Intelligent Fault Detection for Rotating Machinery Using Cyclic Morphological Modulation Spectrum and Hierarchical Teager Permutation Entropy

Junchao Guo, Qingbo He, Senior Member, IEEE, Dong Zhen, and Fengshou Gu

Abstract—Intelligent fault detection of rotating machines is essentially a pattern classification issue. At the same time, effectively obtaining fault features from the measured signals is a key step to timely diagnose the health status of rotating machinery and evaluate the results of fault classification. To accurately obtain effective fault information to enhance fault accuracy, this paper proposes a novel fault detection scheme based on cyclic morphological modulation spectrum (CMMS) and hierarchical Teager permutation entropy (HTPE). In this scheme, firstly, CMMS was developed to analyze the measured signal to obtain a series of CMMS slices with different frequency bands, which solved the deficiencies of manual empirical selection of frequency band bandwidth in the traditional cyclic modulation spectrum (CMS). Subsequently, by integrating Teager energy operator into hierarchical permutation entropy (HPE), an improved feature selection method named hierarchical Teager permutation entropy (HTPE) is presented to obtain fault information of different frequency band slices, which can improve the fault feature extraction capability of HPE. Finally, the acquired HTPE-based vectors are integrated into the extreme learning machine (ELM) classifier to achieve fault classification of rotating machinery under different working conditions. The proposed scheme is validated by experimental cases including cylindrical roller bearings and planetary gearboxes. The analysis results indicate that the proposed scheme not only can effectively obtain the fault features, but also accurately realize the classification and recognition of the fault mode. In addition, the proposed scheme can achieve higher detection accuracy than some existing algorithms.

Index Terms—Cyclic morphological modulation spectrum; Hierarchical Teager permutation entropy; Rotating machinery; Fault detection.

I. INTRODUCTION

Fault diagnosis of rotating machinery is an essential topic in prognostics and health management systems. Unexpected failure of rotating machinery can cause damage to mechanical equipment, resulting in huge economic losses and even casualties [1], [2]. In addition, since the rotating machinery has been working in harsh environments, the measurement signals obtained by sensors are easily affected by the complex transmission path and strong background noise, which also greatly increases the difficulty of rotating machines’ fault diagnosis [3], [4]. Therefore, timely and accurate diagnosis of rotating machines faults is of great significance to prevent serious equipment damage and downtime. Currently, a few signal processing techniques are applied in rotating machinery fault diagnosis. For instance, variational mode decomposition (VMD), maximum correlated kurtosis deconvolution (MCKD), singular value decomposition (SVD) and sparse decomposition, etc. Although these techniques have made certain achievements in improving the accuracy of rotating machinery fault detection, they are based on tracking the amplitude of the fault frequency and neglecting the inherent modulation characteristics in the measurement signal. If the modulation components and carrier components of the measurement signal cannot be effectively separated, it will be difficult to accurately obtain the fault features.

Spectral correlation (SC) analysis is a two-dimensional spectrogram that can simultaneously present modulation components and carrier components for fault detection in rotating machinery [5], [6]. However, the SC contains abundant discrete Fourier transforms in the analysis process of the measurement signals, which results in excessive calculational costs. In order to solve this problem, Antoni et al. [7] developed a cyclic modulation spectrum (CMS), which first performs a short-time Fourier transform (STFT) on the measured signal, and then implements a Fourier transform on the time axis of the spectrum. It not only greatly improves the performance of fault extraction, but also significantly outperforms SC in terms of computational efficiency. With its excellent performance, CMS provides an effective tool for rotating machinery fault diagnosis [8], [9]. However, the maximum cycle frequency \( f_{\text{max}} \) of CMS is chosen to be less than \( F_s/2 \) (Fs represents the sampling frequency), and its cycle frequency resolution is low because the window size applied used in STFT is fixed. Multi-scale combination morphological filter (MCMF) is a nonlinear filtering processing technique based on mathematical morphological transformation. It relies on structural elements (SE) with the same function as the filter window to match the geometric characteristics of the signal to be analyzed to achieve fault feature retention and interference noise removal. However,
because the SE of the same scale is applied, the output of MCMF will be biased [10]. Therefore, an improved MCMF operator with different scales is developed to construct an improved multi-scale combination morphological filter (IMCMF) and integrated into the CMS to replace the STFT window in this study, named cyclic morphological modulation spectrum (CMMS), which can effectively overcome the drawback of low cycle frequency resolution and expand the maximum cycle frequency range to improve the fault detection accuracy. Unfortunately, the choose of fault information bandwidth from CMMS relies on user experience choices. That is, improper selection of frequency bands will also affect the accuracy of the fault features. Sparsity measurement (SM) is a potential method to determine the optimal frequency, which involves the fault impulse component of the measured signal. Moreover, SM has been demonstrated to be an effective frequency band selection scheme [11], [12]. In view of this, SM is utilized to select the optimal frequency band of CMMS to reduce in-band noise.

Apparently, several frequency bands selected by CMMS using SM have different fault signatures. Therefore, it is necessary to adopt accurate measurement schemes to obtain the fault features of each frequency band. In recent years, entropy-based algorithms have been introduced into the fault feature extraction of rotating machinery, such as correlation dimension (CD), sample entropy (SE), approximate entropy (AE) and fuzzy entropy (FE). Although these algorithms have achieved certain benefits in fault feature extraction, they still have some limitations in their applications. The CD requires a longer data set when processing the measured data, so CD is not suitable for practical applications [13]. The SE is generated by the Heaviside step function, which is discontinuous and abrupt at the boundary [14]. The AE largely depends on record length and lower estimation value in fault feature extraction [15]. The FE is defined using the membership function, and the parameters are extremely hard to determine [16]. Recently, a new entropy-based algorithms named permutation entropy (PE) was put forward by Bandit [17], which can amplify weak signals and detect the operating mode of the mechanical devices. On the basis of PE, multi-scale permutation entropy (MPE) is put forward to improve the statistical significance of PE. However, the coarse-grained program applied in MPE indicates linear smoothing, which uses an averaging strategy to obtain low-frequencies while ignoring the feature information submerged in high-frequencies [18]. Therefore, Li et al. [19] developed a new indicator named hierarchical permutation entropy (HPE), which is specifically used to evaluate the dynamics of rotating machinery systems. In view of the excellent performance of HPE, a large number of researchers have made great contributions to promote the development of HPE-based weak fault feature extraction methods [20], [21]. Besides, the Teager energy operator (TEO) was selected as an effective demodulation fault feature method by Randall et al. [22] and Song et al. [23] to recognize fault types under different operating situations. Based on the advantages of HPE and TEO, a novel feature extraction algorithm is presented, called hierarchical Teager permutation entropy (HTPE), which can effectively overcome the modulation information that HPE cannot effectively capture the frequency band features. Hence, CMMS and HTPE are combined to obtain fault features of rotating machine in this study. Finally, the acquired feature vectors are integrated into the extreme learning machine (ELM) classifier to achieve the fault classification of rotating machine under different working situations. It should be pointed out that compared with these classification methods (e.g., back-propagation neural network (BPNN), k-nearest neighbor (KNN) and support vector machine (SVM)), ELM classifier has the advantages of fast learning speed and good generalization ability. Additionally, the weight matrix between the input layer and the hidden layer and the bias of the hidden layer neurons are randomly assigned, and no adjustment is required during the training process [24], [25].

To sum up, the main intention of this paper is to develop a new intelligent fault detection scheme that integrates CMMS, HTPE and ELM classifier to realize automatic fault feature recognition and classification. The main innovations and novelties of this paper are summarized as follows: (1) A new method named CMMS is proposed to decompose the measurement signal to obtain CMMS slices with different frequency bands, which overcomes the shortcomings of traditional CMS that require manual experience to select bandwidth and low cyclic frequency resolution to obtain more comprehensive and abundant fault feature information. (2) A novel fault feature indicator (i.e., HTPE) is developed to construct feature vectors, which can overcome the deficiency that HPE cannot effectively capture the modulation information of frequency band features and improve the fault identification ability of rotating machinery. The results on simulated signals and two experimental cases indicate the proposed scheme realizes the classification and recognition of the fault mode. In addition, compared with some existing methods (e.g., MSB-HTPE, Fast-SC-HTPE, CMMS-HPE, CMMS-HSE and CMMS-HPE), the CMMS-HTPE provides better detection performance when used for rotating machinery fault classification.

The remainder of the paper is organized as follows. In Section II, the theoretical aspects of CMMS are introduced. In Section III, the detailed diagnostic process of the proposed scheme is presented. Vibration signals collected from the cylindrical roller bearings and planetary gearboxes are provided to illustrate the effectiveness of the proposed scheme in Section IV. Finally, the conclusions of the research work are summarized in Section V.

II. CYCLIC MORPHOLOGICAL MODULATION SPECTRUM

A. Cyclic Morphological Modulation Spectrum

Morphological filter (MF) is a non-linear filtering technology. Its core idea is to perform front-to-back translation matching or partial correction to the original signal by constructing specific structural elements, so as to suppress noise while retaining the morphological characteristics of the signal. According to the local morphological characteristics of the signal to be processed, the signal is separated from noise through MF. Let \( g \) is a unit SE, and \( k (k = 1, 2, \ldots, K) \) is the scale, the SE used in the scale \( k \) can be expressed as:

\[
kg = g \oslash g \oslash \cdots \oslash g
\]

(1)

Assume that \( f(n) \) is the input signal. Multi-scale basic morphological operators are defined as:

\[
(f \oslash kg)(n) = f \oslash (g \oslash g \oslash \cdots \oslash g)
\]

(2)

\[
(f \omega kg)(n) = f \omega (g \omega g \omega \cdots \omega g)
\]

(3)

\[
(f \bullet kg)(n) = (f \bullet (g \bullet g \bullet \cdots \bullet g))(n)
\]

(4)

\[
(f \cdot kg)(n) = (f \cdot (g \cdot g \cdot \cdots \cdot g))(n)
\]

(5)

where \( \oslash, \omega, \bullet \) and \( \cdot \) are dilation, erosion, opening and closing operator, respectively. Based on the combination of opening and closing operator, two multi-scale morphological filters, namely, multi-scale opening-closing (\( F_{OC} \)) and multi-scale closing-opening (\( F_{CO} \)), which are defined as follows:

Both \( F_{OC_1}(n) \) and \( F_{OC_2}(n) \) filters can obtain cyclic impulses. However, due to the statistical bias, satisfactory results cannot be obtained by the \( F_{OC_1}(n) \) or \( F_{OC_2}(n) \) filter alone. To solve the problem, the average combination of \( F_{OC_1}(n) \) and \( F_{OC_2}(n) \) (i.e., multi-scale combination morphological filter, MCMF) is proposed, which is expressed as:

\[
y(n)(k) = F_{OC_2}(n) + F_{OC_1}(n)
\]

For the \( F_{OC_1}(n) \) filter, the first opening operator can enhance negative impulses and filter out positive impulses. If the closing operator uses the same length of SE, it will not be able to effectively filter out all negative impulses. Similarly, the \( F_{OC_2}(n) \) filter also has the same issue. To eliminate the above issue, the improved multi-scale opening-closing \( (F_{OC_1}(n)) \) and multi-scale closing-opening \( (F_{OC_2}(n)) \) filters are defined as:

\[
F_{OC_1}(n) = ((f \ast k_1) \ast k_2)(n)
\]
\[
F_{OC_2}(n) = ((f \ast k_2) \ast k_1)(n)
\]

where the lengths of \( k_1 \) and \( k_2 \) are \( L \) and \( 2L-1 \), respectively. The improved multi-scale combination morphological filter (IMCMF), which is denoted as:

\[
y(n)(k) = F_{OC_2}(n) + F_{OC_1}(n)
\]

where \( y(n)(k=1,2,\ldots,K) \) is a two-dimensional matrix \( y(k,n) \), each row of which corresponds to a specified scale \( k \), each column corresponds to a specific point \( n \) across entire scales. Due to the appearance of random noise and interference frequencies, it is difficult to accurately identify periodic impulse features in the time domain. Therefore, the discrete Fourier transform (DFT) is applied to transform \( y(k,n) \) into \( \hat{Y}(k,f) \), which is defined as:

\[
\hat{Y}(k,f) = \sum_{n=0}^{\infty} y(k,n) e^{-j2\pi nf/T}
\]

where \( \hat{Y}(k,f) \) represents spectrum of the IMCMF, \( f \) means spectral frequency. Consequently, cyclic morphological modulation spectrum (CMMS) is calculated from the DFT of IMCMF spectrum, which is defined in Eq. (13):

\[
S_{CMMS}(a,f) = \frac{1}{K} \sum_{k=1}^{K} \hat{Y}(k,f)^{\frac{t - \gamma}{T}}
\]

where \( K \) means the sampling frequency, \( K \) represents the signal length, \( a \) indicates the cyclic frequency. The normalized form of CMMS can be given by Eq. (14), and the result is shown in Fig. 1.

\[
\gamma_{CMMS}(a,f) = \frac{S_{CMMS}(a,f)}{\sqrt{\sum_{f \in \mathcal{F}} S_{CMMS}(0,f)^2 \sum_{f \in \mathcal{F}} S_{CMMS}(0,f,a)}}
\]

Fig. 1. Normalized form of CMMS.

To get the suboptimal \( f_i \) slices, the \( \gamma_{CMMS}(a,f) \) is determined by CMMS slice \( S(f) \), which is computed by averaging the main CMMS peaks:

\[
S(f) = \frac{1}{M-1} \sum_{\alpha=1}^{M} \gamma_{CMMS}(\alpha,f_i)
\]

where \( \alpha \) means the frequency resolution in the \( \alpha \) direction. However, CMMS needs a better method to select the number of frequency bands. In other words, when the CMMS processes the measured signals to extract fault defect features, it needs to choose a reasonable bandwidth. Consequently, a new criterion called sparseness measurement (SM) is proposed to determine the number of CMMS slices to cover the useful resonance frequency band to reduce the loss of effective fault information. According to the analysis of CMMS slice \( S(f) \), the SM can be defined as follows:

\[
SM = \frac{\sum_{f \in \mathcal{F}} |S(f)|^2}{\sum_{f \in \mathcal{F}} |S(f)|^2}
\]

where \( L \) indicates the length of the CMMS slices. The \( |S(f)|^2 \) is \( L_1 \) norm and \( L_2 \) norm, respectively. If the measured signal has good sparsity, the SM will be larger. That is, the larger the SM, the better the CMMS fault feature extraction effect. Therefore, the largest SM is used to determine the number of CMMS slices.

To obtain more robust results, the CMMS detector is constructed by averaging the selected suboptimal slices marked with ‘↓’ in Fig. 1, which is defined as:

\[
S(a) = \frac{1}{P} \sum_{j=1}^{P} S_{CMMS}(a,f_i)
\]

where \( P \) represents the total number of selected CMMS slices, which depends on the number of SM.

**B. Comparison of CMMS, MSB and Fast-SC by simulation signal**

To evaluate the performance of the CMMS, the results are compared with modulation signal bispectrum (MSB) [26] and Fast spectrum correlation (Fast-SC) [27]. The rolling element bearing fault model \( x(t) \) with a local fault is defined as follows [28]:

\[
x(t) = \sum_{i} A_i \cos(2\pi f_i t + \theta_i) + \sum_{j} C_i G(t - iT - \tau_j) + \sum_{j} R_j G(t - Tr_j) + n(t)
\]

where the first part represents the periodic interference components from the rotor or shaft, \( f_i \) indicates the rotational frequencies, set to 10 Hz; \( A_i \) and \( \theta_i \) represent the measured signal amplitude and initial phase of the \( i \)th harmonic, respectively. The second part indicates the periodic impulse signal produced by a bearing defect, \( C_i \) indicates the \( i \)th impulse amplitude. \( G(t) \) means the impulse response. The \( \tau_j \) indicates a uniformly distributed random number for the random slippage effect of the bearing, which is often represented as \( 1\% - 2\% \) of the \( T \). The periodic impulse can be denoted as follow:

\[
G(t) = e^{-\alpha \sin(2\pi f_i t)}
\]

where \( f_i \) stands the resonant frequency, \( \alpha \) represents the decay parameter. The third part indicates random impulse aroused by external interference from the rolling bearing housing. Among them, \( R_j \) stands random impulse amplitude, \( Tr_j \) indicates occurrence time. The last part \( n(t) \) represents Gaussian white noise with a signal-to-noise ratio (SNR) of -15 dB. The parameters of the
bearing fault model are illustrated in Table 1. Meanwhile, the sampling frequency \(f_s=96\) kHz and sampling number \(N=960000\). Fig. 2 displays the time waveform of \(x(t)\) and its frequency spectrum.

<table>
<thead>
<tr>
<th>Rotor or shaft</th>
<th>Periodic impulses</th>
<th>Random shocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A_1)</td>
<td>(\theta_1)</td>
<td>(C_i)</td>
</tr>
<tr>
<td>0.01</td>
<td>(\theta^\ast)</td>
<td>1</td>
</tr>
</tbody>
</table>

The CMMS is utilized to process the simulated signal illustrated in Fig. 2(a) to extract the fault defect frequencies. First of all, the lengths of \(g_1\) and \(g_2\) are 3 and 5 [29], respectively. Subsequently, the CMMS slices \(S(f_i)\) are computed using Eq. (15), and the suboptimal CMMS slices are selected according to the SM method. As illustrated in Fig 3, when the CMMS slice is 3, SM reaches its maximum value. Finally, the CMMS is obtained through averaging 3 suboptimal CMMS slices with \(\ast\ast\) makers in Fig. 4(a), and the detection result is displayed in Fig. 4(b). It is clearly found that the CMMS can accurately distinguish the fault frequency 60 Hz and its first 4 harmonics.

![Fig. 2. Simulation case: (a) time waveform (b) frequency spectrum.](image)

![Fig. 3. The SM value of CMMS slices.](image)

![Fig. 4. Results of the CMMS: (a) CMMS slice (b) CMMS detector.](image)

For comparison, the MSB and Fast-SC are utilized to process the simulated signal of defective bearing. The MSB is a novel modulation signal demodulation algorithm put forward in Ref. [26]. For MSB method, the MSB slice and window length \(N_w\) are set to 3 and 214. The algorithm uses the above parameters to suppress the interference of random noise and non-periodic components to clearly reflect the modulation components in the measurement signal. As depicted in Fig. 5, the detection result obtained by the MSB is difficult to distinguish the fault frequencies. For the Fast-SC algorithm [27], the window length \(N_w\) and the maximum cyclic frequency \(\alpha_{max}\) are equal to 214 and 250 Hz, the Fast-SC using the above parameters to process the fault model. Hereafter, the enhanced envelope spectrum (EES) is utilized to obtain the fault defect frequencies from Fast-SC filtered signals. Although the fault defect frequency 60 Hz and its first 4 harmonics can be extracted, the redundant noise and interference frequencies in higher harmonic frequencies still exist as illustrated in Fig. 6.

![Fig. 5. Results of the MSB: (a) MSB slice (b) MSB detector.](image)

![Fig. 6. Results of the Fast-SC: (a) spectral coherence image (b) EES generated in the frequency band from 4000 Hz to 5000 Hz.](image)

![Fig. 7. The defect index of three algorithms under different SNRs. In order to further highlight the effectiveness of the CMMS algorithm, the defect index \(u\) is presented to evaluate the performance of three algorithms. The defect index \(u\) is defined as follows:](image)
where $A_c$ and $A_r$ indicate the magnitude of the fault frequency and white noise (or random impulse) interferes. Fig. 7 depicts the $u$ of three algorithms under different SNRs. As depicted in Fig. 7, with the SNR decreases, the $u$ of three algorithms also gradually decreases. Moreover, CMMS has stronger fault recognition capabilities than MSB and Fast-SC. Through the above comparison, the CMMS illustrates higher performance in terms of enhanced fault features and signal modulation.

### III. The Proposed Fault Detection Scheme

#### A. Hierarchical Teager permutation entropy for feature extraction

Teager energy operator (TEO) is a novel model of nonlinear operator that forecasts the entire energy required through the signal source to generate a dynamic signal by the nonlinear combination of the instantaneous value of the signal. For a given CMMS slice $S(f_l)f_l=1,2,...P_l$, the TEO can be defined as:

$$\Psi(S(f_l))=S(f_l)−S(f_l−1)$$

where $\Psi(S(f_l))$ means the use of $3$ samples to calculate the signal energy at frequency $f_l$. Then, use phase space reconstruction to obtain the matrix:

$$Y = \begin{bmatrix}
\Psi(S(0)) & \ldots & \Psi(S(m-1)) \\
\Psi(S(1)) & \ldots & \Psi(S(m+1)) \\
\vdots & & \vdots \\
\Psi(S(n)) & \ldots & \Psi(S(n+m-1))
\end{bmatrix}
$$

where $m$ and $\tau$ are embedding dimension and delay time, respectively. Each row $Y(j)$ in the matrix $Y$ is treated as a reconstructed component, with a total of $n$ reconstructed components, and $n=P_l−(m−1)\tau$. Rearrange $Y(j)=\{\Psi(S(0)),\ldots,\Psi(S(m-1))\}$ in ascending order, $j_0,j_1,\ldots,j_{n}$ indicates the index of the column where each element in the reconstructed component is located, that is:

$$\Psi(S(f_l)−j_{l}−1)≤\Psi(S(f_l)−j_{l})≤\ldots≤\Psi(S(f_l)−j_{0})$$

If there are equal values exist in the reconstructed component, which is $\Psi(S(f_l)−j_{l}−1))=\Psi(S(f_l)−j_{l}))$ at this time, the $j_{l}$ and $j_{0}$ values are sorted. When $j_{l}≤j_{l}$ , the arrangement is expressed as:

$$\Psi(S(f_l)−j_{l})≤\ldots≤\Psi(S(f_l−1))$$

Therefore, for each line $Y(j)$ of any time series reconstructed, a set of symbol sequences can be obtained as follow:

$$B(I) = (j_0,j_1,\ldots,j_n)$$

where $l=1,2,\ldots,k$ and $k≤m!$. The $m$ dimensional phase space maps different $m!$ symbol sequences, and symbol sequence $B(I)$ is one of them.

The occurrence probability of each symbol sequence is denoted as $p_{1},p_{2},\ldots,p_{l}$ and the PE of $\Psi(S(f_l))$ can be denoted as:

$$H_{p}(m) = −\sum_{i=1}^{m} p_i \ln p_i$$

where $0≤H_{p}(m)≤\ln(m!)$. When $p_i=1/m!$ , $H_{p}(m)$ reaches the maximum value $\ln(m!)$. The PE value is normalized as:

$$H_{p} = H_{p}(m)/\ln(m!)$$

The value of $H_{p}$ represents the randomness degree of $\Psi(S(f_l))$ . The smaller value of $H_{p}$ , the more orderliness the $\Psi(S(f_l))$ ; otherwise, the more random it is. Therefore, $H_{p}$ reflects a small change in the $\Psi(S(f_l))$ . In order to describe the complexity of $\Psi(S(f_l))$ more accurately, the hierarchical permutation entropy (HPE) method is proposed in Ref. [20]. The averaging operator $Q_{0}$ and high operator $Q_{1}$ are expressed as:

$$Q_{0}(\Psi(S)) = \frac{1}{2}\left[\Psi(S(f_l))+\Psi(S(f_l+1))\right]$$

$$Q_{1}(\Psi(S)) = \frac{1}{2}\left[\Psi(S(f_l))-\Psi(S(f_l+1))\right]$$

Subsequently, the operator $Q_{0}^{(b=0 or 1)}$ at the hierarchical layer $\sigma$ can be formulated as:

$$Q_{0}^{(b)} = \left[\frac{1}{\sigma} \left(\frac{(-1)^{\sigma}}{2^{\sigma-1}} \begin{array}{cccc}
0 & \cdots & 0 & 0 \\
0 & \cdots & 0 & 0 \\
\vdots & \cdots & \vdots & \vdots \\
0 & \cdots & 0 & 1
\end{array}\right) \right]_{2^{\sigma-1}}$$

For a given vector $[\lambda_{1},\lambda_{2},\ldots,\lambda_{n}]$ , the integer $\gamma$ is written as:

$$\gamma = \sum_{\xi=i}^{2^{\sigma-1}} \lambda_{\xi}$$

where $[\lambda_{1},\lambda_{2},\ldots,\lambda_{n}] = [0,1]$ means the $Q_{0}$ or $Q_{1}$ at the $\xi^{\sigma}$ layer.

According to the $[\lambda_{1},\lambda_{2},\ldots,\lambda_{n}]$ , the hierarchical component of $\Psi(S(f_l))$ is written as:

$$X_{\sigma} = \sum_{i=0}^{\gamma} Q_{i}^{(b=0 or 1)}(\Psi(S(f_l)))$$

The $X_{\sigma}$ is used as the input of PE to calculate HTEP, as depicted in Eq. (33):

$$H_{\text{ave}}(m) = \sum_{i=1}^{N} p_i \ln p_i$$

where $H_{\text{ave}}$ can more fully enhance the fault component and eliminate random noise of the measurement signal by combining the TEO and HPE, which is helpful for rotating machinery fault diagnosis.

#### B. Extreme learning machine for fault recognition

Extreme learning machine (ELM) is a classification algorithm based on single hidden layer feedforward neural network (SLFN), which has better generalization performance than traditional learning algorithms. Moreover, ELM is extremely insensitive to dynamic parameters and can complete the classification of data sets faster and more conveniently. Fig. 8 depicts the schematic diagram of ELM.

For $M$ training samples $x_j(j=1,\ldots,M)$ , where $x_j = [x_{j1},x_{j2},\ldots,x_{jn}]^T \in R^n$ indicates the input data, $z_j = [z_{j1},z_{j2},\ldots,z_{jn}]^T \in R^n$ indicates the desired output label. The output function is defined as:

$$\sum_{i=1}^{N} \beta_{j}(\omega_{j}+b) = o_j(j=1,\ldots,M \text{ and } i=1,\ldots,N)$$

(34)
where \( g(\cdot) \) denotes the sigmoid function of ELM, \( w_i = [w_{i1}, w_{i2}, \cdots, w_{in}]^T \) and \( \beta_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{in}]^T \) are the weight vector between the \( i \)th hidden node and the input node, \( b_i \) means the bias of the \( i \)th hidden node and \( o_j \) means the output of ELM for \( j \)th sample. Next, the hidden nodes \( H \) is calculated as:

\[
H = \begin{bmatrix}
g(w_{11}x_1 + b_1) & \cdots & g(w_{1n}x_n + b_1) \\
g(w_{21}x_1 + b_2) & \cdots & g(w_{2n}x_n + b_2) \\
\vdots & \ddots & \vdots \\
g(w_{m1}x_1 + b_m) & \cdots & g(w_{mn}x_n + b_m)
\end{bmatrix}_{m \times n}
\]

(35)

\[
\beta = \begin{bmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_n
\end{bmatrix}
\quad \text{and} \quad
Z = \begin{bmatrix}
z_1 \\
z_2 \\
\vdots \\
z_n
\end{bmatrix}
\]

(36)

where \( \beta \) and \( Z \) denote the output weight vector and target matrix. In view of ELM theories, the hidden nodes \( (w_i, b_i) \) is randomly distributed rather than adjusted. The \( \beta \) of the Eq. (37) is calculated as:

\[
\beta = H^+ Z
\]

(37)

where \( H^+ \) means Moore-Penrose generalized inverse of the hidden nodes \( H \).

Fig. 8. Schematic diagram of ELM.

C. The proposed fault detection scheme

As mentioned above, this section develops a novel fault detection scheme based on CMMS, HTPE and ELM classifier. The framework of the fault detection scheme is illustrated in Fig. 9, and its diagnostic process is elaborated as follows:

IV. EXPERIMENTAL VALIDATION

A. Case I: Bearing Fault Diagnosis

To verify the availability of the proposed scheme for rolling element bearing, the faulty cylindrical roller bearing was performed to collect vibration signals. As depicted in Fig. 10, the test bench is composed of motor, bearing house, flexible type coupling and DC generator. During the experiment, three different fault defects (i.e., outer race, inner race and ball race) on cylindrical roller bearings were produced using the electrical discharge machining, and their depth was 0.5 mm and width was 0.2 mm. The sampling frequency of the faulty cylindrical roller bearing data was 71428 Hz and signal length were 300,000 data points, respectively. Geometric parameters of the faulty cylindrical roller bearing are described in Table II. The faulty cylindrical roller bearing data consists of four working conditions. There are 40 samples for each working condition and a total of 160 samples. The 80 samples are choosing randomly as the training data, and other 80 samples as the testing samples. These data messages are described in Table III.
Diagnosis accuracy is also desirable.

Different working conditions of cylindrical roller bearings, and the CMMS-HTPE and ELM classifier can distinguish the information, the measured signal is pre-processed and demodulated using the CMMS. Subsequently, the TEO of the measured signal is calculated and used as the import of the HTPE vector. Finally, the acquired HTPE vector is integrated into the ELM classifier. Before the implementation of the ELM classifier, the number of hidden neurons is set to 20, and the sigmoidal function is applied [30].

Fig. 12 illustrates the confusion matrix of the proposed scheme in the first trial. As depicted in Fig 12, the minimum detection accuracy rate is 95% and the average detection accuracy rate is 98.75%. This reveals that the CMMS-HTPE and ELM classifier can distinguish the different working conditions of cylindrical roller bearings, and the diagnosis accuracy is also desirable.

The waveform of the cylindrical roller bearing under four working conditions are depicted in Fig. 11. Obviously, the fault feature information is completely overwhelmed by random noise and interference frequencies. To obtain the fault feature information, the measured signal is pre-processed and demodulated using the CMMS. Subsequently, the TEO of the CMMS filtered signal is calculated and used as the import of the HTPE vector. Finally, the acquired HTPE vector is integrated into the ELM classifier. Before the implementation of the ELM classifier, the number of hidden neurons is set to 20, and the sigmoidal function is applied [30].

Table III: Data Messages about Cylindrical Roller Bearing

<table>
<thead>
<tr>
<th>Working conditions</th>
<th>Class label</th>
<th>Number of training data</th>
<th>Number of testing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Outer race</td>
<td>2</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Inner race</td>
<td>3</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Ball race</td>
<td>4</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Fig. 13. Detection accuracy of three methods in 20 trials.

To prove the availability of the proposed scheme, the measured signals of cylindrical roller bearing are applied to contrast the performance between CMMS-HTPE, MSB-HTPE and Fast-SC-HTPE. In addition, considering the randomness of the fault diagnosis results in one trial, the cylindrical roller bearing data is carried out twenty times in this study. Fig. 13 presents the detection accuracy of three algorithms for twenty trials. It can be found that compared with MSB-HTPE and Fast-SC-HTPE, CMMS-HTPE has the most outstanding detection accuracy. Thus, it is deduced that CMMS-HTPE is more robust than MSB-HTPE and Fast-SC-HTPE.

Fig. 14. Average detection accuracy of four methods in 20 trials under different percentages of training samples.

Table IV: Detection Accuracy of the Four Methods under Different Percentages of Training Samples
160 samples. The 80 samples are selected randomly as training samples. There are 40 samples under each working condition for a total of 120 samples. The geometric parameters of the planetary gearbox are listed in Table V. The lengths were 300,000 data points, respectively. Geometric parameters of the experiment data was 96 kHz and signal length were 300,000 data points, respectively. The failure modes mainly include sun gear chipping and sun gear misalignment. The planetary gearbox. In this experiment, the failure modes mainly include sun gear chipping and sun gear misalignment.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy rate acquired under different percentages of training samples (%)</th>
<th>Average CPU time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>30%</td>
</tr>
<tr>
<td>CMMS-H</td>
<td>Max</td>
<td>89.58</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>73.96</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>81.39</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>4.54</td>
</tr>
<tr>
<td>CMMS-H PE [20]</td>
<td>Max</td>
<td>87.84</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>69.10</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>79.08</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>5.66</td>
</tr>
<tr>
<td>CMMS-H SE [14]</td>
<td>Max</td>
<td>80.56</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>62.15</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>70.05</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>4.88</td>
</tr>
<tr>
<td>CMMS-H FE [16]</td>
<td>Max</td>
<td>79.17</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>56.94</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>69.57</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>6.26</td>
</tr>
</tbody>
</table>

For comparison, the advanced fault detection method based on permutation entropy (e.g., Hierarchical permutation entropy (HPE), Hierarchical sample entropy (HSE) and Hierarchical fuzzy entropy (HFE)) are utilized to analyze the same data samples. Moreover, we set the parameters of all algorithms to the same value to process the experimental data, where $m$ means embedding dimension, equal to 6, $r$ indicates time delay, set to 1, $\sigma$ represents standard deviation (SD) of signal, $q$ means hierarchical layer, set to 3, and $r$ indicates the tolerance and is equal to 0.15 times of $c$. Concretely, for HTPE and HPE, we set $m=6$, $r=1$ and $q=3$. For HSE and HFE, we set $m=6$, $q=3$ and $r=0.15\times c\sigma$, according to [19]. Here, we will set the percentage of training samples as: 10%, 30%, 50%, 70% and 90%. Hereafter, the detection accuracy of the above four methods is calculated in 20 trials, as illustrated in Fig. 14. As depicted in Fig. 14, the CMMS-HTPE has the highest detection accuracy compared with the other three algorithms (CMMS-HPE, CMMS-HSE and CMMS-HFE) regardless of the percentage of training samples. In addition, Table IV lists the comparison indicators of the four algorithms in 20 trials, including the maximum (Max), minimum (Min), mean and SD of classification accuracy, as well as the average CPU time. As displayed in Table IV, the maximum and minimum recognition accuracy of the detection result of the proposed scheme is higher than that of the three algorithms (e.g., CMMS-HPE, CMMS-HSE and CMMS-HFE) under different percentages of training samples. At the same time, the average CPU time of the proposed scheme is 4.07 s, which is lower than the CPU time of CMMS-HSE and CMMS-HFE, but higher than that of CMMS-HPE. The main reason that the calculation efficiency of the proposed scheme is lower than that of CMMS-HPE is that TEO is added to HPE. Moreover, the proposed scheme has the smallest SD, which reveals that the proposed scheme has the best detection ability.

### B. Case II: Planetary Gearbox Fault Diagnosis

In this section, the effectiveness of the proposed scheme will be further investigated on the planetary gearbox vibration signal obtained from the test platform presented in Fig. 15. Test platform consists of motor, planetary gearboxes, helical gearbox, load generator and vibration sensor are horizontally fixed on top of the planetary gearbox. In this experiment, the failure modes mainly include sun gear chipping and sun gear misalignment. The sampling frequency of the experiment data was 96 kHz and signal length were 300,000 data points, respectively. Geometric parameters of the planetary gearbox are described in Table V. There are 40 samples under each working condition for a total of 160 samples. The 80 samples are selected randomly as training samples and the other 80 samples as testing samples. The data messages are given in Table VI.

![Fig. 15. The test platform for the planetary gearbox facility.](image)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Number of teeth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun gear</td>
<td>10</td>
</tr>
<tr>
<td>Planet gear (number)</td>
<td>20(3)</td>
</tr>
<tr>
<td>Ring gear</td>
<td>62</td>
</tr>
<tr>
<td>Carrier</td>
<td></td>
</tr>
</tbody>
</table>

**Table VI**

<table>
<thead>
<tr>
<th>Working conditions</th>
<th>Class label</th>
<th>Number of training data</th>
<th>Number of testing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Sun gear chipping</td>
<td>2</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Sun gear with 0.4 mm misalignment</td>
<td>3</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Sun gear with 1.0 mm misalignment</td>
<td>4</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

![Fig. 16. Waveform of the sun gear](image)
In this paper, a novel intelligent detection algorithm based on CMMS, HTPE and ELM is put forward, which can automatically realize fault classification of rotating machinery. Case studies of rolling bearings and planetary gearboxes verify the effectiveness of the proposed scheme in fault classification and identification. The main conclusions of this paper are summarized as follows:

1. A new cyclic morphological modulation spectrum (CMMS) is put forward for fault feature extraction. Simulation and experimental results reveal that CMMS has stronger denoising and demodulation performance than MSB and Fast-SC algorithms, indicating the superiority of using CMMS in fault feature extraction.

2. The hierarchical Teager permutation entropy (HTPE) is proposed to select CMMS features. Experimental analysis shows that CMMS-HTPE has better detection performance than some existing algorithms (e.g., CMMS-HPE, CMMS-HSE and CMMS-HFE).

3. The case study of rotating machinery shows that the proposed scheme realizes the classification and recognition of failure modes under different working conditions, and the detection accuracy is higher than that of the methods (e.g., MSB-HTPE, Fast-SC-HTPE, CMMS-HPE, CMMS-HSE and CMMS-HFE) under the ELM classifier. Thence, the proposed scheme is proved to be effective for identifying different operating conditions of rotating machinery.
REFERENCES


Technology, China, in 2006, received the M.S. degree in signal and information processing from Shandong University of Science and Technology, China, in 2009, received the Ph.D. degree in mechanical engineering from University of Huddersfield, UK, in 2012. He is currently a professor and associate dean of the School of Mechanical Engineering, Hebei University of Technology, Tianjin, China. His main research interests include machinery condition monitoring and fault diagnostics, signal processing and intelligent algorithms development.

Fengshou Gu received the B.S. degree from Taiyuan University of Technology, China, in 1979, received the M.S. degree from Taiyuan University of Technology, China, in 1985, received the Ph.D. degree in mechanical engineering from Manchester University, UK, in 2008. He is currently a professor and the Deputy Dean of the Centre for Efficiency and Performance Engineering (CEPE), University of Huddersfield, UK. His research interests include condition monitoring and fault diagnosis of rotating machines, signal processing, sensing technology and the applications of intelligent diagnosis technology.