



Use of generalized refined composite multiscale fractional dispersion entropy to diagnose the faults of rolling bearing

Jinde Zheng · Haiyang Pan

Received: 22 August 2019 / Accepted: 13 July 2020 / Published online: 13 August 2020
© Springer Nature B.V. 2020

Abstract A typical symptom of vibration signals collected from rolling bearings with local faults is the existence of periodic transients, which makes the intrinsic structure of vibration signals become more and more regular. Generally, the vibration always contains multiple intrinsic oscillatory modes on different scales, which generally is caused by the interaction and coupling of machine components. Therefore, it is necessary to detect the behavior change of complexity of vibration signals in the view of multiple scales for fault information representation. The complexity and nonlinear failure symptom of rolling bearing can be evaluated by the recently proposed nonlinear dynamic tools, such as multiscale entropy (MSE) and its variants. Recently, the improved MSE method, multiscale dispersion entropy (MDE) and its improvement refined composite MDE (RCMDE) are developed to measure the complexity of time domain data. However, the intrinsic shortages of coarse graining approach that used in MDE and RCMDE have limited their application to fault feature representation. In this paper, an improved RCMDE

approach named generalized refined composite multiscale fluctuation-based fractional dispersion entropy (GRCMFDE) is proposed to enhance MDE and RCMDE in complexity measurement of time series. GRCMFDE was compared with MPE, MDE, RCMDE by analyzing synthetic simulation signals to verify its advantages. After that, an intelligent fault diagnosis method was proposed by combining GRCMFDE with supervised multi-clustering feature selection and gray wolf optimized SVM for fault classification of rolling bearing. Lastly, the proposed fault diagnostic method was applied to two experimental data set analysis by comparing with multiscale permutation entropy, MDE- and RCMDE-based fault diagnostic methods and the comparison results indicate that the proposed method can effectively diagnose the fault locations and severities of rolling bearing and get a higher fault identifying rate than the comparative methods.

Keywords Complexity · Fault diagnostic method · Multiscale dispersion entropy · Multiscale permutation entropy · Rolling bearing

J. Zheng (✉) · H. Pan
School of Mechanical Engineering, Anhui University of
Technology, Maanshan 243032, Anhui, China
e-mail: lqdlzheng@126.com

J. Zheng
School of Mechanical and Manufacturing Engineering,
University of New South Wales, Sydney,
NSW 2052, Australia

1 Introduction

Rolling bearing has been a key part of many rotating machines and other equipment; meanwhile, it is also the device most prone to failure. When the rolling

bearing emerges local faults, a typical symptom of their vibration signals is the existence of periodic transients [1], which makes the intrinsic structure of vibration signals become more and more regular. This regularity is presented in the form of self-similarity or complexity of vibration signals. In addition, the vibration signal of rolling bearing always contains multiple intrinsic oscillatory modes on different scales caused by interaction and coupling effects of machine components. Therefore, it is necessary to detect the complexity or irregularity change of vibration signals of rolling bearing in multiple scale views.

Generally speaking, many statistical and nonlinear dynamic parameters can be used to evaluate the complexity and nonlinear features of time domain data of rolling bearing, among which, the entropy-based complexity measure approaches have attracted lots of researchers' attention [2]. Multiscale entropy (MSE) and multiscale permutation entropy (MPE) [3–5] are two kinds of most often used methods for complexity and randomness measure in mechanical fault detection and diagnostics fields, and many fruitful research results have achieved by scholars. For example, the MSE-based statistical features were constructed and employed to reflect the fault information of rolling bearing by Zhang, et al. [6]. However, the computation of sample entropy (SampEn) is time costing and the similarity measurement of template changes suddenly for its use of Heaviside step function. To improve the performance of SampEn used in MSE, in Ref. [7, 8] the improved MSE method called multiscale fuzzy entropy was proposed and used to fault feature extraction of machinery system. MPE was used to represent the nonlinear fault features of rolling bearing by Li et al. [9]. GCMPE was developed for enhancing MPE in entropy fluctuation at large scales and utilized to fault detection of rolling bearing in Ref. [10]. However, in the computation of permutation entropy (PE) used in MPE, the amplitude information of time series is ignored. Besides, the coarse-grained process used in MSE and MPE in nature is linear mean filtering, which will cause important information missing of the analyzed signals [11].

Dispersion entropy (DE) proposed in Ref. [12] was designed as a new nonlinear dynamic indicator to overcome the drawbacks of SampEn and PE for complexity and irregularity measurement. Compared with SampEn and PE, the amplitude size of time series is considered in DE and its computation is much faster.

Most of all, DE has much stronger anti-noise capability than PE and SampEn, as a slight change of amplitude does not change the corresponding class label defined in DE. DE also was expanded to multiscale dispersion entropy (MDE), as well as its improvement, refined composite multiscale dispersion entropy (RCMDE) [13]. However, there are still some problems existed in MDE and RCMDE that need to be solved. First, in the original DE, the fluctuation of patterns was not distinguished. Second, the optimized mapping approach that suitable for the vibration signals of rolling bearing should be selected. Third, for a large scale factor, the coarse graining used in MDE and RCMDE always leads to the length of time series be shorter and the deviation of MDE will be larger accordingly [13].

Aiming at the first and second issues stated above, the fluctuation-based DE (FDE) is used to replace the original DE and is improved in the selection of mapping approach. Also, the original coarse graining is extended to the root-mean-square-based generalized refined composite multiscale way to overcome its intrinsic limitations. Last, the DE is extended to fractional order to improve the anti-noise performance of DE. Based on that, the generalized refined composite multiscale fluctuation-based fractional order dispersion entropy (GRCMFDE) is developed in this paper to enhance the performance of MDE and RCMDE for measuring irregularity and complexity of time series. In GRCMFDE, first, the coarse graining multiscale used in MDE is extended to the generalized coarse graining where, i.e., the first moment (mean) is extended to root-mean-square, which can preserve the information of original data effectively without missing of important information. Second, the refining and composite operation is used to avoid the undefined or imprecise DE and its fluctuation on large scale factors [14]. Third, consider that the fractional order entropy is a useful tool for dynamic description of complex systems and shows higher sensitivity to signal evolution [15, 16], the fractional DE is developed to enhance DE and further fractional order is used for anti-noise.

Because of the interference of background noises, the rolling bearing vibration signals generally are random signals obeying normal distribution without prominent periodic content [17]. The randomness and dynamic complexity will change once the rolling bearing runs with local failures and the characteristic

information of rolling bearing vibration signal related with fault often is distributed over different scales because of the complexity of mechanical system. Therefore, GRMCFDE can be utilized to extract the failure complexity features distributed in multiple scales. The complexity information related with local faults mainly can be extracted. When the GRMCFDE-based fault features are obtained and distributed at different scale factors, which generally contains redundant information unrelated with fault. A suitable feature selection approach is required to map the high dimensional features to a subset that can preserve the most important and intrinsic information of initial features. The selected features of the recently developed dimensionality reduction approach, multi-class feature selection (MCFS), can best preserve the cluster structure of data. Therefore, in this paper MCFS is employed to improve the efficiency of failure mode identification [18].

After that, a multi-classifier should be designed to achieve an intelligent recognition of failure modes of rolling bearings. As a commonly used supervised machine learning tool, support vector machine (SVM) is very suitable for addressing cases of small samples, nonlinear and high dimensional problems. Furthermore, SVM has avoided the local minimum point problem in the structure selection of neural network and the generalization performance can be improved by kernel function and learning. However, the identification of SVM-based multi-classifier heavily depends on the penalty parameter c and the parameter g used in the kernel function (radial basis function, RBF), which is introduced to balance the empirical risks and model complexity. To search the optimal parameters of c and g used in SVM, gray wolf optimization algorithm (GWO) is employed to obtain the best parameters, i.e., GWOSVM is used to construct an optimized multi-classifier for intelligent fault diagnosis [19, 20]. Then, based on GRMCFDE, MCFS and GWOSVM, an intelligent fault diagnosis approach is proposed for rolling bearing. The proposed fault diagnostic method was applied to experimental data analysis of rolling bearing. Also, it was compared with the MPE-, MDE- and RCMDE-based fault diagnostic methods and the comparison results show that it can distinguish the fault severity and classes of rolling bearings effectively and gets better performance of fault diagnostic than the comparative methods.

The rest of this paper is organized as follows. Dispersion entropy, the improved fluctuation-based dispersion entropy, multiscale dispersion entropy and the refined composite multiscale dispersion entropy are reviewed in Sect. 2. Generalized refined composite multiscale normalized dispersion entropy (GRMCFDE) is proposed in Sect. 3, as well as the comparison analysis of GRMCFDE with MPE, MDE and RCMDE. Section 4 introduces the GRMCFDE-, MCFS- and GWOSVM-based fault diagnosis approach for rolling bearing with the applications and comparison with two experiment data cases. The final section concludes the paper.

2 Improved FDE algorithm

DE is a nonlinear irregularity and complexity measure tool of time series, and it takes count into the amplitude value relationship of original data that is ignored in PE; meanwhile, it does not need to rank each embedded vector according to magnitude or calculate the distance between two different delay vectors implemented in SampEn. Besides, DE also has stronger anti-noise ability than SampEn and PE as the small changes of amplitude do not change the corresponding class label of amplitude value. The detailed computation steps of DE can be found in [12, 13] or see “Appendix A” section.

In DE algorithm, when all probability of distribution patterns $p(\pi_{v_0 v_1 \dots v_{m-1}})$ are equal, DE gets the largest entropy value $\ln(c^m)$ and a typical example is Gaussian white noise. In contrast, when the probability of distribution pattern $p(\pi_{v_0 v_1 \dots v_{m-1}})$ is unitary, i.e., only one value is not equal to zero, DE get the smallest value, which indicates that the time series is a completely predictable data and a typical example is the periodic signal with low frequency. In some application, the local or global trend of the analyzed data needs to be removed. In DE algorithm, for $\{1,1,1\}$ and $\{2,2,2\}$, or $\{1,2,4\}$ and $\{2,2,3\}$, these dispersion patterns in fact have no differences, because their fluctuation is equal. In FDE, we consider the differences between adjacent elements of dispersion patterns and generally we term them as fluctuation-based dispersion patterns (FDPs).

The vectors with length $m - 1$, whose element changes from $-c + 1$ to $c - 1$ are obtained and thus we

can obtain $(2c-1)^{m-1}$ potential FDPs. This is the only difference between DE and FDE. Also, the FDE is normalized by dividing $\ln((2c-1)^{m-1})$. In addition, compared with DE, the mapping approach NCDF shown in Eq. (1) is replaced by log-sigmoid (logsig) function defined as $y_j = 1 / (1 + e^{-\frac{y_j - \mu}{\sigma}})$ to get a more accurate complexity assessment, where σ and μ represent the standard deviation and the mean of X , respectively.

The following example is used to illustrate the difference of DE and FDE. The random signal with 20 points is given as $X = \{0.2944, -1.3362, 0.71432, 1.6236, -0.69178, 0.8580, 1.2540, -1.5934, -1.4410, 0.5711, -0.39989, 0.690, 0.8156, 0.7119, 1.2902, 0.86, 1.1908, -1.2024, -0.01979, -0.1567\}$. X is mapped to Y with values belonging to $[0,1]$ by using 'logsig' function. Then, Y is mapped to the class function Z belonging to $\{1, 2, 3\}$ by using Eq. (2), when we set $m = 2$ and $c = 3$ as an example and the results are shown in Fig. 1a. In DE algorithm, we get $c^m = 9$ dispersion patterns, while in FDE we get $(2c - 1)^{m-1} = 5$ FDPs. The probability of the dispersion patterns in DE and the FDP in FDE are shown in Fig. 1b.

3 Generalized refined composite multiscale fluctuation-based dispersion entropy

3.1 GRCMFDE algorithm

MDE has overcome the defects of DE that only measure the complexity in single scale. However, the coarse graining-based multiscale approach used in MDE heavily depends on the length of data and if the scale factor increases, the entropy deviation over multiple scales will increase, correspondingly. In addition, as the amplitudes of coarse-grained time series is obtained by computing the mean of all values in each non-overlapping segment, which inevitably leads to loss of much potentially useful information.

In this subsection, to solve the problems existed in MDE (see "Appendix A" section), GRCMFDE is developed as follows.

- (1) For a given data $\{x(i), i = 1, 2, \dots, N\}$, the generalized coarse-grained time series $x_k^{(\tau)}$ =

$$\left\{ x_{k,j}^{(\tau)} \right\}_{j=1}^{N_\tau} \text{ is defined as}$$

$$y_{k,j}^{(\tau)} = \sqrt{\frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i^2}, \quad 1 \leq j \leq \lfloor N/\tau \rfloor \quad (1)$$

where $\tau = 1, 2, 3, \dots$. When $\tau = 1$, c is defined as the absolute value of original time series. In Eq. (8) of "Appendix A" section, the coarse graining time series is defined by averaging each non-overlapping segment, which inevitably

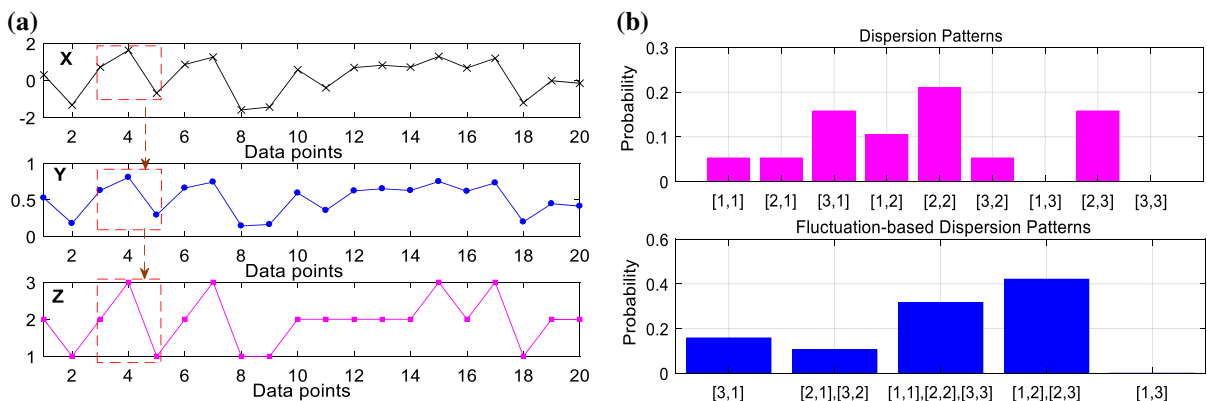


Fig. 1 Illustration of FDE versus DE algorithms. **a** X was mapped to Y belonging to $[0,1]$, then mapped Z belonging to $\{1, 2, 3\}$ and **b** the probabilities of dispersion patterns $\{11, 12, 13, 21, 22, 23, 31, 32, 33\}$ in DE and $\{11, 12, 13, 21, 31\}$ in FDE

leads to the loss of potentially useful information. In the literature [21], the average operation is extended to the second-order statistic by $y_{k,j}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} (x_i - \bar{x}_i)^2$. This approach may be suitable for permutation entropy, which is sensitive to the adjacent amplitude relationship, but this approach is not suitable for DE, as lots of amplitude information will loss in this way.

- (2) For a given $\tau (\geq 2)$, τ generalized coarse-grained data $y_k^{(\tau)} = \{y_{k,j}^{(\tau)}\}_{j=1}^{N/\tau}$ can be obtained by Eq. (7). For each $y_k^{(\tau)} (k = 1, \dots, \tau)$, the FDP $p_{\tau}^k(\pi_{v_0 v_1 \dots v_{m-1}})$ can be estimated according to steps (1) to (4) of DE ($k = 1, 2, \dots, \tau$) by considering the modification of FDE. Next, $\bar{p}(\pi_{v_0 v_1 \dots v_{m-1}}) = \sum_{k=1}^{\tau} p_{\tau}^k(\pi_{v_0 v_1 \dots v_{m-1}})$ is computed as the final average dispersion patterns at scale factor τ .
- (3) Finally, the GRCMFDE of $\{x(i), i = 1, 2, \dots, N\}$ at the scale τ is defined by

$$\text{GRCMFDE}(X, m, c, d, \tau) = - \frac{1}{\ln((2c - 1)^{m-1})} \sum_{\pi=1}^{(2c-1)^{m-1}} \bar{p}(\pi_{v_0 v_1 \dots v_{m-1}}) \cdot \ln \bar{p}(\pi_{v_0 v_1 \dots v_{m-1}}) \tag{2}$$

- (4) Based on the fluctuation-based calculus and generalized expression of Shannon entropy, GRCMFDE is generalized to fractional order domain, and for convenience the generalized one is still noted as GRCMFDE. We define $\text{GRCMFDE}_{\alpha} = -D^{\alpha} \text{GRCMFDE}$ and $D^{\alpha}(\cdot)$ denotes the derivative of order α . Finally, GRCMFDE_{α} of original data $\{x(i), i = 1, 2, \dots, N\}$ is given by

$$\begin{aligned} &\text{GRCMFDE}_{\alpha}(X, m, c, d, \tau) \\ &= \sum_{\pi=1}^{c^m} \left\{ -\frac{\bar{p}^{-\alpha}(\pi_{v_0 v_1 \dots v_{m-1}})}{\Gamma(\alpha + 1)} [\ln(\bar{p}(\pi_{v_0 v_1 \dots v_{m-1}}))] \right. \\ &\quad \left. + \psi(1) - \psi(1 - \alpha) \right\} \bar{p}(\pi_{v_0 v_1 \dots v_{m-1}}) \end{aligned} \tag{3}$$

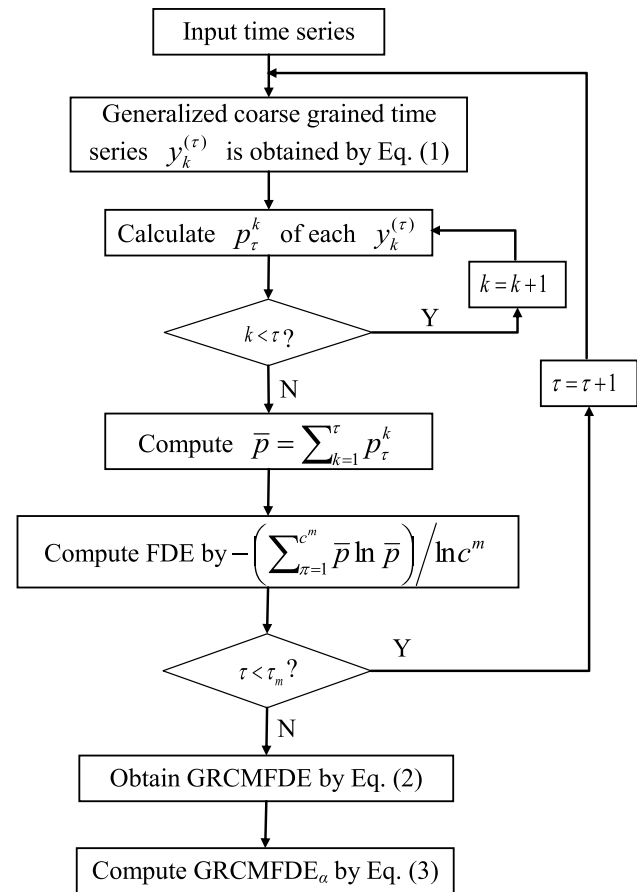
where $-1 < \alpha < 1$, α denotes the fraction order and Eq. (3) is the same as the expression of GRCMFDE in Eq. (2) when $\alpha = 0$. Γ and ψ represent the gamma and digamma functions, respectively. In fact, α less than 0 makes that $D^{\alpha}(\cdot)$ a fractional integral. It should be pointed out that the fractional entropy is a novel expression for entropy inspired in the properties of fractional calculus.

- (5) Let $\tau = \tau + 1$ and repeat the steps (2) and (4) until $k = k + 1$, where τ is the preset largest scale factor.
- (6) All τ GRCMFDE values are taken as a function of scale factor.

The calculation process of GRCMFDE is described and shown in Fig. 2.

In original coarse graining, the mean of each segment is computed, while in [21] the second-order statistic (variance or standard deviation) of each segment is computed. In Eq. (1), the RMS of each segment is computed to enhance original coarse graining. In this part, the simulated vibration signal of rolling bearing with amplitude modulation of $x(t) = (1 + 0.5 \cos(2\pi 8t)) \cos(2\pi 100t)$ is utilized to illustrate the difference of four different coarse graining methods. Figure 3a, f gives the time domain of $x(t)$ and $|x(t)|$. Figure 3b–e gives the mean, standard deviation (STD), variation and root-mean-square (RMS) ways for constructing coarse graining with $\tau = 2$, while Fig. 3g–j gives the results when $\tau = 3$. From Fig. 3, it can be found that the mean computation is similar to down sampling or linear filter, while the results of standard deviation and variation are very similar and lots of amplitude modulated information are lost at high scale factor. The proposed RMS-based generalized coarse graining way can effectively preserve the amplitude modulated information. Therefore, in this paper, RMS-based generalized coarse graining is used for multiscale computation of time series. To sum up, to overcome the shortages of DE

Fig. 2 The flowchart of GRCMFDE



and the traditional coarse graining-based multiscale method, the following improvements were made in GRCMFDE. First, the conventional coarse graining multiscale approach is extended to RMS-based generalized coarse graining multiscale way shown in Eq. (1). Second, when we estimate DE at scale factor τ , the original signal was firstly divided into $\lfloor N/\tau \rfloor$ segments with a length of τ and the starting points of x_1, x_2, \dots, x_τ , and then τ coarse graining sequences are averaged. Third, when calculating GRCMFDE, the probability of fluctuation-based dispersion pattern π of each generalized coarse-grained time series is calculated and then the average of the probabilities of these fluctuation-based dispersion pattern is computed. This refining grained processing can effectively reduce the loss of some statistical information during MDE algorithm. Meanwhile, probability averaging of multiple time series with different initial points can effectively alleviate the influence of large scale factor on fluctuation of entropy curve and reduce the

calculation deviation. Finally, FDE is extended to the fractional order according to Eq. (9) to suppress the impact of noise.

3.2 Parameter selection and comparison analysis

The calculation of GRCMFDE is related to embedding dimension m , class number c , time delay d and fractional order α . The literature [13] recommended that in general m gets a value of 2 or 3, $d = 1$, c gets an integer between 3 and 9. We will study the influence of these parameters in the following part. First, the selection of m is examined. We take the white, pink and blue noises, which though are random signals, have different distribution in the spectrum and thus different intrinsic structure, as examples to see if the proposed method can differ them from each other. The GRCMFDE of 20 different white, pink and blue noises with length 5000 points are computed when $m = 2, 3, 4$ and $c = 6, d = 1, \alpha = 0$ to study the influence of m on

Fig. 3 **a** and **f** are the time domain of $x(t)$ and $|x(t)|$. **b** obtained by using Eq. (8), **c** obtained by using standard deviation in right of Eq. (1), **d** obtained by using variation in right of Eq. (1), **e** obtained by using Eq. (1) for $\tau = 2$ and **g–j** are the corresponding results for $\tau = 3$

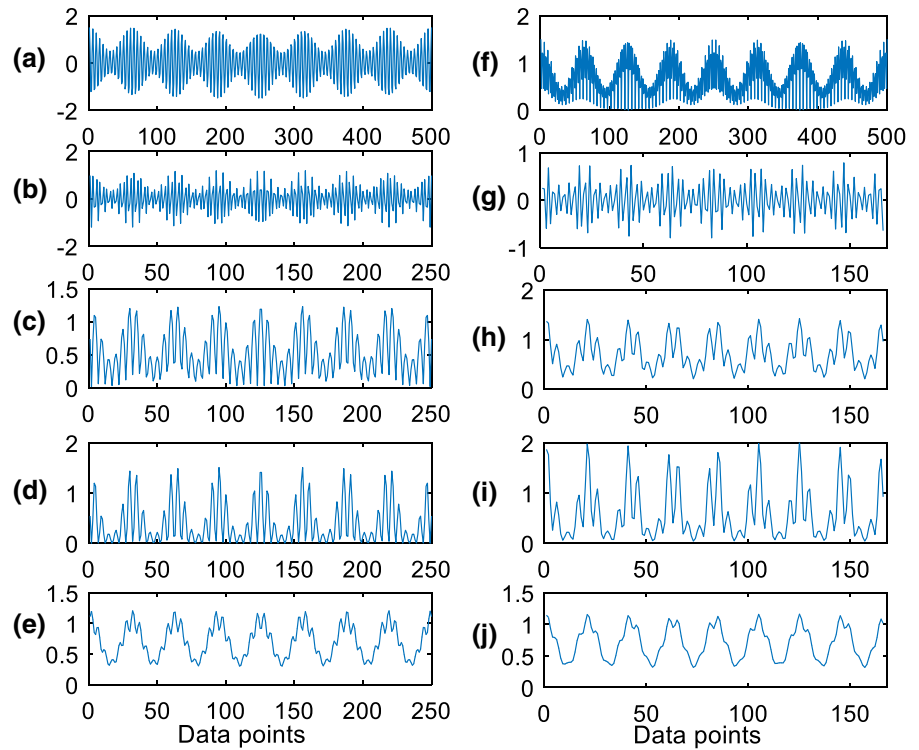


Fig. 4 GRCMFDE of three noise for different m values

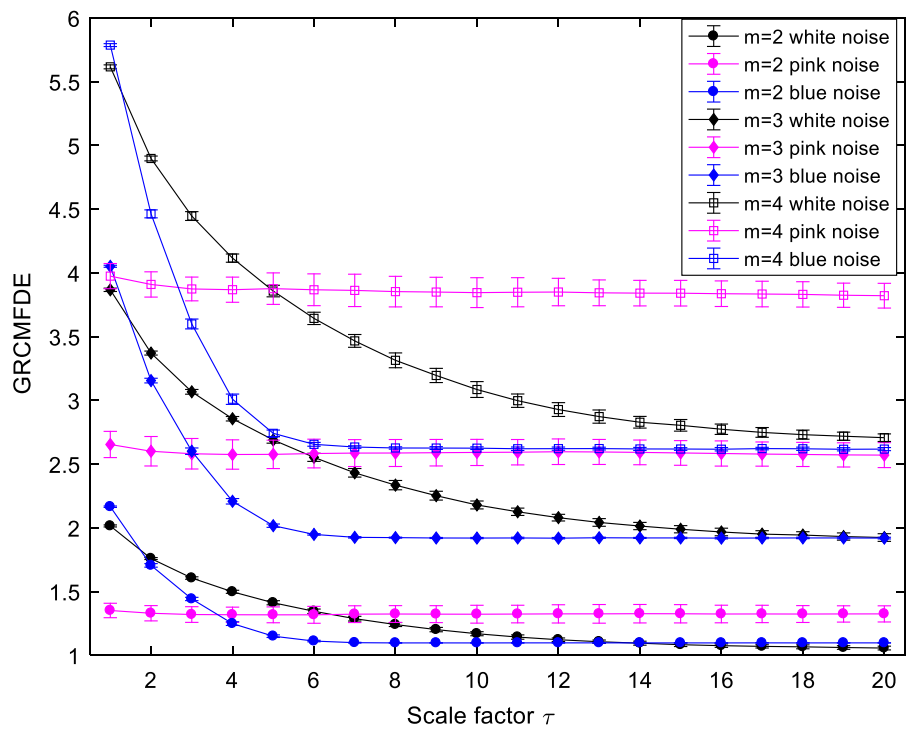
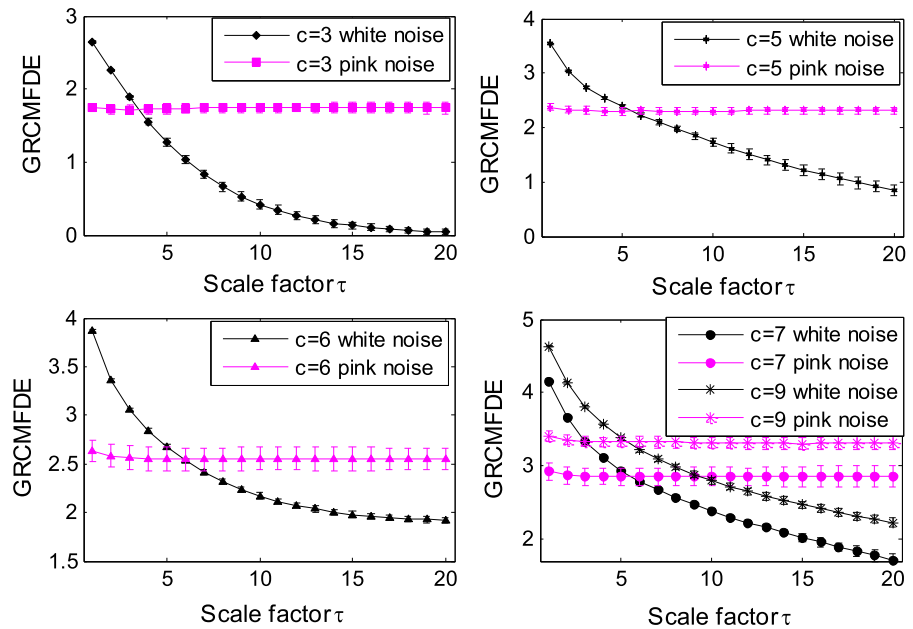


Fig. 5 GRCMFDE of two noise signals for different c



GRCMFDE. The mean and STD curves of GRCMFDE of the three noises are computed and depicted in Fig. 4. From Fig. 4, it can be found that when m is large (for example 4), the FDE values are also large, together with the STD. And thus, generally, we set m as a small integer of 2 or 3.

Next, the GRCMFDE of 20 different white and pink noises with a length of 5000 points are computed when $m = 2$ and $c = 3, 5, 6, 7, 9, d = 1, \alpha = 0$ to study the influence of c on GRCMFDE, and the computed mean and STD curves of GRCMFDE are shown in Fig. 5. From Fig. 5, it can be seen that, like m , a large c will result in large FDE values. If c is too small, the slight change of amplitude will not be detected, and if c is too large, FDE will be sensitive to noise. Therefore, generally, we set c as 5 or 6.

Actually, the research of fractional order entropy on complexity of time series is still in exploratory stage and there are still many issues to be solved before applying it in real applications. Ref. [22] indicates that tuning the fractional order allow an high sensitivity to the signal evolution, which is useful in describing the system dynamics. Also, we compute the FDE (i.e., GRCMFDE with scale factor = 1) for different values of α (changing from -0.1 to 0.8) and the results are shown in Fig. 6, from which with increasing of α , the FDE values of white noise and pink noise are also increasing to a very large value and their difference is

much obvious for a large α , as well as the STD and it is hard to select a suitable α for the following applications. Next, we compute the GRCMFDE in 20 scales when $\alpha = 0.7, 0.5, 0.3, 0.1, 0, -0.1$ and the results are shown in Fig. 7. From Fig. 7, when $\alpha = 0.7$, the FDE of the first scale gets a value nearly 40, while FDE of blue noise gets a negative value when scale factor larger than 6. And when α is larger than 0.3, the values of noises are larger than 10. In fact, for different α , the changing trend of GRCMFDEs is nearly the same for different α for white, pink or blue noises. Therefore, generally we set α as 0.1, 0 or -0.1 to obtain a suitable entropy value.

It should be noted that the newly proposed GRCMFDE algorithm is different from the RCMDE method developed in Ref. [13]. In GRCMFDE, the generalized coarse graining (root-mean-square used in this paper) is different from the coarse graining used in the first moment. In addition, the improved FDE used in GRCMFDE is different from the DE used in RCMDE. Last, we also introduced the fractional order to FDE, which also is different from DE. In the following part, for comparison purpose, we also compute the MDE, MPE, RCMDE, GRCMFDE₀, GRCMFDE_{-0.1} and GRCMFDE_{0.1} of the 20 different white, blue and pink noises when $m = 2, c = 6, d = 1$ and $\alpha = 0.1$ and the results are shown in Fig. 8. Three noises can be differed by MDE, RCMDE,

Fig. 6 FDE for different values of α

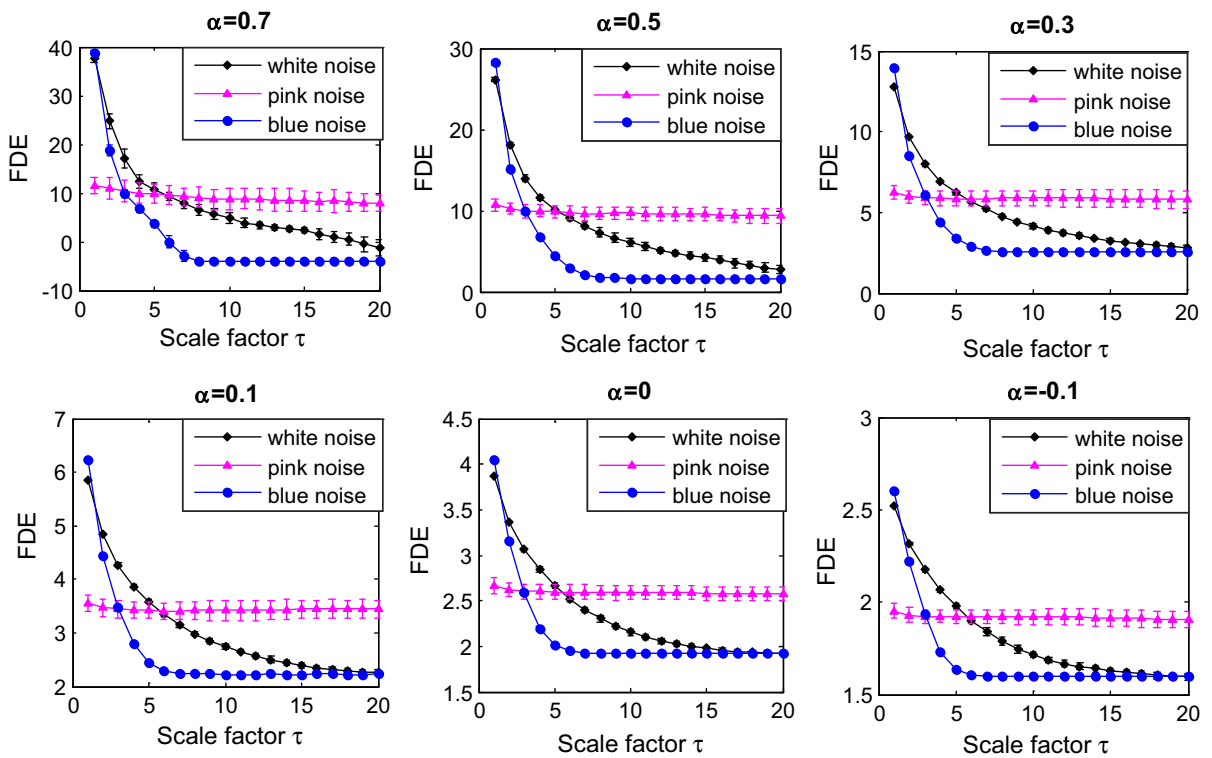
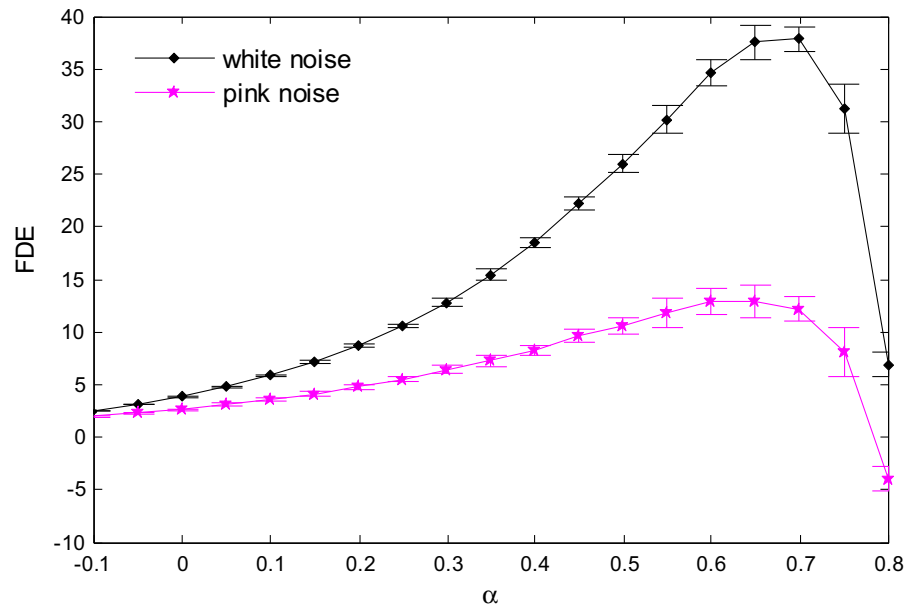


Fig. 7 GRCMFDE when $\alpha = 0.7, 0.5, 0.3, 0.1, 0$ and -0.1

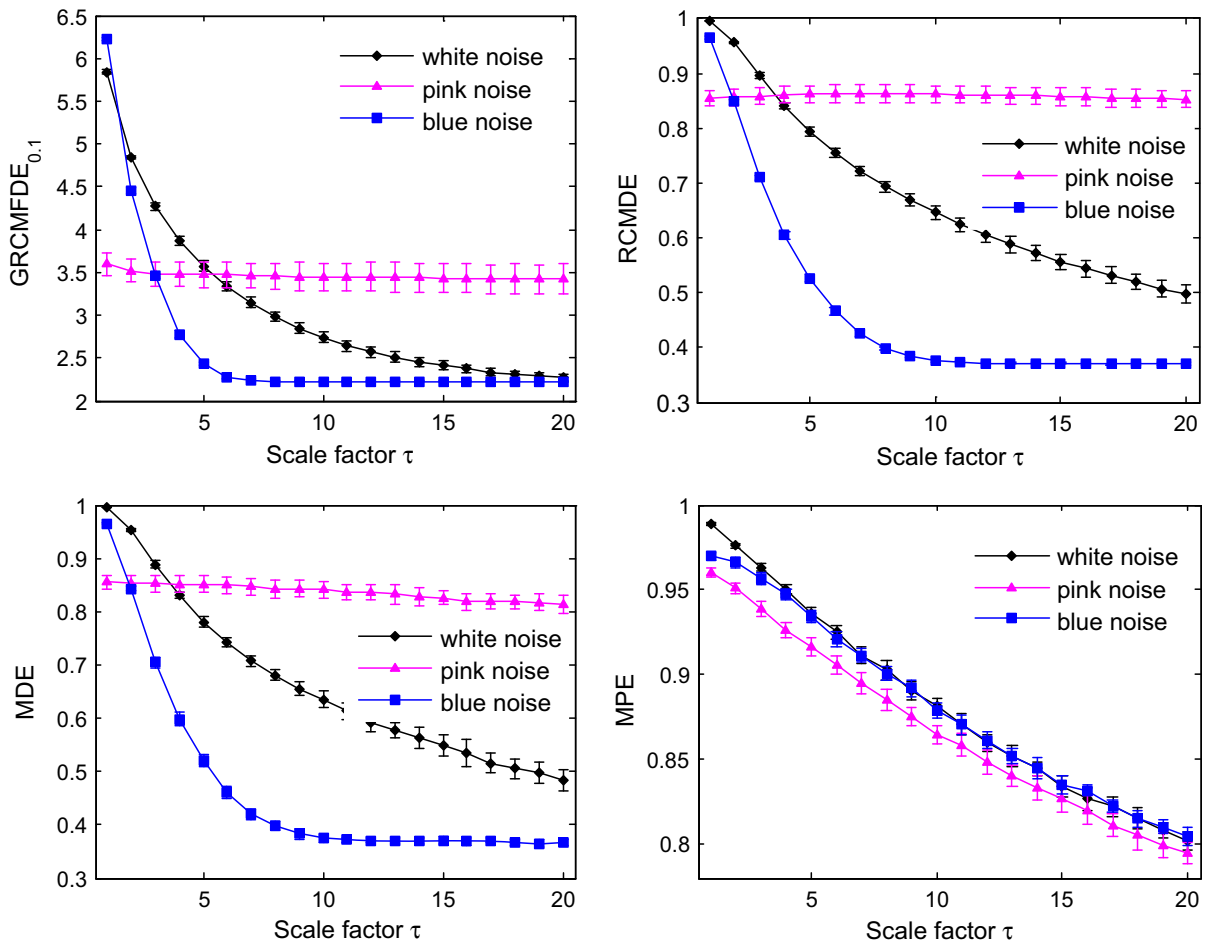


Fig. 8 MDE, MPE, RCMDE and GRCMFDE_{0.1} of 20 groups of white, blue and pink noises

GRCMFDE₀, GRCMFDE_{-0.1} and GRCMFDE_{0.1} obviously while the MPEs of blue and pink noises are very similar and it is difficult to distinguish them.

4 GRCMFDE-based fault diagnosis method

4.1 The proposed fault diagnostic method

Based on GRCMFDE, with combining MCFS and GWOSVM, an intelligent rolling bearing fault diagnostic method is proposed as follows.

- (1) Let rolling bearing contains K states composed of different fault classes and severities. The sample numbers of each state are M_1, M_2, \dots, M_K . All samples of each state are randomly

divided into $M_k/2$ for training and $M_k/2$ for testing ($k = 1, 2, \dots, K$).

- (2) The GRCMFDE of all training and test samples are calculated with scale factor $\tau_m = 20$ and other selected parameters ($m = 3, c = 6$ and $d = 1$). Then, the most important p features of MCFS feature selection are selected to form the K sensitive fault feature sets.
- (3) The sensitivity fault features sets of training samples are employed to train the GWOSVM-classifier.
- (4) Testing sensitive fault feature sets are used for testing. And they are used to verify the trained GWOSVM-classifier for diagnostics. The final output of classifier is used to differentiate the bearing operating conditions.

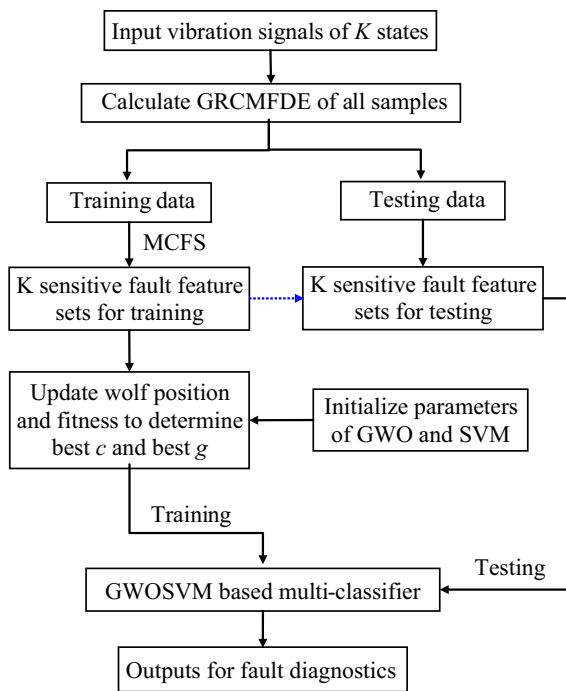


Fig. 9 Flow diagram of the proposed approach

Figure 9 gives the flowchart of the proposed approach.

4.2 Analysis of experiment data

In this subsection, the experimental data provided by Bearing Data Center of Case Western Reserve University (can be download from the website [23]) are used to verify the effectiveness of proposed fault diagnosis method for rolling bearing. In the experimental system of simulation fault (shown in Fig. 10), the electro-discharge machining technology was

employed to seed the single point local faults in each 6205-2RS deep groove ball bearing for test. The sampling frequency of data recorder in data collection system was 12 kHz when the motor speed was set as 1730 r/min with load 3hp. Since the operating frequency of Power system in USA is 60 Hz and the synchronous speed of the 2-pole pair motor is 1800 rpm, the maximum speed of the 2-pole pair asynchronous motor is no higher than 1800 rpm. Therefore, the motor speed used in this test generally is smaller than the critical shaft speed of motor and much smaller than the system resonance frequency. The signals were acquired under ten states including the bearings with inner race faults (IR), outer race faults (OR) and balling element faults (BE) with fault diameter sizes 0.1778, 0.3556 and 0.5334 mm [24, 25], together with the normal rolling bearing. Therefore, the experimental data analysis turns into a ten-class classification issue in consideration of categories and severities.

For each state of the rolling bearing, the data collected from Drive end accelerometer are used. Each state contains 29 samples, and the length of each sample is 4, 096 points. Table 1 shows the label information of rolling bearing experimental data. The waveforms of these signals in the time domain are shown in Fig. 11.

For comparison purpose, the MDE, MPE, RCMDE, GRCMFDE with $\alpha = -0.1, 0$ and 0.1 of all rolling bearing samples are computed. The mean and STD entropy curves of ten classes of all rolling bearing for different methods are depicted in Fig. 12a–f. From Fig. 12d, e the RCMDE and MDE of OR2 and BE2 have large STD and they have nearly the same trends and values. To fulfill an intelligent fault diagnostic of

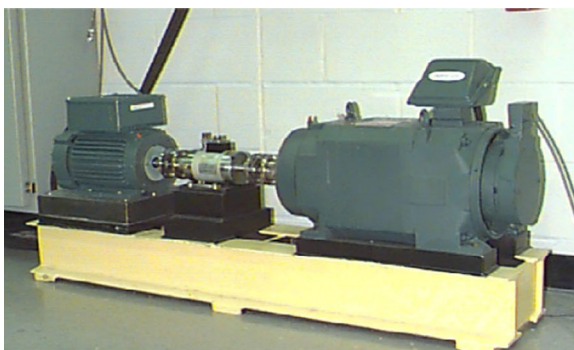


Fig. 10 The experiment system of rolling bearing and its diagram

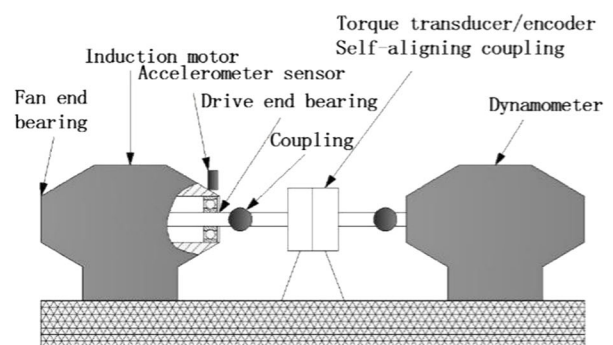


Table 1 Label description of the used data [10]

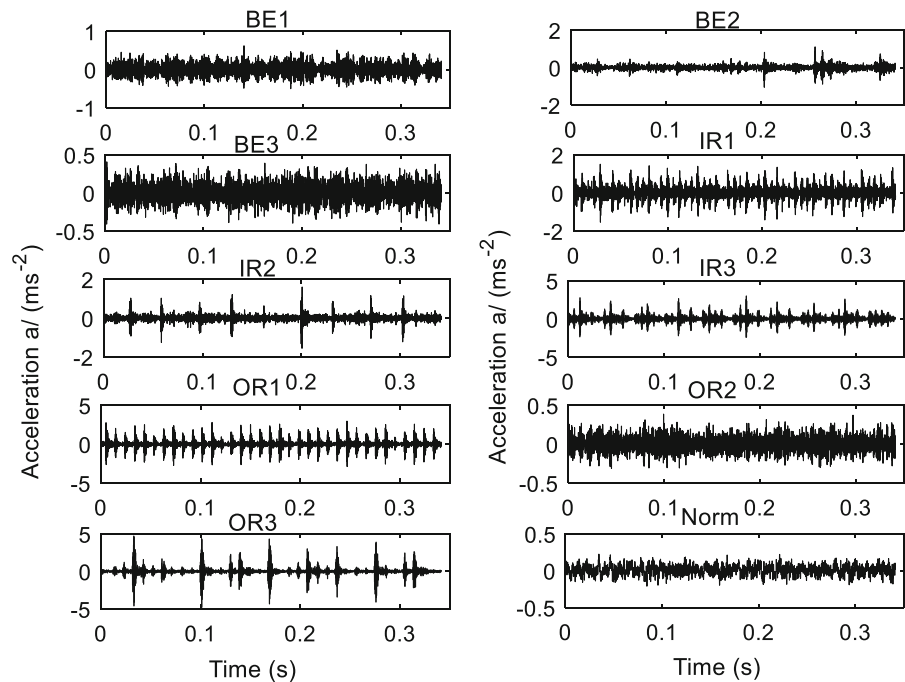
Fault class	Fault size (mm)	Class label	Used data
Norm	0	1	100_normal_3
IR1	0.1778	2	108_IR007_3
IR2	0.3556	3	172_IR014_3
IR3	0.5334	4	212_IR021_3
OR1	0.1778	5	133_OR007@6_3
OR2	0.3556	6	200_OR014@6_3
OR3	0.5334	7	237_OR021@6_3
BE1	0.1778	8	121_B007_3
BE2	0.3556	9	188_B014_3
BE3	0.5334	10	225_B021_3

rolling bearing, a multi-classifier is founded. The $GRCMFDE_0$ ($\alpha = 0$) of all 290 samples with the feature number of 20 are taken as initial fault sets. First, 11 samples of initial fault sets are randomly selected as training samples from 29 ones of each class by using the “randperm” function in Matlab and the remaining 18 ones are taken as testing samples. Thus, totally, the fault feature training sets consisting of 110 samples with dimension (110×20) and the fault feature testing sets consisting of 180 samples with dimension (180×20) can be obtained. Second, the

MCFS method is used to reduce the dimension of fault feature sets 20 to a value of $p = 5$ and the order of the most important several feature values is selected through using MCFS to train the fault feature training data sets. The obtained order by MCFS is utilized to construct new fault feature training and testing data sets. Then, the obtained new fault feature training sets with dimension (110×5) , together with the new fault feature testing sets with dimension (180×5) are employed to train and test the GWOSVM multi-fault classifier. Third, new fault feature sets of training data are used to train the GWOSVM multi-classifier, where the GWO is used to optimize the parameter c and g among the interval $[0.01, 100]$. In GWO algorithm, the number of search agents is set as 10, the maximum number of iterations is set as 100, the cross-validation parameter of SVM is 5. The initial positions of wolves are zeros and the initial objective function value is infinite. The outputs of GWOSVM multi-classifier of testing samples are presented in Fig. 13a.

For comparison purpose, the MPE, MDE, RCMDE, $GRCMFDE_{0,1}$ and $GRCMFDE_{-0,1}$ are also seen as the initial nonlinear fault features. The same number of training and testing data of each class (11 and 18, respectively) is used to form the initial fault feature data sets. Similarly, new fault feature training data sets of dimension 110×5 can be obtained by using

Fig. 11 Time domain waveform of rolling bearing of different classes



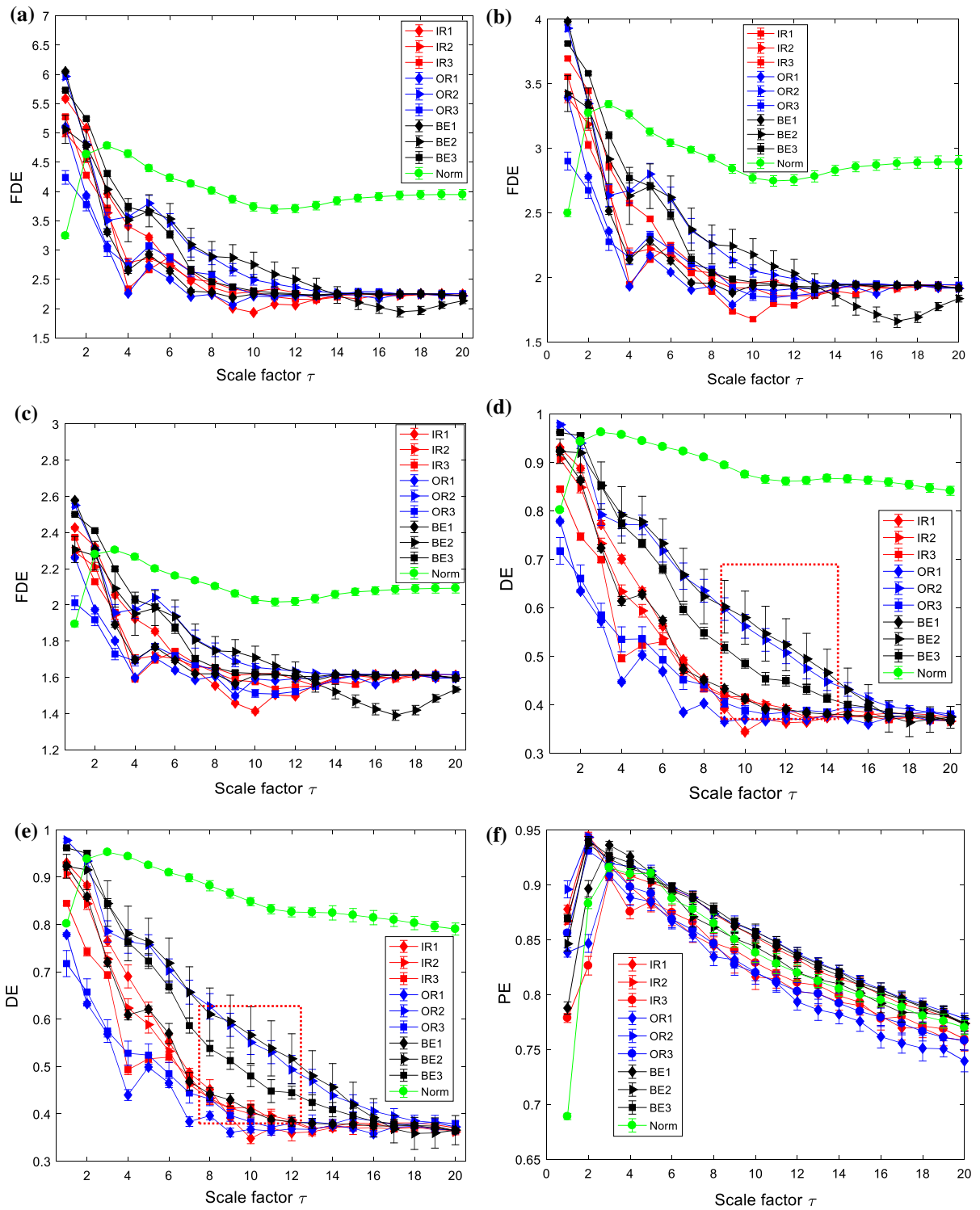


Fig. 12 MDE, MPE, RCMDE and GRCMFDE of rolling bearing data. **a** GRCMFDE_{0.1} **b** GRCMFDE₀ **c** GRCMFDE - _{0.1} **d** RCMDE **e** MDE and **f** MPE

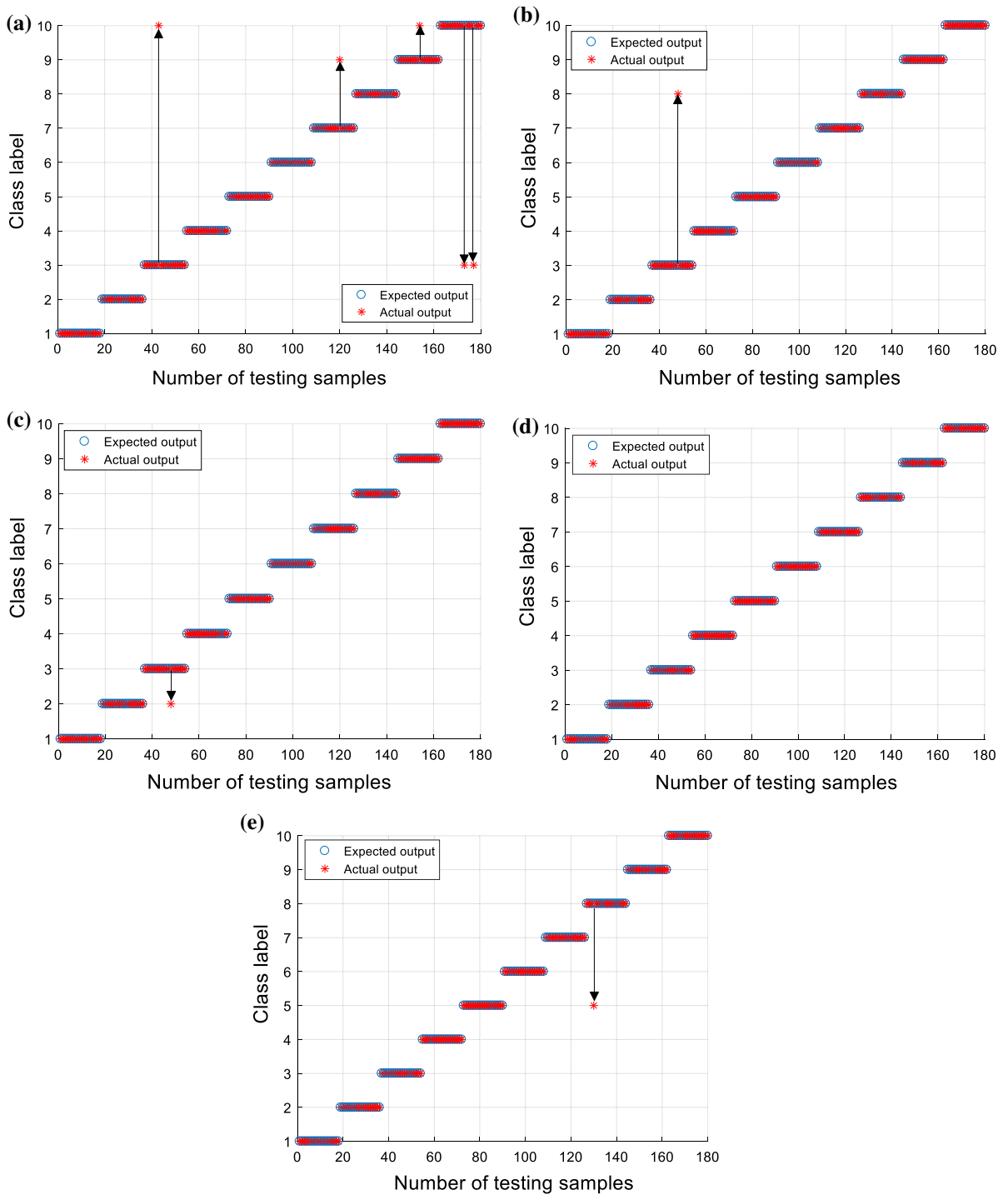


Fig. 13 Identifying rates of different methods. The identifying rate of **a** MPE, **b** MDE **c** RCMDE **d** GRCMFDE₀ and GRCMFDE_{0,1}, and **e** GRCMFDE_{0,1}

MCFS, together with the new fault feature testing data sets with dimension 180×5 . Then, they, respectively, are input to the GWOSVM multi-classifier for training and testing, and the results are shown in Fig. 13a–e. From Fig. 13 first, all testing samples are correctly classified by the proposed GRCMFDE₀ and GRCMFDE_{-0.1}-based fault diagnostics methods and the identifying rate is 100%. The identifying rates of MDE, RMCDE and GRCMFDE_{0.1}-based fault feature extraction method are 99.44%, that is, one sample of class three (IR2) is misclassified to class ten (BE3) when using MDE for fault extraction, and one sample belonging class three (IR2) is misclassified to class two (IR1) by RCMDE, while one sample of class eight (BE1) is misclassified to class five (OR1) when GRCMFDE_{-0.1} is used. The MPE-based fault diagnostic method gets the lowest identifying rate of 97.22% when five samples are misclassified. Therefore, the above results indicate that the proposed GRCMFDE₀ and GRCMFDE_{-0.1}-based fault feature diagnostic methods show much better performance than that of MPE, MDE and GRCMFDE_{0.1}-based methods when the number of new fault feature element is set as 5.

A fundamental problem of the proposed method is to determine the number of selected new fault feature element (p) through MCFS. Without loss of generality, we set the new fault feature elements with a number of $p = 1$ to 20 as the inputting of GWOSVM-based multi-classifier. The corresponding identifying rates of different methods mentioned above for different numbers of new fault feature elements are given in Tables 2 and 6, and the identifying rate comparisons are presented in Fig. 14. By observing Fig. 14 and Tables 2, 3, 4, 5, and 6, we can find that when the number of fault feature elements is set as 6 and 9, the highest identifying rate of MPE, MCFS and GWOSVM-based fault diagnostics method is

97.78%. When the number of fault feature elements is larger than 7, the highest identifying rate of the MDE, MCFS and GWOSVM-based fault diagnostics method is 99.44%. The identifying rate of RCMDE-based method reaches 100% when the number of fault feature elements larger than 5, while that of GRCMFDE₀ and GRCMFDE_{-0.1}-based method reaches 100% when the number of feature elements larger than 4 and 3, i.e., for the GRCMFDE_{-0.1} method, four features are enough for an accurate identifying rate of 100%. The identifying rate of GRCMFDE_{0.1}-based method reaches 100% when the number of fault feature elements is 4 and larger than 6. To obtain the highest identifying rate of 100%, for GRCMFDE₀ and GRCMFDE_{-0.1}-based fault representation methods, we only need 4 or 5 features to reflect the whole fault information, while for other methods, we may need more features.. Therefore, the result indicates that GRCMFDE₀ and GRCMFDE_{-0.1} show much more robustness and better performance than MPE, MDE and RCMDE.

Lastly, also, the necessity of MCFS for feature selection was investigated. With loss of generality, all the 20 features of RCMDE, GRCFMDE₀, GRCFMDE_{0.1}, GRCFMDE_{-0.1} methods are randomly reordered through using Matlab function “randperm.m,” then the first one to ten features are, respectively, selected as fault features and input to the GWOSVM multi-classifier. The identifying rates of different methods for different number of used features are provided in Fig. 15, in which the identifying rates of the proposed method for different number features are given. For a clear comparison, identifying rates of the proposed methods using MCFS mentioned above are also given in Fig. 15. From Table 8 and Fig. 15, it can be observed that for the same number of used features (from 4 to 10), the proposed methods by combining MCFS show better

Table 2 Identifying rate of MPE and MCFS for different number of features

		Number of features	1	2	3	4	5	6	7	8	9	10
MPE + MCFS + GWOSVM	Identifying rate (%)		80.56	92.22	82.22	85.56	97.22	97.78	97.22	96.11	97.78	96.67
	Number of features		11	12	13	14	15	16	17	18	19	20
	Identifying rate (%)		95.56	96.11	96.11	97.22	95.56	96.67	95.56	93.89	94.44	92.78

Fig. 14 The identifying rate comparisons for different number of fault features

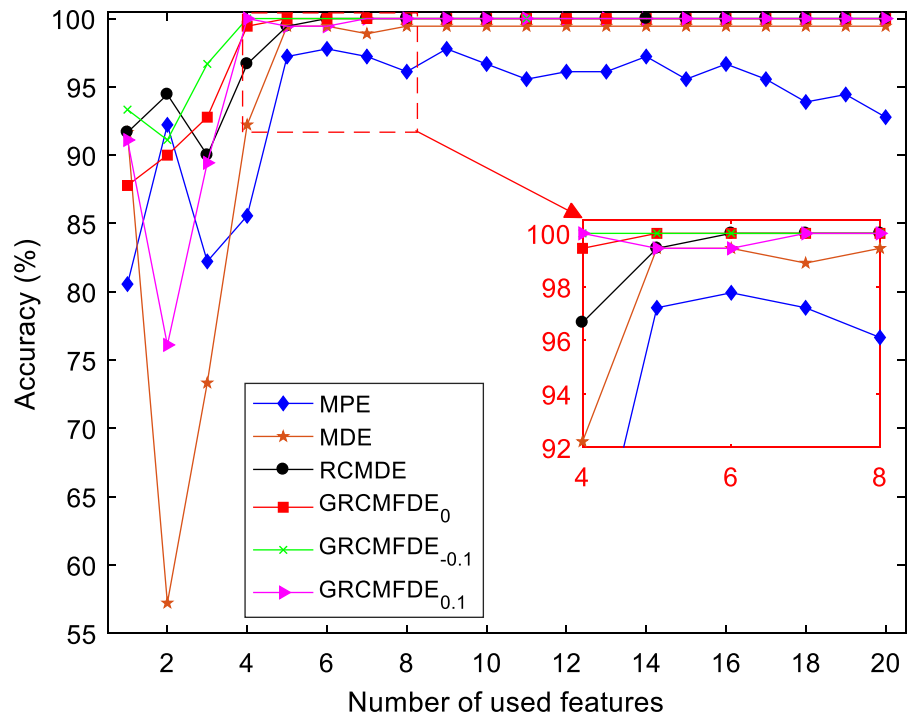


Table 3 Identifying rate of MDE and MCFS for different number of features

		Number of features	1	2	3	4	5	6	7	8	9	10
MDE + MCFS + GWOSVM	Identifying rate (%)		91.67	57.22	73.33	92.22	99.44	99.44	98.89	99.44	99.44	99.44
	Number of features		11	12	13	14	15	16	17	18	19	20
	Identifying rate (%)		99.44	99.44	99.44	99.44	99.44	99.44	99.44	99.44	99.44	99.44

Table 4 Identifying rate of RCMDE and MCFS for different number of features

		Number of features	1	2	3	4	5	6	7	8	9	10
RCMDE + MCFS + GWOSVM	Identifying rate (%)		91.67	94.44	90	96.67	99.44	100	100	100	100	100
	Number of features		11	12	13	14	15	16	17	18	19	20
	Identifying rate (%)		100	100	100	100	100	100	100	100	100	100

performance and higher identifying rates than the methods using random features. This states that it is necessary and superior to use MCFS for feature selection.

4.2.1 Case 2

The experiment data used in this part provided by Anhui University of Technology are employed to illustrate the versatility of the proposed method. As

Table 5 Identifying rate of GRCFMDE₀ and MCFS for different number of features

	Number of features	1	2	3	4	5	6	7	8	9	10
GRCFMDE ₀ + MCFS + GWOSVM	Identifying rate (%)	87.78	90	92.78	99.44	100	100	100	100	100	100
	Number of features	11	12	13	14	15	16	17	18	19	20
	Identifying rate (%)	100	100	100	100	100	100	100	100	100	100

Table 6 Identifying rate of GRCFMDE_{0.1} and MCFS for different number of features

	Number of features	1	2	3	4	5	6	7	8	9	10
GRCFMDE _{0.1} + MCFS + GWOSVM	Identifying rate (%)	91.11	76.11	89.44	100	99.44	99.44	100	100	100	100
	Number of features	11	12	13	14	15	16	17	18	19	20
	Identifying rate (%)	100	100	100	100	100	100	100	100	100	100

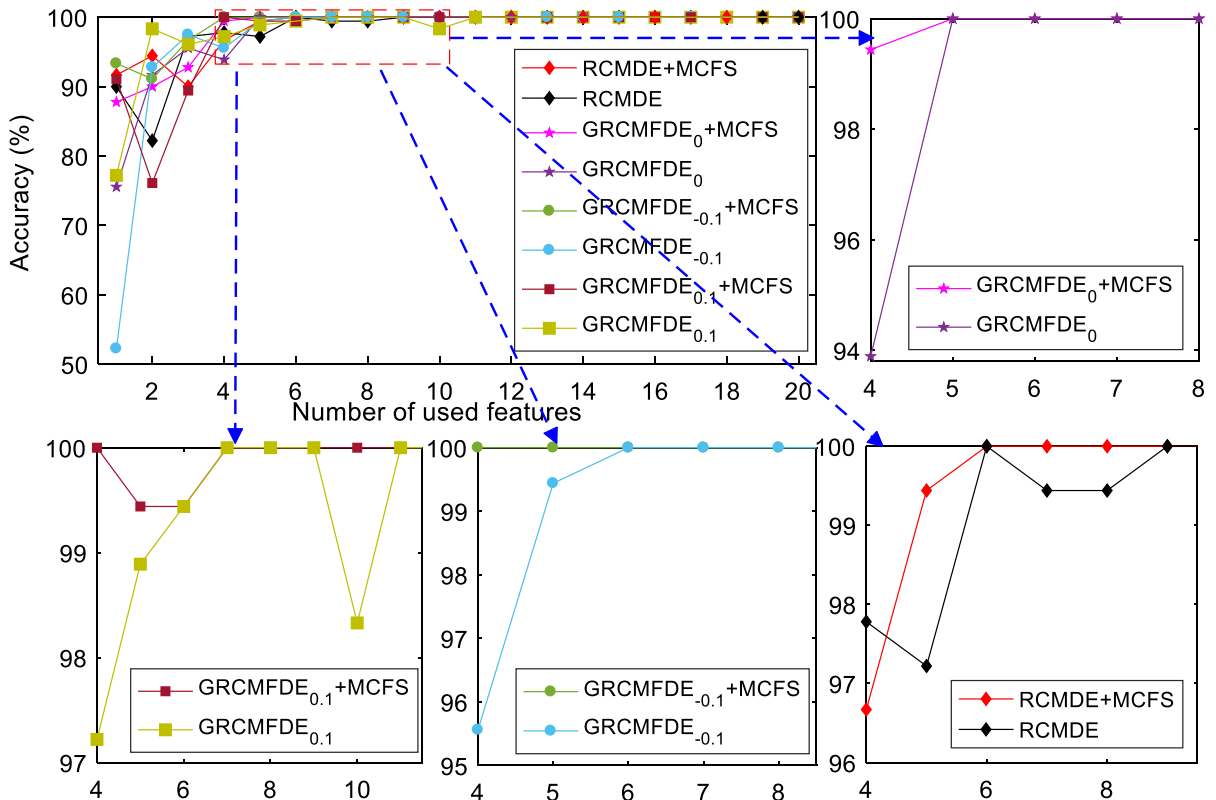


Fig. 15 The accuracy comparison of the methods using MCFS for feature selection and random features

shown in Fig. 16a, b, the test rolling bearings are 6206-2RS deep groove ball bearing and they are seeded local failures with different locations (shown in

Fig. 16c–e). In the test, the load is set as 0 and 5 kN, respectively, with the sampling frequency of 10, 240 Hz. The motor used is AC and asynchronous

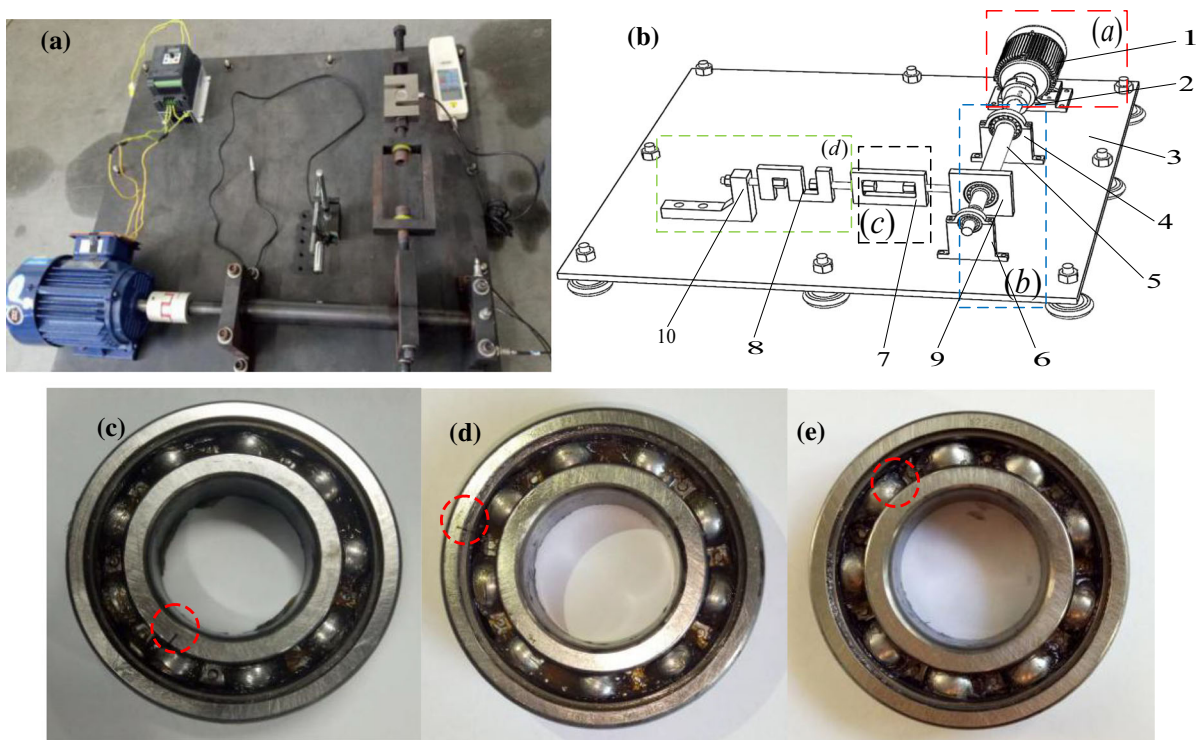


Fig. 16 **a** The bearing test rig, **b** its schematics of case 2, where 1—motor, 2—coupling, 3—base plate, 4—support bearing, 5—spindle, 6—test bearing, 7—buffer device, 8—dynamometer, 9 and 10—loading adjust device, **c** inner fault, **d** outer fault and **e** rolling element fault

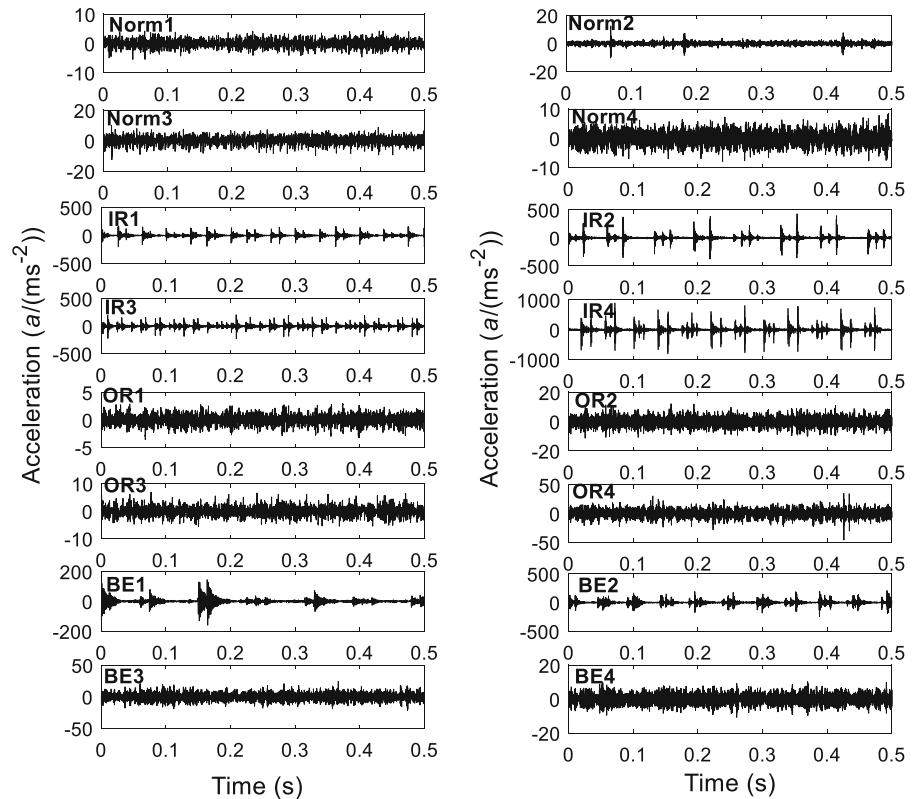
Table 7 Identifying rate of GRCFMDE_{-0.1} and MCFS for different number of features

	Number of features	1	2	3	4	5	6	7	8	9	10
GRCFMDE _{-0.1} + MCFS + GWOSVM	Identifying rate (%)	93.33	91.11	96.68	100	100	100	100	100	100	100
	Number of features	11	12	13	14	15	16	17	18	19	20
	Identifying rate (%)	100	100	100	100	100	100	100	100	100	100

Table 8 Class label and description of AHUT rolling bearing data

Fault class	Fault size (mm)	Load (kN)	Speed (r/min)	Class label	Fault class	Fault size (mm)	Load (kN)	Speed (r/min)	Class label
Norm1	0	0	900	1	OR1	3	0	900	5
Norm3	0	0	1500		OR3	3	0	1500	
Norm2	0	5	900	2	OR2	3	5	900	6
Norm4	0	5	1500		OR4	3	5	1500	
IR1	3	0	900	3	BE1	2	0	900	7
IR3	3	0	1500		BE2	2	0	1500	
IR2	3	5	900	4	BE4	2	5	900	8
IR4	3	5	1500		BE3	2	5	1500	

Fig. 17 The time domain waveforms of used bearing data



motor with adjustable speed produced by China. The rated power is 1500 W, the rated speed is 2800 r/min and supply voltage 220 V. In the test, the motor rotates with two speeds of 900 r/min and 1500 r/min, respectively (both of them smaller than the critical shaft speed of motor). The signals of normal rolling bearing, together with the faulty bearings with a fault depth of 2 and 3 mm are collected [26] using accelerometer from three directions and the data from axial direction are finally used in this paper. For each fault class, four kinds of working conditions are studied (Table 7). The experiment data used in this paper are described in Table 8. The waveforms in time domain of the used rolling bearing data are shown in Fig. 17.

For the 16 states of rolling bearings of AHUT experimental data described in Table 8, 40 samples with a length of 5120 points of each state are used, and thus totally 640 samples are achieved in this test. For comparison purpose, the MPE, MDE, RCMDE, GRCMFDE₀, GRCMFDE_{0.1}, GRCMFDE_{-0.1} and GRCMFDE₀ methods of all samples are computed and the corresponding mean and STD curves are

depicted in Fig. 18, since the results of GRCMFDE_{0.1}, GRCMFDE_{-0.1} and GRCMFDE₀ are similar, only the result of GRCMFDE₀ is given for saving space. From Fig. 18, the RCMDE, MDE and MPE curves have larger STDs than the GRCMFDE₀ method at some scales. Second, the working conditions with speeds 900 and 1500 r/min nearly have the same multiscale entropy curves and trends (MPE, MDE, RCMDE, GRCMFDE₀), which indicates that the entropy-based fault feature extraction methods are sensitive to load but is insensitive to rotating speed. Therefore, in the following, the conditions of rolling bearing with the same load but different speed will be taken as the same class to fulfill an intelligent fault diagnosis (class label information is given in Table 8).

Next, 80 samples of each class (40 samples for the speed of 900 and 1500 r/min, respectively) are randomly arranged by the Matlab function “randperm.” Among them, 50 are used for training and the remaining 30 are used for testing. Then, the MPE, MDE, RCMDE, GRCMFDE_{0.1}, GRCMFDE₀ and GRCMFDE_{-0.1} are taken as initial nonlinear fault features. Next, MCFS is utilized to learn the training

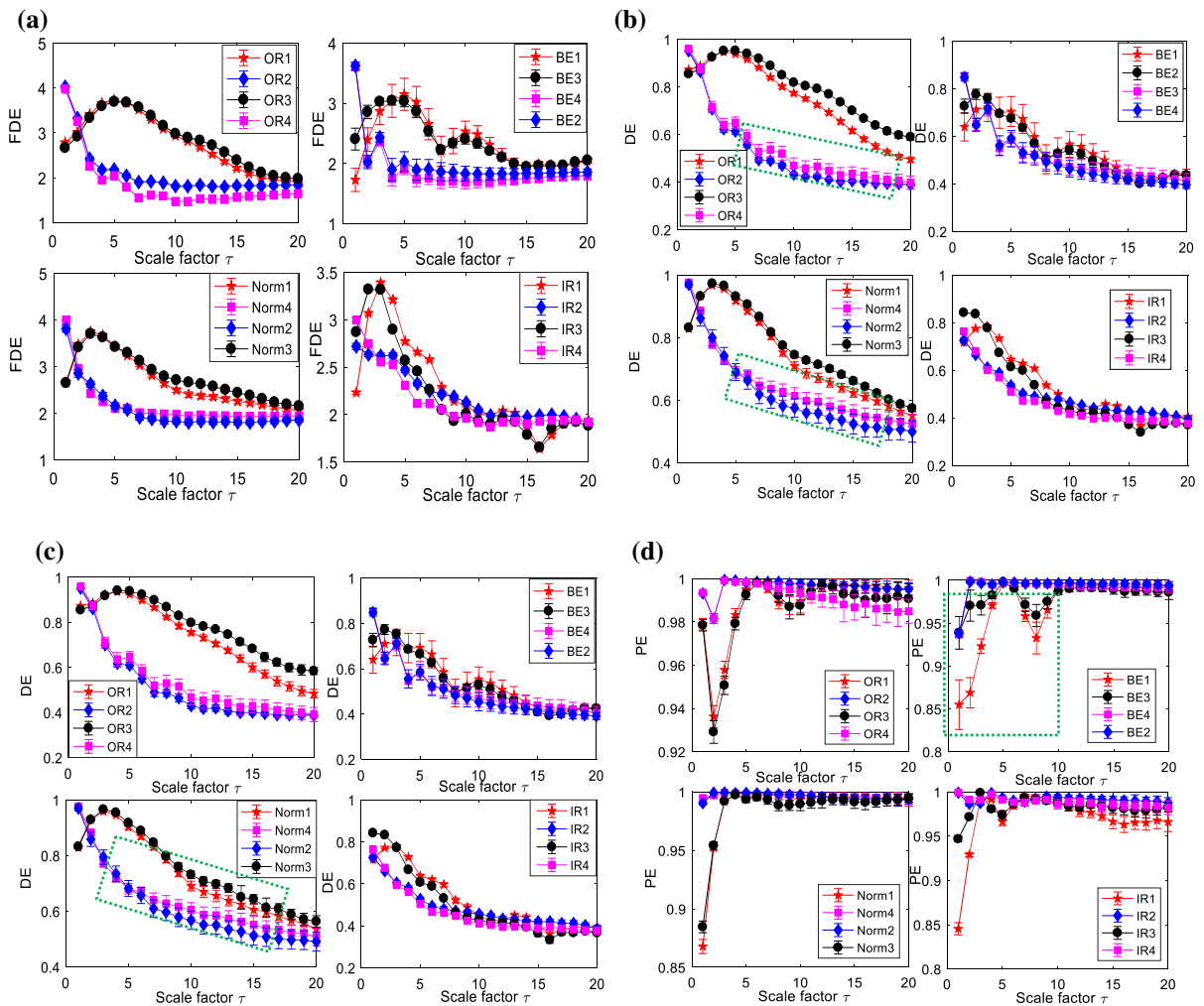


Fig. 18 The multiscale complexity features of rolling bearings for different classes **a** GRCMFDE **b** RCMDE **c** MDE and **d** MPE

data constructed by different methods. New fault feature training data sets with dimension $400 \times p$ ($p = 1, 2, \dots, 10$, is the number of selected features) can be obtained by using MCFS, together with the new fault feature testing data sets with dimension $240 \times p$. Like the process mentioned above, the new fault features of 400 training samples, together with that of 240 testing samples, are, respectively, input to the GWOSVM-based multi-classifier. The output results of classifier of all testing samples for different methods when different number of features used are shown in Fig. 20, and only the outputs of different methods with 5 features used are illustrated in Fig. 19a–d. From Figs. 19 and 20, all testing samples are correctly classified by the

proposed GRCMFDE_{0,1}, GRCMFDE₀ and GRCMFDE_{0,1}-based fault diagnostic methods and the corresponding identifying rates of these three methods are 100% when 5 features are used. However, for the RCMDE method, one sample belonging to class 4 is misclassified to class 8, while for MDE and MPE, 2 and 6 samples, respectively, are misclassified to the wrong classes. This indicated the superiority of the proposed GRCMFDE approach in nonlinear fault representation. In addition, by observing Fig. 20 it can be found that the identifying rate of GRCMFDE₀-based fault diagnostic method reaches 100% when the number of used features larger than 3, while the identifying rate of the GRCMFDE_{0,1}-based method reaches 100% when the number of used features larger

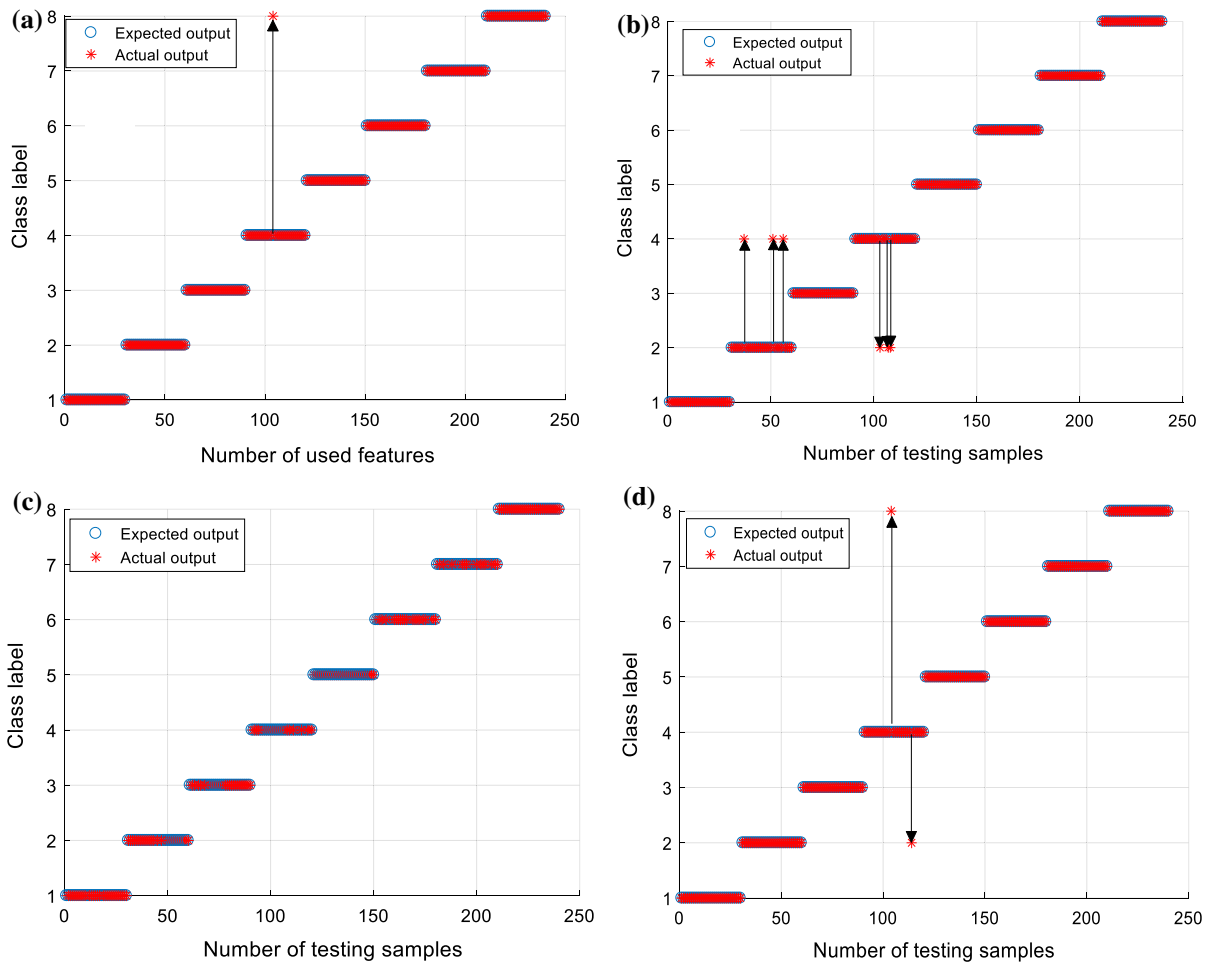


Fig. 19 a RCMDE b MPE c GRCMFDE_{-0.1}, GRCMFDE₀, GRCMFDE_{0.1} and d MDE

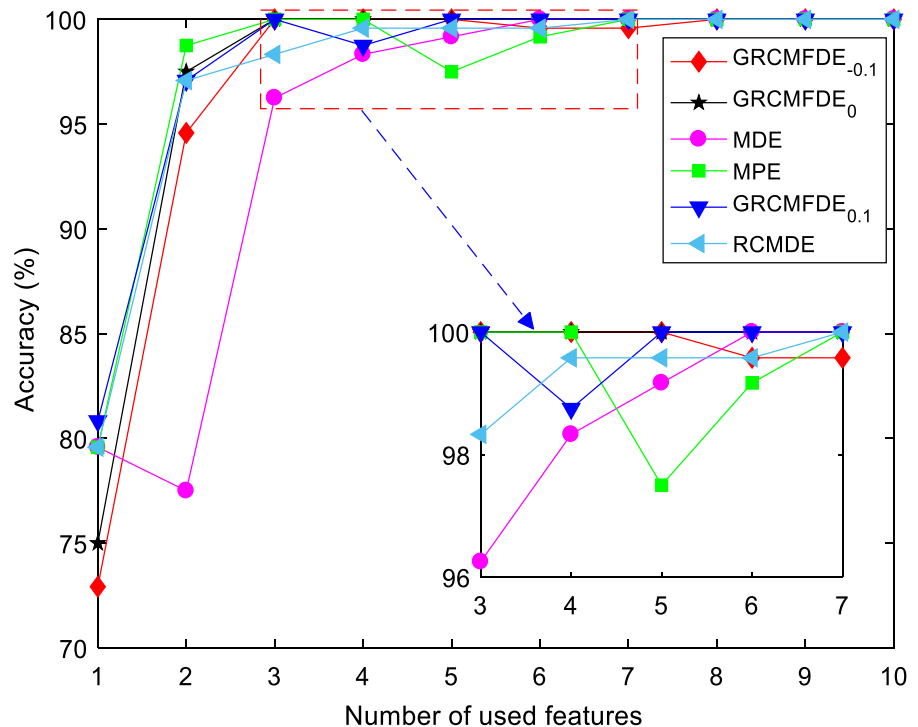
than 5. The above result indicates that the proposed GRCMFDE (especially, GRCMFDE₀)-based fault feature extraction and diagnostic methods show much better performance than that of MPE, MDE and GRCMFDE_{0.1}-based methods.

5 Conclusions

The symptom of periodic or regular transients will exist in corresponding vibration signals when the rolling bearing work with local faults. This intrinsic structural change of vibration signals makes it possible for us to detect and diagnosis the existence of local faults through measuring the complexity and irregularity of vibration time series. For this purpose, a root-mean-square-based generalized composite coarse

graining way is proposed and then the generalized refined composite multiscale fluctuation-based fractional dispersion entropy (GRCMFDE) is proposed to overcome the shortages of multiscale distribution entropy in complexity measure of time series. Also, by analyzing different complex noise signals, GRCMFDE is compared with MDE, RCMDE and MPE, and the results show that GRCMFDE is much more stable than MPE, MDE and RCMDE. Also, based on GRCMFDE for fault feature extraction, MCFS for feature selection and GWOSVM for feature recognition, an intelligent fault diagnostic method was proposed and then employed to analyze two cases of rolling bearing experiment data. The proposed fault diagnosis method was also compared with RCMDE-, MDE- and MPE-based fault diagnosis approaches, together with different fractional orders (0, 0.1,

Fig. 20 Outputs of testing samples of the six methods for different number of used features



– 0.1). The research indicates that the GRCMFDE₀ method generally gets a robust and reliable feature representation and fault diagnosis effect than other fractional orders, as well as other approaches. Besides, the number of features for inputting to the multi-fault classifier is investigated. The comparison results show that the proposed GRCMFDE by synthesizing different fractional order dispersion entropy can effectively distinguish the fault locations and degrees of bearing and the fault identifying rate reaches 100% when the selected number of GRCMFDE features equaling or larger than five. Last, we all investigate the influence of rotating speed and load on the proposed method and the analysis result indicates that the multiscale entropy curve trends of rolling bearing are generally sensitive to load change but are insensitive to speed. There are still several issues that need to be solved, such as the parameter selection and the determination of the number of used features. In the future, we will concentrate to address these problems and perform its related theories to more widely application areas.

We have used two kinds of data sets to verify the effectiveness of the proposed method. For another or a real application, it is really difficult to diagnose the fault states and damage degree without the given

known knowledge. We want to say that the two diagnosis cases are used to verify that whether the proposed method is effective in representing the fault feature information and if it is sensitive to the fault degree or fault locations. In the future work, we will use the proposed method to diagnose the fault of gear and gearbox and also investigate the relationships among the fault degree and the work operation with the feature representation method.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 51975004), the National Key Technologies Research & Development Program of China (No. 2017YFC0805100), the Natural Science Foundation of Anhui Province, China (No. 2008085QE215) and the Key Program of Natural Science Research of Higher Education in Anhui Province of China (No. KJ2019A053, KJ2019A092).

Compliance with ethical standards

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

Appendix A: Reviews of DE and MDE

For a given univariate time series $X = \{x_1, x_2, \dots, x_N\}$ with a length of N , the computation steps of DE are given as follows [12, 13].

- (1) The original time series X is mapped to $Y = \{y_1, y_2, \dots, y_N\}$ by using normal cumulative distribution function (NCDF) shown as

$$y_j = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_j} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt. \tag{4}$$

It is obvious that $y_i \in (0, 1)$. σ is standard deviation and μ is mean of X .

- (2) By using the linear transform shown in Eq. (2), all elements of $Y(y_i, j = 1, 2, \dots, N)$ are mapped to c classes (which is an integer), i.e.,

$$z_j^c = R(c \cdot y_j + 0.5) \tag{5}$$

where R represents the rounding operation and z_j^c represents the j th member of the classified time series. Although the step (2) is linear, the whole mapping way is nonlinear for the use of NCDF in step (1).

- (3) For the given time delay d and embedding dimension m , $z_i^{m,c}$ can be obtained according to $z_i^{m,c} = \{z_i^c, z_{i+d}^c, \dots, z_{i+(m-1)d}^c\}$, $i = 1, 2, \dots, N - (m - 1)d$. Each time series $z_i^{m,c}$ can be mapped to a dispersion pattern $\pi_{v_0 v_1 \dots v_{m-1}}$, where $z_i^c = v_0, z_{i+d}^c = v_1, \dots, z_{i+(m-1)d}^c = v_{m-1}$. The signal has m members, and they all belong to the integer interval from 1 to c . The number of possible dispersion patterns is equal to c^m .
- (4) The relative frequency for each dispersion patterns $\pi_{v_0 v_1 \dots v_{m-1}}$ can be estimated by

$$p(\pi_{v_0 v_1 \dots v_{m-1}}) = \frac{\text{Number}\{i | i \leq N - (m - 1)d, z_i^{m,c} \text{ has type } \pi_{v_0 v_1 \dots v_{m-1}}\}}{N - (m - 1)d} \tag{6}$$

where $p(\pi_{v_0 v_1 \dots v_{m-1}})$ stands for the number of dispersion patterns of $\pi_{v_0 v_1 \dots v_{m-1}}$ assigned to $z_i^{m,c}$

divided by the total number of embedded signals for embedding dimension m .

- (5) The DE of X is computed by

$$\text{DisEn}(X, m, c, d) = - \sum_{\pi=1}^{c^m} p(\pi_{v_0 v_1 \dots v_{m-1}}) \ln(p(\pi_{v_0 v_1 \dots v_{m-1}})) \tag{7}$$

It can be found from the algorithm of DE that when all probability of distribution patterns $p(\pi_{v_0 v_1 \dots v_{m-1}})$ are equal, DE gets the largest entropy value $\ln(c^m)$ and a typical example is Gaussian white noise. In contrast, when the probability of distribution pattern $p(\pi_{v_0 v_1 \dots v_{m-1}})$ is unitary, i.e., only one value is not equal to zero, DE get the smallest value, which indicates that the time series is a completely predictable data and a typical example is the periodic signal with low frequency.

Based on DE, the computation steps of MDE are given as follows.

- (1) For a given univariate data $W = \{w_1, w_2, \dots, w_L\}$ with a length of L , firstly, it is divided into τ non-overlapping segments with length $\lfloor L/\tau \rfloor$. Then, the average of each segment is computed to derive the coarse-grained time series. This process is named coarse graining and is given as follows:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{a=(j-1)\tau+1}^{j\tau} w_a, \quad 1 \leq j \leq \lfloor L/\tau \rfloor \tag{8}$$

τ is called scale factor. $y^{(1)}$ is the original data W and when $\tau > 1$, W is divided into τ coarse-grained time series $y^{(\tau)}$ with a length of $\lfloor N/\tau \rfloor$ (which stands for the largest integral smaller than $\{x(i), i = 1, 2, \dots, N\}$).

- (2) The DE of $y^{(\tau)}$ is estimated by

$$\text{MDE}(W, \tau, m, c, d) = \text{DE}(y^{(\tau)}, m, c, d) \tag{9}$$

The DEs over different time scales are depicted as a function of scale factor, and this process is called MDE analysis.

References

- Xu, X., Qiao, Z., Lei, Y.: Repetitive transient extraction for machinery fault diagnosis using multiscale fractional order entropy infogram. *Mech. Syst. Signal Process.* **103**, 312–326 (2018)
- Zheng, J., Pan, H., Cheng, J.: Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mech. Syst. Signal Process.* **85**, 746–759 (2017)
- Yin, Y., Shang, P.: Multivariate weighted multiscale permutation entropy for complex time series. *Nonlinear Dyn.* **88**(3), 1707–1722 (2017)
- Yin, Y., Shang, P.: Multivariate multiscale sample entropy of traffic time series. *Nonlinear Dyn.* **86**(1), 479–488 (2016)
- Li, Y., Xu, M., Wang, R., et al.: A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy. *J. Sound Vib.* **360**, 277–299 (2016)
- Zhang, L., Xiong, G., Liu, H., et al.: Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. *Expert Syst. Appl.* **37**(8), 6077–6085 (2010)
- Zheng, J., Cheng, J., Yu, Y., et al.: A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination. *Mech. Mach. Theory* **78**(16), 187–200 (2014)
- Zheng, J., Tu, D., Pan, H., et al.: A refined composite multivariate multiscale fuzzy entropy and Laplacian score-based fault diagnosis method for rolling bearings. *Entropy* **19**(11), 585 (2017)
- Li, Y., Xu, M., Wei, Y., et al.: A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. *Measurement* **77**, 80–94 (2016)
- Zheng, J., Pan, H., Yang, S., et al.: Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **99**, 229–243 (2018)
- Heurtier, H.: The multiscale entropy algorithm and its variants: a review. *Entropy* **17**, 3110–3123 (2015)
- Rostaghi, M., Azami, H.: Dispersion entropy: a measure for time-series analysis. *IEEE Signal Process. Lett.* **23**(5), 610–614 (2016)
- Azami, H., Rostaghi, M., Abasolo, D., et al.: Refined composite multiscale dispersion entropy and its application to biomedical signals. *IEEE Trans. Biomed. Eng.* **64**(12), 2872–2879 (2017)
- Azami, Hamed, Arnold, Steven E., Sanei, Saeid, et al.: Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases. *IEEE Access* **7**, 68718–68733 (2019)
- He, S., Sun, K., Wang, R.: Fractional fuzzy entropy algorithm and the complexity analysis for nonlinear time series. *Eur. Phys. J. Special Top.* **227**(7–9), 943–957 (2018)
- Karci, A.: Fractional order entropy: new perspectives. *Opt. Int. J. Light Electron Opt.* **127**(20), 9172–9177 (2016)
- Wahyu, C., Tegoeh, T.: A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slewing bearing. *Machines* **5**(4), 21 (2017)
- Deng, C., Zhang, C., He, X.: Unsupervised feature selection for multi-cluster data. In: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2010)
- Liu, Z., Cao, H., Chen, X., et al.: Multi-fault classification based on wavelet SVM with GWO algorithm to analyze vibration signals from rolling element bearings. *Neurocomputing* **99**(1), 399–410 (2013)
- Rezaei, H., Bozorghaddad, O., Chu, X.: Grey wolf optimization (GWO) algorithm. In: *Advanced Optimization by Nature-Inspired Algorithms* (2018)
- Costa, M.D., Goldberger, A.L.: Generalized multiscale entropy analysis: application to quantifying the complex volatility of human heartbeat time series. *Entropy* **17**(3), 1197–1203 (2015)
- Machado, J.T.: Fractional order generalized information. *Entropy* **16**(4), 2350–2361 (2014)
- <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>. Bearing Data Center, Case Western Reserve University
- Smith, W.A., Randall, R.B.: Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study. *Mech. Syst. Signal Process.* **64–65**, 100–131 (2015)
- Zheng, J., Cheng, J., Yu, Y.: Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.* **40**(1), 136–153 (2013)
- Zheng, J., Dong, Z., Pan, H., et al.: Composite multi-scale weighted permutation entropy and extreme learning machine based intelligent fault diagnosis for rolling bearing. *Measurement* **143**, 69–80 (2019)

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