Measurement 130 (2018) 94–104

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Study on planetary gear fault diagnosis based on variational mode decomposition and deep neural networks

Yong Li [a], Gang Cheng [a,*], Chang Liu [a], Xihui Chen [b]

[a] School of Mechatronic Engineering, China University of Mining and Technology, 221116 Xuzhou, China
[b] College of Mechanical and Electrical Engineering, Hohai University, 213022 Changzhou, China

ARTICLE INFO

Article history:
Received 14 November 2017
Received in revised form 27 May 2018
Accepted 2 August 2018
Available online 3 August 2018

Keywords:
Planetary gear
Fault diagnosis
Variational mode decomposition
Power spectral entropy
Deep neural network

ABSTRACT

Planetary gear failures occur frequently in working conditions at low speeds, large loads, and closed operating environments, which makes the identification of faults a difficult task. A fault diagnosis method for planetary gear based on power spectral entropy of variational mode decomposition (VMD) and deep neural networks (DNN) is proposed herein. The three-axial vibration signals of a planetary gear are collected and decomposed into narrowband components with different frequency centres and bandwidths based on VMD. Power spectral entropy (PSE) is used as the original feature to represent the magnitude and distribution of the spectral amplitude of each component. A DNN based on an automatic encoder (AE) and back propagation neural network is used to realise the reduction of original signal features and the classification of gear states. The achieved overall recognition rate is 100% after the training of neural networks with training samples. The experimental results indicate that the proposed method is capable of extracting the sensitive features and recognising the fault states.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Planetary gears are a common epicyclic transmission system that has the advantages of small volume, high-transmission ratio and efficiency, and high-bearing capacity. They are extensively used in the fields of wind power generation, helicopters, shearer rocker arms, as well as in other engineering machinery. However, gear failures occur frequently in low-speed and large-load conditions, and their closed operating environments make it difficult to detect failures in time [1]. Vibration signal analysis is a common method used for fault diagnosis of planetary gears without requiring disassembly [2]. As an important part of fault diagnosis, fault type identification is used to judge the necessity of gear replacement without disassembling the equipment [3]. In addition, it provides good reference information for equipment maintenance and avoids wasting manpower and material resources. It is of great significance for efficient maintenance.

Compared to spur gears, the transmission paths and composition of vibration signals of planetary gear trains are more complex, thus leading to significant non-linear and non-stationary signals [4]. Currently, the methods of adaptive decomposition, such as empirical mode decomposition (EMD) and local mean decomposition (LMD), are commonly used to deal with non-linear and non-stationary vibration signals. Compared to the common time–frequency analyses methods, such as, short-time Fourier transforms and wavelet analyses, adaptive decomposition is based on the data-driven process of the signal to be decomposed, which does not need to select the base function in advance [5]. In addition, it is effective in dealing with non-linear and non-stationary vibration signals. However, there are a number of problem areas in the recursive adaptive decomposition method, such as modal aliasing and end effects. A number of improved methods are proposed to reduce modal aliasing and other issues by adding Gaussian white noise, such as ensemble empirical mode decomposition (EEMD) and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [6,7]. However, the effect in dealing with complex signals is not obvious. In order to avoid the problems caused by the recursive algorithm and obtain better signal decomposition components, a non-recursive adaptive signal processing method known as VMD, is selected. This method was proposed by Konstantin Dragomiretskiy in 2013 [8]. It is a process used to solve variational problems based on classical Wiener filtering, Hilbert transforms, and frequency mixing, and is different from the recursive filtering patterns of EMD and LMD [9]. This method determines the frequency and bandwidth of the centre of each decomposed component by iteratively searching for the optimal solution to the variational model [10]. It can adaptively decompose

* Corresponding author.
E-mail address: liyong2015@cumt.edu.cn (G. Cheng).

https://doi.org/10.1016/j.measurement.2018.08.002
0263-2241/© 2018 Elsevier Ltd. All rights reserved.

the signal into a sparse component, exhibiting better noise robustness. The vibration signal is decomposed into finite narrowband components with different frequency centres and bandwidths after VMD processing [11]. The small differences in the different original fault signals are amplified and displayed for each component. Feng proposed a planetary gearbox fault diagnosis method via joint amplitude and frequency demodulation analysis based on VMD [12]. Based on VMD decomposition, typical fault bands can be effectively detected. VMD decomposition provides stationary component signals for joint amplitude and frequency demodulation, and significantly improves the accuracy of fault diagnosis. However, the disadvantage of this non-recursive adaptive decomposition is that parameters, such as the decomposition level and the Lagrangian operator, need to be set in advance. The selection of parameters determines the effect of signal decomposition to a great extent. Therefore, it is necessary to analyse the nature of the signal and select the appropriate parameters in order to decompose the signal effectively.

The extraction of signal features is typically the process used to quantify the differences between different fault modes. When the bearing fails, the energy and the distribution of the fault vibration signal will change. These changes in energy and distribution can be used to extract features to distinguish different types of fault states effectively [13]. However, it is impossible to guarantee that the intercepted signal is a complete periodic signal. Compared to the frequency domain signal, the time domain signal is more sensitive to the selection of the interception time and is highly volatile [14]. Therefore, the stability of features extracted from the signal frequency domain is preferable. PSE is a parameter describing the complexity of one-dimensional sequences, which can satisfactorily quantify the magnitude and distribution of the amplitude of the spectrum [15,16]. Chen proposed a planetary gear fault diagnosis method based on entropy feature fusion of the ensemble empirical mode decomposition [6]. As one of the most important entropy features, PSE identifies the differences between different fault signal features. It is effective for the diagnosis of planetary gear faults, and the comprehensive recognition rate can reach 88.8%.

The structure of planetary gear trains is complex, and different types of fault signals contain a great number of common components. The constructed feature matrix contains abundant feature but also redundant information. Redundant information will interfere with the network structure when training the neural networks, thereby leading to unsatisfactory diagnosis results. DNN has received significant attention in the fields of machine learning and artificial intelligence in recent years. Through multi-layer nonlinear transformations, low-level features are combined to form more abstract high-level features [17]. Compared to single-layer neural networks, DNN exhibits strong non-linear expression capabilities. DNN comprises multiple non-supervision layers and one supervision layer. The non-supervised layers are used to mine the deep information of the data layer-by-layer, and the monitoring layer is used to classify the signals [18]. The DNN based on the AE and back propagation neural network (BPNN) constitutes a classic type of neural networks, the calculations are simple, and the recognition effect is good [19]. By encoding and decoding, the AEs can effectively remove redundant information and obtain representative low-dimensional space coding vectors. The DNN based on AEs effectively avoids the problems of poor generalisation abilities and large recognition errors in shallow neural networks [20]. The output layer with classification functions is then added, and the BPNN is used to adjust the parameters of the neural network by using the feedback error [21]. In order to excavate internal characteristics effectively, automatic coding is used to encode the original signal several times [22]. The trained neural network can establish the accurate mapping relationship between the input features and the output categories, and realise the identification and classification of different types of signals. Lei proposed a health monitoring method based on the deep learning theory for large data volumes of mechanical equipment [23]. The recognition rate of the equipment fault types was greatly improved.

In this study, a comprehensive fault diagnosis algorithm for planetary gears is proposed based on VMD–PSE–DNN. In order to describe the gear signals of various faults more accurately, a three-axis vibration sensor is used to obtain the three axial vibration signals [24]. The algorithm first identifies the differences in the frequency band of the fault signal. First, VMD is used to decompose the signal into narrow-band components with independent centre frequencies. The PSE is then used to quantify the size and distribution of the amplitudes of the side band in the components. Finally, DNN based on AEs is used to mine depth features of the feature vectors, and to further classify the fault signals. With these three steps, the relationship between the side band difference characteristics and the fault categories can be established, and fault signal recognition can be determined.

2. Model analysis

The fault diagnosis can be divided into four parts: signal decomposition, feature extraction, feature reduction, and pattern recognition. In this study, a comprehensive fault diagnosis algorithm for planetary gears is proposed, which is based on VMD–PSE–DNN, where VMD corresponds to signal decomposition, PSE corresponds to feature extraction, and DNN corresponds to feature reduction and pattern recognition. To better understand the algorithmic model, the detailed flow chart is shown in Fig. 1 below.

2.1. Variational mode decomposition

VMD is a new method of signal decomposition in which K modal functions are obtained by continually updating the centre frequencies and bandwidths with the smallest sum of estimated bandwidths [25]. The modal functions refer to a series of narrowband components with central frequencies. Combined with the characteristics of machinery fault signals, modal functions represent vibration signal components, such as the rotational frequencies of the motor, sun gear, planetary gear, and planetary frame, as well as the meshing frequency of the planetary gear system. The changes in the energy and distribution of the vibration signals with different faults are decomposed into individual components based on the differentiation of the different signals using VMD, and the highlighting of anomalies. The decomposition steps are as follows:

Step 1: Parameters of VMD should be initialised, including the decomposition layer K, the modal function $u_k(t)$, the centre frequency ω_k, the Lagrange multiplier, cycle times ($n = 0$), and convergence accuracy.

Step 2: The spectrum of each mode is modulated to the corresponding fundamental band. Hilbert transformation is used to transform each modal signal into single-side spectrums. By mixing the predicted centre frequencies of the single-side spectrum, the analytic signal is obtained. The analytic signal of the modal function is as follows,

$$\left[\left(\delta(t)+\frac{j}{\pi t}\right)u_k(t)\right]e^{-j\omega_k t} \tag{1}$$

where t is time, $\delta(t)$ is the impact function, $\{u_k(t)\} = \{u_1(t), u_2(t), ..., u_k(t)\}$ is defined as the set of the decomposed K IMF components, and $\{\omega_k\} = \{\omega_1, .., \omega_K\}$ is defined as the set of the central frequencies of each IMF component.

Step 3: The gradient of the L^2-norm of the demodulated signal is calculated and the bandwidth of each modal component is

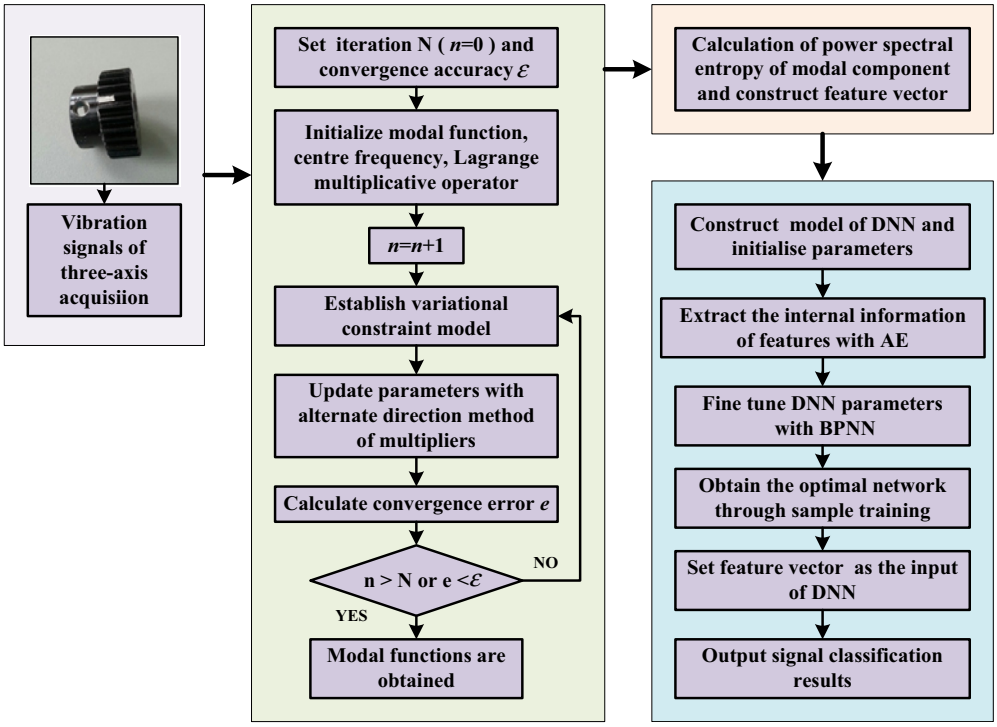

Fig. 1. Flow chart of the proposed VMD–PSE–DNN algorithm.

estimated. The corresponding constraint variational model is expressed as follows,

$$\left.\begin{aligned}&\min_{\{u_k(t)\},\{\omega_k\}}\left(\left\|\frac{\partial\left[\delta(t)+\frac{j}{\pi t}\right]u_k(t)}{\partial t}e^{-j\omega_k t}\right\|^2\right)\\&\sum_{k=1}^{K}u_k(t)=f(t)\end{aligned}\right\}\tag{2}$$

Step 4: The two-penalty factor α and the Lagrangian multiplication operator $\lambda(t)$ are introduced. The extended Lagrange expression is as follows,

$$\begin{aligned}L(\{u_k(t)\},\{\omega_k\},\lambda(t)) = \alpha\sum_{k=1}^{K}\left\|\frac{\partial\left[\delta(t)+\frac{j}{\pi t}\right]u_k(t)}{\partial t}e^{-j\omega_k t}\right\|_2^2\\+\left\|f(t)-\sum_{k=1}^{K}u_k(t)\right\|_2^2\\+\left\langle\lambda(t),f(t)-\sum_{k=1}^{K}u_k(t)\right\rangle\end{aligned}\tag{3}$$

where α can guarantee the reconstruction accuracy of the signal in the presence of Gaussian noise, and the Lagrangian operator $\lambda(t)$ makes the constraint condition strict.

Step 5: The alternate direction method of multipliers (ADMM) is used to compute the variation problem, and the saddle points of the Lagrangian expression can be obtained by updating the parameters $u_k^n(t)$, ω_k^n, and $\lambda^n(t)$, in an alternating manner. The process of solving $u_k^{n+1}(t)$ can be expressed as

$$\begin{aligned}u_k^{n+1}(t)=\arg\min\left\{\left\|\alpha\frac{\partial\left[\delta(t)+\frac{j}{\pi t}\right]u_k(t)}{\partial t}e^{-j\omega_k t}\right\|_2^2+\left\|f(t)-\sum_{i\neq k}^{K}u_i(t)+\frac{\lambda(t)}{2}\right\|_2^2\right\}\\i=\{1,2,\ldots,K\}\quad i\neq k\end{aligned}\tag{4}$$

Using the Parseval–Plancherel Fourier isometry under the L^2-norm, this problem can be solved in the spectral domain,

$$\begin{aligned}\hat{u}_k^{n+1}(t)=\arg\min\left\{\alpha\left\|j\omega\left[1+sgn(\omega+\omega_k)\right)\hat{u}_k(\omega+\omega_k)\right]\right\|_2^2\right.\\\left.+\left\|\hat{f}(\omega)-\sum_i\hat{u}_i(\omega)+\frac{\overset{\wedge}{\lambda}(\omega)}{2}\right\|_2^2\right\}\end{aligned}\tag{5}$$

A change of variables $\omega=\omega-\omega_t$ is then performed that leads to

$$\begin{aligned}\hat{u}_k^{n+1}(t)=\arg\min\left\{\alpha\left\|j(\omega-\omega_k)\left[1+sgn(\omega)\right)\hat{u}_k(\omega)\right]\right\|_2^2\right.\\\left.+\left\|\hat{f}(\omega)-\sum_i\hat{u}_i(\omega)+\frac{\overset{\wedge}{\lambda}(\omega)}{2}\right\|_2^2\right\}\end{aligned}\tag{6}$$

In the reconstruction fidelity term, both terms are written as half-space integrals over the non-negative frequencies by exploiting the Hermitian symmetry of the real signals:

$$\begin{aligned}\hat{u}_k^{n+1}(t)=\arg\min\left\{\int_0^{\infty}4\alpha(\omega-\omega_k)^2\left|\hat{u}_k(\omega)\right|^2\right.\\\left.+2\left|\hat{f}(\omega)-\sum_i\hat{u}_i(\omega)+\frac{\overset{\wedge}{\lambda}(\omega)}{2}\right|^2d\omega\right\}\end{aligned}\tag{7}$$

The first variation for the positive frequencies is removed to solve the quadratic optimisation problem,

$$\hat{u}_k^{n+1}(\omega)=\frac{\hat{f}-\sum_{i\neq k}\hat{u}_i(\omega)+\frac{\hat{\lambda}(\omega)}{2}}{1+2\alpha(\omega-\omega_k)^2}\tag{8}$$

In the formula listed above, $\hat{u}_k^{n+1}(\omega)$ is equivalent to Wiener filtering of $\hat{f}-\sum_{i\neq k}\hat{u}_i(\omega)\hat{f}$, and $\{u_k(t)\}$ is the real part of $\{\hat{u}_k(\omega)\}$ after Fourier transformation.

The centre frequency is updated using the same process,

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d_\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d_\omega}, \quad k \in \{1, 2, ..., K\} \tag{9}$$

The updated formula of the multiplication operator is as follows

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)) \tag{10}$$

where τ is a time constant, typically equal to zero.

Step 6: Update u_k^{n+1}, ω_k^{n+1}, and $\lambda^{n+1}(\omega)$, by repeating step 5 until the convergence accuracy is satisfied, or until $n > N$. The convergence accuracy is typically considered equal to $1e^{-6}$.

$$\sum_{k=1}^{K} \frac{\|\hat{u}_k^{n+1}(\omega) - \hat{u}_k^n(\omega)\|}{\|\hat{u}_k^n\|} < \varepsilon \tag{11}$$

The selection of VMD parameters is important in this study. In order to achieve a better decomposition effect, the following steps are adopted.

Step 1: FFT is used to translate the time domain signal into a frequency domain signal first and the appropriate number of components K is selected based on the peak distribution of the signal spectrum.

Step 2: The appropriate α value is selected according to the sideband width of the demand and the distribution of the central frequency of each component.

2.2. Power spectral entropy

The magnitude and distribution of the vibration signal energy generated by gear meshing differ for different types of faults. Entropy is a parameter describing the chaotic degree of one-dimensional sequences, which can accurately reflect the magnitude and distribution of the spectrum, and the entropy has better stability than the general characteristics [26]. The frequency domain signal is more stable than the time domain signal in the case of incomplete signal interception periods [27]. Power spectral entropy is used to quantify the degree of chaos of various types of fault vibration signals from the magnitude and distribution of the frequency domain amplitudes. The discrete Fourier transform of the single channel signal $\{x_t\}$ can be set to $X(f_i)$, and the power spectrum is then expressed as follows:

$$S(f_i) = |X(f_i)|^2 / N \tag{12}$$

where N is the number of signals sampled.

The power of a certain frequency accounting for a proportion of the total power can be set as

$$p_i = X(f_i)/sum(X(f_i)) \tag{13}$$

According to the definition of entropy, PSE can be expressed as follows

$$H(X) = \sum_{i=1}^{n} p_i \lg p_i \tag{14}$$

2.3. Deep neural network

The differences between DNN and ordinary neural networks lie in the training process. The network training process is divided into two steps: unsupervised pre-training and supervised fine-tuning training [28]. The trained neural network can effectively mine the deep information in the data and realise data classification. A DNN based on AE and BP neural networks is used in this study.

An AE is a type of an unsupervised neural network, which comprises coding and decoding networks [29]. The input data is the same as the output target. The input data in high-dimensional space is transformed into encoding vectors in low-dimensional space with the coding network, and the encoding vector of low-dimensional space is reconstructed back to the original input data by the decoding network. The AE structure is shown in Fig. 2.

As a classical neural network monitoring algorithm, the BP neural network algorithm can feed back errors, and fine-tune network parameters. The training process of the DNN, which comprises AE and BPNN, is as follows:

Step 1: The coding network is used to code the sample $\{x_m\}$ to low-rank coding h_{m1}, $\{x_m\}(m = 1 \sim M)$,

$$h_{m1} = s_f(w_1 \cdot x_m + b_1) \tag{15}$$

$$x'_m = s_g(w'_1 \cdot h_{m1} + b'_1) \tag{16}$$

where sigmoid s_f and s_g are the activation functions of the coding and decoding networks, respectively, and w_1, w_1', b_1, and b_1', are the link weights of the coding and decoding networks, respectively.

Step 2: In order to minimise and reconstruct the error $L(x_m, x'_m)$, the training of a single decoding network is completed by training the samples,

$$L(x_m, x_m') = \frac{1}{M} \|x_m - x_m'\|^2 \tag{17}$$

Step 3: The encoding h_{m1} is defined as the set of input parameters and output targets of the next AE. The coding h_{m2} set can be obtained by repeating steps 1 and 2. In this manner, the coding h_{mn} can be obtained as follows,

$$h_{mn} = s_f(w_{n-1} \cdot h_{m(n-1)} + b_{n-1}) \tag{18}$$

Step 4: The coding h_{mn} is set as the input, and the coding $h_{m(n+1)}$ is obtained by the coding network of the $(n+1)^{th}$ layer. The classification label L_m is set as the output of the final layer. The decoding network with the label of the final layer is expressed as follows

$$h_{m(n+1)} = s_f(w_n \cdot h_{mn} + b_n) \tag{19}$$

$$L_m = s_g(w_n' \cdot h_{m(n+1)} + b_n') \tag{20}$$

Step 5: Using backward propagation, the connection weights between different layers are fine-tuned from the final layer with the training samples.

Step 6: Repeat steps 1–5 until n iterations are completed. The training of the network is completed.

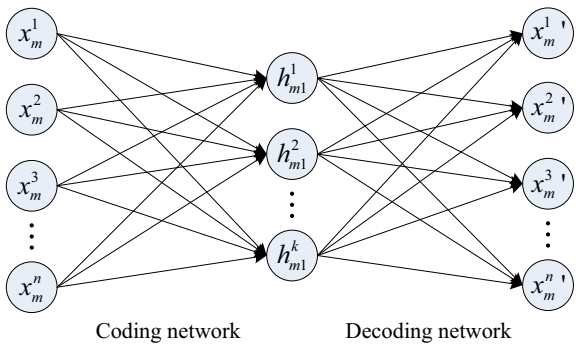

Fig. 2. AE model structure.

3. Simulation analyses

A simulation signal is used to verify the selection of VMD parameters in this article.

$$s = s_1 * s_a + s_2 * s_a + s_3 * s_a + s_4 * s_a + s_1 * s_b + s_2 * s_b + s_3 * s_b + s_4 * s_b + s_n \quad (21)$$

where t represents the sampling time interval, s_1, s_2, s_3, and s_4 are base signals, s_a and s_b are amplitude-frequency modulation signals, s_n is a noise signal, randn represents the function given to the initial value at random and the sampling frequency is 1000 Hz.

$$\begin{aligned} s_1 &= \cos(2 * \pi * 34 * t) \\ s_2 &= \cos(2 * \pi * 168 * t) \\ s_3 &= \cos(2 * \pi * 266 * t) \\ s_4 &= \cos(2 * \pi * 428 * t) \\ s_a &= \cos(2 * \pi * 6 * t) \\ s_b &= \cos(2 * \pi * 12 * t) \\ s_n &= 2 * randn(size(s_1)) \end{aligned} \quad (22)$$

The time domain and frequency domain images of the synthesised signal are shown in Fig. 3.

It can be observed from Fig. 3 that the signal contains four spectral peaks, so the K value is equal to four. When the K value is four, the central frequency distribution for different α values is listed in Table 1.

It can be observed from the table that the appropriate centre frequency can be estimated more accurately with α values ranging between 1000 and 5000. The value of α then needs to be selected according to the bandwidth of the desired sideband. Appropriate bandwidths can effectively avoid excessive noise sideband interference to each component and improve the signal quality of each component. The distribution range of the sidebands is approximately 30 Hz. When $\alpha =$ 4000, the bandwidth control range can be attained. Therefore, the value of α is set to 4000. The signal decomposition effect is shown in the following Fig. 4. The method effectively decomposes the simulation signal.

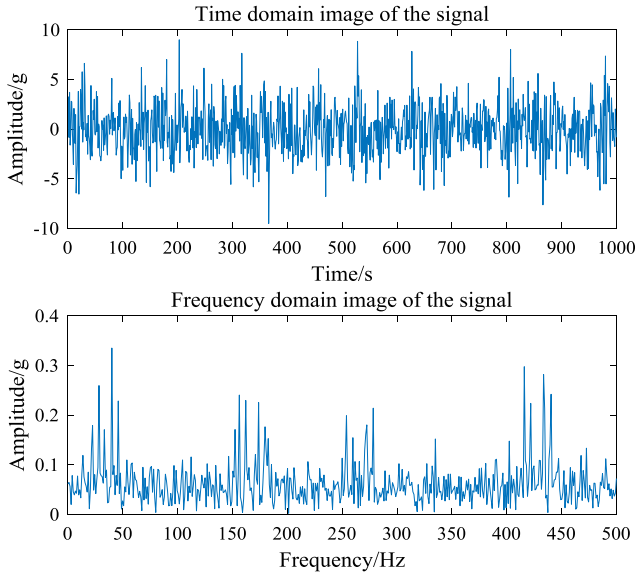

Fig. 3. Time domain and frequency domain diagrams of simulation signals.

Table 1
Central frequency of each component at different α values (K = 4).

Value a	Central frequencies of components/Hz			
	1	2	3	4
100	42.1	165.3	270.2	424.5
1000	33.6	164.0	268.3	428.8
2000	30.2	171.0	268.87	427.6
3000	35.7	166.1	261.2	423.9
4000	36.5	169.0	269.0	422.8
5000	38.1	168.4	270.5	422.6
6000	34.6	173.8	267.7	382.7

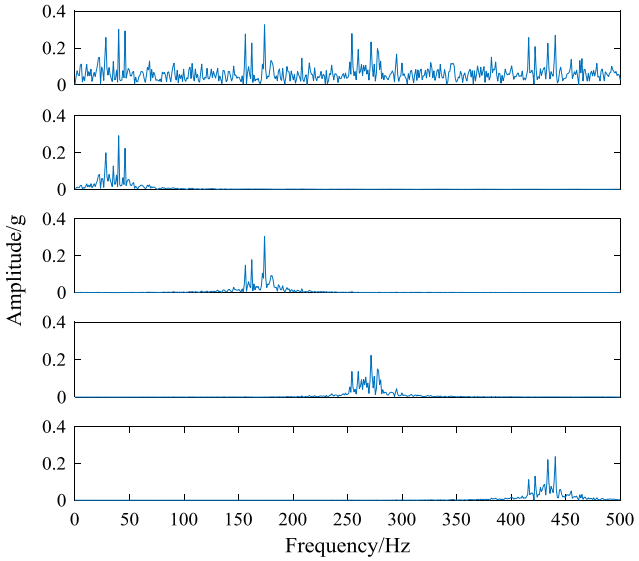

Fig. 4. Decomposition effect diagram of the simulation signal.

4. Experimental analysis

4.1. Experimental equipment

The Spectra Quest synthetic simulation platform for mechanical faults was used in this study, as shown in Fig. 5. The experimental equipment includes a control motor, a planetary gear box, a gear box, a magnetic brake, an acceleration sensor, a data acquisition instrument, and a portable computer. The acceleration sensor and data acquisition instrument are used to measure the planetary gear vibration signals. The portable computer was used to control the motor speed, set the braking torque, and record the vibration signals. The sun gear faults of the second stage planetary gear were also simulated. Four gear states are shown in Fig. 6, namely, a normal gear, a broken gear, a gear with a root crack, and a wear gear. The motor output speed was set at 40 Hz, the brake load was set to 20 Nm, the sampling frequency was set to 12,800 Hz, and each sample comprised 6400 data points.

The experiment was roughly divided into four steps:

Step 1: Set the frequency of the motor at 40 Hz, assign a torque value of 20 Nm and the sampling frequency to 12,800 Hz.
Step 2: Start the motor and wait until the amplitude of the signal waveform is stabilised
Step 3: Replace the fault gear and keep all parameters unchanged. Repeat step 2, and complete all data acquisition.
Step 4: The signal fragment is intercepted at 0.5 s. A total of 400 sets of signal fragments were obtained, and the number of sampling points of each set was 6400.

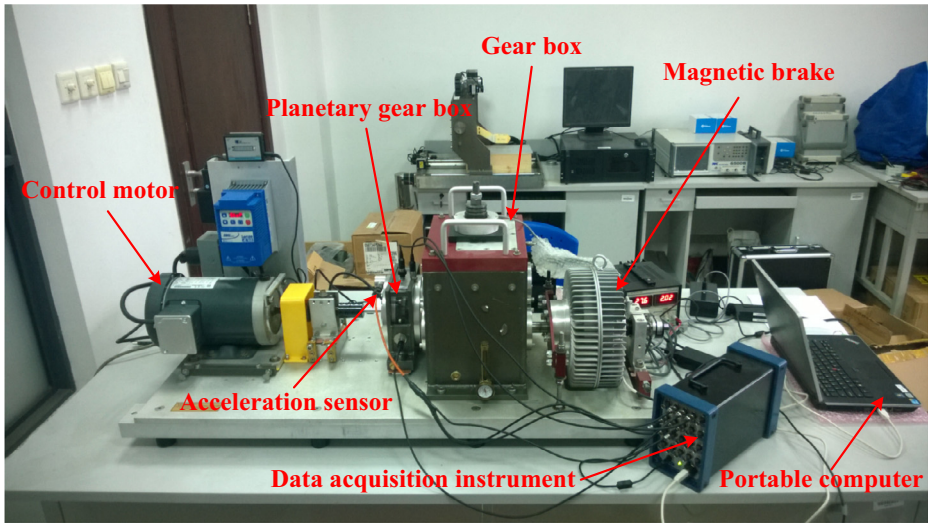

Fig. 5. Spectra Quest synthetic simulation platform for mechanical faults.

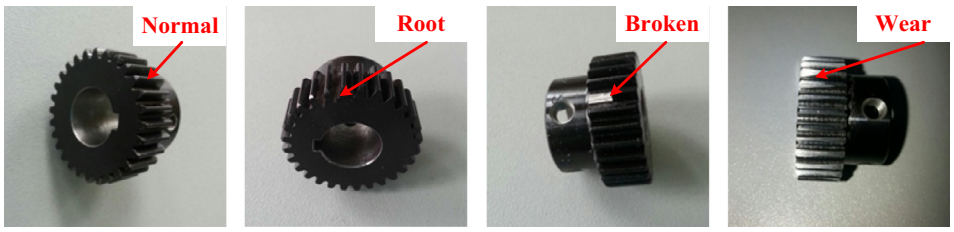

Fig. 6. Gear states.

4.2. Signal decomposition based on VMD

A planetary gear fault diagnosis method based on the PSE of VMD and DNN is proposed. In order to test the correctness and effectiveness of the method, a second-stage solar gear is used in this study, and the method was verified by analysing different fault states. Three axial vibration signals of four faults states were collected. Because of the large number of samples, only the x-axis vibration signals of the fault states were used as examples. When a fault occurs, the magnitude and distribution of the vibration signal energy changes. The time domain and frequency domain signals of the four faults are shown in Fig. 7. These changes are not obvious in the time-domain, and it is difficult to differentiate between different fault vibration signals. However, the differences between different types of fault signals in the signal frequency domain are more obvious, such as the amplitude and side-band distribution of gear meshing frequencies, and the doubling of the frequency.

In order to amplify the differences of the spectrums, the VMD method was used to process the original vibration signals, which were decomposed into K narrow-band components with different frequencies and bandwidths using VMD. The decomposition layer K and the two-penalty factor α are two key parameters in the VMD. The selection of the optimal parameter combinations is critical to achieve the effective separation of signal components. The parameter K determines the decomposition level of the signal, and the parameter α limits the bandwidth of the signal. The smaller the α value is, the greater the bandwidth of each component is. Considering the characteristics of various fault signals, K and α need to be universally applicable, and the decomposition itself is a differentiated process. Therefore, it is not necessary to use correlation optimisation algorithms to accurately calculate their values. The selection process used in this study is as follows.

First, the range of the number of the modal components is determined according to the peak distributions of the signal spectrum. As it can be observed in Fig. 7, the peaks of the spectrum are primarily the motor frequency conversion, the meshing frequency of the planetary gear train, and the doubling of the first seven frequencies. Therefore, a K value that is approximately equal to nine can be used to achieve better signal decomposition. The range of K is from 7 to 11. The cross-validation method is adopted in this study. The central frequency distribution and the degree of edge-band crossing are set as the reference standard of the signal decomposition effect. Various α values from five K values are tested, using significant amounts of experimental data. When K is within the range of 7–8, the meshing frequency of the gear is discarded when α is too large, and a component has two central frequencies when α is too small. When K equals 10 or 11, the sideband frequency increases, the noise increases, and the repetition of the two meshing frequencies occur when α is too small. The same central frequency will appear in different components when α is too large. When K equals 9, the aforementioned problem is eliminated and the signal can be better decomposed. The central frequencies of each component under different α values are shown in Table 2.

It can be observed from Table 2 that when α is 4000, the centre frequency is stable. When α is too small, the bandwidth is too wide and the centre frequencies deviate from the ideal peak. When α is too large, the bandwidth is too narrow and characteristic information is lost.

In order to verify that the effect of VMD is superior to other signal decomposition methods, such as LMD and EEMD, these

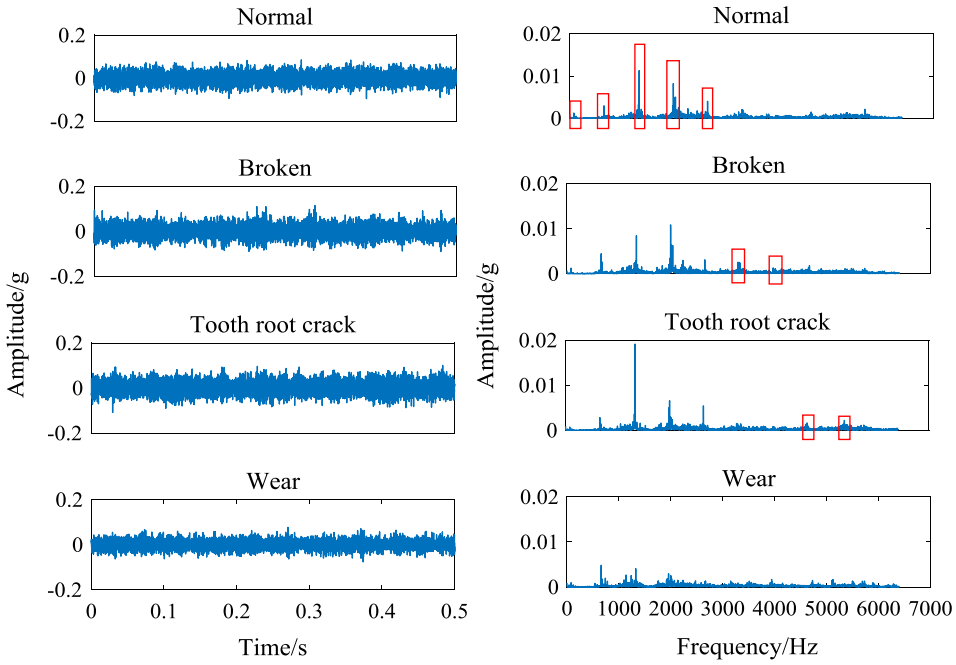

Fig. 7. Time domain and frequency domain of vibration signals obtained from four faults.

Table 2
Central frequency of each component at different α values ($K = 9$).

Valueα	Central frequencies of components/Hz								
	1	2	3	4	5	6	7	8	9
$\alpha = 100$	216.4	1322.5	1363.9	2003.0	2700.9	4009.2	4978.6	5486.5	6015.3
$\alpha = 1000$	119.4	1150.6	1.3545	1982.9	2606.8	3077.7	4034.8	5323.8	5941.1
$\alpha = 2000$	104.9	1158.4	1351.0	1990.9	2597.9	3144.6	4061.4	5284.3	5696.5
$\alpha = 3000$	89.6	680.4	1342.0	1990.2	2602.6	3133.1	4049.2	5273.6	5624.1
$\alpha = 4000$	89.3	676.7	1341.0	1995.3	2608.5	3160.1	4037.6	4946.1	5465.6
$\alpha = 5000$	89.2	674.2	1340.2	1998.6	2613.3	3180.3	4031.4	4932.4	5449.7

methods are tested with the original fault signal in this study. The decomposition results of LMD are shown in Fig. 8, and the decomposition results of VMD are shown in Fig. 9. It can be observed in Fig. 8 that a significant number of modal aliasing phenomena occur in the decomposed components, which will lead to negative separation effects of the original signal. This problem is not evident in VMD. In Fig. 9, the detailed features of each centre frequency and its sidebands are shown in detail, and the differences of the different faults are highlighted.

4.3. Feature extraction based on PSE

The PSE can satisfactorily quantify the magnitude and distribution of the spectrum amplitude. However, different types of fault signals contain numerous similar components. In order to describe signals in detail, three axial vibration signals are processed, and 3×9 fault features are obtained.

Fig. 10 shows the PSE characteristic curves of four faults. It can be observed in this figure that the power spectrum curves of different types of faults have a number of differences, which can allow the distinction of the different types of signals in a better manner. However, there are also a number of insensitive features, which are depicted within the red-coloured frames.

Fig. 11 shows the PSE characteristic curves of four broken samples. It can be observed that the eigenvalues of the bulk of the points are similar, but there are a number of unstable eigenvalues exhibiting greater disparities, as indicated in the red-coloured frames.

4.4. Feature reduction and fault recognition based on DNN

PSE characteristics based on VMD and the three-axis signals proposed in this study show the differences between different fault type signals. However, there are also some insensitive and unstable feature points, which are not conducive to fault-type recognition. In order to eliminate the influence of these features on fault identification, DNN is used in this study, which is based on AE and BPNN. DNN can effectively mine the deep characteristics of data, which can effectively enhance the diversity of different types of signals and the similarity of the same type of signals. Compared to the traditional methods, DNN is simple, requires a few calculations, and has a high-recognition rate. DNN typically comprises multilayer unsupervised networks and a supervised network. The network structure is determined by the number of neurons of the first and the final layers. There are four types of faults and 27 fault features which were extracted from the three axial signals. It can be observed that the neural network structure is simple, which makes the number of neurons in the middle layer insensitive to the overall recognition rate. Several parameters are also tested in the experiment to validate it. This article selects the intermediate value of 12 and completes the structure of the network.

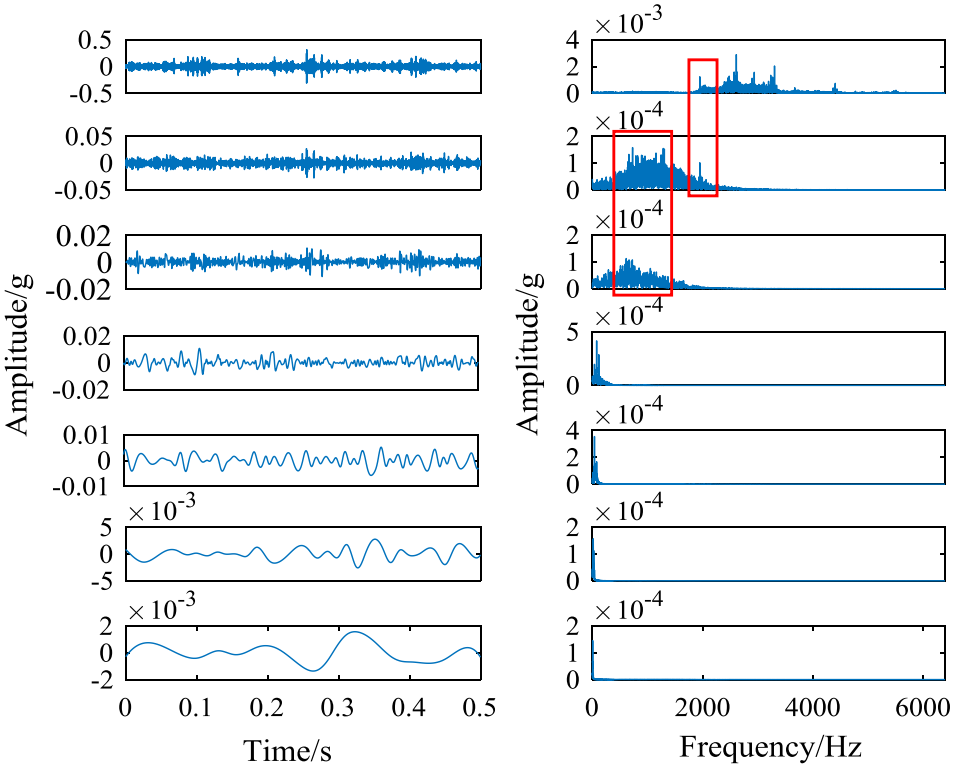

Fig. 8. Decomposition results of LMD.

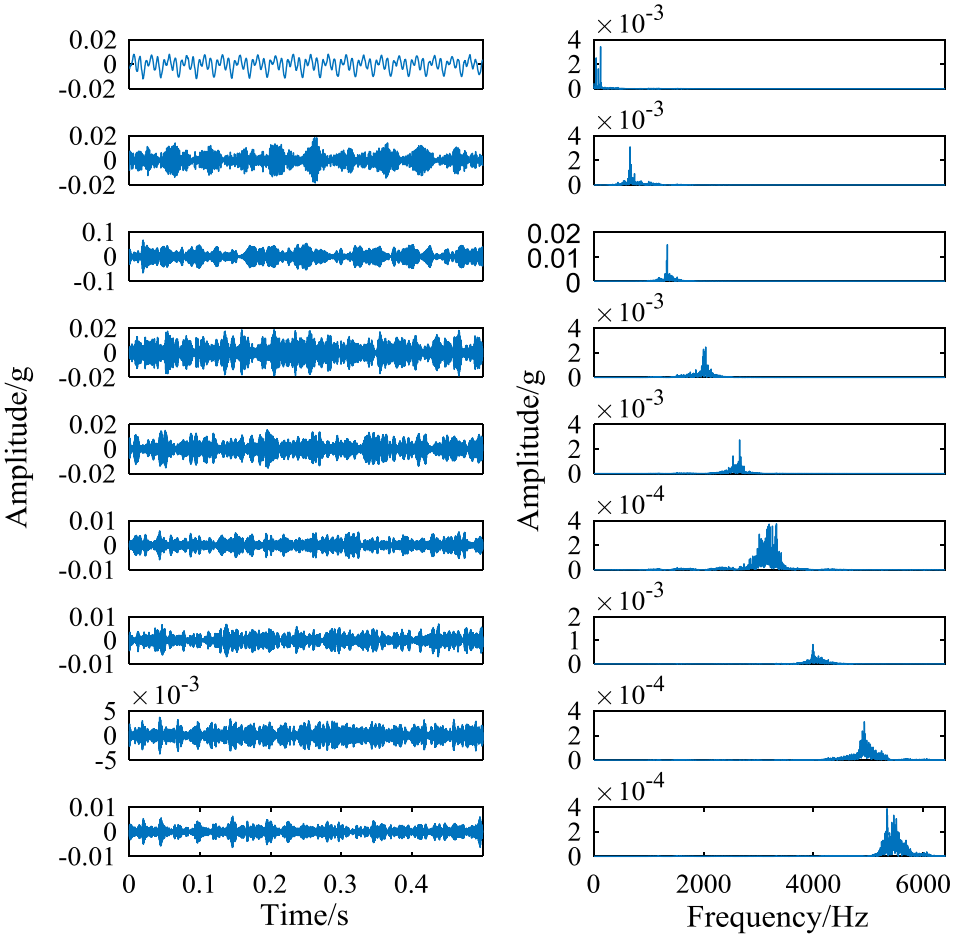

Fig. 9. Decomposition results of VMD ($K = 9$, $\alpha = 4000$).

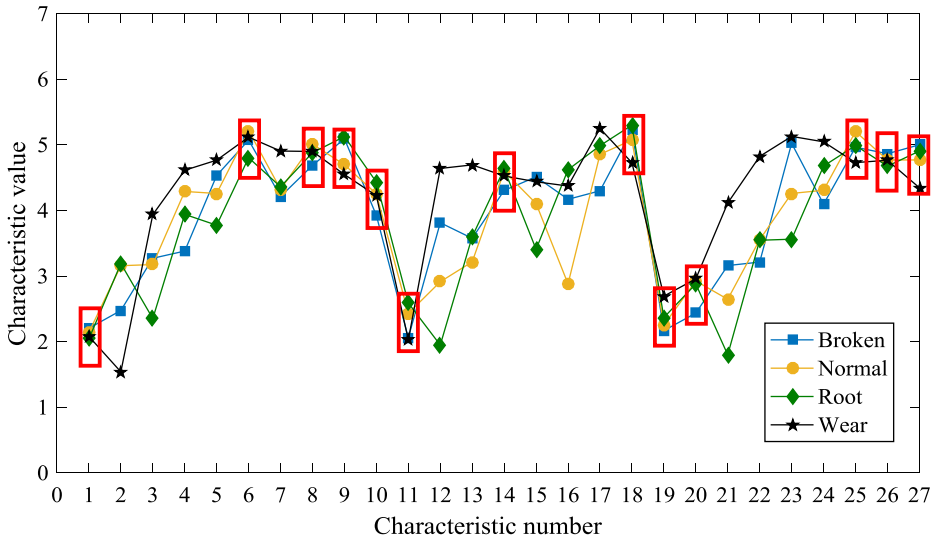

Fig. 10. Power spectrum entropy characteristic curve of four-fault signals.

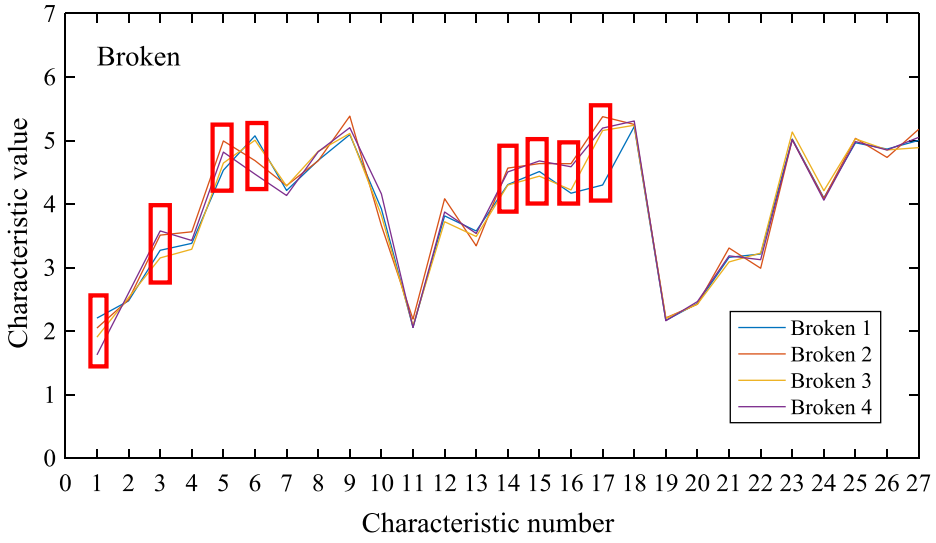

Fig. 11. Power spectral entropy characteristic curves of four-broken samples.

Network training is divided into two steps: unsupervised training and supervised fine tuning. Unsupervised training is mainly accomplished by AE. An AE is divided into three layers: input, coding, and output layers. When there is no supervised training, the input and output layers are consistent, and all of them are high-dimensional feature vectors. Through multiple coding dimensional reduction and decoding reconstruction, the low-dimensional feature vectors with significant differences can be effectively obtained. At the end of the supervised networks, the BP neural network and the practical sample labels are added. The network parameters are further modified through error reverse transmission to improve the overall network quality.

Comparison curves of the coding features of the four states are shown in Fig. 12, and comparison curves of the coding features of the four group signals of the broken tooth fault are shown in Fig. 13. Compared to Fig. 10, it can be observed in Fig. 12 that the insensitivity feature is more concentrated, and the sensitive feature is more discrete. The internal links of feature data are extracted and the feature quality has effectively improved. By comparing Figs. 11 and 13, it can be observed that the characteristic fluctuation of the same signal is improved, which greatly reduces the interference of the unstable feature to the network training.

A total of 400 sets of vibration signal data of four sets of faults were recorded. From these, 60 sets of each fault were used to train the neural network, and 40 sets from each fault were used to verify the recognition rate of the neural network. The normal, broken, cracked root, and the wear gears are represented by [1 0 0 0], [0 1 0 0], [0 0 1 0], and [0 0 0 1], respectively. In order to improve the recognition accuracy and reduce the computational time, the recognition rates of 1–30 sample iterations were tested.

As it can be observed in Fig. 14, the recognition rate increases with increasing iteration training samples. When the number of iterations reaches 17, the recognition rate has reached 100%, which can accurately distinguish the four types of gear states. Therefore, the number of iterations selected in this study is 17.

In order to highlight the superiority of DNN, the experiment was repeated with both BP and support vector machines (SVM). Through iterative feature extraction, DNN reduced the redundancy of data features effectively, and effectively overcame the error recognition problem of traditional neural networks in processing

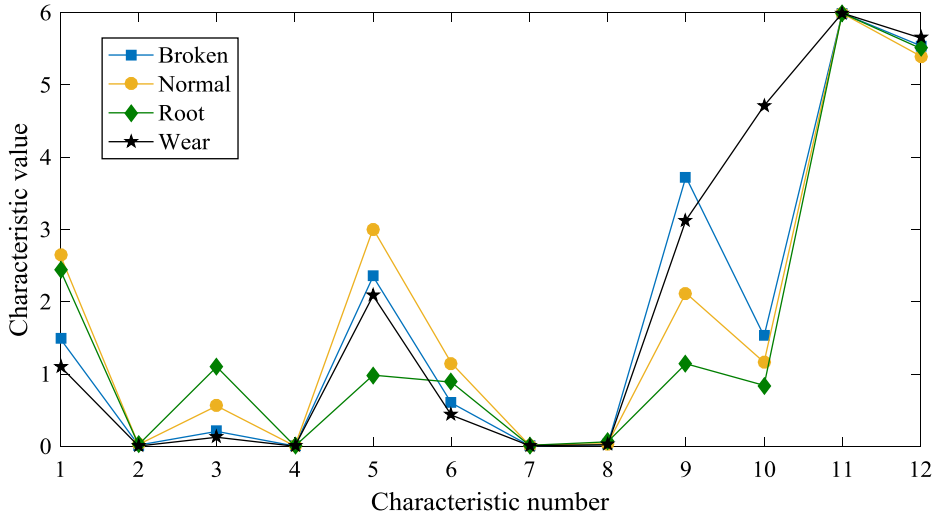

Fig. 12. Comparison curves of coding features of four fault states.

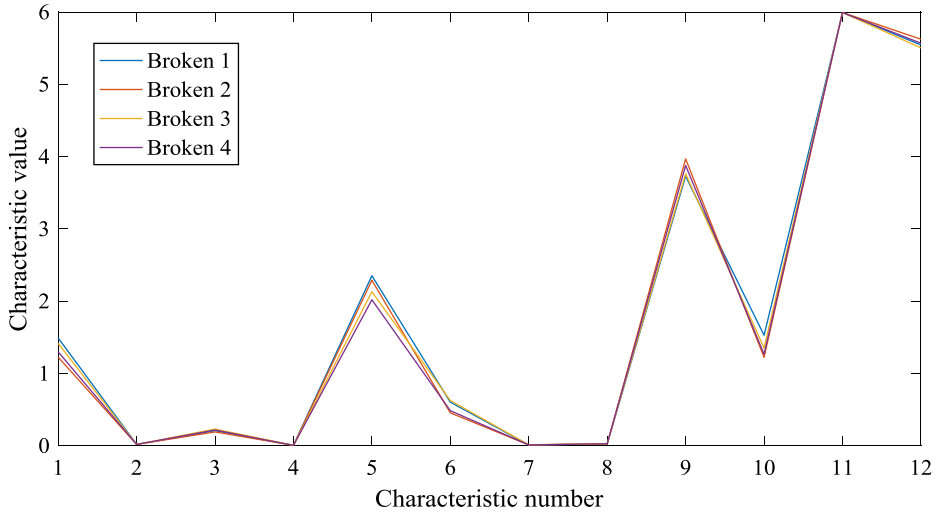

Fig. 13. Comparison curves of coding features of four signals of broken tooth.

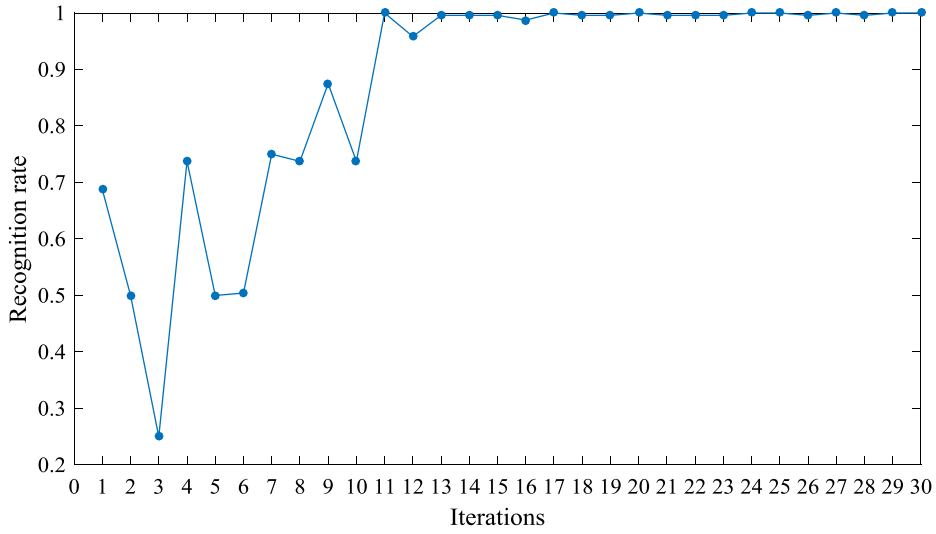

Fig. 14. Overall recognition rate spanning 1–30 iterations.

Table 3
Recognition effects of different neural networks.

Method	Recognition rate				
	Overall	Normal	Broken	Wear	Root
DNN	100%	100%	100%	100%	100%
BP	89.25%	87%	86%	90%	94%
SVM	87.25%	88%	85%	89%	87%

multi-feature inputs. The identification results are presented in Table 3.

DNN constructs a more accurate mapping relationship between feature vectors and fault types. It can be observed from Table 3 that when the signal decomposition method and feature extraction selection remain unchanged, DNN can effectively improve the recognition rate of the fault types compared to other neural networks. The elicited increased recognition rates justify the validation of the method proposed in this study.

5. Conclusion

A planetary gear fault diagnosis method based on power spectrum entropy of VMD and DNN was proposed. The single axial vibration signal of the planetary gear was decomposed into nine narrowband components with different centre frequencies and bandwidths using VMD. The energy and distribution changes caused by different faults were decomposed into individual components, which were defined as characteristic parameters. PSE was then used to quantify the magnitude and distribution of the energy. The PSE values of 27 three-axial components were extracted as fault features. However, there were a number of instabilities or insensitivities in individual features, which interfered with the recognition of gear states. Therefore, DNN was used to eliminate the characteristic interferences, and realised fault recognition. The PSE feature was defined as the input of the DNN model, and the marker value of different planetary gears was defined as the output of the DNN model. The structure of the DNN was 27–12–4. In order to adjust the network connection parameters and obtain the optimal DNN model, the training samples were used to iteratively train the neural network. Test samples were used to evaluate the quality of the trained DNN. The recognition rate of four gear states attained values of 100% simultaneously. This indicated that the trained DNN model performed satisfactorily, and that this method can be used for planetary gear fault diagnosis.

Acknowledgements

This work was supported by a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions and Fundamental Research Funds for the Central Universities (grant number 2015YC02). This support is gratefully acknowledged.

References

[1] Z.P. Feng, M. Liang, F.L. Chu, Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples, Mech. Syst. Sig. Process. 38 (1) (2013) 165–205.
[2] Z.P. Feng, M.J. ZUO, Vibration signal models for fault diagnosis of planetary gearboxes, J. Sound Vib. 331 (2012) 4919–4939.
[3] G. Cheng, Y.L. Cheng, L.H. Shen, J.B. Qiu, S. Zhang, Gear fault identification based on Hilbert-Huang transform and SOM neural network, Measurement 46 (3) (2013) 1137–1146.
[4] Y.G. Lei, W. Tang, D.T. Kong, J. Lin, Vibration Signal Simulation and Fault Diagnosis of Planetary Gearboxes Based on Transmission Mechanism Analysis, J. Mech. Eng. 50 (17) (2014) 61–68.
[5] K. Thameur, T. Marc, G. Raynald, Comparison between the efficiency of LMD and EMD algorithms for early detection of gear defects, Mech. Industry 14 (2) (2013) 121–127.
[6] G. Cheng, X.H. Chen, H.Y. Li, P. Li, H.G. Liu, Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposition, Measurement 91 (2016) 140–154.
[7] M.I. Bouhalais, A. Djebala, N. Ouelaa, CEEMDAN and OWMRA as a hybrid method for rolling bearing fault diagnosis under variable speed, Int. J. Adv. Manuf. Technol. 94 (5–8) (2018) 2475–2489.
[8] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (3) (2014) 531–544.
[9] X.L. An, L.P. Pan, F. Zhang, Analysis of hydropower unit vibration signals based on variational mode decomposition, J. Vib. Control 23 (12) (2017) 1938–1953.
[10] M. Hafida, C. Fakher, F. Ahmed, Detection of gear faults in variable rotating speed using variational mode decomposition (VMD), Mech. Industry 17 (2) (2016), 207 U81.
[11] X.L. An, H.T. Zeng, C.H. Li, Envelope demodulation based on variational mode decomposition for gear fault diagnosis, Proc. Inst. Mech. Eng. 231 (4) (2017) 864–870.
[12] Z.P. Feng, D. Zhang, M.J. Zuo, Planetary Gearbox Fault diagnosis via Joint Amplitude and Frequency Demodulation Analysis Based on Variational Mode Decomposition, Appl. Sci. 7 (8) (2017) 775.
[13] M. Inalpolat, A. Kahraman, A theoretical and experimental investigation of modulation sidebands of planetary gear sets, J. Sound Vib. 323 (2009) 677–696.
[14] M. Li, X.Y. Liu, X. Liu, Infrasound signal classification based on spectral entropy and support vector machine, Appl. Acoust. 113 (2016) 116–120.
[15] T.J. Ji, X.B. Wang, Z.B. Liu, EEMD-based online milling chatter detection by fractal dimension and power spectral entropy, Int. J. Adv. Manuf. Technol. 92 (1–4) (2017) 1185–1200.
[16] X.H. Chen, G. Cheng, H.Y. Li, M. Zhang, Diagnosing planetary gear faults using the fuzzy entropy of LMD and ANFIS, J. Mech. Sci. Technol. 30 (6) (2016) 2453–2462.
[17] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[18] Z.Q. Chen, S.C. Deng, X.D. Chen, et al., Deep neural networks-based rolling bearing fault diagnosis, Microelectron. Reliab. 75 (2017) 327–333.
[19] J. Ma, S.H. Ni, W.J. Xie, Deep Auto-encoder Observer Multiple-model Fast Aircraft Actuator Fault Diagnosis Algorithm, Int. J. Control Autom. Syst. 15 (4) (2017) 1641–1650.
[20] M. Xia, T. Li, L.Z. Liu, L. Xu, C.W. de Silva, Intelligent fault diagnosis approach with unsupervised feature learning by stacked denoising autoencoder, Sci. Measur. Technol. 11 (6) (2016) 687–695.
[21] L.A. Wulandhari, A. Wibowo, M.I. Desa, Condition diagnosis of multiple bearings using adaptive operator probabilities in genetic algorithms and back propagation neural networks, Neural Comput. Appl. 26 (1) (2015) 57–65.
[22] J.D. Sun, C.H. Sun, J.T. Wen, Intelligent Bearing Fault Diagnosis Method Combining Compressed Data Acquisition and Deep Learning, IEEE Trans. Instrum. Meas. 67 (1) (2018) 185–195.
[23] Y.G. Lei, F. Jia, X. Zhou, J. Lin, A Deep Learning-based Method for Machinery Health Monitoring with Big Data 51 (21) (2015) 49–56.
[24] X.Y. Gong, L.L. Ding, W.L. Du, H.C. Wang, Gear Fault Diagnosis Using Dual Channel Data Fusion and EEMD Method, Procedia Eng. 174 (2017) 918–926.
[25] Y.X. Wan, R. Markert, J.W. Xiang, Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system, Mech. Syst. Sig. Process. 60–61 (2015) 243–251.
[26] H.H. Bafroui, A. Ohadi, Application of wavelet energy and Shannon entropy for feature extraction in gearbox fault detection under varying speed conditions, Neurocomputing 133 (2014) 437–445.
[27] W.D. Mark, H. Lee, R. Patrick, A simple frequency-domain algorithm for early detection of damaged gear teeth, Mech. Syst. Sig. Process. 24 (8) (2010) 2807–2823.
[28] S.H. Gao, Y.T. Zhang, K. Jia, Single Sample Face Recognition via Learning Deep Supervised Autoencoders, IEEE Trans. Inf. Forensics Secur. 10 (10) (2015) 2108–2118.
[29] W.J. Sun, S.Y. Shao, R. Zhao, A sparse auto-encoder-based deep neural network approach for induction motor faults classification, Measurement 89 (2016) 171–178.