

## Research Article

# Multiple-Fault Diagnosis Method Based on Multiscale Feature Extraction and MSVM\_PPA

Min Zhang , Zhenyu Cai, and Wenming Cheng

*School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China*

Correspondence should be addressed to Min Zhang; [zhmzhangmin16@126.com](mailto:zhmzhangmin16@126.com)

Received 1 December 2017; Accepted 20 February 2018; Published 30 April 2018

Academic Editor: Evgeny Petrov

Copyright © 2018 Min Zhang 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.

Identification of rolling bearing fault patterns, especially for the compound faults, has attracted notable attention and is still a challenge in fault diagnosis. In this paper, a novel method called multiscale feature extraction (MFE) and multiclass support vector machine (MSVM) with particle parameter adaptive (PPA) is proposed. MFE is used to preprocess the process signals, which decomposes the data into intrinsic mode function by empirical mode decomposition method, and instantaneous frequency of decomposed components was obtained by Hilbert transformation. Then, statistical features and principal component analysis are utilized to extract significant information from the features, to get effective data from multiple faults. MSVM method with PPA parameters optimization will classify the fault patterns. The results of a case study of the rolling bearings faults data from Case Western Reserve University show that (1) the proposed intelligent method (MFE\_PPA\_MSVM) improves the classification recognition rate; (2) the accuracy will decline when the number of fault patterns increases; (3) prediction accuracy can be the best when the training set size is increased to 70% of the total sample set. It verifies the method is feasible and efficient for fault diagnosis.

## 1. Introduction

With the increasing complexity of modern industry, fault diagnosis as accurately and timely plays an important role in industrial applications. Many fault diagnosis analysis methods have been developed to accurately and automatically identify faults in the past two decades. They usually use some basic measurements, like vibration, acoustic, temperature, and wear debris analysis [1, 2]. Although these methods have been beneficial, tests are still quite expensive and time-consuming. Fault diagnosis using the big data is a goal that has not yet been fully implemented. When the machine creates faults, the dynamic signals of the machine structure can be monitored in real time. The most effective fault diagnosis is feature learning from the monitoring information. Numerous researchers focus on fault diagnosis by faults patterns recognition, but most studies are concerned with the single fault patterns recognition [3–5]. However, the fault of the machinery may be the compound, which is influenced by two or three causes in practice.

Compared with single fault, compound faults lead to serious performance degradation and are more difficult to

recognize. This leaves the challenging task to identify multiple faults effectively. There are a few studies on multiple faults patterns recognition [6–8]. Chen et al. integrate extreme-point symmetric mode decomposition with extreme learning machine to identify typical multiple patterns recognition [6]. Ranaee and Ebrahimzadeh use a back-propagation (BP) neural network to recognize multiple-fault patterns [7]. Lu et al. propose a hybrid system that uses independent component analysis (ICA) and support vector machine (SVM) for recognizing mixture patterns. That method initially applies the ICA to get the independent components (ICs), and then the ICs are used as the inputs for the SVM classifier [8].

Feature extraction has become a major technique for multiple-fault patterns recognition. Numerous previous studies have reported about signal processing [9–13], like Fourier Transform [9] and wavelet transform [10]. But these methods should select appropriate base functions in advance. It is difficult to get effective analysis results, because the data from real world machines are nonstationary and nonlinear. Empirical mode decomposition (EMD), as a formidable and effective time–frequency analysis method, is programed to analyze the nonstationary signals and can be adaptive to decompose

the confusion signal into intrinsic mode functions (IMFs) by the inherent characteristics of the signals [11–13]. Features extraction by EMD is appropriate for distinguishing different mechanical signals [14–16]. Wang et al. [14] propose a novel feature extraction method by nonnegative EMD manifold in machinery fault diagnosis. Saidi et al. [15] proposed an EMD-based fault diagnosis module to detect the incipient bearing faults based on the raw vibration signals. Ali et al. [16] use EMD as feature extraction method then select the most important intrinsic mode functions and classify bearings defects by the artificial neural network (ANN). However, due to factors of uncertainty and nonlinearity of the production process, EMD analysis methods may overcome these limitations to implement the multiple faults diagnosis.

Traditionally, fault diagnosis was examined and analyzed manually by some measurements data. With the development of machine learning techniques, expert systems were employed for faults recognition in automatic process monitoring [17–20]. Support vector machine (SVM) has been widely used in recognizing multiple faults for its excellent performance in the practical application. Unlike neural network methods, SVM has great generalization ability of dealing with small samples since its model complexity does not depend on the number of features and thus is suitable for high dimensional data [13, 21, 22].

The above discussion shows that although some methods have shown prospective results in improving multiple-fault diagnosis performance, none has been widely used and still has improvement room to achieve the final goal. Multiple-fault classification method is proposed by a combination of empirical mode decomposition, PCA, and MSVM theory with PPA parameters optimization. Process signals were decomposed into IMFs by EMD method, and Hilbert transformation is utilized to get the instantaneous frequency of decomposed components. Then, the statistical features of intrinsic mode function and instantaneous frequency were calculated. Principal component analysis is utilized to extract significant information from the statistical features, to get effective data of multiple faults. Finally, MSVM with PPA parameters optimization will classify the fault modes.

## 2. Processing Model

The proposed fault patterns recognition model performs in three modules to effectively monitor the multiple faults (Figure 1).

- (i) Features extraction as the first stage: according to the characteristics of the machinery and equipment operating process, process signals were chosen and decomposed into IMFs by EMD method and instantaneous frequency of decomposed components was obtained by Hilbert transformation. Statistical and shape features are extracted from the EMD data. Then PCA is further applied to reduce the feature dimension and the computational complexity.
- (ii) Classify the fault patterns by MSVM in the second stage: the selected features are used as the inputs and

the MSVM classifier should be designed properly for getting the satisfactory recognition performance.

- (iii) Optimization module as the third stage:  $K$ -fold cross-validation and an adaptive mutation particle swarm optimization are combined as the PPA parameters optimization method to select the parameters for MSVM.

### 2.1. Features Extraction

(1) *EMD Method Principle.* Empirical mode decomposition (EMD) is an adaptive frequency analysis method; its main principle is to decompose the complex signals into some IMFs self-adaptively by the inherent characteristics of the signal, and every IMF shows the specific frequency information of the signals. EMD helps to smooth processing of the signal, decompose the fluctuations or trends of signal in different scales gradually, and overcome the pre-filter center frequency and bandwidth problem in traditional envelope analysis. EMD is suitable for nonstationary and nonlinear signal analysis, and it has been widely used in many fields [11–14], and the steps of EMD can be shown as follows.

Firstly, the raw vibration signal was decomposed; see

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t), \quad (1)$$

where  $x(t)$  is the original vibration signal;  $c_i(t)$  is the intrinsic mode function;  $r_n(t)$  is the remaining functions, which represents the overall trend of signals.  $i$  means decomposition time and  $n$  means total decomposition times.

The mode components are separated by the instantaneous frequency from high to low; EMD method can be viewed as a set of high-pass filter from the filter characteristic. EMD method obtained the first few high-frequency IMF components, which can effectively represent the signal characteristics; remaining IMFs belong to the residual component, which is mainly the low-frequency noise. Selected high-frequency IMF components are transformed by the Hilbert method (see (2)), and (3) construct their analytic signals.

$$H[c_i(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_i(\tau)}{t - \tau} d\tau, \quad (2)$$

$$z(t) = c_i(t) + iH[c_i(t)] = a(t)e^{i\phi(t)}. \quad (3)$$

Amplitude function and phase function can be got by

$$\alpha(t) = \sqrt{c_i(t)^2 + H[c_i(t)]^2}, \quad (4)$$

$$\phi(t) = \arctan \frac{H[c_i(t)]}{c_i(t)}.$$

Eq. (5) means the derivative of the phase function, namely, the instantaneous frequency.

$$\omega(t) = \frac{d\phi(t)}{dt}. \quad (5)$$

IMF components largely reflect the true characteristic information of the original signal. However, both ends of the

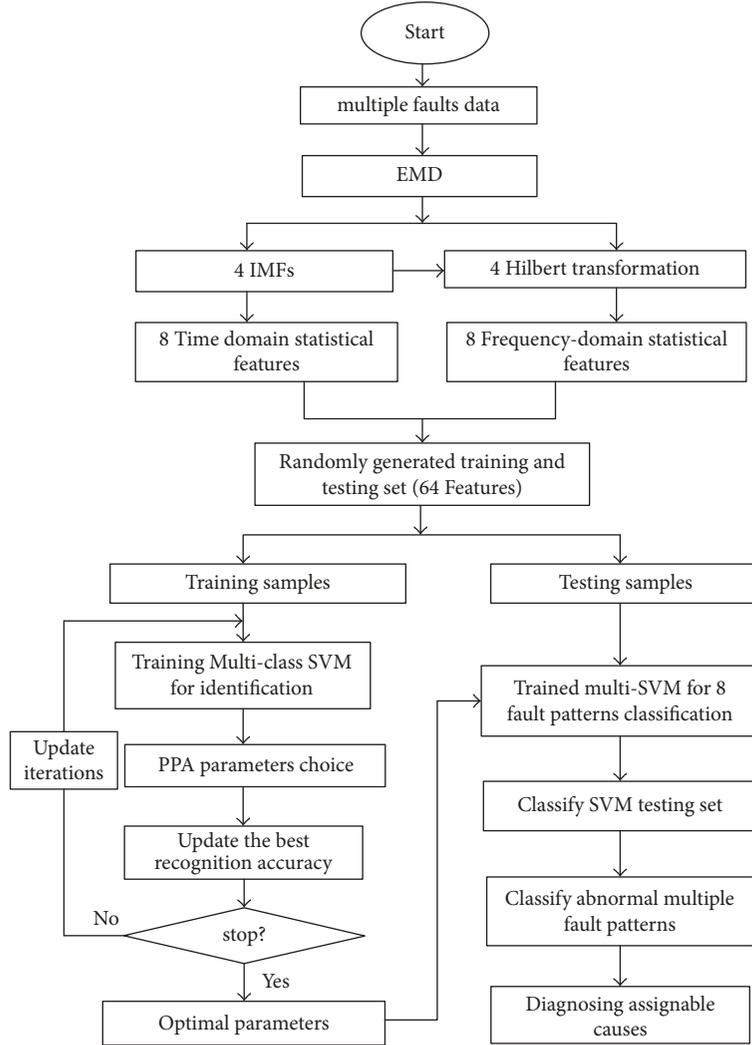


FIGURE 1: EMD-MSVM-PPA classification flowchart.

signal generate divergent phenomenon due to the EMD having used the cubic spline interpolation method. And with the gradual deepening of decomposition, divergent phenomenon has extended to the entire signal and produces modal aliasing. Recent studies show that the longer signal can be selected to reduce endpoint divergence and then select the subsequent IMF component, whose both ends of the signal are intercepted, to reduce the impact of endpoint modal aliasing.

In this paper, IMFs and their corresponding instantaneous frequencies are selected as characteristic variable data, and they effectively extract features value to calculate the corresponding eight statistical feature values as feature data set, respectively, mean, max, range, standard deviation, skewness, kurtosis, coefficient variation, and sum of square (see Table 1).

(2) *Principal Component Analysis (PCA)*. Multivariate statistical analysis is the most commonly used data-driven method for fault diagnosis. The principal component analysis (PCA) as an efficient representative uses linear transformation to obtain fewer features as little as possible. It can make these new features not correlated to each other but maintain the

original information as much as possible. The PCA method is chosen as the second feature extraction step to eliminate duplication of information in this study.

PCA algorithm is assumed that we have a collection of  $n$  unlabeled training data samples organized into the matrix  $X = [x^{(1)}, x^{(2)}, \dots, x^{(n)}]$ . Each data sample is a (column) vector of dimension  $m$ . For example, sample  $i$  can be denoted as  $x^{(i)} = [x_1^{(i)}, x_2^{(i)}, \dots, x_m^{(i)}]^T$ , where, for a matrix (or a vector)  $A$ ,  $A^T$  indicates its transpose. For simplicity of description, we assume each data sample has already been appropriately *scaled and demeaned*.

First, calculate the covariance matrix

$$E = \frac{1}{n} XX^T. \quad (6)$$

The matrix  $E$  describes the overall scattering of the process data. The matrix  $E$  is symmetric and can be orthogonally diagonalized as

$$E = P\Lambda P^T, \quad (7)$$

TABLE 1: Eight statistical features.

Statistical features name	Formula
Mean	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Max	$x_{\max}$
Range	$ x_{\max} - x_{\min} $
Standard deviation	$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
Skewness	$a_3 = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n\sigma^3}$
Kurtosis	$a_4 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n\sigma^4}$
Coefficient variation	$\beta = \frac{\sigma}{\bar{x}}$
Sum of square	$\bar{x}^2 = x_1^2 + x_2^2 + \dots + x_n^2$

where  $P = [p_1, p_2, \dots, p_m]$  is an orthonormal eigenvector matrix, and  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ , is the diagonal matrix of eigenvectors, such that  $(E)(p_i) = \lambda_i$ ,  $1 \leq i \leq m$ . Moreover, the eigenvalues are ordered such that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ . The eigenvectors  $p_i$ ,  $1 \leq i \leq m$  are the principle components of the  $i$ th dimension.

Let

$$\rho_i = \frac{\lambda_i}{\sum_{i=1}^m \lambda_i}, \quad i = 1, 2, \dots, m \quad (8)$$

be the contribution rate of the  $i$ th principal component. Choose the smallest  $k$  principal components such that

$$\sum_{i=1}^k \rho_i \geq \text{threshold}, \quad (9)$$

where the threshold is the desired percent of variance retained; for instance, a threshold can be chosen to be 0.99. The data set shall be approximated using the set of the first  $k$  principal components organized into the matrix

$$P_k = [p_1, p_2, \dots, p_k]. \quad (10)$$

Specifically, for each data sample  $x^{(i)}$ ,  $1 \leq i \leq n$ , the extracted new feature is a  $k$ -dimension vector given by

$$z^{(i)} = P_k^T x^{(i)}, \quad 1 \leq i \leq n. \quad (11)$$

The new data set is subsequently given by

$$Z = [z^{(1)}, z^{(2)}, \dots, z^{(n)}]. \quad (12)$$

Therefore, the original data set is

$$X = P_k Z + E \quad (13)$$

where  $E$  is the residual matrix, mainly caused by noise. Removing residuals will not cause significant impact on the useful information.

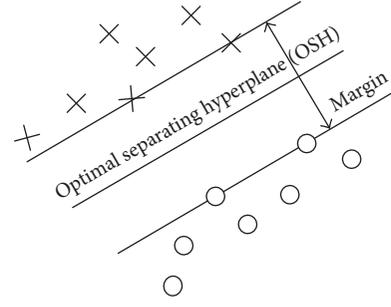


FIGURE 2: The OSH of a binary SVM.

**2.2. Support Vector Machine.** Basic binary SVM is initially designed to deal with two-class problems based on the structural risk minimization theory. It is set up to get the best solution between model complexities and learning ability. However, it has been extended to multiclass problems.

The binary SVM classification method is established by constructing an optimal separating hyperplane (OSH), in order to maximize the margin between two classes of data points (see Figure 2). Suppose that the training set  $(x^{(i)}, y^{(i)})$ ,  $i = 1, 2, \dots, n$ ,  $x^{(i)} \in R^m$ ,  $y^{(i)} \in \{-1, 1\}$ . A binary SVM model with a nonlinear kernel  $K$  is to find the best classifier  $f$  with parameters  $\theta_j$ ,  $j = 0, 1, \dots, n$ , in the form of

$$f(x) = \sum_{l=1}^n \theta_l K(x, x^{(l)}) + \theta_0 \quad (14)$$

in order to

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2} \sum_{l=1}^n \theta_l^2 + C \sum_{l=1}^n \xi_l \\ &\text{Subject to} \quad \forall i = 1, 2, \dots, n, \\ & \quad y^{(i)} \left( \sum_{l=1}^n \theta_l K(x^{(i)}, x^{(l)}) + \theta_0 \right) \geq 1 - \xi_i \\ & \quad \xi_i \geq 0, \end{aligned} \quad (15)$$

where the minimization is over all decision variables  $\theta_0$ ,  $\theta_j$ ,  $\xi_j$ ,  $j = 1, \dots, n$  and  $C > 0$  is a penalty constant. It can be applied to control the trade-off between minimize classification error and maximize margin.

For nonlinear decision boundary, the kernel function is applied to transform the input from a low-dimensional space into a higher dimensional feature space, so that an optimal linear separating hyperplane can be found. Although many researchers proposed several types of kernel functions, radial basis functions (RBF) are the most widely used to solve nonlinear problems in SVM. Its definition can be described as for  $x, x^{(l)} \in R^m$ ,  $l = 1, 2, \dots, n$ :

$$K(x, x^{(l)}) = \exp(-\gamma \|x - x^{(l)}\|^2), \quad \gamma \geq 0, \quad (16)$$

where  $\gamma \geq 0$  denotes the width of the RBF, and  $\|x\|$  indicates the Euclidean norm of  $x$ .

To solve multiclass problems, a MSVM method is applied in the second classifier stage. Two kinds of the MSVM methods are widely used; one is one-against-all (OAA); the other is one-against-one (OAO). In this paper, the OAO is adopted for multiple faults recognition. This method constructs  $M(M-1)/2$  binary SVM classifiers and every sample is trained to separate one class from another class. Testing sample can be got the results by the voting results.

The largest problems encountered in the MSVM are to select the best penalty parameter ( $C$ ) and the kernel function value ( $\gamma$ ). To alleviate this difficulty, PPA parameters optimization method is next used to get the best values of parameters  $C$  and  $\gamma$  in the MSVM classifier.

**2.3. Parameters Optimization.** MSVM is applied to classify the multiple-fault patterns in this paper, but the largest problem encountered is how to select the penalty parameter ( $C$ ) and kernel function parameters value ( $\gamma$ ) of MSVM. In many literatures, these two parameters can be got by  $K$ -fold cross-validation.

The principal of  $K$ -fold cross-validation is make the original sample separated into  $K$  subsamples randomly and then choose one subsample as the testing data for validation, and the rest ( $K-1$ ) subsamples are used as the training data. Then repeat  $K$  times; apply every subsample adopted once as the validation data. An averaged testing result can be got by this  $K$ -folds. The advantage of this method is to ensure that all data are applied for training and testing. Select the optimal penalty parameter ( $C$ ) and kernel function parameters value ( $\gamma$ ) with the highest rate of  $K$ -fold cross-validation as the final MSVM parameters, and then use these two parameters on the entire training set, and finally test unknown testing samples by the trained classification model.

$K$ -fold cross-validation weakness is that elected best parameters obtained from the training data cannot represent the entire training data. The result will be affected when there are only small sizes of training data. Therefore, adaptive mutation particle swarm optimization and  $K$ -fold cross-validation are combined to get the best parameters of MSVM. Firstly, define MSVM regularization parameter ( $C$ ) and kernel parameter ( $\gamma$ ) as a combination of a particle and use the training accuracy by  $K$ -fold cross-validation as the fitness function; 3-fold cross-validation is applied. The steps of the proposed parameters optimization method are followed.

*Step 1.* Set PPA parameters, like population number, swarm size, maximum velocity, and the probability of adaptive mutation rate, parameter ranges (see Table 2).

*Step 2.* Randomly generate the initial particle and set velocity; the particles are used MSVM to get training accuracy as fitness value by 3-fold cross-validation method, *The 3-fold in cross-validation is chosen according to the proportion of the training samples to the testing sample [23]*, the particle's best known position of the initial position.

*Step 3.* Update the individual position and velocity of every particle. Subsequently, renew the best known position ( $P_i$ ) of each particle  $i$  and the best group position  $P_g$ . Specifically,

TABLE 2: The parameters of AMPSO.

Name	Value
Maximum number of iterations $N$	100
Swam size	20
$C$ value	$[10^{-1}, 10^2]$
$\gamma$ value	$[10^{-2}, 10^3]$
Velocity parameters	0.6
Constants $c_1 = c_2$	1.5
Probability of adaptive mutation rate ( $\rho$ )	50%

let  $r_1, r_2$  be two uniformly distributed random numbers in the interval  $[0, 1]$ . The velocity  $V_i^{k+1}$  and position  $X_i^{k+1}$  of each particle at iteration  $k+1, k > 0$ , are updated using the current velocity and the distances from  $P_i$  and  $P_g$  as follows. Specifically, let  $p_j^{\max}$  be the maximum allowed position in the  $j$ th dimension and  $p_j^{\min}$  be the minimum allowed position in the  $j$ th dimension. Define

$$P_g^{\max} = (p_1^{\max}, p_2^{\max}, \dots, p_m^{\max}). \quad (17)$$

And similarly

$$P_g^{\min} = (p_1^{\min}, p_2^{\min}, \dots, p_m^{\min}). \quad (18)$$

Moreover, define

$$V_{\max} = k \cdot p_g^{\max}. \quad (19)$$

Calculate

$$\bar{V}_i^{k+1} = V_i^k + c_1 r_1 (P_i - X_i^k) + c_2 r_2 (P_g - X_i^k). \quad (20)$$

Then calculate each component of  $V_i^{k+1} = (v_1^{k+1}, v_2^{k+1}, \dots, v_m^{k+1})$  by

$$v_j^{k+1} = \begin{cases} \text{sgn}(\bar{v}_j^{k+1}) v_j^{\max}, & \text{if } |\bar{v}_j^{k+1}| > v_j^{\max} \\ \bar{v}_j^{k+1}, & \text{otherwise.} \end{cases} \quad (21)$$

For  $j = 1, 2, \dots, m$ , the notation  $\text{sgn}$  indicates the usual signum function. Then the  $i$ th particle position is updated with

$$X_i^{k+1} = X_i^k + V_i^{k+1}. \quad (22)$$

*Step 4.* For solving PSO's "premature" problem, which is easy to relapse into a local extremum and other particle quickly moves to this local position in the optimization process. AMPSO is used to solve this problem; it makes the algorithm escape from the local optima to find the best solution in the other space.

As can be seen in formula (22), the next position of a particle is determined by both its current position and its new velocity. The new velocity is determined by the immediately previous velocity, individually best  $P_i$ , and group best  $P_g$ , as shown in formula (20). If the algorithm is in premature,

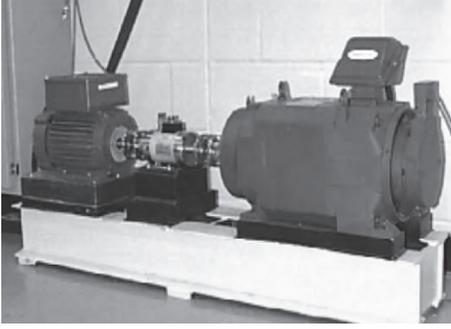


FIGURE 3: Case Western Reserve University Rolling experiment platform.

then the group best  $P_g$  is the local optimal solution. If  $P_g$  is changed, the search direction of particles will be redirected. Thus, the main idea of the AMPSO is by mutating  $P_g$  in hope that the search will get out of a local optimum to explore new individual optimum and group optimum.

The mutation of the PSO is designed as a random operator with a certain probability  $\rho$ . Specifically, for a uniformly distributed random number  $\text{rand} \in (0, 1)$  a mutated new group optimum is obtained as follows:

$$P_g = \begin{cases} (p_g^{\max} - p_g^{\min}) \cdot \text{rand} + p_g^{\min}, & \text{if } \text{rand} > \rho \\ P_g, & \text{otherwise.} \end{cases} \quad (23)$$

*Step 5.* Until meeting a termination criterion, which can be the number of iterations performed, or meeting the accuracy requirements, repeat from Step 2.

*Step 6.* Find the global best position ( $P_g$ ).

Through improved adaptive mutation particle swarm optimization algorithm and  $K$ -fold cross-examination validating, the optimum parameters of MSVM are used to train the entire training set; get the PPA-MSVM classification model. Then trained PPA-MSVM classifier is utilized to train unknown testing set, so that diagnosis of the rolling bearing faults is done intelligently.

### 3. Case Analyses

For verifying the feasibility and effectiveness of this method, the bearing dataset of Case Western Reserve University Bearing Data Center is adopted in this paper. The detailed description of the experimental apparatus is presented in Figure 3; it consists of a 2 hp, three-phase induction motor (left), a torque transducer (middle), and a dynamometer-load (right) [24, 25]. There are four different bearing conditions with four different loads simulated including healthy, inner-race defect, rolling element defect, and outer-race defect. Typical waveforms of the four conditions are illustrated in Figure 4.

*3.1. Data Descriptions.* In this study, the bearings with 1797 r/min in rotating speed at a sampling frequency of 12 kHz for four bearing conditions were selected to evaluate the

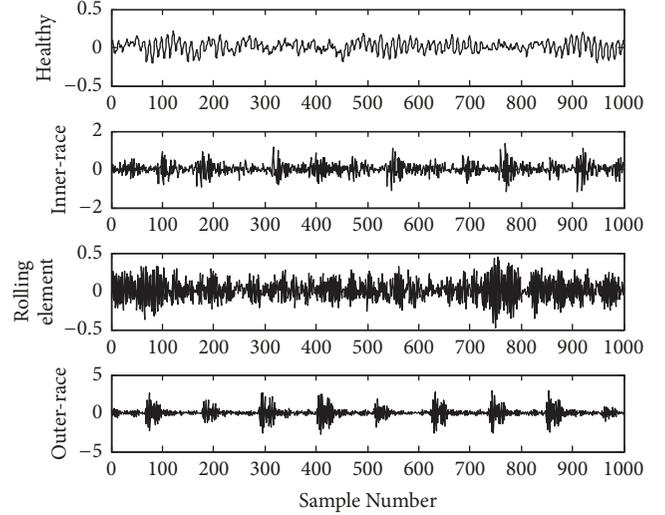


FIGURE 4: Typical waveforms from the four conditions.

TABLE 3: Bearing fault data.

Number	Bearing condition	Diameter/mm
1	Healthy	—
2	Inner-race	0.178
3	Rolling-element	0.178
4	Outer-race	0.178
5	Inner-race	0.356
6	Rolling-element	0.356
7	Outer-race	0.356
8	Inner-race	0.533
9	Rolling-element	0.533
10	Outer-race	0.533

proposed method. Single point faults with different fault diameters of 0.178 mm, 0.356 mm, and 0.533 mm were used to test in this paper. So there are ten fault classes; specific data is shown in Table 3. Each sample set selects 1000 vibration data, 120 groups of data sets in each mode.

#### 3.2. Features Extraction

(1) *Empirical Mode Decomposition.* Firstly, Process vibration signals were decomposed into intrinsic mode function by EMD method. IMF1 ~ 4 after intercepting both ends are selected in this paper; then instantaneous frequency of decomposed components was obtained by Hilbert transformation (seen in Figure 5). A total of eight components are, respectively, four IMF components and four corresponding instantaneous frequency sequences. For suppressing the end effect of EMD, each component signal cut off both ends, taking the remaining 800 data points of IMF and instantaneous frequency.

(2) *Statistical Features.* Calculate eight statistical features (mean, max, range, std, skewness, kurtosis, variation coefficient (VC), and Sum Of Squares), extract 64-dimensional

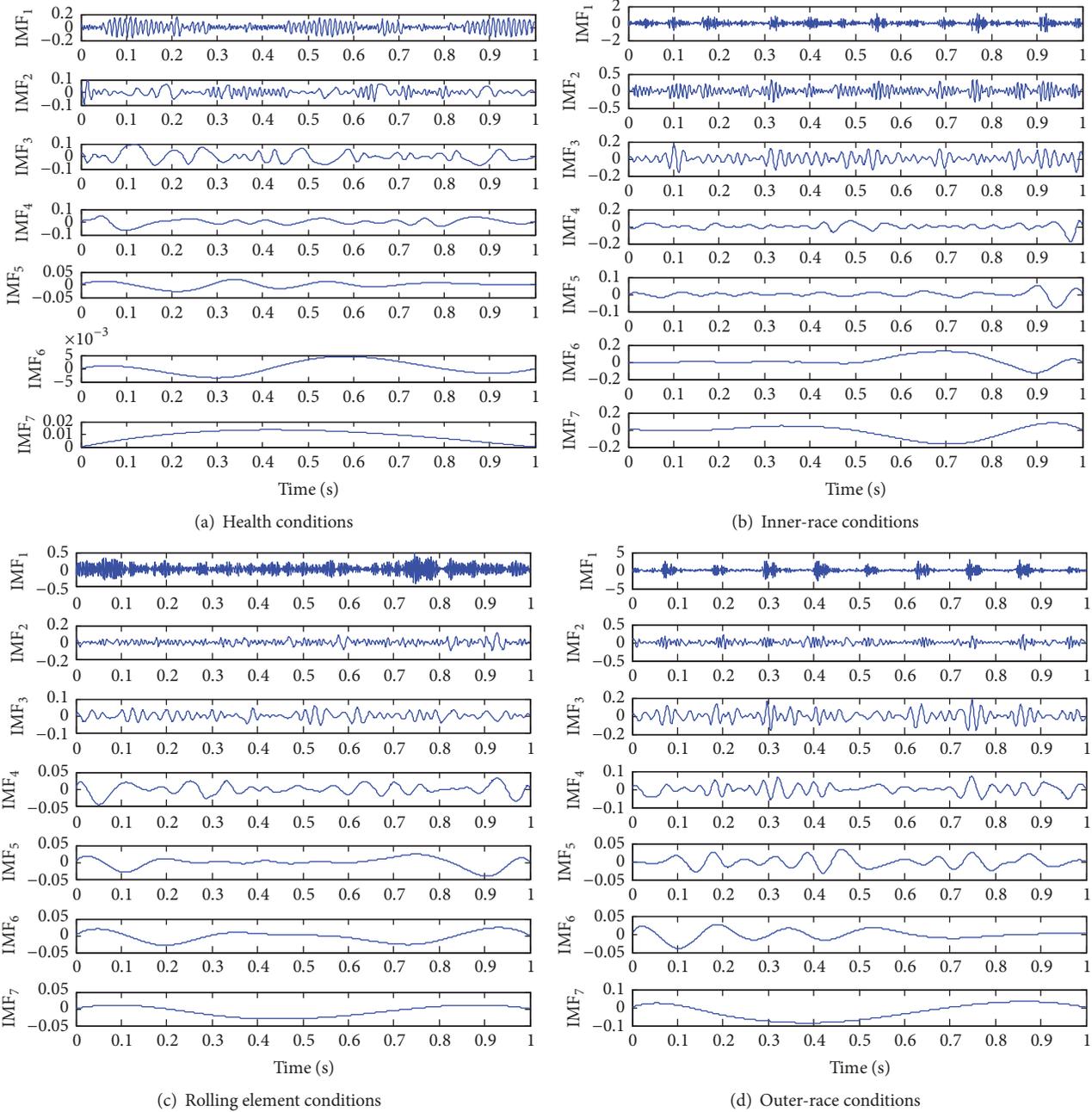


FIGURE 5: IMFs of four kinds of conditions in rolling bearing case.

feature vectors from each sample. Every patterns data form is  $120 \times 64$ .

The feature space box plots of the eight features in 10 different classes are generated by the data after EMD (seen in Figure 6). The features of different patterns present different values, which can be relatively well separated from the other classes.

(3) *Principal Component Analysis*. Principal component analysis can get the fewer variables but maintain the original information as much as possible. So in this study, after choosing the statistical and sharp features, we use the PCA method

as the second feature extraction to eliminate duplication of information. 70% of the sample are used for training and the rest 30% are for testing. The recognition accuracy can be estimated by the testing samples. In PCA, cumulative contribution threshold value is set to 90%, indicating that PCA select owns 90% of data information by the main components; the results are shown in Figure 7 and Table 4.

### 3.3. Result Analysis

(1) *Performance of Recognizer in Optimization*. After the features extraction of bearing fault, MSVM with radial basis

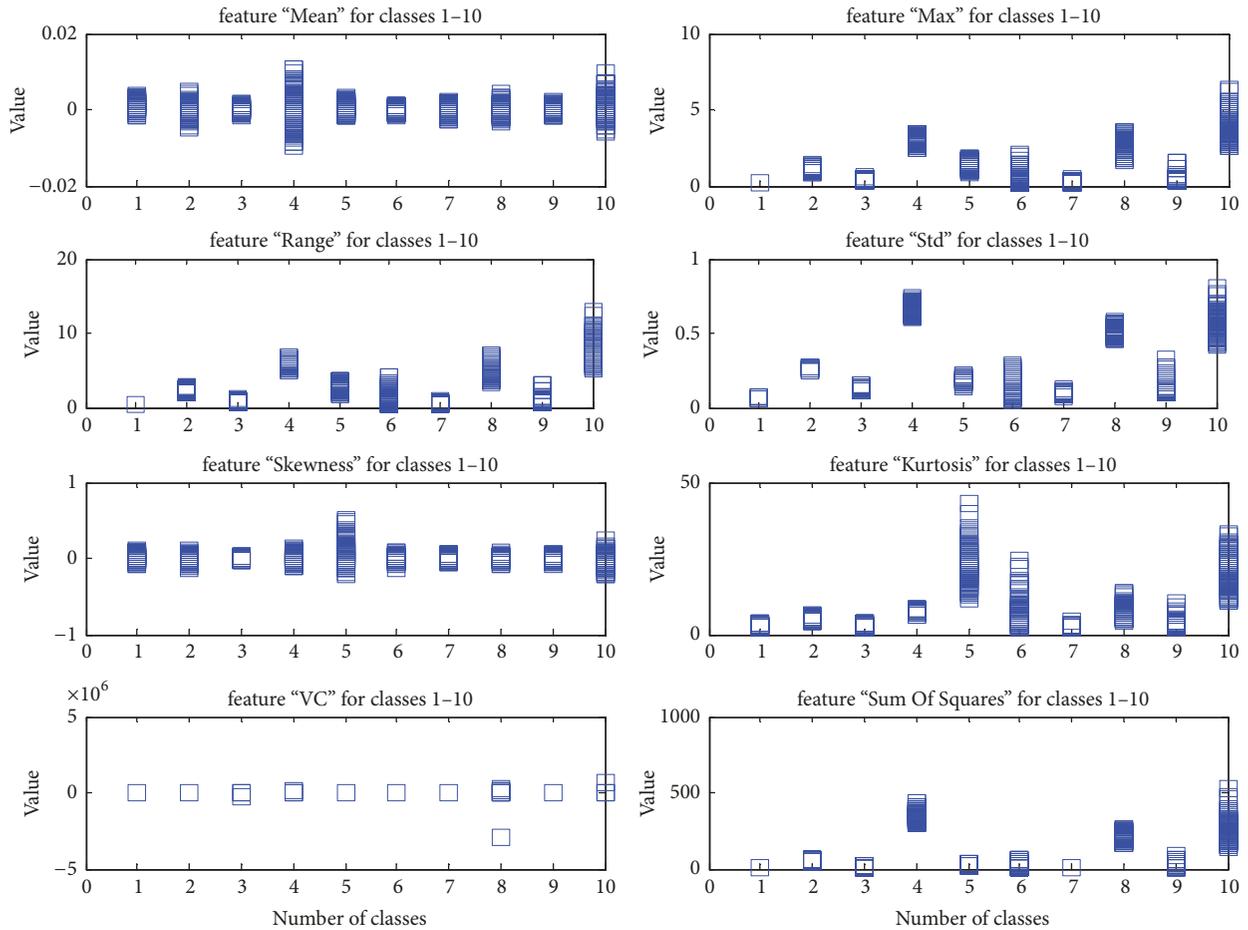


FIGURE 6: Box plots of the eight features for different classes.

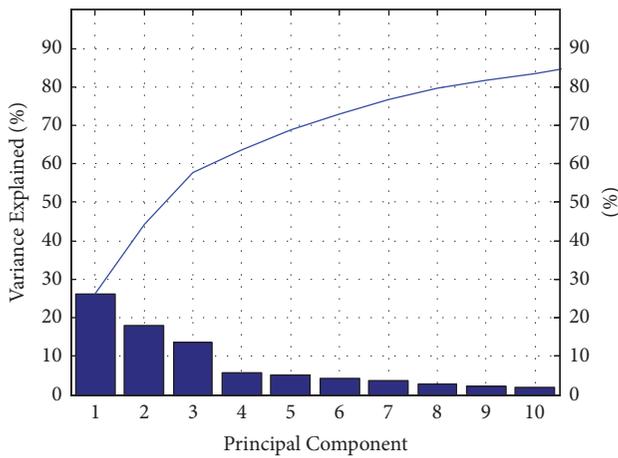


FIGURE 7: Principal components of PCA (Pareto).

function, first use the cross-validation (CV) method to get regularization parameter ( $C$ ) and kernel parameter ( $\gamma$ ), both of the search range  $[10^{-2} 10^2]$ ; the index step is 0.5, the relationship of the trained MSVM accuracy and  $C, \gamma$  is shown in Figures 8 and 9.

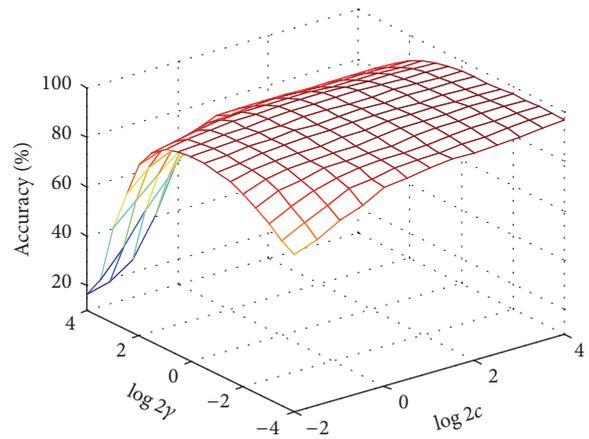
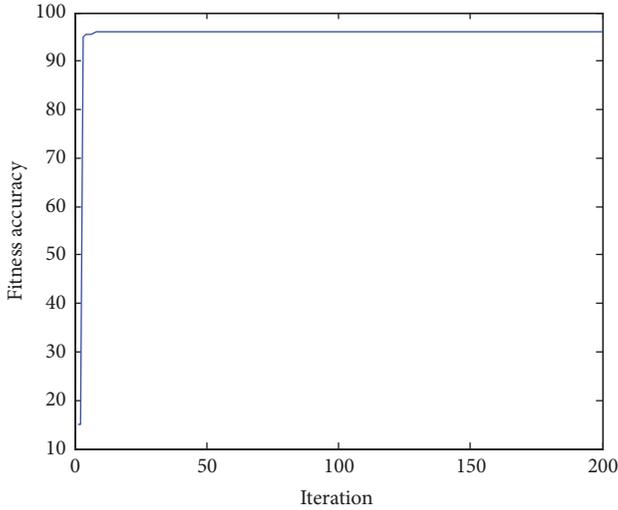


FIGURE 8: Accuracy with  $C, \gamma$ .

From Figures 8 and 9, MSVM classification recognition rate has a greater impact when  $C$  and  $\gamma$  vary over a considerable range. The lowest classification rate is only around 18%. Accordingly, use particle parameter adaptive method to optimize MSVM kernel parameters  $\gamma$  and regularization parameter  $C$ . The parameters of PPA can be seen in Table 5,

TABLE 4: Principal component analysis.

Component	Initial Eigenvalues	Percentage (%)	Cumulative (%)
1	0.5227	26.42	26.42
2	0.3128	15.81	42.23
3	0.2932	14.82	57.05
4	0.1143	5.78	62.82
5	0.1044	5.28	68.10
6	0.0857	4.33	72.44
7	0.0751	3.80	76.23
8	0.0583	2.94	79.18
9	0.0419	2.12	81.29
10	0.0389	1.97	83.26
11	0.0372	1.88	85.14
12	0.0349	1.76	86.90
13	0.0290	1.46	88.37
14	0.0280	1.41	89.78
15	0.0247	1.25	91.03

FIGURE 9: Accuracy with  $C$ ,  $\gamma$ .

and take fold number  $K$  of  $K$ -fold cross-validation as 3 (best cross-validation accuracy = 95.24%, best  $c = 100$ , and best  $g = 0.01$ ).

The average recognition accuracies of GV\_MSVM and PPA\_MSVM show that proposed PPA\_MSVM method plays a significant role in increasing the recognition accuracy. Because GV\_MSVM quite depends on the  $K$  fold results of the training sample data, but not the entire training set, PPA algorithm can solve this problem.

The simulations show that the PCA method is less effective than that of EMD statistical features. The data dimensions of the original feature set will be effectively reduced to improve the efficiency of the identification operation. And only relying on a small number of 15 main element characteristics can effectively identify the type of fault and still maintain a high accuracy rate. But on the other hand, compared with the original feature set, using PCA can cause

information missing to reduce the recognition effect, but this negative effect impact to identify the fault is small; we can accept this result.

(2) *Performance of Recognizer in Different Fault Patterns.* To verify the proposed method and analyze the classify accuracy on different fault pattern, the behavior of the simulated fault patterns number, PCA, and parameters optimization method were examined through a full factorial experiment. It mainly had the following three factors: (1) parameters optimization method with two levels (CV\_MSVM and PPA\_MSVM), (2) fault pattern number with three levels of four patterns (see Table 3, fault numbers 1–4), seven patterns (see Table 3, fault numbers 1–7), and ten patterns (see Table 3, fault numbers 1–10).

Table 6 shows the results of the full factorial experiment. The prediction accuracy can be reached to 100% when it only has 4 fault patterns. It indicates that the proposed method was effective in classifying the fault patterns. However, the accuracy will decline when the numbers of fault patterns increase. The reasons are the multiple-fault patterns information may be confused and more difficult to recognize and so result in serious performance degradation. Besides, when it is 7 fault patterns, PCA method plays an effective role in two different parameters optimization method. Finally, the results show that parameters optimization method of PPA is more effective than that of CV when coupled with the PCA and different fault pattern number.

(3) *Performance of Recognizer in Different Training Samples.* The performance of the optimization methods has been compared with CV for investigating the capability of the proposed REBs multiple-fault patterns method. To indicate the influence of the training sample, we test the accuracies of the PCA\_PPA\_MSVM model based on the proposed EMD and statistical feature extraction, in the cases where the percentages of training samples are 40%, 50%, 60%, 70%, 80%, and 90%. The testing results are presented in Table 7.

TABLE 5: Comparison of the performance of PPA\_MSVM with GV\_MSVM classifiers.

Method	PCA method	Best C	Best $\gamma$	Training accuracy (%)	Prediction accuracy (%)
GV_MSVM	/	16	0.25	98.69	<b>93.33</b>
	PCA	8	0.25	98.21	<b>92.5</b>
PPA_MSVM	/	69.89	0.01	97.62	<b>96.67</b>
	PCA	100	0.01	96.42	<b>95.24</b>

TABLE 6: Comparison of the performance of recognizer in different fault patterns.

Method	Fault pattern number	PCA method	Training accuracy (%)	Prediction accuracy (%)
CV_MSVM	4	/	100	100
	4	PCA	<b>100</b>	<b>100</b>
	7	/	99.32	97.62
	7	PCA	<b>98.30</b>	<b>98.81</b>
	10	/	98.69	93.33
	10	PCA	<b>98.21</b>	<b>92.5</b>
PPA_MSVM	4	/	100	100
	4	PCA	<b>100</b>	<b>100</b>
	7	/	99.49	98.41
	7	PCA	<b>98.30</b>	<b>99.21</b>
	10	/	97.62	96.67
	10	PCA	<b>96.42</b>	<b>95.28</b>

As shown by the experimental results, prediction accuracy can be the best when the training set size is increased to 70% of the total sample set. The reason is the prediction accuracy will be higher when the training model gets the best parameters. But it also has the overfitting problem; we can find the prediction accuracy to be not well when the percentages of training samples are 80% and 90%.

(4) *Performance of Recognizer in Different Feature Extraction Methods.* Feature extraction can lead to faster training and more efficiency in multiple-fault diagnosis method. Thirteen statistical and shape features are utilized as the inputs in this paper. In order to explain its effectiveness, MSVM classifier using the EMD, PCA, and MFE (Combine EMD and PCA) as the feature extraction method is constructed. Table 8 shows the recognition accuracy of three different feature extraction methods.

The average prediction accuracies of EMD\_MSVM (8.33%), PCA\_MSVM (72.5%), and MEF\_MSVM (94.50%) show that feature extraction method plays an important role in improving the recognition accuracy. From the results, we can find that multiple-fault diagnoses are difficult to recognize due to the complex relation, but the result is much better after using multiscale feature extraction (MEF) method, which decomposes the data into intrinsic mode function empirical mode decomposition method and instantaneous frequency of decomposed components was obtained by Hilbert transformation, and then statistical features and principal component analysis are utilized to extract significant information from the features.

## 4. Conclusion

The objective of this study is to propose a fusion approach for the multiple-fault diagnosis with single and coupling faults, by multiscale feature extraction with integrating three information methods (empirical mode decomposition, statistical features extraction, and principal component analysis) of signal progress, respectively, in time domain, frequency domain, and time–frequency domain. MSVM method with particle parameter adaptive (PPA) parameters optimization will classify the fault patterns. From this discussion, the proposed MFE\_MSVM\_PPA method can produce the highest average correct classifier accuracy compared with other methods in experiments. Besides, we analyze the influences of the prediction accuracies under different elements, like parameters optimization method, fault pattern number, PCA, and training sample size. The proposed classification method holds high precision on multiple faults fusion diagnosis and is proved to be a promising diagnosis approach for catering to the increasing characteristic parameters and feature information.

This multiple-fault diagnosis approach is feasible and, as the computational results show, quite effective in improving the compound faults diagnosis of rolling bearing fault patterns. While still immature, the data is from simulation, not relying on significant real-time testing. Because getting field data for validation of the approach is very difficult, we generate the simulated data using the rolling bearings faults original data from the Case Western Reserve University, then to analyze the multiple-fault pattern recognition problem.

TABLE 7: The optimal values of the MSVM classifier parameters for different training samples.

Method	Training set number (each pattern)	Testing set number (each pattern)	Training accuracy (%)	Prediction accuracy (%)
40%	48	72	93.33	86.39
50%	60	60	94.5	93.00
60%	72	48	94.72	90.63
70%	84	36	96.42	<b>95.28</b>
80%	96	24	96.20	88.33
90%	108	12	97.59	89.17

TABLE 8: Prediction accuracy of three different feature extraction methods.

Method	Prediction accuracy (%)
EMD_MSVM	8.33
PCA_MSVM	72.50
MEF_MSVM	94.50

The future work will be focused on the following aspects: (1) employing multiscale feature extraction (MFE) as feature extraction method which we will compare with other excellent feature extraction methods; (2) comparing particle parameter adaptive (PPA) with other intelligent algorithms; (3) researching the fundamental principles of multiple faults diagnosis; and (4) study with real data is certainly our next research task.

## Conflicts of Interest

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

## Acknowledgments

This work is financially supported by the Fundamental Research Funds for the Central Universities under Grant no. 2682016CX031 and National Natural Science Foundation of China (NSFC) under Grant no. 51175442.

## References

- [1] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 108–126, 2013.
- [2] L. Saidi, F. Fnaiech, H. Henao, G. A. Capolino, and G. Cirrincione, "Diagnosis of broken-bars fault in induction machines using higher order spectral analysis," *ISA Transactions*, vol. 52, no. 1, pp. 140–148, 2013.
- [3] V. Ranaee, A. Ebrahimzadeh, and R. Ghaderi, "Application of the PSOSVM model for recognition of control chart patterns," *ISA Transactions*, vol. 49, no. 4, pp. 577–586, 2010.
- [4] H. De La Torre Gutierrez and D. T. Pham, "Estimation and generation of training patterns for control chart pattern recognition," *Computers & Industrial Engineering*, vol. 95, pp. 72–82, 2016.
- [5] G. D. Pelegrina, L. T. Duarte, and C. Jutten, "Blind source separation and feature extraction in concurrent control charts pattern recognition: Novel analyses and a comparison of different methods," *Computers & Industrial Engineering*, vol. 92, pp. 105–114, 2016.
- [6] Z. Chen, S. Lu, and S. Lam, "A hybrid system for SPC concurrent pattern recognition," *Advanced Engineering Informatics*, vol. 21, no. 3, pp. 303–310, 2007.
- [7] V. Ranaee and A. Ebrahimzadeh, "Control chart pattern recognition using an optimized neural network and efficient features," *ISA Transactions*, vol. 49, no. 3, pp. 387–393, 2010.
- [8] C.-J. Lu, Y. E. Shao, and P.-H. Li, "Mixture control chart patterns recognition using independent component analysis and support vector machine," *Neurocomputing*, vol. 74, no. 11, pp. 1908–1914, 2011.
- [9] A. G. Rehorn, E. Sejdić, and J. Jiang, "Fault diagnosis in machine tools using selective regional correlation," *Mechanical Systems and Signal Processing*, vol. 20, no. 5, pp. 1221–1238, 2006.
- [10] P. Konar and P. Chattopadhyay, "Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs)," *Applied Soft Computing*, vol. 11, no. 6, pp. 4203–4211, 2011.
- [11] B. Liu, S. Riemenschneider, and Y. Xu, "Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum," *Mechanical Systems and Signal Processing*, vol. 20, no. 3, pp. 718–734, 2006.
- [12] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.
- [13] X. Liu, L. Bo, and H. Luo, "Bearing faults diagnostics based on hybrid LS-SVM and EMD method," *Measurement*, vol. 59, pp. 145–166, 2015.
- [14] C. Wang, M. Gan, and C. Zhu, "Non-negative EMD manifold for feature extraction in machinery fault diagnosis," *Measurement*, vol. 70, pp. 188–202, 2015.
- [15] L. Saidi, J. Ben Ali, F. Fnaiech, and B. Morello, "Bi-spectrum based-EMD applied to the non-stationary vibration signals for bearing faults diagnosis," in *Proceedings of the 6th International Conference on Soft Computing and Pattern Recognition, SoCPaR 2014*, pp. 25–30, Tunisia, August 2014.
- [16] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, pp. 16–27, 2015.
- [17] C. H. Wang, R. S. Guo, M. H. Chiang, and J. Y. Wong, "Decision tree based control chart pattern recognition," *International Journal of Production Research*, vol. 46, no. 17, pp. 124–134, 2008.

- [18] C.-H. Wang and W. Kuo, "Identification of control chart patterns using wavelet filtering and robust fuzzy clustering," *Journal of Intelligent Manufacturing*, vol. 18, no. 3, pp. 343–350, 2007.
- [19] Z. J. Xiao, R. H. Wang, C. X. Yang et al., "Multiple-fault diagnosis in motion system based on SVM," *International Journal of Machine Learning & Cybernetics*, vol. 3, no. 1, pp. 77–82, 2012.
- [20] S. S. Tayarani-Bathaie, Z. N. Sadough Vanini, and K. Khorasani, "Dynamic neural network-based fault diagnosis of gas turbine engines," *Neurocomputing*, vol. 125, pp. 153–165, 2014.
- [21] P. Xanthopoulos and T. Razzaghi, "A weighted support vector machine method for control chart pattern recognition," *Computers & Industrial Engineering*, vol. 70, no. 1, pp. 134–149, 2014.
- [22] M. Zhang, W. Cheng, and P. Guo, "Intelligent recognition of mixture control chart pattern based on quadratic feature extraction and SVM with AMPSO," *Journal of Coastal Research*, pp. 304–309, 2015.
- [23] R.-H. Hu, P.-H. Lou, D.-B. Tang, and M.-D. Liu, "Fault diagnosis of rolling bearings based on EMD and parameter adaptive support vector machine," *Computer Integrated Manufacturing Systems, CIMS*, vol. 19, no. 2, pp. 438–447, 2013.
- [24] Case Western Reserve University Bearing Data Center. Download a data file [DB/OL]. <http://www.eecs.cast.edu/laboratory/bearing/download.htm>&gt;.
- [25] S. Wu, P. Wu, C. Wu, J. Ding, and C. Wang, "Bearing fault diagnosis based on multiscale permutation entropy and support vector machine," *Entropy*, vol. 14, no. 8, pp. 1343–1356, 2012.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

