IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 16, 2020, accepted October 22, 2020, date of publication October 29, 2020, date of current version November 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3034651

Multi-Domain Extreme Learning Machine for Bearing Failure Detection Based on Variational Modal Decomposition and Approximate Cyclic Correntropy

XIAOHUI WANG[1], GUANGZHOU SUI[1], JIAWEI XIANG[2], (Member, IEEE), GUANGBIN WANG[1], ZHIQIANG HUO[3], AND ZHEN HUANG[4]

[1]Department of Electrical and Mechanical Engineering, Lingnan Normal University, Zhanjiang 524048, China
[2]Department of Electrical and Mechanical Engineering, Wenzhou University, Wenzhou 325035, China
[3]School of Engineering, University of Lincoln, Lincoln LN6 7TS, U.K.
[4]Department of Physics, Lingnan Normal University, Zhanjiang 524048, China

Corresponding authors: Guangzhou Sui (suigzh@lingnan.edu.cn) and Jiawei Xiang (jwxiang@wzu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61705095 and Grant 5157518, and in part by the Non-Funded Science and Technology Public Relations Plan Project Foundation of Zhanjiang under Grant 2017B01092.

ABSTRACT Rolling bearings are critical in industrial mining machinery. Due to strong Gaussian noise, frequent random shocks, and disordered loads in industrial settings, it is usually difficult to detect weak fault symptoms in vibration signals from a bearing. To detect incipient bearing faults, this paper proposes a new multi-domain kernel extreme learning machine (MKELM) based on variational modal decomposition (VMD) and a cyclic correntropy function. A normalized approximation algorithm for a cyclic correntropy function (NACCF) was first built to suppress the impulsive background noise. This approach is suitable for machine learning. To eliminate the Gaussian noise effectively, genetic mutation particle swarm optimization (GMPSO) with cyclic information entropy (CIE) was used to optimize the VMD parameters. The CIE was created as a fitness function in GMPSO to search for the best hyperparameters. It can be used to select effective intrinsic mode functions (IMFs) to reconstruct denoised signals. Then, statistical functions based on NACCF were used to extract the cyclic frequency-domain characteristics of the denoised signal, and the singular values of the IMFs were obtained as time-domain features of the signal. Finally, the multi-dimensional features from the two domains were input into MKELM to classify the health of the bearing. Experimental studies were carried out to investigate the proposed method in bearing fault detection and identification. The results demonstrated the effectiveness of the proposed method in motor-bearing failure detection and its robustness to noise when analyzing bearing vibration signals under different working loads.

INDEX TERMS Cyclic correntropy, GMPSO, KELM, motor bearing, VMD.

I. INTRODUCTION

Induction motors are widely used in modern industrial factories, including manufacturing and petrochemical plants. Motors are typically operated under variable conditions, such as variable rotating speed and overload or overspeed states. These operations may lead to failure, which can result in performance degradation and even unexpected downtime, thereby causing financial losses or safety issues. According to [1], bearing-related failures account for over 40% of all motor failures. Therefore, there is a necessity to detect and identify bearing faults and potential faults as early as possible to ensure high production efficiency in industrial processes [2].

Intelligent failure detection of a motor bearing involves signal preprocessing, feature extraction, and fault identification. Signal preprocessing is used to eliminate noise. Some common methods are wavelet transforms [3], empirical mode decomposition (EMD) [4], local mean decomposition [5], and variational modal decomposition (VMD) [6].

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

An intelligent bearing failure detection method was proposed based on local mean decomposition, singular-value decomposition (SVD), and an extreme learning machine (ELM) [7]. Although local mean decomposition can make up for the shortcomings of EMD, both use recursive mode decomposition with cumulative errors. In 2013, Dragomiretskiy and Zosso [8] introduced the VMD algorithm and a variational model to determine the frequency center and bandwidth of each component, which can avoid the cumulative error and endpoint effect. The inhibition of modal confusion has high decomposition efficiency. Two VMD parameters are commonly optimized. For instance, Zhang *et al*. proposed a VMD parameter optimization method based on the grasshopper algorithm [9].

In 2019, Zhang and Yan proposed a motor diagnosis model using feature extraction for a wind rolling bearing based on VMD with particle swarm optimization [10]. Despite its wide application, this approach has two problems. First, particle swarm optimization can easily become trapped in a local optimum when solving a complex multimodal problem [11]. A hybrid particle swarm optimization algorithm with genetic mutation (GMPSO) was proposed with crossovers and mutations in the legacy algorithm to jump out of the local optimum [12]. Second, the choice of fitness function also affects the optimization [13]. Entropy measures the disorder in a system and is used to describe the state of the system [14]. In this paper, entropy is calculated with the cyclic correntropy function (CCF). Cyclic information entropy (CIE) is proposed as the fitness function. Therefore, this paper proposes a CIE-based GMPSO algorithm to optimize the VMD parameters, which are mutated by a genetic algorithm.

Feature extraction is the key to intelligent failure detection. In fact, different characteristics of a vibration signal are reflected in the time, frequency, and wavelet domains [15], although they are from the same source. Moreover, multiple domain features can be extracted from a smaller sample to give more complete information and more accurate predictions [16]. Wang *et al*. [17] developed a diagnosis method based on statistical features extracted from the time and frequency domains and integrated from vibration signals. Cerrada *et al*. proposed a multi-stage feature selection method for failure detection of gearboxes, using features from the time and frequency domains [18].

The most commonly used statistical features include those from the time domain, fast Fourier transform frequency domain, time–frequency domain, and wavelet domain [19]. Nevertheless, cyclic statistical features have rarely been mentioned or studied. Like other rotating machinery, the vibration signal from a bearing has an obvious periodic stability due to its special working mode and is called a cyclostationary signal. Thus, cyclostationary analysis is widely used for bearing failure detection. Algorithms such as the cyclic autocorrelation function (CAF) [15], spectral correlation density [20], and cyclic bispectrum [21] can demodulate the fault information in bearing signals effectively. However, the classical cyclostationary algorithm mentioned above, like conventional algorithms such as envelope spectrum and spectral kurtosis-based fast kurtogram, are signal analysis methods based on an assumption of Gaussian noise. Thus, its performance declines rapidly, or it even fails, when exposed to impulse noise [22].

In 2017, Fontes *et al*. [23] introduced correlation entropy [24] from information theory into cyclic stationary analysis. They defined the CCF as a generalization of the CAF. It is a new mathematical tool for higher-order cyclostationary analysis. The corresponding cyclic correntropy spectrum obtained by a fast Fourier transformation of the CCF is a generalization of the spectral correlation density. Due to the use of a Gaussian kernel function in CCF, impulse noise can be suppressed effectively [25]. That is, when a pulse type with a large value appears in the signal, the Gaussian kernel function approaches zero. Based on this, in 2019, Zhao *et al*. [22] first proposed using cyclic correntropy spectrum analysis for bearing failure detection. The approach is effective when the spectrum of the fault information is submerged in the impulse noise. Although CCF can effectively extract information despite the background of pulse noise, it is rarely used for cluster regression for the following reasons. First, the choice of the kernel function parameter zeta seriously affects the amplitude and performance of CCF. Second, three-dimensional CCF with information redundancy requires a large number of calculations. These factors are detrimental to machine training and learning.

Deep learning and classical machine learning are popular intelligent failure detection techniques. Deep learning combines low-level features to form more abstract characteristics. Its performance is excellent [26]. To achieve high performance in deep learning, typically, a large amount of data is required to train the optimized model. Comparatively, under certain constraints, small samples can also achieve excellent performance. Classical machine learning is more suitable for the failure detection of mining machinery. Common machine learning algorithms include artificial neural networks, support vector machines [27], ELM, kernel extreme learning machine (KELM) [28], etc. The KELM algorithm has been widely used because it has few parameters, a fast response, high precision, and good generalization.

Fault detection and identification for motor bearings are particularly difficult in actual working conditions for three reasons: 1) the signal is acquired indirectly and has strong noise, 2) there are impulsive noise caused by random shocks, and 3) there are large load changes and frequent switching. To mitigate the first of these, to the best of our knowledge, this is the first use of a CIE-based GMPSO algorithm to optimize the VMD parameters to improve the efficiency of VMD. We used a sensitive intrinsic mode function (IMF) to reconstruct the denoising signal according to the CIE and to eliminate the interference component effectively. For the second point, based on the CAF and CCF, the normalized approximation algorithm for CCF (NACCF) was developed, which can demodulate a signal effectively from a background

of strong impulse noise. For the third point, because of the monotonic relation between the rotation frequency and the load of a three-phase asynchronous motor, the real-time rotation frequency was an input variable when constructing functions to extract cyclic frequency-domain features from NACCF. This can avoid detection failure caused by a load transformation. Finally, to improve the detection accuracy and stability, the time-domain features obtained from the singular values of IMF and the cyclic frequency-domain features obtained from NACCF are fused and input into KELM to form a multi-domain classifier.

The main contributions of this paper are summarized as follows:

1. NACCF is proposed. It is not only simpler and controls the threshold range, but it also retains the performance against impulsive noise of CCF and the flexibility in the selection of the kernel width.

2. The CIE is proposed. It has a monotonic relation with the correlation coefficient (COR) for key IMFs. GMPSO with the CIE as the fitness function is used to optimize the VMD parameters, with satisfactory denoising results.

3. Functions with real-time input rotation frequency are constructed to extract NACCF features, which can effectively avoid the failure of feature classification caused by a load transformation.

4. The multi-domain KELM (MKELM) based on singular values of the VMD component matrix and cyclic statistics achieved good results in bearing failure detection with strong Gaussian noise, frequent load conversion, and impulsive noise interference.

The rest of this paper is arranged as follows. In Section II, NACCF, which is applicable to second-order cyclostationary signals, is developed and verified using a high-moment CCF. Section III proposes the GMPSO parameter optimization method using CIE as the fitness function. In Section IV, the MKELM intelligent failure detection method is proposed based on NACCF and VMD-SVD. In Section V, fault signals from a driving end bearing and a rotating motor bearing are used to verify the effectiveness of the method proposed in this paper. The conclusion is given in the final section.

II. NACCF

In this section, we first introduce the second-order CAF and high-moment CCF and then develop the normalized approximate algorithm in NACCF, which has both CAF mediation and CCF impulsive noise resistance. Finally, the performance of NACCF is verified using a simulated normal-bearing signal and faulty-bearing signals.

A. CYCLIC AUTOCORRELATION AND CYCLIC CORRENTROPY

The CAF proposed by Gardner and Spooner [29] is well suited for demodulating mechanical signals and can extract fault features effectively. From a Fourier series expansion of the time-varying autocorrelation function $R(t,\tau) = E[x(t)x(t+\tau)]$ of the cyclostationary signal $x(t)$, we get:

$$R(t,\tau) = \sum\nolimits_{\alpha} R_x(\alpha,\tau)\exp(j2\pi\alpha t),$$

where α is the cyclic frequency, τ is the time delay, and coefficient $R_x(\alpha,\tau)$ is the CAF of signal $x(t)$ [20], satisfying:

$$\begin{aligned} R_{x(t)}(\alpha,\tau) &= \frac{1}{T}\int_{-T/2}^{T/2} R_x(t,\tau)\exp(-j2\pi\alpha t)dt \\ &= \langle R_x(t,\tau)\exp(-j2\pi\alpha t)\rangle \\ &= \langle x(t)x(t+\tau)\exp(-j2\pi\alpha t)\rangle , \end{aligned} \tag{1}$$

where $\langle\cdot\rangle = 1/T\int_{-T/2}^{T/2}(\cdot)dt$ is the time average operator and T is the time set.

As a generalized correlation function, the correlation entropy is widely used in nonlinear detection [23] for any random process $\{x_t, t \in T\}$. It is defined as:

$$\left.\begin{aligned} V_x(t,\tau) &= E[G_\sigma(x(t)-x(t+\tau)] \\ G_\sigma(\cdot) &= 1/[(2\pi)^{1/2}\sigma]\cdot\exp[-(\cdot)^2/2\sigma^2]) \end{aligned}\right\}, \tag{2}$$

where $G_\sigma(\cdot)$ is the kernel function, $\sigma > 0$ is the kernel length parameter, and $E[\cdot]$ is the expectation.

$V_x(t,\tau)$ can be represented by a Fourier series:

$$V_x(t,\tau) = \sum\nolimits_{\alpha} V_x(\alpha,\tau)\exp(j2\pi\alpha t), \tag{3}$$

where $V_x(\alpha,\tau)$ is the CCF [24] of the signal $x(t)$:

$$V_x(\alpha,\tau) = \langle V_x(t,\tau)\exp(-j2\pi\alpha t)\rangle . \tag{4}$$

The Taylor series can be expanded to give:

$$V_x(\alpha,\tau) = \frac{1}{[(2\pi)^{1/2}\sigma]}\left\langle\sum\nolimits_{n=0}^{\infty}\frac{(-1)^n}{(2^n n!\sigma^{2n})}\Omega\right\rangle, \tag{5}$$

where $\Omega = \left[(x(t)-x(t+\tau))^2 + 2j\sigma^2 2\pi\alpha t\right]^n$.

The CCF contains information regarding the second- and higher-order cyclostationary moments. It extends the classic CAF to the analysis of signals contaminated by non-Gaussian noise, such as impulsive noise.

B. DERIVATION OF THE ALGORITHM PROPOSED IN THIS PAPER

From (3) and (4), we get:

$$\begin{aligned} V_x(\alpha,\tau) &= \langle V_x(t,\tau)\exp(-j2\pi\alpha t)\rangle \\ &= \langle E[G_\sigma(x(t)-x(t+\tau))\exp(-j2\pi\alpha t)\rangle . \end{aligned} \tag{6}$$

Substituting the kernel function into (6) gives:

$$\begin{aligned} V_x(\alpha,\tau) = &\frac{1}{\sqrt{2\pi}\sigma} \\ &\times\left\langle\exp\left(-\frac{(x(t)-x(t+\tau))^2}{2\sigma^2}\right)\exp(-j2\pi\alpha t)\right\rangle . \end{aligned} \tag{7}$$

Using a Taylor series only for

$$\exp\left(-\frac{(x(t)-x(t+\tau))^2}{2\sigma^2}\right)$$

Gives:

$$\begin{aligned} & V_x(\alpha,\tau) \\ &= A\left\langle \sum_{n=0}^{\infty} \frac{[-x(t)-x(t+\tau))^2]^n}{(2^n n!\sigma^{2n})} B \right\rangle \\ &= A\left\{ \langle B\rangle - \frac{\left\langle (x(t)-x(t+\tau))^2 B\right\rangle}{2\sigma^2} + \xi_x(\alpha,\tau) \right\} \\ &= A\left\{ \langle B\rangle - \frac{\left\langle \left(x^2(t)-2x(t)x(t+\tau)+x^2(t+\tau)\right) B\right\rangle}{2\sigma^2} \right. \\ &\quad \left. + \xi_x(\alpha,\tau)\right\}, \end{aligned} \tag{8}$$

where $A = 1/\sqrt{2\pi}\sigma$, $B = \exp(-j2\pi\alpha t)$, and

$$\xi_{x(t)}(\alpha,\tau) = \left\langle \sum_{n=2}^{\infty} \frac{[-(x(t)-x(t+\tau))^2]^n}{(2^n n!\sigma^{2n})} B \right\rangle. \tag{9}$$

According to the definition of the CAF:

$$R_{x(t)}(\alpha,\tau) = \langle x(t)x(t+\tau)\exp(-j2\pi\alpha t)\rangle. \tag{10}$$

When $\tau = 0$, $R_{x(t)}(\alpha,0) = \langle x^2(t)\exp(-j2\pi\alpha t)\rangle$.

Equation (8) can be expressed as:

$$\begin{aligned} & V_{x(t)}(\alpha,\tau) = A\left\{L_{x(t)}(\alpha,\tau) + \xi_{x(t)}(\alpha,\tau)\right\}, \\ & L_{x(t)}(\alpha,\tau) = \langle B\rangle \\ &\qquad - \frac{\left(R_{x(t)}(\alpha,0) - 2R_{x(t)}(\alpha,\tau) + R_{x(t+\tau)}(\alpha,0)\right)}{2\sigma^2}. \end{aligned} \tag{11}$$

$L_{x(t)}(\alpha,\tau)$ mainly contains CAF-related information about the second moment. For sufficiently large values of σ, the traditional second-order statistics are preserved, thus minimizing the influence of term $\xi_{x(t)}(\alpha,\tau)$ [24]. $\xi_{x(t)}(\alpha,\tau)$ contains all higher moments ($n \geq 2$) and depends on the coefficient $(-1)^n/\left[\left(2^{1/2}\sigma\right)^{2n} n!\right]$. We can see that as the nuclear length parameter σ grows, the amplitude of the decay of $\xi_{x(t)}(\alpha,\tau)$ is much greater than that of $L_{x(t)}(\alpha,\tau)$.

This paper selects the optimal parameter according to the Silverman rule, one of the most widely used kernel estimation methods, to ensure the effective demodulation of the CCF [22]:

$$\sigma = 0.9MN^{-1/5}, \tag{12}$$

where N is the signal length and M is the minimum of the empirical standard deviation of the signal. When the impulsive noise is not included in the signal or the strength is not strong, $L_{x(t)}(\alpha,\tau)$, which contains the second moment, can effectively demodulate the frequency component. Conversely, if the included impulsive noise is strong, the signal empirical standard deviation (M) becomes larger. It can be seen from (8) and (9) that if $\sigma > 1/2^{1/2}$, then $|L_{x(t)}(\alpha,\tau) \gg \xi_{x(t)}(\alpha,\tau)|$, so:

$$\begin{aligned} & V_{x(t)}(\alpha,\tau) \\ &\approx A\left\{\langle B\rangle - \frac{R_{x(t)}(\alpha,0) - 2R_{x(t)}(\alpha,\tau) + R_{x(t+\tau)}(\alpha,0)}{2\sigma^2}\right\}. \end{aligned} \tag{13}$$

To reduce the amount of calculation, we take the slice at $\tau = 1/f_s$, where f_s is the sampling frequency:

$$\begin{aligned} & V_{x(t)}(\alpha,\frac{1}{f_s}) \\ &\approx A\left\{\langle B\rangle - \frac{R_{x(t)}(\alpha,0) - 2R_{x(t)}(\alpha,\frac{1}{f_s}) + R_{x(t+\frac{1}{f_s})}(\alpha,0)}{2\sigma^2}\right\}. \end{aligned} \tag{14}$$

After normalization, the formula for the NACCF is obtained:

$$VN_{x(t)}(\alpha) = \frac{V_{x(t)}\left(\alpha,\frac{1}{f_s}\right) - V_{x(t)}^{\min}\left(\alpha,\frac{1}{f_s}\right)}{V_{x(t)}^{\max}\left(\alpha,\frac{1}{f_s}\right) - V_{x(t)}^{\min}\left(\alpha,\frac{1}{f_s}\right)}. \tag{15}$$

When the CCF is calculated, its amplitude depends significantly on σ. After normalization, the amplitude of the NACCF is almost unaffected by the kernel function parameters σ.

C. VERIFICATION OF THE PERFORMANCE OF THE ALGORITHM PROPOSED IN THIS PAPER

In this section, simulation signals (example 1) and bearing signals (example 2) are used to illustrate the performance of NACCF.

Example 1 aims to show that NACCF is still effective in the presence of impulse noise. A bearing vibration signal is a cyclically stable signal similar to an amplitude modulated (AM) signal [30]. An AM signal is $x(t) = (1 + b\cos(2\pi f_t t))\cos(2\pi f_z t + \theta)$. Substituting this into (14) gives:

$$\begin{aligned} & V_x(\alpha, 1/f_s) \\ &= \begin{cases} A + C\left\{1 + \frac{b^2}{2} - \cos(2\pi\frac{f_z}{f_s})\left[1 + \frac{C^2}{2}\cos(2\pi\frac{f_t}{f_s})\right]\right\} \\ \qquad \times \alpha = 0; \\ Cb\left[\frac{D}{2} - \cos(2\pi f_z/f_s)\cos(2\pi f_t/f_s)\right] \qquad \alpha = \pm f_t; \\ Cb^2\left[\frac{D}{2} - \cos(2\pi f_z/f_s)\right] \qquad \alpha = \pm 2f_t; \\ Ce^{\pm j2\theta}\left[\frac{(2+b^2)D}{8} - \frac{1}{2} - \frac{b^2}{4}\cos(2\pi f_z/f_s)\right] \quad \alpha = \pm 2f_z; \\ Cbe^{\pm j2\theta}\left[\frac{D}{4} - \frac{1}{2}\cos(2\pi f_t/f_s)\right] \qquad \alpha = \pm(2f_z \pm f_t); \\ Cb^2 e^{\pm j2\theta}\left(\frac{D}{16} - \frac{1}{8}\right) \qquad \alpha = \pm(2f_z \pm 2f_t); \end{cases} \end{aligned} \tag{16}$$

where $C = -1/2\sqrt{2\pi}\sigma^3$ and $D = 1 + e^{j2\pi\alpha\tau}$.

Let $b = 1.5, f_t = 20, f_z = 90, \theta = \pi/6, x1(t) = x(t) + 2 \times \text{wgn}(t)$, and $x2(t) = x(t) + \text{gua}(t) + \text{alpha}^{0.03}(t)$, where $\text{gua}(t)$ is the Gaussian noise and $\text{alpha}^{0.03}(t)$ is the alpha stable distribution signal ($\beta = 0$, $\lambda = 0.03$, and $\alpha_0 = 1$) for simulating the impulsive noise. Fig.1 is a time-domain diagram, CAF slice diagram ($\tau = 0$), and NACCF slice diagram ($\tau = 1/f$) of these three signals $x(t)x1(t)$ and $x2(t)$, respectively.

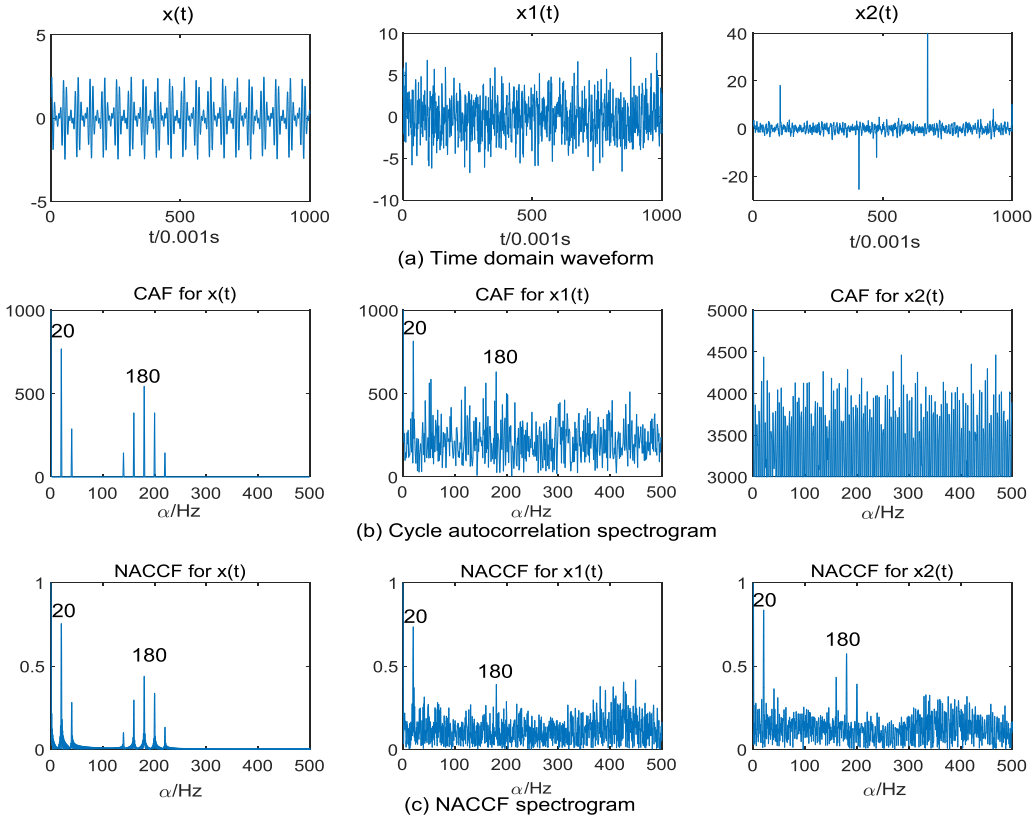


FIGURE 1. CAF and NACCF demodulation of different signals.

It can be seen from Fig. 1 that the NACCF distribution for the signal $x(t)$ is similar to that for CAF and is consistent with (15). Thus, like CCF, NACCF can suppress impulse noise, it can demodulate the modulation frequency and carrier frequency in the signal, and it can also cope with interference from Gaussian noise as well as CAF can.

Bearing signals are cyclostationary like AM signals. If a bearing fails, it will produce a corresponding characteristic fault frequency (f_c), which is not only used as the modulation frequency to modulate the system frequency, but also as the carrier frequency, which is modulated by the rotation frequency [22]. Therefore, in theory, the peak at the low frequency of NACCF will appear at: 1) the characteristic fault frequency (f_c) and its multiples (nf_c), 2) the converted frequency (f_r) and its multiples (nf_r), and 3) also on both sides of multiples of the characteristic fault frequency offset by the same multiple of the rotation frequency ($nf_c \pm nf_r$). As n gets larger, the amplitude gets smaller.

Example 2 used drive-end bearing fault data in deep-groove ball-bearing signals (ball-bearing model 6205-2RS JEM SKF) from the website of Case Western Reserve University Bearing Data Center [31]. The purpose of this example was to demonstrate that NACCF could be used in machine learning. The amplitudes in the NACCF spectrum for the same fault signals are stable even under different loads and different impulse noise backgrounds, whereas the NACCF spectrum for different fault signals shows obvious differences. Fault features that are convenient for clustering regression can be identified in the NACCF spectrum.

The sample includes an inner-ring fault signal and an outer-ring fault signal. The sampling frequency was 12 000 Hz, and the motor speed depended on the load.

The research objects of this paper are single-row angular contact ball bearings. The characteristic fault frequency of this kind of bearing is calculated as follows.

The working rotation frequency is

$$f_r = v/60. \tag{17}$$

The characteristic roll fault frequency is

$$f_{rp} = D\left[1 - (d\cos\beta/D)^2\right] f_r/2d. \tag{18}$$

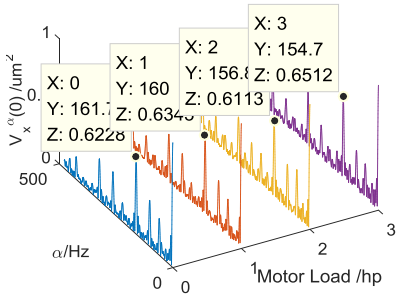
The characteristic inner-ring fault frequency is

$$f_{ci} = Z\,[1 + d\cos\beta/D] f_r/2. \tag{19}$$

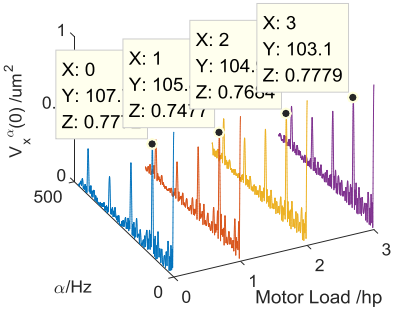
Finally, the characteristic outer-ring fault frequency is

$$f_{co} = Z\,[1 - d\cos\beta/D] f_r/2 \tag{20}$$

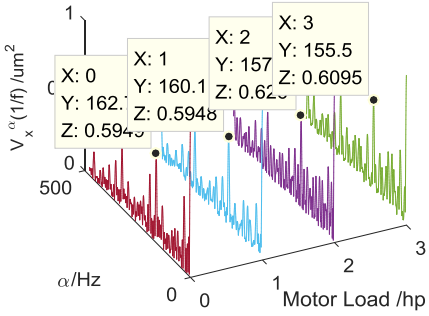
where v is the motor speed, d is the diameter of the scroll, D is the nodal diameter, Z is the number of rolls, and β is the angle of contact. Here, $d = 8$ mm, $D = 44$ mm, $Z = 9$, and $\beta = 0°$. Table 1 was calculated using (17)–(20) for the characteristic bearing frequency.



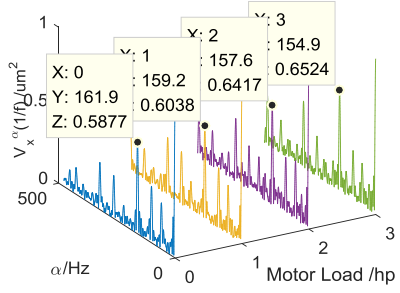
(a) Mild inner-ring fault signal (no impulse noise)



(b) Mild outer-ring fault signal (no impulse noise)



(c) Severe inner-ring fault signal (no impulse noise)



(d) Mild inner-ring fault signal (plus impulse noise)

FIGURE 2. Different degrees of failure.

NACCF demodulation was performed on signals for bearing mode loads of 0, 1, 2, and 3 hp (units of horsepower), as shown in Fig. 2. The sample length of each signal was 2048.

Figure 2(a) shows the NACCF spectrum for a mild inner-ring fault signal under different loads. The x-axis is the load, the y-axis is the cycle frequency, and the z-axis is the NACCF amplitude. When the load is 0 hp, a peak appears

TABLE 1. Characteristic frequencies of bearings under different loads.

Motor load (hp)	v (rpm)	f_r (Hz)	f_{ci} (Hz)	f_{co} (Hz)
0	1797	30.0	162	107
1	1772	29.5	160	106
2	1750	29.2	157	105
3	1730	28.8	155	103

at or near the characteristic frequency of the inner-ring fault ($f_{ci}^{j=0} = 162$, $2f_{ci}^{j=0} = 324$, and $f_{ci}^{j=0} \pm f_r^{j=0} = 162 \pm 30$). Similarly, when the loads are 1, 2, or 3 hp, there are peaks at the characteristic frequency corresponding to the load (f_{ci}^j, $2f_{ci}^j$, $f_{ci}^j \pm f_r^j$, $j = 1$, 2, or 3). Moreover, the amplitude of the peak of the corresponding characteristic fault frequency under different loads is stable. The peak range of the inner-ring fault signal in the figure is 0.62 ± 0.05 at the characteristic inner-ring frequency, and for the outer-ring fault signal, it is 0.75 ± 0.05 at the characteristic outer-ring frequency. Thus, regardless of the load that the bearing is under, for the same failure mode signal, NACCF will provide similar failure information.

These test data contain signals for: 1) the same degree of failure but different faulty parts (Fig. 2(a) and 2(b)), 2) the same faulty part but for different degrees of failure (Fig. 2(a) and 2(c)), and 3) the same faulty part and the same degree of fault but different degrees of impact load (Fig. 2(a) and 2(d)). The characteristics of the NACCF spectra in Fig. 2(a), 2(b), and 2(c) are obvious and have their own unique and stable laws, indicating that NACCF spectra can be used to distinguish between different faulty parts and different degrees of fault. Furthermore, there is almost no difference between the information in Fig. 2(a) and 2(d), which further shows that NACCF has stable and good impulsive noise resistance.

In general, the simulated AM signal and bearing fault data show that NACCF demodulation can indicate the location of a fault and degree of damage of the bearing. It can cope with Gaussian noise and has good impulsive noise resistance and stability. Moreover, the NACCF amplitude for a fault frequency under the same mode is stable, and the NACCF for a fault frequency under different modes shows obvious differences. Therefore, NACCF is suitable as a new feature domain for feature extraction and intelligent detection of bearing failures.

III. VMD WITH OPTIMIZED PARAMETERS

NACCF has good demodulation performance, but it may still fail under actual working conditions with high-intensity noise. Thus, it is necessary to preprocess the signal to reduce the noise. This section first briefly introduces the concept, advantages, and parameters of VMD. It describes the problem, then introduces the GMPSO parameter optimization algorithm and defines CIE. The function was constructed, and

a simulation was run to assess its performance. CIE-GMPSO is proposed to optimize adaptively the VMD parameters. Finally, VMD was used to decompose signals. Sensitive components were selected according to CIE to get the denoising signals.

A. VMD

VMD is essentially a variational problem construction and solution process. VMD uses a decomposed IMF component as an AM-FM signal. If we assume that the original signal can be decomposed into IMF components, an expression for the kth IMF component ($k \in \{1, 2, \ldots, K\}$) is [28]

$$u_k(t) = A_k(t)\cos[\phi_k(t)], \tag{21}$$

where $A_k(t)$ is the instantaneous amplitude of $u_k(t)$, the phase $\phi_k(t)$ is a non-decreasing function, and $d\phi_k(t)/dt = \omega_k(t)$ is the instantaneous frequency. $A_k(t)$ and $\omega_k(t)$ change slowly as $\phi_k(t)$ changes.

The VMD algorithm gets rid of the cyclic screening and stripping signal processing of the EMD algorithm. It transfers the signal decomposition to the variational framework and realizes signal adaptation by searching for the optimal solution of the constrained VMD. The frequency center and bandwidth of each IMF component are continuously updated in each process in VMD. Finally, the adaptive division of the signal band can be completed according to the frequency-domain characteristics of the actual signal, and several narrowband IMF components can be obtained. If we assume that the original signal has been decomposed into K_v IMF components, the corresponding constrained variational model expression is

$$\begin{left}\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ s.t. \sum_k u_k(t) = x(t)\end{left}, \tag{22}$$

where $\{\omega_k\}$ is the center frequency corresponding to the IMF component, $\delta(t) + j/\pi t$ is the unilateral spectrum of $u_k(t)$, and $x(t)$ is the original signal. To obtain the optimal solution of the above constrained variational problem, a quadratic penalty factor α_v and Lagrange operator $\lambda_v(t)$ are introduced to transform the problem into the following unconstrained variational problem:

$$\begin{aligned} &L(\{u_k\}, \{\omega_k\}, \lambda_v) \\ &= \alpha_v \sum\nolimits_k \|\partial_t [(\delta(t) + j/\pi t)\, u_k(t)] \exp(-j\omega_k t)\|_2^2 \\ &\quad + \left\langle \lambda_v(t), x(t) - \sum\nolimits_k u_k(t) \right\rangle + \left\| x(t) - \sum\nolimits_k u_k(t) \right\|_2^2. \end{aligned} \tag{23}$$

The saddle point of the extended Lagrange expression is calculated via the alternating direction method of the multiplier algorithm. In detail, the steps are as follows:

1) Initialize $\{u_k\}, \{\omega_k\}, \lambda_v$, and $n = 0$.
2) Execution cycle: $n = n + 1$.
3) For all $\omega \geq 0$, update $\{u_k\}, \{\omega_k\}$, and λ_v.
4) Repeat steps (2) through (3) until the iteration stop condition is satisfied.
5) Stop the iterations and obtain the IMF components.

The choice of the second penalty factor α_v and number of components K_v affects the decomposition [28], and the parameter selection is irregular. To select the optimal parameter combination, the VMD method can be used to extract the rich features in the signal. This paper uses GMPSO to optimize the parameters.

B. GMPSO BASED ON CIE

This section defines CIE-GMPSO. In a D-dimensional search space, the population X is composed of m particles: $X = [x_1, x_2, \ldots, x_m]$. The position of each particle in the search space can be represented by a D-dimensional vector: $x_i = [x_{i1}, x_{i2}, \ldots, x_{iD}]$. D is the number of parameters to be optimized. The speed of the ith particle $v_i = [v_{i1}, v_{i2}, \ldots, v_{iD}]$. The local extremum of the particle is $p_i = [p_{i1}, p_{i2}, \ldots, p_{iD}]$. The global extreme of this generation of the population is $G_1 = [g_1, g_2, \ldots, g_D]$, and the suboptimal value is $G_2 = [g'_1, g'_2, \ldots, g'_D]$. The maximum optimal retention algebra of an individual is maxAge, and the probability of a mutation is q. To prevent particles from falling into a local optimum, it is necessary to record the optimal retention algebra of individual particles during each iteration. Individual local extreme values and global extreme values are used to update the position and velocity of the next generation:

$$v_i^{n+1} = \omega v_i^n + c_1\eta(p_i - x_i^n) + c_2\eta(G - x_i^n), \tag{24}$$

$$x_i^{n+1} = x_i^n + v_i^{n+1}, \tag{25}$$

where ω is the inertial weight, η is a random number in [0,1], and c_1 and c_2 are learning factors for the local search ability and the global search ability, respectively. The number of iterations n is consistent with the above definition. The inertial weight in the current iteration is determined with the linear decreasing weight method [32]:

$$\omega = \omega_{\max} - (\omega_{\max} - \omega_{\min})\frac{n}{n_{\max}}, \tag{26}$$

where $\omega_{\max}$ and $\omega_{\min}$ are the maximum and minimum inertial weights, respectively, n is the current iteration, and $n_{\max}$ is the defined maximum number of iterations. When the individual optimal algebra reaches maxAge, a genetic mutation updates the particle's position and velocity so that it jumps out of the local optimum. Each time a particle moves to a new position, its fitness value is calculated, and when it reaches the optimal value, the program terminates and outputs the optimal parameter.

When applying GMPSO, it is necessary to select an evaluation criterion for VMD as the fitness function. Generally speaking, the higher the similarity between the component and the original signal without noise, the higher the value, the better the VMD. However, under actual working conditions, pure fault signals without noise are not available, and the COR between noisy signal components and noiseless signals cannot be obtained, so we need to find a function that

can replace the similarity as the fitness function. Information entropy measures the amount of information in a signal and the uniformity of the probability distribution [14]. The more ordered the signal, the lower the information entropy. Guiji and Xiaolong [33] based the envelope entropy (EE) on information entropy. The envelope signal obtained after signal demodulation is processed into a sequence of probability distributions p_j, and the EEs E_p of zero mean signals $x(j)$ $(j = 1, 2, \ldots, N)$:

$$\left.\begin{aligned} p_j &= a(j)/\textstyle\sum_{j=1}^{N} a(j) \\ E_P &= -\textstyle\sum_{f=1}^{f_s} P_f \log(10) P_f \end{aligned}\right\} \tag{27}$$

where p_j is the normalized form of $a(j)$, which is the envelope signal obtained after the signal $x(j)$ is demodulated by a Hilbert transform. The EE reflects the sparse characteristics of the original signal. The EE is calculated for all IMFs obtained by VMD. The IMF with the smallest entropy is the best component with rich fault feature information in the classified groups. In this case, the EE is suitable as a fitness function.

When the signal is in the background of impulsive noise, the EE becomes unstable or even invalid. In this paper, the CIE is proposed based on information entropy and NACCF. Compared with the EE, the CIE effectively reflects the sparse characteristics of the signal under interference from impulsive noise. After the NACCF signal is demodulated, the data are processed into a probability distribution sequence P_f $(f = 1, 2, \ldots, f_s)$, where f_s is the sampling frequency of the signal $x(j)$ $(j = 1, 2, \ldots, N)$. The CIE (E_{ccf}) of $x(j)$ is defined as:

$$\left.\begin{aligned} E_{ccf} &= \frac{[E_p]}{\sum_{j=1}^{N} |x(j)|}, \\ E_P &= -\textstyle\sum_{f=1}^{f_s} P_f \log(10) P_f, \\ P_f &= \frac{|V(f)|}{\sum_{f=1}^{f_s} |V(f)|}, \\ V(f) &= v_{x(j)}^{1/f_s}(f) - \frac{1}{f_s} \textstyle\sum_{f=1}^{f_s} v_{x(j)}^{1/f_s}(f) \end{aligned}\right\} \tag{28}$$

where $v_{x(t)}^{1/f_s}(f)$ is the NACCF of the signal after the Hilbert transform and $V(f)$ is the de-averaged $v_{x(t)}^{1/f_s}(f)$. Using P_f reduces the influence of the different amplitudes of the IMF components in the NACCF. E_P is calculated with the formula for information entropy. E_{ccf} takes into account the energy contribution of the IMF in the time domain and reflects the cyclic stability of the original signal. Within a certain range of accuracy, the lower the entropy, the more cyclically ordered the signal.

C. SIMULATION EFFECT OF VMD BASED ON CIE-GMPSO

As an example, consider the signal for a mild outer-ring fault numbered 130 in the drive-end bearing fault data provided by Case Western Reserve University [31]. This signal is denoted as $\mathrm{sig}_1(t)$. The noise content of this signal is low, so here, $\mathrm{sig}_1(t)$ is regarded as the original signal without noise. To simulate a strong impulsive noise background, we set:

$$\mathrm{sig}_2(t) = \mathrm{sig}_1(t) + 2\mathrm{gau}(t) + \mathrm{alpha}^{0.01}(t),$$

where $\mathrm{gua}(t)$ is Gaussian noise and $\mathrm{alpha}^{0.01}(t)$ is an alpha-stable distribution ($\beta = 0, \lambda = 0.01$, and $\alpha_0 = 1$). In processing $sig_2(t)$ with VMD to get IMF_i $(i = 1, 2, \ldots, K_v)$, we assume the range of K_v is 3–12, and $\alpha_v = 2000$. The CIEs and EEs of IMFs decomposed from $\mathrm{sig}_2(t)$ were calculated as well as the COR between the IMFs and the original signal $\mathrm{sig}_1(t)$. Some of the data are shown in Table 2. The COR reflects the correlation between the noisy signal IMF and the original signal without additional noise. The IMF corresponding to the maximum COR has the least noise and is regarded as the best component. In theory, the smaller the CIE, the larger the corresponding COR. In reality, the COR cannot be calculated, so the IMF corresponding to the smallest CIE is the best component.

It can be seen from Table 2 that within a certain error range, the CIE and COR decrease monotonously, which is basically consistent with the theory, whereas the EE becomes invalid due to the impulsive noise so that it does not accurately reflect the best component. The optimal component IMF 4 is for $K_v = 4$. The COR between this component and the original signal is 0.560. The corresponding CIE is 0.787, which is the lowest CIE in the table. We select the two IMF reconstructions with the smallest CIE as the denoised sig_3. The COR for sig_1 and the original signal is 0.352, and the COR for sig_3 and the original signal after noise reduction increased to 0.597. The NACCF of the original signal sig_1, the noise-including signal sig_2, and the denoised sig_3 are shown in Fig. 3.

The NACCF of the original outer-ring fault signal in Fig. 3 clearly shows that the signal is one to four times the characteristic frequency f_o of the outer-ring fault. After adding Gaussian and alpha-stable impulsive noise, NACCF for sig_2 has some robustness against Gauss noise and impulsive noise. It can be seen that there is an obvious peak at f_o. At the double and triple frequencies ($2f_o$ and $3f_o$), the information is still submerged by noise. After the noisy signal is decomposed by parameter-optimized VMD, the two components with the smallest CIE are selected to reconstruct the denoising signal sig3. The NACCF spectrum of sig3 is similar to the original signal NACCF, and information is clearly visible at one to four times the characteristic frequency (f_o, $2f_o$, $3f_o$, and $4f_o$) of the outer-ring fault. The effect of VMD noise elimination effect based on CIE-GMPSO is obvious.

After repeated verification, the results of the analysis can be summarized as follows. Both the CIE and the COR indicate the optimal IMF. The CIE and COR for different K values within a certain error range have a monotonic relation: the smaller the CIE is for a component, the higher the correlation with the original signal. Therefore, the CIE can be used effectively as the fitness function in GMPSO to evaluate the

TABLE 2. CIE and COR of IMFs with different K values.

K_v		IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6
3	COR	0.014	0.022	0.531			
	CIE	1.015	1.088	0.795			
	EE	4.617	5.150	5.245			
4	COR	0.010	0.011	0.243	0.560		
	CIE	1.031	1.105	0.898	0.787		
	EE	4.458	4.978	4.852	5.125		
5	COR	0.009	0.008	0.011	0.255	0.513	
	CIE	1.054	1.168	1.160	0.901	0.821	
	EE	4.316	4.896	4.677	4.953	4.023	
6	COR	0.001	0.012	0.008	0.011	0.265	0.512
	CIE	1.297	1.169	1.201	1.171	0.899	0.825
	EE	4.112	4.235	4.658	4.667	4.856	3.935

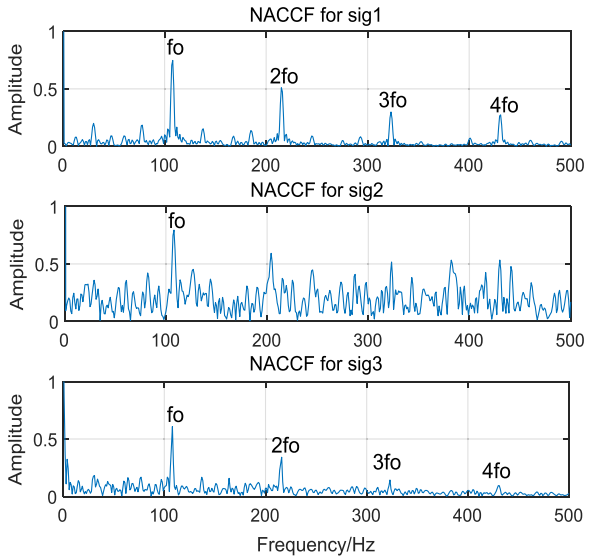


FIGURE 3. Comparison of NACCF demodulation before and after signal denoising.

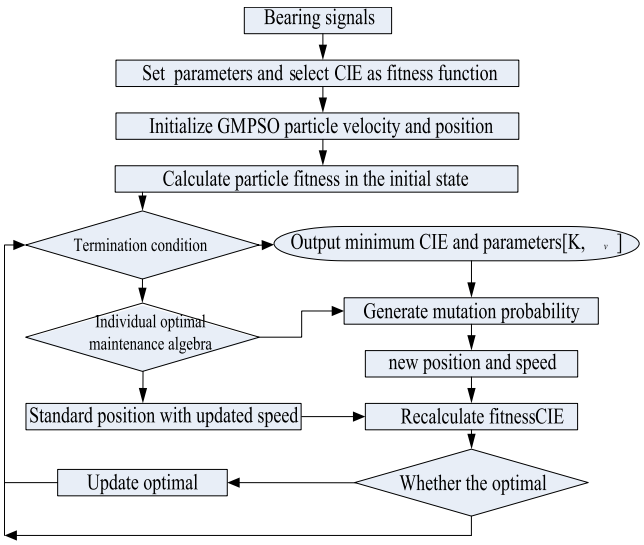


FIGURE 4. Flow chart for optimization of VMD parameters using CIE-GMPSO.

decomposition of VMD. It can also be used as ae component sensitivity coefficient to select components for reconstructing denoising signals. A flow chart for GMPSO based on CIE is shown in Fig. 4.

IV. THE METHOD PROPOSED IN THIS PAPER

Multi-domain feature extraction is helpful for improving the accuracy and stability of classification [16]. This section introduces the proposed MKELM based on NACCF and GMPSO-VMD. First, the feature extraction function is constructed according to the low-frequency demodulation characteristics of NACCF in a bearing signal to extract time-frequency features. Then, the VMD component of the time-domain features are extracted by SVD and the same features are normalized and integrated into the input to the nuclear limit learning machine.

A. FEATURE EXTRACTION WITH NACCF

According to the low-frequency characteristics from NACCF for the rotation frequency and the four characteristic frequencies (as mentioned in Section II.C) and considering the monotonic relation between load and rotational speed, the structural characteristic functions are as follows.

For features 1–5,

$$C(j) = \sum_{\alpha=f_j-1}^{f_j+1} V^{1/f_s}(\alpha) + \sum_{\alpha=f_j+f_r-1}^{f_j+f_r+1} V^{1/f_s}(\alpha) + \sum_{\alpha=f_j-f_r-1}^{f_j-f_r+1} V^{1/f_s}(\alpha) + \sum_{\alpha=2f_j-1}^{2f_j+1} V^{1/f_s}(\alpha), \quad (j = 1, 2, 3, 4, 5) \tag{29}$$

where $V^{1/f_s}(\alpha)$ is the NACCF of the signal, and f_j $(j = 1, 2, 3, 4, 5)$ are the characteristic inner-ring frequency, the characteristic outer-ring frequency, the characteristic

ball- bearing frequency, the characteristic cage frequency, and the rotation frequency, respectively. The frequencies are calculated according to the rotational speed to reduce the influence of load changes.

Feature 6 is the peak of the low frequency of the NACCF. Feature 7 is the cycle frequency corresponding to the peak of the low frequency of the NACCF. These are expressed as follows:

$$[\text{value},\text{index}] = \max_{\alpha\in(0,3f_{inner})}(VN_{x(t)}(\alpha))$$
$$C(6) = \text{value}, \quad C(7) = \text{index} \tag{30}$$

where f_{inner} is the characteristic inner-ring frequency.

Feature 8 is the CIE of the signal. The formula is expressed as in [27].

B. FEATURE EXTRACTION WITH VMD AND SVD

According to matrix theory, the singular value is an inherent feature of a matrix, and it has good stability. That is, if a matrix element changes a little, the singular value of the matrix changes only a little, so it is often used to extract signal features [34]. In this paper, the VMD with optimized parameters can decompose the signal into K_v IMFs adaptively according to the frequency-domain characteristics of the actual signals. The K_v IMFs are formed into an initial feature vector matrix, and this matrix undergoes SVD. The K_v values of different signals after optimization are different. The minimum value is $(K_v)_{\min}$, and the first $(K_v)_{\min}$ singular values of each matrix are normalized and taken as the time-domain features of the signal.

Some commonly used intelligent failure detection algorithms encompass back-propagation neural networks, support vector machines, ELMs, and so on. KELM [35] is an improved ELM proposed by Huang. It combines the kernel function in the support vector machine with an ELM. The learning rate, measurement, and generalization ability of KELM have been greatly improved. In this paper, KELM is used to classify and regress the above NACCF features and VMD singular values to obtain a multi-domain failure detection model.

The steps in the failure detection method proposed in this paper are as follows:

Step 1: Use GMPSO to find the best parameter for VMD.

Step 2: Decompose the signal with VMD and get the singular value of the matrix formed by the IMFs.

Step 3: Reconstruct the first two components with the smallest CIEs to obtain the denoised signal and extract the NACCF feature from the denoised signal.

Step 4: The normalized VMD-SV feature vector and NACCF form a multi-domain feature input to KELM for learning and failure detection.

The flow chart is shown in Fig. 5.

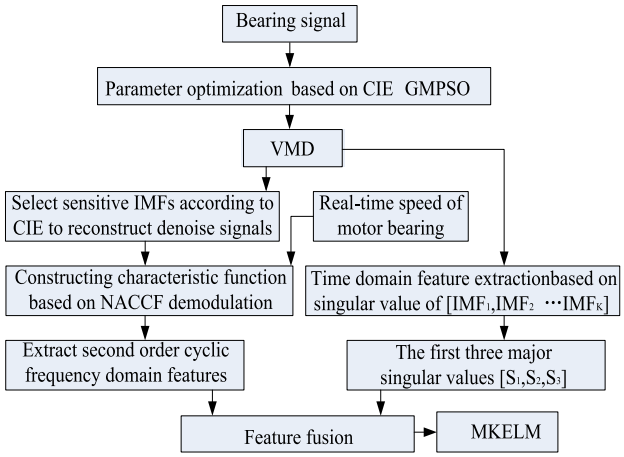


FIGURE 5. KELM based on VMD and NACCF.

TABLE 3. Parameter settings of the feature extraction methods.

Method	Parameters
Proposed method	K_v and α_v are determined by CIE-GMPSO optimization. The kernel length parameter in NACCF is calculated by the formula mentioned above: $\sigma = 0.9MN^{-1/5}$.
EMD-SE	When calculating the sample entropy, the capacity coefficient m = 1 and the threshold r is 0.2 times the standard deviation of the signal.
TD-KELM	Ten commonly used time-domain statistics: standard deviation, variance, skewness index, kurtosis index, peak-to-peak value, peak value, mean square amplitude, average amplitude, root square amplitude, and waveform index.
VMD-SV KELM	Use values based on experience: $K_v = 6$ and $\alpha_v = 2000$.
NACCF-KELM	The kernel length parameter in NACCF is calculated by the formula mentioned above: $\sigma = 0.9MN^{-1/5}$.

V. FAILURE DETECTION APPLICATIONS

Three cases are used to validate the effectiveness of the proposed method, and the following four methods are compared with the method proposed in this paper:

1) EMD-SE KELM is KELM based on sample entropy from an EMD [36].
2) TD-KELM is a KELM using classic time-domain (TD) statistics [37].
3) VMD-SV KELM is a KELM based on singular values after decomposition by traditional VMD [38].
4) NACCF-KELM is a KELM based on features extracted by NACCF.

Details of the parameters for all the feature extraction methods are listed in Table 3.

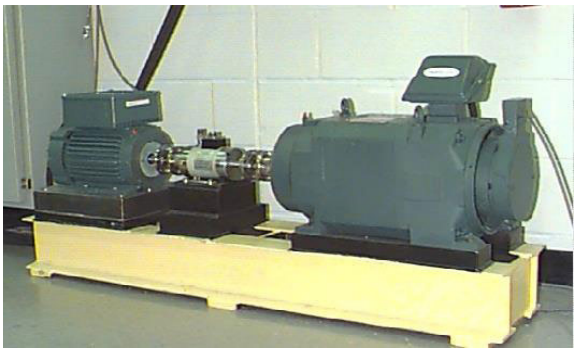


FIGURE 6. Rolling bearing test rig for case 1.

A. CASE STUDY 1

The fault data used to verify the proposed method is a drive-end bearing signal from the website of Case Western Reserve University Bearing Data Center [31]. The experimental signal noise was small, so this section adds different degrees of Gaussian noise and impulsive noise to the original signal to analyze the anti- Gaussian noise and anti- impulsive noise performance of this method.

The bearing model is a drive-end bearing (6205-2RS JEM SKF). As shown in Fig. 6, the test bed consists of a 2-hp motor (left), a torque transducer and encoder (center), a dynamometer (right), and control electronics (not shown). The sampling frequency was 12 000 Hz.

In this case, five fault patterns under different working conditions were simulated:

Pattern 1: Normal bearing

Pattern 2: Mild inner-ring fault with fault diameters of 0.18 mm

Pattern 3: Severe inner-ring fault with fault diameters of 0.53 mm

Pattern 4: Mild double-ball fault with fault diameters of 0.18 mm

Pattern 5: Mild outer-ring fault with fault diameters of 0.18 mm

Each pattern has four working conditions, corresponding to a motor load of 0 to 3 hp. The corresponding relation between motor speed and load is shown in Table 1. We used 15 signals for each working condition, which means there were 60 signals per pattern. The sample length of each signal is 2048. The characteristic frequencies of the bearings in this case under different loads are shown in Table 1. The NACCF spectra for the same fault signals under different loads are shown in Fig. 2. The figure shows that regardless of the load borne by the bearing, the NACCF amplitude at the characteristic fault frequencies for a failure pattern signal are similar. Inputting the load as a parameter into (28)–(30) gives stable cyclic frequency domain features.

So first, according to the rotational speed in Table 1 and the process shown in Fig. 5, the characteristics of all the original signal samples in these five patterns were calculated.

Since this experiment does not include a cage fault signal, the characteristic C4 at the frequency of cage failure does not need to be calculated. The remaining seven NACCF characteristics are shown in Fig. 7(a). Along the abscissa, 1 to 7 indicate C1, C2, C5, C6, C7, C8, and C3, respectively. In Fig. 7(b), along the abscissa, 1 to 3 represent the first three singular values of the VMD matrix (S1, S2, and S3).

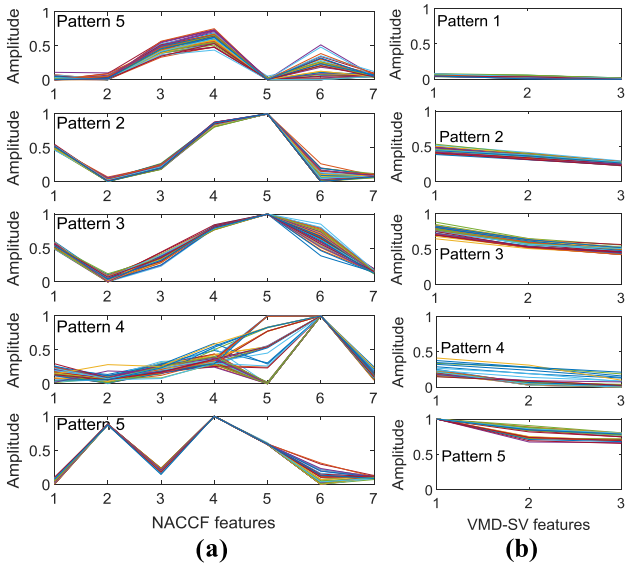


FIGURE 7. Feature distributions of original signal samples. Based on (a) NACCF and (b) VMD-SV.

To study the impulsive noise resistance of the proposed method, we added strong impulsive noise noise (alpha-stable distribution noise with $\beta = 0$, $\lambda = 0.015$, and $\alpha_0 = 1$) to the five pattern samples. We then found the feature values of all the samples for these five patterns. The distributions of the NACCF and VMD-SV features are shown in Fig. 8.

Figure 8(a) demonstrates that NACCF-based feature extraction can distinguish different pattern features under strong impulsive noise. In contrast, the features in Fig. 8(b) for the five modes extracted with VMD-SV are seriously confused with the impulsive noise background. Compared with the faulty bearing signal, the cyclic characteristic of the normal-bearing signal is the weakest. In this example, pattern 4 is a weak double-ball fault and its cyclic spectrum is more complicated and irregular than that of a single-ball fault. It is more easily confused with the normal bearing pattern (pattern 1) and the cyclic characteristics have a large dispersion, as shown in Fig. 8.

By using in KELM the NACCF features, VMD-SV features, and the combined multi-domain features (the method in this paper) shown in Fig. 8, we formed three classifiers: NACCF-KELM, VMD-SV KELM, and VMD-NACCF-KELM. We used 500 test samples (100 for each pattern) for diagnosis. The detection accuracies of the three methods are 79.6%, 62.2%, and 87.8%, respectively. The classification confusion matrix of the method in this paper is shown in Table 4.

The accuracies of patterns 1 to 5 are 89%, 87%, 92%, 76%, and 95%. Pattern 4 is a mild double-ball fault. Its cycle

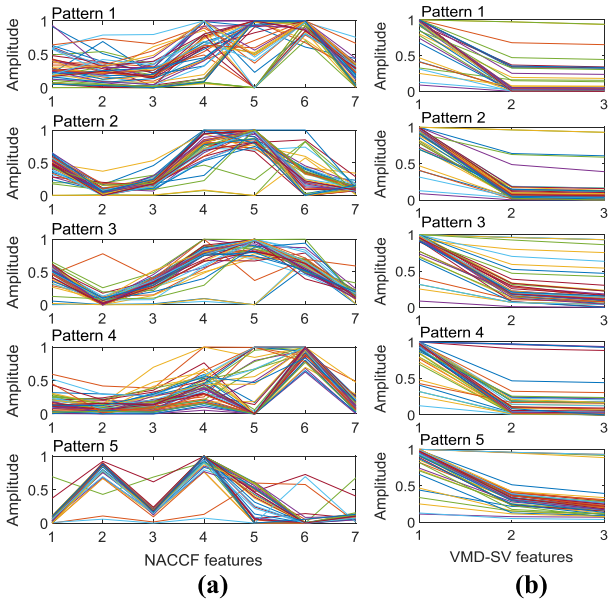


FIGURE 8. Feature distributions of signal samples with a background of shock noise. Based on (a) NACCF and (b) VMD-SV.

TABLE 4. Confusion matrix for the method in this paper (Accuracy = 87.8%).

Real pattern	Predicted pattern				
	1	2	3	4	5
1	89	0	2	**9**	0
2	6	87	5	2	0
3	6	0	92	2	0
4	**16**	0	8	76	0
5	3	1	1	0	95

characteristics are relatively weak, as shown in Fig. 7(a). So, it is most affected by impulsive noise interference and has the lowest detection accuracy. Overall, with a background of impulsive noise, VMD-SV-KELM almost fails and the proposed NACCF feature has obvious advantages in impulsive noise resistance.

To analyze further the impulsive noise resistance difference between this method and the other methods, we added an alpha-stable distribution signal with $\alpha = 1$ and $\lambda = 0.005$, $\lambda = 0.01$, and $\lambda = 0.015$ in sequence. A comparison of the results of the diagnostic accuracy for this method and the other methods as an average of 10 tests is shown in Fig. 9.

As shown in Fig. 9, when the impulse noise was 0, there is no difference in the failure detection accuracy of the other four methods, all reaching an accuracy of more than 99%, except EMD-SE-KELM. With an increase of the impulse noise, the accuracies of the methods with NACCF features (NACCF-KELM and the proposed method) are significantly higher than other methods. Thus, the proposed method can suppress impulsive noise.

Moreover, to verify the anti-interference ability of this method, Gaussian noise generated by $0.3\text{gau}(t)$, $0.6\text{gau}(t)$,

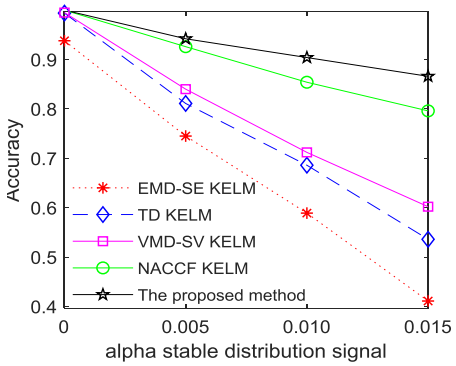


FIGURE 9. Detection accuracy of different methods under impulse noise.

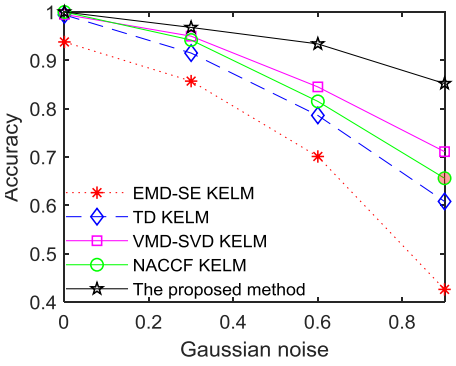


FIGURE 10. Detection accuracy of different methods under Gaussian noise.

and $0.9\text{gau}(t)$ was added in sequence. A comparison of the results of the detection accuracy of this method and the other methods is shown in Fig. 10.

In Fig. 10, as the interference from Gaussian noise increases, it can be clearly seen that the anti-Gaussian noise ability of VMD-SV with a single domain feature is better than the NACCF frequency domain feature. Under Gaussian noise with a factor of 0.9 times, VMD-SV and NACCF with MKELM have the highest detection accuracies, whereas KELM based on traditional time-domain statistics and EMD-SE features have the lowest detection accuracies.

This case shows that NACCF features and VMD-SV features have good shock resistance and anti-Gaussian noise capabilities. The MKELM classifier with values from NACCF or singular features from GMPSO-VMD as feature quantities achieved good classification results. The proposed method is more stable and has better anti-noise performance.

B. CASE STUDY 2

Case 2 is an experiment using data for a bearing damaged by an electric current in the shaft under a background of white noise interference. The analyzed signal contains information about the damage site and different damage methods. Case 2 shows the different advantages of the MD-SV and NACCF features proposed in this paper and their complementary ability to detect defects. Thus, multi-domain joint

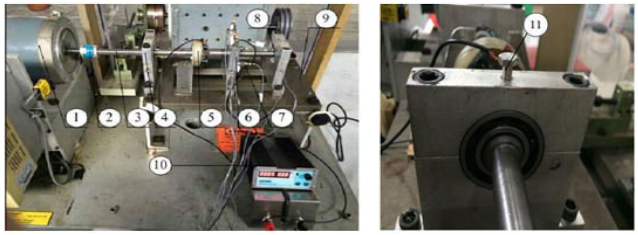


FIGURE 11. Test rig of current damage of bearing. 1 motor, 2 insulated coupling, 3 spindle, 4 support bearing seat, 5 carbon brush, 6 test bearing seat, 7 vibration acceleration sensor, 8 insulated bearing, 9 base, 10 adjustable dc switching power supply, and 11 conductive bolt.

learning can improve detection accuracy and stability. In addition, this case shows that the method proposed in this paper can be generalized to bearing failure detection for different damage mechanisms.

The test rig for simulating damage to a bearing caused by an electric current in the shaft is shown in Fig. 11. The insulated bearings are ceramic. They are installed at both ends of the supporting bearing seat. The bearings block the current and support the spindle. The insulated coupling and insulated bearings together protect the motor. The carbon brushes are fixed on an insulated stand to prevent the entire experimental bench from being charged.

When the voltage applied to a bearing is greater than the threshold voltage of the lubricating oil film, the oil film breaks down, resulting in a current along the shaft. There will then be spark discharges on the inner and outer raceways, causing electrical corrosion to the bearing.

The signal acquisition system was the Pulse data acquisition system produced by B&K. The test bearing was a deep-groove ball bearing, model 6205EKA. The parameters were as follows. The bearing inner diameter was 25 mm, the outer diameter was 52 mm, the rolling element diameter was 7.925 mm, the pitch circle diameter was 39 mm, and the number of rolling elements $n = 9$. The rotation frequency was 20 Hz and the sampling frequency 16 384 Hz. By substituting the above parameters into the theoretical equations, we found that the characteristic frequency of an outer-ring fault is 72 Hz, of an inner-ring fault is 108 Hz, and of a rolling element fault is 47 Hz.

In this case, all patterns were tested under the same conditions: motor load was 500 N and motor speed was 1200 r/m. The five fault patterns are as follows:

1) Severe outer-ring shaft-current damage (The shaft was corroded by the current for 100 hours.)

2) Flaking fault of bearing on outer ring

3) Mild inner-ring shaft-current damage (The shaft was corroded by the current for 50 hours.)

4) Flaking fault of bearing on inner ring

5) Normal bearing

The four faults are shown in Fig. 12.

The inner and outer raceways of a normal bearing are smooth. After being damaged by a shaft current, the loss of material is evenly distributed along the raceway, as shown

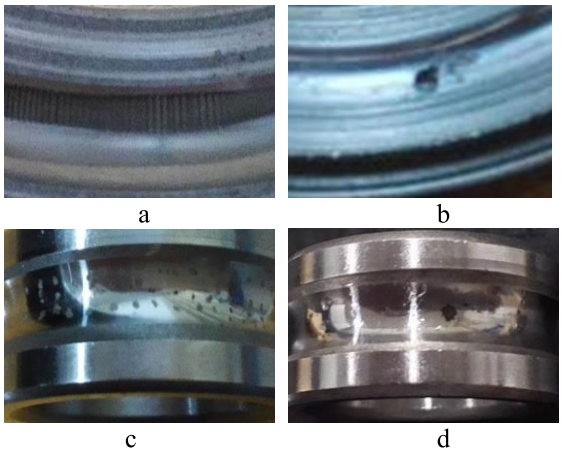


FIGURE 12. Some simulated faults in bearings. (a) Severe outer-ring shaft-current damage. (b) Flaking fault of bearing on outer ring. (c) Mild inner-ring shaft-current damage. (d) Flaking fault of bearing on inner ring.

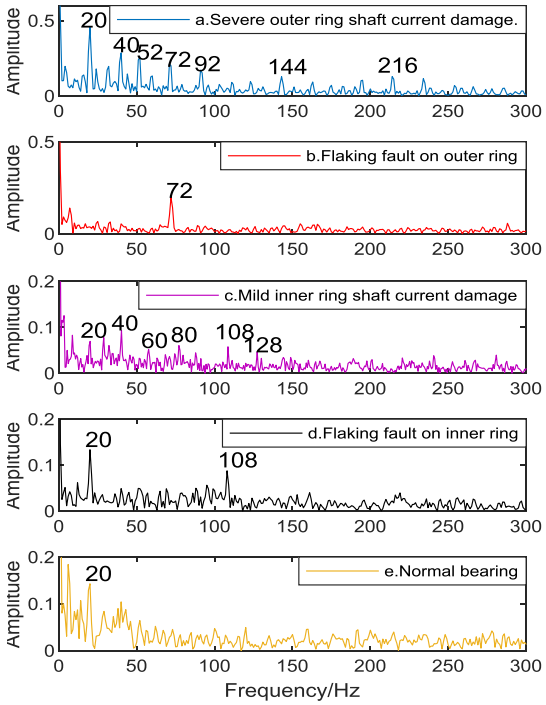


FIGURE 13. NACCF spectra of five signals.

in Fig. 12(a) and 12(c). The loss can even look like a washboard, as shown in Fig. 12(a). Bearings with flaking faults often have a single obvious dent due to the long-term alternating load and impact force, as shown in Fig. 12(b) and 12(d).

We used 250 samples for each run, and each pattern had 50 samples, of which the first 40 were used as a training set and the remaining 10 signals were used as a test set. Each sample length was 4096. The experimental data include bearing fault signals due to different damage mechanisms at the same location, which are difficult to diagnose. The spectra from the NACCF demodulation of the signals for these five patterns are shown in Fig. 13.

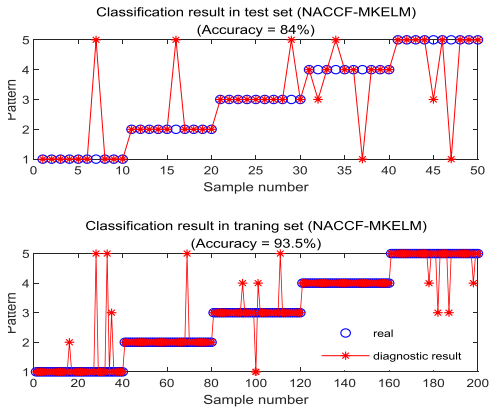


FIGURE 14. Classification results of NACCF-KELM.

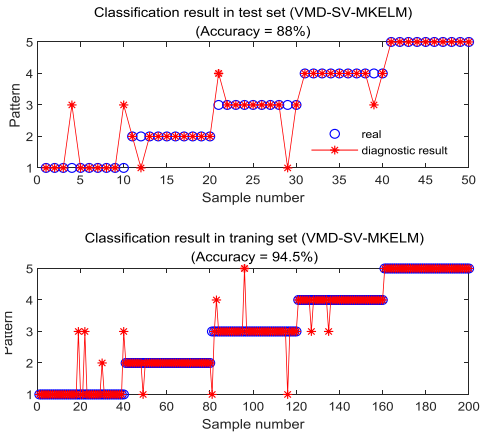


FIGURE 15. Classification results of GMPSO-VMD-SV KELM.

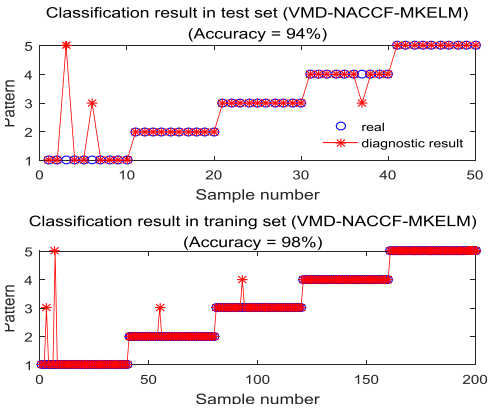


FIGURE 16. Classification results of the proposed method.

NACCF uses the CCF, which is a generalization of the CAF [23]. It retains the demodulation performance of CAF, as long as the outer ring fails, regardless of the damage mechanism. The corresponding characteristic outer-ring fault frequency f_o(72 Hz) appears in the low-frequency NACCF information, as shown in Fig. 13(a) and 13(b). Similarly, the NACCF spectra of all inner-ring fault signals have information at the characteristic frequency f_i (108 Hz) of an inner-ring fault, as shown in Fig. 13(c) and 13(d). This demonstrates once again that NACCF can diagnose well the location of a fault.

Secondly, for a fault at the same location, different damage mechanisms have completely different characteristics. Shaft-current damage shows as multiple-point faults, while flaking faults are generally single-point faults, as shown in Fig. 12. The NACCF spectrum can well reflect this characteristic. In Fig. 13, the low-frequency NACCF information for a flaking fault has a peak only at the characteristic fault frequency, either f_o(72 Hz) in Fig. 13(b) or f_i(108 Hz) in Fig. 13(d). A NACCF spectrum of shaft-current damage signals (Fig. 13(a) and 13(c)) has more peaks due to the multiple faults. These can occur at double the characteristic fault frequency of the signal, at its side frequencies, at the rotation frequency, and even at triple the characteristic fault frequency and its side frequencies. Figure 13(a) and 13(b) are both NACCF spectra of outer-ring fault signals. The difference is that Fig. 13(a) not only has a peak at the outer-ring fault frequency f_o (72 Hz), it also has peaks at the rotation frequency f_r (20 Hz), at double the rotation frequency $2f_r$ (40 Hz), at the side frequencies of an outer-ring fault $f_o \pm f_r$ (72 ± 20 Hz), at double the characteristic frequency of the outer ring $2f_o$ (144 Hz), at its side frequencies $2f_o \pm f_r$(144 ± 20Hz), at triple the characteristic frequency of the outer ring $3f_o$ (216 Hz), and at its side frequencies $3f_o \pm f_r$(216 ± 20Hz). In the same way, the NACCF spectrum in Fig. 13(a) of the signal for inner-ring shaft-current damage also shows more peak frequencies than Fig. 13(d), although it is a mild fault with smaller peak amplitudes.

To evaluate fully the different advantages of the GMPSO-VMD-SV and NACCF features, we followed the steps described in Section IV to obtain the NACCF features and GMPSO-VMD singular values for the samples. We inputted the two sets of data into KELM to form NACCF-KELM and GMPSO-VMD-SV KELM. The NACCF features and GMPSO-VMD singular values were fused to form the multi-domain extreme learning machine (VMD-NACCF-MKELM) proposed in this paper. Then these three classifiers were used to diagnose faults in the samples. In this case, 40 training samples and 10 prediction samples were used. The diagnostic results for the three methods are shown in Figs. 14, 15, and 16, respectively.

We used different samples to test the repeatability. The confusion matrices of the results of 10 tests (500 test samples in total) of the three methods are shown in Tables 5, 6, and 7, respectively. The misjudgment number for NACCF-KELM was 79, and its accuracy was 84.2% [= (500 – 79) / 500], which is slightly lower than the accuracy of GMPSO-VMD-SV KELM (87.4%). This means that in an environment with only white noise interference, compared with the CIE-based GMPSO-VMD-SV KELM, NACCF-KELM does not have an advantage. The confusion matrices of the two methods (Tables 5 and 6) are not the same. The highest detection rate by NACCF-KELM was for pattern 2 (96%), and the

TABLE 5. Confusion Matrix of NACCF-KELM (Accuracy = 84.2%).

Real pattern	Predicted pattern				
	1	2	3	4	5
1	83	0	5	0	12
2	0	96	3	0	1
3	1	0	**77**	7	15
4	0	2	8	90	0
5	4	6	13	2	**75**

TABLE 6. Confusion Matrix of GMPSO-VMD-SV KELM (Accuracy = 87.4%).

Real pattern	Predicted pattern				
	1	2	3	4	5
1	**79**	4	14	0	3
2	7	90	3	0	0
3	13	1	**82**	4	0
4	0	2	9	89	0
5	1	0	0	2	97

TABLE 7. Confusion Matrix of method proposed in this paper (Accuracy = 95.0%).

Real pattern	Predicted pattern				
	1	2	3	4	5
1	96	0	3	0	1
2	0	98	1	1	0
3	2	1	**90**	4	1
4	0	1	6	93	0
5	1	0	1	0	98

TABLE 8. Number of confused patterns for three methods.

NACCF-KELM		GMPSO-VMD SV KELM		Method of this paper	
Confusing patterns	Number	Confusing patterns	Number	Confusing patterns	Number
3 and 5	**28**	**1 and 3**	**27**	3 and 4	10
1 and 5	16	3 and 4	13	**1 and 3**	**6**
3 and 4	15	1 and 2	11	**3 and 5**	**3**
Else	21	Else	12	Else	6
Total	79	Total	63	Total	25

lowest for patterns 3 and 5 (77% and 75%, respectively). The highest detection rate for GMPSO-VMD-SV KELM was for pattern 5 (97%), and the lowest for patterns 1 and 3 (79% and 82%, respectively).

Based on the data behind Tables 5, 6, and 7, we obtained the number of confused patterns for the three methods shown in Table 8.

Note that pattern 3 is for a multi-point mild inner-ring shaft-current damage and its characteristic NACCF peak is relatively small. It is easily confused with a normal bearing (as shown in Fig. 13). Table 8 shows that NACCF-KELM

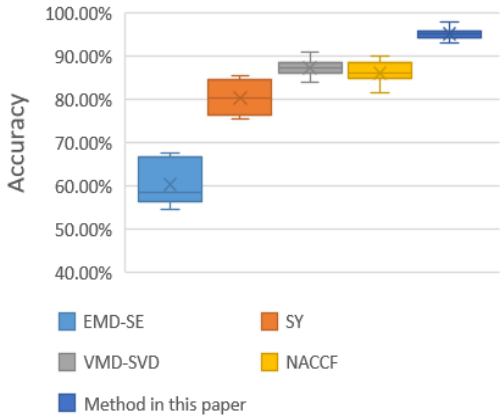


FIGURE 17. Results of 10 tests for the five methods.

is most likely to confuse patterns 3 and 5. There were 15 instances when real pattern 3 was misjudged as pattern 5, and 13 when real pattern 5 was misjudged as pattern 3. The percentage of these misjudgments was 35.44% (= 28 / 79). Patterns 1 and 5 are also prone to confusion. Of the total, there were 20.25% (= 16 / 79).

The most confusing patterns for GMPSO-VMD-SV KELM were 3 and 1. There were 27 confusions, 42.86% of the total. Patterns 3 and 4 were the next most likely to be confused, with 13 or 20.63% of the total. This also proves that the features of different domains can reflect different information in the signal. NACCF features can distinguish outer-ring and inner-ring faults as well as shaft-current damage. These features make up for the deficiencies of using VMD singular values, which can better distinguish between mild inner-ring shaft-current damage and normal bearings, which are difficult to distinguish with NACCF features. Multiple domain features can be extracted from a smaller sample to give more complete information and more accurate predictions.

For the five methods listed in Table 3, we calculated the detection accuracy for the 10 test sets. The resulting box plots are shown in Fig. 17. By applying the MKELM proposed in this paper, the accuracy for the test set improved to 95.0%.

Thus, in the absence of impulsive noise interference, NACCF features are not more advantageous than VMD singular values, but the fusion of the two methods makes the detection results more accurate and more stable.

C. CASE STUDY 3

Case 3 studies motor bearing faults with both ordinary Gaussian white noise and irregular impulsive noise. This example shows that the method in this paper can maintain high detection accuracy in mixed interference in working conditions and has practical applicability.

The experimental equipment provided by Taiyuan University of Technology, China, includes a motor (Y160M2-8), a piezoelectric acceleration sensor coupler (KISTLER5134), a 16-channel signal collector (DEWEtron), a magnetic powder brake, a gearbox, and a loading mechanism. The test rig and

TABLE 9. Means of 10 feature vectors.

Pattern	C1	C2	C3	C5	C6	C7	C8	S1	S2	S3
1	0.181	0.193	0.218	0.424	0.552	0.451	0.372	0.691	0.208	0.097
2	0.327	0.151	0.163	0.218	0.731	0.749	0.156	0.451	0.342	0.108
3	0.242	0.197	0.331	0.267	0.434	0.279	0.225	0.562	0.195	0.114
4	0.367	0.232	0.381	0.362	0.626	0.423	0.430	0.717	0.314	0.152
5	0.148	0.876	0.152	0.251	0.869	0.572	0.149	0.887	0.652	0.237

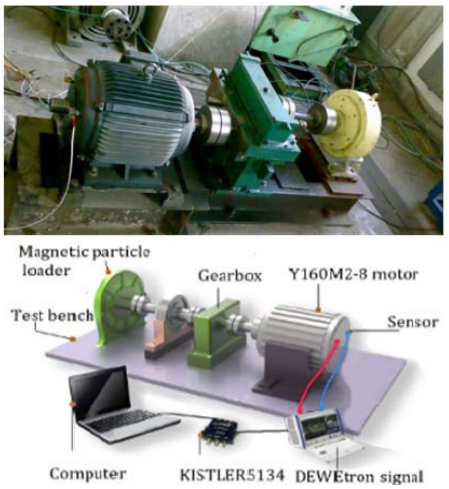


FIGURE 18. Photograph (top) of rolling bearing test rig and a schematic diagram (bottom) for case 3.

TABLE 10. Confusion matrix of method proposed in this paper (Accuracy = 89.2%).

Real pattern	Predicted pattern				
	1	2	3	4	5
1	90	2	3	4	1
2	1	94	1	4	0
3	3	2	87	8	0
4	6	9	7	**78**	0
5	2	0	0	1	**97**

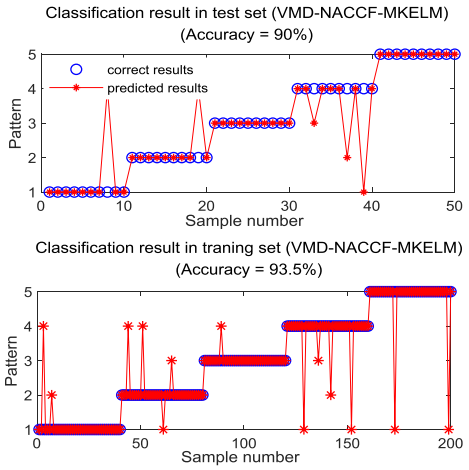


FIGURE 19. Detection results of the proposed method.

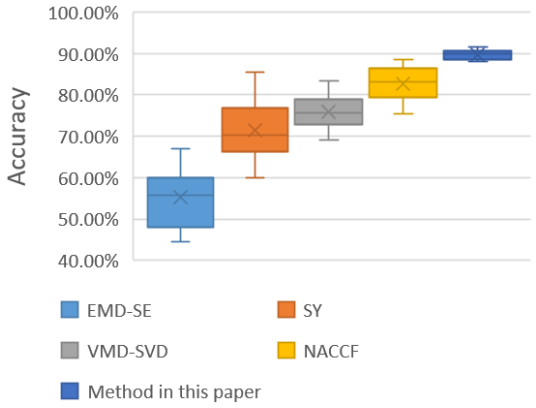


FIGURE 20. Results of 10 tests for five methods.

a schematic diagram are shown in Fig. 18. In this case, the test object was the rolling bearing inside the running motor. The sampling frequency was 5000 Hz. The vibrations of the bearing cannot be directly measured, but they are transmitted directly to the sensor installed on the end cover. Thus, the interference in this case is more complicated than in case 1. The rolling bearing (model 6309) was in the motor. The loader was used to simulate a frequently changing load. A random pulse was added to the collected signal to simulate a shock signal. The inner ring of the motor bearing and the motor shaft have an interference fit. During the experiments, the outer ring was fixed and the spindle speed was measured with an eddy current displacement sensor.

In this case, five failure patterns were simulated under different working conditions: 1) normal bearing, 2) inner-ring failure, 3) ball failure, 4) ball and inner-ring failure, and 5) outer-ring failure. For each pattern, there were 50 samples, of which the first 40 were used as the training set and the remaining 10 signals were used as the test set. After the sample signal was decomposed by VMD, the two IMFs with the smallest CIE were selected to reconstruct the denoising signal, and the eight features of the NACCF were obtained from the sample signal after denoising. The first five statistics with the same dimension were normalized and the last three were normalized separately to get the matched features (C1–C3 and C5–C8). This case does not include a cage fault signal, so the characteristic C4 at the frequency of a cage failure does not need to be calculated. Then, SVD was performed on the matrix composed of the component IMFs in the VMD

signal. The most important of the first three singular values were normalized to give the time-domain features S1–S3. The averages of the 10 features from the NACCF frequency and time domains are shown in Table 9.

Under impulsive noise interference, due to the impulsive noise resistance of CCF, the cyclic features extracted from the NACCF spectrum are more obvious than those extracted from other traditional fields, as shown in Table 9.

We used 10 test sets (500 samples in total) for diagnosis. The classification confusion matrix of the method proposed in this paper is shown in Table 10. The highest accuracy rate is for the outer-ring failure (pattern 5), followed by inner-ring failure (pattern 2). Ball and inner-ring failure (pattern 4) has the lowest accuracy rate. This pattern is easily confused with inner-ring failure (pattern 2), ball failure (pattern 3), and a normal bearing (pattern 1).

For the five methods listed in Table 3, we calculated the detection accuracy of the 10 test sets, as shown in Fig. 20. It can be seen that, with interference, the method proposed in this paper is more accurate and more stable than the other methods.

VI. CONCLUSION

A failure detection method for motor bearings based on VMD and NACCF is introduced in this paper. The proposed method not only can effectively cope with Gaussian noise but also impulse noise. The NACCF is a simplified algorithm suitable for machine learning that evolved from the CCF. The Silverman rule is used to select the optimal kernel size. The method retains the effective mechanism for eliminating impulse noise in the CCF kernel function. The CIE is proposed based on information entropy and NACCF low-frequency demodulation features. The CIE can effectively reflect the cyclostationarity of a signal. In the VMD of a second-order cyclostationary signal, the CIE has a monotonic relation with the COR within a certain accuracy range. Therefore, the selection requirements of the fitness function are met. A GMPSO based on CIE as the fitness function was used to optimize the VMD parameters. After the VMD, salient IMFs were selected to obtain a denoising signal based on the CIE. The simulation results show that the proposed method can achieve adaptive parameter optimization and effectively remove Gaussian noise. Finally, feature functions were constructed to extract the second-order cyclic frequency-domain features with NACCF and for frequency conversion. Moreover, our approach combines VMD-SV time-domain features into MKELM. The frequency conversion used in the function eliminates the interference to failure detection caused by different working conditions. The simulation and tests on motor bearings in working conditions demonstrate that the proposed method is more precise at dealing with noise and more robust for different working loads.

Our subsequent research will mainly focus on combining NACCF with deep learning methods, using deep algorithms such as denoising self-encoding to extract features from the cyclic frequency domain automatically. In addition, since the algorithm proposed in this paper can process second-order cyclostationary signals, it should be useful for failure detection for other types of rotating machinery, such as gears and propellers. Thus, this work may help in the development of deep network classifiers with good generalizability.

ABBREVIATIONS

AM	Amplitude modulated
CAF	Cyclic autocorrelation function
CCF	Cyclic correntropy function
CIE	Cyclic information entropy
COR	Correlation Coefficient
EE	Envelope entropy
ELM	Extreme learning machine
EMD	Empirical mode decomposition
GMPSO	Genetic mutation for particle swarm optimization
IMF	Intrinsic mode function
KELM	Kernel extreme learning machine
MKELM	Multi-domain kernel extreme learning machine
NACCF	Normalized approximation algorithm for CCF
SVD	Singular-value decomposition
VMD	Variational modal decomposition
α	Cyclic frequency
τ	Time delay
σ	Kernel length parameter
α_0	Characteristic exponent
λ	Scale parameter
β	Skewness parameter
α_v	Second penalty factor
K_v	VMD decomposition number

REFERENCES

[1] P. Zhang, Y. Du, T. G. Habetler, and B. Lu, "A survey of condition monitoring and protection methods for medium-voltage induction motors," *IEEE Trans. Ind. Appl.*, vol. 47, no. 1, pp. 34–46, Jan./Feb. 2011, doi: 10.1109/TIA.2010.2090839.

[2] Z. S. Sun and J. K. J. K. Zhu, "Method and prevention of coal mine mechanical bearing lose efficacy analysis," *Coal Mine Mach.*, vol. 9, no. 9, pp. 150–151, 2005.
[3] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zi, J. Yuan, B. Chen, and Z. He, "Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 1–35, Mar. 2016.
[4] E. Alickovic, J. Kevric, and A. Subasi, "Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction," *Biomed. Signal Process. Control*, vol. 39, pp. 94–102, Jan. 2018.
[5] J. S. Smith, "The local mean decomposition and its application to EEG perception data," *J. Roy. Soc. Interface*, vol. 2, no. 5, pp. 443–454, Dec. 2005.
[6] W. Zeng, J. Yuan, C. Yuan, Q. Wang, F. Liu, and Y. Wang, "Classification of myocardial infarction based on hybrid feature extraction and artificial intelligence tools by adopting tunable-Q wavelet transform (TQWT), variational mode decomposition (VMD) and neural networks," *Artif. Intell. Med.*, vol. 106, Jun. 2020, Art. no. 101848.
[7] Y. Tian, J. Ma, C. Lu, and Z. Wang, "Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine," *Mechanism Mach. Theory*, vol. 90, pp. 175–186, Aug. 2015.
[8] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
[9] X. Zhang, Q. Miao, H. Zhang, and L. Wang, "A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery," *Mech. Syst. Signal Process.*, vol. 108, pp. 58–72, Aug. 2018.
[10] P. Zhang and J. Yan, "Fault feature extraction of wind turbine rolling bearing based on PSO-VMD," in *Proc. Chin. Intell. Autom. Conf.* Singapore: Springer, 2019, pp. 638–646.
[11] Z.-H. Zhan, J. Zhang, Y. Li, and H. S.-H. Chung, "Adaptive particle swarm optimization," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 6, pp. 1362–1381, Dec. 2009.
[12] A. A. A. Esmin and S. Matwin, "HPSOM: A hybrid particle swarm optimization algorithm with genetic mutation," *Int. J. Innov. Comput. Inform. Control*, vol. 9, no. 5, pp. 1919–1934, 2013.
[13] S.-Y. Ho, H.-S. Lin, W.-H. Liauh, and S.-J. Ho, "OPSO: Orthogonal particle swarm optimization and its application to task assignment problems," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 38, no. 2, pp. 288–298, Mar. 2008.
[14] Z. Huo, M. Martinez-Garcia, Y. Zhang, R. Yan, and L. Shu, "Entropy measures in machine fault diagnosis: Insights and applications," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2607–2620, Jun. 2020.
[15] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis," *Neurocomputing*, vol. 168, pp. 119–127, Nov. 2015.
[16] S. Pang, X. Yang, X. Zhang, and X. Lin, "Fault diagnosis of rotating machinery with ensemble kernel extreme learning machine based on fused multi-domain features," *ISA Trans.*, vol. 98, pp. 320–337, Mar. 2020.
[17] Z.-Y. Wang, C. Lu, and B. Zhou, "Fault diagnosis for rotary machinery with selective ensemble neural networks," *Mech. Syst. Signal Process.*, vol. 113, pp. 112–130, Dec. 2018.
[18] M. Cerrada, R. Sánchez, D. Cabrera, G. Zurita, and C. Li, "Multi-stage feature selection by using genetic algorithms for fault diagnosis in gearboxes based on vibration signal," *Sensors*, vol. 15, no. 9, pp. 23903–23926, Sep. 2015.
[19] S. Kang, D. Ma, Y. Wang, C. Lan, Q. Chen, and V. I. Mikulovich, "Method of assessing the state of a rolling bearing based on the relative compensation distance of multiple-domain features and locally linear embedding," *Mech. Syst. Signal Process.*, vol. 86, pp. 40–57, Mar. 2017.
[20] L. Li and L. Qu, "Cyclic statistics in rolling bearing diagnosis," *J. Sound Vib.*, vol. 267, no. 2, pp. 253–265, Oct. 2003.
[21] J. Niu, K. Li, W. Jiang, X. Li, G. Kuang, and H. Zhu, "A new method of micro-motion parameters estimation based on cyclic autocorrelation function," *Sci. China Inf. Sci.*, vol. 56, no. 10, pp. 1–11, Oct. 2013.
[22] X. Zhao, Y. Qin, C. He, L. Jia, and L. Kou, "Rolling element bearing fault diagnosis under impulsive noise environment based on cyclic correntropy spectrum," *Entropy*, vol. 21, no. 1, p. 50, Jan. 2019.
[23] A. I. R. Fontes, J. B. A. Rego, A. D. M. Martins, L. F. Q. Silveira, and J. C. Principe, "Cyclostationary correntropy: Definition and applications," *Expert Syst. Appl.*, vol. 69, pp. 110–117, Mar. 2017.
[24] W. Liu, P. P. Pokharel, and J. C. Príncipe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
[25] T. Liu, T. Qiu, and S. Luan, "Cyclic frequency estimation by compressed cyclic correntropy spectrum in impulsive noise," *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 888–892, Jun. 2019.
[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[27] Z. Huo, Y. Zhang, L. Shu, and M. Gallimore, "A new bearing fault diagnosis method based on Fine-to-Coarse multiscale permutation entropy, Laplacian score and SVM," *IEEE Access*, vol. 7, pp. 17050–17066, 2019.
[28] K. Li, L. Su, J. Wu, H. Wang, and P. Chen, "A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine," *Appl. Sci.*, vol. 7, no. 10, p. 1004, Sep. 2017.
[29] W. A. Gardner and C. M. Spooner, "Signal interception: Performance advantages of cyclic-feature detectors," *IEEE Trans. Commun.*, vol. 40, no. 1, pp. 149–159, Jan. 1992.
[30] Z. Fuchang, "Research on fault diagnosis method of rolling element bearing based on cyclostationary signal processing," M.S. thesis, Shanghai Jiaotong Univ., Shanghai, China, 2006, p. 155.
[31] *Case Western Reserve University Bearing Data Center*. Accessed: Dec. 28, 2018. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home
[32] Y. Gao and Z. Ren, "Adaptive particle swarm optimization algorithm with genetic mutation operation," in *Proc. 3rd Int. Conf. Natural Comput. (ICNC)*, Haikou, China, 2007, pp. 211–215.
[33] T. Guiji and W. Xiaolong, "Parameter optimized variational mode decomposition method with application to incipient fault diagnosis of rolling bearing," *J. Xi'an Jiaotong Univ.*, vol. 49, no. 5, pp. 73–82, 2015.
[34] R.-R. Ma and Z.-J. Bai, "A structure-preserving one-sided Jacobi method for computing the SVD of a quaternion matrix," *Appl. Numer. Math.*, vol. 147, pp. 101–117, Jan. 2020.
[35] G.-B. Huang, "An insight into extreme learning machines: Random neurons, random features and kernels," *Cognit. Comput.*, vol. 6, no. 3, pp. 376–390, Sep. 2014.
[36] X. Q. Zhang, J. Liang, X. Zhang, F. Zhang, L. Zhang, and B. Xu, "Combined model for ultra short-term wind power prediction based on sample entropy and extreme learning machine," *Proc. CSEE*, vol. 33, no. 25, pp. 33–40, 2013.
[37] O. O. Ogundile, A. M. Usman, O. P. Babalola, and D. J. J. Versfeld, "A hidden Markov model with selective time domain feature extraction to detect inshore Bryde's whale short pulse calls," *Ecol. Informat.*, vol. 57, May 2020, Art. no. 101087.
[38] M. A. Zengqiang, J. Zhang, A. Zhang, and W. Ruan, "Fault feature extraction of rolling bearings based on VMD-SVD joint de-noising and FSWT," *J. Vib. Shock*, vol. 37, no. 17, pp. 210–217, 2018.

XIAOHUI WANG received the B.S. and M.S. degrees in mechanical and electrical engineering from the Taiyuan University of Technology, Taiyuan, China, in 2006 and 2009, respectively. She is currently an Experimenter with the School of Mechanical and Electrical Engineering, Lingnan Normal University, China. Her research interests include pattern recognition, machine learning, mechanical equipment health monitoring, and fault diagnosis.

GUANGZHOU SUI received the B.S. degree in mechanical engineering from Liaoning Technical University, Fuxin, Liaoning, China, in 2003, and the M.S. degree in pattern recognition and intelligence systems from the Dalian University of Technology, Dalian, Liaoning, China, in 2008. He is currently pursuing the Ph.D. degree in mechanical engineering with the Taipei University of Technology, Taiwan. From 2017 to 2018, he was a Visiting Scholar with the Institute of Engineering, University of Lincoln, Lincoln, U.K. He is currently a Lecturer at Lingnan Normal University. His research interests include mechanical fault diagnosis, signal processing, and ultrasound technology.

JIAWEI XIANG (Member, IEEE) received the Ph.D. degree in mechanical engineering from the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China, in 2006. He is currently a Doctor, a Professor, a master's Supervisor, and a Humboldt Scholar with Wenzhou University. He has presided more than 30 vertical and horizontal research projects. He has published more than 100 research articles in academic journals at home and abroad, and more than 90 articles included in SCI. His research interests include fault diagnosis, monitoring and diagnosis software, instrument development, and so on.

GUANGBIN WANG is currently a Scientific Researcher with the College of Mechanical Engineering, Lingnan Normal University, China. He has hosted two national natural science funds, ten scientific research projects in Hunan Province, and more than 20 scientific and technological projects from enterprise and other agency. He has published more than 50 articles in journals. He holds more than 30 national patents. His main research interests include mechanical system dynamics, mechanical equipment health monitoring, and fault diagnosis.

ZHIQIANG HUO received the B.S. and M.S. degrees from the China University of Geosciences, Beijing, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the University of Lincoln, Lincoln, U.K. His research interests include time series complexity analysis, pattern recognition, machine learning, and wireless sensor networks towards intelligent data-driven fault diagnosis of industrial systems. He received the INISCOM 2017 Best Paper Award. He served as the co-chairs for international conferences/workshops, such as AINIS, in 2015 and 2016, and CollaborateCom, in 2017.

ZHEN HUANG received the Ph.D. degree in instrumentation science and technology from the Harbin Institute of Technology, Harbin, China, in 2014. Her research interests include precision measurement, including selfmixing interference and ultraprecision detection.