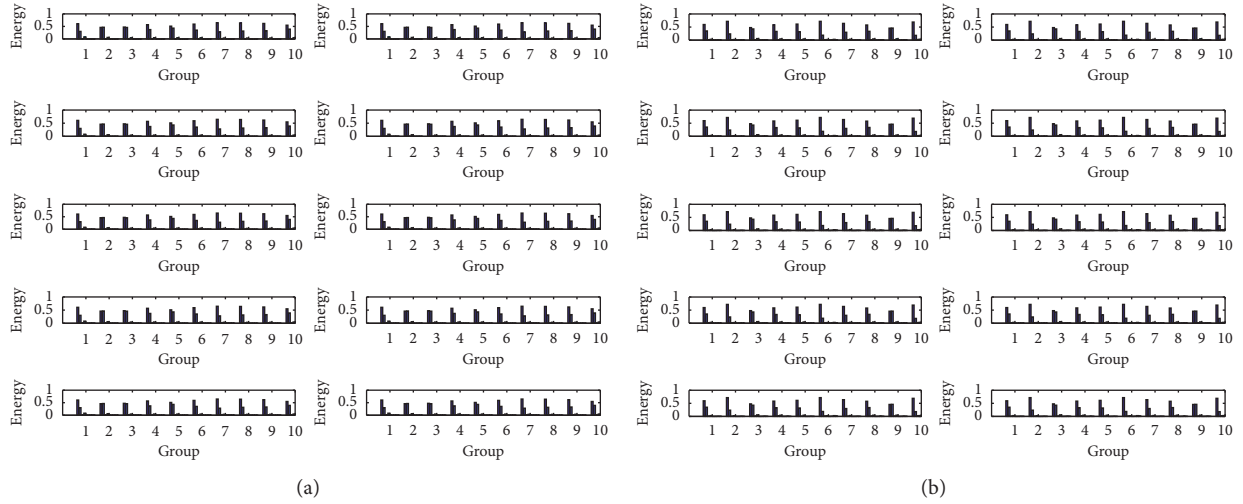


FIGURE 9: (a) Normal signal energy characteristics. (b) Outer crack signal energy characteristics.

TABLE 1: Normal signal energy characteristics.

	$E1$	$E2$	$E3$	$E4$	$E5$	$E6$	$E7$	$E8$
1	0.61	0.30	0.00	0.08	0.00	0.00	0.00	0.00
2	0.46	0.47	0.00	0.06	0.00	0.00	0.00	0.00
3	0.48	0.46	0.01	0.05	0.00	0.00	0.00	0.00
4	0.57	0.37	0.00	0.05	0.00	0.00	0.00	0.00
5	0.51	0.43	0.01	0.05	0.00	0.00	0.00	0.00
6	0.59	0.35	0.00	0.05	0.00	0.00	0.00	0.00
7	0.65	0.28	0.01	0.04	0.01	0.01	0.01	0.00
8	0.64	0.32	0.00	0.04	0.00	0.00	0.00	0.00
9	0.62	0.33	0.01	0.04	0.00	0.00	0.00	0.00
10	0.55	0.39	0.00	0.05	0.00	0.00	0.00	0.00

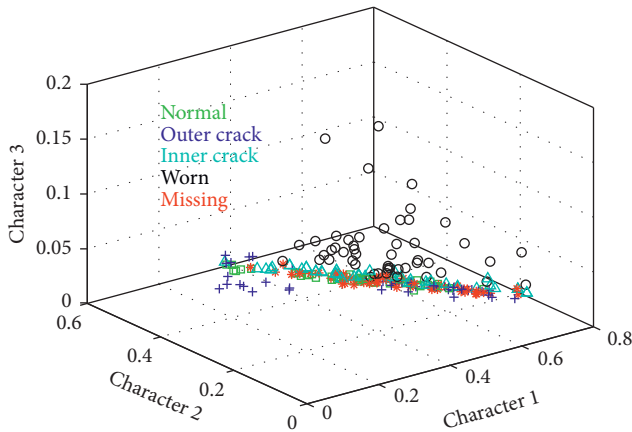


FIGURE 10: Original energy distributions.

In the TCA algorithm, the kernel function maps the data from the source domain and the target domain to the high-dimensional space. Therefore, the choice of the kernel function is related to the data mapping process of the source domain and the target domain. Four different kernel functions, namely, primal, RBF, linear, and SAM, are used to conduct comparative experiments. Under different kernel functions in TCA, the ability to analyze the corresponding

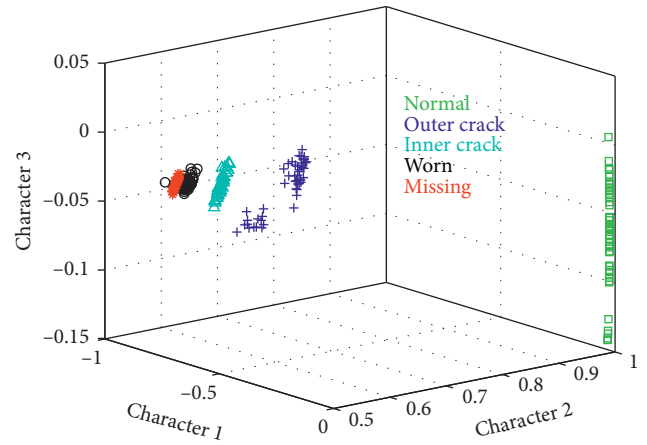


FIGURE 11: Energy distributions after transfer.

energy characteristics is tested. The training accuracy and test accuracy are calculated. Because TCA is a data dimensionality reduction algorithm, the dimension of data reduction is related to the classification accuracy. In this paper, the original data dimension of the energy feature data set is 8, and the dimensionality reduction is varied from 1 to 8 to test the diagnostic accuracy of the fault diagnosis method. Combining the results from Table 2 and Figure 12, the diagnostic accuracy of the RBF kernel function is relatively high and stable. Therefore, the RBF kernel function is used for bearing fault diagnosis analysis.

Based on the above fault diagnosis classification model, each group uses 100 sets of data features for fault identification. The simulation results for the training phase and the test phase are shown in Figure 12(a).

As shown in Figure 13(a), there are 210 training values for the 5 states ((1) normal, (2) internal cracking, (3) outer cracking, (4) wear, and (5) missing). The training accuracy is 100%.

According to the test data in Figure 13(b), there are 120 groups of test data for the 5 states ((1) normal, (2) internal cracking, (3) outer cracking, (4) wear, and (5) missing).

TABLE 2: Fault diagnosis accuracy under different kernel functions.

Different kernel functions	Training accuracy	Test accuracy	Statistical accuracy (%)
Primal	100	98.4	98.7
Linear	99.8	97.6	98
RBF	100	99.6	99
SAM	99.6	96.8	97.6

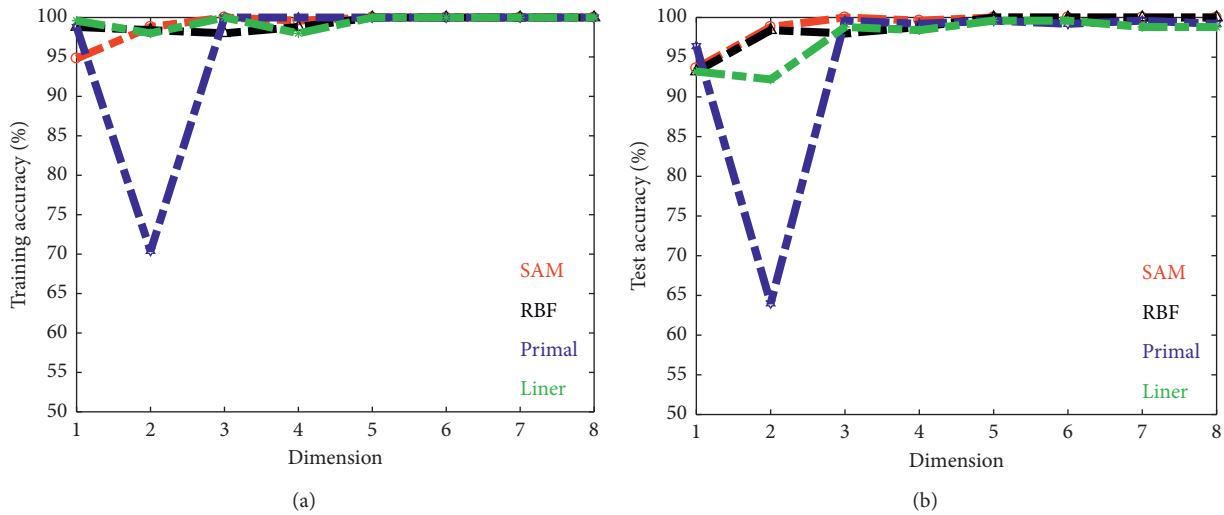


FIGURE 12: (a) Relationship between the feature dimension and training accuracy after mapping. (b) Relationship between the feature dimension and test accuracy after mapping.

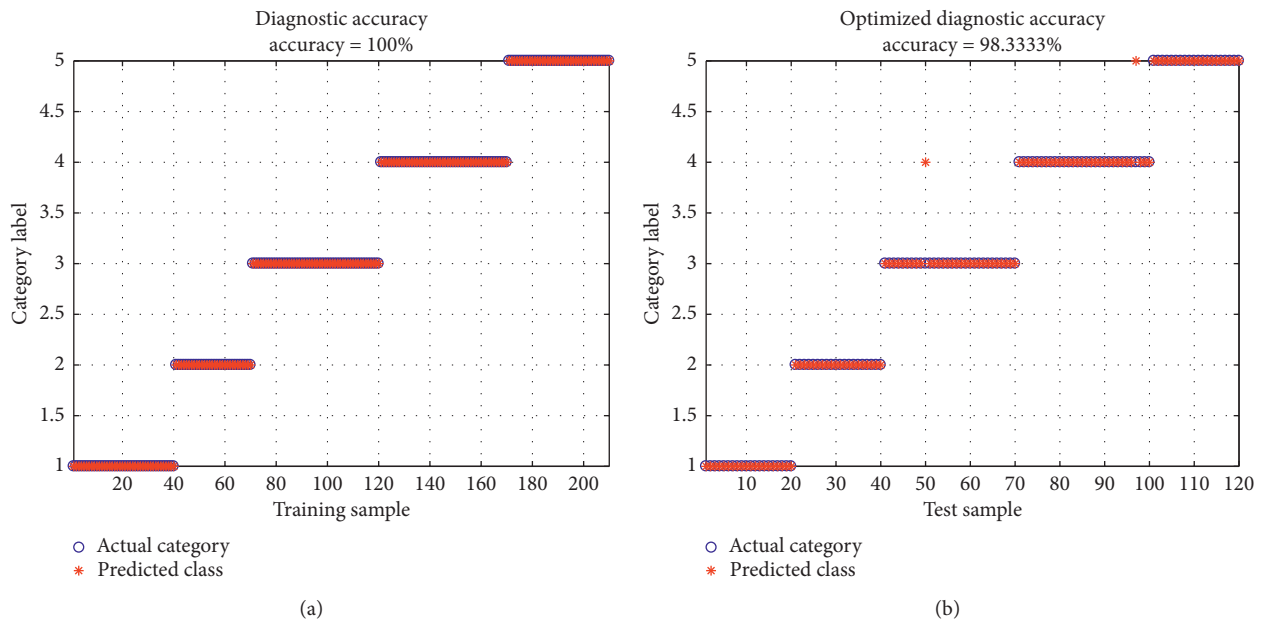


FIGURE 13: (a) Training accuracy. (b) Test accuracy.

Among them, two values are incorrectly classified, so the test accuracy is 98.3%.

To verify the validity and superiority of the algorithm presented in this paper, we compare different unoptimized algorithms with the optimized classification algorithm proposed.

Four different methods are compared under the same experimental environment and the same experimental data. Table 3 shows that the correct rate can reach 100% during the training process using the method developed in this paper.

Additionally, the correct rate can reach 99.8% during the test process. The fault diagnosis accuracy is better

TABLE 3: Accuracy of fault diagnosis in the comparative experiments.

Diagnosis method	Training accuracy (%)	Test accuracy (%)
The method proposed in this paper	100	99.8
EN-TCA-PSO-LSSVM	100	97.5
EN-PSO-LSSVM	98.8	88.4
EN-LSSVM	97.5	87.2

than that of the other three methods. The comparison shows that the TCA algorithm is effective in analyzing the energy characteristics of wavelet packets. Moreover, the optimized classification algorithm is superior to the traditional single classification algorithm and has a better diagnostic ability.

## 7. Conclusions

In this paper, to improve the accuracy of identifying and classifying fault in variable working conditions, a new method based on optimization of multidimension fault energy characteristics and integrate with an improved least-squares support vector machine (LSSVM). The main contributions of this paper are as follows.

- (1) The method of wavelet packet is used to reduce the surrounding noise and decompose the original signal with eight different frequency bands. The energy of every component is calculated to construct a feature set for bearing.
- (2) Because the TCA can amend the distribution of the energy feature, the The distribution of the feature set is optimized, and the data dimension is much closer than before. The optimized feature structure could improve the accuracy of bearing fault diagnosis.
- (3) Particle swarm and the wavelet mutation were integrated to optimize two parameters of LSSVM. Through the real data of bearing, the training accuracy of the proposed method is 100%, and the test accuracy 99.8%. The experiment result shows that the proposed method is effective in the low-precision problem of fault diagnosis for complex bearings in the equipment.
- (4) Unfortunately, there are still two problems to be solved in the next research. First, the complex noise of the original signal brings interference to the fault diagnosis of bearing. Second, the kernel function selected in the TCA algorithm is very single. Therefore, the next step is to focus on signal denoising and TCA construction of multicore kernel functions to further improve the fault accuracy.

## Data Availability

The data used to support the findings of this study are included within the article. The data sets are provided by the Guangdong University of Petrochemical Technology for experimental verification.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The authors thank the Guangdong University of Petrochemical Technology for providing datasets for experimental verification. This work was supported by the National Natural Science Foundation of China under Grant no. 61304104; Chongqing Technology Innovation and Application Special Key Project under Grant no. cstc2019jscx-mbdxX0015; the Innovation Foundation of Chongqing Postgraduate Education under Grant no. CYS20282; and the Science and Technology Project of Power Science Research Institute of State Grid Xinjiang Electric Power Co., Ltd., under Grant no. SGXJDK00JLJS1800161.

## References

- [1] R. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Transactions on Bio-medical Engineering*, vol. 58, no. 1, pp. 121–131, 2010.
- [2] C. Shen, D. Wang, F. Kong, and P. W. Tse, "Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier," *Measurement*, vol. 46, no. 4, pp. 1551–1564, 2013.
- [3] J. Qu, C. Shi, F. Ding, and W. Wang, "A novel aging state recognition method of a viscoelastic sandwich structure based on permutation entropy of dual-tree complex wavelet packet transform and generalized chebyshev support vector machine," *Structural Health Monitoring*, vol. 19, no. 1, pp. 156–172, 2020.
- [4] Z. Ling, D. Jing, D. Huang, M. Bo, K. Lan, and Y. Liu, "The incipient fault feature enhancement method of the gear box based on the wavelet packet and the minimum entropy deconvolution," *Systems Science & Control Engineering*, vol. 6, no. 3, pp. 235–241, 2018.
- [5] X. Li, S. Wu, X. Li, H. Yuan, and D. Zhao, "Particle swarm optimization-support vector machine model for machinery fault diagnoses in high-voltage circuit breakers," *Chinese Journal of Mechanical Engineering*, vol. 33, no. 1, pp. 1–10, 2020.
- [6] H. Shao, J. Cheng, H. Jiang, Y. Yang, and Z. Wu, "Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing," *Knowledge-Based Systems*, vol. 188, p. 105022, 2020.
- [7] A. C. Jahagirdar and K. K. Gupta, "Fractional envelope to enhance spectral features of rolling element bearing faults," *Journal of Mechanical Science and Technology*, vol. 34, no. 2, pp. 573–579, 2020.



- [8] H. Jin, A. Titus, Y. Liu, Y. Wang, and a. Z. Han, "Fault diagnosis of rotary parts of a heavy-duty horizontal lathe based on wavelet packet transform and support vector machine," *Sensors*, vol. 19, no. 19, p. 4069, 2019.
- [9] S. Gao, Y. Wu, and Z. Jiang, "Static and dynamic rubbing positions identification of cryocooler based on wavelet packet analysis and support vector machine," *Journal of Infrared and Millimeter Waves*, vol. 38, no. 5, pp. 627–632, 2019.
- [10] S. Pan, I. Tsang, J. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks a Publication of the IEEE Neural Networks Council*, vol. 22, no. 2, pp. 199–210, 2010.
- [11] S. Zhou, S. Qian, W. Chang, Y. Xiao, and Y. Cheng, "A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier," *Sensors*, vol. 18, no. 6, p. 1934, 2018.
- [12] Y. Zhang, Y. Qin, Z.-y. Xing, L.-m. Jia, and X.-q. Cheng, "Roller bearing safety region estimation and state identification based on LMD-PCA-LSSVM," *Measurement*, vol. 46, no. 3, pp. 1315–1324, 2013.
- [13] X. Gu, F. Deng, X. Gao, and R. Zhou, "An improved sensor fault diagnosis scheme based on TA-LSSVM and ECOC-SVM," *Journal of Systems Science and Complexity*, vol. 31, no. 2, pp. 372–384, 2018.
- [14] X. Jiang, S. Li, and Y. Wang, "A novel method for self-adaptive feature extraction using scaling crossover characteristics of signals and combining with LS-SVM for multi-fault diagnosis of gearbox," *Journal of Vibroengineering*, vol. 17, no. 4, pp. 1861–1878, 2015.
- [15] W. Fu, K. Wang, C. Zhang, and J. Tan, "A hybrid approach for measuring the vibrational trend of hydroelectric unit with enhanced multi-scale chaotic series analysis and optimized least squares support vector machine," *Transactions of the Institute of Measurement and Control*, vol. 41, no. 15, pp. 4436–4449, 2019.
- [16] A. Moosavian, S. Jafari, M. Khazaei, and H. Ahmadi, "A comparison between ANN, SVM and least squares SVM: application in multi-fault diagnosis of rolling element bearing," *International Journal of Acoustics & Vibration*, vol. 23, no. 4, pp. 432–440, 2018.
- [17] M. Sun, H. Wang, P. Liu, S. Huang, and P. Fan, "A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings," *Measurement*, vol. 46, no. 4, pp. 1551–1564, 2019.
- [18] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mechanical Systems and Signal Processing*, vol. 122, pp. 692–706, 2019.
- [19] W. Qian, S. Li, and J. Wang, "A new transfer learning method and its application on rotating machine fault diagnosis under variant working conditions," *IEEE ACCESS*, vol. 6, pp. 69907–69917, 2019.
- [20] H. Kim and B. D. Youn, "A new parameter repurposing method for parameter transfer with small dataset and its application in fault diagnosis of rolling element bearings," *IEEE ACCESS*, vol. 7, pp. 46917–46930, 2019.
- [21] W. Qian, S. Li, P. Yi, and K. Zhang, "A novel transfer learning method for robust fault diagnosis of rotating machines under variable working conditions," *Measurement*, vol. 138, pp. 514–525, 2019.
- [22] J. Wang, L. Qiao, Y. Ye, and Y. Chen, "Fractional envelope analysis for rolling element bearing weak fault feature extraction," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 2, pp. 353–360, 2016.
- [23] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 136–144, 2019.
- [24] Y. Zhang, P. Zhou, and G. Cui, "Multi-model based PSO method for burden distribution matrix optimization with expected burden distribution output behaviors," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1506–1512, 2019.
- [25] B. Long, W. Xian, M. Li, and H. Wang, "Improved diagnostics for the incipient faults in analog circuits using LSSVM based on PSO algorithm with Mahalanobis distance," *Neuro-computing*, vol. 133, pp. 174–278, 2014.
- [26] Z. Lv, L. Wang, Z. Han, J. Zhao, and W. Wang, "Surrogate-assisted particle swarm optimization algorithm with pareto active learning for expensive multi-objective optimization," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 3, pp. 838–849, 2019.
- [27] K. Zhang, K. Peng, S. X. Ding, Z. Chen, and X. Yang, "A correlation-based distributed fault detection method and its application to a hot tandem rolling mill process," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 3, pp. 2380–2390, 2020.
- [28] K. Zhang, H. Hao, Z. Chen, S. X. Ding, and K. Peng, "A comparison and evaluation of key performance indicator-based multivariate statistics process monitoring approaches," *Journal of Process Control*, vol. 33, pp. 112–126, 2015.
- [29] H. Luo, S. Yin, T. Liu, and A. Q. Khan, "A data-driven realization of the control-performance-oriented process monitoring system," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 1, pp. 521–530, 2020.
- [30] H. Luo, X. Yang, M. Krueger, S. X. Ding, and K. Peng, "A plug-and-play monitoring and control architecture for disturbance compensation in rolling mills," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 200–210, 2020.
- [31] H. Luo, K. Li, O. Kaynak, S. Yin, M. Huo, and H. Zhao, "A robust data-driven fault detection approach for rolling mills with unknown roll eccentricity," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 6, pp. 2641–2648, 2020.