

Received July 31, 2019, accepted August 15, 2019, date of publication August 27, 2019, date of current version October 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937828

# A Multi-Stage Hybrid Fault Diagnosis Approach for Rolling Element Bearing Under Various Working Conditions

XIAOAN YAN<sup>1,2</sup>, (Member, IEEE), YING LIU<sup>1</sup>, MINPING JIA<sup>2</sup>, AND YINLONG ZHU<sup>1</sup>

<sup>1</sup>School of Mechatronics Engineering, Nanjing Forestry University, Nanjing 210037, China

<sup>2</sup>School of Mechanical Engineering, Southeast University, Nanjing 211189, China

Corresponding author: Xiaoan Yan (yanxiaoan89@sina.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51675098, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX17\_0059, in part by the Key Research and Development Project of Jiangsu Province under Grant BE2019030637, and in part by the Agricultural Science and Technology Independent Innovation Project of Jiangsu Province under Grant 2018325-SCX (18)2049.

**ABSTRACT** To timely detect bearing operating condition, and accurately identify bearing fault type and fault severity, a novel multi-stage hybrid fault diagnosis strategy for rolling bearing is proposed in this paper, which mainly consists of three stages (i.e. fault initial detection, fault type recognition and fault severity assessment). Firstly, the procedure of permutation entropy (PE)-based fault initial detection is performed to estimate bearing operating condition. If the bearing fault exists, the next two stages are conducted for fault type recognition and fault severity assessment. Specifically, in the second and third stages, for each dataset under different fault conditions, hybrid-domain features including time-domain, frequency-domain and time-frequency domain are firstly extracted to establish high-dimensional feature space based on statistical analysis and variational mode decomposition (VMD). Then, locality preserving projection (LPP) is introduced to compress high-dimensional dataset into low-dimensional feature space which can reflect preferably intrinsic information of the raw signal and remove information redundancy embedded in hybrid-domain features. Finally, the obtained low-dimensional dataset is imported into Fuzzy C-means (FCM) clustering for recognizing bearing fault type and fault severity. The efficacy of the proposed approach is verified by experimental bearing data under different working conditions. The results indicate that our proposed method can both assess effectively bearing health status and recognize accurately bearing fault type and fault severity. In addition, our proposed approach has higher diagnosis precision than traditional single-stage diagnosis method mentioned in this paper.


**INDEX TERMS** Permutation entropy, variational mode decomposition, locality preserving projection, rolling bearing, fault diagnosis.

## I. INTRODUCTION

Research on fault detection of rolling element bearing has drawn much attention in recent years. Rolling element bearings are the major parts of rotating machinery and widely used in many industrial fields. Tiny faults existing in rolling element bearing easily cause the stop running of mechanical system, bring the tremendous economic losses and even give rise to the serious accident and personnel casualties [1]–[5]. Consequently, it is of much concern to explore a novel and

effective fault detection approach for preventing the local fault and reducing downtime of mechanical system [6].

Currently, many algorithms have been reported for detecting bearing defects. For instance, Zhang *et al.* [7] presented a mixed fault diagnosis scheme, where three techniques (i.e. permutation entropy, ensemble empirical mode decomposition and SVM optimized by inter-cluster distance) are combined to identify fault type of bearing. Yan and Jia [8] proposed an improved multiscale dispersion entropy to obtain bearing fault feature information, and then the mRMR method is used for feature selection and finally ELM is adopted for intelligent fault diagnosis of rotating machinery. Zheng *et al.* [9] proposed the multi-scale fuzzy entropy to

The associate editor coordinating the review of this manuscript and approving it for publication was Youqing Wang .

measure the complexity of a time series, then combine Laplacian score (LS) and variable predictive model-based class discrimination to identify bearing fault categories. Li *et al.* [10] combined the local mean decomposition and improved multiscale fuzzy entropy to extract fault features, and then LS is adopted to refine fault features by sorting the scale factor. Finally, improved support vector machine based binary tree is applied to implement fault classification. Ali *et al.* [11] combined empirical mode decomposition and energy entropy to extract fault feature information, then the extracted features is regarded as the input of artificial neural network to identify bearing fault status. However, the aforementioned fault diagnosis approaches concentrated mainly on multiscale feature extraction and single-stage fault diagnosis. That is, their diagnosis procedure is carried out directly over the trained fault identification model. Unfortunately, when we analyze a complex fault diagnosis problem containing multiple different fault conditions, the single-stage method will increase the complexity of model training and reduce diagnostic accuracy and efficiency [12]. Hence, to simplify fault diagnosis problem and improve diagnostic efficiency, research on a new diagnosis scheme which can effectively estimate bearing health status and accurately identify bearing fault type and fault severity is the subject of this paper.

Apparently, most of fault diagnosis approaches mainly include four steps (i.e. signal processing, feature extraction, feature dimension reduction and pattern recognition). The first three steps are prepared for pattern recognition and are directly related to the effect of pattern recognition [13]. Concretely, signal processing and feature extraction are taken as a whole to reveal the intrinsic feature of the original signal. Then, feature dimension reduction is performed for removing information redundancy of the original eigenspace. At present, to improve diagnostic accuracy and efficiency, many works have launched on feature extraction and feature dimension reduction [14]–[16]. In these studies, the frequently-used and representative feature extraction methods are time-domain and frequency-domain analysis. However, due to the various effects of non-linear factors including load, friction and impact, raw vibration signal measured from bearing are often characterized by nonlinear and non-stationary behaviors, the time-domain or frequency-domain features are extracted independently which is inadequate for reflecting fully the inherent properties of the original signal [17]–[19]. Thus, time-frequency analysis method is introduced to extract multiscale feature information [20]. Nowadays, successful results have been achieved by the application of many time-frequency analysis techniques (e.g. Wigner-Ville distribution (WVD), wavelet transforms (WT), empirical mode decomposition (EMD), local mean decomposition (LMD) and intrinsic time-scale decomposition (ITD)) in fault detection [21]. However, these approaches involve some inherent drawbacks for multiscale feature extraction [22]. Such as, WVD has inevitably cross term interference in processing multi-component signal. WT can separate a vibration signal into a set of wavelet

details, but it does not possess self-adapted ability due to the predefined wavelet basis. EMD have some weaknesses (e.g. end effect, mode mixing, overshoot and undershoot). As a perfection of EMD, LMD also suffers from the end effect and scale-mixing issue [23]. ITD is quite appropriate for analysis of non-stationary signals, but it is liable to result in a waveform burr and curve distortion since the utilization of linear transformation. Variational mode decomposition (VMD) is a novel time-frequency analysis algorithm [24], which can decompose adaptively a multi-component signal into several sub-signals termed intrinsic mode function (IMF) and has the advantage of strong noise-robustness. Some applications of VMD for fault detection have reported in [25]–[30]. Unfortunately, capability of VMD rests heavily with its preset mode number. The improper setting of mode number may induce the over or under decomposition issues, which affect the precision of signal decomposition. Hence, VMD is introduced to extract multiscale features and central frequency iteration rule in [31] is employed to determine automatically its mode number. Given the above, hybrid-domain features containing time-domain, frequency-domain and time-frequency domain are extracted and combined to construct high dimensional feature space in this paper.

After obtaining hybrid-domain feature, to further improve the diagnostic accuracy and efficiency, suitable dimension reduction methods are required to remove the redundant and negative features from high dimensional feature space. At present, dimension reduction methods are divided into two types (i.e. the linear and nonlinear). Representative linear dimension reduction methods [32] mainly include principal component analysis (PCA), local feature analysis (LFA) and linear discriminant analysis (LDA). Although these methods are very suitable for dataset with linear structure, they can do nothing for solving nonlinear and complex problems [33]–[35]. On this basis, some nonlinear dimension reduction methods nicknamed manifold learning stand out for overcoming the weakness of linear dimension reduction methods. Familiar manifold learning involves locality preserving projection (LPP), isometric feature mapping (ISOMAP), local tangent space alignment (LTSA), laplacian eigenmap (LE), and locally linear embedding (LLE) etc. High-dimensional eigenvector can be projected into a lower dimensional feature space containing the inherent nature of original high-dimensional space by using these manifold learning methods. Currently, many achievements of manifold learning have been obtained in [36]–[38]. Among these nonlinear methods, LPP is a linear expression of LE by replacing the nonlinear mapping relation to reach the purpose of dimension reduction and is seen as an acceptable technique for its well practicability and fast calculation [39]. Hence, in this paper, LPP is introduced to remove some redundant information hidden in high-dimensional feature space.

Main contribution and novelty of this paper are the presentation of a multi-stage hybrid fault diagnosis approach for identifying different health conditions of rolling bearing. Within our method, three stages are executed separately to

improve the diagnostic accuracy and efficiency. For the first stage, permutation entropy (PE)-based fault initial detection is performed to determine whether there is a fault in bearing. If bearing comes up fault, the next two stages are conducted to identify fault type and fault severities. Specifically, in the stage of the identification of fault type and fault severities, hybrid-domain features containing time-domain, frequency-domain and time-frequency domain are extracted to construct a high-dimensional feature space. Next, LPP is introduced to compress the obtained high-dimensional feature space into low-dimensional feature space, which is aimed at removing the redundant or insensitive feature information. Finally, the obtained low-dimensional feature vector is taken as the input of Fuzzy C-means (FCM) clustering to achieve bearing fault classification. Of course, some other classifiers (e.g. neural network, fuzzy support vector machine [40], [41] and hidden markov model) can also be used to replace FCM to achieve fault recognition. Four cases including 12 health conditions are used respectively to estimate generalization performance of our approach. Results show that our approach can identify effectively bearing fault type and fault severity. Besides, extensive comparisons with traditional single-stage diagnosis approach highlight the superiority of our method.

The rest of this paper is organized as follows. Section II introduces the basic theory of PE and discusses the parameter selection of PE. Besides, in Section II, VMD method is reviewed and decomposition superiority of VMD is verified by using a numerical signal. Section III provides the detailed framework of the proposed approach for fault diagnosis of rolling element bearing under variable conditions. In Section IV, experimental data analysis derived from rolling bearing is used to validate the effectiveness and superiority of the provided method. The conclusion is drawn in Section V and some future work is provided.

## II. THEORY BACKGROUND

### A. PERMUTATION ENTROPY

For any given time series  $\{X(i), i = 1, 2, \dots, N\}$ , the results of phase-space reconstruction can be given by [42]

$$Y = \begin{pmatrix} x(1) & x(1 + \tau) & \dots & x(1 + (m - 1)\tau) \\ x(2) & x(2 + \tau) & \dots & x(2 + (m - 1)\tau) \\ x(j) & x(j + \tau) & \dots & x(j + (m - 1)\tau) \\ \vdots & \vdots & & \vdots \\ x(n) & x(n + \tau) & \dots & x(n + (m - 1)\tau) \end{pmatrix} \quad (1)$$

where  $m$  is the embedding dimension,  $\tau$  is the time delay,  $n$  represents the number of the reconstructed components,  $x(j)$  is the  $j$ -th row of the restructuring matrix  $Y$  and  $j = 1, 2, \dots, n$ . Each row of  $Y$  is a reconstructed component, which implies the matrix  $Y$  has  $n = N - (m - 1)\tau$  reconstructed component. The  $j$ -th reconstructed component of  $Y$  can be arranged in an ascending order as

$$x(i + (j_1 - 1)\tau) \leq x(i + (j_2 - 1)\tau) \leq \dots \leq x(i + (j_m - 1)\tau) \quad (2)$$

If there are two or more equivalent elements in reconstructed component  $x(i + (j_1 - 1)\tau) = x(i + (j_2 - 1)\tau)$ , according to value of  $j_1$  and  $j_2$ , the rank ordering is rewritten. In other words, when the value of  $j_1$  is less than  $j_2$ , the arrangement can be obtained by

$$x(i + (j_1 - 1)\tau) \leq x(i + (j_2 - 1)\tau) \quad (3)$$

For any reconstructed component, a set of symbol sequences are further be given by

$$S(l) = (j_1, j_2, \dots, j_m) \quad (4)$$

where  $l = 1, 2, \dots, n$  and  $n \leq m!$ , there is  $m!$  symbol sequence for  $m$  dimensional embedding space at most,  $S(l)$  is the one of  $m!$  symbol sequence. Probability distribution of each symbol sequence is  $P_1, P_2, \dots, P_k$  and satisfies the equality  $P_1 + P_2 + \dots + P_k = 1$ . Hence, PE value of the time series  $\{X(i), i = 1, 2, \dots, N\}$  can be expressed as below

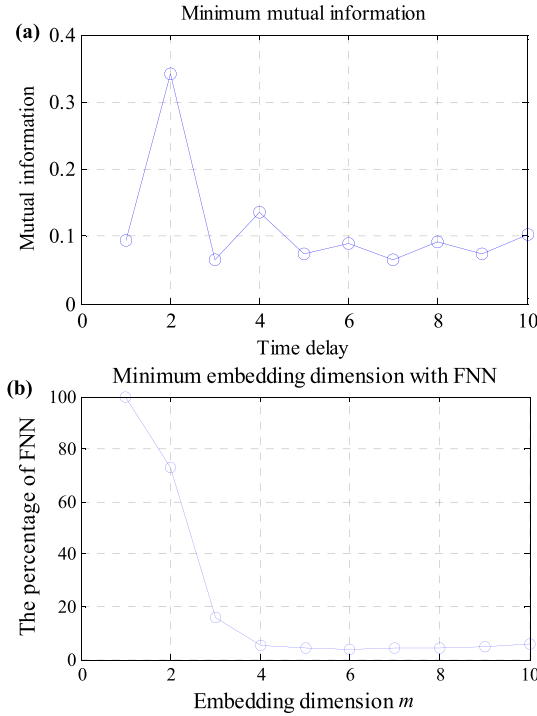
$$H_p(m) = - \sum_{j=1}^{m!} P_j \ln P_j \quad (5)$$

where  $0 \leq H_p(m) \leq \ln(m!)$ , when  $P_j = 1/m!$ ,  $\ln(m!)$  is the largest value of  $H_p(m)$ . PE can be further normalized as

$$H_p = H_p(m) / \ln(m!) \quad (6)$$

In Eq. (6),  $H_p$  can describe the complexity of the signal. The larger  $H_p$  represents the greater signal randomness. According to the definition of PE, two parameters (i.e. time delay  $\tau$  and embedding dimension  $m$ ) need to be set before calculating PE. If embedding dimension  $m$  is set too large, this will bring extra computation time and PE cannot reflect a subtle change of the time series. If embedding dimension  $m$  is set too small, the reconstructed component will contain tiny information and PE cannot detect the dynamic mutation of the time series. Hence, the selection principle of two key parameters (embedding dimension  $m$  and time delay  $\tau$ ) of PE is discussed in this part.

Firstly, for a bearing vibration signal with inner race fault from Case Western Reserve University (CWRU), the mutual information and false nearest neighbor (FNN) method reported in [43] and [44] are employed to determine two key parameters (time delay  $\tau$  and embedding dimension  $m$ ) of PE, respectively. Fig. 1(a) shows the variation relationship between mutual information and time delay  $\tau$ , while the variation relationship between FNN percent and embedding dimension  $m$  is plotted in Fig. 1(b). As can be seen from Fig. 1(a), the mutual information appears the local minimum in time delay  $\tau = 3$ . That is, the reconstructed time series can be as irrelevant as possible when the time delay  $\tau$  is set at 3. Besides, considering that many studies shows that time delay  $\tau$  has little influence on the calculation results of PE. Hence, in this study, according to [43], we empirically select time delay  $\tau = 3$  to analyze bearing vibration signal. Likewise, as shown in Fig. 1(b), percent of FNN maintains a steady trend when the embedding dimension  $m \geq 6$ . Therefore,



**FIGURE 1.** Time delay  $\tau$  and embedding dimension  $m$  of PE determined by different methods for a bearing vibration signal: (a) Mutual information method and (b) false nearest neighbor method.

in this study, according to [44], we choose embedding dimension  $m = 6$  to calculate PE value of bearing vibration signal. It is worth mentioning that determination of time delay  $\tau$  and embedding dimension  $m$  is a difficult problem, more specific and effective selection criteria of PE parameters (time delay  $\tau$  and embedding dimension  $m$ ) for different signals will be the focus of our future work.

**B. VMD METHOD**

VMD is a recently reported time-frequency analysis method, which is equivalent to an adaptive Wiener filter banks. Main idea of VMD is regarded as the solving for a constrained variational problem shown in Eq. (7).

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_k u_k = f \end{cases} \quad (7)$$

where  $u_k$  is the  $k$ -th intrinsic mode function (IMF) and  $\omega_k$  is the center frequency of the  $k$ -th IMF. To address efficiently the constrained variational issue, the penalty parameter  $\alpha$  and Lagrangian multiplier  $\lambda(t)$  are introduced. Consequently, the variational issue are rewritten as

$$\begin{aligned} &L(\{u_k\}, \{\omega_k\}, \lambda) \\ &= \alpha \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &+ \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (8)$$

According to the alternate direction method of multipliers, we can solve Eq. (8), and the generating IMF component and their corresponding center-frequency  $\omega_k$  can be updated in terms of the following Eq. (9).

$$\begin{cases} \hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \\ \omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \end{cases} \quad (9)$$

When the IMF component and its center-frequency are changed, Lagrangian multiplier  $\hat{\lambda}$  is also changed by

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)) \quad (10)$$

The update procedure is conducted repeatedly until the convergence conditions described in Eq. (11) is satisfied.

$$\sum_k \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2 / \left\| \hat{u}_k^n \right\|_2^2 < \varepsilon \quad (11)$$

where  $\varepsilon = 10^{-6}$  and all IMF components can be recovered according to the above-mentioned step. Concrete details of VMD algorithm can be found in the [24].

To show the effectiveness of VMD algorithm, here a numerical signal  $x(t)$  is considered as follows

$$\begin{cases} x(t) = x_1(t) + x_2(t) + x_3(t) \quad \in (0, 0.5) \\ x_1(t) = (1 + 0.5 \cos(18\pi t)) \cos(100\pi t + 40\pi t^2) \\ x_2(t) = \cos(240\pi t) \\ x_3(t) = \cos(360\pi t) \end{cases} \quad (12)$$

In Eq. (12), numerical signal  $x(t)$  consists of an amplitude modulation frequency modulation signal  $x_1(t)$  and two cosine signals ( $x_2(t)$  and  $x_3(t)$ ). Main frequency of three components ( $x_1(t)$ ,  $x_2(t)$  and  $x_3(t)$ ) are 50 Hz, 120 Hz and 180 Hz, respectively. The sampling frequency and sampling number are set at 2048 Hz and 1024, respectively. Numerical signal analysis is conducted on an Intel Pentium G3420 3.20GHz CPU with 4.00 GB RAM, and MATLAB (2010a) platform is employed to implement the simulation. Fig. 2 describes time domain waveform of the simulation signal  $x(t)$  and its three components ( $x_1(t)$ ,  $x_2(t)$  and  $x_3(t)$ ).

VMD is employed to analyze the simulation signal  $x(t)$ . Firstly, central frequency iteration rule in [31] is adopted to determine the decomposition number  $K$  of VMD. Table 1 lists the central frequency value of each sub-signals obtained by VMD when the preset mode number  $K = 2 \sim 6$ . From the Table 1, it is clear that the over decomposition phenomenon occurs in decomposition results when the mode number  $K = 4 \sim 6$ . Thus, mode number  $K$  of VMD is set as 3 in processing the simulation signal  $x(t)$ . Fig. 3(a) plots time-frequency representation generated by VMD. As shown in Fig. 3(a), IMF component obtained by VMD is very close to the real ingredient (50 Hz, 120 Hz and 180 Hz), which implies that VMD is effective in analyzing

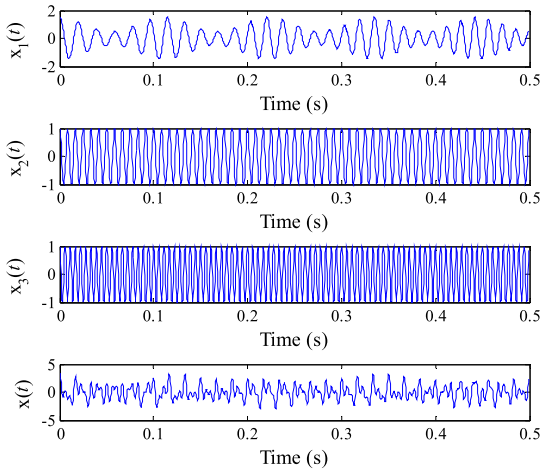


FIGURE 2. Simulation signal  $x(t)$  and its three components.

TABLE 1. The central frequency of each sub-signals obtained by VMD under different mode number.

Mode number $K$	Central frequency of each sub-signal (Hz)					
2	33.6	80.7	—	—	—	—
3	29.2	58.6	87.9	—	—	—
4	26.4	32.0	58.6	87.9	—	—
5	25.5	29.6	33.4	58.6	87.9	—
6	21.6	25.9	29.7	33.2	58.6	87.9

multi-component signal. Besides, three traditional methods (i.e. EMD, WT and WVD) are applied to analyze the same simulation signal  $x(t)$  and the obtained time-frequency representation is displayed in Figs. 3(b)-(d), respectively. It becomes apparent in Fig. 3(b) that EMD-based time-frequency representation appears serious mode mixing problem. Fig. 3(c) shows that time-frequency representation obtained by WT can reveal three components of  $x(t)$ , but it suffers from lower resolution. Besides, there is cross-term interference of 150 Hz and 90 Hz in the time-frequency representation generated by WVD (see the red arrow in Fig. 3(d)). Hence, VMD method is adopted to decompose bearing vibration signal and extract multiscale features in this paper.

C. HYBRID-DOMAIN FEATURE EXTRACTION

1) STATISTICAL FEATURE EXTRACTION

Time-domain and frequency-domain features used in this paper are listed in Table 2, where  $x(n), n = 1, 2, \dots, N$  is a given discrete time series,  $N$  is the number of data points of  $x(n), y(k), k = 1, 2, \dots, K$  is FFT spectrum of  $x(n), K$  is the number of spectrum line in FFT spectrum,  $f_k$  is the frequency value of the  $k$ -th spectrum line,  $I_1 \sim I_{16}$  represents 16 time-domain statistical features (i.e. mean value, standard deviation, square root amplitude, absolute mean value, skewness, kurtosis, variance, maximum, minimum, peak-to-peak value, waveform index, peak index, pulse index, margin index, skewness index, kurtosis index),  $I_{17} \sim I_{29}$  represents 13 frequency-domain statistical features, where  $I_{17}$  denotes

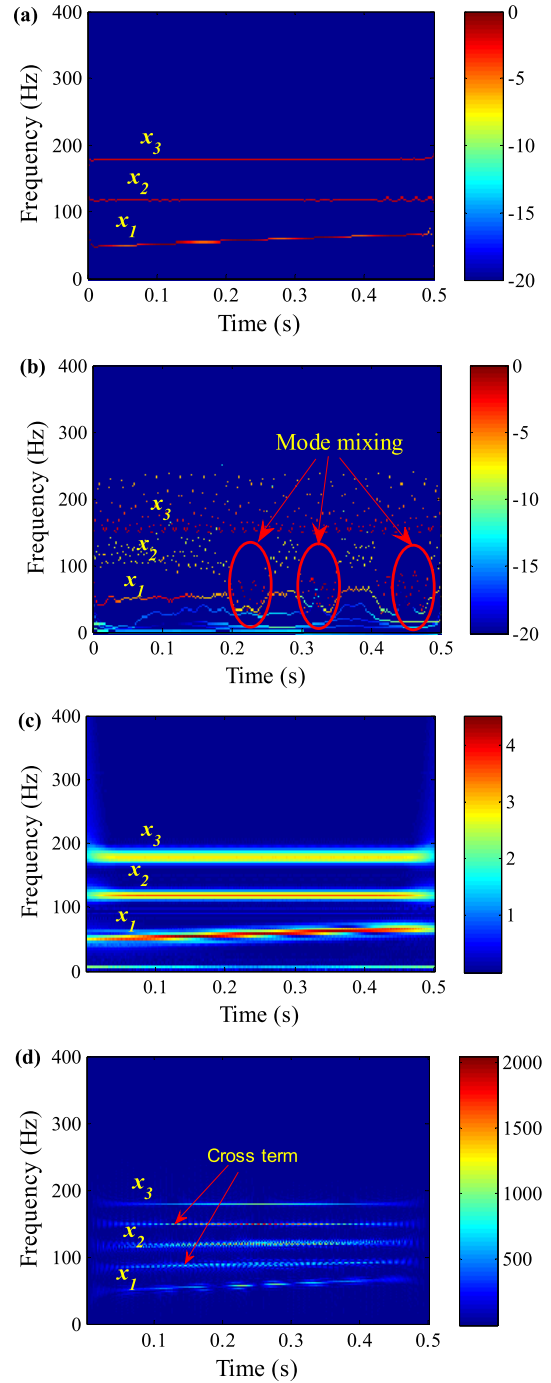


FIGURE 3. Time-frequency representation obtained by different methods: (a) VMD, (b) EMD, (c) WT and (d) WVD for simulation signal.

frequency domain energy of the signal  $x(n), I_{18} \sim I_{20}, I_{22}$  and  $I_{26} \sim I_{29}$  respectively denotes the concentration and dispersion of frequency spectrum of the signal  $x(n)$ , and  $I_{21}$  and  $I_{23} \sim I_{25}$  denotes position variance of main frequency band of the signal  $x(n)$ .

2) MULTISCALE FEATURE EXTRACTION

As is well-known, the real bearing vibration signal is characterized by nonlinear and non-stationary, which means that it

TABLE 2. Expression of statistical features.

Number	Feature expression	Number	Feature expression
1	$I_1 = \frac{1}{N} \sum_{n=1}^N x(n)$	16	$I_{16} = \frac{I_6}{(I_7)^2}$
2	$I_2 = \sqrt{\frac{1}{N-1} \sum_{n=1}^N [x(n) - I_1]^2}$	17	$I_{17} = \frac{1}{K} \sum_{k=1}^K y(k)$
3	$I_3 = \left( \frac{1}{N} \sum_{n=1}^N \sqrt{ x(n) } \right)^2$	18	$I_{18} = \frac{1}{K-1} \sum_{k=1}^K [y(k) - I_{17}]^2$
4	$I_4 = \frac{1}{N} \sum_{n=1}^N  x(n) $	19	$I_{19} = \frac{1}{K(\sqrt{I_{18}})^3} \sum_{k=1}^K [y(k) - I_{17}]^3$
5	$I_5 = \frac{1}{N} \sum_{n=1}^N [x(n)]^3$	20	$I_{20} = \frac{1}{K(I_{18})^2} \sum_{k=1}^K [y(k) - I_{17}]^4$
6	$I_6 = \frac{1}{N} \sum_{n=1}^N [x(n)]^4$	21	$I_{21} = \frac{\sum_{k=1}^K [f_k y(k)]}{\sum_{k=1}^K y(k)}$
7	$I_7 = \frac{1}{N} \sum_{n=1}^N [x(n)]^2$	22	$I_{22} = \frac{1}{K} \sqrt{\sum_{k=1}^K [(f_k - I_{21})^2 y(k)]}$
8	$I_8 = \max  x(n) $	23	$I_{23} = \frac{\sqrt{\sum_{k=1}^K [f_k^2 y(k)]}}{\sum_{k=1}^K y(k)}$
9	$I_9 = \min  x(n) $	24	$I_{24} = \frac{\sqrt{\sum_{k=1}^K [f_k^4 y(k)]}}{\sum_{k=1}^K [f_k^2 y(k)]}$
10	$I_{10} = I_8 - I_9$	25	$I_{25} = \frac{\sum_{k=1}^K (f_k^2 y(k))}{\sqrt{\sum_{k=1}^K (f_k^4 y(k))} \sum_{k=1}^K y(k)}$
11	$I_{11} = \frac{I_2}{I_4}$	26	$I_{26} = \frac{I_{22}}{I_{21}}$
12	$I_{12} = \frac{I_8}{I_2}$	27	$I_{27} = \frac{1}{K(I_{22})^3} \sum_{k=1}^K [(f_k - I_{21})^3 y(k)]$
13	$I_{13} = \frac{I_8}{I_4}$	28	$I_{28} = \frac{1}{K(I_{22})^4} \sum_{k=1}^K [(f_k - I_{21})^4 y(k)]$
14	$I_{14} = \frac{I_8}{I_3}$	29	$I_{29} = \frac{1}{K\sqrt{I_{22}}} \sum_{k=1}^K [\sqrt{ f_k - I_{21} } y(k)]$
15	$I_{15} = \frac{I_5}{(\sqrt{I_7})^3}$	—	—

is difficult to reveal fully the inherent properties only using time-domain or frequency-domain features. Multiscale features obtained by time-frequency analysis method have been

proved to be effective in fault recognition [45]. Hence, VMD is employed to decompose bearing vibration signal into several sub-signals. Afterwards, multiscale features are obtained

through calculating amplitude energy value and instantaneous energy distribution-permutation entropy of each sub-signal. Detailed process is described below:

**Step 1:** For the given signal  $x(n)$ , using VMD to divide  $x(n)$  into several sub-signals  $u_j(n), j = 1, 2, \dots, K$ , where  $K$  is the decomposition mode number and determined by the central frequency iteration rule in [31].

**Step 2:** Calculate the amplitude energy value of each sub-signals and denoted by  $I_{30}^j$ .

$$I_{30}^j = \sum_{n=1}^N [FFT(u_j(n))]^2, \quad j = 1, 2, \dots, K \quad (13)$$

where  $FFT(u_j(n))$  represents fast Fourier transform of each sub-signals and  $N$  is the sampling number.

**Step 3:** According to [46], calculate the instantaneous energy distribution-permutation entropy  $I_{31}^j$  of each sub-signals. Instantaneous energy distribution of each sub-signal  $u_j(n)$  is expressed as follows

$$e_j(n) = \frac{1}{2} |a_j(n)|^2, \quad j = 1, 2, \dots, K \quad (14)$$

where  $a_j(n)$  is the instantaneous amplitude function of sub-signal  $u_j(n)$ . Note that  $e_j(n)$  is regarded as the input of PE to calculate the corresponding instantaneous energy distribution-permutation entropy and denoted by  $I_{31}^j$ .

Finally, high-dimensional feature space for training set can be constructed according to hybrid-domain features, of which each features is an element of it, as shown in Eq. (15).

$$H_{train} = \begin{bmatrix} I_{1 \times 16} & I'_{1 \times 13} & I''_{1 \times K} & I'''_{1 \times K} \\ \vdots & \vdots & \vdots & \vdots \\ I_{p \times 16} & I'_{p \times 13} & I''_{p \times K} & I'''_{p \times K} \\ \vdots & \vdots & \vdots & \vdots \\ I_{q \times 16} & I'_{q \times 13} & I''_{q \times K} & I'''_{q \times K} \end{bmatrix}_{q \times (16+13+2 \times K)} \quad (15)$$

where  $q$  denotes the total number of data sample, the feature dimension of  $H_{train}$  is  $16 + 13 + 2 \times K$ ,  $I_{p \times 16}$  indicates the 16 dimensional time-domain feature vector of the  $p$ -th sample,  $I'_{p \times 13}$  denotes the 13 dimensional frequency-domain feature vector of the  $p$ -th sample,  $I''_{p \times K}$  indicates the  $K$  dimensional amplitude energy feature vector of the  $p$ -th sample,  $I'''_{p \times K}$  implies the  $K$  dimensional instantaneous energy distribution-permutation entropy feature vector the  $p$ -th sample. The dimension of high-dimensional feature space  $H_{test}$  for the testing set is the same as that of the training set.

#### D. LOCALITY PRESERVING PROJECTION

Locality preserving projection (LPP) is regarded as a linear approximation procedure of Laplacian eigenmaps. It avoided problems that cannot effectively handle testing samples and improved generalization ability of the algorithm. Assuming  $X = (x_1, x_2, \dots, x_n)$  is the original high-dimensional space dataset and its dimension is  $m$ ,  $Y = (y_1, y_2, \dots, y_n)$  is the dataset after dimension reduction and its dimension is  $d$ , the implementation procedure of LPP is outlined below

**Step 1:** Calculate the neighbor points. According to Eq. (16), calculate Euclidean distance between each sample point  $x_i(i, j = 1, 2, \dots, n)$  and the remainder sample points, find out its nearest  $k$  points and construct a neighbor map.

$$d(x_i, x_j) = \|x_i - x_j\| \quad (16)$$

**Step 2:** Select the weighting value. If  $x_i$  and  $x_j$  is neighboring, the weighting value  $W_{ij} = \exp(-\|x_i - x_j\|^2/t)$ , otherwise the weight value  $W_{ij} = 0$ , where  $t$  is the heat kernel parameter,

**Step 3:** Compute the eigenvector  $a$  and eigenvalue  $\lambda$  from the formula  $XLX^T a = \lambda XDX^T a$ , where  $D$  is the diagonal matrix and meets  $D_{ii} = \sum_j W_{ji}$ ,  $L$  is Laplacian matrix and meets  $L = D - W$ ,  $W$  is the weight matrix and its elements is  $W_{ij}$ , symbol  $a = a_1, \dots, a_d$  is the corresponding eigenvector of eigenvalue  $\lambda_1 < \dots < \lambda_d$ . Hence, low-dimensional space sample  $y_i = A^T x_i$ , where  $A = (a_1, \dots, a_d)$  is the transfer matrix. Firstly, transfer matrix  $A$  of high-dimensional space mapped to low-dimensional space for the training sample is obtained, and then low-dimensional space for the testing samples can be obtained just by the transfer matrix  $A$ .

### III. THE PROPOSED DIAGNOSIS SCHEME

To improve the diagnostic accuracy of rolling bearing and computational efficiency, a neoteric multi-stage hybrid fault diagnosis scheme is proposed in this paper. Fig. 4 shows the whole process of the proposed approach for bearing fault detection, which is mainly divided into three stages (i.e. PE-based fault initial detection, fault type recognition and fault severity assessment).

#### A. PE-BASED FAULT INITIAL DETECTION

As reported in [42], PE has been proved to be effective in identifying bearing health status, so PE is used for fault initial detection in the first stage of our method. However, when we apply PE to monitor bearing health status, a key problem is to determine the suitable threshold. According to chebyshev's theorem, for any dataset, it will meet the inequality shown in Eq. (17).

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2} \quad (17)$$

where  $X$  denotes the given dataset,  $\mu$  and  $\sigma$  indicates the mean value and standard deviation of  $X$ , and  $k$  is a constant. When  $k = 5$ , at least data of 24/25 (i.e. 96%) is located in the scope of plus and minus five standard deviations of mean. That is, the probability of PE value of data sample lies within the interval  $[\mu - 5\sigma, \mu + 5\sigma]$  is 96%. Besides, prior studies have shown that PE value of the data sample under normal state is lesser than those of the data sample under fault state. Thereby, in this study, data sample under normal state is adopted to calculate the PE threshold, and the upper value  $\mu + 5\sigma$  can be deemed as the PE threshold for achieving fault initial detection accuracy of 96%.

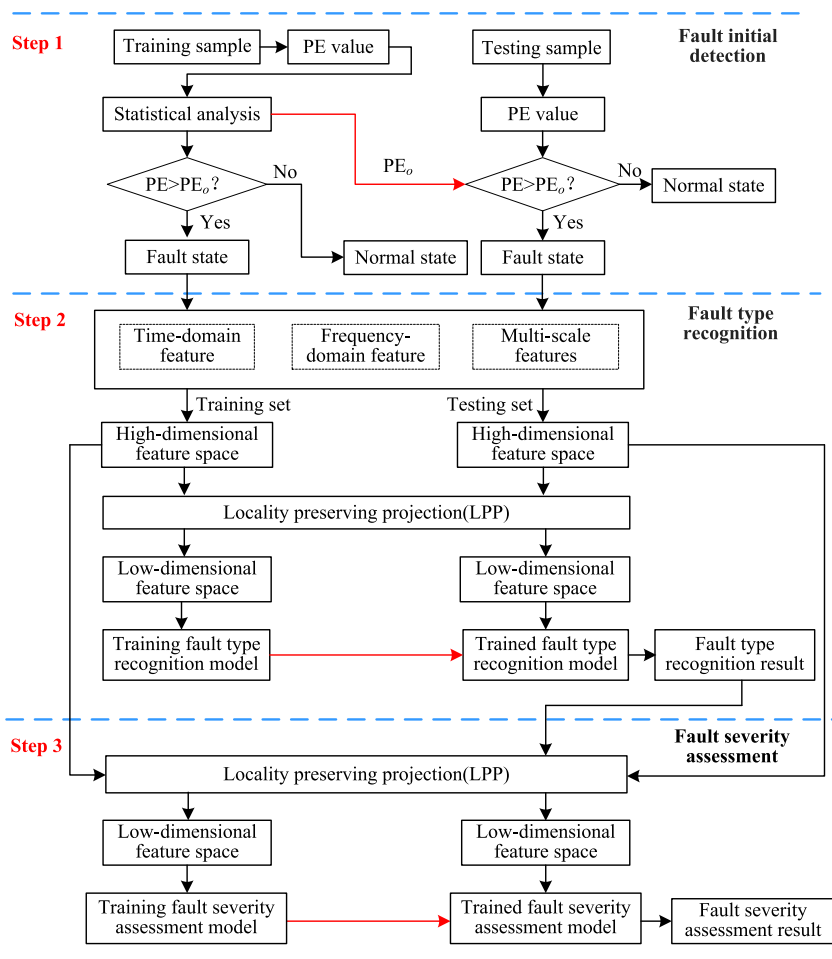


FIGURE 4. Flowchart of the proposed diagnosis scheme.

**B. FAULT TYPE RECOGNITION**

After fault initial detection, if a fault occurs, we will conduct fault type recognition. Based on statistical method and VMD, hybrid-domain features containing time-domain, frequency-domain and time-frequency domain are firstly extracted to obtain a high-dimensional feature space equipped with more comprehensive and richer fault feature signatures. Next, high-dimensional feature space is normalized to unify the statistical distribution of samples. Meanwhile, for purpose of reducing the information redundancy of high-dimensional feature space, the dimension reduction algorithm called LLP is adopted to obtain the low-dimensional feature space which has higher discrimination ability for different bearing fault types. Finally, the low-dimensional feature space is input to the trained FCM classifier to identify automatically bearing fault type and output the diagnostic results.

**C. FAULT SEVERITY ASSESSMENT**

According to the result of fault type recognition, fault severity assessment is performed for each type of fault which is correctly classified. Firstly, for the training and testing set, high-dimensional feature space established in fault type

recognition is divided into several sub-spaces according to fault types. Each sub-space of the training and testing set is projected to low-dimensional feature space by using LPP. Afterwards, low-dimensional feature space of the training set is regarded as the input of FCM classifier to train fault severity assessment model. Finally, low-dimensional feature space of the testing set is imported into the trained fault severity assessment model to carry out fault severity assessment of each type of fault and output the results of fault severity assessment.

**IV. EXPERIMENTAL STUDY**

**A. DATA DESCRIPTION**

Motor bearing fault data derived from Case Western Reserve University (CWRU) is devoted to prove the validity of the proposed method. Experimental device is depicted in Fig. 5, which is mainly composed of an induction motor, a torque transducer and a dynamometer. Three kinds of single-point defects (i.e. inner race (IR), outer race (OR) and ball (B)) are generated respectively on normal bearing by using spark machining, where IR and B have four fault severities (0.007, 0.014, 0.021 and 0.028 inches), while OR only contains three



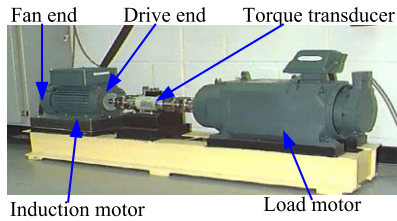


FIGURE 5. Bearing fault simulation test-bed.

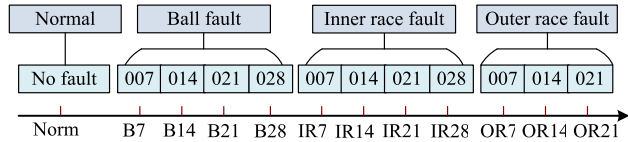


FIGURE 6. Detailed description of the considered health condition in the four cases (0-3hp).

fault severities (0.007, 0.014 and 0.021 inches) of 6 o'clock positions. Dataset under four states (i.e. normal, IR defect, OR defect and ball defect) were collected by using an accelerometer installed on motor drive end of the laboratory bench. Also, bearing vibration data was acquired under four different motor loads (0, 1, 2 and 3hp), where four constant motor loads (0, 1, 2 and 3hp) respectively correspond to four constant motor speeds (1797, 1772, 1750 and 1730 rpm).

To illustrate the applicability of the proposed method, four cases (the dataset of 0-3hp) are considered in this study. The investigative health conditions for each case are described in Fig. 6. For each case (i.e. each motor load or speed), the dataset under 12 health status (i.e. Norm, IR7, IR14, IR21, IR28, OR7, OR14, OR21, B7, B14, B21, B28) is established, where "Norm" is the abbreviation of normal state and "IR7" is the abbreviation of inner race fault with 0.007 inches. Sampling frequency during the experiment was set at 12 kHz. The first 12,000 points under each health status are divided averagely into 40 data samples (i.e. each data sample contains 3000 points), where 20 data samples under each health status are randomly selected for training and the remainder 20 samples is taken as testing. Thus, 240 training samples and 240 test samples can be acquired. Essentially it is a twelve-class classification issue need to be solved. Fig. 7 shows the raw bearing vibration signal of different health status in case 1 (0hp). As shown in Fig. 7, it is difficult to detect different fault types directly by observing time domain waveform. Hence, to improve the diagnostic accuracy and computing efficiency, we first need to determine whether faults or not (i.e. fault initial detection). If the bearing appears fault, the second stage (i.e. fault type recognition) is performed and then conducts the third stage (i.e. fault severity assessment). Concretely, the proposed method is implemented in accordance with the following three stages.

**B. PE-BASED FAULT INITIAL DETECTION**

According to the first stage of the proposed method, PE-based fault initial detection is conducted to estimate whether bearing fault occurs. For four cases under 0-3hp, PE value of each data samples in the training and testing set is calculated

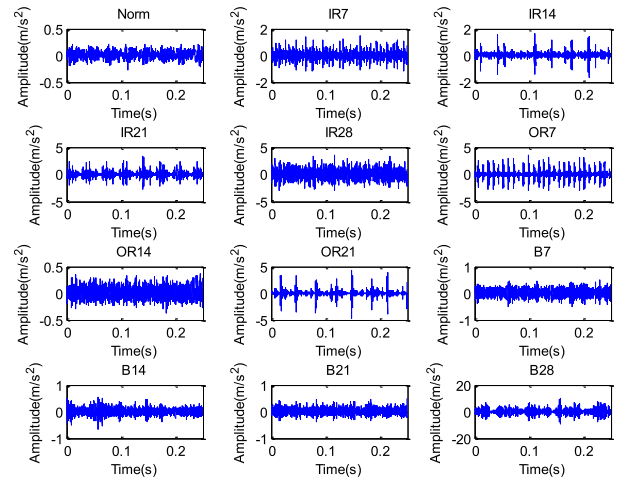


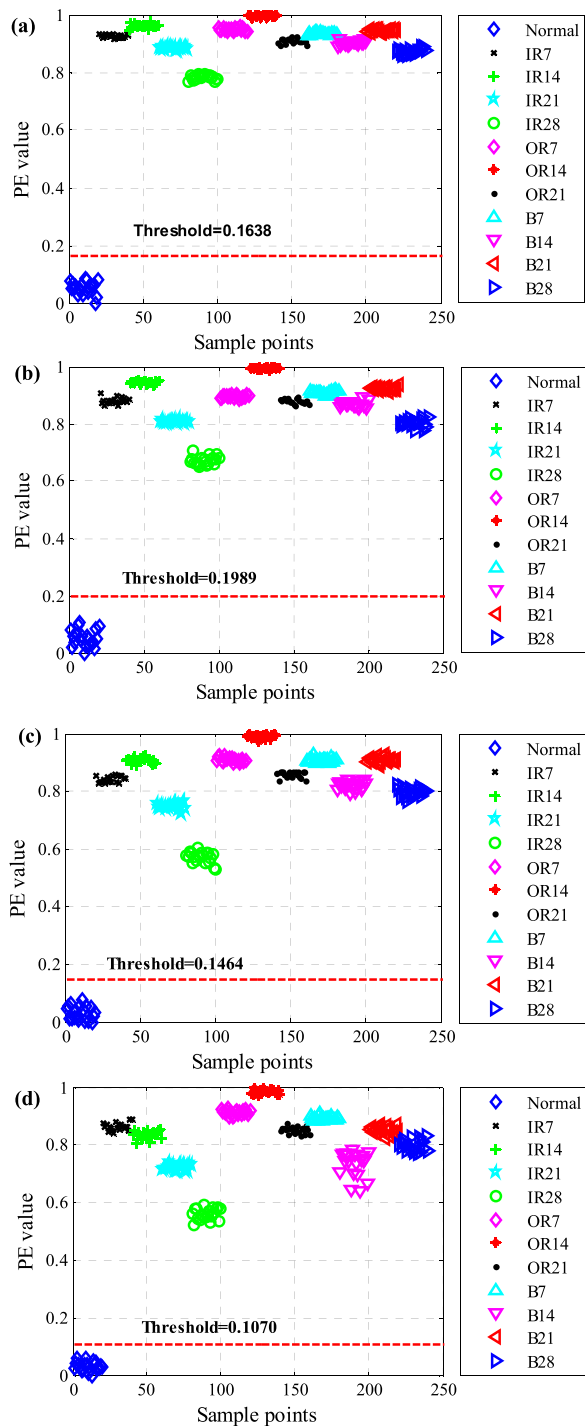
FIGURE 7. Time domain waveform of the original bearing vibration signal under different health conditions in case 1 (0hp).

according to Fig. 4, and the computed results for the training and testing set are plotted in Figs. 8 and 9, respectively. One can see clearly from Figs. 8 and 9 that PE of the data sample under fault state is considerably greater than those of the data sample under normal state. For the sake of intuition, a key point in this stage is the setting up of threshold for assessing rapidly and accurately health state of rolling bearing. Here, based on chebyshev's theorem, the red dotted line value in Figs. 8 and 9 is regarded as the PE threshold for obtaining detection accuracy of 96%. If PE value of one data sample is greater than the PE threshold, it indicates the fault state, otherwise denotes the normal state.

As shown in Figs. 8 and 9, PE value of the data sample under fault state for four cases is greater than the PE threshold, whereas PE value of the data sample under normal state is smaller than the PE threshold. This indicates that bearing health status can be evaluated effectively by the application of PE. In other words, fault preliminary detection is executed successfully in four cases under 0-3hp. It is noteworthy that some overlapping phenomenon occurs in the data sample under fault states, which mean that the final diagnostic target cannot be achieved just by calculating PE value. The following two stages are required to implement.

**C. FAULT TYPE RECOGNITION**

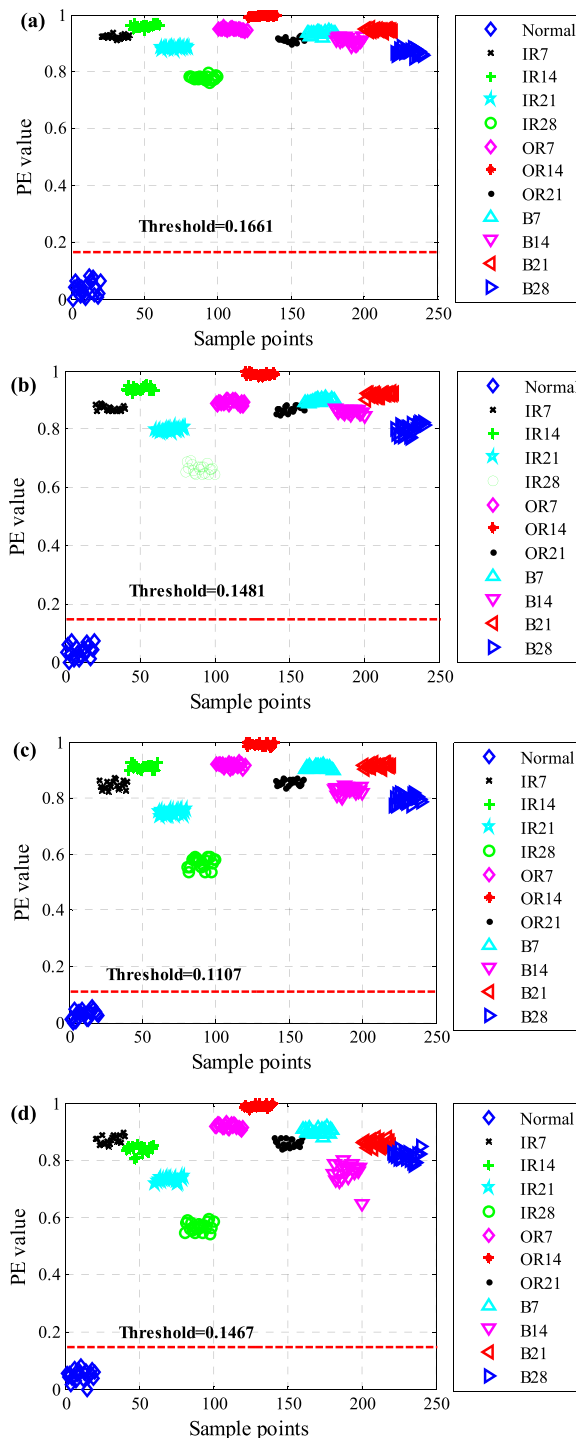
Similarly, as shown in Fig. 4, the second stage is fault type recognition. Firstly, hybrid domain features containing time-domain, frequency-domain and time-frequency domain are extracted to construct a high-dimensional feature space. Specifically, for each sample, 16 time-domain features and 13 frequency-domain features in Table 2 are calculated. Meanwhile, VMD is used to extract 8 multiscale features. In VMD, the balancing parameter *a* is set at 2000, and mode number *K* of VMD is determined by the central frequency iteration rule in [31]. For one data sample with inner race fault, Table 3 shows the central frequency of each sub-components obtained by VMD under different mode number *K*. As is clear in Table 3, similar central frequency will



**FIGURE 8.** Scatter plots of PE value of the training set in four cases: (a) case 1: 0hp, (b) case 2: 1hp, (c) case 3: 2hp and (d) case 4: 3hp.

occur in VMD results when decomposition mode number  $K$  is set at 5 to 10. This indicates that the over decomposition phenomenon occurs. Hence, decomposition mode number  $K$  of VMD is set at 4 in this study.

Figs. 10(a) and (b) show each sub-components obtained by VMD and their corresponding FFT spectrum, respectively. Besides, four sub-components obtained by EMD and their corresponding FFT spectrum is shown in Figs. 11(a) and (b),



**FIGURE 9.** Scatter plots of PE value of the testing set in four cases: (a) case 1: 0hp, (b) case 2: 1hp, (c) case 3: 2hp and (d) case 4: 3hp.

respectively. As can be seen from Figs. 10 and 11, compared with EMD, each sub-component obtained by VMD has better narrow-band property, which implies VMD is more suitable for signal decomposition. Hence, for each sample, four amplitude energy features and four instantaneous energy distribution-permutation entropy features will be obtained in multiscale feature extraction. That is, for the training and

TABLE 3. Central frequency of each sub-signal obtained by VMD with different decomposition mode number.

Mode number $K$		Central frequency corresponding to each sub-signal (Hz)									
2	1353	2808	—	—	—	—	—	—	—	—	—
3	1176	2283	2999	—	—	—	—	—	—	—	—
4	479	1132	2264	2984	—	—	—	—	—	—	—
5	473	1124	2244	2778	3037	—	—	—	—	—	—
6	473	1106	2094	2359	2807	3035	—	—	—	—	—
7	468	1105	2092	2356	2793	3014	3387	—	—	—	—
8	64	524	1109	2095	2358	2794	3015	3409	—	—	—
9	63	524	1108	2096	2357	2782	2989	3142	3734	—	—
10	52	516	1071	1555	2125	2366	2787	2989	3142	4360	—

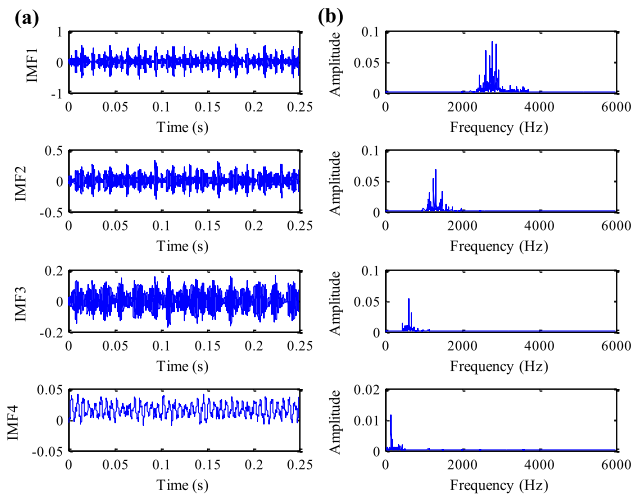


FIGURE 10. Decomposition results obtained by VMD for one sample with IR fault: (a) sub-signal and (b) the corresponding FFT spectrum.

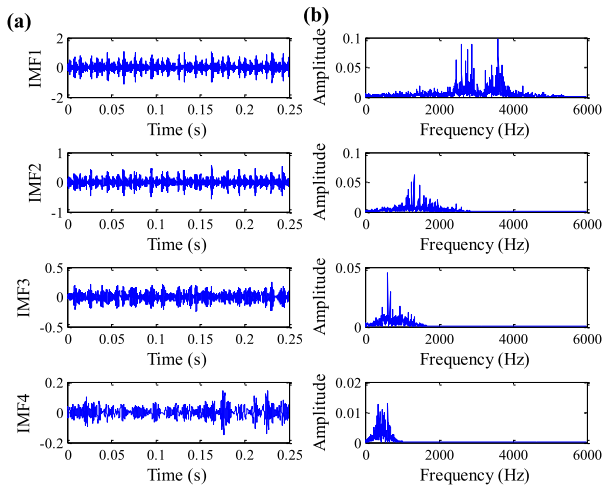


FIGURE 11. Decomposition results obtained by EMD for one sample with IR fault: (a) sub-signal and (b) the corresponding FFT spectrum.

testing sets, the high-dimensional feature dataset with dimension size of  $220 \times 37$  are constructed in this stage.

Secondly, to remove the information redundancy of high-dimensional feature space, LPP is employed to reduce the dimension of high-dimensional feature space to 3 dimensions (3D). Take the testing set of case1 (0hp) as an example,

Fig. 12(a) shows scatter plots of 3D feature space obtained by LPP for the testing set. As shown in Fig. 12(a), three types of fault (i.e. IR, OR and ball fault) are fully differentiated.

For a comparison, five dimension reduction methods (i.e. PCA, LLE, ISOMAP, LTSA and LS) are adopted to process the same high-dimensional feature space. Figs. 12(b)-(f) are scatter plots of 3D feature space obtained by PCA, LLE, ISOMAP, LTSA and LS, respectively. As can be seen from Figs. 12(b)-(f), there are overlapping regions among data samples with different faults, which mean the identification of bearing fault type is impeded. Namely, the result of dimension reduction of LPP outperforms other five methods (i.e. PCA, LLE, ISOMAP, LTSA and LS) in this case.

To further evaluate the clustering performance of different dimension reduction methods, 3D feature space obtained by different dimension reduction methods is regarded as the input of FCM clustering classifier to calculate the corresponding classification coefficient and average fuzzy entropy. According to [31], the bigger the classification coefficient is, the better the clustering performance is. That is, the larger classification coefficient indicates the greater discriminating degree and the higher identification precision. Besides, due to average fuzzy entropy is expected to be close to zero, so the smaller average fuzzy entropy shows the better clustering results. Note that, to compare the computational efficiency of different methods, CPU time of each dimension reduction method is calculated on MATLAB 2010a platform. Table 4 lists the comparison result of different dimension reduction methods.

It can be seen clearly from Table 4 that LPP has highest identification rate of 100% compared with other approaches. Specifically, the classification coefficient obtained using LPP is higher than that of other five methods (i.e. PCA, LLE, ISOMAP, LTSA and LS). Moreover, average fuzzy entropy obtained using LPP is lesser than that of other five methods (i.e. PCA, LLE, ISOMAP, LTSA and LS), and computation time of LPP is the smallest among six methods. This means that LPP is more effective in feature dimension reduction and more suitable for online monitoring diagnosis.

To investigate the influence of feature dimension on the diagnosis results, we calculate the diagnostic accuracy obtained by different dimension reduction methods under different feature dimension. The detailed diagnosis result

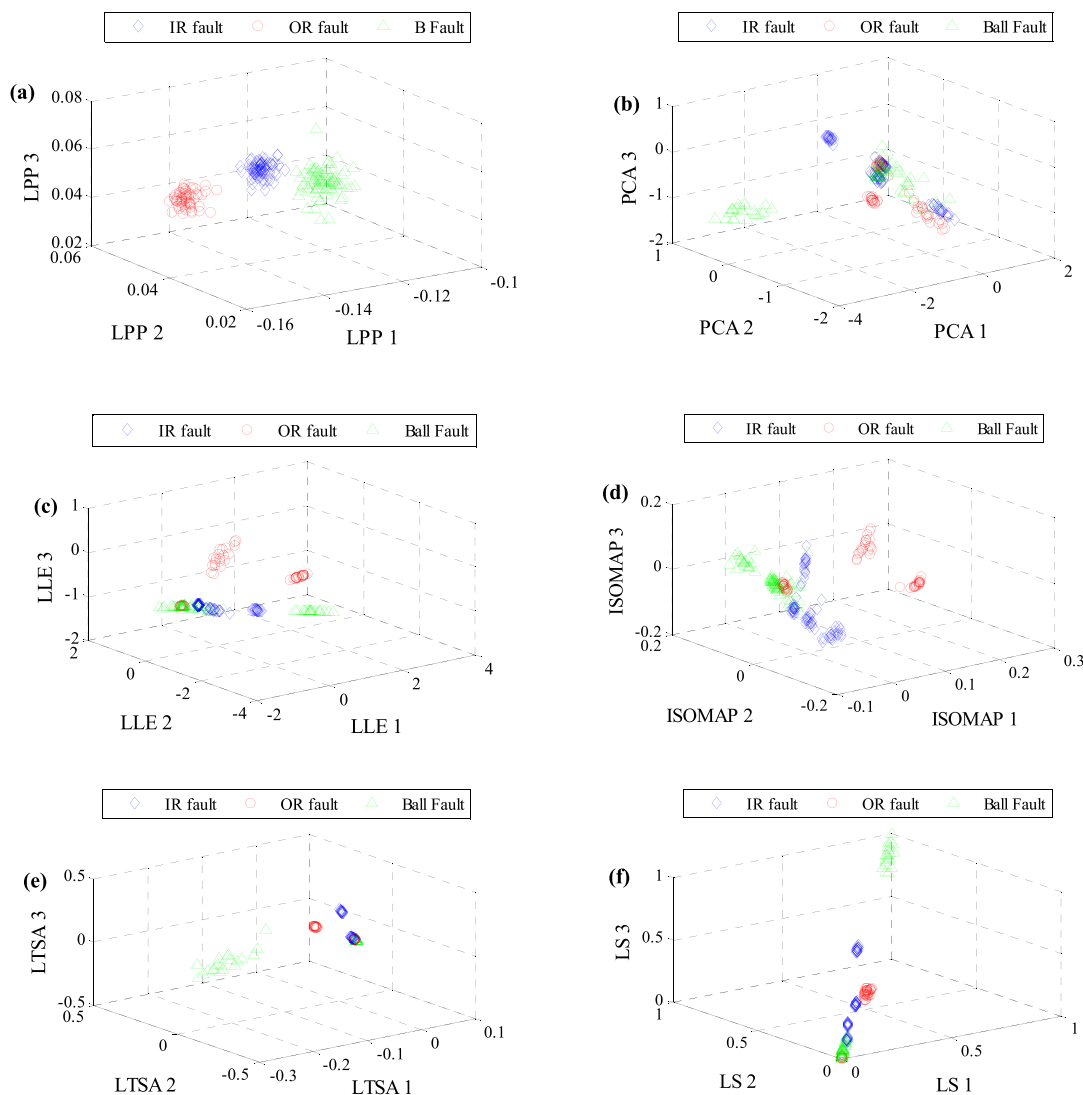


FIGURE 12. Scatter plots of 3D feature space obtained by different dimensionality reduction methods for testing set under ohp: (a) LPP, (b) PCA, (c) LLE, (d) ISOMAP, (e) LTSA and (f) LS.

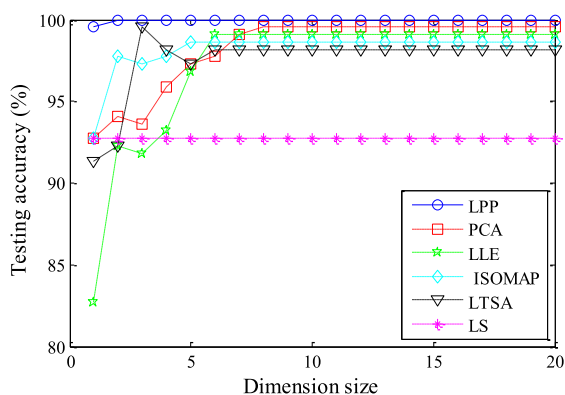


FIGURE 13. Diagnosis results obtained by different dimensionality reduction methods under different dimension sizes in case 1.

of different methods is shown in Fig. 13. It can be seen clearly from Fig. 13 that diagnosis accuracy obtained by different methods is inclined to smooth and steady increasing

with dimension size. However, when the dimension size is between 1 and 3, diagnosis accuracy of LPP is obviously superior to that of other dimension reduction methods, which implies that the low-dimensional feature space obtained using LPP can better remove the redundant and unnecessary features, preserve the intrinsic property of the raw signal and has preferable discrimination.

#### D. FAULT SEVERITY ASSESSMENT

After finishing fault type recognition, the next stage is the fault severity assessment. That is, the identification of fault severity of three conditions (IR, OR and B) is conducted respectively. Similarly, take case 1 as an example, after establishing high-dimensional feature space, LPP is utilized to remove the redundancy and reduce the dimension. Figs. 14(a)-(c) shows representation of 3D feature obtained by LPP for the testing set of three conditions (IR, OR and B), respectively. It is very obvious in Fig. 14 that fault

**TABLE 4. Diagnostic capability of different dimension reduction methods for the testing set in four cases (0-3hp).**

Case	Method	Classification coefficient	Average fuzzy entropy	CPU time (s)	Recognition rate (%)
Case 1 (0hp)	LPP	0.9591	0.1882	0.0658	100
	PCA	0.7535	0.4384	0.1028	97.73
	LLE	0.8485	0.2985	0.0636	98.63
	ISOMAP	0.7117	0.5319	1.3438	96.82
	LTSA	0.7027	0.5351	1.9821	92.73
	LS	0.8898	0.2036	1.9753	98.18
Case 2 (1hp)	LPP	0.9318	0.1419	0.0419	99.54
	PCA	0.7457	0.4554	0.0129	95.91
	LLE	0.7288	0.4740	0.1386	94.54
	ISOMAP	0.6549	0.6089	1.4309	92.73
	LTSA	0.8241	0.3320	1.2509	97.73
	LS	0.8563	0.2563	2.0848	98.18
Case 3 (2hp)	LPP	0.9525	0.1148	0.0471	100
	PCA	0.7666	0.4194	0.0217	93.18
	LLE	0.8269	0.3077	0.9632	95.91
	ISOMAP	0.6680	0.5971	2.5056	92.73
	LTSA	0.8300	0.3076	0.9131	96.82
	LS	0.8663	0.2432	2.1637	97.27
Case 4 (3hp)	LPP	0.9296	0.1610	0.0388	99.54
	PCA	0.7766	0.3965	0.0185	94.54
	LLE	0.8691	0.2490	0.0955	97.27
	ISOMAP	0.6417	0.6342	1.2133	93.18
	LTSA	0.8816	0.2351	1.3709	98.18
	LS	0.8582	0.2526	2.4204	95.45

**TABLE 5. Diagnostic capability of the proposed method for the testing set at stage of fault severity assessment.**

Case	Fault severity assessment	Classification coefficient	Average fuzzy entropy	Average recognition rate (%)
Case 1 (0hp)	IR7, IR14, IR21, IR28	0.9841	0.0129	98.75
	OR7, OR14, OR21	0.9995	0.0025	98.99
	B7, B14, B21, B28	0.9974	0.0097	99.25
Case 2 (1hp)	IR7, IR14, IR21, IR28	0.9703	0.0739	99.50
	OR7, OR14, OR21	0.9998	0.0012	99.33
	B7, B14, B21, B28	0.9920	0.0231	98.75
Case 3 (2hp)	IR7, IR14, IR21, IR28	0.9985	0.0062	99.25
	OR7, OR14, OR21	0.9995	0.0024	99.67
	B7, B14, B21, B28	0.9765	0.0628	99.50
Case 4 (3hp)	IR7, IR14, IR21, IR28	0.9980	0.0081	99.75
	OR7, OR14, OR21	0.9996	0.0020	99.33
	B7, B14, B21, B28	0.9833	0.0460	99.25

**TABLE 6. Comparison results among different methods.**

Methods	Case 1 (0hp)		Case 2 (1hp)		Case 3 (2hp)		Case 4 (3hp)	
	Accuracy	CPU time	Accuracy	CPU time	Accuracy	CPU time	Accuracy	CPU time
Multi-stage approach (Proposed method)	99.67%	26.78s	99.25%	25.16s	99.67%	27.02s	99.25%	25.93s
Single-stage approach with LPP	95.83%	29.99s	94.17%	30.15s	95.41%	29.55s	95.00%	28.93s
Single-stage approach without LPP	92.91%	58.77s	90.83%	59.15s	91.67%	57.64s	92.08%	58.03s

severity of three conditions (IR, OR and B) has good discrimination. Concretely, intra-class of the sample of the same class has good aggregation, whereas inter-class distance of the sample of different class is very large. Finally, FCM classifier is employed to evaluate the results of fault severity identification.

To avoid the contingency of diagnostic results, here five calculation trials are conducted in this stage. Take 0hp as an example, for fault severity identification of IR, the recognition rate obtained in five trials is 100%, 98.75%, 97.50%, 98.75% and 98.75%, respectively. For fault severity identification of OR, the recognition rate obtained in five trials

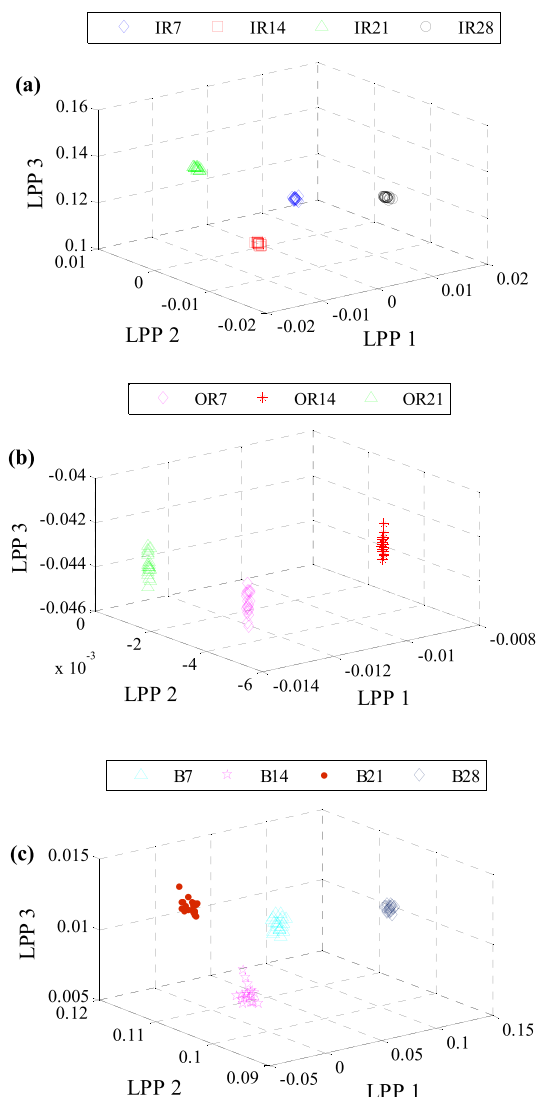


FIGURE 14. 3D feature distribution of the testing set in case 1 (0hp): (a) Inner race fault, (b) outer race fault, (c) ball fault.

is 100%, 98.33%, 98.33%, 98.33% and 100%, respectively. For fault severity identification of ball, the recognition rate obtained in five trials is 100%, 98.75%, 98.75%, 100% and 98.75%, respectively. Note that the average accuracy of five diagnosis results is taken as the final identification result. The detailed analysis results are displayed in Table 5. As shown in Table 5, the proposed approach can achieve average recognition accuracy of more than 98% under different loads. Moreover, the obtained classification coefficient is greater than 0.9, and average fuzzy entropy is less than 0.1. This indicates that the proposed approach has an excellent fault recognition performance and further verifies the effectiveness of the proposed method.

**E. COMPARISON BETWEEN THE PROPOSED METHOD AND TRADITIONAL SINGLE-STEP APPROACH**

In order to further validate the effectiveness of the proposed method, more comparisons with traditional single-stage

approach are performed in this section. Within traditional single-stage approach with dimension reduction, hybrid-domain features of each sample are first extracted, and then LPP is utilized to obtain the meaningful low-dimensional sensitive features from hybrid-domain features. Finally, low-dimensional feature dataset is fed into FCM classifier for fault recognition. It is worthy of note that dimension size of fault feature of single-stage approach is the same as that of the proposed method in this paper. The comparative aspect includes the CPU computation time and average diagnostic accuracy. Besides, the overall accuracy of recognition rate of the proposed approach is calculated according to the following weighted formula:

$$P = 1 - \frac{\alpha e_1 + \beta e_2 + \gamma e_3}{N} \tag{18}$$

where  $e_1, e_2$  and  $e_3$  are identification error of the three stages, respectively.  $\alpha, \beta$  and  $\gamma$  respectively are the number of testing sample of the three stages, and  $N$  is the total number of testing sample in this paper.

The comparison results of different methods are listed in Table 6. As is clear in Table 6, the proposed method (multi-stage diagnosis approach) can obtain recognition rate of 99% above for different load conditions, whereas diagnostic accuracy obtained by single-stage approaches with or without LPP is kept below 96%, which indicates that the presented multi-stage fault diagnosis approach can achieve higher identification accuracy than traditional single-stage approach mentioned in this paper. That is to say, the proposed multi-stage diagnosis approach is promising for fault identification of rolling bearing under various loads conditions. Note that this experimental study was carried out at a constant speed. However, the effectiveness and applicability of the proposed method at variable speed are unknown for us. Therefore, in the future work, it is valuable to introduce the proposed method into the bearing fault diagnosis under variable rotation speed in practical engineering, which is regarded as the focus of our future research.

**V. CONCLUSION**

In this paper, a neoteric multi-stage hybrid fault diagnosis scheme is proposed for identifying different health conditions of rolling bearing, which is divided into three stages (i.e. fault initial detection, fault type recognition and fault severity assessment). Initially, PE-based statistical analysis is used to achieve fault initial detection. Then, hybrid-domain features are constructed and combined with LPP for reflecting more completely intrinsic property of raw vibration signal and removing redundant information. Finally, the obtained low-dimensional dataset are fed into FCM classifier for identifying bearing fault type and fault severity. Contribution and novelty of this paper are summarized as follows:

- (1) Hybrid-domain features are extracted to reveal intrinsic characteristics of bearing fault signal more comprehensively.
- (2) A multi-stage diagnosis strategy for rolling bearing is proposed for health status identification.

(3) The performance of the presented approach is validated on experimental data under various load conditions.

It turned out that our approach is suitable and efficient in estimating bearing operating condition and identifying multi-class fault of bearing. Besides, our approach outperforms traditional single-stage diagnosis technique. However, ability of the proposed approach is unknown for health condition recognition of bearing under variable speed. For the future work, we intend to apply this method to solve multi-class fault identification issue under variable speed, including our own experimental data.

## ACKNOWLEDGMENT

The authors would like to appreciate CWRU for their kind permission to use their bearing data.

## REFERENCES

- Z. Wang, J. Zhou, J. Wang, W. Du, J. Wang, X. Han, and G. He, "A novel fault diagnosis method of gearbox based on maximum kurtosis spectral entropy deconvolution," *IEEE Access*, vol. 7, pp. 29520–29532, 2019.
- X. Yan and M. Jia, "Application of CSA-VMD and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 122, pp. 56–86, May 2019.
- L. Cui, J. Huang, and F. Zhang, "Quantitative and localization diagnosis of a defective ball bearing based on vertical-horizontal synchronization signal analysis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 11, pp. 8695–8706, Nov. 2017.
- J. Tian, Y. Ai, C. Fei, M. Zhao, F. Zhang, and Z. Wang, "Fault diagnosis of intershaft bearings using fusion information exergy distance method," *Shock Vib.*, vol. 2018, Aug. 2018, Art. no. 7546128.
- J. Tian, Y. T. Ai, C. W. Fei, F. L. Zhang, and Y. S. Choy, "Dynamic modeling and simulation of inter-shaft bearings with localized defects excited by time-varying displacement," *J. Vib. Control*, vol. 25, no. 8, pp. 1436–1446, Jan. 2019.
- J. Xiang, Y. Lei, Y. Wang, Y. He, C. Zheng, and H. Gao, "Structural dynamical monitoring and fault diagnosis," *Shock Vib.*, vol. 2015, Jul. 2015, Art. no. 193831.
- X. Zhang, Y. Liang, and J. Zhou, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, Jun. 2015.
- X. Yan and M. Jia, "Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection," *Knowl.-Based Syst.*, vol. 163, pp. 450–471, Jan. 2019.
- 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 Mach. Theory*, vol. 78, no. 16, pp. 187–200, 2014.
- Y. Li, M. Xu, R. Wang, and W. Huang, "A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy," *J. Sound Vib.*, vol. 360, pp. 277–299, Jan. 2016.
- Z. Liu, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Appl. Acoust.*, vol. 89, no. 3, pp. 16–27, Mar. 2015.
- L. Wang and J. Xiang, "A two-stage method using spline-kernelled chirplet transform and angle synchronous averaging to detect faults at variable speed," *IEEE Access*, vol. 7, pp. 22471–22485, 2019.
- Y. Wang, Y. Si, B. Huang, and Z. Lou, "Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008–2017," *Can. J. Chem. Eng.*, vol. 96, no. 10, pp. 2073–2085, May 2018.
- Q. Gao, H.-S. Tang, J.-W. Xiang, and Y. Zhong, "A multi-sensor fault detection strategy for axial piston pump using the Walsh transform method," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 4, Apr. 2018, Art. no. 1550147718772531.
- C. Li, J. Zheng, H. Pan, J. Tong, and Y. Zhang, "Refined composite multivariate multiscale dispersion entropy and its application to fault diagnosis of rolling bearing," *IEEE Access*, vol. 7, pp. 47663–47673, 2019.
- 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, 2017.
- Y. Ai, J.-Y. Guan, C.-W. Fei, J. Tian, and F.-L. Zhang, "Fusion information entropy method of rolling bearing fault diagnosis based on n-dimensional characteristic parameter distance," *Mech. Syst. Signal Process.*, vol. 88, pp. 123–136, May 2017.
- C. W. Fei, Y. S. Choy, C. G. Bai, and W. Z. Tang, "Multi-feature entropy distance approach with vibration and acoustic emission signals for process feature recognition of rolling element bearing faults," *Struct. Health Monit.*, vol. 17, no. 2, pp. 156–168, Mar. 2018.
- X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, Nov. 2018.
- Z. Feng, M. Liang, and F. Chu, "Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples," *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 165–205, Jul. 2013.
- L. Xiang and X. Yan, "A self-adaptive time-frequency analysis method based on local mean decomposition and its application in defect diagnosis," *J. Vib. Control*, vol. 22, no. 4, pp. 1049–1061, Mar. 2016.
- X. Yan and M. Jia, "Improved singular spectrum decomposition-based 1.5-dimensional energy spectrum for rotating machinery fault diagnosis," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 41, p. 50, Jan. 2019.
- Z. Wang, J. Wang, W. Cai, J. Zhou, W. Du, J. Wang, G. He, and H. He, "Application of an improved ensemble local mean decomposition method for gearbox composite fault diagnosis," *Complexity*, vol. 2019, Art. no. 1564243.
- K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
- Y. Wang, R. Markert, J. Xiang, and W. Zheng, "Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system," *Mech. Syst. Signal Process.*, vols. 60–61, pp. 243–251, Aug. 2015.
- M. Zhang, Z. Jiang, and K. Feng, "Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump," *Mech. Syst. Signal Process.*, vol. 93, pp. 460–493, Sep. 2017.
- Z. Li, J. Chen, Y. Zi, and J. Pan, "Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive," *Mech. Syst. Signal Process.*, vol. 85, pp. 512–529, Feb. 2017.
- M. Yonghao, Z. Ming, and L. Jing, "Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition," *ISA Trans.*, vol. 84, pp. 82–95, Jan. 2019.
- Z. Wang, J. Wang, and W. Du, "Research on fault diagnosis of gearbox with improved variational mode decomposition," *Sensors*, vol. 18, no. 10, p. 3510, Oct. 2018.
- H. Liu and J. Xiang, "A strategy using variational mode decomposition, L-kurtosis and minimum entropy deconvolution to detect mechanical faults," *IEEE Access*, vol. 7, pp. 70564–70573, 2019.
- C. Liu, Y. Wu, and C. Zhen, "Rolling bearing fault diagnosis based on variational mode decomposition and fuzzy C means clustering," *Proc. CSEE*, vol. 35, no. 13, pp. 3358–3365, Jul. 2015.
- N. Li, S. Guo, and Y. Wang, "Weighted preliminary-summation-based principal component analysis for non-Gaussian processes," *Control Eng. Pract.*, vol. 87, pp. 122–132, Jun. 2019.
- Z. Liu, W. Guo, J. Hu, and W. Ma, "A hybrid intelligent multi-fault detection method for rotating machinery based on RSGWPT, KPCA and Twin SVM," *ISA Trans.*, vol. 66, pp. 249–261, Jan. 2017.
- Z. Su, B. Tang, Z. Liu, and Y. Qin, "Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine," *Neurocomputing*, vol. 157, pp. 208–222, Jun. 2015.
- G. Chen, F. Liu, and W. Huang, "Sparse discriminant manifold projections for bearing fault diagnosis," *J. Sound Vib.*, vol. 399, pp. 330–344, Jul. 2017.
- F. Li, J. Wang, B. Tang, and D. Tian, "Life grade recognition method based on supervised uncorrelated orthogonal locality preserving projection and K-nearest neighbor classifier," *Neurocomputing*, vol. 138, pp. 271–282, Aug. 2014.
- Y. Li, D. Luo, and S. Liu, "Orthogonal discriminant linear local tangent space alignment for face recognition," *Neurocomputing*, vol. 72, pp. 1319–1323, Jan. 2009.
- Q. Jiang, M. Jia, J. Hu, and F. Xu, "Machinery fault diagnosis using supervised manifold learning," *Mech. Syst. Signal Process.*, vol. 23, no. 7, pp. 2301–2311, 2009.

[39] X. Ding, Q. He, and N. Luo, "A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification," *J. Sound Vibrat.*, vol. 335, pp. 367–383, Jan. 2015.

[40] C. Fei, G.-C. Bai, W.-Z. Tang, and S. Ma, "Quantitative diagnosis of rotor vibration fault using process power spectrum entropy and support vector machine method," *Shock Vib.*, vol. 2014, Mar. 2014, Art. no. 957531.

[41] C.-W. Fei and G.-C. Bai, "Wavelet correlation feature scale entropy and fuzzy support vector machine approach for aeroengine whole-body vibration fault diagnosis," *Shock Vib.*, vol. 20, no. 2, pp. 341–349, 2013.

[42] R. Yan, Y. Liu, and R. X. Gao, "Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines," *Mech. Syst. Signal Process.*, vol. 29, pp. 474–484, May 2012.

[43] X. Zhang and J. Zhou, "Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines," *Mech. Syst. Signal Process.*, vol. 41, nos. 1–2, pp. 127–140, Dec. 2013.

[44] Y. Gao, F. Vilecco, M. Li, and W. Song, "Multi-scale permutation entropy based on improved LMD and HMM for rolling bearing diagnosis," *Entropy*, vol. 19, no. 4, p. 176, 2017.

[45] C. Zhao and Z. Feng, "Application of multi-domain sparse features for fault identification of planetary gearbox," *Measurement*, vol. 104, pp. 169–179, Jul. 2017.

[46] X. Yan, M. Jia, and Z. Zhao, "A novel intelligent detection method for rolling bearing based on IVMD and instantaneous energy distribution-permutation entropy," *Measurement*, vol. 130, pp. 435–447, Dec. 2018.



**YING LIU** was born in 1965. She received the B.S. degree in chemical machinery and equipment from Jiangsu Chemical Engineering Institute (currently Changzhou University), Changzhou, China, in 1988, and the M.S. and Ph.D. degrees in mechanical engineering from Southeast University, Nanjing, China, in 1998 and 2006, respectively. She is currently a Full Professor with the School of Mechatronics Engineering, Nanjing Forestry University, Nanjing. Her research interests include wood non-destructive testing, artificial intelligence, and pattern recognition.



**MINPING JIA** was born in 1960. He received the B.S. and M.S. degrees from Nanjing Institute of Technology (currently Southeast University), Nanjing, China, in 1982 and 1985, respectively, and the Ph.D. degree from Southeast University, Nanjing, in 1991, all in mechanical engineering, where he is currently a Full Professor with the School of Mechanical Engineering. His research interests include dynamic signal processing, mechanical fault diagnosis, and vibration engineering applications.



**XIAOAN YAN** was born in 1989. He received the M.S. degree in mechanical engineering from North China Electric Power University, Baoding, China, in 2015, and the Ph.D. degree in mechanical engineering from Southeast University, Nanjing, China, in 2019. He is currently a Lecturer with the School of Mechatronics Engineering, Nanjing Forestry University, Nanjing. His research interests include signal processing, intelligent fault diagnosis, pattern recognition, health status assessment, and residual life prediction.



**YINLONG ZHU** received the B.S. degree from Yangzhou University, China, in 2002, and the M.S. and Ph.D. degrees from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2007 and 2012, respectively, all in mechanical engineering. He is currently an Associate Professor with the School of Mechatronics Engineering, Nanjing Forestry University, China. His current research interests include smart materials, soft robotics, and mechatronic system control.

...