



Intelligent fault diagnosis of wind turbine gearboxes based on refined generalized multi-scale state joint entropy and robust spectral feature selection

Wei Dong · Shuqing Zhang · Mengfei Hu · Liguozhang · Haitao Liu

Received: 12 May 2021 / Accepted: 28 October 2021 / Published online: 25 January 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract The fault diagnosis of gearboxes and bearings in wind turbines is crucial to extend their service life and reduce maintenance costs. This paper proposes a novel fault diagnosis method combining the refined generalized composite multi-scale state joint entropy (RGCMSSJE), robust spectral feature selection (RSFS) unsupervised learning framework, and extreme learning machine (ELM). The method enables feature extraction, dimensionality reduction, and pattern recognition to identify the different condition states of gearboxes. In this method, multi-scale average Euclidean divergence is adopted to assist the RGCMSSJE in parameter selection. Then, the RGCMSSJE is utilized to extract multi-scale features from the gearbox vibration signal and

construct a high-dimension feature set. The RSFS method is subsequently used to reduce the dimensionality of the RGCMSSJE feature set. Finally, the reduced low-dimensional features are fed into an ELM classifier for fault pattern recognition. The effectiveness of the proposed fault diagnosis method is verified using two experimental datasets, for which the average accuracy reached 99.9% and 99.3%, respectively. The analysis results show that the method can effectively and accurately identify different fault types in wind turbine gearboxes.

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

W. Dong · S. Zhang (✉) · M. Hu · L. Zhang · H. Liu
School of Electrical Engineering, Key Laboratory of Measurement Technology and Instrumentation of Hebei Province, Engineering Research Center of the Ministry of Education for Intelligent Control System and Intelligent Equipment, Yanshan University, Qinhuangdao 066004, China
e-mail: zhshq-yd@163.com

W. Dong
e-mail: 1962899053@qq.com

M. Hu
e-mail: 2233988692@qq.com

L. Zhang
e-mail: zlgtime@163.com

H. Liu
e-mail: 499137183@qq.com

Abbreviations

AA	Average accuracy
EED	Entropy Euclidean distance
AM-FM	Amplitude modulation and frequency modulation
ANOVA	Analysis of variance
ARI	Adjusted rand index
ARX	Average refined generalized composite multi-scale state joint entropy
SDED	Standard deviation Euclidean distance

AWOA	Adaptive whale optimization algorithm	RSFS	Robust spectral learning framework for unsupervised feature selection
CM	Confusion matrix	SE	Sample entropy
CMSPE	Composite multi-scale permutation entropy	SJE	State joint entropy
CMSWPE	Composite multi-scale weighted permutation entropy	SHLFFNN	Single-hidden layer feed forward neural network
CS	Classified states	SP	Spalling
CVA	Cluster validity assessment	SR	Softmax regression
DE	Dispersion entropy	SVM	Support vector machine
ELM	Extreme learning machine	STFT	Short time Fourier transform
FM	Frequency modulated	GWN	Gaussian white noise
FS	Fisher score	WTGF	Wind turbine gearbox fault
FE	Fuzzy entropy	WT	Wind turbine
GA	Genetic algorithm		
GRCMS	Generalized refined composite multi-scale		
HC	Healthy condition		
KNN	K-nearest neighbor		
LS	Laplacian score		
MSAED	Multi-scale average Euclidean divergence		
MEP	Maximum entropy partition		
MIP	Precision		
MIR	Recall		
MLM	Modified linear mapping		
MRMR	Max-relevance and min-redundancy		
MSD	Multi-scale standard deviations		
MT	Missing tooth		
NCDF	Normal cumulative distribution function		
NMI	Normalized mutual information		
O&M	Operation and maintenance		
ORF	Outer race fault		
PE	Permutation entropy		
RC	Root crack		
RCFOA	Reverse cognitive fruit fly optimization algorithm		
RCHFE	Refined composite hierarchical fuzzy entropy		
RCMSA	Refined composite multi-scale analysis		
RCMSDQC	Refined composite multi-scale dispersion q-complexity		
RF	Random forest		
RGCMSSJE	Refined generalized composite multi-scale state joint entropy		

1 Introduction

As a clean and environment-friendly renewable energy source, 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]. Unfortunately, wind turbine (WT) failures occur frequently because of the complex and harsh environments they operate in. This leads to high operation and maintenance (O&M) costs and significant economic losses. In particular, gearboxes and bearings are some of the most critical components of WTs, which are also most susceptible to failures. [3] Therefore, it is crucial to develop reliable fault diagnosis methodologies for timely and accurate monitoring of the operation status of gearboxes and bearings in order to reduce O&M costs [4, 5]. To that end, health monitoring and identification systems have widely been adopted in engineering practice [6–8].

It is advisable to apply the detection and identification technologies which have been widely used in other fields to WTs [9]. The analysis of gearbox condition usually uses vibration signals, which can be more easily acquired and processed than photoelectric, acoustic emission, temperature and other types of signals [10]. However, due to the instability of wind speed and strong background noise, the vibration signals are usually nonlinear, non-stationary, and noisy. The weak characteristics of vibration signals usable for fault identification are difficult to extract, especially in different working conditions [11, 12].

Nevertheless, various useful approaches have been proposed utilizing data-driven and machine learning methods. These schemes comprise two main steps: feature extraction and pattern recognition [13], where the former step significantly affects the latter [14]. Traditional time–frequency analysis methods require advanced knowledge and detailed information of mechanical components, which greatly limits their applications [15–18]. Recently, with the development of nonlinear methods, feature extraction based on entropy has gradually developed into a promising alternative for WT fault diagnosis [19, 20].

Entropy is a measure of the dynamic characteristics of time series. Various types of entropy, including sample entropy (SE) [21], permutation entropy (PE) [22], fuzzy entropy (FE) [23], and dispersion entropy (DE) [24], have been widely applied to damage feature extraction for monitoring gearboxes and bearings. Among them, the time consumption of SE is large, especially when dealing with long time series [25]. PE is markedly faster than SE, but it ignores the time series amplitude information. The feature extraction ability of FE is limited by the fuzzy membership function it adopts [26]. On the other hand, DE has exhibits significantly higher computational efficiency compared to SE, PE, and FE, and its entropy estimation is stable and effective. This can be attributed to the symbolic mapping based on statistics and the probability distribution based on embedding pattern adopted in the DE algorithm [27]. However, all these entropy algorithms ignore the dynamic characteristics of time series, i.e., the transition probability of data from the current state to the next state. Therefore, harnessing the advantages of DE, we propose a new entropy estimation algorithm called state joint entropy (SJE). The algorithm considers the joint probability distribution of the current state and the subsequent state. Analyses of numerically simulated and experimental signals showed that SJE not only inherits the efficiency and stability of the DE algorithm, but can also extract richer fault information. Moreover, a large number of studies demonstrated that these entropy algorithms only estimate the irregularity and complexity of the signal at a single time scale and hence some important fault information at other scales will be missed [28, 29]. In order to extract fault features at multiple time scales, refined composite multi-scale analysis (RCMSA) was proposed by Azami et al. [30]. It not only addresses the problem of undefined entropy

in the original composite multi-scale analysis, but also increases the stability of results for long time series. However, the RCMSA mainly considers the average process, which means that with the increase in scale factor the variance of entropy values will increase rapidly and become statistically unstable. Therefore, this paper proposes a refined generalized composite multi-scale analysis (RGCMSA) to ameliorate the above shortcomings by using the mean and variance of the whole time series. Finally, this paper combines the SJE and RGCMSA into the refined generalized composite multi-scale state joint entropy (RGCMSSE) method, which is used for feature extraction for monitoring gearboxes and bearings.

It is well known that after extracting high-dimensional multi-scale features it is usually necessary to reduce the computational burden. At present, the Laplacian score (LS) [31], Fisher score (FS) [32], and max-relevance and min-redundancy (MRMR) [33] have been widely used for feature selection in various fields. However, the LS only focuses on the similarity of adjacent samples and ignores the global information separation. On the other hand, the FS has the opposite characteristics, i.e., it focuses on the global separation of samples and does not consider the similarity of adjacent samples [34]. Therefore, the features selected by the FS and LS cannot effectively represent the separability of multi-class samples. The MRMR approach is based on the principle of maximizing the between-class distance and minimizing the within-class distance, but it is computationally demanding and not suitable for fast calculations. In order to address these problems, this paper introduces the robust spectral feature selection (RSFS) method into the feature selection process. This method uses a robust local learning method to address the cluster assignment error to improve the local information retention, while simultaneously taking into account the global information [35]. Two experimental cases show that the multi-class features extracted by RSFS are more separable compared to methods that do not use RSFS, while computational efficiency of feature selection is ensured.

The next key step is to design a multi-classifier for intelligent fault identification in WT gearboxes [36]. Peña et al. [37] used k-nearest neighbor (KNN) for bearing fault diagnosis, and analysis of variance (ANOVA) and cluster validity assessment (CVA) for feature extraction. In Zheng [38], the softmax

regression (SR) algorithm was used to identify the fault features of rotating machinery extracted by the short time Fourier transform (STFT). Zheng et al. [39] adopted the extreme learning machine (ELM) to identify the fault features of rotating machinery extracted by composite multi-scale weighted permutation entropy (CMSWPE) method. Li et al. [40], used a genetic algorithm optimized support vector machine (GA-SVM) method to diagnose fault of rotating machinery, and refined composite multi-scale Lempel–Ziv algorithm for feature extraction. Wei [41] proposed the random forest (RF) and the refined composite hierarchical fuzzy entropy (RCHFE) method for intelligent identification and extraction of features. However, after conducting a comparison we concluded that the CPU running times of the SVM and RF algorithms were excessive, and the accuracy of the SR and KNN algorithms was low. The ELM algorithm is superior in terms of both computational time and accuracy. It not only has a higher generalization ability compared to the SR, SVM, RF and KNN algorithms, but also requires less manual intervention [42, 43]. Hence, we adopted the ELM in this study for the identification of different fault types in gearboxes and bearings. Table 1 provides a detailed comparison of the soft computing and machine learning methods discussed in this section.

The contributions of this paper can be summarized as follows:

1. The SJE method is proposed to extract the current and the subsequent states of the time series, which enables extracting more fault features while maintaining the efficiency of DE algorithm.
2. The RGCMSJE method is proposed to extract fault features at multiple time scales, which addresses the deficiency of the RCMSA. Multi-scale average Euclidean divergence (MSAED) is proposed for automatic selection of the parameters for the RGCMSJE method.
3. The RSFS is introduced for the first time to select fault sensitive features with higher distinguishability as the input for the ELM intelligent classifier.
4. The RGCMSJE-RSFS-ELM fault diagnosis method is systematically developed, and its effectiveness is enhanced using MSAED parameter selection and verified through comparative studies using experimental data.

The reminder of the paper is organized as follows. Section 2 introduces the theory of SJE and discusses its performance, such as computational efficiency, stability, and information richness of extracted features. Section 3 describes in detail the RGCMSJE and MSAED procedural steps. Section 4 introduces the

Table 1 Literature-based comparison between soft computing and machine learning methods

Study	Applications	Classifiers and variables	CS	AA (%)	Advantages	Limitations
[37]	Rolling bearings	KNN and ANOVA, CVA	10	100	Richer feature information	Tedious parameter debugging
[38]	Rotary machinery	SR and STFT	10	91.1	Simple structure	Low accuracy
[39]	Rolling bearings	ELM and CMSWPE	7	99.04	Fast calculation speed and high accuracy	Undefined entropy appears
[40]	Rotary machinery	RCMS-Lempel–Ziv and GA-SVM	9	98.87	Richer feature information	Long computational time
[41]	Planetary gearboxes	RCHFE and RF	6	100	High accuracy	Long computational time
[42]	Rolling bearings	RCFOA-ELM and CMPE	10	98.34	High accuracy	Long computational time
[43]	Rolling bearings	AWOA-ELM and RCMSDQC	6	99.1	High accuracy and richer feature information	Long computational time

Peña et al. [37], Zhang et al. [38], Zheng et al. [39], Li et al. [40], Wei et al. [41], He et al. [42], Dong et al. [43]

CS classified states, AA average accuracy, RCFOA reverse cognitive fruit fly optimization algorithm, AWOA adaptive whale optimization algorithm, RCMSDQC refined composite multi-scale dispersion q-complexity

RSFS of fault sensitive features. Section 5 explains the ELM algorithm and the procedural steps of the proposed RGCMSSJE-RSFS-ELM method. Section 6 introduces two experiments and discusses the results of the application of the proposed method to the experimental data. Finally, conclusions are drawn in Sect. 7.

2 State joint entropy

Table 2 provides the control variables and their roles and values discussed in this section.

2.1 State joint entropy calculation procedure

For a given univariate discrete time series of length N , denoted by $x_i(i = 1, 2, \dots, N)$, the SJE algorithm comprises the following six main steps:

1. Mapping from the original time series to a symbolic series. First, the range of the original time series is divided into c categories, labeled from 1 to c . Each value in the time series is mapped into a unique category. Then the normal cumulative distribution function (NCDF) is employed to map each x_i into set $\varepsilon_i = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N\}$, where $0 \leq \varepsilon_i \leq 1$. Finally, the modified linear mapping (MLM) is employed to map ε_i to z_i^c that ranges from 1 to c . The MLM mapping process comes from the DE, which eliminates the undefined mapping by adding constraints as follows [44]:

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

where z_i^c represents the i -th member of the classified time series.

2. Constructing the embedding and state vectors from the symbolic series $z_i^c(i = 1, 2, \dots, N)$ as follows:

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

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

where $\mathbf{z}_k^{m,c}$ are the embedding vectors and $\mathbf{z}_k^{T(m,c)}$ are the state vectors, m is the embedding dimension, and d is the time delay. The dispersion patterns of $\mathbf{z}_k^{m,c}$ and $\mathbf{z}_k^{T(m,c)}$ are denoted as π_{V_β} and $\pi_{V_\beta}^T$, respectively, where $\beta = 1, 2, \dots, N - m$. The number of possible dispersion patterns is c^m , because $\mathbf{z}_k^{m,c}$ or $\mathbf{z}_k^{T(m,c)}$ have m elements and each element may have c possible values.

3. The corresponding state value from embedded vector $\mathbf{z}_k^{m,c}$ to state vector $\mathbf{z}_k^{T(m,c)}$ is denoted as $q_\beta^{a,m,d}(a = 1, 2, \dots, c)$. Therefore, we can 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:

$$p(\mathbf{z}_k^{T(m,c)}) = \frac{\text{Number}\{k | k \leq N - (m - 1)d, \mathbf{z}_k^{T(m,c)} \text{ has type}(\pi_{V_\beta}^T)\}}{N - (m - 1)d - 1} \quad (3)$$

where $\text{type}(\cdot)$ represents the type of the dispersion pattern.

Table 2 Roles and values of control variables

Control variables	Role	Value
c	Dividing time series range	[2, 8]
m	Defining embedded vector dimension	[2, 8]
d	Limiting time series dispersion	1
A_0	Controlling signal amplitude	3
B	Controlling signal damping coefficient	0.1
τ_i	Controlling simulated signal time lag	0
f_0	Limiting failure frequency	124 Hz
f_s	Limiting sampling frequency	12 kHz
L	Controlling data length	10,000
T	Limiting sampling interval	0.01
M	Limiting signal cycles	6

- Constructing a state values matrix, $Q_\beta^{a,m,d}$, according to the state values $q_\beta^{a,m,d}$ as follows:

$$Q_\beta^{a,m,d} = \begin{bmatrix} q_1^{1,m,d} & q_1^{2,m,d} & \dots & q_1^{c^m,d} \\ \vdots & \vdots & & \vdots \\ q_r^{1,m,d} & q_r^{2,m,d} & \dots & q_r^{c^m,d} \\ \vdots & \vdots & & \vdots \\ q_{c^m}^{1,m,d} & q_{c^m}^{2,m,d} & \dots & q_{c^m}^{c^m,d} \end{bmatrix} \tag{4}$$

where $1 < r < c^m$. Then, the state transition matrix conditional on $Q_\beta^{a,m,d}$ can be obtained as follows:

$$T_\beta^{a,m,d} = \begin{bmatrix} (q_1^{1,m,d} | \pi_{V_1}^T) & (q_1^{2,m,d} | \pi_{V_1}^T) & \dots & (q_1^{c^m,d} | \pi_{V_1}^T) \\ \vdots & \vdots & & \vdots \\ (q_r^{1,m,d} | \pi_{V_r}^T) & (q_r^{2,m,d} | \pi_{V_r}^T) & \dots & (q_r^{c^m,d} | \pi_{V_r}^T) \\ \vdots & \vdots & & \vdots \\ (q_{c^m}^{1,m,d} | \pi_{V_{c^m}}^T) & (q_{c^m}^{2,m,d} | \pi_{V_{c^m}}^T) & \dots & (q_{c^m}^{c^m,d} | \pi_{V_{c^m}}^T) \end{bmatrix} \tag{5}$$

where $\pi_{V_r}^T$ is the r -th dispersion pattern. The state transition matrix describes how the symbolic time series changes from one state to another with time.

- Calculating the probability of the state transition matrix, as follows:

$$P(q_\beta^{a,m,d} | \pi_{V_\beta}^T) = \frac{\text{Number}\{k | k \leq N - (m - 1)d, (q_\beta^{a,m,d} | \pi_{V_\beta}^T) \text{ has type } (\pi_{V_\beta}^T)\}}{\text{sum}(\text{Number})} \tag{6}$$

- Using the formulas for joint entropy and conditional entropy [45], the following can be obtained:

$$H(\mathbf{z}_k^{m,c}, q_\beta^{a,m,d}) = H(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)}) + H(\mathbf{z}_k^{m,c}) \tag{7}$$

Finally, the value of SJE is calculated as follows:

$$\begin{aligned} \text{SJE}(x, m, d, c) &= - \sum_{k=1}^{c^m} P(\mathbf{z}_k^{T(m,c)}, q_\beta^{a,m,d}) \\ &\quad - \sum_{k=1}^{c^m} p(\mathbf{z}_k^{m,c}) \ln(p(\mathbf{z}_k^{m,c})) \end{aligned} \tag{8}$$

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

$$P(\mathbf{z}_k^{T(m,c)}, q_\beta^{a,m,d}) = P(\mathbf{z}_k^{T(m,c)}) P(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)}) \ln(P(\mathbf{z}_k^{T(m,c)}) P(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)})) \tag{9}$$

$$P(\mathbf{z}_k^{m,c}) = \frac{\text{Number}\{k | k \leq N - (m - 1)d, \mathbf{z}_k^{m,c} \text{ has type } (\pi_{V_\beta})\}}{N - (m - 1)d - 1} \tag{10}$$

A flowchart of the SJE calculation procedure is shown in Fig. 1.

In order to explain the details of the calculation process of SJE, an illustrative example is given. Consider time series $\{x\} = \{1.2, -2.1, -0.5, 0.4, -1.7, 1.4, 3.6, 1, 0.9, -1, -2.5, 2\}$ with $d = 1$, $m = 2$, and $c = 2$. First, the dispersion space ε_i is obtained according to the NCDF algorithm. Then the symbolized time series $z_i^2 = \{2, 1, 1, 2, 1, 2, 2, 2, 1, 1, 2\}$ can be obtained using Eq. (1). There are $c^m = 2^2 = 4$ possible dispersion patterns, i.e. $(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$. Then the embedded and the state vectors can be obtained as follows:

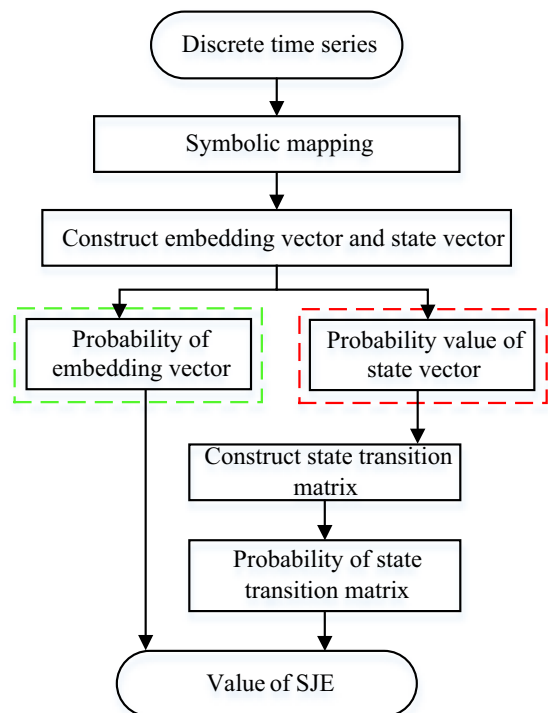


Fig. 1 Flow chart of SJE calculation procedure

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

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

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

$$\begin{aligned} & \left[p(\mathbf{z}^{T(1,1)}), p(\mathbf{z}^{T(1,2)}), p(\mathbf{z}^{T(2,1)}), p(\mathbf{z}^{T(2,2)}) \right] \\ &= \left[\frac{2}{10}, \frac{3}{10}, \frac{2}{10}, \frac{3}{10} \right] \end{aligned} \tag{11}$$

The probability matrix of the state transition matrix can be obtained using Eq. (6) as follows:

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

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

$$\begin{aligned} \text{SJE}(x, m, d, c) &= - \sum_{k=1}^{c^m} P(\mathbf{z}_k^{T(m,c)}, q_\beta^{a,m,d}) \\ &\quad - \sum_{k=1}^{c^m} p(\mathbf{z}_k^{m,c}) \cdot \ln(p(\mathbf{z}_k^{m,c})) \\ &= 3.2585 \end{aligned} \tag{13}$$

From Eqs. (7) and (8), we can find that the total number of all possible dispersion modes of the

embedded and state vectors are c^m and c^{m+1} , respectively. The normalized SJE is calculated as follows:

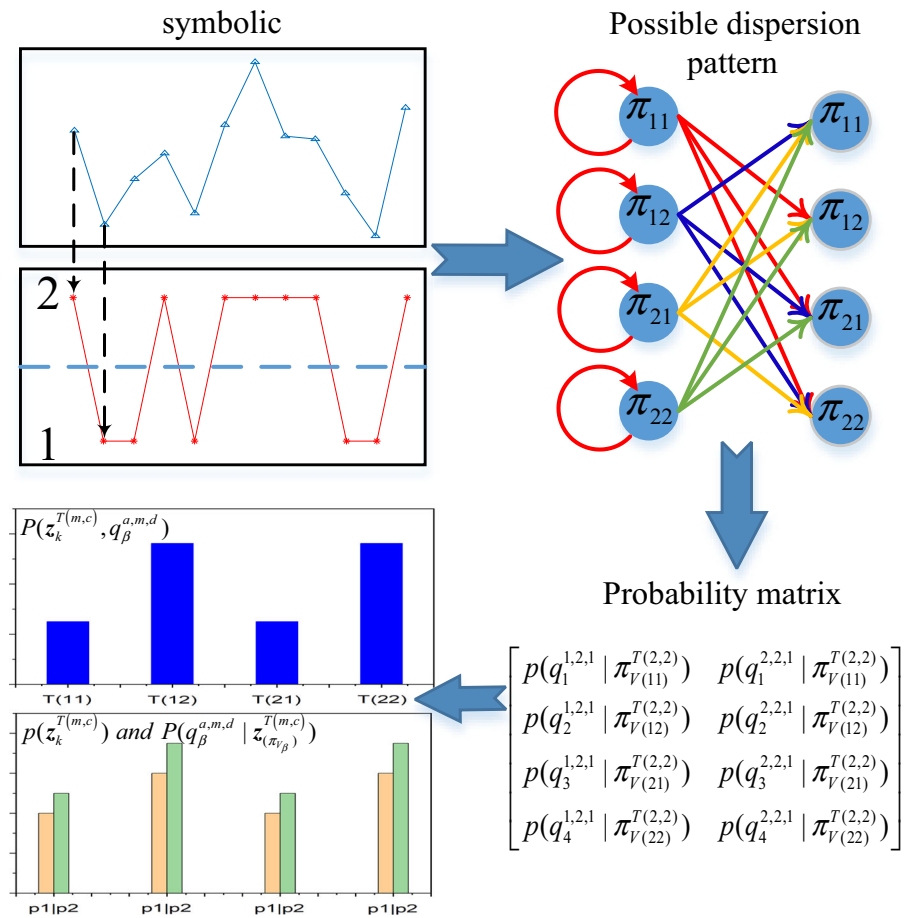
$$\begin{aligned} \text{NSJE}(x, m, d, c) &= \frac{H(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)})}{\ln(\varepsilon^{m+1})} + \frac{H(\mathbf{z}_k^{m,c})}{\ln(\varepsilon^m)} \\ &= 1.8955 \end{aligned} \tag{14}$$

According to Eq. (8), the greater the irregularity of the time series, the bigger the value of its SJE. In order to further illuminate the above example, it is illustrated in Fig. 2, where the process of obtaining the values of $P(\mathbf{z}_k^{T(m,c)}, q_\beta^{a,m,d})$, $p(\mathbf{z}_k^{T(m,c)})$, and $p(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)})$ is shown. Figure 3 shows all possible dispersion patterns of the embedded vectors and state vectors. The number of state dispersion patterns is $c^{m+1} = 8$, and each of them is presented in Fig. 3b, where blue dots represent the original state values, red dots the newly emerged state values, and red lines the newly emerged state dispersion mode, respectively.

2.2 Symbolization performance comparison

To explore the performance of the MLM algorithm for mapping time series, a numerical fault signal model of rolling bearing from Ref. [46] was employed to simulate the outer race fault (ORF). The model signal was symbolized using a mapping algorithm [47, 48], and then the time-domain and frequency-domain analyses were performed on the symbolized sequence. Finally, the ability of different mapping algorithms to retain the original time signal information was compared using time-domain and frequency-domain

Fig. 2 Example of SJE calculations for symbol numbers $c = 2$, embedding number $m = 2$, and time delay $d = 1$



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

$$\begin{cases} x(t) = \sum_{i=1}^M A_i s(t - iT - \tau_i) + w(t) \\ A_i = A_0 \cos(2\pi Q t + \phi_A) + C_A \\ s(t) = e^{-Bt} \sin(2\pi f_n t + \phi_w) \end{cases} \quad (15)$$

where $s(t)$ is a harmonic function, $T = 0.01$ is the interval time, A_i is the amplitude modulation signal with a frequency $Q = 33$ Hz, M is the number of signal harmonics ($M = 6$), A_0 is the amplitude of the modulation signal ($A_0 = 3$), C_A is a constant $C_A > A_0$ ($C_A = 4$), B is the signal exponential decay coefficient ($B = 0.1$), f_n is the natural frequency of the system ($f_n = 3000$ Hz), τ_i is the time lag ($\tau_i = 0$), $w(t)$ is a Gaussian white noise (GWN) (which was ignored in our simulations), and ϕ_A and ϕ_w are the initial phases of A_i and $s(t)$, respectively. The other parameters of the problem are set as follows: the fault frequency of outer race f_0 is 124 Hz, the sampling frequency

$f_s = 12$ kHz, and the data length used for analysis is 10,000 samples.

The time-domain waveform of the simulation series and the symbolic series obtained by the MLM and MEP algorithms are shown in Fig. 4. For the comparative analysis in this section, without loss of generality, the following parameters of the MLM and MEP algorithms were used: number of symbol categories $c = 3$, embedding dimension $m = 3$, and time delay $d = 1$. As shown in Fig. 4, the MLM algorithms captured more details of the bearing failure simulation series than the MEP algorithms. The MEP algorithm mapped too many data points to the maximum symbol category 3, resulting in redundant amplitude information suppression of the original series periodicity. It should be noted that the symbolic sequence obtained by the MLM algorithm not only retained the amplitude information, but also protected the periodicity of the signal.

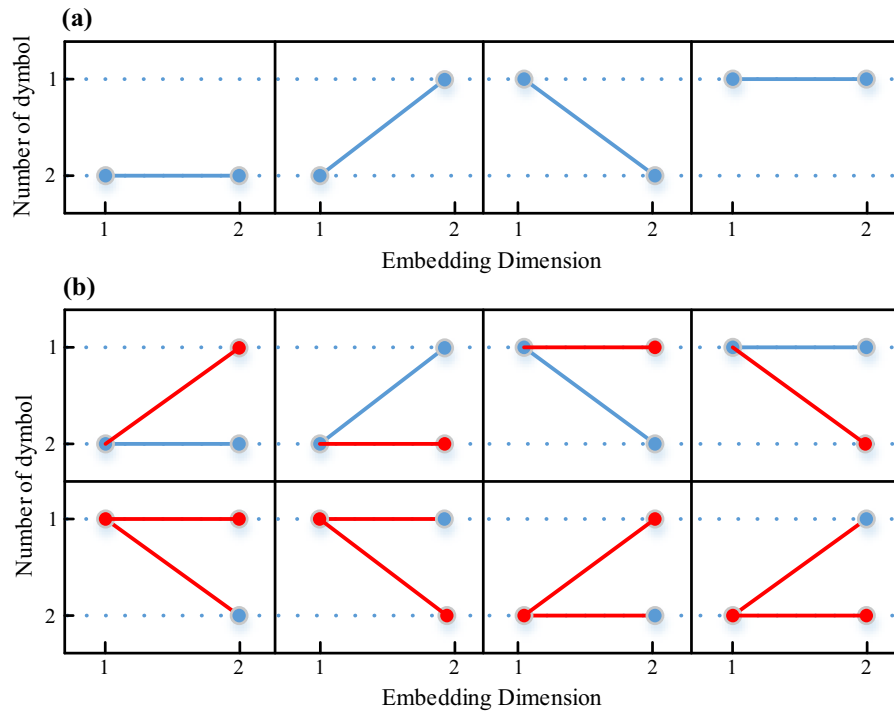


Fig. 3 Illustration of dispersion patterns: **a** Possible dispersion patterns of embedded vectors. **b** Possible dispersion patterns of state vectors

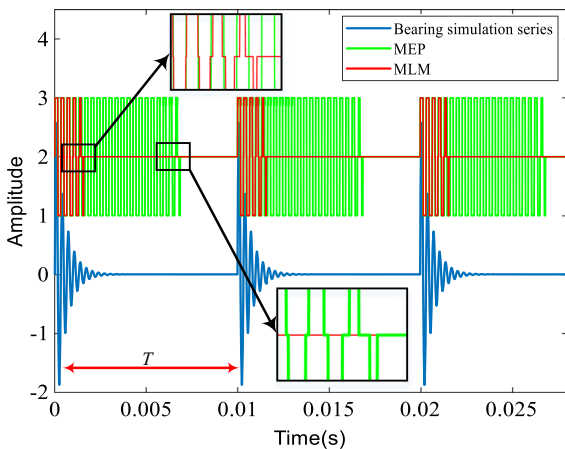


Fig. 4 Comparison of symbolic capabilities of MLM and MEP recorded at sampling time of 0.025 s from ORF simulation time series

In order to further explore the information contained in the symbolic series obtained by the two mapping algorithms, we calculated their fast Fourier transform and envelope spectra and displayed them in Figs. 5 and 6, respectively. From Fig. 5b, it can be concluded that the frequency spectrum of the symbolic

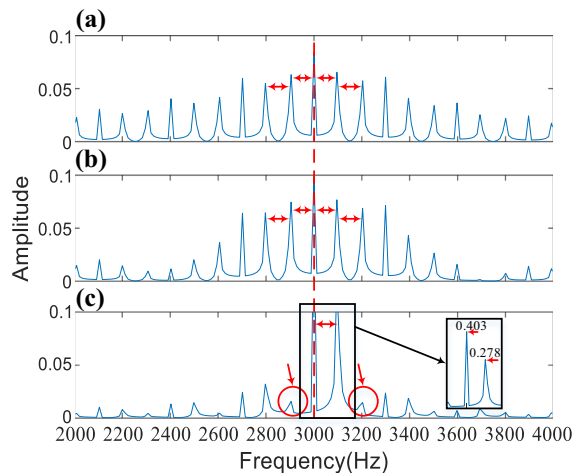


Fig. 5 Fast Fourier Transform spectra: **a** Simulated bearing fault time series. **b** MLM symbolic series **c** MEP symbolic series

sequence obtained by the MLM algorithm was almost the same as the frequency spectrum of the simulated bearing fault time series, and the natural frequency and its sidebands were clearly visible. However, in Fig. 5c it can be observed that the spectral amplitude at the natural frequency and the first right-hand sideband of

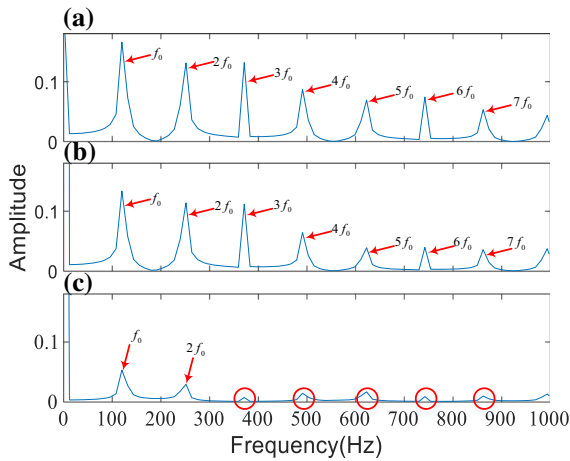


Fig. 6 Envelope spectra: **a** Simulated bearing fault time series. **b** MLM symbolic series **c** MEP symbolic series

the MEP symbolic sequence were enhanced, while the remaining sidebands were strongly attenuated. This was caused by the redundancy and periodicity weakening of the amplitude information. It should be noted that the fault frequency was the sideband interval frequency. The envelope spectra in Fig. 6b and c also confirm these observations. The MLM symbolic sequence retained the fault frequency of 124 Hz, while in the envelope spectrum of the MEF symbolic sequence it was strongly suppressed. Thus, the MLM mapping effectively protected the amplitude and fault information in the original signal and its performance was 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 were analyzed, and their DE, PE, and SE were compared to SJE. All the simulated signals had a length of 360 s and a sampling frequency of 150 Hz. A 12-s long sliding window with 75% overlap was adopted to divide the data into 3-s-long segments.

Without loss of generality, the embedding dimension $m = 3$ and the time delay $d = 1$ were adopted for the four methods. For the SE algorithm, the tolerance was $r = 0.15\sigma$. At the same time, in order to ensure the same symbolic mapping range, the number of categories $c = 3$ was adopted in the SJE and DE algorithms.

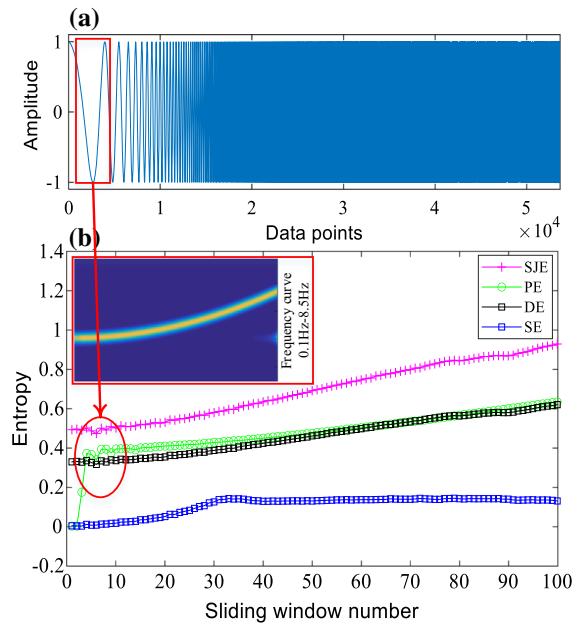


Fig. 7 Comparison of SJE, PE, DE, and SE algorithm performance for constant amplitude FM signal: **a** Constant amplitude FM signal. **b** Entropy curves

To compare the sensitivity of the SJE method for variable to frequency signals, a frequency modulated (FM) signal was used in this study, whose frequency

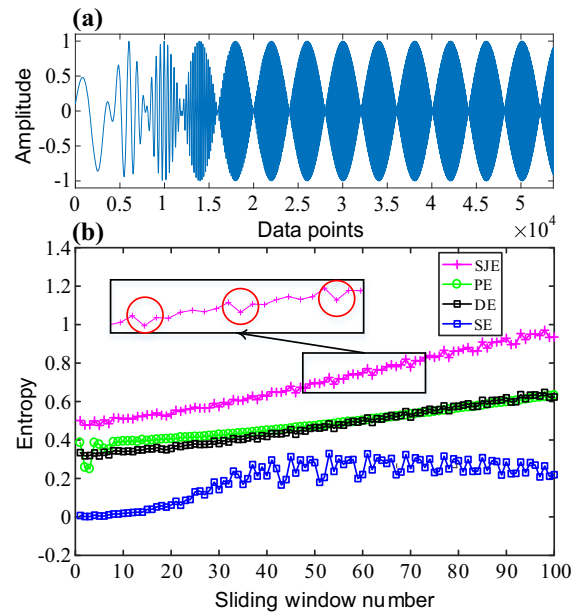


Fig. 8 Comparison of SJE, DE, PE, and SE algorithm performance for AM-FM signal: **a** AM-FM signal. **b** Entropy curves

increased from 0.1 to 8.5 Hz in 360 s. The time-domain waveform and the comparison results of FM signal are shown in Fig. 7. We observed that the SJE had a highest accuracy in detecting frequency changes compared to PE, DE, and SE. The SE curve tended to be stable beyond the 35th sliding window, which indicated that SE did not detect the frequency changes with time. Although the PE and DE detected the frequency changes, the PE plot fluctuated significantly at the beginning, and the growth rate of the two curves was smaller than that of SJE curve.

To study the sensitivity of the SJE method for variable frequency and amplitude signals, an amplitude and frequency-modulated (AM-FM) signal shown in Fig. 8a was studied. The signal, whose frequency increased nonlinearly from 0.1 to 8.5 Hz in 360 s, was created by using pure sine wave modulation with a frequency of 0.1 Hz. It can be observed that after the 30th sliding window, the SE plot did not show an upward trend and fluctuated noticeably. This indicates that SE did not detect the frequency changes, and its performance was unstable. Although the PE values increased monotonously with frequency, compared to the SJE and DE, it was insensitive to the amplitude change. Both SJE and DE plots showed an

upward trend, but the upward trend and periodic fluctuations of the SJE were more obvious. Thus, the SJE accurately detected the changes in both frequency and amplitude.

To explore the sensitivity of the SJE method to the level of noise, a quasi-periodic signal with different levels of noise shown in Fig. 9a was created. The signal was modulated by two sinusoidal signals with frequencies of 0.5 and 1 Hz, respectively. There was no noise in the first 24 s of the sequence. Then, GWN sequences with gradually increasing noise levels were added to the signal every 12 s. The analysis results are shown in Fig. 9b. The PE value remained constant from the 10th sliding window onward, which indicates that the method did not detect the changes in noise level. The SE plot increased monotonously with the increase in noise level, but it fluctuated fast from the 40th sliding window onward, which indicated that the SE method was vulnerable to noise. It is worth noting that the SJE and DE plots increased steadily with the increase in noise level, which indicates that both SJE and DE methods detected noise level and maintained stable performance.

To investigate the ability of SJE method to detect sudden changes in signal amplitude, a signal comprising a series of impulses added to a GWN sequence was

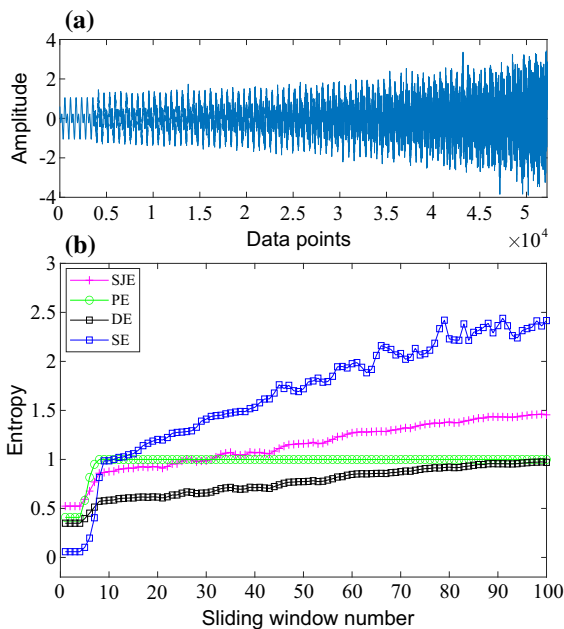


Fig. 9 Comparison of SJE, DE, PE, and SE algorithm performance for quasi-periodic signal with noise: **a** Quasi-periodic signal with noise. **b** Entropy curves

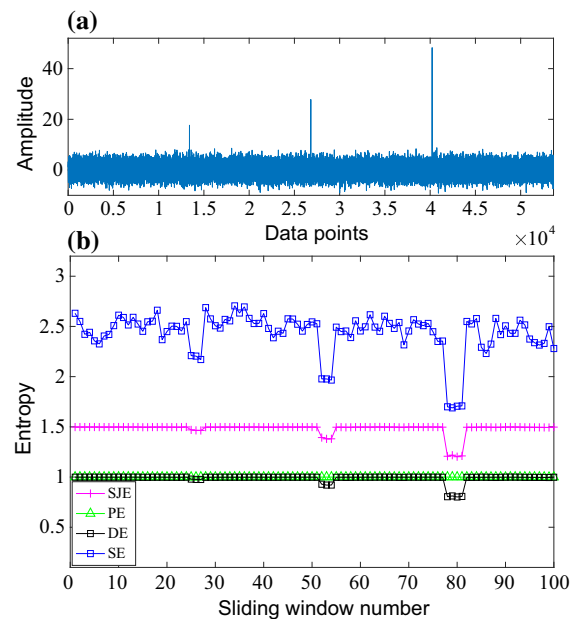


Fig. 10 Comparison of SJE, DE, PE, and SE algorithm performance for GWN signal with impulses: **a** GWN signal with impulses. **b** Entropy values

created, as shown in Fig. 10a. Three impulses with amplitudes of 20, 30, and 50, respectively, were added at 80-s intervals. The entropies are compared in Fig. 10b. It can be observed that the PE values were constant, which indicates that the method did not detect the impulses. The SE method detected the amplitude changes, but its plot fluctuated violently, which indicated an unstable performance of the method. The SJE and DE methods effectively detected the changes in signal amplitude, but the SJE method was more sensitive. These results show that the SJE method not only inherited the stability of DE method, but also had stronger amplitude sensing ability.

To compare the computational complexity of the SJE, DE, SE, and PE methods, we analyzed their CPU times. The hardware used for the simulations comprised a 2.80 GHz CPU (Intel (R) CoreTM i5-8400), a PC with 8 GB RAM, which run Windows 10 (64-bit) operating system and Matlab software version R2016a. The CPU times listed in Table 3 are the averages of 50 runs. It can be seen from Table 3 that the SJE and DE methods consumed very little CPU time, with average CPU times for the four simulation signals in the order of milliseconds. The SE method required the longest CPU times exceeding 13 s, while the PE method consumed more than 1 s. This indicates that the SJE method not only inherited the high efficiency of the DE method, but had also lower computational complexity.

3 Refined generalized composite multi-scale state joint entropy algorithm

Table 4 provides the control variables and their roles and values discussed in this section.

Table 3 CPU times of SJE, DE, SE, and PE methods

Simulated signal	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	0.285	0.234	15.103	2.329
Impulse series	0.307	0.254	15.658	1.042

3.1 Refined generalized composite multi-scale analysis

The RCMSA effectively addresses the problem of entropy uncertainty in multi-scale analysis. However, in the coarse-grained RCMSA, the variance of entropy increases with the scale factor, which leads to the low stability and distinguishability of the feature space. In order to overcome these shortcomings, a new RGCMSSE method is proposed. The detailed steps of the RGCMSSE method are as follows:

- (1) For a given time series $x_i (i = 1, 2, \dots, N)$, in order to obtain a stable feature space, calculate the mean, μ_x , and standard deviation, σ_x , of the time series. 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 follows:

$$y_{k,j}^{(s)} = \frac{1}{s} \sum_{i=k+s(j-1)}^{k+s \cdot j - 1} x_i, \quad 1 \leq j \leq N, \quad 1 \leq k \leq s \tag{16}$$

- (2) Compute the average probability of embedding vector $\bar{P}(\mathbf{z}_k^{m,c})$, average probability of state vector, $\bar{P}(\mathbf{z}_k^{T(m,c)})$, and average probability of state transition matrix, $\bar{P}(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)})$, for all embedding vectors, $\mathbf{z}_k^{m,c}$, and state vectors, $\mathbf{z}_k^{T(m,c)}$, as follows.

$$\bar{P}(\mathbf{z}_k^{m,c}) = \frac{1}{s} \sum_{k=1}^s P(\mathbf{z}_k^{m,c}) \tag{17}$$

$$\bar{P}(\mathbf{z}_k^{T(m,c)}) = \frac{1}{s} \sum_{k=1}^s P(\mathbf{z}_k^{T(m,c)}) \tag{18}$$

$$\bar{P}(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)}) = \frac{1}{s} \sum_{k=1}^s P(q_\beta^{a,m,d} | \mathbf{z}_k^{T(m,c)}) \tag{19}$$

- (3) At scale factor s , the RGCMSSE is defined as the Shannon entropy of the joint state model after shifting the time series as follows:

Table 4 Values and roles of control variables discussed

Control variable	Role	Value
c	Dividing time series range	[2, 8]
m	Defining embedded vector dimension	[2, 8]
d	Limiting time series dispersion	1
s	Limiting multi-scale factor	20
N	Controlling data length	$c^{m+1} \ll N$

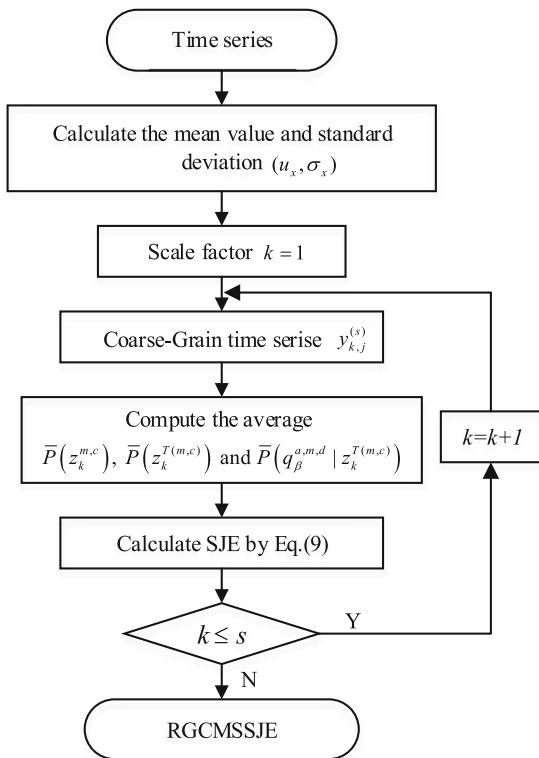


Fig. 11 Flowchart of RGCMSSE calculations

$$\begin{aligned}
 &RGCMSSE(y, m, c, d, \tau) \\
 &= - \sum_{k=1}^c \bar{P}(\mathbf{z}_k^{T(m,c)}, q_{\beta}^{a,m,d}) \\
 &\quad - \sum_{k=1}^c \bar{p}(\mathbf{z}_k^{m,c}) \ln(\bar{p}(\mathbf{z}_k^{m,c})) \quad (20)
 \end{aligned}$$

The calculation process of RGCMSSE is illustrated in Fig. 11.

It is noteworthy that in Ref. [49], the generalized multi-scale processes are extended to second-order statistics by $y_{k,j}^{(s)} = 1/s \sum_{b=(j-1)s+k}^{js+k-1} (x_b - \bar{x}_b)^2$. This approach has proven suitable for the SE and PE,

which are sensitive to the relationship between adjacent amplitudes. However, it was found that this method is not suitable for the SJE, because a significant portion of amplitude information will be lost after adopting the second moment, which will lead to the loss of some useful information [49]. In this paper, the RGCMSSE method is introduced to address these problems. The RGCMSSE method has been improved in two aspects. First, the RCMSA in Step (1) can stably extract the changes in time series at multiple scales and avoid the appearance of undefined entropy. Secondly, in order to reduce the fluctuations of the variance of time series at large scales, Step (1) uses the unified mapping of standard deviation and mean value instead of step-by-step mapping, thus providing the RGCMSSE with a stronger fault feature extraction ability.

3.2 Parameter selection of RGCMSSE

The selection of parameters for an entropy algorithm is very important, because appropriately selected parameters will produce more informative fault features. The traditional entropy parameter selection method is to traverse the parameter space and then observe the entropy curve to determine the parameter values. However, this approach is only qualitative and depends on the analyst's experience. 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 for entropy parameter selection [50], but that method did not consider the multi-scale factors and the stability of feature space. In the multi-scale analysis, the scale factor is an important parameter that needs to be selected. At the same time, we found that when the stability of feature space was poor, even if the AED value was large, the correct detection rate of gearbox operation state was greatly reduced. Therefore, a new index for the selection of multi-scale entropy

parameters, namely the MSAED, is proposed in this paper by considering jointly multi-scale analysis, feature space stability, and AED.

There are five parameters in the RGCMSSE calculation process, including sample data length, N , scale factor, s , embedding dimension, m , number of categories, c , and time delay, d . Without loss of generality, in order to simplify the calculation process, the sample data length, N , and time delay, d , are usually set to 2048 and 1, respectively. Suppose that the dataset to be tested has τ different classes, and each class has n samples of length n_l . The detailed steps of MSAED calculation are as follows:

- (1) Initialize parameters (c, m) in RGCMSSE. Set $c \in [2, 8]$ and $m \in [2, 8][2, 8]$, while c and m also need to satisfy $c^{m+1} \ll n_l$.
- (2) Calculate the average RGCMSSE (ARX) and multi-scale standard deviation (MSSD) of sample for i -th class and s -th scale as follows:

$$ARX_{i,s}(p) = \frac{1}{n} \sum_{p=1}^n RGCMSSE_{i,s}(p) \tag{21}$$

$$MSSD_{i,s}(p) = \sqrt{\frac{\sum_{p=1}^n (RGCMSSE_{i,s}(p) - ARX)^2}{n - 1}} \tag{22}$$

where $1 \leq s \leq 20$.

- (3) Calculate the Euclidean distances and standard deviation between i -th and j -th classes as follows:

$$EED_{i,s}(i, j) = \sqrt{\sum_{p=1}^n (ARX_{i,s}(p) - ARX_{j,s}(p))^2} \tag{23}$$

$$SDED_{i,s}(i, j) = \sqrt{\sum_{p=1}^n (MSSD_{i,s}(p) - MSSD_{j,s}(p))^2} \tag{24}$$

where EED is the entropy Euclidean distance and SDED is the standard deviation Euclidean distance.

- (4) Calculate MSAED as follows:

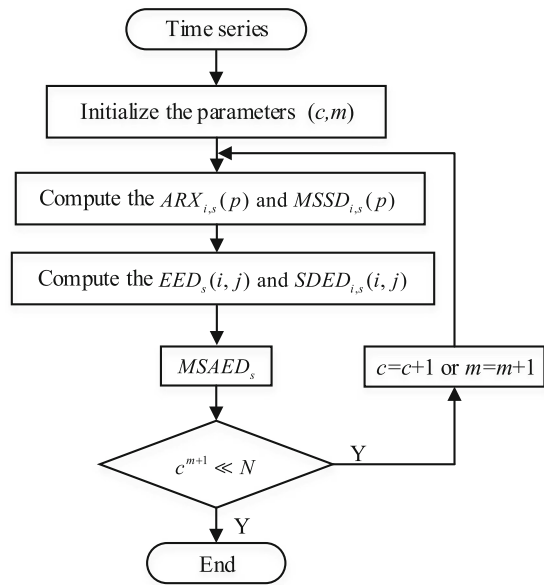


Fig. 12 Flowchart of RGCMSSE parameter selection process

$$MSAED_s = \frac{\sum_{i=1}^{\tau} \sum_{j=1, j \neq i}^{\tau} EED_s(i, j)}{\sum_{i=1}^{\tau} \sum_{j=1, j \neq i}^{\tau} SDED_s(i, j)} \tag{25}$$

- (5) Update parameters c and m , and then repeat Steps (1)–(4) to calculate the required MSAED value.

The flowchart of the RGCMSSE parameter selection process is shown in Fig. 12.

The RGCMSSE method can extract the information of different gearbox health state signals at multiple scales. The RGCMSSE values obtained for different parameters can be quantified using the MSAED index. The larger the MSAED value, the greater the distinguishability between different health states at multiple scales, and the better the stability of entropy values. This means that RGCMSSE can extract more information from the gearbox vibration signals and offers higher distinguishability. In the process of traversing parameters c and m , we can determine the optimum combination of c and m yielding the larger MSAED value.

4 Unsupervised feature selection using RSFS

High-dimensional fault features can be extracted by the RGCMSSE method. However, some of these features are redundant and will unnecessarily increase the computational time of the classification algorithm and harm the classification accuracy. Therefore, in order to improve the accuracy of fault recognition, it is necessary to select a limited number of features from the high-dimensional features. In this paper, the RSFS method is proposed for the sensitive feature selection.

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

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

- (1) Construct a local kernel regression, $p_{il}(\cdot)$, based on data point, x_i as follows:

$$p_{il}(x_i) = \frac{\sum_{x_j \in G_i} K(x_i, x_j) y_{jl}}{\sum_{x_j \in G_i} K(x_i, x_j)} \tag{26}$$

where G_i is the neighborhood of x_i , and $K(\cdot)$ is the kernel function. Matrix $S = [s_{ij}] \in \mathbb{R}^{N \times N}$ is defined as follows:

$$s_{ij} = \begin{cases} \frac{K(x_i, x_j)}{\sum_{x_j \in G_i} K(x_i, x_j)} & x_j \in G_i \\ 0 & x_j \notin G_i \end{cases} \tag{27}$$

- (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 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 , are then given as follows:

$$W = \left(XX^T + \frac{\beta}{\alpha} D \right)^{-1} X(F - Z) \tag{28}$$

$$Z_{ij} = \begin{cases} 0 & \text{if } |E_{ij}| \leq \frac{\gamma}{2\alpha} \\ \left(1 - \frac{\gamma}{2\alpha|E_{ij}|} \right) E_{ij}, & \text{otherwise} \end{cases} \tag{29}$$

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

- (4) Update F as follows:

$$F_{ij} \leftarrow F_{ij} \sqrt{\frac{[M^- F + vF + \alpha A^+]_{ij}}{[M^+ F + \alpha F + vF F^T F + \alpha A^-]_{ij}}} \tag{30}$$

where $A = X^T W + Z$, and v is input parameter.

- (5) Update the diagonal matrix $D_{ii} = \frac{1}{2\|w_i\|}$, and sort all features in descending order according to $\|w_i\|$, and then select the feature with the top rank.

The RSFS is an unsupervised feature selection algorithm based on spectral regression and sparse graph embedding. We found that the traditional algorithm based on spectral feature selection had two major problems: (1) Noise and uncorrelated features might adversely affect the estimated Laplacian, and (2) noise was inevitably introduced into the estimated clustering tags in the process of mapping discrete tags into continuous embedding. The RSFS method uses a robust local learning method to reduce the adverse effects of noise and uncorrelated features on the Laplacian operator, and a robust spectral regression method to deal with the effect of noise on clustering labels.

5 Proposed fault diagnosis method

5.1 Extreme learning machine

After extracting the feature set, an intelligent classifier is used to identify the gearbox fault type. The ELM is a new strong classifier based on single-hidden layer feed forward neural network (SHLFFNN), which can overcome some disadvantages of the traditional SHLFFNN, such as low learning efficiency, overfitting, and tendency to be attracted to local minima. The input weights and biases of the algorithm are randomly generated, and the calculation process remains unchanged; thus, it has a fast learning speed and requires minimum manual intervention. The ELM calculation procedure can be summarized as follows:

- (1) Given a training dataset $\{x_i, y_i\}, (i = 1, 2, \dots, M)$, 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 input weights, ω_i , and the input biases, b_i .

- (2) Calculate the hidden layer output matrix, H :

$$H = \begin{bmatrix} g(\omega_1, b_1, x_1) & \cdots & g(\omega_L, b_L, x_1) \\ \vdots & \cdots & \vdots \\ g(\omega_1, b_1, x_M) & \cdots & g(\omega_L, b_L, x_M) \end{bmatrix}_{M \times L} \tag{31}$$

where M is the number of samples.

- (3) Calculate the output weight, β :

$$H\beta = T \rightarrow \beta = H^+T = H^T \left(\frac{I}{C} + HH^T \right)^{-1} T \tag{32}$$

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) Define the ELM outputs as follows:

$$Y_i = \sum_{i=1}^K \beta_i g(\omega_i x_i + b_i), \quad i = 1, 2, 3, \dots, L \tag{33}$$

The structure of an ELM is illustrated in Fig. 13.

5.2 Proposed fault diagnosis method

Because of the advantages of the RGCMSSJE, RSFS, and ELM, a fault diagnosis method for WT gearboxes that combines the three algorithms is presented in this paper. Figure 14 shows the flowchart of the proposed method, which can be summarized as follows:

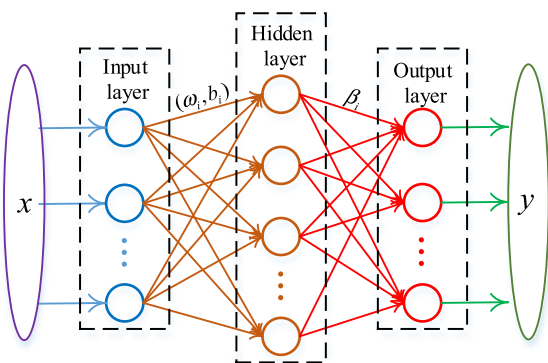


Fig. 13 Structure of ELM

- (1) Collect and store the vibration data from different health states of WT gearboxes.
- (2) Use the RGCMSSJE method to extract fault features through quantifying the complexity of gearboxes vibration signals for different health conditions. The parameter combination of the RGCMSSJE method is determined by the MSAED method. In this paper, we use the time delay $d = 1$ and scale factor $s = 20$, and find the best parameter combination of embedding dimension, m , and number of symbol categories, c .
- (3) Use the RSFS method to select the most sensitive features to construct the sensitive feature vector.
- (4) Input the selected sensitive feature vectors into the ELM multi-fault classifier to train to identify different operation conditions. In order to eliminate the influence of data randomness on the recognition accuracy of the ELM method, the tenfold cross-validation method is applied.

6 Experimental verification

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

6.1 Case 1: Experimental gearbox vibration data from simulated fault test

6.1.1 Experimental system and input dataset.

The experimental gearbox vibration signal dataset from the University of Connecticut was adopted in this study [51, 52]. The specifications of the experimental platform are described as follows. The experimental platform, shown in Fig. 15, used a benchmark two-stage gearbox, whose speed was controlled by a motor. The torque was supplied by a magnetic brake, which was adjusted by changing its input voltage. A 32-tooth pinion and an 80-tooth gear were installed on the first stage input shaft. The second stage consisted of a 48-tooth pinion and 64-tooth gear. The input shaft speed was measured by a tachometer, and gear vibration signal was measured by an accelerometer.

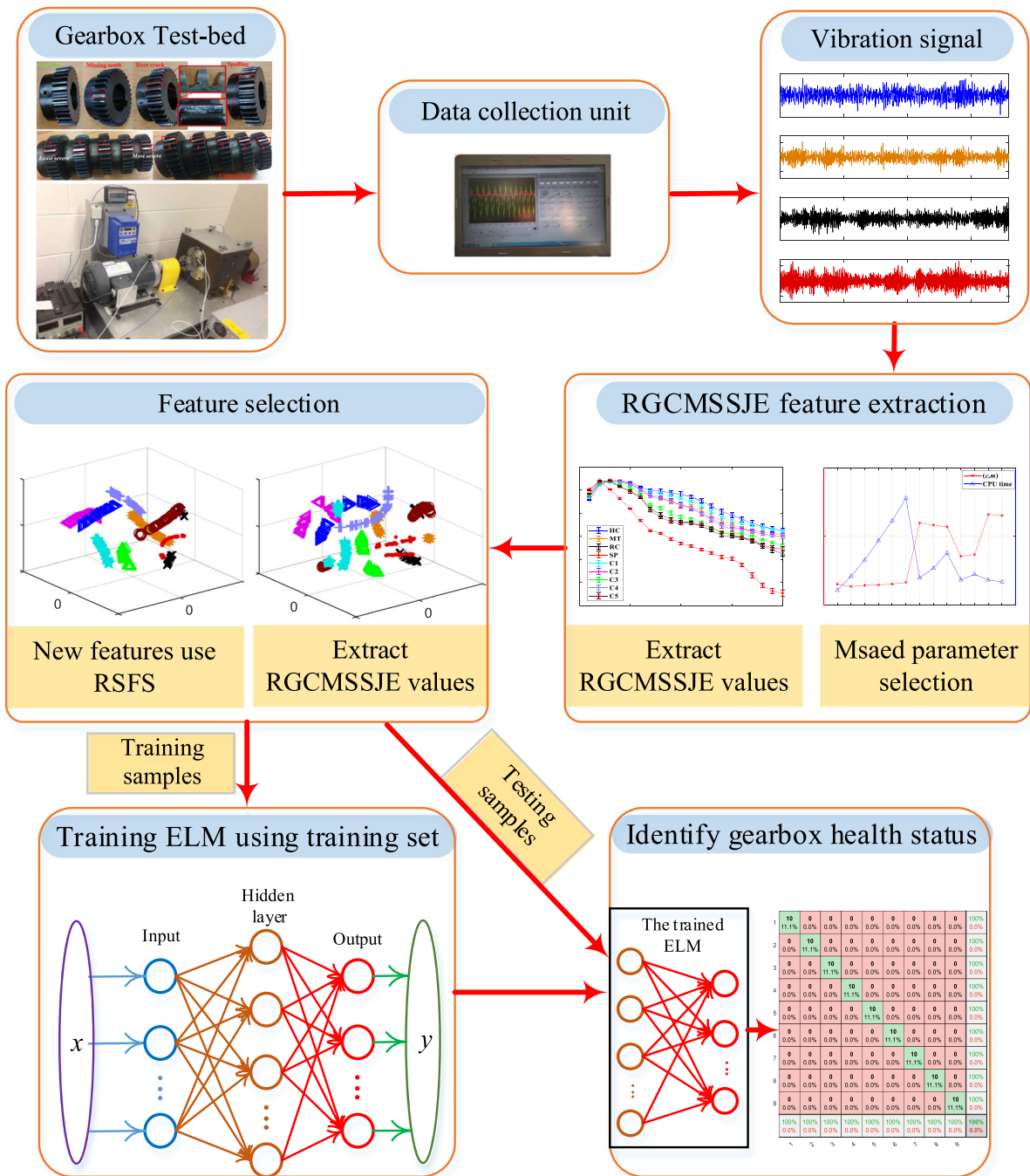


Fig. 14 Flowchart of proposed WT gearbox fault diagnosis method

The signals were recorded using a dSPACE system (DS1006 processor board, dSPACE Inc.) at a sampling frequency of 20 kHz. The gearbox faults were simulated on the pinion of the first stage input shaft, while the second stage gear of the gearbox was kept intact.

Nine different gear damage states were introduced into the input shaft pinion, including five fault types and five wear degrees. For each gear state, 104 vibration signals were recorded, each containing 3600 data points. In order to eliminate the non-stationary caused

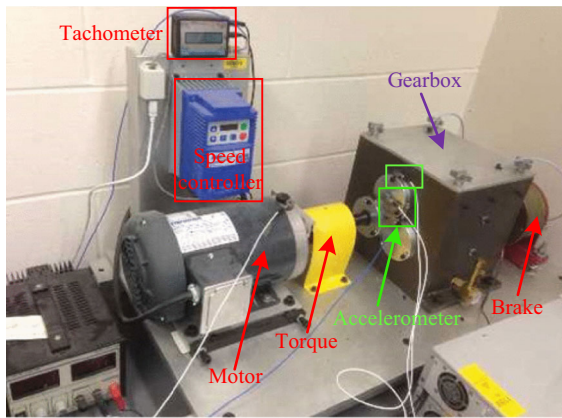


Fig. 15 Laboratory gearbox experimental platform

by the variable speed, the experimental platform converted the vibration signal from the time domain to the angular domain and averaged it in the angular domain. More details of the dataset can be obtained from Ref. [52].

In this test, all nine types of faulty gears were applied to the test equipment. The experimental vibration signals were designated as healthy condition (HC), missing tooth (MT), root crack (RC), spalling (SP), and chipping tip with five different levels of severity (C1–C5), and 100 non-overlapping samples were taken from each state to form an input dataset. Table 5 lists all the experimental datasets and input datasets.

6.1.2 Parameter selection and feature set determination

Figure 16 shows the time series and frequency spectra of the gearbox vibration signals for the nine different fault conditions. From Fig. 16, we can find that when the gear failed, the amplitude in the high frequency band increased because of the defect resonance. However, the frequency of the high frequency resonance caused by the nine health conditions was the same in the spectrum, and the change in amplitude was irregular. Because, the vibration signals usually had nonlinear and non-stationary characteristics, it was difficult to accurately identify different fault types and severities directly from the time series and frequency spectra. In addition, relying on time–frequency analysis for fault diagnosis requires significant experience of the analyst. This not only reduces the efficiency, but may also lead to significant errors. Therefore, it is necessary to develop an effective feature extraction method to realize automatic and efficient fault diagnosis.

The RGCMSSE method was used to extract more fault information from the gearbox vibration signals, to extract the features. According to the flow of the method for gearbox health status identification proposed in Sect. 5.2, first, the MSAED method was adopted to select the appropriate parameter combination. 10% of the data were used for parameter determination. Under the condition $c^{m+1} \ll N$, 13 (c, m) parameter combinations were selected to calculate the MSAED values, which are shown in Fig. 17 alongside the corresponding CPU times. It can

Table 5 List of laboratory experimental datasets and the input datasets

Fault class	Class label	(Data points × Signal samples)	
		Experimental datasets	Input datasets
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

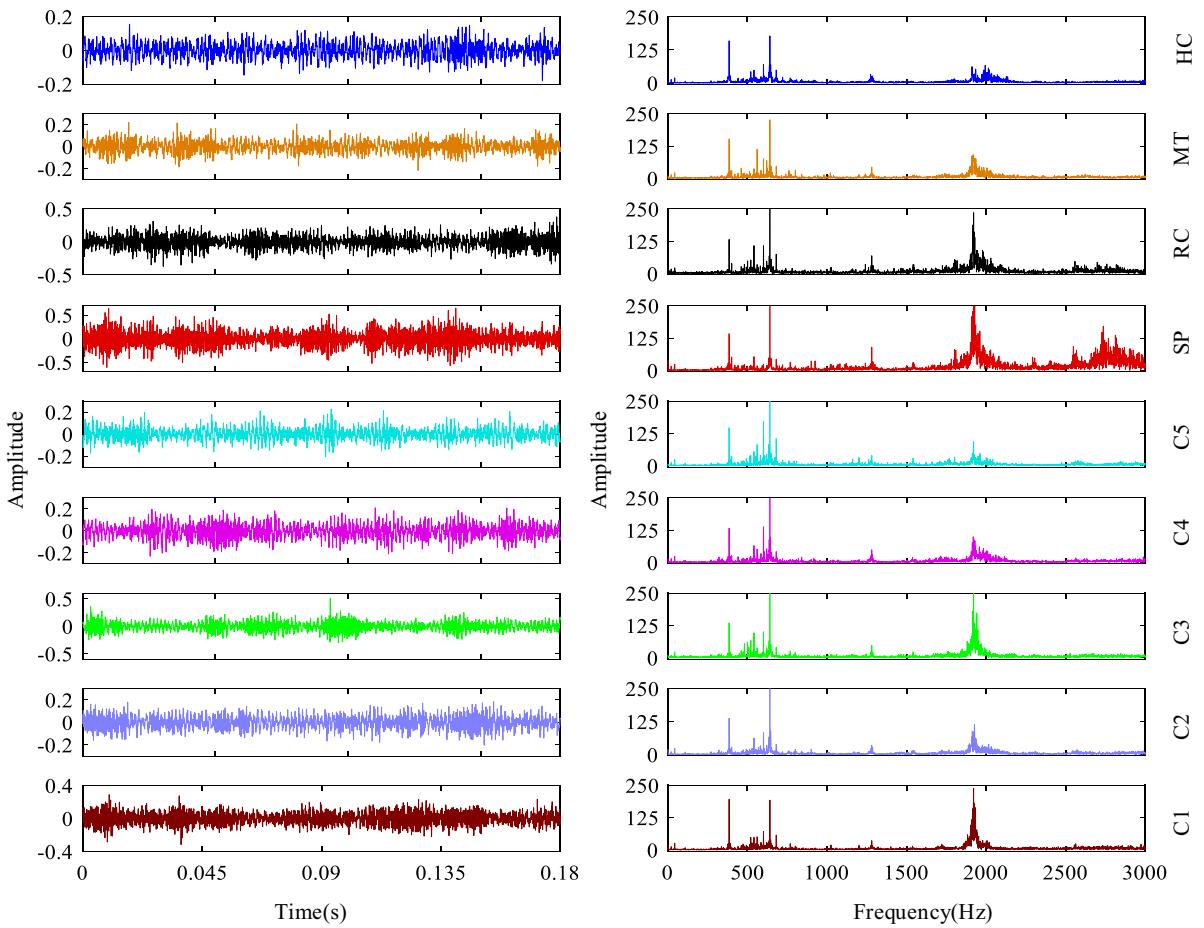


Fig. 16 Laboratory gearbox fault experimental signals

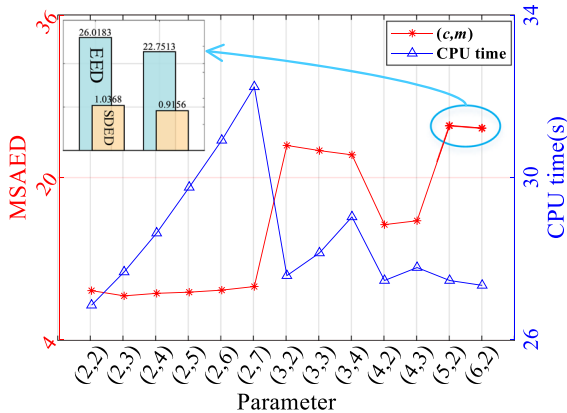


Fig. 17 MSAED using different (c, m) parameter combinations for laboratory gearbox data

be seen from Fig. 17 that the value of MSAED increased with the increase in c when m was fixed. At

Table 6 Comparison of different feature extraction methods for laboratory gearbox data

Method	Parameters	MSAED	CPU time (s)
RGCMSSJE	$c = 5, m = 2$	25.0945	27.77
GRCMSSJE	$c = 5, m = 2$	9.9686	52.609
RCMSSJE	$c = 4, m = 2$	11.7236	27.989
GRCMSDE	$c = 5, m = 2$	10.0981	48.938
GRCMSPE	$m = 3$	3.31	93.256
GRCMSSE	$m = 2, r = 0.15$	6.3455	78.841

the same time, when c was fixed, the CPU time increased rapidly with the increase in m . This provides a useful guidance for the selection of parameter m . It should be noted that the maximum value of MSAED of 25.0945 was achieved for the combination of

parameters (5,2). This shows that the RGCMSSE method with the parameter combination (5,2) could not only extract more fault information in a stable way and with higher efficiency than other parameter combinations. Therefore, considering also the favorable CPU time for this parameter combination, we selected it for further data processing.

We comprehensively compared the advantages of the feature extraction methods proposed in this paper to other five methods, using their optimal parameter values determined by the MSAED method. The results are shown in Table 6, where GRCMSSE is the generalized refined composite multi-scale SSE entropy based on the second-order moment. It can be seen from Table 6 that the MSAED value of the RGCMSSE was the largest, while the CPU time consumption the least, which indicates that this method could extract more useful information from the time series and its extraction efficiency was high. A graph comparing how the multi-scale entropy values change with the scale factor for the six feature extraction methods is shown in Fig. 18. It can be seen that when the scale factor s was greater than 5, the

dispersion of the RGCMSSE values was the best, and there were no large fluctuations. This shows that the multi-scale features extracted by the RGCMSSE method were the most distinguishable and stable.

6.1.3 Dimension reduction and comparative analysis of feature space

As can be seen from Fig. 18a, not all 20-dimensional features did not confidently distinguish between the fault states of the gearbox. This shows that there was redundant information contained in the 20-dimensional features, which was not conducive to achieving the efficient performance of intelligent algorithms. In order to further improve the classification accuracy and eliminate redundant information, it was necessary to select the features that contained the most useful information for fault classification. Therefore, this study used the RSFS algorithm to select the most relevant features. The RSFS method uses local robust learning to rank high-dimensional features. According to their importance, the RGCMSSE features were sorted as follows:

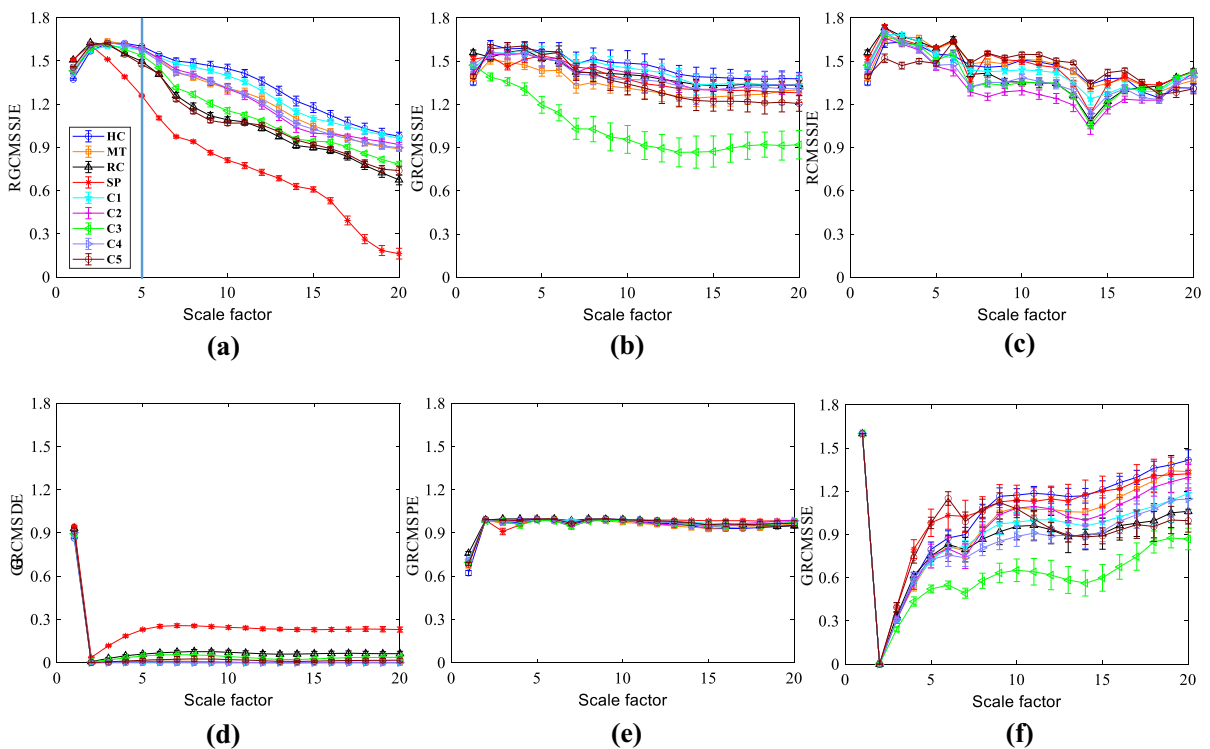


Fig. 18 Entropy of laboratory gearbox vibration signals versus scale factor for optimal parameters: **a** RGCMSSE; **b** GRCMSSE; **c** RCMSSJE; **d** GRCMSDE; **e** GRCMSPE; **f** GRCMSSE

$$\begin{aligned}
 R_6 > R_{20} > R_1 > R_{10} > R_9 > R_{11} > R_{17} > R_3 > R_2 \\
 > R_{14} > R_{16} > R_{13} > R_{18} > R_{12} > R_{15} > R_4 > \\
 R_5 > R_{19} > R_8 > R_7
 \end{aligned}
 \tag{34}$$

where R_s represents the feature selected by RSFS and s is the scale factor. At the same time, in order to comprehensively evaluate the performance of RSFS method, we adopted the LS, FS, and MRMR indices for comparison [53]. Then, the dimension of the features to be retained was set to $n_s - 1$, where n_s was the number of fault states of gearbox (i.e., $n_s = 9$). In other words, the first eight features of the four dimensionality reduction methods were selected as the most sensitive features to compose a new feature set. Here a description of the feature space is given. In this experiment, the feature dataset extracted by the RGCMSSE method was a 900×20 matrix, where the number of feature samples was 900 and the feature dimension was 20. In addition, the experimental data came from nine gearbox fault states, and thus the category labels were 1–9. In other words, after the four feature selection algorithms selected eight features

from the 900×20 feature dataset, the reduced feature dataset was 900×8 . The performance comparison of the four feature selection algorithms is shown in Fig. 19 and Table 7.

The visualization of the original signal and all feature distributions is given in Fig. 19. In order to intuitively analyze the feature space, we applied the t-distributed stochastic neighbor embedding (t-SNE) algorithm to project the features onto a three-dimensional space. It can be seen from Fig. 19a that the original signals of the nine fault states overlapped with each other, which indicated that it would be very difficult to classify the faults by using the original features directly. At the same time, Fig. 19 shows that the RGCMSSE algorithm clustered the features of the same category, while separating the features of different categories from one another, which is convenient for the classification of different fault conditions. It is worth noting that the features selected by the RSFS method were more clearly clustered. Although the other three feature selection algorithms separated the nine states well, the clustering within the same category was not strong enough. This shows that

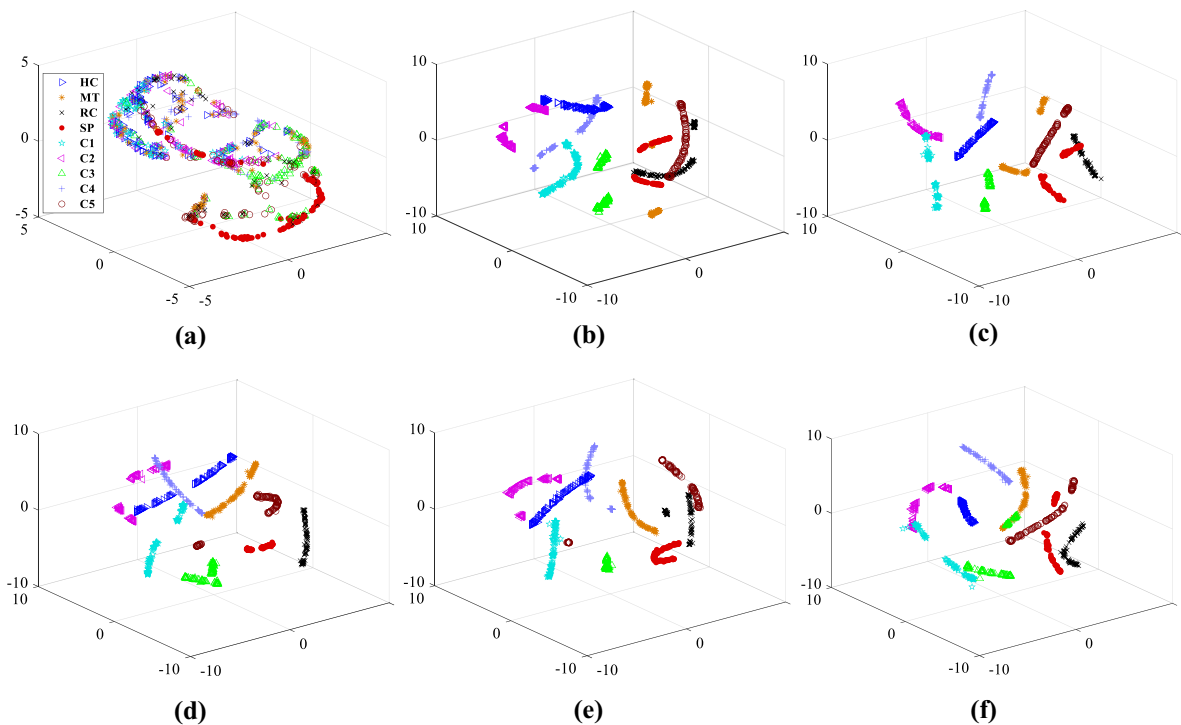


Fig. 19 Feature visualization for laboratory gearbox signal via t-SNE: **a** Raw signal. **b** RGCMSSE. **c** RGCMSSE + RSFS. **d** RGCMSSE + LS. **e** RGCMSSE + FS. **f** RGCMSSE + MRMR

Table 7 Comparison of different dimensionality reduction algorithms using laboratory experimental data

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	0.0900	28.875	0.162
FS	4.4935	0.1428	31.459	0.108
MRMR	6.1143	0.1624	37.648	42.83

the RSFS algorithm extracted features with more fault information.

In order to quantify the feature selection performance of the above four methods, the between-class scatter, within-class scatter, and CPU running times were adopted for comprehensive analysis, as shown in Table 7. We took the ratio of between-class scatter and within-class scatter as the performance index of feature selection. The larger the ratio, the more concentrated the similar features and the more distant the heterogeneous features. From Table 7, compared to the LS, FS, and MRMR, the RSFS had the largest feature selection performance index, while its CPU time was only twice those of the LS and FS algorithms, which demonstrated the superiority of RSFS algorithm in feature selection.

6.1.4 Analysis of fault diagnosis results

As the final step, after the appropriate features had been extracted, the ELM classifier was used to distinguish between the fault states of the gearbox. It

is worth noting that before the ELM classification, we need to decide the number of hidden layer neurons and select a suitable kernel function. This analysis is shown in Fig. 20, where the means and standard deviations are shown for 20 trials. Figure 20a shows the effect of different kernel functions on the ELM classification accuracy. It can be seen that the sin function achieved the highest classification accuracy in a wide range of the numbers of hidden layer neurons, and the stability of ELM classification was gradually enhanced with the increase in the number of hidden layer neurons. The preset kernel function of Fig. 20b was the sin function. As can be seen from Fig. 20b, when the number of hidden layer neurons was 70, respectively, the classification accuracy was the highest, and stability the best, while the CPU time was still moderate. Therefore, we used the sin function as the kernel function in the ELM and set the number of hidden layer neurons to 70.

In order to improve the reliability and accuracy of the fault classification results, the tenfold cross-validation algorithm was adopted in the training and

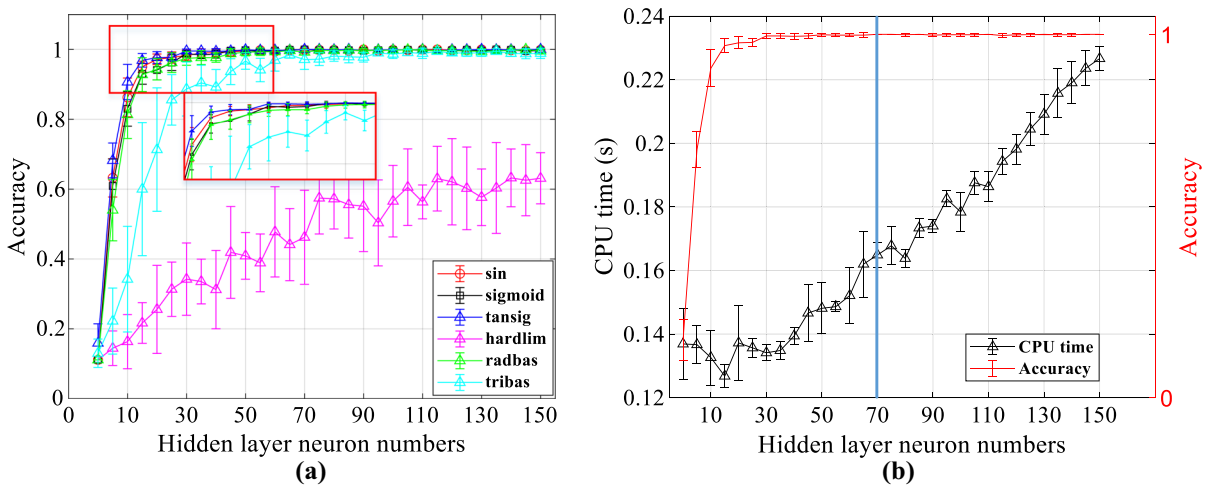


Fig. 20 ELM parameter selection: **a** Classification accuracy versus number of hidden layer neurons for different kernel functions. **b** CPU time versus number of hidden layer neurons for sin kernel function

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%	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%	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%	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%	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%	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%	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%	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%	0 0.0%	0 0.0%	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%	100% 0.0%	100% 0.0%	99.0% 1.0%	100% 0.0%	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. 21 ELM CM for RGCMSSE + RSFS method

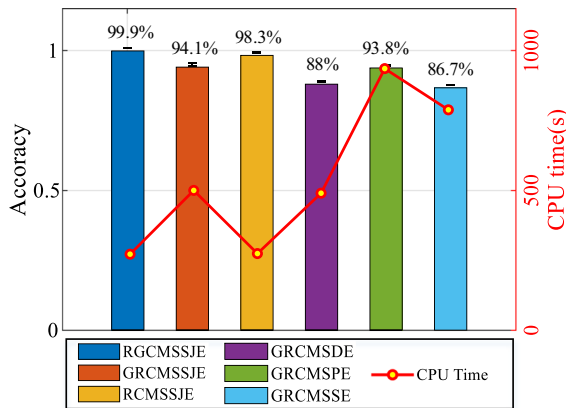


Fig. 22 Performance comparison of six feature extraction algorithms

testing process of the intelligent classification algorithms in this section. The confusion matrix (CM) for the ELM intelligent classifier is shown in Fig. 21, which shows the result that repeated most often in the 20 experiments. It can be seen from Fig. 21 that the overall recognition accuracy of the fault diagnosis method proposed in this paper reached 99.9%, and only the recognition rate of C4 fault was less than 100%. This indicates that the method effectively identified the different gearbox fault types and severities, and had an excellent overall classification accuracy. Although fault C4 signal was very similar to that of fault C2, only one sample was misclassified

in this trial. In order to highlight the advantages of the RGCMSSE feature extraction method proposed in this paper, we also compared the classification performance of the features extracted by the RGCMSSE, GRCMSSE, RCMSSJE, GRCMSDE, GRCMSPE, and GRCMSSE methods. The performance comparison of the six feature extraction algorithms, all using the ELM, is shown in Fig. 22, from which the following conclusions can be drawn. First, the RGCMSSE method had a higher accuracy than the other entropy methods combined with the generalized refined multi-scale analysis methods (GRCMSSE, RCMSSJE, GRCMSDE, GRCMSPE, and GRCMSSE). This was mainly because the joint state analysis extracted the current and subsequent state characteristics of the fault from the original signal, whereas the other entropy methods only considered the current fault information contained in the signal. Furthermore, compared to the other methods, the proposed RGCMSSE method achieved the highest test classification accuracy with less CPU time. The CPU time here refers to the total time including the RGCMSSE feature extraction, RSFS feature selection, and ELM fault classification. This further highlights the advantages of RGCMSSE for fault feature extraction, which reduced the CPU time and extracted more reliable feature information.

In order to verify the classification performance of ELM, it was compared to four different classifiers, including the SR, SVM, RF, and KNN. Four common evaluation indices, including accuracy, F-score, adjusted Rand index (ARI), and normalized mutual information (NMI), were used to verify the performance of different models. Accuracy represents the proportion of all correctly classified results in the total results. F-score is the weighted harmonic mean of both the precision and recall. It reaches its best value at one and worst at zero. ARI takes into account the number of instances that exist in the same cluster and different clusters. The higher the value of ARI, the more data instances are correctly clustered. NMI is adopted to estimate the quality of the clustering with respect to a given class label of the data. The formulas for the four indices are given in Table 8, where tp, tn, fp, and fn refer to true positives, true negatives, false positives, and false negatives, respectively. Precision =

$$\sum_{i=1}^k tp_i / \sum_{i=1}^k (tp_i + fp_i) \quad \text{and}$$

Table 8 Performance evaluation metrics

Metric	Formula	Range
Accuracy	$Accuracy = \frac{tp+tn}{tp+fp+tn+fn}$	[0,1]
F-score	$F = 2 \frac{Recall \times Precision}{Recall + Precision}$	[0,1]
ARI	$ARI = \frac{n_{11} + n_{00}}{C_n^2}$	[0,1]
NMI	$NMI = \frac{\sum_{\varphi \in \Phi} \sum_{\omega \in \Omega} p(\varphi, \omega) \log(p(\varphi, \omega) / p(\varphi)p(\omega))}{\sqrt{(\sum_{\varphi \in \Phi} p(\varphi) \log(p(\varphi))) (\sum_{\omega \in \Omega} p(\omega) \log(p(\omega)))}}$	[0,1]

Table 9 Parameters of four classification methods

Method	Parameters	Value
ELM	Kernel function	sin
	Number of hidden layer neurons	70
SR	Lambda	0.00001
	Maximum number of iterations	200
	Theta	310
SVM	Kernel function	RBF
	Penalty factor	1
	Kernel parameter	10
RF	Estimation mode	Out-of-bag
	Number of trees	100
KNN	Number of nearest neighbors	1
	Distance metric	Euclidean

respectively. Then n_{11} is the number of the sample pairs with overlapping labels in Φ and Ω , n_{00} is the number of sample pairs with non-overlapping labels in Φ and Ω , C_n^2 is the total number of sample pairs, $p(\cdot)$ is the probability distribution function, and $p(\varphi, \omega)$ is the joint probability distribution function of Φ and Ω . The larger the value of the above evaluation indices, the stronger the comprehensive classification ability of the classifier.

Table 9 lists the parameter values of the five classifiers. The feature sets of the five classification algorithms were obtained by the RGCMSSE and RSFS methods. In order to reduce randomness, each method was tested using tenfold cross-validation. The final verification results are shown in Fig. 23 and Table 10. From the radar diagram in Fig. 23, it can be seen that the curve of the RGCMSSE-RSFS-ELM model proposed in this paper is farthest from the center. At the same time, Table 10 shows that the ELM classifier training and testing consumed the least CPU time. Based on the above advantages of the ELM classifier, it was selected as the intelligent classifier of the proposed model. The results show that the proposed model not only achieved the best diagnostic results, but also exhibited good stability compared to the other models (RGCMSSE-RSFS-SR, RGCMSSE-RSFS-SVM, RGCMSSE-RSFS-RF, and RGCMSSE-RSFS-KNN).

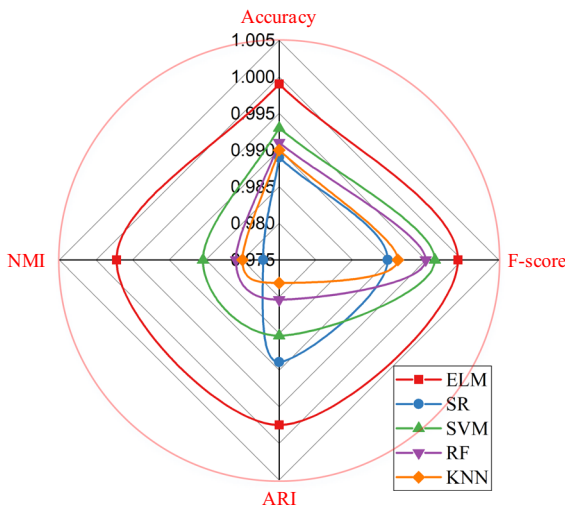


Fig. 23 Radar chart of classifier evaluation

$Recall = \sum_{i=1}^k tp_i / \sum_{i=1}^k (tp_i + fn_i)$, where k is the number of classes to be classified. Suppose Φ and Ω are the sets of actual labels and classified labels,

6.2 Case 2: WT gearbox vibration data

In what follows, in order to verify the fault identification and generalization abilities of the proposed method, we study experimental data of WT gearbox with faults (WTGFs) under different working conditions [54].

Table 10 Comparison of classifier performance

Method	Performance index				
	Accuracy (%)	F-score	ARI	NMI	CPU time (s)
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

Table 11 Description of working conditions of WT gearbox

Fault class			
Healthy		Broken tooth	
Condition label	Dataset	Condition label	Dataset
h30hz0	88,832 × 4	b30hz0	88,320 × 4
h30hz10	92,928 × 4	b30hz10	111,616 × 4
h30hz20	108,544 × 4	b30hz20	114,432 × 4
h30hz30	106,240 × 4	b30hz30	89,856 × 4
h30hz40	100,608 × 4	b30hz40	94,464 × 4
h30hz50	110,848 × 4	b30hz50	94,208 × 4
h30hz60	99,840 × 4	b30hz60	95,488 × 4
h30hz70	101,377 × 4	b30hz70	100,865 × 4
h30hz80	99,841 × 4	b30hz80	110,336 × 4
h30hz90	106,752 × 4	b30hz90	105,728 × 4

6.2.1 Experimental system and input dataset

The WTGF dataset was created by researchers engaged in fault diagnosis of WT gearboxes. Details of this dataset are as follows [54]. The dataset contains two gearbox operating states: healthy and broken tooth. Each operating states contains 10 loading modes representing 10 working conditions. Each working condition has four attributes, which represent synchronous data collected by four sensors installed at four locations on the gearbox. The data were recorded at a sampling frequency of 30 Hz with the load changing from 0 to 90%. The data collected in the healthy operation state included $1,015,808 \times 4$ data points, while the data collected in the broken tooth operation state included $1,005,311 \times 4$ data points (see Table 11).

For this study, we selected three working conditions from the healthy and broken teeth states each (h30hz0, h30hz20, h30hz90, b30hz0, b30hz20, and b30hz90). Similar to Case 1, 102,400 data points of two sensors were divided into 50 data samples (each data sample

containing 2048 points). The tenfold cross-validation algorithm was also applied in the training and testing processes.

6.2.2 Parameter selection and feature set determination

The time- and frequency-domain representations of the six vibration signals are shown in Fig. 24, where it is easy to distinguish between the healthy and the broken tooth states. The amplitudes of the broken tooth fault time series are clearly smaller, and a clear low frequency peak appears in the spectra. However, it is difficult to distinguish between the three working conditions in each of the two gear states by directly observing the time- and frequency-domain plots. Therefore, we propose to use the RGCMSJE method to extract the vibration signal features for that purpose.

Similar to Case 1, this paper uses the RGCMSJE method to analyze the recorded experimental data to verify the effectiveness and generalization ability of the method. First, the MSAED method was used to

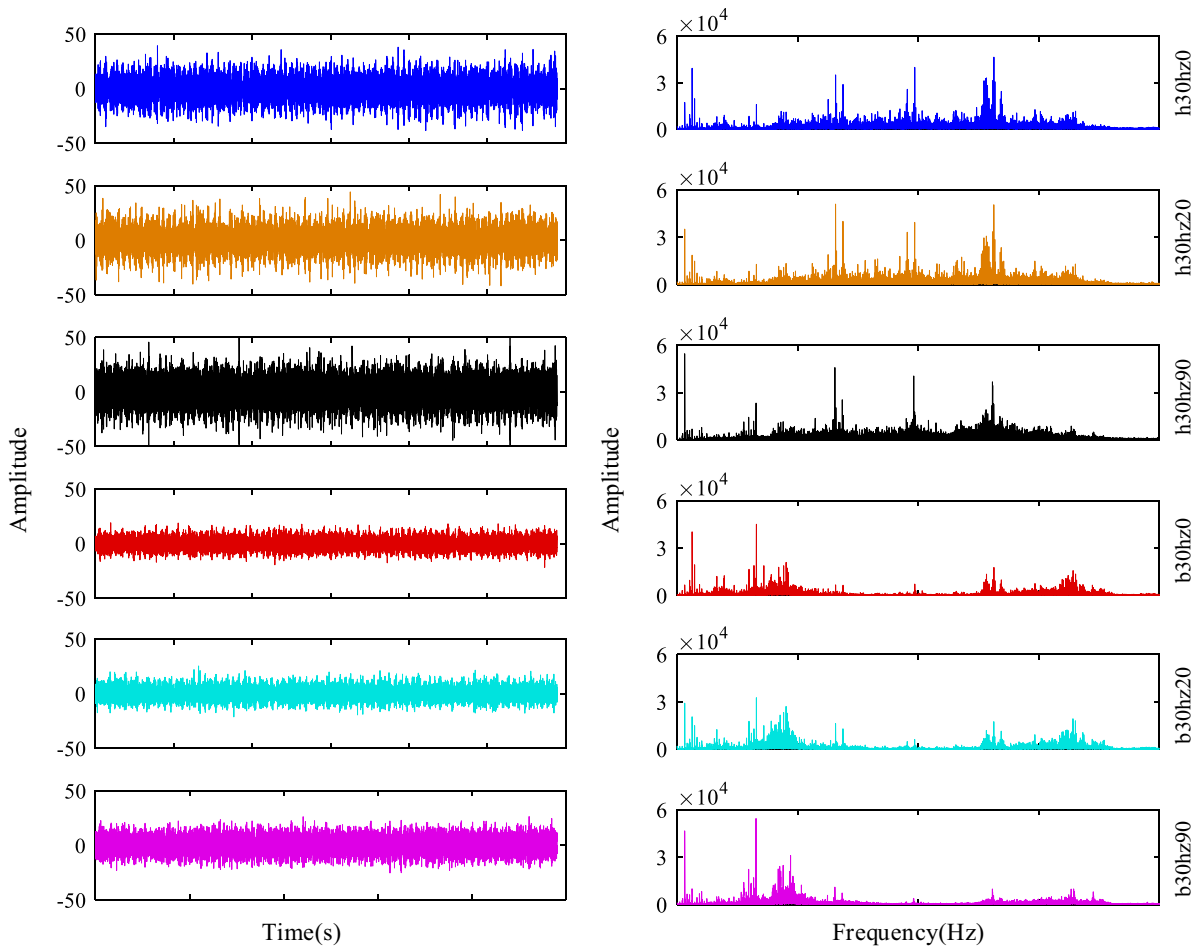


Fig. 24 WT gearbox experimental vibration data

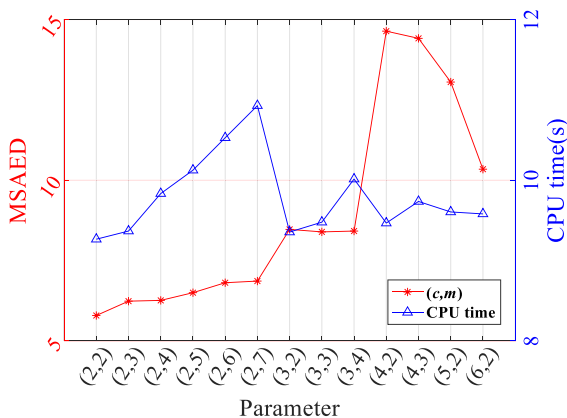


Fig. 25 MSAED using different (c, m) parameter combinations for WTGF dataset

select the optimal parameter combination, and then RGCMSSE method was applied to extract the multi-

Table 12 Comparison of different feature extraction methods for WTGF dataset

Method	Parameters	MSAED	CPU time (s)
RGCMSSE	$c = 4, m = 2$	14.6402	9.849
GRCMSSE	$c = 2, m = 7$	5.9415	19.571
RCMSSE	$c = 2, m = 7$	6.8231	12.277
GRCMSDE	$c = 2, m = 7$	5.0401	19.135
GRCMSPE	$m = 4$	3.6734	43.479
GRCMSSE	$m = 2, r = 0.15$	3.5557	26.721

scale feature vector set of dimension 20 using the optimal parameter combination. The experimental steps in this section were the same as those in Case 1, and the MSAED values and the corresponding CPU

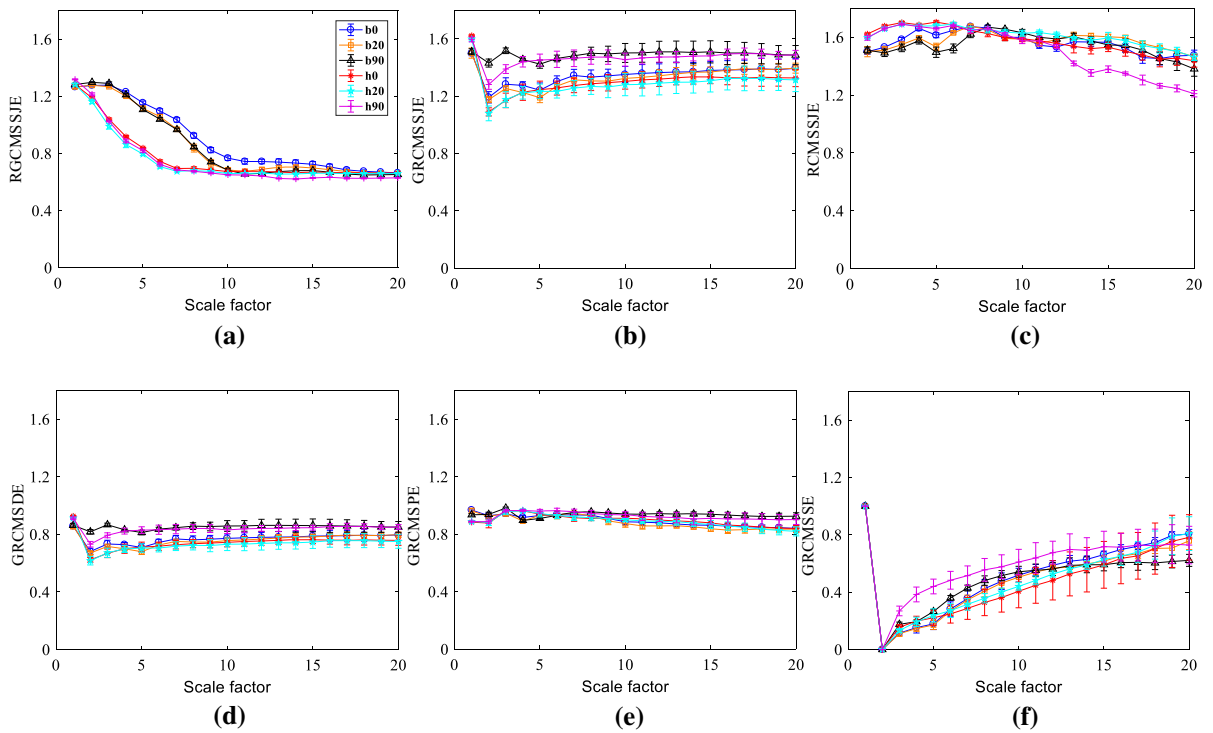


Fig. 26 Entropy of gearbox vibration signal from WTGF dataset versus scale factor for optimal parameters: **a** RGCMSJE, **b** GRCMSJE, **c** RCMSSJE, **d** GRCMSDE, **e** GRCMSPE, **f** GRCMSSE

times are shown in Fig. 25. Figure 25 demonstrates that the maximum value of MSAED was 14.6402, for the parameter combination (4,2). This shows that for the WTGF dataset, the RGCMSJE algorithm extracted more abundant fault information when the parameter combination was $c = 4$ and $m = 2$; therefore, considering also the CPU time, we chose these parameter values. Subsequently, the RGCMSJE method was compared to the other five feature extraction algorithms, and the comparison results are shown in Table 12 and Fig. 26. It can be seen from Table 12 that the results are the same as in Case 1. The RGCMSJE method gave the largest MSAED value and the shortest CPU time. The entropy plots in Fig. 26 also supported this claim, because the RGCMSJE curve had the largest discrimination power and the smallest standard deviation.

6.2.3 Dimension reduction and comparative analysis of feature space

Similar to Case 1, in order to improve the efficiency of fault recognition and eliminate the redundant

information from the feature space, we used the RSFS algorithm to reduce the dimension of the feature space. First, using the RSFS algorithm, the 20-dimensional features were sorted as follows:

$$\begin{aligned}
 R_9 &> R_4 > R_1 > R_6 > R_2 > R_7 > R_{19} > R_{20} > R_{18} \\
 &> R_{17} > R_{10} > R_8 > R_5 > R_{14} > R_3 > R_{11} \\
 &> R_{16} > R_{13} > R_{15} > R_{12}
 \end{aligned}
 \tag{35}$$

Subsequently, following [50], the feature dimension retained after dimensionality reduction was set to $n_s - 1$, where n_s is the number of gearbox operating conditions ($n_s = 6$). In other words, the first five features were selected as sensitive features to construct a new 6-dimensional feature vector set. Lastly, as the feature selection method adopted in this study, the RSFS approach was compared to the LS, FS, and MRMR methods. Figure 27 depicts the raw signal and feature distributions, which were obtained by applying the t-SNE algorithm to project the features onto a three-dimensional space. From Fig. 27, one can observe that the separation of different classes and

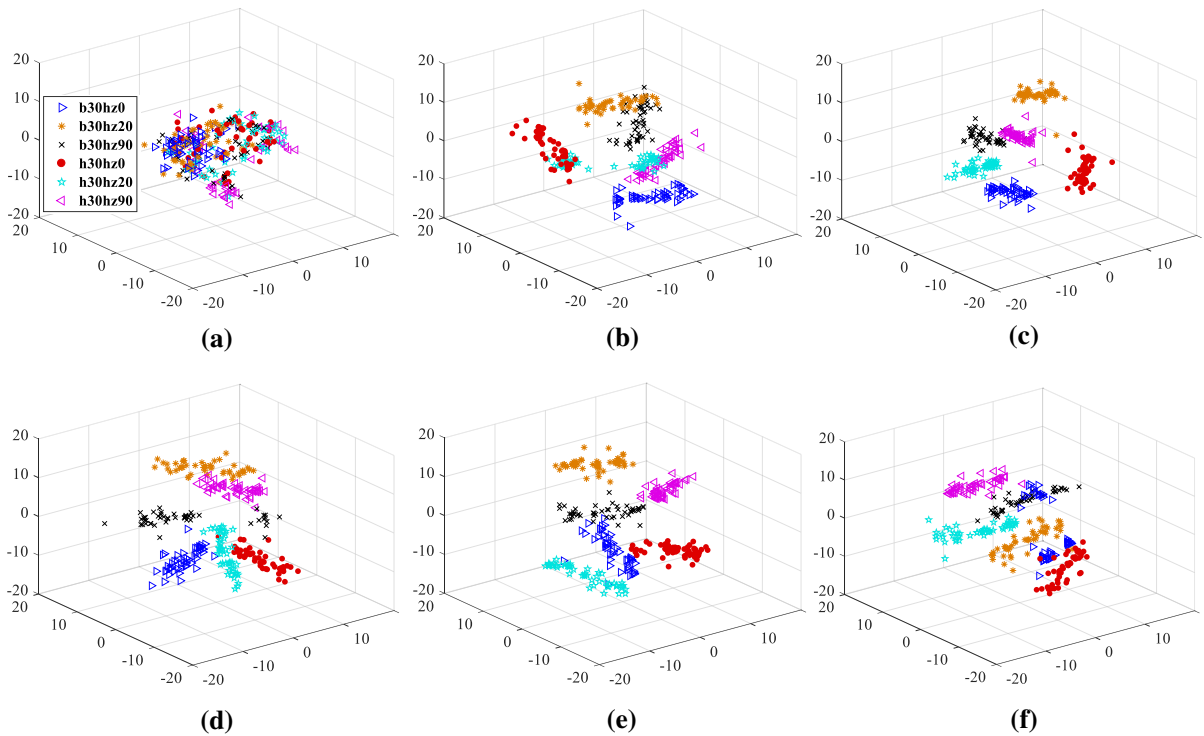


Fig. 27 Feature visualization for WTGF data via t-SNE: **a** Raw signal. **b** RGCMSSE. **c** RGCMSSE + RSFS. **d** RGCMSSE + LS. **e** RGCMSSE + FS. **f** RGCMSSE + MRMR

Table 13 Comparison of different dimensionality reduction algorithms using WTGF data

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

the clustering within the same class are both clearer after 6 sensitive features had been selected using the RSFS approach. By comparing Fig. 27a–f, the RGCMSSE method clearly extracted recognizable features from cluttered and inseparable signals.

At the same time, the gearbox state distinguishing ability of the features selected by the LS, FS, and MRMR was significantly lower than that of RSFS. To further illustrate the performance of the RSFS feature selection, the ratio of S_b/S_w is shown in Table 13. From Table 13, it can be concluded the RSFS method had the largest feature selection performance index compared to the LS, FS, and MRMR, and the CPU time of only 0.0905 s, which further demonstrates the

superiority of RSFS algorithm for feature selection. Therefore, this study adopted the RSFS algorithm as the dimension reduction algorithm for the high-dimensional feature space of the WTGF dataset.

6.2.4 Analysis of fault diagnosis results

Similar to Case 1, the reduced-dimension features were fed into the ELM intelligent classifier for gearbox fault classification. The parameters of the ELM adopted in this section were the same as in Case 1. The kernel function was the sin function and the number of hidden layer neurons was 70.

Output Class	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%
		1	2	3	4	5	6	
		Target Class						

Fig. 28 ELM CM for RGCMSSE + RSFS

Similarly, to improve the reliability and accuracy of the fault classification results, the tenfold cross-validation was used during training and testing of the ELM. The average results of 20 simulations were used to verify the accuracy of fault identification, as illustrated by CM shown in Fig. 28. 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. 28, it can be seen that except for working condition 3, the recognition accuracy of the other five working conditions reached 100%. The overall classification accuracy of the method proposed in this paper reached 99.3%,

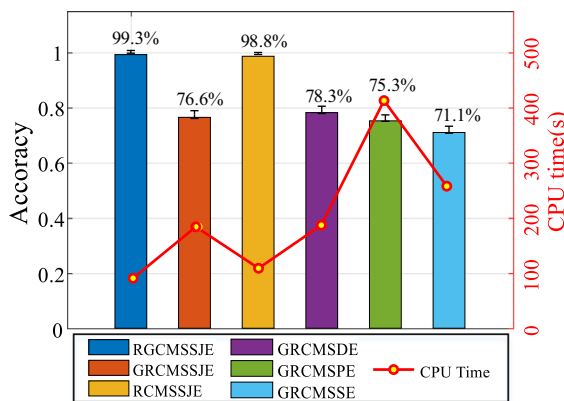


Fig. 29 Performance comparison of six feature extraction algorithms for WTGF data

which means that the method effectively detected the different working conditions of the gearbox transmission. The CM shows that one sample of working conditions 1 and 2 each was wrongly classified as working condition 3, which indicates that it was not conducive to distinguish the change of working condition in case of gear failure.

Similarly, we compared the classification accuracies and CPU times of the RGCMSSE, GRCMSSE, RCMSSJE, GRCMSDE, GRCMSPE, and GRCMSSE when using the ELM intelligent classifier. It is clear from Fig. 29 that the features extracted by the RGCMSSE method had the highest classification accuracy, while the RCMSSJE method was the second.

This shows that the features extracted by the joint state analysis effectively classified the gearbox states and the RGCMSSE method achieved the highest recognition accuracy while using less CPU time. At the same time, it was demonstrated that the RGCMSSE feature extraction method had good generalization ability.

Similar to Sect. 6.1, we compared comprehensively the performance of the five classifiers to verify the efficiency of the ELM. The parameters adopted by the five classifiers were the same as in Table 9, and the input feature set of the classifiers were obtained by the RGCMSSE and RSFS methods. Each algorithm was run 20 times, and detailed comparison results are given in Fig. 30 and Table 14. From the radar plot

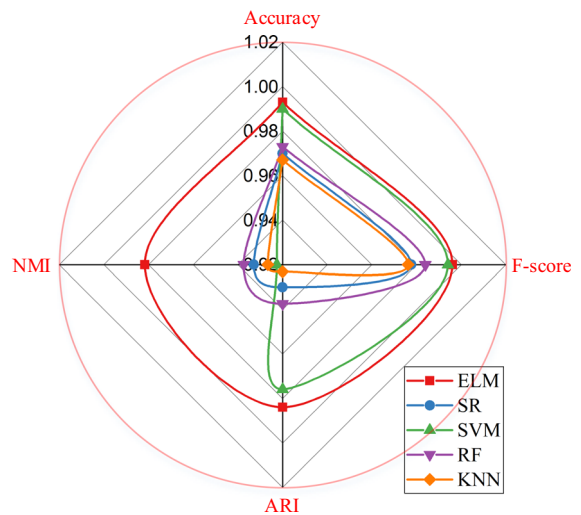


Fig. 30 Radar chart of different classifier evaluation for WTGF data

Table 14 Comparison of classifier performance for WTGF data

Method	Performance index				
	Accuracy (%)	F-score	ARI	NMI	CPU time (s)
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

shown in Fig. 30, it can be clearly observed that the four evaluation parameters of the RGCMSJE-RSFS-ELM method proposed in this paper were farthest from the center, which indicates that the overall classification ability of ELM was the best. It also shows that the RGCMSJE-RSFS-ELM method had the highest fault recognition ability. Furthermore, the data in Table 14 show that the training and testing of ELM classifier consumed the least CPU time.

6.3 Further discussions

By comparing the above methods, we can conclude that the proposed method combines the advantages of RGCMSJE, MSAED, RSFS, and ELM so that the gearbox component fault state can be effectively identified. The proposed method utilizes the following four concepts: the refined generalized multi-scale analysis, state joint entropy, MSAED aided parameter determination, and RSFS. At the same time, after the comparative analysis of intelligent classifier recognition accuracy and efficiency, we adopted the ELM as the fault identifier. The results of experimental Cases 1 and 2 showed that the fault diagnosis method proposed in this paper not only identified various gear faults under the same working conditions, but also gear faults under variable working conditions and had good generalization ability and stability. The reasons behind the advantages of the proposed model are summarized as follows:

- (1) In previous studies, the DE algorithm has been demonstrated to have some advantages in computing time and reliability due to its unique symbolization and linear mapping rules. However, the DE algorithm is similar to the SE and PE algorithms, which only consider the mode probability of current state and ignore the state transition probability of fault information from one state to another. Because of these shortcomings, we proposed the SJE entropy, which inherits the advantages of the DE algorithm and considers the mode and transition probabilities of the current state. Therefore, the SJE can extract fault information efficiently and in a stable fashion.
- (2) The generalized composite multi-scale analysis method effectively addressed the problem of missing mutation in the composite multi-scale analysis. However, it was found unsuitable for the DE and SJE algorithms, because significant amounts of useable information would vanish with the transformation to the second moment, leading to instable entropy values. Therefore, we proposed the RGCMSA method to address these problems. The refined generalized analysis and composite algorithm avoided information loss caused by transforming data to the second moment and extracted information at multiple time scales.
- (3) Previous studies on the parameters of entropy algorithms were based on the traversal simulation of analog signals. Therefore, the entropy parameters for specific fault data could not be obtained, and the parameter selection process required an experienced analyst, which inevitably led to errors. In this paper, the MSAED algorithm was proposed to determine the parameters of the RGCMSJE method, which adaptively optimized the parameters for the data at hand.
- (4) In the feature selection process, the LS algorithm was found to lack the capability of global information separation. On the other hand, the FS algorithm lacked the capability of local information preservation. Therefore, we applied the RSFS method to reduce the dimension of feature space. This approach adopted a robust

local learning method to deal with clustering errors so as to improve the local information retention ability while taking into account the global information.

7 Conclusions

In this paper, a novel intelligent fault diagnosis method for WT gearboxes based on the RGCMSSE, MSAED, RSFS, and ELM was presented. First, the concept of SJE entropy was defined, and the advantages of SJE were verified by comparing it to DE, SE, and PE. Then, a new fault feature extraction method RGCMSSE was proposed based on the RCMSA. Furthermore, the MSAED adaptive algorithm was proposed for automatically selecting the optimal parameters (c, m). Compared to the RGCMSSE, RCMSSE, GRCMSDE, GRCMSSE, and GRCMSPE methods, the proposed RGCMSSE method adopting the optimal parameters had the advantages of faster calculation speed and more stable feature extraction. Subsequently, the RSFS algorithm was introduced to automatically select salient features from the multi-scale features using robust local learning. Finally, a gearbox fault diagnosis method combining RGCMSSE, MSAED, RSFS, and ELM was constructed. The analysis results of two experimental cases indicated that the method effectively extracted the operating characteristics of the gearbox and successfully identified a variety of gearbox local faults and working conditions. More importantly, compared to the existing methods (such as the GRCMSDE, GRCMSSE and GRCMSPE), the proposed diagnostic scheme exhibited significantly enhanced diagnostic accuracy and computational efficiency. In future work, more industrial data should be employed to verify the generalization ability of the proposed method. Furthermore, the selection of MSAED-assisted entropy parameters and the combination of multi-scale analysis and RSFS method are very interesting directions to explore.

Acknowledgements 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. Authors are grateful to the University of Connection for allowing us to use their data. The authors would like to express their gratitude to EditSprings

(<https://www.editsprings.com/>) for the expert linguistic services provided.

Authors' contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by WD, SZ, MH, LZ, and HL. The first draft of the manuscript was written by WD, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The “Gear Fault Data” data that support the findings of this study are available from repository “figshare,” hyperlink to datasets “https://figshare.com/articles/dataset/Gear_Fault_Data/6127874/1.” And the “Gearbox data” data that support the findings of this study are available from repository “Github,” hyperlink to datasets <https://github.com/Gearboxdata/Gear-Box-Fault-Diagnosis-Data-Set>.

Declarations

Conflict of interest No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on the behalf of my co-authors that the work described was original research that has not been published previously and not under consideration for publication elsewhere. The authors declare that they have no conflict of interest.

Code availability (software application or custom code)

The core code used in this work is custom code.

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. *IEEE Trans. Ind. Inform.* **17**, 6875–6884 (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. *IEEE Trans. Ind. Inform.* **17**, 4613–4623 (2020)
3. Yang, Y., Zheng, H., Li, Y., et al.: A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network. *ISA Trans.* **91**, 235–252 (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. *Appl. Soft Comput.* **95**, 106515 (2020)
5. Morales-Valdez, J., Alvarez-Icaza, L., Escobar, J.A., et al.: Identification system for structural health monitoring in buildings. In: Dervilis, N. (ed.) *Special topics in structural dynamics and experimental techniques*, vol. 5, pp. 31–38. Springer, Cham (2020)
6. Chen, Y., Mendoza, A.S.E., Griffith, D.T.: Experimental and numerical study of high-order complex curvature mode shape and mode coupling on a three-bladed wind turbine assembly. *Mech. Syst. Signal Process.* **160**, 107873 (2021)

7. Ghaderi, P., Amini, F.: Development of a new method for online parameter identification in seismically excited smart building structures using virtual synchronization and adaptive control design. *Appl. Math. Model.* **87**, 203–221 (2020)
8. Zeng, M., Tan, B., Ding, F., et al.: An experimental investigation of resonance sources and vibration transmission for a pure electric bus. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **234**(4), 950–962 (2020)
9. Morales-Valdez, J., Alvarez-Icaza, L., Escobar, J.A.: Damage localization in a building structure during seismic excitation. *Shock Vib.* **2020**, 8859527 (2020)
10. Miao, Y., Zhao, M., Liang, K., et al.: Application of an improved MCKDA for fault detection of wind turbine gear based on encoder signal. *Renew. Energy* **151**, 192–203 (2020)
11. 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. *Renew. Energy* **155**, 1312–1327 (2020)
12. Chen, Y., Avitabile, P., Dodson, J.: Data consistency assessment function (DCAF). *Mech. Syst. Signal Process.* **141**, 106688 (2020)
13. Chen, Y., Logan, P., Avitabile, P., et al.: Non-model based expansion from limited points to an augmented set of points using Chebyshev polynomials. *Exp. Tech.* **43**(5), 521–543 (2019)
14. 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. *Neurocomputing* **315**, 246–260 (2018)
15. 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. *Renew. Energy* **152**, 754–769 (2020)
16. Yan, X., Jia, M.: A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neurocomputing* **313**, 47–64 (2018)
17. Zhang, S., Zhang, Y.D.: Low-rank Hankel matrix completion for robust time-frequency analysis. *IEEE Trans. Signal Process.* **68**, 6171–6186 (2020)
18. Zhang, S., Zhang, Y.D.: Robust time-frequency analysis of multiple FM signals with burst missing samples. *IEEE Signal Process. Lett.* **26**(8), 1172–1176 (2019)
19. 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. *Struct. Health Monit.* **20**, 151–172 (2020)
20. 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. *Renew. Energy* **168**, 972–980 (2021)
21. Wang, C.: A sample entropy inspired affinity propagation method for bearing fault signal classification. *Digit. Signal Process.* **102**, 102740 (2020)
22. 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. *Measurement* **172**, 108891 (2021)
23. 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. *IEEE Access* **6**, 350 (2018)
24. Rostaghi, M., Ashory, M.R., Azami, H.: Application of dispersion entropy to status characterization of rotary machines. *J. Sound Vib.* **438**, 291–308 (2019)
25. Zheng, J., Tu, D., Pan, H., et al.: A refined composite multivariate multiscale fuzzy entropy and Laplacian score-based fault diagnosis method for rolling bearings. *Entropy* **19**(11), 585 (2017)
26. Zhu, K., Jiang, X., Chen, L., et al.: Performance degradation assessment of rolling element bearings using improved fuzzy entropy. *Meas. Sci. Rev.* **17**(5), 219 (2017)
27. Rostaghi, M., Azami, H.: Dispersion entropy: a measure for time-series analysis. *IEEE Signal Process. Lett.* **23**(5), 610–614 (2016)
28. Zheng, J., Pan, H., Yang, S., et al.: Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **99**, 229–243 (2018)
29. Wang, Z., Yao, L., Cai, Y.: Rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine. *Measurement* **156**, 107574 (2020)
30. Azami, H., Rostaghi, M., Abásolo, D., et al.: Refined composite multiscale dispersion entropy and its application to biomedical signals. *IEEE Trans. Biomed. Eng.* **64**(12), 2872–2879 (2017)
31. 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. *ISA Trans.* **81**, 329–341 (2018)
32. Berredjem, T., Benidir, M.: Bearing faults diagnosis using fuzzy expert system relying on an improved range overlaps and similarity method. *Expert Syst. Appl.* **108**, 134–142 (2018)
33. Gan, X., Lu, H., Yang, G.: Fault diagnosis method for rolling bearings based on composite multiscale fluctuation dispersion entropy. *Entropy* **21**(3), 290 (2019)
34. 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. *Struct. Health Monit.* **20**, 1169–1195 (2020)
35. Shi, L., Du, L., Shen, Y.D.: Robust spectral learning for unsupervised feature selection. In: 2014 IEEE International Conference on Data Mining, pp. 977–982. IEEE (2014)
36. Lei, Y., Yang, B., Jiang, X., et al.: Applications of machine learning to machine fault diagnosis: a review and roadmap. *Mech. Syst. Signal Process.* **138**, 106587 (2020)
37. Peña, M., Cerrada, M., Alvarez, X., et al.: Feature engineering based on ANOVA, cluster validity assessment and KNN for fault diagnosis in bearings. *J. Intell. Fuzzy Syst.* **34**(6), 3451–3462 (2018)
38. 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. Cent. South Univ.* **26**(6), 1607–1618 (2019)
39. Zheng, J., Dong, Z., Pan, H., et al.: Composite multi-scale weighted permutation entropy and extreme learning machine based intelligent fault diagnosis for rolling bearing. *Measurement* **143**, 69–80 (2019)

40. Li, Y., Wang, S., Deng, Z.: Intelligent fault identification of rotary machinery using refined composite multi-scale Lempel-Ziv complexity. *J. Manuf. Syst.* (2020). <https://doi.org/10.1016/j.jmsy.2020.05.004>
41. Wei, Y., Yang, Y., Xu, M., et al.: Intelligent fault diagnosis of planetary gearbox based on refined composite hierarchical fuzzy entropy and random forest. *ISA Trans.* **109**, 340–351 (2021)
42. He, C., Wu, T., Gu, R., et al.: Rolling bearing fault diagnosis based on composite multiscale permutation entropy and reverse cognitive fruit fly optimization algorithm-extreme learning machine. *Measurement* **173**, 108636 (2021)
43. 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. *Measurement* **176**, 108977 (2021)
44. 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. *Entropy* **23**(2), 259 (2021)
45. Yao, W., Wang, J.: Double symbolic joint entropy in nonlinear dynamic complexity analysis. *AIP Adv.* **7**(7), 075313 (2017)
46. Li, Y., Liang, X., Xu, M., et al.: Early fault feature extraction of rolling bearing based on ICD and tunable Q-factor wavelet transform. *Mech. Syst. Signal Process.* **86**, 204–223 (2017)
47. Ghaderi, P., Amini, F.: A new method for online identification of civil structures: virtual synchronization. *Int. J. Adapt. Control Signal Process.* **33**(1), 16–38 (2019)
48. Chen, Y., Avitabile, P., Page, C., et al.: A polynomial based dynamic expansion and data consistency assessment and modification for cylindrical shell structures. *Mech. Syst. Signal Process.* **154**, 107574 (2021)
49. Zheng, J., Pan, H.: Use of generalized refined composite multiscale fractional dispersion entropy to diagnose the faults of rolling bearing. *Nonlinear Dyn.* **101**(2), 1417–1440 (2020)
50. 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. Sound Vib.* **428**, 72–86 (2018)
51. Cao, P., Zhang, S., Tang, J.: Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning. *IEEE Access* **6**, 26241–26253 (2018)
52. Xu, G., Liu, M., Jiang, Z., et al.: Online fault diagnosis method based on transfer convolutional neural networks. *IEEE Trans. Instrum. Meas.* **69**(2), 509–520 (2019)
53. Yan, X., Jia, M.: Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection. *Knowl. Based Syst.* **163**, 450–471 (2019)
54. Available online: <https://github.com/Gearboxdata/Gear-Box-Fault-Diagnosis-Data-Set>. Accessed 05 Nov 2020

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