RBFNN Fault Diagnosis Method of Rolling Bearing Based on Improved Ensemble Empirical Mode Decomposition and Singular Value Decomposition

Cheng Zhong, Yu Liu *, Jie-Sheng Wang, Zhong-Feng Li

Abstract—**The ensemble empirical mode decomposition (EEMD) can not completely eliminate the mode mixing phenomenon due to the noise of the working environment and weak early fault signals of rolling bearings. When the upper and lower envelope fitting is carried out by selecting the extreme points, EEMD can not completely eliminate the mode mixing. There are some extreme points that affect the fitting performance and the mean curve. The IMF components decomposed by EEMD has the same high dimension as the original fault signal, which leads to the result accuracy of the classifier. Therefore, a rolling bearing radial basis neural network (RBFNN) fault diagnosis method based on improved EEMD-SVD was proposed. When using the EEMD algorithm, the extreme points affecting the fitting envelope are removed, and then the fault signals are decomposed to obtain several sets of intrinsic mode function (IMF). Then these IMF components as well as the largest energy ratio of several IMF components are constructed into a new matrix, and then singular value decomposition (SVD) was adopted to realize the matrix SVD decomposition so as to obtain a set of low dimensional singular values. This set of singular values is then used to replace the original RBFNN fault signal input to perform bearing fault diagnosis. Finally, simulation experiments based on bearing failure data from Case Western Reserve University verify the proposed method.**

Index Terms—**Rolling bearing; Fault diagnosis; Ensemble empirical mode decomposition; Singular value decomposition; Intrinsic mode function; Radial basis function neural network**

I. Introduction

As an important part of all kinds of modern machinery and equipment, the' operation safety and operation condition of rolling bearings are related to workers' safety and enterprise's economic benefits. Statistics show that 30%-40% of the fault sources of equipment involving rotating machinery come from rolling bearings, Therefore, it is very necessary to monitor the running process of rolling bearings [1]. The monitor system is mainly to obtain signal features by analyzing the original fault signals, which can be roughly divided into four main processes: Collect status information, analyze information, extract information characteristics, classification [2]. Generally speaking, when the rolling bearing vibrates, the signal generated by its vibration has a lot of information, such as normal vibration information, fault location and degree information, so how to analyze and process vibration signals to extract effective fault information has always been the focus and hot issue of research [3]. Traditional signal analysis methods are linear and stationary signals have certain effect. But in practical engineering applications, due to equipment malfunction, load, speed, change and the impact of noise, Because the vibration signals obtained in the general industrial environment are not stationary or linear signals [4-5], the old signal analysis methods have certain limitations. In comparison, The time-frequency domain analysis method has better performance.

Huang proposed a signal analysis method called Empirical Mode Decomposition, which decomposes the signal into the sum of IMF components according to the characteristics of the signal, so as to effectively extract the characteristics of the signal. The IMF's intentions vary. Due to the advantages of this method, EMD has been widely used in machine diagnosis [6]. However, EMD algorithm has many disadvantages, such as stop condition, end effect and modal mixing. In order to reduce the influence of patterns on EMD signal analysis, some researchers proposed an ensemble Empirical Mode Decomposition (EEMD) method. The main feature of this method is to add white noise to suppress mode stratification [7]. Currently, in the study of vibration failure signal analysis based on EMD and EEMD, some researchers have proposed an improved HHT and sensitive FMI method to effectively extract the early friction failure characteristics of large units [8]. Lu et al. proposed to use EEMD and energy density in signal analysis and applied them to defect diagnosis of inner ring of rotating mechanical bearings, and achieved good results [9]. In Ref. [10], IMF components are used to obtain envelope signals through Hilbert transform. After calculating envelope spectrum, high-frequency amplitude modulation signals generated by locally damaged rolling bearings are extracted, which achieved better diagnosis effect than traditional envelope demodulation

Manuscript received October 16, 2021; revised January 9, 2022. This work was supported by the Basic Scientific Research Project of Institution of Higher Learning of Liaoning Province (Grant No. LJKZ0293), and the Project by Liaoning Provincial Natural Science Foundation of China (Grant No. 2019-ZD-0031).

Cheng Zhong is a postgraduate student in the School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, 114051, P. R. China (e-mail: 1728318375@qq.com).

Yu Liu is a lecturer of School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, 114051, P. R. China (Corresponding author, phone: 86-0412-2538246; fax: 86-0412-2538244; lnasacl@126.com).

Jie-Sheng Wang is a professor of School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, 114051, P. R. China (e-mail: wang_jiesheng@126.com).

Zhong-Feng Li is a senior engineer of Yingkou Institute of Technology, Yingkou, 115000, P. R. China (e-mail: afeng0601@163.com).

methods. Zhao et al proposed the application of the EMD based correlation dimension diagnosis method and AR model based on IMF components [11]. Ref. [12] studied different pattern recognition methods of bearing fault identification based on EMD marginal spectrum. An algorithm for automatic selection of IMF components is proposed in literature [13], and its advantages are verified by the detection results of bearing vibration signals.

EEMD algorithm is relatively optimal for signal decomposition. When it is used for fault signal decomposition, it can effectively decompose components with a single frequency and effectively solve the mode mixing phenomenon of signal components. However, a large amount of noise will be generated in the working environment of rolling bearings [14], and the early fault signals are relatively weak. Under the huge noise intervention coverage, the EEMD algorithm still cannot completely eliminate the phenomenon of mode aliasing. In addition, each IMF component has the same high dimension as the original fault signal. If the obtained IMF component is directly input into the RBFNN for training without processing, it will inevitably increase the difficulty of feature extraction and ultimately affect the diagnosis performance. Aiming at the existence of above two problems, when using EEMD algorithm, the part of envelope extreme points affecting the fitting out the envelope are deleted and the IMF components to select the largest energy ratio of several IMF component are used to construct another matrix. Then singular value decomposition (SVD) was adopted to realize the matrix SVD decomposition in order to get a set of low dimensional singular value. The obtained singular values are used to replace the original fault signals and be fed into RBFNN for training. Simulation results demonstrate the feasibility of the above methods.

II. Analysis on Rolling Bearing Fault

Rolling bearings exist in many large mechanical equipment. Due to the influence of various external working environment factors and its own working intensity, rolling bearings have a relatively high probability of failure [15]. At the same time, the running state of rolling bearings is of critical significance on the whole mechanical equipment, Therefore, it is very important to analyze the causes of the problems of bearings and study how to monitor and diagnose the location and degree of fault damage [16].

A. Bearing Internal Structure and Fault Types

Bearing is mainly composed of inner ring, outer ring, bearing body and cage. Its internal structure is shown in Fig. 1. In particular, the relative failure rate of the first three parts is the highest, the outer ring is over-matched with the bearing seat or mechanical shell hole to play a supporting role. The bearing body, also called a ball, roller or needle roller. Due to the fastening of the outer ring, the motion of the rolling body is carried out by rolling friction rather than sliding friction, which greatly reduces the loss of the rolling body. The shape, size and number of rolling bodies directly affect the composite capacity and performance of bearings. In addition to the orderly arrangement of the rolling body, the cage can also guide the rotation of the rolling body and improve the internal lubrication performance of the bearing. In practical engineering applications, rolling bearings also have different structures due to different working conditions, objectives and requirements for bearing structure, load capacity, service performance and other aspects [17]. The internal parameters of rolling bearings are listed in Table 1.

B. Fault Characteristic Frequency of Rolling Bearings

When a rolling bearing fails, the contact between the fault part and other parts of the bearing is periodic. The impact frequency between such contacts is called the fault characteristic frequency, and the characteristic frequency is different when the fault occurs at different parts. The characteristic frequencies of different parts can be calculated according to the relevant parameters of the internal structure of the bearing as shown in Fig. 1 and Table 1. The empirical formula of failure frequency of each part can be obtained.

The rotation frequency of the cage f_{bc} :

$$f_{bc}=\frac{1}{2}\left[\left(1-\frac{d}{D}\cos\alpha\right)f_{ir}+\left(1+\frac{d}{D}\cos\alpha\right)f_{or}\right] \tag{1}$$

The outer ring fault characteristic frequency f_o :

$$f_o=\frac{N}{2}\left|f_{or}-f_{ir}\right|\left(1-\frac{d}{D}\cos\alpha\right)=\frac{N}{2}f_r\left(1-\frac{d}{D}\cos\alpha\right) \tag{2}$$

The inner ring fault characteristic frequency f_i :

$$f_i=\frac{N}{2}\left|f_{or}-f_{ir}\right|\left(1+\frac{d}{D}\cos\alpha\right)=\frac{N}{2}f_r\left(1+\frac{d}{D}\cos\alpha\right) \tag{3}$$

The rolling body fault characteristic frequency f_b :

$$f_b=\frac{1}{2}\frac{D}{d}\left|f_{or}-f_{ir}\right|\left[1-\left(\frac{d}{D}\cos\alpha\right)^2\right]=\frac{1}{2}\frac{D}{d}f_r\left[1-\left(\frac{d}{D}\cos\alpha\right)^2\right] \tag{4}$$

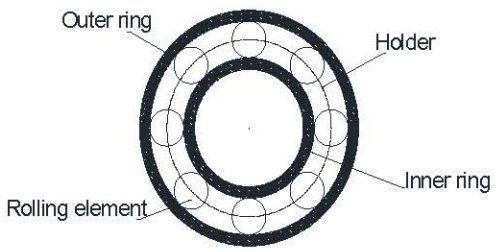


Fig. 1 Schematic diagram of inner structure of rolling bearing.

TABLE 1. INTERNAL PARAMETERS OF ROLLING BEARINGS

Name of parameter	Comment declaration
Outer orbit radius (r_2)	Mean diameter of outer orbit.
Pitch diameter (D)	Diameter of the circle where the center of the rolling body is located.
Inner orbit radius (r_1)	Mean diameter of inner orbit.
Roller diameter (d)	Average diameter of rolling body.
Contact angle (α)	The Angle between the force direction of the rolling body and the vertical line of the inner and outer raceway.

C. Reasons of Bearing Faults

(1) Raceway surface metal peeling. When the raceway surface of rolling bearing is subjected to repeated impact force, it is easy to produce constantly changing contact stress. When the stress reaches the limit, the internal structure surface is prone to fatigue spalling. Too large load of the rolling bearing, the not correct installation, the shaft bending and other reasons will also cause the raceway spalling.

(2) Plastic deformation. When the rolling bearing works under a very strong static load or impact load, the local stress on its surface will easily exceed the yield strength limit of the material, eventually leading to plastic deformation.

(3) The cage is damaged. Improper assembly or use will lead to cage deformation, resulting in increased friction between cage and ball, and even some rolling can not roll.

(4) The creep. Due to insufficient interference in the inner fit of the bearing, insufficient fastening of the sleeve, abnormal temperature rise, excessive load of the main engine and other factors, the inside or outside diameter surface skid, which results in mirror or discoloration.

(5) Agglutination. Gluing occurs on the contact surface of the ball and raceway or between the ball, that is to say that the metal spalling on one surface adheres to another surface.

(6) Rust corrosion. Partial or total surface rust, rolling body change line rust mainly because of poor storage environment, improper packaging, insufficient anti-rust agent, water acid solvent intrusion, or directly with the hand bearing, etc.

(7) Bearing burn. Due to insufficient lubrication or the use of inferior lubricating oil, as well as too compact when assembling bearings and other reasons, it is easy to cause the surface burn of each part of the bearing and the appearance of backfire color.

III. EEMD Signal Processing Analysis

Hilbert-Huang transformation is the general term for EMD and Hilbert time spectrum. The main feature of this transform is to adaptively transform the signals into several eigenmode functions by using the original characteristics of the signals. In order to make up for the old method of frequency domain analysis, the problem of forming a fixed basis function by recombining signals with multiple basis functions and selecting the best basis is solved. Therefore, if the signal is not linear or stationary, HHT has obvious advantages compared with other methods. Based on its unique advantages of adaptive decomposition, such as seismic, marine, remote sensing image processing, rotating machinery fault diagnosis and nonlinear system research. However, this method also has some problems such as break-point extension, determination of decomposition criteria, and inherent limitations of Hilbert demodulation.

A. Instantaneous Frequency

The concept of frequency plays an important role in signal analysis. A single component signal source has only one frequency value at any time, and this frequency is called instantaneous frequency. If it is a multi-component signal, then there are different instantaneous frequency values at different times. At present, instantaneous frequency is defined as:

If there is a signal $x(t)=a(t)cos\Phi(t)$, its instantaneous frequency can be defined as $f(t)=(d(\arg z_{(t)})/dt)/2\pi$, where $z_{(t)}$ is the analytic signal associated with $x_{(t)}$. For a given function $x(t)$, its Hilbert transform $y(t)$ can be calculated by:

$$y_{(t)}=\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x_{(\theta)}}{t-\theta}d_{(\theta)} \tag{5}$$

Construct an analytic function $z(t)$:

$$z_{(t)}=x_{(t)}+jy_{(t)}=a_{(t)}e^{j\varphi(t)} \tag{6}$$

where:

$$a_{(t)}=\sqrt{x_{(t)}^{\ 2}+y_{(t)}^{\ 2}},\varphi(t)=\arctan(\frac{y_{(t)}}{x_{(t)}}) \tag{7}$$

where, $a(t)$ is the instantaneous amplitude function to reflect the change of signal energy with time; $\varphi(t)$ is a function of the instantaneous phase to reflect the change of signal frequency with time.

Theoretically, the imaginary part of the analytic function $z(t)$ can be defined in different ways, and the Hilbert transform is the convolution of signals $x(t)$ and $1/t$ so as to effectively highlight the local properties of the signal $x(t)$. The instantaneous frequency $\omega(t)=d\varphi(t)/d(t)$ is obtained by taking the derivative of the instantaneous phase function.

B. EMD Decomposition

Since the general non-stationary signal does not satisfy the basic conditions of the inherent modal function, some scholars make two assumptions: all complex signals can be composed of several IMF components that do not affect each other. Each IMF component does not need to limit its linear characteristics. The difference between the number of extreme points and zeros of the curve corresponding to the eigenvector function is equal to or less than 1, and the mean values of the maximum and minimum extreme points corresponding to any point in the curve are 0. The EMD decomposition process is described as follows:

Step 1: Find all the local maximum and minimum points of the signal $x(t)$.

Step 2: Cubic spline interpolation method was used to perform cubic spline interpolation fitting for all maximum and minimum points respectively to obtain the upper and lower envelops. Then the mean curve m1 of the two envelops was calculated and $h_1=x(t)-m_1$ was calculated.

Step 3: h_1 is the first IMF component obtained by signal decomposition. If not, it should replace the above $x_{(t)}$ and continue the above two operations k times until the first IMF, h_{1k} component is found. At this time, the first inherent mode function is $c_{1(t)}=h_{1k}$. Generally, the termination criteria of EMD algorithm is defined as:

$$SD=\frac{\sum_{t=0}^{T}\left|h_{1,k-1}(t)-h_{1,k}(t)\right|^2}{\sum_{t=0}^{T}\left|h_{1,k-1}(t)\right|^2} \tag{8}$$

where, this value is between $(0.2, 0.3)$.

Step 4: Separate $c_1(t)$ from signal $x(t)$ to get:

$$r_1(t) = x(t) - c_1(t) \tag{9}$$

Take $r_1(t)$ as the $x(t)$ and repeat above three steps for n times to get the next IMF component $c_2(t)$ until the n-th IMF component $c_n(t)$, then there will be:

$$\begin{cases} r_2(t) = r_1(t) - c_2(t) \\ \cdots\cdots \\ r_n(t) = r_{n-1}(t) - c_n(t) \end{cases} \tag{10}$$

Step 5: After $r_n(t)$ is transformed into a monotone function, the cycle ends, and the remaining $r(t)$ is the average level of the $x_{(t)}$, which is called the residual component. All IMF components and residual components are superimposed to form the original signal $x(t)$.

$$x(t) = \sum^{n} c_i(t) + r_n(t) \tag{11}$$

EMD decomposition theory can be adaptive to the signal into a single frequency component of the IMF, which solves the problem of needing to select basis function in advance to a certain extent of conventional time-frequency analysis methods, such as wavelet transform and Fourier transform because the basic functions are generated according to the characteristics of the signal itself from adaptation. However, in practical engineering application, due to strong noise interference and weak early fault signals, the IMF components obtained by simply using EMD algorithm to decompose the fault signals often have the phenomenon of mode mixing. On the other hand, in the upper and lower envelope, due to the uncertainty of the extreme at the end point, it is easy to cause the end point effect. Flowchart of EMD algorithm is shown in Fig. 2.

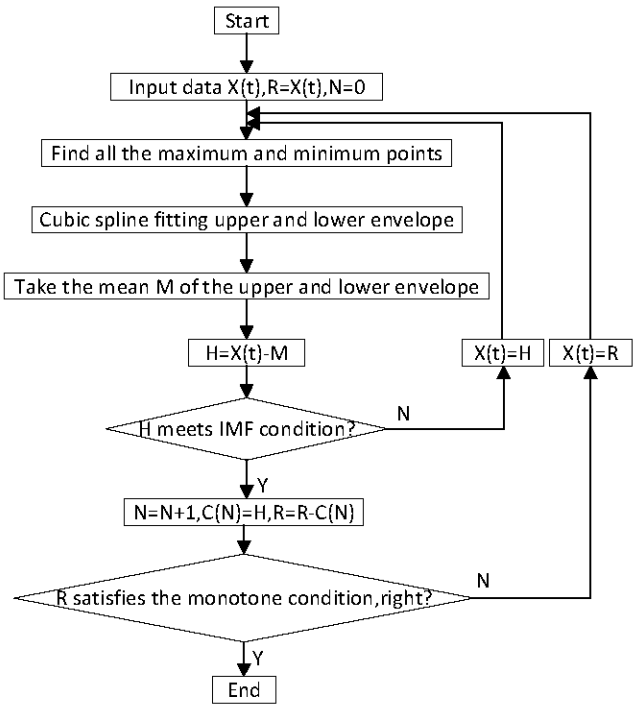


Fig. 2 Flowchart of EMD algorithm.

C. EEMD Decomposition

EEMD decomposition method is mainly to add white noise into the signals according to the characteristic that the average value of white noise is zero. EEMD is essentially an improvement of EMD algorithm, and its core is still to use EMD for decomposition. The result of decomposition is processed on average, and the more times of average processing, the less influence the noise brings to the result of decomposition. Set the signal as $x(t)$, and the specific decomposition steps are described as follows.

Step 1: Set the number of equalization for $x(t)$ M and the initial value $i = 1, 2, ..., M$.

Step 2: Add a certain copy of random white noise $n_i(t)$ to signal $x(t)$ to form a new set of signal $x_i(t)$.

$$x_i(t) = x(t) + n_i(t) \qquad i = 1, 2, ..., M \tag{12}$$

Step 3: Perform EMD decomposition of this new set of signals $x_i(t)$ to obtain:

$$x_i(t) = \sum_{n=1}^{n} c_{in}(t) + r_{in}(t) \tag{13}$$

where, n is the number of IMF, $c_{in}(t)$ is IMFs, and $r_{in}(t)$ is the participating component.

Step 4: Repeat Step 2-3 M times, each time adding a certain amplitude of white noise, decompose a series of IMFs. The IMF component $c_n(t)$ of EEMD decomposition was obtained by means processing the set of IMFs. IMFs:

$$\left[\{c_{1,n}(t)\} \quad \{c_{2,n}(t)\} \quad \cdots \quad \{c_{M,n}(t)\}\right] \qquad n = 1, 2, ..., N$$

$$c_n(t) = \frac{1}{M}\sum_{i=1}^{M} c_{i,n}(t) \qquad i = 1, 2, ..., M; n = 1, 2, ..., N \tag{14}$$

The flowchart of EEMD algorithm is shown in Fig. 3. For EEMD, the introduction of white noise makes its performance very good, but this does not mean that the introduction of white noise can be arbitrary. When the amplitude of the introduced white noise is too large, false components will be generated. If the amplitude is too small, the local extreme value of the signal will change frequently, resulting in the phenomenon of mode aliasing. Although there is no reliable theory to prove it at present, the standard deviation criterion of white noise energy can be summarized as:

$$\alpha = \sqrt{\frac{e_n}{e_o}} \qquad \beta = \sqrt{\frac{e_h}{e_o}} \tag{15}$$

where, e_h represents the standard deviation of the high-frequency component, e_o represents the standard deviation of the energy of the original signal, and e_n represents the standard deviation of the energy of the white noise. Generally, the first component is taken as the effective high-frequency component. α and β satisfy $0 < \alpha < \beta / 2$. In addition, the average processing times M is also a very critical parameter, which satisfies the following relationship with the relative error e of IMF components.

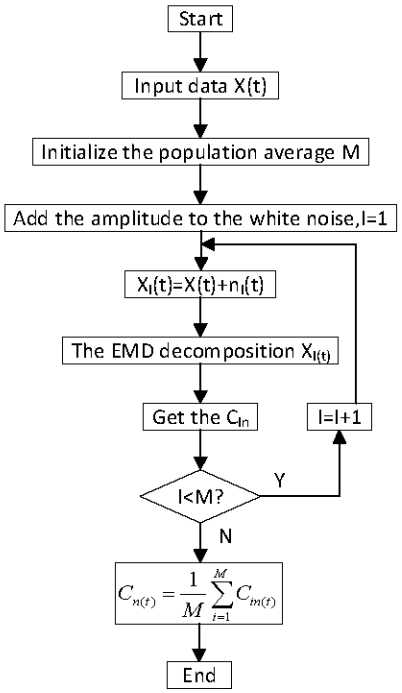


Fig. 3 Flowchart of EEMD algorithm.

$$e=\frac{\alpha}{\sqrt{M}} \tag{16}$$

where, α is the ratio of white noise to the standard deviation of the energy of the original signal, and M is the average number.

In order to verify the difference between EMD, EEMD and wavelet decomposition (WD) algorithms in the performance of fault signal decomposition, the simulation experiments were designed to verify that EEMD algorithm has better anti-mode aliasing ability. Construct the simulation signals:

$$\begin{aligned} x(t)&=sin(2\pi\cdot 350t)+sin(2\pi\cdot 150t)\\ &+sin(2\pi\cdot 50t)+0.1\cdot randn(1,length(t)) \end{aligned} \tag{17}$$

The signal consists of three sinusoidal signals with different frequencies and a random noise signal. Experimental sampling frequency fs =1024, sampling point N =1024. EMD, WD and EEMD were used to decompose the signals, and the time domain and frequency domain waveform of each component were plotted.

Fig. 4 shows the result of EMD decomposition. Fig. 4(a) shows the time domain waveform of the fault signal and the first four IMF components, and Fig. 4(b) shows the frequency domain waveform of the first four IMF components. It can be seen from the frequency domain waveform that IMF1, IMF2 and IMF3 correspond exactly to the frequency values 350,150 and 5 of the original signal expression. Their physical meanings are very clear, fully illustrating the effectiveness of EMD algorithm for signal decomposition. However, a careful observation of the frequency domain waveform shows that IMF1 and IMF3 components also have a peak value at the frequency of about 25. In other words, the same component has two different frequency values, which is the phenomenon of mode aliasing. The so-called mode mixing refers to the failure to effectively separate different mode components according to the time scale, resulting in the original different modes appearing in one mode, and the mode mixing phenomenon of one component will also affect the following components, which results in the loss of the actual physical significance of the components. Fig. 5 shows the distribution of three-layer wavelet packets of analog signals. Fig. 5(a) shows the waveform reconstructed by three-layer wave packet decomposition of analog signals, and Fig. 5(b) is the spectrum of it. The time domain diagram shows that the reconstructed signal deviates greatly from the sinusoidal component of the original signal and loses its physical significance. It is found that there are many peaks in the frequency of clutter in the spectrum diagram. The main reason is that the wavelet decomposition is represented by the fixed wavelet basis function. When the original signal differs greatly from the basis function, the reconstructed signal may lose the physical meaning of the original signal. The EEMD decomposition results in Fig. 6 show that it not only has the advantages of EMD decomposition, original signals of each component according to the high frequency to low frequency decomposition in turn out, and first three IMF components are almost non-existent modal aliasing phenomenon. The same modal component appears only a peak frequency, therefore the EEMD modal aliasing resistance is stronger. The performance is better than EMD, wavelet decomposition and other algorithms.

IV. Influence Analysis of Extreme Points on Upper and Lower Envelope

The above simulation signals are relatively ideal. But in the practical engineering applications, because the discrete degree of different data acquisition or affected by the noise is too big, some false extreme value point and redundancy extreme value point will be produced. This kind of extreme value points will cause great influence on cubic fitting to form envelope, which easily leads to a phenomenon of envelope. An actual simulation signal is shown as follows. The sampling frequency f_s in this experiment is 800Hz, and the sampling point N is 100. The simulation signal $x(t)$ is composed of two amplitude modulation signals.

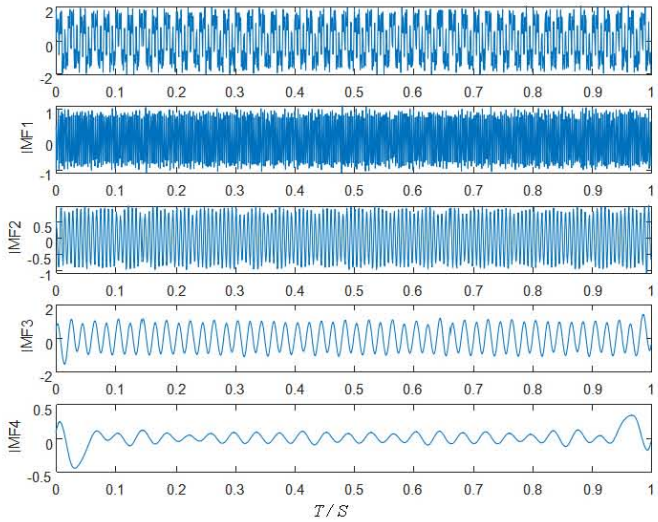
$$x_{1(t)}=\cos(2\pi\cdot 500\cdot t+0.8\cdot\sin(2\pi\cdot 90\cdot t)) \tag{18}$$

$$x_{2(t)}=\cos(2\pi\cdot 1200\cdot t+0.6\cdot\sin(2\pi\cdot 300\cdot t)) \tag{19}$$

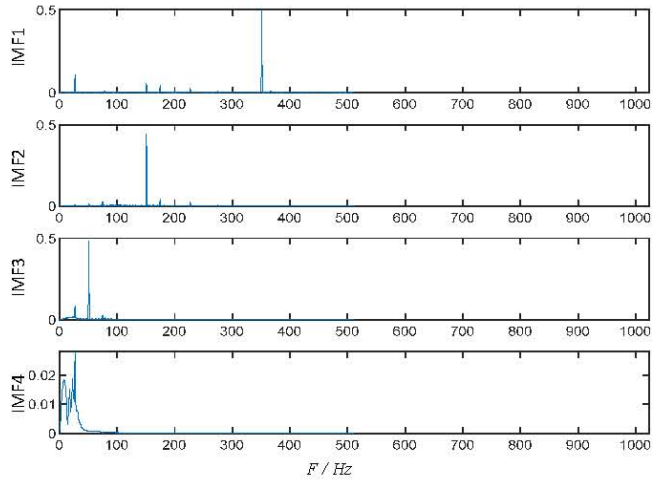
$$x_t=x_{1(t)}+x_{2(t)} \tag{20}$$

Fig. 7 is the time domain waveform of the signal. It can be seen from Fig. 7 that there are some more obvious redundant extreme points, such as three points with red star label. If this kind of extreme value points are used as normal extreme value points for cubic spline interpolation, which makes it easy to cause over envelope phenomenon when fitting the upper and lower envelope. Fig. 8 shows the simulation of fitting the upper and lower envelope by cubic spline interpolation without removing the redundant extreme points. As can be seen from Fig. 8, due to the influence of the

redundancy extreme as shown in Fig. 7, the upper and lower envelop lines cross at multiple points, that is to say that the over-envelope phenomenon occurs. For example, between the interval [0,10] and [70,100], the upper and lower envelops cross heavily. Therefore, as shown in Fig. 9, the mean curve will fluctuate greatly in the over-envelope region and will no longer be stable.

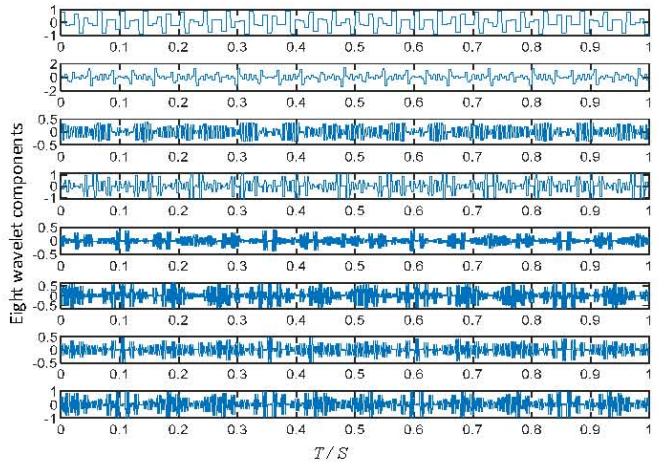


(a) Time domain waveform of original signal and first four IMF components

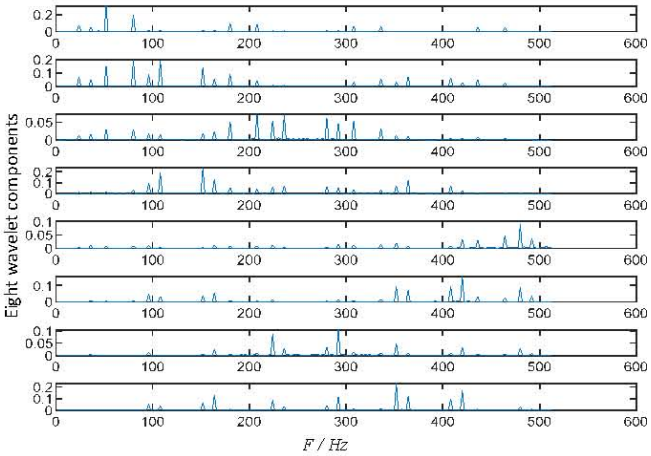


(b) Frequency domain waveform of four IMF components

Fig. 4 Decomposition results of EMD.

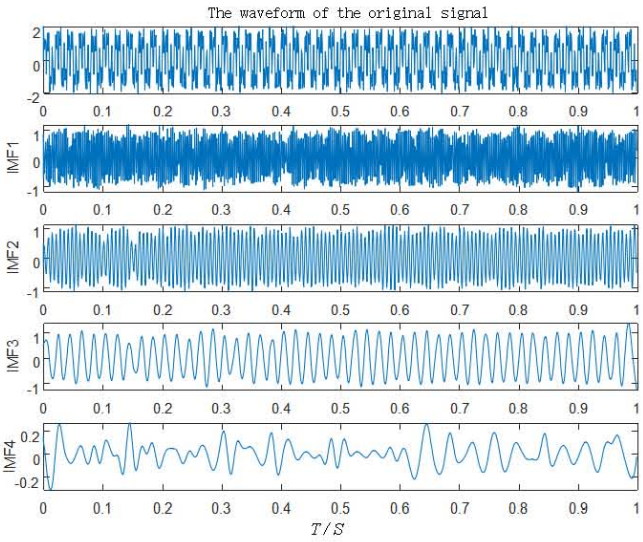


(a) Reconstruction waveform of signal three-layer wavelet packet decomposition

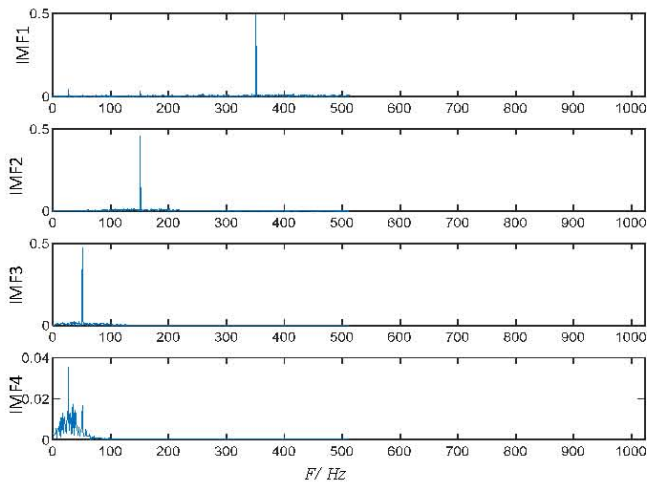


(b) Corresponding spectrum diagram

Fig. 5 Decomposition results of wavelet transform.



(a) Time domain waveform of original signal and first four IMF components



(b) Frequency domain waveform of four IMF components

Fig. 6 EEMD decomposition results.

Fig. 10 and Fig. 11 show the envelope line and mean curve after removing the pseudo extreme points and redundant extreme points. It can be seen that the over-envelope phenomenon has been effectively solved after removing these extreme points, and the mean curve is relatively more stable, and its variation trend is basically consistent with the

original signal. Therefore, the performance of EEMD algorithm can be effectively improved if the above extreme points affecting the mean curve and the envelope are removed when EEMD algorithm searches for extreme points to fit the upper and lower envelope.

V. Rolling Bearing Fault Diagnosis Based on RBFNN

A. Singular Value Decomposition (SVD)

EEMD method can suppress the mode aliasing phenomenon in EMD algorithm. It can also solve the problem of selecting the basis function in the wavelet decomposition and the problem of constant multi-resolution in the traditional time-frequency analysis method. However, the decomposed intrinsic mode components are characterized by high dimension and excessively long data. Direct input to RBFNN for training will affect its extraction of fault features and the learning rate of neural network. Therefore, the singular value decomposition (SVD) method is adopted to process the intrinsic mode components of signals. Finally, RBFNN is used for fault diagnosis.

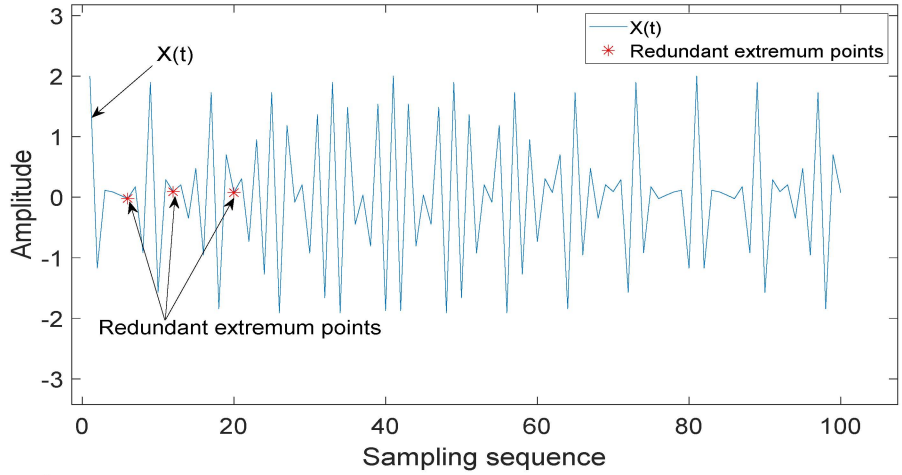


Fig. 7 Simulation signal waveform.

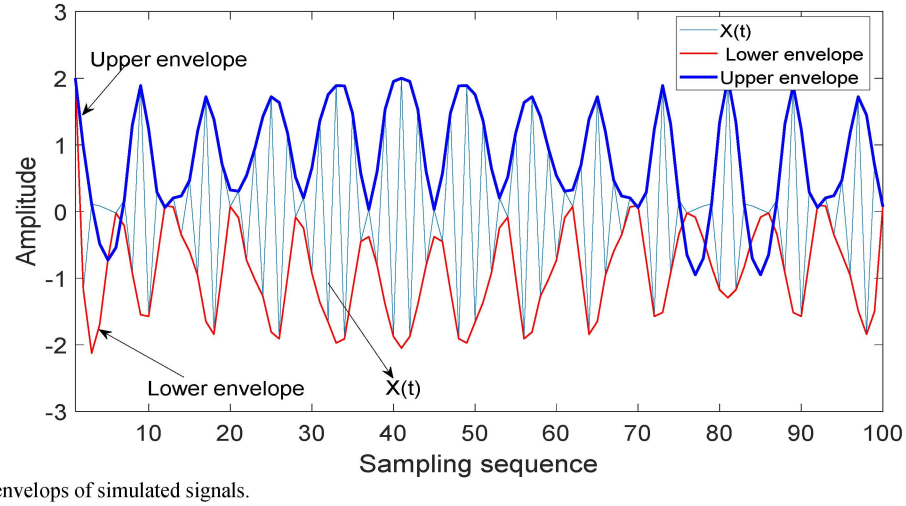


Fig. 8 Upper and lower envelops of simulated signals.

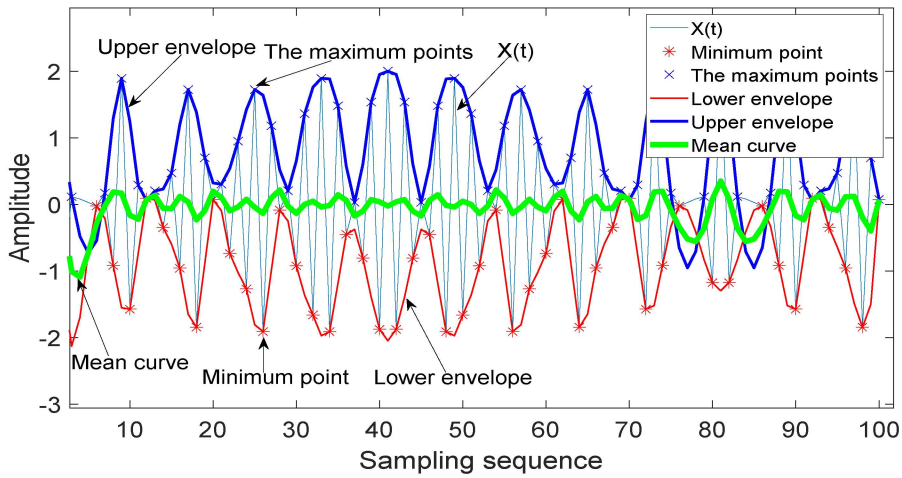


Fig. 9 Mean curve and envelope.

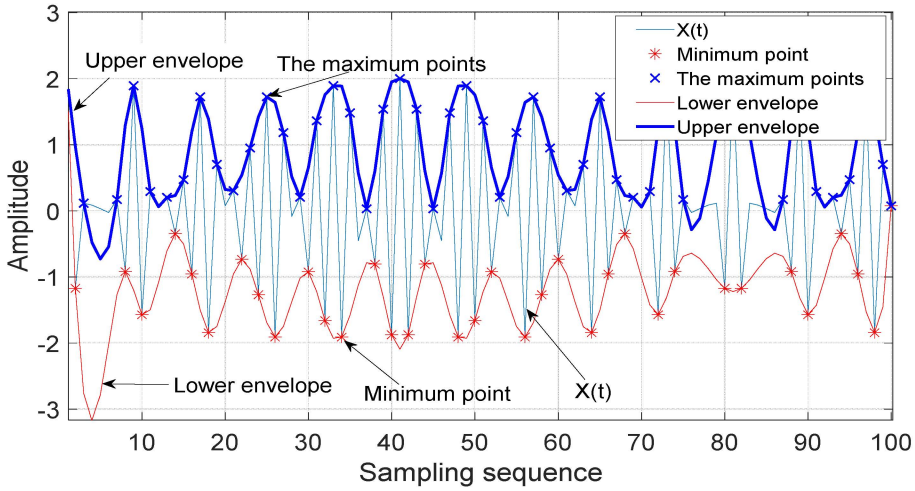


Fig. 10 Upper and lower envelope lines of simulation signals with removed pseudo extreme points.

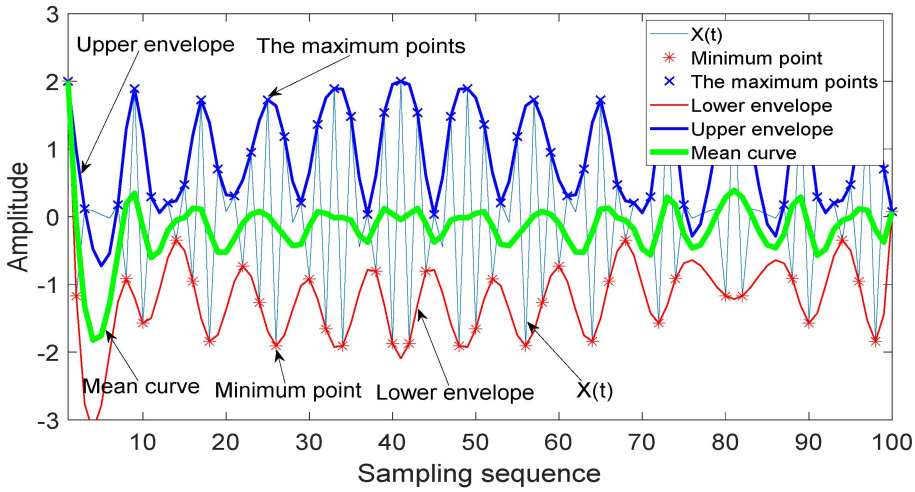


Fig. 11 Mean value curve and envelope line after removing pseudo extreme points.

Because the singularity value of any matrix is inherent to the matrix and has good stability, It also has good application in image processing and signal analysis. SVD is a method of orthogonalization of matrices. For any matrix, whether related to rows or columns, its left and right sides can be multiplied by an orthogonal matrix for transformation, so that the linearly dependent rows or columns in the original matrix can be transformed into linearly independent ones. For any real matrix $A_{m\times n}$, assuming its rank is r, singular value decomposition is performed on $A_{m\times n}$, that is, there are two orthonormal matrices U and V and diagonal matrix D to satisfy:

$$A = UDV^T \tag{21}$$

where , $U_{m\times m} = [u_1, u_2, \ldots\ldots, u_m]$, $D_{m\times n} = \begin{bmatrix} \Delta_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}$, $V_{n\times n} = [v_1, v_2, \ldots\ldots, v_n]$, $r = \min(m, n)$, $\Delta_{r\times r} = diag(\sigma_1, \sigma_2, \ldots\ldots \sigma_r)$, $\sigma_i (i = 1, 2, \cdots\cdots, r)$ are called singular values of matrix $A_{m\times n}$, and $\sigma_i = \sqrt{\lambda_i}, \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$ is an eigenvalue of matrix $A^T A$. Under the constraint $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$, the singular values of the matrix are unique.

The singular value of a matrix has the following two characteristics.

(1) The singular values of the matrix have good stability;

(2) Singular values of matrices are inherent features of matrices [18].

Based on the characteristics of the singular value of the matrix, several components of the signal can be decomposed, and the first five components with the largest energy proportion can be selected as the eigenvectors of the fault signal, which are superposed to form an eigenvector, and the singular value of the matrix can be obtained by singular value decomposition.

B. Radial Basis Function Neural Network (RBFNN)

Radial basis function neural network (RBFNN) is a forward network based on function approximation theory. Its structure is mainly composed of input layer, hidden layer and output layer. Several signal source nodes constitute the input layer and the structure of the hidden layer is mainly determined by the complexity of the problem described. The transformation function is RBF. The third layer is the output layer, which is mainly used to give results for input parameters [19-20]. The structure of RBFNN is shown in Fig. 12. The radial basis function usually uses the Gaussian function. RBF neural network must be provided before training, including an input vector, a corresponding output vector and an extended RBF constant. The training process can be divided into unsupervised learning (training weights

for determining entry and exit levels) and supervised learning (training weights for determining entry and exit levels).

C. Fault Diagnosis Simulation Experiments

The early fault signals of rolling bearings are relatively weak, and the fault features are difficult to extract. However, the diagnosis of the early fault signals can reduce the loss in time, so as to avoid the fault developing to a more serious point in the later stage. Therefore, in this experiment, bearing fault signal data of Case Western Reserve University were used to simulate early, middle and late faults by measuring different damage diameters. The damage diameters were 0.007mm, 0.014mm and 0.021mm, respectively. In order to verify the effectiveness of the algorithm, bearing early failure signal data, namely damage diameter of 0.007mm, were selected to verify the effect. The experimental data information is listed in Table 2. A total of 4 types of fault are normal data (represented by 1), rolling body fault (represented by 2), inner ring fault (represented by 3), and outer ring fault (represented by 4). A total of 160 groups of data with damage diameter of 0.007mm are selected, and 40 groups of data are selected for each category. Data sampling point is 10000, sampling frequency is 48KHz. The complete flow chart of signal processing is shown in Fig. 13. EEMD algorithm was used to decompose the signal into 14 IMF components and eliminate redundant poles and pseudo-poles. The ratio of energy of the first five IMF components reached 97% through calculation. Therefore, the first five IMF components were selected as the eigenvector matrix, and the five eigenvalues $\sigma_i (i = 1, 2, 3, 4, 5)$ of each fault signal were obtained according to the process shown in Fig. 13. 160 sets of data were processed according to the process shown in Fig. 13, and the data sets were randomly scrambled. Randomly selects 60 samples from the dataset to train the RBF neural network, and the remaining 100 samples are used to test the accuracy of the system.

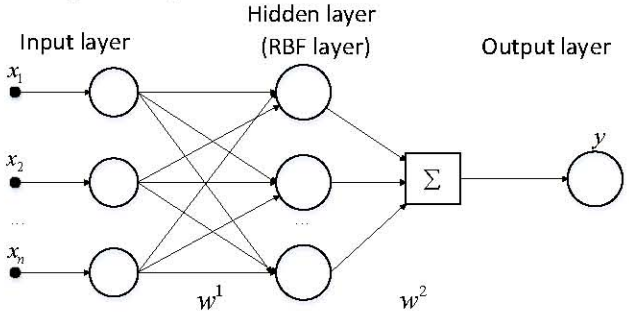


Fig. 12 Structure of radial basis function neural network.

TABLE 2. EXPERIMENTAL DATA INFORMATION TABLE

Bearing fault	Number of samples	Single sample data points	Damage to the diameter	Category
Normal	40	10000	0.000	1
Rolling body fault	40	10000	0.007	2
Inner ring fault	40	10000	0.007	3
Outer ring fault	40	10000	0.007	4

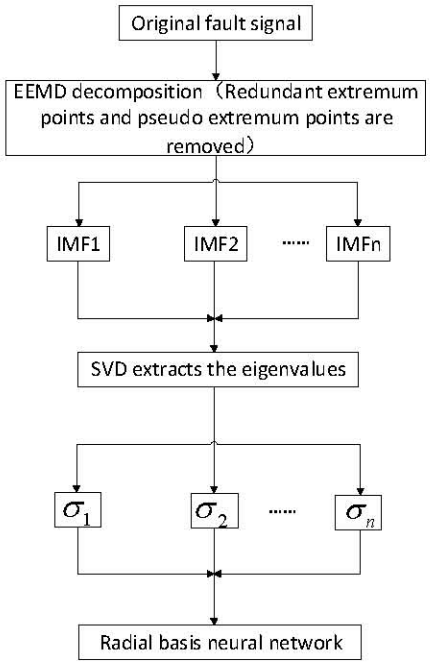


Fig. 13 Fault diagnosis flowchart of improved EEMD-SVD-RBFNN.

Since the output value predicted by the RBFNN for samples is a floating point number close to the sample label, So for the intuition of the result, the predicted value is specified as follows. Suppose the predicted value of the network for the sample type is x, the sample type is $c_j (j = 1, 2, 3, 4)$. If $c_j - 0.5 < x < c_j + 0.5$, the value of x is j. In other words, if the predicted value falls within the range of 0.5 in the left and right fields of a certain category, it is considered that the predicted value of the sample belongs to this category. Such treatment can improve the feasibility of the classification system.

First of all, above data sets are used to directly calculate the IMF components based on EEMD decomposition, wavelet decomposition and EMD decomposition respectively. Then they are fed into RBFNN for training without any processing. The test results are shown in Fig. 14-16. The accuracy was 70%, 48% and 45%, respectively. Fig. 17 shows the test results of 100 samples after the fault signal components obtained by improved EEMD decomposition algorithm based on SVD and RBFNN. The blue cross is the sample predicted value, and the red circle is the sample label. As can be seen from Fig. 17, there were 7 sets of data prediction errors, that is to say, the classification accuracy was 93%. Fig. 18 and Fig. 19 show the test results of fault signal components obtained from wavelet decomposition and EMD decomposition respectively based on SVD and RBFNN. The test accuracy is 88% and 82% respectively, which is lower than that in Fig. 17, which indicates that improved EEMD has better performance than wavelet decomposition and EMD algorithm. In addition, the fault diagnosis accuracy shown in Fig.17-19 is much higher than that of Fig. 14-16. Therefore, the proposed EEMD-SVD-RBFNN fault diagnosis model by eliminating pseudo extreme points can effectively improve the diagnosis accuracy.

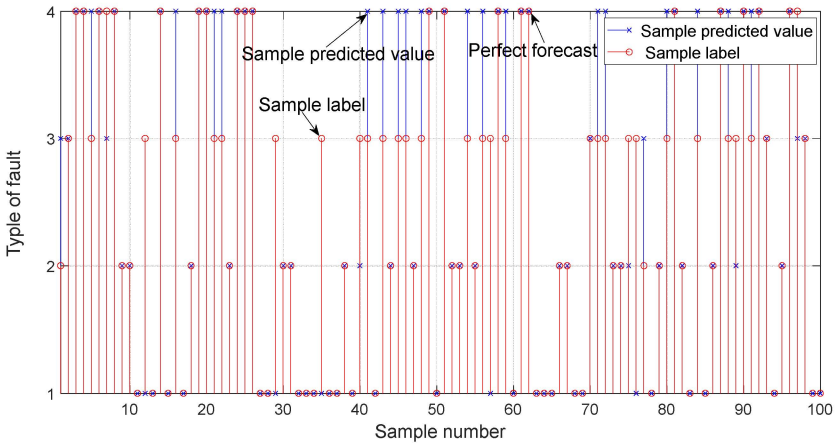


Fig. 14 Fault diagnosis results of EEMD-RBFNN.

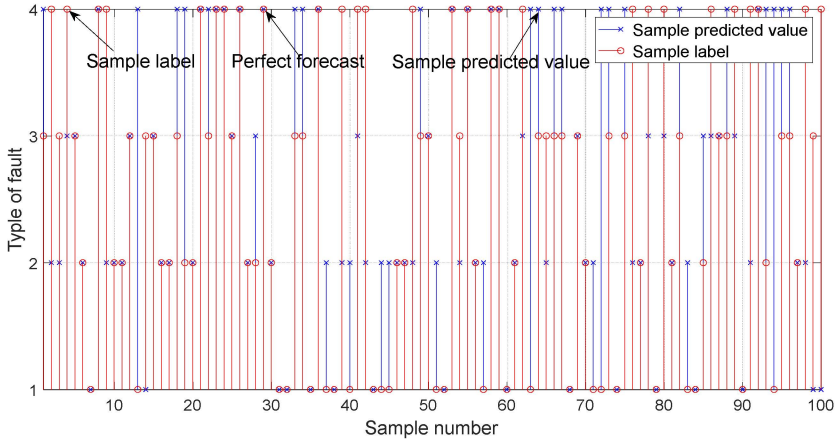


Fig. 15 Fault diagnosis results of WD-RBFNN.

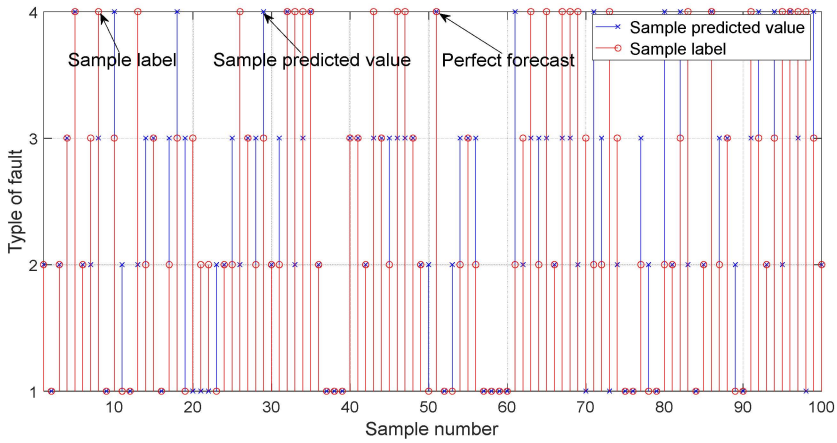


Fig. 16 Fault diagnosis results of EMD-RBFNN.

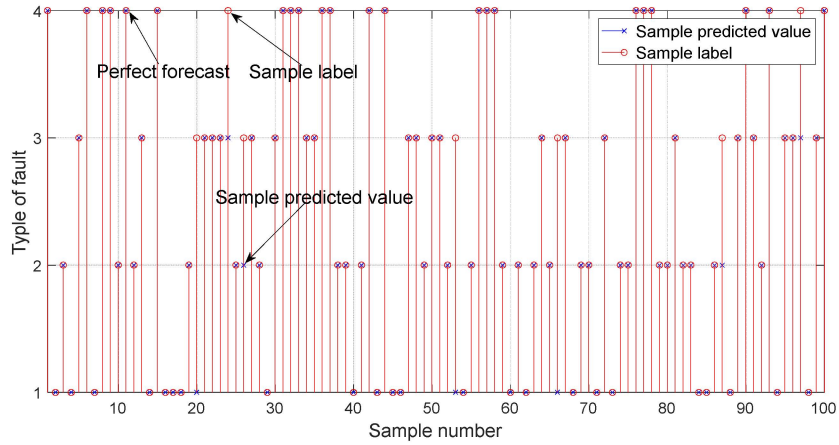


Fig. 17 Fault diagnosis results of improved EEMD-SVD-RBFNN.

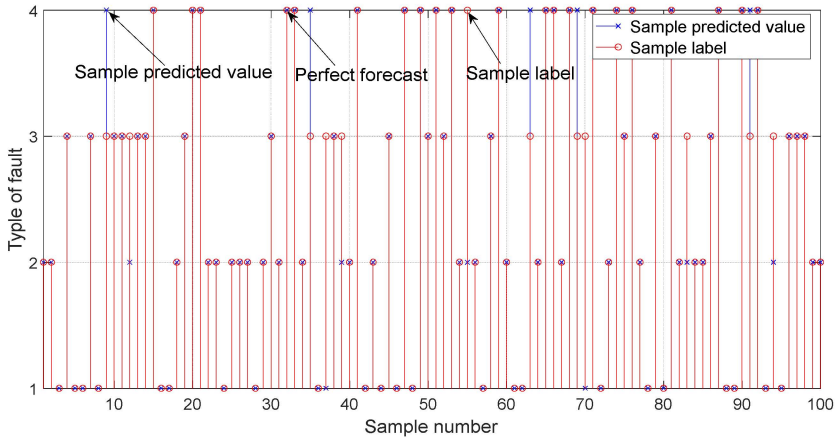


Fig. 18 Fault diagnosis results of WD-SVD-RBFNN.

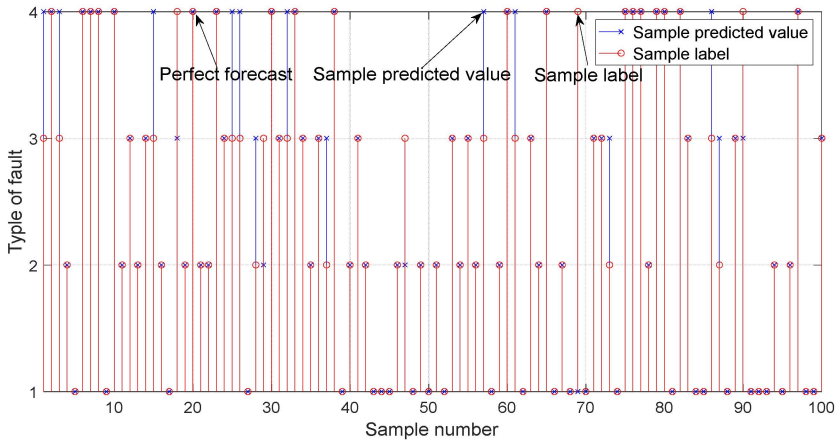


Fig. 19 Fault diagnosis results of EMD-SVD-RBFNN.

VI. Conclusion

In this paper, the advantages of EEMD algorithm are verified by comparing the effect of several time-frequency analysis algorithms on signal decomposition. However, the EEMD algorithm is still subject to the influence of uncontrollable factors such as noise and generates the phenomenon of mode mixing. By analyzing the influence of extreme points on the upper and lower envelope and the mean curve, it is found that some pseudo extreme points and redundant extreme points can cause the over-envelope phenomenon, thus affecting the mean curve. Therefore, the EEMD performance can be optimized by eliminating these extreme points. Based on the improved EEMD algorithm, the IMF component matrix was decomposed into the fault signals and the SVD was performed to obtain several eigenvalues representing the fault signals. Finally, the eigenvalues obtained from the decomposition of each signal are fed into the RBFNN for training. By comparing with the original processing methods, the simulation results show that the proposed rolling bearing fault diagnosis method is effective and feasible.

References

[1] Y. M. Yue, Y. L. Mao, and J. Q. Wang, "Research about Fault Intelligent Diagnosis of High Speed Rotating Machinery," *Journal of Shenyang Architectural & Civil Engineering Institute*, vol. 12, no. 6, pp. 770-773, 2005.

[2] C. Sheng, Z. Li, Q. Li, Z. Guo, and Y. Zhang, "Recent Progress on Mechanical Condition Monitoring and Fault Diagnosis," *Procedia Engineering*, vol. 15, no. 18, pp. 142-146, 2011.

[3] F. W. Guo, B. L. Yu, and G. L. Zhi, "Fault Classification of Rolling Bearing Based on Reconstructed Phase Space and Gaussian Mixture Model," *Journal of Sound & Vibration*, vol. 323, no. 3-5, pp. 1077-1089, 2009.

[4] Z. Gui, H. Zhang, H. Sun, and F. Zhang, "Research and Application of Early Warning Model of Vibration Deterioration for Hydroelectric-Generator Unit." *Journal of Hydraulic Engineering*, vol. 49, no. 2, pp. 216-222, 2018.

[5] A. Rai, and S. H. Upadhyay, "A Review on Signal Processing Techniques Utilized in the Fault Diagnosis of Rolling Element Bearings," *Tribology International*, vol. 96, pp. 289-306, 2016.

[6] C. Damerval, S. Meignen, and V. Perrier, "A Fast Algorithm for Bidimensional EMD," *IEEE Signal Processing Letters*, vol. 12, no. 10, pp. 701-704, 2005.

[7] M. Rezaee, and A. T. Osguei, "Improving Empirical Mode Decomposition for Vibration Signal Analysis," *ARCHIVE Proceedings of the Institution of Mechanical Engineers Part C Journal of Mechanical Engineering Science*, vol. 203-210, pp. 1989-1996, 2016.

[8] Y. G . Lei, "Mechanical Fault Diagnosis Based on Improved Hilbert-huang Transform," *Journal of Mechanical Engineering*, vol. 47, no. 5, pp. 71-77, 2011.

[9] J. M. Lu, F. L. Meng, H. Shen, L. B. Ding, and S. N. Bao, "Fault Diagnosis of Roiling Bearing Based on EEMD and Instantaneous Energy Density Spectrum," *Applied Mechanics and Materials*, vol. 97-98, pp. 80-89, 2011.

[10] D. Yu, J. Cheng, and Y. Yang, "Fault Diagnosis Approach for Roller Bearing Based on Empirical Mode Decomposition Method and Hilbert Transform," *Chinese Journal of Mechanical Engineering*, vol. 18, no. 2, pp. 267-270, 2005.

[11] Y. Zhao, J. Zhu, B. Zhang, and F. Hanal. "Rub-impact Fault Diagnosis of Rotating Machinery Based on Hilbert-huang Transform," *Journal of Computational and Theoretical Nanoscience*, vol. 13, no. 11, pp. 8123-8129, 2016.

[12] H. Li, Y. P. Zhang, and H. Q. Zheng, "Hilbert–Huang Transform and Marginal Spectrum for Detection and Diagnosis of Localized Defects in Roller Bearings," *Journal of Mechanical Science and Technology*, vol. 23, no. 2, pp. 291-301, 2009.

[13] J. Terrien, C. Marque, and B. Karlsson, "Automatic Detection of Mode Mixing in Empirical Mode Decomposition Using Non-stationarity Detection: Application to Selecting IMFs of Interest and De-noising," *EURASIPJ Journal on Advances in Signal Processing*, vol. 2011, no. 1, pp. 37, 2011.