Research Article

Intelligent Roller Bearing Fault Diagnosis in Industrial Internet of Things

Ji Xu, Hong Zhou, and Yanjun Fang

Department of Automation, Wuhan University, Wuhan 430072, China

Correspondence should be addressed to Hong Zhou; hzhouwuhee@whu.edu.cn

Received 26 November 2021; Accepted 13 January 2022; Published 30 March 2022

1. Introduction

With the rapid development of Internet of things (IoT) [1, 2] and Industry 4.0 [3], there are increasingly massive real-time data from various types of mechanical equipment [4]. The availability of these data that contain abundant information about machine health has attracted more and more enterprises’ attention. It has been proved that large volume, high velocity, and diversity mechanical big data are the major properties of mechanical big data [5, 6]. Effective feature extraction from these data and accurate machinery health state evaluation with ever-accelerated updating of schemes have become hot research issues in the prognostic and health management systems in the era of industrial IoT [7].

For roller bearing, a well-known mechanical component most extensively used in the industry, its operating status directly affects the operation of the entire machinery and equipment. Roller bearing failure is an important factor leading to mechanical equipment failure, so timely detection and fault diagnosis are of great significance, and analysing the vibration signal collected by the sensors to determine its failure is a commonly adopted scheme [8]. As the roller bearings are usually working in the vibration source environment, with the destruction of complex forms, the vibration signal is typically nonlinear and nonstationary. How to excavate the fault feature based on the vibration signals has been a research hotspot [9, 10].

The widely used signal decomposing schemes include wavelet analysis [11, 12] and EMD [13]. However, the wavelet transformation is essentially nonadaptive due to the configuration of wavelet basis and decomposing layers. In EMD, the complicated signal can be self-adaptively decomposed into some IMFs with a residual component. Nevertheless, mode mixing is a stumbling block in EMD. To deal with this challenge, EEMD is proposed [14], in which a complex signal can be disintegrated into IMFs in terms of...
the local time-scale characteristic of the signal. In recent, EEMD has been extensively employed in fault detection [15]. Using the EMD or EEMD schemes for fault feature extraction has been widely concerned. However, these two schemes have some disadvantages. The recursive mode decomposition in EMD/EEMD will propagate the envelope estimation error continuously. The signal contains no noise or intermittent signal, which leads to the decomposition of the mode mixing. Although the white noise scheme was used to suppress the mode mixing, the scheme needs to be carried out several hundreds of times of EMD/EEMD operation and will break out the signal composed of more than a real component. Moreover, EMD and EEMD cannot be separated correctly from the close frequencies, and it is quite challenging to select the suitable number of IMFs when applying EMD and EEMD.

To address the abovementioned problems, a neoteric scheme of signal decomposition estimation called VMD was proposed [16–18]. The whole framework of the alternating direction multiplier scheme was resorted to consecutively refresh the modes and their centre frequencies and gently demodulate the modes into the corresponding base frequency bands. Ultimately, each mode and corresponding centre frequencies were abstracted together. In comparison with the recursive filter pattern of EMD and EEMD, VMD converts the signal into variational and non-recursive decomposition patterns, and it consists of adaptive Wiener filter groups in nature. VMD can isolate two pure harmonic signals with similar frequencies [16–18]. The IMFs obtained by EEMD denote the natural oscillatory pattern inserted into the signals, where the entropy values of every IMF were often abstracted as a feature to discover the properties of the vibration signals [19, 20].

Entropy was one of the schemes to evaluate the time-series complexity. S. M. Pincus proposed the approximate entropy (AE) [21]. However, the length of the time series makes a critical influence on the performance of AE scheme. Hence, the value of AE is conformably lower than the intended one and fails relative coherence particularly when the data length is short. Based on AE, SE was proposed to deal with this challenge [22]. It has the advantages of short data, stable and low noise and interference capacity, and good consistency in the region of large parameter range. It can be noticed that the irregularity of time series can be only reflected by AE and SE individually. When the roller bearings fail, not only the frequencies but also the corresponding complexity of vibration signals has tremendous deviations. Accordingly, ME can be viewed as a property index for fault diagnosis [23]. Considering the MSE, the feature of the vibration signals can be abstracted under all kinds of conditions, and the eigenvectors were counted as the input of adaptive neuro-fuzzy inference system (ANFIS) for roller bearing fault recognition [22]. Moreover, the SE, to some extent, was undefined as no standard vectors, which were in accordance with one another. Undefined or imprecise SE leads to the degradation of authenticity of MSE algorithm. To overcome this drawback, the RCMMSE was revolved by Zhang et al. [24] to overcome these challenges. It is demonstrated that RCMMSE can not only ascend the precision of entropy assessment but also descend the probability of inducing undefined entropy.

However, after EEMD and VMD, a necessary step is how to select the number of modes. For example, the number of patterns was usually determined by the correlation coefficient and mutual information scheme between each component and the original signals [18, 25]. In addition, MSE and RCMMSE can only get the single-channel signal. Therefore, it cannot precisely display the overall signal information. The multivariate multiscale sample entropy (MMSE) was a scheme that gets different time series and conducts various embedding aspects, delay time, and amplitude ranges of data channels in a strict way. Hence, it can directly analyse multichannel data [26]. Therefore, a scheme that relied on VMD and RCMMSE was proposed in this study to get the vibration feature of roller bearings.

With the advent of computer techniques, a lot of fault recognition algorithms, such as SVM, RF, and PNN [27, 28], have been broadly applied to fault diagnosis. Therefore, the SVM, RF, and PNN were resorted to fulfill the fault recognition.

As mentioned above, a scheme based on RCMMSE/VMD and VMD is designed. At first, the vibration signals were disintegrated into series IMF/BL-IMF modes by the EEMD/VMD. Secondly, the RCMMSE was utilized to figure out the entropy values. Finally, the SVM/RF/PNN models were utilized to achieve the fault recognition.

The rest of the study is organized as follows. It is shown that the review of VMD and RCMMSE schemes, respectively, is presented in Section 2. The comparison of MSE/RCMMSE and EEMD/VMD is presented in Section 3. Section 4 gives the procedures of the proposed scheme, experimental data sources, and parameter selection. Section 5 provides the experimental validation. Conclusions are given in Section 6.

2. Theoretical Framework of VMD and RCMMSE Models

2.1. Basic Principle of Variational Mode Decomposition

The VMD process is divided into establishment and solution of variational constraint problems. Assuming there is a limited bandwidth in each mode, the variational problem is formulated as $k$ mode functions $u_k(t)$, which could minimize the estimated bandwidth. The summation of each mode is constrained to be corresponded to the input signal $f$. In particular, the structure of this problem could be divided into the following steps.

2.1.1. Variational Problem Formulation

① Hilbert transform: the analytic signal of every mode function is obtained, whose purpose is to get its spectrum:

\[
\left[ \hat{\delta}(t) + \frac{j}{\pi t} \right] \ast u_k(t).
\]  

(1)

② The centre frequency $e^{-j\omega_k t}$ of each modal signal is estimated. The spectrum of each mode is modified to be adaptive to its own baseband.
\[ [\delta(t) + \frac{j}{\pi t}] * u_k(t) ] e^{-j\omega_k t}. \]  

The \( L^2 \) is derived, and it denotes the signal gradient, and the signal’s bandwidth is obtained. Therefore, constrained variational questions can be obtained as follows:

\[
\begin{align*}
\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \| \partial_t (\delta(t) + \frac{j}{\pi t} \mu_k(t)) e^{-j\omega_k t} \|^2 \right\} \\
\text{s.t.} \sum_k \mu_k = f
\end{align*}
\]

where \( \{\mu\} \) is the decomposition of the \( k \) mode components, \( \{\mu\} = \{\mu_1, \mu_2, \ldots, \mu_k\}, \{\omega\} \) is the centre frequencies of each mode, and \( \{\omega\} = \{\omega_1, \omega_2, \ldots, \omega_k\} \).

2.1.2. The Solution of the Problem

The above constraint problem could be modified as a nonbinding problem by adding it to the penalty factor \( \delta \) and the Lagrangian multiplication operator \( \lambda(t) \), and here, the second penalty factor is used to guarantee the accuracy of the signal with the Gaussian noise, and the Lagrangian operator is adopted to ensure the constraint condition satisfied strictly, and the extended Lagrangian description is as follows:

\[
L(\{u_k\}, \{\omega_k\}, \lambda) = \| \sum_k \partial_t (\delta(t) + \frac{j}{\pi t} \mu_k(t)) e^{-j\omega_k t} \|^2 \\
+ \| f(t) - \sum_k u_k(t) \|^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle.
\]

VMD used multiplication operator alternating direction method of multipliers (ADMMs) to solve the above variable problem. By alternately updating \( A, B, \) and \( C \) to seek the “saddle” of the Lagrangian expression:

\[
u_k^{n+1} = \arg \min_{u_k \in \mathbb{C}} \left\{ \| \partial_t (\delta(t) + \frac{j}{\pi t} \mu_k(t)) e^{-j\omega_k t} \|^2 \\
+ \| f(t) - \sum_{i \neq k} u_i(t) + \lambda(t) \|^2 \right\},
\]

where \( \omega_k \) is equal to \( \omega_k^{n+1} \), and \( \sum_i u_i(t) \) is equal to \( \sum_{i \neq k} u_i(t)^{n+1} \).

Equation (5) is transformed to another domain by frequency by the Parseval/Plancherel Fourier equidistant transformation:

\[
\hat{u}_k^{n+1} = \arg \min \left\{ \| f(\omega - \omega_k)(1 + \text{sgn}(\omega - \omega_k)) \hat{\mu}_k(\omega) \|^2 + \| \hat{\mu}_k(\omega) + \hat{\lambda}(\omega) \|^2 \right\}.
\]

The first item of \( \omega \) with \( \omega - \omega_k \) instead is as follows:

\[
\hat{\omega}_k^{n+1} = \arg \min \left\{ \| f(\omega - \omega_k)(1 + \text{sgn}(\omega - \omega_k)) \hat{\mu}_k(\omega) \|^2 + \| \hat{\mu}_k(\omega) + \hat{\lambda}(\omega) \|^2 \right\}.
\]

Equation (7) is converted into a nonnegative integral form with frequency interval:

\[
\hat{\omega}_k^{n+1} = \arg \min \left\{ \int_0^\infty 4a(\omega - \omega_k)^2 |\hat{\mu}_k(\omega)|^2 + \| \hat{\mu}_k(\omega) + \hat{\lambda}(\omega) \|^2 d\omega \right\}.
\]

In this case, the solution of the quadratic optimization problem is as follows:

\[
\hat{\omega}_k^{n+1} = \frac{\hat{\omega}_k^{n+1} \omega_k}{1 + 2a(\omega_k - \omega_k)^2}.
\]

Based on the above process, the centre frequency value is converted to the frequency domain:

\[
\omega_k^{n+1} = \arg \min \left\{ \int_0^\infty (\omega - \omega_k)^2 |\hat{\mu}_k(\omega)|^2 d\omega \right\}.
\]

The centre frequency updated scheme is as follows:

\[
\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{\mu}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{\mu}_k(\omega)|^2 d\omega}.
\]
where \( \tilde{u}_k^{m+1}(\omega) \) is equivalent to the current remaining amount of \( \tilde{f}(\omega) - \sum_k \tilde{u}_k(\omega) \) Wiener filtering. \( \lambda \) is the central gravity of the spectrum of the current mode function. For inverse Fourier transmission of \( \tilde{u}_k(\omega) \), then the real part is \( \{u_k(t)\} \).

The procedure of VMD algorithm is as follows:

1. Initialize the \( \{\tilde{u}_k^{(1)}\}, \{\omega_k^{(1)}\}, \{\lambda^{(1)}\}, \) and \( n. \)
2. Update \( u_k \) and \( \omega_k \) followed by (9) and (11).
3. Update \( \lambda \), and
   \[
   \lambda^{m+1}(\omega) - \lambda^m(\omega) + \tau \left( \tilde{f}(\omega) - \sum_k \tilde{u}_k^{m+1}(\omega) \right).
   \] (12)
4. For a given accuracy \( e > 0, \) if \( \sum_k \| \tilde{u}_k^{m+1} - \tilde{u}_k^m \|^2 / \| \tilde{u}_k^m \|^2 < e, \) stop iteration; otherwise, return to Step 2.

2.2. Basic Principle of Refined Composite Multivariate Multiscale Sample Entropy

2.2.1. Sample Entropy. The SE uses the measurement of the exponential function for two sequences with a tolerance \( r \) from \( m \) points to remain \( r \) of each other at the next point. For a time series with \( N \) sample points, the procedure of SE calculation is shown as follows:

1. The \( m \) length vectors \( X_m(i) \) are formed:
   \[
   \left[ X_m(i) = x(i), x(i + 1), \ldots, x(i + m - 1) \right] - u(i),
   \] (13)
   where \( m \) represents embedding dimension and \( X_m(i) \) has \( m \) consecutive values. Commencing with the \( i \)th point and generalized by removing a baseline:
   \[
   u(i) = \frac{1}{m} \sum_{j=0}^{m-1} x(i + j).
   \] (14)
2. For each \( X_m(i) \), the similarity degree between the \( X_m(i) \) and its neighboring vector \( X_m(j) \) is calculated by \( D_{i,j}^m \):
   \[
   D_{i,j}^m = \text{dist}(d_{i,j}^m, r),
   \] (15)
   where \( d_{i,j}^m \) denotes the maximum absolute difference in the corresponding scalar components of \( X_m(i) \) and \( X_m(j) \).
3. For each \( X_m(i) \) and the fixed tolerance \( r \), let \( A_i \) be the number of vectors that satisfy, and then, \( B_i^m(r) \) is denoted as follows:
   \[
   B_i^m(r) = \frac{A_i}{N - m + 1}, \quad 1 \leq i \leq N - m.
   \] (16)
4. The average of the \( B_i^m(r) \) is designated as follows:
   \[
   B^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} B_i^m(r),
   \] (17)
where \( r \) represents the boundary width of the exponential function.

(5) Increasing dimension \( m \) to \( m + 1 \), steps (1)-(4) are repeated to calculate the corresponding of SE values to find \( B_i^{m+1}(r) \), and SE is defined as follows:

\[
\text{SampleEn}(m, r, N) = \sum_{N \rightarrow \infty} -\ln \frac{B_i^{m+1}(r)}{B_i^m(r)}.
\] (18)

When \( N \) is finite, the SE can be estimated as follows:

\[
\text{SampleEn}(m, r, N) = -\ln \left[ \frac{B_i^{m+1}(r)}{B_i^m(r)} \right] = B^m(r) - \ln B^{m+1}(r).
\] (19)

SE determines the time sequence irregularity on the single scale. The smaller the value of FE is, the higher the time sequence self-similarity can be achieved. Conversely, the greater the FE value is, the more complicated time sequences without rules can be achieved.

2.2.2. Multiscale Entropy and Multiscale Sample Entropy. ME could be represented as a series of times with different scales, which is obtained through a coarse-grained process. The irregularity of time series is then presented by the ME, where the self-similarity of various scales is reflected. Considering the scale entropy, when there is a sequence with a higher entropy than another sequence, the corresponding complexity will also be higher than the other. That is, if there is a case that the increasing scale of a time series conversely results in the decreased entropy value, there are more chances that the time series holds a relatively simple sequence structure.

The MSE is obtained as follows:

1. Consider a discrete-time series \( \{X_m(i)\} \) \( 1 \leq i \leq N \). Firstly, the value of embedding dimension \( m \) and similar tolerance \( r \) of the SE are set. Another auxiliary time sequence is then constructed in a vector form, which is presented as \( y_k^r = \left[ y_{k,1}^r, y_{k,2}^r, \ldots, y_{k,p}^r \right] \) and named coarse-grained vector:
   \[
   y_k^r = \frac{1}{\tau} \sum_{i=1}^{\tau} X_m(i), 1 \leq j \leq \frac{N}{\tau}, 1 \leq k \leq \tau,
   \] (20)
   where \( \tau \) is the scale factor. Actually, we can find that when \( \tau = 1 \), the coarse-grained time series is equal to the previous one \( \{X_m(i)\} \), which means that by dividing the length of time series, we can decompose the original time series into a coarse-grained vector series \( y_k^r \). It is worth mentioning that the number of coarse-grained time series \( y_k^r \) is \( \tau \) and the length is \( N/\tau \).
2. Based on the different scales \( \tau \) of the time sequence, the SE values are obtained. Generally, the \( r \) in SE takes the standard deviation of the primary time
series when calculating the value of SE. This procedure is called MFE analysis.

\[ \text{MSE}(X, \tau, m, r) = \text{SampleEn}(m, r, y_i^r). \]  

(21)

2.2.3. Refined Composite Multiscale Sample Entropy. In the process of applying coarse graining, the length of the original time series is declined by introducing a factor of \( \tau \), as shown in equation (13). For example, the authors developed the RCMMSE algorithm with the aim to boost the accuracy of the MSE. In particular, the SE of all the time series is obtained at a certain scale of factor \( \tau \). The RCMMSE value is defined as follows:

\[ \text{RCMMSE}(X, \tau, m, r) = -\ln \frac{B^{m+1}(r)}{B^m(r)}, \]  

(22)

where \( B^{m+1}(r) = 1/\tau \sum_{k=1}^r B^{m+1}(r) \) and \( B^m(r) = 1/\tau \sum_{k=1}^r B^m(r) \). Thus, the RCMMSE value can be obtained as follows:

\[ X_m^p(i) = [x_1(i), \ldots, x_1(i + (m-1)\lambda), x_2(i), \ldots, x_2(i + (m-1)\lambda), \ldots, x_1(i), \ldots, x_1(i + (m-1)\lambda)], \]  

(24)

then, the probability \( B^m(r) \) is defined as follows:

\[ B^m(r) = \frac{p_i}{N - m + 1}, \quad 1 \leq i \leq N - m, \]  

(26)

where \( m = \max[M] * \max(\lambda) \).

(4) The mean of the \( B^m_i(r) \) is obtained as follows:

\[ \bar{B}^m_i(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} B^m_i(r). \]  

(27)

Increasing dimension \( m \) to \( m + 1 \), steps (1)-(4) are repeated to get the corresponding SE values to find \( B^{m+1}_i(r) \) and \( \bar{B}^{m+1}_i(r) \) of the multiple embedding vectors.

\[ B^{m+1}_i(r) = \frac{1}{p(N - m - 1)} \sum_{i=1}^{p(N-m-1)} B^{m+1}_i(r). \]  

(28)

The calculation of the RCMMSE is defined as follows:

\[ \text{RCMMSE}(X, \tau, m, r) = -\ln \frac{B^{m+1}(r)}{B^m(r)} = -\ln \frac{\sum_{k=1}^r B^{m+1}(r)}{\sum_{k=1}^r B^m(r)}. \]  

(23)

2.2.4. Refined Composite Multivariate Multiscale Sample Entropy. Given time series \( X_m^p(i) \) with \( p \) variables, the RCMMSE is calculated according to the following procedures:

(1) Each coarse-grained time series \( X_{m}^p(i) \) requires to compute the RCMMSE, the multiple embedding vectors \( M \), and therefore, the original \( X_m^p(i) \) is reconstructed using the multiple embedding vectors \( M \).

\[ \text{RCMMSE}(X, \tau, m, r) = -\ln \frac{\bar{B}^{m+1}_i(r)}{\bar{B}^m_i(r)}, \]  

(29)

3. Scheme Comparisons

3.1. Comparison of EEMD/VMD. To compare the EEMD and VMD models, a simulation signal is used to prove that the VMD model is better than EEMD. The simulation signal \( X(t) \) is superimposed by the multifrequency signal and the random noise signal with the standard deviation of 1, and the following formula is the signal definition.

\[ X(t) = [1 + \cos(2\pi \times 30t)] \times \cos(2\pi \times 125t) \]
\[ + [1 + \cos(2\pi \times 30t)] \times \cos(2\pi \times 155t + \cos(2\pi \times 5t)] \]
\[ + [1 + \cos(2\pi \times 30t)] \times \cos(2\pi \times 185t). \]

(30)

The time-domain waveforms and its envelope spectrum of the original signal are given in Figure 1. As shown in Figure 1(b), the main frequency of the original signal focuses on 0–200 Hz, especially 30.27 Hz. Then, EEMD and VMD were used to decompose the aforementioned signal.

To achieve the VMD, firstly, the number of modes \( k \) should be determined. It should be noted that if the value of \( k \) in VMD is chosen too tiny, the original signal with time frequency could not be fully captured by the decomposition of the mode. Larger \( k \) values produce the similar frequency between the BL-IMF components, and they may be over-decomposition. Therefore, we use the observing centre frequency of the signal scheme to select the applicable \( k \) in
3.2. Comparison of Validity of MSE/RCMMSE. In this section, the \(1/f\) noise signal was employed to demonstrate the superior performance of the RCMMSE compared with MSE. Resorting to MSE and RCMMSE, the probabilities of inducing undefined entropy were presented to address the white noise and \(1/f\) noises with 200 samples. For each sample, the length is chosen as 1000. The results of the corresponding probabilities are shown in Table 2. From Table 2, we can see that the probability of undefined entropy showed the same tendency as the time scale increased, while there was a decrease in the length of the time series. As given in Table 1, the probability of undefined entropy is about zero when they were used to analyse white noise in the MSE model. On the contrary, the probability of inducing undefined entropy is 0.01 when \(r = 4\) for all 200 \(1/f\) noise samples. In the MSE algorithm, the entropy is undefined when \(B_m^m(r)\) or \(B_m^{m+1}(r)\) is zero. However, the RCMMSE can successfully get the values of entropy from 1 to 20 when the white and \(1/f\) noises are considered. Hence, the RCMMSE model is superior to MSE.

3.3. Comparison of Accuracy in MSE/RCMMSE. In this section, the efficiency of the MSE/RCMMSE models was verified through case studies. In each simulation study, we employ noise samples with 1000 data points containing white and \(1/f\) noises. The results of MSE/RCMMSEs are presented in Figure 4. Meanwhile, Figures 5 and 6 present the results of means and standard deviations of entropies.

As seen in Figure 4(a), the overall recently RCMMSE model shows higher performance compared with the MSE model in the condition of white noise, resulting from the tolerance \(r\) in SE, which is utilized to evaluate the similarity between any two time series. Note that \(r\) was often chosen as \(0.15 \times \text{SD of the original time series}\).

It can be seen from Figures 5(a) and 5(b) that the means of the entropy values obtained using the MSE/RCMMSE are nearly equal in white noise. Nevertheless, the means of the entropy values of the MSE are higher than that of the RCMMSE (see Figure 5(b)). Figure 6 shows that the standard deviation of RCMMSE is all lower than that of MSE, and this result indicates that the entropies obtained using the RCMMSE algorithm were more consistent than those obtained using the MSE algorithm.

With the purpose of studying the relationship between the data length and the effectiveness of the MSE/RCMMSE models, several different data lengths \((N = 600, 1200, 1800,\) and 2400) are used to get the entropies. In Table 3, the means and standard deviation of the entropies are presented. It can be found that the results of the means of entropies by MSE and RCMMSE are nearly equivalent with the long primary time series \((N = 2400)\). However, with short original time series, the results of means show a significant difference. On the other hand, the results of the RCMMSE algorithm hold the lowest error when the discrimination between the
entropy of $1/f$ noise with 500 data points and the entropy of
$1/f$ noise with 2000 data points is considered. That is to say,
the RCMMSE has higher performance compared with other
algorithms at any length of the time series. In conclusion, the
RCMMSE is more reliable than MSE according to the above
discussions.

### 4. Experimental Setup

#### 4.1. Dataset Source.

The performance of the proposed scheme is verified through experiments in this part. Case
Western Reserve University Bearing Data are used, which was obtained from an induction motor. It should be noted
that the data are generated by accelerometers at the drive end and fan end. Besides, EDM is employed to model the possible faults in the motor bearings. In particular, the fault tolerance is set as 0.1778 mm. After the fault detection, the data located in the outer race were aggregated in a state, which holds the time configuration at 6:00 clock. To ensure
the data acquisition system suitable for the vibration signals, the amplifier is particularly designed with a high bandwidth. On the other hand, to improve the accuracy, a sampling frequency of 12000 Hz of each channel was set for the data recorder.

Table 4 demonstrates different working conditions, which are considered in this experiment. Note that 0.1778 mm is the fault tolerance. “Normal” and “slight” denote the fault severity. In addition, considering the drive end of the motor 1785 rpm is chosen as the motor revolving speed. In total, 200 data samples are adopted with each fault condition consisting of 51 samples. In particular, in each data sample, there are 2048 data points.

| Scale factor | MSE White noise | 1/f noise | RCMMSE White noise | 1/f noise |
|-------------|----------------|-----------|--------------------|-----------|
| 1–3         | 0              | 0         | 0                  | 0         |
| 4           | 0              | 0.01      | 0                  | 0         |
| 5           | 0              | 0.015     | 0                  | 0         |
| 6           | 0              | 0.045     | 0                  | 0         |
| 7           | 0              | 0.11      | 0                  | 0         |
| 8           | 0              | 0.155     | 0                  | 0         |
| 9           | 0              | 0.275     | 0                  | 0         |
| 10          | 0              | 0.35      | 0                  | 0         |
| 11          | 0              | 0.36      | 0                  | 0         |
| 12          | 0              | 0.4824    | 0                  | 0         |
| 13          | 0              | 0.4949    | 0                  | 0         |
| 14          | 0              | 0.5606    | 0                  | 0         |
| 15          | 0              | 0.5377    | 0                  | 0         |
| 16          | 0              | 0.5714    | 0                  | 0         |
| 17          | 0              | 0.7059    | 0                  | 0         |
| 18          | 0              | 0.7211    | 0                  | 0         |
| 19          | 0              | 0.6402    | 0                  | 0         |
| 20          | 0              | 0.6667    | 0                  | 0         |

4.2. Procedure of Our Proposed Scheme. The procedure of the proposed scheme can be shown as follows:

1. Resorting to EEMD and VMD models, the vibration signals under different cases were disintegrated into a sequence of IMFs and BL-IMFs.

2. Using the RCMMSE model to get the entropy values, the entropy values were regarded as the input of SVM/RF/PNN models for training and testing.

3. Testing samples were regarded as the input of the trained SVM/RF/PNN classifier, while the operating conditions can be recognized by the output of the SVM/RF/PNN classifiers. The procedure of our

![Figure 4: MER/RCMMSE values of white and 1/f noises. (a) White noise. (b) 1/f noise.](image-url)
Table 3: Means and standard deviations of the MSE and RCMMSE of 1/f noise at a scale factor of 20.

| Data length | MSE | RCMMSE |
|-------------|-----|--------|
| 500         | —   | 2.1323 | 0.5234 |
| 1000        | —   | 2.0185 | 0.2667 |
| 1500        | —   | 2.0185 | 0.1761 |
| 2000        | —   | 1.9373 | 0.1230 |

Table 4: Experimental data of the roller bearings.

| Fault category | Fault tolerance (mm) | Motor speed (rpm) | Number of samples | The fault severity |
|----------------|----------------------|-------------------|-------------------|--------------------|
| NR             | 0                    | 1785              | 51                | Normal             |
| IRF            | 0.1778               | 1785              | 51                | Slight             |
| BF             | 0.1778               | 1785              | 51                | Slight             |
| ORF            | 0.1778               | 1785              | 51                | Slight             |
proposed fault diagnosis scheme is shown in Figure 7.

4.3. Parameter Selection. There are some parameters in different models that should be present before its calculation:

(1) EEMD: considering EEMD, there are two parameters that need to be determined, namely the ensemble number \( m \) and \( n_i \) related to the white noise. In particular, \( n_i \) denotes the amplitude. It should be mentioned that there is a one-to-one correspondence between the result and the ensemble number. Note that the ensemble number should be chosen as a few hundred. In addition, consider the situation that the input signal holds the standard deviation, and a fraction of one percent of error will be caused by the remaining noise with the added noise. In particular, \( m \) is set as 100.

(2) VMD: the number of BL-IMF modes \( k \) is according to the centre frequency. The second penalty factor \( \vartheta \) in equation (4) is often set as 2000.

(3) SE/RCMMSE:

(1) SE: before the process of SE, there are three parameters that must be chosen suitably. Firstly, considering the different lengths of sequences in SE, embedding dimension \( m \) should be designed. In particular, the dynamic process will be reconstructed in a more detailed manner with a larger \( m \). However, it should also be mentioned that if \( m \) is chosen too large, the need of \( N = 10^m - 30^m \) would be greatly limited, which will cause the loss of important information, and the general condition will be hard to satisfy. In the literature, \( m \) is often set as 2. Another two parameters, namely, similarity tolerance \( n \) and \( nr \), are dependent on the gradient of the exponential function and its bound, respectively. Experimentally, \( r = (0.1 - 0.25)SD \), and \( r \) is chosen as 0.15SD in SE.

(2) RCMMSE: the multiple embedding vector in RCMMSE scheme is set as \( M = [2_1, 2_2, 2_3, \ldots, 2_k] \), and here, \( k \) is the number of the BL-IMF components, and therefore, the time delay is set as 1.

(4) SVM: the kernel function in SVM is selected as the radial basis function (RBF).

(5) PNN: the distribution density of PNN is set as 1.5.

(6) RF: there are two parameters that need to be confirmed before using RF model, such as the number of input variables \( mtry \) is selected randomly based on the \( M \) input variables. Regarding the scale factor \( \tau = 20 \) and EEMD/VMD-RCMMSE as feature, the number of input variables \( M = \tau = 20 \), and the parameters often meet the condition \( mtry \leq \sqrt{M} \) [15]. Therefore, we let \( mtry = 4 \) and the number of the DT is set as 4 and 1000 in this simulation.

5. Experimental Results and Analysis

5.1. EEMD/VMD. The data are selected from the experiments in which SKF bearings are used. The number of each sample is set as NR:1–50, IRF:51–100, BF:101–150, ad ORF:151–200. As limited space, here with a sample of each state for an example, the time-domain waveforms of vibration signals under different working conditions are shown in Figure 8.

The vertical axis is the acceleration vibration amplitude. Because of the influence of noise, it is difficult to find significant differences in different states. As shown in Figure 8, it is hard to distinguish the four signals; in particular, there is no obvious regularity in two states of NR and BF signals.
The EEMD and VMD were used to disintegrate the vibration signals into series modes (IMF and BL-IMF). Here, we use the IRF vibration signals to observe the frequency change situation, and the waveforms of one IRF vibration signal and its spectrum envelope are shown in Figure 9. As shown in Figure 5(b), the IRF signal fault frequency is 164.1 Hz. The results of the centre frequency under different $k$ values with VMD are given in Table 5. There are some modes, such as BL-IMF4 (0.2787 kHz) and BL-IMF5 (0.3071 kHz), and with similar centre frequencies, when $k = 5$ in Table 5, it is considered that there is over-decomposition. Therefore, the number of $k$ values is selected as 4 to decompose the vibration signals.

Employing EEMD/VMD, the roller bearing primary signals in Table 4 are decomposed into IMF1-10 and BL-IMF1-4. Figures 10 and 11 show the VMD and the corresponding spectrum of each mode for an IRF signal.

It can be shown from Figure 6(b) that the VMD can not only effectively remove the dummy components, but also each IMF in the range of a certain scale, and there is no mode mixing problem between them. Therefore, the scale characterization and decomposition effect of VMD are better than EEMD, and this result reveals that the VMD owns a good robustness.

The correlation coefficient scheme is utilized to prove the degree of correlation between each mode and the primary signal to further prove the effectiveness of VMD scheme. In Tables 6 and 7, the overall average correlation values of BL-IMF component and original IRF signals are higher than IMF. Therefore, the decomposition effect of VMD is superior to EEMD.

After the vibration signal feature extraction using VMD and EEMD, the RCMMSE was employed to get the entropy value. The value of VMD-RCMMSE values under different $k$ values and scale factors is shown in Figure 12.

Figure 12 shows that the RCMMSE values of the VMD effective mode decrease with the increasement in the scale factor and the number of modes $k$, and the overall RCMMSE values are close to the steady state when the value of $k$ is more than 3, because the VMD algorithm decomposes the original signals into several BL-IMFs with information of different frequency bands. This can also be supported from Table 5. The correlation coefficient of most EEMD modes and the original signal is much smaller than that of VMD mode. This indicates that the VMD algorithm can overcome the mode mixing shortcoming in EEMD. The entropy values of EEMD-RCMMSE and VMD-RCMMSE ($k = 4$) are shown in Figure 13.

As shown in Figure 13(a), it is difficult to distinguish the four types of roller bearing signals with EEMD-RCMMSE, especially the NR and IRF, BF, and ORF. On the one hand, the complexity of signals under normal states and the conditions with ball fault is similar. In other words, the feature of the two kinds of signals is alike. On the other hand, much noise hidden in the signals may affect the identification of them. However, the RCMMSE curves of BL-IMF components can recognize the four states clearly in Figure 13(b).

5.2. Fault Identification. Regarding the extracted RCMMSE vectors as the inputs of the chosen fault classifier, such as SVM/RF/PNN, there are four kinds of data with 200 samples in total. 10, 20, and 30 samples are used under different statuses as the training samples, and the rest samples (20, 30, 40) as the testing samples are used to compare the
The time domain waveforms of original signal

(a)

The envelope spectrum of original signal

(b)

Figure 9: Waveforms and the envelope spectrum of the original signal in time domain. (a) Time domain. (b) Envelope spectrum.

Table 5: Centre frequency under different $k$ in VMD.

| $k$ | BL-IMF1  | BL-IMF2  | BL-IMF3  | BL-IMF4  | BL-IMF5  | BL-IMF6  |
|-----|----------|----------|----------|----------|----------|----------|
| 2   | 0.0922   | 0.2631   | —        | —        | —        | —        |
| 3   | 0.0895   | 0.2246   | 0.2873   | —        | —        | —        |
| 4   | 0.0494   | 0.1140   | 0.2268   | 0.2878   | —        | —        |
| 5   | 0.0490   | 0.1137   | 0.2239   | 0.2787   | 0.3071   | —        |
| 6   | 0.0489   | 0.1136   | 0.2228   | 0.2730   | 0.2890   | 0.3144   |

The BLIMFs obtained from VMD

(a)

Figure 10: Continued.
The envelope spectrum of each BLIMF with VMD

Figure 10: Signal decomposition based on VMD. (a) The BL-IMF components. (b) The spectrum of every BL-IMF.

The IMFs obtained from EEMD

Figure 11: Continued.
classification accuracy. Part of the experiment classification results is shown in Figure 14 and Table 8.

As given in Table 8, the highest classification precision is up to 100% when \( k = 4 \) in VMD-RCMMSE model. On the contrary, the lowest accuracy is 63.75% in EEMD-RCMMSE model. The overall classification accuracy in VMD-RCMMSE model is higher than EEMD-RCMMSE model. The overall classification accuracy in SVM model is higher than RF/PNN models. As mentioned above, the experimental results show that VMD-RCMMSE combination model can recognize different working conditions of the roller bearings effectively.

Table 6: Total average correlation values of each IMF with EEMD.

| Mode | IMF1 | IMF2 | IMF3 | IMF4 | IMF5 | IMF6 | IMF7 | IMF8 | IMF9 | IMF10 |
|------|------|------|------|------|------|------|------|------|------|-------|
| EEMD |      |      |      |      |      |      |      |      |      |       |
| NR   | 0.54 | 0.62 | 0.46 | 0.45 | 0.45 | 0.27 | 0.12 | 0.022| 0.0047| 0.0089 |
| IRF  | 0.91 | 0.38 | 0.178| 0.101| 0.052| 0.0201| 0.00149| 0.00086| 0.00025| 2.46e−05 |
| BF   | 0.84 | 0.406| 0.328| 0.266| 0.085| 0.021 | 0.004 | 0.003 | 0.0013| 0.0013 |
| ORF  | 0.86 | 0.308| 0.366| 0.189| 0.125| 0.053 | 0.0041| 0.00268| 0.0013| 0.0001 |

Table 7: Total average correlation values of each BL-IMF with VMD.

| Mode | Mode | BL-IMF1 | BL-IMF2 | BL-IMF3 | BL-IMF4 |
|------|------|---------|---------|---------|---------|
| VMD (\( k = 4 \)) | NR   | 0.6152  | 0.4463  | 0.4940  | 0.5384  |
|     | IRF  | 0.2275  | 0.3718  | 0.5773  | 0.7702  |
|     | BF   | 0.3849  | 0.4141  | 0.4943  | 0.7376  |
|     | ORF  | 0.4201  | 0.5187  | 0.7421  | 0.2708  |
Figure 12: RCMMSE values under different $k$ values and scale factors.

Figure 13: RCMMSE values with EEMD and VMD ($k = 4$) modes.
EEMD-RCMMSE-SVM
(The number of test samples is 120, Accuracy=73.33%)

VMD-RCMMSE-SVM
(k=4, The number of test samples is 120, Accuracy=99.17%)

EEMD-RCMMSE-RF
(The number of test samples is 120, Accuracy=69.17%)

VMD-RCMMSE-RF
(k=4, The number of test samples is 120, Accuracy=96.67%)

EEMD-RCMMSE-PNN
(The number of test samples is 120, Accuracy=74.17%)

VMD-RCMMSE-PNN
(k=4, The number of test samples is 120, Accuracy=92.5%)

Table 8: Mean and standard deviation of the MSE/MFE/RCMSE and RCMMSE of 1/f noise at a scale factor of 20.

| Mode                  | Precision (%) | Total number of testing samples |
|-----------------------|---------------|---------------------------------|
|                       | 160           | 120                             | 80                      |
| EEMD-RCMMSE-SVM       | 69.375        | 73.33                           | 86.25                   |
| EEMD-RCMMSE-RF        | 63.75         | 69.17                           | 80                      |
| EEMD-RCMMSE-PNN       | 71.25         | 74.16                           | 71.25                   |
| VMD-RCMMSE-SVM (k = 3)| 85            | 88.3                            | 91.25                   |
| VMD-RCMMSE-RF (k = 3) | 81.25         | 88.3                            | 91.25                   |
| VMD-RCMMSE-PNN (k = 3)| 83.125        | 85                              | 86.25                   |
| VMD-RCMMSE-SVM (k = 4)| 98.125        | 99.1667                         | 100                     |
| VMD-RCMMSE-RF (k = 4) | 96.25         | 96.67                           | 97.5                    |
| VMD-RCMMSE-PNN (k = 4)| 93.75         | 90.83                           | 91.25                   |
| VMD-RCMMSE-SVM (k = 5)| 98.75         | 97.5                            | 98.75                   |
| VMD-RCMMSE-RF (k = 5) | 97.5          | 95                              | 96.25                   |
| VMD-RCMMSE-PNN (k = 5)| 86.875        | 90                              | 91.25                   |

Figure 14: Waveforms in the time domain.
6. Conclusion
The VMD analysis model and RCMMSE are proposed in this study. The VMD model is utilized to analyze the complex vibration signals. Therefore, BL-IMF can be obtained from the decomposition of the roller bearing vibration signals, the RCMMSE can deal with multichannel data, and the EEMD/VMD-RCMMSE values are regarded as eigenvectors. Because the VMD can achieve adaptive subdivision of each component in the frequency domain of signals, the roller bearing vibration signals are distinguished. Finally, the VMD-RCMMSE-SVM/RF/PNN combination models can distinguish roller bearing fault types effectively.

With the rapid development of Internet of things (IoT) [1, 2] and Industry 4.0 [3], there are increasingly massive real-time data from various types of mechanical equipment [4]. The availability of these data that contain abundant information about machine health has attracted more and more enterprises’ attention. It has been proved that large volume, high velocity, and diversity mechanical big data are the major properties of mechanical big data [5, 6]. Effective feature extraction from these data and accurate machinery health state evaluation with ever-accelerated updating of schemes have become hot research issues in the prognostic and health management systems in the era of industrial IoT [7].

7. Future Ideas
Most related roller bearing signal fault diagnoses depend on manual aids and expert knowledge. To diagnose more smartly, convolutional neural network (CNN), as well as long short-term memory (LSTM), will be considered for an end-to-end scenario. For ensuring high recognition accuracy, the fault classifier based on CNN may be used to extract the effective signal from the severely distorted signal; the LSTM could cope with time-varying bearing failure. Additionally, the aforementioned SVM fault classifier performs well when the data samples are insufficient. In the future, we may consider the generative adversarial networks (GANs) and transfer learning for lack of data. By simulating the distribution of fault samples, GAN could obtain more data. Besides, the transfer learning-related diagnosis model can reuse the previous information in the new task. Consequently, small-sized samples could achieve accurate fault identification.

Data Availability
The prior studies and data are cited at relevant places within the text as references [20].

Conflicts of Interest
The authors declare that they have no conflicts of interest.

References
[1] L. Zhao, J. Li, A. Al-Dubai, A. Y. Zomaya, G. Min, and A. Hawbani, “Routing schemes in software-defined vehicular networks: design, open issues and challenges,” IEEE Intelligent Transportation Systems Magazine, vol. 13, no. 4, pp. 217–226, 2021.
[2] L. Zhao, W. Zhao, A. Hawbani et al., “Novel online sequential learning-based adaptive routing for edge software-defined vehicular networks,” IEEE Transactions on Wireless Communications, vol. 20, no. 5, pp. 2991–3004, 2021.
[3] C. Chen, L. Liu, S. Wang, F. Hui, and Q. Pei, “Data dissemination for industry 4.0 applications in Internet of vehicles based on short-term traffic prediction,” ACM Transactions on Internet Technology, vol. 22, no. 1, pp. 1–18, 2021.
[4] J. Lu, H. Liu, Z. Zhang, J. Wang, S. K. Goudos, and S. Wan, “Towards fairness-aware time-sensitive asynchronous federated learning for critical energy infrastructure,” IEEE Transactions on Industrial Informatics, vol. 18, no. 5, pp. 3462–3472, 2021.
[5] L. Wu, C. Quan, C. Li, Q. Wang, B. Zheng, and X. Luo, “A context-aware user-item representation learning for item recommendation,” ACM Transactions on Information Systems, vol. 37, no. 2, pp. 1–29, 2019.
[6] S. Liu, J. Yu, X. Deng, and S. Wan, “FedCSP: an efficient communication federated learning approach for vehicular edge computing in 6G communication networks,” IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 2, pp. 1616–1629, 2021.
[7] J. Wang, L. Wu, K.-K. R. Choo, and D. He, “Blockchain-based anonymous authentication with key management for smart grid edge computing infrastructure,” IEEE Transactions on Industrial Informatics, vol. 16, no. 3, pp. 1984–1992, 2020.
[8] M. P. Paidousis and E. B. Deksnis, “Articulated models of cantilevers conveying fluid: the study of a paradox,” Journal of Mechanical Engineering Science, vol. 12, no. 4, pp. 288–300, 1970.
[9] J. Liu, L. Dai, and L. J. Zhao, “Modeling and simulation of flexible multi-body dynamics of concrete pump truck arm,” Chinese Journal of Mechanical Engineering, vol. 43, no. 11, pp. 131–134, 2007.
[10] G. Cazzulani, C. Ghielemetti, H. Giberti, F. Resta, and F. Ripamonti, “A test rig and numerical model for investigating truck mounted concrete Pumps,” Automation in Construction, vol. 20, no. 8, pp. 1133–1142, 2011.
[11] J. Rafiee, M. A. Rafiee, and P. W. Tse, “Application of mother wavelet functions for automatic gear and bearing fault diagnosis,” Expert Systems with Applications, vol. 37, no. 6, pp. 4568–4579, 2010.
[12] X. Lou and K. A. Loparo, “Bearing fault diagnosis based on wavelet transform and fuzzy inference,” Mechanical Systems and Signal Processing, vol. 18, no. 5, pp. 1077–1095, 2004.
[13] N. E. Huang, Z. Shen, S. R. Long et al., “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, vol. 454, no. 1971, pp. 903–995, 1998.
[14] Z. Wu and N. E. Huang, “Ensemble empirical mode decomposition: a noise-assisted data analysis method,” Advances in Adaptive Data Analysis, vol. 1, no. 1, pp. 1–41, 2009.
[15] X. Zhang and J. Zhou, “Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines,” Mechanical Systems and Signal Processing, vol. 41, no. 1–2, pp. 127–140, 2013.
[16] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” IEEE Transactions on Signal Processing, vol. 62, no. 3, pp. 531–544, 2014.
[17] N. Huang, H. Chen, G. Cai, L. Fang, and Y. Wang, “Mechanical fault diagnosis of high voltage circuit breakers based on variational mode decomposition and multi-layer classifier,” Sensors (Basel, Switzerland), vol. 16, no. 11, pp. 1–19, 2016.

[18] X. An, L. Pan, and P. Luo, “Bearing fault diagnosis of a wind turbine based on variational mode decomposition and permutation entropy,” Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, vol. 231, no. 2, pp. 200–206, 2017.

[19] K. Zhu, X. Song, and D. X. Xue, “Fault diagnosis of rolling bearings based on IMF envelope sample entropy and support vector machine,” Journal of Information and Computational Science, vol. 10, no. 16, pp. 5189–5198, 2013.

[20] F. Xu, Y. J. Fang, and R. Zhang, “A fault diagnosis method combined with ensemble empirical mode decomposition, base-scale entropy and clustering by fast search algorithm for roller bearings,” Journal of Vibro engineering, vol. 18, no. 7, pp. 04472–04490, 2016.

[21] M. Costa, A. L. Goldberger, and C. K. Peng, “Multiscale entropy analysis of complex physiologic time series,” Physical Review Letters, vol. 89, no. 6, pp. 068102–068118, 2002.

[22] M. Costa, A. L. Goldberger, and C. K. Peng, “Multiscale entropy analysis of biological signals,” Physical review. E, Statistical, nonlinear, and soft matter physics, vol. 71, no. 5, pp. 021906–021918, 2005.

[23] J. D. Zheng, J. S. Cheng, and Y. Yang, “A rolling bearing fault diagnosis approach based on multiscale entropy,” Journal of Hunan University (Natural Sciences), vol. 39, no. 5, pp. 38–41, 2012.

[24] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. Guo, “Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference,” Expert Systems with Applications, vol. 37, no. 8, pp. 6077–6085, 2010.

[25] S.-D. Wu, C.-W. Wu, S.-G. Lin, K.-Y. Lee, and C.-K. Peng, “Analysis of complex time series using refined composite multiscale entropy,” Physics Letters A, vol. 378, no. 20, pp. 1369–1374, 2014.

[26] C. L. Liu, Y. G. Wu, and C. G. Zhen, “Rolling bearing fault diagnosis based on variational mode decomposition and fuzzy C means clustering,” Proceedings of the CSEE, vol. 35, no. 13, pp. 3358–3365, 2015.

[27] M. U. Ahmed and D. P. Mandic, “Multivariate multiscale entropy analysis,” IEEE Signal Processing Letters, vol. 19, no. 2, pp. 91–94, 2012.

[28] B. Gu, V. S. Sheng, and K. Y. Tay, “Incremental support vector learning for ordinal regression,” IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 7, pp. 1403–1416, 2015.