A Novel Fault Detection Scheme Using Improved Inherent Multiscale Fuzzy Entropy With Partly Ensemble Local Characteristic-Scale Decomposition

SONGRONG LUO [1,2], WENXIAN YANG [3], AND YOUXIN LUO[1,2]

[1]Hunan Provincial Cooperative Innovation Center for the Construction and Development of Dongting Lake Ecological Economic Zone, Changde 415000, China
[2]College of Mechanical Engineering, Hunan University of Arts and Science, Changde 415000, China
[3]School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.

Corresponding author: Songrong Luo (luosongrong@126.com)

ABSTRACT At present, the multiscale fuzzy entropy has been verified to be an excellent measure of the complexity for dynamic time series. However, when using to short-time time series collected in practical application, the conventional multiscale fuzzy entropy may result in undefined or unreliable value. In this work, improved multiscale fuzzy entropy, named moving-average based multiscale fuzzy entropy (MA_MFE), is presented at first to potentially characterize the complexity of short-term time series. The MA_MFE algorithm can successfully produce more template vectors to overcome the problem of shortening the samples in the procedure of the existing approaches. The analysis experiments for both white noise signal and $1/f$ noise signal are made and the results show MA_MFE method is more effective for the short-term datasets. Then, a novel fault detection scheme has been developed. After using non-local mean approach to reduce background noise, the non-stationary vibration signals are decomposed into several intrinsic scale components (ISCs) by a newly developed time-frequency signal analysis method– partly ensemble local characteristic-scale decomposition (PELCD); The ISCs with higher correlation coefficients are used to reconstruct into a new signal and the inherent MA_MFEs are extracted to quantify the complexity of the collected vibration signal. At last, the multiSVM and improved variable predictive model based class discrimination (VPMCD) are employed as small-sample classifiers to achieve fault detection. Two experiments have been conducted, which include both rolling bearing as vital component in rotating machinery and a piston pump as typical reciprocation machinery in hydraulic system. The comparison results show that the proposed fault detection scheme is more effective and reliable and suitable for real-time online fault detection.

INDEX TERMS Multiscale analysis, moving-average, partly ensemble local characteristic-scale decomposition, fault detection.

I. INTRODUCTION

Condition monitoring and fault detection are crucial to ensure the high reliability, safety and maintainability of the whole mechanical system. Local faults in mechanical system will cause extra vibration and noise, and the vibration signals collected by sensors would change with different fault types and fault severity. Moreover, the structure is extremely complex and many factors including working load, friction, and damping capacity vary instantaneously in the running condition, and strong coupling also exists among the fault features. Simultaneously, the samples for fault states are rare and they

The associate editor coordinating the review of this manuscript and approving it for publication was Tianhua Xu.

always are short-term time series. As a result, it is still a huge challenge to detect the running states and fault types effectively and quickly [1]–[5].

The validity of a data-driven condition monitoring and fault detection system greatly depends on the fault features extracted and the performance of the classifier. This procedure mainly involves two key steps: feature extraction and class recognition [2], [4]–[6]. At present, various methods have been developed and applied to fault detection [2], [7]–[10]. In the stage of feature extraction, entropy-based measures have been widely used to characterize the complexity of time series from dynamical system. These commonly entropy-based complexity indexes mainly include sample entropy (SampEn) [11], fuzzy entropy (FuzzyEn) [12], [13], permutation entropy (PE) [14], [15]. However, the results of most traditional single-scale entropy methods are not always consistent with the complexity of real-world signals [16]. For example, the SampEn of white noise is higher than that of $1/f$ noise, but the latter has more complex structure [16]. Aiming at this issue, Costa *et al* proposed multiscale entropy (MSE) to represent the complexity of time series [16], [17]. In order to quantify the complexity at different time scales, the original time series are coarse-grained with time scales and then the multiscale fuzzy entropy for each coarse-grained time series is computed. In the multiscale technique, for the scale factor τ, coarse-graining procedure is achieved by calculating the arithmetic mean of τ neighboring values without overlapping. The original data with length N is divided into N$/\tau$ segments and obtain the coarse-grained time series with length N$/\tau$, which is used to compute an entropy value with regard to scale factor τ. Integrated with the MSE conception, the multiscale fuzzy entropy (MFE) [18], [19] and the refined composite multiscale fuzzy entropy (RCMFE) were proposed and used successfully to analysis complex time series in different research fields [20], [21]. However, from the perspective of signal processing, the coarse-graining procedure averages the data within a window and the data is down-sampled by multiple time scales. The length of coarse-grained time series is shorten and may be not long enough to calculate an accurate value and obtain either undefined or unreliable values [22]. Hence, as far as the short-term vibration signals from dynamic mechanical system are concerned, it is more difficult to obtain accurate multiscale entropies using large scale factor. Targeting at this problem, Wu et al introduced a moving-average approach to complete the coarse-graining procedure so that large number of template vectors can be obtained to better MSE performance [22]. Inspired by this idea, Li et al introduced moving-average approach to improve multiscale fuzzy entropy for fault feature extraction [23].

At the same time, vibration signals from mechanical system are always non-stationary, including some superimposed trend oscillations which could influence the standard deviation of the estimated values. Hence, it is necessary to deal with the original signals using data-driven signal decomposition method before feature extraction. Empirical mode

decomposition (EMD) method is widely-used time-frequency analysis technique for non-stationary time series [24]. But, EMD has some shortcomings such as mode-mixing phenomenon, end-point effect [25] et.al. The complementary ensemble empirical mode decomposition (CEEMD) is an improved EMD method and can overcome mode-mixing problem to a certain degree [26]. However, the CEEMD needs large amount of calculation. In [23], the local mean decomposition (LMD) method was also served as an improved EMD method to extract multiscale fuzzy entropy for rolling bearing. However, the LMD method has also the problem of mode mixing, distorted components and large time-consuming [27]. Local characteristic-scale decomposition (LCD) was proposed by Cheng et. al in [27]. Compared with EMD-based method, LCD shows much higher computation efficiency. However, mode-mixing and end-point effect [28]–[30] still exist. In the latest years, as an improved LCD method, partly ensemble local characteristic-scale decomposition (PELCD) has been developed and applied successfully for fault diagnosis in [31], [32]. Accordingly, in this paper, PELCD is employed as data-driven signal decomposition method to eliminate the superimposed trend oscillations. Firstly the original signal is decomposed into several intrinsic scale components (ISCs) by PELCD. Then, inherent ISCs are selected out using spearman correlation coefficients to reconstruct a new signal. Lastly, the MA_MFEs of the reconstructed signal, defined as inherent MA_MFEs, are extracted to evaluate the complexity of the vibration signal. Moreover, novel fault detection models are built. It is noted that though deep learning network and its improved techniques are proved to be promising intelligent identification approaches and have been widely applied in fault diagnosis field [4], [5], a large number of samples are generally required. In this paper, multiSVM and improved VPMCD [33] are used as small-sample multiclass recognition techniques to verify the proposed inherent MA_MFE method. At the end of this paper, the proposed scheme is applied not only to fault detection experiment for rolling bearing but also to multi-fault detection experiment for hydraulic piston pump.

The remaining part is organized as follows. In Section II, the algorithm of MA_MFE is developed and verified with two simulation signals. A novel fault detection scheme is presented In Section III, and applied in Section IV. The conclusions are given in Section V.

II. BACKGROUNDS
A. MULTISCALE FUZZY ENTROPY
Sample entropy and fuzzy entropy are two most prevalent approaches and have been commonly utilized to quantify the dynamical complexity of time series. As an improved sample entropy, fuzzy entropy is more reliable and less dependent on the parameters and data length [12], [13], [29]. The main conception is described as follows.

A time series is written as $\mathbf{y} = [y_1, y_2, \cdots, y_i, \cdots, y_N]$ with the length N. Given embedding dimension m, a template

vector is reconstructed as

$$y_i^m = \{y_i, y_{i+1}, \cdots, y_{i+m-1}\} - \bar{y}_i^m \quad (1)$$

where $\bar{y}_i^m = \frac{1}{m} \sum_{j=0}^{m-1} y_{i+j}, i = 1, 2, \cdots, N-m+1$

Then, define the Euclidean distance between the vectors y_i^m and y_j^m as

$$d_{ij}^m = d[y_i^m, y_j^m] = max\left\{\left|y_i^m - y_j^m\right|\right\} \quad (2)$$

where $0 \le k \le m-1$ and $i \ne j$

In the conventional sample entropy algorithm, a match happens while $d_{ij}^m \le r$; Accordingly, the similarity of the vectors y_i^m and y_j^m is defined by the Heaviside Function. But, in the fuzzy entropy, a family of exponential function is employed as the fuzzy function to calculate the similarity between the two vectors. That is,

$$D_{ij}^m = u(d_{ij}^m, n, r) = \exp(-(d_{ij}^m)^n / r) \quad (3)$$

For the vector y_i^m, all the similarities of its neighboring vectors can be averaged as the following formula,

$$\emptyset_i^m(n, r) = \frac{1}{N-m-1} \sum_{j=1, j \ne i}^{N-m} D_{ij}^m \quad (4)$$

And define a function as

$$\varphi^m(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \emptyset_i^m(n, r) \quad (5)$$

Repeat the above procedure, then we get $\varphi^{m+1}(n, r)$ as

$$\varphi^{m+1}(n, r) = \frac{1}{N-m-1} \sum_{i=1}^{N-m-1} \emptyset_i^{m+1}(n, r) \quad (6)$$

The fuzzy entropy value of the original time series with the data length N can be estimated statistically as

$$FuzzyEn(\mathbf{X}, m, n, r) = -\ln \frac{\varphi^{m+1}(n, r)}{\varphi^m(n, r)} \quad (7)$$

In order to quantify the complexity in different time scales, the multiscale entropy (MSE) algorithm have been developed in the literature [16]. A coarse-graining procedure is used to extract the subsequence of the original time series in different time scales. To obtain the coarse-grained time series, the original data $X = [x_1, x_2, \cdots, x_i, \cdots, x_N]$ is separated into segments of length τ and the data points are averaged in each segment. Each element z_j^τ of the coarse-grained time series z^τ is defined according to the formula as below.

$$z_j^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \le j \le N/\tau \quad (8)$$

The multiscale analysis was introduced to the fuzzy entropy to produce the conception of the multiscale fuzzy entropy (MFE) [19]. However, from the formula (8), it can be seen that the coarse-grained time series z^τ is greatly shorten with the scale factor τ increasing. Therefore, for short-term time series, both the MSE and MFE could cause inaccurate and even undefined entropy values.

It is worthwhile noting that a refined composite multiscale fuzzy entropy (RCMFE) has been developed as an improved

multiscale fuzzy entropy algorithm recently. The RCMFE utilizes different start points to produce different coarse-grained time series in the same scale and obtain the final results by averaging the entropy values [20]. Please refer to the literatures [20], [21] for the details. The relevant research have shown that RCMFE is more reliable than MFE technique and has widely used in varying fields [12], [21], [34]. However, the shorten-data problem still exists unavoidably in the procedure. As a result, to address the issue of the MFE and RCMFE, a moving-average based multiscale fuzzy entropy (MA_MFE) algorithm is proposed as below.

B. MOVING-AVERAGE BASED MULTISCALE FUZZY ENTROPY

In this paragraph, moving-average based multiscale fuzzy entropy (MA_MFE) algorithm is proposed, including the following steps.

Step 1: Conduct coarse-graining procedure using moving-average algorithm at a scale factor τ. A coarse-grained time series at a scale factor τ can be written as $\mathbf{z}^\tau = \{z_j^\tau(j = 1, 2, \cdots, N-\tau+1\}$, which can be constructed according to the following Equation (9). And the schematic illustration of the coarse-graining procedure based on moving-average is given in Fig. 1.

$$z_j^\tau = \frac{1}{\tau} \sum_{i=j}^{j+\tau-1} y_i, \quad k \le j \le N-\tau+1 \quad (9)$$

Step 2: For a defined scale factor τ, embedding dimension m and tolerant r, $\varphi_\tau^m(n, r)$ and $\varphi_\tau^{m+1}(n, r)|$ are calculated for each of the moving-average time series $\mathbf{Z}^\tau$ according to Equation (5)-(6).
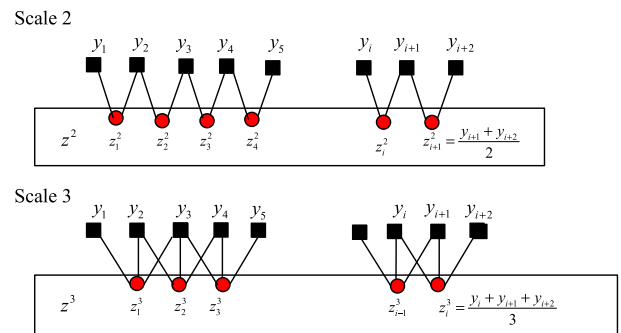

FIGURE 1. Schematic illustration of the coarse-graining procedure based on moving- average algorithm.

Step 3: Thus the MA_MFE values can be obtained as

$$MA\_MFE(\mathbf{X}, \tau, m, n, r) = -\ln \frac{\bar{\varphi}_\tau^{m+1}(n, r)}{\bar{\varphi}_\tau^m(n, r)} \quad (10)$$

From the description above, both the MFE and the RCMFE employ non-lapping windows to construct the template vectors, while the MA_MFE method uses lapping windows to acquire new coarse-graining time series. Thus, the template vectors used in the MA_MFE algorithm are much more than ones in the MFE and the RCMFE so that the yielding
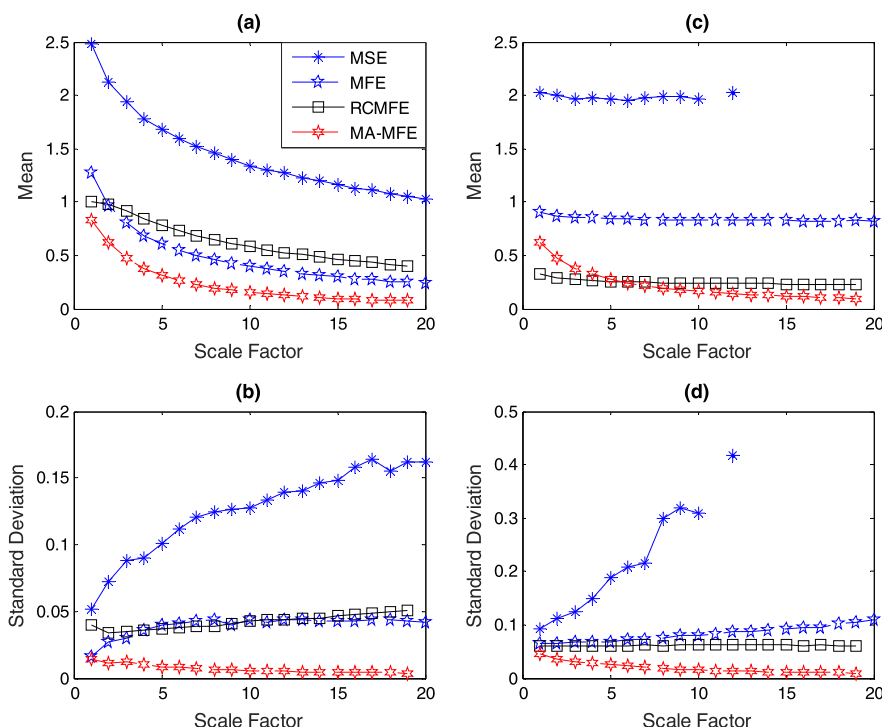
FIGURE 2. Entropy- based analysis of simulated signals under different scale factors. (a) Mean of analysis results for 200 white noise. (b) Standard deviation of analysis results for 200 white noise. (c) Mean of analysis results for 200 1/*f* noise. (d) Standard deviation of analysis results for 200 1/*f* noise.

probability of undefined entropy would greatly decreased even in the case of big scale factors.

In this paragraph, the MA_MFE will be compared with the MSE, the MFE and the RCMFE to verify its advantages. According to the definition of various multiscale entropy, their calculation are commonly related to the length N of a time series, embedding dimension m, similarity tolerance r and gradient n. Generally speaking, the bigger m is, the richer the information obtained from the time series would be. But, while adopting a bigger m, a bigger data length N would be required. Generally speaking, $N = 10^m - 30^m$. In order to be fair for comparison in this paper, the embedding dimension is set $m = 2$. Similarity tolerance r represents the boundary width of comparison window and mainly controls the similarity of template matching. Hence similarity tolerance r directly involves in the accessibility of templates matching and the accuracy of the statistical information in the calculating process. Too large r would make matching more difficult, but too small r could make the estimated value sensitive to noise and become inaccurate. In order to obtain better effect, it is usually set $r = 0.15SD$ (SD is standard deviation of the analyzed time series). Regarding to gradient n, detailed information will be lost more as n rises up; therefore, a small $n=2$ is set. In the end, considering a fact that there is a limited data length N in the engineering practice, the maximum scale is set by $\tau_{\max} = 20$. In summary, the parameters are set $m = 2, r = 0.15SD, n = 2, \tau_{max} = 20$.

The white noise and 1/*f* noise with different length N ranged from N = 500 to N = 3500 at the interval of 500 as two synthetic noise signals are represented to evaluate the advantages of the proposed MA_MFE method. In this simulation analysis, 200 independent noise samples are taken to compare four multiscale entropy-based feature extraction techniques–MSE, MFE, RCMFE and MA_MFE.

The entropy curves of white noise and 1/*f* noise are shown in Fig. 2. The MSE curve of white noise decreases with the scale factors increasing whereas those of 1/*f* noise remains constant [17], which can be seen in Fig. 2. As mentioned above, fuzzy entropy algorithm is the improvement of sample entropy, so it can be found that the same trend as MSE value is presented for the results of MFE, RCMFE and MA_MFE. In the meantime, the means and standard deviations of the two synthetic noise signals are given in Table 1 and 2, respectively. Moreover, whatever kind of noise signals is used, it can be observed that the MA_MFE has the smallest standard deviations (SDs), which indicates that its results are the most reliable. In addition, for 1/*f* noise, the MSE method causes undefined entropy values when the data length is smaller than 1500, whereas the MFE and the RCMFE can still achieve reliable entropy values, especially using the MA_MFE method. In the end column of Table 1, the SDs of mean values using the MA_MFE method are smaller in both cases, which shows the MA_MFE method is robust to the data

TABLE 1. Mean of analysis results at $\tau = 20$ obtained by different algorithms with various data length *N*.

Method		Data length N							Std
		500	1000	1500	2000	2500	3000	3500	(with N)
White noise	MSE	1.0887	1.0330	1.0306	1.0204	1.0113	1.0139	1.0158	0.0269
	MFE	0.2344	0.2406	0.2406	0.2394	0.2394	0.2381	0.2366	0.0023
	RCMFE	0.3990	0.4047	0.4048	0.4040	0.4011	0.4051	0.3992	0.0027
	MA_MFE	0.0796	0.0753	0.0742	0.0732	0.0729	0.0723	0.0723	0.0026
1/*f* noise	MSE	☆	☆	☆	1.9224	1.9153	1.9037	1.8964	☆
	MFE	0.8756	0.8211	0.8016	0.8020	0.7862	0.7896	0.7834	0.0322
	RCMFE	0.2825	0.2304	0.2195	0.2098	0.2045	0.2006	0.1951	0.0299
	MA_MFE	0.1097	0.0985	0.0947	0.0930	0.0912	0.0907	0.0892	0.0071

Note: ☆ refers to undefined value .

TABLE 2. Table 2 SDs of analysis results at $\tau = 20$ obtained by different algorithms with various data length *N*.

Method		Data length						
		500	1000	1500	2000	2500	3000	3500
White noise	MSE	0.3319	0.1621	0.1151	0.0880	0.0786	0.0764	0.0679
	MFE	0.0589	0.0422	0.0319	0.0265	0.0246	0.0234	0.0229
	RCMFE	0.0759	0.0506	0.0376	0.0324	0.0276	0.0267	0.0238
	MA_MFE	0.0051	0.0038	0.0033	0.0025	0.0023	0.0022	0.0021
1/*f* noise	MSE	☆	☆	☆	0.2752	0.2431	0.1936	0.1710
	MFE	0.1604	0.1091	0.0871	0.0689	0.0638	0.0588	0.0551
	RCMFE	0.0906	0.0612	0.0565	0.0455	0.0468	0.0428	0.0421
	MA_MFE	0.0118	0.0098	0.0082	0.0075	0.0070	0.0067	0.0068

Note: ☆ refers to undefined value.

length N. Therefore, the MA_MFE method is suitable for the short-term complex signals like 1/*f* noise signal.

C. PARTLY ENSEMBLE LOCAL CHARATERISTIC-SCALE DECOMPOSITION

In order to improve LCD method, following the main idea of CEEMD, the literature**s** [31], [32] proposed a new noise assisted data analysis method, termed as partly ensemble local characteristic-scale decomposition, whose procedure is summarized as follows.

Step 1: Add a pair of noises with opposite signs, whose amplitudes are α to a raw signal S(t). Generally, $\alpha(t) = 0.05\text{k}SD$, $k = 1, 2, 3, \cdots, 8$, SD is the standard deviate of the raw signal $s(t)$

$$S_i^+(t) = S(t) + \alpha \times n_i(t) \quad (11a)$$
$$S_i^-(t) = S(t) - \alpha \times n_i(t) \quad (11b)$$

$i = 1, 2, \cdots, Ne$ indicates the number of noise pairs, $n_i(t)$ is the noise to add.

Step 2: Starting with $j = 1$, perform LCD decomposition on $S_i^+(t)$ and $S_i^-(t)$ respectively to obtain a series of $\left\{I_{ij}^+(t)\right\}$ and $\left\{I_{ij}^-(t)\right\}$, and $I_j(t)$ can be obtained by integrating the

average of the experiments

$$I_j(t) = \frac{1}{2Ne} \sum_{i=1}^{Ne} \left[I_{ij}^+(t) + I_{ij}^-(t) \right] \quad (12)$$

Step 3: Calculate the permutation entropy (PE) value E_{pj} of $I_j(t)$ according to the literature [14], and check if $E_{pj} > \theta_0$ (θ_0 is the threshold value of PE, generally , θ_0 is set to [0.4, 0.6]). If the answer is 'yes', then $I_j(t)$ is high frequency intermittent signal or noise, then, let $j = j + 2$ and return and repeat Step 2 until $E_{pj} < \theta_0$.

Step 4: Separate the first $(p - 1)$ components from the raw signal $S(t)$ as

$$R(t) = S(t) - \sum_{j=1}^{p-1} I_j(t) \quad (13)$$

Step 5: Perform LCD decomposition on $R(t)$, then acquire N intrinsic scale components (ISCs) and a residual component $r(t)$, written as

$$S(t) = \sum_{i=1}^{N} ISC_i + r(t) \quad (14)$$

PELCD adds white noise in pair to reduce noise residue and ensure decomposition completeness. At the same time, it avoids unnecessary ensemble averaging by using PE to

detect high frequency intermittent signal or noise in time. It is noted that the key point of PELCD is to detect the high frequency intermittent signal or noise, which greatly influence the decomposition result. PE is used to check dynamic changes of the time series. The relevant theory regarding to PE can be referred in [32]. Once those high frequency intermittent signal or noise is eliminated, the normal LCD is applied to decomposition the rest signal. which will not only constrain mode-mixing to increase the analysis accuracy but also improve the computing speed.

D. PROPOSED FEATURE EXTRACTION METHOD

Although PELCD is excellent signal processing technique, some irrelevant ISCs are inevitably caused by background noise or decomposition itself. Therefore, it is essential to select the fault-relevant ISCs, which contain rich fault information. A selection method for relevant intrinsic mode functions (IMFs) using spearman correlation coefficient was proposed in the literature [35]. Recently, ZHENG *et al* proposed a quantified IMFs selection method based on complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and achieved a promising results [36]. However, CEEMDAN needs a large amount of computation cost. Inspired by those literatures, a self-adaptive ISCs selection technique is proposed in this paper. The details are described as below.

Step 1: The false ISCs are eliminated. The product of the energy density of mode component and its average period is constant for a white noise signal [25], [37]. Consequently, the product can be used to identify whether a ISC is a noise signal. The product of the energy density of the *jth* ISC and its average period should be computed as

$$P_j = E_j \times T_j \tag{15}$$

where $E_j = \frac{1}{N} \sum_{i=1}^{N} [A_j(i)]^2$, $T_j = \frac{2N}{O_j}$, O_j is the number of extra points including in the ISC and N is the length of the ISC. Then, a coefficient RP_j can be obtained, shown as,

$$RP_j = \left| \frac{P_j - \frac{1}{j} \sum_{i=1}^{N} P_i}{\frac{1}{j} \sum_{i=1}^{N} P_i} \right| \tag{16}$$

If $RP_j \geq 1$, then the first $(j-1)$ ISCs can be regarded as false components, whose products of the energy density of mode components and their average periods are constant. So they should be removed together with the residue $r(t)$.

Step 2: Correlation coefficients between the rest ISCs and the original signal are calculated and The ISCs with high correlation coefficient are selected out. Correlation analysis is a powerful tool to measure the correlation degree between two time series. Considering that Spearman Correlation Coefficient is superior to the commonly-used Pearson correlation coefficient, it is used as a criterion to discriminate the ISCs, which carry more relevant information with the original signal. For two variables $X = \{x_i\}$ and $Y = \{y_i\}$ with the data length N, they can be converted into ranks $g(x_i)$ and $g(y_i)$

respectively, then the Spearman Correlation Coefficient is defined as [35]:

$$r_s = 1 - \frac{6 \sum d_i^2}{N^2(N^2 - 1)} \tag{17}$$

where $d_i = g(x_i) - g(y_i)$.

The greater the correlation coefficient is, the more relevant the two variables are. Hence, if the correlation coefficient of an ISC is high, then the ISC is highly relevant with the original signal. If the is low, then the ISCs is a false ISC, produced by signal decomposition procedure or by background noise. In the end, m ISCs with high correlation coefficient can be obtain.

Step 3: The selected m ISCs are reconstructed into a new signal, written as,

$$\check{S}(t) = \sum_{i=1}^{m} ISC_i \tag{18}$$

Step 4: In the end, the inherent MA_MFEs are extracted according to Equation (10) from the reconstructed signal as features prepared for intelligent recognition.

III. PROPOSED FAULT DETECTION SCHEME

The vibration signals generated by the local faults of machinery can be easily converted to electrical signal and collected by sensors with a large bandwidth and contain extremely rich fault information. The intrinsic oscillations included in the vibration signals will change under the working states, representing the non-linear and non-stationary characteristics and the energy distribution variety of the different frequency components in the spectrum. In the other words, the complexity of time series will greatly vary with different working states.

The vibration signals contain unavoidably background noises and interference signals. The non-local means (NLM) denoising method has become an increasing popular approach in imagine denoising area due to its excellent performance for Gaussian additive and multiplicative noise [38], [39]. The NLM method can be also utilized to the vibration signals. Hence, the NLM technique is served as a preprocessing technique in this work. At the same time, the vibration signals include also superimposed trend oscillations which could influence the standard deviation of the estimated values. Therefore, following the NLM denoising, in order to eliminate the superimposed trend oscillations, the original vibration signal $x(t)$ can be decomposed into several ISCs and a residue $r(t)$ by the PELCD technique. Next, m ISCs with higher spearman correlation coefficients will be selected out to reconstruct into a new signal $y(t)$ as Equation (18). The superimposed trend oscillations are eliminated after reconstructing and the new signal only contains the inherent ISCs. So the inherent MA_MFEs can be extracted from $y(t)$ to valuate the complexity as fault features.

When a large scale factor is applied, the dimension of features is high. However, some features are not irrelevant for classes and slow down the computing speed or calculation accuracy. If these irrelevant features are removed and

TABLE 3. Details of datasets for rolling bearing experiment.

Class	Fault size /in.	Fault severity	Number of sample	Data length	Class label
Inner race fault	0.007	slight	55	2048	1-IRF1
	0.014	moderate	55	2048	2-IRF2
	0.021	severe	55	2048	3-IRF3
Ball fault	0.007	slight	55	2048	4-BF1
	0.014	moderate	55	2048	5-BF2
	0.021	severe	55	2048	6-BF3
Outer race fault	0.007	slight	55	2048	7-ORF1
	0.014	moderate	55	2048	8-ORF2
	0.0021	severe	55	2048	9-ORF3
Normal			55	2048	10-Norm

the information for effective class recognition is presented with fewer features, classifier would have less computational expense and higher accuracy. Therefore, it is vital to obtain the optimal features as input variables. Here, the ReliefF [42] method is employed to obtain the optimal features according to their importance values. These optimal features are used to train classifiers. Once the classifiers are achieved, the fault features from testing samples will be input into the trained classifier and produce output for detection results. The flowchart of the proposed fault detection scheme is shown in Fig. 3.
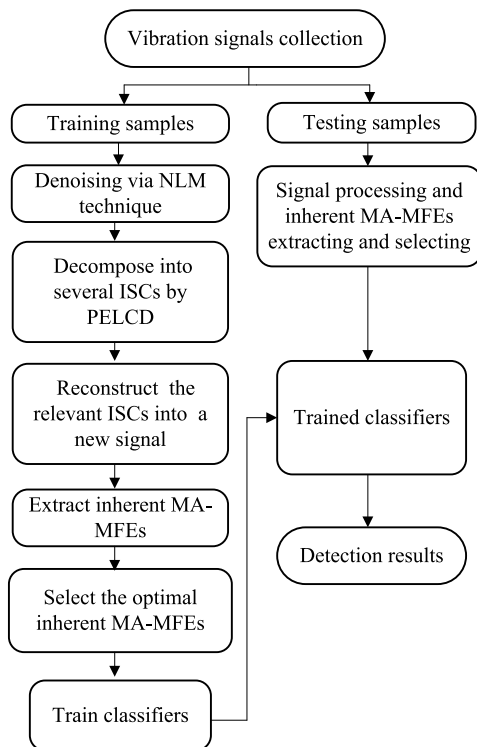

FIGURE 3. Flowchart of proposed fault detection scheme.

In the following section, two fault diagnosis experiments are made. One is for rolling bearings, which is one of the most key components in rotating machines and its failure is one of the most frequent reasons for machine breakdown. The other is for a piston pump which is a typical reciprocating machinery and plays vital role in the hydraulic system of some large-scale equipments. Moreover, in the second experiment, multi-fault diagnosis will be completed

IV. APPLICATION TO FAULT DETECTION
A. EXPERIMENT 1
In this section, the experimental datasets shared by Case Western Reverse Bearing Data Center [40] are used to verify the fault detection scheme proposed in Section **III**. The single point faults were set to the driven-end bearings with fault diameters of 0.007 to 0.021 in. The sampling frequency is 12kHz. The datasets include inner race fault (IRF), ball fault (BF), outer race fault (ORF) and normal state. The motor loads is 2hp, the shaft rotation speed is $f_r = 1750 rpm$. Ten classes of vibration signals are obtained. In this paper, the datasets are divide into 55 segments as samples with the length $N = 2048$. The details of datasets are given in the literature [41] and listed in Table 3.

In order to show the necessity of denoising procedure envelop spectrum analysis is made for inner race fault. The left top figure of Fig. 4 shows that the vibration signal with inner race fault contains plenty of background noise, and the right top figure is denoised signal after applying the NLM approach. According to the theoretical calculation, the fault frequency for inner race is $f_i = 5.415 f_r$ [41], equaling to 158Hz. As shown in the bottom figures, though the fault frequency can be found in the left bottom figure, more than frequency components exist in the envelop spectrum graphic of the original signal and confuse the operator. Whereas, as seen in the right bottom figure, the fault frequency f_i and $2f_i$ can outstand more clearly in the envelop spectrum graphic of the denoised signal. Subsequently, each signal is decomposed into several ISCs and the correlation coefficients were

TABLE 4. Comparison results for intelligent fault diagnosis on roller bearings.

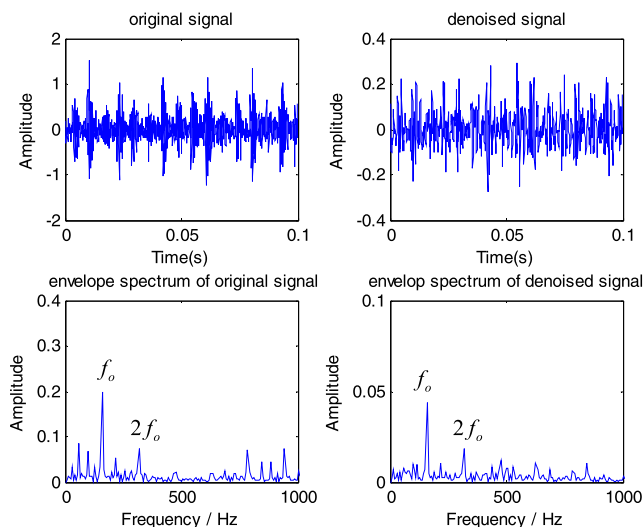| Feature extraction method	Number of test samples	Optimal Features	multiSVM			VPMCD		
			Mean (%)	Std(%)	Time cost (s)	Mean (%)	Std (%)	Time cost (s)
MA_MFE	450	F_{10} F_{11} F_9 F_5 F_{12} F_2 F_1	100.00	0	0.28	100.00	0	0.01
RCMFE[21]	450	F_1 F_2 F_3 F_6 F_5 F_4 F_7 F_8 F_{10}	96.52	0.22	0.45	98.52	0.05	0.06
MFE[19]	450	F_1 F_2 F_5 F_4 $F_3$$F_6$ F_7 F_8 F_9	96.41	0.34	0.70	97.09	0.12	0.12

FIGURE 4. Vibration signal with inner race fault and its envelop spectrum.

computed and the ISCs with the higher correlation coefficients are built into a new signal in which superimposed trend oscillations are eliminated. Lastly, the inherent MA_MFEs of the signals are extracted as fault features. According the length of dataset, the scale factor is set to maximum 20 here. In order to make a comparison, RCMFEs are computed at the same time. In this experiment, 550 samples (55 samples of each class) are used and the results of inherent MA_MFE and RCMFE under ten different states are demonstrated in Fig. 5, from which some conclusions can be obtained. First of all, the inherent MA_MFE values are almost monotonically decreasing with the scale factor growing, which is consistent with the definition of multiscale entropy and the nature of the vibration signals of rolling bearing. However, the RCMFE values do not meet the basic law. Hence, the inherent MA_MFE method has more physical sense than the RCMFE technique and be more suitable to analysis the dynamic characteristics of rolling bearings. Moreover, from the observation to the inherent MA_MFE curves, it can be found that the inherent MA_MFE values from normal state are much larger than those from the normal state so that the faulty states are easily distinguished out. This can be explained by the fact that the self-similarity of the vibration signals increase when the local faults occur, which cause the entropy

decrease correspondingly. The main reason can be explained that the periodic impact characteristics generated by the faults in the vibration signals become more intense with the fault deepening and the self-similarity is more evident. The vibration signals present more regularity and smaller entropy values, which are consistent with the dynamics of rolling bearings.

In this experiment, there are 10 classes and 55 samples in each class as described in Table 3. These samples are randomly divided to $10 \times 10 = 100$ samples for training and $10 \times 45 = 450$ samples for test. After feature extraction by the MA_MFE method, initial fault features can be obtained. The training sample matrix is noted as $\text{Train}_{100 \times 19}$, and the test matrix as $\text{Test}_{450 \times 19}$. These features under different scale factors are able to discover some fault information in different views. However, some features are less relevant with the fault type and are likely to make computation time increase, even become inaccurate. Here, the ReliefF technique is employed to reach the optimal features to boost the performance of classifiers. For the training matrix $\text{Train}_{100 \times 19}$ obtained above, the ReliefF computes the importance values of all features and sorts them from low to high. Seven features with higher importance values are facilitated as the optimal features. They were F_{10}, F_{11}, F_9, F_5, F_{12}, F_2 and F_1 (F_i stands for the MA_MFE value under the scale factor $\tau = i$), which are more sensitive to reflect the fault information from vibration signals. Now, an intelligent classifier is desired to achieve fault detection. Here, two kinds of class discrimination methods are employed. One is widely-used multiclass support vector machine (multiSVM) in the case of small samples. In this work, one-vs-all technique is utilized to extend the basic SVM with linear kernel to multiSVM. Besides, an improved variable predictive model based class discrimination (VPMCD) is served as the other multiclass technique [33]. Table 4 shows the results of intelligent fault detection and comparison.

From Table 4, the proposed multi-fault diagnosis model presents the best performance—the highest detection accuracy, the smallest standard deviate (Std) and the lowest time consumption. When using the MA_MFEs as input variables, both the multiSVM classifier and the improved VPMCD classifier have achieved 100% fault detection rate, more importantly, and the time cost is greatly short, which is vital for online fault detection. In comparison, the Std of both
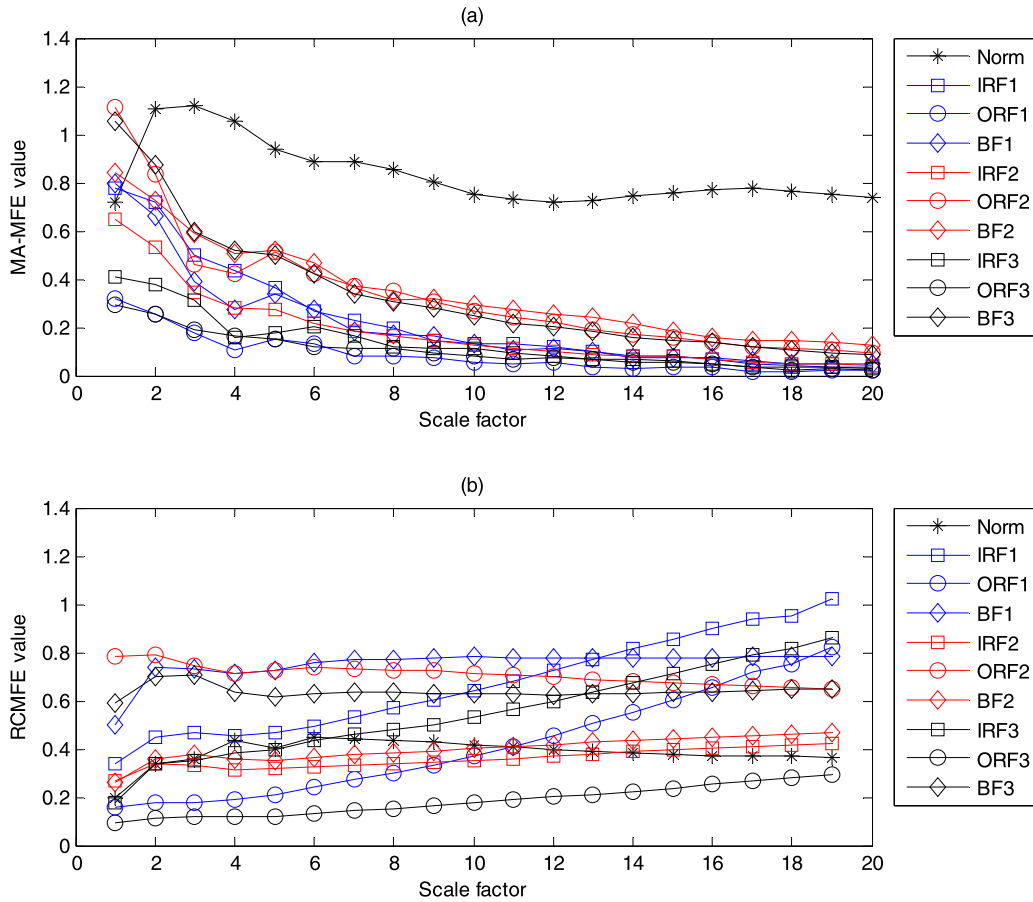
FIGURE 5. Results of (a) MA_MFE and (b) RCMFE under ten different states.

the RCMFE and the MFE are bigger, and the time costs are longer, especially when using the multiSVM classifier. On the other hand, from the comparison between multiSVM and VPMCD, the latter is proved faster and more robust, which indicates that the latter is more suitable to address the classification problem with small samples.

B. EXPERIMENT 2

In Experiment 2, the proposed scheme is applied to realize incipient abrasion detection for a piston pump. The experimental datasets are collected from axial piston pump A11VLO190 with nine pistons. The speed of driving shaft was 1600 RPM and the pressure of the main hydraulic circuit was kept at 10MPa. The accelerometer was installed to the axial piston pump with magnetic base to collect the vibration signal using NI9233 data acquisition card with a sampling frequency of 10 kHz. The axial piston pump was tested under six different conditions; They are noted as class 1 to class 6: class 1– normal condition (Normal), class 2–one piston abrasion (OPA), class 3–swash pate abrasion (SPA), class 4–valve plate abrasion (VPA), class 5–both swash plate and valve plate abrasion (BSVA), class 6–counter-position pistons abrasion (CPPA). There are three single fault types and two compound fault types. The vibration signals collected for different
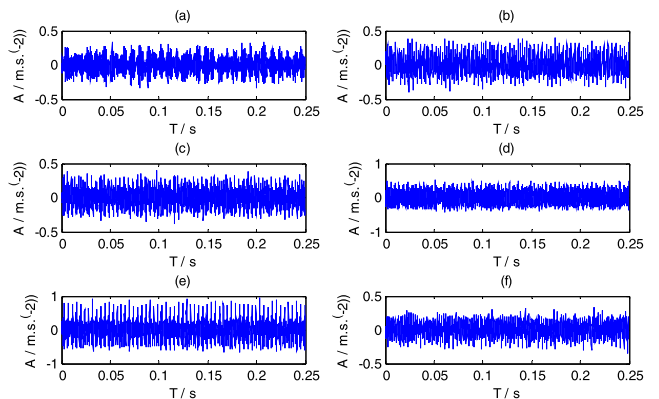

FIGURE 6. Time domain waveform. (a) Normal condition. (b) OPA. (c) SPA. (d) VPA. (e)BSVA. (f) CPPA.

condition are given in Fig.6. In this experiment, we collected 55 raw vibration signals under each condition and 330 raw vibration signals were collected in all. The length of each vibration signal is N = 2500. The samples are divided into two groups, one group is used for training and the other for test.

Similarly, after the NLM denoising, the signal is decomposed into several ISCs by PELCD, partly shown in Fig. 7.

TABLE 5. Comparison results for fault diagnosis on a piston pump.

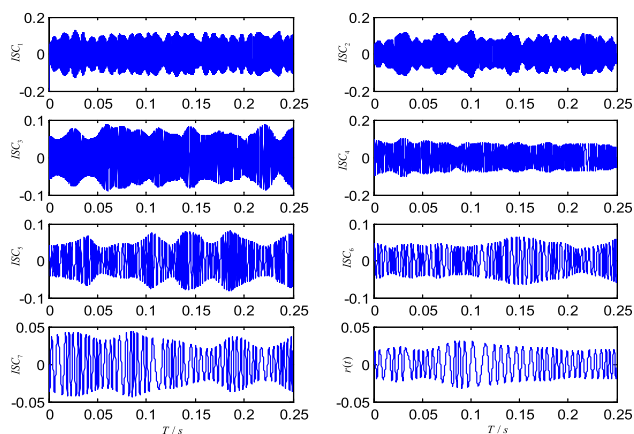| Feature extraction method	Number of test samples	multiSVM			VPMCD		
		Mean (%)	Std (%)	Time cost (s)	Mean (%)	Std (%)	Time cost (s)
MA_MFE	270	100.00	0	3.36	100.00	0	0.02
RCMFE[21]	270	99.26	0.05	4.45	99.67	0.62	0.05
MFE[19]	270	98.93	0.70	4.60	93.70	1.13	0.05

FIGURE 7. PELCD results of vibration signal under OPA.
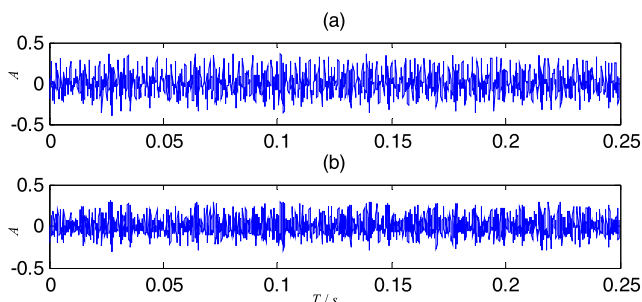
FIGURE 8. Waveforms of a vibration signal under OPA (a) original signal (b) reconstructed signal.
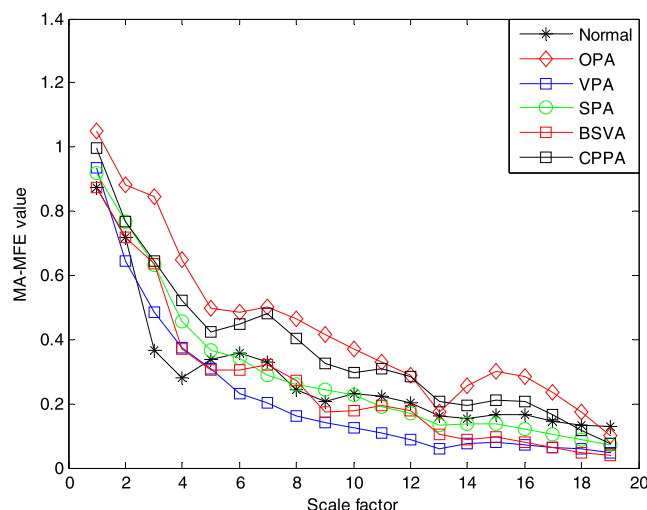

FIGURE 9. Inherent MA_MFE values of vibration signals for a piston pump with abrasion faults.
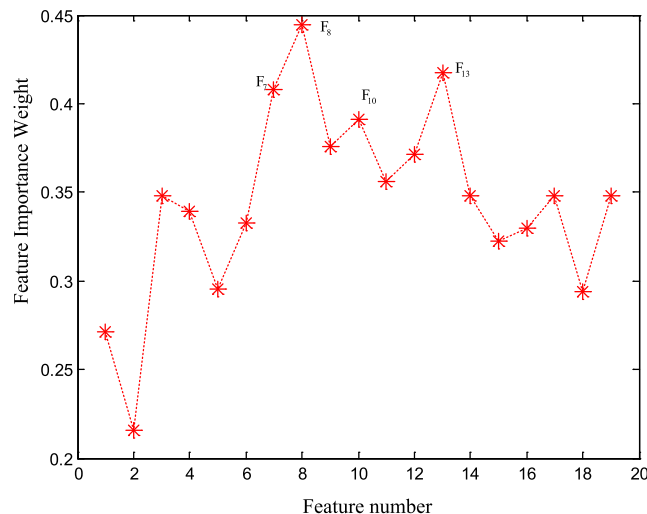
FIGURE 10. Feature importance weights ranked by ReliefF method.

Then, the correlation coefficients between the ISCs and the denoised signal are computed. In the experiment, the maximum correlation coefficient is 0.75 and the threshold is set to one tenth of the maximum correlation coefficient, that is 0.075. The ISCs with higher correlation coefficient are used to reconstruct into a new signal, shown in Fig. 8, from which it can be found that superimposed trend oscillations are reduced. In the end, the fault features are acquired by the MA_MFE method with the maximum scale factor $\tau_{max} = 20$. Fig. 9 shows the MA_MFE values. the multiSVM and the improved VPMCD are employed to make the multi-fault detection again. The first four features are selected out using the ReliefF technique, which are F_8, F_{13}, F_7, F_{10}, shown in Fig.10 as optimal features. The comparison

results are given in Table 5. The detection rates of both multiSVM and VPMCD classifiers are 100%. Simultaneously, the cost time for VPMCD training and classification is 0.02s. Therefore, the results show that the proposed model is suitable for real-time online fault detection again.

V. CONCLUSION

Multiscale entropy is a powerful tool to analysis complex time series. However, it is difficult to characterize the short-term datasets using the existing algorithms. Hence, an improved multiscale fuzzy entropy MA_MFE is developed. In the technique, the moving-average algorithm is utilized to obtain the coarse-grains for short-term complex time series. The results from simulation analysis show that MA_MFE is more accurate and robust than the existing MFE and RCMFE techniques. Meantime, combining the new data-driven signal processing method – partly ensemble local characteristic-scale decomposition, inherent MA_MFE feature extraction method is proposed. Furthermore, a novel fault detection scheme has been developed for mechanical system. The comparison results on roller bearings show that the proposed method works more effectively, reliably and quickly. In addition, the proposed model can also be successfully employed to identify the incipient multiple abrasion of a piston pump, which shows that the proposed method can achieve incipient multi-fault diagnosis online even if only short-term samples and less training samples are collected. Our further work will focus on how to solve real–time online fault diagnosis for large-scale machinery under variable working conditions using the proposed method.

REFERENCES

[1] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 35, nos. 1–2, pp. 108–126, Feb. 2013.

[2] Y. Li, Y. Yang, X. Wang, B. Liu, and X. Liang, "Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine," *J. Sound Vib.*, vol. 428, pp. 72–86, Aug. 2018.

[3] X. Jiang, J. Wang, J. Shi, C. Shen, W. Huang, and Z. Zhu, "A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines," *Mech. Syst. Signal Process.*, vol. 116, pp. 668–692, Feb. 2019.

[4] Q. Wu, Y. Guo, H. Chen, X. Qiang, and W. Wang, "Establishment of a deep learning network based on feature extraction and its application in gearbox fault diagnosis," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 125–149, Jun. 2019.

[5] Y. Qin, X. Wang, and J. Zou, "The optimized deep belief networks with improved logistic sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3814–3824, May 2019, doi: 10.1109/tie.2018.2856205.

[6] W. Huang, J. Cheng, and Y. Yu, "Rolling bearing performance degradation assessment based on convolutional sparse combination learing," *IEEE Access*, vol. 7, pp. 17834–17846, 2019.

[7] J. Ben Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Appl. Acoust.*, vol. 89, pp. 16–27, Mar. 2015.

[8] J. Du, S. Wang, and H. Zhang, "Layered clustering multi-fault diagnosis for hydraulic piston pump," *Mech. Syst. Signal Process.*, vol. 36, no. 2, pp. 487–504, Apr. 2013.

[9] H. Li, J. Sun, H. Ma, Z. Tian, and Y. Li, "A novel method based upon modified composite spectrum and relative entropy for degradation feature extraction of hydraulic pump," *Mech. Syst. Signal Process.*, vol. 114, pp. 399–412, Jan. 2019.

[10] Y. Lan, J. Hu, J. Huang, L. Niu, X. Zeng, X. Xiong, and B. Wu, "Fault diagnosis on slipper abrasion of axial piston pump based on extreme learning machine," *Measurement*, vol. 124, pp. 378–385, Aug. 2018.

[11] 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, Jun. 2000.

[12] M. Li, R. Wang, J. Yang, and L. Duan, "An improved refined composite multivariate multiscale fuzzy entropy method for MI–EEG feature extraction," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–12, Mar. 2019, doi: 10.1155/2019/7529572.

[13] Z. Cao and C.-T. Lin, "Inherent fuzzy entropy for the improvement of EEG complexity evaluation," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 1032–1035, Apr. 2018.

[14] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, Jul. 2002, Art. no. 174102.

[15] Z. Chen, Y. Li, H. Liang, and J. Yu, "Improved permutation entropy for measuring complexity of time series under noisy condition," *Complexity*, vol. 2019, pp. 1–12, Mar. 2019, doi: 10.1155/2019/1403829.

[16] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, Jul. 2002, Art. no. 068102.

[17] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 2, pp. 1–18, Feb. 2005.

[18] 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, Jun. 2007.

[19] J. Zheng, J. Cheng, Y. Yang, and S. Luo, "A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination," *Mechanism Mach. Theory*, vol. 78, pp. 187–200, Aug. 2014.

[20] J. Zheng, H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mech. Syst. Signal Process.*, vol. 85, pp. 746–759, Feb. 2017.

[21] H. Azami, A. Fernandez, and J. Escudero, "Refined multiscale fuzzy entropy based on standard deviation for bimedical signal analysis," *Med. Biol. Eng. Comput.*, vol. 55, no. 11, pp. 2037–2052, 2017.

[22] S.-D. Wu, C.-W. Wu, K.-Y. Lee, and S.-G. Lin, "Modified multiscale entropy for short-term time series analysis," *Phys. A, Stat. Mech. Appl.*, vol. 392, no. 23, pp. 5865–5873, Dec. 2013.

[23] Y. Li, M. Xu, R. Wang, and W. Huang, "A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy," *J. Sound Vib.*, vol. 360, pp. 277–299, Jan. 2016.

[24] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. A*, vol. 454, pp. 903–995, 1998.

[25] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise–assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[26] J.-R. Yeh, J.-S. Shieh, and N. E. Huang, "Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method," *Adv. Adapt. Data Anal.*, vol. 2, no. 2, pp. 135–156, Apr. 2010.

[27] J. Cheng, J. Zheng, and Y. Yang, "A nonstationary signal analysis approach—The local characteristic-scale decomposition method," *J. Vib. Eng.*, vol. 25, no. 2, pp. 215–220, 2012.

[28] S. Luo, J. Cheng, and K. Wei, "A fault diagnosis model based on LCD-SVD-ANN-MIV and VPMCD for rotating machinery," *Shock Vib.*, vol. 2016, Jul. 2016, Art. no. 5141564, doi: 10.1155/2016/5141564.

[29] J. Zheng, J. Cheng, and Y. Yang, "A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy," *Mechanism Mach. Theory*, vol. 70, pp. 441–453, Dec. 2013.

[30] Y. Hao, L. Song, M. Wang, L. Cui, and H. Wang, "Underdetermined source separation of bearing faults based on optimized intrinsic characteristic–scale decomposition and local non–negative matrix factorization," *IEEE Access*, vol. 7, pp. 11427–11435, 2019.

[31] J. Cheng, J. Zheng, Y. Yang, and S. Luo, "Fault diagnosis model for rolling bearing based on partly ensemble local characteristic-scale decomposition and Laplacian score," *J. Vib. Eng.*, vol. 27, no. 6, pp. 942–950, 2014.

[32] J. Zheng, J. Cheng, and Y. Yang, "Partly ensemble local characteristic-scale decomposition: A new noise assisted data analysis method," *Acta Electron. Sinica*, vol. 41, no. 5, pp. 1030–1035, 2013.

[33] S. Luo, J. Cheng, M. Zeng, and Y. Yang, "An intelligent fault diagnosis model for rotating machinery based on multi-scale higher order singular spectrum analysis and GA-VPMCD," *Measurement*, vol. 87, pp. 38–50, Jun. 2016.

[34] M. Hortelano, R. B. Reilly, F. Castells, and R. Cervigón, "Refined multiscale fuzzy entropy to analyse post-exercise cardiovascular response in older adults with orthostatic intolerance," *Entropy*, vol. 20, no. 11, p. 860, Nov. 2018.

[35] Y. Duan and C. Song, "Relevant modes selection method based on Spearman correlation coefficient for laser signal denoising using empirical mode decomposition," *Opt. Rev.*, vol. 23, no. 6, pp. 936–949, 2016.

[36] H. Zheng, C. Dang, S. Gu, D. Peng, and K. Chen, "A quantified self-adaptive filtering method: Effective IMFs selection based on CEEMD," *Meas. Sci. Technol.*, vol. 29, no. 8, Aug. 2018, Art. no. 085701.

[37] R. Chen, B. Tang, and J. Ma, "Adaptive denoising method based on ensemble empireical mode decompositin for vibration," *J. Vib. shock*, vol. 31, no. 15, pp. 82–86, 2012.

[38] J. Antoni and R. B. Randall, "Unsupervised noise cancellation for vibration signals: Part II—A novel frequency-domain algorithm," *Mech. Syst. Signal Process.*, vol. 18, no. 1, pp. 103–117, 2004.

[39] A. Tristán-Vega, V. García-Pérez, S. Aja-Fernández, and C.-F. Westin, "Efficient and robust nonlocal means denoising of MR data based on salient features matching," *Comput. Methods Programs Biomed.*, vol. 105, no. 2, pp. 131–144, Feb. 2012.

[40] *Case Western Reserve University Bearing Data Center Website*. Accessed: May 4, 2016. [Online]. Available: http://cscgroups.case.edu/bearingdatacenter/home

[41] W. A. Smith and R. B. Randal, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.

[42] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, pp. 23–69, Oct. 2003.

WENXIAN YANG received the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, in 1999. Then, he did research with the City University of Hong Kong, from 2001 to 2002, Nottingham Trent University, from 2002 to 2004, Cranfield University, from 2004 to 2007, the University of Durham, from 2007 to 2009, and the U.K. Offshore Renewable Energy CATAPULT Centre, from 2009 to 2013. In May 2013, he joined Newcastle University as a Lecturer, where he was promoted to a Senior Lecturer, in 2018. With the expertise in machine condition monitoring, fault diagnosis, signal processing, pattern recognition, artificial intelligence, system reliability analysis, and reliability design, he has been consistently striving to use advanced condition monitoring and reliability analysis techniques to help the enterprises to increase economic benefits by lowering the downtime, and operation and maintenance costs of their production equipment.

SONGRONG LUO received the Ph.D. degree from Hunan University, in 2015. She is currently an Associate Professor with the College of Mechanical Engineering, Hunan University of Arts and Science. She has been an Academic Visitor with Newcastle University, U.K., since September 2018. Her main research interests include pattern recognition and dynamics signal processing machinery fault diagnosis. Her research articles have been published in *Measurement*, *Shock and Vibration*, and *Cluster Computing*.

YOUXIN LUO has been the Vice President, a Professor, and the master Supervisor of Mechanical Engineering College, Hunan University of Arts and Science, since 2004. As an expertise in mechanical engineering, his main research works focus on grey theory and its application to parameter optimization, digital signal processing, and fault diagnosis for construction machine. His scientific research articles have been published in *Kybernetes*, the *Journal of Mechanical Engineering*, *Advances in Systems Science and Application*, and the *International Journal of Plant Engineering and Management*.

• • •