



A Multi-scale Fuzzy Measure Entropy and Infinite Feature Selection Based Approach for Rolling Bearing Fault Diagnosis

Keheng Zhu¹ · Liang Chen² · Xiong Hu¹

Received: 18 February 2019 / Accepted: 16 August 2019 / Published online: 12 September 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Fuzzy measure entropy (FuzzyMEn) is a recently improved non-linear dynamic parameter for evaluating the signals' complexity. In comparison with fuzzy entropy (FuzzyEn), which only emphasizes the local characteristics of the signal but neglects its global trend, FuzzyMEn can reflect not only the local but also the global characteristics of the signal. Therefore, by calculating the FuzzyMEn values in different scales, the multi-scale fuzzy measure entropy (MFME) method is put forward in this paper and used for extracting the fault features from vibration signals of rolling bearing. After the feature extraction, the newly developed infinite feature selection (Inf-FS) method is employed to choose the most representative features from the original ones of high dimension. Finally, a new rolling bearing fault diagnosis approach is presented based on MFME, Inf-FS and support vector machine (SVM). The experimental analysis indicates that the presented approach can realize the rolling bearing fault diagnosis effectively.

Keywords Multi-scale fuzzy measure entropy · Infinite feature selection · Bearing fault diagnosis

1 Introduction

To guarantee safe operation in production and reduce economic losses resulted from equipment failure, fault diagnosis for the rotary machines are becoming more and more important. Rolling bearing has been extensively used in the rotatory machineries, and bearing failure is one of most frequent fault sources. This vital component needs working highly reliably to reduce failure occurrences and unexpected machine breakdowns [1, 2]. Hence, it is significant to effectively inspect the occurrence of bearing fault. The vibration signal analysis method has been widely utilized to inspect and diagnose bearing faults because the vibration signals contain rich information for comprehending phenomenon associated with the working conditions of rolling bearing [3–5].

However, the measured vibrational signals are considered as nonlinear and non-stationary [6–8], which makes the fault

feature extraction from these complicated signals a challenge. In view of the characteristics of nonlinearity and non-stationarity, nonlinear parameter estimation methods have been applied to extract state-related information buried in the bearing vibrations. Therein, appropriate entropy (ApEn) [9] was introduced by Yan and Gao [10] for the health monitoring of rolling bearing because of containing rich time-related information and efficient computation. Unfortunately, the calculation of ApEn is effected heavily by length of the data and the corresponding estimated value is uniformly lower than the expected one [11, 15]. To address the limitations of ApEn, Richman and Moorman [11] put forward sample entropy (SampEn), which can obtain a better performance than ApEn. Nevertheless, Heaviside function was utilized to define the similarity degree of vectors in both SampEn and ApEn, which has a rigid and discontinuous boundary [12]. Consequently, Chen et al. [12, 13] developed fuzzy entropy (FuzzyEn) by substituting the exponential function for the Heaviside function. Compared with SampEn, FuzzyEn has a better statistical stability and is more suitable to measure the signal complexity because of continuous boundaries of fuzzy functions. Hence, Zheng et al. [14] and Zhu et al. [15] used it to assess the bearing vibration signals' complexity. However, FuzzyEn only emphasizes the local characteristics of the signal and neglects its global fluctuation

✉ Keheng Zhu
khzhdl@gmail.com

¹ School of Logistics Engineering, Shanghai Maritime University, Shanghai 201306, China

² College of Energy and Power Engineering, Dalian University of Technology, Dalian 116023, China

[16]. Therefore, Liu et al. [16] proposed the fuzzy measure entropy (FuzzyMEn) algorithm, which can reflect not only the local but also the global characteristics of the signal. Compared with FuzzyEn, FuzzyMEn achieved better discrimination ability and was successfully utilized to analyze the heart rate variability [16]. Considering the advantage of FuzzyMEn over FuzzyEn, we introduce FuzzyMEn to the field of bearing fault diagnosis.

Nevertheless, SampEn, FuzzyMn and FuzzyMEn are all designed to measure the signal complexity in one single scale. In order to analyze the signal complexity on multiple time scales rather than on only one scale, the multi-scale analysis method was proposed by Costa et al. [17, 18]. The multi-scale entropy approach was then used for the fault diagnosis of rolling bearing [19, 20]. Similarly, the multi-scale fuzzy entropy (MFE) was developed by Zheng et al. [21]. Based on MFE, Li et al. [22] proposed an improved MFE method for extracting bearing fault features. In this paper, combining the merits of both FuzzyMEn and multi-scale analysis, we presented a multi-scale fuzzy measure entropy (MFME) method to acquire condition-related information buried in the bearing vibration signals.

Naturally, after feature extraction using MFME, an intelligent classifier is required to complete the fault recognition based on the extracted features. The classification performance is highly dependent on the quality of the used features which are input to the classifier. The fault feature set obtained from the bearing signals using MFME method is of high dimension and it contains irrelevant or redundant features, which makes the recognition process time-consuming and decrease the identification rate. Therefore, it is critical to select the most informative features to improve the classification efficiency as well as recognition accuracy [23]. Recently, an effective approach called infinite feature selection (Inf-FS) [24] was proposed to select the best features according to their importance. The most appealing characteristic of Inf-FS is that it evaluates the importance of a given feature while considering all the possible subsets of features [24]. For this reason, in this study, the Inf-FS method is adopted to choose the most representative features from the MFME features of bearing vibration signals.

Finally, after feature selection using the Inf-FS method, a multi-class classifier is exploited to automatically identify different bearing working conditions. Over the past decades, a variety of classification techniques have been used in mechanical fault diagnosis field. Among them, support vector machine (SVM) [25, 26] is the most extensively used ones. On the basis of statistical learning theory, SVM is suitable to deal with situations with small-quantity samples. At the same time, SVM has good generalization ability and can ensure the local and global optimal solution exactly the same [27]. Because it has high accuracy and good generalization

ability for a small number of samples [15, 27], SVM is adopted to fulfill the fault diagnosis of rolling bearing.

To sum up, a new fault diagnosis method for rolling bearing is proposed based on MFME, Inf-FS and SVM in this paper. First, the MFME values of vibration signals of rolling bearing in various scales are calculated and treated as the original fault features. Then Inf-FS is employed to choose the most informative features from the original ones and the selected features are fed into the multi-class SVM classifier. Subsequently, the different working conditions of rolling bearings are identified by means of the outputs of the trained classifier.

2 Multi-scale Fuzzy Measure Entropy

2.1 Fuzzy Entropy

The calculation steps of FuzzyEn are as follows [12, 15].

- (1) For a time series with length $N\{u(i) : 1 \leq i \leq N\}$, construct the m -dimensional vectors X_i^m as

$$X_i^m = \{u(i), u(i+1), \dots, u(i+m-1)\} - u_0(i) \quad 1 \leq i \leq N-m+1 \quad (1)$$

where X_j^m stands for a new time series, being generalized by subtracting the average of the m consecutive u values

$$u_0(i) = m^{-1} \sum_{j=0}^{m-1} u(i+j) \quad (2)$$

- (2) The distance between X_i^m and X_j^m is designated as

$$d_{ij}^m = d[X_i^m, X_j^m] = \max_{k \in [0, m-1]} |(u(i+k) - u_0(i)) - (u(j+k) - u_0(j))| \quad (3)$$

- (3) The similarity degree D_{ij}^m can be obtained by using a fuzzy function

$$D_{ij}^m = \mu(d_{ij}^m, r) \quad (4)$$

- (4) Denote $\phi_i^m(r)$ as

$$\phi_i^m(r) = (N-m-1)^{-1} \sum_{j=1, j \neq i}^{N-m} D_{ij}^m \quad (5)$$

- (5) The function $\phi^m(r)$ is defined as

$$\phi^m(r) = (N-m)^{-1} \sum_{i=1}^{N-m} \phi_i^m(r) \quad (6)$$

(6) Similarly, the $\phi^{m+1}(r)$ is obtained by repeating the above procedures

$$\phi^{m+1}(r) = (N - m)^{-1} \sum_{i=1}^{N-m} \phi_i^{m+1}(r) \tag{7}$$

(7) Then define FuzzyEn of the sequence as

$$FuzzyEn(m, r) = \lim_{N \rightarrow \infty} [\ln \phi^m(r) - \ln \phi^{m+1}(r)] \tag{8}$$

(8) Finally, for a N with finite length, FuzzyEn can be computed by

$$FuzzyEn(m, r, N) = \ln \phi^m(r) - \ln \phi^{m+1}(r) \tag{9}$$

The fuzzy function used in FuzzyEn was designated as

$$\mu(d, r, n) = e^{-d/r^n} \tag{10}$$

2.2 Fuzzy Measure Entropy

FuzzyEn subtracts the mean of the original sequence segment, which causes it to neglect the global fluctuation in the signal. The FuzzyMEN uses both the fuzzy local and global measure entropy to assess the signal complexity and thus has better discrimination ability [16]. Considering this advantage, FuzzyMEN is introduced to evaluate the bearing vibration signals' complexity in this paper. The original FuzzyEn is named as the fuzzy local measure entropy (FuzzyLEn). The difference between fuzzy global measure entropy (FuzzyGEN) and FuzzyLEn is the constitution of the vectors of m dimension [16], which is defined as

$$X_i^m = \{u(i), u(i + 1), \dots, u(i + m - 1)\} - \bar{u}_0(i) \quad 1 \leq i \leq N - m + 1 \tag{11}$$

where $\bar{u}_0(i)$ is the average of the entire sequence

$$\bar{u}_0(i) = N^{-1} \sum_{i=0}^{N-1} u(i) \tag{12}$$

The other calculation steps of FuzzyGEN are the same as FuzzyLEn. Then, the FuzzyMEN of the sequence is defined as

$$FuzzyMEN = FuzzyLEn + FuzzyGEN \tag{13}$$

2.3 Multi-scale Fuzzy Measure Entropy

The multi-scale fuzzy measure entropy (MFME) method was proposed in this paper. The steps of MFME method are listed as follows.

(1) Given the original sequence $\{u(i) : 1 \ll i \ll N\}$, pre-determine the embedding dimension m and the similar

tolerance r . Then, construct the consecutive coarse-grained time series $y_j^{(\tau)}$ based on the equation

$$y_j^{(\tau)} = 1/\tau \sum_{i=(j-1)\tau+1}^{j\tau} u(i), \quad 1 \ll j \ll \frac{N}{\tau} \tag{14}$$

where τ represents the scale factor.

(2) For the same r , compute FuzzyMEN value for each coarse-grained time series. The FuzzyMEN values in different scales can be presented as the function of scale factors, and this process is called MFME analysis.

2.4 Parameter Selection

Before the computation of MFME, five parameters need to be chosen, i.e., $m, r, N, n,$ and τ , respectively. The detailed reconstruction of the dynamic process requires large embedding dimension m . However, a too large m will result in information loss and it needs a very large N (10^m-30^m). Generally, m is set to 2. N is fixed to 2048 because the computation of entropy values is less dependent on data length. The width of the fuzzy function boundary is determined by the parameter r while the boundary gradient is determined by the parameter n . Based on the previous researches [9, 12], r is in the range of 0.1–0.25 multiplied by standard deviation (SD) and small integers be assigned to n . Here, $r=0.2SD$ is fixed while $n=2$ is selected. Finally, set the max scale factor τ in MFME as 20.

3 Infinite Feature Selection

In this study, the feature vectors are comprised of the fault features obtained using MFME in 20 scales. However, some of these extracted features may be relevant or redundant. Meanwhile, the high feature dimensionality will increase the time consumption and lower the classification accuracy. Hence, it is necessary to construct the low-dimensional feature vectors through selecting the most discriminating features. The merit of Inf-FS lies that it assesses the importance of a given feature while taking into account all possible subsets of the features [24]. Therefore, Inf-FS is used to complete the bearing fault feature selection.

Given a set of features $F = \{f^{(1)}, \dots, f^{(n)}\}$, an graph G that is indirected fully connected can be built, in which each vertex corresponds to a feature and edges model pairwise relations among features. Let an adjacency matrix A representing G , the nature of the edges can be specified [24]: each element a_{ij} of A represents a pairwise energy term defined as

$$a_{ij} = \alpha \max(\sigma^{(i)}, \sigma^{(j)}) + (1 - \alpha) \left(1 - \left| Spearman(f^{(i)}, f^{(j)}) \right| \right) \tag{15}$$

where α is a loading coefficient $\in [0, 1]$, $\sigma^{(i)}$ is the standard deviation over the samples $\{x\} \in f^{(i)}$, and *Spearman* indicates Spearman’s rank correlation coefficient [24]. A high pairwise energy shows that at least one feature among $f^{(i)}$ and $f^{(j)}$ is discriminative and there is a low correlation degree between them.

Let a finite path between vertices i and j be denoted as $\gamma = \{v_0 = i, v_1, \dots, v_{l-1}, v_l = j\}$, which is simply a subset of the feature pairs along the path. Then, define the energy of γ as

$$\epsilon_\gamma = \prod_{k=0}^{l-1} a_{v_k, v_{k+1}} \tag{16}$$

By expanding length of the path to infinity, the energy of the i th feature can be calculated

$$s(i) = \sum_{l=1}^{\infty} \sum_{j \in V} \left(\sum_{\gamma \in P_{ij}^l} \epsilon_\gamma \right) = \sum_{l=1}^{\infty} \sum_{j \in V} A^l(i, j) \tag{17}$$

where V designates the vertices set, and P_{ij}^l contains all the paths of length l between i and j [24]. However, the sum of infinite A^l terms may lead to divergence. To guarantee the convergence of the infinite sum, a real-valued regularization factor is needed as follows

$$\check{s}(i) = \sum_{l=1}^{\infty} \sum_{j \in V} r^l A^l(i, j) \tag{18}$$

Consequently, by using the convergence property of the geometric power series of a matrix [24], $\check{s}(i)$ can be efficiently calculated

$$\check{S} = (I - rA)^{-1} - I \tag{19}$$

For each feature, the final energy scores can be obtained by marginalization of the quantity

$$\check{s}(i) = [\hat{S}e]_i \tag{20}$$

It should be noted that the higher the final energy score, the more discriminative the feature, namely, the more important the feature. Hence, a rank for the selected features can be obtained by sorting the $\check{s}(i)$ energy scores in descending order. In this study, Inf-FS is utilized to choose the most significant features from the original feature set. The parameter α is initially set as 0.5 and its influence on the selection performance is investigated.

4 The Proposed Bearing Fault Diagnosis Approach

By combining the merits of MFME, Inf-FS and SVM, a new fault diagnosis approach for rolling bearing is proposed as follows.

- (1) The vibration signals of rolling bearing under different running states are acquired by using an accelerometer.
- (2) The MMFE values for each bearing vibration signal are calculated over 20 scales, and then the 20 features are used to form the original feature vector.
- (3) Inf-FS is utilized to rank the primary 20 features based on their importance. The first few features are selected as the most discriminative features and treated as the new feature vector of low dimension.
- (4) The new low-dimensional feature subsets are input into the multi-class SVM classifier and the rolling bearing fault diagnosis is fulfilled automatically.

5 Experimental Verification

5.1 Experimental Data

The vibration data obtaining from bearing data center of Case Western Reserve University [28] are used in this paper. The description of the data acquisition apparatus can be found in detail in the aforementioned literature. The bearing vibration data under the 0 horsepower load with the corresponding speed of 1797 rpm were used for analysis. The data description is presented in detail in Table 1.

5.2 Results and Analysis

To demonstrate the effectiveness of the proposed bearing fault diagnosis method, experimental analyses are carried out. In view of various fault types and fault diameters, the

Table 1 Experimental data description

| Fault type | Fault severities (mm) | Number of training data | Number of testing data | Label of classification |
|------------------|-----------------------|-------------------------|------------------------|-------------------------|
| Normal | 0 | 20 | 30 | 1 |
| Inner race fault | 0.1778 | 20 | 30 | 2 |
| | 0.5334 | 20 | 30 | 3 |
| Out race fault | 0.1778 | 20 | 30 | 4 |
| | 0.5334 | 20 | 30 | 5 |
| Ball fault | 0.1778 | 20 | 30 | 6 |
| | 0.7112 | 20 | 30 | 7 |

bearing fault diagnosis is actually a seven-class identification problem. The data set contains totally 350 data samples. 20 samples for each bearing state, a total of 140 samples are selected randomly as training data while the rest 210 are treated as testing data.

The temporal waveforms of bearing vibration signals are given in Fig. 1. As can be seen, it is difficult to identify various bearing states accurately based only on the time domain waveform, so it is necessary to further process the raw vibration signals. Since MFE focused only on the signals' local characteristics and neglected the global trend in it, it may exhibit its limitation in bearing fault feature extraction. Based on this consideration, the MFME is introduced to acquire more condition-related information from bearing vibration signals.

The MFME values over 20 scales corresponding to the vibration data illustrated in Fig. 1 are given in Fig. 2. It can be observed that, for most scales, the largest entropy values of vibration signals appear when the bearing runs under healthy state, which shows that the healthy signals are more complex than the faulty ones. This phenomenon could be interpreted in the following way [21]. The vibration over

most scales has self-similarity, and is random and irregular when the rolling bearing runs under good state. On the contrary, the self-similarity and regularity of the vibration signals will arise when bearing runs with fault, leading to lower entropy values.

However, if all the MFME values of 20 scales are employed to form the feature vector, the redundant fault information will be inevitably contained in the high-dimensional fault feature set. Meanwhile, features of high dimension will make the process of fault diagnosis time-consuming. For this reason, in order for dimension reduction, Inf-FS is introduced to select the most representative features by ranking the original 20 MFME values. Based on Inf-FS, the rearranged order of the 20 features is: 4, 17, 3, 18, 16, 8, 7, 5, 19, 15, 20, 14, 13, 2, 6, 11, 12, 9, 10 and 1. The new order of the features over different scales is presented in Fig. 3.

After the feature selection, the selected features of training samples are utilized to form the new feature vectors and train the multi-class SVM classifier, where the kernel of radial basis function (RBF) is adopted because of its merits [29]. The optimal parameters in RBF-SVM are obtained by means of the cross-validation approach. After training, the

Fig. 1 Temporal waveforms of bearing vibration signals

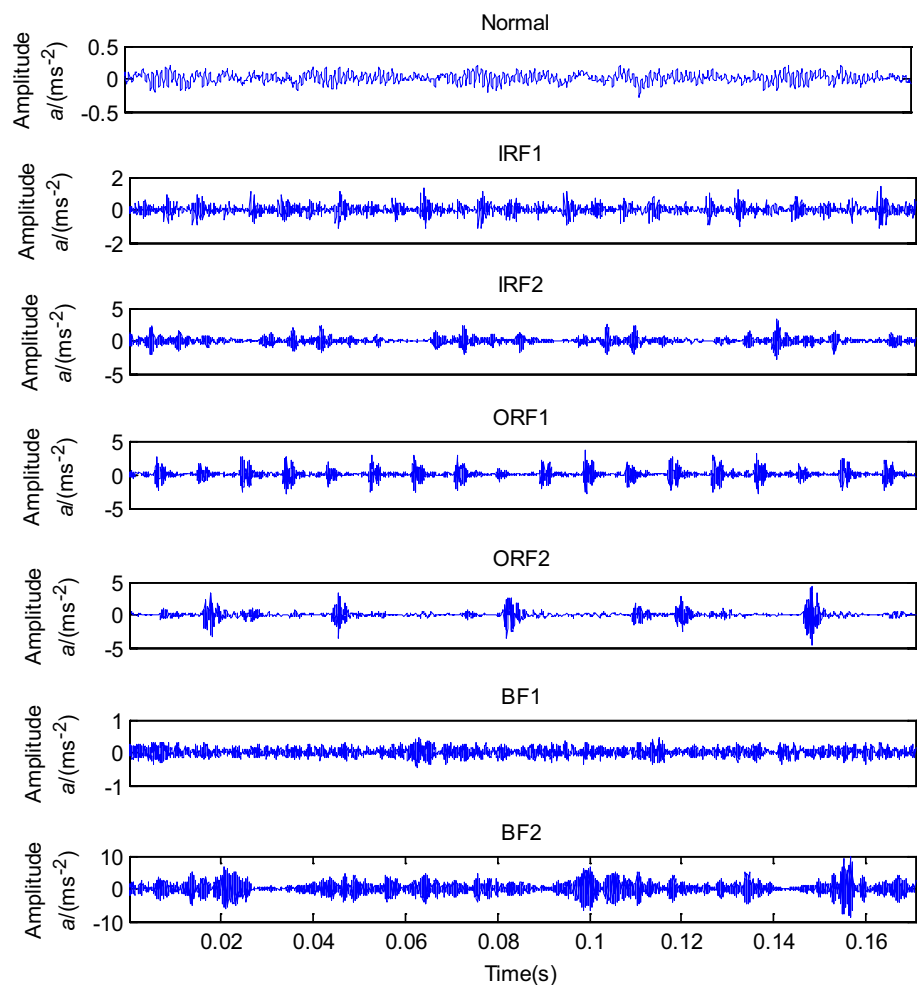


Fig. 2 MFME of vibration signals over 20 scales under seven bearing conditions

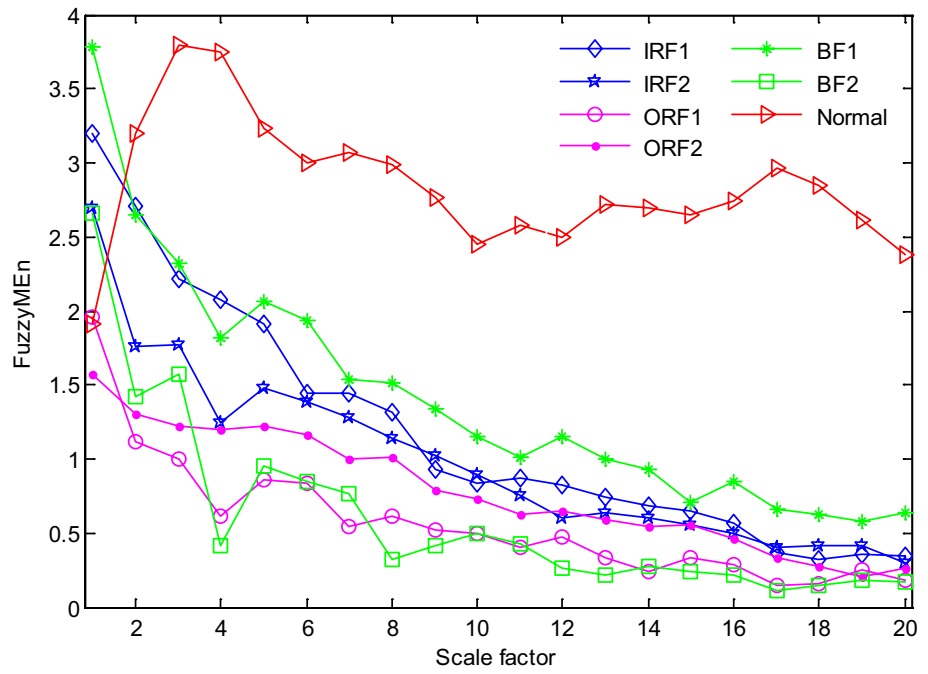
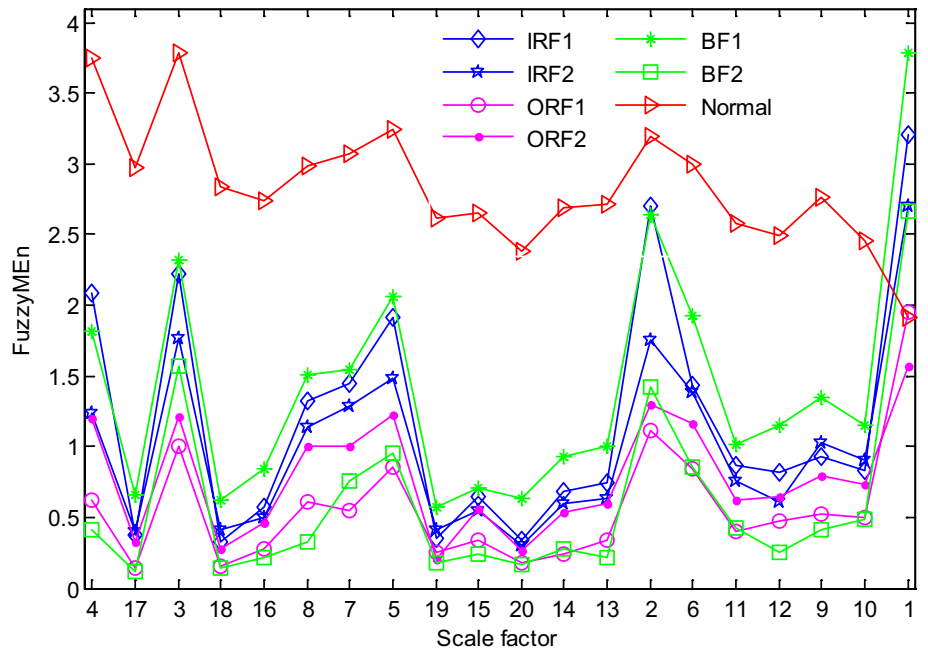


Fig. 3 New order of MFME features ranked by Inf-FS



newly selected features of test data are fed into the trained classifier, and the recognition results of test data versus the number of selected features are given in Table 2. As is presented in Table 2, the classification accuracy can achieve 100% when the first 5 to 20 features reordered by Inf-FS are used. In view of computational complexity and recognition performance, the first 5 rearranged features are treated as the discriminative features for the fault diagnosis of rolling bearing.

For comparison purpose, MFE is also used for fault feature extraction from the same vibration data and Inf-FS is

Table 2 Recognition results of testing data based on MFME and Inf-FS versus different number of features

| Method | Selected features | Number of misclassified samples | Identification rate (%) |
|---------------------|-------------------|---------------------------------|-------------------------|
| MFME + Inf-FS + SVM | First 1 | 19 | 90.95 |
| | First 2 | 17 | 91.90 |
| | First 3 | 2 | 99.05 |
| | First 4 | 1 | 99.52 |
| | First 5, ..., 20 | 0 | 100 |

applied for feature selection. The new order of MFE according to Inf-FS is: 4, 18, 17, 19, 20, 16, 15, 14, 8, 13, 5, 3, 7, 12, 11, 6, 9, 10, 2 and 1. The classification results using MFE features are shown in Table 3. It can be observed that the identification rate achieves 100% when the number of features reaches 19. Compared with five features of MFME, this will increase the complexity of computation and make the training process time-consuming. If the same number of features (five features) as MFME is used for fault pattern recognition, four samples are misclassified and the classification accuracy is 98.10%, which is below 100%. In order to show the comparison more clearly between MFE and MFME based methods, the accuracy comparison versus the number of selected features is illustrated in Fig. 4. This comparison indicates that classification performance of MFME is superior to that of MFE, which may mean that

MFME can obtain more fault-related information from the vibration signals of rolling bearing than MFE.

The coefficient α in Inf-FS is a constant in the range of 0 and 1. Generally, the selection of α values will have an effect on the identification performance. Therefore, the influence of various α values on the classification results is investigated. First, the feature selection results (MFME and MFE) using Inf-FS with varying α values are described in Table 4. Then, classification results based on MFME and Inf-FS with different α values are presented in Table 5, from which we can see that the recognition rate can reach 100% using five sensitive features in most cases. The accuracy of 100% can be obtained using at most 10 features. Correspondingly, the identification results based on MFE and Inf-FS with various α values are shown in Table 6, from which we can see that for an identification rate of 100%, the least number of used features is 9 when α equals 0.2. The above analysis further indicates the superiority of MFME over MFE.

To further validate the effectiveness of Inf-FS, another feature selection method named Laplacian Score [30] is used for comparison, which was recently applied for bearing fault feature selection [21, 22]. For the same features obtained using MFME, new order of the features sorted by LS is: 4, 19, 16, 14, 18, 17, 8, 3, 20, 7, 2, 15, 13, 12, 5, 11, 9, 1, 10 and 6. Classification results based on LS and Inf-FS are both illustrated in Fig. 5. As can be seen, the Inf-FS based identification rates are higher than or equal to those based on LS in most cases except when the selected feature number is 2. The LS based method can achieve an accuracy of 100% as the dimension of the selected features is 8, which is larger than 5 when using Inf-FS. The above analysis indicates that

Table 3 Recognition results of testing data based on MFE and Inf-FS versus different number of features

| Method | Selected features | Number of misclassified samples | Identification rate (%) |
|--------------------|-------------------|---------------------------------|-------------------------|
| MFE + Inf-FS + SVM | First 1 | 17 | 91.90 |
| | First 2 | 8 | 96.19 |
| | First 3 | 6 | 97.14 |
| | First 4,8,9 | 5 | 97.62 |
| | First 5,6 | 4 | 98.10 |
| | First 7,18 | 3 | 98.57 |
| | First 10,...,17 | 2 | 99.05 |
| | First 19,20 | 0 | 100 |

Fig. 4 Recognition rate comparison of MFE and MFME versus number of features

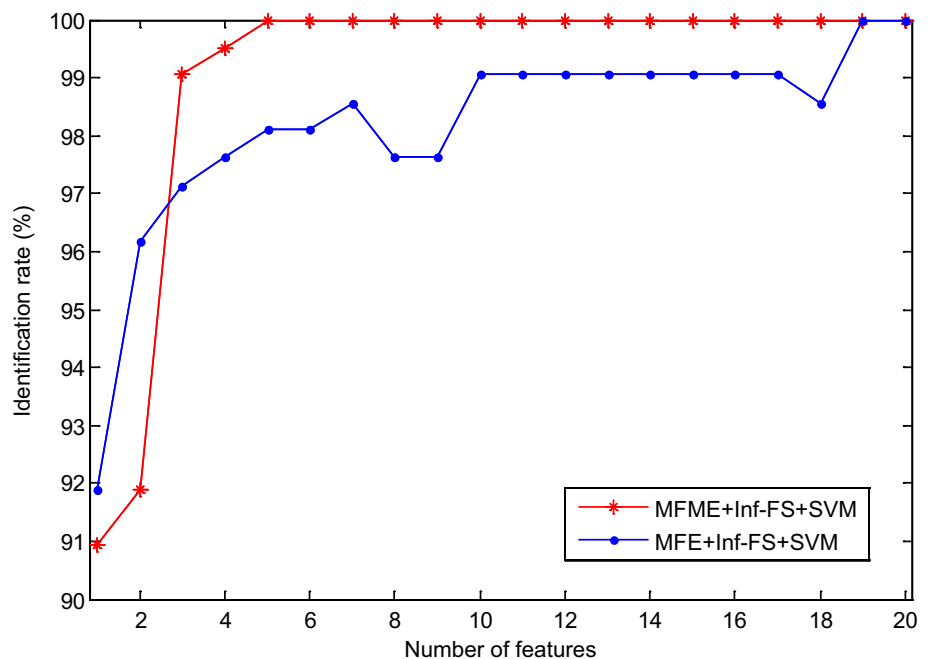


Table 4 Selection results using Inf-FS versus different α values

| α value | Features | New order of the features ranked by inf-FS |
|----------------|----------|----------------------------------------------------|
| $\alpha=0.2$ | MFME | 4,3,17,18,16,19,8,7,5,20,15,2,14,13,9,6,12,11,10,1 |
| | MFE | 4,18,17,19,20,3,16,15,1,14,8,2,13,7,5,12,9,11,6,10 |
| $\alpha=0.3$ | MFME | 4,3,17,18,16,8,19,7,5,20,15,14,2,13,9,6,12,11,10,1 |
| | MFE | 4,18,17,19,20,16,3,15,14,8,13,7,5,2,12,11,9,6,10,1 |
| $\alpha=0.4$ | MFME | 4,17,3,18,16,8,7,19,5,20,15,14,2,13,6,9,12,11,10,1 |
| | MFE | 4,18,17,19,20,16,15,3,14,8,13,7,5,12,11,6,2,9,10,1 |
| $\alpha=0.5$ | MFME | 4,17,3,18,16,8,7,5,19,15,20,14,13,2,6,11,12,9,10,1 |
| | MFE | 4,18,17,19,20,16,15,14,8,13,5,3,7,12,11,6,9,10,2,1 |
| $\alpha=0.6$ | MFME | 4,17,18,16,3,7,8,5,19,15,14,20,13,6,11,12,2,9,10,1 |
| | MFE | 4,18,17,16,19,20,15,14,8,5,13,7,12,6,11,10,9,3,2,1 |
| $\alpha=0.7$ | MFME | 4,17,18,16,7,5,8,14,3,15,20,19,13,6,11,12,10,9,2,1 |
| | MFE | 4,18,17,16,20,19,14,15,5,8,13,6,7,12,11,10,9,2,3,1 |
| $\alpha=0.8$ | MFME | 4,17,16,18,5,7,8,14,15,6,13,11,12,20,19,10,9,3,2,1 |
| | MFE | 4,18,16,17,14,15,20,5,19,8,6,13,11,12,7,10,9,2,3,1 |

in contrast with LS, the Inf-FS can select more discriminative features from the primary feature set, and the Inf-FS based bearing fault diagnosis method can acquire better recognition performance.

Furthermore, to further demonstrate the necessity of feature selection with Inf-FS, the randomly selected 5 features are used for fault pattern recognition. The two groups of features selected at random are MFME values at scales 4,

8, 12, 15 and 18, and MFME values at scales 1, 3, 5, 7 and 9. The corresponding classification accuracies are 97.14% and 99.52%, which are both lower than 100% when using Inf-FS for feature selection. This comparison results show the necessity and effectiveness of Inf-FS from another point of view.

Finally, in order to assess the effects of parameters N and m , the experimental analysis was carried out with the same data set. The effect of parameter N on the MFME values is shown in Tables 7, 8, 9, 10, 11, 12 and 13 (considering the space limitation, only the first eight scales of entropy values are presented), which correspond to the aforementioned seven bearing conditions, respectively. From Tables 7, 8, 9, 10, 11, 12 and 13, it can be observed that the MFME values are stable for all the seven states with $N=1024, 2048, 4096$ and 8192 , as indicated by the corresponding standard deviations. Such results are consistent with the discussion given in [12, 13], which is that the calculation of fuzzy entropy values depend less on the data length.

Subsequently, the effect of parameter m on the experimental results was investigated with $N=2048$. Except for the above discussed result with $m=2$ (Fig. 4), the classification results with $m=1, 3$ and 4 are illustrated in Figs. 6, 7 and 8, respectively. From Figs. 6, 7 and 8, it can be seen that the MFME-based classification rates are higher than or equal to those based on MFE in most cases. These results further validate the ability of MFME for bearing feature extraction.

6 Conclusion

A new fault diagnosis approach of rolling bearing is put forward by combining MFME, Inf-FS and SVM. The MFME algorithm is proposed for the fault feature extraction of rolling bearing, which is used to form the original feature vectors of high dimension. To reduce feature dimension and improve classification accuracy, the Inf-FS method is adopted to choose the most discriminative features. The

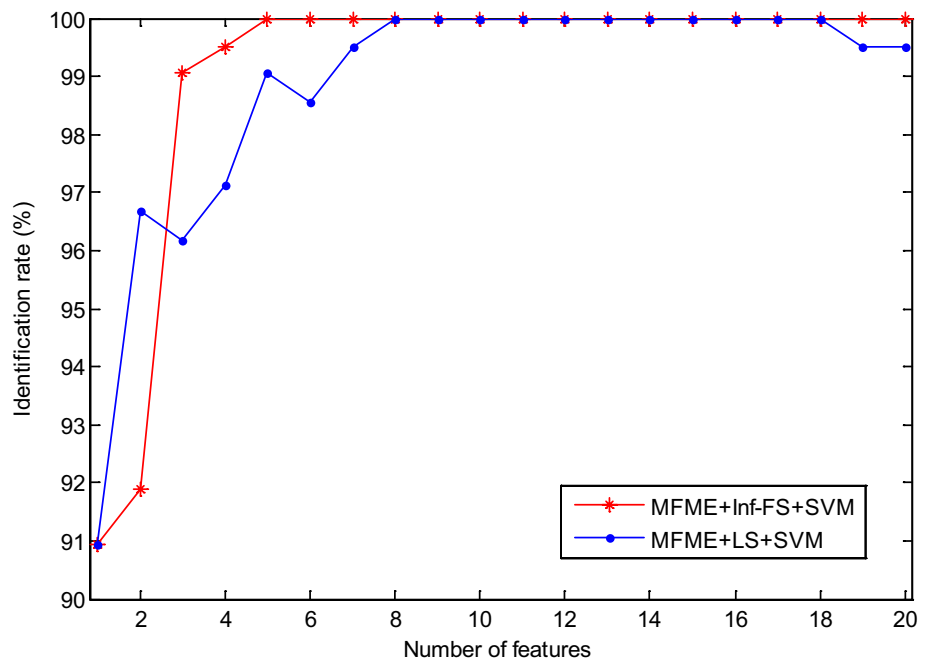
Table 5 Classification results of testing data based on MFME and Inf-FS with different number of features and varying α values

| Method | Selected features | Identification rate (%) | | | | | | |
|---------------------|-------------------|-------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | $\alpha=0.2$ | $\alpha=0.3$ | $\alpha=0.4$ | $\alpha=0.5$ | $\alpha=0.6$ | $\alpha=0.7$ | $\alpha=0.8$ |
| MFME + Inf-FS + SVM | First 1 | 90.95 | 90.95 | 90.95 | 90.95 | 90.95 | 90.95 | 90.95 |
| | First 2 | 97.62 | 97.62 | 91.90 | 91.90 | 91.90 | 91.90 | 91.90 |
| | First 3 | 99.05 | 99.05 | 99.05 | 99.05 | 97.14 | 97.14 | 90.95 |
| | First 4 | 99.52 | 99.52 | 99.52 | 99.52 | 96.19 | 96.19 | 96.19 |
| | First 5 | 100 | 100 | 100 | 100 | 100 | 97.62 | 99.05 |
| | First 6 | 100 | 100 | 100 | 100 | 100 | 99.52 | 99.52 |
| | First 7 | 100 | 100 | 100 | 100 | 100 | 99.52 | 99.52 |
| | First 8 | 100 | 100 | 100 | 100 | 100 | 99.52 | 99.52 |
| | First 9 | 100 | 100 | 100 | 100 | 100 | 100 | 99.52 |
| | First 10...20 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Table 6 Classification results of testing data based on MFE and Inf-FS with different number of features and varying α values

| Method | Selected features | Identification rate (%) | | | | | | |
|--------------------|-------------------|-------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | $\alpha=0.2$ | $\alpha=0.3$ | $\alpha=0.4$ | $\alpha=0.5$ | $\alpha=0.6$ | $\alpha=0.7$ | $\alpha=0.8$ |
| MFE + Inf-FS + SVM | First 1 | 91.90 | 91.90 | 91.90 | 91.90 | 91.90 | 91.90 | 91.90 |
| | First 2 | 96.19 | 96.19 | 96.19 | 96.19 | 96.19 | 96.19 | 96.19 |
| | First 3 | 97.14 | 97.14 | 97.14 | 97.14 | 97.14 | 97.14 | 97.14 |
| | First 4 | 97.62 | 97.62 | 97.62 | 97.62 | 96.67 | 96.67 | 96.67 |
| | First 5 | 98.10 | 98.10 | 98.10 | 98.10 | 97.14 | 97.62 | 96.67 |
| | First 6 | 98.57 | 98.10 | 98.10 | 98.10 | 98.10 | 98.10 | 96.19 |
| | First 7 | 98.10 | 98.57 | 98.57 | 98.57 | 98.57 | 98.10 | 97.62 |
| | First 8 | 98.57 | 98.57 | 98.57 | 97.62 | 97.62 | 97.62 | 98.10 |
| | First 9 | 100 | 98.57 | 98.57 | 97.62 | 97.62 | 98.10 | 98.10 |
| | First 10 | 100 | 98.57 | 98.57 | 99.05 | 98.57 | 98.57 | 98.57 |
| | First 11 | 100 | 99.05 | 99.05 | 99.05 | 99.05 | 99.05 | 98.57 |
| | First 12 | 100 | 98.10 | 98.10 | 99.05 | 99.52 | 99.05 | 99.05 |
| | First 13 | 100 | 99.05 | 99.05 | 99.05 | 99.05 | 99.05 | 98.57 |
| | First 14 | 100 | 100 | 99.05 | 99.05 | 99.05 | 99.05 | 99.05 |
| | First 15 | 100 | 100 | 99.05 | 99.05 | 99.05 | 99.05 | 99.05 |
| | First 16 | 100 | 100 | 99.05 | 99.05 | 98.57 | 98.57 | 98.57 |
| | First 17 | 100 | 100 | 100 | 99.05 | 98.57 | 98.57 | 98.57 |
| | First 18 | 100 | 100 | 100 | 98.57 | 98.57 | 100 | 100 |
| | First 19,20 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Fig. 5 Recognition results comparison of Inf-FS and LS versus number of features



selected features construct the low-dimensional feature vectors and then are applied for fault pattern recognition. For comparison purpose, MFE is also applied to the same vibration data, and the comparison results demonstrate that MFME can acquire more state-related information hidden in the vibration signals and achieve a better recognition

performance. In addition, the necessity and effectiveness of Inf-FS is verified through experimental analysis, and the comparison with LS shows the superiority of Inf-FS for feature selection. Finally, the experimental analysis demonstrates that the proposed diagnosis method can effectively identify different working conditions of rolling bearings.

Table 7 The MFME values of first 8 scales with different N values under normal condition

| N value | Normal condition | | | | | | | |
|---------|------------------|----------|----------|----------|----------|----------|----------|----------|
| | $\tau=1$ | $\tau=2$ | $\tau=3$ | $\tau=4$ | $\tau=5$ | $\tau=6$ | $\tau=7$ | $\tau=8$ |
| N=1024 | 1.9024 | 3.1887 | 3.8624 | 3.7663 | 3.1968 | 3.202 | 3.0528 | 3.0651 |
| N=2048 | 1.9136 | 3.1984 | 3.7895 | 3.7535 | 3.2382 | 3.002 | 3.0714 | 2.9908 |
| N=4096 | 1.8898 | 3.1763 | 3.7546 | 3.8238 | 3.211 | 2.953 | 3.1085 | 2.9387 |
| N=8192 | 1.8937 | 3.1688 | 3.7452 | 3.8081 | 3.1867 | 2.8715 | 3.0999 | 2.8998 |
| Std | 0.0106 | 0.0131 | 0.0532 | 0.0334 | 0.0224 | 0.1406 | 0.0257 | 0.0715 |

Table 8 The MFME values of first 8 scales with different N values under IRF1 condition

| N value | IRF1 condition | | | | | | | |
|---------|----------------|----------|----------|----------|----------|----------|----------|----------|
| | $\tau=1$ | $\tau=2$ | $\tau=3$ | $\tau=4$ | $\tau=5$ | $\tau=6$ | $\tau=7$ | $\tau=8$ |
| N=1024 | 3.3023 | 2.7833 | 2.363 | 2.1949 | 1.8121 | 1.424 | 1.3455 | 1.4091 |
| N=2048 | 3.2024 | 2.7023 | 2.217 | 2.0813 | 1.9149 | 1.4393 | 1.4457 | 1.3195 |
| N=4096 | 3.1899 | 2.6845 | 2.2034 | 2.0619 | 1.8908 | 1.4401 | 1.4249 | 1.2898 |
| N=8192 | 3.2207 | 2.7421 | 2.2082 | 2.0786 | 1.8735 | 1.478 | 1.4523 | 1.294 |
| Std | 0.0506 | 0.0440 | 0.0769 | 0.0611 | 0.0439 | 0.0230 | 0.0491 | 0.0556 |

Table 9 The MFME values of first 8 scales with different N values under IRF2 condition

| N value | IRF2 condition | | | | | | | |
|---------|----------------|----------|----------|----------|----------|----------|----------|----------|
| | $\tau=1$ | $\tau=2$ | $\tau=3$ | $\tau=4$ | $\tau=5$ | $\tau=6$ | $\tau=7$ | $\tau=8$ |
| N=1024 | 2.7337 | 1.7318 | 1.7836 | 1.2233 | 1.507 | 1.3734 | 1.324 | 1.1209 |
| N=2048 | 2.7006 | 1.7594 | 1.7716 | 1.241 | 1.48 | 1.3855 | 1.2806 | 1.1377 |
| N=4096 | 2.6754 | 1.7279 | 1.7455 | 1.2219 | 1.441 | 1.4082 | 1.2644 | 1.1082 |
| N=8192 | 2.6732 | 1.7561 | 1.7742 | 1.24 | 1.4025 | 1.4323 | 1.2832 | 1.1236 |
| Std | 0.0282 | 0.0162 | 0.0163 | 0.0104 | 0.0457 | 0.0260 | 0.0254 | 0.0121 |

Table 10 The MFME values of first 8 scales with different N values under ORF1 condition

| N value | ORF1 condition | | | | | | | |
|---------|----------------|----------|----------|----------|----------|----------|----------|----------|
| | $\tau=1$ | $\tau=2$ | $\tau=3$ | $\tau=4$ | $\tau=5$ | $\tau=6$ | $\tau=7$ | $\tau=8$ |
| N=1024 | 2.0213 | 1.074 | 0.9903 | 0.6038 | 0.877 | 0.8299 | 0.5136 | 0.5886 |
| N=2048 | 1.9556 | 1.1113 | 1.0058 | 0.6174 | 0.8588 | 0.8389 | 0.5466 | 0.6131 |
| N=4096 | 1.9277 | 1.0945 | 0.9644 | 0.5914 | 0.8369 | 0.8056 | 0.5411 | 0.5904 |
| N=8192 | 1.914 | 1.1259 | 0.9893 | 0.6038 | 0.8307 | 0.7946 | 0.5821 | 0.5707 |
| Std | 0.0477 | 0.0223 | 0.0171 | 0.0106 | 0.0212 | 0.0206 | 0.0281 | 0.0174 |

Table 11 The MFME values of first 8 scales with different N values under ORF2 condition

| N value | ORF2 Condition | | | | | | | |
|---------|----------------|----------|----------|----------|----------|----------|----------|----------|
| | $\tau=1$ | $\tau=2$ | $\tau=3$ | $\tau=4$ | $\tau=5$ | $\tau=6$ | $\tau=7$ | $\tau=8$ |
| N=1024 | 1.568 | 1.2397 | 1.1665 | 1.1246 | 1.1669 | 1.0805 | 0.9149 | 0.9487 |
| N=2048 | 1.5683 | 1.3012 | 1.2174 | 1.1998 | 1.2259 | 1.1583 | 1.0011 | 1.0084 |
| N=4096 | 1.6084 | 1.3721 | 1.2454 | 1.2112 | 1.2165 | 1.1387 | 0.9904 | 0.968 |
| N=8192 | 1.5841 | 1.3155 | 1.1583 | 1.1245 | 1.1386 | 1.0399 | 0.9424 | 0.8758 |
| Std | 0.0190 | 0.0544 | 0.0416 | 0.0470 | 0.0414 | 0.0542 | 0.0406 | 0.0555 |

Table 12 The MFME values of first 8 scales with different N values under BF1 condition

| N value | BF1 condition | | | | | | | |
|---------|---------------|----------|----------|----------|----------|----------|----------|----------|
| | $\tau=1$ | $\tau=2$ | $\tau=3$ | $\tau=4$ | $\tau=5$ | $\tau=6$ | $\tau=7$ | $\tau=8$ |
| N=1024 | 3.7483 | 2.6196 | 2.2804 | 1.7578 | 1.9475 | 1.9177 | 1.4875 | 1.4772 |
| N=2048 | 3.78 | 2.6437 | 2.3212 | 1.8149 | 2.0603 | 1.9326 | 1.542 | 1.51 |
| N=4096 | 3.7675 | 2.662 | 2.3037 | 1.7955 | 2.0381 | 1.8902 | 1.5414 | 1.4731 |
| N=8192 | 3.7864 | 2.7135 | 2.2778 | 1.8093 | 2.0343 | 1.93 | 1.5465 | 1.5145 |
| Std | 0.0168 | 0.0398 | 0.0206 | 0.0257 | 0.0497 | 0.0194 | 0.0280 | 0.0216 |

Table 13 The MFME values of first 8 scales with different N values under BF2 condition

| N value | BF2 condition | | | | | | | |
|---------|---------------|----------|----------|----------|----------|----------|----------|----------|
| | $\tau=1$ | $\tau=2$ | $\tau=3$ | $\tau=4$ | $\tau=5$ | $\tau=6$ | $\tau=7$ | $\tau=8$ |
| N=1024 | 2.8238 | 1.6314 | 1.7463 | 0.4577 | 1.0086 | 0.9885 | 0.8342 | 0.351 |
| N=2048 | 2.6641 | 1.4232 | 1.5703 | 0.4126 | 0.9514 | 0.8528 | 0.7626 | 0.3206 |
| N=4096 | 2.6647 | 1.5046 | 1.6631 | 0.4502 | 0.9654 | 0.9163 | 0.7948 | 0.3833 |
| N=8192 | 2.7292 | 1.592 | 1.7705 | 0.4716 | 0.9756 | 0.9895 | 0.8093 | 0.4365 |
| Std | 0.0754 | 0.0930 | 0.0907 | 0.0252 | 0.0243 | 0.0656 | 0.0299 | 0.0496 |

Fig. 6 Recognition rate comparison of MFE and MFME versus number of features ($m=1$)

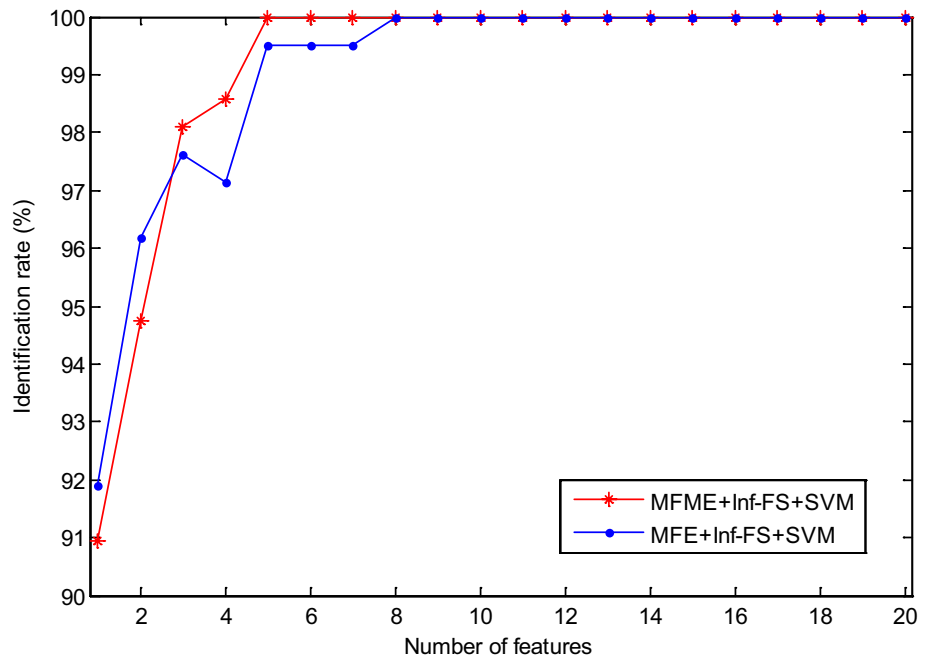


Fig. 7 Recognition rate comparison of MFE and MFME versus number of features ($m=3$)

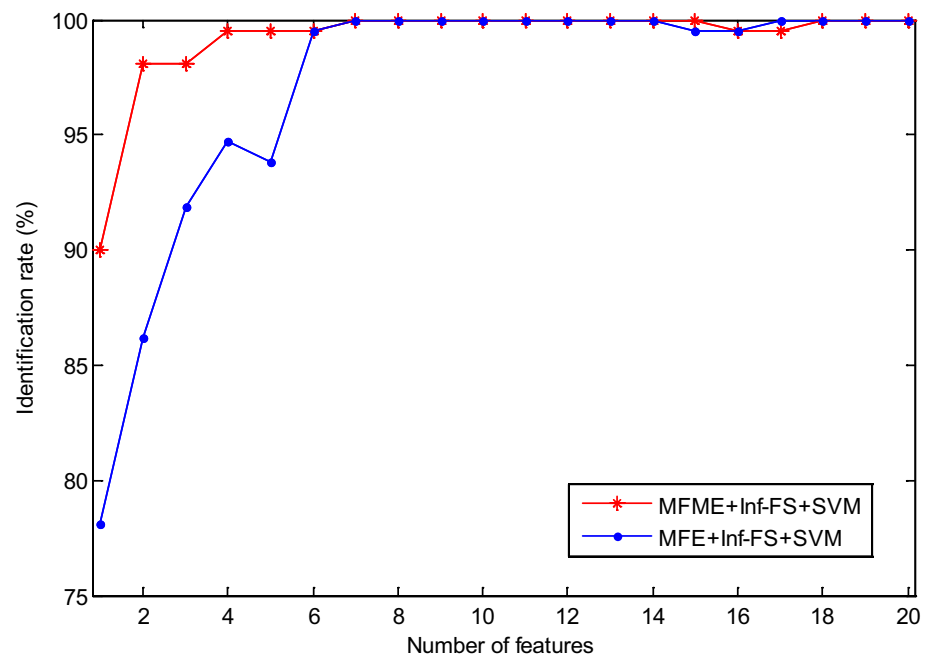
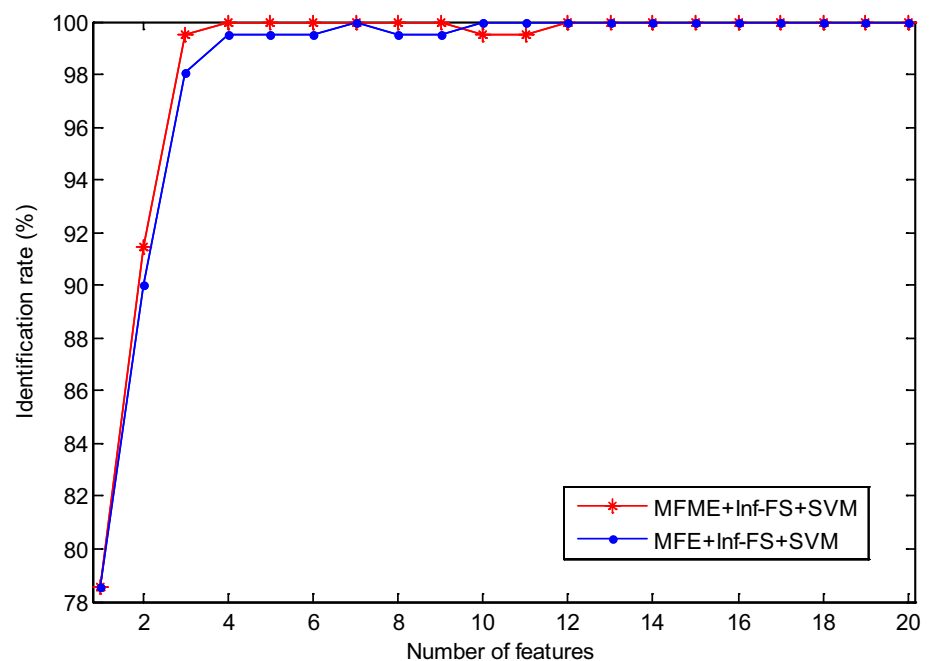


Fig. 8 Recognition rate comparison of MFE and MFME versus number of features ($m=4$)



References

1. Tiwari, R., Gupta, V.K., Kankar, P.K.: Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier. *J. Vib. Control* **21**, 461–467 (2015)
2. Kumar, A., Kumar, R.: Role of signal processing, modeling and decision making in the diagnosis of rolling element bearing defect: a review. *J. Nondestr. Eval.* **38**(1), 5 (2019)
3. Lei, Y.G., Lin, J., He, Z.J., Zi, Y.Y.: Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mech. Syst. Signal. Process.* **25**, 1738–1749 (2011)
4. Sharma, A., Amarnath, M., Kankar, P.K.: Novel ensemble techniques for classification of rolling element bearing faults. *J. Braz. Soc. Mech. Sci. Eng.* **39**(3), 709–724 (2017)
5. Dong, G., Chen, J., Zhao, F.: Incipient bearing fault feature extraction based on minimum entropy deconvolution and k-singular value decomposition. *J. Manuf. Sci. Eng.* **139**(10), 101006 (2017)

6. Vakharia, V., Gupta, V.K., Kankar, P.K.: Efficient fault diagnosis of ball bearing using ReliefF and Random Forest classifier. *J. Braz. Soc. Mech. Sci. Eng.* **39**(8), 2969–2982 (2017)
7. Feng, Z., Chen, X.: Adaptive iterative generalized demodulation for nonstationary complex signal analysis: principle and application in rotating machinery fault diagnosis. *Mech. Syst. Signal. Process.* **110**, 1–27 (2018)
8. Yang, Y., Yu, D.J., Cheng, J.S.: A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *J. Sound Vib.* **294**, 269–277 (2006)
9. Pincus, S.M.: Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. U.S.A.* **88**(6), 2297–2301 (1991)
10. Yan, R., Gao, R.X.: Approximate entropy as a diagnostic tool for machine health monitoring. *Mech. Syst. Signal. Process.* **21**, 241–250 (2007)
11. Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **278**, H2039–H2049 (2000)
12. Chen, W.T., Wang, Z.Z., Xie, H.B., Yu, W.: Characterization of surface EMG signal based on fuzzy entropy. *IEEE Trans. Neural Syst. Rehabil. Eng.* **15**(2), 266–272 (2007)
13. Chen, W.T., Zhuang, J., Yu, W.: Measuring complexity using FuzzyEn, ApEn and SampEn. *Med. Eng. Phys.* **31**, 61–68 (2009)
14. Zheng, J.D., Cheng, J.S., Yang, Y.: A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy. *Mech. Mach. Theory* **70**, 441–453 (2013)
15. Zhu, K.H., Li, H.L.: A rolling element bearing fault diagnosis approach based on hierarchical fuzzy entropy and support vector machine. *Proc. IMechE Part C J. Mech. Eng. Sci.* **230**(13), 2314–2322 (2016)
16. Liu, C.Y., Li, K., Zhao, L.N., Liu, F., Zheng, D., Liu, C., Liu, S.: Analysis of heart variability using fuzzy measure entropy. *Comput. Biol. Med.* **43**, 100–108 (2013)
17. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale entropy analysis of physiologic time series. *Phys. Rev. Lett.* **89**, 062102 (2002)
18. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale entropy analysis of biological signals. *Phys. Rev. E* **71**, 021906 (2005)
19. Liu, H.H., Han, M.H.: A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings. *Mech. Mach. Theory* **75**, 67–78 (2014)
20. Zhang, L., Xiong, G.L., Liu, H.S.: Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. *Expert Syst. Appl.* **37**, 6017–6085 (2010)
21. Zheng, J.D., Cheng, J.S., Yang, Y., Luo, S.: A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination. *Mech. Mach. Theory* **78**, 187–200 (2014)
22. Li, Y.B., Xu, M.Q., Wang, R.X., Huang, W.: A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy. *J. Sound Vib.* **360**, 277–299 (2016)
23. Cerrada, M., Sánchez, R.V., Pacheco, F., Cabrera, D., Zurita, G., Li, C.: Hierarchical feature selection based on relative dependency for gear fault diagnosis. *Appl. Intell.* **44**(3), 687–703 (2016)
24. Giorgio, R., Melzi, S., Cristani, M.: Infinite feature selection. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4202–4210 (2015)
25. Zhu, K.H., Song, X.G., Xue, D.X.: A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm. *Measurement* **47**, 669–675 (2014)
26. Rapur, J.S., Tiwari, R.: On-line time domain vibration and current signals based multi-fault diagnosis of centrifugal pumps using support vector machines. *J. Nondestr. Eval.* **38**(1), 6 (2018)
27. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)
28. Lou, X., Loparo, K.A.: Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mech. Syst. Signal. Process.* **18**(5), 1077–1095 (2004)
29. Widodo, A., Yang, B.S.: Support vector machine in machine condition monitoring and fault diagnosis. *Mech. Syst. Signal. Process.* **21**(6), 2560–2574 (2007)
30. He, X., Cai, D., Niyogi, P.: Laplacian score for feature selection. *Adv. Neural. Inf. Process. Syst.* **18**, 507–514 (2005)

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