Journal of Manufacturing Systems 61 (2021) 725–735

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Technical Paper

Intelligent fault identification of rotary machinery using refined composite multi-scale Lempel–Ziv complexity

Yongbo Li *, Shun Wang, Zichen Deng

MIIT Key Laboratory of Dynamics and Control of Complex System, School of Aeronautics, Northwestern Polytechnical University, Xi'an, Shanxi 710072, China

ARTICLE INFO

Keywords:
Lempel–Ziv complexity
Rotary machinery
Intelligent fault diagnosis
Feature extraction

ABSTRACT

Fault diagnosis of rotary machinery plays a significant role in the prognostic and health management system, which aims to identify the root causes of system failures and provide effective information for prognostics and maintenance. Recently, Lempel–Ziv complexity (LZC) method has been employed for fault diagnosis of rotary machinery. However, one actual problem is that LZC fails to account for the multiscale information inherent in measured vibration signals. We first introduce a method to compute the multi-scale LZC for a signal. However, the variance of LZC values becomes larger as the scale factor increases. To solve this actual problem, this paper proposes refined composite multi-scale Lempel-Ziv complexity (RCMLZC) to estimate the complexity. We find that the proposed RCMLZC method consistently yields better performance when analyzing three simulated noisy and impulsive signals. Based on RCMLZC, a novel intelligent fault diagnosis method is designed to recognize various fault types of rotating machinery. Comparative experiments are performed to confirm the effectiveness of proposed method including single fault and compound fault working conditions. Experimental results indicate that RCMLZC is more accurate than multi-scale LZC, multi-scale entropy, and LZC in extracting fault features from vibration signal and that RCMLZC performs best to recognize the various fault types of rotary machinery.

1. Introduction

The maintenance plays a vitally important role in industrial applications, which has attracted deep attention from both expert and practical maintenance [1–3]. Prognostic and health management (PHM) through the collaborative community has been applied for more effective and efficient maintenance. Due to its undeniable importance, the past decades have witnessed a rapid progress of PHM in condition monitoring and fault diagnosis [4–6]. Fault diagnostics is one of the major tasks in the PHM system, which aims to diagnose and identify the root causes of system failures. The root causes identified can provide effective information for prognostics and maintenance as well as feedback for system design optimization [7].

Rotary machinery has been widely used in civilian, industrial, and military applications. However, their key components, such as bearings and gears are prone to damage due to the tough working conditions [8,9]. The fault diagnostics of rotary machinery allows properly scheduled shutdowns to reduce unnecessary maintenance operations and enhance the reliability of rotary machinery. Therefore, in view of the high impact and extreme costs usually associated with failures, the fault diagnostics of rotary machinery has attracted considerable interests from both academia and industry fields [9–13].

The key step in the fault detection and isolation (FDI) is the fault feature extraction [14,15]. In order to separate the periodical impulses from the measured vibration signals, several advanced signal processing approaches have drawn great attention, such as spectral kurtosis [16,17], correlation coefficient analysis [18], wavelet transform [19], empirical mode decomposition [20], entropy-based methods [21,22], and deep learning techniques [23,24]. In recent years, the theory of complexity proposed by Lempel and Ziv [25] has received considerable attention. The Lempel–Ziv complexity (LZC) method has been successfully applied to analyze time series generated from various nonlinear dynamics systems [26–29]. For example, Hong et al. [30] combined continuous wavelet transform with LZC to construct a new index to recognize the bearing fault severities of both inner race and outer race faults. Kedadouche et al. [31] demonstrated LZC is an effective tool to detect the early fault of gears compared with approximate entropy and sample entropy methods. Cui et al. [32] proposed an FDI method for bearing based on sparsogram and LZC indicator, which gives better diagnosing results than the traditional LZC method. Yin et al. [33] proposed an improved Lempel–Ziv method based on symbolic aggregate

* Corresponding author.
E-mail address: yongbo@nwpu.edu.cn (Y. Li).

https://doi.org/10.1016/j.jmsy.2020.05.004
Received 13 December 2019; Received in revised form 7 May 2020; Accepted 8 May 2020
Available online 4 June 2020

Nomenclature

WGN	white Gaussian noise
LZC	Lempel–Ziv complexity
MLZC	multi-scale Lempel–Ziv complexity
RCMLZC	refined composite multi-scale Lempel–Ziv complexity
MSE	multi-scale entropy
SVM	support vector machines
PHM	Prognostic and Health Management
FDI	fault detection and isolation

approximation for FDI of bearing, which can reflect the modulation information of bearing and enhance the diagnosing performance compared with traditional LZC method.

However, these conventional LZC-based FDI methods have one inherent disadvantage. The fault information is usually embedded in different time-scale domains [34]. The LZC-based FDI methods only perform single-scale analysis, which cannot describe the fault characteristics comprehensively. In order to match the fault characteristic distribution of rotary machinery, we combined the multi-scale analysis [35] and LZC named multi-scale LZC (MLZC) to extract the fault features over multiple scales. However, the length of time series will decrease dramatically as the scale factor increases, which will cause the poor stability performance of MLZC in analyzing the short time series. To overcome this shortcoming, we further propose refined composite multi-scale LZC (RCMLZC) to reduce the variance of the LZC values at larger scales. After the feature extraction, we utilize support vector machines (SVM) [36,37] to recognize several fault types of rotary machinery. The performance of proposed RCMLZC method is validated using both synthetic signals and experimental signals. To illustrate the advantage of our proposed RCMLZC in feature extraction, multi-scale entropy (MSE) [35,38] is also applied to process the experimental signals. Results demonstrate that our proposed RCMLZC method outperforms MLZC, MSE, and LZC in extracting fault characteristics with high stability.

The remainder of this paper is organized as follows. Section 2 briefly reviews the LZC algorithm and introduces our proposed RCMLZC algorithm. Moreover, the superiority of RCMLZC is validated using three simulated noisy and impulsive signals and compared with MLZC and LZC. Section 3 describes main steps of proposed intelligent fault diagnosis method. Section 4 provides experimental validations using two case studies. Section 5 gives concluding remarks of this paper.

2. The proposed RCMLZC method

2.1. Lempel–Ziv complexity

LZC was firstly proposed by Lempel and Ziv [25] in 1976, which has been proven to be an efficient tool to measure the complexity of a given time series. LZC consists of two basic operations: copy and insert [30]. For a given finite sequence $x(t)$ with length N, five steps are needed in the LZC calculation process:

(1) Cover the finite sequence $x(t)$ into 0–1 sequence by comparing with the median value T_d using Eq. (1).

$$s_i = \begin{cases} 0, & \text{if } x(i) < T_d \\ 1, & \text{otherwise} \end{cases} \tag{1}$$

The median is chosen for the symbolic process due to its robustness to any deviant value within the time series [28,39]. Then, we can get the symbol series $S_N = \{s_1 s_2 \dots s_N\}$.

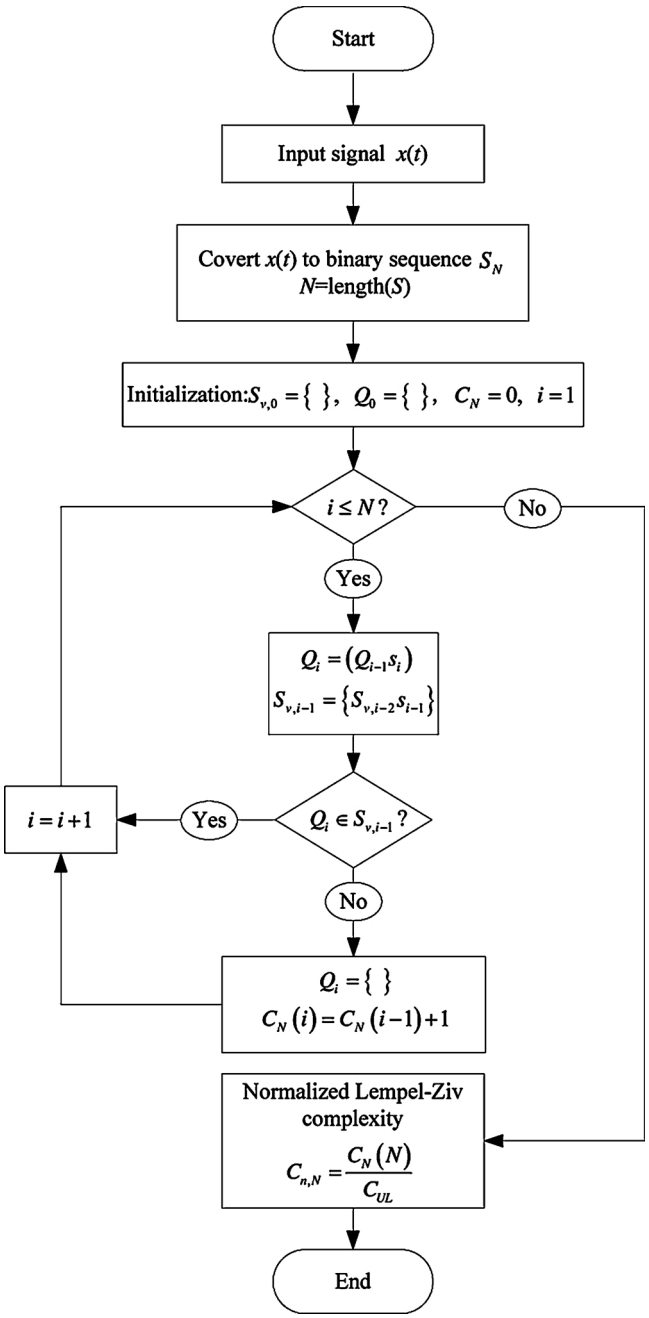

Fig. 1. Flowchart of the Lempel–Ziv complexity.

(2) Set the initial value $S_{v,0} = \{\}, Q_0 = \{\}, C_N(0) = 0$, and $i = 1$. Note that S_v and Q represent the substrings of the symbol series S_N, and C_N represents complexity counter.

(3) Let $Q_i = \{Q_{i-1} s_i\}$ and determine whether Q_i belongs to $S_{v,i-1} = \{S_{v,i-2} s_{i-1}\}$. If so, set $C_N(i) = C_N(i-1)$ and $i = i+1$. Otherwise, set $Q_i = \{\}, C_N(i) = C_N(i-1) + 1$, and update $i = i+1$.

(4) Repeat Step (3) until the end of symbol series $\{s_1 s_2 \dots s_N\}$, and then the $C_N(N)$ can be obtained. $C_N(N)$ is the last complexity counter, which represents the total number of distinct words in the whole sequence S_N.

(5) Normalize the $C_N(N)$ to get relatively independent indicator $C_{n,N}$ using Eqs. (2) and (3):

$$C_{n,N} = \frac{C_N(N)}{C_{\mathrm{UL}}} \tag{2}$$

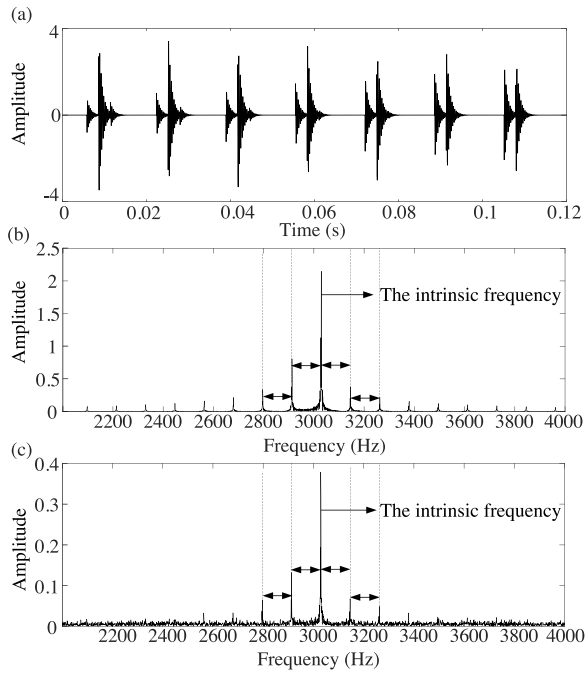

Fig. 2. (a) The waveforms of original vibration signal in Eq. (4), (b) the FFT spectrum of raw signal, (c) the FFT spectrum of 0–1 sequence.

$$C_{\mathrm{UL}} = \lim_{N\to\infty} C_N(N) \approx \frac{N}{log_2 N} \tag{3}$$

The above calculating procedures of LZC are illustrated in Fig. 1. It should be noted that simple amplitude compression of a signal can reserve the main useful information of the original frequency distribution characteristics. A fault bearing model reported in [40] is employed as an example. The mathematical expression of the bearing model is expressed in Eq. (4):

$$\begin{cases} x(t) = \sum_{i=1}^{M} A_i s(t - \mathrm{iT} - \tau_i) + w(t) \\ A_i = A_0 cos(2\pi \mathrm{Qt} + \phi_A) + C_A \\ s(t) = e^{-\mathrm{Bt}} sin(2\pi f_n t + \phi_w) \end{cases} \tag{4}$$

Note that A_i is the amplitude modulation signal with the period $1/Q$, A_0 is the amplitude of the signal, C_A is a constant with restriction $C_A > A_0$, $s(t)$ is the discrete oscillating impulse signal with an interval time T between two adjacent impacts, B is the damping coefficient, f_n is the natural frequency of the system, τ_i is the time lag derived from the random slip of rolling elements, and $w(t)$ is the white Gaussian noise.

The time-domain waveforms of the simulated signal and its corresponding frequency spectrum are displayed in Fig. 2(a) and (b), respectively. From Fig. 2(b), the intrinsic frequency and its sidebands can be clearly observed. Note that the fault frequency is equal to the sideband interval frequency. We convert the time series into 0–1 sequence and conduct FFT analysis. The obtained frequency spectrum is shown in Fig. 2(c). As can be seen, the sidebands of intrinsic frequency in the raw signal are not discarded and the frequency distribution characteristics are well reserved. This phenomenon shows that the binarization process can reserve the fault information effectively.

2.2. Refined composite multi-scale Lempel–Ziv complexity

Although LZC shows powerful performance in estimating the complexity for a given time series [32,33], LZC is facing challenge in processing the vibration signals measured from complex rotary machinery. Because LZC is a single-analysis approach, however, the fault features of complex rotary machinery are embedded in multiple time-scale domains. To match the fault features well, we firstly combine the concept of multi-scale [35] with LZC, called multi-scale LZC (MLZC).

However, multi-scale analysis of MLZC can shorten the length of time series. For example, for a given time series $x(t)$ with $N = 2000$, the length of will decrease into $N = 100$ when the scale factor $\tau = 20$. This phenomenon will cause the variance of MLZC becomes larger as the scale factor τ increases. To address this shortcoming, the refined composite multi-scale LZC (RCMLZC) is introduced in this paper. Four steps are required in computing the RCMLZC value.

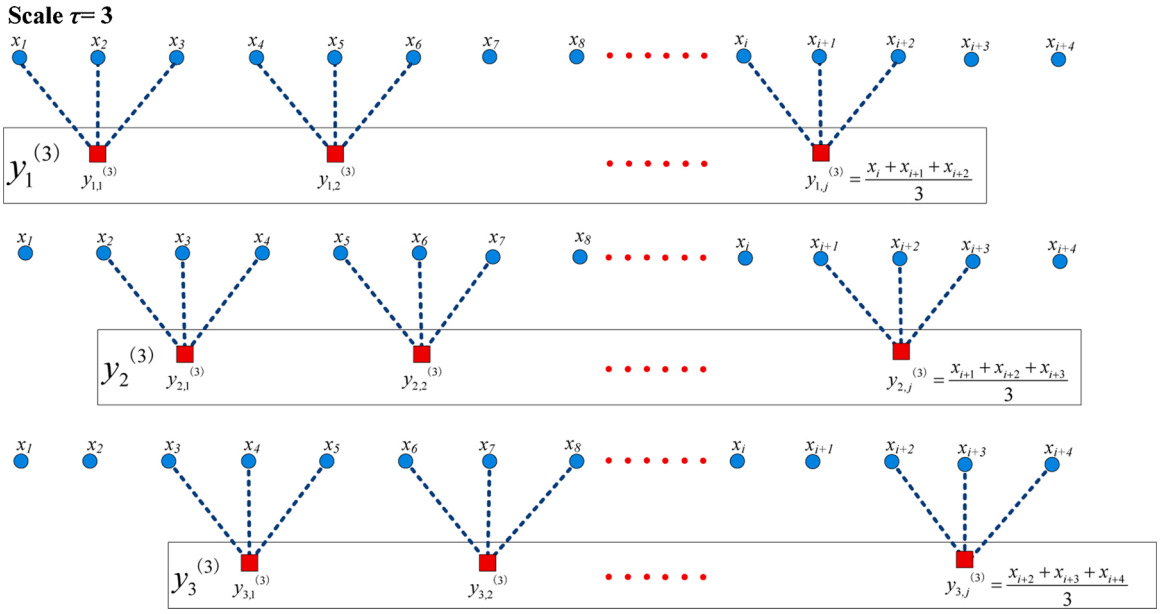

Fig. 3. Schematic illustration of the coarse-grained procedure for $\tau = 3$.

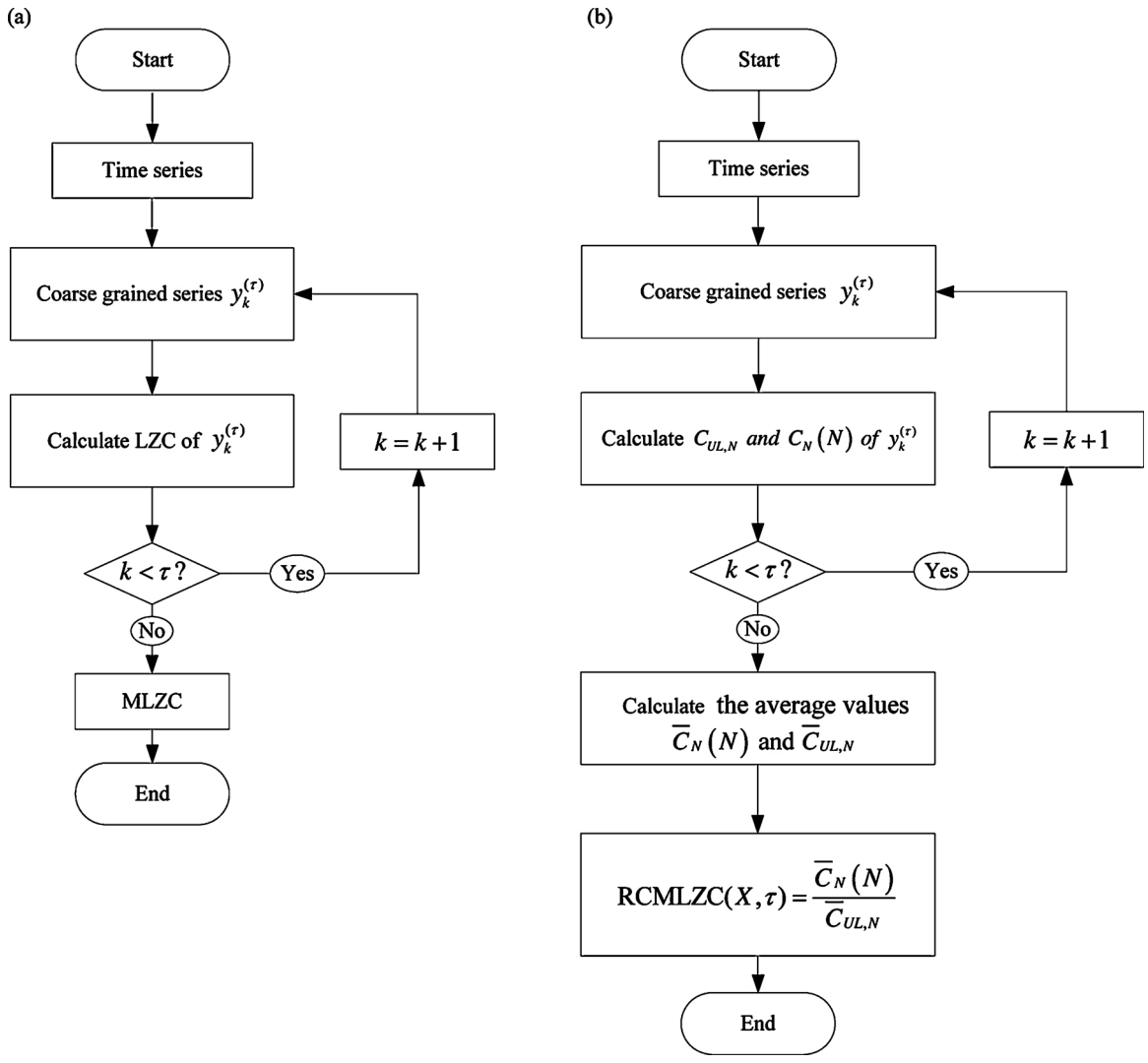

Fig. 4. (a) Flowchart of the MLZC method, (b) flowchart of the RCMLZC method.

(5) Conduct the refined composite multi-scale analysis. For a given time series $X\{x(k), k = 1, 2, \ldots, N\}$, we first conduct the coarse-graining analysis using Eq. (5). Then, the original time series $\{x_k\}$ is divided into τ coarse-grained time series $y_k^{(\tau)} = \{y_{k,1}^{(\tau)}\ y_{k,2}^{(\tau)}\ \cdots\ y_{k,p}^{(\tau)}\}, 1 \le k \le \tau$, where τ is the scale factor. Fig. 3 shows that how to calculate the τ coarse-grained time series when $\tau = 3$. Note that when $\tau = 1$, the time series $y^{(1)}$ is the original time series $\{x_k\}$:

$$y_{k,j}^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i, \quad 1 \le j \le \frac{N}{\tau}, \quad 1 \le k \le \tau \tag{5}$$

(6) Convert the obtained coarse-grained time series $y_k^{(\tau)} = \{y_{k,1}^{(\tau)}\ y_{k,2}^{(\tau)}\ \cdots\ y_{k,p}^{(\tau)}\}$ into symbolic series. Following the procedures in Fig. 1, we can obtain τ $C_N(N)$ and C_{UL} values over different scales.

(7) Calculate the average values $\overline{C}_N(N)$ and $\overline{C}_{\mathrm{UL}}$ for all $C_N(N)$ and C_{UL} using Eqs. (6) and (7), respectively:

$$\overline{C}_N(N) = \frac{1}{\tau}\sum_{k=1}^{\tau} C_N(N, y_k^{(\tau)}) \tag{6}$$

$$\overline{C}_{\mathrm{UL}} = \frac{1}{\tau}\sum_{k=1}^{\tau} C_{\mathrm{UL}}(y_k^{(\tau)}) \tag{7}$$

(8) The definition of RCMLZC can be expressed as:

$$\mathrm{RCMLZC}(X,\tau) = \frac{\overline{C}_N(N)}{\overline{C}_{\mathrm{UL}}} \tag{8}$$

For a better understanding of RCMLZC, Fig. 4(b) shows the main calculation process of RCMLZC. For comparison purpose, the flowchart of MLZC is also given in Fig. 4(a). Compared with MLZC, our RCMLZC method has the averaging process, thereby, the variance at lager scales can be significantly reduced.

2.3. Validation using simulated signals

In this subsection, three noisy and impulsive signals are adopted to verify the advantage of proposed RCMLZC in complexity estimation.

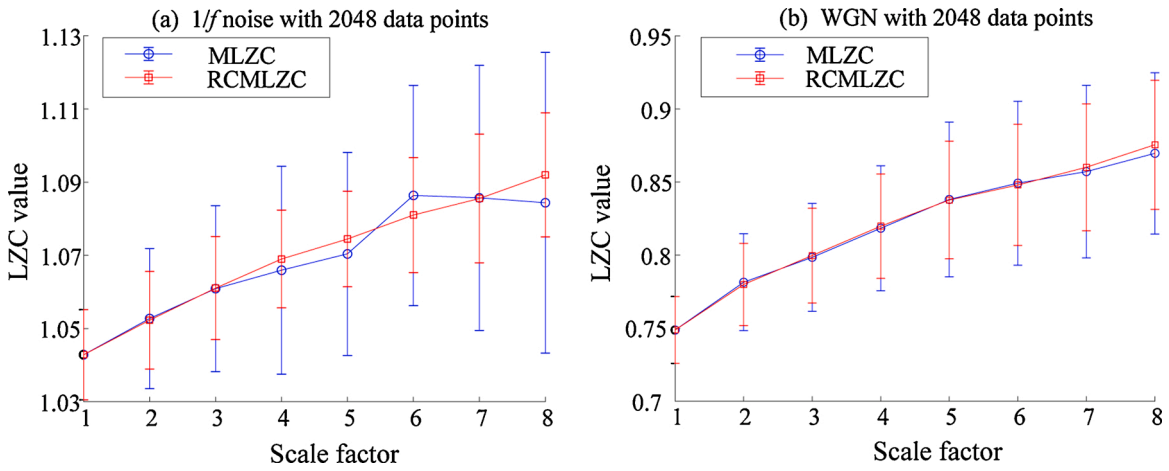

Fig. 5. Analysis results of 100 independent WGN and 1/*f* noise with 2048 data points using MLZC and RCMLZC method: (a) analysis results of 1/*f* noise, (b) analysis results of WGN.

Table 1
Standard deviations of the RCMLZC and MLZC of 1/*f* noise and WGN.

Noises	Method	Scale factor τ							
		1	2	3	4	5	6	7	8
1/*f* noise	RCMLZC	0.0123	0.0133	0.0141	0.0133	0.0131	0.0157	0.0176	0.0169
	MLZC	0.0123	0.0191	0.0226	0.0284	0.0277	0.0301	0.0362	0.0411
WGN	RCMLZC	0.0229	0.0281	0.0324	0.0357	0.0402	0.0414	0.0434	0.0441
	MLZC	0.0229	0.0332	0.0369	0.0427	0.0528	0.0561	0.0591	0.0552

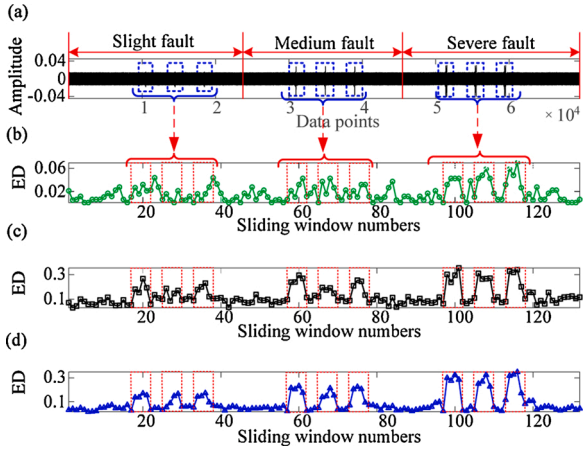

Fig. 6. Performance comparison results of LZC, MLZC and RCMLZC methods: (a) the simulated impulsive signal (b) the ED value of LZC, (b) the ED value of MLZC, (d) the ED value of RCMLZC.

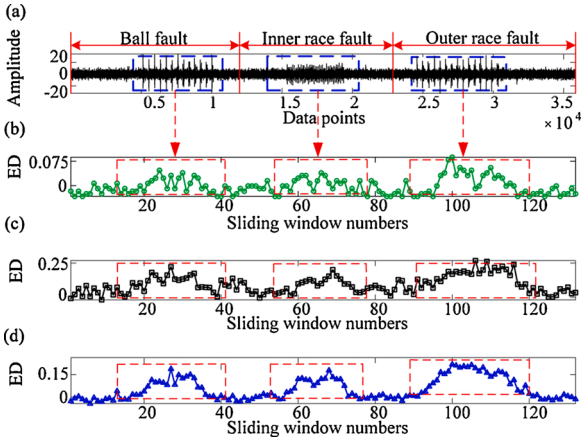

Fig. 7. Performance comparison results of LZC, MLZC and RCMLZC methods: (a) the simulated impulsive signal (b) the ED value of LZC, (b) the ED value of MLZC, (d) the ED value of RCMLZC.

2.3.1. *Simulated noisy signals*

White Gaussian noise (WGN) and 1/*f* noisy signals are used to show the stability advantage of proposed RCMLZC compared with MLZC. There are 100 groups of WGN and 1/*f* noise and each one has 2048 sample points. For comparison, we set scale factor $\tau = 8$ for MLZC and RCMLZC methods. The obtained results are shown in Fig. 5 and Table 1, respectively. Seen from Fig. 5 and Table 1, two conclusions can be drawn as follows. First, the proposed RCMLZC curve is smoother and steadier than the MLZC method, as depicted in Fig. 5. Second, the RCMLZC method has smaller error bars than the MLZC method, especially at larger scales. This phenomenon validates the proposed RCMLZC has the merit of stability.

2.3.2. *Simulated impulsive signals*

To validate the effectiveness of the RCMLZC in detecting gear fault severities, we use the simulated gear faulty signals with three crack fault severities, including slight fault, medium fault, and severe fault. The synthetic gear signal has 69,632 points, which is cut out using sliding windows of 2048 with a moving step of 512. Therefore, the number of sliding windows when the impulses occur can be calculated. The sliding window numbers for three fault severities are 18 to 37, 58 to 69, and 98 to 117, respectively. The time domains of the three simulated bearing faulty signals are illustrated in shown in Fig. 6(a). For comparison purposes, the LZC, MLZC, and RCMLZC are all utilized to process the impulsive signals. Euclidean distance (ED) value between the average of the first 10 samples (normal samples) and each sample is computed to estimate their fault detection ability. Here we set $\tau = 8$ for MLZC and

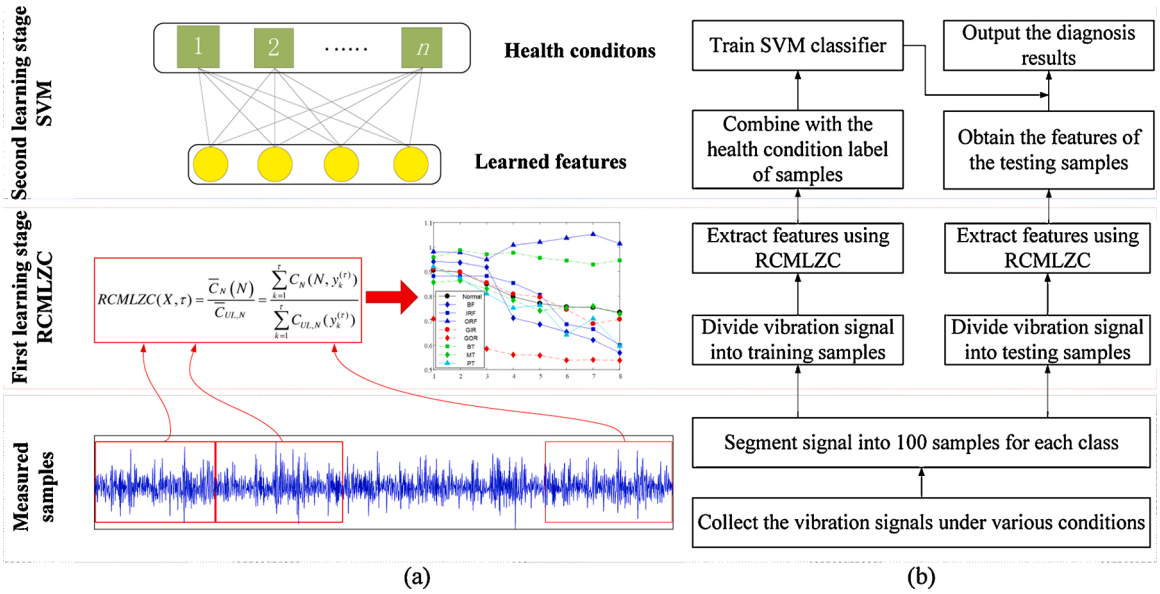

Fig. 8. Flowchart of the proposed RCMLZC-based intelligent fault diagnosis method: (a) illustration and (b) the flowchart of recognition process.

RCMLZC methods.

The obtained results are shown in Fig. 6. Seen from Fig. 6(b), it can be found that the original LZC method cannot detect the impulsive derived from three different fault severities. By contrast, MLZC and RCMLZC both generate higher ED values when the impulses occur. Moreover, our proposed RCMLZC has the least fluctuation in recognizing different gear faulty severities, as shown in Fig. 6(d). The phenomenon validates the ability of the proposed RCMLZC in impulsive detection.

The third simulated signal aims to test the performance of the proposed RCMLZC in recognizing bearing fault types. Here, we utilize three different bearing fault types, including ball fault (BF), inner race fault (IRF), and outer race fault (ORF). The number of sample points of synthetic signal is 35,840, as depicted in Fig. 7(a). Like above, the bearing fault signal is also cut out using sliding windows of 2048 with a moving step length of 256. The impulsive sliding window numbers for three bearing fault types are 14–40, 54–76, and 90–119, respectively. Also, we set the scale $\tau = 8$ for MLZC and RCMLZC methods.

The obtained results using three methods are shown in Figure 7(b)–(d), respectively. Seen from Fig. 7 (b), when periodical impulses are generated by ball fault and inner race fault, the LZC values represent large fluctuation. This phenomenon indicates the original LZC method has poor performance in distinguishing the noise and periodical impulses. Compared with LZC, MLZC and RCMLZC are able to track the periodical impulses. However, our proposed RCMLZC performs best with least fluctuation and highest sensitivity in impulse detection among three methods, as shown in Fig. 7(d). Therefore, we can draw the conclusion that the fault detection ability of the three methods is: RCMLZC > MLZC > LZC.

3. The proposed intelligent fault diagnosis method

There are two stages in the proposed RCMLZC-based intelligent fault diagnosis method. First, RCMLZC is employed to extract the fault features from the vibration signals of rotary machinery. Then, the classifier SVM is applied to classify different fault types. Five steps are required in the proposed fault diagnosis method as below.

Step 1. Collect vibration data for various conditions of rotary machinery and segment the signal into 100 samples for each class.
Step 2. Divide the datasets into training datasets and testing datasets.

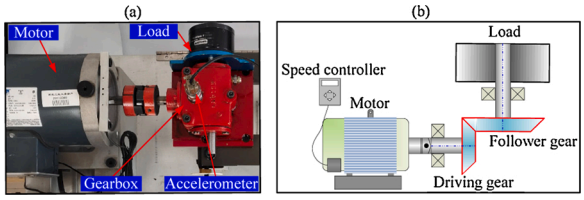

Fig. 9. (a) The machinery fault simulator system, (b) the layout of the test rig.

Step 3. Apply RCMLZC to extract fault information and obtain the fault features. Note that only one parameter of RCMLZC is set as $\tau = 8$.
Step 4. Train the SVM classifier using training features.
Step 5. Test the trained SVM classifier and accomplish the intelligent fault diagnosis of rotary machinery.

The flowchart of the proposed fault diagnosis method is shown in Fig. 8. It should be noted that, in this study, the radial basis function kernel is selected due to its universal application and good performance [41]. The kernel function plays a crucial role in SVM, which not only reduces the computational load but also solves the high-dimensional transformation effectively. The radial basis function kernel is expressed as $K(x_1, x_2) = exp(-\gamma\|x_1 - x_2\|^2)$, where $\gamma > 0$, and γ is the kernel parameter. It should be noted that, in the actual data process, the penalty parameter C and the kernel parameter γ of SVM are optimized by genetic algorithm for each multi-class SVM method. The parameters are set in the following intervals: C (0.1, 1000) and γ (0.001,10) following Ref. [34].

4. Experimental results

4.1. Description of lab setup and experimental setting

Experiments were conducted on rotary machinery called Spectra-Quest machinery fault simulator. The test rig is shown in Fig. 9(a) and (b). It consists of a reliance electric motor, a three-way gearbox with straight cut bevel gears, and rolling bearings. A magnetic clutch was also mounted at the rear of the gearbox for load generation. An accelerometer was installed on the top of the gearbox to collect the vibration signals. The sampling frequency was set 12,800 Hz and the rotating speed was 3000 rpm. In the experiments, the load torque was 5 in-lbs.

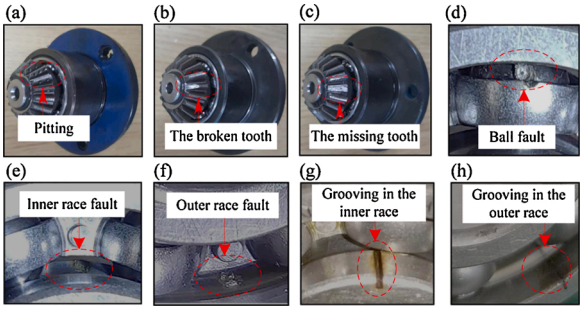

Fig. 10. Faulty gears and bearings: (a) pitting in the driving gear, (b) broken tooth in the driving gear, (c) missing tooth in the driving gear, (d) ball fault, (e) inner race fault, (f) outer race fault, (g) grooving in the inner race, (h) grooving in the outer race.

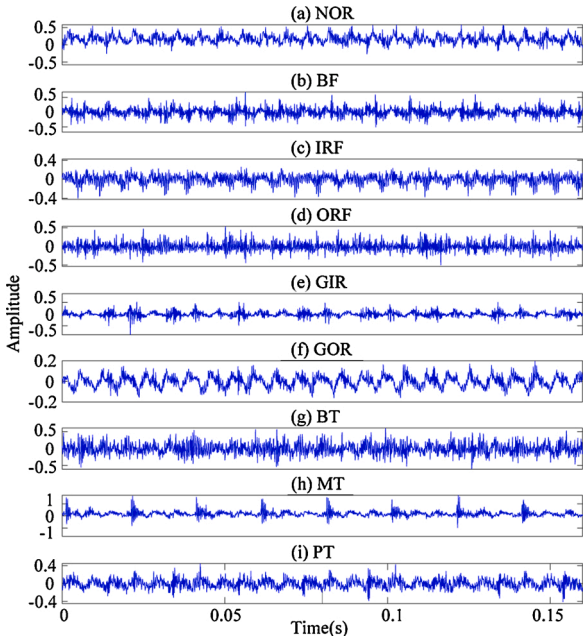

Fig. 11. The waveforms under nine healthy conditions in Experiment 1: (a) normal condition (NOR), (b) ball fault (BF), (c) inner race fault (IRF), (d) outer race fault (ORF), (e) grooving in the inner race(GIR), (f) grooving in the outer race(GOR), (g) broken tooth in the driving gear (BT), (h) missing tooth in the driving gear (MT), (i) pitting in the driving gear (PT).

Different faults were simulated by replacing the fault gear (including the pitting in the driving gear, the broken tooth in the driving gear, and the missing tooth in the driving gear, as shown in Fig. 10(a)–(c), respectively) and the fault bearing (including inner race fault, outer race fault, ball fault, grooving in the inner race, and grooving in the outer race, as shown in Fig. 10(d)–(h), respectively).

In this study, two experiments were designed for validations, which aim to test the diagnostic performance of RCMLZC in recognizing single faults and compound faults, respectively. In the two case studies, the MSE method is also used for comparison. For the purpose of a fair comparison, we set the scale $\tau = 8$ for the MSE, MLZC, and RCMLZC method. Meanwhile, the SVM is trained by 50% of samples, and the rest samples are utilized for performance test.

4.2. Experiment 1

Experiment 1 consists of one healthy condition and eight single fault conditions, including pitting in the driving gear (PT), broken tooth in the driving gear (BT), missing tooth in the driving gear (MT), grooving in the inner race(GIR), grooving in the outer race(GOR), inner race fault (IRF), outer race fault (ORF), and ball fault (BF). Each class owns 100 samples and there are total 900 samples (100 samples × 9 fault types). Meanwhile, the length of each sample was 2048 points. The waveforms under nine healthy conditions are showed in Fig. 11. Table 2 gives detailed information of nine healthy conditions, including class label, damage diameter, and the numbers of training and testing data.

Following the steps in Section 3, we utilize the proposed RCMLZC-SVM method to extract fault features. The obtained RCMLZC values with an average of 100 samples under nine working conditions are

Table 2
The detailed description of the experimental data with nine health conditions.

Health conditions	Class label	Damage diameter(mm)	Number of training data	Number of testing data
NOR	1	0	50	50
BF	2	0.01	50	50
IRF	3	0.01	50	50
ORF	4	0.01	50	50
GIR	5	0.2	50	50
GOR	6	0.2	50	50
BT	7	–	50	50
MT	8	–	50	50
PT	9	–	50	50

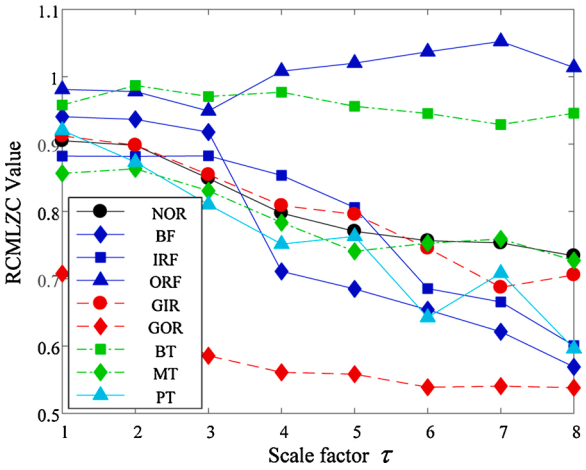

Fig. 12. RCMLZC over 8 scales with the average of 100 trials.

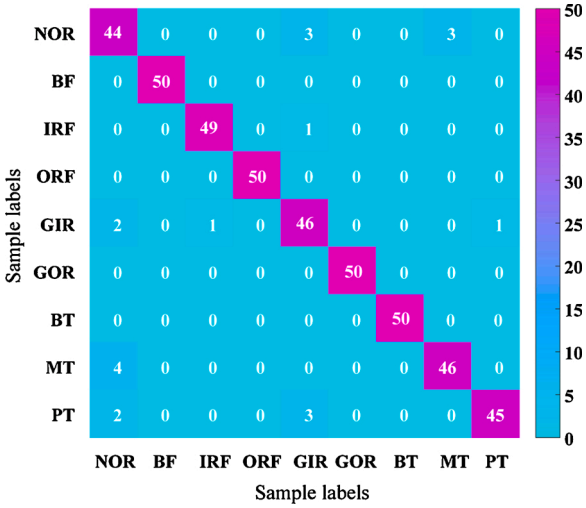

Fig. 13. Classification results of Experiment 1 using proposed RCMLZC-SVM method.

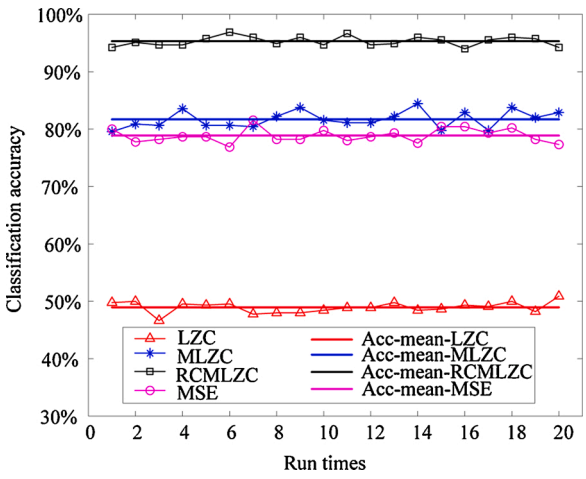

Fig. 14. Classification accuracies using four methods for 20 trial runs in Experiment 1.

presented in Figure 12. As can be seen, the RCMLZC at scale $\tau = 1$ (namely original LZC) cannot distinguish these single faulty conditions with overlapping values. By contrast, the RCMLZC values at larger scales (such as $\tau = 7$) can separate these working conditions well. This phenomenon demonstrates that the fault information is embedded in multiple time-scale domains, and only single-scale analysis is ineffective to extract the comprehensive fault information.

Then, the obtained features are fed into SVM for classification, and the obtained results are shown in Figure 13. It can be found that 20 samples are misclassified with the final accuracy of 95.56%. For comparison, the MLZC-SVM, MSE-SVM, and LZC-SVM are also tested. To avoid randomness, each method runs 20 times. Figure 14 and Table 3 illustrate the detailed recognition results using four methods. First, the proposed RCMLZC-SVM method obtains the highest average classification accuracy of 95.31% (ranging from 94% to 96.89%). This indicates RCMLZC has the best performance in extracting fault features among the four methods. Second, MLZC-SVM method has the second highest average classification accuracy of 81.70% (ranging from 79.56% to 84.44%), which is lower than that of the RCMLZC. This is because that RCMLZC has stability advantage comparing with the MLZC. Third, the LZC-SVM method has the lowest classification accuracies (ranging from 46.67% to 50.89%) due to the ineffectiveness of single-scale analysis. Lastly, the diagnosing ability of MSE-SVM is moderate with average accuracy of 78.88%.

4.3. Experiment 2

Experiment 2 aims to investigate the performance of RCMLZC in compound-fault diagnosis of rotating machinery. Experiment 2 is composed of seven compound-fault types, including normal condition (NOR), normal tooth in the driving gear with inner race fault (NI), normal tooth in the driving gear with outer race fault(NO), broken tooth in the driving gear with inner race fault (BI), broken tooth in the driving gear with outer race fault (BO),missing tooth in the driving gear with inner race fault (MI), and missing tooth in the driving gear with outer race fault (MO). Each class owns 100 samples and there are total 700 samples (100 samples × 7 fault types). Also, the length of each sample was 2,048 points. The waveforms under seven working types are showed in Figure 15. The detailed information of seven compound-fault conditions is shown in Table 4.

First, the proposed RCMLZC method is utilized to extract fault features. Fig. 16 shows the obtained RCMLZC values with an average of 100 samples. Seen from Fig. 16, RCMLZC values can be separated significantly at scale $\tau = 8$ compared with $\tau = 1$. This phenomenon demonstrates that the refined composite multi-scale procedure can enhance the ability of fault detection effectively.

Like Experiment 1, the classifier SVM is applied for classification. The classification results are shown in Fig. 17. Seen from Fig. 17, there are total 13 samples misclassified with an accuracy of 96.29%. For comparison, the MSE, LZC, and MLZC method are also tested and the testing accuracies of four methods are shown in Table 3 and Fig. 18, respectively. As can be seen, first, the existing MSE-SVM method has a mediocre performance with an average classification accuracy of 70.43%. Second, the single-scale analysis (namely LZC-SVM method), is not effective with an average classification accuracy of 40.77%. Third, combined with multi-scale analysis, the MLZC-SVM method is improved with the average classification accuracy of 85.21%. Lastly, the proposed RCMLZC-SVM method has the highest average classification accuracy of 97.24%. The above two experiments have proven that our proposed RCMLZC-based intelligent fault diagnosis method is effective in the FDI of rotary machinery.

Moreover, two-dimensional visualization of three methods: RCMLZC, MLZC, and MSE via the PCA method are depicted in Fig. 19. Seen from Fig. 19 (a), it can be found that the samples with the RCMLZC method are obviously separated from each other by the two-dimensional histograms. Nevertheless, the features are distributed without a nice

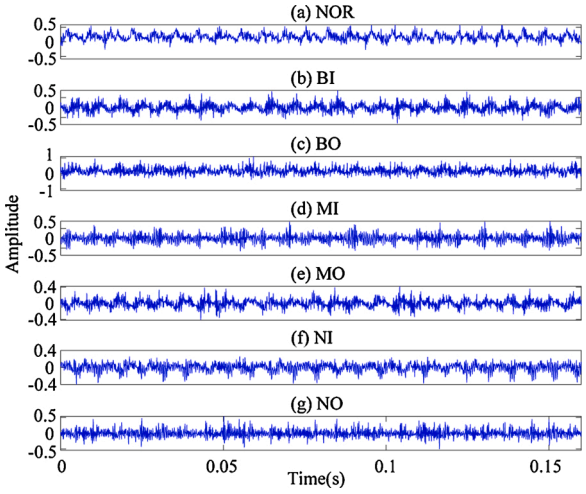

Fig. 15. The waveforms under seven healthy conditions in Experiment 2: (a) normal condition (NOR), (b) broken tooth in the driving gear with inner race fault (BI), (c) broken tooth in the driving gear with outer race fault (BO) (d) missing tooth in the driving gear with inner race fault (MI), (e) missing tooth in the driving gear with outer race fault (MO), (f) normal tooth in the driving gear with inner race fault (NI), (g) normal tooth in the driving gear with outer race fault(NO).

Table 3
Detailed classification accuracy of the experimental data sets in Experiment 1 and Experiment 2.

Experiments	RCMLZC-SVM			MLZC-SVM			LZC-SVM			MSE-SVM		
	Accuracy(%)			Accuracy(%)			Accuracy(%)			Accuracy(%)		
	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean
1	96.89	94	95.31	84.44	79.56	81.70	50.89	46.67	48.97	81.56	76.89	78.88
2	98.87	96.29	97.24	88.29	83.43	85.21	44.57	38	40.77	75.71	65.43	70.43

Table 4
The detailed description of the experimental data with seven health conditions.

Health conditions	Class label	Damage diameter (mm)	Number of training data	Number of testing data
NOR	1	0	50	50
BI	2	0.01	50	50
BO	3	0.01	50	50
MI	4	0.01	50	50
MO	5	0.01	50	50
NI	6	0.01	50	50
NO	7	0.01	50	50

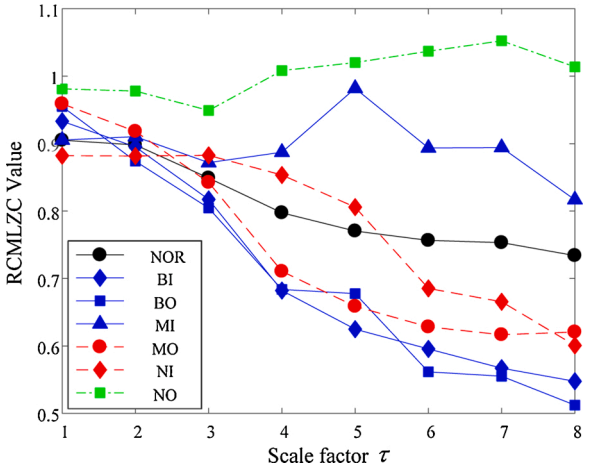

Fig. 16. RCMLZC over 8 scales with the average of 100 trials.

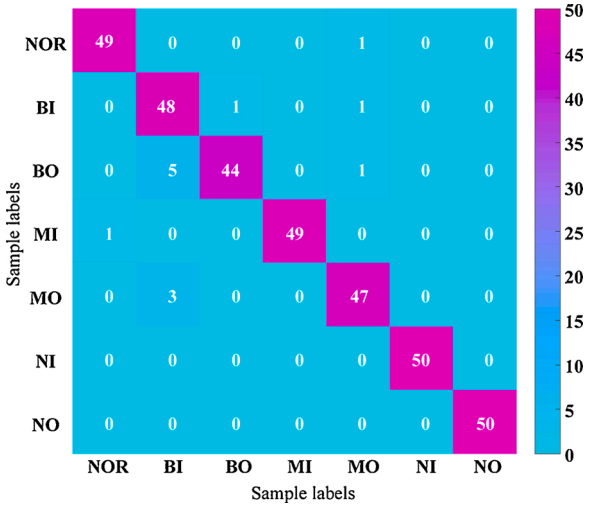

Fig. 17. Classification result of Experiment 2 using the proposed RCMLZC method.

cluster using MLZC and MSE method, as shown in Fig. 19 (b) and (c), respectively. This phenomenon indicates that the fault features extracted by RCMLZC contain much more useful fault information than the other two methods.

We next discuss the effect of scale factor τ in RCMLZC method. We test the classification accuracies and CPU time using different scale factor τ. Nine scale factors are tested: 1, 2, 3, 5, 7, 8, 9, 10, and 11. The obtained results are illustrated in Table 5. It can be found that a smaller scale τ will eliminate the classification accuracy, while a larger scale τ will enhance the CPU time. The scale $\tau = 8$ achieves the highest classification accuracy. Considering the computation efficiency, we select the

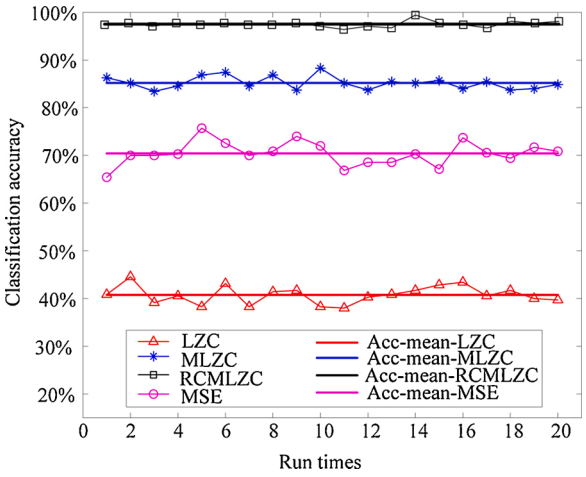

Fig. 18. Classification accuracy of the four methods for 20 trial runs in Experiment 2.

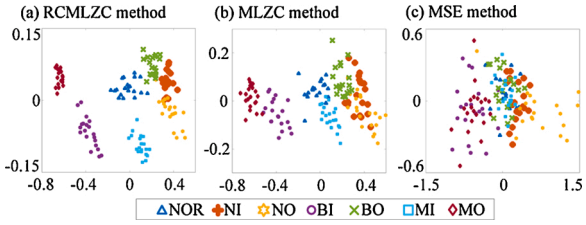

Fig. 19. Two-dimensional visualization using three methods in Experiment 2: (a) proposed RCMLZC method, (b) MLZC method, (c) MSE method.

$\tau = 8$ in this study.

We also test the performance of our proposed RCMLZC method using different percentages of samples for training (the remaining samples will be considered as testing samples). Eight percentages are tested: 10%, 20%, 30%, 40%, 50%, 60%, 70%, and 90%. To reduce randomness, 20 trials are conducted for each percentage. The averaging training and testing accuracies are calculated and their corresponding standard deviations are illustrated in Table 6 and Fig. 20, respectively. Three conclusions can be drawn from Table 6 and Fig. 20. First, since the fault features extracted using our proposed RCMLZC method represent good distinguishing ability, a high classification accuracy can be still obtained even under small sample condition. Second, as the training percentage increases, the average training accuracy shows an increasing trend, while the standard deviation shows a downward trend. For example, when the training percentage is 10%, the average training accuracy is 97.07% with the standard deviation 0.0166. When the training percentage increases to 90%, the average training accuracy is 98.35% with the standard deviation 0.0024. Note that standard deviation values may show small fluctuations for some training percentages due to the random repeating selection of training samples. This phenomenon illustrates that the classification model does not overfit the data. Third, it can be observed that the testing accuracy increases at a modest pace with the increase of training samples. When the percentage increases to 50%, the testing accuracy of the proposed method is 97.24% with the standard deviation 0.0032, which is a good recognition rate with a slight standard deviation. Considering the average accuracy and standard deviation, we select 50% of samples to demonstrate the advantage of our proposed RCMLZC method.

Finally, we calculate the time complexity of RCMLZC. Time complexity is an indicator to reflect the time consuming of the algorithm by countering the main calculation loop [42]. The time complexity of RCMLZC is related to the length N of time series and the scale parameter

Table 5
The mean classification accuracy and CPU time using RCMLZC with different scale factor τ.

Scale factor τ	1	2	3	5	7	8	9	10	11
Accuracy(%)	40.77	60.14	68.73	86.89	95.49	97.24	97.21	97.18	97.18
CPU time(s)	120.47	195.78	279.18	443.14	580.37	668.41	752.49	835.35	940.98

Table 6
The classification results of the proposed RCMLZC method using different percentages of samples for training.

Percentages of samples for training		10%	20%	30%	40%	50%	60%	70%	90%
Training accuracy	Average (%)	97.07	97.36	97.71	97.66	98.17	97.89	98.23	98.35
	Standard deviation	0.0166	0.0132	0.0105	0.0092	0.0048	0.0074	0.0037	0.0024
Testing accuracy	Average (%)	94.78	96.25	96.78	97.09	97.24	97.43	97.63	97.51
	Standard deviation	0.0098	0.0082	0.0086	0.0062	0.0032	0.0052	0.0049	0.0048

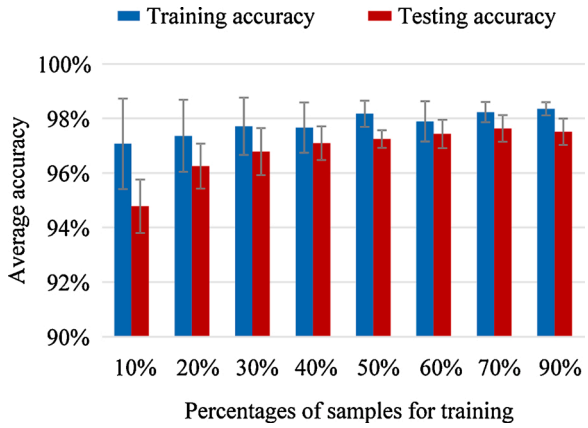

Fig. 20. The classification results of the proposed method using different percentages of samples for training.

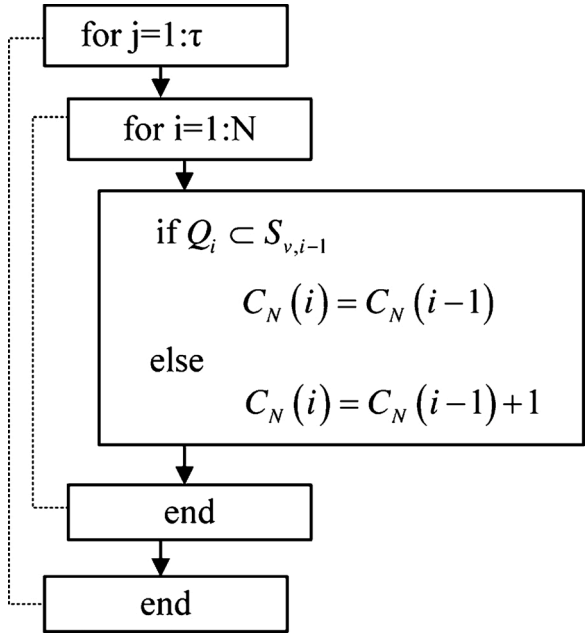

Fig. 21. Main calculation loop of RCMLZC method.

τ. The main calculation loop of RCMLZC methods is provided in Fig. 21. As can be seen, we can get $T = \tau N \sim O(N)$. Since the original LZC method has the complexity of $O(N)$, our proposed RCMLZC method takes more time than the original LZC method in estimating the complexity of time series.

5. Conclusions

A novel complexity analysis algorithm called RCMLZC is proposed to extract the fault characteristics of rotary machinery in this paper. The concept of RCMLZC is defined, and the shortcomings of the original LZC method caused by single analysis can be avoided by performing the refined composite multi-scale analysis. The effectiveness of RCMLZC is validated using a simulation and experimental signals. Results demonstrate that the proposed RCMLZC method has the best performance in the fault diagnostics of rotary machinery compared with MLZC, MSE, and LZC methods. In addition, this method might provide practical hints on application of LZC in condition monitoring for collaborative PHM.

Conflict of interest

None declared.

Acknowledgments

The research was supported by the National Natural Science Foundation of China under Grant 51805434, the China Postdoctoral Innovative Talent Plan, China under Grant BX20180257, the Postdoctoral Science Funds, China under Grant 2018M641021, and theKey Research Program, Shaanxi Province under Grant2019KW-017.

References

[1] Mobley RK. An introduction to predictive maintenance. Elsevier; 2002.
[2] Lin J, Kumar U, In2cloud:. A novel concept for collaborative management of big railway data. Front Eng Manag 2017;4(4):428–36.
[3] Ruschel E, Santos EAP, Loures EdFR. Industrial maintenance decision-making: a systematic literature review. J Manuf Syst 2017;45:180–94.
[4] Lee J, Wu F, Zhao W, Ghaffari M, Liao L, Siegel D. Prognostics and health management design for rotary machinery systems – reviews, methodology and applications. Mech Syst Signal Process 2014;42(1–2):314–34.
[5] Wang P, Yan R, Gao RX, et al. Virtualization and deep recognition for system fault classification. J Manuf Syst 2017;44:310–6.
[6] Zhang L, Lin J, Liu B, Zhang Z, Yan X, Wei M. A review on deep learning applications in prognostics and health management. IEEE Access 2019;7:162415–38.
[7] Tsui KL, Chen N, Zhou Q, Hai Y, Wang W. Prognostics and health management: a review on data driven approaches. Math Probl Eng 2015.
[8] Zhang L, Lin J, Karim R. Adaptive kernel density-based anomaly detection for nonlinear systems. Knowl-Based Syst 2018;139:50–63.
[9] Haidong S, Junsheng C, Hongkai J, Yu Y, Zhantao W. Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing. Knowl-Based Syst 2020;188:105022.
[10] Abdeljaber O, Avci O, Kiranyaz S, Gabbouj M, Inman DJ. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. J Sound Vibr 2017;388:154–70.
[11] Cai B, Liu H, Xie M. A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks. Mech Syst Signal Process 2016;80:31–44.

[12] Li Y, Wang X, Liu Z, Liang X, Si S. The entropy algorithm and its variants in the fault diagnosis of rotating machinery: a review. IEEE Access 2018;6:66723–41.
[13] Zhao R, Yan R, Gao RX. Dual-scale cascaded adaptive stochastic resonance for rotary machine health monitoring. J Manuf Syst 2013;32(4):529–35.
[14] Jardine AK, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 2006;20(7):1483–510.
[15] Zhang Y, Li X, Gao L, Wang L, Wen L. Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning. J Manuf Syst 2018;48:34–50.
[16] Sawalhi N, Randall R, Endo H. The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. Mech Syst Signal Process 2007;21(6):2616–33.
[17] Antoni J, Randall R. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. Mech Syst Signal Process 2006;20(2):308–31.
[18] Zhang Z, Verma A, Kusiak A. Fault analysis and condition monitoring of the wind turbine gearbox. IEEE Trans Energy Convers 2012;27(2):526–35.
[19] Peng Z, Chu F. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. Mech Syst Signal Process 2004;18(2):199–221.
[20] Lei Y, Lin J, He Z, Zuo MJ. A review on empirical mode decomposition in fault diagnosis of rotating machinery. Mech Syst Signal Process 2013;35(1–2):108–26.
[21] Zheng J, Cheng J, Yang Y, Luo S. A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination. Mech Mach Theory 2014;78:187–200.
[22] Li Y, Liang X, Wei Y, Wang X. A method based on refined composite multi-scale symbolic dynamic entropy and ISVM-BT for rotating machinery fault diagnosis. Neurocomputing 2018;315:246–60.
[23] Chen L, Xu G, Zhang S, Yan W, Wu Q. Health indicator construction of machinery based on end-to-end trainable convolution recurrent neural networks. J Manuf Syst 2020;54:1–11.
[24] Wang J, Ma Y, Zhang L, Gao RX, Wu D. Deep learning for smart manufacturing: Methods and applications. J Manuf Syst 2018;48:144–56.
[25] Lempel A, Ziv J. On the complexity of finite sequences. IEEE Trans Inform Theory 1976;22(1):75–81.
[26] Cui L, Gong X, Zhang J, Wang H. Double-dictionary matching pursuit for fault extent evaluation of rolling bearing based on the Lempel–Ziv complexity. J Sound Vibr 2016;385:372–88.
[27] Kedadouche M, Kidar T, Thomas M, Tahan A, El Badaoui M, Guilbault R. Combining EMD and Lempel–Ziv complexity for early detection of gear cracks. Proceedings of the international conference surveillance, vol. 7 2013:100–10.
[28] Aboy M, Hornero R, Abá:rtdcsolo D, Álvarez D. Interpretation of the lempel-ziv complexity measure in the context of biomedical signal analysis. IEEE Trans Biomed Eng 2006;53(11):2282–8.
[29] Zhang X-S, Roy RJ, Jensen EW. Eeg complexity as a measure of depth of anesthesia for patients. IEEE Trans Biomed Eng 2001;48(12):1424–33.
[30] Hong H, Liang M. Fault severity assessment for rolling element bearings using the Lempel–Ziv complexity and continuous wavelet transform. J Sound Vibr 2009;320(1–2):452–68.
[31] Kedadouche M, Thomas M, Tahan A, Guilbault R. Nonlinear parameters for monitoring gear: comparison between Lempel–Ziv, approximate entropy, and sample entropy complexity. Shock Vibr 2015.
[32] Cui L, Li B, Ma J, Jin Z. Quantitative trend fault diagnosis of a rolling bearing based on sparsogram and Lempel–Ziv. Measurement 2018;128:410–8.
[33] Yin J, Xu M, Zheng H. Fault diagnosis of bearing based on symbolic aggregate approximation and Lempel–Ziv. Measurement 2019;138:206–16.
[34] Li Y, Yang Y, Li G, Xu M, Huang W. A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mrmr feature selection. Mech Syst Signal Process 2017;91:295–312.
[35] Costa M, Goldberger AL, Peng C-K. Multiscale entropy analysis of complex physiologic time series. Phys Rev Lett 2002;89(6):68102.
[36] Widodo A, Yang B-S. Support vector machine in machine condition monitoring and fault diagnosis. Mech Syst Signal Process 2007;21(6):2560–74.
[37] Wang J, Liu S, Gao RX, Yan R. Current envelope analysis for defect identification and diagnosis in induction motors. J Manuf Syst 2012;31(4):380–7.
[38] Gao Q, Liu W, Tang B, Li G. A novel wind turbine fault diagnosis method based on intergral extension load mean decomposition multiscale entropy and least squares support vector machine. Renew Energy 2018;116:169–75.
[39] Abásolo D, Simons S, Morgado da Silva R, Tononi G, Vyazovskiy VV. Lempel-ziv complexity of cortical activity during sleep and waking in rats. J Neurophysiol 2015;113(7):2742–52.
[40] Li Y, Liang X, Xu M, Huang W. Early fault feature extraction of rolling bearing based on ICD and tunable q-factor wavelet transform. Mech Syst Signal Process 2017;86:204–23.
[41] Scholkopf B, Sung K-K, Burges CJ, Girosi F, Niyogi P, Poggio T, et al. Comparing support vector machines with gaussian kernels to radial basis function classifiers. IEEE Trans Signal Process 1997;45(11):2758–65.
[42] Wang Y-H, Yeh C-H, Young H-WV, Hu K, Lo M-T. On the computational complexity of the empirical mode decomposition algorithm. Physica A: Stat Mech Appl 2014;400:159–67.