Fault Diagnosis of Rolling Bearings Using Dual-Tree Complex Wavelet Packet Transform and Time-Shifted Multiscale Range Entropy

TAO HAN, JIAN-CHENG GONG, XIAO-QIANG YANG, AND LI-ZHOU AN
College of Field Engineering, Army Engineering University of PLA, Nanjing 210007, China

ABSTRACT Most existing fault diagnosis methods for rolling bearings are single-stage; these methods can only judge the fault type but cannot detect the existence of a fault. Moreover, the uncertainty in pattern recognition may lead to misclassification of healthy bearings as faulty ones. This paper proposes a multistage fault detection scheme for rolling bearings. In the first stage, the sensitivity of the range entropy to bearing failure is used to define a threshold, based on which the health status of the bearing is judged. If the unknown bearing is judged to be faulty, the next stage is implemented. In the second stage, a fault feature extraction method based on dual-tree complex wavelet packet transform (DTCWPT), time-shifted multiscale range entropy (TSMRE), and t-distributed stochastic neighbor embedding (t-SNE) is proposed, and a random forest (RF) discriminator is used for fault classification. To achieve the desired performance of fault classification, a new coarsening approach for complexity measurement called TSMRE is developed on the basis of the range entropy (RE). First, the RE value of each time-shifted coarse-grained time series is calculated, and the TSMRE is obtained by averaging the entropy values. The TSMRE improves the coarse-graining processing of the MRE and enhances the stability and reliability of the algorithm. In addition, it can obtain more information from short time series using the time-shifted coarse-grained technology. Therefore, it is less dependent on the length of the original time series. Two sets of rolling bearing data are used for this experiment. The fault recognition rate of each category of samples is 100%. Therefore, the proposed multistage fault diagnosis method can pre-screen healthy bearings and accurately identify the failure types of faulty bearings.

INDEX TERMS Dual-tree complex wavelet packet transform, time-shifted multiscale range entropy, t-distributed stochastic neighbor embedding, random forest, rolling bearing, two-stage fault diagnosis.

I. INTRODUCTION
A rolling bearing is a widely used component and plays an important role in all types of rotating machinery. High-speed and high-temperature working environments make it prone to various types of failures [1]. Once failure occurs, it can easily cause equipment breakdown or even damage, resulting in serious economic losses and even accidents. Therefore, the intelligent monitoring and diagnosis of the health status of rolling bearings has significant engineering application value [2], [3].

Mechanical vibration signals are commonly used to diagnose faults [4], [5]. However, because of the complex working environments (such as friction, impact, and structural deformation), the gathered vibration signals for the most part have significant nonlinear and nonstationary qualities [6]. Conventional linear signal processing methods are inefficient when it comes to fault information acquisition [7]. Many nonlinear analysis approaches for vibration signals, such as empirical mode decomposition (EMD) [8], local mean decomposition (LMD) [9], and envelope spectrum analysis (ESA) [10], have been presented and applied to identify fault types of rolling bearings. These approaches have a few intrinsic drawbacks. For example, EMD is inclined to mode mixing, and LMD has a low computation efficiency. Other techniques, such as the ensemble empirical mode decomposition (EEMD) [11] and variational mode decomposition (VMD) [12], have been introduced; however, the operators used in these methods need to have a specific knowledge hold, which restricts the work efficiency and the applied scope of
these techniques. Therefore, applying more effective methods to extract fault features from nonlinear vibration signals is a key point of current research in the field of fault diagnosis.

The entropy is widely used in vibration signal processing of mechanical equipment [13], [14]. For example, Yan used the approximate entropy (AE) to monitor the running state of rolling bearings for the first time [15]. However, the signal analysis strongly depends on the data length of the time series, with short time series having a particularly adverse effect. The sample entropy (SE) is an improved version of the AE in that it can overcome the data length issue. Therefore, it is widely used in the fault diagnosis of rotating machinery. Han used the SE to extract the fault information of rolling bearings [16]; however, the SE is extremely sensitive to the amplitude of the signal, because of which it may yield inaccurate estimation and undefined entropy. Considering the drawbacks of SE, Omidvarnia et al. developed the range entropy (RE) in 2018 and provided two versions, versions A and B, of which version B was improved based on the SE [17]. By calculating the RE of a vibration signal, an entropy value can be obtained to represent the signal.

The above methods are single-scale analysis of vibration signals, ignoring the fault information at other scales [18]. The vibration signal of a faulty bearing contains multiple inherent oscillation modes at different scales, caused by the interaction and coupling of the various parts of the machine. Therefore, it is necessary to evaluate vibration signals from multiple scales. Multiscale RE was developed to assess the complexity of time series at different scales [19], [20]. It divides time series into groups of coarse-grained sequences and calculates the entropy of each sequence. However, the conventional coarse-grained segmentation method has some drawbacks that restrict its performance. When the scale factor is high, the coarse-grained time series for high scale factors will become short, leading to the loss of significant effective information. In addition, the conventional coarse-grained analysis is based on the average value of the debris and therefore cannot be used to analyze the high-frequency information contained in signals, resulting in inadequate and incomplete feature extraction. Based on the above analysis, the present coarse-grained method has significant drawbacks, making it necessary to modify it to improve the stability and reliability. Inspired by the operation of time-shifting [21], this study developed a new complexity measurement, called time-shifted multiscale entropy (TSMRE), to characterize the complexity of fault vibration signals of rolling bearings. This method overcomes the shortcomings of the existing coarse granulation methods and is more reliable for quantifying vibration signals.

Generally, the direct application of the TSMRE for the fault feature extraction of original vibration signals is associated with the following drawbacks. (1) An original vibration signal typically contains a large number of nonrelevant components, such as noise; this will inhibit feature extraction, and the extracted features may not accurately represent the fault state. (2) The components of vibration signals with inherent characteristics, such as impact and harmonics, typically have a weak intensity, and an appropriate technology should be adopted to strengthen these components [22]. A signal decomposition algorithm is required to preprocess the data, decomposing signals into multiple components and reconstructing the components selectively [23]. Currently, wavelet transform (WT) is widely used in processing mechanical vibration signals. It is a high-resolution analysis method that can enhance the instantaneous changes in signals through window functions; however, it lacks adaptability, and its performance is limited by parameters. Kingsbury proposed dual-tree complex wavelet transform (DTCWT) [24], which has the advantages of shift invariance and excellent direction selectivity; however, it cannot section the high-frequency components of a signal and provide a sufficient analysis [25]. The dual-tree complex wavelet packet transform (DTCWPT) [26] is proposed to overcome the poor signal decomposition ability of the DTCWT. In recent years, many researchers have used DTCWT for fault diagnosis [27], [28]. In this study, it was used to decompose bearing fault vibration signals and alleviate the interference in feature extraction.

Due to the inevitable existence of redundant information, a direct pattern recognition of original high-dimensional TSMRE feature vectors will not only increase the computational cost, but also affect the recognition accuracy [29]. Therefore, sensitive low-dimensional fault features should be selected to enhance the classification performance. t-distributed random neighborhood embedding (t-SNE) is a nonlinear dimensionality reduction algorithm for high-dimensional data; it can capture the internal structure of samples and discover the relationship between data [30]. Hence, in this study, t-SNE was used to compress the dimension of the extracted features to generate a low-dimensional fault feature matrix. Subsequently, a suitable classifier was required to identify the fault types of samples intelligently. Currently, typical classifiers, such as the extreme learning machine (ELM) and support vector machine (SVM) [31], are widely used in the field of pattern recognition; however, these classifiers have evident drawbacks. For example, the SVM parameters need to be optimized, and the ELM lacks a kernel function, because of which it cannot effectively classify high-dimensional nonlinear samples. Random forest (RF), which is a classification algorithm based on the decision tree, can accurately classify high-dimensional data by integrating the voting results of multiple decision trees [32], [33]. Hence, it was used for the fault identification of test samples in this study.

Most existing fault diagnosis methods for vibration signals include three steps: feature extraction, feature selection, and pattern recognition [34]–[36]. However, the three steps have uncertainty in that they misidentify normal samples as faulty ones. Therefore, this study proposes a multistage diagnostic scheme. In the first stage, an additional step is used to check the bearing condition in advance, screening out normal samples to avoid any misclassification due to the uncertainty in pattern recognition. Since the RE can detect whether there
is a bearing failure, it was used as a reliable tool in this study to determine if there is a bearing failure. In the second stage, a method based on DTCWPT, TSMRE, t-SNE, and RF is implemented to identify the fault type. First, the fault vibration signal is decomposed into multiple frequency band components using the DTCWPT. Subsequently, the TSMRE is used to mine the hidden fault features in each frequency band component to generate an initial high-dimensional feature matrix. The t-SNE is then used to compress the features and construct a low-dimensional feature matrix. Finally, the fault types of the test samples are intelligently identified based on an RF classifier. Two sets of experimental data from different types of rolling bearings are used to validate the proposed method. The accuracy with which the proposed method can screen out healthy bearings and judge the failure types of faulty bearings is evaluated.

The remainder of this paper is organized as follows: Section II primarily presents the theories of DTCWPT and TSMRE. A bearing simulation model is used to verify the performance of the TSMRE. Section III describes the process of the proposed fault diagnosis model in detail. In Section IV, the drive-end and fan-end bearing datasets are taken as research objects, and the excellent performance of the proposed method is fully verified through a series of experiments. The conclusions are given in Section V.

II. CORRELATION THEORY

A. DUAL-TREE COMPLEX WAVELET PACKET TRANSFORM (DTCWPT)

The dual-tree complex wavelet transform (DTCWT) was first proposed by Kingsbury and thereafter improved by Selesnick by developing the decomposition and reconstruction algorithm. In the process of signal decomposition and reconstruction using the DTCWT, the sampling position of the imaginary tree is always at the middle of the real tree, so as to complement the information between the imaginary tree and the real tree, thus avoiding information loss. However, the DTCWT is sensitive to the singularities of signals and cannot effectively mine the characteristic frequencies of signals.

The down-sampling of the conventional discrete wavelet packet transform (DWPT) reduces the sampling frequency and time resolution of the signal by half after each decomposition. Therefore, once the number of decomposition layers is given, the frequency resolution of the band will be fixed. The greater the number of decomposition layers, the higher the frequency resolution, but the lower the time resolution. In addition, the down-sampling process does not have a shift-invariance property and can cause serious frequency aliasing problems.

Based on the DTCWT and DWPT, the DTCWPT has been proposed to analyze nonlinear vibration signals. The decomposition and reconstruction of the DTCWPT are simple, comprising two parallel DWPT units with various filters, one set of high-pass filters and another set of low-pass filters. In the two parallel DWPT units, as shown in Figure 1, one DWPT unit can be viewed as an imaginary tree and the other as a real tree. The two DWPT units complement each other in the signal decomposition process and reduce information loss, by gaining the near-shift invariance and overcoming the drawback in that the DTCWT cannot section the high-frequency components of a signal.

Here, $u(t)$ and $\tilde{u}(t)$ represent the original input data and the reconstructed data, respectively. The real tree wavelet packet is decomposed by the First_1 filter group, and the imaginary tree wavelet packet is decomposed by the First_2 filter group in the first decomposition layer. The two filter groups have two parts each: $f_{1-1}$ and $g_{1-1}$ are high-pass filters, and $f_{1-0}$ and $g_{1-0}$ are low-pass filters. For the decomposition process of more than two layers, it is ensured that the total difference in the delay generated by the two trees in the local layer and all the previous layers is one sampling period compared with the input of the original signal. In other words, a delay of half a sampling period between the phase and frequency responses of the respective filters of the two trees is required. In addition, the amplitude–frequency responses of the two filters are equal. First_1 filter bank $h$ is used to decompose the real tree wavelet packet alternately. Likewise, First_2 filter bank $g$ is used to decompose the imaginary wavelet packet alternately. In the decomposition process of each layer, the coefficient bisection is adopted to remove redundant calculation and improve the efficiency of the signal analysis. The decomposition process of the DTCWPT is the inverse transformation of the decomposition Time-shifted Multiscale Range Entropy (TSMRE)

1) RANGE ENTROPY (RE)

The RE is a nonlinear dynamic method developed by Amir in 2018 for the measurement of complex nonlinear time series. Figure 2 shows the flowchart of its implementation. The detailed implementation procedure is as follows:

(1) Based on the 1D series $\{x_1, x_2, \cdots, x_L\}$ of length $L$, $m$ data points are selected to reconstruct the series $Z_i^{m,d}$ in their original order as follows:

$$Z_i^{m,d} = \{x_{i}, x_{i+d}, x_{i+2d}, \cdots, x_{i+(m-1)d}\}$$

(1)

Here, $m$ is the embedding dimension of the reconstructed phase space, and $d$ denotes the time delay. In the RE algorithm, where $d = 1$, the original data are split into $L-m+1$ new sequences, as shown below:

$$Z_i^m = \{x_i, x_{i+1}, x_{i+2}, \cdots, x_{i+m-1}\}$$

(2)

(2) The distance between each $Z_i^m$ and other subsequences $Z_j^m$ is calculated, and the $L-m$ distances for each $Z_i^m$ are obtained. In the sample entropy and approximate entropy analyses, the distance between the sequences is solved using the Chebyshev function. Subsequently, Omidvaria updated the distance function, which is defined as follows:

$$D_{\text{range}}(Z_i^m, Z_j^m) = \left| x_{i+k} - x_{j+k} \right| - \left| x_{i+k} - x_{j+k} \right|$$

$$k = 0, 1, 2, \cdots, m-1$$

(3)
Assuming that there are two 2D subsequences $Z_1^2 = (2, 7)$ and $Z_2^2 = (1, 5)$, the distance between them can be expressed as follows:

$$D_{\text{range}}(Z_1^2, Z_2^2) = |7 - 5| - |2 - 1| = \frac{1}{3}$$  \hspace{1cm} (4)

(3) Comparing each $D_{\text{range}}(Z_m^i, Z_m^j)$ with the tolerance $r$, $B_{m,r}^i$ is the amount of $D_{\text{range}}(Z_m^i, Z_m^j)$ meet $D_{\text{range}}(Z_m^i, Z_m^j) > r$, as follows:

$$B_{m,r}^i = \sum_{j=1,2,3,\cdots, L-M; j \neq i}^L \Psi(r - D_{\text{range}}(Z_m^i, Z_m^j))$$  \hspace{1cm} (5)

where $\Psi(\cdot)$ is the Heaviside function, defined as:

$$\Psi(r - D_{\text{range}}(Z_m^i, Z_m^j)) = \begin{cases} 0, & r - D_{\text{range}}(Z_m^i, Z_m^j) < 0 \\ 1, & r - D_{\text{range}}(Z_m^i, Z_m^j) \geq 0 \end{cases}$$  \hspace{1cm} (6)

$B_{m,r}^i$ is the average value of all $B_{m,r}^i$, expressed as follows:

$$B_{m,r}^i = \frac{1}{L-M} \sum_{i=1}^{L-m} B_{m,r}^i$$  \hspace{1cm} (7)

(4) Obtaining $B_{m+1,r}^r$ by repeating (1)–(3) at dimension $(m + 1)$.

(5) Based on the Shannon entropy, the RE can be defined as:

$$RE = -\ln \frac{B_{m+1,r}^r}{B_{m,r}^i}$$  \hspace{1cm} (8)

2) MULTISCALE RANGE ENTROPY (MRE)

The RE can not only improve the performance of the SE but can also measure the complexity of time series more
accurately and effectively. However, for the vibration signal of a rolling bearing, which is typically distributed on multiple scales, the RE describes the complexity of the time series only on one scale. In the single RE analysis of a signal, the fault information at other scales is lost, and a comprehensive analysis cannot be made, making the analysis less reliable. Therefore, Zheng et al. proposed the multiscale range entropy (MRE) based on coarse-grained processing, so as to realize a comprehensive and sufficient analysis of the time series. The implementation of the MRE is as follows:

1) For 1D data \( x_1, x_2, \cdots, x_L \) of length \( L \), the coarse-grained time series \( y^j \) can be obtained using the following equation:

\[
y^j = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq \left[ \frac{L}{\tau} \right]
\]

(9)

Here, \( \tau (\tau \in N^*) \) is the scale factor, \( j \) is the index of the coarse-grained time series, and \( [ ] \) is the rounding function. When the scale factor \( \tau = 1 \), the coarse-grained time series is the original time series.

2) Calculating the RE of the coarse-grained time series. The multiscale RE of the original time series can be obtained using the following equation:

\[
MRE = RE \left( y^j, m, d, r \right)
\]

(10)

From the coarse-grained processing of Equation (9), when the scale factor increases, the number of data points contained in the sequence decreases, causing a significant deviation in the calculated entropy. Figure 3 shows the coarse granulation process with a scale factor of 2; this process can only analyze the information between \( x_1 \) and \( x_2 \) and \( x_3 \) and \( x_4 \), but does not consider the information between \( x_2 \) and \( x_3 \) and \( x_4 \) and \( x_5 \). Therefore, a large amount of valid information cannot be used in an integrated manner, because of which the information analysis is not sufficiently comprehensive.

3) TIME-SHIFTED MULTISCALE RANGE ENTROPY (TSMRE)

To obtain fault features from the original signals more comprehensively and sufficiently, this paper proposes TSMRE by combining RE with the time-shifted and coarse-grained multiscale processing method, which can obtain the foremost fault characteristics from the vibration signals. The TSMRE is implemented as follows:

\[
Y_{k,\beta} = (x_{k \beta + 1}, x_{2\beta + 1}, \cdots, x_{\Delta k(\beta + 1)})
\]

(11)

Here, positive integers \( k (1 \leq k \leq s) \) and \( \beta (\beta = s) \) are respectively the beginning of the time series and interval. \( \Delta k = \frac{L - \beta}{k} \) is a rounded integer representing the upper limit.

(2) On the basis of Equation (10), for scale factor \( s \geq 2 \), the RE value of each time-shifted coarse-grained time series is calculated, and the entropy values are averaged to obtain the TSMRE as follows:

\[
TSMRE = \frac{1}{s} \sum_{k=1}^{s} TSRE \left( Y_{k,\beta}, m, d, r \right)
\]

(12)

The TSMRE improves the coarse-graining processing of the MRE, thus enhancing the stability and reliability of the algorithm. In addition, the TSMRE can obtain more information from short time series using the time-shifted coarse-grained technology. Therefore, it has less dependence on the length of the original time series.

4) PARAMETER SETTINGS

In the DTCWPT algorithm, an excessive number of layers of the dual-tree complex wavelet packet decomposition can lead to partial loss of useful signals, making it important to determine the decomposition level; however, there is no specific basis for selecting the decomposition levels. In this study, the method reported in literature [37] was used to determine the optimal number of decomposition layers. This method first decomposes the signal into different layers by means of DTCWPT and obtains the sub-bands of different frequencies. Subsequently, the means of RE are calculated from these sub-bands from different decomposition levels. The minimum mean of RE corresponds to the optimal decomposition layer number, which was set to 3 in this study.

In the algorithm of the TSMRE, the performance is influenced by three parameters, namely the length \( L \) of the time series, the embedding dimension \( m \), and tolerance \( r \). Typically, \( m \) is 2, and \( r \), which is set to 0.2 \( \times \) std, has a significant impact on the performance of the algorithm, where std represents the standard deviation of the original time series. In addition, the length \( L \) of the data cannot be too
small. Although the TSMRE can significantly alleviate the interference of the data length on the algorithm performance using the time-shifted coarse-grained processing technology, fewer data points will still distort the result and cause an inaccurate estimation of the entropy value. However, the length of the time series cannot be too large, because it leads to a low computation efficiency and limited improvement in the algorithm performance.

Therefore, it is necessary to determine an appropriate data length to obtain stable algorithm performance and high computational efficiency. Since the method developed in this study was finally used for the vibration signal of a rolling bearing, the signal of a slight inner ring fault from the CWRU was used for the analysis [38]. To verify the superiority of the TSMRE algorithm over the MRE algorithm in analyzing short time series, an inner ring fault time series (IR7) with lengths of \( L = 512, 1024, 2048, 3072, 4096, 5120, 6144 \) was used to calculate the TSMRE and MRE. As shown in Figure 5, the data length has a significant impact on the performance of the algorithms; particularly when the data length is short, the results obtained using the two algorithms are unreliable. When \( L = 2048 \), the TSMRE achieves a relatively stable effect, which is consistent with the trend of the other curves. On the contrary, the MRE algorithm still has evident fluctuations. When the length \( L = 3072 \), the MRE algorithm produces more reliable results, which shows that the MRE is prone to significant fluctuations in the analysis of the short time series, and this drawback must be alleviated by increasing the data length. Based on the above analysis, \( L = 2048 \) is selected for the subsequent analysis.

5) EFFECTIVENESS VALIDATION

The developed TSMRE algorithm was used to extract the fault features of rolling bearings, which contain a large number of impact components. The advantage of the TSMRE algorithm can be confirmed by its superiority over other methods in extracting the impact components of signals. To simulate the actual inner ring fault of rolling bearings, a bearing simulation model is established as follows:

\[
\begin{align*}
x(t) &= \sum_{i=1}^{M} A_i s(t - iT - \tau_i) + w(t) \\
A_i &= A_0 \cos(2\pi f_i t) - \left[1 - \frac{1}{2e}(1 - \cos(2\pi f_i t))\right]^n \\
s(t) &= e^{-Bt} \sin(2\pi f_0 t + \phi_w)
\end{align*}
\]

Here, \( A_i \) is the amplitude modulation signal with period \( 1/f_r \), \( s(t) \) represents the discrete oscillation pulse signal with interval \( T \), \( \tau_i \) is the time delay caused by the random sliding of the roller elements, \( w(t) \) is the white Gaussian noise (WGN), \( A_0 \) represents the signal amplitude, \( e \) represents the damping parameter, and \( f_n \) represents the natural frequency of the system. The parameters were set as follows: speed \( f_s = 50 \text{ Hz} \), signal amplitude \( A_0 = 0.15 \), time delay \( \tau_i = 0 \), damping parameter \( e = 1000 \), system natural frequency \( f_n = 296.1 \text{ Hz} \), and sampling frequency \( f_s = 20480 \text{ Hz} \). In addition, a WGN with \( SNR = -10.06 \text{ dB} \) was added to the analog signal.

The TSMRE, TSMSE, TSMPE (time-shifted multiscale permutation entropy), and MRE algorithms were used to analyze the bearing simulation signals and test their performance in detecting impact signals. A sliding window with a length of \( 2048 \) points was used to divide the signal into different segments. For each method, the distance between the data of the first window and the data of each other window was calculated using the Euclidean distance. Figure 6 shows that each method has a different sensitivity to the impact components. The ED value of the TSMRE method is the highest, indicative of its sensitivity to shock signals. The TSMSE method has a low ED value, which indicates that its performance for impact components is unstable. When the impact component in the signal is weak, the TSMSE method may miss it. Generally, the TSMRE method outperforms the other methods.

III. PRESENTED MODEL

Based on the above analysis, a multistage scheme, including two subroutines of health detection and fault detection, was established for rolling bearings.
A. HEALTH DETECTION
The RE is a nonlinear dynamic method capable of measuring the complexity of the vibration signals. The RE can characterize the entropy value of the vibration signal under normal and fault conditions, which is the basis for health detection. When the bearing is running normally, the vibration generated is mainly composed of the natural vibration of the machine and those of the coupled components of the environmental noise, and the vibration signal has a certain regularity. Therefore, the RE will be relatively low. When the bearing fails, the vibration signal exhibits a periodic shock characteristic. The high-frequency shock vibration and the natural vibration of the bearing are mixed with each other, resulting in a relatively complex vibration signal and therefore a high RE value.

The RE values of the vibration signals in each fault state are significantly higher than those in the normal states when $\tau = 1$, which can be used for the health detection of unknown bearings. To strictly determine the detection standard, a threshold is set up. When the RE value of the detected bearing vibration signal is less than the threshold, it is detected as normal. Otherwise, it is detected as a fault.

B. FAULT DETECTION
After a bearing fault is detected in the health check of the first stage, the second stage is performed to detect the type and severity of the bearing failure. To excavate features with higher quality and more evident discrimination from fault vibration signals, a multicomponent and multiscale feature mining model based on the DTCWPT and TSMRE is developed. Currently, two types of feature mining models are widely used: (1) mining the MRE of a single component; (2) obtaining multiple subcomponents by decomposing the vibration signal and then calculating the single RE for each subcomponent. Compared with these two multiscale feature mining models, the multicomponent and multiscale feature mining model can highlight fault information and reduce interference by preprocessing the signal, and then extract the fault information from multiple scales. It can extract multiscale features from multiple components to obtain more comprehensive and effective features. The main mechanism of the model is as follows: First, the vibration signal is divided into several frequency band components containing the information of the original signal at different timescales using the DTCWPT. Subsequently, the fault information contained in these channel components is mined based on the TSMRE to form initial high-dimensional features. Next, since the dimension of the original feature is typically too high, the feature is compressed based on t-SNE to obtain a simplified feature matrix. Finally, a random forest classifier is used to intelligently detect the type and severity of failures.

Figure 7 shows the implementation route of the developed fault detection scheme.

The complete process is described as follows:
1. For a given sampling frequency, the vibration data of the rolling bearing under different working conditions are collected, and the vibration signal of length $M$ is obtained.
(2) Divide the vibration signal under different operating modes into \( H \) groups un overlapped samples of length \( L \), in which the \( P \) group is used as a training sample, and the \( Q \) group is used as a testing sample.

(3) Calculate the RE values of all the samples and then provide a threshold based on the RE value for sample health detection. If the RE value of the vibration signal of the bearing to be tested is lower than the threshold, the output state is normal, and the detection procedure is terminated. Otherwise, the input state is faulty, and the next step is executed.

(4) If the bearing is faulty, the DTCWPT is used to separate the vibration signal to obtain a batch of sub-band components.

(5) The fault information of all the sub-band components is mined based on the TSMRE to generate the original high-dimensional features.

(6) t-SNE is used to compress the original high-dimensional features to obtain reduced features.

(7) The \( P \) group samples are used to train the RF classifier, and a generalized optimal model is established. The \( Q \)-group samples are then used to test the model.

(8) The detection results of the test model are inputted, and the fault type and severity are judged.

IV. EXPERIMENT VALIDATION

The public dataset provided by Case Western Reserve University was used in the experiment to verify the effectiveness of the proposed model. Most existing methods only use data of the drive-end bearing for the experiments, and there are few studies on the vibration signals of fan-end bearings. This is mainly because the signal from the fan-end bearing contains significant noise and interference components, which conceal the fault information easily. The proposed model can alleviate the influence of these components on the analysis results by using the DTCWPT to decompose the signal, so as to obtain more robust results. Therefore, to study the generalizability of the proposed model, the data of the fan-end bearing were used to test the reliability of the model.

A. DRIVE-END BEARING EXPERIMENT

1) DATA INTRODUCTION

This experiment is based on the data provided by the CWRU to verify the generalizability of the model. Figure 8 shows a rolling bearing test bench comprising a drive-end rolling bearing, a fan-end bearing, an acceleration sensor, a torque sensor, and a motor. The rolling bearing used was the 6205-2RS JEM SKF deep groove ball bearing. The motor was unloaded, and the speed was 1797 rpm. The vibration signals used in the experiment were collected by two sensors installed on the drive end and the fan end, and the sampling frequency was 12 kHz. Nine different fault states and one normal state were simulated in the experiment. The nine fault states comprise three types of faults with different severities: inner ring fault, rolling element fault, and outer ring fault.

(2) The detection results of the test model are inputted, and the fault type and severity are judged.

FIGURE 8. Rolling bearing platform. Health detection.
First, the DTCWPT is used to decompose each faulty sample into three layers to extract the fault information hidden in the signal and eliminate the effects of noise as well as other interference components. Subsequently, eight sub-band components with different timescale information of the original signal can be obtained. Taking the limitation of the length into consideration, only the DTCWPT decomposition results of the IR7 samples are taken as an example, as shown in Figure 11.

After the signal is decomposed by the DTCWPT, the TSMRE algorithm is used to mine the hidden fault features of each sub-band signal to form the initial high-dimensional fault features. Figure 12 shows the TSMRE values of each sub-band component at the scale of 20 under different fault conditions. The TSMRE could amplify the fault features in the signals at different scales and improve the recognition of features. Figure 12 shows that for the sub-band components under different fault states, the TSMRE has evident differences at some scales. For example, sub-band component 1 and sub-band component 4 have significantly different TSMRE values under different fault conditions, while the difference in the sub-band component 3 is not evident and difficult to distinguish effectively. However, it is difficult to directly judge the fault types of rolling bearings from the curves shown in Figure 12, and the separability of
some of the fault features is not sufficiently evident. It is necessary to quantify the feature extraction effect of the proposed model.

4) FAULT IDENTIFICATION

After the feature extraction, 522 × 160 features were obtained, and the fault types of the samples need to be identified. However, the primitive features are high dimensional, with 160 features per sample, as mentioned in the above analysis. Therefore, using many features to describe the fault type of the samples may produce redundant information, which will not only interfere with the reliability of the identification but also reduce the classification efficiency. Therefore, it is necessary to reduce the dimension of the primitive features to obtain low-dimensional and high-quality feature matrices. T-SNE is used to compress high-dimensional samples for sensitive low-dimensional features. After the dimensionality reduction, the dimension of the feature vector is reduced to 522 × 3. The 252 sets of training samples were fed into an RF classifier for training to generate the best RF mode. The remaining 270 test samples were then fed into a trained RF classifier for fault type identification. Figure 13 shows the identification results of the proposed fault diagnosis model for fault samples in one test. As shown, all the samples have been accurately identified along with their fault types, indicating a 100% identification accuracy for this classification.

A fuzzy matrix scheme is used for evaluating the model performance in the pattern recognition experiment, which can reflect the divergences between the expected and actual outputs, and the results of this method can be accurately reflected for different samples. Figure 14 shows the fuzzy matrix of this method for single-shot classification. By observing the confusion matrix, the identification result of the various fault types is clear. The fault types of each category of samples could be accurately identified. The fault recognition rate of each category of samples is 100%. Therefore, this method can reliably identify different categories of faults.

To prove the necessity and superiority of the DTCWPT in signal decomposition over other methods, several methods, namely the EMD–TSMRE, WPT–TSMRE, DTCWT–TSMRE, DTCWPT–RE, and TSMRE, were used for comparison and validation. The parameters were set as follows: the number of WPT and DTCWT decomposition layers was set to three, the wavelet basis function was db4, and the first eight IMF components were used in the EMD for feature extraction. The fault characteristics extracted from the five methods were fed into the RF classifier to be trained and tested after the same treatment. The results are categorized in Figure 15: the obtained identification rates of 96.67%, 95.19%, 98.52%, 89.63%, and 92.96% are lower than those
The low recognition performances of EMD, WPT, and DTCWT in signal decomposition is due to the shortcoming of signal decomposition, which leads to poor signal composition and limited feature extraction. For the DTCWPT–RE model, only one entropy value was extracted from multiple components without considering the multi-scale information of each component, resulting in incomplete feature extraction and neglecting a large amount of valid information. However, when the original vibration signal was analyzed directly using the TSMRE, a large amount of fault information contained in the extracted fault characteristic sample was lost because the fault frequency and other frequency domain information contained in the fault vibration signal could not be mined, and the fault characteristic analysis was incomplete. Compared with the above five methods, the DTCWPT–TSMRE is a feature extraction model with good signal decomposition performance and feature extraction ability. By amplifying the effective information components in the fault signal and mining multiscale features, the fault mechanism of the signal can be reflected more comprehensively, and the fault features can be distinguished better, so as to facilitate fault identification.

To test the advantage of the developed complexity measurement method TSMRE in analyzing actual vibration signals over other methods, typical signal complexity measurement methods were used for comparison, namely the time-shifted multiscale sample entropy, time-shifted multiscale permutation entropy, and multiscale sample entropy. Table 2 presents the parameter settings of each method. Each method uses DTCWPT to pre-decompose the fault signal and extract the top eight sub-band components of the fault characteristics. It is then subjected to t-SNE dimensionality reduction and RF classification. Each method is repeated 15 times to overcome the potential error encountered in single experiments. Figure 16 and Table 3 present the outcomes of 15 classifications obtained by the five methods. Evidently, the developed TSMRE data complexity measurement method has the best fault recognition results compared with the other four entropy-based methods. It has an average accuracy of 100% and a high stability. The other four methods have a fluctuating classification accuracy curve and lower average accuracy, which proves the validity of the DTCWPT–TSMRE model compared with the other methods. In addition, the multiscale feature extraction method based on time-shifted coarse-graining produces better results than the conventional multiscale method. This is because the former is based on the average value of debris and therefore cannot be used to analyze the high frequency information contained in signals. At the same time, the RE-based feature extraction method outperforms the PE-based feature extraction method. This is because the former ignores the amplitude characteristic of the vibration signal, leading to the loss of important information.

In this part, the necessity of using t-SNE in reducing the dimension of the data is studied. For an intuitive comparison, the first three features of the new data and the first two features of the original high-dimensional data are drawn in Figures 17 and 18, respectively. Figure 17 shows that all the fault samples can be clearly classified and have excellent clustering. This is consistent with the previous finding that this method is consistently 100% accurate in identification. Figure 18 shows that the samples not treated with t-SNE have poor separability, except for the OR21, IR21, and IR7 samples, which have evident cluster centers; the rest of the samples are more scattered. This shows that the random selection of the three features cannot fully characterize the bearing fault information, and compressed features are necessary to obtain a higher degree of integration. To further study the need for data dimensionality reduction, three fault features were randomly selected from the original high-dimensional data and fed into the RF classifier for training and testing to examine the effect of data dimensionality reduction on the fault recognition rate. Table 4 presents the classification results of the five fault diagnosis methods in 15 trials without t-SNE dimensionality reduction. In Table 4, the classification accuracy of the five fault diagnosis methods without t-SNE
The advantage of the RF classifier for the classification is analyzed in this part. The RF classifier is compared with the SVM, ELM, back propagation (BP) neural networks, and artificial neural networks (ANNs). Table 5 presents the diagnostic results of the five methods using different classifiers, with the same proportion of training and testing samples. When different feature extraction models are used, the reduction of high dimensional feature matrices can reduce the loss of redundant information and greatly improve fault identification performance.
T. Han et al.: Fault Diagnosis of Rolling Bearings Using DTCWPT and TSMRE

FIGURE 16. Classification results of six approaches after 15 test iterations.

FIGURE 17. Plotting of the first three features in new feature adopting t-SNE.

TABLE 4. Diagnosis results from five approaches without using t-SNE.

| Different approaches       | Max  | Min  | Mean | SD  |
|----------------------------|------|------|------|-----|
| DTCWPT-TSMRE              | 95.19| 91.85| 93.68| 1.185|
| DTCWPT-TSMSRE             | 93.70| 90.37| 92.12| 0.955|
| DTCWPT-TSMPE              | 90.00| 86.30| 88.22| 1.213|
| DTCWPT-MRE                | 87.78| 83.33| 85.43| 1.314|
| DTCWPT-MSE                | 82.96| 78.89| 81.04| 1.260|

used, the RF classification model achieves the best recognition results, with an average accuracy of 93.65%, which is higher than those of the other classifiers. Furthermore, the DTCWPT–TSMRE achieves the highest fault recognition rate after inputting different feature extraction models into each classifier, which further proves the generality of the proposed feature extraction model.

B. FAN-END BEARING EXPERIMENT

1) DATA INTRODUCTION

The open dataset provided by CWRU electrical laboratory was again used to confirm the generality of the model.

FIGURE 18. Plotting of the first three features without adopting t-SNE.

TABLE 5. Diagnostic results obtained by combining diverse methods with four classifiers.

| Classifier | Recognition rate of classifiers with diverse approaches (%) | Average accuracy (%) |
|------------|------------------------------------------------------------|----------------------|
|            | TSMRE | TSMSE | TSMPT | TSMRE | MSE |         |
| RF         | 100   | 97.78 | 93.43 | 92.59 | 84.44| 93.65  |
| SVM        | 99.26 | 98.15 | 92.60 | 91.85 | 84.07| 93.19  |
| ELM        | 100   | 96.67 | 93.33 | 92.22 | 81.48| 92.74  |
| BP         | 98.15 | 94.07 | 90.37 | 92.59 | 83.33| 91.70  |
| ANN        | 97.03 | 96.30 | 92.96 | 91.11 | 83.70| 92.22  |
| Average    | 98.89 | 96.59 | 92.54 | 92.07 | 83.40|        |

Table 6 presents the model number of the bearing at the fan-end (6203-2RS JEM SKF, deep groove ball bearing) and the bearing specifications. Taking the vibration signal of the fan-end bearing as the experimental data, the bearing conditions in this experiment were the same as those listed in Table 1. The other parameters and equipment were consistent with experiment 1.

2) HEALTH DETECTION

FIGURE 19 shows the vibration signals of the fan-end rolling bearing under various working conditions. The time-domain waveform of the vibration signal is complex, which is locally distinguishable, but inseparable globally. Similarly, the RE is used to evaluate the health of rolling bearings. Figure 20 shows the RE values of all the rolling bearing samples. As shown in Figure 20, faulty bearings have higher RE values whereas healthy bearings have lower RE values. The maximum RE value of the healthy bearing sample is 0.94701, which is much lower than the RE value of the faulty bearing sample. Therefore, the RE can indeed be used to...
evaluate the bearing health. The RE value indicated by the red dotted line was defined as the threshold (0.94701) for health monitoring. If the RE of the sample is less than this value, it proves that the sample is healthy. Otherwise, the sample is faulty. This is consistent with the previous conclusions of the health detection of the rolling bearing at the drive end, proving that the proposed health detection model is universal and can effectively evaluate the health status of diverse rolling bearing data.

3) FAULT FEATURE EXTRACTION
After bearing health check, the same DTCWPT decomposition was used for the IR7 sample of the fan-side rolling bearing to obtain eight sub-band components containing information at different timescales of the original signal. Figure 21 shows the results.

The initial high-dimensional fault features were also obtained from each sub-band signal using the TSMRE algorithm. Figure 22 shows the TSMRE values of each sub-band component at 20 scales under different fault states. Similar to the drive-end rolling bearing, the sub-band components under different fault conditions have evident differences in terms of the TSMRE at some scales.

4) FAULT IDENTIFICATION
After feature extraction, $522 \times 160$ features can be obtained. Similarly, t-SNE was used to compress the sample to obtain $522 \times 3$ sensitive low-dimensional features. For fault type identification training and testing, 252 and 270 groups of samples were respectively fed into the RF classifier. The confusion matrix results are consistent with the results of the rolling bearing at the drive end. Therefore, this method can also reliably identify different fault types of the fan-end bearing.

To verify the necessity of using DTCWPT for data decomposition and its advantages over other signal decomposition methods, the EMD–TSMRE, WPT–TSMRE, DTCWT–TSMRE, DTCWPT–RE, and TSMRE were used for comparison. The parameter settings and various operations remained the same. The fault characteristics extracted using these five methods were fed into the RF classifier for training and testing. Figure 25 shows the results of the five classification methods, and the identification accuracy rates are 97.04%, 98.15%, 99.26%, 91.11%, and 94.81%, respectively, which are all lower than that of the proposed method. Therefore, the proposed feature extraction model has excellent performance and strong generality for fan-end bearing data.

This section serves to verify the advantage of the developed TSMRE complexity measurement method in analyzing actual vibration signals over other methods. Typical signal complexity measurement methods were used for the comparison. The various operation steps were consistent with the previous experiments; Table 2 presents the parameter settings. To overcome the error induced in single experiments, each method was repeated 15 times.
and Table 7 present the classification results of the five methods. Clearly, the developed TSMRE data complexity measurement method produces the best fault identification results. The feature extraction method based on the DTCWPT–TSMRE achieves an average accuracy of 99.85%, with a relatively high stability. This proves the effectiveness of the DTCWPT–TSMRE model in solving fault diagnosis problems.

This part mainly studies the necessity of using the t-SNE method for dimensionality reduction of the fan-end bearing data. The first three features of the dimensionality-reduced new features and the first two features of the original high-dimensional features are respectively plotted in Figures 27 and 28. The categories of all the fault samples can be distinguished clearly after the dimensionality reduction and have excellent clustering, which is consistent with the previous conclusions. The samples not processed with t-SNE had poor separability. This indicates that features need to be compressed to obtain more integrated features. To further study the necessity of data dimensionality reduction, three fault features were randomly selected from the original high-dimensional data and fed into the RF classifier for training and testing to examine the effect of the data dimensionality reduction on the fault recognition rate. The classification accuracy of the five fault diagnosis methods without the t-SNE dimensionality reduction, listed in Table 8, is significantly lower than that listed in Table 7, which proves the necessity of reducing the dimension of the original features.
In addition, without using the t-SNE method for dimensionality reduction, the DTCWPT–TSMRE feature extraction method can still achieve an average accuracy rate of 93.31%, which is higher than those of the other four methods. This shows that the proposed model can efficiently mine high-quality fault features.

This part mainly verifies the advantage of the RF classifier for the classification. The SVM, ELM, BP, and ANN were used to compare with the RF. The proportion of training and test samples for each classifier was the same as that in the previous experiment. Table 9 presents the diagnostic results of the five methods using different classifiers. When different feature extraction models were used, the RF classification model produced the best recognition results, with an average accuracy of 95.11%, which was higher than those of the other classifiers. Furthermore, after inputting the different feature
extraction models into each classifier, the DTCWPT–TSMRE was found to exhibit the highest fault recognition rate, further proving the generality of the proposed feature extraction model.

V. CONCLUSION

This study developed a multistage detection method that combines health status detection and fault type detection to detect and judge the current state and fault types of rolling bearings. In the health detection stage, by calculating the RE value of the bearing vibration signal in different health states, a threshold is given to accurately separate the healthy and faulty bearings. In the fault identification stage, first, the fault vibration signal is decomposed into multiple frequency band components using the DTCWPT method. Subsequently, the TSMRE is used to mine the hidden fault features in each frequency band component to generate
an initial high-dimensional feature matrix. The t-SNE is then used to compress the features and construct a low-dimensional feature matrix. Finally, the fault types of the test samples are intelligently identified based on an RF classifier. The main contributions of this paper can be summarized as follows:

1. A multistage diagnosis scheme is proposed, which checks the bearing condition in advance and avoids missing normal samples for faulty samples due to the uncertainty in pattern recognition. RE was used as a reliable tool in this study to determine the presence of bearing failure.

2. Aiming at the drawbacks of insufficient and incomplete feature extraction of conventional coarse-grained segmentation method, a new complexity measurement considering the RE, namely TSMRE, is proposed. Compared with the other methods, such as TSMSE, TSMPE, MRE, and MSE, TSMRE showed better feature extraction performance, which was proven by both simulation signals and measured bearing vibration signals.

3. Combined with DTCWPT, t-SNE, and RF, a new bearing fault diagnosis method was established. The performance of the proposed method was verified by the measured bearing data. The results showed that this method can effectively determine the failure types of different bearings. Ablation experiments were performed to highlight the advantages of our method. The proposed combined method has evident advantages in each substage of the fault diagnosis, producing a better overall performance.

4. The method was also used for bearing fault diagnosis under two different working conditions. The results showed that the proposed method can achieve an accuracy of 100% under these working conditions. This indicates that the proposed method can effectively diagnose bearing faults under variable working conditions and has good practicability and generality.

In conclusion, compared with existing fault diagnosis methods, this paper introduces a health status detection step to reduce the potential uncertainties in pattern recognition, improves the reliability of the diagnosis, and enhances the efficiency of fault diagnosis. Therefore, it has important engineering application value and is suitable for engineering applications.

More studies are still required to expand the application scope of similar methods used in the fault diagnosis of mechanical equipment such as hydraulic pumps and planetary gearboxes [39], [40].

REFERENCES

[1] Z. K. Peng, P. W. Tse, and F. L. Chu, “A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing,” Mech. Syst. Signal Process., vol. 19, no. 5, pp. 974–988, Sep. 2005, doi: 10.1016/j.ymssp.2004.01.006.

[2] J. Dybala and R. Zimroz, “Rolling bearing diagnosing method based on empirical mode decomposition of machine vibration signal,” Appl. Acoust., vol. 77, pp. 195–203, Mar. 2014, doi: 10.1016/j.apacoust.2013.09.001.

[3] C. Liu, X. Wu, J. Mao, and X. Liu, “Acoustic emission signal processing for rolling bearing running state assessment using compressive sensing,” Mech. Syst. Signal Process., vol. 91, pp. 395–406, Jul. 2017, doi: 10.1016/j.ymssp.2016.12.010.

[4] Z. He, J. Chen, T. Wang, and F. Chu, Theories and Applications of Machinery Fault Diagnostics. Beijing, China: Higher Education Press, 2010.

[5] W. Deng, H. Liu, J. Xu, H. Zhao, and Y. Song, “An improved quantum-inspired differential evolution algorithm for deep belief network,” IEEE Trans. Instrum. Meas., vol. 69, no. 10, pp. 7319–7327, Oct. 2020, doi: 10.1109/TIM.2020.2983233.

[6] Z. Feng, X. Chen, and T. Wang, “Time–varying demodulation analysis for rolling bearing fault diagnosis under variable speed conditions,” J. Sound Vib., vol. 400, pp. 71–85, Jul. 2017, doi: 10.1016/j.jsv.2017.03.032.

[7] Z. Zhang, M. Jia, M. Wang, and S. Wang, “Inertial ltd. change and sliding window symbol sequence statistics-based early fault diagnosis for hydraulic pump on hydraulic tube tester,” Mech. Syst. Signal Process., vol. 23, pp. 1573–1585, Jul. 2009, doi: 10.1016/j.ymssp.2009.01.011.

[8] Q. Gao, C. Duan, H. Fan, and Q. Meng, “Rotating machine fault diagnosis using empirical mode decomposition,” Mech. Syst. Signal Process., vol. 22, no. 5, pp. 1072–1081, Jul. 2008, doi: 10.1016/j.ymssp.2007.10.003.

[9] J. Cheng, Y. Yang, and Y. Yang, “A rotating machinery fault diagnosis method based on local mean decomposition,” Digit. Signal Process., vol. 22, no. 2, pp. 356–366, Mar. 2012, doi: 10.1016/j.dsp.2011.09.008.

[10] P. Nguyen, M. Kang, J.-M. Kim, B.-H. Ahn, J.-M. Ha, and B.-K. Choi, “Robust condition monitoring of rolling element bearings using denoising and envelope analysis with signal decomposition techniques,” Expert Syst. Appl., vol. 42, no. 22, pp. 9024–9032, Dec. 2015, doi: 10.1016/j.eswa.2015.07.064.

[11] M. A. M. Agarwal, “Ensemble empirical mode decomposition: An adaptive method for noise reduction,” IJSCR Electron. Commun. Eng., vol. 5, no. 5, pp. 60–65, 2013, doi: 10.9790/2834-0556065.

[12] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” IEEE Trans. Signal Process., vol. 62, no. 3, pp. 531–544, Feb. 2014, doi: 10.1109/TSP.2013.2278602.

[13] Y. Liu, X. Wang, Z. Liu, X. Liang, and S. Si, “The entropy algorithm and its variants in the fault diagnosis of rotating machinery: A review,” IEEE Access, vol. 6, pp. 66723–66741, 2018, doi: 10.1109/ACCESS.2018.2873782.

[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, doi: 10.1109/TIM.2020.2981229.

[15] R. Yan and R. X. Gao, “Approximate entropy as a diagnostic tool for machine health monitoring,” Mech. Syst. Signal Process., vol. 21, no. 2, pp. 824–839, Feb. 2007, doi: 10.1016/j.ymssp.2006.02.009.

[16] M. Han and J. Pan, “A fault diagnosis method combined with LMD, sample entropy and energy ratio for roller bearings,” Measurement, vol. 76, pp. 7–19, Dec. 2015, doi: 10.1016/j.measurement.2015.08.019.

[17] A. Omidvarnia, M. Mesbah, M. Pedersen, and G. Jackson, “Range entropy: A bridge between signal complexity and self-similarity,” Entropy, vol. 20, no. 12, p. 962, Dec. 2018, doi: 10.3390/e20120962.

[18] Li, Gao, and Wang, “Reverse dispersion entropy: A new complexity measure for sensor signal,” Sensors, vol. 19, no. 23, p. 5203, Nov. 2019, doi: 10.3390/s19125203.

[19] L. Zheng, Y. He, and X. Chen, “Research on a fault diagnosis method for rolling bearing based on improved multiscale range entropy and hierarchical prototype,” Measurement, vol. 32, no. 9, Sep. 2021, Art. no. 095003, doi: 10.1088/1361-6501/abfbaa.

[20] X. Wang, S. Si, and Y. Li, “Multiscale diversity entropy: A novel dynamical measure for fault diagnosis of rotating machinery,” IEEE Trans. Ind. Informat., vol. 17, no. 8, pp. 5419–5429, Aug. 2021, doi: 10.1109/TII.2021.3078747.

[21] T. D. Pham, “Time-shift multiscale entropy analysis of physiological signals,” Entropy, vol. 19, no. 6, p. 257, Jun. 2017, doi: 10.3390/e19060257.

[22] Y. Chen, T. Zhang, W. Zhao, Z. Luo, and H. Lin, “Rotating machinery fault diagnosis based on improved multiscale amplitude-aware permutation entropy and multiclass relevance vector machine,” Sensors, vol. 19, no. 20, p. 4542, Oct. 2019, doi: 10.3390/s19204542.

[23] Y. Lei, J. Lin, Z. He, and M. J. Zuo, “A review on empirical mode decomposition in fault diagnosis of rotating machinery,” Mech. Syst. Signal Process., vol. 35, nos. 1–2, pp. 108–126, Feb. 2012, doi: 10.1016/j.ymssp.2012.09.015.
[24] I. Bayram and I. W. Selesnick, “On the dual-tree complex wavelet packet and M-band transforms,” IEEE Trans. Signal Process., vol. 56, no. 6, pp. 2298–2310, Jun. 2008, doi: 10.1109/tsp.2007.916129.

[25] N. Kingsbury, “The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters,” in Proc. IEEE Digit. Signal Process. Workshop, vol. 86, Aug. 1998, pp. 120–131.

[26] W. Liu, X. Yang, and S. Jinxing, “An integrated fault identification approach for rolling bearings based on dual-tree complex wavelet packet transform and generalized composite multiscale amplitude-aware permutation entropy,” Shock Vib., vol. 2020, pp. 1–18, Nov. 2020, doi: 10.1155/2020/8851310.

[27] J. Qu, Z. Zhang, and T. Gong, “A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion,” Neurocomputing, vol. 171, pp. 837–853, Jan. 2016, doi: 10.1016/j.neucom.2015.07.020.

[28] L. Wang, Z. Liu, H. Cao, and X. Zhang, “Subband averaging kurtogram with dual-tree complex wavelet packet transform for rotating machinery fault diagnosis,” Mech. Syst. Signal Process., vol. 142, Aug. 2020, Art. no. 106755, doi: 10.1016/j.ymssp.2020.106755.

[29] F. Zhou, J. Gong, X. Yang, T. Han, and Z. Yu, “A new gear intelligent fault diagnosis method based on refined composite hierarchical fluctuation dispersion entropy and manifold learning,” Measurement, vol. 186, Dec. 2021, Art. no. 110136, doi: 10.1016/j.measurement.2021.110136.

[30] N. Rogovschi, J. Kitazono, N. Grozavu, T. Omori, and S. Ozawa, “Distributed stochastic neighbor embedding spectral clustering,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), May 2017, pp. 1629–1632, doi: 10.1109/ijcnn.2017.7966046.

[31] S. Zhou, S. Qian, W. Chang, Y. Xiao, and Y. Cheng, “A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier,” Sensors, vol. 18, no. 6, p. 1934, 2018, doi: 10.3390/s18061934.

[32] L. Breiman, “Random forests,” Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/a:1010933404324.

[33] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, “Random forest: A classification and regression tool for compound classification and QSAR modeling,” J. Chem. Inf. Comput. Sci., vol. 43, no. 6, pp. 1947–1958, Nov. 2003, doi: 10.1021/ci034160g.

[34] R. Liu, B. Yang, E. Zio, and X. Chen, “Artificial intelligence for fault diagnosis of rotating machinery: A review,” Mech. Syst. Signal Process., vol. 108, pp. 33–47, Aug. 2018, doi: 10.1016/j.ymssp.2018.02.016.

[35] Z. Feng, M. Liang, and F. Chu, “Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples,” Mech. Syst. Signal Process., vol. 38, no. 1, pp. 165–205, Jul. 2013, doi: 10.1016/j.ymssp.2013.01.017.

[36] H. Shao, H. Liang, F. Wang, and Y. Yang, “Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet,” ISA Trans., vol. 69, pp. 187–201, Jul. 2017, doi: 10.1016/j.isatra.2017.03.017.

[37] S. Gu, H. Wang, and Y. Chen, “Application of double tree complex wavelet packet and adaptive permutation entropy in bearing fault diagnosis,” Mech. Sci. Technol. Aerosp. Eng., vol. 108, pp. 33–47, Jul. 2018, doi: 10.1021/ci034160g.

[38] Accessed: Oct. 15, 2018. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home

[39] K. Feng, K. Wang, M. Zhang, Q. Ni, and M. J. Zuo, “A diagnostic signal selection scheme for planetary gearbox vibration monitoring under non-stationary operational conditions,” Meas. Sci. Technol., vol. 28, no. 3, Mar. 2017, Art. no. 035003, doi: 10.1088/1361-6501/aa543e.

[40] K. Feng, K. Wang, Q. Ni, M. Zuo, and D. Wei, “A phase angle based diagnostic scheme to planetary gear faults diagnostics under non-stationary operational conditions,” J. Sound Vib., vol. 408, pp. 190–209, Nov. 2017, doi: 10.1016/j.wav.2017.07.030.

TAO HAN was born in 1989. He received the bachelor’s degree in mechanical engineering and automation from the PLA University of Science and Technology, China. He is currently pursuing the master’s degree with the College of Field Engineering, Army Engineering University of PLA. His current research interests include mechanics and fault diagnosis.

JIAN-CHENG GONG is currently pursuing the master’s degree with Army Engineering University of PLA, Nanjing, Jiangsu, China. He is also studying the application of machine learning in mechanical fault diagnosis. His research interest includes mechanical fault diagnosis.

XIAO-QIANG YANG was born in 1967. He received the Ph.D. degree in mechanical engineering from the Army Engineering University of PLA. His research interests include mechatronics, electronics, and fault diagnosis.

LI-ZHOU AN was born in 1979. He received the bachelor’s degree from the Army Engineering University of PLA. He is currently a Lecturer in mechanical engineering at the Army Engineering University of PLA. His research interests include mechatronics, maintenance, and fault diagnosis.