

Hierarchical Diversity Entropy for the Early Fault Diagnosis of Rolling Bearing

Xianzhi Wang

Northwestern Polytechnical University

Shubin Si

Northwestern Polytechnical University

Yongbo Li (✉ yongbo@nwpu.edu.cn)

Northwestern Polytechnical University

Research Article

Keywords: Hierarchical diversity entropy, feature extraction, random forest, intelligent fault diagnosis

Posted Date: March 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-202027/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nonlinear Dynamics on February 15th, 2022. See the published version at <https://doi.org/10.1007/s11071-021-06728-1>.

Hierarchical diversity entropy for the early fault diagnosis of rolling bearing

Xianzhi Wang^a, Shubin Si^a, Yongbo Li^{b,*}

a School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, China.

b School of Aeronautics, Northwestern Polytechnical University, Xi'an, 710072, China.

Tel.: +86118809183733.

Fax: +0-000-000-0000.

E-mail address: yongbo@nwpu.edu.cn.

Abstract: Intelligent fault diagnosis provides great convenience for the prognostic and health management of the rotating machinery. Recently, the entropy-based feature extraction method has aroused researchers' attentions due to its independence with prior knowledge, unnecessary of preprocessing, and easy to perform. The multiscale diversity entropy has been proven to be a promising feature extraction method for the intelligent fault diagnosis. Compared to the existing entropy methods, the multiscale diversity entropy has advantages of high consistency, strong robustness and high calculation efficiency. However, the multiscale diversity entropy encounters the challenge to extract features for early fault diagnosis due to the weak fault symptoms and strong noise. This can be attributed to the multiscale diversity entropy only concerns the fault information embedded in the low frequency, which ignores the information hidden in the high frequency. To address this defect, the hierarchical diversity entropy (HDE) is proposed, which can synchronously extract fault information hidden in both high and low frequency. Based on HDE and random forest, a novel intelligent fault diagnosis frame has been proposed. The effectiveness of the proposed method has been evaluated through simulated and experimental bearing signals. The results show that the proposed HDE has the best feature extraction ability compare to multiscale sample entropy, multiscale permutation entropy, multiscale fuzzy entropy, and multiscale diversity entropy.

Keywords: Hierarchical diversity entropy, feature extraction, random forest, intelligent fault diagnosis

Nomenclature

SE	sample entropy
FE	fuzzy entropy
PE	permutation entropy
DE	diversity entropy
MSE	multiscale sample entropy
MPE	multiscale permutation entropy

MFE	multiscale fuzzy entropy
MDE	multiscale diversity entropy
HDE	hierarchical diversity entropy
RF	random forest

1. Introduction

Intelligent fault diagnosis provides great convenience for health monitoring of the rotating machinery. Since the intelligent fault diagnosis independent with expert knowledge, fast response, and low cost, it has received great graces from the modern industry. The intelligent fault diagnosis can be divided into three steps: data acquisition, feature extraction, and fault classification. Among the three steps, feature extraction is the most crucial step. The cluster ability of the feature extraction method will directly affect the final diagnosis result. Thanks to the rapid development of technology, various feature extraction methods have been proposed, such as FFT, wavelet transform, sparse decomposition. Among these measures, the entropy-based method has become a hot topic due to its advantages as follows:

- (1) Independent on prior knowledge such as bearing size, rotating speed, or empirical spectral knowledge;
- (2) The entropy-based methods can be directly used without any preprocessing technique such as de-noise, transformation, filter, or demodulation;
- (3) Few parameters to be set and easy to be implemented in practical applications.

The entropy theory originates from information entropy [1]. The information entropy is a physical quantity to quantify the complexity of information. Unfortunately, the entropy theory can be applied to calculate the dynamical complexity of arbitrary time series until the approximate entropy is proposed [2]. Until sample entropy (SE) has been proposed [3], entropy theory has been widely used in fault feature extraction, image texture recognition, human health diagnosis, etc. SE measures the dynamical complexity of the time series by calculating the ratio of new patterns appearing. Fuzzy entropy (FE) is an improvement of SE which owns more stability than SE [4]. FE employs fuzzy theory to count the state probability. Unlike SE and FE methods, permutation entropy (PE) utilizes the permutation of orbits to estimate the dynamical complexity, which brings a new sight for the

dynamical complexity analysis [5]. Above all, these commonly used entropy-based methods and their applications in fault diagnosis of rotating machinery have been summarized in Ref. [6]. Moreover, a comparison and benchmark study of these methods can be referred to Ref. [7].

Above-mentioned entropy methods measure the dynamical complexity from single scale, which is insufficient for the fault feature extraction. To solve this problem, Costa et al. proposed multiscale sample entropy (MSE) to measure the dynamical complexity through different scales [8]. In ref. [8], the raw signal is decomposed into several multiscale time series by the coarse-gaining procedure. Then the multiscale sample entropy value can be obtained by calculating the sample entropy value of each multiscale time series. The most significant improvement of the multiscale sample entropy is that the multiscale procedure can enrich the extracted features, which provides adequate useful information for fault classification. Benefiting from the multiscale theory, the multiscale sample entropy has been widely applied in the intelligent fault diagnosis of rotating machinery.

Zhang et al. used the multiscale sample entropy to extract the fault features from vibrational signals. Then, the adaptive neuro-fuzzy inference system was used to automatically recognize different bearing fault [9]. Pan et al. extracted the fault features of the induction motors by multiscale sample entropy. Then the mutual information was used to choose the first five most important features. Last, the selected features were put into the support vector machine to classify the motor fault [10]. Cui et al. proposed a fault diagnosis frame based on multiscale sample entropy and fuzzy support vector machine. The multiscale sample entropy was used to extract fault features of the switchgear, and the fuzzy support vector machine was used to identify different fault types of switchgear [11].

Inspired by multiscale sample entropy (MSE), the PE and FE are also extended into multiscale permutation entropy (MPE) and multiscale fuzzy entropy (MFE). Minhas et al. used the multiscale fuzzy entropy to extract fault features, then the support vector machine classifier was used to distinguish different bearing faults [12]. Zhu et al. combined the multiscale fuzzy entropy with the fractal theory to extract features. Then Laplacian support vector machine was utilized to identify the bearing faults [13].

Du et al. used the multiscale permutation entropy to extract fault features, linear discriminant analysis (LDA) was used to select the fault features, and the proposed

harmonic mean difference self-organizing fuzzy logic classifier was used to identify the bearing faults [14]. Tiwari et al. used the multiscale permutation entropy to extract fault features, then adaptive neuro fuzzy classifier was used to classify the bearing faults [15]. Yasir et al. used the local mean decomposition (LMD) to decompose the signal into separate product functions. Then MPE value of the product functions was used as an indicator to distinguish the bearing faults [16].

Recently, a novel entropy based method called multiscale diversity entropy (MDE) has been proposed [17]. Compare to the MFE, MSE, and MPE, the multiscale diversity entropy has advantages as follows:

(1) High consistency. The diversity entropy can estimate the dynamical complexity with high consistency for some certain systems, while PE, SE, and FE may yield undefined value.

(2) High calculation efficiency. The diversity entropy has the lowest calculation complexity, which saves at least half of time compared to other entropy-based methods.

(3) Strong robustness. The diversity entropy has the strongest robustness in FM signal, FM-AM signal and sinusoid signal test.

Essentially, the multiscale procedure used in MDE is a low pass filter based on the Haar wavelet [18]. However, the multiscale analysis has a certain defect that the coarse-gaining procedure only considers the fault information hidden in the low frequency, which ignores the fault information hidden in the high frequency. This induces the MDE extracting incomplete features, which offers insufficient fault information for the early fault diagnosis. When localized fault occurs at early stage, the fault symptoms are weak and the back ground noise is strong. Ignoring the fault information hidden in the high frequency, the MDE can hardly extract accurate fault information for the early fault diagnosis.

To make up such shortage, hierarchical diversity entropy (HDE) is proposed in this paper. The proposed HDE decomposes the raw signal into high frequency components and low frequency components, which can synchronously reveal the fault information hidden in both high and low frequency. Based on HDE and random forest classifier, a novel intelligent fault diagnosis frame has been proposed. The proposed intelligent fault diagnosis frame can be divided into two stages. First, use the HDE to extract fault features from the vibrational signals. Second, take the obtained features as the input of the RF classifier to recognize the fault types. The effectiveness of the

proposed method is evaluated through simulated and experimental signals. The results show that the proposed HDE has the best feature extraction ability compare to multiscale sample entropy, multiscale permutation entropy, multiscale fuzzy entropy, and multiscale diversity entropy.

The rest of this paper is organized as follows. Section 2 describes the basic definition of the proposed method. Section 3 illustrates the simulated bearing signals and evaluates the performance of the proposed method using the simulated signals. Section 4 evaluates the performance of the proposed method using the experimental bearing signals. Last, Section 5 summarizes the conclusions.

2. Methodology

2.1. Diversity Entropy

For an arbitrary time series $X=\{x_1, x_2, \dots, x_i, \dots, x_N\}$, $i \in [1, N]$, N indicates the data length. The diversity entropy (DE) can be computed by four steps.

Step 1. Reconstructed the time series X into a series of orbits with embedding dimension m as expressed in Eq. (1) [19].

$$\begin{aligned}
 Y(m) &= \{y_1(m), y_2(m), \dots, y_{N-m+1}(m)\} \\
 &= \begin{Bmatrix} x_1 & x_2 & \dots & x_m \\ x_2 & x_3 & \dots & x_{1+m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-m+1} & x_{N-m} & \dots & x_N \end{Bmatrix} \quad (1)
 \end{aligned}$$

Step 2. Calculate the cosine similarity $D(m)$ between the adjacent orbits to obtain a series of cosine similarities as Eq. (2). The cosine similarity between two orbits is defined as expressed in Eq. (3).

$$\begin{aligned}
 D(m) &= \{d_1, \dots, d_{N-m}\} \\
 &= \{d(y_1(m), y_2(m)), d(y_2(m), y_3(m)), \dots, d(y_{N-m-1}(m), y_{N-m}(m))\} \quad (2)
 \end{aligned}$$

$$d(y_i(m), y_j(m)) = \frac{\sum_{k=1}^m y_i(k) \times y_j(k)}{\sqrt{\sum_{k=1}^m y_i(k)^2} \times \sqrt{\sum_{k=1}^m y_j(k)^2}} \quad (3)$$

Noted that the range of the cosine similarity is $[-1, 1]$. The defined cosine similarity d is a quantity of similarity between two orbits, which measures the cosine of the angle in geometry. A big cosine similarity value indicates a similar, predictable, or periodic dynamical change between two orbits. Oppositely, a small cosine similarity value represents a diversity, stochastic, or chaotic dynamic behavior.

Step 3. Partition the scope $[-1, 1]$ into ε intervals. Then, count the state probability $(P_1, P_2, \dots, P_\varepsilon)$ that represents the probability of cosine similarities falling into

each interval. Obviously, it can be concluded $\sum_{k=1}^{\varepsilon} P_k = 1$.

Step 4. The DE can be calculated based on the obtained state probability as expressed in Eq. (4).

$$DE(m, \varepsilon) = -\frac{1}{\ln \varepsilon} \sum_{k=1}^{\varepsilon} P_k \ln P_k \quad (4)$$

where the m represents the embedding dimension. The ε represents the number of symbols.

The physical meaning of DE is the expected coding length encoded with ε symbols. Since the state probability is calculated from cosine similarity, the DE can be also regarded as the expectation of the diversity between the orbits. From the perspective of the information entropy theory [1], the DE is monotone increasing with the scope of $[0, 1]$. Physically, when DE tends to 0, the time series is treated with a low complexity that represents the dynamical system having a constant, deterministic, or periodic phenomenon. When DE tends to 1, it represents the dynamical system containing high complexity with a chaotic, stochastic, or irregular phenomenon.

2.2. Multiscale Diversity Entropy

The coarse graining process contains two main calculation steps: first, conduct multiple scale time series by coarse graining process; second, calculate DE value of each coarse-grained time series. The detailed steps are listed below:

Step 1. For a given time series $X = \{x_1, x_2, \dots, x_i, \dots, x_N\}$, $i \in [1, N]$, N indicates the data length., segment it into multiple scale time series:

$$y_j^\tau = \frac{1}{\tau} \sum_{i=j}^{j+\tau-1} x_i \quad 1 < j < N - \tau + 1 \quad (5)$$

where τ represents the scale factor in the multiscale analysis, which should be a positive integer. The scale factor τ aims to quantify the dynamical characteristics of time series over different scales. For $\tau=1$, the time series $\{y^{(1)}\}$ is the original time series.

Step 2. Calculate the entropy value of each multiple scale time series.

$$MDE(X, m, \varepsilon) = DE(y_j^\tau, m, \varepsilon) \quad (6)$$

where the m represents the embedding dimension. The ε represents the number of symbols.

2.3. Hierarchical Diversity Entropy

Essentially, the multiscale procedure used in MDE is a low pass filter based on the Haar wavelet [18]. However, the multiscale analysis has a certain defect that the coarse-gaining procedure only considers the fault information hidden in the low frequency, which ignores the fault information hidden in the high frequency. This induces the MDE extracting incomplete features, which offers insufficient fault information to achieve an accurate diagnosis result. To make up such shortage, hierarchical diversity entropy (HDE) is proposed in this paper. The proposed HDE decomposes the raw signal into high frequency components and low frequency components, which can synchronously reveal the fault information hidden in both high and low frequency. The proposed HDE can be calculated from the following steps:

Step 1: For arbitrary time series $X=\{x_1, x_2, \dots, x_i, \dots, x_N\}$, $i \in [1, N]$, N indicates the data length. Define the low frequency component operator Q_0 and high frequency component operator Q_1 as Eq. (7) and Eq. (8), respectively:

$$Q_0(x) = \frac{x_i + x_{i+1}}{2} \quad i = 1, 2, \dots, N-1 \quad (7)$$

$$Q_1(x) = \frac{x_i - x_{i+1}}{2} \quad i = 1, 2, \dots, N-1 \quad (8)$$

Step 2: Using a matrix to illustrate the operators Q_j ($j=0$ or 1), that can be:

$$Q_j^n = \begin{bmatrix} \frac{1}{2} & \frac{(-1)^j}{2} & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{(-1)^j}{2} & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & \frac{1}{2} & \frac{(-1)^j}{2} \end{bmatrix} \quad (9)$$

Step 3: The hierarchical components $X_{k,e}$ for the k th layer can be obtained as Eq. (10) by using the operators in step 2 repeatedly.

$$X_{k,e} = Q_{r_k}^k \bullet Q_{r_{k-1}}^{k-1} \bullet \dots \bullet Q_{r_1}^1 \bullet x \quad (10)$$

where e is hierarchical node number. For $k \in N$, e can be expressed according to Eq. (11).

$$e = \sum_{m=1}^k 2^{k-m} r_m \quad (11)$$

where $[r_1, r_2, \dots, r_k]$ is the unique vector corresponding to the integer e , and

$\{r_m, m=1, \dots, k\} \in \{0,1\}$ denotes the averaging or differential operator at the m th layer.

Step 4: Repeat steps 1-3 for each channel until obtain all the hierarchical component. According to the definition of DE, the HDE can be obtained using Eq. (12).

$$HDE(X, k, e, m, \varepsilon) = DE(X_{k,e}, m, \varepsilon) \quad (12)$$

where the m represents the embedding dimension. The ε represents the number of symbols.

Since the averaging and differential processes are both used in the hierarchical decomposition procedure, the HPE can capture fault information from both low and high frequency component. The calculation process of the HPE is shown in Fig. 1(a). A diagrammatic sketch of low frequency component and high frequency component is shown in Fig. 1 (b).

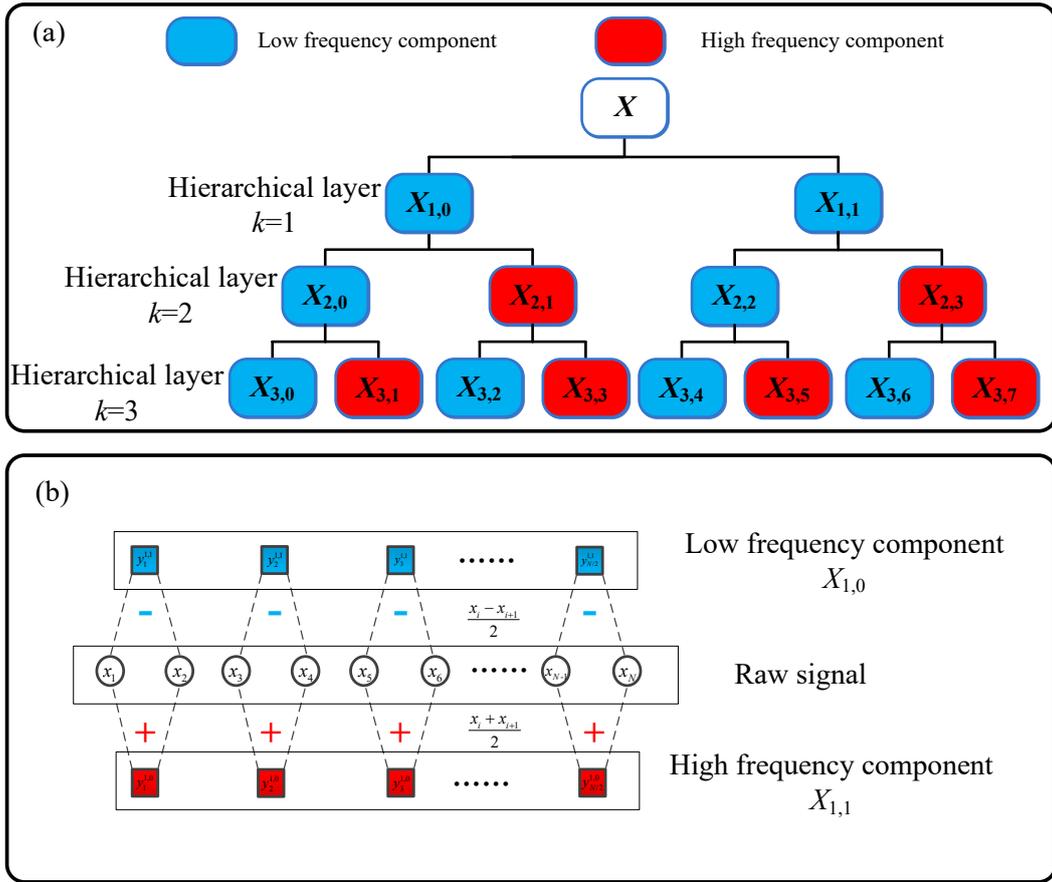


Fig. 1. The diagrammatic sketch of hierarchical decomposition: (a) hierarchical nodes; (b) low frequency component and high frequency component.

2.4. Random forest classifier

Random forest (RF) classifier is a commonly used ensemble learning classifier which has the advantage of high calculation efficiency, easy implementation, and strong resistance to overfitting, etc. The calculation of the RF can be summarized in four steps:

Step 1. Bagging repeatedly selects a random sample with replacement of the training set to construct K subsets. In each subset, randomly choose $s \leq \tau$ features.

Step 2. Construct L decision trees using the s features.

Step 3. The final classification result is determined by the vote of L decision trees.

The diagrammatic sketch of random forest can be seen in Fig. 2. In this paper, the parameters of RF are set according to Ref. [20]. The number of trees is set as 50. Out-of-bag samples are selected to estimate the generalization accuracy. "gini" is applied as the split criterion for each tree. The minimum number of samples required

to split is set to be 2. The minimum number of samples required to be at a leaf node is set to be 1.

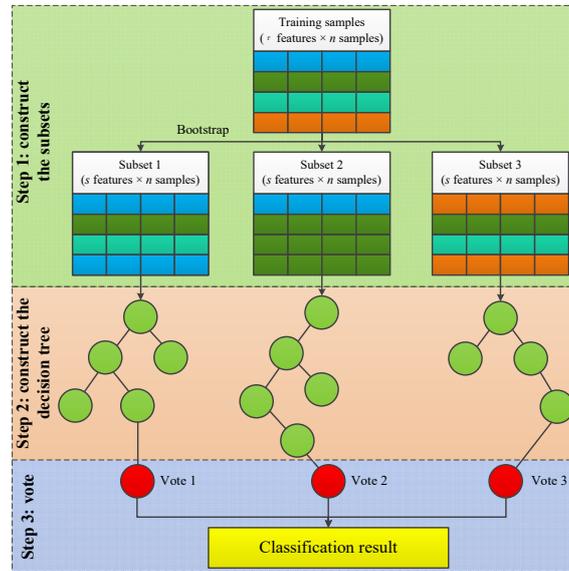


Fig. 2. The diagrammatic sketch of random forest

2.5. HDE and RF based intelligent fault diagnosis

The proposed intelligent fault diagnosis frame can be divided into two stages. First, use the HDE to extract fault features from the vibrational signals. Second, take the obtained features as the input of the RF classifier to recognize the different fault types. There are five steps as follows:

Step 1. Collect the vibrational signals under different health conditions.

Step 2. The HDE is employed to extract the fault features from the measured vibrational signals.

Step 3. Divide the obtained HDE features into the training set and testing set.

Step 4. Use the training set to train the RF classifier.

Step 5. Use the testing set to test the trained RF classifier to recognize the different fault types

3. Simulation evaluation

In this section, three types of simulated bearing fault signals are designed to evaluate the effectiveness of the proposed HDE: outer race fault, inner race fault and ball fault. The simulated rotating speed is 3000 rpm. The sampling frequency is 10240 Hz. The simulated bearing type is N205 cylindrical roller bearing. The bearing size is

listed in Table. 1. The diagrammatic sketch of the simulated bearing is shown in Fig. 3.

Parameter	Value
Pitch circle diameter	35.5 mm
Roller diameter	6.5 mm
Rotating speed	3000 rpm
Number of rollers	12
Sample frequency	10240 Hz
Natural frequency of bearing	4000 Hz
Contact angle	0°

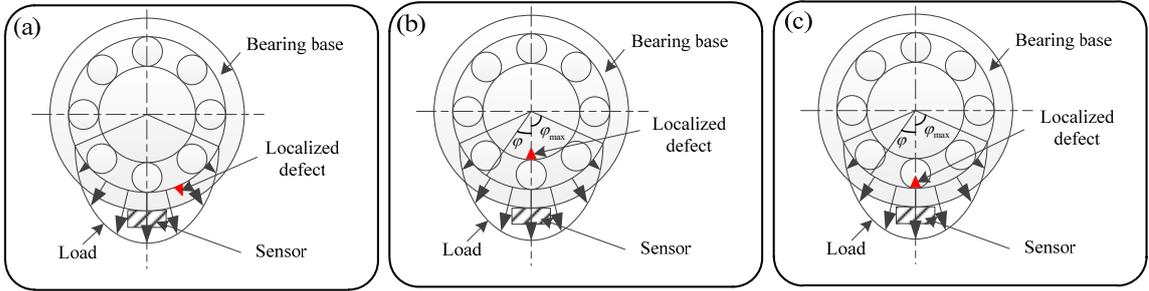


Fig. 3. The diagrammatic sketch of the simulated bearing: (a) outer race fault; (b) inner race fault; (c) ball fault.

3.1. Simulated bearing signals

(1) Localized defect on the outer race

The fault model of the outer race fault is shown in Fig. 3(a). Assume that the pitting defect occurs in the load zone of the bearing. The sensor is mounted in the maximum load density of the load zone. Since the position of the localized defect does not vary with time, the impulsive force is an ideal force. Assuming that at time $t = 0$, the damage point starts contacting with the rolling element, then the impulsive force excited by the localized damage on the outer ring during rotation can be expressed by Eq. (13).

$$\Delta_o(t) = \sum_{t=-\infty}^{+\infty} d_o \delta(t - kT_o) \quad (13)$$

where the d_o indicates the intensity of the impulsive force. The $\delta(t)$ represents the unit impulse function. The k indicates the number of impulses. The $T_o = 1/f_o$ is the interval between two impulses and the f_o is the fault characteristic frequency of the outer race fault.

The damp function generated by the impulsive force is expressed as Eq. (14).

$$e(t) = \begin{cases} \exp(-t/T), & t > 0 \\ 0 & t \leq 0 \end{cases} \quad (14)$$

Thus, the simulated outer race fault signal can be expressed as Eq. (15).

$$v_o(t) = \sum_{t=-\infty}^{+\infty} d_o \delta(t - kT_o) * e(t) \quad (15)$$

The time wave of the simulated outer race fault and the corresponding spectrum are plotted in Fig. 4 (a) and (b) respectively. To simulate the practical working condition, the simulated outer race fault is added with white Gaussian noise, and the signal noise ratio is 10 dB. Time wave of the simulated outer race fault with added white Gaussian noise and the corresponding spectrum are plotted in Fig. 4 (c) and (d) respectively.

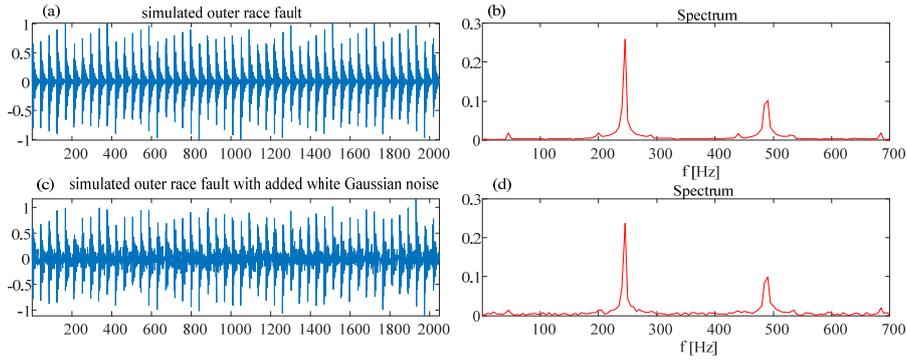


Fig. 4. Time wave of the simulated outer race fault; (b) Spectrum of the simulated outer race fault; (c) Time wave of the simulated outer race fault with added white Gaussian noise; (d) Spectrum of the simulated outer race fault with added white Gaussian noise.

(2) Localized defect on the inner race

The fault model of the inner race fault is shown in Fig. 3(b). Suppose at time $t=0$, the first impulse is generated when the rolling element contacting with the localized defect on inner ring at the peak of the load zone. As the localized defect rotates with the rotation of the inner race, the contacting location of the rolling element in contact with the localized defect will vary. The impulsive force will be generated only when such contact occurs in the load zone. The impulsive force excited by the localized defect on the inner ring during rotation can be expressed as Eq. (16).

$$\Delta_i(t) = \sum_{t=-\infty}^{+\infty} d_i \delta(t - kT_i) \quad (16)$$

where the d_i indicates the intensity of the impulsive force. The $\delta(t)$ represents the unit impulse function. The k indicates the number of impulses. The $T_i=1/f_i$ is the interval between two impulses and the f_i is the fault characteristic frequency of the inner race fault.

Fig. 3(b) shows the radial load distribution. The load distribution is defined as Eq. (17).

$$q(\varphi) = q_{\max} \left[1 - \frac{1}{2\sigma} (1 - \cos \varphi) \right]^n \quad (17)$$

where the σ is the load distributing coefficient. In this paper $\sigma=0.5$. $n=1.1$ for cylindrical roller bearing, $n=1.5$ for rolling ball bearing. In this paper $n=1.1$.

As shown in Fig. 3(b), when the localized defect contacts with the rolling element at φ , the impulsive force acquired by the sensor is the projection of the impulsive force along the axis $\varphi=0$. Therefore, the influence coefficient of localized defect is shown in Eq. (18).

$$p(\varphi) = \cos \varphi \quad (18)$$

Let $\varphi=2\pi f_r t$, where the f_r is the rotating frequency. Thus the impulsive force on the axis of the sensor can be written as Eq. (19).

$$f_i(t) = \Delta_i(t) q(2\pi f_r t) p(2\pi f_r t) \quad (19)$$

Thus, the simulated inner race fault signal can be expressed as Eq. (20).

$$v_i(t) = A_i [\Delta_i(t) q(2\pi f_r t) p(2\pi f_r t)] * e(t) \quad (20)$$

The time wave of the simulated inner race fault and the corresponding spectrum are plotted in Fig. 5 (a) and (b) respectively. To simulate the practical working condition, the simulated inner race fault is added with white Gaussian noise, and the signal noise ratio is 10 dB. Time wave of the simulated inner race fault with added white Gaussian noise and the corresponding spectrum are plotted in Fig. 5 (c) and (d) respectively.

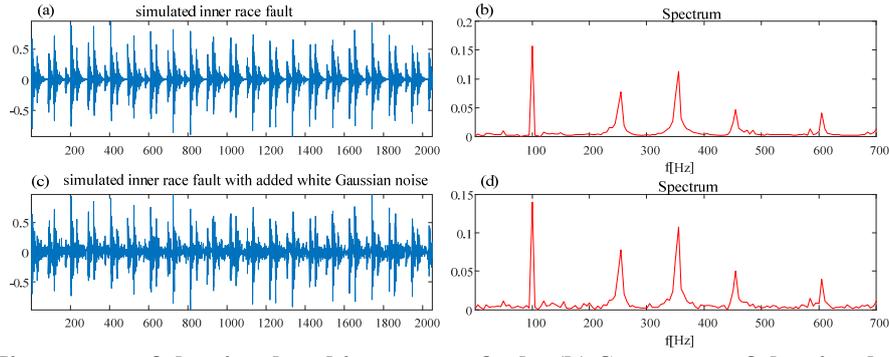


Fig. 5. Time wave of the simulated inner race fault; (b) Spectrum of the simulated inner race fault; (c) Time wave of the simulated inner race fault with added white Gaussian noise; (d) Spectrum of the simulated inner race fault with added white Gaussian noise.

(3) Localized defect on the rolling element

The fault model of the ball fault is shown in Fig. 3(c). Suppose at time $t = 0$, the first impulse is generated when the localized defect on the rolling element contacting with the outer ring at the peak of the load zone. When the rolling element rotates with the shaft, the impulses are generated from the localized defect on the rolling element continuously contacting with the outer ring and inner ring. Therefore, the impulsive force generated by the rolling element during rotation can be expressed as Eq. (21).

$$\Delta_b(t) = \Delta_{bo}(t) + \Delta_{bi}(t) \quad (21)$$

$$\Delta_{bo}(t) = \sum_{t=-\infty}^{+\infty} d_{bo} \delta(t - kT_b) \quad (22)$$

$$\Delta_{bi}(t) = \sum_{t=-\infty}^{+\infty} d_{bi} \delta(t - kT_b - 0.5 * T_b) \quad (23)$$

where the d_{bo} indicates the intensity of the impulsive force of the rolling element contacting with the outer race, and the d_{bi} indicates the intensity of the impulsive force of the rolling element contacting with the inner race. The $T_b = 1/f_b$ is the spin period of the rolling element and the f_o is the fault characteristic frequency of the ball fault.

The hypothesis of the influence of load is the same as the inner race fault. Thus, the simulated ball fault signal can be expressed as Eq. (24).

$$v_b(t) = A_b [\Delta_{bo}(t)q(2\pi f_r t)p(2\pi f_r t)]^* e(t) + A_b [\Delta_{bi}(t)q(2\pi f_r t)p(2\pi f_r t)]^* e(t) \quad (24)$$

The time wave of the simulated ball fault and the corresponding spectrum are plotted in Fig. 6 (a) and (b) respectively. To simulate the practical working condition, the simulated ball fault is added with white Gaussian noise, and signal noise ratio is 10 dB. Time wave of the simulated ball fault with added white Gaussian noise and the corresponding spectrum are plotted in Fig. 6 (c) and (d) respectively.

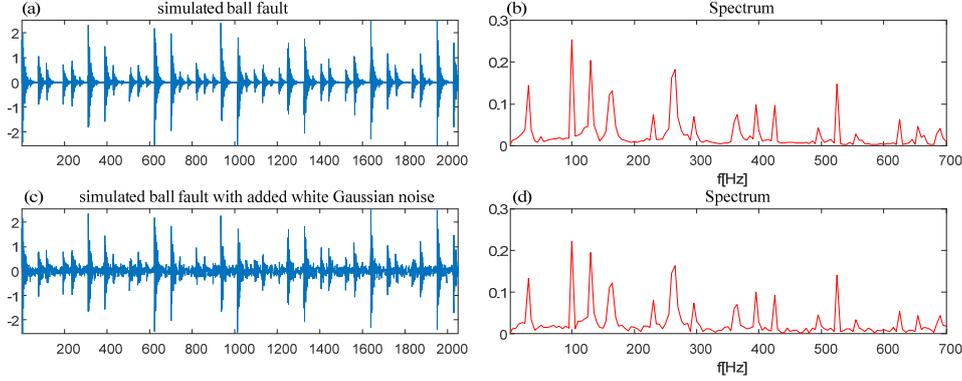


Fig. 6. Time wave of the simulated ball fault; (b) Spectrum of the simulated ball fault; (c) Time wave of the simulated ball fault with added white Gaussian noise; (d) Spectrum of the simulated ball fault with added white Gaussian noise.

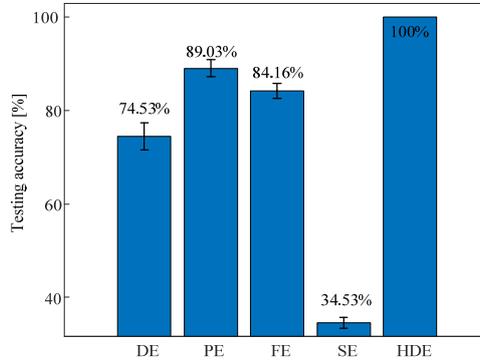
3.2. Simulation results and analysis

In this section, the three types of simulated signals are used to evaluate the effectiveness proposed HDE-RF method in feature extraction. The simulation evaluation procedure can be divided into five steps. First, generate the simulated signals as described in Section 3.1. For each fault type, a slide window with 2048 points is used to extract a slice from the raw signal as a sample. Each fault type contains 100 samples. Second, divide the dataset into training dataset and testing dataset. Each dataset contains 50 samples from each fault type, for a total of 150 samples. Third, the MFE, MSE, MPE, MDE and HDE are used to extract the fault features from both training dataset and testing dataset. The parameters of the five entropy-based methods are shown in Table. 2. Four, use the obtained features from the training dataset to train the random forest classifier. Last, put the obtained features from the testing dataset into the trained random forest classifier to distinguish the different bearing faults. The final testing accuracy can be regarded as the indicator to reflect the feature extraction ability of the entropy-based methods. A high testing accuracy indicates a better feature extraction ability. The final testing accuracy of the five methods is shown in Fig. 7.

Table 2. Parameters of the five entropy-based methods

Method	Embedding dimension m	Scale τ	Layer k	Tolerance r	Symbols ε
MSE	2	31	-	0.15	-
MFE	2	31	-	0.15	-
MPE	6	31	-	-	-
MDE	4	31	-	-	30
HDE	4	-	4	-	30

Noted that the embedding dimension, tolerance, and symbols are set as the best parameters which can be referred to Ref. [3], [4], [5], and [17]. To have a fair comparison, the scale $\tau=31$ and layer $k=4$. Thus, each entropy method will extract 31 features from one sample.

**Fig. 7. Testing accuracy of simulation evaluation**

In Fig. 7, the proposed HDE achieves the highest testing accuracy. To further analyze, the visualized features are plotted in Fig. 8. The visualized features is plotted by using t-SNE to reduce the entropy values into 2 dimensions [21]. In Fig. 8, the cluster ability shows the feature extraction ability of the method: the smaller inner-class distance among samples within the same cluster and the larger inter-class distance among clusters, the better the feature extraction ability. First, in Fig. 8 (e), the MFE is failed to distinguish the outer race fault and the inner race fault. The inter-class distance between the outer race fault and the inner race fault is too close, resulting in the overlapping of two cluster centers. Besides, the inner-class distance is too large, resulting in each class being too scattered to form into a cluster. Second, in Fig. 8 (a), for MDE, the situation of the outer race fault and the inner race fault is similar as the MFE. But the inter-class between ball fault and other classes is closer than MFE. Third, for Fig. 8 (d), for MSE, all the inter-class distances are too close, resulting in the overlapping of three cluster centers. Besides, the inner-class distances of each class are also too large. Four, in Fig. 8 (c), for MPE, the inner-class distance is

close enough to form into a cluster. However, the inter-class distance between the outer race fault and the inner race fault is still too close, resulting in the overlapping of these two clusters. Last, in Fig. 8 (b), for HDE, the inter-class distance between each class is far enough to distinguish the three types of faults, and the inner-class distance is small enough to have a clear cluster boundary.

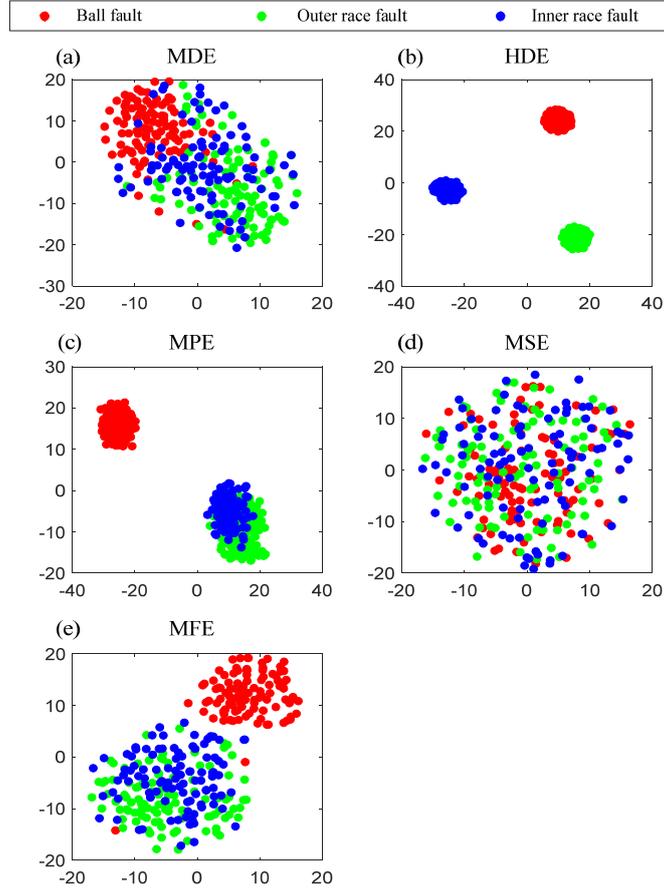


Fig. 8. Feature visualization of the simulated bearing: (a) MDE; (b) MFE; (c) MPE; (d) MSE; (e) HDE.

The better feature extraction ability of HDE can be attributed to the high frequency components enrich the extracted features. To have a clear vision, take the outer race fault as example, the time wave and the corresponding spectrum of the high frequency component and low frequency component are plotted in Fig. 9. From Fig. 9 (b) and (d), can be seen that the high frequency component exhibits different characteristics from low frequency component in both time and frequency domain. This indicates the high frequency component presents a different perspective which can enrich the extracted features. Furthermore, to show the contribution of the high frequency components to the feature extraction, an out-of-bag error method called the random forest importance is adopted [22]. The random forest importance of each

component is shown in Fig. 10. The higher the random forest importance value, the greater the component contributes to classification. In Fig. 10, first, some low frequency components contribute negatively to distinguish the three types of faults, for example, X_{10} . Second, some high frequency components contribute greater than some low frequency components, for example, X_{23} . This indicates the high frequency components can provide indispensable information in the feature extraction, which makes up for the incompleteness of low frequency components in feature extraction.

To conclude, the high frequency components of HDE provide indispensable information to distinguish three types of simulated bearing faults, which makes the HDE achieving the highest classification accuracy.

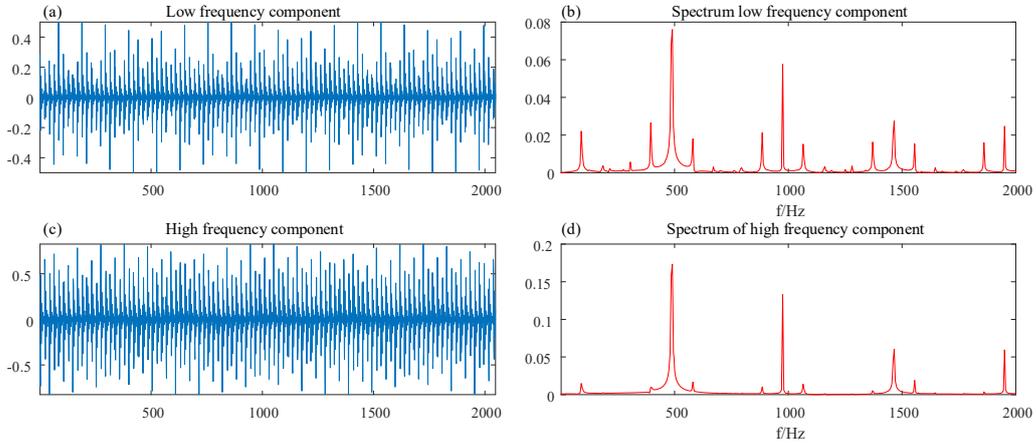


Fig. 9. (a) Time wave of the low frequency component; (b) Spectrum of the low frequency component; (c) Time wave of the high frequency component; (d) Spectrum of the high frequency component.

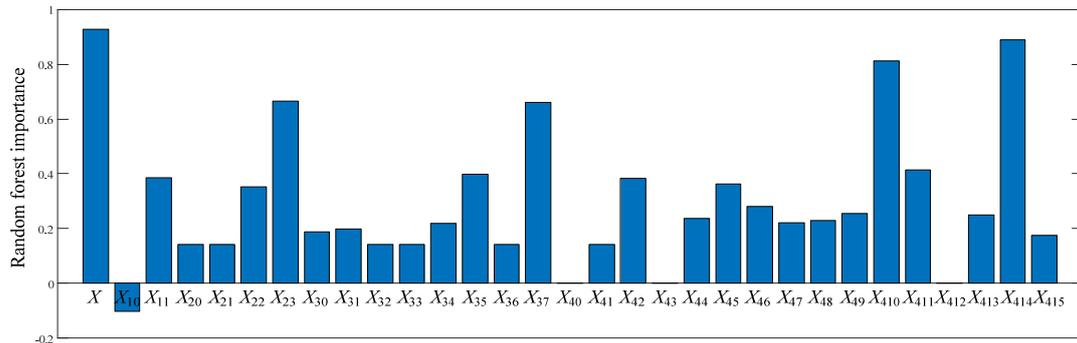


Fig. 10. Random forest importance of HDE.

4. Experiment evaluation

To evaluation the superiority of the proposed method in the early fault diagnosis of rolling bearing, the proposed HDE-RF method is used to distinguish different fault

types from the experimental vibration signals. For comparison, MDE, MFE, MSE and MPE are also used to extract fault features. The experiment procedure consists of five steps: first, collect the vibrational signals from the test rig. Second, divide the acquired signals into training dataset and testing dataset. Third, use the entropy-based method to extract the fault features from both training dataset and testing dataset. Four, use the obtained fault features from training dataset to train the random forest classifier. Last, use the fault features from testing dataset to test the trained random forest classifier. The testing accuracy of the random forest classifier can be regard as the indicator to evaluate the feature extraction ability of the entropy-based method. A high testing accuracy indicates a better feature extraction ability.

4.1. Experimental setting

The experiment was conducted on the test rig as shown in Fig. 11. The test rig consists of a motor, tachometer, test bearing, and magnetic damping. The rotating speed is 3000 rpm. The radial load was generated by the magnetic damping. The radial load is set as 5 N. The accelerometers were mounted on the bearing case. In this paper, the vertical accelerometer was used to collect vibration signals. The sampling frequency was 10240 Hz. The bearing type was N205 cylindrical roller bearing. The specific bearing size is shown in Table. 3. In this experiment, three types of faults were designed including inner race fault, outer race fault and ball fault. The different bearing fault was realized by replacing the testing bearing. The localized defects on the fault bearing were fabricated by electric discharge machining to simulate the pitting defect at early stage. The diameter of the artificial pitting defect was 1 mm and the depth was 0.2 mm. The fault bearings are shown in Fig. 12. The time wave and corresponding spectrums of the collected vibrational signals are shown in Fig. 13.

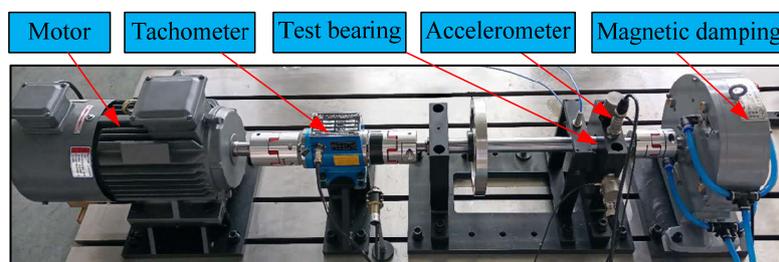


Fig. 11. The test rig

Table. 3. Bearing size of the test bearing

Parameter	Value
Bearing type	N205

Pitch diameter	35.5 mm
Roller diameter	6.5 mm
Rotating speed	3000 rpm
Number of rollers	12
Sample frequency	10240 Hz
Contact angle	0°

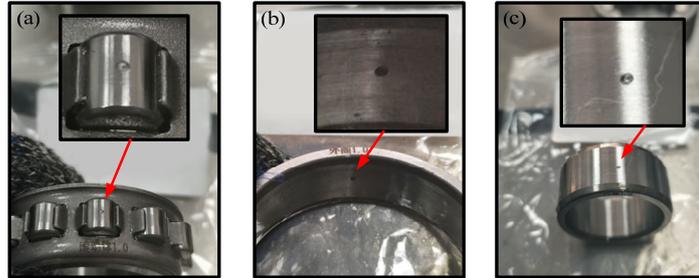


Fig. 12. Bearings with artificial pitting defect: (a) Ball fault; (b) Outer race fault; (c) Inner race fault.

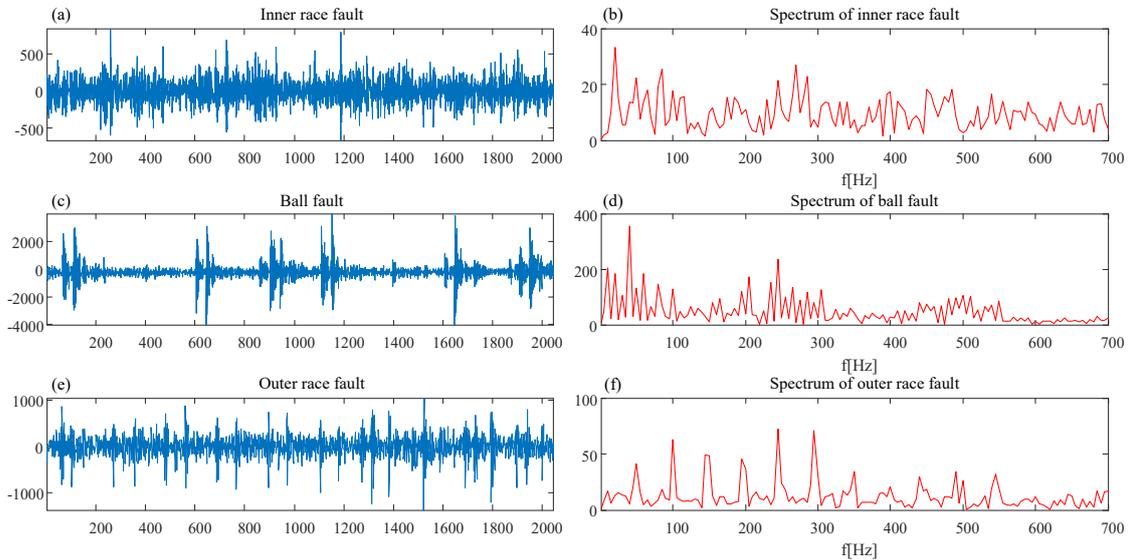


Fig. 13. (a) Time wave of the inner race fault; (b) Spectrum of the inner race fault; (c) Time wave of the ball fault; (d) Spectrum of the ball fault; (e) Time wave of the outer race fault; (f) Spectrum of the outer race fault.

4.2. Results and discussion

The testing accuracy of HDE, MDE, MSE, MFE and MPE are plotted in Fig. 14. Can be seen from Fig. 14, the proposed HDE achieves the highest testing accuracy with the smallest standard deviation. To further investigate, the visualized features of the experimental results by t-SNE are plotted in Fig. 15. First, in Fig. 15(c), for MPE, the features are formed into three clusters. However, the smaller inter-class distance and larger inner-class distance result in the overlapping of the samples near the cluster boundary. This induces a blurring boundary between the inner race fault and outer

race fault, so does the outer race fault and ball fault. Second, in Fig. 15(e), for MFE, the inter-class distance between the inner race fault and outer race fault is smaller than MPE. This results in the overlapping of the two clusters, which makes it difficult to distinguish these two types of fault. Third, in Fig. 15(d), for MSE, the situation is even worse. The inter-class distances between each class are too close, causing the three types of faults mixing together. Fourth, in Fig. 15(a), for MDE, the situation is similar as MPE. Although the inner-class distance is closer than MPE, the inter-class distance is still not far enough to have a nice boundary. Last, in Fig. 14 (b), for HDE, the inter-class distance between each class is far enough to distinguish the three types of faults, and the inner-class distance is small enough to establish cluster boundary.

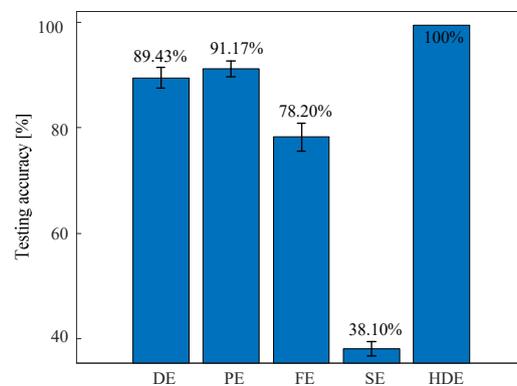


Fig. 14. Testing accuracy of experiment evaluation.

This can be attributed to the high components of HDE, which offers the indispensable fault information to enrich the extracted features. To show the contribution of the high frequency components, the random forest importance of each component is shown in Fig. 16. The higher the random forest importance value, the greater the component contributes to classification. In Fig. 16, first, some low frequency components contribute negatively to distinguish the three types of faults, for example, X_{46} . Second, some high frequency components contribute greater than some low frequency components, for example, X_{23} . This indicates the high frequency components can provide indispensable information in the feature extraction, which makes up for the incompleteness of low frequency components in feature extraction.

Besides, it can be found that the experimental results coincide with the simulation results. First, for MDE, MPE and MFE, they are both difficult to distinguish the inner race fault and outer race fault of the N205 cylindrical roller bearing. Second, the testing accuracies of the MDE, MPE and MFE are almost the

5. Conclusion

A novel feature extraction method called the hierarchical diversity entropy (HDE) is developed for the early fault diagnosis of rolling bearing. The HDE can synchronously extract fault information hidden in both high and low frequency, which enriches the fault features and provide indispensable fault information. Based on HDE and RF, a novel intelligent fault diagnosis frame is proposed to automatically recognize the bearing faults. Both simulated and experimental signal demonstrate that HDE performs the best in classifying bearing faults compare to MSE, MFE, MPE, and MDE. Moreover, the results also show that the high frequency components contribute greatly to the early fault feature extraction. The proposed HDE method presents a new perspective for intelligent fault diagnosis, and potentially offers a promising tool to extract features from other types of signals.

Acknowledgements

This research was funded by the National Natural Science Foundation of China, China [Grant No. 51805434]

Conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.
- [2] S. M. Pincus, "Approximate entropy as a measure of system complexity.," *Proc. Natl. Acad. Sci.*, vol. 88, no. 6, pp. 2297–2301, Mar. 1991, doi: 10.1073/pnas.88.6.2297.
- [3] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am. J. Physiol.-Heart Circ. Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000, doi: 10.1152/ajpheart.2000.278.6.H2039.
- [4] Weiting Chen, Zhizhong Wang, Hongbo Xie, and Wangxin Yu, "Characterization of Surface EMG Signal Based on Fuzzy Entropy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 2, pp. 266–272, Jun. 2007, doi: 10.1109/TNSRE.2007.897025.
- [5] C. Bandt and B. Pompe, "Permutation Entropy: A Natural Complexity Measure for Time Series,"

- Phys. Rev. Lett.*, vol. 88, no. 17, Apr. 2002, doi: 10.1103/PhysRevLett.88.174102.
- [6] Y. Li, X. Wang, Z. Liu, X. Liang, and S. Si, “The Entropy Algorithm and Its Variants in the Fault Diagnosis of Rotating Machinery: A Review,” *Ieee Access*, vol. 6, pp. 66723–66741, 2018, doi: 10.1109/ACCESS.2018.2873782.
- [7] Y. Li, X. Wang, S. Si, and S. Huang, “Entropy Based Fault Classification Using the Case Western Reserve University Data: A Benchmark Study,” *IEEE Trans. Reliab.*, pp. 1–14, 2019, doi: 10.1109/TR.2019.2896240.
- [8] M. Costa, A. L. Goldberger, and C.-K. Peng, “Multiscale Entropy Analysis of Complex Physiologic Time Series,” *Phys. Rev. Lett.*, vol. 89, no. 6, Jul. 2002, doi: 10.1103/PhysRevLett.89.068102.
- [9] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. Guo, “Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference,” *Expert Syst. Appl.*, vol. 37, no. 8, pp. 6077–6085, Aug. 2010, doi: 10.1016/j.eswa.2010.02.118.
- [10] S. Pan, T. Han, A. C. C. Tan, and T. R. Lin, “Fault Diagnosis System of Induction Motors Based on Multiscale Entropy and Support Vector Machine with Mutual Information Algorithm,” *Shock Vib.*, vol. 2016, p. 5836717, 2016, doi: 10.1155/2016/5836717.
- [11] J. Cui, Q. Zheng, Y. Xin, C. Zhou, Q. Wang, and N. Zhou, “Feature extraction and classification method for switchgear faults based on sample entropy and cloud model,” *Iet Gener. Transm. Distrib.*, vol. 11, no. 11, pp. 2938–2946, Aug. 2017, doi: 10.1049/iet-gtd.2016.1459.
- [12] A. S. Minhas, N. Sharma, G. Singh, P. K. Kankar, and S. Singh, “Improvement in classification accuracy and computational speed in bearing fault diagnosis using multiscale fuzzy entropy,” *J. Braz. Soc. Mech. Sci. Eng.*, vol. 42, no. 11, p. 586, Oct. 2020, doi: 10.1007/s40430-020-02671-1.
- [13] X. Zhu, J. Zheng, H. Pan, J. Bao, and Y. Zhang, “Time-Shift Multiscale Fuzzy Entropy and Laplacian Support Vector Machine Based Rolling Bearing Fault Diagnosis,” *Entropy*, vol. 20, no. 8, p. 602, Aug. 2018, doi: 10.3390/e20080602.
- [14] W. Du *et al.*, “A New Fuzzy Logic Classifier Based on Multiscale Permutation Entropy and Its Application in Bearing Fault Diagnosis,” *Entropy*, vol. 22, no. 1, p. 27, Jan. 2020, doi: 10.3390/e22010027.
- [15] R. Tiwari, V. K. Gupta, and P. K. Kankar, “Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier,” *J. Vib. Control*, vol. 21, no. 3, pp. 461–467, Feb. 2015, doi: 10.1177/1077546313490778.
- [16] M. N. Yasir and B.-H. Koh, “Data Decomposition Techniques with Multi-Scale Permutation Entropy Calculations for Bearing Fault Diagnosis,” *Sensors*, vol. 18, no. 4, p. 1278, Apr. 2018, doi: 10.3390/s18041278.
- [17] X. Wang, S. Si, and Y. Li, “Multiscale diversity entropy: a novel dynamical measure for fault diagnosis of rotating machinery,” *IEEE Trans. Ind. Inform.*, pp. 1–1, 2020, doi: 10.1109/TII.2020.3022369.
- [18] J.-J. Ding, *Time frequency analysis and wavelet transform class note*. Taipei, Taiwan: National Taiwan University (NTU), 2007.
- [19] F. Takens, “Detecting strange attractors in turbulence,” in *Dynamical Systems and Turbulence*,

- Warwick 1980*, vol. 898, Springer, pp. 366–381.
- [20] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
- [21] L. van der Maaten and G. Hinton, “Visualizing Data using t-SNE,” *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
- [22] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, “Variable selection using random forests,” *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2225–2236, Oct. 2010, doi: 10.1016/j.patrec.2010.03.014.

Figures

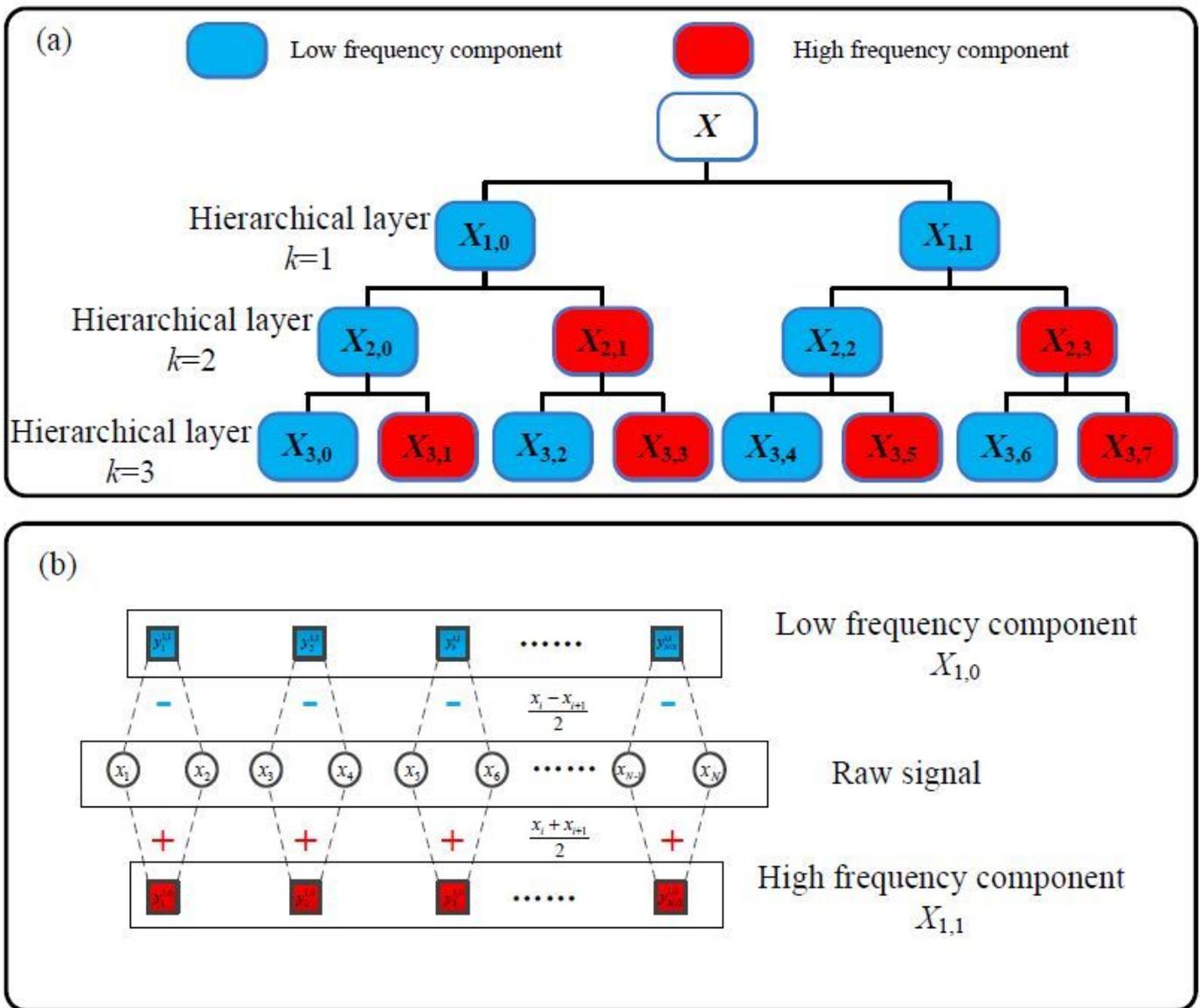


Figure 1

The diagrammatic sketch of hierarchical decomposition: (a) hierarchical nodes; (b) low frequency component and high frequency component.

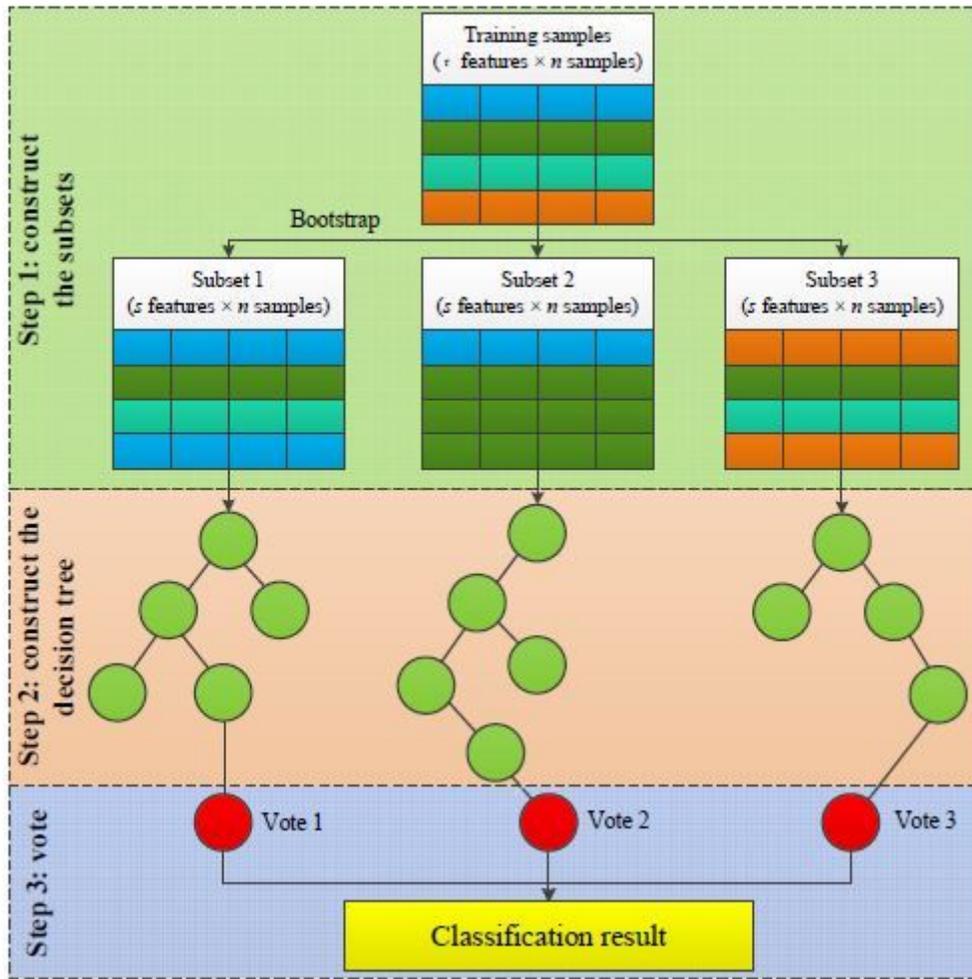


Figure 2

The diagrammatic sketch of random forest

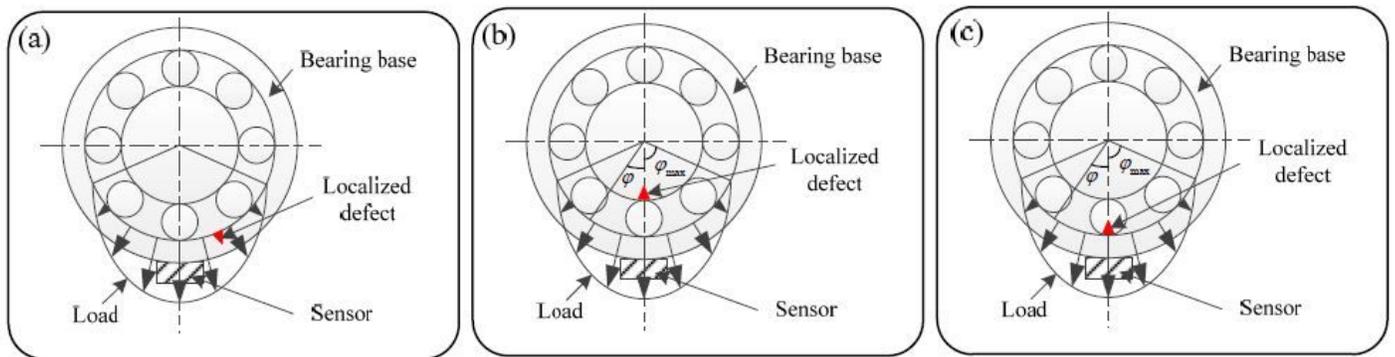


Figure 3

The diagrammatic sketch of the simulated bearing: (a) outer race fault; (b) inner race fault; (c) ball fault.

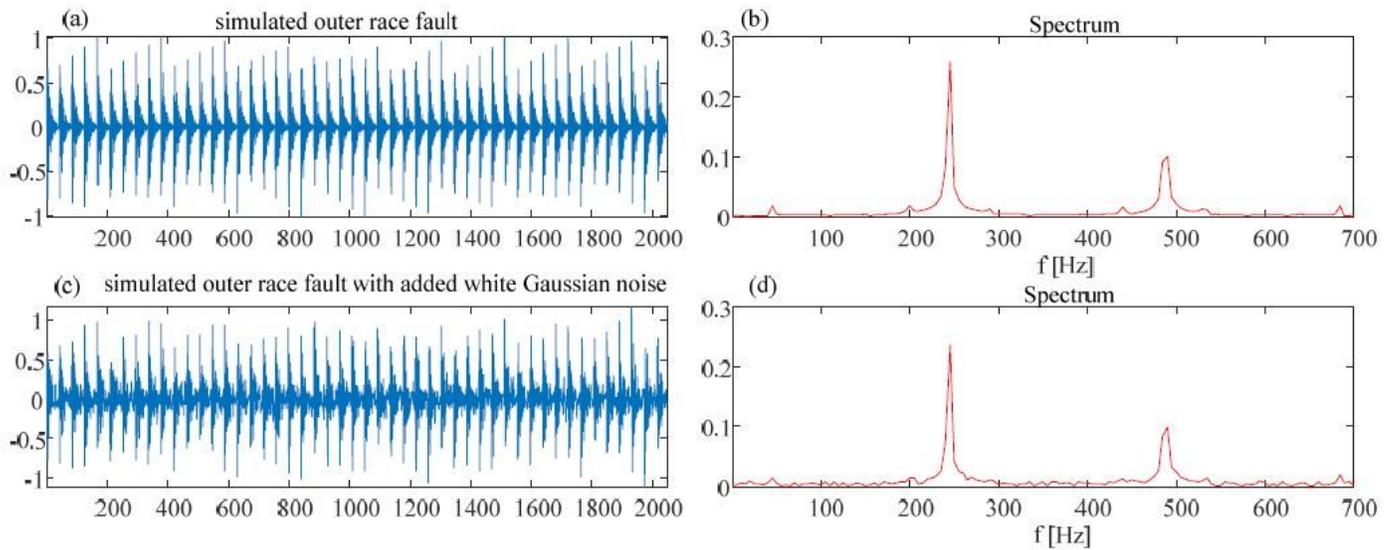


Figure 4

Time wave of the simulated outer race fault; (b) Spectrum of the simulated outer race fault; (c) Time wave of the simulated outer race fault with added white Gaussian noise; (d) Spectrum of the simulated outer race fault with added white Gaussian noise.

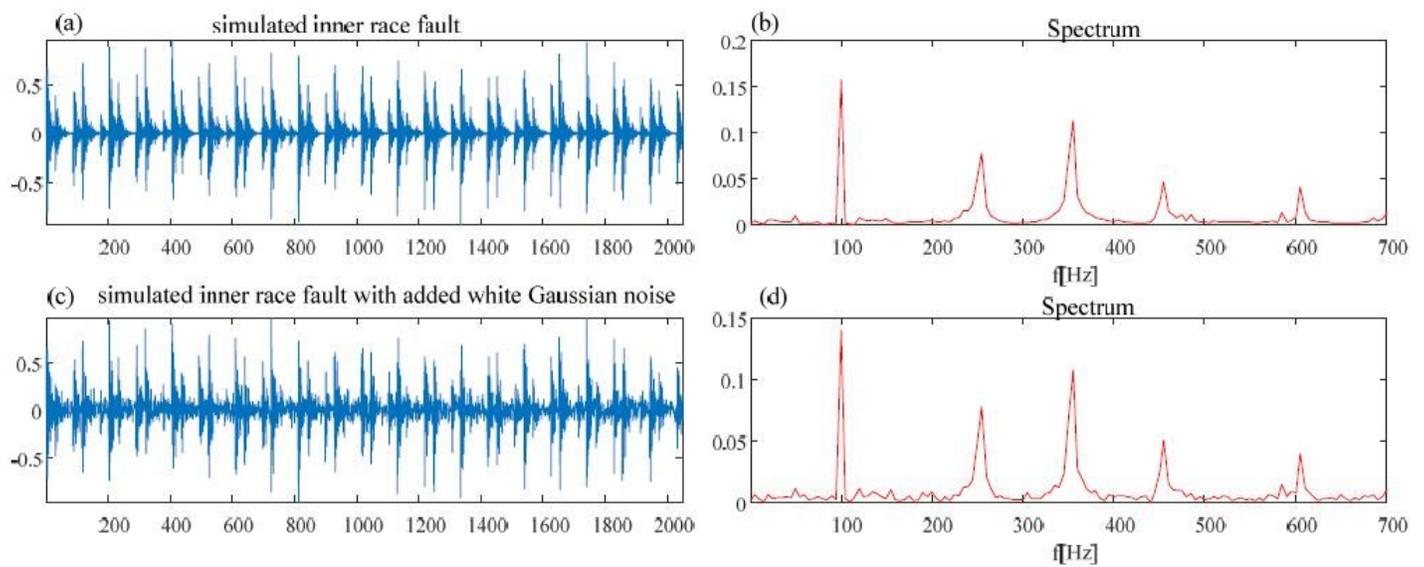


Figure 5

Time wave of the simulated inner race fault; (b) Spectrum of the simulated inner race fault; (c) Time wave of the simulated inner race fault with added white Gaussian noise; (d) Spectrum of the simulated inner race fault with added white Gaussian noise.

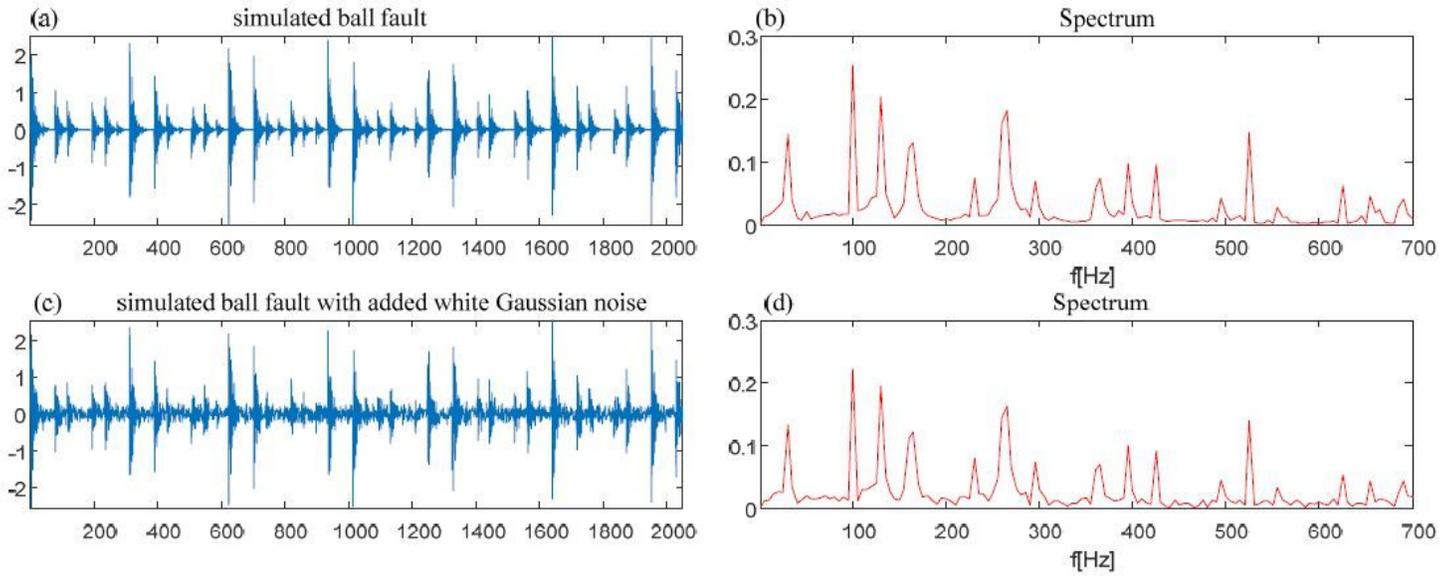


Figure 6

Time wave of the simulated ball fault; (b) Spectrum of the simulated ball fault; (c) Time wave of the simulated ball fault with added white Gaussian noise; (d) Spectrum of the simulated ball fault with added white Gaussian noise.

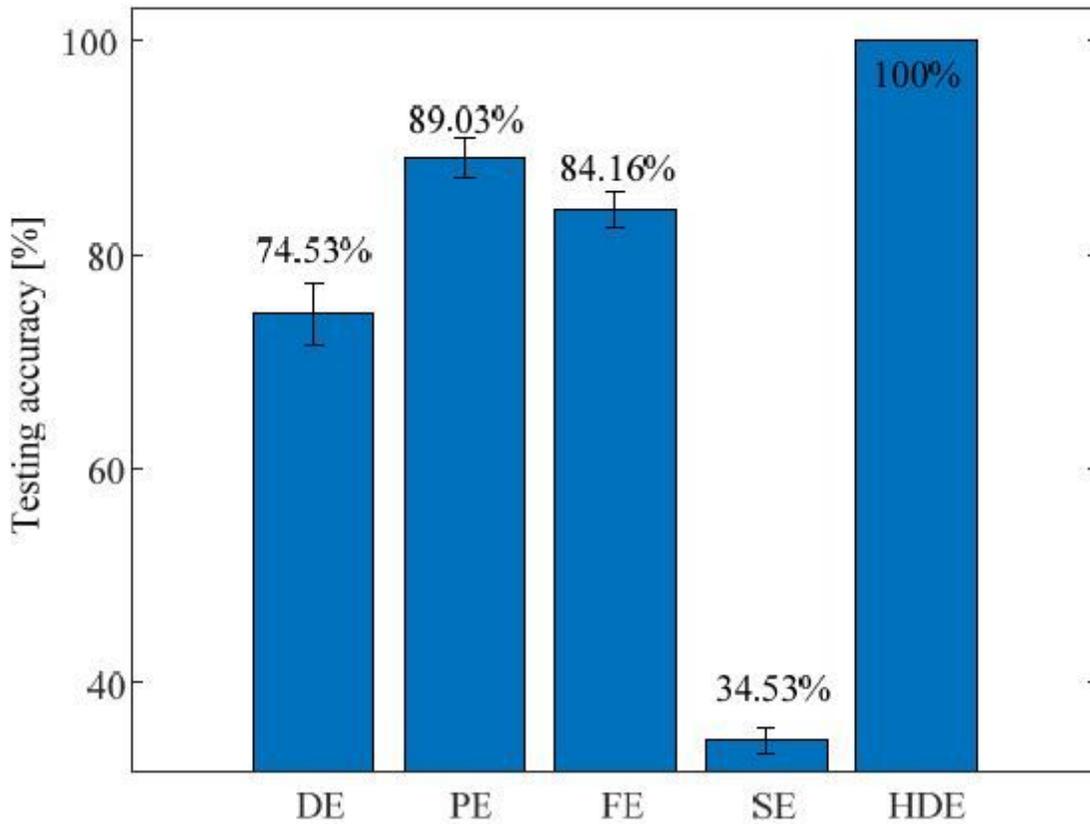


Figure 7

Testing accuracy of simulation evaluation

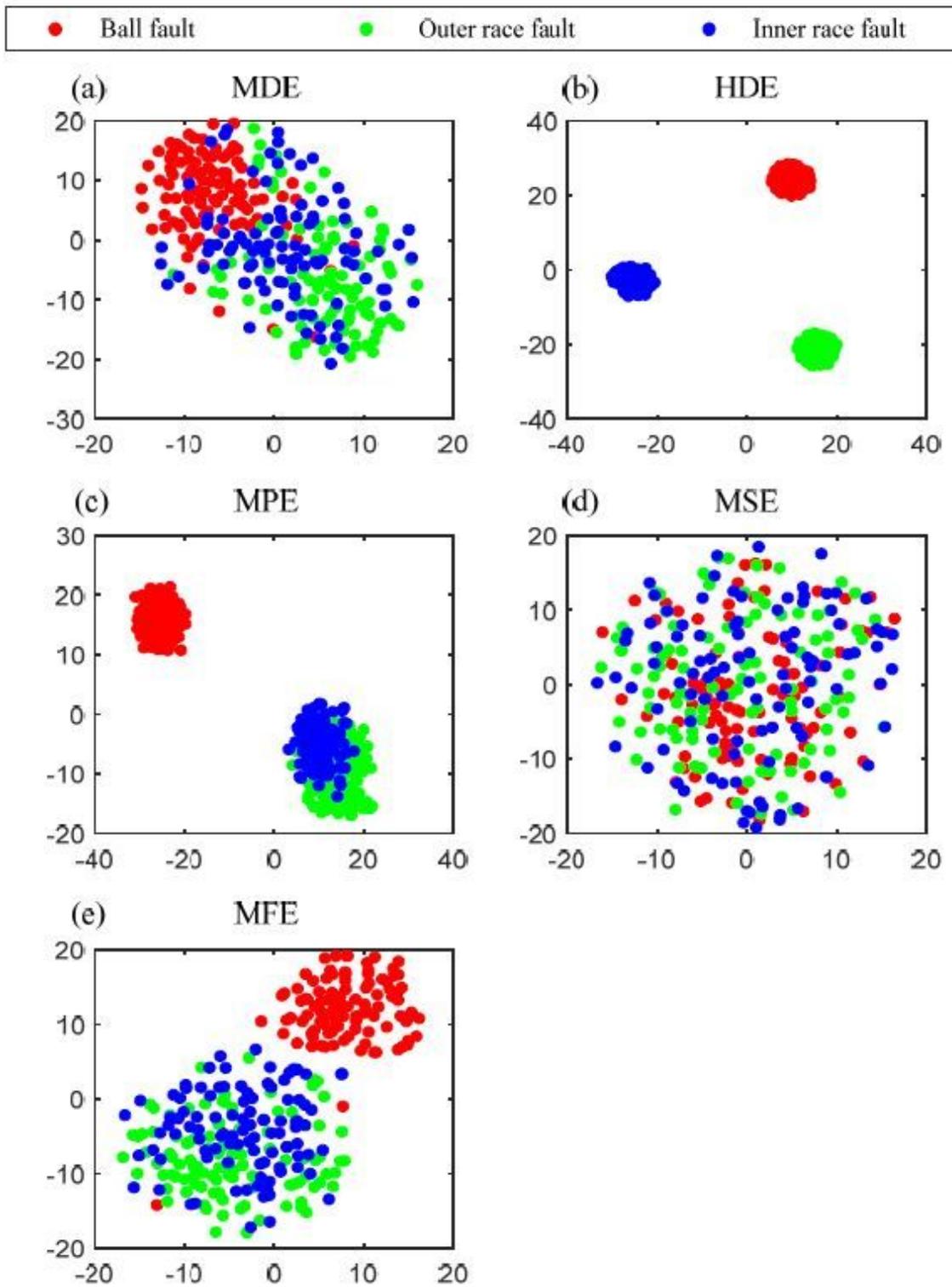


Figure 8

Feature visualization of the simulated bearing: (a) MDE; (b) MFE; (c) MPE; (d) MSE; (e) HDE.

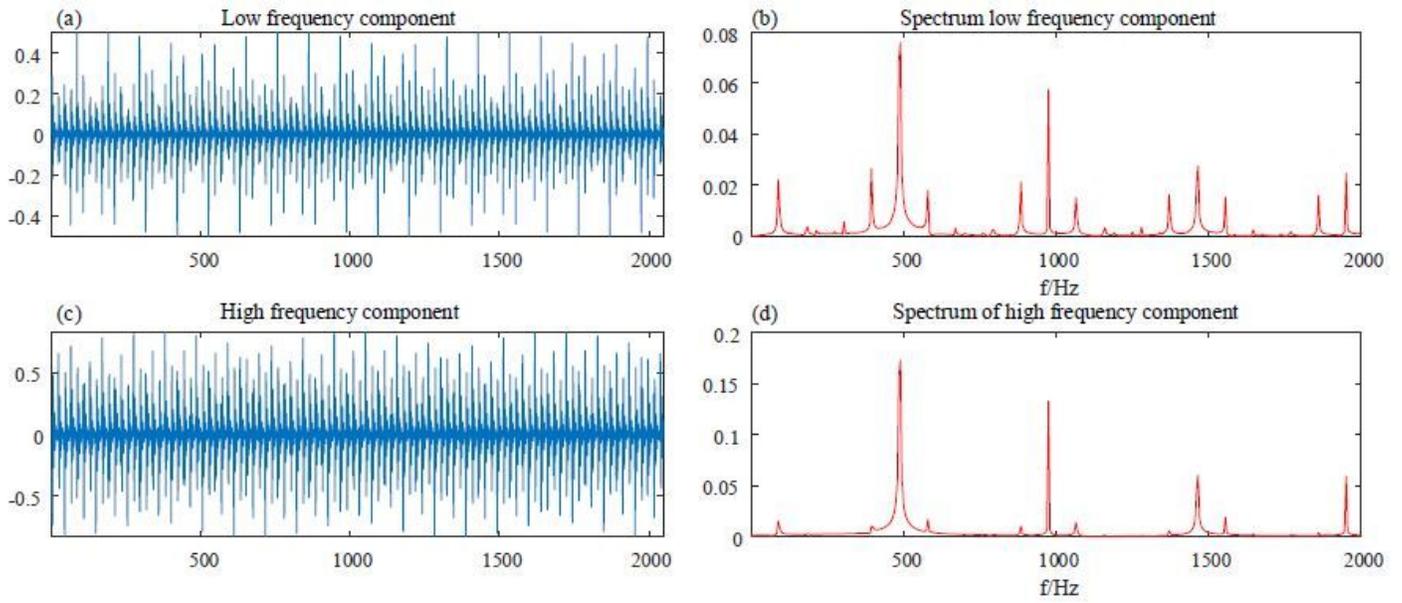


Figure 9

(a) Time wave of the low frequency component; (b) Spectrum of the low frequency component; (c) Time wave of the high frequency component; (d) Spectrum of the high frequency component.

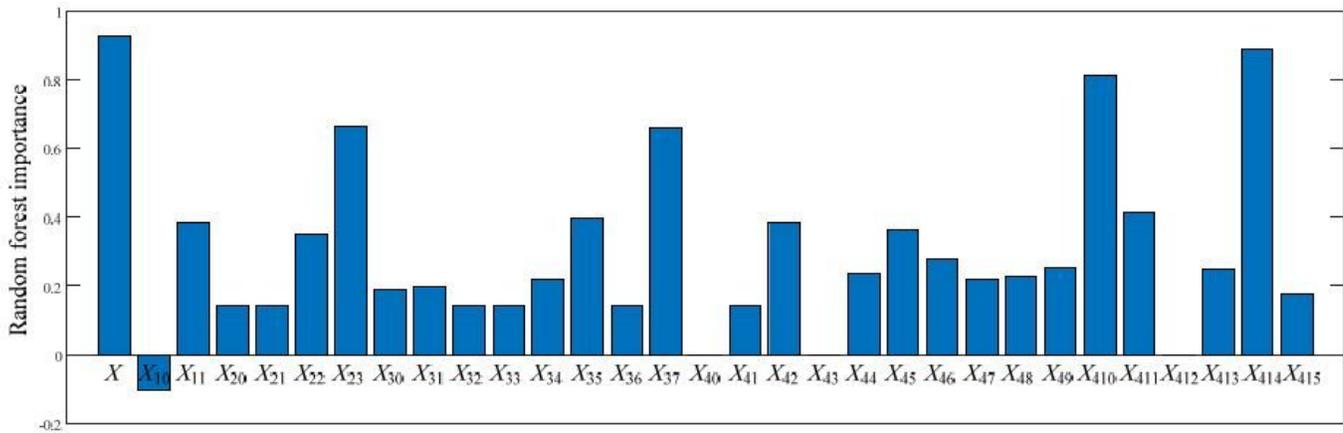


Figure 10

Random forest importance of HDE.

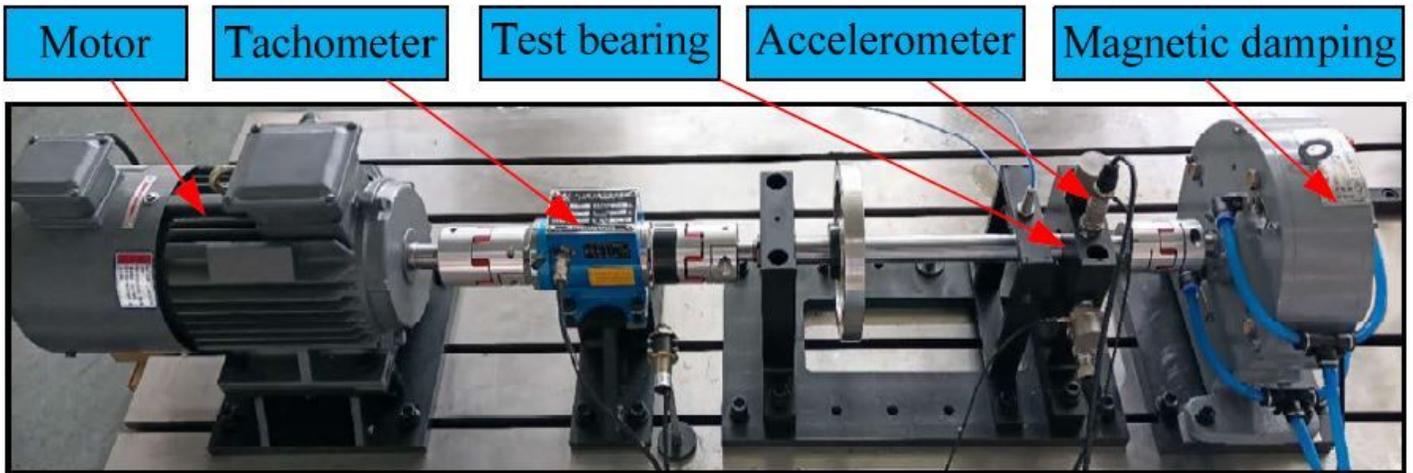


Figure 11

The test rig

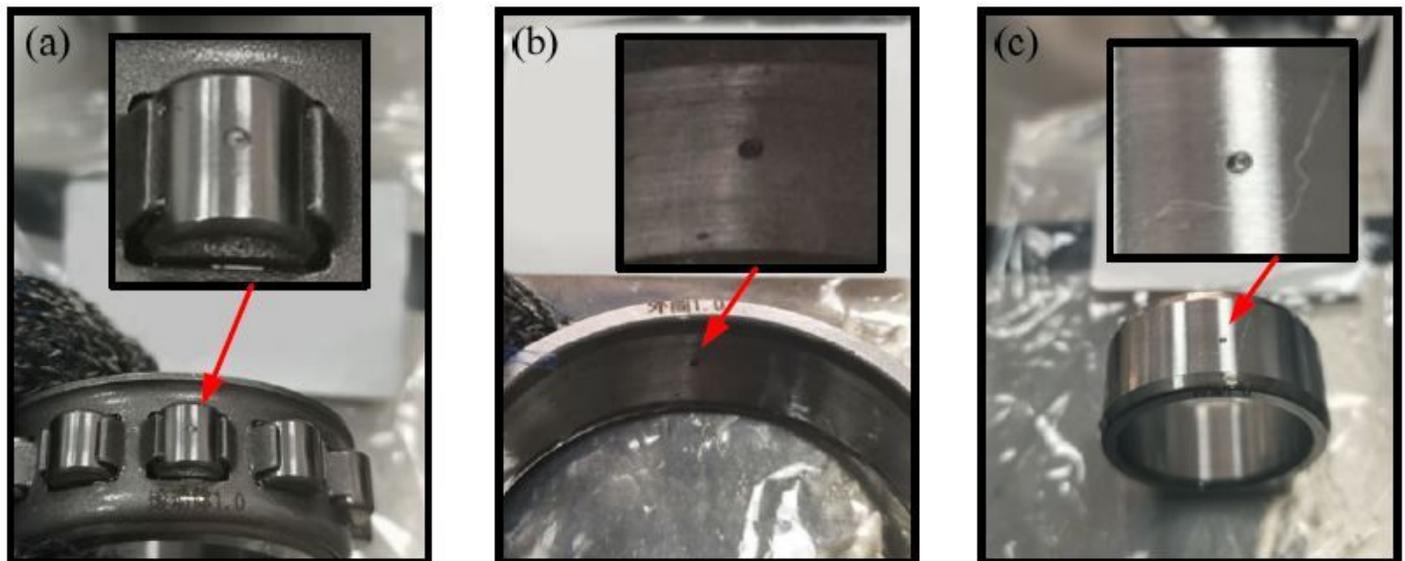


Figure 12

Bearings with artificial pitting defect: (a) Ball fault; (b) Outer race fault; (c) Inner race fault.

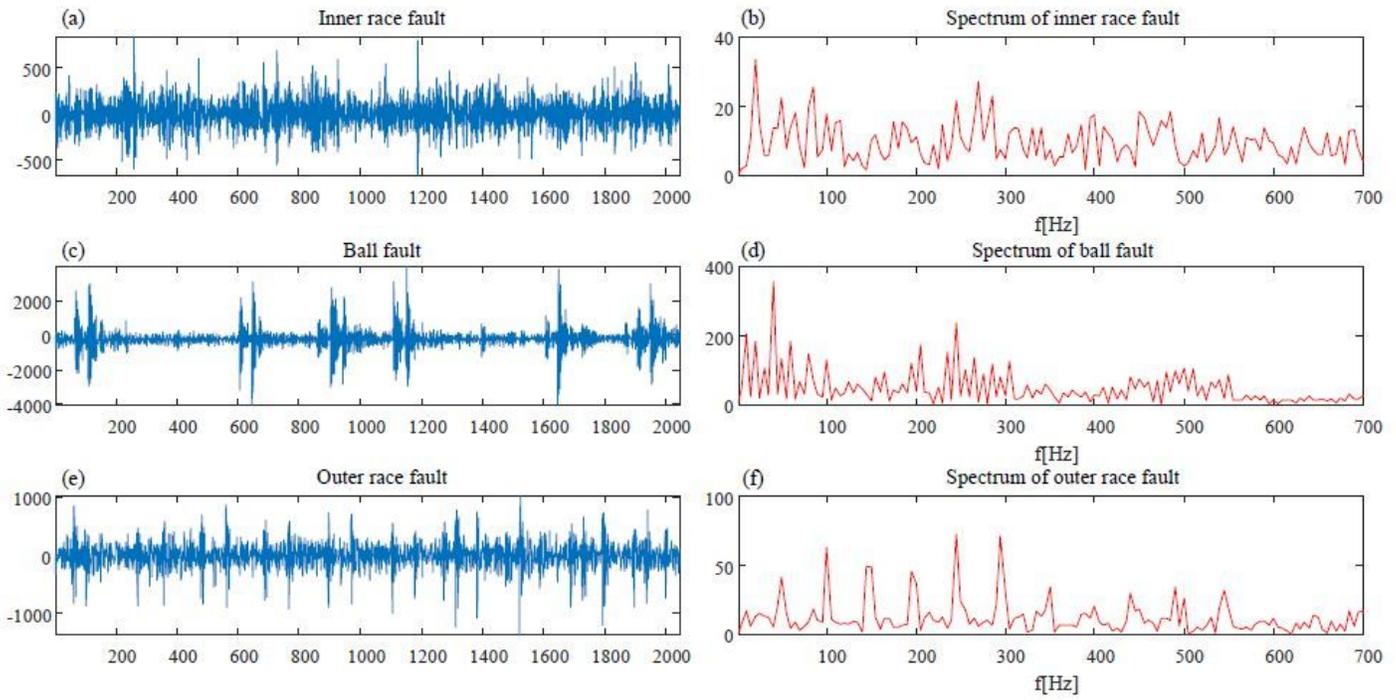


Figure 13

(a) Time wave of the inner race fault; (b) Spectrum of the inner race fault; (c) Time wave of the ball fault; (d) Spectrum of the ball fault; (e) Time wave of the outer race fault; (f) Spectrum of the outer race fault.

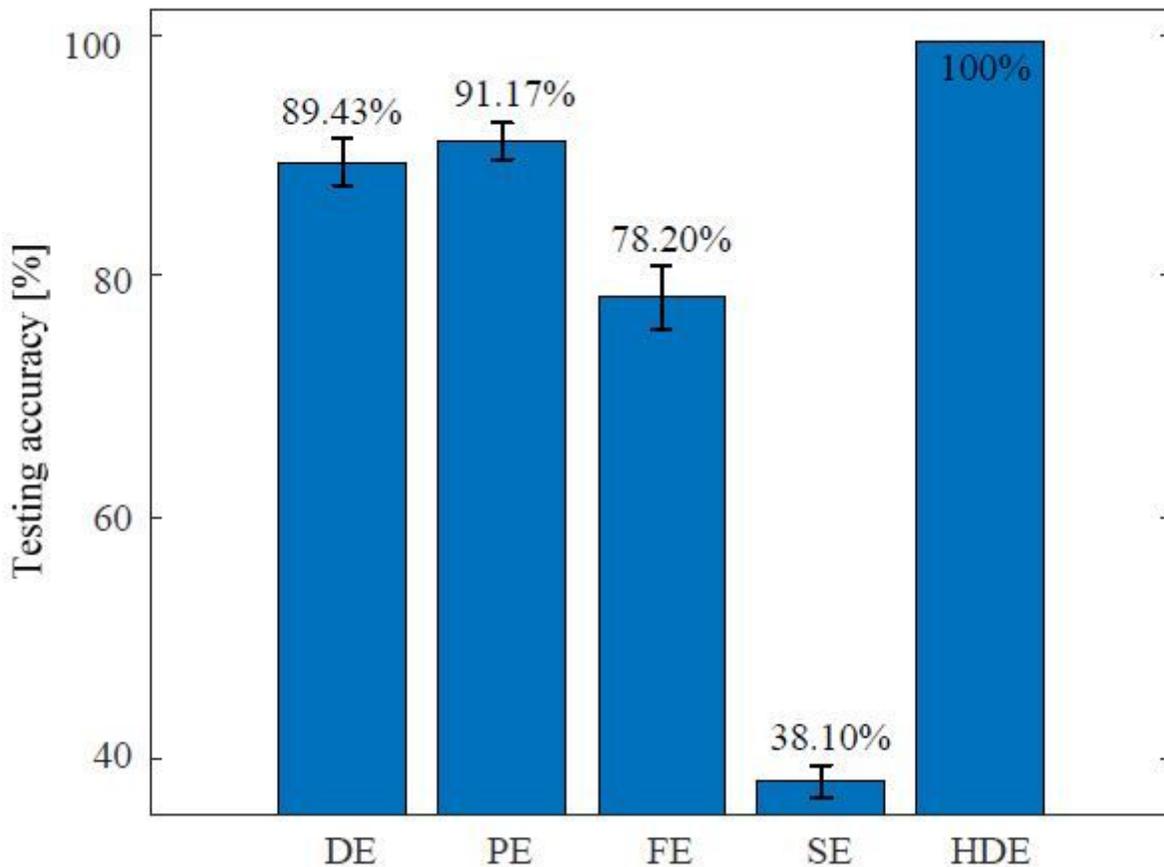


Figure 14

Testing accuracy of experiment evaluation.

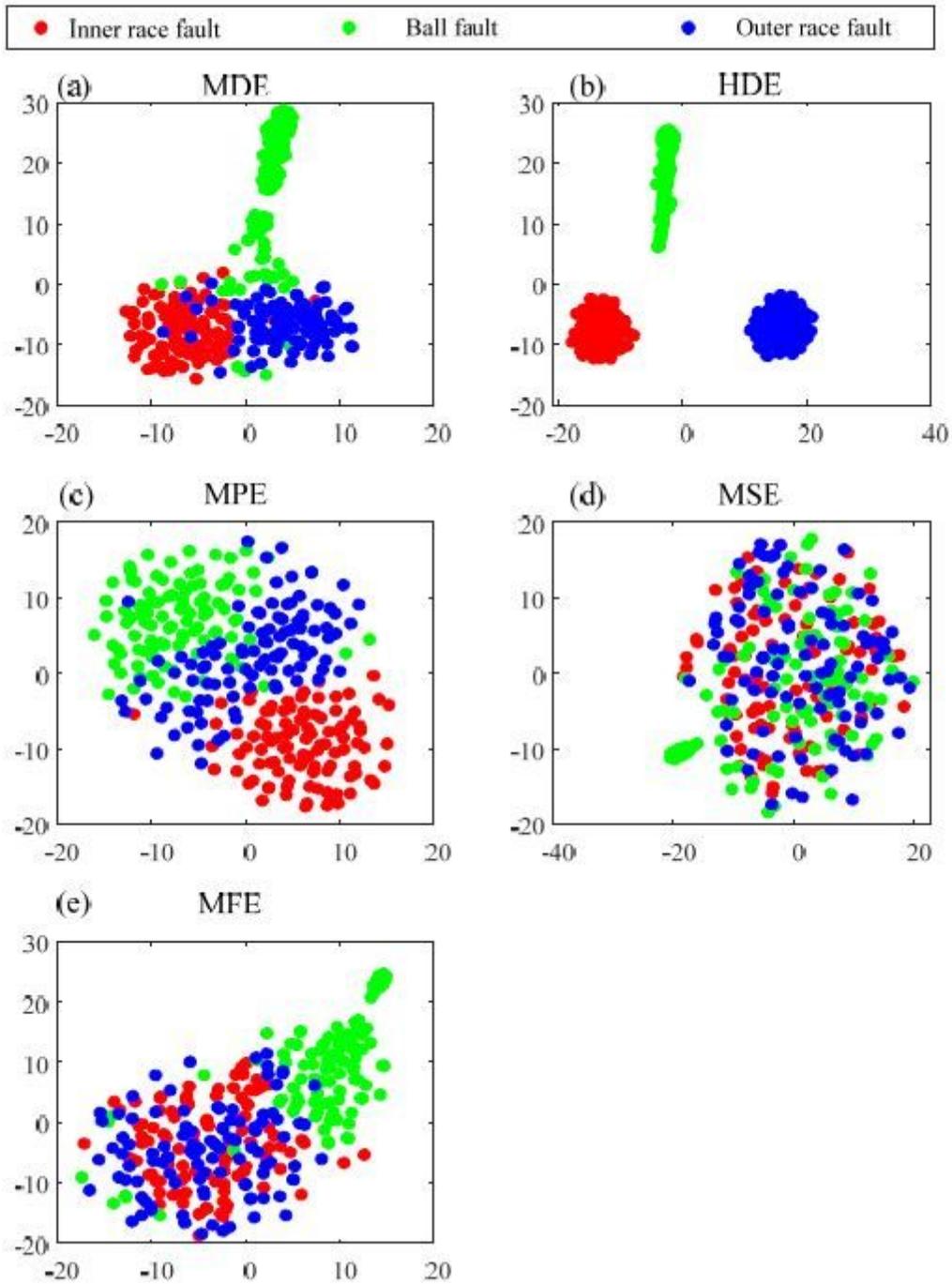


Figure 15

Visualized features of the experimental results

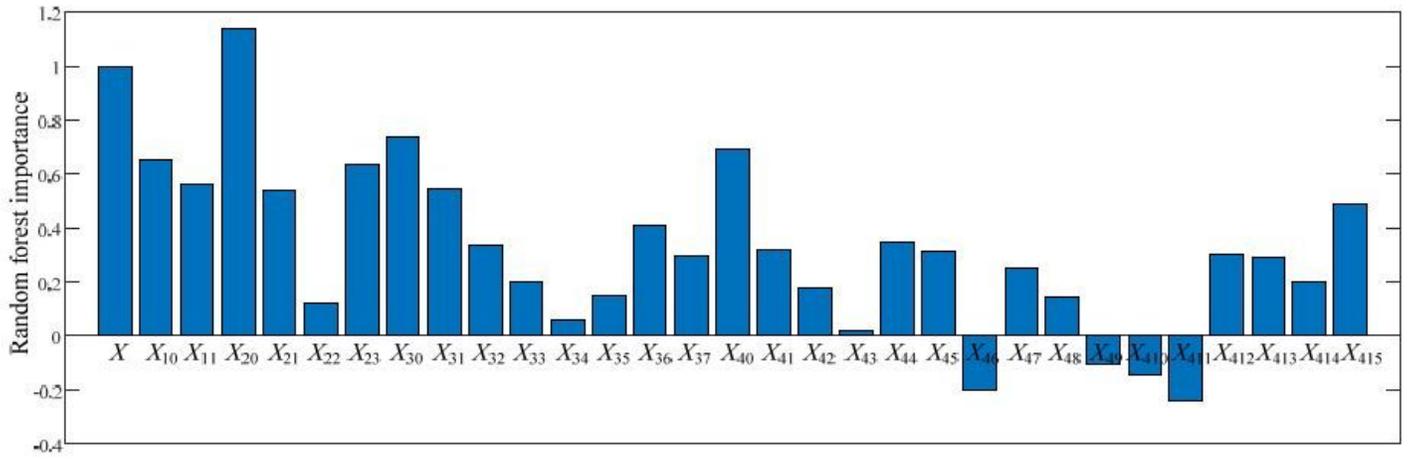


Figure 16

Random forest importance of the experiment results