

Intelligent Fault Diagnosis of Wind Turbine Gearbox based on Refined Generalized Multi-scale State Joint Entropy and RSFS Feature Selection

Wei Dong

Yanshan University

shuqing zhang (✉ zhshp_yd@163.com)

Yanshan University <https://orcid.org/0000-0002-2022-7136>

Mengfei Hu

Yanshan University

Liguo Zhang

Yanshan University

Haitao Liu

Yanshan University

Research Article

Keywords: State joint entropy, refined generalized composite multi-scale state joint entropy, multi-scale average Euclidean Divergence, robust spectral learning framework for unsupervised feature selection, feature selection

Posted Date: June 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-521872/v1>

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

[Read Full License](#)

Intelligent fault diagnosis of wind turbine gearbox based on refined generalized multi-scale state joint entropy and RSFS feature selection

Wei Dong, Shuqing Zhang, Mengfei Hu, Liguang Zhang, Haitao Liu

Institute of Electrical Engineering, Engineering Research Center of the Ministry of Education for Intelligent Control System and Intelligent Equipment, Yanshan University, Qinhuangdao 066004, China

Corresponding author: Shuqing Zhang (e-mail: zhshq-yd@163.com).

This work was supported in part by the National Natural Science Foundation of Hebei Province under Grant F2020203058, and the Key R&D projects of Hebei Province under Grant 18211833D

ABSTRACT The fault diagnosis of gearbox and bearing in wind turbine is crucial to improve service life and reduce maintenance cost. This paper proposes a novel fault diagnosis method based on refined generalized composite multi-scale state joint entropy (RGCMSJE), robust spectral learning framework for unsupervised feature selection (RSFS) and extreme learning machine (ELM) to identify the different health conditions of gearboxes, including feature extraction, feature reduction and pattern recognition. In this method, MAED is firstly adopted to assist RGCMSJE in parameter selection. Second, RGCMSJE is utilized to extract the multi-scale features of gearbox vibration signal and construct high-dimension feature set. Thirdly, RSFS method is used to reduce the dimension of high-dimensional RGCMSJE feature set. In the end, the obtained low-dimensional features are input to the ELM classifier to realize fault pattern recognition. Through two gearbox fault diagnosis experiments, the effectiveness of the fault diagnosis method is verified. The analysis results show that this method can effectively and accurately identify different fault types of wind turbine gearbox.

Keywords: State joint entropy, refined generalized composite multi-scale state joint entropy, multi-scale average Euclidean Divergence, robust spectral learning framework for unsupervised feature selection, feature selection.

1 Introduction

As a clean and environment-friendly renewable energy, wind power has become the focus of global search for new energy, and has achieved large-scale development and utilization all over the world [1, 2]. By the end of 2019, GWEC

data show that the total installed capacity of global wind energy is 651GW, a year-on-year increase of 10%, and it is expected to reach 2930GW in 2040 [3]. Unfortunately, with the increasing number of wind turbines (WTs) installed, due to its complex and harsh operating environment, WTs failures occur frequently. This will lead to high operation and maintenance (O&M) costs and huge economic losses. It is worth noting that gearbox and bearing are one of the most critical components of wind turbine, and also the main place of failure. Therefore, it is crucial to find a reliable fault diagnosis methodology to timely and accurately monitor the operation status of gearbox and bearing in order to reduce the cost of operation and maintenance (O&M) [4].

Gearbox and bearing, as an important part of WTs, have a great influence on the life of the whole wind power system [5]. The analysis methods for gearbox and bearing are usually based on vibration signals, which provide more useful information than photoelectric signals, acoustic emission signals, temperature signals and other types of signals [6]. However, due to the instability of wind speed and strong background noise, the vibration signals are usually nonlinear, nonstationary and mixed with noise. Especially in different working conditions, the weak characteristics of vibration signal are difficult to obtain, so it is difficult to identify the type of fault [7].

Fortunately, many scholars have proposed very interesting solutions based on data-driven and machine learning methods. These schemes mainly include two steps: feature extraction and pattern recognition. The key step affecting the performance of the latter is feature extraction [8]. Traditional time-frequency analysis method needs the professional knowledge of investigators and the detailed information of mechanical components as support, which greatly limits its application [9,10]. Recently, with the development of nonlinear technology, feature extraction based on entropy method has gradually become an important research topic in WTs fault diagnosis [11, 12, 13].

Entropy, as an index to measure the dynamic characteristics of time series, has been widely applied to feature extraction of gearbox and bearing [14], [15], such as sample entropy (SE) [16], permutation entropy (PE) [17], fuzzy entropy (FuzzyEn) [18], and dispersion entropy (DE) [19]. Among them, the SE time consumption is huge, especially when dealing with long time series [20]. PE is obviously faster than SE, but the amplitude relationship between time series is not

taken into account [21]. The feature extraction ability of FuzzyEn is limited by fuzzy membership function [22]. Significantly, DE has the most prominent advantage, its computational efficiency is significantly higher than SE, PE and FuzzyEn, and entropy estimation is stable and effective. This is attributed to the symbol mapping based on statistics and the probability distribution based on embedding pattern in DE method [23]. However, these entropy algorithms ignore the dynamic characteristics of time series, that is, the probability of time series changing from the current state to the next state. Therefore, based on the advantages of DE, we propose a new entropy estimation algorithm called state joint entropy (SJE). The algorithm considers the joint distribution of the current state and the next state. The test results of simulation and experimental signals show that SJE not only inherits the efficiency and stability of DE algorithm, but also can extract more fault information. Moreover, a large number of studies show that the entropy algorithm only estimates the irregularity and complexity of the signal on a single time scale and hence some important fault information in other scales may be discarded [24,25,26]. In order to extract fault features from multiple time scales, refined composite multi-scale analysis is proposed by Azami et al [27]. It not only solves the problem of undefined entropy in composite multi-scale analysis, but also increases the stability of results for long time series. However, refined composite multi-scale analysis mainly considers the average process, which means that with the increase of scale factor, the variance of entropy value will increase rapidly and lose statistical stability. Therefore, this paper proposes a refined generalized composite multi-scale analysis to make up for the above shortcomings by defining the mean and variance of the whole time series. Finally, this paper combines state joint entropy and refined generalized composite multi-scale analysis into refined generalized composite multi-scale state joint entropy (RGCMSJE), which is used as the feature extraction method of gearbox and bearing.

As is well known, after extracting high-dimensional features by multi-scale analysis, feature selection is usually needed to reduce the computational burden. At present, Laplacian score (LS) [28], Fisher score (FS) [29] and Max-relevance and min-redundancy (mRMR) [30] have been widely used in the feature selection tasks in various fields. However, LS only focuses on the similarity of adjacent samples, and ignores the ability of global information separation. Moreover, FS

shows the opposite characteristics. It focuses on the global separation performance of samples, and does not consider the similarity of adjacent samples [31]. Therefore, the features selected by FS and LS cannot effectively represent the separability of multi class samples. Max-relevance and min-redundancy (mRMR) is proposed based on the principle of maximizing the between-class distance and minimizing the within-class distance, but it has a large CPU time consumption, which is not conducive to fast calculation. In order to solve the above problems, this paper introduces RSFS method into feature selection process. This method uses robust local learning method to deal with the noise on the clustering label, so as to improve the local information retention ability while taking into account the global information [32]. Two experimental results show that the multi class features extracted by RSFS are more separable while ensuring the efficiency. Finally, the widely used extreme learning machine (ELM) is used to identify different fault types of gearbox and bearing. ELM not only has higher generalization ability, compared with softmax regression (SR) [33], support vector machine (SVM) [34], random forest (RF) [35] and k-nearest neighbor (KNN) [36] algorithms, but also has the advantages of fast calculation speed and less manual intervention [37-38].

In summary, the contributions of this paper are as follows:

(1) SJE method is proposed to extract the current state and the next state of the time series, which can extract more fault features while inheriting the efficiency of DE algorithm.

(2) RGCMSJE method is proposed to extract features on multiple time scales, which makes up for the deficiency of refined composite multi-scale analysis. Multi-scale average Euclidean Divergence (MAED) is proposed to automatically select the parameters of RGCMSJE method.

(3) RSFS is introduced for the first time to select sensitive features with higher distinguishability as the input of ELM intelligent classifier.

(4) The fault diagnosis method of RGCMSJE-RSFS-ELM is proposed systematically, and its effectiveness is objectively verified through MAED parameter selection and comparative experiments.

The rest of this paper is organized as follows. Section 2 introduces the theoretical derivation and performance comparison of SJE. Section 3 gives the detailed steps of RGCMSJE and MAED. Section 4 introduces the RSFS for the

sensitive feature selection. Section 5 illustrates the ELM algorithm and the process of the proposed method based on RGCMSJE-RSFS-ELM. Section 6 introduces two detailed experiments and discusses the results. Finally, conclusions are drawn in Section 7.

2 State joint entropy

2.1 The procedures of SJE

For a given univariate discrete time series of length $N : x_i (i = 1, 2, \dots, N)$, the SJE algorithm includes eight main steps as follows:

(1). Mapping process from time series to symbol series. First, the fluctuation range of time series is divided into c categories, labeled from ε_1 to ε_c . Each value in the time series corresponds to a unique category. Therefore, a symbol mapping domain $\varepsilon_j (j = 1, 2, \dots, c)$ can be obtained. The symbol series $z_i (i = 1, 2, \dots, N)$ can be obtained by a series of symbol domain elements. There are several mapping algorithms that can be used in the symbolization process, such as linear mapping (LM) [39], maximum entropy partition (MEP) [40], unified quantization (IAUQ) [41] and exact cumulative distribution (NCDF) [42]. Although the LM algorithm is the fastest one, when the maximum value of time series is far greater than the average value, the z_i approaches a straight line. MEP algorithm needs to set standard data, which increases the difficulty of symbolization. The fluctuations of signals are not taken into account by the IAUQ. The NCDF mapping process comes from dispersion entropy, which maps the original time series to a distribution of 0 to 1. Then a linear algorithm is used to map the distribution to an integer from ε_1 to ε_c . The mapping process is expressed as follows [42]:

$$z_i^c = \text{round}(c \cdot \text{NCDF}(x_i) + 0.5) \quad (1)$$

However, there is an unsolvable problem in the above linear mapping process, when $\text{NCDF}(x_i) = 1$, $z_i^c = c + 1$. This is beyond the definition of the range of mapping 1 to c , so the undefined symbol $c + 1$ appears. Therefore, we propose a modified linear mapping process mNCDF, which eliminates the undefined mapping by adding constraints. The revised mapping process is as follows:

$$\begin{cases} z_i^c = \text{round}(c \cdot \text{NCDF}(x_i) + 0.5) & \text{if } 0 \leq \text{NCDF}(x_i) < 1 \\ z_i^c = c & \text{if } \text{NCDF}(x_i) = 1 \end{cases} \quad (2)$$

where z_i^c represents the j^{th} member of the classified time series.

(2). Construct the embedding vectors and state vector based on the symbol series $z_i (i=1, 2, \dots, N)$ according to Eq. (3)

$$\begin{cases} \mathbf{z}_k^{m,c} = \{z_k^c, z_{k+d}^c, \dots, z_{k+(m-1)d}^c\} \\ \mathbf{z}_k^{T(m,c)} = \{z_{k+d}^c, z_{k+2d}^c, \dots, z_{k+(m-1)d}^c\} \end{cases} \quad (3)$$

$$k = 1, 2, \dots, N - (m-1)d$$

where $\mathbf{z}_k^{m,c}$ is the embedding vectors and $\mathbf{z}_k^{T(m,c)}$ is the state vectors, m is the embedding dimension and d is delay time. The dispersion pattern of $\mathbf{z}_k^{m,c}$ and $\mathbf{z}_k^{T(m,c)}$ are π_{V_β} where $\beta = 1, 2, \dots, c^m$. The number of possible dispersion patterns is equal to c^m . Because $\mathbf{z}_k^{m,c}$ or $\mathbf{z}_k^{T(m,c)}$ has m element and each element may have c possible symbols.

(3). The state value corresponding to the embedded vector $\mathbf{z}_k^{m,c}$ to the state vector $\mathbf{z}_k^{T(m,c)}$ is expressed as $q_\beta^{a,m,d} (a = 1, 2, \dots, c)$. Therefore, we could calculate the probability of state vector $\mathbf{z}_k^{T(m,c)}$ corresponding to state value $q_\beta^{a,m,d} (a = 1, 2, \dots, c)$ as follows:

$$\begin{aligned} p(\mathbf{z}_k^{T(m,c)}) \\ = \frac{\text{Number}\{k \mid k \leq N - (m-1)d, \text{type}(\mathbf{z}_k^{T(m,c)}) = q_\beta^{a,m,d}\}}{N - (m-1)d - 1} \end{aligned} \quad (4)$$

where $\text{type}(\bullet)$ represents the corresponding state value of the state vector.

(4). Construct a state values transition matrix $Q_\beta^{a,m,d}$ according to the state values $q_\beta^{a,m,d}$ as follows:

$$Q_{\beta}^{a,m,d} = \begin{bmatrix} q_1^{1,m,d} & q_1^{2,m,d} & \cdots & q_1^{c,m,d} \\ \vdots & \vdots & & \vdots \\ q_r^{1,m,d} & q_r^{2,m,d} & \cdots & q_r^{c,m,d} \\ \vdots & \vdots & & \vdots \\ q_{c^m}^{1,m,d} & q_{c^m}^{1,m,d} & \cdots & q_{c^m}^{c,m,d} \end{bmatrix} \quad (5)$$

Where $1 < r < c^m$. Then the state transition matrix under $Q_{\beta}^{a,m,d}$ condition could be obtained as follows:

$$T_{\beta}^{a,m,d} = \begin{bmatrix} (q_1^{1,m,d} | \mathbf{z}_{(\pi_{1_1})}^{T(m,c)}) & (q_1^{2,m,d} | \mathbf{z}_{(\pi_{1_1})}^{T(m,c)}) & \cdots & (q_1^{c,m,d} | \mathbf{z}_{(\pi_{1_1})}^{T(m,c)}) \\ \vdots & \vdots & & \vdots \\ (q_r^{1,m,d} | \mathbf{z}_{(\pi_{1_r})}^{T(m,c)}) & (q_r^{2,m,d} | \mathbf{z}_{(\pi_{1_r})}^{T(m,c)}) & \cdots & (q_r^{c,m,d} | \mathbf{z}_{(\pi_{1_r})}^{T(m,c)}) \\ \vdots & \vdots & & \vdots \\ (q_{c^m}^{1,m,d} | \mathbf{z}_{(\pi_{c^m})}^{T(m,c)}) & (q_{c^m}^{1,m,d} | \mathbf{z}_{(\pi_{c^m})}^{T(m,c)}) & \cdots & (q_{c^m}^{c,m,d} | \mathbf{z}_{(\pi_{c^m})}^{T(m,c)}) \end{bmatrix} \quad (6)$$

Where $\pi_{V_{\beta}}$ is the β th dispersion mode. $\mathbf{z}_{(\pi_{V_{\beta}})}^{T(m,c)}$ is the dispersion mode of the $\mathbf{z}_k^{T(m,c)}$.

The state transition matrix shows that the symbol time series moves from one state to another with time.

(5). According to the state transition matrix, using Eq. (7) to calculate its probability as follows:

$$P\left(q_{\beta}^{a,m,d} | \mathbf{z}_{(\pi_{V_{\beta}})}^{T(m,c)}\right) = \frac{\text{Number}\left\{k \mid k \leq N - (m-1)d, \text{type}(\mathbf{z}_{(\pi_{V_{\beta}})}^{T(m,c)}) = q_k^{a,m,d}, T_{\beta}^{a,m,d}\right\}}{\text{sum}(\text{Number})} \quad (7)$$

(6). According to the derivation formulas of joint entropy and conditional entropy [43], Eq. (8) can be obtained as follows:

$$H(\mathbf{z}_k^{m,c}, q_{\beta}^{a,m,d}) = H(q_{\beta}^{a,m,d} | \mathbf{z}_k^{T(m,c)}) + H(\mathbf{z}_k^{m,c}) \quad (8)$$

Finally, the value of SJE is calculated according to Eq. (9) as follows:

$$SJE(x, m, d, c) = -\sum_{k=1}^c P(\mathbf{z}_k^{T(m,c)}, q_\beta^{a,m,d}) - \sum_{k=1}^c p(\mathbf{z}_k^{m,c}) \ln(p(\mathbf{z}_k^{m,c})) \quad (9)$$

Where the value of $P(\mathbf{z}_k^{T(m,c)}, q_\beta^{a,m,d})$ is defined as follows:

$$P(\mathbf{z}_k^{T(m,c)}, q_\beta^{a,m,d}) = p(\mathbf{z}_k^{T(m,c)}) p(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)}) \ln(p(\mathbf{z}_k^{T(m,c)}) p(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)})) \quad (10)$$

$$p(\mathbf{z}_k^{m,c}) = \frac{\text{Number}\{k | k \leq N - (m-1)d, \mathbf{z}_k^{m,c} \text{ type}(\pi_{V_\beta})\}}{N - (m-1)d - 1} \quad (11)$$

where $\text{type}(\pi_{V_\beta})$ is the dispersion pattern of $\mathbf{z}_k^{a,m,c}$.

In order to understand the calculation process of SJE in detail, an example is given to illustrate. For a time series x with $d=1$, $m=2$ and $c=2$ ($x = \{1.2, -2.1, -0.5, 0.4, -1.7, 1.4, 3.6, 1, 0.9, -1, -2.5, 2\}$). First, the dispersion space y is obtained according to the NCDF algorithm. Then the symbolized time series $z = \{2, 1, 1, 2, 1, 2, 2, 2, 2, 1, 1, 2\}$ could be obtained according to for Eq. (2). There are $c^m = 2^2 = 4$ possible dispersion patterns ($\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$). Then the embedded vector and the state vector can be obtained as follows:

$$\begin{aligned} \mathbf{z}_1^{2,2} &= \{2, 1\}(\pi_{21}), \quad \mathbf{z}_2^{2,2} = \{1, 1\}(\pi_{11}), \quad \mathbf{z}_3^{2,2} = \{1, 2\}(\pi_{12}), \quad \mathbf{z}_4^{2,2} = \{2, 1\}(\pi_{21}), \\ \mathbf{z}_5^{2,2} &= \{1, 2\}(\pi_{12}), \quad \mathbf{z}_6^{2,2} = \{2, 2\}(\pi_{22}), \quad \mathbf{z}_7^{2,2} = \{2, 2\}(\pi_{22}), \quad \mathbf{z}_8^{2,2} = \{2, 2\}(\pi_{22}), \\ \mathbf{z}_9^{2,2} &= \{2, 1\}(\pi_{21}), \quad \mathbf{z}_{10}^{2,2} = \{1, 1\}(\pi_{11}). \\ \mathbf{z}_1^{T(2,2)} &= \{1, 1\}(\pi_{11}), \quad \mathbf{z}_2^{T(2,2)} = \{1, 2\}(\pi_{12}), \quad \mathbf{z}_3^{T(2,2)} = \{2, 1\}(\pi_{21}), \quad \mathbf{z}_4^{T(2,2)} = \{1, 2\}(\pi_{12}), \\ \mathbf{z}_5^{T(2,2)} &= \{2, 2\}(\pi_{22}), \quad \mathbf{z}_6^{T(2,2)} = \{2, 2\}(\pi_{22}), \quad \mathbf{z}_7^{T(2,2)} = \{2, 2\}(\pi_{22}), \quad \mathbf{z}_8^{T(2,2)} = \{2, 1\}(\pi_{21}), \\ \mathbf{z}_9^{T(2,2)} &= \{1, 1\}(\pi_{11}), \quad \mathbf{z}_{10}^{T(2,2)} = \{1, 2\}(\pi_{12}). \end{aligned}$$

The number is $N - (m-1)d - 1 = 10$ of embedded vectors and state vectors. The probability of state vector as follows:

$$[p(\mathbf{z}^{T(1,1)}), p(\mathbf{z}^{T(1,2)}), p(\mathbf{z}^{T(2,1)}), p(\mathbf{z}^{T(2,2)})] = [\frac{2}{10}, \frac{3}{10}, \frac{2}{10}, \frac{3}{10}]$$

Third, generate the probability matrix of the state transition matrix according to Eq. (6) as follows:

$$\begin{bmatrix} P(q_1^{1,2,1} | \mathbf{z}_{(\pi_{11})}^{T(2,2)}) & P(q_1^{2,2,1} | \mathbf{z}_{(\pi_{11})}^{T(2,2)}) \\ P(q_2^{1,2,1} | \mathbf{z}_{(\pi_{12})}^{T(2,2)}) & P(q_2^{2,2,1} | \mathbf{z}_{(\pi_{12})}^{T(2,2)}) \\ P(q_3^{1,2,1} | \mathbf{z}_{(\pi_{21})}^{T(2,2)}) & P(q_3^{2,2,1} | \mathbf{z}_{(\pi_{21})}^{T(2,2)}) \\ P(q_4^{1,2,1} | \mathbf{z}_{(\pi_{22})}^{T(2,2)}) & P(q_4^{2,2,1} | \mathbf{z}_{(\pi_{22})}^{T(2,2)}) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \\ 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix}$$

Finally, the SJE value is calculated according to Eq. (8).as follows:

$$SJE(x, m, d, c) = -\sum_{k=1}^{c^m} P(\mathbf{z}_k^{T(m,c)}, q_\beta^{a,m,d}) - \sum_{k=1}^{c^m} p(\mathbf{z}_k^{m,c}) \cdot \ln(p(\mathbf{z}_k^{m,c})) = 3.2585$$

From the Eq. (8) and the Eq. (9), we can find that the total number of all possible dispersion modes of embedded vector and state vector is c^m and c^{m+1} respectively.

The normalized SJE is also calculated as follows:

$$NSJE(x, m, d, c) = \frac{H(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)})}{\ln(\varepsilon^{m+1})} + \frac{H(\mathbf{z}_k^{m,c})}{\ln(\varepsilon^m)} = 1.8955$$

According to the Eq. (8) of SJE, it can be concluded that the greater the irregularity of the time series, the bigger value of SJE. At the same time, the more regular the time series, the smaller the value of SJE. In order to explain the process of the above example in more detail, the image description is given through Fig.1. It should be noted that Fig. 1 finally gets the values of $P(\mathbf{z}_k^{T(m,c)}, q_\beta^{a,m,d})$, $p(\mathbf{z}_k^{T(m,c)})$ and $p(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)})$. Fig. 2 graphically shows all possible dispersion patterns of embedded vectors and state vectors. The number of state dispersion patterns is $c^m = 4$, each of them is presented in Fig. 2(b). Where, blue points represent the original state value, red points represent the newly emerged state value, and red lines represent the newly emerged state dispersion mode.

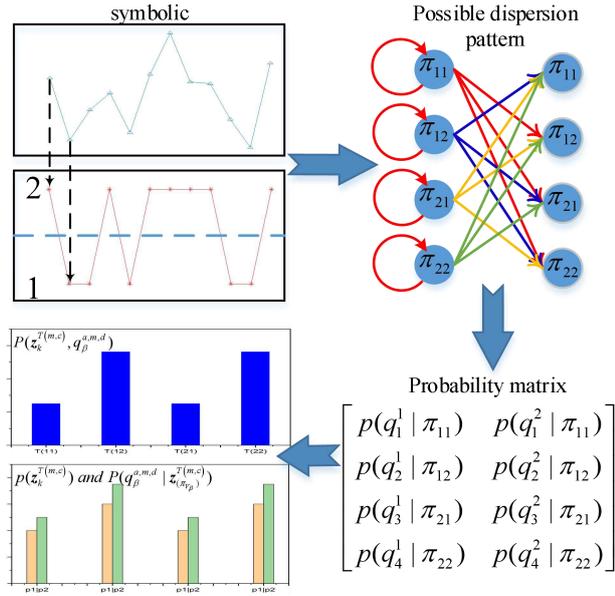


Fig. 1 An example of the SJE method with symbol number $c = 2$, embedding number $m = 2$ and time delay $d = 1$.

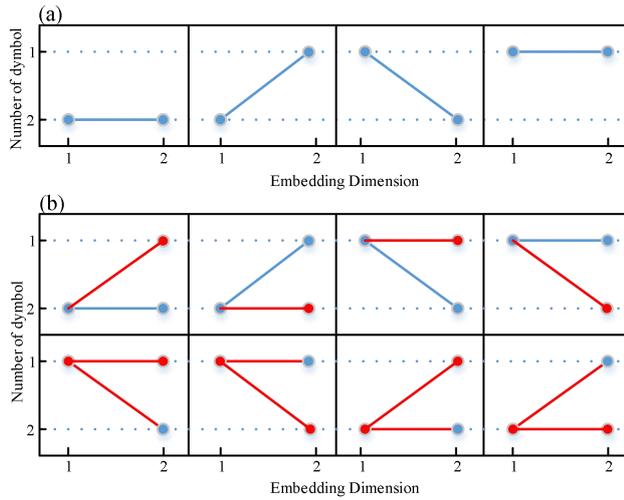


Fig. 2 Illustration of the dispersion patterns: (a) The possible dispersion patterns of embedded vectors. (b) The possible dispersion patterns of state vectors.

2.2 Symbolization performance comparison

To validate the advantages of mNCDF algorithm for mapping time series, a fault signal model of rolling bearing in Ref [44] is employed to simulate the outer race fault (ORF). Symbolize the model signal through a mapping algorithm. Then the time domain and frequency domain analysis is performed on the symbolized sequence signal. Finally, the ability of different mapping algorithms to retain the original time signal information could be compared through time domain and

frequency domain images. The mathematical expression of the fault signal model as follows:

$$\begin{cases} x(t) = \sum_{i=1}^M A_i s(t - iT - \tau_i) + w(t) \\ A_i = A_0 \cos(2\pi Q t + \phi_A) + C_A \\ s(t) = e^{-Bt} \sin(2\pi f_n t + \phi_w) \end{cases} \quad (12)$$

where A_i is the amplitude modulation signal with the frequency $Q = 33$ Hz, A_0 is the amplitude of the signal with $A_0 = 3$, C_A is a constant with $C_A > A_0$, B is the damping coefficient with $B = 0.1$, f_n is the natural frequency of the system with $f_n = 3000$ Hz. τ_i is the time lag with $\tau_i = 0$ and $w(t)$ is the white Gaussian noise. The other parameters values in Eq. (12) are set as follows, the fault frequency of outer race f_0 is 124 Hz, the sampling frequency $f_s = 12$ kHz, the data length used for analysis is $L = 10000$, $w(t) = 0$ and the interval time is $T = 0.01$.

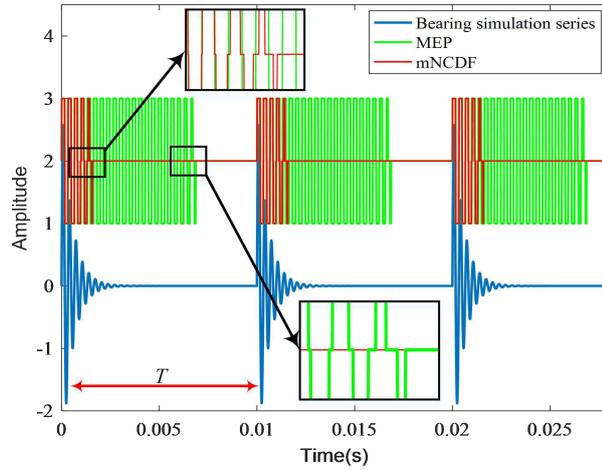


Fig. 3 The comparison of symbolic capabilities between mNCDF and MEP, recorded over 0.025s sample points. The bearing simulation series is the outer race fault (ORF).

The time-domain waveform of the simulation series and the symbol series obtained by mNCDF and MEP algorithms are shown in Fig. 3. For comparative analysis in this section, without loss of generality, set the parameters of mNCDF and MEP algorithm as symbol type $c = 3$, embedding dimension $m = 3$ and time delay $d = 1$. As shown in Fig. (3), mNCDF algorithms could reflect more details of bearing failure simulation series than the MEP algorithms. The MEP algorithm

maps too many bearing failure simulation series to the maximum number of symbols 3. Therefore, redundant amplitude information will be caused, and the original series periodicity will be destroyed. It should be noted that the symbol sequence obtained by the mNCDF algorithm not only appropriately retains the amplitude information but also protects the periodicity of the signal.

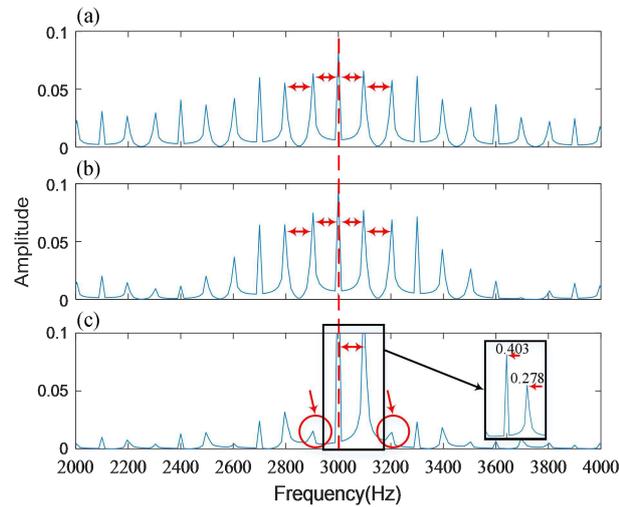


Fig. 4 The FFT spectrum: (a) Bearing simulation series. (b) mNCDF symbol series (c) MEP symbol series.

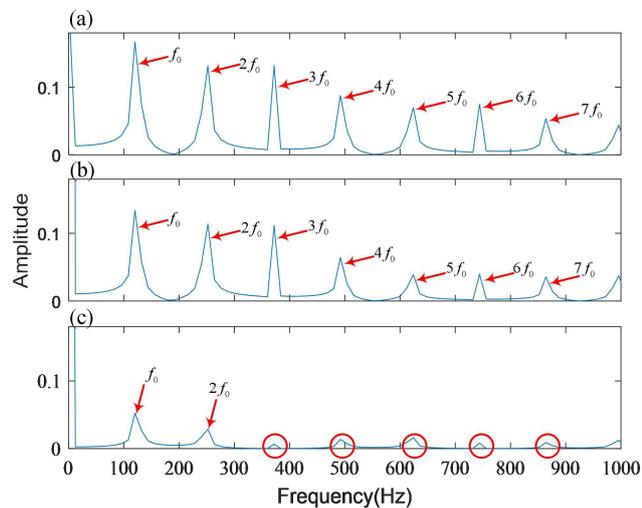


Fig. 5 The envelope spectrum: (a) Bearing simulation series. (b) mNCDF symbol series (c) MEP symbol series.

In order to further compare the information contained in the symbol series obtained by the two mapping algorithms, we obtained the spectrum and envelope spectrum of the symbol sequence through simulation are displayed in Fig. 4 and Fig. 5, respectively. From Fig. 4(b), the frequency spectrum of the symbol

sequence obtained by the mNCDF algorithm is almost the same as the frequency spectrum of the Bearing simulation series, and the natural frequency and its sidebands can be clearly observed. From Fig. 4(c), it can be observed that the amplitude of the natural frequency and the first right sideband of the MEP symbol sequence are enhanced, while the surrounding sidebands are severely weakened. This is caused by the redundancy and periodic destruction of amplitude information. It should be noted that the fault frequency is the sideband interval frequency. The envelope spectra in Figures 5(b) and 5(c) also verify the above analysis. The mNCDF symbol sequence retains the fault frequency of 124Hz, while the fault frequency in the envelope spectrum of the MEF symbol sequence is severely weakened. This phenomenon shows that the mNCDF mapping process can effectively retain the amplitude and fault information of the original signal and its performance is better than the MEF algorithm.

2.3 Comparison between SJE, DE, SE and PE

To study the performance of the proposed SJE method in describing the complexity of time series, four simulated signals are used for analysis. At the same time, the three methods of DE, PE and SE are used for comparative analysis. All these simulation signals have a length of 360s and a sampling frequency of 150 Hz. Design a 12s sliding window to divide the data with 75% overlapping steps, which means that each movement step is 3s.

In order to compare different entropy algorithms reasonably, the parameters of the four methods need to be set first. Without loss of generality, the embedding dimension $m = 3$ and the delay time $d = 1$ of the four methods. For the SE algorithm, the tolerance $r = 0.15\sigma$. At the same time, in order to ensure the same symbol mapping range, the category parameter $c = 3$ in the SJE and DE algorithms.

To compare the sensitivity of the SJE method to frequency varying signals, a chirp signal is applied in this paper. The frequency of the linear frequency modulated (FM) signal increases linearly from 0.1 Hz to 8.5 Hz in 360 s. The time waveform and the corresponding results based on chirp signal, in comparison with PE, DE and SE ones, are shown in Fig. 6. We observed that the SJE can not only detect frequency changes more accurately than PE, DE and SE, but also performs best. The SE curve tends to be stable after the 35th sliding window, which

indicates that SE cannot describe the frequency change with time. Although the PE and DE curves can detect the frequency change, the amplitude of PE curve fluctuates greatly at the beginning, and the growth rate of the two curves is smaller than that of SJE curve.

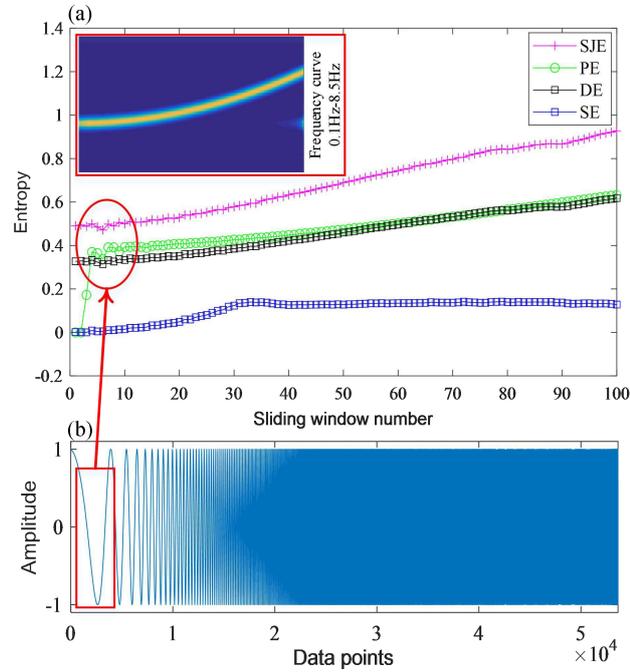


Fig. 6 (a) The entropy value of SJE, PE, DE and SE. (b) chirp signal with constant amplitude.

To study the sensitivity of the SJE method to frequency and amplitude varying signals, the amplitude modulated chirp signal is applied, depicted in Fig. 7(b). The signal, whose frequency is linearly increasing from 0.1 Hz to 8.5 Hz in 360 s, was created by using pure sinusoid wave modulation with a frequency of 0.1 Hz. It can be observed that after the 30th sliding window, the curve of SE does not show an upward trend and fluctuates greatly. This indicates that SE cannot show the frequency change, and its performance is unstable. Although PE value increases monotonously with the increase of frequency, compared with SJE and DE, which is insensitive to the amplitude change. Both SJE and DE curves show an upward trend, but the upward trend and periodic fluctuation of SJE curve are more obvious. The results show that SJE can accurately evaluate the frequency and amplitude change.

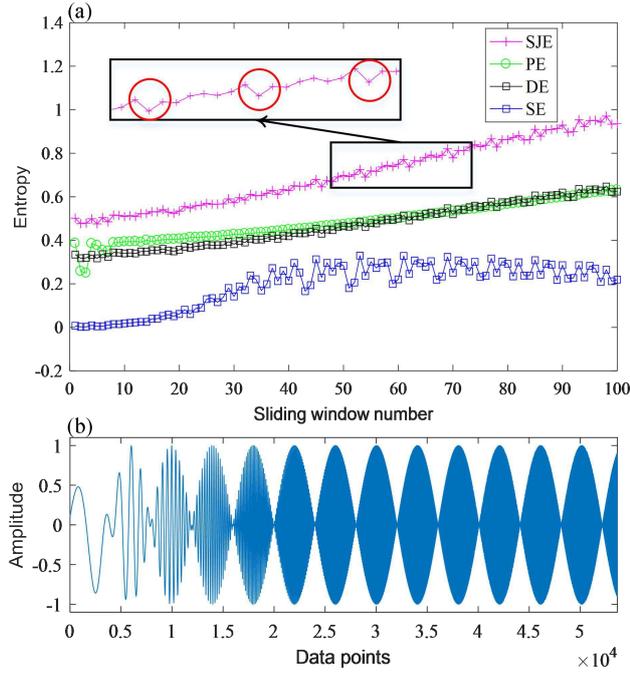


Fig. 7 (a) The entropy value of SJE, DE, PE and SE. (b) amplitude-modulated chirp signal.

To detect the sensitivity of the SJE method to the level of noise, the quasi-periodic signal with different level of noise power is created, depicted in Fig. 8(b). The signal is modulated by two sinusoidal signals with frequencies of 0.5 and 1 Hz. There is no noise in the first 24 seconds of the sequence. Then, the Gaussian white noise (WGN) is added to the signal every 12s, with the noise power increasing gradually. The simulation results are shown in Fig 8 (a). The entropy value of PE is constant from the 10th sliding window, which indicates that the method cannot detect the change of noise power. The SE entropy curve increases monotonously with the increase of noise power, but it fluctuates sharply from the 40-th sliding window, which indicates that the SE method is vulnerable to noise. It is worth noting that the curves of SJE and DE value increase monotonously and steadily with the increase of noise power, which indicates that both SJE and DE methods can detect noise power and maintain stable performance.

To investigate the recognition ability of SJE method for amplitude jump signal, a signal consisting of impulse and WGN is created, depicted in Fig. 9(b). The impulse signals are added every 80s, and their amplitudes are 20, 30 and 50 respectively. The comparison results are shown in Fig. 9 (a).

It can be observed that the PE curve is constant, which indicates that the method cannot detect the change of impulse amplitude. The SE method can detect the amplitude change, but the SE curve fluctuates violently, which indicates

that the performance of the method is unstable and the anti-noise ability is poor. SJE and DE methods can effectively detect the change of amplitude, and SJE method is more sensitive. The above results show that SJE method not only inherits the stability of DE method, but also has stronger amplitude sensing ability.

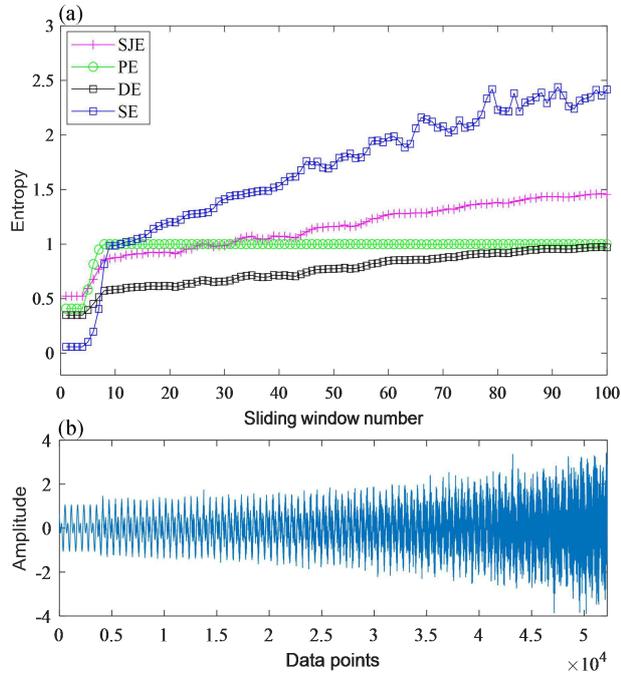


Fig. 8 (a) The entropy value of SJE, DE, PE and SE. (b) a quasi-periodic signal with noise.

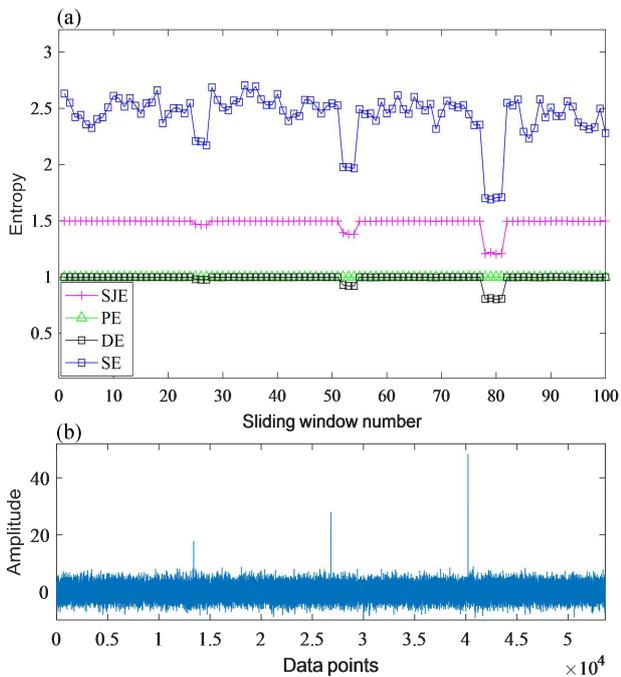


Fig. 9 (a) The entropy value of SJE, DE, PE and SE. (b) a WGN signal with impulse abrupt.

Table 1 CPU time of the SJE, DE, SE and PE methods.

Simulated signals	CPU time (s)			
	SJE	DE	SE	PE
FM	0.307	0.239	13.649	1.074
AM-FM	0.308	0.245	13.451	1.032
Quasi-periodic signal	0.285	0.234	15.103	2.329
Amplitude jump signal	0.307	0.254	15.658	1.042

To compare the complexity of the SJE, DE, SE, and PE methods, we counted their CPU time consumption. The CPU time in Table 1 is the average of 50 statistical results. It can be seen from table 1 that SJE and DE methods consume very little time, and their average time damages in the four simulation signals are all at the millisecond level. The SE method takes the longest time, which is more than 13s, while PE method also takes more than 1s. This indicates that SJE method not only inherits the high efficiency of DE method, but also has better performance of complexity estimation. The hardware system for all these emulation processes is a 2.80 GHz CPU (Intel (R) CoreTM i5-8400), a PC with 8GB RAM. The PC operating system is Windows 10 (64-bit), and Matlab software version is R2016a.

3 Refined generalized composite multiscale state joint entropy algorithm

3.1 Refined generalized composite multiscale analysis

The refined composite multi-scale analysis (RCMSA) effectively solves the problem of uncertain entropy in multi-scale analysis. However, in the coarse-grained process of RCMSA, the variance of entropy values increases with the increase of scale factor, which leads to the low stability and distinguishability of feature space. In order to overcome these shortcomings, a new refined generalized composite multiscale state joint entropy (RGCMSJE) method is proposed. The detailed description of RGCMSJE is given as follows:

- (1) For a given time series $\{x(i), i = 1, 2, \dots, N\}$, in order to obtain a stable feature space, the mean value μ_x and standard deviation σ_x of the time series are calculated.

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (13)$$

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)^2} \quad (14)$$

The k^{th} coarse-grained time series $y_{k,j}^{(s)} = \{y_{k,1}^{(s)}, y_{k,2}^{(s)}, \dots, y_{k,N/s}^{(s)}\}$ is defined as:

$$y_{k,j}^{(s)} = \frac{1}{S} \sum_{i=k+s(j-1)}^{k+s(j-1)+S-1} x_i, \quad 1 \leq j \leq N/s, 1 \leq k \leq S \quad (15)$$

(2) According to the coarse-grained time series $y_{k,j}^{(s)}$, the unified mapping is carried out. The $NCDF_x$ maps $y_{k,j}^{(s)}$ into $\varepsilon = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N\}$, from 0 to 1 as follows:

$$NCDF_x = \frac{1}{\sigma_x \sqrt{2\pi}} \int_{-\infty}^{y_{k,j}^{(s)}} e^{-\frac{(t-\mu_x)^2}{2\sigma_x^2}} dt \quad (16)$$

(3) For scale factor s , the RGCMSJE value is described as Shannon entropy of the second-order moment state joint model obtained after the time series is shifted.

$$RGCMSJE(y, m, c, d, \tau) = -\sum_{\pi=1}^{c^m} \bar{p}(\pi_{v_0 v_1 \dots v_{m-1}}) \ln(\bar{p}(\pi_{v_0 v_1 \dots v_{m-1}})) \quad (17)$$

Where $\bar{p}(\pi_{v_0 v_1 \dots v_{m-1}}) = \frac{1}{S} \sum_{k=1}^S p_k^s$ is the average of the probability of the dispersion pattern π in the coarse-grained sequence $y_{k,j}^{(s)} (1 \leq k \leq S)$, and m is the embedding dimension.

It is noteworthy that in reference [45], generalized multiscale processes are extended to second-order statistics by $y_{k,j}^{(s)} = \frac{1}{S} \sum_{b=(j-1)S+k}^{jS+k-1} (x_b - \bar{x}_b)^2$. This approach has been proved to be suitable for the sample entropy and permutation entropy which are sensitive to the relationship between adjacent amplitudes. However, it is found that this method is not suitable for SJE because a lot of amplitude information will deviate with the change of second moment. This leads to the loss of useful information to some extent [45]. In this paper, the RGCMSJE method is introduced to solve the above problems. RGCMSJE has been improved

in two aspects. First, the process of refined composite multi-scale analysis in step (1) can stably extract the changes of time series from multiple scales and avoid the appearance of undefined entropy. Second, in order to reduce the variance fluctuation of time series with large scale factor, step (2) uses the mapping method of unified standard deviation and mean value instead of step-by-step mapping, so that RGCMSJE has stronger fault feature extraction ability.

3.2 The parameter selection of refined generalized composite multi-scale state joint entropy

In the entropy algorithm, the selection of parameters is a very important problem. The appropriate parameters can make the feature information extracted by entropy more abundant. The traditional entropy parameter selection method is to traverse the parameter interval, and then observe the entropy curve to determine the parameter value. However, the above method is only qualitative analysis and cannot provide quantitative data guidance, and it needs experience as support. This not only reduces the efficiency of fault diagnosis, but also restricts the fault recognition rate. Recently, the average Euclidean Distance (AED) was proposed as an index of entropy parameter selection in reference [46], but this method did not consider the multi-scale factors and the stability of feature space. When multi-scale analysis is used, the scale factor is an important factor in the selection of entropy parameters. At the same time, we find that when the stability of feature space is poor, even if the AED value is large, the recognition rate of gearbox operation state is greatly reduced. Therefore, a new index for the selection of multi-scale entropy parameters, namely multi-scale average Euclidean Divergence (MAED), is proposed in this paper by combining multi-scale analysis, feature space stability and average Euclidean Distance. The detailed steps of MAED calculation are described as follows:

There are five parameters in the RGCMSJE calculation process, including sample data length N , scale factor s , embedding dimension m , category parameter c and time delay d . Without loss of generality, in order to simplify the calculation process, the sample data length N and delay time d are usually set to 2048 and 1, respectively. Suppose that the data set to be tested has k different classes, and each class has n samples with N length.

- (1) Initialize the parameters (c, m) in RGCMSJE. Set $s = 20$, $c \in [2, 8]$ and $m \in [2, 8]$, while c and m also need to satisfy the qualification $c^{m+1} \ll N$.
- (2) Calculate the average refined generalized composite multi-scale state joint entropy (ARX) and multi-scale standard deviations (MSD) of sample in i^{th} class and s^{th} scale as:

$$ARX_{i,s}(p) = \frac{1}{n} \sum_{p=1}^n RGCMSJE_{i,s}(p) \quad (18)$$

$$MSD_{i,s}(p) = \sqrt{\frac{\sum_{p=1}^n (RGCMSJE_{i,s}(p) - ARX)^2}{n-1}} \quad (19)$$

Where $1 \leq s \leq 20$.

- (3) The Euclidean distance of entropy space and standard deviation space between i^{th} and j^{th} classes are calculated respectively as follows:

$$AED_s(i, j) = \sqrt{\sum_{p=1}^n (ARX_{i,s}(p) - ARX_{j,s}(p))^2} \quad (20)$$

$$ASD_{i,s}(i, j) = \sqrt{\sum_{p=1}^n (MSD_{i,s}(p) - MSD_{j,s}(p))^2} \quad (21)$$

- (4) The MAED value is calculated as follows:

$$MAED_s = \sum_{i=1}^k \sum_{j=1, j \neq i}^k AED_s(i, j) / \sum_{i=1}^k \sum_{j=1, j \neq i}^k ASD_s(i, j) \quad (22)$$

- (5) Update the parameters c and m , and then repeat the above steps (1) - (4) to calculate the required MAED value.

The RGCMSJE method can extract the information of different health state signals of the gearbox from multiple scales. The parameter value is quantified by the MAED index between different health conditions of the gearbox obtained based on the RGCMSJE value. A larger MAED value, a greater distinguishability between different health states on multiple scales, and a better the stability between different samples. This means that RGCMSJE can extract more information from gearbox vibration signal, and has higher distinguishability.

Therefore, in the process of traversing c and m , we determine the optimum combination of c and m according to the larger MAED value.

4 Feature selection using robust spectral learning framework for unsupervised feature selection (RSFS)

High dimensional fault features are extracted by RGCMSJE method. However, some of these features are redundant, which will increase the time consumption of classification algorithm and reduce the classification accuracy. Therefore, in order to improve the accuracy of fault recognition, it is necessary to select some features from high-dimensional features. In this paper, RSFS method is proposed to select sensitive features.

Suppose that the dataset $X = \{x_1, x_2, \dots, x_N\} \in \mathbb{R}^{f \times N}$ has n samples, k classes, and each sample has f features. $Y = [y_1, y_2, \dots, y_k] = [y_{il}] \in \{0, 1\}^{N \times k}$ is the partitioned set of the dataset X .

The detailed calculation steps of RSFS algorithm can be summarized as follows:

- (1) A local kernel regression $p_{il}(\bullet)$ is constructed based on the data point x_i .

$$p_{il}(x_i) = \frac{\sum_{x_j \in G_i} K(x_i, x_j) y_{jl}}{\sum_{x_j \in G_i} K(x_i, x_j)} \quad (23)$$

where G_i is the neighborhood of x_i , $K(\bullet)$ is the kernel function. The matrix $S = [s_{ij}] \in \mathbb{R}^{N \times N}$ is given as,

$$s_{ij} = \begin{cases} \frac{K(x_i, x_j)}{\sum_{x_j \in G_i} K(x_i, x_j)} & x_j \in G_i \\ 0 & x_j \notin G_i \end{cases} \quad (24)$$

- (2) Calculate the degree matrix B of $(S + S^T)$ and matrix $M = (B - S - S^T)$.

- (3) Initialize $F = Y(Y^T Y)^{-\frac{1}{2}} \in \mathbb{R}^{N \times k}$ and a sparse matrix $Z \in \mathbb{R}^{N \times k}$. Set $D \in \mathbb{R}^{f \times f}$ as an identity matrix. The spectral regression coefficients W is given as,

$$W = (XX^T + \frac{\beta}{\alpha}D)^{-1}X(F - Z) \quad (25)$$

$$Z_{ij} = \begin{cases} 0 & \text{if } |E_{ij}| \leq \frac{\gamma}{2\alpha} \\ (1 - \frac{\gamma}{2\alpha|E_{ij}|})E_{ij}, & \text{otherwise} \end{cases} \quad (26)$$

where α, β and γ are input parameters. $E = F - X^TW$.

(4) The update rules of F are as follows:

$$F_{ij} \leftarrow F_{ij} \sqrt{\frac{[M^-F + vF + \alpha A^+]_{ij}}{[M^+F + \alpha F + vFF^T F + \alpha A^-]_{ij}}} \quad (27)$$

where $A = X^TW + Z$, v is input parameters.

(5) Update the diagonal matrix $D_{ii} = \frac{1}{2\|w_i\|}$, all features are sorted according to

$\|w_i\|$, and then the feature with the top rank is selected.

The RSFS is an unsupervised feature selection algorithm based on spectral regression and sparse graph embedding. Through research, we find that the traditional algorithm based on spectrum feature selection has two major problems: 1) in the process of constructing Laplacian, noise and uncorrelated features may adversely affect the estimated Laplacian; 2) in the process of approximating discrete tags to continuous embedding, noise is inevitably introduced into the estimated clustering tags. The RSFS method uses robust local learning method to reduce the adverse effects of noise and uncorrelated features on Laplacian operator, and robust spectral regression method to deal with the noise on clustering labels.

5 Proposed fault diagnosis method

5.1 Extreme learning machine

After extracting feature set, an intelligent classifier is used to identify the fault type of gearbox. Extreme learning machine (ELM) is a new strong classifier based on single-hidden layer feed forward neural network (SLFN), which can overcome some disadvantages of the traditional SLFN, such as of low learning efficiency,

over fitting and easy to fall into local minimum. The input weights and biases of the algorithm are randomly generated, and the calculation process remains unchanged, so it has fast learning speed and less manual intervention [47]. The main procedure of ELM is briefly explained as follows.

(1) Given an training data set $\{x_i, y_i\}, (i=1, 2, \dots, M)$, where x_i is the network input and y_i is the network output. Determine the activation function $g(x)$ and the number of hidden layer nodes L , and randomly generate the input weight ω_i and the input bias b_i .

(2) The hidden layer output matrix H is calculated.

$$H = \begin{bmatrix} g(\omega_1, b_1, x_1) & \cdots & g(\omega_L, b_L, x_1) \\ \vdots & \cdots & \vdots \\ g(\omega_1, b_1, x_Q) & \cdots & g(\omega_L, b_L, x_Q) \end{bmatrix}_{Q \times L} \quad (28)$$

where Q is the number of samples.

(3) The output weight β is calculated.

$$H\beta = T \rightarrow \beta = H^+T = H^T \left(\frac{I}{C} + HH^T \right)^{-1} T \quad (29)$$

Where H^+ is the Moore-Penrose generalized inverse matrix of H , C represents the penalty factor, I is the identity matrix, and T is the expected output matrix.

(4) The output equation of ELM is defined as,

$$Y_i = \sum_{i=1}^K \beta_i g(\omega_i x_i + b_i), i=1, 2, 3, \dots, L \quad (30)$$

The structure of ELM is illustrated in Fig. 10.

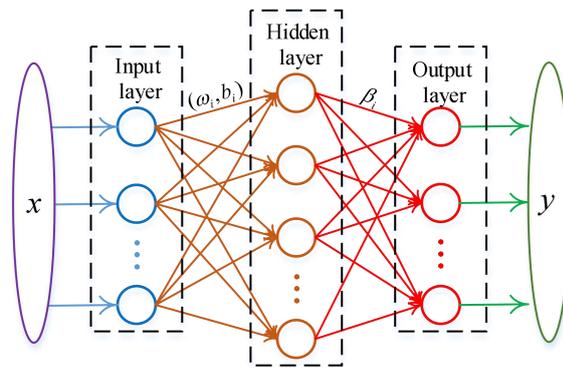


Fig. 10 The structure of ELM.

5.2 The proposed fault diagnosis method

Based on the superiorities of RGCMSJE, RSFS, and ELM, a fault diagnosis method for gearbox of wind turbine is presented in this paper. Figure 11 displays the flowchart of the proposed method and it can be summarized as follows:

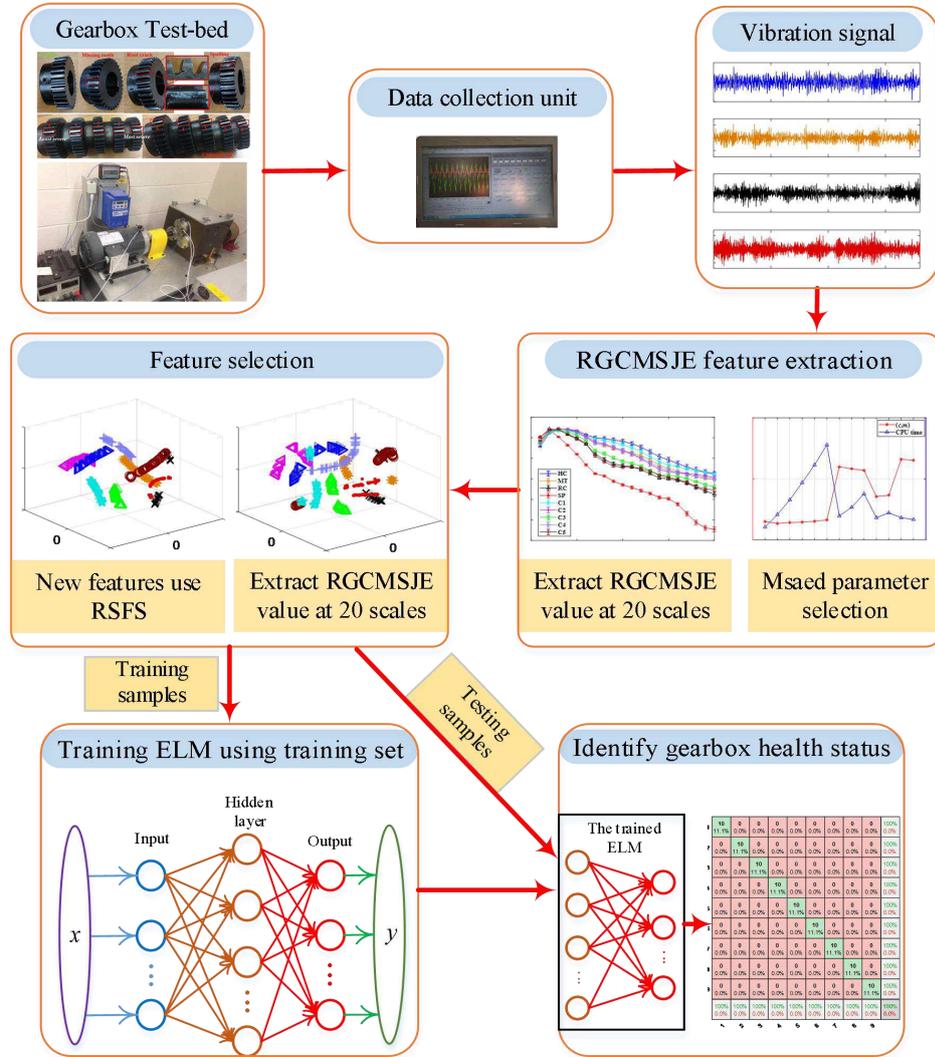


Fig. 11 Flowchart of the proposed method.

- (1) Collect and store the vibration data of different healthy types of wind turbine gearbox;
- (2) Apply RGCMSJE method to extract fault features through quantifying the complexity of gearboxes vibration signals under different health conditions. The parameter combination of RGCMSJE method is determined by MAED method. In this paper, we set the time delay $d=1$ and scale factor $s=20$. And on this basis, find the best parameter combination of embedding dimension m and scale factor c .

- (3) RSFS method is used to select the most sensitive feature to construct the state feature vector;
- (4) The selected state feature vectors are input into the multi fault classifier ELM to train and identify different operation conditions. In order to eliminate the influence of data randomness on the recognition accuracy of ELM method, the 10 times cross validation method is applied to the recognition process.

6 Experimental verification

In this section, the proposed intelligent diagnosis method is applied to the analysis of a laboratory gearbox dataset and a wind turbine gearbox dataset to verify its effectiveness.

6.1 Case 1: gearbox vibration data from artificial fault test

6.1.1 *Experimental system description and input dataset introduction.*

In this case, the experimental datasets of gearbox vibration signal were collected from the University of Connecticut [48-49]. The experimental platform is shown in Fig. 12, which contains a benchmark two-stage gearbox. The gearbox fault occurs on the pinion of the first stage input shaft, which contains a 32-tooth pinion and an 80-tooth gear. There is no fault on the second stage of the gearbox, which consists of a 48-tooth pinion and 64-tooth gear. The speed of the gearbox can be adjusted freely by the motor. The vibration signal of gearbox is recorded by dSPACE system, and the sampling frequency of accelerometer is 20kHz. Nine different gear states are introduced into the input shaft pinion, including five fault types and five wear degrees. For each gear state, 104 vibration signals are recorded, and each signal contains 3600 data points. In order to eliminate the non-stationary issue caused by speed varying, the experimental platform converts the vibration signal from time domain to angular domain and averaged in angular domain. More details of the dataset can be obtained in Reference [49].

In this test, all nine types of faulty gears were applied to the test equipment. The vibration signals for the experiment are divided into healthy condition (HC), missing tooth (MT), root crack (RC), spalling (SP), and chipping tip with five different levels of severity (C1~C5). And they are successively denoted as: (a)

healthy condition (HC), (b) missing tooth (MT), (c) root crack (RC), (d) spalling (SP), (e) chipping tip with degree level 1 of severity (C1), (f) chipping tip with degree level 2 of severity (C2), (g) chipping tip with degree level 3 of severity (C3), (h) chipping tip with degree level 4 of severity (C4), (i) chipping tip with degree level 5 of severity (C5). After the above process, 9 rolling bearing states (a) ~ (i) are obtained, and 2048*100 non-overlapping samples are taken from each state to form the input data set. Table 2 shows the total experimental data sets and the input data sets.



Fig. 12 The gearbox experimental platform.

Table 2 The numbers of the experimental data sets and the input data sets.

Fault class	Class label	Experimental data sets	Input data sets
		(Data points* Samples)	
HC	1	3600*104	2048*100
MT	2	3600*104	2048*100
RC	3	3600*104	2048*100
SP	4	3600*104	2048*100
C1	5	3600*104	2048*100
C2	6	3600*104	2048*100
C3	7	3600*104	2048*100
C4	8	3600*104	2048*100
C5	9	3600*104	2048*100
Total	9	3600*936	2048*900

6.1.2 Parameter selection and feature sets determination.

Figure 13 shows the time domain and frequency spectrum of gearbox vibration signal under nine different health conditions. From Fig. 13, we can find that when the gear fails, the amplitude in the high frequency band will increase which is

caused by the resonance of the defect. But the frequency of high frequency resonance caused by nine kinds of health conditions is the same in the spectrum, and the change of amplitude is irregular. Therefore, it is difficult to accurately identify different fault types and severity just by time domain and frequency domain analysis, because vibration signals usually have nonlinear and non-stationary characteristics. In addition, relying on time-frequency analysis for fault diagnosis requires a lot of expert experience and artificial identification. This not only reduces the efficiency, but also has a big error. Therefore, it is necessary to develop an effective feature extraction method to realize automatic and efficient fault diagnosis.

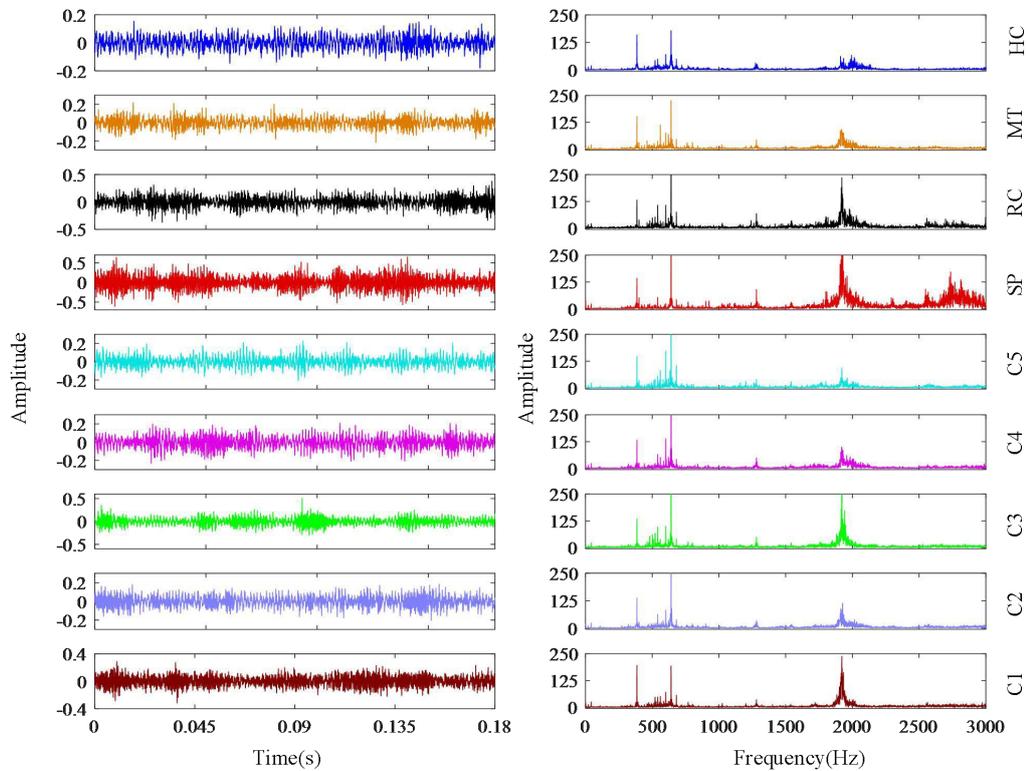


Fig. 13 The gearbox experimental platform.

In order to extract more fault information, RGCMSJE method is used to extract the feature of gearbox vibration signal. According to the method flow of gearbox health status identification proposed in section 5.2, firstly, MAED method is used to select the appropriate parameter combination. At the same time, in order to improve the efficiency of finding parameter combination, 10% of the data is used for parameter finding. Under the condition of $c^{m+1} \ll N$, 13 pairs of (c, m) parameter combinations are selected to calculate the MAED value. The results of MAED curve and corresponding CPU time loss are shown in Fig. 14. It can be

seen from the Fig. 14 that the value of MAED increases with the increase of c value when m value is fixed. At the same time, when c value is fixed, CPU time loss increases rapidly with the increase of m value. This provides a good reference for the selection of parameter m . It should be noted that when the combination of parameters is (5, 2), the maximum value of MAED curve is 25.0945. This shows that the RGCMSJE corresponding to the parameter combination (5, 2) can not only extract more fault information stably, but also has higher efficiency. Therefore, considering the time loss and MAED value, we select the category parameter $c = 5$ and the embedding dimension $m = 2$.

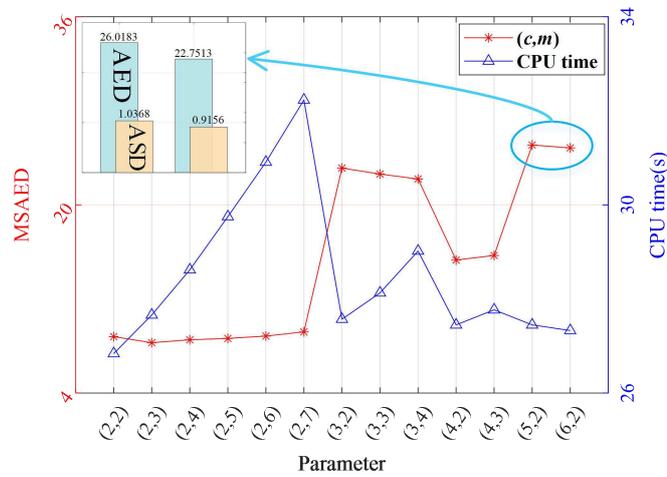


Fig. 14 Multi-scale average Euclidean Distance using different parameter combinations (c, m) .

Table 3 The comparison results of different feature extraction methods.

Methods	Parameter	MSAED	CPU time(s)
RGCMSJE	$c = 5, m = 2$	25.0945	27.77
GRCMSJE	$c = 5, m = 2$	9.9686	52.609
RCMSJE	$c = 4, m = 2$	11.7236	27.989
GRCMDE	$c = 5, m = 2$	10.0981	48.938
GRCMPE	$m = 3$	3.31	93.256
GRCMSE	$m = 2, r = 0.15$	6.3455	78.841

In order to comprehensively compare the advantages of the feature extraction methods proposed in this paper, we introduce five feature extraction methods for comparison. As shown in Table 3, GRCMSJE is the generalized refined composite multi-scale SJE entropy based on the second-order moment. At the same time, the parameters of all methods are the optimal values selected by MAED method. It can be seen from Table 3 that the MAED value of RGCMSJE is the largest, and the CPU time consumption is the least, which indicates that this

method can extract more useful information from time series, and the extraction efficiency is higher. A graph of all six feature extraction methods is shown in Fig. 15. It can be found from Fig. 15 that when the scale factor s is greater than 5, the dispersion of RGCMSJE curve is the best, and there is no large fluctuation. This shows that the multi-scale features extracted by RGCMSJE method are the most distinguishable and stable.

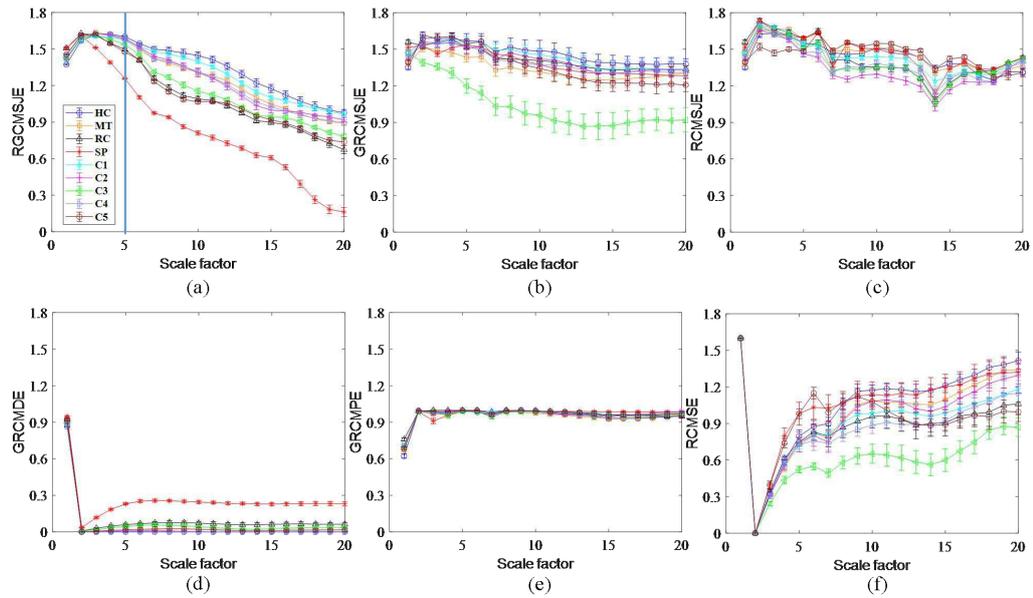


Fig. 15 The entropy curve of gearbox vibration signal under the optimal parameters: (a) RGCMSJE; (b) RGCMSJE; (c) RCMSJE; (d) RCMDE; (e) RCMPE; (f) RCMSE.

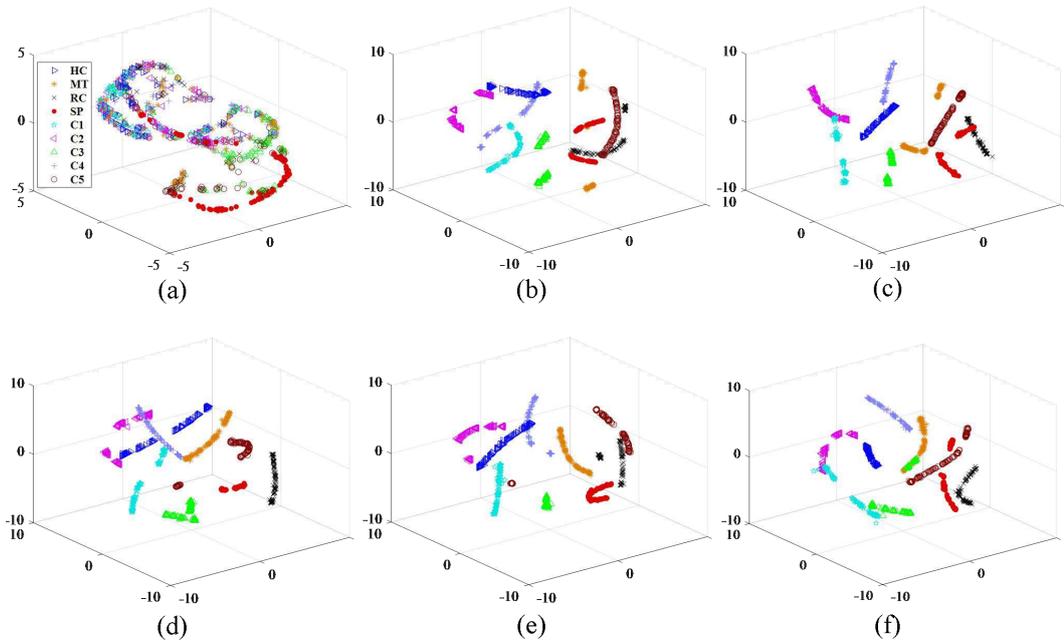


Fig. 16 The Feature visualization via t-SNE: (a) Raw signal; (b) RGCMSJE; (c) RGCMSJE+RSFS; (d) RGCMSJE+LS; (e) RGCMSJE+FS; (f) RGCMSJE+mRMR.

6.1.3 Dimension reduction and comparative analysis of feature space.

As can be seen from the feature curve in Fig. 15 (a), not all 20-dimensional features can highly distinguish the running state of the gearbox. This shows that there is redundant information in the 20-dimensional features, which is not conducive to the recognition of intelligent algorithm and reduces the recognition efficiency. In order to further improve the classification accuracy and eliminate redundant information, it is necessary to select the feature that contains the most gear information. Therefore, this paper uses the spectral learning framework for unsupervised feature selection (RSFS) of unsupervised feature selection to select the most important features. RSFS method uses local robust learning to rank high-dimensional features. According to the importance of features, RGCMSJE features are sorted as follows:

$$\begin{aligned} R_6 > R_{20} > R_1 > R_{10} > R_9 > R_{11} > R_{17} > R_3 > R_2 > R_{14} \\ > R_{16} > R_{13} > R_{18} > R_{12} > R_{15} > R_4 > R_5 > R_{19} > R_8 > R_7 \end{aligned} \quad (31)$$

Where R_s represents the feature selected by RSFS and s is the scale factor. At the same time, in order to comprehensively analyze the performance of RSFS method, we introduce Laplacian score (LS), Fisher score (FS) and max correlation min redundancy (RMR) methods for comparative study. According to the research of literature [50], the feature dimension retained after dimension reduction is set to be $n-1$, where n is the number of healthy states of gearbox and $n=9$ in this study. That is, the first eight features of the four dimensionality reduction methods are selected as sensitive features to compose a new feature set. Here an explanation about feature space is given to describe four feature selection methods more clearly. In this experiment, the feature space extracted by RGCMSJE is a $900 * 20$ matrix, where the number of data samples is 900 and the feature dimension is 20. In addition, this experiment has 9 gearbox health status, so the category label is 1~9. In other words, the four feature selection algorithms select eight features from the feature space of $900 * 20$, and the final dimension reduction feature space is $900 * 8$. A detailed description of the feature space is shown in Table 5. The performance comparison results of the four feature selection algorithms are shown in Fig. 16 and Table 4.

The visualization results of the original signal and all feature distribution is given in Fig. 16. In order to intuitively analyze the feature space, we apply t-SNE algorithm to project the feature into the three-dimensional space. It can be seen from Fig. 16 (a) that the original signals of 9 health state are overlapped with each other, which indicates that it is very difficult to classify by using the original signals directly. At the same time, Figure 16 shows that the RGCMSJE algorithm clusters the features of the same category, and the features of different categories are separated from each other, which is convenient for the classification of different health conditions. It is worth noting that the features selected by RSFS method have stronger clustering. Although the other three feature selection algorithms can separate the nine states well, the clustering of the same category is not strong enough. This shows that RSFS algorithm can extract features with more health status information.

Table 4 The numbers of the experimental data sets and the input data sets.

Fault class	Class label	RGCMSJE feature	Dimension reduction
		space (Samples * Feature dimension)	feature space
HC	1	100*20	100*8
MT	2	100*20	100*8
RC	3	100*20	100*8
SP	4	100*20	100*8
C1	5	100*20	100*8
C2	6	100*20	100*8
C3	7	100*20	100*8
C4	8	100*20	100*8
C5	9	100*20	100*8
Total	9	900*20	900*8

Table 5 The comparison results of different dimensionality reduction algorithms.

Method	Feature selection performance			CPU time(s)
	between-class scatter (S_b)	within-class scatter (S_w)	S_b / S_w	
RSFS	4.1885	0.1006	41.625	0.248
LS	2.5993	0.09	28.875	0.162
FS	4.4935	0.1428	31.459	0.108
mRMR	6.1143	0.1624	37.648	42.83

In order to quantify the feature selection performance of the above four methods, the between-class scatter, within-class scatter and CPU running time are introduced for comprehensive analysis, as shown in Table 5. We take the ratio of between-class scatter and within-class scatter as the performance index of feature selection. The larger the index is, the more concentrated the similar features are and the more distant the heterogeneous features are. From Table 4, compared with LS, FS and mRMR, RSFS has the largest feature selection performance index, and the CPU time loss is only twice that of LS and FS algorithm, which proves the superiority of RSFS algorithm in feature selection.

6.1.4 Fault diagnosis results analysis.

Finally, when the appropriate features are extracted, ELM classifier is used to classify the health status of wind turbine gearbox. It is worth noting that before ELM classification, we need to set the number of hidden neurons and select the appropriate activation function. The detailed results are shown in Figure 16, where the mean and standard deviation curves are the results of 20 calculations. Figure 16 (a) shows the effect of different activation functions on ELM classification accuracy. It can be seen that sin function achieves the highest classification accuracy when the number of hidden layers remains unchanged, and the stability of ELM classification is gradually enhanced with the increase of the number of hidden layers. As can be seen from Figure 16 (b), when the number of hidden layers is set to 70, the classification accuracy is the highest, the stability is the best, and the CPU time loss is moderate. Therefore, we set sin function as the activation function of ELM, and set the number of hidden layers to 70.

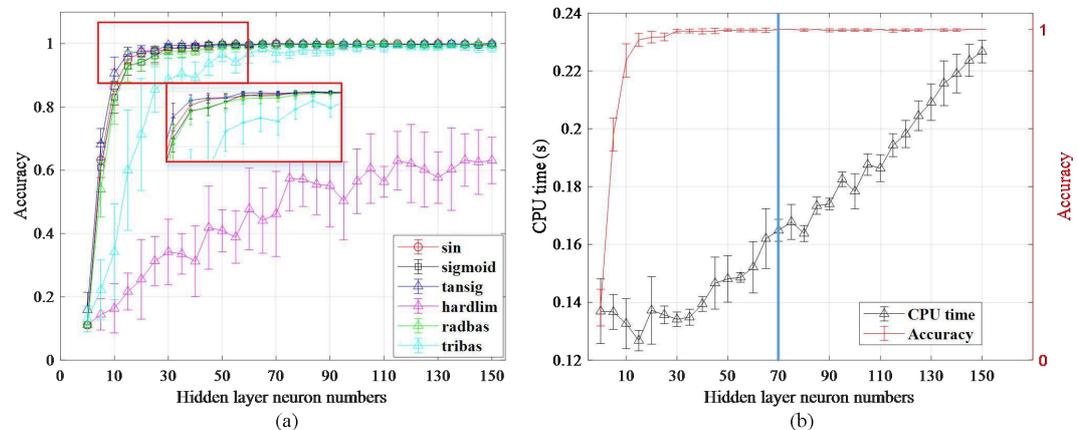


Fig. 16 ELM parameters: (a) selection of activation functions, and (b) selection of hidden layer numbers.

Output Class	1	100 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	2	0 0.0%	100 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	3	0 0.0%	0 0.0%	100 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	4	0 0.0%	0 0.0%	0 0.0%	100 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100 11.1%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99 11.0%	0 0.0%	0 0.0%	100% 0.0%
	7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100 11.1%	0 0.0%	0 0.0%	100% 0.0%
	8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	0 0.0%	100 11.1%	99.0% 1.0%
	9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100 11.1%	100% 0.0%
		100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	99.0% 1.0%	100% 0.0%	100% 0.0%	100% 0.0%	99.9% 0.1%
		1	2	3	4	5	6	7	8	9
		Target Class								

Fig. 17 ELM Confusion Matrix (CM) for RGCMSJE+RSFS

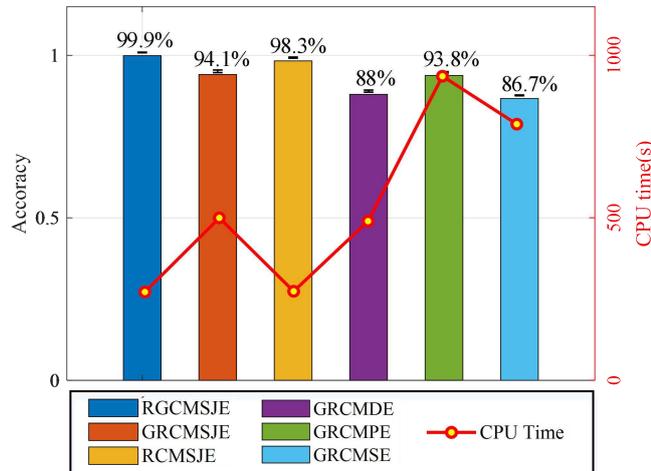


Fig. 18 The performance comparison results of the six feature extraction algorithms.

In order to improve the reliability and accuracy of the classification results, the 10-fold cross validation algorithm is introduced in the training and testing process of intelligent classification algorithms in this section. The confusion matrix (CM) obtained by ELM intelligent classifier is shown in Fig.17. Figure 17 is that result that is repeated most often in the 20 experiments. It can be seen from Fig. 16 that the total recognition accuracy of the fault diagnosis method proposed in this paper reaches 99.9%, and only the recognition rate of No. 8 fault is less than 100%. This indicates that the method can effectively identify different fault types and severity of gearbox, and has satisfactory classification accuracy as the whole. Although C4

fault signal is very similar to C2 fault signal, only one sample is misclassified in this method. At the same time, in order to compare the advantages of the RGCMSJE feature extraction method proposed in this paper, we compare the classification performance of the features extracted by RGCMSJE, GRCMSJE, RCMSJE, GRCMDE, GRCMPE and GRCMSE. When all the intelligent classifiers are ELM, the performance comparison results of the six feature extraction algorithms are shown in Fig 18. The following conclusion can be drawn from Figure 18. First, the RGCMSJE method has higher accuracy than other entropy methods combined with generalized refined multi-scale analysis (GRCMSJE, RCMSJE, GRCMDE, GRCMPE and GRCMSE). This is mainly because the state joint analysis can extract the current and subsequent state characteristics of the fault from the original signal; however, other entropy analysis only considers the current fault information in the signal. On the other hand, compared with other methods, the proposed RGCMSJE method achieves the highest test classification accuracy with less CPU time loss. The CPU time here refers to the total time including RGCMSJE feature extraction, RSFS feature selection and ELM fault classification. This further proves the advantages of RGCMSJE in fault feature extraction, which can reduce time loss and provide more reliable feature information.

In order to verify the superiority of ELM in classification performance, four different classifiers, softmax regression (SR), support vector machine (SVM), random forest (RF) and k-nearest neighbor (KNN), are studied in this paper. And four common evaluation indices, containing Accuracy (ACC), F-score (F), Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI) are used to verify the performance of different models. The ACC is an overall metric to measure the accuracy of model classification. F-score is a comprehensive metric to evaluate the overall performance of all classification results. ARI is the quality of correctly clustered samples in the same class. NMI is the estimated clustering quality relative to a given data class. The equation and function description of the above four indices are given in Table 6, where tp , tn , fp and fn refer to true

positives, true negatives, false positives, respectively. $MIP = \frac{\sum_{i=1}^k tp_i}{\sum_{i=1}^k (tp_i + fp_i)}$ and

$MIR = \frac{\sum_{i=1}^k tp_i}{\sum_{i=1}^k (tp_i + fn_i)}$, where k is the number of classes to be classified. Suppose

Φ and Ω are the sets of actual labels and classified labels respectively; n_{11} is the sample pairs of overlapping labels in Φ and Ω ; n_{00} is the sample pairs of non-overlapping labels in Φ and Ω ; C_n^2 is the total number of sample pairs; $p(\bullet)$ is the probability function and $p(\varphi, \omega)$ is the joint probability function of Φ and Ω . The larger the value of the above evaluation index, the stronger the comprehensive classification ability of the classifier.

Table 6 The performance evaluation metrics.

Metric	Equation	Range
ACC	$ACC = \frac{tp + tn}{tp + fp + tn + fn}$	[0,1]
F	$F = 2 \frac{MIP \times MIR}{MIP + MIR}$	[0,1]
ARI	$ARI = \frac{n_{11} + n_{00}}{C_n^2}$	[0,1]
NMI	$NMI = \frac{\sum_{\varphi \in \Phi} \sum_{\omega \in \Omega} p(\varphi, \omega) \log(p(\varphi, \omega) / p(\varphi)p(\omega))}{\sqrt{\left(\sum_{\varphi \in \Phi} p(\varphi) \log(p(\varphi)) \right) \left(\sum_{\omega \in \Omega} p(\omega) \log(p(\omega)) \right)}}$	[0,1]

Table 7 Parameters of four classifiers.

Method	Parameters	Value
ELM	Kernel function	Sigmoid
	Number of hidden layers	200
SR	Lambda	0.000,01
	Max iterations	200
SVM	Thera	310
	Kernel function	RBF
	Penalty factor	1
RF	Kernel parameter	10
	The estimation mode	Out-of-bag
KNN	Number of trees	100
	Number of nearest neighbors	1
	Distance type	Euclidean distance

Table 7 lists the parameter settings of the five classifiers. The feature sets of the five classification algorithms are obtained by RGCMSJE method and RSFS

method. At the same time, in order to reduce randomness, each method is tested with a 10-fold cross-validation method. The final verification results are shown in Fig. 19 and Table 8. From the radar diagram in Fig. 19, it can be found that the curve of RGCMSJE-RSFS-ELM model proposed in this paper is farthest from the center. At the same time, Table 8 shows that ELM classifier training and testing consumes the shortest CPU time. Based on the above advantages of ELM classifier, this paper selects ELM as the intelligent classifier of the model. The results show that the proposed model can not only obtain the best diagnosis results, but also has good stability compared with other models (RGCMSJE-RSFS-SR, RGCMSJE-RSFS-SVM, RGCMSJE-RSFS-RF and RGCMSJE-RSFS-KNN).

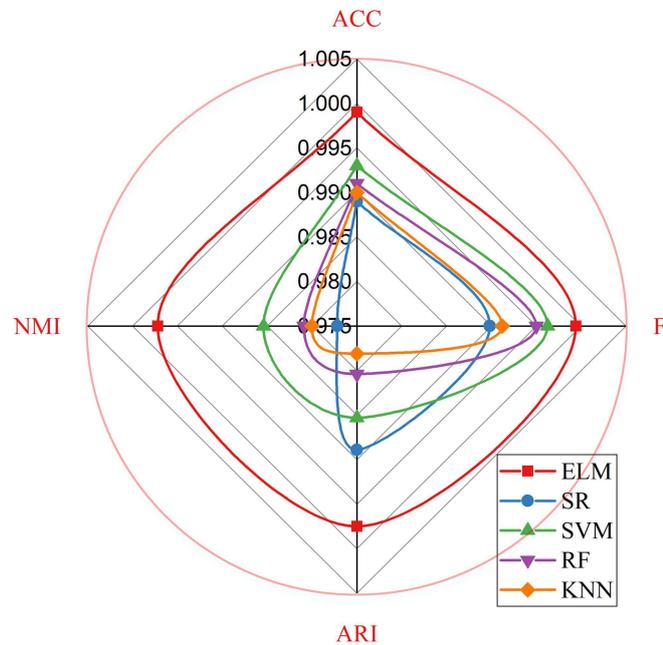


Fig. 19 The radar chart on evaluation results of different classifiers.

Table 8 The results of classifier performance evaluation are compared.

Mode	Result evaluation				
	ACC	F	ARI	NMI	CPU time
ELM	99.9%	0.99938	0.9975	0.9972	0.1624
SR	98.9%	0.98997	0.9889	0.9772	3.2633
SVM	99.3%	0.99624	0.9853	0.9854	17.031
RF	99.1%	0.99498	0.9804	0.9809	7.0882
KNN	99.0%	0.99118	0.9781	0.9800	2.2723

6.2 Case 2: gearbox vibration data from wind turbine

In what follows, in order to verify the good fault identification ability and generalization ability of the proposed method, we introduce the vibration data set of simulating wind turbine gearbox fault under different working conditions (WTGF) [51]. The effectiveness and superiority of the method are verified by analyzing the data under different working conditions.

6.2.1 Experimental system description and input dataset

introduction.

The WTGF data set was created by researchers engaged in fault diagnosis of wind turbine gearbox. The data set consists of two gearbox operating states: Healthy and Broken Tooth. And each operation state contains 10 load modes, representing 10 working conditions respectively. Each working condition contains four attributes, which represent the synchronous data collected by four sensors installed in four directions of the gearbox. The data set was recorded at a frequency of 30 Hz with a percentage change in load from "0" to "90". The data under healthy operation state includes 1015808*4 data points, and the data under broken tooth operation state includes 1005311*4 data points, and their details are listed in Table 9.

Table 9 Description of different working conditions of gearbox data set.

Fault class			
Healthy		Broken tooth	
Condition label	Data sets	Condition label	Data sets
h30hz0	88832*4	b30hz0	88320*4
h30hz10	92928*4	b30hz10	111616*4
h30hz20	108544*4	b30hz20	114432*4
h30hz30	106240*4	b30hz30	89856*4
h30hz40	100608*4	b30hz40	94464*4
h30hz50	110848*4	b30hz50	94208*4
h30hz60	99840*4	b30hz60	95488*4
h30hz70	101377*4	b30hz70	100865*4
h30hz80	99841*4	b30hz80	110336*4
h30hz90	106752*4	b30hz90	105728*4

In this section, we select three working conditions of healthy and broken teeth as the research objects, and their condition labels are: h30hz0, h30hz20, h30hz90, b30hz0, b30hz20 and b30hz90, respectively. Similar to experiment 1, 102400 points of sensor 1 and sensor 2 are divided into 50 data samples (each data sample contains 2048 points). And the 10-fold cross validation algorithm is also applied to the training and testing process of feature set.

6.2.2 Parameter selection and feature sets determination

The time and frequency domain waveforms of the six vibration signals are shown in Fig. 20. Seen from Fig. 20, it is easy to distinguish the healthy state and the broken tooth state. The time domain amplitude of the broken tooth fault is obviously smaller, and the obvious low frequency peak appears in the frequency domain diagram. However, the three working conditions of the two gear states are very similar, which means that it is difficult to identify different working conditions of the gearbox by directly observing the time and frequency domain waveforms. Therefore, we propose to use RGCMSJE method to extract the vibration signal features of gearbox.

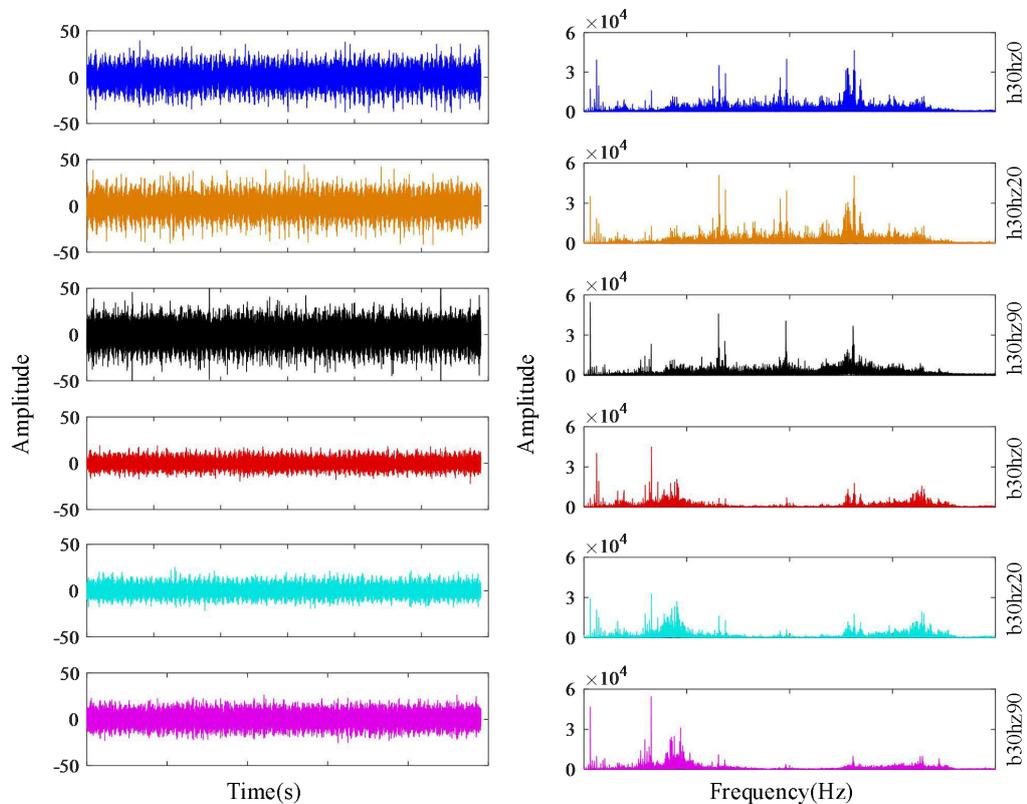


Fig. 20 The gearbox experimental platform.

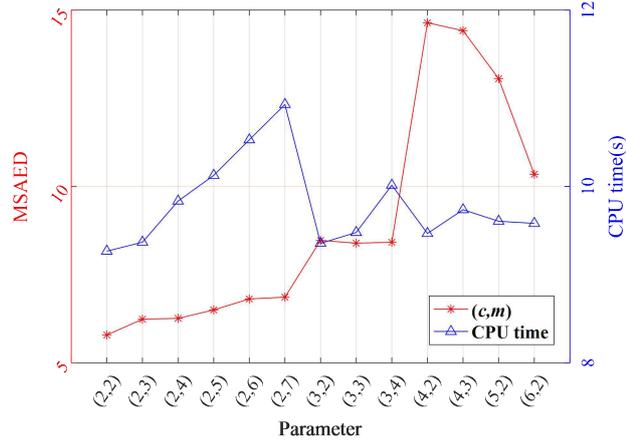


Fig. 21 Multi-scale average Euclidean Distance using different parameter combinations (c, m) .

Table 10 The comparison results of different feature extraction methods.

Methods	Parameter	MSAED	CPU time(s)
RGCMSJE	$c = 4, m = 2$	14.6402	9.849
GRCMSJE	$c = 2, m = 7$	5.9415	19.571
RCMSJE	$c = 2, m = 7$	6.8231	12.277
GRCMDE	$c = 2, m = 7$	5.0401	19.135
GRCMPE	$m = 4$	3.6734	43.479
GRCMSE	$m = 2, r = 0.15$	3.5557	26.721

Similar to experiment 1, in order to extract feature information efficiently, this paper uses RGCMSJE method to analyze the recorded experimental data to verify the effectiveness and generalization ability of this method. Firstly, MAED method is used to select the appropriate parameter combination, and then RGCMSJE method is applied to extract the multi-scale feature vector set with dimension of 20 based on the optimal parameter combination. The constraints in this section are the same as those in Experiment 1, and the MSAED curve and the corresponding CPU time loss are shown in Fig. 21. Figure 21 shows that the maximum value of MAED curve is 14.6402, and the parameter combination is (4,2). This shows that for the WTGF data set, the RGCMSJE algorithm can extract more abundant fault information efficiently when the parameter combination is $c = 4$ and $m = 2$. Therefore, considering CPU time loss and MAED, we choose category parameter $c = 4$ and embedding dimension $m = 2$. Subsequently, RGCMSJE is compared with other five feature extraction algorithms, and the comparison results are shown in Table 10 and Fig. 22. It can be seen from table 9 that the experimental results are the same as experiment 1. The MAED value of RGCMSJE is the

largest and the CPU time consumption is the smallest, which indicates that the feature extraction method has good generalization ability. The 20 dimensional characteristic curve in Fig. 22 also proves this point. The RGCMSJE curve has the largest discrimination and the smallest standard deviation.

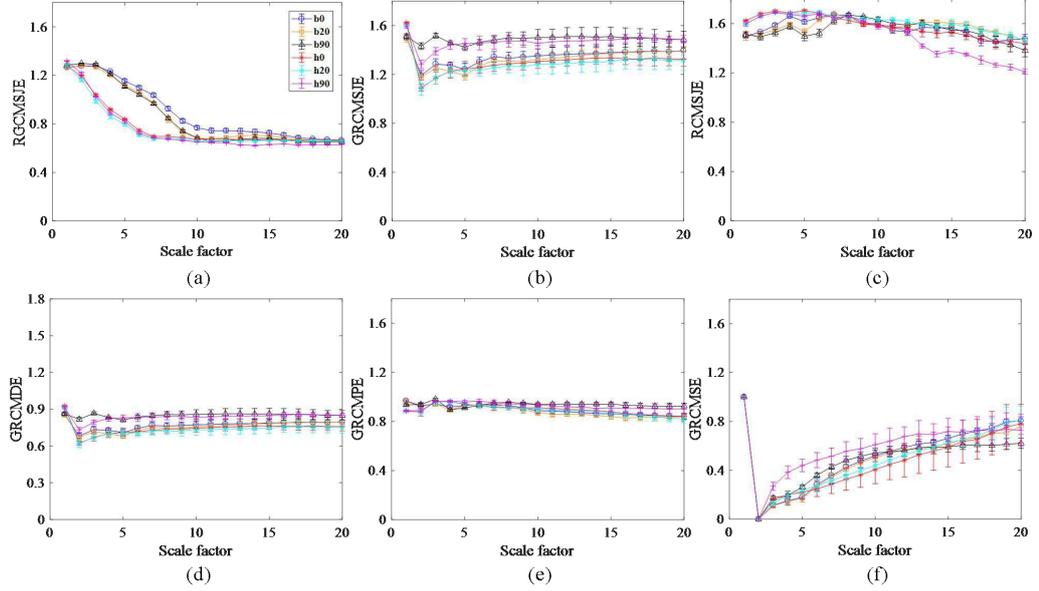


Fig. 22 The entropy curve of gearbox vibration signal under the optimal parameters: (a) RGCMSJE; (b) RGCMSJE; (c) RCMSJE; (d) RCMDE; (e) RCMPE; (f) RCMSE.

6.2.3 Dimension reduction and comparative analysis of feature space

Similar to experiment 1, in order to improve the efficiency of fault recognition and eliminate the redundant information in feature space, we use RSFS algorithm to reduce the dimension of feature space. Firstly, according to RSFS algorithm, the 20 dimensional features of feature space are sorted as follows:

$$\begin{aligned}
 R_9 > R_4 > R_1 > R_6 > R_2 > R_7 > R_{19} > R_{20} > R_{18} > R_{17} \\
 > R_{10} > R_8 > R_5 > R_{14} > R_3 > R_{11} > R_{16} > R_{13} > R_{15} > R_{12}
 \end{aligned} \quad (32)$$

Subsequently, according to literature research [50], the feature dimension retained after dimensionality reduction is set as $n-1$, where n is the number of operating conditions of gearbox, $n=6$. This is, the first six features are selected as sensitive features to get a new 6-dimensional feature vector set. Lastly, as the feature selection method selected in this paper, RSFS approach is also used to compare with the LS, FS and mRMR methods. Figure 23 visually depicts the visualization results of the raw signal and the distribution of all features, which is obtained by applying t-SNE algorithm to project the features into the three-

dimensional space. From Fig. 23, one can clearly see that the distance between different classes is larger and the clustering between similar classes is clearer when 6 sensitive features are selected using RSFS approach. By comparing Fig. 23 (a), (b), (d), (E) and (f), obviously, the RGCMSJE can extract recognizable features from cluttered and inseparable signals.

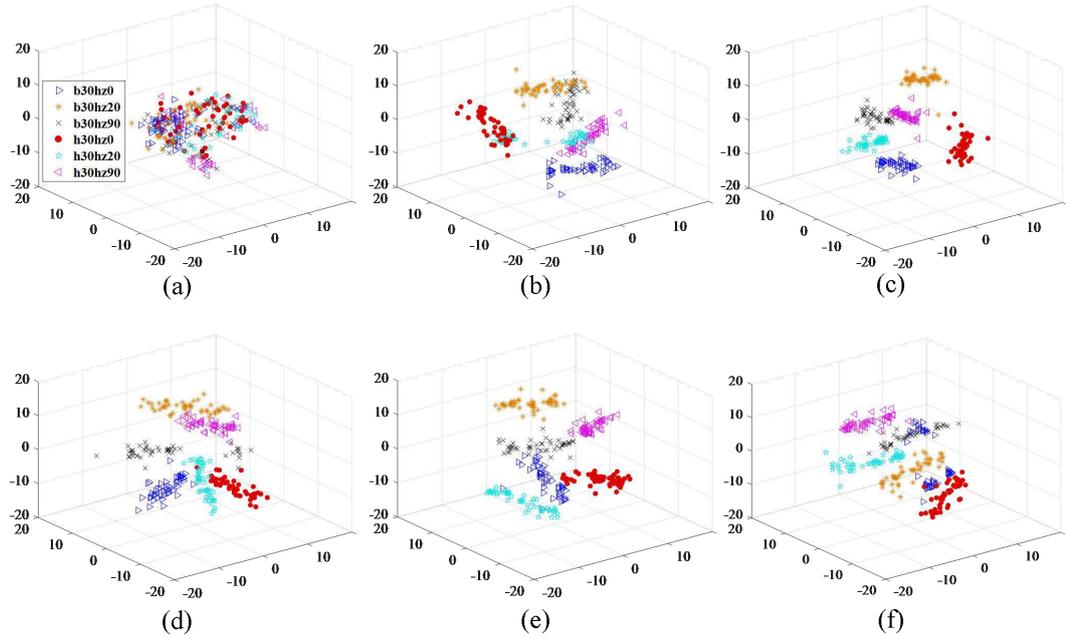


Fig. 23 The Feature visualization via t-SNE: (a) Raw signal; (b) RGCMSJE; (c) RGCMSJE+RSFS; (d) RGCMSJE+LS; (e) RGCMSJE+FS; (f) RGCMSJE+mRMR.

Table 11 The comparison results of different dimensionality reduction algorithms.

Method	Feature selection performance			CPU time(s)
	between-class scatter (S_b)	within-class scatter (S_w)	S_b / S_w	
RSFS	1.4249	0.1336	10.666	0.0905
LS	0.3802	0.1248	3.127	0.0632
FS	0.4203	0.1171	3.589	0.0356
mRMR	0.0721	0.0641	1.125	13.751

At the same time, the distinguishing ability of features selected by LS, FS and mRMR was significantly lower than that of RSFS. To further illustrate the importance of RSFS feature selection, the quantitative index S_b / S_w of feature selection performance is shown in Table 11. From Table 11, compared with LS, FS and mRMR, RSFS has the largest feature selection performance index, and the CPU time loss is only 0.0905s, which further proves the superiority of RSFS

algorithm in feature selection. Therefore, this paper chooses RSFS algorithm as the dimension reduction algorithm of high-dimensional feature space.

6.2.4 Fault diagnosis results analysis

Similar to experiment 1, the obtained dimension reduced features are fed into the ELM intelligent classifier for gearbox fault classification. The parameter setting of ELM in this section is the same as experiment 1. The active function is set to sin function and the number of hidden layers is set to 70.

1	49 16.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	49 16.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	1 0.3%	1 0.3%	50 16.7%	0 0.0%	0 0.0%	0 0.0%	96.2% 3.8%
4	0 0.0%	0 0.0%	0 0.0%	50 16.7%	0 0.0%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	50 16.7%	0 0.0%	100% 0.0%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	50 16.7%	100% 0.0%
	98.0% 2.0%	98.0% 2.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	99.3% 0.7%
	1	2	3	4	5	6	
	Target Class						

Fig. 24 ELM Confusion Matrix (CM) for RGCMSJE+RSFS

Similarly, to avoid the contingency of diagnostic results of the proposed method, the 10-fold cross validation algorithm is used in the training and testing process of ELM. And the average results of 20 experiments are used to verify the accuracy of fault identification. The confusion matrix in Fig. 24 shows the average accuracy of 20 experiments. The category labels defined by six working conditions b30hz0, b30hz20, b30hz90, h30hz0, h30hz20 and h30hz90 are 1, 2, 3, 4, 5 and 6 respectively. From Fig. 24, it is clearly that except for working condition 3, the recognition accuracy of the other five working conditions has reached 100%. The overall classification accuracy of the method proposed in this paper reaches 99.3%, which means that the method proposed in this paper is effective for the detection of different working conditions of transmission. The confusion matrix shows that one sample of working condition 1 and 2 is wrongly

classified into working condition 3, which indicates that it is not conducive to distinguish the change of working condition in case of gear failure.

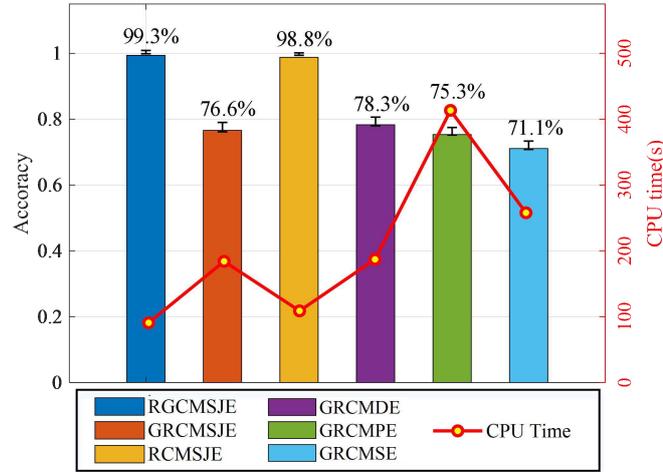


Fig. 25 The performance comparison results of the six feature extraction algorithms.

Similarly, we compare RGCMSJE, GRCMSJE, RCMSJE, GRCMDE, GRCMPE and GRCMSE from classification accuracy and CPU time loss using ELM intelligent classifier. A clear comparison result is shown in Fig. 25 that the feature extracted by RGCMSJE has the highest classification accuracy, while RCMSJE is second only to it.

This shows that the features extracted by state joint analysis are effective, and the RGCMSJE method achieves the highest recognition accuracy with lower CPU time loss. At the same time, it proves that RGCMSGE feature extraction method has good generalization ability.

Similar to Section 6.1, we compare the comprehensive performance of the five classifiers to verify the efficiency of ELM. The parameter setting of the five classifiers is the same as Table 7, and the feature set of the input classifier is obtained by RGCMSJE and RSFS methods. Each algorithm is operated 20 times, and detailed comparison results are given in Fig. 26 and Table 12. From the radar distribution Figure 26, it is clearly observed that the four evaluation parameters of the RGCMSJE-RSFS-ELM curve proposed in this paper are farthest from the center, which indicates that the comprehensive classification ability of ELM is the best, and it also proves that the RGCMSJE-RSFS-ELM model proposed in this paper has the highest fault recognition ability. Besides, the data in Table 12 shows that the training and testing of ELM classifier consumes the least CPU time.

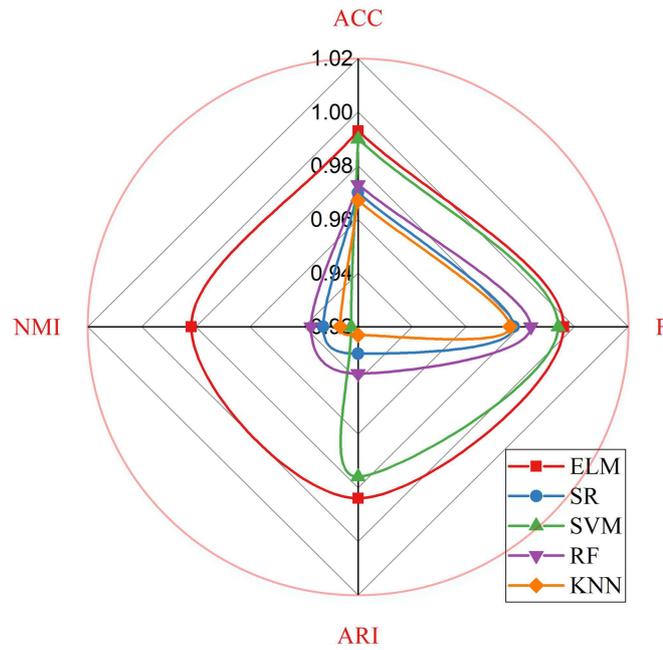


Fig. 26 The radar chart on evaluation results of different classifiers.

Table 12 The results of classifier performance evaluation are compared.

Mode	Result evaluation				
	ACC	F	ARI	NMI	CPU time
ELM	99.3%	0.99599	0.9842	0.9817	0.0983
SR	97%	0.97753	0.9803	0.9329	1.6757
SVM	99%	0.99398	0.9759	0.9226	3.2421
RF	97.3%	0.98388	0.9375	0.9374	2.2705
KNN	96.7%	0.97615	0.9231	0.9266	0.2929

6.3 Further discusses

Through the comparative analysis of the above methods, we can conclude that the proposed method combines the advantages of RGCMSJE, MAED, RSFS and ELM, so that the fault state of gearbox components can be effectively identified. The emphasis of the proposed method focuses on three aspects: refined generalized multi-scale analysis, state joint entropy, MAED aided parameter determination and RSFS. At the same time, through the comparative analysis of ELM intelligent classifier recognition accuracy and efficiency, we take ELM as the final fault identifier. The experimental results of section A and section B show that the fault diagnosis method proposed in this paper can not only identify various gear faults under the same working conditions, but also can identify gear faults under various working conditions, and has good generalization ability and

stability. Here, the reasons leading to the advantages of the proposed model are summarized as follows:

- (1) In previous reports, DE has been proved to have some advantages in computing time and reliability due to its unique symbolization and linear mapping rules. However, DE is the same as SE and PE algorithms, which only considers the mode probability of current state and ignores the state transition probability of fault information from one state to another. Based on these facts, SJE entropy, which combines the advantages of DE algorithm and considering the mode probability and transfer probability of current state, is proposed. Therefore, SJE entropy can extract fault information stably and efficiently.
- (2) Generalized composite multi-scale analysis method effectively solves the problem of missing mutation behavior in composite multi-scale analysis, but the former is not suitable for DE and SJE algorithm, because many useful information will disappear with the transformation of second moment, which will lead to the instability of entropy value. Therefore, we propose a refined generalized composite multiscale analysis method to solve the above problems. Refined generalized analysis and composite operation can avoid the problem of information loss caused by second moment and extract information on multiple time scales.
- (3) Previous studies on the parameters of entropy algorithm are based on the traversal simulation of analog signals. Therefore, the entropy parameters for specific fault data cannot be obtained, and the parameter selection process needs the experience support of researchers, which inevitably leads to errors. In this paper, MAED method is proposed to determine the parameters of RGCMSJE method, which can realize the analysis of specific data and get the best parameters adaptively.
- (4) In the feature selection process, LS algorithm lacks the ability of global information separation, on the contrary, FS algorithm lacks the ability of local information preservation. Therefore, we apply a new feature selection method RSFS to reduce the dimension of multi-dimensional feature space. This method uses robust local learning method to deal with the noise on the clustering label, so as to improve the local information retention ability while taking into account the global information.

7 Conclusion

In this paper, a neoteric intelligent fault diagnosis method of wind turbine gearbox based on RGCMSJE, MAED, RSFS and ELM is presented. First, the concept of SJE entropy is defined, and the advantages of SJE are verified by comparing SJE with DE, SE and PE. Then a new fault feature extraction method RGCMSJE is proposed based on the refined composite multi-scale process, which is used to extract the fault features of wind turbine gearbox. At the same time, the MAED adaptive auxiliary RGCMSJE algorithm proposed in this paper selects the parameters to avoid the error caused by artificial experience. After obtaining the best parameters, compared with GRCMSJE, RCMSJE, GRCMDE, GRCMSE and GRCMPE methods, the proposed RGCMSJE method has the advantages of faster calculation speed and more stable feature extraction. Subsequently, the RSFS algorithm is introduced to automatically select salient features from the multi-scale features based on robust local learning. Finally, a gearbox fault diagnosis method based on RGCMSJE, MAED, RSFS and ELM is constructed. The analysis results of two experimental cases indicate that the method can extract the operating characteristics of the gearbox concisely and effectively, and can successfully identify a variety of local faults and working conditions of the gearbox. More importantly, compared with the existing methods (such as GRCMDE, GRCMSE and GRCMPE), the proposed diagnostic scheme has significant advantages in diagnostic accuracy and computational efficiency. In future work, more industrial site data should be employed to verify the generalization ability of the proposed method. In addition, the selection of MAED assisted entropy parameters and the combination of multi-scale analysis and RSFS method are very interesting.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of Hebei Province under Grant F2020203058 and E2018203339, and the Key R&D projects of Hebei Province under Grant 18211833D. Authors are grateful to the University of Connection for allowing us to use their data.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- [1] He Q, Pang Y, Jiang G, et al. A spatio-temporal multiscale neural network approach for wind turbine fault diagnosis with imbalanced SCADA data[J]. *IEEE Transactions on Industrial Informatics*, pp. 1-1, 2020.
- [2] Liu Z, Tang X, Wang X, et al. Wind turbine blade bearing fault diagnosis under fluctuating speed operations via bayesian augmented lagrangian analysis[J]. *IEEE Transactions on Industrial Informatics*, pp. 1-1, 2020.
- [3] Lee J, Zhao F, Dutton A, et al. Global wind report 2019[J]. Brussels: Global Wind Energy Council (GWEC), 2020. [Online]. Available: <https://gwec.net/globalwind-report-2019/>
- [4] Xu Z, Li C, Yang Y. Fault diagnosis of rolling bearing of wind turbines based on the Variational Mode Decomposition and Deep Convolutional Neural Networks[J]. *Applied Soft Computing*, 2020, 95: 106515.
- [5] Yu X, Tang B, Zhang K. Fault Diagnosis of Wind Turbine Gearbox Using a Novel Method of Fast Deep Graph Convolutional Networks[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-14.
- [6] Miao Y, Zhao M, Liang K, et al. Application of an improved MCKDA for fault detection of wind turbine gear based on encoder signal[J]. *Renewable Energy*, 2020, 151: 192-203.
- [7] Wang Z, Yao L, Cai Y, et al. Mahalanobis semi-supervised mapping and beetle antennae search based support vector machine for wind turbine rolling bearings fault diagnosis[J]. *Renewable Energy*, 2020, 155: 1312-1327.
- [8] Li Y, Liang X, Wei Y, et al. A method based on refined composite multi-scale symbolic dynamic entropy and ISVM-BT for rotating machinery fault diagnosis[J]. *Neurocomputing*, 2018, 315: 246-260.
- [9] Kong Y, Wang T, Feng Z, et al. Discriminative dictionary learning based sparse representation classification for intelligent fault identification of planet bearings in wind turbine[J]. *Renewable Energy*, 2020, 152: 754-769.
- [10] Yan X, Jia M. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing[J]. *Neurocomputing*, 2018, 313: 47-64.
- [11] Wang X, Si S, Li Y, et al. An integrated method based on refined composite multivariate hierarchical permutation entropy and random forest and its application in rotating machinery[J]. *Journal of Vibration and Control*, 2020, 26(3-4): 146-160.
- [12] Yang C, Jia M. Health condition identification for rolling bearing based on hierarchical multiscale symbolic dynamic entropy and least squares support tensor machine-based binary tree[J]. *Structural Health Monitoring*, 2020: 1475921720923973.

- [13] Ren H, Liu W, Shan M, et al. A novel wind turbine health condition monitoring method based on composite variational mode entropy and weighted distribution adaptation[J]. *Renewable Energy*, 2021, 168: 972-980.
- [14] Yang Y, Zheng H, Li Y, et al. A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network[J]. *ISA transactions*, 2019, 91: 235-252.
- [15] Yin J, Xu M, Zheng H. Fault diagnosis of bearing based on Symbolic Aggregate approximation and Lempel-Ziv[J]. *Measurement*, 2019, 138: 206-216.
- [16] Wang C. A sample entropy inspired affinity propagation method for bearing fault signal classification[J]. *Digital Signal Processing*, 2020, 102: 102740.
- [17] Noman K, Wang D, Peng Z, et al. Oscillation based permutation entropy calculation as a dynamic nonlinear feature for health monitoring of rolling element bearing[J]. *Measurement*, 2021, 172
- [18] Deng W, Zhang S, Zhao H, et al. A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing[J]. *IEEE Access*, 2018, 6: 350
- [19] Rostaghi M, Ashory M R, Azami H. Application of dispersion entropy to status characterization of rotary machines[J]. *Journal of Sound and Vibration*, 2019, 438: 291-308.
- [20] 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[J]. *Entropy*, 2017, 19(11): 585.
- [21] Li Y, Yang Y, Wang X, et al. Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine[J]. *Journal of Sound and Vibration*, 2018, 428: 72-86.
- [22] Zhu K, Jiang X, Chen L, et al. Performance degradation assessment of rolling element bearings using improved fuzzy entropy[J]. *Measurement Science Review*, 2017, 17(5): 219.
- [23] Rostaghi M, Azami H. Dispersion entropy: A measure for time-series analysis[J]. *IEEE Signal Processing Letters*, 2016, 23(5): 610-614.
- [24] Zheng J, Pan H, Yang S, et al. Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis[J]. *Mechanical Systems and Signal Processing*, 2018, 99: 229-243.
- [25] Wang Z, Yao L, Cai Y. Rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine[J]. *Measurement*, 2020, 156: 107574.
- [26] Dong W, Zhang S, Jiang A, et al. Intelligent fault diagnosis of rolling bearings based on refined composite multi-scale dispersion q-complexity and adaptive whale algorithm-extreme learning machine[J]. *Measurement*, 2021, 176: 108977.
- [27] Azami H, Rostaghi M, Abásolo D, et al. Refined composite multiscale dispersion entropy and its application to biomedical signals[J]. *IEEE Transactions on Biomedical Engineering*, 2017, 64(12): 2872-2879.
- [28] Li Y, Li G, Wei Y, et al. Health condition identification of planetary gearboxes based on variational mode decomposition and generalized composite multi-scale symbolic dynamic entropy[J]. *ISA transactions*, 2018, 81: 329-341.

- [29] Berredjem T, Benidir M. Bearing faults diagnosis using fuzzy expert system relying on an improved range overlaps and similarity method[J]. *Expert Systems with Applications*, 2018, 108: 134-142.
- [30] Gan X, Lu H, Yang G. Fault diagnosis method for rolling bearings based on composite multiscale fluctuation dispersion entropy[J]. *Entropy*, 2019, 21(3): 290.
- [31] Yan X, Liu Y, Huang D, et al. A new approach to health condition identification of rolling bearing using hierarchical dispersion entropy and improved Laplacian score[J]. *Structural Health Monitoring*, 2020: 1475921720948620.
- [32] Shi L, Du L, Shen Y D. Robust spectral learning for unsupervised feature selection[C]//2014 IEEE International Conference on Data Mining. IEEE, 2014: 977-982.
- [33] Zhang Z, Chen H, Li S, et al. A novel sparse filtering approach based on time-frequency feature extraction and softmax regression for intelligent fault diagnosis under different speeds[J]. *Journal of Central South University*, 2019, 26(6): 1607-1618.
- [34] Zhang X, Li C, Wang X, et al. A novel fault diagnosis procedure based on improved symplectic geometry mode decomposition and optimized SVM[J]. *Measurement*, 2021, 173: 108644.
- [35] Han T, Jiang D, Zhao Q, et al. Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery[J]. *Transactions of the Institute of Measurement and Control*, 2018, 40(8): 2681-2693.
- [36] Peña M, Cerrada M, Alvarez X, et al. Feature engineering based on anova, cluster validity assessment and knn for fault diagnosis in bearings[J]. *Journal of Intelligent & Fuzzy Systems*, 2018, 34(6): 3451-3462.
- [37] Yu H, Sun C, Yang X, et al. ODOC-ELM: Optimal decision outputs compensation-based extreme learning machine for classifying imbalanced data[J]. *Knowledge-Based Systems*, 2016, 92: 55-70.
- [38] Haidong S, Hongkai J, Xingqiu L, et al. Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine[J]. *Knowledge-Based Systems*, 2018, 140: 1-14.
- [39] Sarkar S, Mukherjee K, Jin X, et al. Optimization of time-series data partitioning for parameter identification[C]//Dynamic Systems and Control Conference. 2010, 44175: 867-874.
- [40] Li Y, Xu M, Wei Y, et al. Health condition monitoring and early fault diagnosis of bearings using SDF and intrinsic characteristic-scale decomposition[J]. *IEEE Transactions on Instrumentation and Measurement*, 2016, 65(9): 2174-2189.
- [40] Li Y, Xu M, Wei Y, et al. Health condition monitoring and early fault diagnosis of bearings using SDF and intrinsic characteristic-scale decomposition[J]. *IEEE Transactions on Instrumentation and Measurement*, 2016, 65(9): 2174-2189.
- [41] Chen Z, Li Y, Liang H, et al. Improved permutation entropy for measuring complexity of time series under noisy condition[J]. *Complexity*, 2019, 2019.
- [42] Ma J, Li Z, Li C, et al. Rolling Bearing Fault Diagnosis Based on Refined Composite Multi-Scale Approximate Entropy and Optimized Probabilistic Neural Network[J]. *Entropy*, 2021, 23(2): 259.

- [43] Yao W, Wang J. Double symbolic joint entropy in nonlinear dynamic complexity analysis[J]. *Aip Advances*, 2017, 7(7): 075313.
- [44] Li Y, Liang X, Xu M, et al. Early fault feature extraction of rolling bearing based on ICD and tunable Q-factor wavelet transform[J]. *Mechanical Systems and Signal Processing*, 2017, 86: 204-223.
- [45] Zheng J, Pan H. Use of generalized refined composite multiscale fractional dispersion entropy to diagnose the faults of rolling bearing[J]. *Nonlinear Dynamics*, 2020, 101(2): 1417-1440.
- [46] Li Y, Yang Y, Wang X, et al. Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine[J]. *Journal of Sound and Vibration*, 2018, 428: 72-86.
- [47] Chen Z, Gryllias K, Li W. Mechanical fault diagnosis using convolutional neural networks and extreme learning machine[J]. *Mechanical systems and signal processing*, 2019, 133: 106272.
- [48] Cao P, Zhang S, Tang J. Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning[J]. *Ieee Access*, 2018, 6: 26241-26253.
- [49] Xu G, Liu M, Jiang Z, et al. Online fault diagnosis method based on transfer convolutional neural networks[J]. *IEEE Transactions on Instrumentation and Measurement*, 2019, 69(2): 509-520.
- [50] Yan X, Jia M. Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection[J]. *Knowledge-Based Systems*, 2019, 163: 450-471.
- [51] Available online: <https://github.com/Gearboxdata/Gear-Box-Fault-Diagnosis-Data-Set> (accessed on 05 November 2020).

Figures

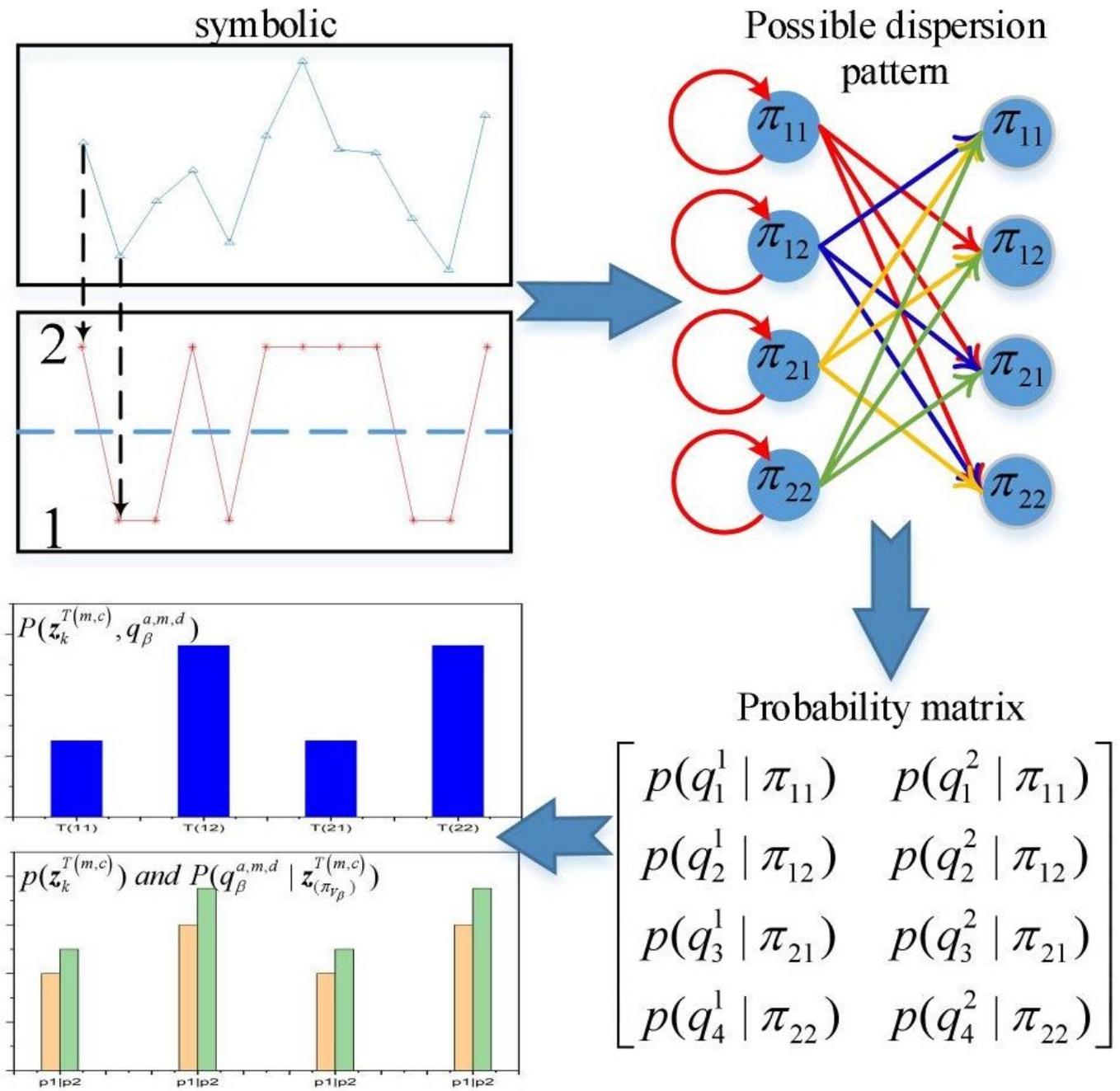


Figure 1

Please see the Manuscript PDF file for the complete figure caption

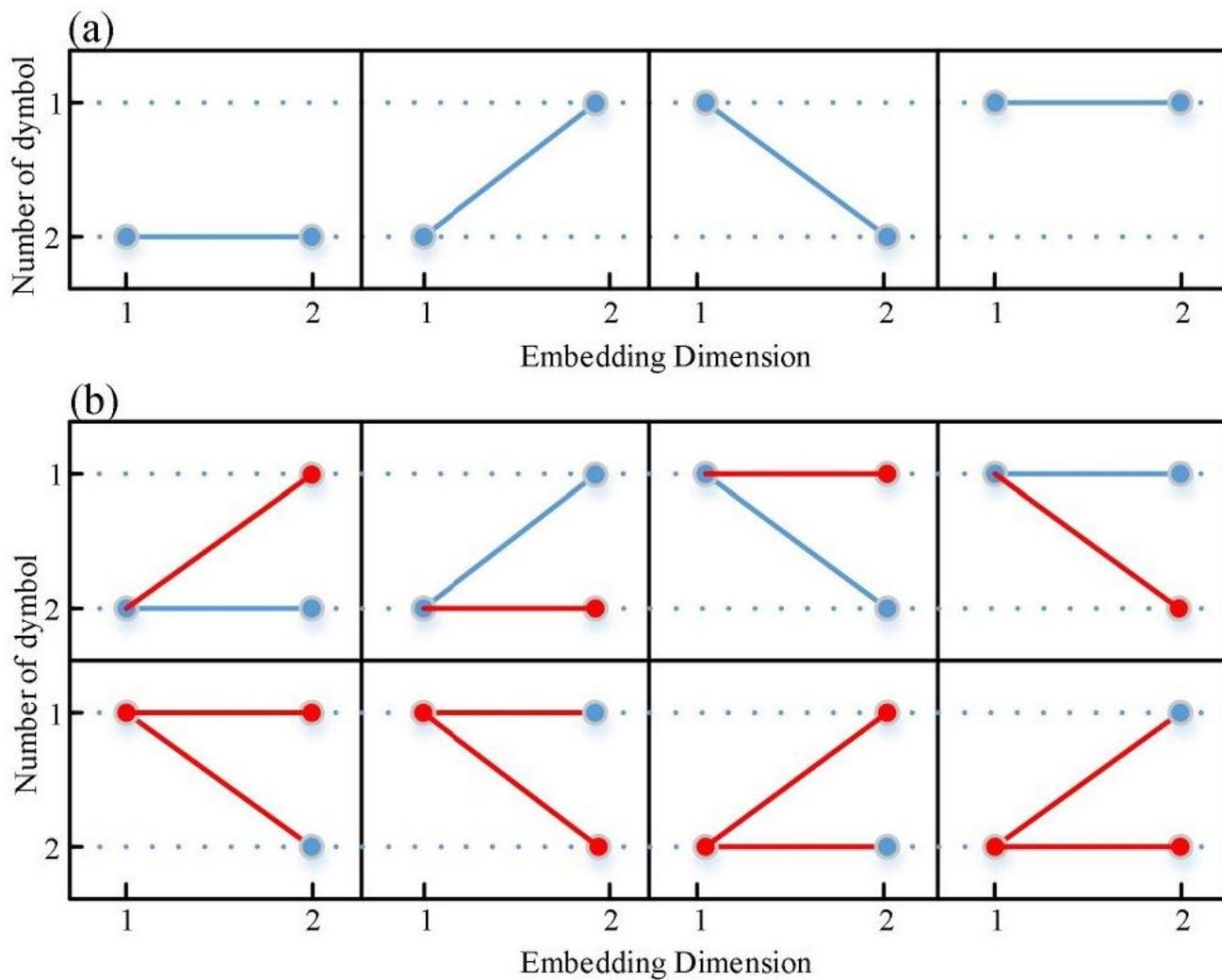


Figure 2

Please see the Manuscript PDF file for the complete figure caption

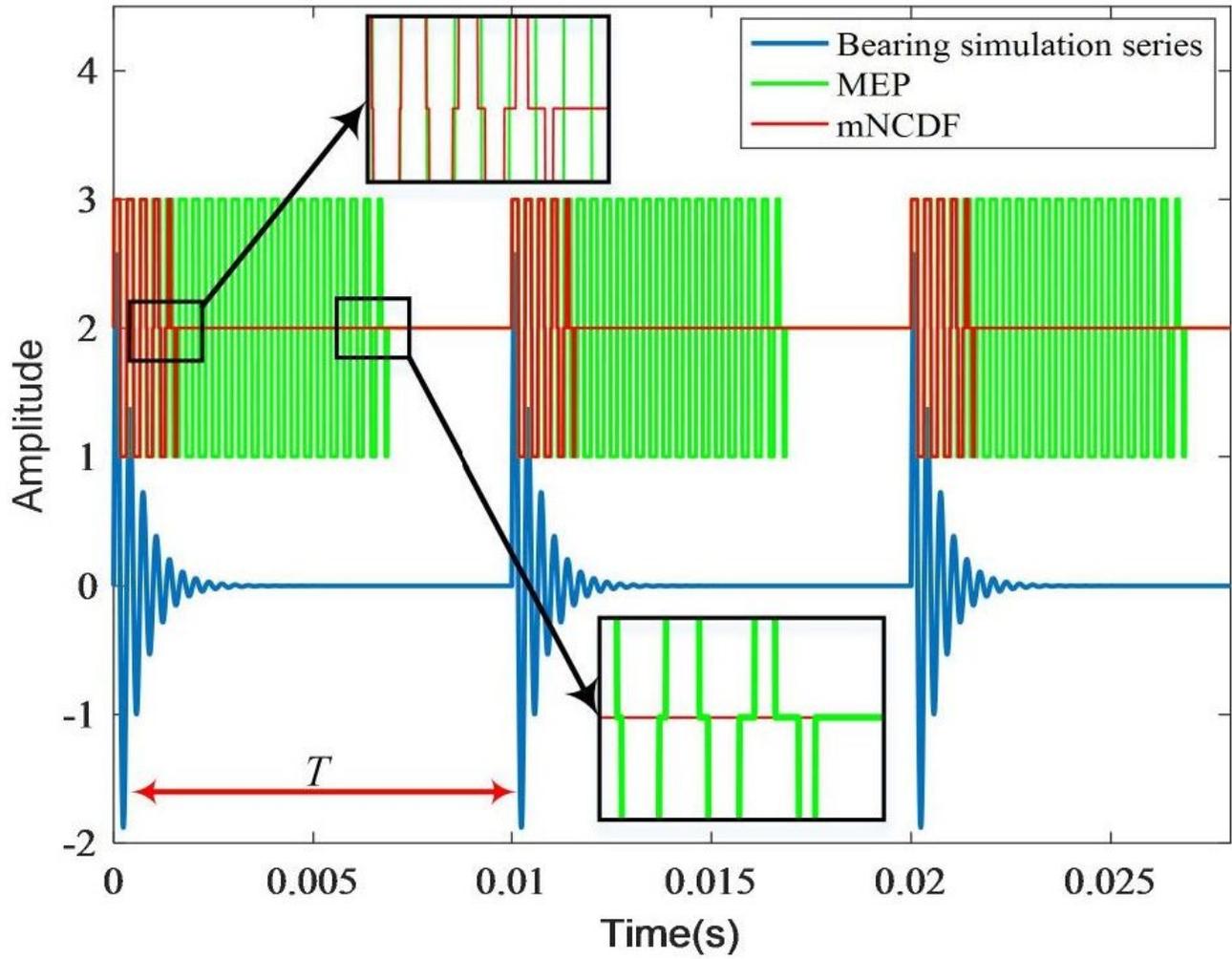


Figure 3

Please see the Manuscript PDF file for the complete figure caption

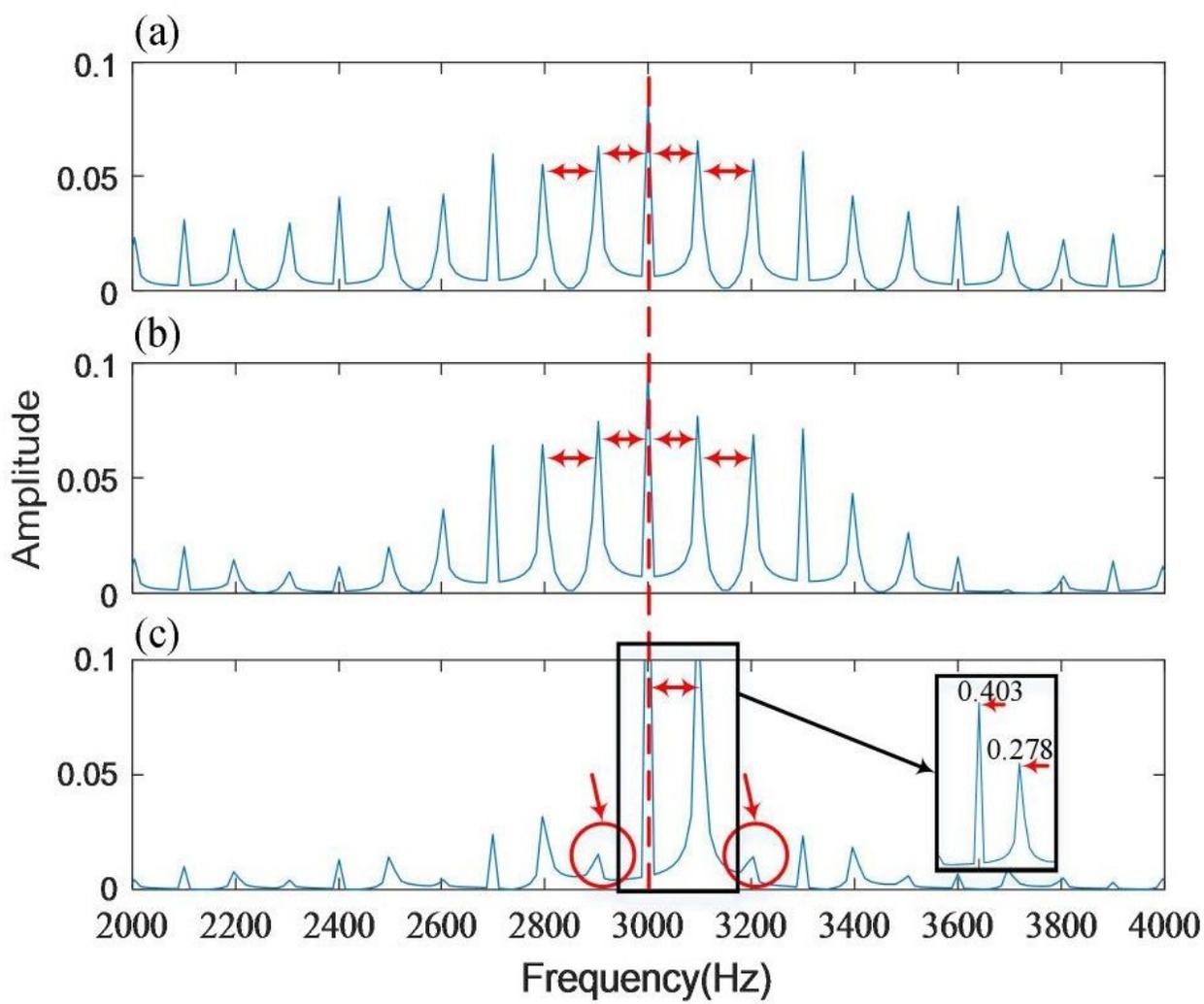


Figure 4

Please see the Manuscript PDF file for the complete figure caption

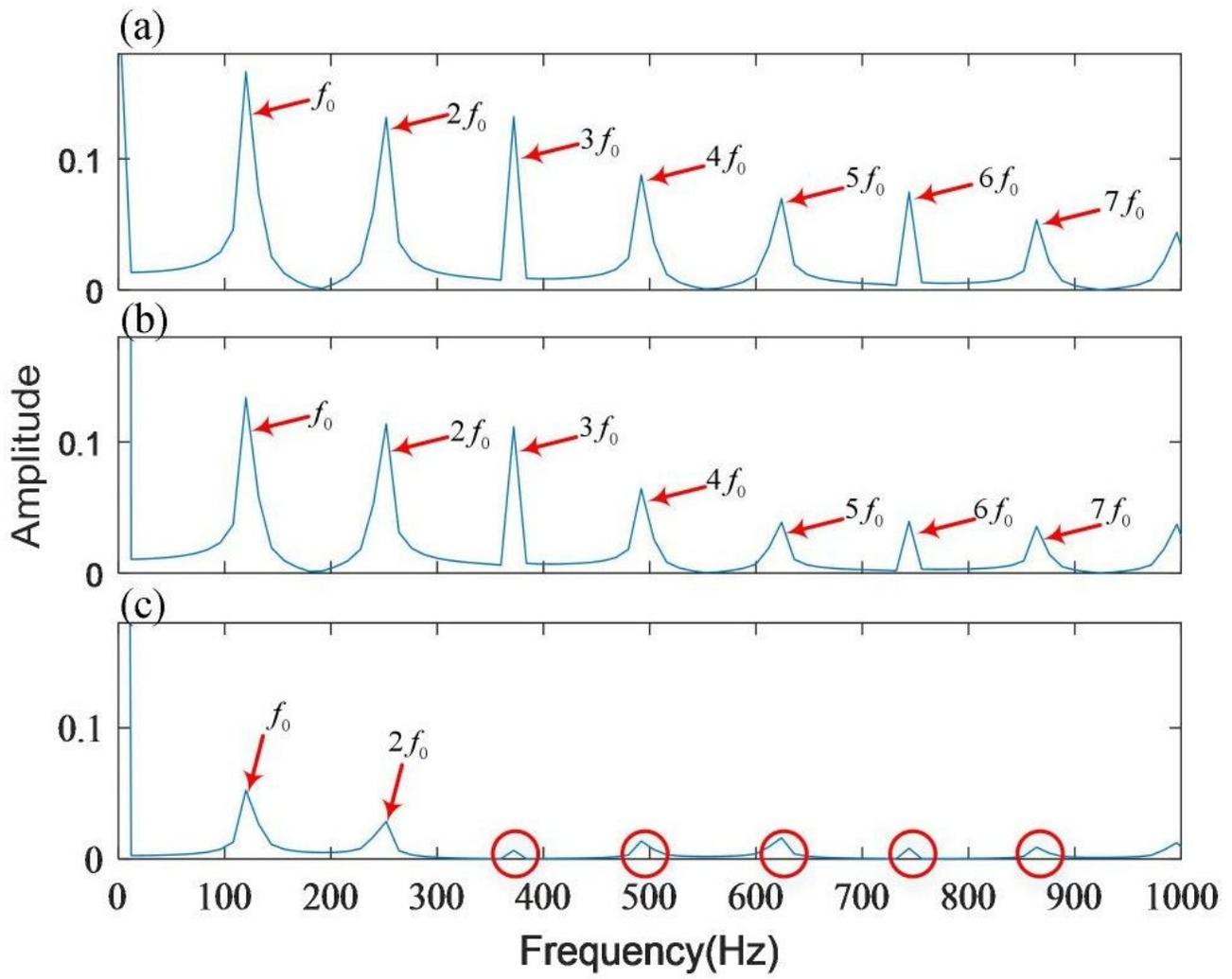


Figure 5

Please see the Manuscript PDF file for the complete figure caption

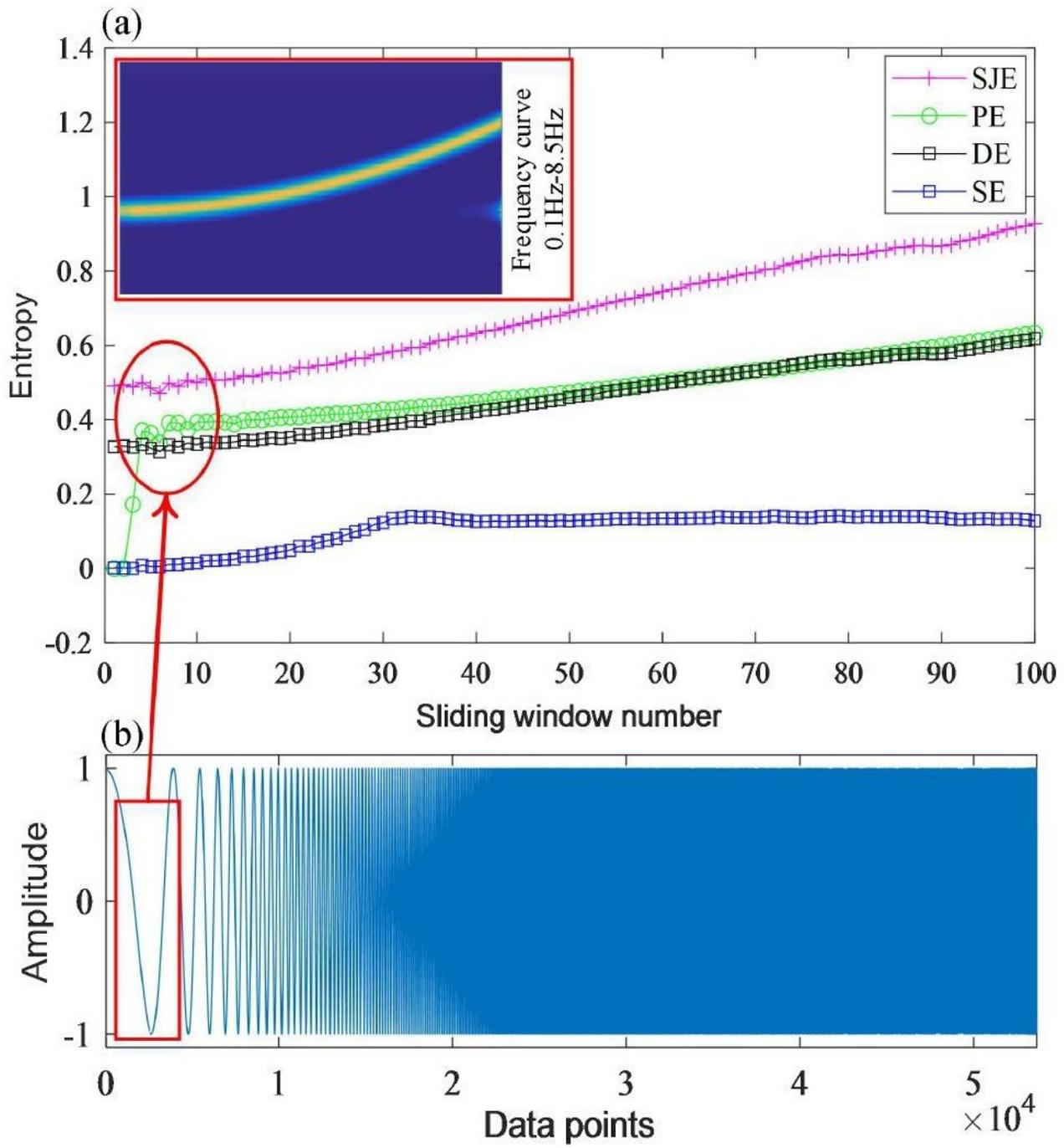


Figure 6

Please see the Manuscript PDF file for the complete figure caption

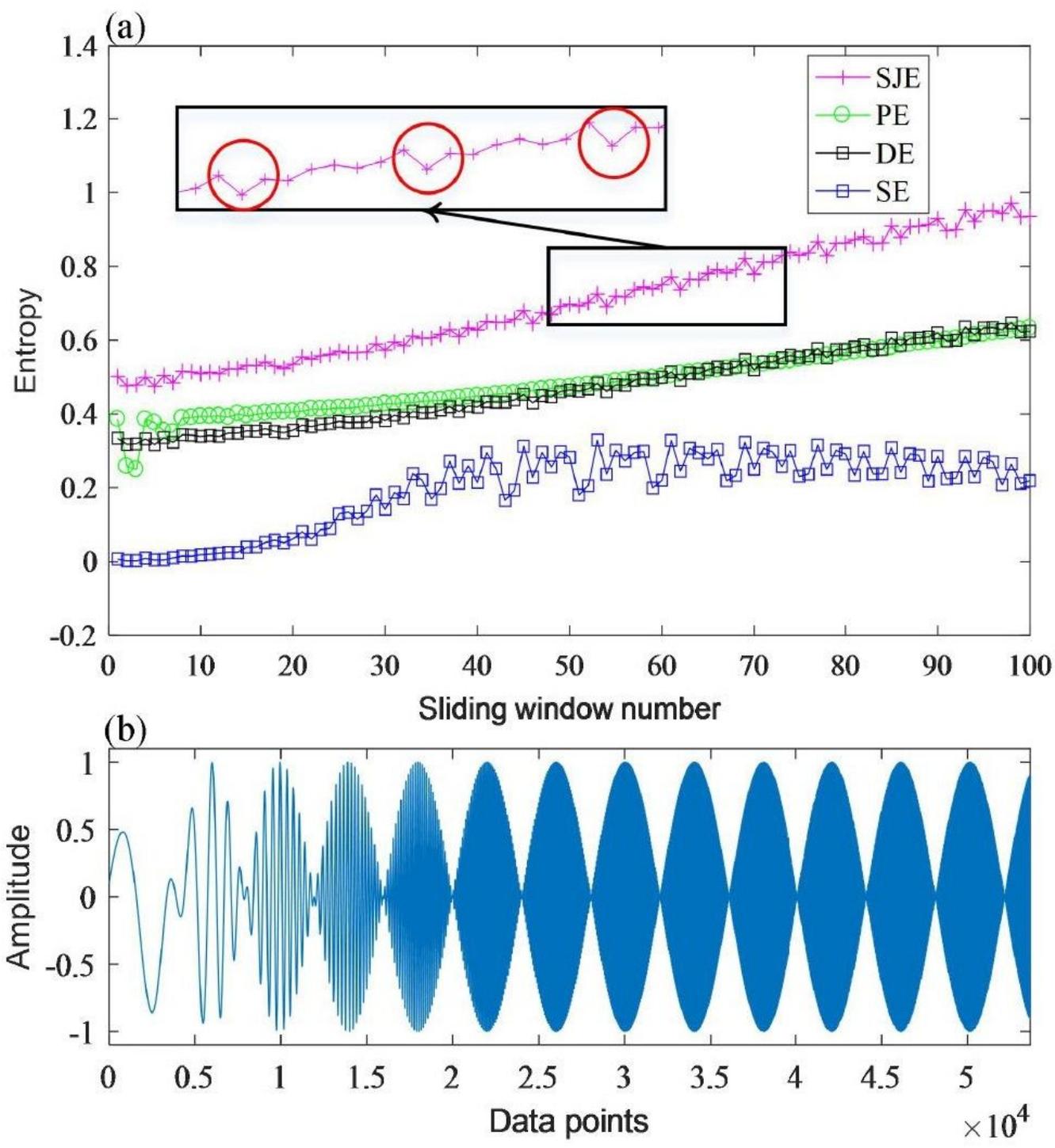


Figure 7

Please see the Manuscript PDF file for the complete figure caption

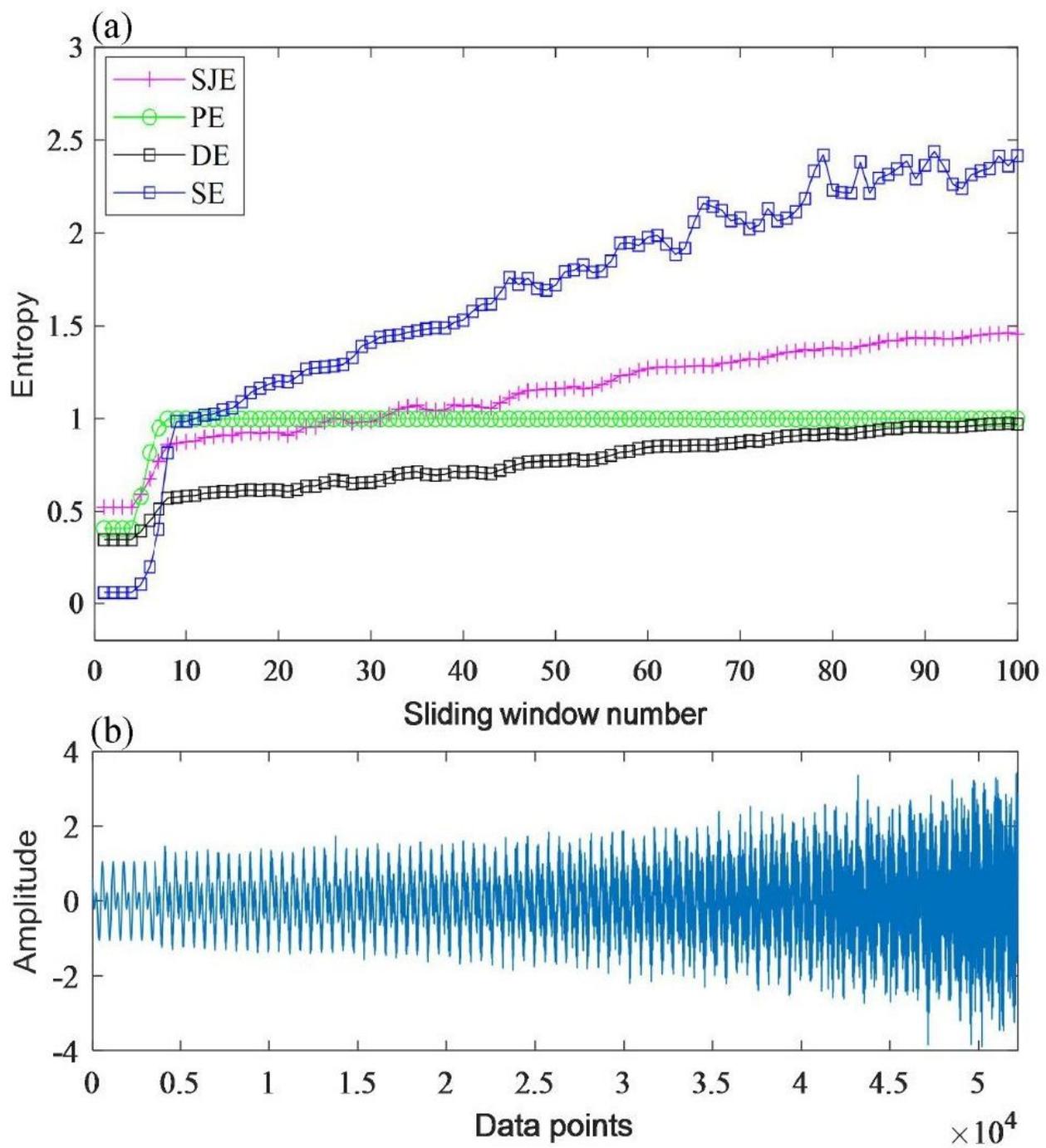


Figure 8

Please see the Manuscript PDF file for the complete figure caption

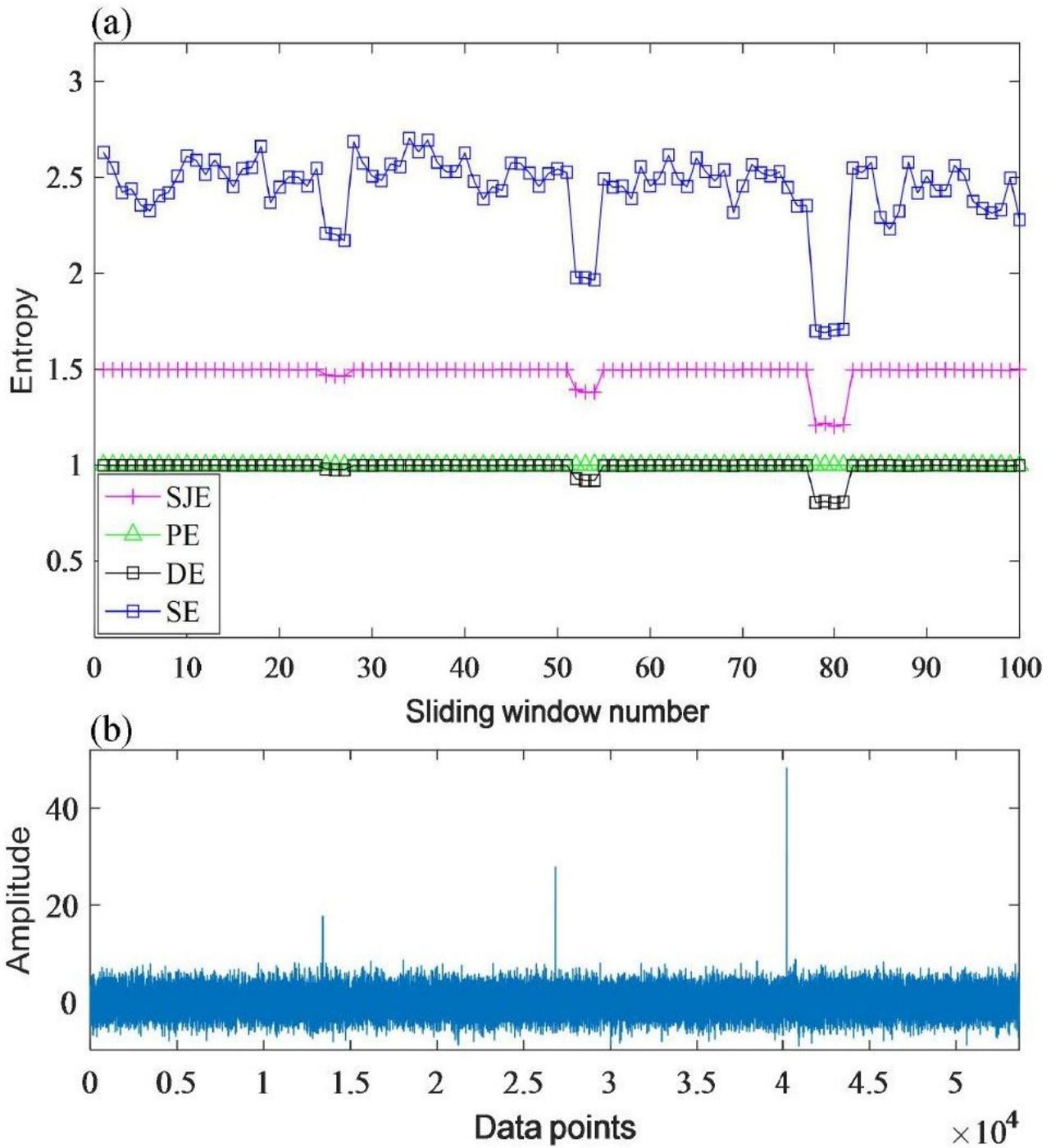


Figure 9

Please see the Manuscript PDF file for the complete figure caption

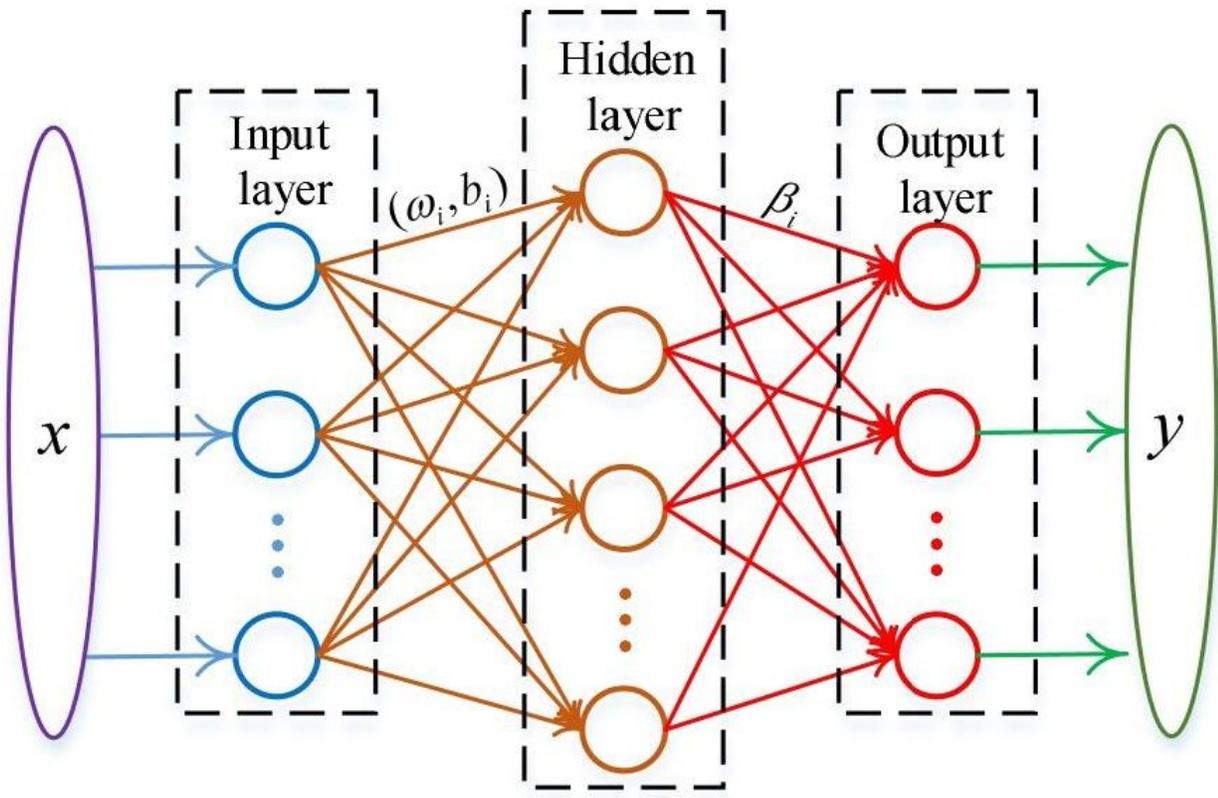


Figure 10

Please see the Manuscript PDF file for the complete figure caption

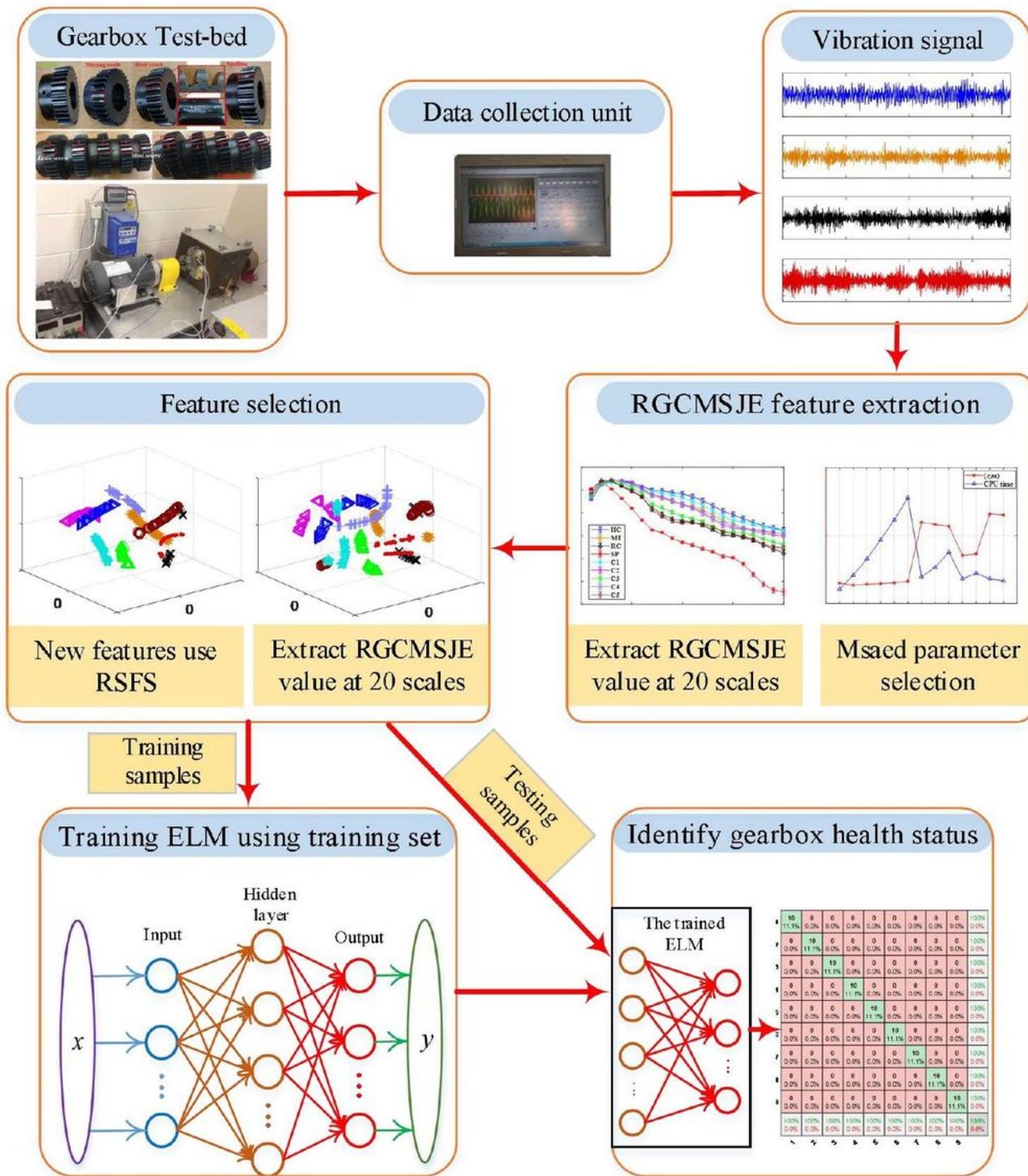


Figure 11

Please see the Manuscript PDF file for the complete figure caption



Figure 12

Please see the Manuscript PDF file for the complete figure caption

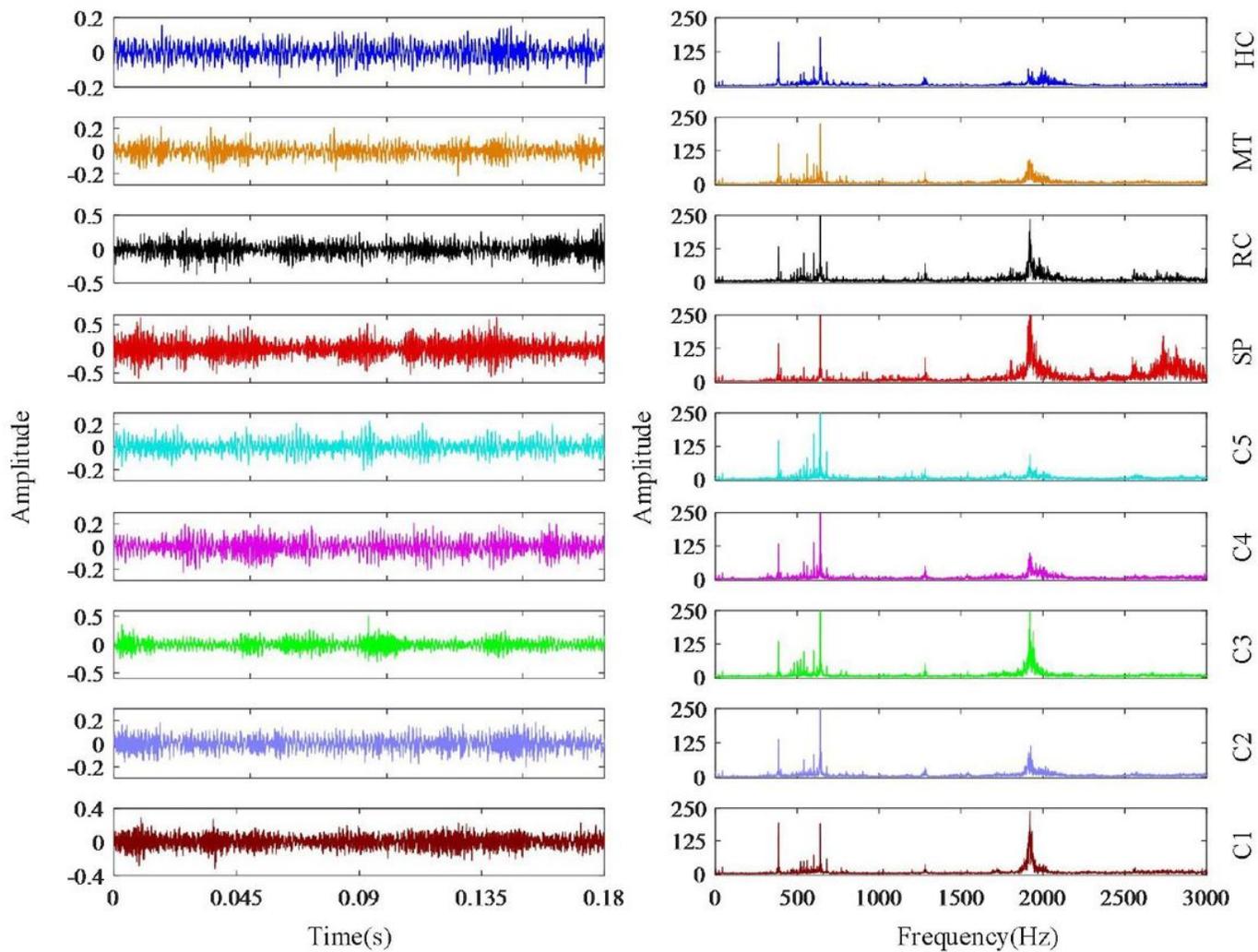


Figure 13

Please see the Manuscript PDF file for the complete figure caption

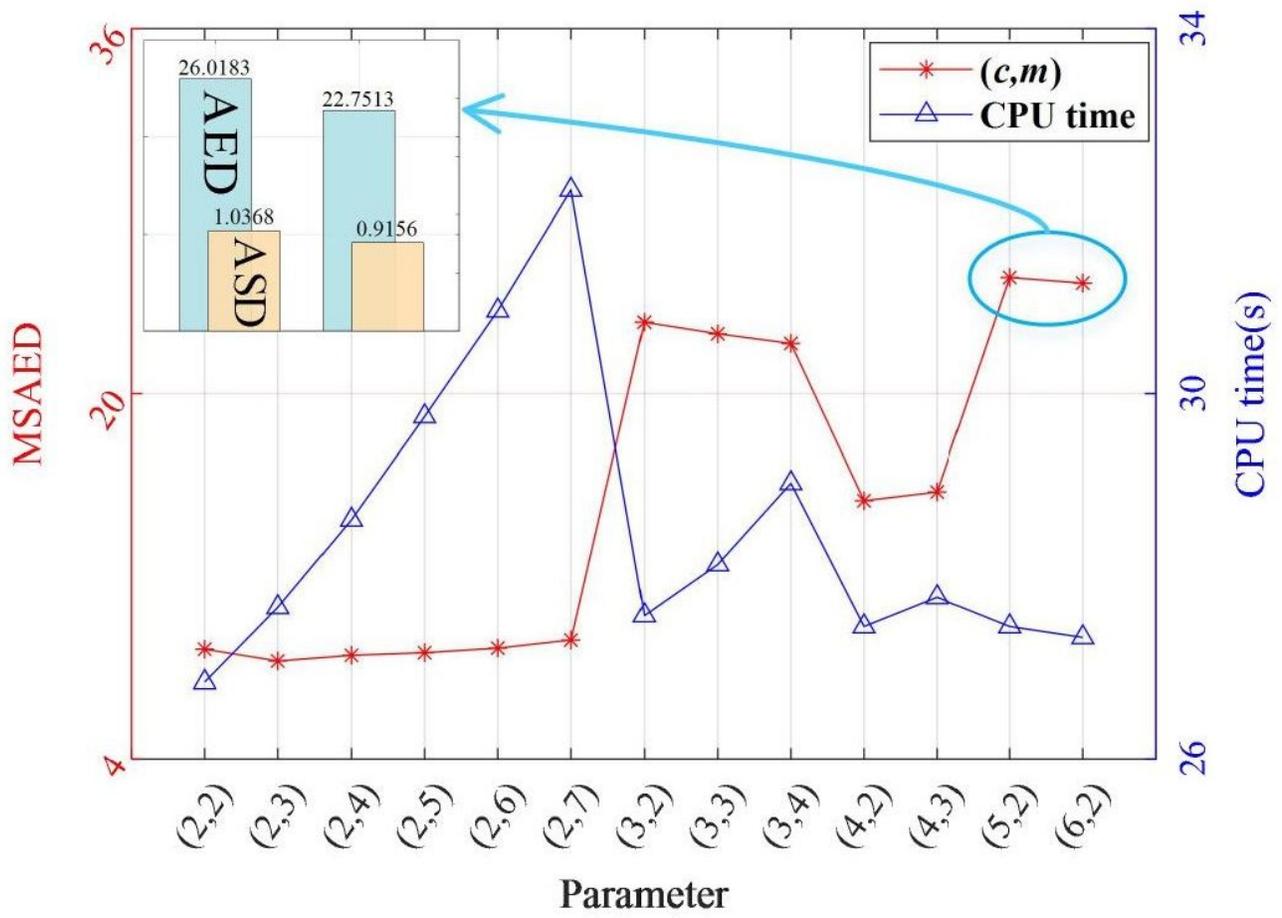


Figure 14

Please see the Manuscript PDF file for the complete figure caption

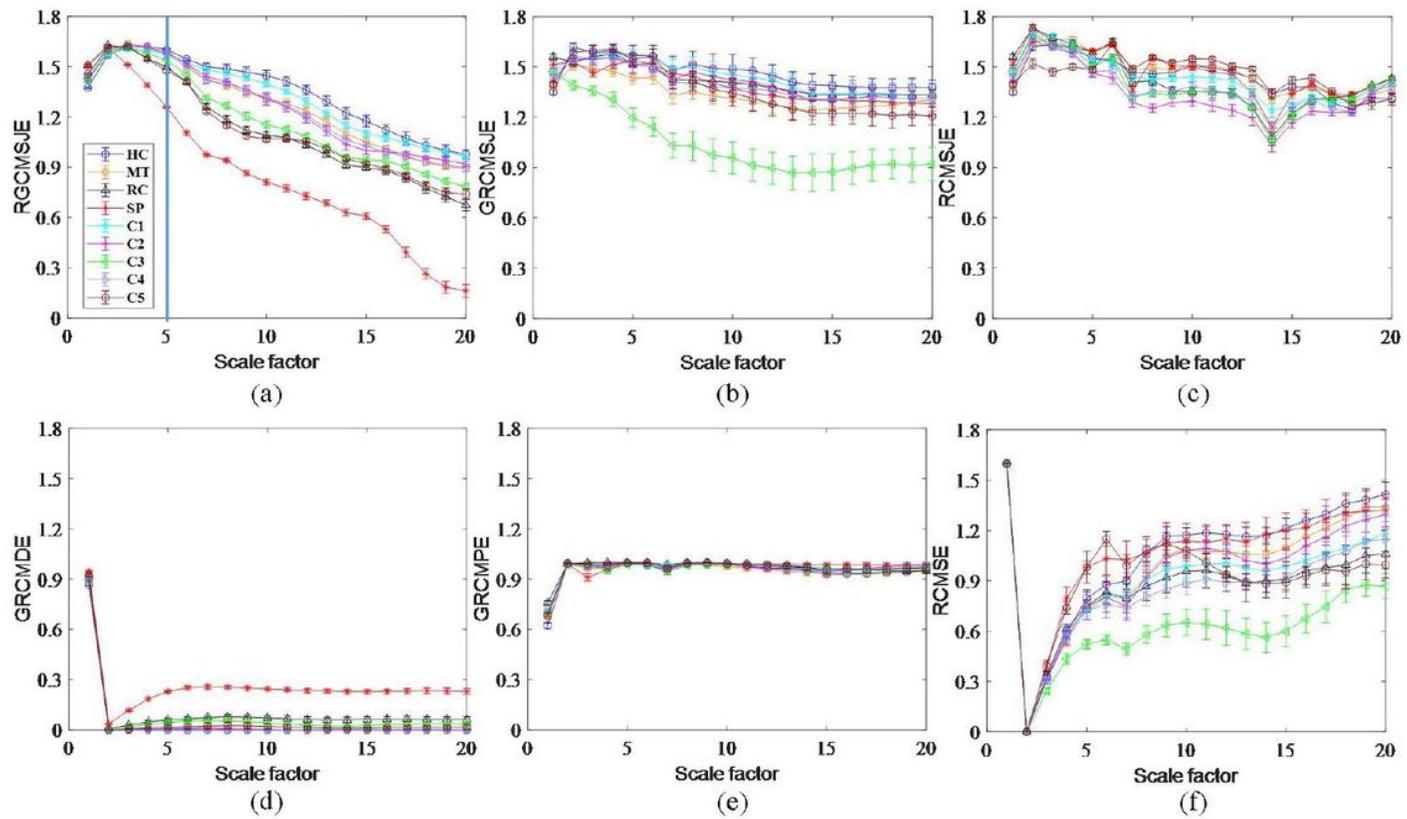


Figure 15

Please see the Manuscript PDF file for the complete figure caption

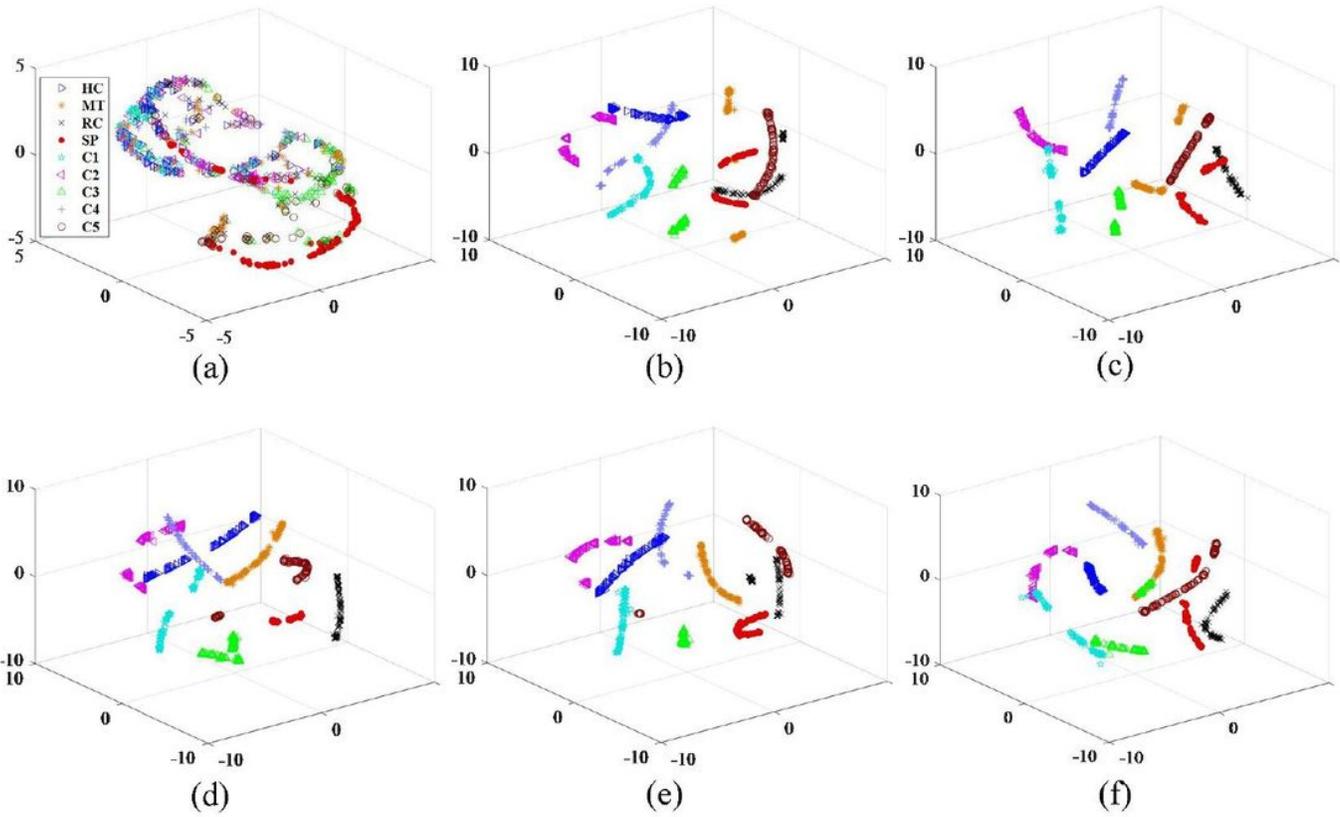


Figure 16

Please see the Manuscript PDF file for the complete figure caption

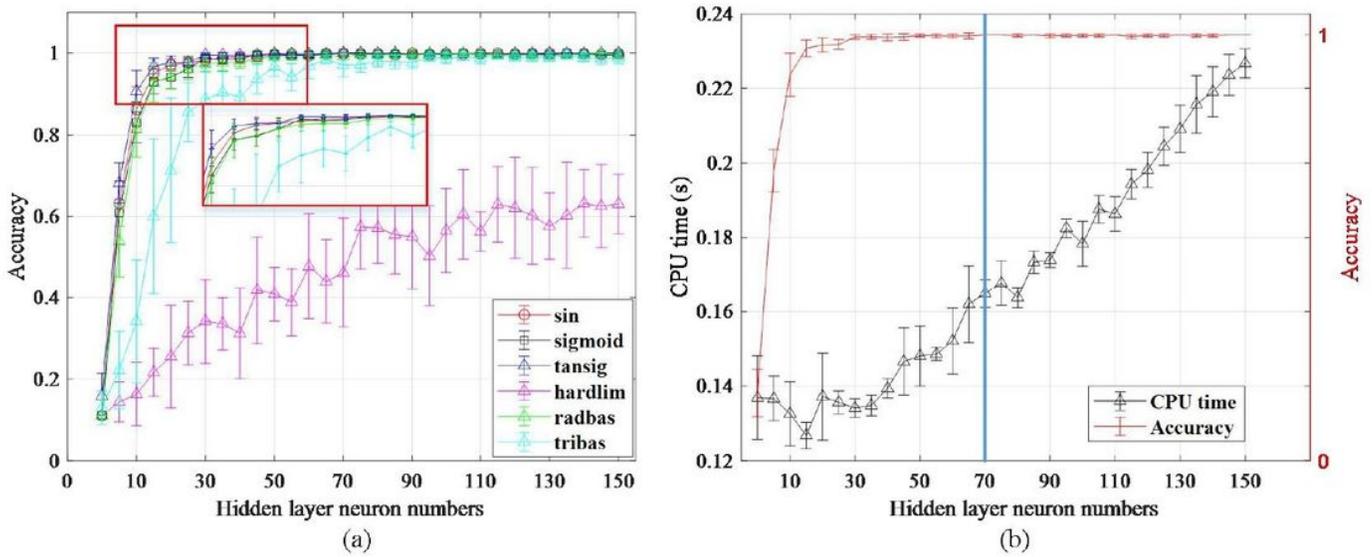


Figure 17

Please see the Manuscript PDF file for the complete figure caption

Output Class	1	100 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	2	0 0.0%	100 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	3	0 0.0%	0 0.0%	100 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	4	0 0.0%	0 0.0%	0 0.0%	100 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100 11.1%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99 11.0%	0 0.0%	0 0.0%	100% 0.0%
	7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100 11.1%	0 0.0%	100% 0.0%
	8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	0 0.0%	100 11.1%	99.0% 1.0%
	9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 11.1%
			100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	99.0% 1.0%	100% 0.0%	100% 0.0%	100% 0.0%
		1	2	3	4	5	6	7	8	9
		Target Class								

Figure 18

Please see the Manuscript PDF file for the complete figure caption

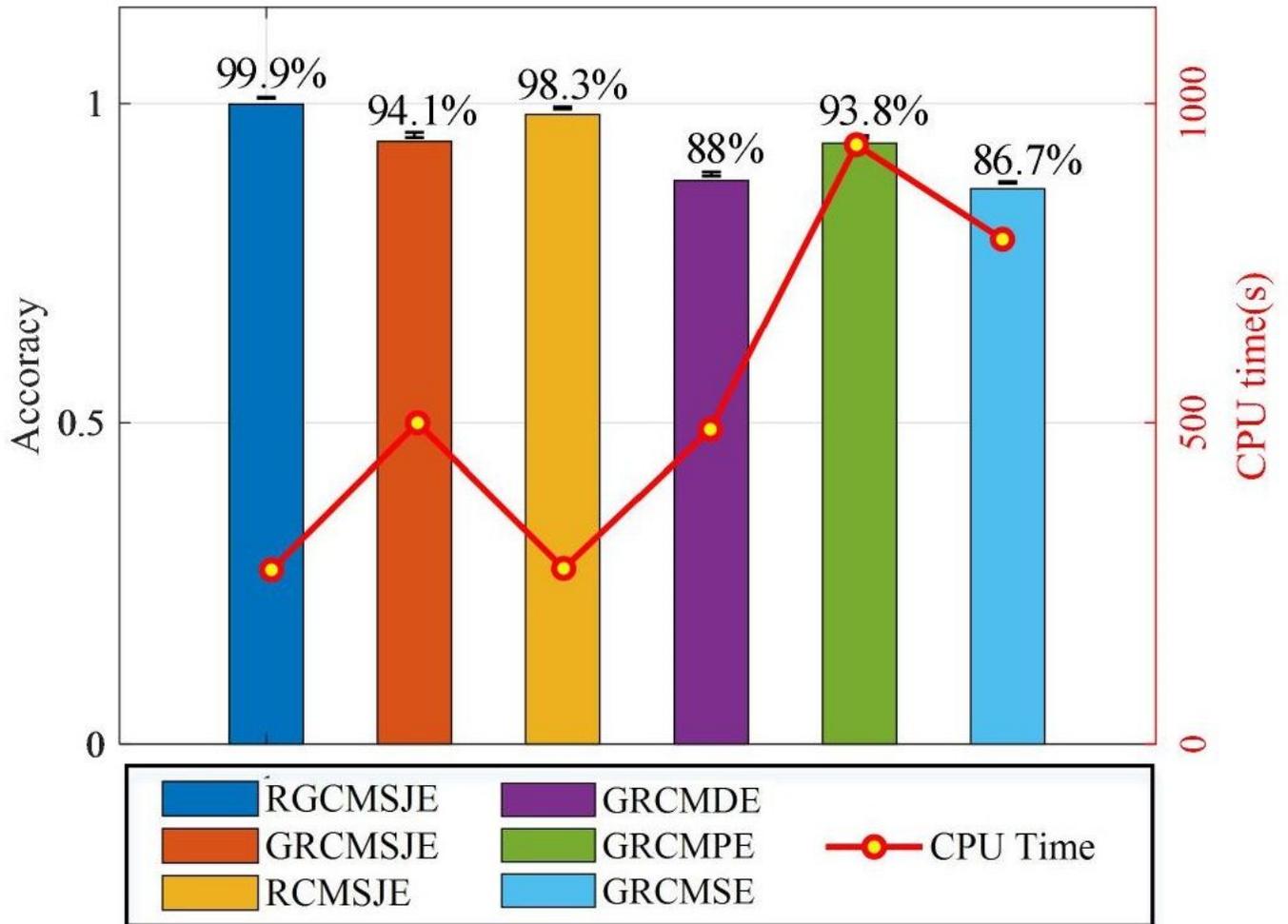


Figure 19

Please see the Manuscript PDF file for the complete figure caption

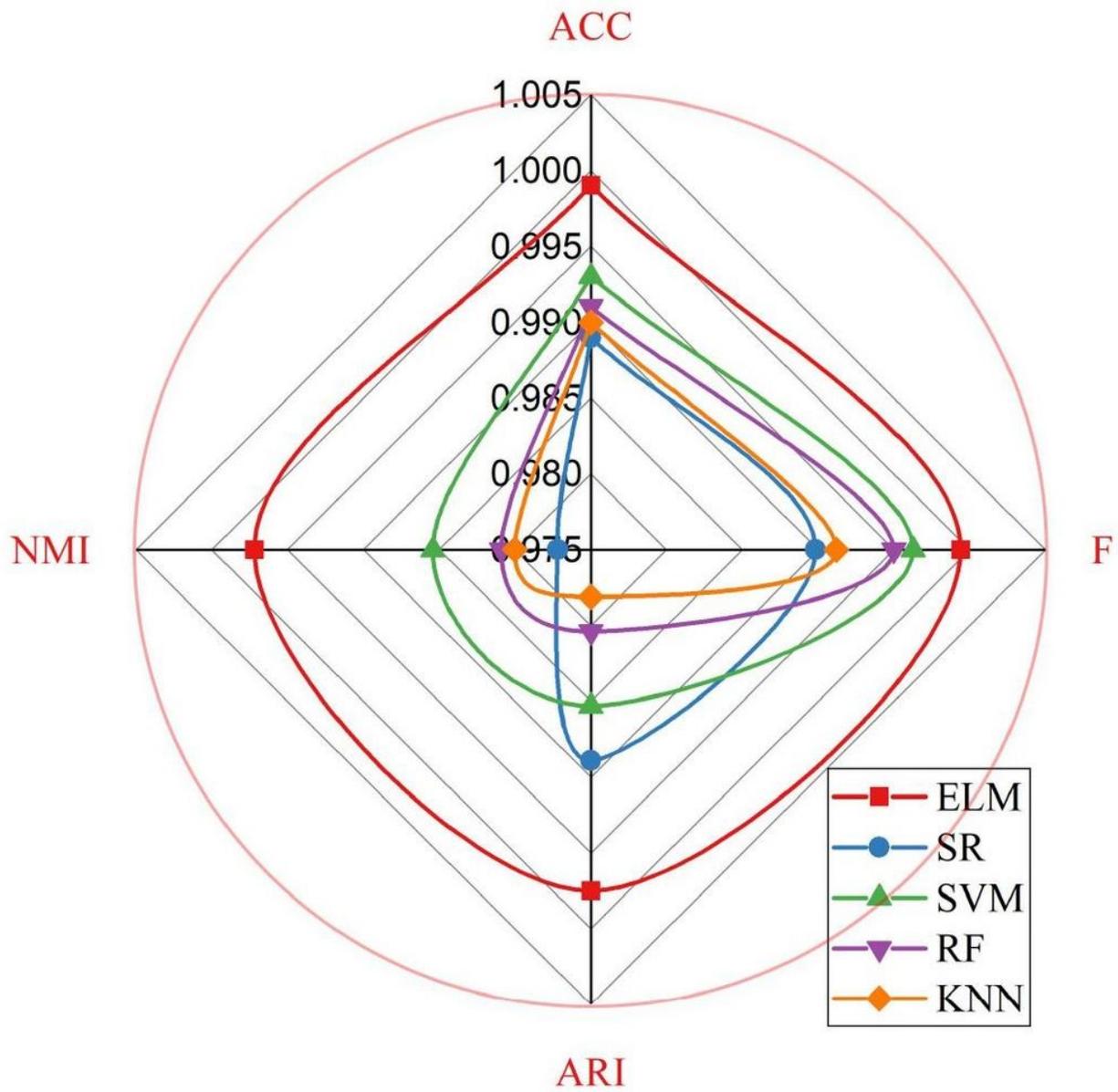


Figure 20

Please see the Manuscript PDF file for the complete figure caption

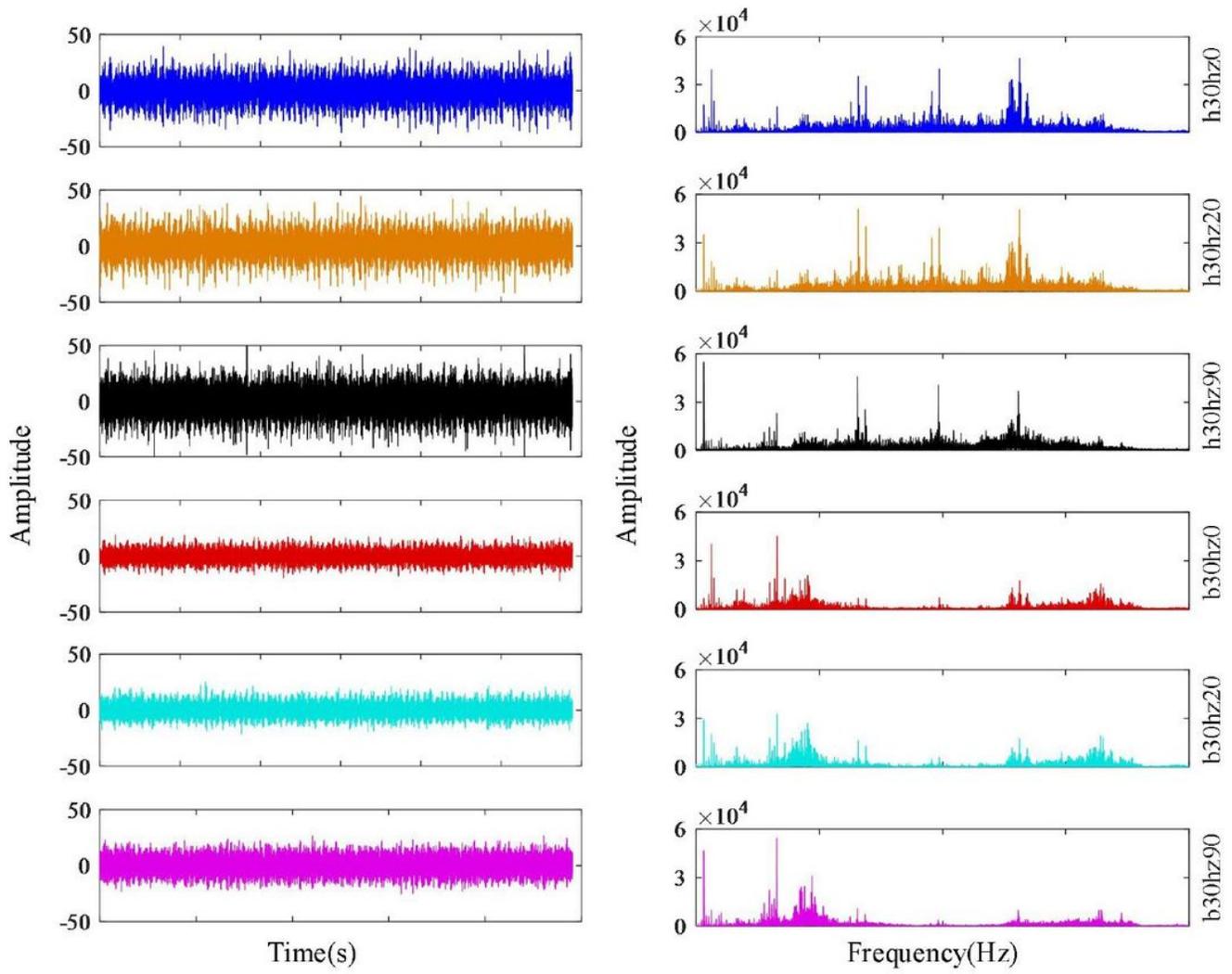


Figure 21

Please see the Manuscript PDF file for the complete figure caption

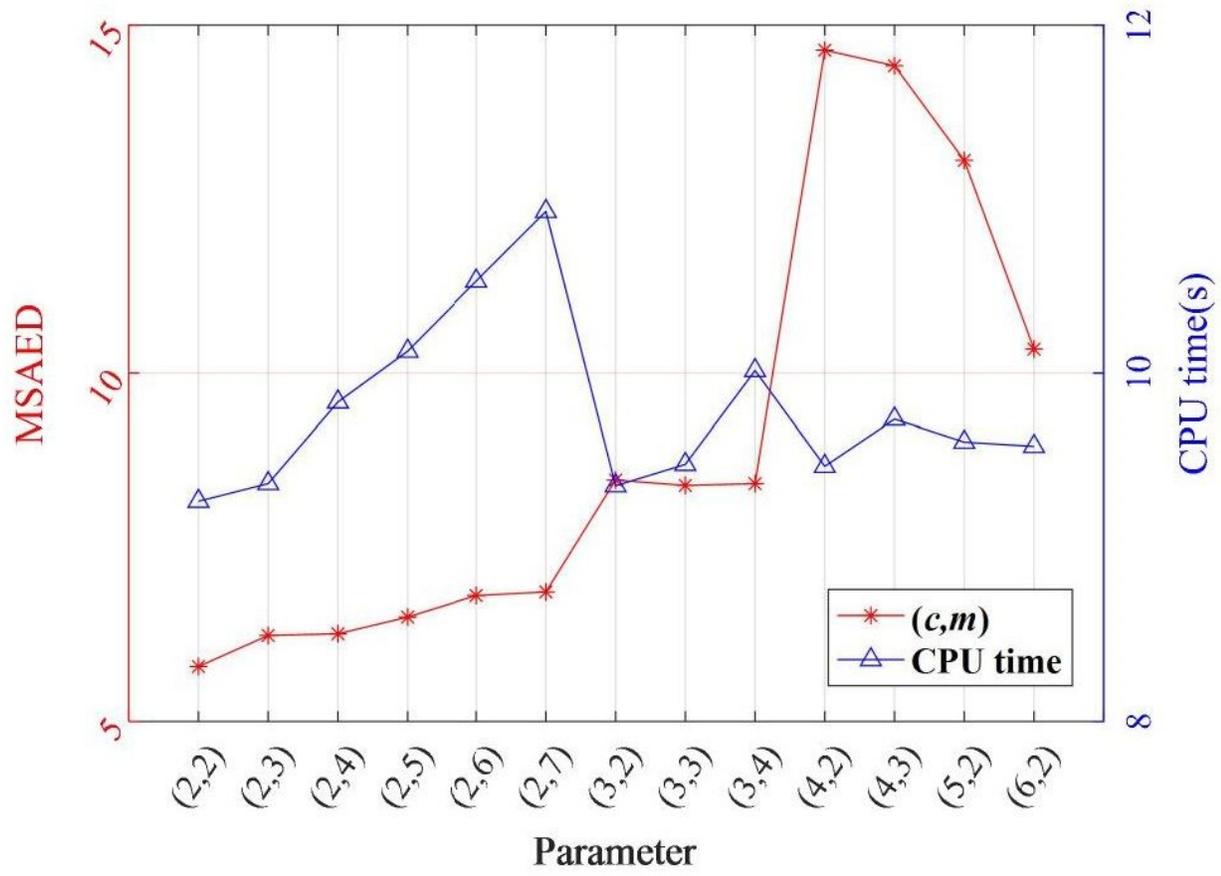


Figure 22

Please see the Manuscript PDF file for the complete figure caption

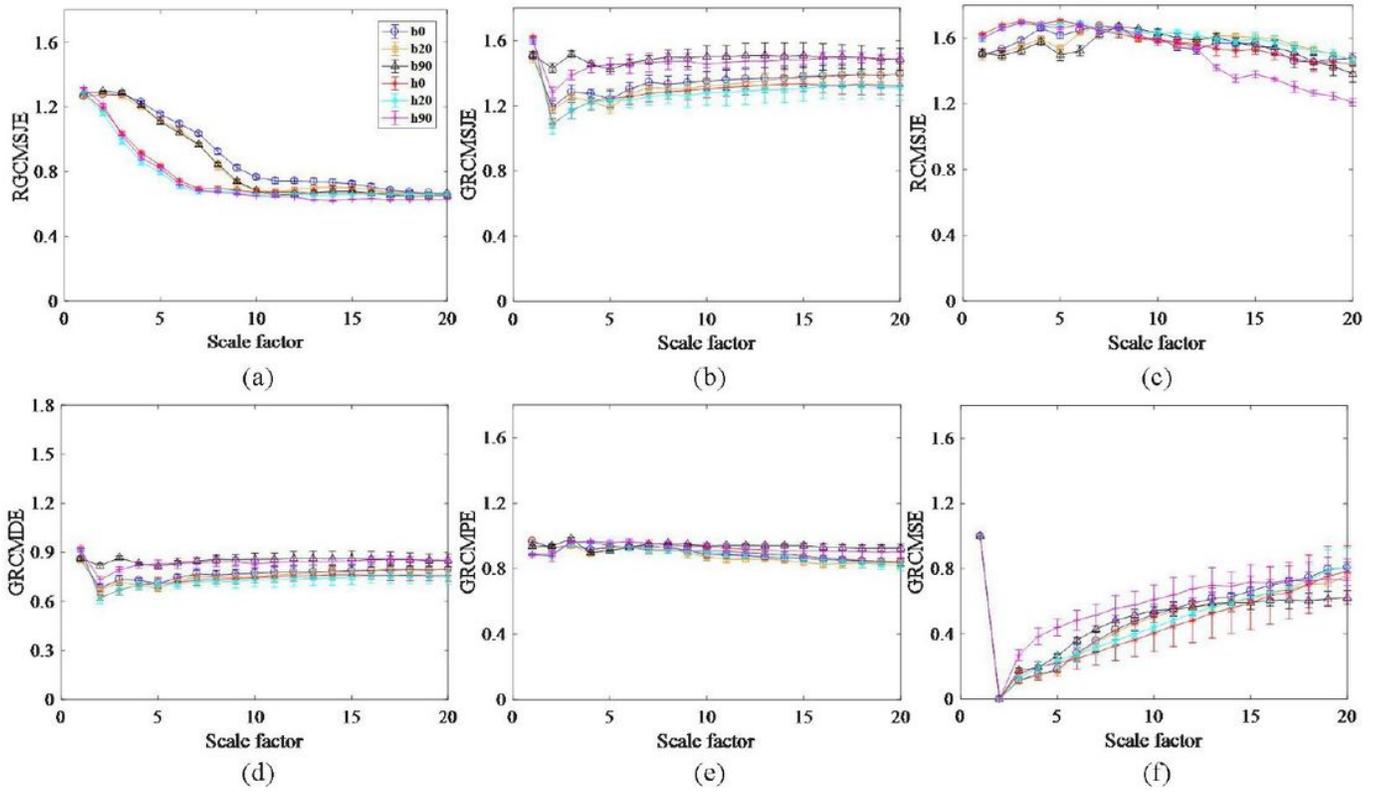


Figure 23

Please see the Manuscript PDF file for the complete figure caption

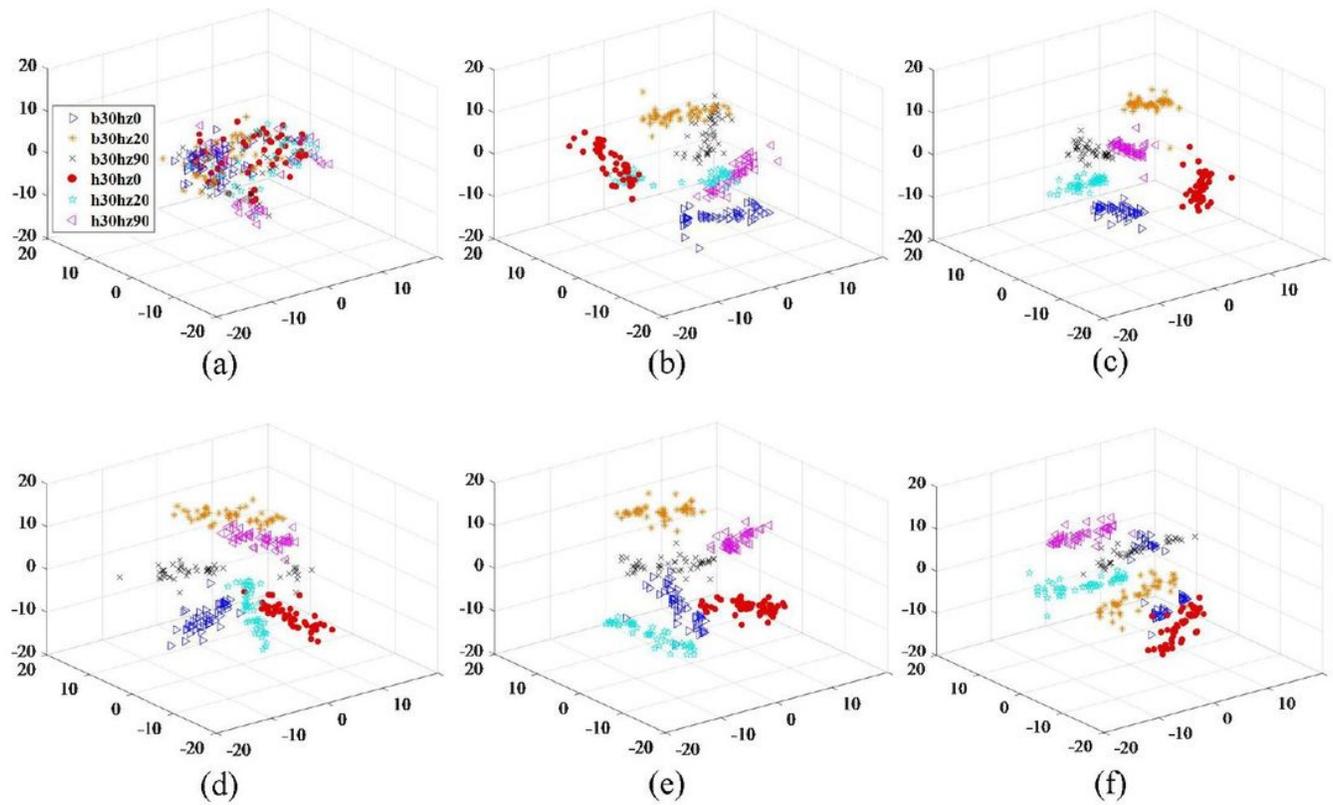


Figure 24

Please see the Manuscript PDF file for the complete figure caption

1	49 16.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	49 16.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	1 0.3%	1 0.3%	50 16.7%	0 0.0%	0 0.0%	0 0.0%	96.2% 3.8%
4	0 0.0%	0 0.0%	0 0.0%	50 16.7%	0 0.0%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	50 16.7%	0 0.0%	100% 0.0%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	50 16.7%	100% 0.0%
	98.0% 2.0%	98.0% 2.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	99.3% 0.7%
	1	2	3	4	5	6	

Figure 25

Please see the Manuscript PDF file for the complete figure caption

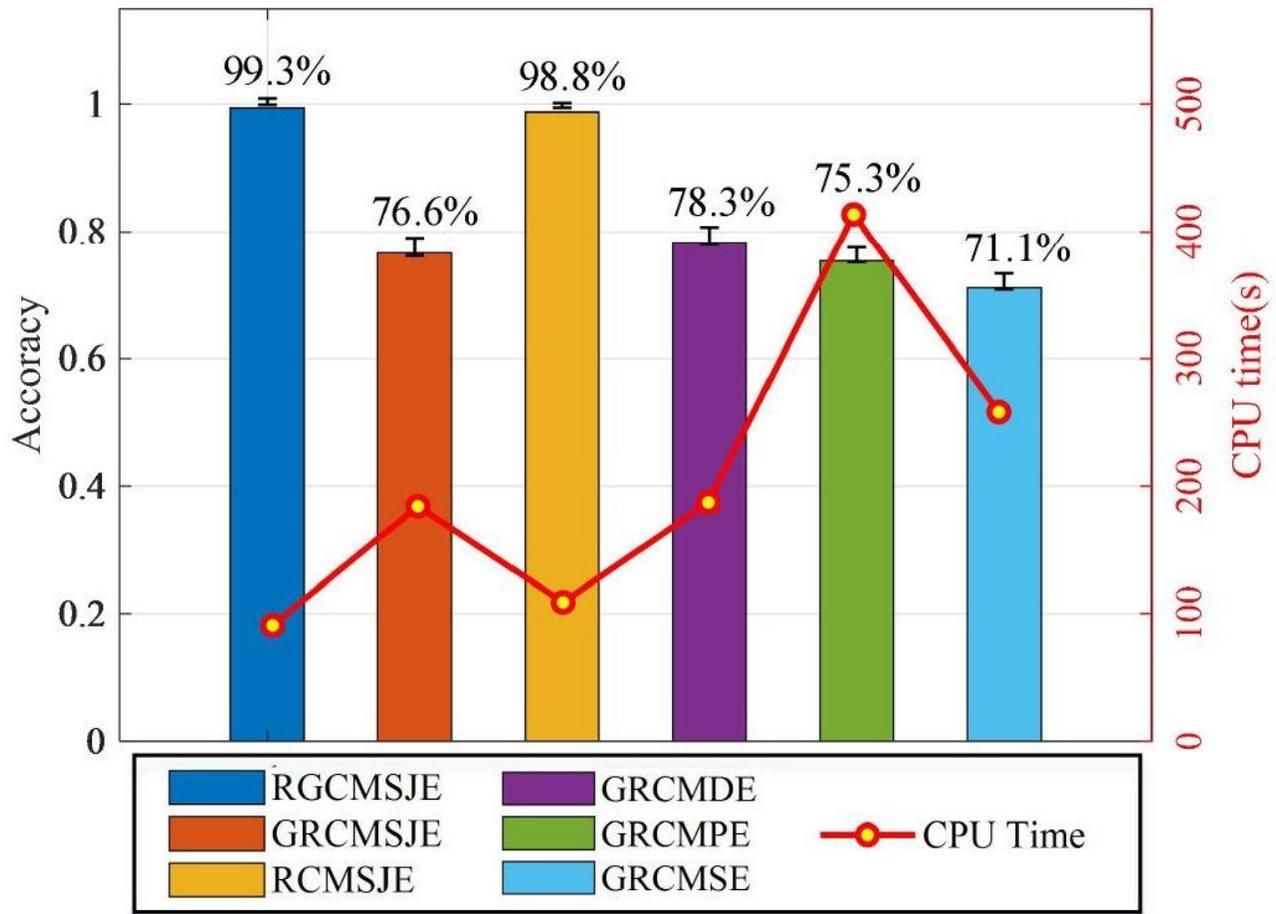


Figure 26

Please see the Manuscript PDF file for the complete figure caption

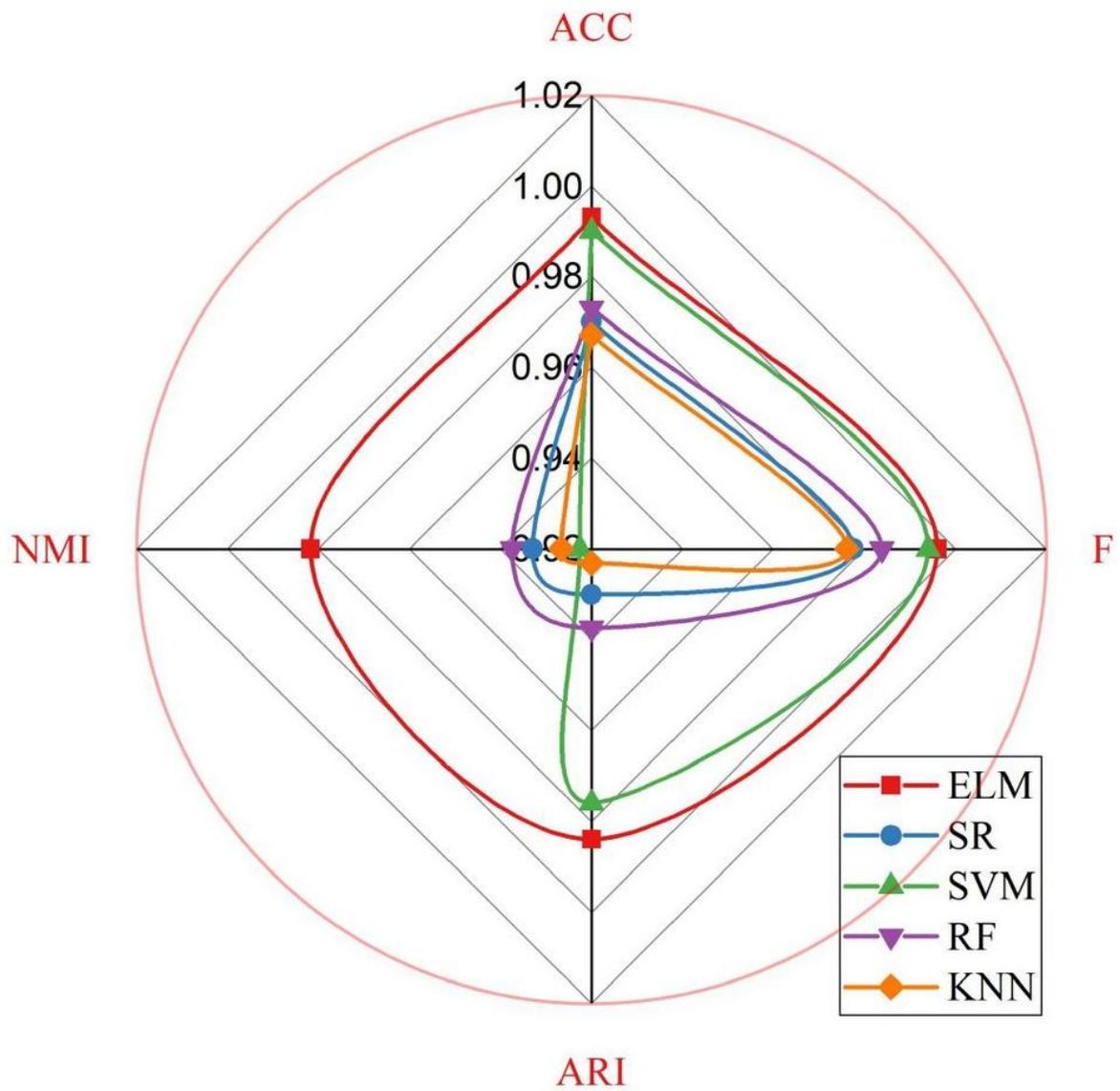


Figure 27

Please see the Manuscript PDF file for the complete figure caption