100 year anniversary
Measurement Science and Technology

PAPER • OPEN ACCESS

Bi-filter multiscale-diversity-entropy-based weak feature extraction for a rotor-bearing system

View the article online for updates and enhancements.

Bi-filter multiscale-diversity-entropy-based weak feature extraction for a rotor-bearing system

Yongbo Li[1] **, Xinyue Wang**[1]**, Jinde Zheng**[2]**, Ke Feng**[3,]***** **and J C Ji**[4]

[1] School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, People's Republic of China
[2] School of Mechanical Engineering, Anhui University of Technology, Maanshan 243032, People's Republic of China
[3] School of Engineering, University of British Columbia, Kelowna, BC V1V 1V7, Canada
[4] School of Mechanical and Mechatronic Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia

E-mail: ke.feng@ubc.ca

CrossMark

Abstract

Multiscale-based entropy methods have proven to be a promising tool for extracting fault information due to their high feature extraction ability and easy application. Despite multiscale analysis showing great potential in extracting fault characteristics, it has some drawbacks, such as cutting the data length and neglecting high-frequency information. This paper proposes a bi-filter multiscale diversity entropy (BMDE) to filter comprehensive fault information and address the data length problem. First, the low-frequency information is filtered out by moving average in a multi-low procedure and the high-frequency information is filtered out by an adjacent subtraction in a multi-high procedure. Second, a modified coarse-grained process is introduced to overcome the issue of data length. The validity of the BMDE method is evaluated using both simulation signals and experimental measurements. Results demonstrate that the proposed method offers optimal feature extraction capability with the highest diagnostic accuracy compared with four other traditional entropy-based diagnosis methods.

Keywords: entropy theory, rotor system, bearing fault diagnosis, feature extraction

(Some figures may appear in colour only in the online journal)

1. Introduction

Rotor-bearing systems play an important role in modern production and manufacturing fields [1–4]. Timely maintenance strategies can guarantee a long service life for a rotor-bearing system, enhance its economic benefits and reduce unplanned downtime [5–7]. Feature extraction is the core step for the condition monitoring of rotor-bearing systems. To this end, entropy theory emerged from information entropy, as originally proposed by Shannon. Information entropy reflects uncertainty and complexity in a system, which laid the foundations for modern information theory and digital communication. Recently, entropy-based methods have been demonstrated to be a promising tool in extracting hidden information from measured vibration signals [8–14]. They have the advantages of independence of prior knowledge, there being no reprocessing manipulations required, and easy application [13, 14]. The most commonly used entropy methods include approximate entropy (AE) [10], sample entropy (SE) [15–17], fuzzy

***** Author to whom any correspondence should be addressed.

entropy (FE) [12, 13, 18] and permutation entropy (PE) [19]. SE can be regarded as an improvement of AE with more reliability. Seeking greater stability, FE was then developed as an improvement for SE. Using the amplitude permutation of orbits to estimate dynamical complexity, PE provided a new perspective for dynamic complexity analysis. These methods each have their own intrinsic merits and drawbacks.

In order to extract multi-dimensional features, Rostaghi *et al* [20] proposed a multiscale procedure using a course-graining operation. By virtue of the merits of this multiscale analysis, multiscale sample entropy (MSE), multiscale permutation entropy (MPE) [21] and multiscale fuzzy entropy (MFE) [18, 22] were developed, respectively, for comprehensive feature extraction [23–28].

Unfortunately, there are two main drawbacks to using multiscale analysis. From one aspect, the course-graining operation can shorten the length of the time series [29], leading to large fluctuations and inaccurate complexity evaluation [27]. From another aspect, the coarse-graining procedure involved can be considered as a low-pass filter based on a Haar wavelet [29]. By performing this process, only the low-frequency information is preserved, resulting in the loss of high-frequency information.

Aiming at overcoming the inherent drawback of shortening time series in coarse-grained algorithms, composite multiscale entropy was proposed [25]. In composite multiscale sample entropy (CMSE), at a factor of τ, the entropy value was computed for all processed time series and the CMSE value is defined as the means of τ value. High-frequency information loss was not addressed, although the data-length problem was addressed. Hierarchical procedure [30] can be regarded as an improvement of coarse-grained procedure which creates two frequency operators to multiply in the following layers. However, the signals in different layers do not distribute as high-frequency components or low-frequency components. There are also some other methods to enhance the coarse-grained method, but its drawbacks have not been satisfactorily addressed [31–33]. Therefore, it is necessary to introduce a new method to overcome these drawbacks.

This paper proposes bi-filter multiscale analysis, which can decompose a raw signal into multiple-scale time series by a multi-high procedure and multi-low procedure, simultaneously. Note that the multi-low procedure aims to extract the low-frequency information over different scales by overlapping means. Meanwhile, high-frequency information can be gained through the multi-high procedure by transplacement subtraction. The multi-low procedure and multi-high procedure use a fixed-step-size sliding window to achieve relatively invariable data length. By doing this, the comprehensive fault information embedded in the low and high spectra can be extracted using bi-filter multiscale analysis, and thereby the drawbacks of the original coarse-graining process can be mitigated. Based on the superiority of bi-filter multiscale analysis, bi-filter multiscale diversity entropy (BMDE) is generated by calculating the diversity entropy value of the processed signals.

According to the whole fault diagnosis method, the BMDE is applied to fault-feature extraction from vibrational signals first. Then, a random forest (RF) [30, 34] classifier is selected to identify the fault types using the features from the BMDE. Section 2 introduces the theoretical framework of the proposed method and the related theories. Section 3 presents a simulation model and simulated signal used to evaluate the performance of the BMDE. Section 4 validates the effectiveness of the proposed method with experimental bearing signals. Finally, section 5 summarizes the conclusions.

2. Methodology

2.1. The shortcomings of the original multiscale analysis

The BMDE method is composed of two main steps. First, to generate multiple time series by a multiscale procedure; second, to calculate the entropy value of each coarse-grained time series. Figure 1 demonstrates the original multiscale procedure. The implementation steps of the multiscale process are as follows.

Step 1. For a given time series $X = \{x_1, x_2, \ldots, x_i, \ldots, x_N\}$, where $i = 1, 2, \ldots, N$, and N indicates the data length, segment the original time series into multiple-scale time series using equation (1):

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

where τ indicates the scale factor, which should be fixed as a positive integer. The parameter τ represents the strength of the procedure.

Step 2. Calculate the BMDE value of the multiscale time series using equation (2). The detailed computational procedure can be found in [30]. The diversity entropy algorithm can be described via

$$\mathrm{BMDE}(X, m, \varepsilon) = \mathrm{DiversityEntropy}(y_j^\tau, m, \varepsilon), \qquad (2)$$

$$\begin{aligned} \mathrm{FMS} = {}& \sin(2\pi) + \sin(4\pi) + \sin(6\pi) + \sin(8\pi) \\ & + \sin(10\pi) + \sin(12\pi). \end{aligned} \qquad (3)$$

For a given time series of synthesized signals, the data length is 8196. The simulated signal is given by equation (3). The sampling frequency is fixed at 63 Hz. Figure 2 illustrates the waveforms of the multiple series. By comparing the three waveforms ($\tau = 1$, $\tau = 3$, $\tau = 5$), it is noted that the envelope of the signal tends to flatten as the scale factor increases. The length of the data becomes shorter with a larger scale factor. Figure 3 represents spectrograms of multiple series ($\tau = 1$, $\tau = 3$, $\tau = 5$). The low-frequency components are retained, and the high-frequency components are suppressed. It has been mentioned that the multiscale procedure is effectively a low-pass filter [29]. The resulting spectra
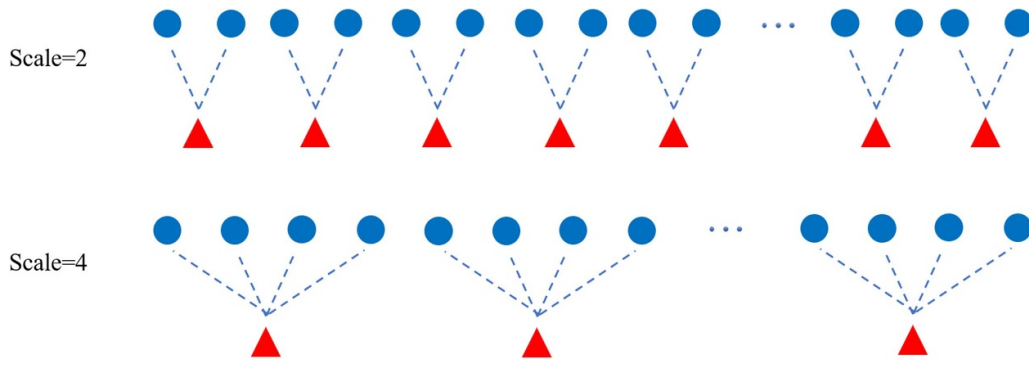
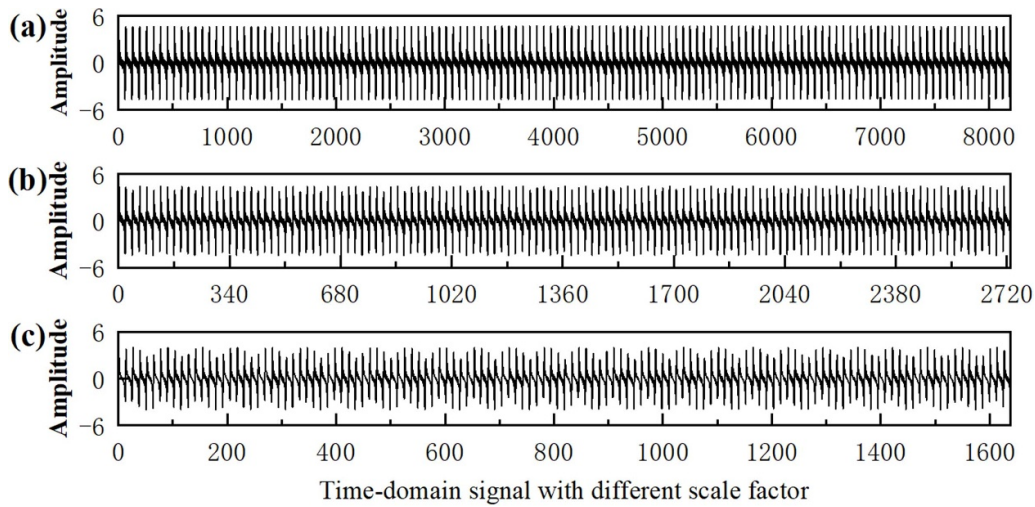**Figure 1.** Schematic of multiscale graining procedure.

Figure 2. Time-domain signal with different scale factors: (a) $\tau = 1$, (b) $\tau = 3$, (c) $\tau = 5$.

in figure 2 are consistent with multiscale theory. It demonstrates that the multiscale procedure excavates low-frequency information.

2.2. Improvements to multiscale procedure

To tackle the data-length problem, a new strategy, called bi-filter multiscale analysis, is developed in this paper. From one aspect, a modified coarse-grained process is developed to solve the difficulty of data length when using original multiscale analysis. In this paper, the BMDE utilizes a sliding window as one step, and then most of the data points are reused in the generation of the next multiple series' data points. The schematic of new strategy shows great stable in data-length in figure 4. For example, the data length of the original time series is 1000. In the traditional multiscale procedure, the length of multiple series is reduced to 100, while the new strategy generates a data length for its multiple series of 980.

From another aspect, bi-filter multiscale analysis adopts a multi-high operator to generate high-frequency information. In this strategy, some of the data points are reused to generate adjacent multiple-series data points. This can enhance the relevance of the multiple series. Traditionally, via subtraction with adjacent signal points, the noise signal is offset and high-frequency information is retained. The strategy to gain

high-frequency information is given in equation (4). In the multi-high operator, the x_j constantly subtracts with the next $\tau - 1$ points. Then, it calculates the mean value of the $\tau - 1$ results:

$$\text{multi-high operator}: z_j^\tau = x_j - \frac{1}{(\tau - 1)} \sum_{i=j+1}^{j+\tau-1} x_i, \quad 1 \leqslant j \leqslant N. \tag{4}$$

While the traditional process just considers the influence of the adjacent data points, the multi-high operator calculated by equation (4) contains N data points after x_j. The frequency-loss problem is also considered in the new strategy.

2.3. The proposed BMDE method

In the proposed BMDE method, more frequency information can be captured and the length of data kept constant. The computational procedures are introduced as follows:

Step 1. For a given time series $X = \{x_1, x_2, \ldots, x_i, \ldots, x_N\}$, $i = 1, 2, \ldots, N$, where N indicates the data length and τ represents the scale factor. When the scale factor is fixed, X will be processed with the multi-low procedure and multi-high procedure, which are given by the multi-low operator and multi-high operator as below:
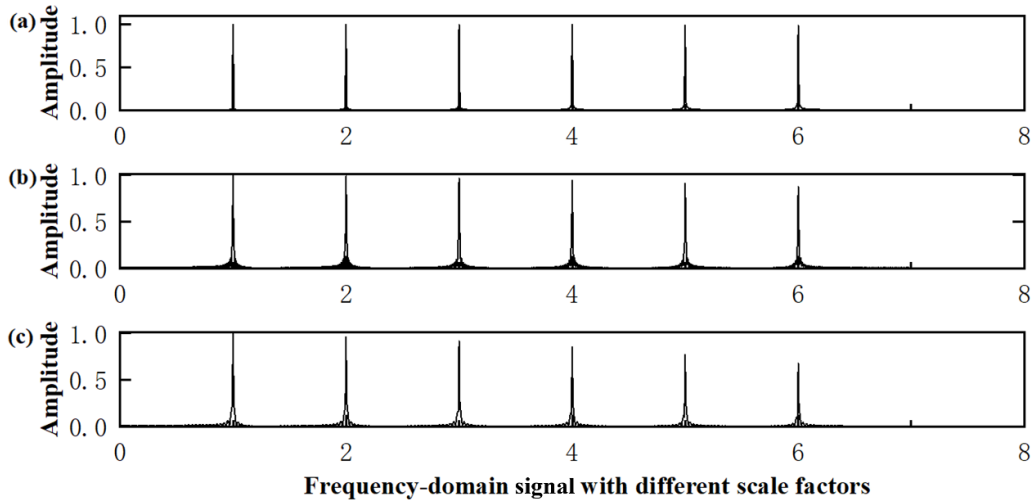
Figure 3. Signal with different scale factors in the frequency domain: (a) $\tau = 1$, (b) $\tau = 3$, (c) $\tau = 5$.
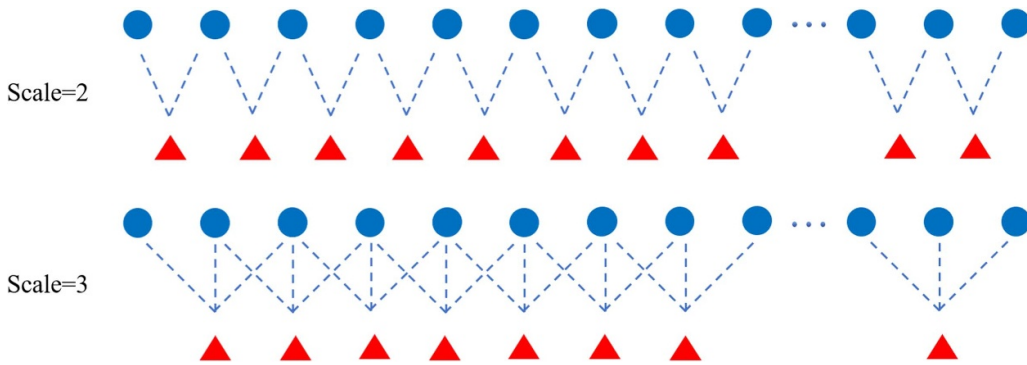

Figure 4. Schematic of the new strategy, which stabilizes the length of the signal.

$$\text{multi-low operator}: y_j^\tau = \frac{1}{\tau} \sum_{i=j}^{j+\tau-1} x_i, \quad 1 \leqslant j \leqslant N, \qquad (5)$$

$$\text{multi-high operator}: z_j^\tau = x_j - \frac{1}{\tau-1} \sum_{i=j+1}^{j+\tau-1} x_i, \quad 1 \leqslant j \leqslant N. \qquad (6)$$

Step 2. Calculate the entropy value of the two types of series in each multiple-scale time series. The expression of the BMDE is given by

$$\text{BMDE}(X, m, \varepsilon) = \text{DE}(y_j^\tau, m, \varepsilon), \text{DE}(z_j^\tau, m, \varepsilon). \qquad (7)$$

A flow chart of the proposed method is given in figure 5.

A diagram of the proposed method is shown in figure 6. It is worth noting that the process length of the time series will not be reduced dramatically. Two columns of data can be obtained using the proposed method, which represents the high-frequency components and low-frequency components, respectively.

Overall, a novel method that considers both the length of data and more frequency information is here proposed.

The BMDE values calculated by the procedure can be considered as fault features. After feature extraction by the proposed BMDE method, the separability can be evaluated by RF classification.

3. Simulation evaluation

To validate the reliability of the proposed method, a simulation bearing-structure model is constructed in this section. In this case, three types of faults (inner race fault, outer race fault and rolling element fault) are introduced. The detailed parameters of the tested bearing are listed in table 1. In the simulation, the rotation speed is set to 6000 rpm and the sampling frequency to 10 240 Hz. Structure diagrams of the three types of simulated faults are shown in figure 7.

3.1. Simulated bearing signals

3.1.1. Outer race fault model for rolling bearings. Assume that the damage point is located in the load zone of the rolling bearings and the sensor is located in the load zone with the maximum load density, as shown in figure 7(a). Suppose the damage point comes into contact with the rolling element at
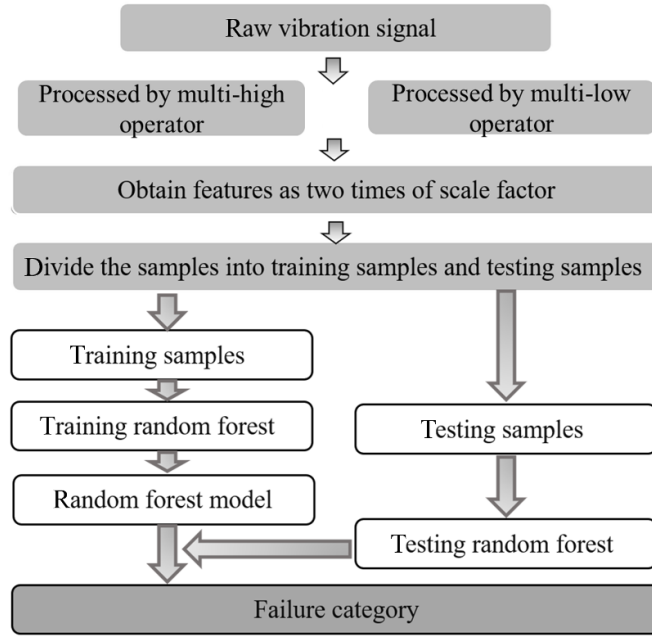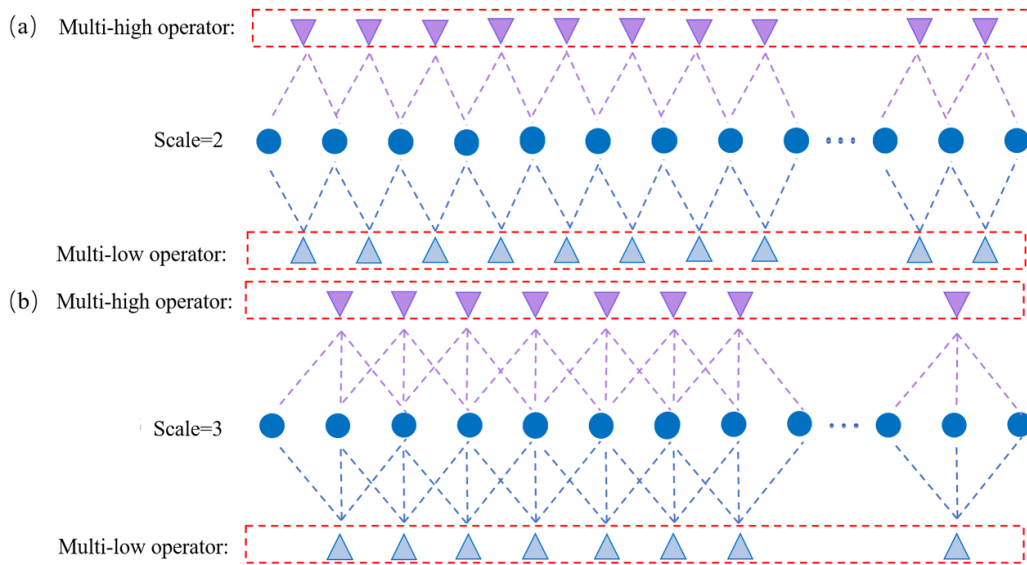
Figure 5. Flow chart of the proposed method.

Figure 6. Schematic of two procedures: (a) $\tau = 2$, (b) $\tau = 3$.

Table 1. Parameters of simulated bearing fault signal.

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

$t = 0$; then, the impulse forces induced by the local damage point on the outer race during rotation can be expressed as equation (8):

$$\Delta_{\text{o}}(t) = \sum_{t=-\infty}^{+\infty} d_{\text{o}}\delta(t - kT_{\text{o}}), \tag{8}$$

where d_{o} denotes the impulse intensity, $\delta(t)$ denotes the unit impulse function, k is the number of generated pulses, f_{o} signifies the characteristic frequency of the outer race fault and $T_{\text{o}} = 1/f_{\text{o}}$ is the time interval between two impulses.

The damping vibration function aroused by the impulse force is written as

$$e(t) = \begin{cases} \exp(-t/T), & t > 0 \\ 0 & t \leqslant 0 \end{cases}. \tag{9}$$
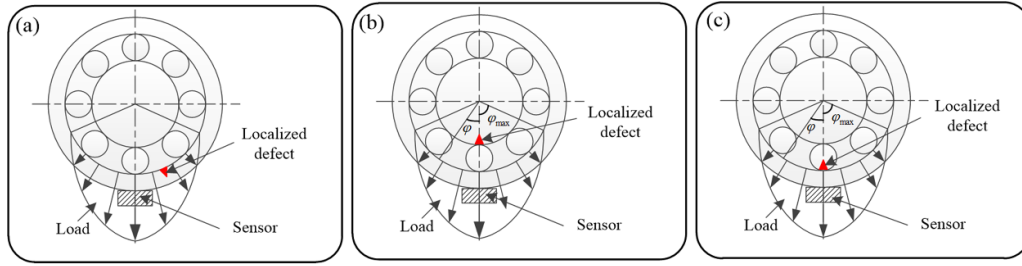
Figure 7. Structure diagrams of simulated faults. (a) Outer race fault, (b) inner race fault, (c) rolling element fault.

Because the time required for attenuation of the bearing vibration is much smaller than T_o, the damage point of the outer race will continuously generate impact pulses at frequency f_o during rotation. The outer race fault signal can be calculated by equation (10):

$$v_o(t) = \sum_{t=-\infty}^{+\infty} d_o \delta(t - kT_o) \times e(t). \tag{10}$$

3.1.2. Inner race fault model for rolling bearings. The inner race fault model of rolling bearings is shown in figure 7(b). It is assumed that the rolling elements come into contact with the damage point in the inner race at the point of maximum load and the first pulse is aroused at $t = 0$. Upon ignoring the influence of load distribution and damage-point position, a series of impulsive forces generated by the damage point can be expressed as equation (11), which is similar to the outer race fault:

$$\Delta_i(t) = \sum_{t=-\infty}^{+\infty} d_i \delta(t - kT_i), \tag{11}$$

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

The load distribution can be expressed as equation (12):

$$q(\varphi) = q_{max} \left[1 - \frac{1}{2\sigma}(1 - \cos\varphi) \right]^n. \tag{12}$$

In this paper, it is set as $n = 1.1$ and $\sigma = 0.5$.

As shown in figure 7(b), when the damage point of the inner race contacts with the rolling element at a certain angle φ, the impulsive force collected by the sensor is the projection of which on the axis of the sensor. The expression of the influence coefficient of the damage-point location is as follows:

$$p(\varphi) = \cos\varphi, \quad \varphi = 2\pi f_r t, \tag{13}$$

where f_r indicates the rotation frequency. In this way, the impulsive force on the axis of the sensor is given by equation (14):

$$f_i(t) = \Delta_i(t) q(2\pi f_r t) p(2\pi f_r t). \tag{14}$$

Finally, the simulated inner race fault signal can be expressed as equation (15):

$$v_i(t) = A_i [\Delta_i(t) q(2\pi f_r t) p(2\pi f_r t)] \times e(t), \tag{15}$$

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

3.1.3. Rolling element fault model. The simulated rolling element fault signal is similar to the bearing inner and outer simulations, and which is shown in equation (16):

$$v_b(t) = A_b [\Delta_{bo}(t) q(2\pi f_r t) p(2\pi f_r t)] \times e(t) \\ + A_b [\Delta_{bi}(t) q(2\pi f_r t) p(2\pi f_r t)] \times e(t) \tag{16}$$

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

After constructing fault models and obtaining corresponding simulated signals, figure 8 illustrates the raw time-domain signals and the corresponding spectra. The signals used are consistent with the built models.

3.2. Simulation results and analysis

To illustrate the superiority of the proposed method, the proposed BMDE is compared with hierarchical diversity entropy (HDE) and traditional multiscale diversity entropy (MDE) using the simulated signals. The length of the data used is fixed at 4096. To simulate a realistic noisy operating environment of the rotation machine, white noise interference is added to the simulated signals. The signal-to-noise ratio is set as 8. The parameters of the three methods are listed in table 2. Figure 9 shows the Euclidean distance of each set of entropy for different methods. The mean value of the features of the first 10 samples is taken as the base vector, and the vertical axis represents the Euclidean distance between the samples and the reference vector. The closer the Euclidean distance between the same fault sample and the reference vector, the better the stability of the entropy method. Furthermore, the larger the Euclidean distance between different fault samples is, the better the possible performance of the feature extraction and fault classification of the entropy method. The points numbered 1–50, 51–100 and 101–150 in (a)–(c) are generated by the simulated outer race fault signal, inner race fault signal and rolling element fault signal, respectively.
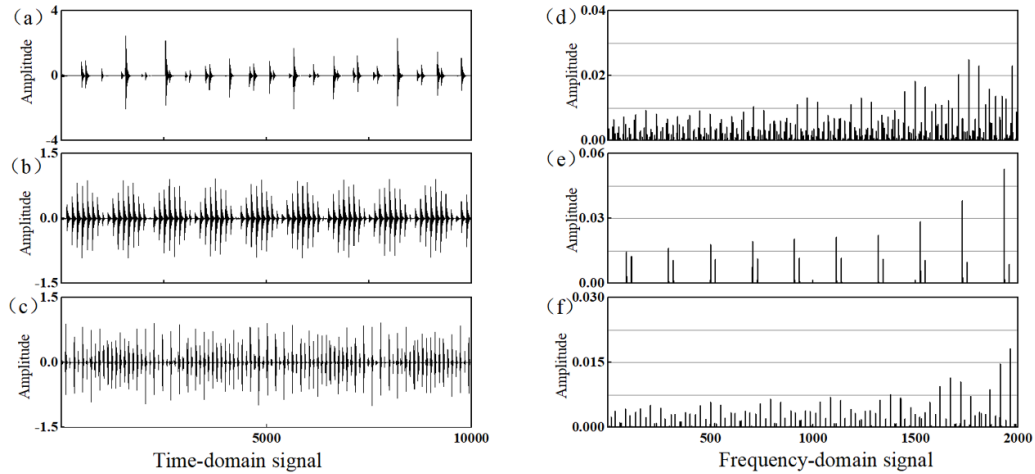
Figure 8. Simulated time-domain signals and frequency spectrum. (a) and (d) Outer fault, (b) and (e) inner fault, (c) and (f) rolling element fault.

Table 2. Parameters of applied entropy methods.

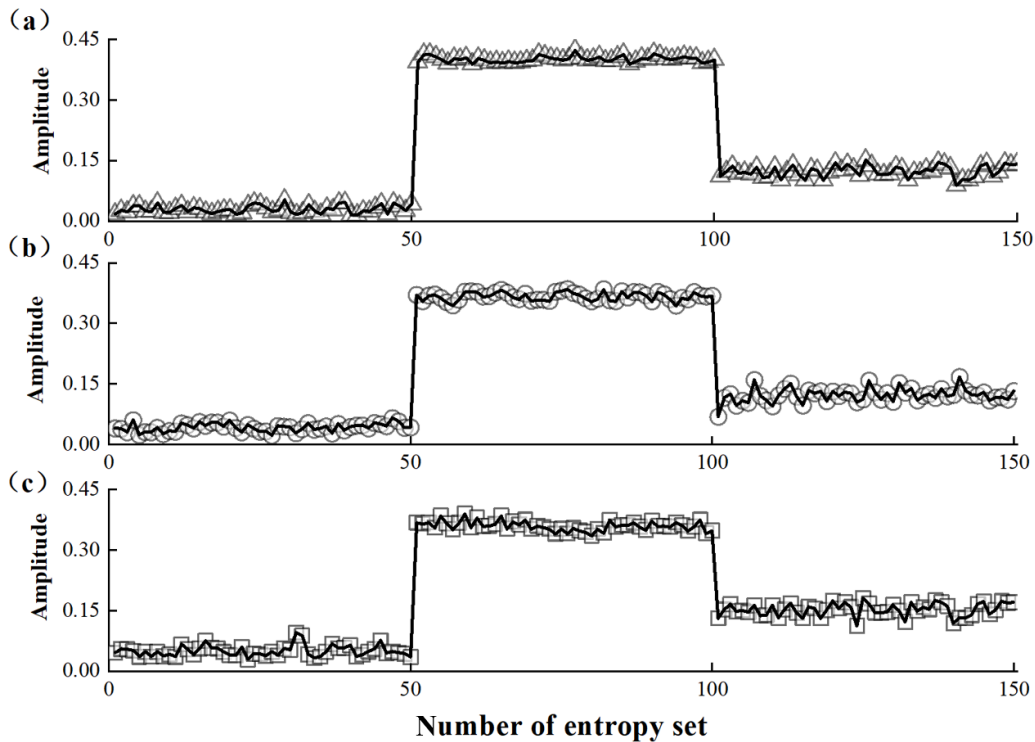| Method	m	τ	σ
BMDE	4	1	30
HDE	3	1	30
MDE	4	1	30

Figure 9. Comparison results using simulated bearing fault signals. (a) BMDE method, (b) HDE method, (c) traditional MDE method.

From figure 9, by comparing the first working condition (points numbered 1–50) in figures 9(a)–(c), the variability of BMDE is similar (the standard deviation of which is 0.0138) to HDE (the standard deviation of which is 0.0102) but the variability of MDE is much worse. While in the second health condition (points numbered 51–100), the fluctuation of BMDE, HDE and MDE is gradually exacerbated. Moreover, regarding on the third health condition (points numbered 101–150), the proposed method performs best in stability. Based on the aforementioned analysis,
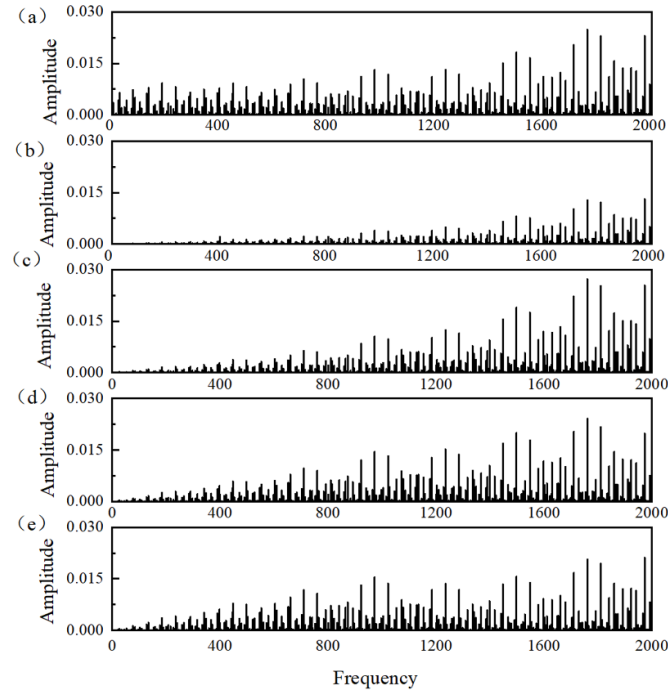
Figure 10. Spectra of signals processed by multi-high operator in different scales. (a) $\tau = 1$, (b) $\tau = 2$, (c) $\tau = 4$, (d) $\tau = 6$, (e) $\tau = 8$.
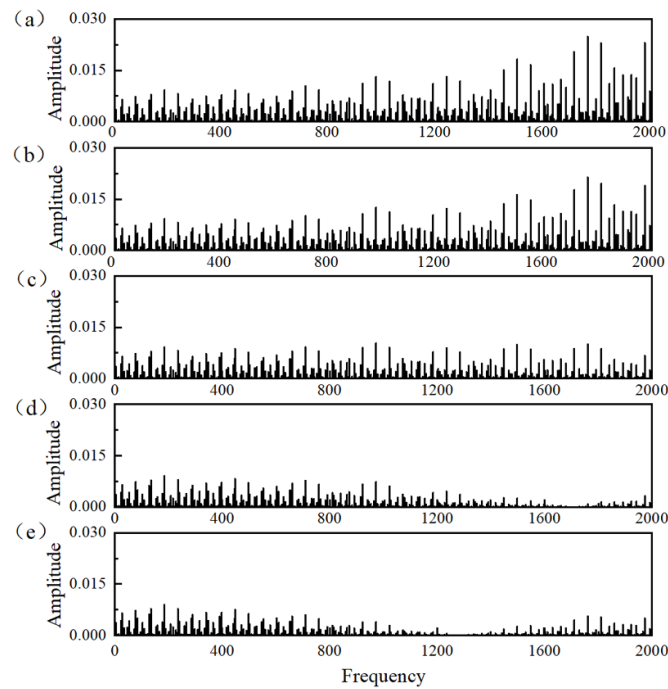

Figure 11. Spectra of signals processed by multi-low operator. (a) $\tau = 1$, (b) $\tau = 2$, (c) $\tau = 4$, (d) $\tau = 6$, (e) $\tau = 8$.

the proposed method demonstrates better distinguishability, stability and dependability than the other two entropy methods.

Figure 10 presents frequency spectra of the data used through the multi-high process. The data used are simulated based on the constructed outer race fault model. Considering

the length of the data, the scale is equal to 1, 2, 4, 6 and 8, respectively. It should be noted that figure 10(a) represents the spectrum of the original data when $\tau = 1$.

As mentioned above, the multi-high process is designed to gain high-frequency information. Compared with the raw signal spectrum in figure 10(a), the frequency component in the
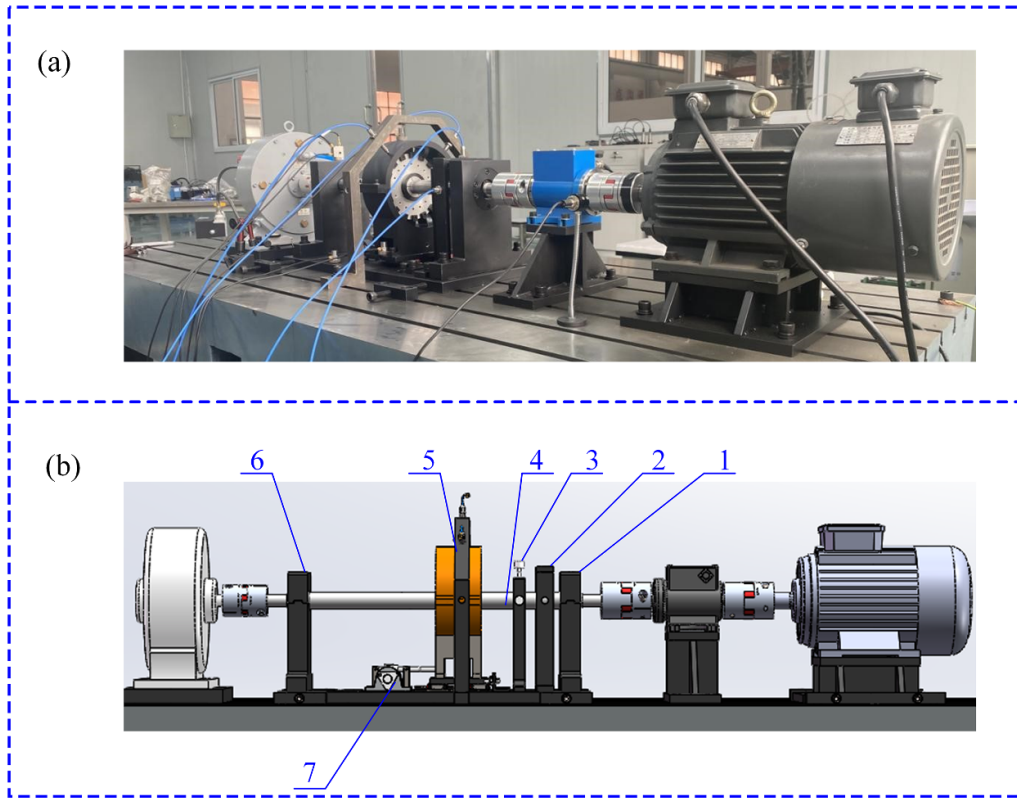
Figure 12. The rolling bearing experiment and the layout of the test rig.

Table 3. Components of the test rig.

No.	Component
1	Support bearing pedestal
2	Displacement sensor bracket
3	Friction assembly and bracket
4	Axis
5	Casing friction support and blade disc
6	Test bearing pedestal
7	Worm and worm gear

Table 4. Parameters of the experiment.

Bearing specs	6205
Number of rolling elements	7
Pitch diameter	52 mm
Rolling elements diameter	15 mm
Contact angle	0°
Sampling frequency	10 240 Hz
Rotating speed	2000 rpm

low area is attenuated and the frequency spectrum is boosted in the high area ((b)–(e) in figure 10) when the number of scales increases.

Similarly, figure 11 presents frequency spectra of the data used in the multi-low progress, which is also obtained from the constructed outer race fault model. Compared with the spectrum of the raw signal when $\tau = 1$, the frequency components gradually gather in the low-frequency area with the increase of the number of scales. The appearances of these multi-low and multi-high spectra are consistent with the theoretical derivation.

In this section, the simulation model was first established and the proposed method has been verified by comparing it with other well-known entropy methods. Then, the capacity of BMDE to capture more frequency information in spectrum processing has also been analyzed and proven.
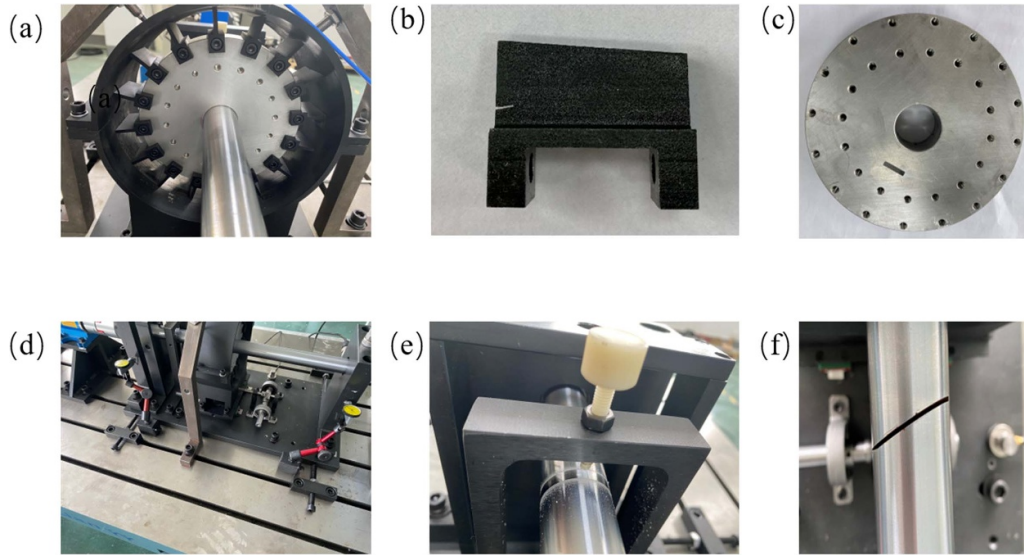
4. Experimental evaluation

To evaluate the feasibility and advantage of the proposed method in practical working conditions, an experimental test rig was set up, as shown in figure 12. The experimental system is based on a variable frequency motor, rotor shaft, rolling bearing seat, shaft system load disc, radial loading device, grinding installation bracket and coupling. The type of bearing is a 6205 bearing, and each component of the test rig in figure 12 is described in table 3.

4.1. Experimental setting

Faults are simulated by substituting the faulty parts or increasing the extent of the axis shift. More than 20 types of faults are simulated by the system, which includes single-portion failures and combined faults. Single-point failures and combined faults will be discussed, respectively. The parameters of the experiment are given in table 4. In order to obtain the

Table 5. Parameters of methods set in the experiment.

Method	Scale	Layer	Time delay	Threshold	Embedding dimension
HDE	—	3	1	—	3
MPE	20	—	1	—	6
MSE	20	—	—	—	2
MFE	20	—	—	0.15	2
BMDE	20	—	1	—	3
MDE	20	—	1	—	3

Figure 13. Practical faults diagram: (a) full annular rubbing, (b) blade crack, (c) blisk crack, (d) shaft coupling fault, (e) shaft rubbing, (f) shaft crack.

experimental data, an acceleration sensor was placed on the axis, which is labeled in figure 12. The constant rotational speed was 2000 rpm. The load was set to 40% and the sampling frequency to 10 240 Hz.

4.2. Results and discussion

The performance of the proposed method has been validated by comparing it with other classical entropy models. According to correlated references, the optimal parameter settings of the methods applied are listed in table 5 [5].

4.2.1. Single-point failure analysis.
The signal-point failures considered are a healthy condition (He), full annular rubbing (FAR), blade crack (BlaC), blisk crack (BliC), shaft coupling fault (SCF), shaft rubbing (SR) and shaft crack (SC). The blade crack faults are divided into four kinds of faults, according to the damage extent. The fault sets in the experiment rig are shown in figure 13.

To validate the performance in a practical environment, a shaft rubbing fault signal is used. In the simulated portion, the frequencies in the low-frequency band are limited, while the frequencies in the high-frequency band are emphasized. In the experimental shaft-rubbing signal, this condition still pertains.

The performance in the low-frequency procedures is similar, as well. The detailed performance is shown in figure 14.

To examine the quality of the features, features from six methods are first projected onto a two-dimensional plane with the *t*-distributed stochastic neighbor embedding(*t*-SNE) method, as shown in figure 15. The clustering shows the feature-extraction ability of the method: the smaller the intra-class distance among samples within the same cluster, and the larger the inter-class distance among clusters, the better performance the feature extraction has.

The features from the proposed BMDE method display the largest inter-class distance and smallest intra-class distance in figure 15(a). There is only a little mix in the margin of class-D and class-E, which means that the BMDE has good separability among features. In figures 15(b) and (e), the features of HDE indicate better inter-class distance than those of MPE, but the intra-class distance of MPE is smaller than that of HDE. In figures 15(c) and (d), the clusters of MDE and MFE are dispersed, and even show a relatively big mix in MFE. Furthermore, the clusters obtained by MSE mix together and can hardly be distinguished, as is shown in figure 15(f). Based on the analysis above, the BMDE method generates the most distinctive features, which are superior to other entropy methods.

Following this, the original features from the six methods are processed using the RF classifier mentioned above. Half of the data will be used to train the classifier and the rest of
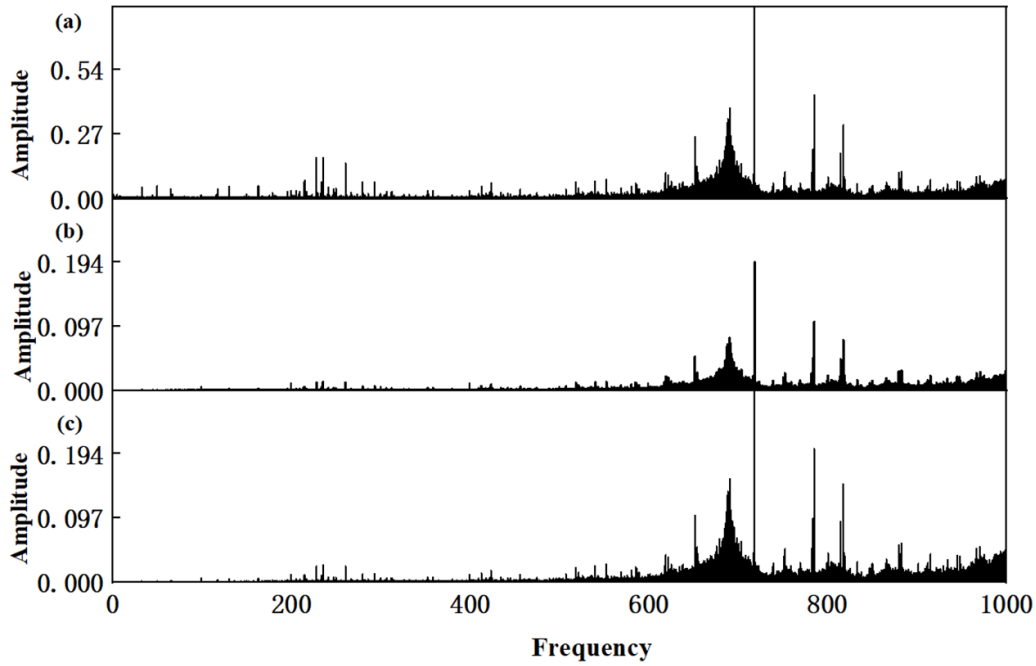
Figure 14. Analysis with experimental shaft-rubbing signal in multi-high procedure: $\tau = 1$ to 3, (a)–(c).
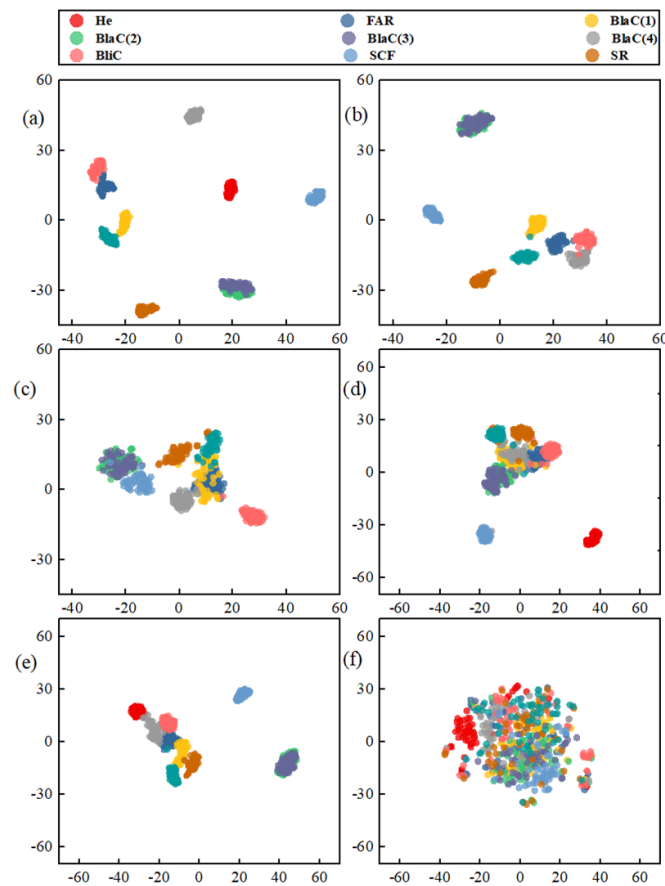

Figure 15. Scatter diagrams of features from different methods: (a) BMDE, (b) HDE, (c) MDE, (d) MFE, (e) MPE, (f) MSE.

the data used to test the identification rate. As figure 16 illustrates, the BMDE method can provide data that has the most distinctive features among the six methods. Significantly, the

identification rate of BMDE can reach up to 95.1%, which is favorable in industrial practice. The performance of HDE is notable among the other methods, the identification rate of
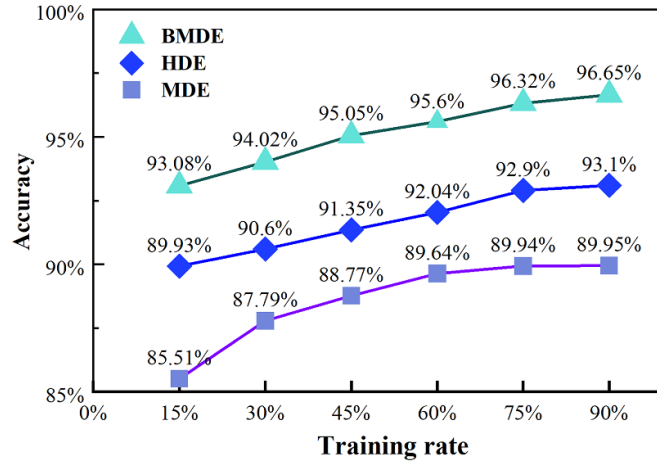
Figure 16. Identification rate distribution with the increase of training data portion.
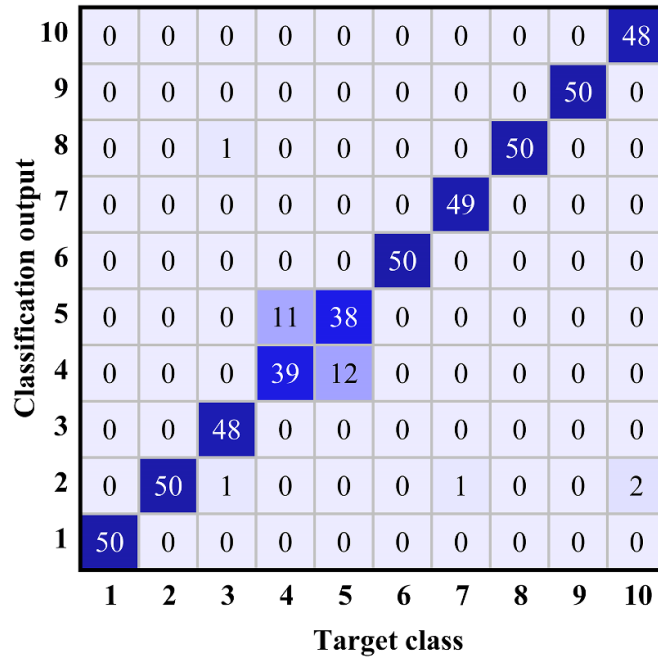

Figure 17. Confusion matrix of rotor-bearing system under 10 health conditions.

which reaches 91.76%. MDE, MPE and MFE generally show the same ability. The identification rate of MDE is barely satisfactory, according to the reported 85.51%.

In reality, the recognition accuracy will increase with augmentation of the proportion of training data. It is noticeable that when only 15% of the data is set as training data, the proposed method shows high accuracy, which is the equivalent to 90% of the data being trained in HDE, and much higher than that of MDE. This means the proposed method works accurately, especially when the data set is insufficient. While dealing with examples of real-world industrial diagnosis, data are usually limited. In this regard, BMDE demonstrates outstanding application prospects.

Figure 17 shows the identification results of the proposed BMDE. Only 28 samples are misclassified. Based on the

confusion matrix, samples of class-4 and samples of class-5 tend to be misclassified, which displays a high degree of consistency. Further, the method performs well in feature extraction with the RF classifier.

4.2.2. Combined failure analysis. In a real working environment, combined malfunctions are commonly found along with single faults, and so should be considered in any analysis. As such, 11 kinds of combined failures and health working data are collected from the experiment bench shown in figure 12. The combined failures include full annular rubbing with a shaft coupling fault and shaft crack (FARCFSF), full annular rubbing and shaft rubbing (FARSR), a blade crack(4) with shaft coupling fault (BC(4)SCF), a blisk crack with full
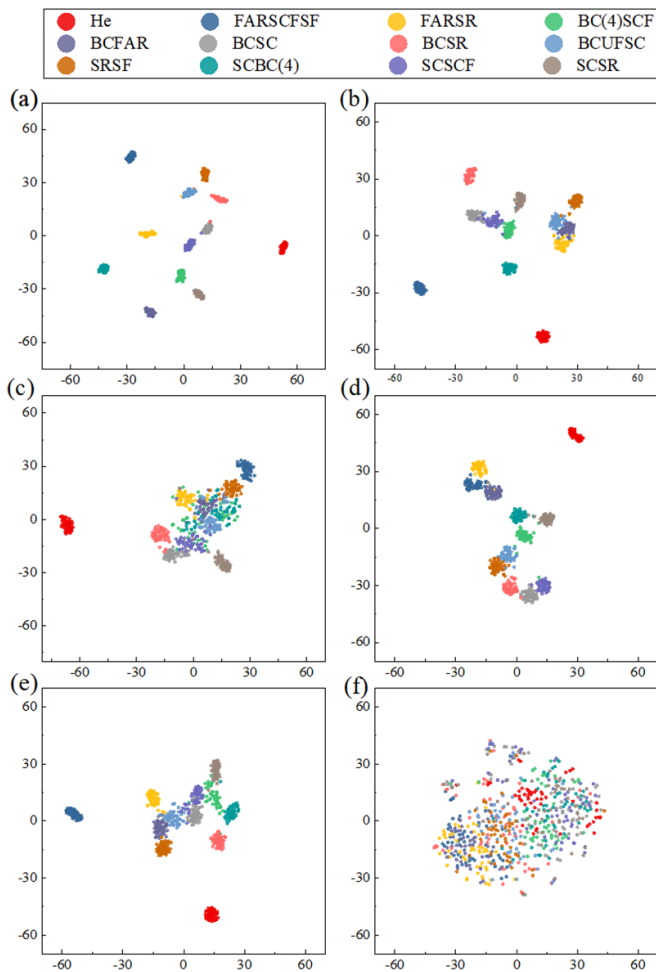
Figure 18. Scatter diagrams of features using different methods: (a) BMDE, (b) HDE, (c) MDE, (d) MFE, (e) MPE, (f) MSE.

Figure 19. Fault identification rates of the six methods.

5. Conclusion

To address the characteristic issues of entropy-based fault diagnosis using multiscale analysis, this paper proposed a bi-filter multiscale analysis method. To solve the frequency-limitation problem, a multi-low and multi-high procedure were used to filter out low-frequency and high-frequency information. To deal with the data-length loss problem, a bi-filter multiscale analysis was performed to reset the step of the sliding window to overcome the drawbacks of coarse-graining.

The bi-filter analysis is combined with diversity entropy, which thus constitutes the new BMDE method. Finally, a fault-diagnosis framework was developed based on BMDE and RF to achieve accurate fault diagnosis for a rotor-bearing system. The performance of the proposed method was verified by using simulated and experimental data. By comparing with the HDE, MDE, MFE, MPE and MSE methods, it was found that the BMDE could extract more features and provide accurate entropy values with long data lengths. Thus, it can generate the highest diagnostic accuracy and shows superior performance in resisting environmental noises compared with the aforementioned methods.

In future work, bi-filter multiscale analysis will be combined with other entropy-based methods which can provide accurate and low-volatility features. The application of bi-filter multiscale analysis shows great potential in weak feature extraction for fault diagnosis.

Data availability statement

The data cannot be made publicly available upon publication because they contain commercially sensitive information. The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Yongbo Li https://orcid.org/0000-0003-2699-9951
Ke Feng https://orcid.org/0000-0003-2338-5161

annular rubbing (BCFAR), a blisk crack with a shaft crack (BCSC), a blisk crack with shaft rubbing (BCSR), a shaft coupling unbalanced fault with a shaft crack (BCUFSC), shaft rubbing with a shaft fault (SRSF), a shaft crack with a blade crack(4) (SCBC(4)), a shaft crack with a shaft coupling fault (SCSCF) and a shaft crack with shaft rubbing (SCSR). The parameter settings for data collection remain unchanged.

First, the raw features from six methods are processed by data dimension-reduction methods to obtain a visual data distribution. The sample clusters of the proposed BMDE shown in figure 18(a) display a small intra-distance and large inter-distance, with few clusters mixed. In figures 18(b), (d) and (e), the clusters of HDE, MFE and MPE are nearly the same, with larger intra-distances and closer inter-distances than that of BMDE. However, the clusters of HDE, MFE and MPE are basically distinguishable. In addition, the result of the MSE method is far from ideal, as shown in figure 18(f).

When 50% of the data is used for training, the recognition rate can reach up to 98.2% with 12 kinds of combined failures by the RF classifier, which is shown in figure 19. The recognition rate achieved by the classifier shows high uniformity, as confirmed in the cluster diagrams. HDE, MFE, and MPE show practical unanimity after the classifier with scattergram.
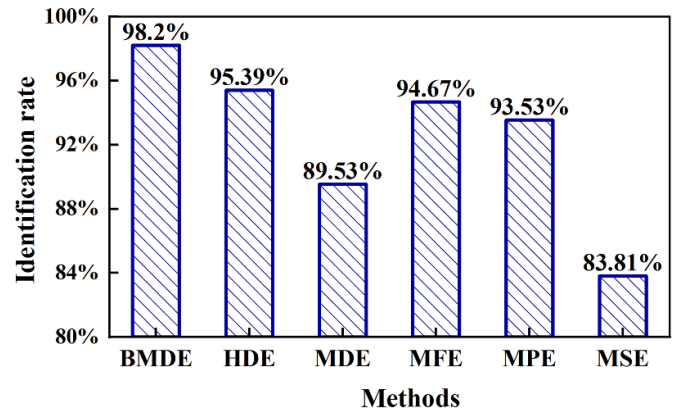
References

[1] Gao Y, Tian Y L, Liu H and Hu J W 2021 Entropy based inverse design of aircraft mission success space in system-of-systems confrontation *Chin. J. Aeronaut.* **34** 99–109

[2] Li B Y, Zhang L G, Zhu X J, Yu X Q and Ma X D 2017 Reliability analysis based on a novel density estimation method for structures with correlations *Chin. J. Aeronaut.* **30** 1021–30

[3] Fidkowski K J, Ceze M A and Roe P L 2012 Entropy-based drag-error estimation and mesh adaptation in two dimensions *J. Aircr.* **49** 1485–96

[4] Zhang W, Yan X A and Jia M P 2022 Sparse enhancement based on the total variational denoising for fault feature extraction of rolling element bearings *Measurement* **195** 111163

[5] Li Y B, Wang X Z, Liu Z B, Liang X H and Si S B 2018 The entropy algorithm and its variants in the fault diagnosis of rotating machinery: a review *IEEE Access* **6** 66723–41

[6] Feng K, Ni Q, Chen Y J, Ge J and Liu Z 2023 A cyclostationarity-based wear monitoring framework of spur gears in intelligent manufacturing systems *Struct. Health Monit.* 1–17

[7] Rostaghi M and Azami H 2016 Dispersion entropy: a measure for time-series analysis *IEEE Signal Process. Lett.* **23** 610–4

[8] Ye G D, Pan C, Huang X L, Zhao Z Y and He J Q 2018 A chaotic image encryption algorithm based on information entropy *Int. J. Bifurcation Chaos* **28** 1850010

[9] Cui T J, Liu S and Li L L 2016 Information entropy of coding metasurface *Light Sci. Appl.* **5** e16172

[10] Wang D, Zhong J J, Shen C Q, Pan E, Peng Z K and Li C 2021 Correlation dimension and approximate entropy for machine condition monitoring: revisited *Mech. Syst. Signal Process.* **152** 107497

[11] Sobczyk K 2001 Information dynamics: premises, challenges and results *Mech. Syst. Signal Process.* **15** 475–98

[12] Cao Z H, Lin C T, Lai K L, Ko L W, King J T, Liao K K, Fuh J L and Wang S J 2020 Extraction of SSVEPs-based inherent fuzzy entropy using a wearable headband EEG in migraine patients *IEEE Trans. Fuzzy Syst.* **28** 14–27

[13] Tan A H, Shi S W, Wu W Z, Li J J and Pedrycz W 2022 Granularity and entropy of intuitionistic fuzzy information and their applications *IEEE Trans. Cybern.* **52** 192–204

[14] Müller L M, Dupuis F, Szehr O, Fehr S and Tomamichel M 2013 On quantum Rényi entropies: a new generalization and some properties *J. Math. Phys.* **54** 122203

[15] Delgado B A and Marshak A 2019 Approximate entropy and sample entropy: a comprehensive tutorial *Entropy* **21** 541

[16] Feng K, Wang K S, Ni Q, Zuo M J and Wei D 2017 A phase angle based diagnostic scheme to planetary gear faults diagnostics under non-stationary operational conditions *J. Sound Vib.* **408** 190–209

[17] Yao P, Zhou K and Zhu Q 2017 Quantitative evaluation method of arc sound spectrum based on sample entropy *Mech. Syst. Signal Process.* **92** 379–90

[18] Li Y B, Wang S, Yang Y and Deng Z C 2022 Multiscale symbolic fuzzy entropy: an entropy denoising method for weak feature extraction of rotating machinery *Mech. Syst. Signal Process.* **162** 118052

[19] Yan R Q, Liu Y B and Gao R X 2012 Permutation entropy: a nonlinear statistical measure for status characterization of rotary machines *Mech. Syst. Signal Process.* **29** 474–84

[20] Rostaghi M, Ashory M R and Azami H 2019 Application of dispersion entropy to status characterization of rotary machines *J. Sound Vib.* **438** 291–308

[21] Zheng J D, Pan H Y, Yang S B and Cheng J S 2018 Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis *Mech. Syst. Signal Process.* **99** 229–43

[22] Zheng J D, Pan H Y and Cheng J S 2017 Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines *Mech. Syst. Signal Process.* **85** 746–59

[23] Li Y B, Li G Y, Yang Y T, Liang X H and Xu M Q 2018 A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy *Mech. Syst. Signal Process.* **105** 319–37

[24] Shao K X, Fu W L, Tan J W and Wang K 2021 Coordinated approach fusing time-shift multiscale dispersion entropy and vibrational Harris hawks optimization-based SVM for fault diagnosis of rolling bearing *Measurement* **173** 108580

[25] Wu S D, Wu C W, Lin S G, Lee K Y and Peng C K 2014 Analysis of complex time series using refined composite multiscale entropy *Phys. Lett.* A **378** 1369–74

[26] Yan X A and Jia M P 2019 Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection *Knowl.-Based Syst.* **163** 450–71

[27] Wang X Z, Si S B and Li Y B 2022 Variational embedding multiscale diversity entropy for fault diagnosis of large-scale machinery *IEEE Trans. Ind. Electron.* **69** 3109–19

[28] Li Y B, Yang Y T, Wang X Z, Liu B B and Liang X H 2018 Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine *J. Sound Vib.* **428** 72–86

[29] Wu S D, Wu C W, Lin S G, Wang C C and Lee K Y 2013 Time series analysis using composite multiscale entropy *Entropy* **15** 1069–84

[30] Dou J, Yunus A P, Tien B D, Merghadi A, Sahana M, Zhu Z F, Chen C-W, Khosravi K, Yang Y and Pham B T 2019 Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan *Sci. Total Environ.* **662** 332–46

[31] Feng K, Ji J C, Ni Q and Beer M 2023 A review of vibration-based gear wear monitoring and prediction techniques *Mech. Syst. Signal Process.* **182** 109605

[32] Feng K, Ji J C, Zhang Y C, Ni Q, Liu Z and Beer M 2023 Digital twin-driven intelligent assessment of gear surface degradation *Mech. Syst. Signal Process.* **186** 109896

[33] Li Y, Feng K, Liang X and Zuo M J 2019 A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy *J. Sound Vib.* **439** 271–86

[34] Li T Y and Zhou M 2016 ECG classification using wavelet packet entropy and random forests *Entropy* **18** 285