

## Research Article

# Bearing Fault Diagnosis with Kernel Sparse Representation Classification Based on Adaptive Local Iterative Filtering-Enhanced Multiscale Entropy Features

Jinbao Zhang <sup>1</sup>, Yongqiang Zhao <sup>1</sup>, Xinglin Li,<sup>2</sup> and Ming Liu<sup>1</sup>

<sup>1</sup>School of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150001, China

<sup>2</sup>Hangzhou Bearing Test and Research Center, Hangzhou 310022, China

Correspondence should be addressed to Yongqiang Zhao; 13b308008@hit.edu.cn

Received 3 January 2019; Revised 31 March 2019; Accepted 30 April 2019; Published 4 June 2019

Academic Editor: Marcelo A. Savi

Copyright © 2019 Jinbao 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.

To improve the bearings diagnosis accuracy considering multiple fault types with small samples, a new approach that combined adaptive local iterative filtering (ALIF), multiscale entropy features, and kernel sparse representation classification (KSRC) is put forward in this paper. ALIF is used to adaptively decompose the nonlinear, nonstationary vibration signals into a sum of intrinsic mode functions (IMFs). Multiple entropy features such as sample entropy, fuzzy entropy, and permutation entropy with multiscale are computed from the first three IMFs and a total of one hundred and eighty features are obtained. After normalization, the features are employed to train and test the classifier KSRC, respectively. Finally, the proposed approach is evaluated with two experimental tests. One is concerned with different types of bearing faults from the centrifugal pump; and the other is from Case Western Reserve University (CWRU) considering 12 bearing fault states. Experimental results have proved that the proposed approach is efficient for bearing fault diagnosis, and high accuracy will be obtained with high dimensional features through small samples.

## 1. Introduction

The rolling bearings are mostly used in rotating machinery and their working conditions are concerned with maintenance of machines and safety of workers. Since the faults in bearings are always accompanied with the vibration which is easy to measure, many works are focused on the fault diagnosis based on the vibration analysis. Generally, the procedure for bearing fault diagnosis is composed of four steps: (1) preprocessing based on adaptive mode decomposition for nonlinear and nonstationary vibration signals, (2) extract features that are relatively insensitive to the data length and immune to the noise, (3) dimension reduction of the feature matrix based on principal component analysis (PCA) or Laplacian scores (LS), etc., and (4) fault pattern identification with the classifier. For example, Zhao et al. [1] computed multiscale permutation entropy of subbands by wavelet packet decomposition (WPD) and employed hidden Markov model (HMM) to identify the fault pattern of the rolling bearing. Yang et al. [2] extracted the energy entropy

from the intrinsic mode functions (IMFs) by empirical mode decomposition (EMD) [3] as features and employed artificial neural network (ANN) to identify the fault types. Li et al. [4] utilized local mean decomposition (LMD) [5] for preprocessing, improved multiscale fuzzy entropy as features, Laplacian scores for feature selection, and improved support vector machine based binary tree for bearing fault diagnosis. Yang et al. [6] combined variational mode decomposition (VMD) [7], local linear embedding (LLE) with support vector machine (SVM) to diagnose mechanical faults of the rotor-bearing-casing system. Good effects in bearing fault diagnosis have been realized to some extent from the above description; however, some problems still exist and need to be investigated further.

The first problem is concerned with adaptive mode decomposition [8]. WPD needs to prespecify the basis function and could not decompose the signals adaptively. The representative approach for adaptive decomposition is EMD, which could decompose a complicated signal into the sum of some certain IMFs, yet it is subjected to the problems of end

effect and mode mixing. Some modifications are proposed following EMD, such as LMD and VMD. Recently, a new approach called adaptive local iterative filtering (ALIF) was proposed by Cicone in 2016 [9, 10]. It follows the structure of EMD and has advantages of the adaptive filter adjusted with the Fokker-Planck (FP) equation and an adaptive filter length. Authors in [11] have successfully applied ALIF and approximate entropy for wind turbine bearing. Consequently, ALIF will be more suitable to process the faulty vibration signals of the rolling bearings.

The next problem is how to extract efficient features for classification with high accuracy [12]. The traditional features are from time domain, frequency domain, and time-frequency domain. To deal with the nonlinear dynamic characteristics of bearing fault signals, entropy is introduced [13], such as approximate entropy (ApEn) [11, 14], sample entropy (SaEn) [15, 16], fuzzy entropy (FuEn) [17, 18], and permutation entropy (PE) [19, 20]. However, they all estimate the complexity of signals at a single scale, which may be not conducive to the extraction of signal features. To overcome this drawback, multiscale entropy (MSE) was proposed by Costa et al. to measure the complexity of signals over a range of scales [21, 22]. Based on MSE, multiscale sample entropy (MSaE), multiscale fuzzy entropy (MFE), and multiscale permutation entropy (MPE) are proposed, which have been proved to have better performance compared with SaEn, FuEn, and PE in application of diagnosis on rolling bearing fault [23–25]. However, it is not ideal to use the entropy features directly for classification because of the influence of noise and interference harmonics in the vibration signals. Hence, ALIF is utilized to decompose the original signals into a sum of IMFs, which reduce the interference of noise and harmonics and highlight the effect of the fault information. In the following, the three multiscale entropy features are computed with the IMFs containing the most fault information. Considering the advantages of the three multiscale entropy features in feature extraction and characteristics of the following classifier, all of them are employed in this paper.

Before classification, feature selection like LS or dimensionality reduction like PCA and LLE should be performed. In the following, classifiers such as HMM [1], ANN [2], SVM [6], and multiclass relevance vector machine [20] are carried out for identification of the fault type. Though theories of them are well established, the inherent limitations have confined them to some extent. For example, ANN and VPMCD [26] need large training samples to obtain high classification accuracy; also, the SVM is a binary classifier which requires a classification strategy such as one against one, one versus all, and binary tree. Nevertheless, this is a two-stage combined feature reduction and classifiers. In addition, the training samples in the practical application are small, but with multiple features. Hence, a sparse representation classifier [27] is introduced to achieve the two stages at one time and realize feature selection through regularization. The classifier is firstly proposed to recognize human faces viewed in front considering cases of varying expression and illumination. Its advantage lies in the requirement of a sufficiently large number of features for high classification accuracy, but the number of samples. To modify the classification

accuracy with high dimensional features, kernel approach is introduced and KSRC is proposed and applied in face recognition [28, 29]. Hence, KSRC is employed in this paper to identify bearing fault states combined with ALIF-enhanced multiple entropy features.

The organization of this paper is as follows. Theoretical backgrounds including ALIFD MMPE, and KSRC are briefly introduced in Sections 2, 3, and 4. The illustration concerning the proposed method based on the theoretical backgrounds is presented in Section 5. Experiment datasets are employed to verify the proposed method in Section 6, and in Section 7 the conclusions are finally drawn.

## 2. Adaptive Local Iterative Filtering

Given a nonstationary, nonlinear signal  $x(t)$ , it could be reconstructed as the sum of several IMFs and the residue:

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

where  $c_i(t)$  represents the IMF,  $N$  is the number of IMFs and the IMF should satisfy two conditions [3]: (1) extrema in the whole data set must have the same number with zero crossings or differ by one at most; and (2) at any point, the mean value of the upper envelope connecting all the local maxima and the lower envelope connecting all the local minima is zero. Generally, the decomposition process consists of two loops: the inner loop and the outer loop, where the former is used for IMF extraction, while the latter is used to determine the number of IMFs and the residual. In EMD algorithm, the cubic spline interpolation is employed for the upper and the lower envelope functions, which will be susceptible to singularities. Consequently, iterative filtering computes the moving average  $\Theta(x(t))$  of the signal  $x(t)$  by the convolution

$$\Theta(x(t)) = x(t) * f(t) = \int_{-l}^l x(t + \tau) f(\tau) d\tau \quad (2)$$

in lieu of the envelop functions. In (2),  $*$  represents the convolution operator,  $f(t)$  constrained with  $\int_{-l}^l f(\tau) d\tau = 1$  is a low pass filter, and  $l$  is the mask length. Afterwards by the sifting process, the first IMF is generated:

$$c_1(t) = \lim_{n \rightarrow \infty} \Theta_{1,n}(x_n(t)), \quad (3)$$

where  $n$  is the iterative number,  $x_1(t) = x(t)$ , and  $x_n(t) = \Theta_{1,n-1}(x_{n-1}(t))$ . Since the number  $n$  is impossible to achieve infinite in (3), so (4) is adopted as a stop criterion for iterations:

$$\frac{\|\Theta_{i,n} - \Theta_{i,n-1}\|_2}{\|\Theta_{i,n-1}\|_2} \leq \xi, \quad (4)$$

where  $\Theta_{i,n}$  represents the moving average of the  $n$ -th iteration of the  $i$ -th IMF and  $\xi$  is a prespecified parameter. If  $\xi$  is large, rough decomposed results may be obtained. However, if  $\xi$  is too small, the computation will be expensive and noise will

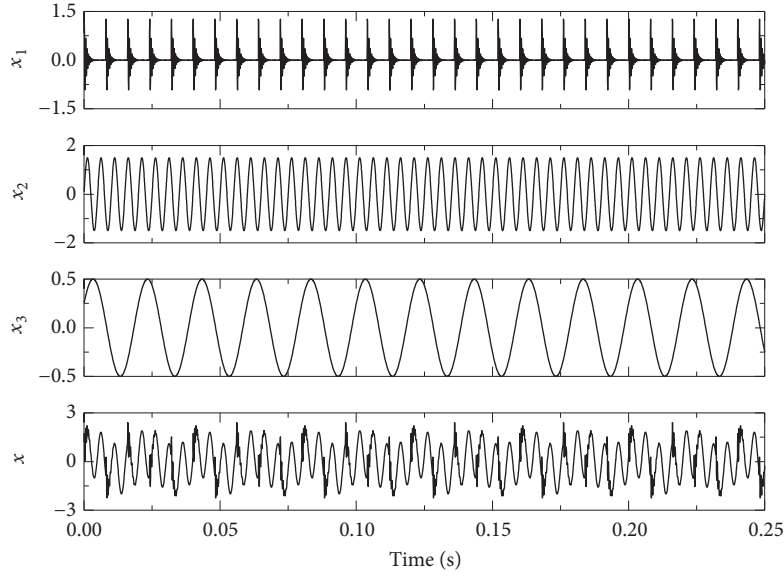


FIGURE 1: Time-domain waveforms of simulation signals.

be introduced. Finally, 0.001 is determined to  $\xi$  after trials. In the next step, the second IMF will be obtained by repetition of the previous iterative process to the residual signal  $r(t) = x(t) - c_1(t)$ . With the same manner, all the subsequent IMFs are produced by

$$c_k(t) = \lim_{n \rightarrow \infty} \Theta_{k,n}(x_n(t)), \quad (5)$$

Finally, if  $r(t)$  does not satisfy the two conditions of IMF, then treat it as the residual and stop the iteration.

The ALIF method is improved from the iterative filtering technique, which could adaptively adjust the filter with the FP equation and adaptively compute the filter length. Consequently, the above equation (2) can be rewritten as

$$\Theta(x(t)) = x(t) * f(t) = \int_{-l(t)}^{l(t)} x(t+\tau) f(t,\tau) d\tau, \quad (6)$$

which is subjected to

$$\int_{-l(t)}^{l(t)} f(t,\tau) d\tau = 1, \quad (7)$$

where  $f(t,\tau)$ ,  $\tau \in [-l(t), l(t)]$ , is the filter at time  $t$ , and  $l(t)$  is the mask length varying with  $t$ .

To show the advantage of ALIF in signal decomposition, a simulation is performed. Signals of the rolling bearing are written as

$$x(t) = x_1(t) + x_2(t) + x_3(t), \quad (8)$$

in which  $x_1(t) = 2e^{-1256t \times \text{mod}(t, T_0)} \sin(4000\pi t)$  is the periodic exponential decay signals,  $\text{mod}(\cdot, \cdot)$  is the surplus function, and  $T_0 = 1/20000$ ;  $x_2(t) = 1.5 \sin(400\pi t)$  and  $x_3(t) = 0.5 \sin(100\pi t + \pi/6)$  are harmonic interferences. Then the mixed signals with its components in the time interval  $t = 0 : 0.25$  are shown in Figure 1.

The decomposition results of ALIF and EMD are shown in Figures 2 and 3. In Figure 2, IMF1, IMF2, and IMF3 correspond to the components  $x_1$ ,  $x_2$ , and  $x_3$ . Moreover, the decomposed components from IMF4 to IMF8 are residuals. However, the IMF1 by EMD corresponding to the  $x_1$  is distorted in Figure 3, and IMF2 and IMF4 are corresponding to  $x_2$  and  $x_3$  because IMF3 is a false component. The absolute error is employed to compare the decomposition results of ALIF with EMD. From the comparison in Figure 4, ALIF outperforms EMD.

### 3. Multiscale Entropy Features

Following the preprocess of signals with ALIF, entropy features will be extracted from the IMFs for the preparation of fault diagnosis. Since a fault type with more features could be better represented, yet with the consideration of computation efficiency, hence sample entropy, fuzzy entropy, and permutation entropy are just introduced.

**3.1. Sample Entropy.** Considering a time series  $\{x(k), k = 1, 2, \dots, N\}$ , then the  $m$  dimensional vector at time  $i$  can be constructed as

$$\mathbf{x}_i^{(m)} = \{x(i), x(i+\tau), \dots, x(i+(m-1)\tau)\}, \quad (9)$$

$$i = 1, 2, \dots, N - (m-1)\tau,$$

where  $\tau$  is the time delay. The distance between  $\mathbf{x}_i^{(m)}$  and  $\mathbf{x}_j^{(m)}$  is defined as

$$d_{ij}^m = d[\mathbf{x}_i^{(m)}, \mathbf{x}_j^{(m)}]$$

$$= \max_{k=0,1,2,\dots,m-1} [|x(i+k) - x(j+k)|], \quad (10)$$

$$(i = 1, 2, \dots, N - m, j \neq i).$$

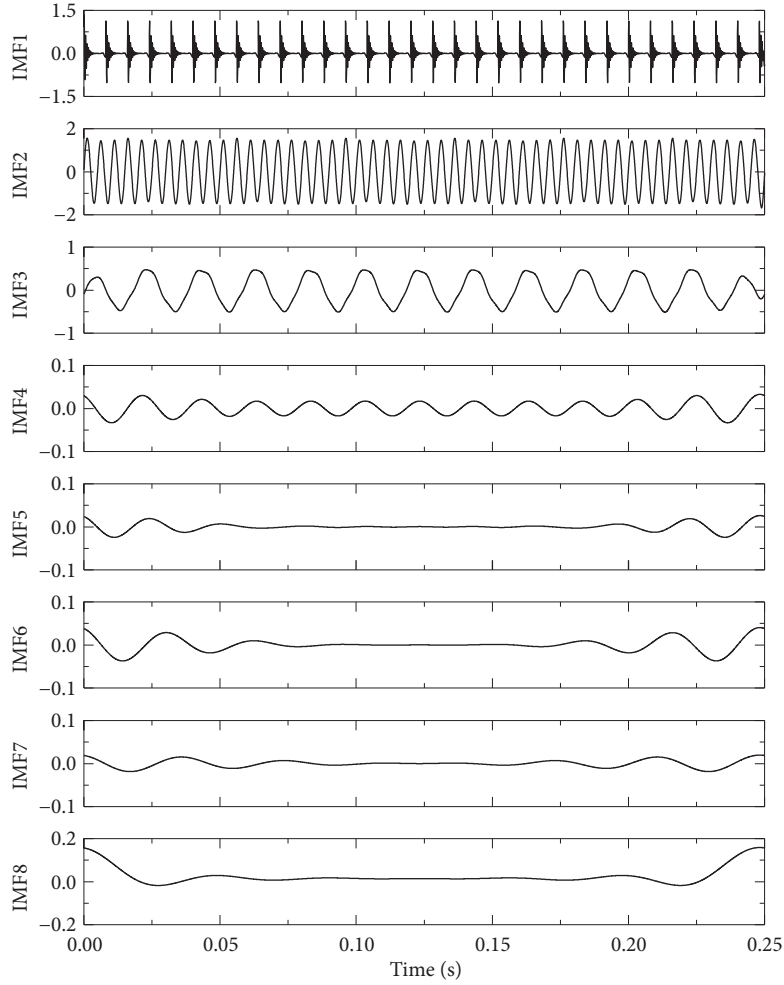


FIGURE 2: Decomposition results of ALIF.

Set the threshold  $r$ , and the ratio of distance less than  $r$  is defined as

$$B_i^{(m)}(r) = \frac{\#\{d[\mathbf{x}_i^{(m)}, \mathbf{x}_j^{(m)}] < r\}}{N - m - 1}, \quad (11)$$

with the mean

$$B^{(m)}(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} B_i^{(m)}(r). \quad (12)$$

Repeat the above steps for  $B_i^{(m+1)}(r)$  and obtain the mean  $B^{(m+1)}(r)$ ; then the sample entropy is

$$SaEn(m, r) = \lim_{N \rightarrow \infty} \left[ -\ln \frac{B^{(m+1)}(r)}{B^{(m)}(r)} \right]. \quad (13)$$

Since  $N$  is a finite value, (13) can be rewritten as

$$\begin{aligned} SaEn(m, r, N) &= \left[ -\ln \frac{B^{(m+1)}(r)}{B^{(m)}(r)} \right] \\ &= \ln B^{(m)}(r) - \ln B^{(m+1)}(r). \end{aligned} \quad (14)$$

3.2. *Fuzzy Entropy*. The distance in (10) is used to measure the fuzzy similarity as follows:

$$D_{ij}^{(m)} = \mu(d_{ij}^{(m)}, n, r) = \exp \left[ -\left( \frac{d_{ij}^{(m)}}{r} \right)^n \right]. \quad (15)$$

Define the function at  $m$ :

$$\phi^{(m)}(n, r) = \frac{1}{N - m} \sum_{i=1}^{N-m} \left[ \frac{1}{N - m - 1} \sum_{j=1, j \neq i}^{N-m} D_{ij}^{(m)} \right], \quad (16)$$

and the function at  $m + 1$ :

$$\begin{aligned} \phi^{(m+1)}(n, r) \\ &= \frac{1}{N - m} \sum_{i=1}^{N-m} \left[ \frac{1}{N - m - 1} \sum_{j=1, j \neq i}^{N-m} D_{ij}^{(m+1)} \right], \end{aligned} \quad (17)$$

and then the fuzzy entropy is

$$\begin{aligned} FuzzyEn(m, n, r) \\ &= \lim_{N \rightarrow \infty} \left[ \ln \phi^{(m)}(n, r) - \ln \phi^{(m+1)}(n, r) \right]. \end{aligned} \quad (18)$$

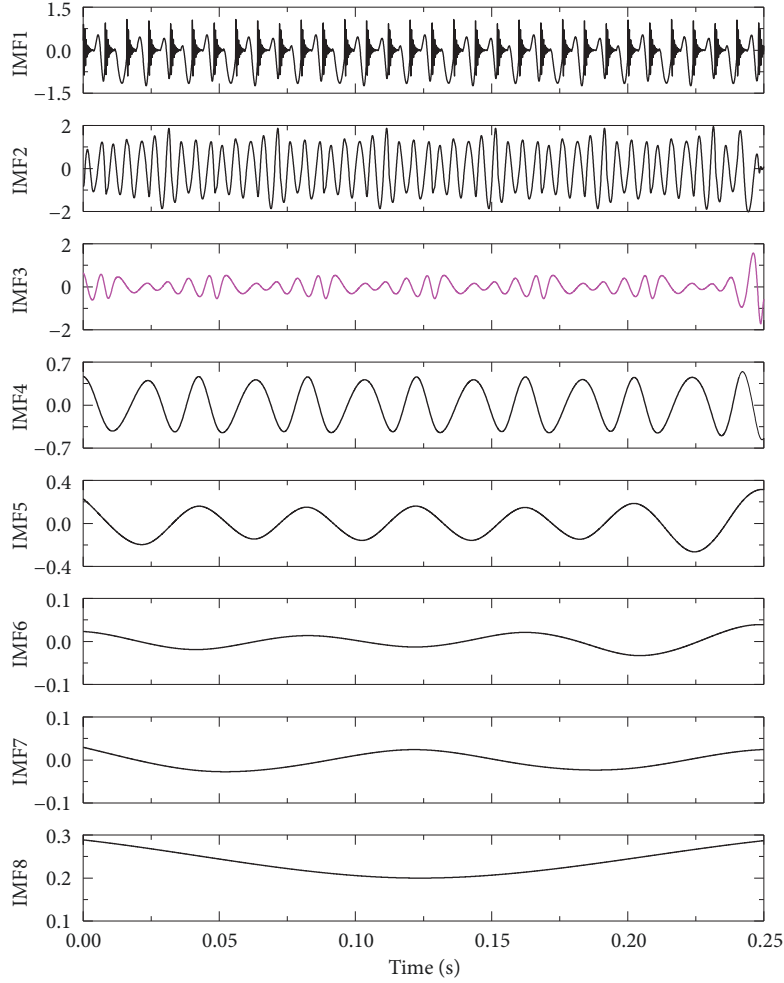
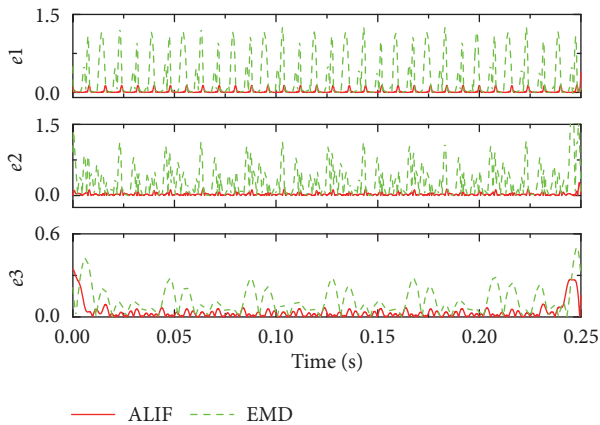


FIGURE 3: Decomposition results of EMD.


 FIGURE 4: The absolute errors of  $(x_1)$ - $(x_3)$  with their corresponding decomposed components by ALIF and EMD.

When  $N$  is a finite number, (18) is rewritten as

$$\begin{aligned} \text{FuzzyEn}(m, n, r, N) &= \ln \phi^{(m)}(n, r) \\ &\quad - \ln \phi^{(m+1)}(n, r). \end{aligned} \quad (19)$$

3.3. *Permutation Entropy.* Let  $\pi_j = (r_0, r_1, \dots, r_{m-1})$ ; then  $\mathbf{x}_i^{(m)}$  has a permutation  $\pi_j$  if it satisfies the fact that

$$x(i + r_0\tau) \leq x(i + r_1\tau) \leq \dots \leq x(i + r_{m-1}\tau), \quad (20)$$

where  $0 \leq r_n \leq m-1$ , and  $r_{n-1} < r_n$  when  $x(t + r_{n-1}\tau) = x(t + r_n\tau)$ .

For each permutation  $\pi_j$ ,  $1 \leq j \leq m!$ , the relative frequency can be defined as

$$p(\pi_j) = \frac{\#\{i \mid i \leq N - (m-1)\tau, \mathbf{x}_i^{(m)} \text{ has type } \pi_j\}}{N - (m-1)\tau}, \quad (21)$$

where  $\#$  represents the number of  $\mathbf{x}_i^{(m)}$  belonging to the type  $\pi_j$ . Then the definition of PE with  $m$  dimension can be written as

$$PE(m) = -\sum_{j=1}^{m!} p(\pi_j) \ln [p(\pi_j)]. \quad (22)$$

3.4. *Coarse Grained Process.* The multiple scales are realized through the coarse-grained process for better feature extraction. Further, the length of the coarse-grained time series

TABLE 1: Prespecified parameters in the entropy computation.

Entropy	$N$	$s$	$m$	$\tau$	$r$	$n$
Sample			2	1	$0.2 \times \text{SD}$	--
Fuzzy	2048	20	2	1	$0.2 \times \text{SD}$	2
Permutation			4	1	--	--

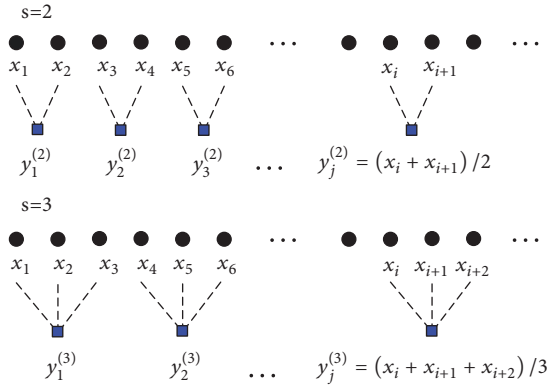


FIGURE 5: Illustration of the coarse-grained process.

depends on the length of the original time series divided by the corresponding scale factor, which is illustrated in Figure 5. Hence, the coarse grained time series  $y_j^{(s)}$  at a scale factor of  $s$  can be constructed according to

$$y_j^{(s)} = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} x(i), \quad 1 \leq j \leq \frac{N}{s}. \quad (23)$$

Then SaEn, FuEn, and PE of each coarse-grained time series are calculated based on (14), (19), and (22) and, respectively, plotted them as functions of the scale factor  $s$ :

$$MSaE(x, s, m, \tau, r) = SaEn(y_j^{(s)}, m, \tau, r) \quad (24)$$

$$MFE(x, s, m, \tau, r, n) = FuEn(y_j^{(s)}, m, \tau, r, n) \quad (25)$$

$$MPE(x, s, m, \tau) = PE(y_j^{(s)}, m, \tau) \quad (26)$$

In this paper, the prespecified parameters are set in Table 1. Especially, SD represents the standard deviation (std.) of the original signals.

## 4. Kernel Sparse Representation Classifier

**4.1. Sparse Representation Classification.** Let a matrix  $\mathbf{A}_i$  represent features of the  $i$ th class for auxiliary training samples, namely,  $\mathbf{A}_i = [\mathbf{v}_{i1}, \mathbf{v}_{i2}, \dots, \mathbf{v}_{in_i}] \in \mathbf{R}^{m \times n_i}$ , where  $m$  is the feature dimension, and  $n_i$  is the number of auxiliary training samples of the  $i$ th class. The auxiliary testing samples  $\mathbf{y}_i \in \mathbf{R}^m$  from the same class could be approximately expressed as

$$\mathbf{y}_i = \alpha_{i1} \mathbf{v}_{i1} + \alpha_{i2} \mathbf{v}_{i2} + \dots + \alpha_{in_i} \mathbf{v}_{in_i}. \quad (27)$$

Considering the overall  $q$  object classes and the whole training sets of the  $n$  auxiliary training samples with  $n = n_1 + n_2 + \dots + n_q$ , then the total matrix  $\mathbf{A}$  can be formed as

$$\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_q] = [\mathbf{v}_{11}, \mathbf{v}_{12}, \dots, \mathbf{v}_{qn_q}]. \quad (28)$$

Consequently, the linear representation of  $\mathbf{y}$  along with all auxiliary training samples is expressed as

$$\mathbf{y} = \mathbf{A}\boldsymbol{\alpha}_0, \quad (29)$$

where  $\boldsymbol{\alpha}_0 = [0, \dots, 0, \alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in_i}, 0, \dots, 0]^T \in \mathbf{R}^n$  is a coefficient vector, in which the entries are zero if they do not belong to the  $i$ th class.

The sparse solution to  $\mathbf{y} = \mathbf{A}\boldsymbol{\alpha}$  can be achieved by optimizing the following  $l^1$ -minimization problem:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_1 &= \arg \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1, \\ \text{subject to } & \mathbf{A}\boldsymbol{\alpha} = \mathbf{y}. \end{aligned} \quad (30)$$

When considering small noise, a noise term  $\mathbf{z} \in \mathbf{R}^m$  with  $\|\mathbf{z}\|_2 < \varepsilon$  is introduced to (29) and the formula can be modified as

$$\mathbf{y} = \mathbf{A}\boldsymbol{\alpha} + \mathbf{z}. \quad (31)$$

The flexible  $l^1$ -minimization problem for a sparse solution  $\boldsymbol{\alpha}$  is

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_1 &= \arg \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1, \\ \text{subject to } & \|\mathbf{A}\boldsymbol{\alpha} - \mathbf{y}\|_2 \leq \varepsilon. \end{aligned} \quad (32)$$

When a new sample is for testing, it could be expressed as  $\hat{\mathbf{y}}_i = \delta_i(\hat{\boldsymbol{\alpha}}_1)$ , where  $\delta_i(\hat{\boldsymbol{\alpha}}_1) \in \mathbf{R}^n$  is a vector in which parts of entries associated with class  $i$  are nonzero but the rest are zeros, and  $\delta_i(\cdot) : \mathbf{R}^n \rightarrow \mathbf{R}^n$  is a function that achieves coefficients selection related to the  $i$ -th class. Finally, the object class of the new testing sample could be identified with the residual between  $\mathbf{y}$  and  $\hat{\mathbf{y}}_i$ :

$$\min_i r_i(\mathbf{y}) = \|\mathbf{y} - \mathbf{A}\delta_i(\hat{\boldsymbol{\alpha}}_1)\|_2, \quad (33)$$

The algorithm for SRC is summarized as follows.

- (1) Input: a matrix of auxiliary training samples  $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_q] \in \mathbf{R}^{m \times n}$  for  $q$  classes, auxiliary testing samples  $\mathbf{y} \in \mathbf{R}^m$ , and an optional error tolerance  $\varepsilon > 0$ .
- (2) Normalize the columns of  $\mathbf{A}$  to have unit  $l^2$ -norm.

- (3) Solve the  $l^1$ -minimization problem:  $\hat{\alpha}_1 = \arg \min_{\alpha} \|\alpha\|_1$  subject to  $\|\mathbf{A}\alpha - \mathbf{y}\|_2 \leq \varepsilon$ .
- (4) Compute the residual  $r_i(\mathbf{y}) = \|\mathbf{y} - \mathbf{A}\delta_i(\hat{\alpha}_1)\|_2$ , for  $i = 1, \dots, q$ .
- (5) Output:  $\text{identity}(\mathbf{y}) = \arg \min_i r_i(\mathbf{y})$ .

**4.2. Kernel Sparse Representation Classification.** By means of the kernel trick, SRC is extended to KSRC for nonlinearity. Suppose a nonlinear mapping  $\phi: \mathbf{R}^m \rightarrow H$ ,  $\mathbf{v} \rightarrow \phi(\mathbf{v})$ , which realizes the transformation of auxiliary training samples from the original feature space  $\mathbf{R}^m$  into the kernel feature space  $H$ . Similar to SRC, the  $l_1$ -norm minimization problem of (32) can be reformulated as

$$\begin{aligned} \hat{\beta}_1 &= \arg \min_{\beta} \|\beta\|_1, \\ \text{subject to } &\|\Phi\beta - \phi(\mathbf{y})\|_2 \leq \varepsilon. \end{aligned} \quad (34)$$

Since  $\Phi$  and  $\phi(\mathbf{y})$  are unknown, (34) cannot be solved directly. But, according to Theorem 1, (34) can be transformed as

$$\hat{\beta}_1 = \arg \min_{\beta} \|\beta\|_1, \quad (35)$$

$$\text{subject to } \|\mathbf{K}\beta - \mathbf{Y}\|_2 \leq \delta,$$

where

$$\mathbf{K} = \Phi^T \Phi = \begin{bmatrix} k(\mathbf{v}_1, \mathbf{v}_1) & k(\mathbf{v}_2, \mathbf{v}_1) & \cdots & k(\mathbf{v}_n, \mathbf{v}_1) \\ k(\mathbf{v}_1, \mathbf{v}_2) & k(\mathbf{v}_2, \mathbf{v}_2) & \cdots & k(\mathbf{v}_n, \mathbf{v}_2) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{v}_1, \mathbf{v}_n) & k(\mathbf{v}_2, \mathbf{v}_n) & \cdots & k(\mathbf{v}_n, \mathbf{v}_n) \end{bmatrix}, \quad (36)$$

$$\mathbf{Y} = \Phi^T \phi(\mathbf{y}) = [k(\mathbf{y}, \mathbf{v}_1) \ k(\mathbf{y}, \mathbf{v}_2) \ \cdots \ k(\mathbf{y}, \mathbf{v}_n)]^T, \quad (37)$$

and a kernel function is defined as  $k(\mathbf{v}_i, \mathbf{v}_j) = \langle \phi(\mathbf{v}_i), \phi(\mathbf{v}_j) \rangle = \phi(\mathbf{v}_i)^T \phi(\mathbf{v}_j)$ . In this paper, the linear kernel  $k(\mathbf{v}_i, \mathbf{v}_j) = \mathbf{v}_i^T \mathbf{v}_j$  is used. Solution of (35) is achieved with the software package of disciplined convex programming [35]. Finally, the new testing sample can be assigned to one prespecified class by minimizing the residual between  $\Phi^T \phi(\mathbf{y})$  and  $\Phi^T \Phi \hat{\beta}_1^i$ , ( $i = 1, 2, \dots, q$ ):

$$\min_i R_i(\mathbf{y}) = \|\Phi^T \phi(\mathbf{y}) - \Phi^T \Phi \hat{\beta}_1^i\|_2. \quad (38)$$

**Theorem 1.** For any  $\varepsilon \geq 0$ , there must exist  $\delta \geq 0$  such that we have  $\|\Phi\beta - \phi(\mathbf{y})\|_2 \leq \varepsilon$ , as long as  $\|\Phi^T \Phi \beta - \Phi^T \phi(\mathbf{y})\|_2 \leq \delta$  is satisfied.

The algorithm for KSRC is summarized as follows.

- (1) Input: a matrix of auxiliary training samples  $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_q] \in \mathbf{R}^{m \times n}$  for  $q$  classes, auxiliary testing samples  $\mathbf{y} \in \mathbf{R}^m$ , and an optional error tolerance  $\varepsilon > 0$ .
- (2) Normalize the columns of  $\mathbf{A}$  to have unit  $l^2$ -norm.
- (3) Calculate  $\mathbf{K}$  and  $\mathbf{Y}$  by (36) and (37).

- (4) Solve the  $l^1$ -minimization problem in (35).
- (5) Compute the residuals  $R_i(\mathbf{y})$  defined in (38).
- (6) Output:  $\text{identity}(\mathbf{y}) = \arg \min_i R_i(\mathbf{y})$ .

## 5. Illustration of the Proposed Method

Since the proposed method could simultaneously perform feature selection and multiclass classification, the corresponding procedure based on ALIF enhanced multiscale entropy features and KSRC is set up, and the steps are as follows.

- (1) Collect vibration signals of bearings with healthy and different defective types, in addition to different defect sizes for each defective type.
- (2) Decompose the vibration signals into a sum of IMFs with ALIF. The first three IMFs containing prominent fault information are selected to extract multiscale entropy features and they are used to construct feature vectors after normalization with

$$\Gamma_{ij} = \frac{(\rho_{ij} - (\rho_{i \cdot})_{\min})}{((\rho_{i \cdot})_{\max} - (\rho_{i \cdot})_{\min})}, \quad (39)$$

where  $\rho$  represents all features in one sample,  $i$  denotes the sequence of samples, and  $j$  is the sequence of features in each sample.

- (3) Set the number of training samples and testing samples. The training samples are randomly selected for KSRC. It is noted that the number of training samples includes the number of auxiliary training sample  $n$  and the number of auxiliary testing samples.
- (4) After successful training, KSRC is used to test samples and identify the fault patterns with different severity levels.

The illustration of the proposed approach is shown in Figure 6.

## 6. Experimental Verification

To validate the capability of the proposed approach, two cases concerning bearing faults are investigated. One is about the bearing in the centrifugal pump considering different fault types [16]. The other is about rolling bearings in the test rig from CWRU with different fault categories and severity levels [36].

**6.1. Bearing Fault of the Centrifugal Pump.** The centrifugal pump test system is shown in Figure 7, and the experimental details can be found in [16]. Five commonly occurring faults in the centrifugal pump were set, including normal, bearing roller wearing (BRW), bearing inner race wearing (BIRW), bearing outer race wearing (BORW), and centrifugal pump impeller wearing (PIW). Vibration signals at the five fault states are shown in Figure 8, and the corresponding first five IMFs by ALIF are shown in Figure 9. From the comparison,

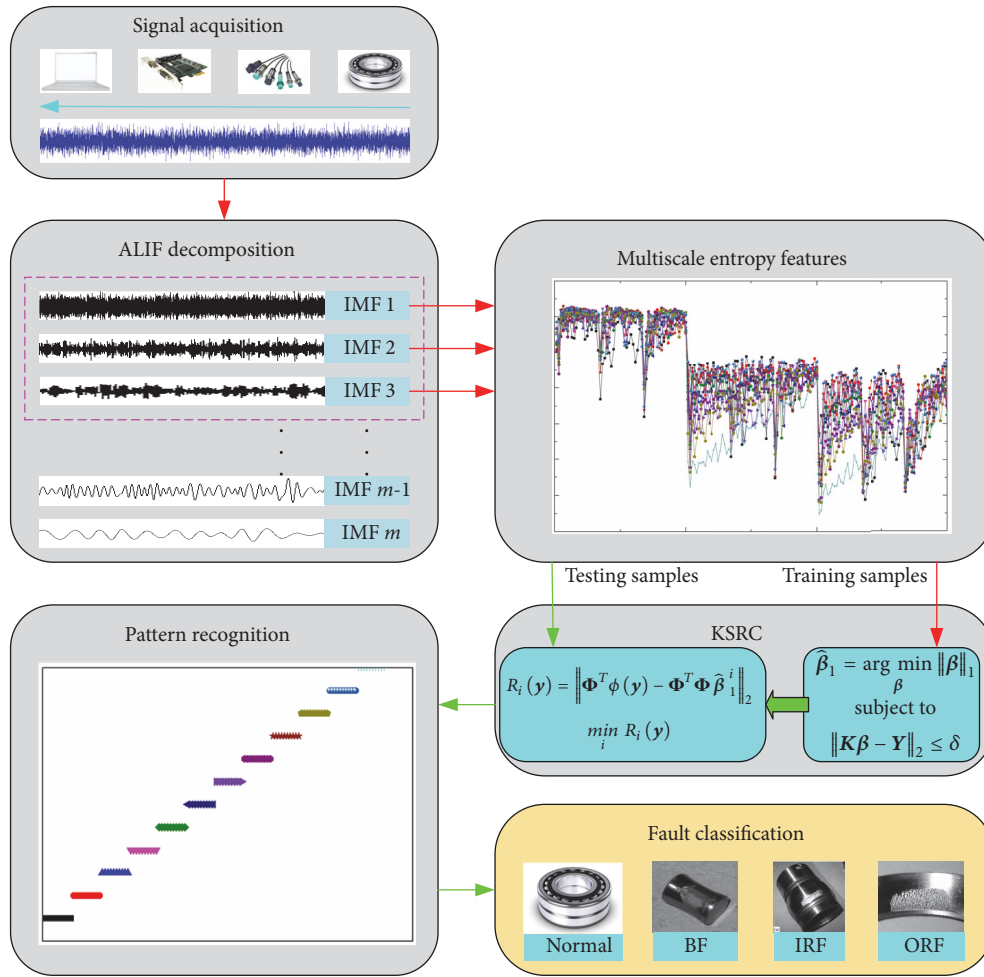


FIGURE 6: Illustration of bearing fault classification.

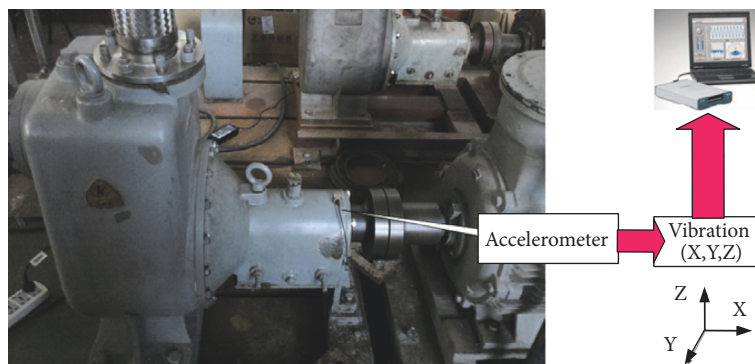


FIGURE 7: Centrifugal pump data acquisition system.

it is shown that the first three IMFs contain the most part of energy. Hence, multiscale entropy features including MSaE, MFE, and MPE over 20 scales of the first three components by ALIF are extracted according to the parameters prespecified in Table 1, and a total of 180 entropy features are obtained corresponding to one sample.

Labels of fault types are specified for classification, and description of bearing fault states can be found in Table 2. Each fault type has 50 samples, and the average of each entropy feature of fifty samples in each fault type is shown in Figure 10. In each test, 10 random samples are selected to train the SRC classifier, and the rest 40 samples are utilized



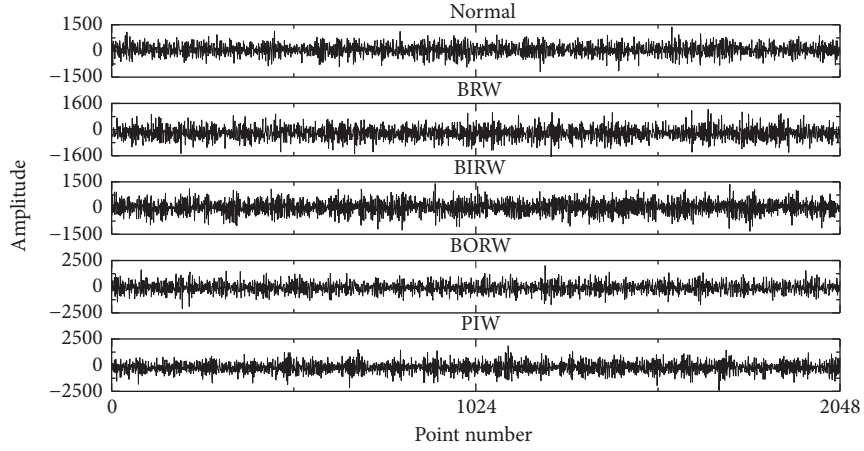


FIGURE 8: Bearing vibration signals with different fault types.

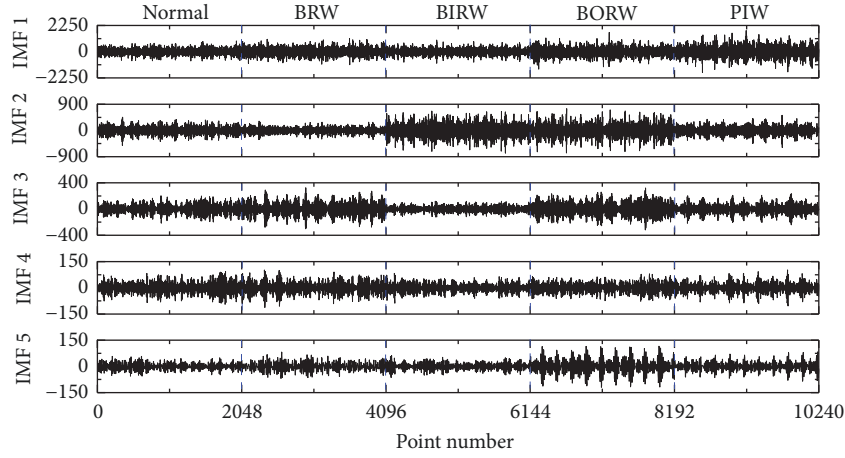


FIGURE 9: The first five IMFs by ALIF decomposition.

TABLE 2: Description of the experimental data.

Fault type	Training samples	Testing samples	Classification labels
Normal	10	40	1
BRW	10	40	2
BIRW	10	40	3
BORW	10	40	4
PIW	10	40	5
Number of samples	$T_{tr} = 50$	$T_{te} = 200$	

for test. The accuracy formula of testing samples is defined as

$$A_{te} = \frac{R_{te}}{T_{te}}, \quad (40)$$

and the accuracy formula of training samples is

$$A_{tr} = \frac{R_{tr}}{T_{tr}}, \quad (41)$$

where  $R_{te}$  is the number of right classified testing samples;  $T_{te}$  is the number of testing samples;  $R_{tr}$  is the number of

TABLE 3: Result of bearing fault diagnostic accuracy.

		Train	Test
Accuracy (%)	Mean	100	96.95
	Std.	0	0.98

right classified training samples;  $T_{tr}$  is the number of training samples. As listed in Table 3, mean of the testing classification accuracy by (40) with repetitions of ten times is 96.95% with std. 0.98%, and mean of the training classification accuracy

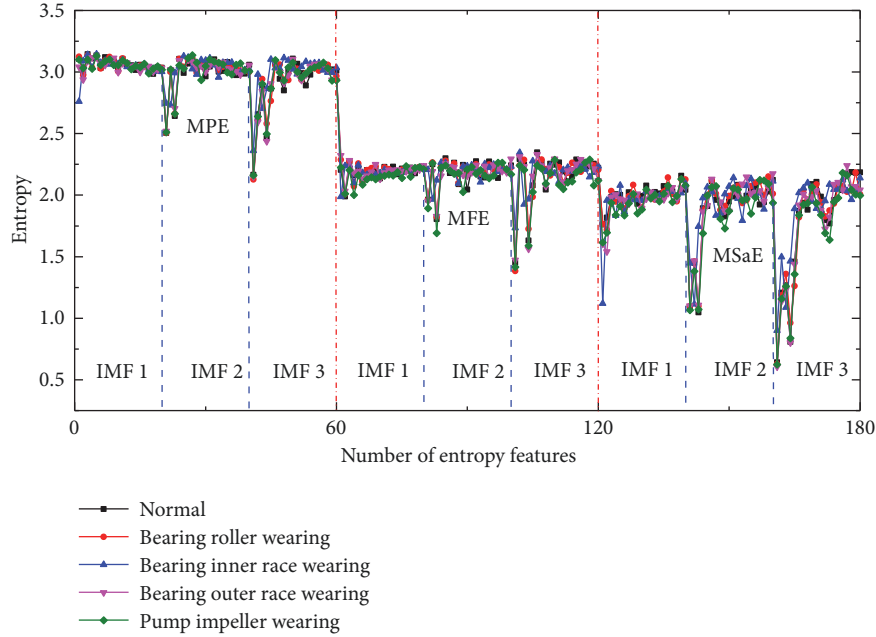


FIGURE 10: Multiscale entropy features over 20 scales of the first three IMFs by ALIF with the average of fifty samples.

TABLE 4: Diagnostic accuracy of ten times.

Times	Classification labels					Total	Test Accuracy (%)
	1	2	3	4	5		
1	37	37	40	40	40	194	97
2	38	33	40	40	40	191	95.5
3	38	38	40	40	40	196	98
4	38	36	40	39	39	192	96
5	37	39	40	39	40	195	97.5
6	39	38	40	40	39	196	98
7	38	36	40	40	40	194	97
8	38	38	40	40	40	196	98
9	38	34	40	40	39	191	95.5
10	39	37	40	40	38	194	97

TABLE 5: A comparative study of different features on the effect of diagnostic accuracy.

Features		MSaE	MFE	MPE	MSaE+MFE	MSaE+MPE	MFE+MPE	MSaE+MFE+MPE	
Accuracy (%)	Train	Mean	100	100	100	100	100	100	
		Std.	0	0	0	0	0	0	
	Test	Mean	79.7	88.45	92.65	92.90	94.95	96.35	96.95
		Std.	3.55	3.88	1.93	2.12	1.62	1.55	0.98

by (41) is 100% with std. 0. In Table 4, accuracies of the ten tests are listed, and the maximum accuracy could reach 98%. The corresponding classification result of the proposed method at accuracy 98% is shown in Figure 11. Compared with [16], their mean accuracy varies from 94.58% to 97.08% according to ratio of the std. of the added noise in ensemble empirical mode decomposition (EEMD); moreover, the ratio of training samples in [16] is 40%, yet it is 20% in our paper. To show the advantage of high dimensional features in KSRC for accuracy improvement of bearing fault diagnosis,

a comparison is performed as listed in Table 5. The sequence of the effect in accuracy from small to large, respectively, is MSaE, MFE, and MPE, and in pairs.

**6.2. Artificially Seeded Damage Bearing.** The bearing data are obtained from Bearing Data Centre of CWRU, and the bearing test system is shown in Figure 12. The drive end bearing 6205-2RS JEM SKF is investigated, which is seeded with single point faults using electrodischarge machining. There are four states, including norm, ball fault (BF), inner

TABLE 6: Description of the experimental data.

Fault type	Defect size (inch)	Number of training samples			Number of testing samples		Classification labels
		0 HP	0HP	2HP	0HP	2HP	
Normal	0	10	40	50			1
	0.007	10	40	50			2
	0.014	10	40	50			3
	0.021	10	40	50			4
	0.028	10	40	50			5
IRF	0.007	10	40	50			6
	0.014	10	40	50			7
	0.021	10	40	50			8
	0.028	10	40	50			9
ORF	0.007	10	40	50			10
	0.014	10	40	50			11
	0.021	10	40	50			12
Number of samples		$T_{tr} = 120$	$T_{te} = 480$	$T_{te} = 600$			

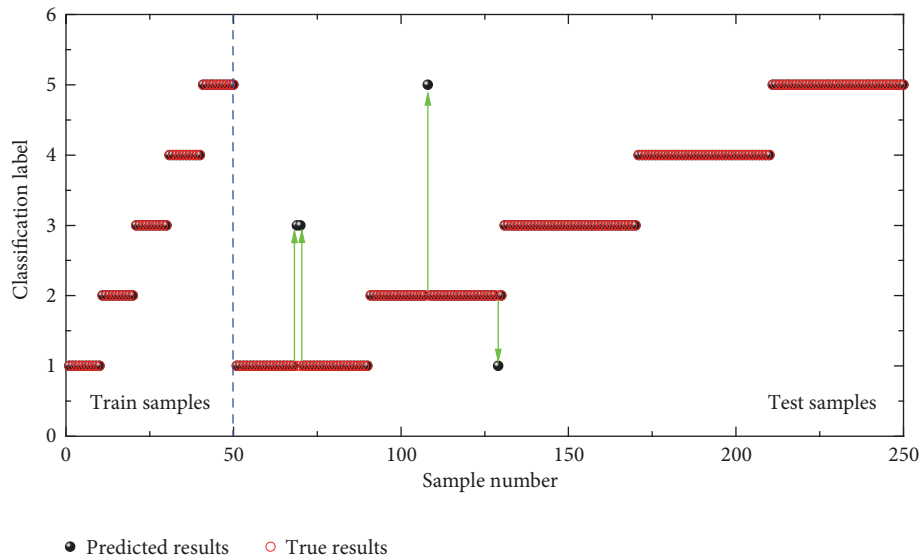


FIGURE 11: Classification result of the proposed method at accuracy 98%.

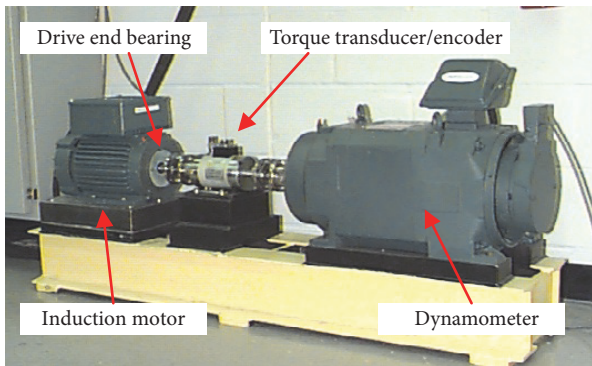


FIGURE 12: Bearing experiment system.

race fault (IRF), and outer race fault (ORF) (at the 6 o'clock position). Vibration signals are collected from the

accelerometers placed at the drive end of the motor housing with the sampling frequency 12 kHz. The defective bearing at 0 HP is investigated in this case. The sampling time is 10s in each state, and the overall length of the collected vibration signals is divided into nonoverlapping segments. Vibration signals of normal, BF, IRF, and ORF at defect size 0.007 inch and load 0 HP are shown in Figure 13, as well as the first five IMFs by ALIF in Figure 14.

Considering the defect size, labels of fault types corresponding to 12 bearing fault states are specified for classification, and description of bearing fault states can be found in Table 6. Each state has 50 samples, and MSaE, MFE, and MPE over 20 scales of the first three components by ALIF are averaged from the fifty samples as shown in Figure 15. It is shown that the distances among different types of the three entropy features are not distinct; hence a combination is considered. To prove the accuracy of the proposed approach, ten repetitive tests are performed with randomly selected

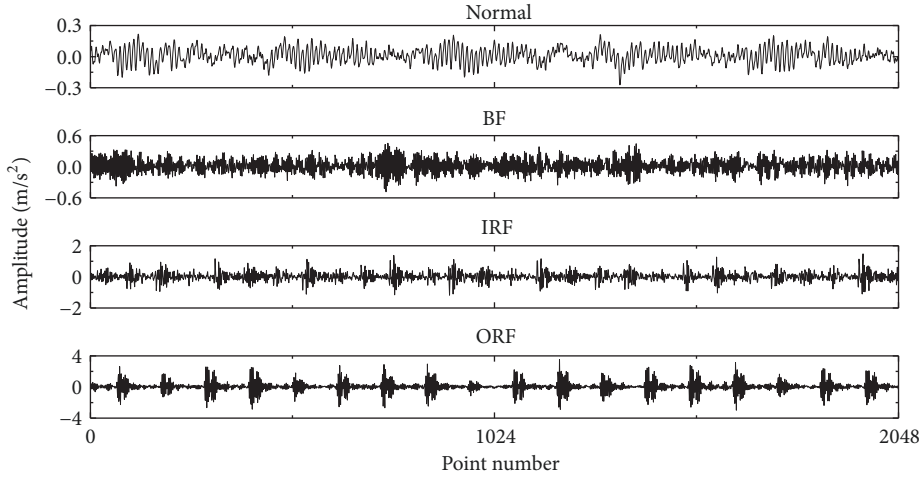


FIGURE 13: Bearing vibration signals with different fault types at defect size 0.007 inch and load 0 HP.

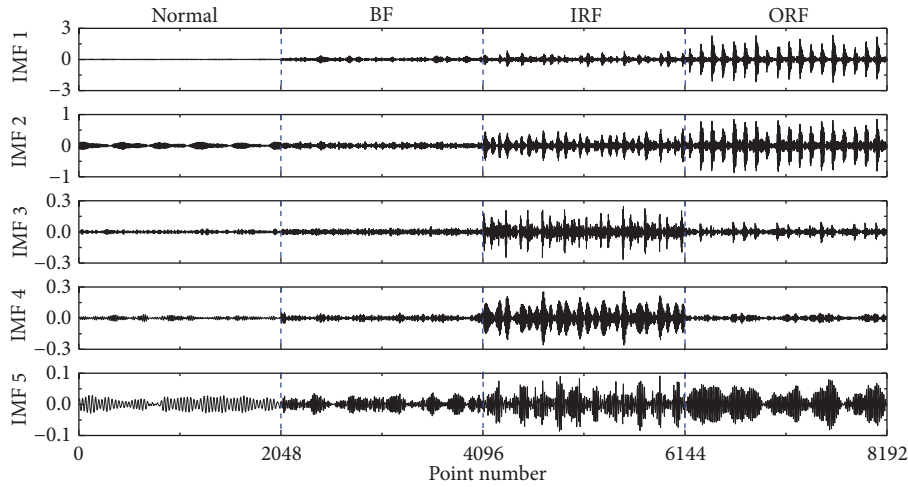


FIGURE 14: The first five IMFs by ALIF decomposition with defect size 0.007 inch and load 0 HP.

TABLE 7: Result of bearing fault diagnostic accuracy.

		Train (0 HP)	Test (0 HP)	Test (2 HP)
Accuracy (%)	Mean	100	98.48	89.73
	Std.	0	0.7	2.02

samples. As listed in Table 7, mean of the testing classification accuracy by (40) with repetitions of ten times is 98.48% with std. 0.7%, and mean of the training classification accuracy by (41) is 100% with std. 0. Furthermore, in each time, 10 random samples are selected to train the SRC classifier, and the rest 40 samples are utilized for test. In Table 8, accuracies of the ten tests with mean 98.48% and std. 0.7% are listed, and the maximum accuracy could reach 99.38% at 0 HP. For better illustration, the corresponding classification result of the proposed method at accuracy 99.38% with 0 HP is shown in Figure 16.

In addition, another condition at 2 HP is considered to test the flexibility of the proposed approach with different

loads. Ten random samples of load 0 HP are used to train KSRC and all samples of load 2 HP are used to test. Ten tests are performed as above and the results are listed in Table 9. The mean diagnostic accuracy at 2 HP is 89.73% with std. 2.02%, and the maximum accuracy is 92.83% with the corresponding illustration in Figure 17. The high accuracy rate of diagnosis demonstrates the usefulness of the proposed approach under different loads.

Since the multiscale entropy features in Figure 15 are not easy to be distinguished from others due to the multiple faults, a comparison is performed considering different combinations of features in Table 10. Ten tests with different random testing samples as above are carried out as well. The mean and the std. corresponding to different features are listed in Table 10. The results have verified the advantage of the combination of MSaE, MFE, and MPE.

Besides, a list of literatures using the CWRU bearing data is collected in Table 11, and they are arranged according to the classified states. Based on the comparisons in Table 11 and our work, respectively, it is shown that the

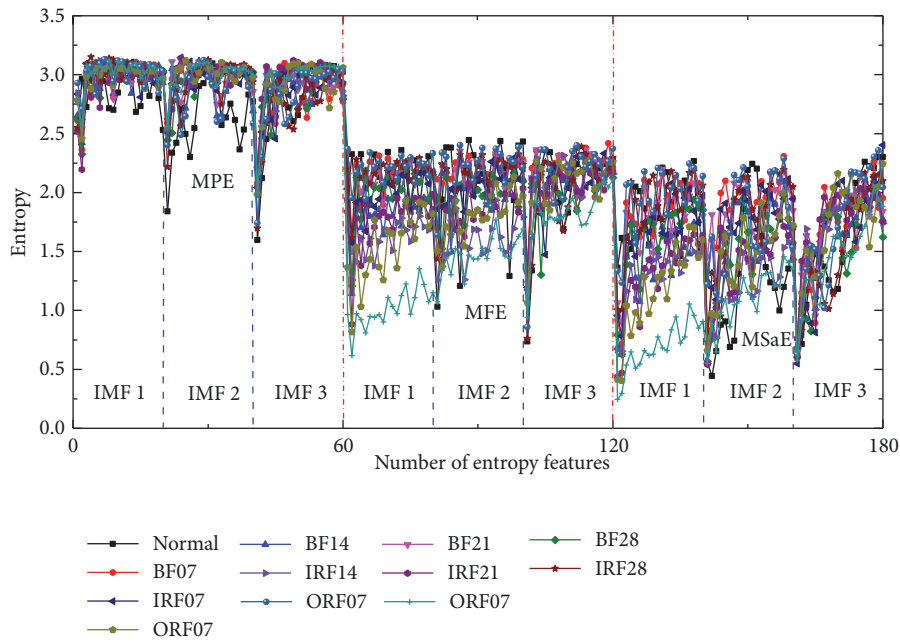


FIGURE 15: Multiscale entropy features over 20 scales of the first three IMFs by ALIF with the average of fifty samples.

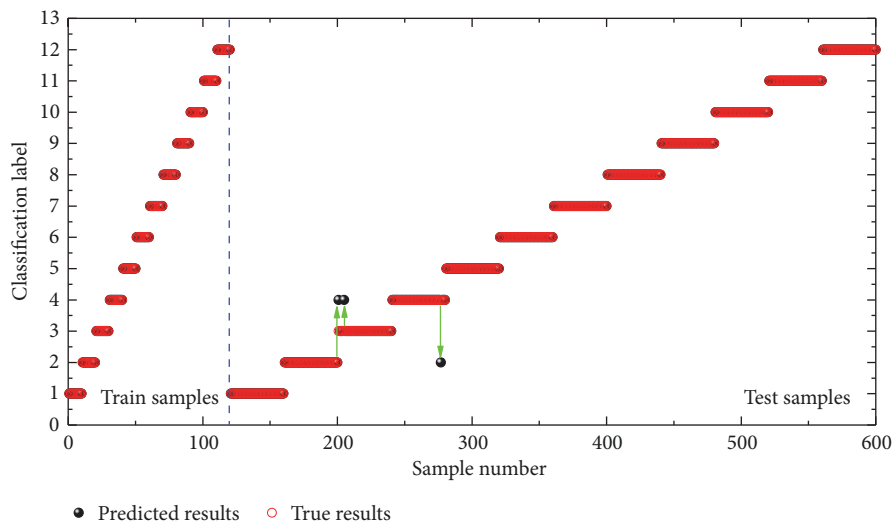


FIGURE 16: Classification result of the proposed method at classification accuracy 99.38% and load 0 HP.

TABLE 8: Diagnostic accuracy of ten times at 0 HP.

Times	Classification labels												Total	Test accuracy (%)
	1	2	3	4	5	6	7	8	9	10	11	12		
1	40	38	39	30	37	40	40	40	40	40	40	40	464	96.67
2	40	40	37	39	37	40	40	40	40	40	40	40	473	98.54
3	40	40	39	37	37	40	40	40	40	40	40	40	473	98.54
4	40	38	38	39	39	40	40	40	40	40	40	40	474	98.75
5	40	39	39	39	40	40	40	40	40	40	40	40	477	99.38
6	40	40	40	38	37	40	40	40	40	40	40	40	475	98.96
7	40	40	40	35	38	40	40	40	40	40	40	40	473	98.54
8	40	39	38	39	37	40	40	40	40	40	40	40	473	98.54
9	40	40	35	39	40	40	40	40	40	40	40	40	474	98.75
10	40	39	37	39	37	40	40	40	40	40	40	40	472	98.33

TABLE 9: Diagnostic accuracy of ten times at 2 HP.

Times	1	2	3	4	5	6	7	8	9	10
Accuracy (%)	85.83	91.17	90.17	88.83	92.83	91.33	90.17	90.17	89.50	87.33

TABLE 10: A comparative study of different features on the effect of diagnostic accuracy.

Features		MSaE	MFE	MPE	MSaE+MFE	MSaE+MPE	MFE+MPE	MSaE+MFE+MPE
Accuracy (%)	Train	Mean	100	100	100	100	100	100
	(0 HP)	Std.	0	0	0	0	0	0
	Test	Mean	95.69	96.27	97.48	97.13	97.9	98.15
	(0 HP)	Std.	1.09	1.23	0.99	0.81	1.32	0.74
	Test	Mean	74.83	78.22	77.42	82.67	87.18	88.33
	(2 HP)	Std.	4.99	2.64	4.24	4.25	5.30	2.75

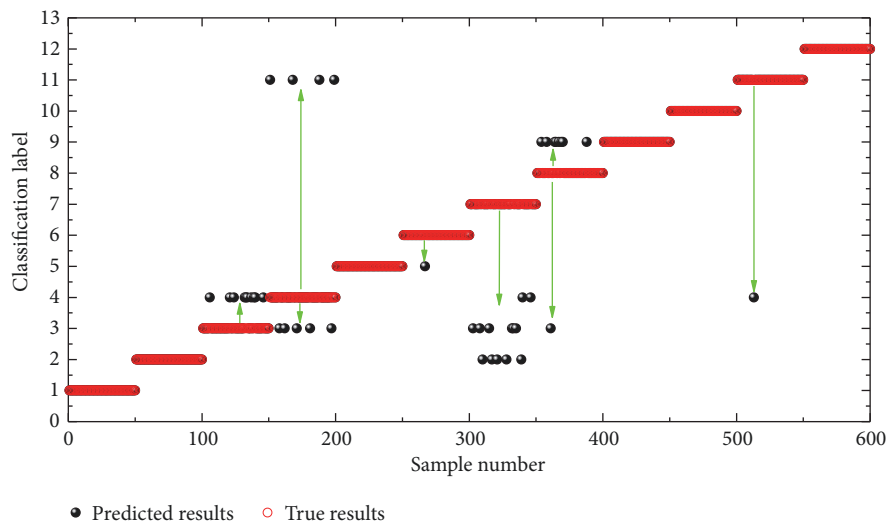


FIGURE 17: Classification result of the proposed method at classification accuracy 92.83% and load 2 HP.

entropy-based features, especially MPE, are mostly employed for bearing fault diagnosis and good results could be obtained. When considering more classified classes, combinations of multiscale entropy features will be a solution. Though the result of the proposed approach could not reach 100% like in [34] with 12 classified states as well, the proposed approach avoids the problem of feature selection and parameter optimization of SVM. Compared with the remaining researches, the proposed approach could deal with more classified states with high accuracy.

## 7. Conclusion

To improve accuracy of the bearing fault diagnosis considering multiple fault states with small samples, a novel bearing fault diagnosis method based on ALIF-enhanced multiscale entropy features and KSRC is proposed in this paper. Adaptive local iterative filtering could decompose the nonlinear and nonstationary vibration signals adaptively into a sum of IMFs with different scales. MSaE, MFE, and MPE values of the first three IMFs by ALIF are computed and normalized. Further, KSRC could accurately identify

multiple faulty types of roller bearings with the normalized entropy features and realized features selection through regularization. Eventually, the proposed method is evaluated with experimental data concerning bearing faults in the centrifugal pump and multiple bearing faults from CWRU. The comparison shows that high dimensional features through small samples could achieve high accuracy of bearing fault diagnosis at 0 HP as well as varying working condition 2 HP. The results demonstrate that the proposed method is feasible and effective in bearing fault diagnosis.

## Data Availability

The bearing data used to support the findings of this study have been, respectively, deposited in the following: 6.1 Bearing fault of the centrifugal pump [https://www.researchgate.net/profile/Chen\\_Lu15/publications](https://www.researchgate.net/profile/Chen_Lu15/publications) 6.2 Artificially Seeded Damage Bearing <http://csegroups.case.edu/bearingdatacenter/pages/apparatus-procedures>. In addition, the references concerning the data have been clearly cited in the paper.

TABLE 11: Previous work concerning bearing fault diagnosis published in the literatures.

Preprocessing	Feature extraction	Feature selection	Classifier	Classified states	Maximum accuracy (%)	Ref.
WPD	MPE	N/A	HMM	4	94.20	[1]
EEMD	PE	N/A	M-RVM	4	99.58	[20]
N/A	MPE	N/A	SVM	4	100.00	[23]
LMD	MPE	LS	SVM-BT	4	100.00	[30]
N/A	MSE, MPE, MSRMS, and MBSE	MD	SVM	4	99.79	[31]
LCD	FuEn	N/A	ANFIS	7	100.00	[18]
LMD	IMFE	LS	SVM-BT	8	100.00	[4]
N/A	MFE	LS	VPMCD	8	100.00	[24]
N/A	RQA	ARSFS	AOR-VPMCD	10	90.00	[26]
DTCWPT	IMPE	LLTSA	ELM	10	100.00	[32]
N/A	Entropy, holder coefficient, and improved fractal box-counting dimension	N/A	GRPR	11	98.41	[33]
EMD	Multiple domains	SM-LFDA	SVM	12	100.00	[34]

Note: local characteristic-scale decomposition (LCD); support vector machine-binary tree (SVM-BT); multiclass relevance vector machines (M-RVM); recurrence Quantification Analysis (RQA); Mahalanobis distance (MD); dual tree complex wavelet packet transform (DTCWPT); improved multiscale permutation entropy (IMPE); linear local tangent space alignment (LLTSA); extreme learning machine (ELM); support margin-local fisher discriminant analysis (SM-LFDA); grey relation pattern recognition (GRPR); multiscale root-mean-square(MSRMS); multiband spectrum entropy (MBSE).

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 51505100). The authors would like to thank the editors and the anonymous reviewers for their valuable suggestions which have greatly improved the paper.

## References

- [1] L.-Y. Zhao, L. Wang, and R.-Q. Yan, "Rolling bearing fault diagnosis based on wavelet packet decomposition and multiscale permutation entropy," *Entropy*, vol. 17, no. 9, pp. 6447–6461, 2015.
- [2] Y. Yu, Y. Dejie, 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.
- [3] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [4] Y. B. Li, M. Q. Xu, R. X. Wang, and W. H. Huang, "A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy," *Journal of Sound and Vibration*, vol. 360, Article ID 12638, pp. 277–299, 2016.
- [5] 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.
- [6] Y. Yang and D. Jiang, "Casing vibration fault diagnosis based on variational mode decomposition, local linear embedding, and support vector machine," *Shock and Vibration*, vol. 2017, Article ID 5963239, 14 pages, 2017.
- [7] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [8] Z. Feng, D. Zhang, and M. J. Zuo, "Adaptive mode decomposition methods and their applications in signal analysis for machinery fault diagnosis: a review with examples," *IEEE Access*, vol. 5, pp. 24301–24331, 2017.
- [9] A. Cicone, J. Liu, and H. Zhou, "Adaptive local iterative filtering for signal decomposition and instantaneous frequency analysis," *Applied and Computational Harmonic Analysis*, vol. 41, no. 2, pp. 384–411, 2016.
- [10] M. Piersanti, M. Materassi, A. Cicone, L. Spogli, H. Zhou, and R. G. Ezquer, "Adaptive local iterative filtering: a promising technique for the analysis of non-stationary signals," *Journal of Geophysical Research: Space Physics*, vol. 123, no. 1, pp. 1031–1046, 2018.
- [11] X. An and L. Pan, "Wind turbine bearing fault diagnosis based on adaptive local iterative filtering and approximate entropy," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 231, no. 17, pp. 3228–3237, 2017.
- [12] W. Caesarendra and T. Tjahjowidodo, "A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slewing bearing," *Machines*, vol. 5, no. 4, article 21, 2017.
- [13] A. Humeau-Heurtier, "The multiscale entropy algorithm and its variants: a review," *Entropy*, vol. 17, no. 5, pp. 3110–3123, 2015.
- [14] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 6, pp. 2297–2301, 1991.
- [15] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.
- [16] Y. Wang, C. Lu, H. Liu, and Y. Wang, "Fault diagnosis for centrifugal pumps based on complementary ensemble empirical mode decomposition, sample entropy and random forest" in *Proceedings of the 12th World Congress on Intelligent Control and Automation, WCICA '16*, pp. 1317–1320, China, 2016.
- [17] W. Chen, J. Zhuang, W. Yu, and Z. Wang, "Measuring complexity using FuzzyEn, ApEn, and SampEn," *Medical Engineering & Physics*, vol. 31, no. 1, pp. 61–68, 2009.
- [18] J. Zheng, J. Cheng, and Y. Yang, "A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy," *Mechanism and Machine Theory*, vol. 70, pp. 441–453, 2013.
- [19] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, Article ID 174102, 2002.
- [20] X. Liu, Y. Chen, and J. Yang, "A novel fault diagnosis method for rolling bearing based on EEMD-PE and multiclass relevance vector machine," in *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference, I2MTC '17*, 2017.
- [21] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, Article ID 068102, 2002.
- [22] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. Guo, "Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference," *Expert Systems with Applications*, vol. 37, no. 8, pp. 6077–6085, 2010.
- [23] Y. Lv, R. Yuan, T. Wang, H. Li, and G. Song, "Health degradation monitoring and early fault diagnosis of a rolling bearing based on CEEMDAN and improved MMSE," *Materials*, vol. 11, no. 6, article 1009, 2018.
- [24] J. Zheng, J. Cheng, Y. Yang, and S. Luo, "A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination," *Mechanism and Machine Theory*, vol. 78, pp. 187–200, 2014.
- [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.
- [26] 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.
- [27] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [28] J. Yin, Z. Liu, Z. Jin, and W. Yang, "Kernel sparse representation based classification," *Neurocomputing*, vol. 77, no. 1, pp. 120–128, 2012.



- [29] B. Gaonkar and C. Davatzikos, "Analytic estimation of statistical significance maps for support vector machine based multivariate image analysis and classification," *NeuroImage*, vol. 78, pp. 270–283, 2013.
- [30] Y. B. Li, M. Q. Xu, Y. Wei, and W. H. Huang, "A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree," *Measurement*, vol. 77, pp. 80–94, 2016.
- [31] S. D. Wu, C. W. Wu, T. Y. Wu, and C. C. Wang, "Multi-scale analysis based ball bearing defect diagnostics using mahalanobis distance and support vector machine," *Entropy*, vol. 15, no. 2, pp. 416–433, 2013.
- [32] G. Tang, X. Wang, and Y. He, "A novel method of fault diagnosis for rolling bearing based on dual tree complex wavelet packet transform and improved multiscale permutation entropy," *Mathematical Problems in Engineering*, vol. 2016, Article ID 5432648, 13 pages, 2016.
- [33] C. Yang, H. Wang, Z. Gao, and X. Cui, "Improving rolling bearing online fault diagnostic performance based on multi-dimensional characteristics," *Royal Society Open Science*, vol. 5, no. 5, article 180066, 2018.
- [34] X. Yu, F. Dong, E. Ding, S. Wu, and C. Fan, "Rolling bearing fault diagnosis using modified LFDA and EMD with sensitive feature selection," *IEEE Access*, vol. 6, pp. 3715–3730, 2018.
- [35] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," 2013, <http://cvxr.com/cvx>.
- [36] Bearing Data Center, *Case Western Reserve University*, 2006, <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.

