Multiscale Symbolic Lempel–Ziv: An Effective Feature Extraction Approach for Fault Diagnosis of Railway Vehicle Systems

Yongbo Li, Fulong Liu, Shun Wang, and Jiancheng Yin

***Abstract*—In this article, a novel intelligent fault diagnosis method based on multiscale symbolic Lempel-Ziv (MSLZ) is proposed to identify several faults of railway vehicle systems (RVSs). The proposed MSLZ is essentially for the purpose of estimating the irregularity of a given time series. In the proposed MSLZ method, the symbolization and multiscale techniques are combined with Lempel–Ziv (LZ) to enhance its feature extraction ability. First, the symbolization can facilitate LZ to remove the noises and reserve the fault information. Second, multiscale analysis can extend LZ to multiple time scales, which can further enhance the description ability of dynamic characteristics. Using numerical data and experimental signals collected from RVSs, the performance of the MSLZ method is demonstrated to be sensitive to periodical impulses and robust to environmental noise. Moreover, it has been demonstrated that MSLZ has superiority in extracting fault information of the RVS compared with LZ, symbolic LZ, and multiscale LZ methods.**

***Index Terms*—Fault diagnosis, impulse detection, Lempel–Ziv (LZ), multiscale analysis, symbolic analysis.**

I. Introduction

CONDITION monitoring (CM) of railway vehicles has received much attention from both academic and industrial researchers. Due to high service load, harsh operation condition, or inevitable fatigue, a fault may occur on railway vehicle systems (RVS). CM of RVSs can allow operators to take remedial actions, reduce the economic loss, and avoid catastrophes [1]–[7].

In the past few decades, there has been a rapid development of vibration-based CM methods for RVSs. For instance, the short-time Fourier transform, Wigner–Ville transform, and wavelet transform (WT) were all employed to detect the wheel flat by analyzing vibrations of axle boxes [8]. Recently, an adaptive multiscale morphology filter (AMMF) was applied for wheel flat detection [9]. Results demonstrated that the AMMF can extract fault characteristics from axle-box vibration signals and diagnose wheel flat faults effectively. In addition, Kalman-filter-based methods were also popular in CM of RVSs [10]–[13]. Recently, a deep neural network has been constructed and utilized to diagnose eight conditions of a bogie belonging to a high-speed train [14].

However, most of these methods considered an RVS as a linear system, which is untrue in reality as the vehicle suspension systems are highly nonlinear. The measured vibration signals of the RVS often represent nonlinear and nonstationary characteristics due to the large dynamic force at the wheel–rail interface generated by rail and wheel faults, such as rail surface crack, corrugation, wheel flat [8], [15], and damping fault [16]. Furthermore, the suspension fault features are easily masked by environment noise and other interferences. Recently, some nonlinear dynamic approaches methods, such as sample entropy (SE) [17], permutation entropy (PE) [18], [19], and fuzzy entropy (FE) [20], have been developed to extract the fault information via estimating the complexity of time series [21], [22]. However, these commonly used entropy-based methods have their own disadvantages. SE and FE have low computation efficiency when processing big data [18]. PE, though computationally efficient, only considers the amplitude information of time series [23], [24].

Nowadays, increasing attentions have been paid on the application of Lempel–Ziv (LZ) for fault feature extraction. LZ was originally proposed by Lempel and Ziv, which is able to estimate the complexity and effectively detect the dynamic change of time series [25]. The main idea of LZ can be described into two basic operations: copy and insert [26]. Yan and Gao [27] applied LZ to extract bearing fault features. Hong and Liang [26] combined continuous WT and LZ to accomplish the fault severity identification of a bearing. Empirical mode decomposition and LZ were integrated for bearing fault identification by Dou and Zhao [28], which proposed local mean decomposition and LZ to recognize different bearing fault severities. LZ was also combined with machine learning methods for pattern identification. Xia *et al.* [29] introduced Hilbert vibration decomposition, LZ, and support vector machine to identify different bearing

Manuscript received October 16, 2019; revised January 22, 2020 and February 13, 2020; accepted March 5, 2020. Date of publication March 23, 2020; date of current version October 23, 2020. This work was supported by the National Natural Science Foundation of China under Grant 51805434. Paper no. TII-19-4649. *(Corresponding author: Yongbo Li.)*

Yongbo Li and Shun Wang are with the MIIT Key Laboratory of Dynamics and Control of Complex System, School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: yongbo@nwpu.edu.cn; 18809183733@163.com).

Fulong Liu is with the Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, U.K. (e-mail: u1553439@hud.ac.uk).

Jiancheng Yin is with the Department of Astronautical Science and Mechanics, Harbin Institute of Technology, Harbin 150001, China (e-mail: wdyydy@163.com).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2020.2980923

faults. Cui *et al.* [30] combined LZ and the double-dictionary matching pursuit to assess the fault severity of rolling bearings. Sparsogram and LZ were also integrated to diagnose bearing faults [31].

Although LZ shows remarkable effectiveness in the field of fault diagnosis, it has two main shortcomings. First, LZ converts the time series into 0–1 sequence using a predefined threshold [31], which will undoubtedly discard the amplitude and frequency information hidden in raw vibration signals. Second, LZ is a single-scale analysis method, which is difficult to estimate the complexity of a time series comprehensively [32]. To solve these drawbacks of the original LZ method, this article proposes a novel method, namely multiscale symbolic Lempel–Ziv (MSLZ). First of all, LZ is combined with symbolic dynamic and information theory [33], [34] to facilitate its ability of noise cancellation and fault information reservation (called symbolic Lempel–Ziv—SLZ). Then, we combine multiscale analysis with SLZ to extract features over different multiple scales [32], called MSLZ method, which can further enforce the description ability of dynamic characteristics. After the fault extraction using the MSLZ method, the multiclassifier extreme learning machine (ELM) is applied to accomplish the fault pattern identification [35]. The ELM is a supervised learning linear classification algorithm with good generalization ability. In addition, the ELM has a high computing efficiency, which is simple and easy to be implemented [35].

In order to validate the effectiveness of the proposed method, experiments were conducted on a one-fifth railway vehicle bogie at the University of Huddersfield. Multiple suspension faults were manually created in the experiments. It is worth noting that RVSs are highly nonlinear systems. When the fault occurs on RVSs, such as rail surface crack, corrugation, wheel flat [14], [15], and damping fault [16], the measured vibration signals of RVSs contain strong environment noises due to the long transfer path. Meanwhile, the vibration signals of RVSs generally show great variations in multiple observation scales. The multiscale characteristics (such as multiple time and frequency scales) are inherent in the measured vibration signals of RVSs. Since the original LZ method is a single-scale analysis approach with of the 0–1 encoding strategy, it is ineffective to extract the fault characteristics of RVSs. In our proposed MSLZ method, the symbolization and multiscale techniques are combined with LZ to enhance its feature extraction ability. Simulation and experimental results demonstrated that the proposed MSLZ outperforms LZ [25], SLZ, multiscale Lempel–Ziv (MLZ), and multiscale sample entropy (MSE) [32] in fault diagnosis of the RVS. In summary, we have three main contributions in this article.

1) The symbolization for noise cancellation is proposed for LZ to enhance its denoising ability.
2) The multiscale analysis is used to extend LZ to multiple time scales.
3) The proposed method yields the best fault diagnosis ability by comparing with MLZ-ELM, SLZ-ELM, LZ-ELM, and MSE-ELM methods.

The remainder of this article is organized as follows. Section II describes detailed steps of the proposed MSLZ method and illustrates the superiority of MSLZ via numerical simulations. Section III gives detailed procedures of the MSLZ and ELM methods. Section IV introduces the test rig and validates its effectiveness via fault diagnosis of the RVS. Finally, Section V concludes this article.

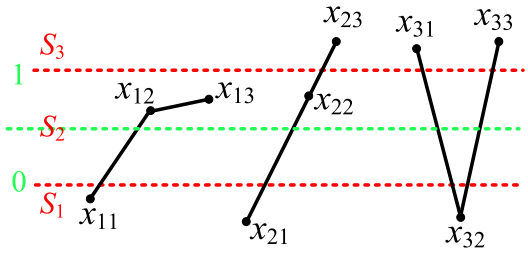

Fig. 1. Illustration of the shortcoming of the LZ method.

II. Multiscale Symbolic Lempel–Ziv

A. Shortcomings of Lempel–Ziv

The LZ method transforms the time series into a sequence of 0 and 1 [25]. However, the operation of an RVS may generate complicated dynamic responses due to various contacting modes between wheels and rails. The first shortcoming is that the usage of 0–1 encoding of the LZ method cannot effectively describe the amplitude and frequency information hidden in raw vibration signals [8].

For better understanding, three different vectors $X_1 = \{x_{11}, x_{12}, x_{13}\}$, $X_2 = \{x_{21}, x_{22}, x_{23}\}$, and $X_3 = \{x_{31}, x_{32}, x_{33}\}$ are selected for illustration, as shown in Fig. 1. By using the traditional LZ method with a preset threshold, as denoted in green line, the above three vectors have the same symbol time series $\{0, 1, 1\}$. It can be found that the information about the amplitude difference is discarded. Meanwhile, the usage of 0–1 encoding of the LZ method can only reserve the main frequency-domain characteristics; some frequency distribution characteristics related to the fault information may be discarded in raw vibration signals. A fault-bearing model reported in [36] is employed to illustrate the loss of frequency distribution characteristics. The mathematical expression of this bearing model is given as

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

where A_i is the amplitude modulation (AM) signal with a period time $1/Q$, 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

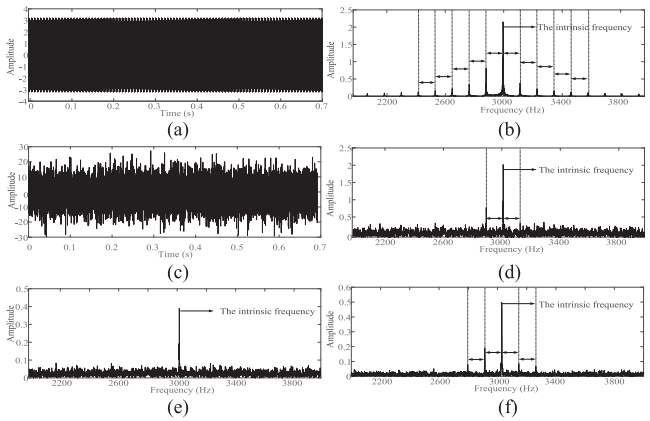

Fig. 2. (a) Waveforms of the original vibration signal in (1). (b) FFT spectrum of the raw signal. (c) Original vibration signal with additional noises. (d) FFT spectrum of the noisy signal. (e) FFT spectrum using the 0–1 encoding method. (f) FFT spectrum using the symbolization process.

frequency is equal to the sideband interval frequency. First, we convert the time series into 0–1 sequence and conduct fast Fourier transform (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 discarded using 0–1 encoding of the LZ method. 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. First, we add the additional noise with SNR $= -18$ dB on the raw signal and conduct FFT analysis. The time-domain waveforms of the noisy signal and its corresponding frequency spectrum are plotted in Fig. 2(c) and (d), respectively. As can be seen from Fig. 2(d), the sidebands of intrinsic frequency are heavily masked by the background noise, and only first harmonic of modulation frequency can be detected. This is because the energy of the periodic impacts is very low, which is buried in interference noise, leading to the spectrum line disappearing. Second, we convert the time series into 0–1 sequence and conduct FFT analysis. The obtained frequency spectrum is shown in Fig. 2(e). As can be seen, the sidebands of intrinsic frequency in the raw signal are discarded using 0–1 encoding of the LZ method.

Second, direct application of the LZ method may generate an unreliable result using a single scale to quantify the complexity of time series [37]. As known that the measured vibration signal from the RVS consists of multiple temporal scale structures, the feature extraction ability of the traditional LZ method will be largely weakened when processing such complicated vibration signals [38].

B. Multiscale Symbolic Lempel–Ziv

The key point for fault diagnosis of the RVS is to extract the fault characteristic [6], [12]. LZ is able to estimate the complexity and effectively detect the dynamic change of time series. However, LZ converts the time series into 0–1 sequence [31], which will undoubtedly discard the amplitude and frequency information. To overcome this shortcoming, we combine the symbolization process with LZ (called SLZ) to facilitate its ability of noise cancellation and fault information reservation. However, the multiscale characteristics (such as multiple time and frequency scales) are inherent in the measured vibration signals of the RVS. Direct application of SLZ will undoubtedly omit the fault information inherent in the raw vibration signal. There is a need to take multiscale information into consideration. Therefore, the MSLZ method is proposed by combining symbolization and multiscale analysis to facilitate its ability of feature extraction. In this article, we also demonstrate that the proposed MSLZ method outperforms SLZ and MLZ alone using both synthetic and experimental signals. Detailed explanations are given as follows.

1) SLZ Method: For a given time series $X\{x(k), k = 1, 2, \ldots, N\}$, five steps are required in the proposed SLZ method as follows.

1) Convert the obtained coarse-grained time series $X = \{x_1, x_2, \ldots, x_n\}$ into symbolic time series. For ε symbols, we can obtain a symbolic time series $S = \{s_1, s_2, \ldots, s_\varepsilon\}$ after symbolization. In this article, the maximum entropy partitioning [21], [39] is applied to accomplish the symbolization due to its adaptive segmentation. Fig. 1 gives insights into the advantage of symbolization. Here, we set the symbol number $\varepsilon = 3$ as denoted by the red line in Fig. 1. The obtained vectors will generate different patterns $\{S_1, S_2, S_2\}$, $\{S_1, S_2, S_3\}$, and $\{S_1, S_3, S_3\}$ using the symbolic dynamic process. In order to show the advantage of symbolization, Fig. 2(f) shows the FFT spectrum using the symbolization process. We find that the first and second harmonics of modulation frequency are well reserved with less interference frequencies. This phenomenon shows that the symbolization can facilitate LZ to remove the background noise and reserve the fault information.
2) Initialize $S_{v,0} = \{\}$, $Q_0 = \{\}$, $C_N = 0$, and $r = 1$. Let $Q_r = (Q_{r-1}s_r)$ and determine whether Q_r belongs to $S_{v,r-1} = \{S_{v,r-2}s_{r-1}\}$.
3) If Q_r belongs to $S_{v,r-1} = \{S_{v,r-2}s_{r-1}\}$, set $C_N(r) = C_N(r-1)$. Then, go to step 2 with $r = r + 1$. Else, set $Q_r = \{\}$ and $C_N(r) = C_N(r-1) + 1$. Then, go to step 2 with $r = r + 1$.
4) Repeat steps 2 and 3 until the whole symbol time series $S = \{s_1, s_2, \ldots, s_\varepsilon\}$ is covered.
5) Normalize the LZ complexity as follows:

$$0 \le C_{nN} = \frac{C_N(N)}{C_{UL}} \le 1 \tag{2}$$

where $C_{UL} = \lim_{N\to\infty} C_N(N) = \lim_{N\to\infty} \frac{N}{(1-\beta)\log_k N} \approx \frac{N}{\log_k N}$ and k is the number of symbols.

For clarity, a flowchart of the SLZ method is given in Fig. 6.

In order to validate the superiority of the SLZ method in the detection of dynamical change of time series, a widely used dynamical system, i.e., logistic map [40]–[42], is investigated.

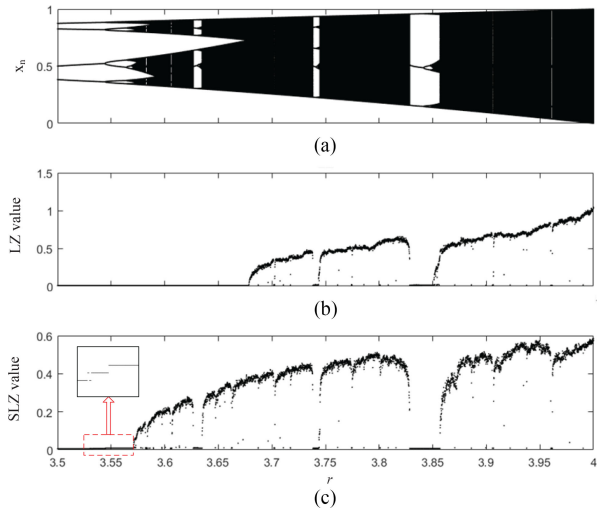

Fig. 3. Logistic map and its LZ and SLZ complexity evaluation values with control parameter r. (a) Logistic map. (b) LZ method of the logistic map. (c) SLZ method of the logistic map.

The definition of logistic map is expressed as follows:

$$x_{i+1} = r \times x_i \times (1 - x_i), \quad x_1 = 0.65 \tag{3}$$

where r is the control parameter with $3.5 \leq r \leq 4$.

Fig. 3(a) displays the logistic map. It is shown that the chaotic dynamic behavior occurs in the range of $3.573 \leq r \leq 4$. The results of LZ and SLZ in describing logistic map are displayed in Fig. 3(b) and (c), respectively. It can be found that when the chaotic dynamic behavior occurs in the range of $3.573 \leq r \leq 4$, SLZ can detect the dynamic change accurately. Moreover, the behavior of SLZ matches well with the positive Lyapunov exponents. However, the traditional LZ method is unchanged with zero from $3.5 \leq r \leq 3.68$. Moreover, the SLZ method is also sensitive to some small complexity fluctuations. For example, the time series of the logic map bifurcates from one status (period: -4) into another status (period: -8), when $r = 3.53$, we can find that the SLZ value increases correspondingly. The comparison results have demonstrated that the SLZ method has certain superiority in detecting the dynamic change of logic map compared with the original LZ method

$$X(t) = \begin{cases} 1.8 \times e^{\frac{1.2}{2048} \times t} \times \sin\left(\frac{100}{2048} \times t\right), & 0 < t < 400 \\ 1.8 \times e^{\frac{1.2}{2048} \times t} \times \sin\left(\frac{100}{2048} \times t\right) & \\ \quad + 0.12 \times \sin\left(\frac{800}{401} \times t\right), & 400 \leq t \leq 800 \\ 1.8 \times e^{\frac{1.2}{2048} \times t} \times \sin\left(\frac{100}{2048} \times t\right), & 800 < t < 2048 \end{cases} \tag{4}$$

To illustrate the advantage of the symbolization process in noise cancellation and information reservation, an AM signal with different frequency fluctuation is simulated to compare the performance of SLZ and traditional LZ. The expression of the AM signal is provided in (4). The time domain of the simulated signal is displayed in Fig. 4(b). Here, we set $\varepsilon = 6$, and the final obtained results using SLZ and LZ methods are shown in Fig. 4(a). As can be seen, the obtained values change along with the increase of amplitude and frequency. However,

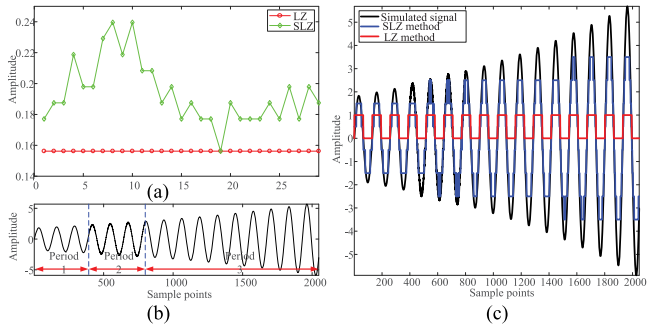

Fig. 4. Dynamic change tracking results using proposed SLZ and LZ methods. (a) SLZ and LZ values. (b) Time domains of the AM signal. (c) 0–1 coded using the LZ method and symbolization using the proposed SLZ method.

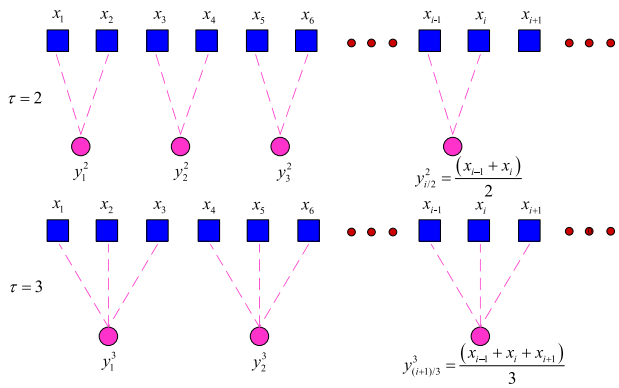

Fig. 5. Scheme illustrating the coarse-graining with $\tau = 2$ and $\tau = 3$.

the traditional LZ method keeps an approximately constant value. The comparison analysis results further demonstrate the superiority of the SLZ method in feature extraction.

To demonstrate better performance of SLZ in quantifying the complexity, we provide the coding results of SLZ and LZ methods, as shown in Fig. 4(c). As can be seen from Fig. 4(c), the SLZ method works well in tracking the dynamic change of simulated signals using the proposed encoding strategy. On the contrary, the 0–1 encoding form of the traditional LZ method is not effective in detecting the dynamic change, which is almost unchanged in the whole time series.

2) MSLZ Method: Vibration signals generally show great variations in multiple observation scales; thereby, it is necessary to take multiscale information into consideration [43]. It is well known that the low-frequency periodical impulses will be generated and modulated with the high-frequency component when the localized damage occurs. SLZ with single-scale analysis will undoubtedly omit the fault information inherent in the raw vibration signal.

To better extract the multiscale fault characteristics, the multiscale analysis proposed by Costa *et al.* [37] is combined with SLZ to assess the dynamical characteristics of time series over multiple scales. Fig. 5 illustrates the coarse-grained procedure. With the merits of multiscale analysis, MSLZ is able to perform the multiscale feature extraction. In summary, two steps are involved in the proposed MSLZ method as follows.

1) For a given time series $X\{x(k), k = 1, 2, \ldots, N\}$, it can be divided into several coarse-grained time series $\{y_j^s\}$ (s is a positive integer) in the form of (5). It can be found that the scale factor specifies the number of data points averaged to obtain each element of the coarse-grained time series. Therefore, the coarse-grained process is actually the averaging process, which can obtain the low-frequency components of the original time series. In other words, multiscale analysis can assess the dynamical characteristics of time series over different frequency bands. For $\tau = 1$, the time series $\{y^{(1)}\}$ is the original time series with single-scale analysis

$$y_j^s = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \le j \le \left\lceil \frac{N}{\tau} \right\rceil. \tag{5}$$

2) Calculate the SLZ value of each coarse-grained time series $\{y_j^\tau\}$ using steps 1–5 referred in the SLZ method, which can be expressed as follows:

$$\mathrm{MSLZ}(x, d, n, \tau) = \mathrm{SLZ}(y_j^\tau, d, n). \tag{6}$$

Fig. 6 illustrates the calculation process of MSLZ. Meanwhile, the MATLAB code of the MSLZ is provided in Appendix A for convenient application and validation.

C. MSLZ Performance Verification Using Synthetic Signal

To validate the advantage of the proposed MSLZ algorithm, a bearing fault model in [44] is applied to simulate the rolling bearing with inner race fault, outer race fault, and rolling element fault. The bearing model considers the effects of bearing geometry, shaft speed, the load, decaying exponential, and so on [44]. The mathematical model aims to generate three different types of periodical impulses as follows:

$$X(t) = \begin{cases} \left[\sum_{k=-\infty}^{+\infty} d_o \delta(t - kT_o)\right] * e(t), & 0.3 < t < 0.5 \\ \left\{\left(\sum_{k=-\infty}^{+\infty} (d_i \delta(t - kT_i)\right) \right. \\ \quad \left. \cdot\, q(2\pi f_r t) \cdot p(2\pi f_r t)\right\} * e(t), & 0.8 \le t \le 1.0 \\ \left(\sum_{k=-\infty}^{+\infty} (d_{bo} \delta(t - kT_b)\right. \\ \quad \left. + \; d_{bi} \delta(t - kT_b - \frac{1}{2}T_b)\right)\right) & 1.3 < t < 1.5 \\ \quad \cdot\, q(2\pi f_r t) \cdot p(2\pi f_r t) * e(t), \end{cases} \tag{7}$$

where d_o, $d_o = i$, d_{bo}, and d_{bi} are the amplitudes of impact produced by the defects on outer race, defects on inner race, defects on rolling elements striking the outer race, and defects on rolling element striking the inner race, respectively. T_o, T_i, and T_b are the reciprocal values of fault characteristics of outer race, inner race, and rolling elements, respectively. $q(\varphi)$ is the radial load distribution of rolling element, f_r is the rotating frequency, $e(t)$ is the damping function, $\delta(t)$ is the impulse function, and k is the number of impulses.

The parameters are given as follows: the amplitudes of impact produced by the defects d_o, $d_o = i$, d_{bo}, and d_{bi} are all set as 4, the rotating frequency f_r is 50 Hz, the reciprocal value of fault frequency of outer race T_o is 0.0049, the reciprocal value of fault frequency of inner race T_i is 0.0034, the reciprocal value of fault frequency of rolling element T_b is 0.0038 Hz, the sampling frequency $f_s = 20\,480$ Hz, and the number of impulses $k = 11$. In addition, the white Gaussian noise SNR $= -2$ dB is added to simulate the noisy environment. The time-domain waveform of the simulated signal is displayed in Fig. 7(b).

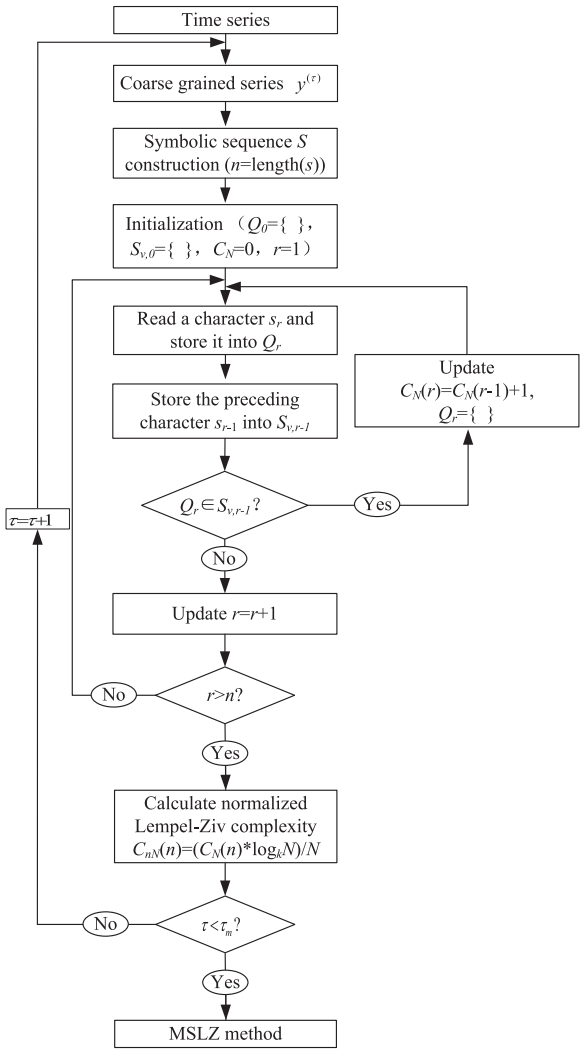

Fig. 6. Framework of the proposed MSLZ method.

A comprehensive comparison analysis is conducted between the LZ, MLZ, SLZ, and MSLZ methods. Note that we combine multiscale analysis with the traditional LZ method (called MLZ) for comparison. The synthetic signals are 128 s with a sampling frequency of 256 Hz. A sliding window of 7 s is applied to cut out the data with a step length of 2 s, which means a sliding window data with 75% overlap moving along the signal. Here, we set the symbolization number $\varepsilon = 3$ for MSLZ and SLZ methods and the scale $\tau = 10$ for MLZ and MSLZ methods.

It is worth noting that we apply four methods to calculate the Euclidean distance (ED) value between the first ten samples (taken as normal condition) and each of other samples to evaluate their performance of fault detection. The obtained results are shown in Fig. 7(a). As seen from Fig. 7(a), we can observe that the LZ, SLZ, MLZ, and MSLZ methods are all able to detect the

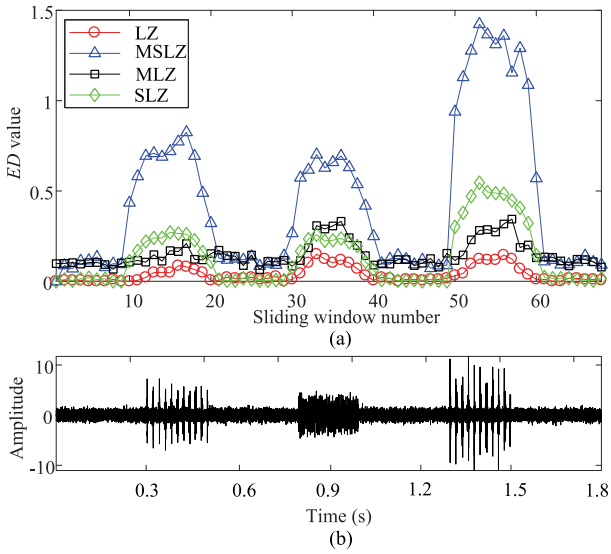

Fig. 7. Performance comparison results of LZ, SLZ, MLZ, and MSLZ methods. (a) ED values of LZ, SLZ, MLZ, and MSLZ. (b) Simulated signal.

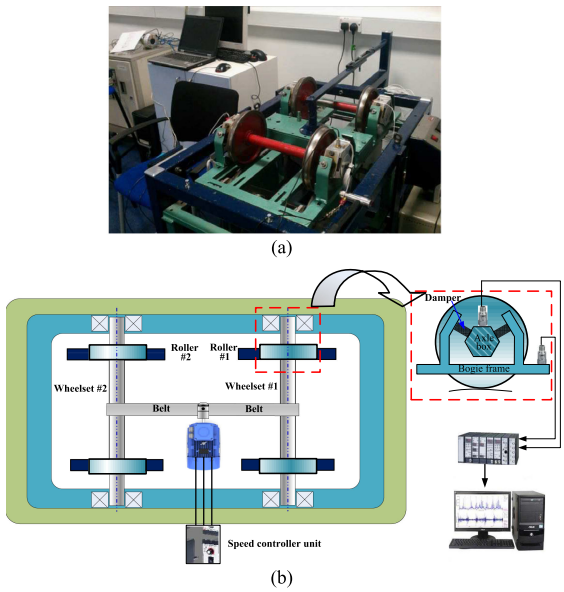

Fig. 8. (a) One-fifth scaled roller test rig [14]. (b) Schematic of the roller rig.

impulses derived from different bearing fault types. However, the proposed MSLZ generates the largest ED value around the three impulses. The fault detection performance of four methods is listed as MSLZ>SLZ>MLZ>LZ. The comparison results indicate that the proposed MSLZ method can greatly enhance the impulse detecting ability. Furthermore, the MSLZ method can even clearly differentiate three bearing fault types.

III. Procedure of MSLZ-Based Fault Diagnosis

A fault diagnosis scheme is designed to conduct CM of the RVS using the proposed MSLZ and ELM classifier. Five steps are summarized as follows.

1) Collect vibration signals of the RVS using a data acquisition (DAQ) system and segregate the whole data into several samples.
2) Divide the initial samples into training samples and testing samples.
3) Apply the MSLZ method to extract the fault characteristics from the samples. It is worthy to note that the parameters of MSLZ are set as follows: $\varepsilon = 3$ and $\tau = 10$ in the following case study. Then, SLZ values over ten scales can be obtained for constructing initial fault features.
4) Utilize the training samples to build the classification model based on the ELM classifier.
5) Feed the testing samples into the well-trained classification model to automatically classify the several fault types of the RVS.

IV. Experiment Validation

In this section, the proposed MSLZ method is validated using an experiment on a one-fifth scaled bogie of railway vehicles [8], [15], [16]. We first describe the test rig, and then, the detailed experimental data description is introduced. Finally, the performance comparison and analysis between LZ, SLZ, MLZ, MSE, and the proposed MSLZ methods is presented.

A. Experimental Setup

A one-fifth scaled bogie at the University of Huddersfield was utilized to collect the data, as shown in Fig. 8(a). The scaled bogie is popular in railway vehicle dynamic analysis, since it has two main advantages: 1) the scaled bogie can realistically simulate the dynamic behavior of a full-size railway vehicle bogie without losing generality under various operation conditions due to its flexible design and 2) the scaled bogie is easy to be controlled with the small occupation of land. The designed roller rig has two rollers, which is driven by an ac motor. The rubber belt is utilized to connect the ac motor and the shaft using two belt pulleys, as shown in Fig. 8(b). In this experiment, eight mount bushings are designed as the primary suspension to connect the bogie frame and the wheelsets. Meanwhile, the rotational speed of this experiment can be controlled by a speed controller unit and the speed is set with 1000 r/min. In our experiments, the wheel and roller surface is unlubricated; thereby, the coefficient of friction is constant. Besides, the data were collected when the speed of the roller arrived at a low stable speed. Therefore, the slip ratio between wheel and rail should be small. On account of this reason, the effect of friction slip is not considered following [5] and [11], in which the fault diagnosis of wheel and roller is conducted using the same test rig in Fig. 8.

The DAQ system includes two acceleration sensors and a DAQ card, as shown in Fig. 8(b). In this experiment, the data from transducer #2 are selected as the main data source due to its direct detection of the fault-induced impulses. The sampling frequency is 24 kHz, and 720 samples were collected. The designed test rig can simulate various common faults occurred on the bogie, such as damping fault, wheel surface fault (WSF), roller surface fault, and compound fault.

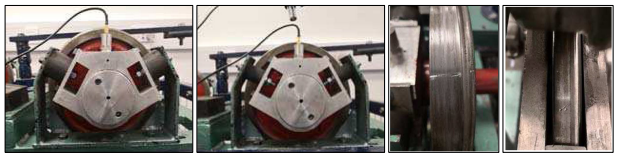

Fig. 9. (a) Suspension with a small damping fault. (b) Suspension with a medium damping fault. (c) Wheel surface fault. (d) Combination of the wheel fault and the roller fault.

To simulate the fatigue of the suspension components, we replace both two normal stud mounts with two softer ones, since the damper stiffness will decrease before the fatigue failure. Besides, two harder stud mounts were also employed to simulate the aging of the primary suspension. In the first case, we aim to recognize the mount bushings with different stiffness and damping ratio severity: harder, normal, and softer. The suspension with harder and softer dampers is displayed in Fig. 9(a) and (b), respectively. The second suspension fault case was fault injection on the wheel surface and roller surface, as shown in Fig. 9(c) and (d), respectively.

It is well known that a softer vehicle suspension can lead to a smoother driving due to the vehicle response is smaller. This phenomenon is same in the scaled roller rig experiment, which will have smaller response with a softer stud mount. In addition, the WSF and the roller fault will result in impulse response. However, the impulse signal signature will become undistinguishable when the suspension (stud mount) becomes soft. Above all, the stiffness reduction of the suspension system will lead to less shock and vibration but will increase the difficulty of diagnosing WSF and roller fault. In this article, the harder stud mount is denoted as small damping fault and the softer case is denoted as medium fault.

B. Experimental Data Description

In this case study, seven faults of a bogie are tested, including suspension with a small damping fault (SSDF), suspension with a medium damping fault (SMDF), the WSF, combination of the wheel fault and the small damping fault (WFDF), combination of the wheel surface fault and the roller fault (WFRF), and combination of the wheel surface fault, the roller fault, and the damping fault (WFRFDF). The above faults are tested on wheelset #1, as marked in red circle in Fig. 8(b). In this article, six suspension, wheel, and roller faults and normal conditions are all considered for validation. Note that the damage on roller aims to simulate the to the rail fault. There are two main advantages of conducting the fault diagnosis of wheel and rail faults. First, many research studies still focus on the fault diagnosis of wheel and rail faults [5], [6], [11], [15]. Conducting the research on wheel and rail failure in laboratory environment could be beneficial in preventing such severe failures occurring. Second, the study that covers six failures will be more comprehensive than the one only including damping fault [16].

For each health condition, the collected data are divided into several nonoverlapping samples. Each sample contains 4096 data points. Each health condition has 100 samples, and there are

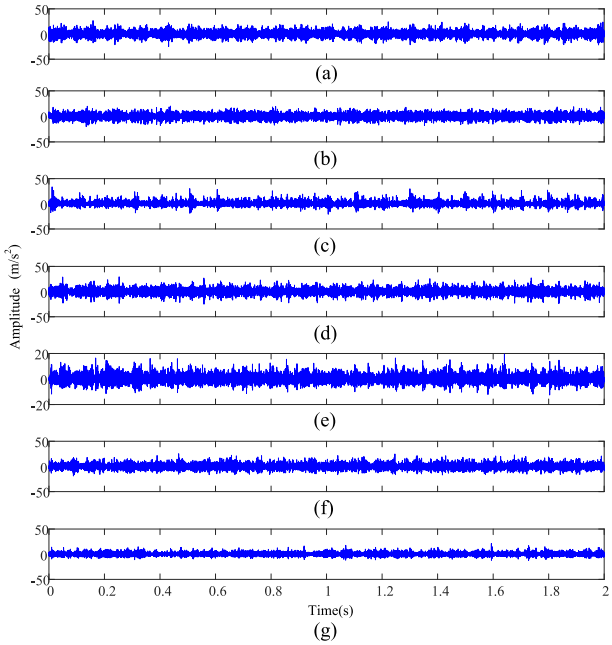

Fig. 10. Vibration acceleration signals of each suspension: (a) normal, (b) SSDF, (c) SFDF, (d) WSF, (e) WFDF, (f) WFRF, and (g) WFRFDF.

total 700 samples. The time-domain waveforms of suspension with seven different health conditions are shown in Fig. 10. As can be seen, the measured vibration signal of the RVS represents the nonlinear characteristics, which is masked by measurement noise and other interferences. For this reason, it is necessary to extract fault features from the vibration signals using the proposed MSLZ method.

For data subset, 350 samples are randomly selected from the whole data to train the ELM classifier, and the remaining samples are taken as testing data. Following the steps in Section III, the MSLZ is employed to calculate the features with ten scales. Then, the obtained MSLZ values are fed into the multiclassifier ELM for pattern recognition.

To demonstrate the superiority of the proposed method, ELM using SLZ (simplified into SLZ-ELM), ELM using LZ (simplified into LZ-ELM), and ELM using popular entropy method MSE (simplified into MSE-ELM) are all employed to conduct the fault diagnosis of the RVS.

C. Performance Comparison and Analysis

1) Comparison of Diagnosis Results: In order to reduce the impact of randomness, we perform 20 runs for each method, and the obtained classification accuracies are shown in Fig. 11 and Table I, respectively. As can be seen, four conclusions can be drawn as follows. First, the proposed method obtains the highest mean classification accuracy (97.5–100%) with 98.90%. This contributes to the superior ability of the proposed MSLZ method in the feature extraction. Second, the MLZ-ELM method has the second highest mean accuracy (90–94.00%) with 91.31%, which is lower than that of the proposed MSLZ method. Moreover, the LZ-ELM method has the least classification accuracy (66–72.00%) with 67.74%, which is lower than SLZ-ELM

TABLE I
DIAGNOSIS RESULTS OF THE FOUR METHODS

MSLZ-ELM			MLZ-ELM			SLZ-ELM			LZ-ELM			MSE-ELM		
Accuracy (%)			Accuracy (%)			Accuracy (%)			Accuracy (%)			Accuracy (%)		
Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean
100	96.00	98.90	94.00	90.00	91.31	79.00	73.00	75.64	72.00	66.00	67.61	92.00	87.00	89.16

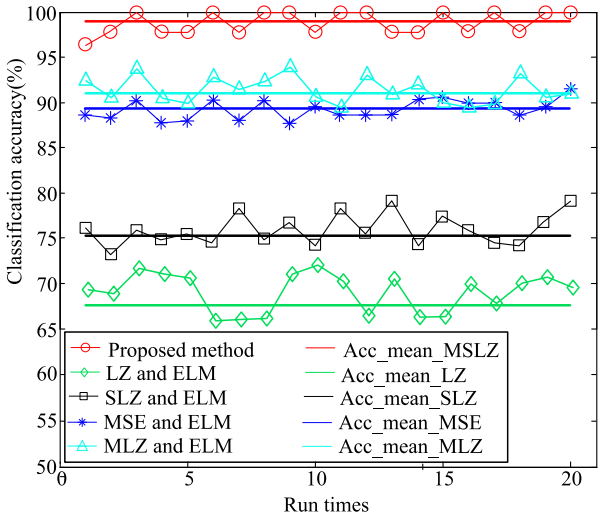

Fig. 11. Classification accuracy of the four methods for 20 runs.

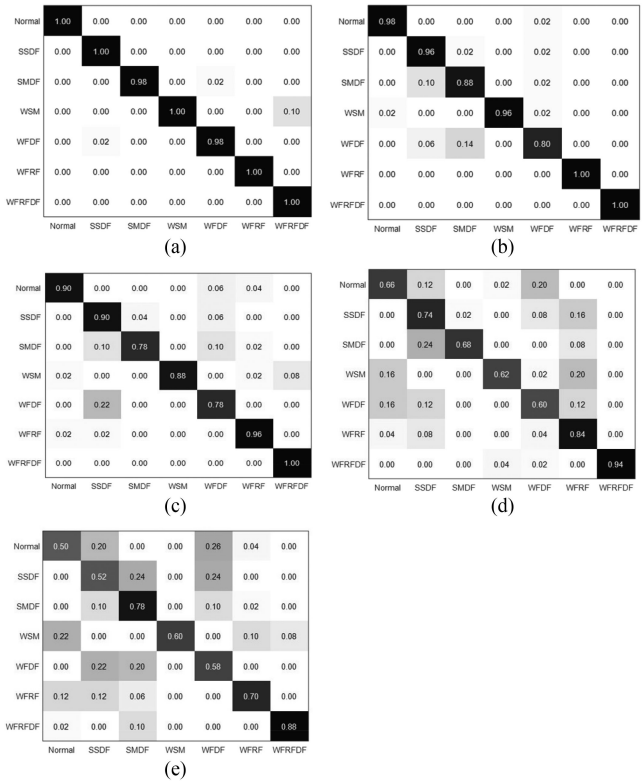

Fig. 12. Confusion matrix of MSLZ-ELM, SLZ-ELM, MLZ-ELM, and MSE-ELM methods.

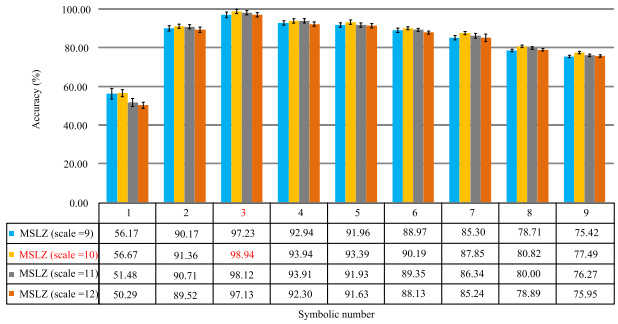

Fig. 13. Diagnosis performance using the proposed MSLZ with different symbols and scales.

(73–79.00%) with 75.64%. This reinforces the superiority of symbolization process in noise canceling. Third, the classification accuracy of SLZ-ELM is lower than the proposed MSLZ-ELM method, which validates the advantage of multiscale analysis in assessing the dynamical characteristics. Last, our proposed method has a higher accuracy compared with the MSE-ELM method (87–92.00%) with 89.16%. This illustrates that the proposed MSLZ method has a better feature extraction ability.

Fig. 12 shows the confusion matrix of the MSLZ-ELM, SLZ-ELM, MLZ-ELM, and MSE-ELM methods. It can be observed that the proposed MSLZ-ELM method has the highest testing accuracy with two testing samples misclassified into wrong samples. This validates that the proposed MSLZ-ELM method is effective in extracting fault characteristics from the RVS.

2) Discussions on the Effects of Symbols and Scales: The performance of the proposed MSLZ method can be influenced by the number of symbols and scales. To estimate the effects of scales and the number of symbols on the MSLZ in classification performance, different symbols ranging from one to nine are considered. The obtained results are displayed in Fig. 13. It can be clearly observed that the proposed MSLZ with symbol $\varepsilon = 3$ outperforms MSLZ with other number of symbols. In general, as more symbols are incorporated, the antinoise ability of MSLZ can be reduced, while less symbols will result in the fact that MSLZ cannot extract enough fault information.

To quantify the performance of scale in feature extraction, the MSLZ with scales ranging from nine to twelve are also studied. Note that when scale $\tau = 1$, the MSLZ degenerates into SLZ. As seen in Fig. 13, it can be found that the MSLZ with scale $\tau = 10$ has the highest classification accuracy. Because a smaller scale number will eliminate the performance in feature extraction, while a larger scale will result in dimension disaster and enhance

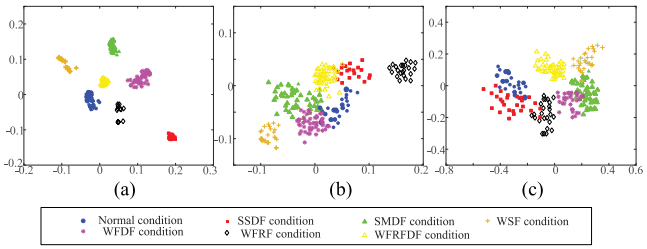

Fig. 14. Two-dimensional visualization using three methods: (a) proposed MSLZ method, (b) MLZ method, and (c) MSE method.

TABLE II
INTRACLASS DISTANCE AND AVERAGE INTERCLASS DISTANCE USING PROPOSED MSLZ, MLZ, AND MSE METHODS

Methods	Intra-class distance							Average inter-class distance
	1	2	3	4	5	6	7	
MSLZ	8.33	7.30	8.24	6.69	5.85	7.31	4.43	166.75
MLZ	35.11	31.09	45.21	32.46	33.73	29.61	29.68	101.93
MSE	43.25	50.05	58.85	45.89	39.10	36.46	32.15	94.96

the CPU time. Therefore, the comparison results demonstrate that MSLZ with $\tau = 10$ and can learn more discriminative and robust features from different scales of raw vibration signals and generate a higher classification accuracy.

3) Empirical Analysis: To illustrate the superiority of the proposed MSLZ in the feature extraction, two-dimensional visualization of MSLZ, MLZ, and MSE via the principal component analysis method under different health conditions is depicted in Fig. 14. As seen in Fig. 14(a), we can observe that when using the MSLZ method, the obtained samples in the same class are much closer than the MLZ and MSE methods. Moreover, the samples between different classes are easier to be separated. Compared with the MSLZ method, the features obtained using MLZ and MSE methods are distributed without a nice cluster. This phenomenon indicates that MSLZ can extract more discriminating and stable features compared with MLZ and MSE methods. In addition, we use the ED to calculate the intraclass distance and average interclass distance in Fig. 14. The obtained results are listed in Table II. It can be observed from Table II that the proposed MSLZ method has a smaller intraclass distance than MLZ and MSE, which means that the feature obtained using MSLZ has a stronger capability of maintaining the local information. Meanwhile, the MSLZ method has the largest interclass distance, which indicates that the feature has a stronger discriminant ability.

For comparison, four recently published methods, convolutional neural network (CNN) [45], deep belief network (DBN) [46], symbolic aggregate approximation and Lempel–Ziv (SAX-LZ) [47], and sparsogram and Lempel–Ziv (SPLZ) [31], are all applied to analyze the same vibration signal. We use the classification accuracy to evaluate the performance of the five methods. For fair comparison, 50% of the original data were randomly selected as the training dataset, and the remaining is taken as the testing dataset. To reduce the effect of randomness, 20 trials were conducted for each method. The averaging training and testing accuracies are calculated, and their corresponding standard deviations are represented using the positive error bars.

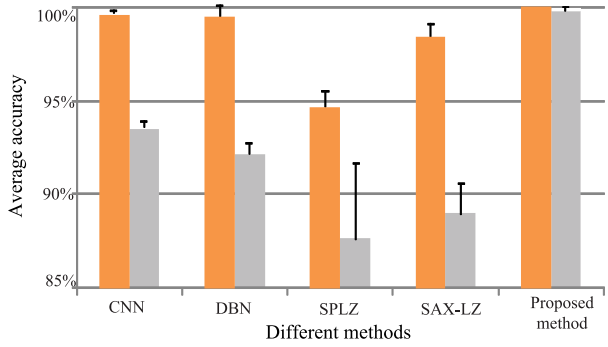

Fig. 15. Diagnosis results of the six trails using five methods.

TABLE III
PERFORMANCE COMPARISON OF FOUR ALGORITHMS BASED ON RVS DATASETS

Feature extraction Methods	Training accuracy (%)	Testing accuracy (%)	Reference
CNN	99.62	93.51	[45]
DBN	99.48	92.15	[46]
SPLZ	94.62	87.64	[31]
SAX-LZ	98.40	88.96	[47]
Proposed method	100.00	98.90	-

The results are shown in Fig. 15. Meanwhile, Table III shows the classification accuracy of each method using the RVS datasets. It can be observed from Fig. 15 and Table III that the proposed MSLZ method has the highest testing accuracy with 98.90%.The accuracies of deep learning algorithms exceed 90%. The CNN and the DBN have second and third highest testing accuracy, respectively. The SPLZ method has the lowest testing accuracy with 87.64%. It can be concluded that the proposed MSLZ method performs best in extracting the hidden fault information compared with other four methods.

V. CONCLUSION

In this article, a new fault identification method based on MSLZ was developed. The major contribution of MSLZ was the incorporation of the symbolization process and multiscale analysis into the traditional LZ. The proposed MSLZ method can enhance the denoising ability through the symbolization process and further estimate the complexity of time series over a series of scales. Therefore, the MSLZ method can extract rich fault information from vibration signals and enhance the fault identification performance. Synthetic and experimental results demonstrated that MSLZ-based method achieved much better performance than SLZ, LZ MLZ, and MSE in diagnosing various faults of the RVS. Moreover, the proposed method brings new vitality for the traditional LZ method and can be easily extended to deal with different machines and industrial systems.

The proposed MSLZ method was tested and demonstrated to be effective for the RVS. However, the effectiveness of the proposed method for other rotating machinery was unknown. Further tests using other machines will be considered in our future work. Meanwhile, the effect of the friction slip between two surfaces was not considered. Further tests considering the slip will also be conducted in our future work.

ACKNOWLEDGMENT

The authors would like to thank Dr. F. Gu and the Centre for Efficiency and Performance Engineering, University of Huddersfield, for providing the experimental data.

REFERENCES

[1] K. Liang, N. Qin, D. Huang, and Y. Fu, "Convolutional recurrent neural network for fault diagnosis of high-speed train bogie," *Complexity*, vol. 2018, 2018, Art. no. 4501952.
[2] G. Xiantai, M. Shiheng, J. Weidong, and L. Xiao, "Fault type recognition of high-speed train bogie based on dual-channel integration of information entropy," in *Proc. 2nd IEEE Int. Conf. Comput. Commun.*, 2016, pp. 1880–1884.
[3] C. Tang, G. Zhou, Z. Gao, X. Shu, and P. Chen, "A novel rail inspection robot and fault detection method for the coal mine hoisting system," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 2, pp. 110–121, Summer 2019.
[4] L. Cui, J. Wang, and S. Lee, "Matching pursuit of an adaptive impulse dictionary for bearing fault diagnosis," *J. Sound Vib.*, vol. 333, no. 10, pp. 2840–2862, 2014.
[5] L. Su, L. Ma, N. Qin, D. Huang, and A. H. Kemp, "Fault diagnosis of high-speed train bogie by residual-squeeze net," *IEEE Trans. Ind. Inform.*, vol. 15, no. 7, pp. 3856–3863, Jul. 2019.
[6] A. Jaschinski, H. Chollet, S. Iwnicki, A. Wickens, and J. Würzen, "The application of roller rigs to railway vehicle dynamics," *Veh. Syst. Dyn.*, vol. 31, no. 5-6, pp. 345–392, 1999.
[7] L. Cui, J. Huang, F. Zhang, and F. Chu, "HVSRMS localization formula and localization law: Localization diagnosis of a ball bearing outer ring fault," *Mech. Syst. Signal Process.*, vol. 120, pp. 608–629, 2019.
[8] B. Liang, S. Iwnicki, Y. Zhao, and D. Crosbee, "Railway wheel-flat and rail surface defect modelling and analysis by time–frequency techniques," *Veh. Syst. Dyn.*, vol. 51, no. 9, pp. 1403–1421, 2013.
[9] Y. Li, M. J. Zuo, J. Lin, and J. Liu, "Fault detection method for railway wheel flat using an adaptive multiscale morphological filter," *Mech. Syst. Signal Process.*, vol. 84, pp. 642–658, 2017.
[10] X. Wei, L. Jia, and H. Liu, "A comparative study on fault detection methods of rail vehicle suspension systems based on acceleration measurements," *Veh. Syst. Dyn.*, vol. 51, no. 5, pp. 700–720, 2013.
[11] X. Wei, L. Jia, K. Guo, and S. Wu, "On fault isolation for rail vehicle suspension systems," *Veh. Syst. Dyn.*, vol. 52, no. 6, pp. 847–873, 2014.
[12] X. Liu, S. Alfi, and S. Bruni, "An efficient recursive least square-based condition monitoring approach for a rail vehicle suspension system," *Veh. Syst. Dyn.*, vol. 54, no. 6, pp. 814–830, 2016.
[13] D. Lebel, C. Soize, C. Funfschilling, and G. Perrin, "High-speed train suspension health monitoring using computational dynamics and acceleration measurements," *Veh. Syst. Dyn.*, pp. 1–22, 2019.
[14] Y. Zhao, Z. H. Guo, and J. M. Yan, "Vibration signal analysis and fault diagnosis of bogies of the high-speed train based on deep neural networks," *J. Vibroeng.*, vol. 19, no. 4, pp. 2456–2474, 2017.
[15] B. Liang, S. Iwnicki, A. Ball, and A. E. Young, "Adaptive noise cancelling and time–frequency techniques for rail surface defect detection," *Mech. Syst. Signal Process.*, vol. 54, pp. 41–51, 2015.
[16] F. Liu, H. Zhang, X. He, Y. Zhao, F. Gu, and A. D. Ball, "Correlation signal subset-based stochastic subspace identification for an online identification of railway vehicle suspension systems," *Veh. Syst. Dyn.*, pp. 1–21, 2019.
[17] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.—Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, 2000.
[18] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, 2002, Art. no. 174102.
[19] R. Yan, Y. Liu, and R. X. Gao, "Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines," *Mech. Syst. Signal Process.*, vol. 29, pp. 474–484, 2012.
[20] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 2, pp. 266–272, 2007.
[21] Y. Li, Y. Yang, G. Li, M. Xu, and W. Huang, "A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection," *Mech. Syst. Signal Process.*, vol. 91, pp. 295–312, 2017.
[22] Y. Li, X. Wang, Z. Liu, X. Liang, and S. Si, "The entropy algorithm and its variants in the fault diagnosis of rotating machinery: A review," *IEEE Access*, vol. 6, pp. 66723–66741, 2018.
[23] M. Zanin, L. Zunino, O. A. Rosso, and D. Papo, "Permutation entropy and its main biomedical and econophysics applications: A review," *Entropy*, vol. 14, no. 8, pp. 1553–1577, 2012.
[24] M. Rostaghi and H. Azami, "Dispersion entropy: A measure for time-series analysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 610–614, May 2016.
[25] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 75–81, Jan. 1976.
[26] H. Hong and M. Liang, "Fault severity assessment for rolling element bearings using the Lempel–Ziv complexity and continuous wavelet transform," *J. Sound Vib.*, vol. 320, nos. 1/2, pp. 452–468, 2009.
[27] R. Yan and R. X. Gao, "Complexity as a measure for machine health evaluation," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 4, pp. 1327–1334, Aug. 2004.
[28] D.-Y. Dou and Y.-K. Zhao, "Fault severity assessment for rolling element bearings based on EMD and Lempel-Ziv index," *J. Vib. Shock*, vol. 29, no. 3, pp. 5–8, 2010.
[29] P. Xia, X. U. Hua, M. A. Zaichao, M. Lei, S. Pei, and Z. Qiao, "Assessment method of early damage severity for rolling bearing based on the improved HVD and Lempel-Ziv index," *J. Xian Jiaotong Univ.*, vol. 51, no. 6, pp. 8–13, 2017.
[30] L. Cui, X. Gong, J. Zhang, and H. Wang, "Double-dictionary matching pursuit for fault extent evaluation of rolling bearing based on the Lempel-Ziv complexity," *J. Sound Vib.*, vol. 385, pp. 372–388, 2016.
[31] L. Cui, B. Li, J. Ma, and Z. Jin, "Quantitative trend fault diagnosis of a rolling bearing based on sparsogram and Lempel-Ziv," *Measurement*, vol. 128, pp. 410–418, 2018.
[32] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E, Statist. Nonlinear Soft Matter Phys.*, vol. 71, no. 2, 2005, Art. no. 021906.
[33] V. Rajagopalan, S. Chakraborty, and A. Ray, "Estimation of slowly varying parameters in nonlinear systems via symbolic dynamic filtering," *Signal Process.*, vol. 88, no. 2, pp. 339–348, 2008.
[34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.
[35] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.
[36] Y. Li, X. Liang, M. Xu, and W. Huang, "Early fault feature extraction of rolling bearing based on ICD and tunable Q-factor wavelet transform," *Mech. Syst. Signal Process.*, vol. 86, pp. 204–223, 2017.
[37] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, 2002, Art. no. 068102.
[38] S.-D. Wu, P.-H. Wu, C.-W. Wu, J.-J. Ding, and C.-C. Wang, "Bearing fault diagnosis based on multiscale permutation entropy and support vector machine," *Entropy*, vol. 14, no. 8, pp. 1343–1356, 2012.
[39] V. Rajagopalan and A. Ray, "Symbolic time series analysis via wavelet-based partitioning," *Signal Process.*, vol. 86, no. 11, pp. 3309–3320, 2006.
[40] G.-C. Wu, D. Baleanu, H.-P. Xie, and F.-L. Chen, "Chaos synchronization of fractional chaotic maps based on the stability condition," *Phys. A: Statist. Mech. Appl.*, vol. 460, pp. 374–383, 2016.
[41] G.-C. Wu and D. Baleanu, "Chaos synchronization of the discrete fractional logistic map," *Signal Process.*, vol. 102, pp. 96–99, 2014.
[42] G.-C. Wu and D. Baleanu, "Discrete fractional logistic map and its chaos," *Nonlinear Dyn.*, vol. 75, nos. 1/2, pp. 283–287, 2014.
[43] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3196–3207, Apr. 2019.
[44] P. McFadden and J. Smith, "Model for the vibration produced by a single point defect in a rolling element bearing," *J. Sound Vib.*, vol. 96, no. 1, pp. 69–82, 1984.
[45] Y. Ma *et al.*, "A new fault diagnosis method based on convolutional neural network and compressive sensing," *J. Mech. Sci. Technol.*, vol. 33, no. 11, pp. 5177–5188, 2019.
[46] Y. Zhang *et al.*, "A cable fault recognition method based on a deep belief network," *Comput. Elect. Eng.*, vol. 71, pp. 452–464, 2018.
[47] J. Yin, M. Xu, and H. Zheng, "Fault diagnosis of bearing based on symbolic aggregate approximation and Lempel-Ziv," *Measurement*, vol. 138, pp. 206–216, 2019.