

Research Article

Fault Diagnosis for Reducer via Improved LMD and SVM-RFE-MRMR

Xiaoguang Zhang , Zhenyue Song , Dandan Li, Wei Zhang ,
Zhike Zhao, and Yingying Chen

School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China

Correspondence should be addressed to Zhenyue Song; cumtsongzhenyue@163.com

Received 5 February 2018; Revised 13 May 2018; Accepted 10 June 2018; Published 15 July 2018

Academic Editor: Daniele Baraldi

Copyright © 2018 Xiaoguang 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.

The vibration signals are usually characterized by nonstationary, nonlinearity, and high frequency shocks, and the redundant features degrade the performance of fault diagnosis methods. To deal with the problem, a novel fault diagnosis approach for rotating machinery is presented by combining improved local mean decomposition (LMD) with support vector machine–recursive feature elimination with minimum redundancy maximum relevance (SVM-RFE-MRMR). Firstly, an improved LMD method is developed to decompose vibration signals into a subset of amplitude modulation/frequency modulation (AM-FM) product functions (PFs). Then, time and frequency domain features are extracted from the selected PFs, and the complicated faults can be thus identified efficiently. Due to degradation of fault diagnosis methods resulting from redundant features, a novel feature selection method combining SVM-RFE with MRMR is proposed to select salient features, improving the performance of fault diagnosis approach. Experimental results on reducer platform demonstrate that the proposed method is capable of revealing the relations between the features and faults and providing insights into fault mechanism.

1. Introduction

Rotating machines are the heart of some modern equipment such as aircraft engine, gas turbine, and reducer and often a common source of machine faults. However, most vibration signals are nonstationary, nonlinear, and disturbed with heavy background noises, and the machinery components are seriously coupled, providing a challenge to detect and identify specific rotating machine faults accurately [1, 2]. For accurate diagnosis of rotating machine fault, a number of techniques have been developed in recent years [3–5].

Generally, vibration signals from mechanical equipment are the most effective and common way to identify the states and faults of rotating machine [3]. According to the domains to which the extracted features belong, machinery fault diagnosis methods can be categorized into time domain-based, frequency domain-based, and time-frequency domain-based diagnosis methods [4]. The time domain-based features have intuitive physical meaning with low computation complexity while the classification, location, and severity of fault cannot

be identified through these features. The frequency domain-based methods like Fourier transform and its variants detect the fault according to the spectrum of vibration signal. However, Fourier transform is not capable of dealing with nonstationary and nonlinear signal. Time-frequency domain diagnosis methods can decompose signal into multiple-scale components; the local characteristics of signal can be revealed. Moreover, the components are approximately stationary and linear, and the time or frequency features are then extracted to detect machinery faults.

Due to poor working environment, fault features are usually immersed in heavy background noise. Especially for incipient fault, the fault features are too weak to be extracted [5]. Time-frequency analysis techniques, such as Wigner-Ville distribution [6], wavelet transform [7], and EMD [8], are effective in suppressing signal noise. In Wigner-Ville distribution, cross-term interferences among multiple component signals impair fault feature because of quadratic form of Wigner-Ville distribution. Due to the limited length of wavelet basis function, energy leakage may happen for

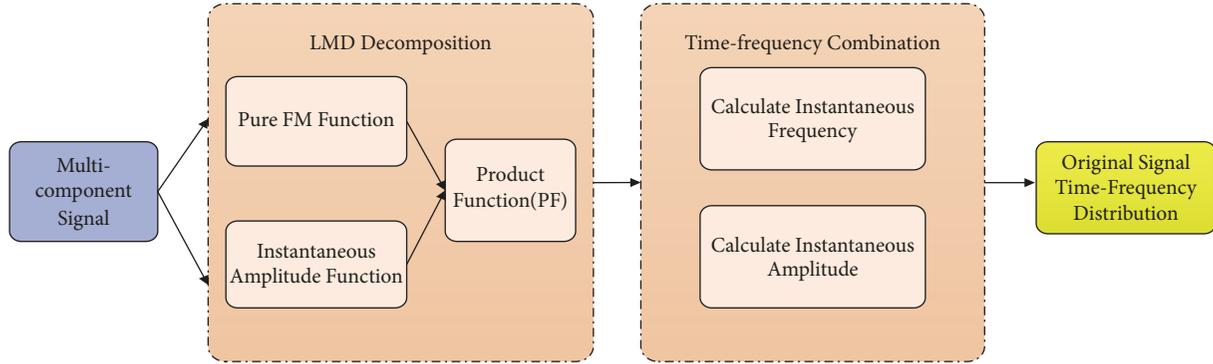


FIGURE 1: Schematic of local mean decomposition.

wavelet transformation, degrading the resolution signals in time domain and frequency domain. Empirical mode decomposition (EMD) suffers from mode mixing and end effect. SMITH [9] proposed the LMD to compute instantaneous frequency and the distribution of time-frequency energy of electroencephalogram signals in 2005. The method suppresses end effect, eliminates over-envelope and under-envelope, and reduces computational complexity. Compared with EMD, LMD achieves instantaneous amplitude and frequency without Hilbert transformation and fitting the upper and lower envelopes. Therefore, LMD not only overcomes the problems existing in EMD, but also concentrates more useful information in a few decomposition components, which is suitable to analyze nonlinear and nonstationary signal [10]. At present, the LMD integrating with machine learning methods has been developed to enhance fault features and identify fault patterns [11–14].

Fault feature extraction from vibration signals is a critical step of fault diagnosis. However, in many real-world applications, there are some redundant features in high-dimensional data, reducing generalization performance of fault detection model. Therefore, removing redundant features from high-dimensional data has become an effective way to enhance the performance of fault diagnosis model [15, 16]. Aiming at the insufficient stability when variable predictive model class discrimination (VPMCD) was applied to small samples or in multi-correlative feature space, Tang et al. [15] proposed ARSFS method based on affinity propagation (AP) clustering, RReliefF, and sequential forward search for feature selection. The experimental results showed that ARSFS effectively identified the faults of the rolling bearing. Intelligence optimization methods have been commonly applied for feature selection, which achieves better identification accuracy than most existing feature selection approaches [16]. Support vector machine–recursive feature elimination (SVM-RFE) based on one-to-one or one-to-many strategy provides a feasible way to identify multiple machinery faults [17–19]. Tang et al. [18] proposed a two-stage SVM-RFE, in which feature extraction and the feature subset selection were conducted sequentially. However, SVM-RFE barely takes the degree of the features effect on the classifier into account and ignores redundant features. Sometimes, the classification accuracy on subset consisting of M features from N ($M < N$) salient features

may be higher than that on the subset consisting of N salient features in some applications.

Taking the above problem into account, minimum redundancy maximum relevance (MRMR) algorithm is introduced to evaluate the feature subset selected by SVM-RFE, and the optimal feature subset is thus yielded. In MRMR, mutual information is used to measure the relevance and redundancy of features, and the feature subset is achieved through two cost functions called information difference and information entropy [20]. This algorithm is called support vector machine–recursive feature elimination–minimum redundancy maximum relevance (SVM-RFE-MRMR). As seen later, the new method is capable of effectively dealing with the problem of traditional feature selection algorithm.

The structure of this paper is organized as follows: Section 2 briefly introduces the improved LMD method, and the elimination of end effect is illustrated. SVM-RFE-MRMR algorithm is proposed and the details are given in Section 3. Experiments are performed on a reducer test system to verify the effectiveness of the proposed method in Section 4. Finally, some conclusions are drawn in Section 5.

2. Improved Local Mean Decomposition

LMD is an adaptive time-frequency analysis method similar to empirical mode decomposition (EMD). Both methods can decompose a complex signal into several single component signals with physical meaning. Compared with EMD, LMD can diminish end effect and eliminate over-envelope and under-envelope with low computational complexity and less information loss [21].

LMD decomposes complicated signals into a set of product functions (PF). The instantaneous frequency of each PF is the product of an envelope signal and a frequency modulated signal. The time-frequency distribution of original signal can be yielded by combining the instantaneous amplitude and instantaneous frequency of all PF components. The schematic of LMD is shown in Figure 1.

Local extreme value is the premise for evaluating product functions. Due to the limited length of a signal, the endpoint of the signal may not be the extreme point. Then false components will gradually “pollute” the whole signal sequence inward from two endpoints, resulting in the divergence of

envelope function and degradation of the decomposition results, namely, end effect. To deal with the problem, an improved LMD method is proposed. Firstly, support vector regression (SVR) model is utilized to extend both ends of the original signal, and then the extended signal is decomposed into a set of PFs via LMD.

Assuming a set of data $\{(x_i, y_i)\}_{i=1}^N$, where $x_i \in R^D$, $y_i \in R$, and N is the length of the samples, a SVR model is formulated as follows:

$$f(x) = w^T \phi(x) + b \quad (1)$$

where $\phi(x)$ is a nonlinear mapping function to map the sample data into the high-dimensional feature space, w is the weight vector, and b is the offset.

Introducing penalty parameter C and insensitive loss function, the SVR model can be achieved by solving the following convex optimization problem:

$$\begin{aligned} \min_{w,b} \quad & P = \frac{1}{2} w^T w + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & y_i - w^T \phi(x) - b \leq \varepsilon + \xi_i \\ & w^T \phi(x) + b - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \dots, N \end{aligned} \quad (2)$$

where ξ_i and ξ_i^* are the slack variables.

This above optimization problem can be transformed into the dual problem by introducing kernel trick; i.e.,

$$\begin{aligned} \min_{\alpha, \alpha^*} \quad & D \\ & = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N K(x_i, x_j) (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \\ & \quad - \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) + \sum_{i=1}^N \varepsilon (\alpha_i + \alpha_i^*) \end{aligned} \quad (3)$$

$$\text{s.t.} \quad \alpha_i, \alpha_i^* \in [0, C], \quad i = 1, 2, \dots, N$$

$$\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0$$

where α_i and α_i^* are the Lagrangian multipliers. Finally, the SVR model is obtained.

$$y = f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (4)$$

Here, radial basis function (RBF) is selected as the kernel function.

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) \quad (5)$$

where σ and C is required to be prespecified.

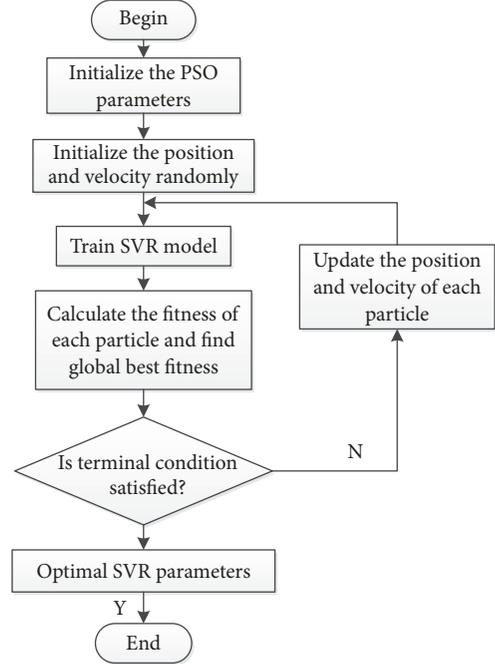


FIGURE 2: The process of optimized PSO-SVR.

Penalty parameter C and the kernel parameter σ play a significant role in the model generalization. Particle swarm optimization (PSO) algorithm is introduced to find the optimal parameters. The hybrid PSO-SVR method treats the parameters C and σ as the particles, and the velocity and location are constantly updated.

In above 2D searching space, population $X = (X_1, X_2, \dots, X_n)$ consists of n particles, where the i th particle is expressed as a 2D vector $X_i = (C_i, \sigma)^T$.

Let $V_i = (V_{i1}, V_{i2})^T$ be the velocity of i th particle and denote individual and global extremums by $P_{i,best} = (P_{i1}, P_{i2})^T$ and $g_{best} = (g_{best1}, g_{best2})^T$, respectively. In each iteration, the velocity and position of each particle are updated according to the following formula:

$$V_{i,d}^{k+1} = \omega V_{i,d}^k + c_1 r_1 (P_{i,best,d}^k - X_{i,d}^k) + c_2 r_2 (g_{best,d}^k - X_{i,d}^k) \quad (6)$$

$$X_{i,d}^{k+1} = X_{i,d}^k + V_{i,d}^{k+1} \quad (7)$$

where ω represents the inertia factor, d represents d -dimensional space, $i = 1, 2, \dots, n$, k represents the current number of iterations, $V_{i,d}$ represents the velocity of i th particle, $c_1 = c_2 = 2$ are acceleration factors, and r_1 and r_2 are random numbers between $[0, 1]$.

The process of optimized PSO-SVR is illustrated in Figure 2.

The fitness function is mean square error (MSE).

$$MSE_i = \frac{1}{l} \sum_{i=1}^l (f(x_i) - y_i)^2 \quad (8)$$

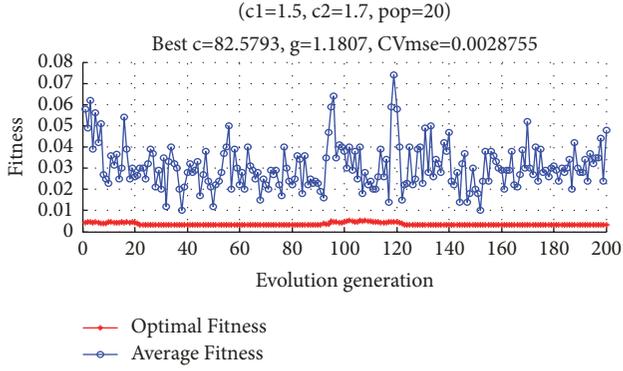


FIGURE 3: Optimal parameter results of PSO.

where MSE_i represents the fitness of i th particle, $f(x_i)$ is the output of SVR with input x_i , y_i is the expected output, and l is the number of samples.

In order to demonstrate the efficiency of the improved LMD method, end effect evaluation index δ is introduced, which is calculated as follows:

$$\delta = \frac{1}{R_0} \left| R_0 - \sqrt{R_u^2 + \sum_{p=1}^k R_p^2} \right| \quad (9)$$

where R_0 is the effective value of original signal, R_p is the effective value of the p th PF, R_u is the effective value of residual component, and

$$R = \sqrt{\frac{1}{l} \sum_{i=1}^l s^2(i)} \quad (10)$$

where $s(i)$ is the signal to be calculated. The larger the δ ($\delta > 0$) is, the greater the influence of end effect is.

Based on the above specifications, a synthesis signal is generated as

$$y = 1.8 * [\cos(2\pi * 20t + 2 \sin(2\pi * t))] + \sin(0.6\pi * t) * \sin(2\pi * 6t). \quad (11)$$

The sampling frequency is 500 Hz; the time interval is $[0.056, 0.654]$ and it contains 300 points in total. Optimal parameter results of PSO with generations are shown in Figure 3. The optimized penalty parameter is $C = 82.6$, and the kernel function parameter is $\sigma = 1.181$.

The extreme mirror extension method and the SVR-based extension method are performed on the simulation signal. LMD results of the extreme mirror extension and SVR-based extension methods are shown in Figures 4 and 5; the corresponding end effect evaluation indexes are $\delta = 0.0348$ and $\delta = 0.0144$, respectively. It can be seen from Figure 4 that PF2 has obvious distortion in its left and right endpoint, so the method cannot characterize the trend of the original signal. Based on the comparison of two methods, it is induced that the SVR-based extension approach outperforms the other method.

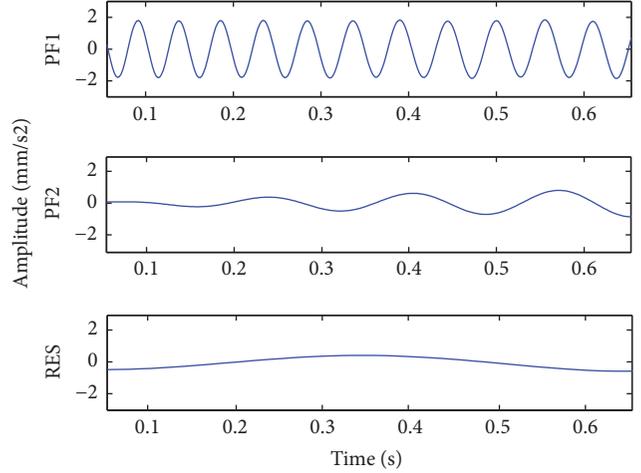


FIGURE 4: LMD results of the extremal mirror extension method.

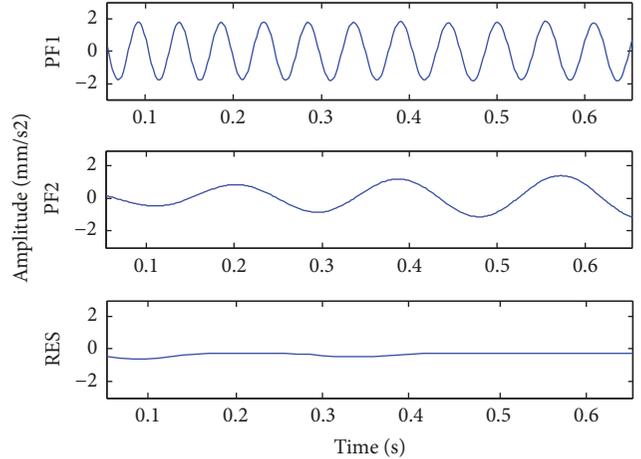


FIGURE 5: LMD results of the SVR-based extension method.

3. Proposed SVM-RFE-MRMR Method

3.1. Support Vector Machine-Recursive Feature Elimination (SVM-RFE). SVM-RFE was firstly introduced to rank genes from gene expression data for cancer classification [22]. SVM-RFE performs backward feature elimination. All features are firstly sorted in terms of the ranking scores of features, and the feature with smaller ranking score will be removed from candidate feature subset. Similar to the above procedure, SVM is iteratively trained on the new candidate feature subset, and the remaining features are resorted until one feature is left in candidate feature subset.

More specifically, given training dataset $\{(x_i, y_i)\}_{i=1}^N$, where x_i is the training sample, $x_i \in R^D$, $y_i \in \{-1, +1\}$ is the class label of x_i , N is the number of training samples, and D is the feature dimension of the sample, the decision function of SVM is expressed as $f(x) = w^T x + b$, where $w = [w_1, w_2, \dots, w_m]^T$ is a vector and b is a scalar. By introducing

Input: Training samples $\{x_i, y_i\}_{i=1}^N$, $y_i \in \{-1, +1\}$.
Output: Ranked feature list R .
(A) Initialize: The original feature subset $S = \{f_1, f_2, \dots, f_i\}$ and the ranked feature list $R = \emptyset$.
(B) Loop following procedures until $S = \emptyset$.
(1) **while** $S \neq []$ **do**
(2) Train a SVM with features in S
(3) **for** all i th feature in S **do**
(4) Compute c_i using equation (14)
(5) **end for**
(6) $p = \arg \min_i c_i$
(7) $R = \{p\} \cup R$
(8) $S = S - \{p\}$
(9) **end while**
(10) **return** R

ALGORITHM 1: SVM-RFE.

the kernel trick, the dual optimization problem of SVM can be written as

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) - \sum_{i=1}^N \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^N y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, N \end{aligned} \quad (12)$$

where C is a trade-off between the training accuracy and the model complexity.

In SVM, the importance of features is measured by their weight coefficients:

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad (13)$$

where α_i is achieved by solving the optimization problem in (12).

In SVM-RFE, the ranking score of i th feature is defined as

$$c_i = w_i^2. \quad (14)$$

The detailed algorithm of SVM-RFE is shown in Algorithm 1.

From Algorithm 1, it is induced that the ranking criteria can measure the correlation between features and decisions by judging the weight of features. Although SVM-RFE can eliminate those unimportant features one by one, it is noted that whether the deleted features are redundant is a problem that requires attention. To address the issue, MRMR is introduced in the paper.

3.2. Minimum Redundancy Maximum Relevance (MRMR) Method. Ding and Peng introduced a criterion firstly to measure relevance and redundancy of features by using

mutual information called minimum redundancy maximum relevance (MRMR) [23]. MRMR measures the maximal sample information and the minimal relevance among features by defining maximum relevance criteria and minimum redundancy criteria, respectively. MRMR can achieve the ranking score of each feature. Specifically, given two feature vectors x and y with probability density $p(x)$ and $p(y)$, let $p(x, y)$ be the joint probability density of x and y . Their mutual information (MI) which measures the redundancy among features is defined as follows:

$$I(x, y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy. \quad (15)$$

The minimum redundancy and maximum relevance are calculated as

$$\min R(F), \quad R = \frac{1}{|F|^2} \sum_{f_r, f_o \in F} I(f_r, f_o) \quad (16)$$

$$\max D(F, c), \quad D = \frac{1}{|F|} \sum_{f_r \in F} I(f_r, c) \quad (17)$$

where F and $|F|$ represent feature set and the number of features in F , respectively, c is class label, $I(f_r, c)$ represents the MI of feature f_r and class label c , $I(f_r, f_o)$ represents the MI of feature f_r and feature f_o , D is the mean of MI, and R represents the MI among the features.

MRMR intends to yield the features with minimum redundancy and maximum relevance through the following two criteria:

$$\begin{aligned} \max \Phi_1(D, R), \quad \Phi_1 &= D - R \\ \max \Phi_2(D, R), \quad \Phi_2 &= \frac{D}{R}. \end{aligned} \quad (18)$$

3.3. Support Vector Machine-Recursive Feature Elimination-Minimum Redundancy Maximum Relevance (SVM-RFE-MRMR). According to the above introduction, SVM-RFE takes the correlation between features and decisions into

Input: Training samples $\{x_i, y_i\}_{i=1}^N$, $y_i \in \{-1, +1\}$.
Output: Ranked feature list R .
(A) Initialize: The original feature subset $S = \{f_1, f_2, \dots, f_i\}$ and the ranked feature list $R = \emptyset$.
(B) Loop following procedures until $S = \emptyset$:
(1) **while** $S \neq []$ **do**
(2) Train a SVM with features in S
(3) **for** all i th feature f_i in S **do**
(4) Compute c_i using equation (14)
(5) Compute r_i using equation (19)
(6) **end for**
(7) Find \bar{c}_i and \bar{r}_i using equation (20) and (21) respectively
(8) **for** all i th feature f_i in S **do**
(9) Compute $s(i) = \bar{c}_i + \bar{r}_i$
(10) **end for**
(11) $p = \arg \min_i s_i$
(12) $R = \{p\} \cup R$
(13) $S = S - \{p\}$
(14) **end while**
(15) **return** R

ALGORITHM 2: SVM-RFE-MRMR.

account, ignoring the relation among features while MRMR can yield the features with minimum redundancy and maximum relevance. Therefore, SVM-RFE-MRMR is proposed to yield a subset of features with minimum redundancy and maximum relevance. Besides the ranking criterion defined by SVM-RFE, SVM-RFE-MRMR defines a new criterion for feature f_i , which is formulated by

$$r_i = \frac{I(f_r, c)}{(1/|F|)} \sum_{f_o \in F} I(f_r, f_o) \quad (19)$$

where F is the set of features that remain in each iteration. A ranking score is defined by normalizing SVM-RFE criterion and MRMR criterion of each feature; i.e.,

$$\bar{c}_i = \frac{c_i}{c_i^*} \quad (20)$$

$$\bar{r}_i = \frac{r_i}{r_i^*} \quad (21)$$

where $c_i^* = \max_i c_i$, $r_i^* = \max_i r_i$.

The overall algorithm of SVM-RFE-MRMR is described in Algorithm 2.

4. Fault Identification and Analysis for Reducer

4.1. Experimental Device. Fault diagnosis for reducer is conducted on vibration data while it is difficult to acquire multiple fault signals in practice. Therefore, a reducer simulation platform is constructed to acquire fault signals. The mechanical system of the whole device is shown in Figure 6(a), and the number of each component is shown in Figure 6(b). This system mainly consists of motor, reducer, magnetic powder brake, vibration sensor, etc. The vibration sensor is installed at the 4# bearing. The magnetic powder brake is used to

simulate mechanical load by adjusting the voltage. In the experiment, a tooth of gear b, a part of inner circle in 4# bearing, and a part of outer circle in 4# bearing were cut off by line cutting. Figure 7 shows broken tooth fault, inner circle fault, and outer circle fault, respectively. The sampling frequency was 4 kHz, and the motor rotating speed was 1420 r/min. Original waveforms of various states are shown in Figure 8.

4.2. Fault Feature Generation. The fault diagnosis process for reducer under different states is shown in Figure 9. After acquiring the original vibration waveform, we select reasonable signal segments that can reflect fault. Then we use the improved LMD algorithm to decompose the 4 kinds of vibration signals and obtain corresponding PFs (Figures 10–13). It is particularly important to extract effective feature parameters. Common fault features include amplitude domain features, time domain features, frequency domain features, and time-frequency features. Approximate entropy as a time domain characteristic has good anti-noise, anti-interference ability and good robustness. Power spectral entropy as a frequency domain characteristic represents the distribution state of vibration energy in frequency domain. The more uniform the distribution of vibration energy, the less concentrated the signal distribution; namely, the signal is more complex. Therefore, we select the PF that contains fault feature frequency to calculate the root mean square, kurtosis, approximate entropy, and power spectral entropy in this paper. All these parameters are combined to form the original feature set as shown in Table 1. Only part of the data is listed due to limited space.

The feature parameters extracted from vibration signals under 4 states are arranged in sequence into a matrix of 120 rows and 8 columns, and each column of the matrix is normalized to $[0, 1]$.

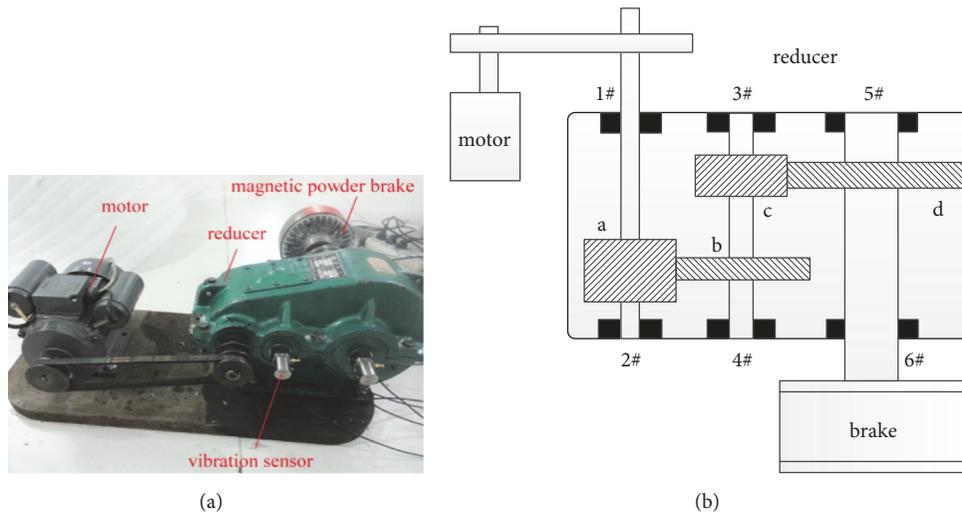


FIGURE 6: Experimental device: (a) part name, (b) part number.

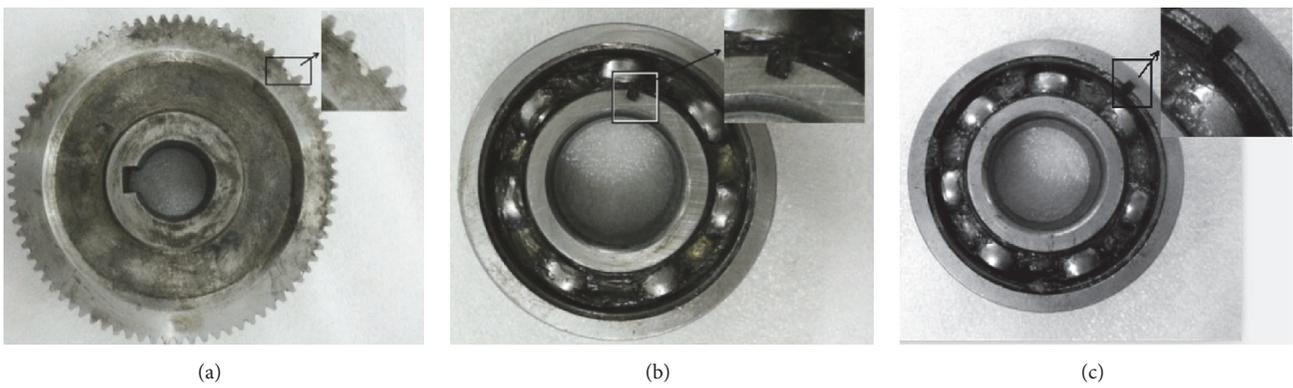


FIGURE 7: Fault pictures of (a) broken tooth fault, (b) inner circle fault, and (c) outer circle fault.

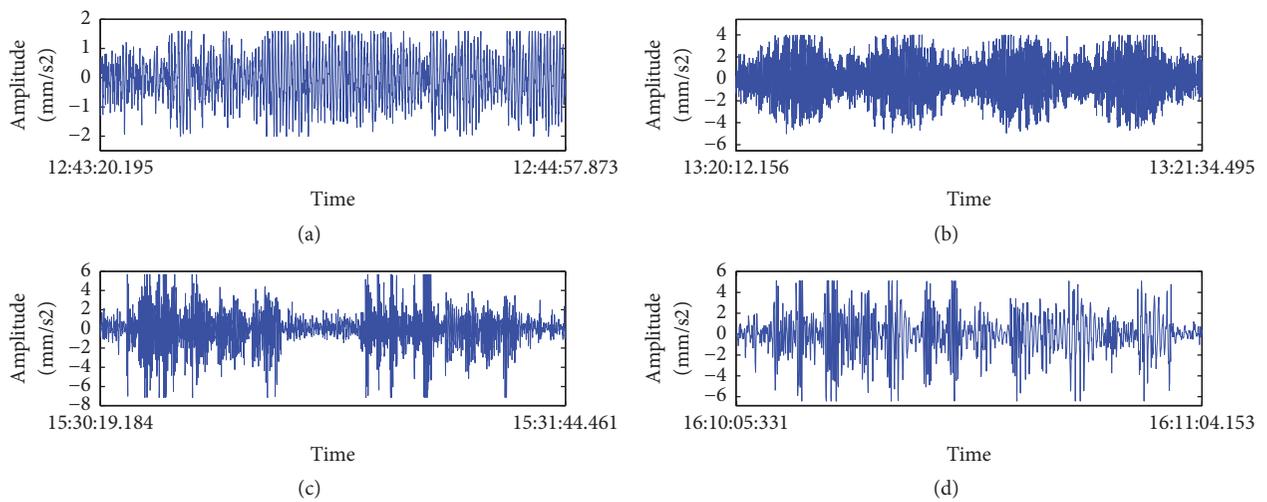


FIGURE 8: Original waveforms of various states of the gearbox: (a) vibration signal of normal state, (b) vibration signal of broken tooth fault, (c) vibration signal of 4# bearing inner circle fault, (d) vibration signal of 4# bearing outer circle fault.

TABLE I: Feature parameters list.

Feature parameters	X_{rms} rms1	β kpf1	$ApEn$ app_pfl	HF HF1	X_{rms} rms2	β kpf2	$ApEn$ app_pf2	HF HF2
Type		Normal PFI				Normal PF2		
1	2.51	2.77	0.13	3.45	1.07	0.79	0	1.86
1	2.67	2.67	0.14	3.62	1.01	0.87	0.01	1.82
1	2.83	3.13	0.12	2.94	1.17	1.13	0.01	1.9
Type		Broken tooth PFI				Broken tooth PF2		
2	7.28	6.39	0.5	5.92	1.99	1.39	0.31	3.6
2	10.3	7.98	0.5	5.81	2.15	1.69	0.22	3.56
2	10.78	7.57	0.61	6.7	1.54	3	0.14	3.51
Type		Inner circle fault PFI				Inner circle fault PF2		
3	5.89	8.9	0.38	5.83	2.15	1.63	0.19	3.1
3	6.85	7.99	0.42	5.41	1.54	1.85	0.18	3.6
3	5.98	5.12	0.47	4.99	1.47	2.06	0.21	3.16
Type		Outer circle fault PFI				Outer circle fault PF2		
4	7.8	8.45	0.25	4.19	2.05	1.31	0.2	3.17
4	7.5	8.32	0.21	4.42	1.53	1.38	0.25	3.19
4	8.38	8.37	0.33	4.22	1.88	1.26	0.23	3.15

Notes. X_{rms} : RMS; β : Kurtosis, $ApEn$: Approximate Entropy, HF: Power Spectrum Entropy.

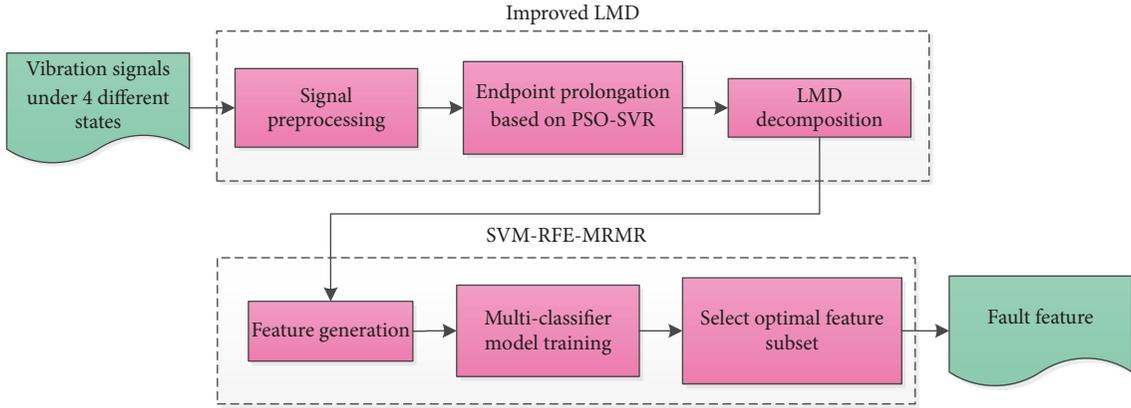


FIGURE 9: Flowchart of fault diagnosis for reducer.

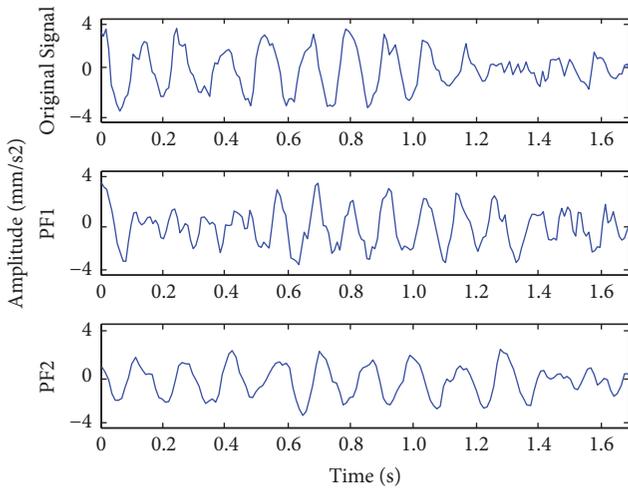


FIGURE 10: LMD results of normal signal.

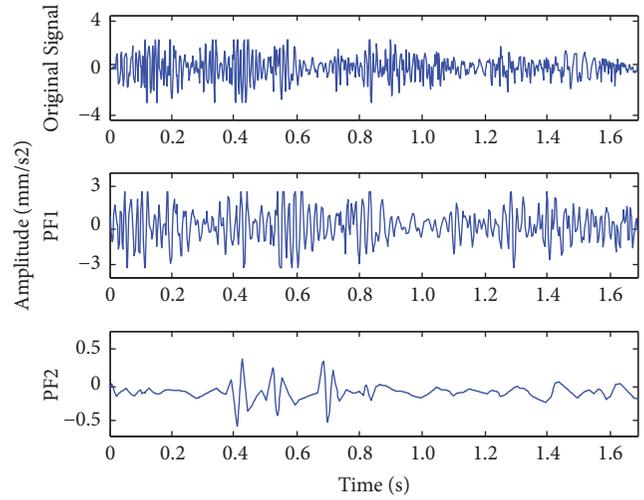


FIGURE 12: LMD results of inner circle fault signal.

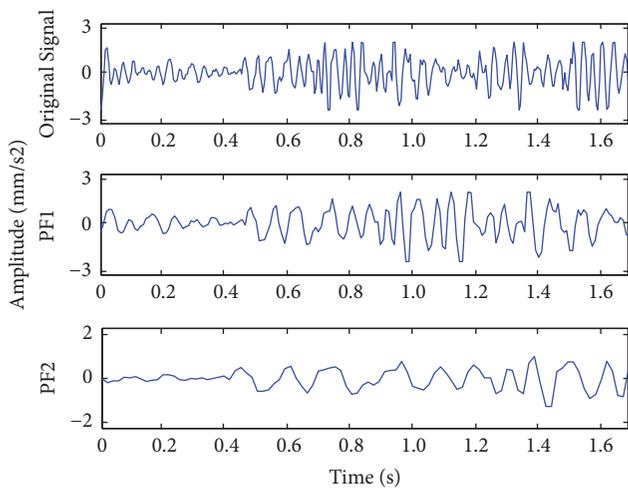


FIGURE 11: LMD results of broken tooth fault signal.

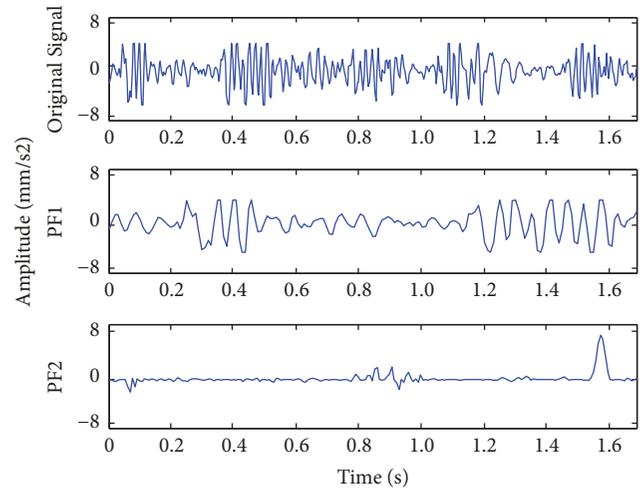


FIGURE 13: LMD results of outer circle fault signal.

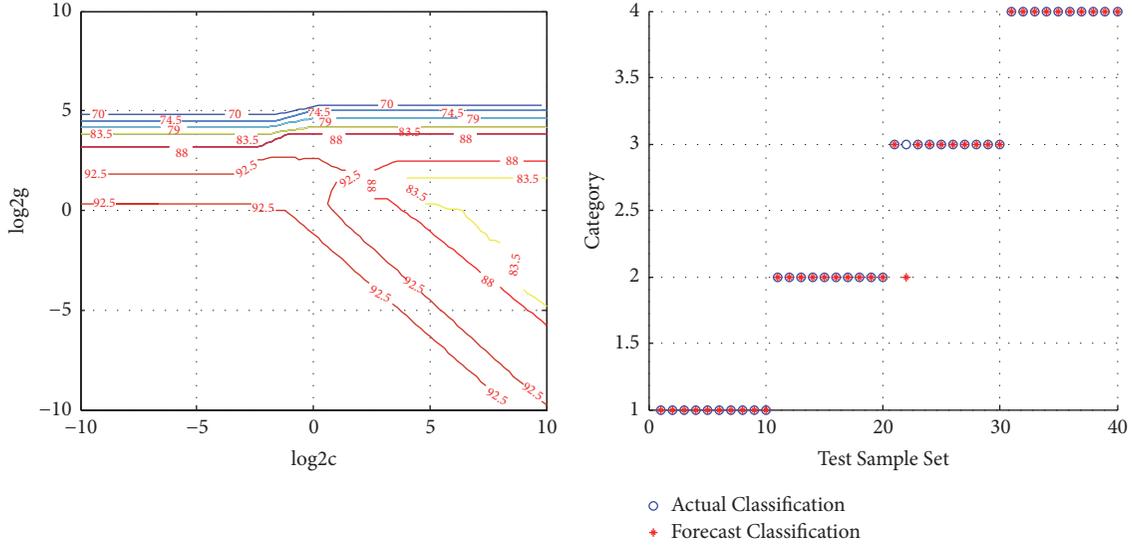


FIGURE 14: Classification training and test results based on grid method.

The normalized equation can be expressed as

$$f: y \rightarrow h = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \quad (22)$$

where $y, h \in R^n$, y_{\min} is the minimum value of y , and y_{\max} is the maximum value of y .

4.3. Multifault Classification Model Training. There are 4 kinds of fault states in this experiment, which belongs to multiclassification problem. The training of multifault classification model is shown as below:

Firstly, we label the different operation status as Table 1. The normal state is defined as 1, broken tooth fault is defined as 2, inner circle fault is defined as 3, and outer circle fault is defined as 4.

Secondly, 6 binary classifiers need to be constructed, such as (1v2), (1v3), (1v4), (2v3), (2v4), and (3v4). Corresponding vector is selected as the training set in the process of training each binary classifier, and we obtain 6 kinds of training models. Then we use corresponding test set to test the 6 kinds of results, respectively. The entire feature set contains 120 samples; each sample contains 8 feature characteristics. Lines 1–30 represent normal state (label 1); lines 31–60 represent broken tooth state (label 2); lines 61–90 represent the inner circle fault (label 3); and lines 91–120 represent the outer circle fault (label 4). We randomly select two-thirds of the samples in each label as the training samples, and the remaining as the test samples.

Finally, we apply the test samples to test the accuracy of classification model.

Optimal parameters need to be searched in the process of classification training. Common parameter optimization methods include grid method, genetic algorithm (GA) method, and PSO method. The process of searching optimal parameters via three methods and corresponding test results are shown in Figures 14, 15, and 16, respectively. The classification accuracy of each method is shown in Table 2. As

TABLE 2: Optimized parameters C and σ and corresponding classification accuracy.

Optimization method	Penalty factor C	Kernel function parameter σ	Classification accuracy Acc
Grid method	5	1.6	97.5%
GA	4.12	1.62	97.5%
PSO	4.12	1.62	97.5%

can be seen from the table, two model parameters achieved by these methods are very close and their classification accuracy is consistent. The advantage of grid method is that it can search multiple parameters at the same time. For independent parameter pairs, it is easy to search in parallel and it takes less time when searching for fewer optimization parameters. Besides, the approach can find global optimal solution when the optimization interval is large enough and the step distance is short enough [24]. GA method has good global search performance, but the search speed is slow and the solution efficiency is low. PSO method has strong local search performance and fast convergence speed, but it is prone to premature convergence and then fall into local optimal solution [25, 26]. Therefore, the SVM model parameters are selected via grid method in this paper.

4.4. Fault Feature Selection. The process for selecting optimal feature subset via SVM-RFE-MRMR is shown in Figure 17. ① We use the training set of overall feature set to train support vector classifier (SVC) and get corresponding classification model. ② The classification model is used to calculate the ranking score. The feature attribute with minimum ranking score is eliminated according to the ranking criteria. ③ SVM is used to retrain the remaining features to obtain a new ranking score until a feature ranked table is found. ④ The ranked table is used to train SVM by defining several

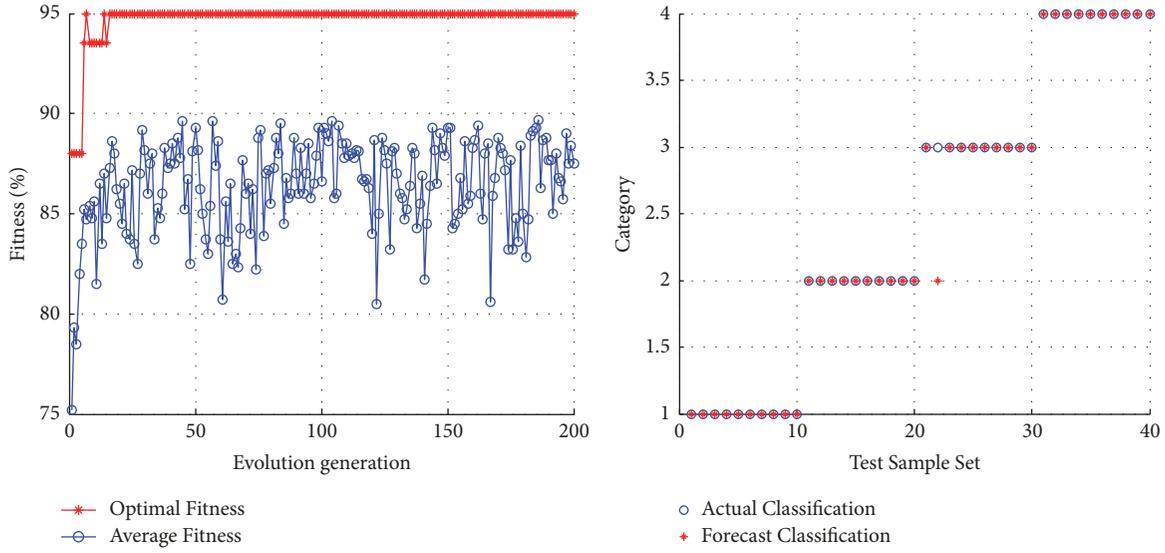


FIGURE 15: Classification training and test results based on GA.

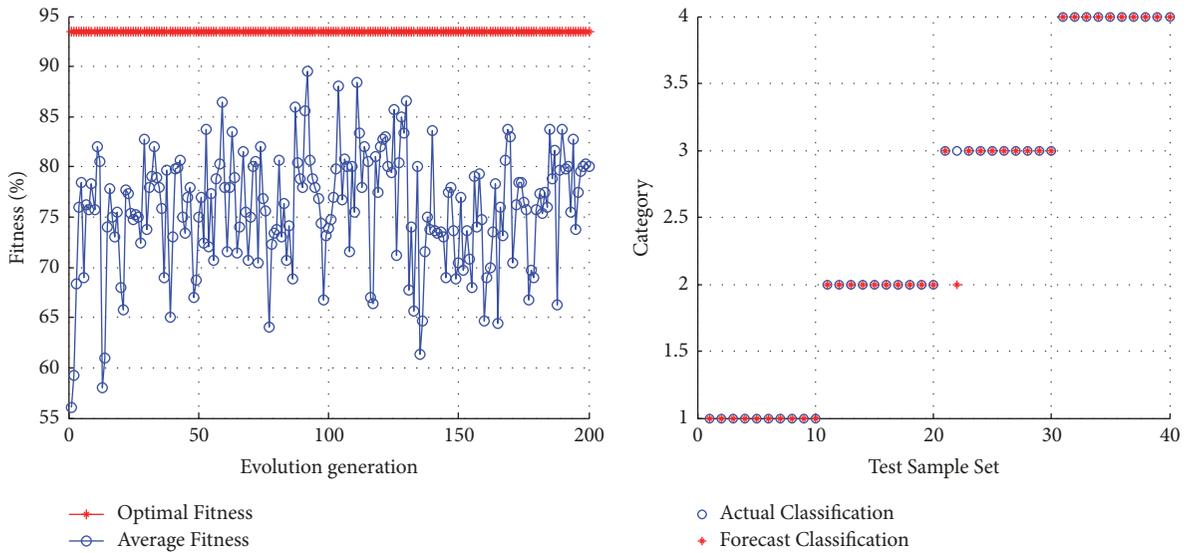


FIGURE 16: Classification training and test results based on PSO.

nested feature subsets $F_1 \subset F_2 \subset \dots \subset F_n$. Prediction accuracy of SVM is used to evaluate the performance of these subsets so as to obtain the optimal feature subset. ⑤ We use two criteria, training set leave-one-out cross validation error recognition rate (Loo Error Rate) and independent test set error recognition rate (Test Error Rate), to determine the optimal feature subset.

Following the steps outlined above, we use the training set designed in Section 4.3 to derive the SVC training and get optimized parameters $C_1 = 0.7579$ and $\sigma_1 = 1.3195$. These parameters are then applied to rank the influence degree of each feature; the result is shown in Table 3. At the same time, the nested feature subset is obtained as shown in Table 4. In this table, the relationship among each subset is $F_1 \subset F_2 \subset$

$\dots \subset F_8$. Next, calculate the classification accuracy on the basis of optimized parameters and nested feature subset in first step, and use the Loo Error Rate criterion to determine the optimal feature subset. The result is shown in Table 5. Finally, select the nested feature subset to train SVC and get optimized parameters $C_2 = 9.7656 \times 10^{-4}$, $\sigma_2 = 4$. The two parameters and the test set are used to calculate the classification accuracy and Test Error Rate as before, which can evaluate the performance of predictive model. It can be seen from Table 6 that the Test Error Rate of feature subset F_8 is the lowest.

Through contrasting and analyzing Tables 5 and 6, we can conclude the following information: (a) different combinations of features can achieve the same effect; (b) some

TABLE 3: SVM-RFE-MRMR feature ranked table.

Influence Degree	1	2	3	4	5	6	7	8
Feature	app_pf1	kpf1	HF1	app_pf2	rms1	HF2	kpf2	rms2
Characteristic Code	3	2	4	7	1	8	6	5

TABLE 4: Nested feature subset.

Characteristic number	1	2	3	4	5	6	7	8
Characteristic code	3	3,2	3,2,4	3,2,4,7	3,2,4,7,1	3,2,4,7,1,8	3,2,4,7,1,8,6	3,2,4,7,1,8,6,5
Subset symbol	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8

TABLE 5: Loo Error Rate.

Subset Symbol	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
Non Normalization Classification Accuracy	87.5%	72.5%	82.5%	82.5%	87.5%	82.5%	80%	77.5%
Normalization Classification Accuracy	87.5%	87.5%	92.5%	95%	97.5%	97.5%	97.5%	97.5%
Loo Error Rate	12.5%	12.5%	7.5%	7.5%	2.5%	2.5%	2.5%	2.5%

TABLE 6: Test Error Rate.

Subset Symbol	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
Normalization Classification Accuracy	25%	25%	25%	25%	25%	80%	90%	97.5%
Test Error Rate	75%	75%	75%	75%	25%	20%	10%	2.5%

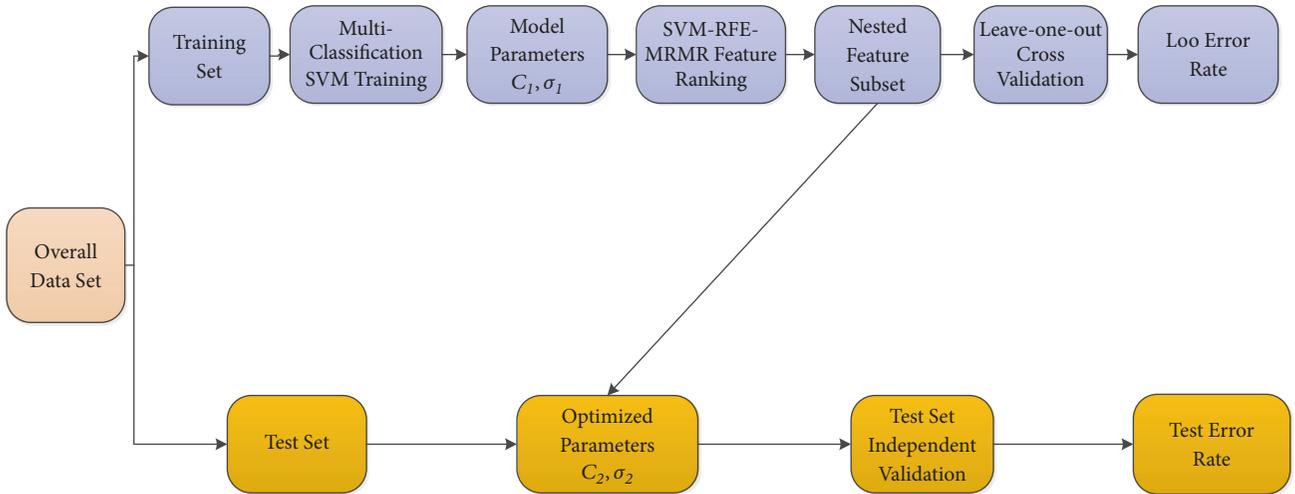


FIGURE 17: Diagram of searching the optimal feature subset.

fault features contain analogous information; (c) the optimal feature subset contains the least number of features.

5. Conclusions

This paper presents a fault diagnosis method based on the combination of improved LMD and support vector machine–recursive feature elimination–minimum redundancy maximum relevance (SVM-RFE-MRMR) for feature redundancy of reducer vibration signals. The approach

reduces feature dimensionality of original feature set through discarding those features that contribute less to classification or are insensitive to fault and preserving the best feature set made up of the optimal feature parameters. The results of the experiment verify that the proposed approach has good reliability and achieves high classification accuracy, which provides good reference value for condition monitoring of rotating machinery. In addition, this method can also be applied to multichannel fault signals processing and overcome the problem of high data bits and small samples.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

This work has been supported in part by the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions, University-Industry Cooperation Research Project in Jiangsu Province under Grant BY2016026-02, and the State Key Laboratory of Integrated Services Networks under Grant ISN10-10.

References

- [1] S. Riaz, H. Elahi, K. Javaid, and T. Shahzad, "Vibration Feature Extraction and Analysis for Fault Diagnosis of Rotating Machinery-A Literature Survey," *Asia Pacific Journal of Multidisciplinary Research*, vol. 5, no. 51, pp. 103–110, 2017.
- [2] X. Zhang, D. Jiang, T. Han, N. Wang, W. Yang, and Y. Yang, "Rotating Machinery Fault Diagnosis for Imbalanced Data Based on Fast Clustering Algorithm and Support Vector Machine," *Journal of Sensors*, vol. 2017, pp. 1–15, 2017.
- [3] Y. Yu and C. Junsheng, "A roller bearing fault diagnosis method based on EMD energy entropy and ANN," *Journal of Sound and Vibration*, vol. 294, no. 1-2, pp. 269–277, 2006.
- [4] L. Li, H. H. Liu, and H. S. Zhao, "Review of Vibration Monitoring and Fault Diagnosis Methods of Wind Turbines," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 33, pp. 94–100, 2017.
- [5] H. Jiang, C. Li, and H. Li, "An improved EEMD with multi-wavelet packet for rotating machinery multi-fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 36, no. 2, pp. 225–239, 2013.
- [6] Q. Guo, B.-Y. Liu, L.-B. Shi, and B. Li, "Experimental study and fault signals analysis of rotating machinery based on dual EEMD and Wigner-Ville distribution," *Zhendong yu Chongji*, vol. 31, no. 13, pp. 129–153, 2012.
- [7] H. Li, X. Lian, and S. Zhou, "Application on weak information classification by using wavelet scalogram synchronous averaging," *Jixie Gongcheng Xuebao/Journal of Mechanical Engineering*, vol. 49, no. 5, pp. 32–38, 2013.
- [8] H. Sun, Y. Zi, J. Yuan, Z. He, K. Li, and X. Chen, "Undecimated multiwavelet and Hilbert-Huang time-frequency analysis and its application in the incipient fault diagnosis of planetary gearboxes," *Journal of Mechanical Engineering*, vol. 49, no. 3, pp. 56–62, 2013.
- [9] J. S. Smith, "The local mean decomposition and its application to EEG perception data," *Journal of the Royal Society Interface*, vol. 2, no. 5, pp. 443–454, 2005.
- [10] W. Jiang, Z. Zheng, Y. Zhu, and Y. Li, "Demodulation for hydraulic pump fault signals based on local mean decomposition and improved adaptive multiscale morphology analysis," *Mechanical Systems and Signal Processing*, vol. 58, pp. 179–205, 2015.
- [11] J. S. Hu, S. X. Yang, and D. Q. Ren, "Spline-based local mean decomposition methods for vibration signal," *Journal of Data Acquisition & Processing*, vol. 24, no. 1, pp. 78–83, 2009.
- [12] K. Zhang, J.-S. Cheng, and Y. Yang, "Product function criterion in local mean decomposition method," *Zhendong yu Chongji*, vol. 30, no. 9, pp. 84–88, 2011.
- [13] 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.
- [14] S. Zhang, G. Sun, L. Li, X. Li, and X. Jian, "Study on mechanical fault diagnosis method based on LMD approximate entropy and fuzzy C-means clustering," *Yi Qi Yi Biao Xue Bao/Chinese Journal of Scientific Instrument*, vol. 34, no. 3, pp. 714–720, 2013.
- [15] T. Tang, L. Bo, X. Liu, B. Sun, and D. Wei, "Variable predictive model class discrimination using novel predictive models and adaptive feature selection for bearing fault identification," *Journal of Sound and Vibration*, vol. 425, pp. 137–148, 2018.
- [16] L. Lu, J. Yan, and C. W. De Silva, "Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition," *Journal of Sound and Vibration*, vol. 344, pp. 464–483, 2015.
- [17] K. B. Duan, J. C. Rajapakse, and M. N. Nguyen, "One-versus-one and one-versus-all multiclass SVM-RFE for gene selection in cancer classification," *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, pp. 47–56, 2007.
- [18] Y. Tang, Y.-Q. Zhang, and Z. Huang, "Development of two-stage SVM-RFE gene selection strategy for microarray expression data analysis," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 4, no. 3, pp. 365–381, 2007.
- [19] X. Li, S. Peng, J. Chen, B. Lü, H. Zhang, and M. Lai, "SVM-T-RFE: a novel gene selection algorithm for identifying metastasis-related genes in colorectal cancer using gene expression profiles," *Biochemical and Biophysical Research Communications*, vol. 419, no. 2, pp. 148–153, 2012.
- [20] M. H. Yao, N. Wang, M. Qi, and Y. Li, "Study on feature selection method of modified maximal relevance minimal redundancy," *Computer Engineering and Applications*, vol. 50, no. 9, pp. 116–122, 2014.
- [21] J. Cheng, "Local Characteristic-scale Decomposition Method and Its Application to Gear Fault Diagnosis," *Journal of Mechanical Engineering*, vol. 48, no. 09, p. 64, 2012.
- [22] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, 2002.
- [23] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185–205, 2005.
- [24] J. F. Wang, L. Zhang, G. X. Chen, and X. W. He, "A parameter optimization method for an SVM based on improved grid search algorithm," *Applied Science and Technology*, vol. 39, no. 3, pp. 28–31, 2012.
- [25] Y.-T. Kao and E. Zahara, "A hybrid genetic algorithm and particle swarm optimization for multimodal functions," *Applied Soft Computing*, vol. 8, no. 2, pp. 849–857, 2008.
- [26] X. Chen, J. Qiu, and G. Liu, "Optimal test selection based on hybrid BPSO and GA," *Chinese Journal of Scientific Instrument*, vol. 30, no. 8, pp. 1674–1680, 2009.



Hindawi

Submit your manuscripts at
www.hindawi.com

