

Research Article

Fault Diagnosis of Rolling-Element Bearing Using Multiscale Pattern Gradient Spectrum Entropy Coupled with Laplacian Score

Xiaoan Yan ^(b),¹ Ying Liu,¹ Peng Ding,² and Minping Jia ^(b)

¹School of Mechatronics Engineering, Nanjing Forestry University, Nanjing 210037, China ²School of Mechanical Engineering, Southeast University, Nanjing 211189, China

Correspondence should be addressed to Xiaoan Yan; yanxiaoan89@sina.com

Received 31 August 2019; Revised 29 December 2019; Accepted 21 January 2020; Published 17 February 2020

Academic Editor: Yong Xu

Copyright © 2020 Xiaoan Yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Feature extraction is recognized as a critical stage in bearing fault diagnosis. Pattern spectrum (PS) and pattern spectrum entropy (PSE) in recent years have been smoothly applied in feature extraction, whereas they easily ignore the partial impulse signatures hidden in bearing vibration data. In this paper, the pattern gradient spectrum (PGS) and pattern gradient spectrum entropy (PGSE) are firstly presented to improve the performance of fault feature extraction of two approaches (PS and PSE). Nonetheless, PSE and PGSE are only able to evaluate dynamic behavior of the time series on a single scale, which implies there is no consideration of feature information at other scales. To address this problem, a novel approach entitled multiscale pattern gradient spectrum entropy (MPGSE) is further implemented to extract fault features across multiple scales, where its key parameters are determined adaptively by grey wolf optimization (GWO). Meanwhile, a Laplacian score- (LS-) based feature selection strategy is employed to choose the sensitive features and establish a new feature set. Finally, the selected new feature set is imported into extreme learning machine (ELM) to identify different health conditions of rolling bearing. Performance of our designed algorithm is tested on two experimental cases. Results confirm the availability of our proposed algorithm in feature extraction and show that our method can recognize effectively different bearing fault categories and severities. More importantly, the designed approach can achieve higher recognition accuracies and provide better stability by comparing with other entropy-based methods involved in this paper.

1. Introduction

With the rapid progress and improvement of the level of industrial modernization, large-scale mechanical equipment has been widely used in electric power, petrochemical, aerospace, marine, and other fields, including generator, wind turbine, compressor, aeroengine, and hydroturbine [1]. Rolling bearings are regarded as an important joint of modern mechanical equipment; due to the influence of various factors (e.g., manufacturing error, material defect, improper installation, inadequate lubrication, poor sealing, high speed, and heavy load working), its running state will inevitably change and even appear with various faults (e.g., wear, notch, scratch, smudginess, and fretting corrosion). Microlesion emerged in rolling bearing will bring great security risks to enterprises directly, light will cause some economic losses and heavy will result in personnel casualties [2-5]. On the one hand, owing to the unstable operation condition of mechanical system and the impact of some nonlinear elements (e.g., friction, clearance, and stiffness), the practical bearing fault data has the properties of nonlinear, nonstationary, and low signal-to-noise ratio (SNR), which implies that it is difficult to make effective diagnosis directly through fast Fourier transform (FFT) and its variants [6]. On the other hand, according to relevant data statistics, in the key parts (e.g., bearing, gear, and rotor) of mechanical equipment, the equipment failure rate caused by the damage of bearing is up to 30%, about 20% of the faults in the gearbox are caused by the damage of bearing, and the damage rate of bearing in the motor is up to 40%, which indicates that rolling bearing is considered to be one of the most easily damaged parts in mechanical equipment. Accordingly, exploring efficient damage detection approach is an urgent and challenging task in engineering application.

Mathematical morphology (MM) is a nonlinear signal processing algorithm and known for its simplicity and practicality [7]. Currently, there are many applications of morphological filter in mechanical condition monitoring, including feature extraction and intelligent fault diagnosis, where feature extraction is a critical stage in intelligent fault diagnosis of rolling-element bearing. For instance, Bai and Zhou [8] put forward a modified top-hat transformation to capture periodic impulse characteristics related to bearing fault. Dong et al. [9] proposed a modified morphological method based on the average operator of opening and closing and signal-to-noise ratio (SNR) to extract bearing fault feature information. Li and Liang [10] presented a continuous-scale mathematical morphology to identify impulsive fault feature information located in the optimal scale band. Shen et al. [11] presented a fast and adaptive varying-scale morphological analysis approach, and an effective detection result can be obtained. Li et al. [12] proposed a diagonal slice spectrum assisted optimal scale morphological filter for bearing feature extraction and achieved a good diagnosis result. Lv and Yu [13] proposed an average combination difference morphological filter (ACDIF) and combined Teager energy kurtosis to select the length of structuring element (SE) to extract bearing fault feature.

On the other hand, the studies of the MM-based intelligent fault diagnosis approach are also universal (e.g., one-dimensional adaptive rank-order morphological filter [14], the combination of morphological filter and k-nearestneighbor classifier [15], the combination of morphological operators and fuzzy inference [16], the combination of MM and support vector machine [17], the combination of morphological filter and local tangent space alignment [18], and the combination of morphological filter and grey relational degree [19]). Meanwhile, most MM-based fault diagnosis studies focus on three aspects including morphological fractal dimension [20], morphological particle [21], and pattern spectrum (PS) [22]. Thereinto, PS can reveal shape characteristics of bearing vibration signal based on the morphological opening or closing operation with multiscale SE, while pattern spectrum entropy (PSE) is defined based on PS, which is regarded as a newly spawned index of complexity estimation [23]. At present, some applications with respect to the PS and PSE in fault detection have been conducted. Hao et al. [24] applied PSE to extract the feature vectors from bearing fault signal, and then the extracted fault feature is imported into support-vector machine (SVM) to complete the identification of different fault types of bearing. Zhang et al. [25] combined local mean decomposition (LMD) and pattern spectrum to extract fault signatures and recognize different health conditions of bearings. Zheng et al. [26] firstly adopted ensemble empirical mode decomposition (EEMD) to decompose vibration signal into several components named intrinsic mode function (IMF), then calculate PS of the first three IMF to build a feature set, and finally apply the SVM classifier to

provide the identification results of different health conditions of bearing. However, the traditional PS and PSE are developed based on morphological opening or closing operation. Obviously, morphological opening or closing operation is very suitable for extracting unilateral fault pulse characteristics of rolling bearing, but it is easy to neglect feature information hidden on another side [27]. According to theoretical knowledge of MM available, it can be known that the morphological gradient operation (MGO) can extract effectively bilateral impulse signatures (the positive and negative direction) from bearing fault signal [28, 29]. Meanwhile, MGO has the advantages of noise reduction and detail preserving. Therefore, based on MGO and PSE, a modified indicator called pattern gradient spectrum entropy (PGSE) is introduced in this paper, which can improve fault feature extraction performance of PSE and measure dynamic change of various time series. Unfortunately, whether PSE or PGSE, they assess the complexity of a time series within a single scale, which indicates that the relevant characteristics on other scales are ignored. Consequently, it is necessary to design an effective nonlinear dynamic algorithm to address this issue.

Multiscale entropy (MSE) proposed by Costa et al. [30] is an effective algorithm for estimation of complexity of signal, which can obtain different feature information over multiple scales by applying a coarse-grained procedure and overcome the weakness of inadequate feature extraction of single-scale entropy. Besides, MSE has been utilized in various fields including biomedical engineering [31], EEG data denoising [32], and mechanical fault detection [33]. For instance, Zhang et al. [34] firstly applied improved multiscale entropy to extract fault features, and then the extracted features are taken as the input of SVM classifier for achieving the identification of different health conditions of bearing. Liu and Han [35] firstly adopted LMD to separate bearing fault signal to a series of product function (PF) and then take multiscale entropy of PF as the inputs of BP neural network to recognize different fault categories of rolling bearing. Hsieh et al. [36] firstly used empirical mode decomposition (EMD) to decompose a vibration signal into several IMF components and then adopted the MSE curve of IMF component to identify the local fault of high-speed spindle. Nevertheless, the conventional MSE easily appears as an undefined entropy value for short-term time series and possesses an underlying instability with the increase of the scale factor. Furthermore, computational efficiency of traditional MSE is not high, so it will cost significantly for the application of real-time condition monitoring of industrial scene [37]. Consequently, to solve these problems existing in traditional MSE and capture more comprehensively fault symptoms over multiple scales, through the combination of the coarse-grained procedure and two entropies (i.e., PSE and PGSE), two new multiscale entropies (i.e., multiscale pattern spectrum entropy (MPSE) and multiscale pattern gradient spectrum entropy (MPGSE)) are proposed in this paper to quantify the complexity of time series, which has the performance of multiresolution analysis and can extract fault characteristics over a range of scales.

As is known to all, because the large feature dimension easily leads to the low computational efficiency and even brings about dimension disaster, it is required for feature selection after using entropy-based method to extract multiscale features [38]. Laplacian score (LS) is a newfangled feature selection approach [39], which estimates the importance of different features according to the score of the alternative features and then complete effective selection of sensitive features. Compared with some other feature selection approaches (e.g., Fisher score [40] and data variance [41]), the LS approach is simpler and easier to understand and has advantages of reserving immensely local and global structure information. Hence, the LS approach in this paper is employed to feature selection, which is aimed at reducing information redundancy and improving calculating efficiency of the whole fault diagnosis. A novel sensitive feature set can be obtained after using the LS algorithm for feature selection. Ultimately, in order to realize automatically the health status identification of bearings, an appropriate classification model needs to be employed. Existing classification models are divided into supervised and unsupervised approaches. Generally speaking, the supervised method with the mapping relation between the sample and label can achieve better classification results than the unsupervised method without the label information [42]. Various supervised classifiers in recent years are devoted to fault detection in rolling bearing, such as artificial neural network (ANN) [43], SVM [44], and Bayesian classification model [45]. However, these approaches are equipped with some shortcomings. For example, the ANN is easy to result in a local minimum, slow convergence speed, and overfitting due to the empirical risk minimization principle. Two important parameters (i.e., penalty factor c and kernel parameter q) of SVM are demanded to be defined in advance and have the very tremendous influence on its classification ability. The Bayesian classification model needs to assume that the distribution is independent, but complete independence distribution is difficult to satisfy in reality. Compared with the supervised approach above, extreme learning machine (ELM) has some advantages, including the higher computational efficiency, stronger generalization ability, and lesser human intervention [46]. Hence, in this paper, the selected new feature set is regarded as the input of ELM to identify different fault patterns of rolling bearing. In a word, the main contributions and originality of this paper are to put forward a novel bearing intelligent diagnosis scheme based on MPGSE and LS. Compared with the previous MSE, the proposed MPGSE method has some merits, which are mainly expressed in three aspects. Firstly, because of the simpleness of morphological operation, the calculation of MPGSE is faster than MSE, especially for long-term time series. Secondly, MPGSE can avoid the undefined entropy value existing in MSE and provide a more reliable and

accurate estimation of complexity of the nonlinear signal. Thirdly, MPGSE is relatively insensitive to noise and is more stable than MSE in assessing the irregularity of the signal. Of course, the proposed MPGSE also has some areas for improvement. The biggest drawback is that its computational efficiency still needs to be improved if MPGSE is applied to real-time health monitoring of mechanical equipment under variable conditions, which is planned for the focus of future research work. Specific novelties of this paper are summarized as follows:

- (1) A modified method called PGSE is introduced for improving dynamic change detection of the signal
- (2) A neoteric approach termed as MPGSE is developed to extract more abundant features over multiple scales
- (3) The LS-based feature selection strategy is employed to choose sensitive characteristics with a higher discrimination
- (4) Simulation data and experimental examples are performed to verify the feasibility of our presented approach

The remainder of this paper is organized as follows. Section 2 reviews the theory of PS and PSE. Besides, the concept of PGS and PGSE is presented in Section 2 and simulation analysis is conducted to compare their performance. In Section 3, a new method named MPGSE in which coarse-grained procedure and PGSE are combined is proposed. Section 4 describes the flowchart of the proposed approach in detail. In Section 5, two experimental cases are analyzed to show the effectiveness of the proposed approach. Finally, conclusions and some future works are given in Section 6.

2. Theory Background

2.1. Multiscale Morphological Operation. For a given signal f(n), set morphological operation g as a unit SE; λg is the SE at scale λ and is defined as

$$\lambda g = \underbrace{g \oplus g \oplus \cdots \oplus g}_{\lambda - 1 \text{ times}} = \underbrace{((g \oplus \cdots \oplus g) \oplus g) \oplus g}_{\lambda - 1 \text{ times}}, \quad (1)$$

where \oplus represents dilation operation. Multiscale dilation and erosion are expressed as

$$(f \oplus \lambda g)(n) = f \oplus \underbrace{(g \oplus g \oplus \dots \oplus g)}_{\lambda-1 \text{ times}},$$

$$(f \Theta \lambda g)(n) = f \Theta \underbrace{(g \Theta g \Theta \dots \Theta g)}_{\lambda-1 \text{ times}},$$
(2)

where Θ denotes erosion operation. Multiscale opening or closing operation can be obtained by

$$(f \circ \lambda g)(n) = ((f \Theta \lambda g) \oplus \lambda g)(n),$$

(f \cdot \lambda g)(n) = ((f \operatorname{\lambda} \lambda g) \Operatorname{\lambda} \lambda g)(n). (3)

Correspondingly, multiscale morphological gradient operation (MGO) can be written as [47]

$$\mathrm{MGO}(f(n)_{\lambda g}) = (f \oplus \lambda g)(n) - (f \Theta \lambda g)(n). \tag{4}$$

2.2. PS and PSE. If f(n) is a given one-dimensional signal and g(m) is the SE, PS is defined as

$$PS(f,\lambda,g) = \begin{cases} -\frac{dA(f \circ \lambda g)}{d\lambda}, & \lambda \ge 0, \\ \\ -\frac{dA(f \bullet (-\lambda)g)}{d\lambda}, & \lambda < 0, \end{cases}$$
(5)

where λ is the SE scale, A(f) is a finite area in the domain of definition, and \circ and \bullet are the opening and closing operation, respectively. When $\lambda \ge 0$, equation (5) is called as PS for the open operation and denoted by PS₊ (f, λ, g) , while when $\lambda < 0$, equation (5) is called as PS for the closing operation and denoted by PS₋ (f, λ, g) . PS is usually composed of the positive and negative intervals, where PS₊ (f, λ, g) represents intrinsic information of the signal and PS₋ (f, λ, g) represents background information of the signal. Because the positive and negative intervals are consistent, PS₋ (f, λ, g) is usually used for calculating the PS. If f(n) is an one-dimensional discrete signal, equation (5) is rewritten as

$$\begin{cases} PS_{+}(\lambda, g) = A[(f \circ \lambda g) - (f \circ (\lambda + 1)g)], & 0 \le \lambda \le \lambda_{\max}, \\ PS_{-}(\lambda, g) = A[(f \bullet (-\lambda)g) - (f \bullet (-\lambda + 1)g)], & \lambda_{\min} \le \lambda \le 0, \end{cases}$$
(6)

where $A(f) = \sum_{n} f(n)$. According to the definition of Shannon entropy, PSE is defined as

$$PSE(f/g) = -\sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} q(\lambda) \ln q(\lambda).$$
(7)

2.3. PGS and PGSE. PS and PSE are defined based on the opening or closing operation, whereas the opening or closing operation only can extract negative impulses, which is easy to ignore some key information. Because MGO can disclose simultaneously the positive and negative impulse sequences, PGS and PGSE are presented in this section by combining MGO and two methods (PS and PSE). If f(n) is an one-dimensional signal and g(m) is the SE, PGS is expressed as

$$\operatorname{PGS}(f,\lambda,g) = \begin{cases} -\frac{\mathrm{dA}(f \oplus \lambda g - f \Theta \lambda g)}{d\lambda}, & \lambda \ge 0, \\ -\frac{\mathrm{dA}(f \Theta(-\lambda)g - f \oplus (-\lambda)g)}{d\lambda}, & \lambda < 0. \end{cases}$$
(8)

If f(n) is an one-dimensional discrete signal, equation (8) is rewritten as

$$\begin{cases} PGS_{+}(\lambda,g) = A[(f \oplus (\lambda+1)g) - (f \Theta (\lambda+1)g) - (f \oplus \lambda g - f \Theta \lambda g)], & 0 \le \lambda \le \lambda_{\max}, \\ PGS_{-}(\lambda,g) = A[(f \Theta (-\lambda+1)g) - (f \oplus (-\lambda+1)g) - (f \Theta (-\lambda)g - f \oplus (-\lambda)g)], & \lambda_{\min} \le \lambda \le 0, \end{cases}$$
(9)

where $A(f) = \sum_{n} f(n)$. According to the definition of Shannon entropy, PGSE is defined as

$$PGSE(f,\lambda) = -\sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} q_1(\lambda) \ln q_1(\lambda).$$
(10)

Apparently, the larger the PGSE is, the more the complexity and uncertainty of the signal is.

2.4. Simulation Analysis. To verify the effectiveness of PGS and PGSE, a simplified model with bearing local fault is established as follows:

$$x(t) = \exp\left(-\alpha t_0\right) \left(\cos\left(2\pi f_c t\right)\right) + r(t), \tag{11}$$

where $t_0 = \text{mod}(k/f_s, 1/f_m), k = 0, 1, 2, \dots, 4999, \alpha$ and f_c denote attenuation coefficient and carrier frequency of the system, respectively, and f_s and f_m denote sampling frequency and fault frequency of the signal, respectively. Because the noise contained in the practical measured data is often colored noise, r(t) in the above fault model should be considered as a nonwhite noise (i.e., pink noise). To study

more comprehensively the efficacy of the proposed method, here we investigate two cases (i.e., simulation signal containing a white noise/nonwhite noise (pink noise)). Specifically, in equation (11), the sampling frequency $f_s = 10 \text{ kHz}$, the sampling length L = 5000 points, the attenuation coefficient $\alpha = 800 \text{ rad/s}$, the carrier frequency f_c is set to be 2000 Hz, 3000 Hz, and 4000 Hz, respectively, r(t) is a white noise/nonwhite noise, and the fault frequency f_m is set to be 60 Hz, 120 Hz, and 180 Hz, respectively. Hence, three simulation signals containing a white or nonwhite noise can be obtained, which are, respectively, considered as the vibration signal generated by the outer race fault (ORF), inner race fault (IRF), and ball fault (BF) of rolling bearing. Figure 1 shows temporal waveform of three simulation signals under two noises (i.e., white noise and pink noise).

Apparently, the familiar SE includes the flat, triangle, and semicircle. The flat SE is simple in structure and its computing time is lesser compared with the triangle and semicircle SE, so the flat SE g = [0, 0, 0] is selected directly to perform PGS and PGSE. In the process of calculating the PGS and PGSE, the largest SE scale λ_{max} is set as 20. Figures 2(a)–2(d) plot the PS,



FIGURE 1: Temporal waveform of three simulation signals under different noises: (a) white noise and (b) pink noise.

PGS, PSE, and PGSE of three simulation signals under white noise, respectively. Note that 20 data samples of each simulation signal are selected in the process of calculating PSE and PGSE. As shown in Figure 2(a), PS has inferior smoothness and upand-down fluctuating phenomenon, whereas PGS has good smoothness and distinction degree, as shown in Figure 2(b). As seen in Figures 2(c) and 2(d), the stability of PGSE is better than that of PSE. Besides, PGSE has better differentiation among different simulation signals compared with PSE, which indicates feature extraction performance of PGSE is better than that of PSE. Figures 3(a)-3(d) show the PS, PGS, PSE, and PGSE of three simulation signals under pink noise, respectively. By comparing Figures 3(a) and 3(b), we can observe that the smoothness of PGS is also better than that of PS when the colored noise is added into the fault model. Besides, it can be seen clearly from Figures 3(c) and 3(d) that the separability and stability of using PGSE for three simulation signals is superior to that of using PS. This further verifies the effectiveness of the proposed PGS and PGSE methods. To observe the computational efficiency, CPU running time of different methods under white noise and pink noise is listed in Table 1. All methods are conducted on a computer with Intel Pentium CPU G3420 @ 3.20 GHz/4.00 GB RAM. As given in Table 1, whether white noise or nonwhite noise, computation time of PGS and PGSE is slightly lower than that of PS and PSE. Generally speaking, because four algorithms (i.e., PS, PSE, PGS, and PGSE) are implemented on the basis of dilation and erosion operation, their computational efficiency is very close in most applications.

3. MPSE and MPGSE

3.1. Multiscale Entropy. MSE presented by Costa et al. [30] can generate a sequence of scale time series and effectively describe the irregularity of the nonstationary signal. To be specific, MSE is considered as the set of sample entropy of a time series at different scales, but MSE easily cause the

undefined value for short-term time series. The concept of MSE is summarized as follows:

(1) For a given time series $\{x(i): 1 \le i \le N\}$, the coarsegrained time series at different scale factors τ can be obtained by equation (12). Taking scale factor $\tau = 2$ and $\tau = 3$ as an example, the corresponding coarsegrained procedure is displayed in Figure 4. As shown in Figure 4, coarse-grained procedure of MSE is briefly interpreted as the process of averaging the raw time series in the window with a length of τ , and the downsampling is subsequently carried out at a scale factor of τ :

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x(i), \quad 1 \le j \le \left\lfloor \frac{N}{\tau} \right\rfloor,$$
 (12)

where $\tau = 1, 2, \ldots$ denotes the scale factor. Specifically, when $\tau = 1$, the coarse-grained time series is equivalent to the raw signal. For $\tau > 1$, the raw signal is separated into τ coarse-grained time series with a length of $\lfloor N/\tau \rfloor$, where $\lfloor \bullet \rfloor$ denotes the round-off number.

(2) Computing sample entropy of each coarse-grained time series $y_j^{(\tau)}$ $(1 \le j \le N/\tau)$ and then plotting whole sample entropy as a function of scale factor τ ,

$$MSE(x, m, r, \tau) = SE(y_j^{(\tau)}, m, r), \qquad (13)$$

where *m* means the embedding dimension, *r* represents the similarity tolerance, and τ denotes the scale factor. The flowchart of the MSE algorithm is shown in Figure 5, where τ_m represents the maximum scale factor.

3.2. MPSE and MPGSE. To extract different fault features over multiple scales and avoid the drawbacks of the undefined entropy value existing in MSE, two new multiscale



FIGURE 2: The analysis results of different simulation signals under white noise: (a) PS, (b) PGS, (c) PSE, and (d) PGSE.

entropies (i.e., MPSE and MPGSE) are proposed by combining the coarse-grained process and two entropies (i.e., PSE and PGSE), which are defined as follows:

 For a time series {x (i): 1 ≤ i ≤ N}, the coarse-grained time series at different scale factors τ is obtained using the following formula:

$$y_{j}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x(i), \quad 1 \le j \le \frac{N}{\tau}, \ \tau = 1, 2, \ \dots, \ \tau_{\max},$$
(14)

where τ_{\max} stands for the defined largest scale factor. Note that τ_{\max} coarse-grained time series with the length of $\lfloor N/\tau \rfloor$ are obtained when $\tau = \tau_{\max}$, where $\lfloor \bullet \rfloor$ denotes the round-off number. (2) Compute PS and PGS of each coarse-grained time series y_j^(τ) (1 ≤ j ≤ N/τ), that is,

$$PS_{+}(y_{j}^{(\tau)},\lambda,g) = A[(y_{j}^{(\tau)}\circ\lambda g) - (y_{j}^{(\tau)}\circ(\lambda+1)g)],$$

$$PGS_{+}(y_{j}^{(\tau)},\lambda,g) = A[(y_{j}^{(\tau)}\oplus(\lambda+1)g) - (y_{j}^{(\tau)}\Theta(\lambda+1)g) - (y_{j}^{(\tau)}\Theta\lambda g - y_{j}^{(\tau)}\Theta\lambda g)],$$

$$-(y_{j}^{(\tau)}\oplus\lambda g - y_{j}^{(\tau)}\Theta\lambda g)],$$

(15)

where $\lambda = 0, 1, ..., \lambda_{max}$ and λ_{max} is the largest SE scale.

(3) According to the definition of Shannon entropy, PSE and PGSE of each coarse-grained time series $y_j^{(\tau)}$ can be obtained, and then plotting whole PSE and PGSE as a function of scale factor τ , we obtain the following:



7



FIGURE 3: The analysis results of different simulation signals under pink noise: (a) PS, (b) PGS, (c) PSE, and (d) PGSE.

TABLE 1: CPU running time of different methods under white noise and pink noise.

Noise type	PS	PSE	PGS	PGSE
White noise	3.3335 s	3.3808 s	3.1778 s	3.2130 s
Pink noise	4.1667 s	4.2376 s	4.1083 s	4.1392 s

$$MPSE(x, \lambda, \tau) = PSE(y_j^{(\tau)}, \lambda, \tau),$$

$$MPGSE(x, \lambda, \tau) = PGSE(y_j^{(\tau)}, \lambda, \tau),$$
(16)

where λ means the SE scale and τ denotes the scale factor. Figures 6(a) and 6(b) describe the flowchart of two methods (MPSE and MPGSE), respectively. Apparently, the key idea of MPSE and MPGSE is the coarse-grained procedure and the morphological operation containing the SE scale, which indicates MPSE and MPGSE do not yield the undefined value for short-term time series. That is, compared with MSE, MPGSE

has some superiority in avoiding the undefined entropy value and can provide a more reliable and accurate estimation of dynamic change. Nonetheless, two new multiscale entropies (i.e., MPSE and MPGSE) and MSE are similar in theory. Firstly, two new multiscale entropies (i.e., MPSE and MPGSE) and MSE have the same coarse-grained process. Secondly, the SE scale λ of two multiscale entropies (i.e., MPSE and MPGSE) can be interpreted as embedding dimension *m* of MSE. That is, MSE can obtain ten-dimensional reconstructed vector when embedding dimension m = 10, while two multiscale entropies (i.e., MPSE and MPGSE) can acquire different temporal components on 10 scales when the SE scale $\lambda = 10$.

3.3. Parameter Selection of MPSE and MPGSE. Two important parameters (i.e., SE scale λ and scale factor τ) are required to be set beforehand when two multiscale entropies



FIGURE 4: Exhibition of coarse-graining procedure for scale factor $\tau = 2$ and $\tau = 3$.



FIGURE 5: Flowchart of the MSE method.

(i.e., MPSE and MPGSE) are introduced to deal with nonlinear and nonstationary data.

Firstly, because the flat SE is simple in structure and its computational efficiency is higher than that of the triangle and semicircle SE, we select the flat SE in this study. Moreover, according to properties of multiscale morphological analysis, for the selection of SE scale λ , if λ is selected as too large, two multiscale entropies (i.e., MPSE and MPGSE) have a good noise immunity, but they will have poor capability of preserving detailed information and are more time-consuming. Conversely, if λ is selected as too small, two multiscale entropies (i.e., MPSE and MPGSE) have a superior detail-preserving ability and high calculating efficiency, but their antinoise property is inferior. Hence, based on these facts, we suggest the flat SE scale $\lambda = 1 \sim \left| f_s / f_q \right| - 2$ for two multiscale entropies (i.e., MPSE and MPGSE), where $\lfloor \bullet \rfloor$ is the round-off number, and f_s and f_{q} are sampling frequency and fault frequency, respectively. Here, there are several reasons for the choice of the flat SE scale. On the one hand, according to Dong's suggestion [9], the maximal analysis length of flat SE is normally set as $[f_s/f_g]$, which is meaningful and can cover feature information of one fault repetition period completely. On the other hand, according to [48, 49], the relationship between the length and scale of flat SE satisfies $L = \lambda + 2$. That is, the maximal analysis scale of flat SE can be set as $|f_s/f_q| - 2$. Therefore, we recommend the range of the flat SE scale is located between 1 and $|f_s/f_q| - 2$, which can achieve a tradeoff between noise reduction and detail preserving.

Secondly, according to related properties of MSE, for the selection of scale factor τ , if τ is set as too small, two multiscale entropies (i.e., MPSE and MPGSE) will not be able to extract complete and comprehensive fault feature signatures [50]. On the contrary, if τ is set as too big, two multiscale entropies (i.e., MPSE and MPGSE) will give rise to the instability entropy value at the larger scale factor. Besides, the bigger τ will increase computation time of two multiscale entropies (i.e., MPSE and MPGSE). Consequently, according to the recommendation of [51], the maximum scale factor τ_{max} is usually chosen as 20, which can achieve an accurate evaluation of complexity of the signal and is sufficient to handle the actual data.

3.4. Comparison between MPGSE and MPSE Using Simulation Signal. To validate the efficacy of the presented MPGSE algorithm, two stochastic signals (white noise and 1/f noise)



FIGURE 6: Flowchart of (a) MPSE and (b) MPGSE.

containing 2000 data points are used to compare the performance of MPGSE and MPSE. In the process of comparison, we select the parameters $\lambda = 4$ and $\tau = 20$. Figures 7(a) and 7(b) depicts temporal waveform of the white noise and 1/*f* noise, respectively. Figures 7(c) and 7(d) plots the FFT spectrum of the white noise and 1/*f* noise, respectively. As shown in Figure 7, the 1/*f* noise has stronger complexity compared with white noise due to the long-range correlation properties. Meanwhile, white noise is more uncertain compared with 1/*f* noise. In other words, the entropy value of white noise is theoretically larger than that of 1/*f* noise.

Firstly, the MPGSE and MPSE are applied to deal with the white noise and 1/*f* noise over 20 scales and the acquired results are displayed in Figure 8. It can be found from Figure 8 that the entropy curve obtained using MPGSE is more stable and smoother than that obtained using MPSE. In addition, for MPGSE and MPSE, the entropy curve of 1/*f* noise is greater than that of white noise across most scales, which are consistent with the intuitive result of spectrum analysis. Preliminary results of the comparison show that MPGSE can give a stable estimation of complexity of nonlinear and nonstationary time series, which indicates MPGSE is suitable for detecting dynamic variation of the complex signal.

To further compare the performance of the MPGSE and MPSE, we apply the MPGSE and MPSE to analyze 100 independent white noise and 1/f noise, each of which contains 2000 data points. Figure 9(a) shows the error bar of white noise for two entropies (MPGSE and MPSE), while Figure 9(b) shows the error bar of 1/f noise for two entropies

(MPGSE and MPSE). As can be seen from Figures 9(a) and 9(b), the mean value curve of MPSE has bigger fluctuations than MPGSE. In addition, the standard deviation (SD) of MPGSE at each scale is smaller than that of MPSE, which implies MPGSE can offer a more accurate and stable estimation of entropy.

To investigate the impact of the SE scale λ on MPGSE, we calculate two entropies (MPGSE and MPSE) of two noise signals (white noise and 1/f noise) on different SE scales λ , and Figure 10 shows the obtained results. It is obvious in Figure 10 that the MPSE curve of two noise signals fluctuates significantly, whereas the MPGSE curve of two noise signals has a stabilization changing trend, which indicates that MPGSE overmatches the MPSE method in estimation accuracy of entropy. Besides, as the SE scale increases, the entropy value of two noise signals at the same scale factor gradually grows larger. According to multiscale morphological theory, the larger SE scale implies the more computing time. Consequently, to make a tradeoff between reliable estimation and computing efficiency, according to [48], the SE scale λ should not be set as too large or too small and its value is commonly recommended as 1 to $\lfloor f_s/f_g \rfloor - 2$, where $\lfloor \bullet \rfloor$ is the rounding operation and f_s and f_{q} are the sampling frequency and fault frequency, respectively.

To study the effect of data size of the signal on MPGSE, we apply MPGSE and MPSE, respectively, to analyze two noise signals (white noise and 1/f noise) containing five data sizes (i.e., N = 2000, 4000, 6000, 8000, and 10000). Figure 11 reveals the analyzed results of two noise signals under different data sizes. As shown in Figure 11, as the data size



FIGURE 7: Waveform and spectrum of the simulation signal. (a) Temporal waveform of white noise, (b) temporal waveform of 1/f noise, (c) FFT spectrum of white noise, and (d) FFT spectrum of 1/f noise.



FIGURE 8: Two entropies' analysis of white noise and 1/f noise.

increases, the results of MPGSE basically remain unchanged, which indicates MPGSE is barely affected by data size of the signal. Thus, in this case, the signal with a data size of 4000 points is usually sufficient for the implementation and calculation of MPGSE. However, in practice, it should be noted that we may need more data size for more accurate estimation of complexity of the signal. Considering that there are no specific criteria to accurately determine the data size of the signal, we intend to choose different data sizes to evaluate the complexity of signals according to the specific signals, which is seen as the focus of our future research. Besides, it can be observed in Figure 11 that the results of MPGSE have better stability than MPSE. Namely, MPGSE has more superior performance in complexity estimation of signal compared with MPSE.

To evaluate the computational efficiency of two entropies (MPGSE and MPSE), Tables 2 and 3 list the CPU running time of MPGSE and MPSE for two noise signals containing different data sizes, respectively. As shown in Tables 2 and 3, the CPU time of MPGSE is slightly smaller than that of MPSE. In addition, when using MPGSE and MPSE to detect the dynamic mutation of the signal, the bigger data length in Tables 2 and 3 implies the more CPU time and a smaller CPU time indicates the better feature extraction ability. Hence, according to the overall comparison result above, in this paper, MPGSE is utilized to extract multiscale fault features from the raw vibration data in priority.

4. The Proposed Fault Detection Scheme

4.1. Laplacian Score for Feature Selection. As is well known, a high-dimensional feature space will be constructed when MPGSE is applied to excavate the fault signatures across 20 scales. Nevertheless, the more features represent the more information redundancy and the lower computing efficiency. Laplacian score (LS) is an effective feature selection algorithm which can refine the sensitive feature information from the acquired multiscale features and improve fault classification accuracy. Hence, this paper selects LS for feature selection. Theories of LS are summarized as follows.



FIGURE 9: (a) MPGSE and MPSE curves of white noise and (b) MPGSE and MPSE curves of 1/f noise.



FIGURE 10: MPGSE and MPSE of the noise signal under different SE scales. (a) MPGSE of white noise, (b) MPSE of white noise, (c) MPGSE of 1/*f* noise, and (d) MPSE of 1/*f* noise.



FIGURE 11: MPGSE and MPSE of the noise signal under different data sizes. (a) MPGSE of white noise, (b) MPSE of white noise, (c) MPGSE of 1/*f* noise, and (d) MPSE of 1/*f* noise.

TABLE 2	2: CPU	running	time	of I	MPGSE	and	MPSE	for	white	noise.
INDEL 2		1 anning	unit	01 1	UL OOL	unu	THILOT	101	Willie C	110100.

Methods		The CPU time of	of two entropies under di	fferent data sizes	
	2000	4000	6000	8000	10000
MPGSE	0.4471 s	0.9166 s	1.3082 s	1.7549 s	2.2377 s
MPSE	0.4624 s	0.9368 s	1.3813 s	1.8932 s	2.4232 s

TABLE 3: CPU running ti	ne of MPGSE and	MPSE for $1/f$ noise.
-------------------------	-----------------	-----------------------

Methods		The CPU time of	of two entropies under di	fferent data sizes	
	2000	4000	6000	8000	10000
MPGSE	0.4376 s	0.8352 s	1.2991 s	1.7751 s	2.2995 s
MPSE	0.4716 s	0.8919 s	1.3792 s	1.9065 s	2.4881 s

Assume that *m* data samples are collected and each data sample contains *n* features. Suppose that L_r indicates the Laplace score of the *r*-th feature, where r = 1, 2, ..., n, and f_{ri} indicates the *r*-th feature of the *i*-th sample, where i = 1, 2, ..., m. Specific procedure of LS for feature selection is as follows:

- (1) Establish a nearest neighbor graph G with m nodes, where the *i*-th node is corresponding to x_i. If x_i and x_j are "close," such as x_i is among k-nearest neighbors of x_j or x_j is among k-nearest neighbors of x_i, there are edge joins; otherwise, there is no edge join. When the node label is known, one can connect a line between two nodes of the same label.
- (2) If nodes *i* and *j* are connected, weight matrix S_{ij} of the graph models is defined as

$$S_{ij} = e^{-\|x_i - x_j\|^2 / t},$$
(17)

where t represents a befitting constant. Otherwise, $S_{ii} = 0$.

(3) For the *r*-th feature, define f_r as

$$f_{r} = [f_{r1}, f_{r2}, \dots, f_{rm}]^{\mathrm{T}},$$

$$D = \operatorname{diag}(SI),$$

$$I = [1, 1, \dots, 1]^{\mathrm{T}},$$

$$L = D - S,$$

(18)

where the matrix L is known as the Laplacian matrix of the graph G. To avoid the influence of some dimensional data with significant difference on construction of the neighbor graph, each feature is averaged by

$$\tilde{f}_r = f_r - \frac{f_r^T DI}{I^T DI} I.$$
(19)

(4) Obtain Laplacian score of the *r*-th feature as

$$L_r = \frac{\sum_{ij} (f_{ri} - f_{rj})^2 S_{ij}}{\operatorname{Var}(f_r)} = \frac{\tilde{f}_r^T L \tilde{f}_r}{\tilde{f}_r^T D \tilde{f}_r},$$
 (20)

where $Var(f_r)$ represents the variance of the *r*-th feature.

According to a previous research study, the greater the S_{ij} value is, the smaller $(f_{ri} - f_{rj})^2$ value is and the smaller the Laplacian score L_r is, which indicates the smaller the difference between adjacent samples in this feature is, the stronger the local information preserving ability of this feature is. Besides, the greater $Var(f_r)$ value represents the smaller Laplacian score L_r , which means the larger the difference between different samples of this feature is. In concrete terms, a superior and important feature is equipped with a small LS value. Hence, for feature selection, we first calculate the LS value of each feature, then rank these LS values from low to high, and finally choose the first several features containing smaller LS values as the sensitive features to fault pattern identification.

4.2. ELM for Fault Pattern Identification. After using the LS algorithm for feature selection, an appropriate method needs

to be adopted to identify automatically fault categories and severities of bearing. ELM proposed by Huang et al. [52] is an effective pattern recognition method, which can avoid some shortcomings (e.g., the local minimum, improper learning rate, and overfitting) of the conventional feedforward neural network. Hence, this paper selects ELM for bearing fault pattern identification. Theories of ELM are summarized as follows.

Suppose that *N* different training samples $\{x_j, y_j\}$, j = 1, 2, ..., N are available, where $x_j = [x_{j1}, x_{j2}, ..., x_{jd}]^T \in \mathbb{R}^d$ is the network input vector and $t_j = [t_{j1}, t_{j2}, ..., t_{js}] \in \mathbb{R}^s$ denotes the target output vector. The mathematical model of the ELM model containing *L* single hidden layer nodes is expressed as

$$\sum_{i=1}^{L} \beta_i g(\omega_i \cdot x_j + b_i) = o_j, \quad (j = 1, 2, \dots, N, i = 1, 2, \dots, L),$$
(21)

where $g(\cdot)$ represents the sigmoid activation function, $\omega_i = [\omega_{i1}, \omega_{i2}, \ldots, \omega_{id}]^T$ indicates the weight vector between the input layer and the *i*-th hidden layer neuron, $\beta_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{is}]^T$ represents the connection weight of the *i*-th hidden node to the output layer, b_i denotes the bias of the *i*-th hidden node, and o_j implies the output of the ELM model of the *j*-th sample. Another form of equation (21) can be described as the following equality:

$$H\beta = T,$$
 (22)

where $\beta = [\beta_1, \beta_2, ..., \beta_L]^T$ represents the output weight matrix, $T = [t_1, t_2, ..., t_N]^T$ denotes the target matrix, and H indicates the hidden layer output matrix of network, which is given by

$$H = \begin{bmatrix} g(\omega_1 \cdot x_1 + b_1) & \cdots & g(\omega_L \cdot x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\omega_1 \cdot x_N + b_1) & \cdots & g(\omega_L \cdot x_N + b_L) \end{bmatrix}_{N \times L}$$
(23)

The purpose of training of the ELM model is to find the parameters β which can reduce the error between output matrix and target matrix to minimum. The output weight β is also expressed as

$$\beta = H^+ T, \tag{24}$$

where H^+ is the Moore–Penrose generalized inverse matrix of H. It can be seen that there is no need to select the input weight and bias of the hidden layer in ELM. Moreover, in ELM, the problem of local optimal solution will not appear. Hence, in this paper, it is feasible to adopt ELM for pattern recognition after feature extraction and selection.

4.3. The Proposed Fault Detection Scheme. To identify different health status of rolling bearing accurately, a novel fault diagnosis scheme based on MPGSE and LS is presented. The flowchart of our algorithm is illustrated in Figure 12, which mainly consists of three steps (i.e., multiscale fault feature extraction, feature selection, and fault classification). Overall procedures of our algorithm are given as follows:



FIGURE 12: Flowchart of the proposed algorithm.

Step 1: bearing vibration data are collected from a mechanical fault simulator by using a data collection unit

Step 2: the MPGSE algorithm is proposed to obtain multiscale fault signatures from the collected bearing vibration data. Concretely, in this step, to overcome the disadvantage of relying on experience to select the parameters of MPGSE, a novel intelligent optimization algorithm termed as grey wolf optimization (GWO) is firstly employed to determine self-adaptively two important parameters (i.e., SE scale λ and scale factor τ) of MPGSE. Subsequently, MPGSE with the optimized parameters is applied for multiscale feature extraction. Figure 13 shows the flowchart of parameter optimization of MPGSE. A specific optimization procedure is described as follows:

- Input data sample, initialize the population X_i (i = 1, 2, ..., m), and set the parameters of GWO. Concretely, define population size of wolves m = 30 and maximum iterations T = 10. Because the preoptimization parameters only involve two variables (λ, τ), each wolves are expressed as X_i = (x_λ, x_τ), i = 1, 2, ..., m, where x_λ and x_τ represent the SE scale and scale factor, respectively.
- (2) Calculate and compare the fitness value according to the misclassification rate between the number of misclassified samples and the amount of training samples, find the optimal position X_{best}^i of the individual grey wolf, and determine the global optimal position X_{best}^g of whole wolves.
- (3) Update the position of each wolf in terms of grey wolf movement pattern shown in the following equation:

$$X_{i}(t+1) = \frac{X_{i,\alpha}(t) + X_{i,\beta}(t) + X_{i,\delta}(t)}{3},$$
 (25)

$$\begin{cases} X_{i,\alpha}(t+1) = X_{\alpha}(t) - A \cdot |C \cdot X_{\alpha}(t) - X_{i}(t)|, \\ X_{i,\beta}(t+1) = X_{\beta}(t) - A \cdot |C \cdot X_{\beta}(t) - X_{i}(t)|, \\ X_{i,\delta}(t+1) = X_{\delta}(t) - A \cdot |C \cdot X_{\delta}(t) - X_{i}(t)|, \end{cases}$$
(26)

where *A* denotes the convergence factor and meets $A = 2d \cdot r_1 - d$, *C* denotes the swing factor and meets $C = 2 \cdot r_2$, *d* denotes the range control parameter which linearly decays from 2 to 0 over whole iteration, and r_1 and r_2 are the random numbers between 0 and 1.

(4) Determine whether the stopping condition is satisfied. Concretely, judge whether the current iterations are less than maximum iterations (i.e., t≤T) or whether the minimum error is small enough. If iteration conditions are satisfied, stop the iteration and output the best wolves X_{best} (i.e., the best parameters x_λ^{best} and x_τ^{best} of MPGSE). Otherwise, let t=t+1 and go back to Step 2 to continue to run until the iteration condition is satisfied.

Step 3: the LS method is applied to calculate the score of each feature and sort the obtained fault feature according to the size of Laplacian score of each feature, and then establish a new low-dimensional eigenvector on the basis of the first four sensitive features containing the lower Laplacian score.



FIGURE 13: Flowchart of parameter optimization of MPGSE.

Step 4: the obtained low-dimensional eigenvector is fed into the ELM for recognizing different fault categories and severities of rolling bearing, and the final diagnosis results and reports are given automatically.

5. Experimental Investigation

As everyone knows, each algorithm has advantages and disadvantages. In other words, each algorithm has certain limitations in the application scenario and sphere. Given this, in this subsection, two experimental cases are conducted, respectively, to show the effectiveness of our proposed algorithm in detecting different working conditions of rolling bearing. Meanwhile, the proposed algorithm is also compared with other entropy-based methods to reveal the benefits of our proposed algorithm.

5.1. Case 1: Benchmark Data from CWRU

5.1.1. Description of the Dataset. The presented algorithm is first used to process bearing vibration data collected from Case Western Reserve University (CWRU) [53]. Laboratory equipment and its sketch are displayed in Figure 14, which is mainly composed of a three-phase induction motor (left), a torque transducer (middle), and a load motor (right). During this experiment, four loads (0, 1, 2, and 3 hp) are added



FIGURE 14: The experimental device and its sketch.

separately to the test bearing. Single-point faults with four sizes (0.007, 0.014, 0.021, and 0.028 inches) were manufactured, respectively, on normal bearing by using electric discharge machining. Figure 15 is the description of data type of bearing under different health conditions. As shown in Figure 15, the inner race, outer race, and ball fault is abbreviated to "IR," "OR," and "B," respectively. The number after the abbreviation indicates the fault size, for example, "7" in IR7 denotes the damage size of 0.007 inches. The test bearing is located at the motor drive end. The accelerometer is mounted on the 12 o'clock position of the drive end of motor housing for collecting the vibration data. All data samples were obtained at a sampling frequency of 12 kHz. Table 4 lists the detailed specification of testing bearing.

In order to verify the recognition effect of our algorithm on bearing fault type and degree, three datasets (A, B, and C) under 1797 rpm are investigated, which are illustrated detailedly in Table 5. Concretely, in this case, we only investigated three kinds of bearing faults (i.e., bearing inner race, outer race, and ball fault). Each fault is manufactured on normal bearing by using electric discharge machining. The fault size of each fault is divided into three kinds (i.e., 0.007, 0.014, and 0.021 inches), where 1 inch roughly equals 25.4 mm. Besides, the bearing outer race fault is fixed, so the bearing outer race faults located at 6 o'clock (orthogonal to the load zone) are only analyzed in this experiment. A more specific description of the bearing fault can be found in [53]. For data collection, 200 samples for each dataset are obtained through a 2048-point nonoverlapping window. For each working state (i.e., normal, IR fault, OR fault, and ball fault), there are total 50 samples, where 25 data samples under each working status are selected randomly for training and the remainder 25 samples are taken as testing. Therefore, for each dataset, 100 training samples and 100 test samples are established. Taking A dataset as an example, Figure 16 shows the temporal waveforms and their corresponding frequency spectra of four running status. As shown in Figure 16, although fault type of bearing can be identified using time domain and frequency domain analysis, this approach is not automatic and has high requirements for professional knowledge. Thus, it is necessary to apply the intelligence identification method to process these vibration data.

5.1.2. Experimental Results and Analysis. Taking the A dataset as an example, the presented algorithm is firstly

applied to analyze the A dataset. In the first step, the GWO method is firstly employed to select automatically the parameters of MPGSE and MPSE as $\lambda = 4$ and $\tau = 20$. Meanwhile, MPGSE with the optimized parameters is used to extract the multiscale fault feature of the A dataset and obtain a high-dimensional feature matrix with a size of 200×20 . Figure 17 shows the PS and PGS curve of bearing vibration signal under different health conditions. It is obvious from Figure 17 that the PGS curve of four working conditions possesses higher distinguish degree than the PS curve. Besides, the PGS curve has better stability in feature extraction. Then, the LS algorithm is employed to sort the obtained fault feature from a low score to a high score, and the rank situations of the obtained features are shown in equation (27). According to equation (27), the first four features $\tau = 2, 1, 4$, and 3 with lower score are selected as important features to establish a new fault feature set. Figures 18(a) and 18(b), respectively, show the two-dimensional distribution of MPGSE of different health conditions before and after applying LS. Finally, the selected new feature set is fed into the ELM to identify different health conditions of bearings. Figure 19 gives the diagnosis results of our method for the A dataset. As can be seen in Figure 19, our method can achieve an accuracy rate of 100%, which validates the feasibility and effectiveness of our method in intelligent fault detection of bearings.

$$\begin{split} \mathrm{LS}_2 < \mathrm{LS}_1 < \mathrm{LS}_4 < \mathrm{LS}_3 < \mathrm{LS}_5 < \mathrm{LS}_{14} < \mathrm{LS}_{15} < \mathrm{LS}_{16} < \mathrm{LS}_{13} \\ < \mathrm{LS}_{12} < \mathrm{LS}_{17} < \mathrm{LS}_9 < \mathrm{LS}_8 < \mathrm{LS}_{11} < \mathrm{LS}_{10} < \mathrm{LS}_7 \\ < \mathrm{LS}_6 < \mathrm{LS}_{18} < \mathrm{LS}_{19} < \mathrm{LS}_{20}. \end{split}$$

To illustrate the superiority of our proposed approach, the comparisons among three methods (MPGSE and LS; MPSE and LS; and MSE and LS) are performed by analyzing the same experimental data. For each algorithm, 20 trials are conducted to avoid the randomness of diagnostic results. For the sake of fairness of comparison, the important parameters of MSE are also determined by the GWO method, which are, respectively, determined as m = 4, $\tau = 20$, and $r = 0.15\sigma$, where σ is the standard deviation of the original signal. Figure 20 plots the identification results obtained by different methods, and the detailed diagnosis results are listed in Table 6, including the maximum, minimum, mean, and



FIGURE 15: Detailed description of bearing fault states.

TABLE 4: The specification of test bearing.

Bearing type	Roller diameter (mm)	Pitch diameter (mm)	Number of the roller	Contact angle (°)
SKF6205-2RS	7.94 mm (0.3126 inches)	39.04 mm (1.537 inches)	9	0

TABLE 5: Description of three datasets.

Dataset	Working state	Damage size (inches)	Number of training samples	Number of testing samples	Category labels
	Normal	0	25	25	1
٨	IR fault	0.007	25	25	2
A	OR fault	0.007	25	25	3
	Ball fault	0.007	25	25	4
	Normal	0	25	25	1
D	IR fault	0.014	25	25	2
D	OR fault	0.014	25	25	3
	Ball fault	0.014	25	25	4
	Normal	0	25	25	1
C	IR fault	0.021	25	25	2
C	OR fault	0.021	25	25	3
	Ball fault	0.021	25	25	4



FIGURE 16: Temporal waveform and FFT spectrum of bearing vibration data under different health conditions.



FIGURE 17: Curves of different bearing vibration signal in case 1. (a) PS and (b) PGS.



FIGURE 18: (a) The original distribution of MPGSE and (b) the new distribution of MPGSE after applying the LS method in case 1.



FIGURE 19: Classification results of the first trial of the proposed approach in case 1.



FIGURE 20: Diagnostic accuracy of 20 trials of three methods in case 1.

TABLE 6: Diagnostic results of different methods with LS feature selection for the A dataset in case 1.

Different methods	Diagnost	ic accuracy meth	obtaine ods (%	d using)	different
	Maximum	Minimum	Mean	SD	CPU time (s)
MPGSE and LS	100	100	100	0	90.04
MPSE and LS	98	95	96.75	0.7164	97.77
MSE and LS	98	90	95.50	1.9868	132.78

standard deviation (SD) of accuracies and the average CPU time. As can be seen, our method has the highest recognition accuracy (100%). Besides, the proposed approach has the lowest SD and CPU time, which indicates our method has better stability and higher computing efficiency. Experimental results show that our method is effective in identifying different fault patterns of bearing.

To validate the effectiveness of combining MPGSE and LS, Figures 21(a)–21(c) show plotting of the first two most important features obtained by the three methods (MPGSE and LS; MPSE and LS; and MSE and LS), respectively. As can be seen, the sensitive features selected by the LS method are discriminated clearly. Besides, feature aggregation and differentiation of the proposed algorithm are superior to that of other two methods, which means that the LS algorithm can select fault features with higher discrimination. As a comparison, two randomly selected features in the three methods (MPGSE, MPSE, and MSE) are shown in Figures 22(a)–22(c), respectively. From Figure 22, we can find that distinguish ability of the features selected randomly is not good enough, which indicates that it is essential to apply LS to refine the sensitive features.

To show the efficacy of applying the LS method, 20 trials are conducted in the three methods (MPGSE, MPSE, and MSE) without LS-based feature selection. Specifically, four random features ($\tau = 1, 3, 7, \text{ and } 9$) in a multiscale feature

set are directly inputted to an ELM classifier to identify different fault patterns of bearing. Table 7 shows the classification results of the three methods (MPGSE, MPSE, and MSE). As can be seen in Table 7, the average accuracy of MPGSE without using LS is 99%, which is slightly lower than the average accuracy obtained using MPGSE and LS, which indicates the advantage of applying the LS method is not obvious. Here, we give a specific explanation for this issue. Concretely, in case 1, the difference between benchmark bearing data under different fault patterns is relatively obvious, which indicates the complexity of bearing vibration data under different fault patterns is diverse from each other (i.e., MPGSE features of different bearing vibration data have a great difference), so it becomes easier to distinguish. In other words, for the benchmark bearing data, the inner-class distance in the sample of the same category is small enough and the between-class distance in the sample of the different category is large enough, so the average accuracy obtained by MPGSE with and without LS has a small difference. In addition, it is important to note that benchmark bearing vibration data in case 1 suffer from less noise interference compared with the actual vibration data and MPGSE of benchmark data can be identified effectively whether we use LS or not. This further verifies the powerful feature extraction ability of MPGSE. Nonetheless, from another point of view, the average accuracy (96.75%) of combining MPSE and LS is apparently higher than that of the average accuracy (73.55%) of using only MPSE, which indicates LS-based feature selection is still a necessary step for improving diagnostic accuracy. Moreover, the average accuracy rate of MPGSE is still higher than that of MPSE and MSE, which further verify the advantage of the proposed method in fault recognition.

To show the influence of feature dimension on the diagnosis result, Figure 23 plots the relation curve between accuracy and feature dimension selected by LS. As shown in Figure 23, as the feature dimension increases, accuracy of different methods has an upward tendency. However, the greater feature dimension implies the more computing time. Thus, to make a tradeoff between diagnosis accuracy and computational efficiency, this paper selects the first four most important features for fault identification, which meets the engineering requirements.

Our approach is adopted to further analyze the B and C dataset, and 20 trials are also performed. The detailed diagnosis results of the three methods (MPGSE and LS; MPSE and LS; and MSE and LS) are given in Table 8. As can be seen in Table 8, mean accuracies of our approach reach 94.70% and 100% for the B and C dataset, respectively, which is higher than that of other two methods. This indicates that our approach can effectively recognize different fault patterns of bearing.

Similarly, for the B and C dataset, classification performance of three methods (MPGSE, MPSE, and MSE) is investigated when LS is not used for feature selection. Table 9 lists the detailed diagnostic results. As shown in Table 9, whether B or C dataset, the average accuracy obtained without using LS is less than that of using LS. That is, the LS method is effective in improving fault identification rate.



FIGURE 21: The plotting of the best two features selected by LS in case 1. (a) MPGSE and LS, (b) MPSE and LS, and (c) MSE and LS.

5.2. Case 2: Experimental Data from Laboratory

5.2.1. Description of the Dataset. The presented method is utilized to deal with bearing vibration data from Research Center of Condition Monitoring and Fault Diagnosis (RCCMFD), Southeast University. Figure 24 displays global picture of the experimental bench, which primarily consists of loading equipment, bearing test module, driving system, electrical control system, and computer monitoring equipment. In this experiment, four kinds of faults are separately manufactured on normal bearing by using spark machining, including outer race fault (ORF), inner race fault (IRF), outer race-ball compound fault (ORBF), and outer-inner race compound fault (OIRF). Specifically, whether bearing inner race or outer race, their single local fault size is 0.5 mm in depth and 0.1 mm in width. Besides, a scratch fault is machined on the ball surface of bearing. Figure 25 shows bearing with different defects. The specific size of the testing bearing is displayed in Table 10. A PCB accelerometer with a sensitivity of 100 mV/g was installed near the testing bearing block to gather the bearing fault signal. For each fault pattern, motor speed is stable at 1050 rpm, and bearing fault data are sampled with a sampling frequency of 10240 Hz.

In this experiment, for every failure state, 50 samples with the data length of 2048 points are selected, which means that a total of 200 data samples are generated. More specifically, in the process of verification of the algorithm, 25 samples for every failure state are randomly chosen as the training data and the remainder 25 samples are regarded as the testing data. Apparently, 100 training samples and 100 testing samples can be acquired to show the effectiveness of our algorithm, respectively. It is essentially a four-classification problem to be addressed. Figure 26 plots the timedomain graph of different fault signals and their corresponding FFT spectrum. As illustrated in Figure 26, because the collected bearing vibration signal contains heavy background noise and other external interference and the raw bearing vibration data under different fault patterns have some similarity in their waveform and spectrum, the bearing fault pattern cannot be exactly recognized directly through the waveform and the spectrum, which implies that



FIGURE 22: The plotting of the two randomly selected features in case 1. (a) MPGSE, (b) MPSE, and (c) MSE.

TABLE 7: Diagnostic results of different methods without LS feature selection for the A dataset in case 1.

Different methods	Diagnostic accuracy obtained using different methods (%)				
	Maximum	Minimum	Mean	SD	
MPGSE	100	96.00	99	1.026	
MPSE	76.00	70.00	73.55	1.394	
MSE	91.00	85.00	88.55	1.905	

an appropriate technology should be introduced to complete the efficient recognition of each fault pattern.

5.2.2. Experimental Results and Analysis. The proposed approach is utilized to analyze the abovementioned fault data. Firstly, we apply GWO-based parameter-optimized MPGSE to extract multiscale fault features over different scales. Note that the optimal parameters of MPGSE are selected as $\lambda = 4$ and $\tau = 20$, respectively. Figures 27(a) and

27(b) show the PS and PGS curves of bearing fault data under four states, respectively. As can be seen in Figure 27, discrimination degree of PGS under different fault status is better than that of PS. Meanwhile, as the SE scale increases, the PGS value of fault data tends to stabilization, whereas the PS value of fault data fluctuates largely, which has a negative influence on fault classification and complexity estimation. Secondly, the LS algorithm is employed to rank the extracted multiscale features according to their importance and sensitivity. The new order of multiscale features is shown in equation (28). According to equation (28), the first four features ($\tau = 3, 8, 1, \text{ and } 7$) containing the richest fault information are selected to build a new feature set with the size of 200×4 . Figures 28(a) and 28(b) display the distribution of multiscale features after and before applying LS, respectively. Finally, the new feature set is regarded as the input of ELM to recognize fault categories of bearings. Figure 29 plots the diagnostic results of our method. As can be seen in Figure 29, our method can obtain an accuracy rate of 100%, which verifies the validity of our method.



FIGURE 23: Diagnostic accuracy of three methods under different feature dimension in case 1.

TABLE 8: Diagnostic results of different methods with LS feature selection for the B dataset and C dataset in case 1.

		MPGSE a	and LS		MPSE a	nd LS		MSE ar	nd LS
Dataset	Accur	acy (%)	CDU times (a)	Accur	acy (%)	CDU time (a)	Accur	acy (%)	CDU time (a)
Max	Max	Mean	CPU time (s)	Max	Mean	CPU time (s)	Max	Mean	CPU time (s)
В	98	94.70	84.88	95	91.75	88.49	92	89.20	128.28
С	100	100	87.04	100	98.95	90.45	100	98.65	127.54

TABLE 9: Diagnostic results of different methods without LS feature selection for the B dataset and C dataset in case 1.

		MPGSE			MPSE			MSE	
Dataset	set Diagnostic accuracy (%) Diagnostic accuracy (%)			Diagnostic accuracy (%)					
	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean
В	93	92.25	92.95	77	71	73.85	75	67	70.35
С	97	94	96.40	95	89	92.15	97	90	94.00



FIGURE 24: Structure of the experimental platform.

$$\begin{split} \mathrm{LS}_{3} &< \mathrm{LS}_{8} < \mathrm{LS}_{1} < \mathrm{LS}_{7} < \mathrm{LS}_{2} < \mathrm{LS}_{9} < \mathrm{LS}_{4} < \mathrm{LS}_{5} < \mathrm{LS}_{6} \\ &< \mathrm{LS}_{20} < \mathrm{LS}_{12} < \mathrm{LS}_{18} < \mathrm{LS}_{11} < \mathrm{LS}_{19} < \mathrm{LS}_{15} < \mathrm{LS}_{10} \\ &< \mathrm{LS}_{16} < \mathrm{LS}_{17} < \mathrm{LS}_{14} < \mathrm{LS}_{13}. \end{split}$$

To highlight the advantages of the proposed approach, the same fault data are analyzed on two comparison methods (MPSE and MSE). Note that the important parameters of MPSE and MSE are also determined by the GWO method, which are optimized as m = 4 and $\tau = 20$. To reduce the

contingency of the diagnosis results, 20 trials are conducted on each method. The identification results obtained using different methods are shown in Figure 30, and the detailed results are given in Table 11. As you can see, the proposed method (MPGSE and LS) achieves the highest average accuracy (100%), the second method (MPSE and LS) has a diagnostic accuracy of 95% to 98%, and the third method (MSE and LS) has a diagnostic accuracy of 90% to 98%. In addition, SD of recognition results of the proposed approach is lower than those in the second and third methods, which means that the recognition rate of MPGSE has better stability than that of MPSE and MSE. Meanwhile, The CPU running time of the proposed MPGSE is lesser than that of MPSE and MSE. Hence, superiority of the presented MPGSE method in fault classification is demonstrated by the comparative analysis.

To show the necessity of integrating the LS method, Figures 31(a)-31(c) plot the first two elements of the new feature obtained by three methods (i.e., MPGSE and LS; MPSE and LS; and MSE and LS), respectively. For comparison, in the three methods (MPGSE and LS; MPSE and LS; and MSE and LS), two randomly selected features



FIGURE 25: Three fault conditions of rolling bearing. (a) ORF, (b) IRF, and (c) OIRF.



TABLE 10: The parameters of rolling bearing.

FIGURE 26: Temporal waveform and FFT spectrum of different bearing fault signals.

without using LS are shown in Figures 32(a)-32(c), respectively. As can be seen in Figure 31(a), features selected by LS have a nice cluster and recognition degree, whereas the randomly selected features in Figure 32(a) cannot be recognized clearly. This implies that the combination of MPGSE and LS is helpful in fault classification. Besides, by comparing Figures 31(b) and 32(b), we can find that features obtained by MPSE before and after applying LS have no good differentiation. By comparing Figures 31(c) and 32(c),

it can also be found that features obtained by MSE before and after applying LS are also not good enough.

To further show the effectiveness of using the LS approach, we randomly select four features ($\tau = 1, 7, 9, 11$) as the input of the ELM classifier to identify different bearing fault patterns. For each approach, 20 trials are also carried out to avoid the randomness of their classification results. Detailed diagnosis results of different approaches are given in Table 12. It can be observed that our method can achieve



FIGURE 27: Curves of different bearing fault signals in case 2: (a) PS and (b) PGS.



FIGURE 28: (a) The original distribution of MPGSE and (b) the new distribution of MPGSE after applying the LS method in case 2.



FIGURE 29: Classification results of the first trial of the proposed approach in case 1.



FIGURE 30: Diagnostic accuracy of 20 trials of three methods in case 2.

TABLE 11: Diagnostic results of different methods with LS feature selection in case 2.

D'fferrent method		Diagnostic accuracy	y obtained using dif	ferent methods (%)	
Different methods	Maximum	Minimum	Mean	SD	CPU time (s)
MPGSE and LS	100	100	100	0	106.272
MPSE and LS	92	81	88.15	2.5603	107.926
MSE and LS	98	84	91.45	3.8041	147.881

an average accuracy of 90.30%, which is higher than that of MPSE and MSE. In other words, MPGSE has better performance in feature extraction than MPSE and MSE. However, it is very obvious that the average accuracy of Table 12 is lower than those of Table 11, which validates the efficacy of LS in feature selection.

5.3. Results and Discussion. According to the experimental analysis results of two datasets above, the proposed fault diagnosis scheme is demonstrated to be effective for bearing fault identification. Concretely, the proposed method (i.e., MPGSE and LS) can achieve a classification accuracy of 99% and above at the same time to ensure that its computational efficiency is higher than that before improvement. In addition, various combination comparisons are performed to verify the superiority of the proposed diagnosis algorithm. Concretely, several metrics (i.e., maximum, minimum, mean, and SD of classification accuracy) and CPU running time are utilized to compare the diagnosis performance of different methods, which prove that the classification accuracy of the proposed approach is larger than that of other comparison methods (i.e., MPSE and LS; MSE and LS). Meanwhile, the CPU running time of the proposed approach is smaller compared with other comparison methods (MPSE and LS; MSE and LS). Nevertheless, some challenges remain when the proposed diagnosis approach is applied to address the problems of the identification of different health conditions.

 Although the proposed MPGSE approach can overcome the shortcomings of the undefined entropy value existing in traditional multiscale entropy and has superior feature extraction performance for rolling element bearing, the selection of two parameters (i.e., SE scale λ and scale factor τ) of the proposed MPGSE is still empirical. Hence, except for GWO used in this paper, for the future work, we can also adopt other swarm intelligent optimizers (e.g., genetic algorithm (GA), particle swarm optimization (PSO), cuckoo search algorithm (CSA), firefly algorithm (FA), fruit fly optimization algorithm (FOA), and whale optimization algorithm (WOA)) to adaptively optimize the parameters of MPGSE.

- (2) The core idea of this paper is that the presented MPGSE and LS are combined to achieve intelligent fault diagnosis of rolling element bearing. That is, except for MPGSE-based feature extraction, feature selection is also very important in the proposed method. Hence, in the future research, it is necessary to introduce other effective feature extraction techniques to analyze and compare the experimental data, such as hybrid feature selection scheme [54], local and global principal component analysis (LGPCA), minimum redundancy maximum relevance (mRMR) [55], partial maximum correlation information (PMCI) [56], and multicluster feature selection (MCFS) [57]. The research and comparison of these methods are the focus of the follow-up research.
- (3) Another important point worth mentioning is that the proposed fault diagnosis scheme increases running time due to the fusion of three stages (MPGSE-



FIGURE 31: The plotting of the best two features selected by LS in case 2. (a) MPGSE and LS, (b) MPSE and LS, and (c) MSE and LS.



FIGURE 32: Continued.



FIGURE 32: The plotting of the two randomly selected features in case 2. (a) MPGSE, (b) MPSE, and (c) MSE.

Table	12:	Diagnostic	results	of	different	methods	without	LS
feature	sele	ection in cas	e 2.					

Different methods	Diagnostic accuracy obtained using different methods (%)					
	Maximum	Minimum	Mean	SD		
MPGSE	99	83	90.30	4.2932		
MPSE	84	60	74.00	4.9736		
MSE	77	60	71.05	4.3465		

based feature extraction, LS-based feature selection, and ELM-based fault classification). Running time of the algorithm is not a big issue, thanks to the development of computer science; in order to make our approach quickly applicable to the on-line condition monitoring and diagnosis of machinery, the improvement of computational efficiency of our approach is regarded as our future research direction.

6. Conclusions

In this paper, a novel dynamical indicator named PGSE is proposed for evaluating complexity and uncertainty of the time series. To extract different fault signatures over multiple scales, the coarse-grained procedure and PGSE are combined to design a new algorithm called MPGSE, where its key parameters are selected by GWO. Secondly, LS approach is utilized to select the sensitive features and establish a new feature set, which can remove the redundant or irrelevant feature information and improve computational efficiency. Ultimately, the acquired new feature dataset is entered into the ELM classifier to identify automatically different health conditions of rolling bearing. According to the experimental analysis results from two examples above, our designed algorithm is proved to be effective in identifying different fault categories and severities of rolling element bearing. Innovations and main contributions of this article are summarized as follows:

- A modified method called PGSE is formulated, which can improve detection of dynamic change of time series
- (2) A neoteric algorithm termed as MPGSE is developed to extract abundant fault feature over multiple scale.
- (3) The LS method is employed to select several sensitive characteristics with a higher discrimination
- (4) The advantages of our algorithm in fault classification are validated by the application of two experimental cases

The abovementioned results show that our method is satisfactory and promising in intelligent fault detection of rolling bearing. However, performance of our method is unknown for health status recognition of bearing under variable speed. It will be very valuable to apply MPGSE to diagnose different fault patterns of rolling bearing when rolling bearing is running at variable speed. This research point will be carried out in the future work.

Nomenclature

SNR: Signal-to-noise ratio FFT: Fast Fourier transform MM: Mathematical morphology PS: Pattern spectrum SE: Structuring element PSE: Pattern spectrum entropy SVM: Support-vector machine EEMD: Ensemble empirical mode decomposition IMF: Intrinsic mode function MGO: Morphological gradient operation PGSE: Pattern gradient spectrum entropy MSE: Multiscale entropy LMD: Local mean decomposition

PF:	Product function
EMD:	Empirical mode decomposition
MPSE:	Multiscale pattern spectrum entropy
MPGSE:	Multiscale pattern gradient spectrum entropy
LS:	Laplacian score
ANN:	Artificial neural network
ELM:	Extreme learning machine
GWO:	Grey wolf optimization.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare there are no conflicts of interests regarding the publication of this paper.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China under Grant 51675098, Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant KYCX17_0059, Jiangsu Provincial Key Research and Development Program under Grant BE2019030637, and Jiangsu Agricultural Science and Technology Independent Innovation Fund under Grant 2018325-SCX(18)2049. The authors would like to appreciate CWRU for their kind permission to use their bearing data.

Supplementary Materials

MATLAB code of the proposed MPGSE is published in the supplementary material. (*Supplementary Materials*)

References

- X. Yan, Y. Liu, and M. Jia, "Research on an enhanced scale morphological-hat product filtering in incipient fault detection of rolling element bearings," *Measurement*, vol. 147, Article ID 106856, 2019.
- [2] Z. Wang, J. Zhou, J. Wang et al., "A novel fault diagnosis method of gearbox based on maximum kurtosis spectral entropy deconvolution," *IEEE Access*, vol. 7, no. 1, pp. 29520–29532, 2019.
- [3] X. Yan, Y. Liu, M. Jia, and Y. Zhu, "A multi-stage hybrid fault diagnosis approach for rolling element bearing under various working conditions," *IEEE Access*, vol. 7, pp. 138426–138441, 2019.
- [4] S. Yan, S. K. Nguang, and L. Zhang, "Nonfragile integralbased event-triggered control of uncertain cyber-physical systems under cyber-attacks," *Complexity*, vol. 2019, 2019.
- [5] Y. Huang, R. Lu, and K. Chen, "Detection of internal defect of apples by a multichannel Vis/NIR spectroscopic system," *Postharvest Biology and Technology*, vol. 161, Article ID 111065, 2020.
- [6] X. Yan, Y. Liu, and M. Jia, "A feature selection frameworkbased multiscale morphological analysis algorithm for fault diagnosis of rolling element bearing," *IEEE Access*, vol. 7, pp. 123436–123452, 2019.

- [7] A. Hu and L. Xiang, "Selection principle of mathematical morphological operators in vibration signal processing," *Journal of Vibration and Control*, vol. 22, no. 14, pp. 3157– 3168, 2016.
- [8] X. Bai and F. Zhou, "Analysis of different modified top-hat transformations based on structuring element construction," *Signal Processing*, vol. 90, no. 11, pp. 2999–3003, 2010.
- [9] Y. Dong, M. Liao, X. Zhang, and F. Wang, "Faults diagnosis of rolling element bearings based on modified morphological method," *Mechanical Systems and Signal Processing*, vol. 25, no. 4, pp. 1276–1286, 2011.
- [10] C. Li and M. Liang, "Continuous-scale mathematical morphology-based optimal scale band demodulation of impulsive feature for bearing defect diagnosis," *Journal of Sound and Vibration*, vol. 331, no. 26, pp. 5864–5879, 2012.
- [11] C. Shen, Q. He, F. Kong, and P. W. Tse, "A fast and adaptive varying-scale morphological analysis method for rolling element bearing fault diagnosis," *Proceedings of the Institution* of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, vol. 227, no. 6, pp. 1362–1370, 2013.
- [12] Y. Li, X. Liang, and M. J. Zuo, "Diagonal slice spectrum assisted optimal scale morphological filter for rolling element bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 85, pp. 146–161, 2017.
- [13] J. Lv and J. Yu, "Average combination difference morphological filters for fault feature extraction of bearing," *Mechanical Systems* and Signal Processing, vol. 100, pp. 827–845, 2018.
- [14] H. Li and D.-y. Xiao, "Fault diagnosis using pattern classification based on one-dimensional adaptive rank-order morphological filter," *Journal of Process Control*, vol. 22, no. 2, pp. 436–449, 2012.
- [15] B. Li, P.-l. Zhang, Z.-j. Wang, S.-s. Mi, and D.-s. Liu, "A weighted multi-scale morphological gradient filter for rolling element bearing fault detection," *ISA Transactions*, vol. 50, no. 4, pp. 599–608, 2011.
- [16] A. S. Raj and N. Murali, "Early classification of bearing faults using morphological operators and fuzzy inference," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 2, pp. 567–574, 2013.
- [17] X. Yu, F. Chu, and R. Hao, "Fault diagnosis approach for rolling bearing based on support vector machine and soft morphological filters," *Journal of Mechanical Engineering*, vol. 45, no. 7, pp. 75–80, 2009.
- [18] S. Dong, L. Chen, B. Tang, X. Xu, Z. Gao, and J. Liu, "Rotating machine fault diagnosis based on optimal morphological filter and local tangent space alignment," *Shock and Vibration*, vol. 2015, pp. 1–9, 2015.
- [19] C. Wen and C. Zhou, "Morphological filter based on grey relational degree and its application in rolling bearing fault diagnosis," *Journal of Vibration and Shock*, vol. 34, no. 14, pp. 51–55, 2015.
- [20] Z. Zheng, W. Jiang, Z. Wang, Y. Zhu, and K. Yang, "Gear fault diagnosis method based on local mean decomposition and generalized morphological fractal dimensions," *Mechanism and Machine Theory*, vol. 91, pp. 151–167, 2015.
- [21] H. Li, Y. Wang, B. Wang, J. Sun, and Y. Li, "The application of a general mathematical morphological particle as a novel indicator for the performance degradation assessment of a bearing," *Mechanical Systems and Signal Processing*, vol. 82, pp. 490–502, 2017.
- [22] E. R. Dougherty and Y. Cheng, "Morphological patternspectrum classification of noisy shapes: exterior granulometries," *Pattern Recognition*, vol. 28, no. 1, pp. 81–98, 1995.

- [23] P. Maragos, "Pattern spectrum and multiscale shape representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 701–716, 1989.
- [24] R. Hao, Z. Peng, Z. Feng, and F. Chu, "Application of support vector machine based on pattern spectrum entropy in fault diagnostics of rolling element bearings," *Measurement Science* and Technology, vol. 22, no. 4, p. 045708, 2011.
- [25] K. Zhang, J. Cheng, and Y. Yang, "Rotating machinery fault diagnosis based on local mean decomposition and pattern spectrum," *Journal of Vibration and Shock*, vol. 32, pp. 135– 140, 2013.
- [26] Z. Zheng, W. Jiang, H. Hu et al., "Research on rolling bearings fault diagnosis method based on EEMD morphological spectrum and kernel fuzzy C-means clustering," *Journal of Vibration Engineering*, vol. 28, no. 2, pp. 324–330, 2015.
- [27] X. Yan and M. Jia, "Application of CSA-VMD and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 122, pp. 56–86, 2019.
- [28] B. Li, P.-l. Zhang, Z.-j. Wang, S.-s. Mi, and Y.-t. Zhang, "Gear fault detection using multi-scale morphological filters," *Measurement*, vol. 44, no. 10, pp. 2078–2089, 2011.
- [29] M. H. Khakipour, A. A. Safavi, and P. Setoodeh, "Bearing fault diagnosis with morphological gradient wavelet," *Journal of the Franklin Institute*, vol. 354, no. 6, pp. 2465–2476, 2017.
- [30] M. Costa, A. Goldberger, and C. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, Article ID 068102, 2002.
- [31] M. Hu and H. Liang, "Adaptive multiscale entropy analysis of multivariate neural data," *IEEE Transactions on Bio-Medical Engineering*, vol. 59, no. 59, pp. 12–15, 2012.
- [32] R. Mahajan and B. I. Morshed, "Unsupervised eye blink artifact denoising of EEG data with modified multiscale sample entropy, kurtosis, and Wavelet-ICA," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 1, pp. 158–165, 2015.
- [33] 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.
- [34] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. Guo, "Applying improved multi-scale entropy and support vector machines for bearing health condition identification," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 224, no. 6, pp. 1315–1325, 2010.
- [35] H. Liu and M. Han, "A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings," *Mechanism and Machine Theory*, vol. 75, pp. 67–78, 2014.
- [36] N.-K. Hsieh, W.-Y. Lin, and H.-T. Young, "High-speed spindle fault diagnosis with the empirical mode decomposition and multiscale entropy method," *Entropy*, vol. 17, no. 4, pp. 2170–2183, 2015.
- [37] Y. Li, G. Li, Y. Yang, X. Liang, and M. Xu, "A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy," *Mechanical Systems and Signal Processing*, vol. 105, pp. 319–337, 2018.
- [38] X. Yan, Y. Liu, and M. Jia, "Health condition identification for rolling bearing using a multi-domain indicator-based optimized stacked denoising autoencoder," *Structural Health Monitoring*, Article ID 1475921719893594, 2019.
- [39] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in Advances in Neural Information Processing Systems, vol. 18, Vancouver, Canada, 2005.
- [40] K. Zhang, Y. Li, P. Scarf, and A. Ball, "Feature selection for high-dimensional machinery fault diagnosis data using

multiple models and radial basis function networks," *Neurocomputing*, vol. 74, no. 17, pp. 2941–2952, 2011.

- [41] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018.
- [42] V. Havlíček, A. Córcoles, K. Temme et al., "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, 2019.
- [43] J. Ali, B. Chebel-Morello, L. Saidi et al., "Accurate bearing remaining useful life prediction based on weibull distribution and artificial neural network," *Mechanical Systems and Signal Processing*, vol. 56, pp. 150–172, 2015.
- [44] R. Liu, B. Yang, X. Zhang, S. Wang, and X. Chen, "Timefrequency atoms-driven support vector machine method for bearings incipient fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 75, pp. 345–370, 2016.
- [45] M. Y. Asr, M. M. Ettefagh, R. Hassannejad, and S. N. Razavi, "Diagnosis of combined faults in rotary machinery by nonnaive Bayesian approach," *Mechanical Systems and Signal Processing*, vol. 85, pp. 56–70, 2017.
- [46] Z.-X. Yang, X.-B. Wang, and J.-H. Zhong, "Representational learning for fault diagnosis of wind turbine equipment: a multi-layered extreme learning machines approach," *Energies*, vol. 9, no. 6, p. 379, 2016.
- [47] Y. Li, M. J. Zuo, Y. Chen, and K. Feng, "An enhanced morphology gradient product filter for bearing fault detection," *Mechanical Systems and Signal Processing*, vol. 109, pp. 166–184, 2018.
- [48] X. Yan, M. Jia, W. Zhang, and L. Zhu, "Fault diagnosis of rolling element bearing using a new optimal scale morphology analysis method," *ISA Transactions*, vol. 73, pp. 165–180, 2018.
- [49] Y. Li, X. Liang, and M. J. Zuo, "A new strategy of using a timevarying structure element for mathematical morphological filtering," *Measurement*, vol. 106, pp. 53–65, 2017.
- [50] H. Azami, M. Rostaghi, D. Abasolo, and J. Escudero, "Refined composite multiscale dispersion entropy and its application to biomedical signals," *IEEE Transactions on Bio-Medical En*gineering, vol. 64, no. 64, pp. 2872–2879, 2017.
- [51] J. Zheng, Z. Jiang, and H. Pan, "Sigmoid-based refined composite multiscale fuzzy entropy and t-SNE based fault diagnosis approach for rolling bearing," *Measurement*, vol. 129, pp. 332–342, 2018.
- [52] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.
- [53] http://csegroups.case.edu/bearingdatacenter/home.
- [54] Y. Yang, Y. Liao, G. Meng, and J. Lee, "A hybrid feature selection scheme for unsupervised learning and its application in bearing fault diagnosis," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11311–11320, 2011.
- [55] X. Yan and M. Jia, "Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection," *Knowledge-Based Systems*, vol. 163, pp. 450–471, 2019.
- [56] M. Yuan, Z. Yang, and G. Ji, "Partial maximum correlation information: a new feature selection method for microarray data classification," *Neurocomputing*, vol. 323, pp. 231–243, 2019.
- [57] C. Li, J. Zheng, H. Pan, J. Tong, and Y. Zhang, "Refined composite multivariate multiscale dispersion entropy and its application to fault diagnosis of rolling bearing," *IEEE Access*, vol. 7, pp. 47663–47673, 2019.