Hierarchical diversity entropy for the early fault diagnosis of rolling bearing

Xianzhi Wang · Shubin Si · Yongbo Li

Abstract Intelligent fault diagnosis provides great convenience for the prognostic and health management of the rotating machinery. Recently, the multiscale diversity entropy has been proven to be a promising feature extraction tool for the intelligent fault diagnosis. Compared with 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 from early fault signals with weak fault symptoms and strong noise. This can be attributed to the multiscale diversity entropy that 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 frequencies. 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 compared with 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 is a complexity indicator which defines the ratio of new patterns appearing.

FE Fuzzy entropy is a complexity indicator which defines the negative natural logarithm of the conditional probability.

PE Permutation entropy is a complexity indicator which defines the distribution of the permutation of orbits.

DE Diversity entropy is a complexity indicator which defines the expectation of the diversity between the orbits.

MSE Multiscale sample entropy is a multiscale complexity indicator which combines the multiscale analysis with the SE.

MPE Multiscale permutation entropy is a multiscale complexity indicator which combines the multiscale analysis with the PE.

MFE Multiscale fuzzy entropy is a multiscale complexity indicator which combines the multiscale analysis with the FE.
1 Introduction

Intelligent fault diagnosis provides great convenience for the safety operation of the rotating machinery. Since the intelligent fault diagnosis has merits of independent with expert knowledge, fast response, and low cost, it has been widely used in the modern industry. Generally, the intelligent fault diagnosis can be divided into three steps: data acquisition, feature extraction, and fault classification. Among these three steps, feature extraction is the most crucial step. To effectively extract features from the measured signal, the entropy-based method has become a hot topic due to the advantages of independence with prior knowledge, unnecessary of preprocessing, and easy to perform.

The entropy theory originates from information entropy [1]. In Ref. [1], Shannon defined the concept of information entropy, which measures the complexity from message [1]. Then, Pincus proposed approximate entropy to quantify the dynamical complexity from arbitrary time series [2]. Sample entropy (SE) emerged as an improvement method of approximate entropy [3], which makes more reliable dynamical complexity than approximate entropy. Fuzzy entropy (FE) has been proposed to improve the stability of SE [4]. The emergence of permutation entropy (PE) brings a new sight for dynamic complexity analysis, which utilizes the amplitude permutation of orbits to estimate the dynamical complexity [5]. Above all, these commonly used entropy-based methods and their applications in fault diagnosis of rotating machinery have been summarized in Ref. [6], [7].

The above-mentioned entropy methods can only extract unidimensional feature from time series, which can be hardly applied in the practical engineering. To extract multidimensional features, Costa et al. extended the SE into multiscale sample entropy (MSE) [8]. In Ref. [8], the raw signal is decomposed into multiple scale time series by the coarse-gaining procedure. Then, the MSE can be obtained by calculating the SE value under different scales. Since the MSE can provide rich information for fault classification, many researchers have applied MSE in the fault feature extraction of rotating machinery.

Zhang et al. used the MSE 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 by MSE. Then, the mutual information was used to choose the five most important features. Last, the selected features were put into the support vector machine to classify the induction motor fault [10]. Cui et al. used the MSE to extract fault features. Then, the fuzzy support vector machine was used to identify different fault types of switchgear [11].

Inspired by multiscale sample entropy (MSE), PE and FE are also extended into multiscale permutation entropy (MPE) and multiscale fuzzy entropy (MFE). Minhas et al. used the MFE to extract fault features; then, the support vector machine classifier was used to distinguish different bearing faults [12]. Zhu et al. combined the MFE 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 MPE to extract fault features. Then, the linear discriminant analysis (LDA) was used to select the fault features. At last the selected features were put into harmonic mean difference self-organizing fuzzy logic classifier to identify the bearing faults [14]. Tiwari et al. used the MPE 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 values of the product functions were used as fault features to distinguish the bearing faults [16].

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

MDE Multiscale diversity entropy is a multiscale complexity indicator which combines the multiscale analysis with the DE.

HDE Hierarchical diversity entropy is a multiscale complexity indicator which combines the hierarchical analysis with the DE.

RF Random forest is an ensemble learning classifier which utilizes multiple decision trees for training and predicting.
(1) High consistency. The diversity entropy can estimate the dynamical complexity with high consistency, while PE, SE, and FE may yield inconsistent results in estimating some periodic systems.

(2) High calculation efficiency. The diversity entropy has the lowest calculation complexity, which saves at least half of time compared to PE, SE, and FE.

(3) Strong robustness. Compared to PE, SE, and FE, the diversity entropy has the strongest robustness when the FM signal, the FM signal and the sinusoidal signal are interfered by Gaussian white noise.

Although the MDE has such merits, it encounters challenge of extracting adequate fault features for the early fault diagnosis of rotating machinery. Essentially, the multiscale procedure used in MDE is a low-pass filter based on the Haar wavelet [18]. However, the multiscale procedure only considers the fault information hidden in the low frequency, which ignores the fault information hidden in the high frequency. This induces incomplete feature extraction of MDE, which offers insufficient fault information for the early fault diagnosis. When localized fault occurs at early stage, the fault symptoms are weak and the background noise is strong. Ignoring the fault information hidden in high frequency, the MDE can hardly extract accurate fault information for the early fault diagnosis.

To make up for above defect, 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 compared to MSE, MPE, MFE, and MDE. Moreover, the results also show that the high-frequency components used in HDE can provide indispensable information to enrich the extracted features, which makes up for the incomplete feature extraction of low-frequency components.

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, Sect. 5 summarizes the conclusions.

2 Methodology

2.1 Diversity entropy

For an arbitrary time series \( X = \{x_1, x_2, \ldots, x_i, \ldots, x_N\} \), \( i \in [1, N] \), \( N \) 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].

\[
Y(m) = \{y_1(m), y_2(m), \ldots, y_{N-m+1}(m)\} = \begin{bmatrix}
  x_1 & x_2 & \cdots & x_m \\
  x_2 & x_3 & \cdots & x_{m+1} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{N-m+1} & x_{N-m} & \cdots & x_N
\end{bmatrix}
\]

**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).

\[
D(m) = \{d_1, \ldots, d_{N-m}\} = \{d(y_1(m), y_2(m)), d(y_2(m), y_3(m)), \ldots, d(y_{N-m-1}(m), y_{N-m}(m))\}
\]

\[
d(y_i(m), y_j(m)) = \frac{\sum_{k=1}^{m} y_i(k) \times y_j(k)}{\sqrt{\sum_{k=1}^{m} y_i(k)^2} \times \sqrt{\sum_{k=1}^{m} y_j(k)^2}}
\]

Here, we set 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. Opposi-
tely, a small cosine similarity value represents a
diversity, stochastic, or chaotic dynamic behavior.

**Step 3** Partition the scope \([-1, 1]\) into \(\varepsilon\) intervals. Then, count the state probability \((P_1, \ldots, P_\varepsilon)\) 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).

\[
DE(m, \varepsilon) = -\frac{1}{\ln \varepsilon} \sum_{k=1}^\varepsilon P_k \ln P_k
\]

(4)

where the \(m\) represents the embedding dimension. The \(\varepsilon\) represents the number of symbols.

The physical meaning of \(DE\) is the expected coding length encoded with \(\varepsilon\) 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 multiscale diversity entropy can be calculated as following steps:

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

\[
y_j^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, 1 \leq j \leq N/\tau
\]

(5)

where the \(\tau\) represents the scale factor in the multiscale analysis, which should be a positive integer. The scale factor \(\tau\) 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.

\[
MDE(X, m, \varepsilon) = DE(y_j^\tau, m, \varepsilon)
\]

(6)

where the \(m\) represents the embedding dimension. The \(\varepsilon\) represents the number of symbols.

2.3 Hierarchical diversity entropy

To make up for incomplete feature extraction of MDE, hierarchical diversity entropy (HDE) is proposed in this paper. The proposed HDE can be calculated from the following steps:

**Step 1** For arbitrary time series \(X = \{x_1, x_2, \ldots, x_i, \ldots, x_N \}, i \in [1, N], N\) 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:

\[
Q_0(x) = \frac{x_i + x_{i+1}}{2}, i = 1, 2, \ldots, N - 1
\]

(7)

\[
Q_1(x) = \frac{x_i - x_{i+1}}{2}, i = 1, 2, \ldots, N - 1
\]

(8)

**Step 2** Use a matrix to illustrate the operators \(Q_j\) \((j = 0 \text{ or } 1)\) that can be expressed as:

\[
Q_j^\tau = \begin{bmatrix}
\frac{1}{2} & \frac{(-1)^j}{2} & 0 & 0 & \ldots & 0 & 0 \\
0 & 0 & \frac{1}{2} & \frac{(-1)^j}{2} & \ldots & 0 & 0 \\
0 & 0 & 0 & 0 & \ldots & \frac{1}{2} & \frac{(-1)^j}{2}
\end{bmatrix}
\]

(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.

\[
X_{k,e} = Q_{r_k}^e \cdot Q_{r_{k-1}}^{e-1} \cdot \ldots \cdot Q_{r_1}^1 \cdot x
\]

(10)

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

\[
e = \sum_{m=1}^k 2^{k-m} r_m
\]

(11)

where \([r_1, r_2, \ldots, r_k]\) is the unique vector corresponding to the integer \(e\) and \(\{r_m, m = 1, \ldots, k\} \in \{0, 1\}\) denotes the averaging or differential operator at the \(m\)th layer.

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

\[
HDE(X, k, \epsilon, m, \bar{\epsilon}) = DE(X_{k, \epsilon}, m, \bar{\epsilon})
\] (12)

where the \( m \) represents the embedding dimension. The \( \epsilon \) 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. 1a. A diagrammatic sketch of low-frequency component and high-frequency component is shown in Fig. 1b.

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 three steps:

Step 1 Construct \( K \) subsets via drawing from the training set with the bootstrap method. In each subset, randomly choose \( s \leq \tau \) features.

Step 2 Construct \( L \) decision trees using the selected \( s \) features from Step 1.

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.

2.5 Proposed fault diagnosis frame

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 rolling element fault. The simulated bearing type is N205 cylindrical roller bearing. The detailed bearing size is listed in Table 1. The simulated rotating speed is 3000 rpm. The sampling frequency is 10240 Hz. The diagrammatic sketch of the simulated bearing is shown in Fig. 3.

#### Table 1  Bearing size

| Parameter                  | Value       |
|----------------------------|-------------|
| Pitch circle diameter      | 35.5 mm     |
| Roller diameter            | 6.5 mm      |
| Rotating speed             | 3000 rpm    |
| Number of rollers          | 12          |
| Sample frequency           | 10,240 Hz   |
| Natural frequency of bearing | 4000 Hz   |
| Contact angle              | 0°          |

![Diagrammatic sketch of random forest](image-url)
3.1 Simulated bearing signals

3.1.1 Localized defect on the outer race

The fault model of the outer race fault is shown in Fig. 3a. The distribution of load zones is shown in Fig. 3a. The sensor is mounted at the maximum load density of the load zone. Since the position of the localized defect does not time-varying, the impulsive force can be regarded as ideal force. Assuming at time \( t = 0 \), the localized defect starts contacting with the rolling element, then the impulsive force excited by the localized defect on the outer ring can be expressed as Eq. (13).

\[
D_o(t) = \sum_{t=-\infty}^{+\infty} d_o \delta(t - kT_o)
\]  

(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 \( f_o \) is the fault characteristic frequency of the outer race fault.

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

\[
e(t) = \begin{cases} 
\exp\left(-t/T\right), & t > 0 \\
0, & t \leq 0
\end{cases}
\]  

(14)

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

\[
v_o(t) = \sum_{t=-\infty}^{+\infty} d_o \delta(t - kT_o) * e(t)
\]  

(15)

The time wave of the simulated outer race fault and the corresponding envelope spectrum are plotted in Fig. 4a 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 envelope spectrum are plotted in Fig. 4c and d, respectively.

3.1.2 Localized defect on the inner race

The fault model of the inner race fault is shown in Fig. 3b. The basic assumption is the same as the outer race fault. Suppose at time \( t = 0 \), the first impulse is generated by the rolling element contacting with the localized defect at the peak of the load zone. As the localized defect rotates with the inner race, the contacting location between the roller and inner race is time-varying. The impulsive force will be generated only by such contact occurring in the load zone. The impulsive force excited by the localized defect on the inner ring can be expressed as Eq. (16).

\[
D_i(t) = \sum_{t=-\infty}^{+\infty} d_i \delta(t - kT_i)
\]  

(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 \( f_i \) is the fault characteristic frequency of the inner race fault.

The load distribution is defined as Eq. (17).

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

(17)

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

Fig. 3 The diagrammatic sketch of the simulated bearing: a outer race fault; b inner race fault; c rolling element fault
As shown in Fig. 3b, when the localized defect contacts with the rolling element at $\phi$, the impulsive force acquired by the sensor is the projection along the axis $u=0$. Therefore, the influence coefficient of localized defect is shown in Eq. (18).

$$p(\phi) = \cos \phi$$  \hspace{1cm} (18)

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

$$f_i(t) = \Delta_i(t)q(2\pi f_r t)p(2\pi f_r t)$$  \hspace{1cm} (19)

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

$$v_i(t) = A_i [\Delta_i(t)q(2\pi f_r t)p(2\pi f_r t)] * e(t)$$  \hspace{1cm} (20)

where the $A_i$ is the conversion coefficient between impulsive force and vibration. In this paper, $A_i = 1$.

The time wave of the simulated inner race fault and the corresponding envelope spectrum are plotted in Fig. 5.
3.1.3 Localized defect on the rolling element

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

$$\Delta_b(t) = \Delta_{bo}(t) + \Delta_{bi}(t)$$  \hspace{1cm} (21)

$$\Delta_{bo}(t) = \sum_{i=-\infty}^{+\infty} d_{bo} \delta(t - kT_b)$$  \hspace{1cm} (22)

$$\Delta_{bi}(t) = \sum_{i=-\infty}^{+\infty} d_{bi} \delta(t - kT_b - 0.5 * T_b)$$  \hspace{1cm} (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.

| Method | Embedding dimension $m$ | Scale $\tau$ | Layer $k$ | Tolerance $r$ | Symbols $\varepsilon$ |
|--------|-------------------------|--------------|-----------|--------------|-----------------|
| MSE    | 2                       | 31           | –         | 0.15         | –               |
| MFE    | 2                       | 31           | –         | 0.15         | –               |
| MPE    | 6                       | 31           | –         | –            | –               |
| MDE    | 4                       | 31           | –         | –            | 30              |
| HDE    | 4                       | –            | 4         | –            | 30              |

Table 2 Parameters of the five entropy-based methods

Fig. 6 Time wave of the simulated rolling element fault; b envelope spectrum of the simulated rolling element fault; c time wave of the simulated rolling element fault with added white Gaussian noise; d envelope spectrum of the simulated rolling element fault with added white Gaussian noise

3.1.3 Localized defect on the rolling element

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

$$\Delta_b(t) = \Delta_{bo}(t) + \Delta_{bi}(t)$$  \hspace{1cm} (21)

$$\Delta_{bo}(t) = \sum_{i=-\infty}^{+\infty} d_{bo} \delta(t - kT_b)$$  \hspace{1cm} (22)

$$\Delta_{bi}(t) = \sum_{i=-\infty}^{+\infty} d_{bi} \delta(t - kT_b - 0.5 * T_b)$$  \hspace{1cm} (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.

Fig. 7 Testing accuracy of simulation evaluation
element, and the $f_b$ is the fault characteristic frequency of the rolling element fault.

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

$$v_b(t) = A_b [\Delta \delta_b(t) q(2\pi f_r, t) p(2\pi f_r, t)] * e(t) + A_b [\Delta \delta_r(t) q(2\pi f_r, t) p(2\pi f_r, t)] * e(t)$$

where the $A_b$ is the conversion coefficient between impulsive force and vibration. In this paper, $A_b = 1$.

The time wave of the simulated rolling element fault and the corresponding envelope spectrum are plotted in Fig. 6.

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. The simulation evaluation procedure can be divided into five steps. First, generate the simulated
signals as described in Sect. 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. We randomly choose 50% samples from each fault type for training and residual samples for testing. 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 fault features of the training dataset to train the random forest classifier. Last, put the fault features of 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 compare 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.

Note that the embedding dimension, tolerance, and symbols are set as the best parameters which can be referred to Ref. [3–5], 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.

In Fig. 7, the proposed HDE achieves the highest testing accuracy. This indicates the proposed HDE has the best feature extraction ability compared to MFE, MSE, MPE, and MDE. To further analyze, the visualized features using t-SNE [21] are plotted in

**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| 10,240 Hz        |
| Contact angle   | $0^\circ$        |
Fig. 8. In Fig. 8, the cluster ability shows the feature extraction ability: 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 of the testing method.

First, in Fig. 8e, 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 two clusters overlapping. Besides, the large inner-class distance makes the samples too scattered to form into a cluster. Second, in Fig. 8a, for MDE, the situation of the outer race fault and the inner race fault is similar as the MFE. But the inter-class between rolling element fault and other classes is closer than MFE. Third, in Fig. 8d, for MSE, all three clusters mix together. Four, in Fig. 8c, for MPE, the inner-class distance is close enough to form into clusters. However, the inter-class distance between the outer race fault and the inner race fault is still too close, resulting in the two clusters overlapping. Last, in Fig. 8b, 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 a clear cluster boundary.

This can be attributed to indispensable information provided by the high-frequency components. To prove
this point, the random forest importance is employed to show the contribution of each components used in HDE [22]. The result is shown in Fig. 9. In Fig. 9, the higher the random forest importance value, the greater the component contributes to classification. First, some low-frequency components contribute negatively to distinguish the three types of faults, e.g., $X_{10}$. This indicates that not all low-frequency components are beneficial to classification. Second, some high-frequency components contribute greater than some low-frequency components, e.g., $X_{23}$. This indicates the high-frequency components can provide indispensable information, which makes up for the incomplete feature extraction of low-frequency components.

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

4 Experiment evaluation

To evaluate 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 is the same as the simulation described in Sect. 3.2, and the testing accuracy is regarded as the indicator to evaluate the performance of each 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. 10. The test rig is HD-FD-H-03X produced by Wuxi Houde Automation Meter Co., Ltd. The test rig consists of 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 10,240 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 rolling element 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. 11. The time wave and corresponding envelope spectrums of the collected vibrational signals are shown in Fig. 12.

4.2 Results and discussion

The testing accuracy of HDE, MDE, MSE, MFE, and MPE is plotted Fig. 13. As can be seen from Fig. 13, the proposed HDE achieves the highest testing accuracy with the smallest standard deviation. This indicates the proposed HDE has the best feature extraction ability among these six methods.

To further investigate, the visualized features by t-SNE are plotted in Fig. 14. First, in Fig. 14c, 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 rolling element fault. Second, in Fig. 14e, for MFE, the inter-class distance between the inner race fault and outer race fault is smaller than MPE. This makes these samples even less separable. Third, in Fig. 14d, for
MSE, the situation is even worse. Fourth, in Fig. 14a, 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 build a nice boundary. Last, in Fig. 14b, 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.

In this section, we also conduct the random forest importance test to evaluate the contribution of the high-frequency components. The result is shown in Fig. 15. In Fig. 15, the result coincides with the simulation in Sect. 3.2: first, some low-frequency components contribute negatively to distinguish the three types of faults, e.g., $X_{46}$. Second, some high-frequency components contribute greater than some low-frequency components, e.g., $X_{23}$. This result also
demonstrates the effectiveness of the high-frequency components for classification.

5 Conclusion

A novel feature extraction method called 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 frequencies, which enriches the fault features to improve the fault diagnosis accuracy. Based on HDE and RF, a novel intelligent fault diagnosis frame is proposed to automatically recognize the bearing faults. Both simulated and experimental signals demonstrate that HDE performs the best in classifying bearing faults among the MSE, MFE, MPE, and MDE methods. Moreover, the comparison results also show that the high-frequency components used in HDE can provide indispensable information, which makes up for the incomplete feature extraction of low-frequency components. The proposed HDE method presents a new perspective for the early fault diagnosis and potentially offers a promising tool to extract features from other types of signals.

Acknowledgements This study was funded by the National Natural Science Foundation of China, China (Grant No. 51805434)

Data availability Data of this study will be made available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. 27(3), 379–423 (1948). https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
2. Pincus, S.M.: Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. 88(6), 2297–2301 (1991). https://doi.org/10.1073/pnas.88.6.2297
3. Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. Am. J. Physiol.-Heart Circ. Physiol. 278(6), H2039–H2049 (2000). https://doi.org/10.1152/ajpheart.2000.278.6.H2039
4. Chen, W., Wang, Z., Xie, H., Wangxin, Yu.: Characterization of surface EMG signal based on fuzzy entropy. IEEE Trans. Neural Syst. Rehabil. Eng. 15(2), 266–272 (2007). https://doi.org/10.1109/TNSRE.2007.897025
5. Bandt, C., Pompe, B.: Permutation entropy: a natural complexity measure for time series, Phys. Rev. Lett., 88(17), (2002) doi: https://doi.org/10.1103/PhysRevLett.88.174102.
6. Li, Y., Wang, X., Liu, Z., Liang, X., Si, S.: The entropy algorithm and its variants in the fault diagnosis of rotating machinery: a review. IEEE Access 6, 66723–66741 (2018). https://doi.org/10.1109/ACCESS.2018.2873782
7. Li, Y., Wang, X., Si, S., Huang, S.: Entropy based fault classification using the case western reserve university data: a benchmark study, IEEE Trans. Reliab., pp. 1–14, 2019, doi: https://doi.org/10.1109/TR.2019.2896240.
8. Costa, M, Goldberger, A. L., Peng, C.-K.: Multiscale entropy analysis of complex physiologic time series, Phys. Rev. Lett., 89(6), (2002) doi: https://doi.org/10.1103/PhysRevLett.89.068102.
9. Zhang, L., Xiong, G., Liu, H., Zou, H., Guo, W.: Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. Expert Syst. Appl. 37(8), 6077–6085 (2010). https://doi.org/10.1016/j.eswa.2010.02.118

Fig. 15 Random forest importance of the experiment results
10. Pan, S., Han, T., Tan, A.C.C., Lin, T.R.: Fault diagnosis system of induction motors based on multiscale entropy and support vector machine with mutual information algorithm. Shock Vib. 2016, 5836717 (2016). https://doi.org/10.1155/2016/5836717

11. Cui, J., Zheng, Q., Xin, Y., Zhou, C., Wang, Q., Zhou, N.: Feature extraction and classification method for switchgear faults based on sample entropy and cloud model. Iet Gener. Transm. Distrib. 11(11), 2938–2946 (2017). https://doi.org/10.1049/iet-gtd.2016.1459

12. Minhas, A.S., Sharma, N., Singh, G., Kankar, P.K., Singh, S.: Improvement in classification accuracy and computational speed in bearing fault diagnosis using multiscale fuzzy entropy. J. Braz. Soc. Mech. Sci. Eng. 42(11), 586 (2020). https://doi.org/10.1007/s40430-020-02671-1

13. Zhu, X., Zheng, J., Pan, H., Bao, J., Zhang, Y.: Time-shift multiscale fuzzy entropy and laplacian support vector machine based rolling bearing fault diagnosis. Entropy 20(8), 602 (2018). https://doi.org/10.3390/e20080602

14. Du, W., Guo, X., Wang, Z., Wang, J., Yu, M., Li, C., Wang, G., Wang, L., Guo, H., Zhou, J., Shao, Y., Xue, H., Yao, X.: A new fuzzy logic classifier based on multiscale permutation entropy and its application in bearing fault diagnosis. Entropy 22(1), 27 (2020). https://doi.org/10.3390/e22010027

15. Tiwari, R., Gupta, V.K., Kankar, P.K.: Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier. J. Vib. Control 21(3), 461–467 (2015). https://doi.org/10.1177/1077546313490778

16. Yasir, M.N., Koh, B.-H.: Data decomposition techniques with multi-scale permutation entropy calculations for bearing fault diagnosis. Sensors 18(4), 1278 (2018). https://doi.org/10.3390/s18041278

17. Wang, X., Si, S., Li, Y.: Multiscale diversity entropy: a novel dynamical measure for fault diagnosis of rotating machinery, IEEE Trans. Ind. Inform., pp. 1–1, (2020), doi: https://doi.org/10.1109/TII.2020.3022369.

18. Ding, J.-J.: Time Frequency Analysis And Wavelet Transform Class Note. Taipei, Taiwan: National Taiwan University (NTU) (2007).

19. Takens, F.: “Detecting strange attractors in turbulence,” In: Dynamical Systems and Turbulence, Warwick 1980, vol. 898, Springer, pp. 366–381.

20. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G.: Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011)

21. van der Maaten, L., Hinton, G.: Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605 (2008)

22. Genuer, R., Poggi, J.-M., Tuleau-Malot, C.: Variable selection using random forests. Pattern Recognit. Lett. 31(14), 2225–2236 (2010). https://doi.org/10.1016/j.patrec.2010.03.014

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.