

ORIGINAL ARTICLE

Open Access



Fault Diagnosis Based on BP Neural Network Optimized by Beetle Algorithm

Maohua Xiao^{1*} , Wei Zhang¹, Kai Wen¹, Yue Zhu¹ and Yilidaer Yiliyasi²

Abstract

In the process of Wavelet Analysis, only the low-frequency signals are re-decomposed, and the high-frequency signals are no longer decomposed, resulting in a decrease in frequency resolution with increasing frequency. Therefore, in this paper, firstly, Wavelet Packet Decomposition is used for feature extraction of vibration signals, which makes up for the shortcomings of Wavelet Analysis in extracting fault features of nonlinear vibration signals, and different energy values in different frequency bands are obtained by Wavelet Packet Decomposition. The features are visualized by the K-Means clustering method, and the results show that the extracted energy features can accurately distinguish the different states of the bearing. Then a fault diagnosis model based on BP Neural Network optimized by Beetle Algorithm is proposed to identify the bearing faults. Compared with the Particle Swarm Algorithm, Beetle Algorithm can quickly find the error extreme value, which greatly reduces the training time of the model. At last, two experiments are conducted, which show that the accuracy of the model can reach more than 95%, and the model has a certain anti-interference ability.

Keywords: Rolling bearing, BP neural network, Beetle algorithm, Wavelet packet transform

1 Introduction

According to statistics [1], 70% of rotating machinery failures are related to bearing failures. Once a bearing fails, it will cause a series of chain failures, which will seriously affect the operation safety of the entire equipment [2]. At present, the diagnosis of bearing faults is mostly based on vibration signals. Data-driven methods generally extract features of vibration signals of limited length, such as mean square error, kurtosis, and non-linear dynamic parameters for fault diagnosis [3]. In order to obtain richer features of fault diagnosis, generally, the signal samples are first decomposed, and then the corresponding parameters of each sub-signal are calculated as the fault characteristics. Fourier Analysis, Wavelet Transform, Empirical Mode Decomposition, Local Mean Decomposition [4, 5] are all very effective linear system analysis methods in the past. In recent years, many new

signal processing methods have been produced, such as the Wavelet Packet Analysis, which provides a broader development space for bearing fault diagnosis research [6, 7].

The classification of faults is essentially a process of identifying the size and type of fault. Scholars at home and abroad have proposed many supervised models for fault classification and recognition. Yan et al. [8] extracted signal time domain and frequency domain features and based on the particle swarm optimization support vector machine (PSO-SVM) classification model to realize the recognition of rolling bearing multi-fault states. Kankar et al. [9] used continuous wavelet transform to extract statistical features, and compared support vector machines with artificial neural networks and self-organizing graph models. The results showed that the features selected by Meyer wavelet can achieve higher fault classification efficiency through SVM classifier. Zhou et al. [10] extracted bearing fault features based on the NCA method, and merged multi-channel data through the CHMM method, which effectively removed redundant

*Correspondence: xiaomaohua@njau.edu.cn

¹ College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

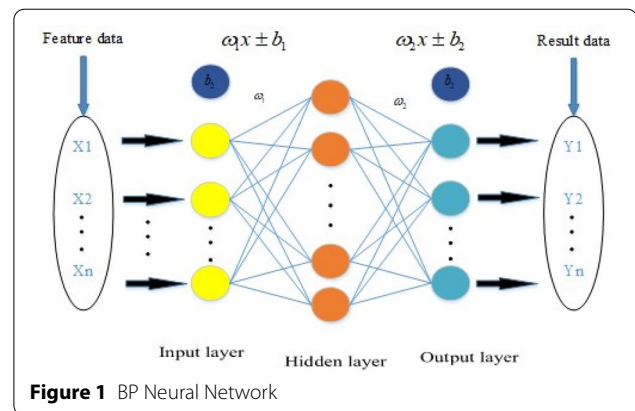
Full list of author information is available at the end of the article

information and improved the fault diagnosis effect. Zaragoza [11] and others combined hidden Markov model and deep perceptron to correctly model variable-length symbol sequences through non-segmented training. The results show that the results obtained by the proposed hybrid MLP-HMM method are to a large extent. It is better than previous work. Chen [12] combined the classification ability of SVM and the ability of HMM to distinguish dynamic time series, and built bearing fault diagnosis models of SVM and HMM with the help of Sigmoid function and Gaussian model, and established the feature vector of AR parameters for diagnosis, which effectively improved. The accuracy of bearing fault diagnosis is improved. Although these supervised models can achieve fault classification, they are not suitable for situations where the number of samples is large. In response to this defect, some scholars have proposed unsupervised learning methods such as deep learning and deep neural networks [13]. The deep neural network model is implemented on the basis of shallow neural networks by changing the number and structure of intermediate layers. Shallow neural networks mainly include BP neural networks, LVQ neural networks, application of which is relatively extensive. Pu [14] used a three-layer BP Neural Network for bearing fault diagnosis. Wu [15] used signal processing techniques to extract the crest factor, waveform factor, pulse factor, margin factor and kurtosis of rolling bearing vibration time domain signals as feature to diagnose bearing failure. Jiang [16] introduced the signal multi-resolution technology into the analysis of bearing fault vibration signals. A fault diagnosis model for rolling bearings is established by using a Learning Vector Quantization (LVQ) Neural Network. Zhao proposed a new BP Neural Network model based on an improved frog jumping algorithm and applied it to identify bearing failures which has good generalization ability and strong robustness [17]. BP Neural Network has been widely used in many fields, but there are still some shortcomings in the application [18]. In this paper, a fault diagnosis model based on BP Neural Network optimized by Beetle Algorithm is proposed to identify the bearing faults, which greatly improves the performance of BP network and the accuracy can reach more than 95%.

2 Review of BP Neural Network and Beetle Algorithm

2.1 BP Neural Network

BP Neural Network [19] is called Multi-layer front feedback network in most cases which is composed of three blocks: input layer, output layer and hidden layer. Figure 1 shows the structure of the BP Neural Network. The BP Neural Network [20] is trained and learned through error back feedback, which can train and store a large number



of mapping relationships without explicitly knowing the mathematical equations reflecting the mapping relationship. After the training data enters the input layer of the network, the weights and activation functions of the hidden layer and the input layer are used to take corresponding calculations to obtain the actual output value. Then the actual output value expected is used to calculate the error value under actual conditions [21]. If the actual error value exceeds the expected error value, the feedback of this error value should be reversed and new training data will be entered. The condition for stopping the calculation is either to meet the accuracy requirement or to reach the maximum number of iterations. The algorithm process of BP Neural Network is shown in the Figure 1.

2.2 Beetle Algorithm

In 2017, the concept of Beetle Algorithm was first proposed. Compared with the particle swarm optimization algorithm, the Beetle Algorithm requires [22] only one individual, so its computational efficiency is much higher than that of the Particle Swarm Algorithm. Even in the absence of specific expressions of functions and gradient information, such Beetle Algorithm can achieve the desired optimization requirements.

Beetle Algorithm can be described as follows.

- (1) The main part of the beetle is simplified into a centroid, and the two sides of the centroid are the two tentacles of the beetle.
- (2) The ratio of the step size of the beetle to the distance d_0 between the two antennae is a fixed constant. Therefore, it can be obtained that the long distance between the two antennae is the big beetle taking a big step, while the shorter distance between the two antennae is the small beetle, which takes a small step.

- (3) After the Beetle flies to the next step, the head orientation is random.

The specific modeling steps of the Beetle Algorithm are as follows:

- (1) To optimize a specific n -dimensional space, ' x_l ' can be chosen as the sign of the left antenna coordinates, ' x_r ' as the sign of the coordinates of the right antenna and ' x ' as the coordinate symbol of the centroid. The distance between the two antennas is represented by d_0 . According to the theory above, the vector direction of the right horn to the left should be also unrestricted since the orientation of the head of the celestial is not limited. So the random vector ' $dir = rands(n, 1)$ ' can be used to represent it. Then $dir = dir / norm(dir)$ the conclusion of $x_l - x_r = d_0 \cdot dir$ can be got from normalizing the random vector; obviously, ' x_l ', ' x_r ' can also be obtained through the position of the centroid:

$$x_l = x + d_0 \cdot dir / 2, \tag{1}$$

$$x_r = x - d_0 \cdot dir / 2. \tag{2}$$

- (2) Because the function ' f ' to be optimized, the function value of the pair of antennae are calculated and judged.

When $f_{left} < f_{right}$, the beetle moves forward in the direction of the left tentacles:

$$x = x + step \cdot normal(x_l - x_r). \tag{3}$$

When $f_{left} > f_{right}$, the beetle moves forward in the direction of the right tentacles:

$$x = x - step \cdot normal(x_l - x_r). \tag{4}$$

The two cases described above are all for searching for the minimum value of ' f ', which can be uniformly written by using the symbol function sign:

$$x = -step \cdot normal(x_l - x_r) \cdot sign(f_{left} - f_{right}) \tag{5}$$

$$step \cdot dir \cdot sign(f_{left} - f_{right}).$$

3 Feature Extraction Based on Wavelet Packet Transform

3.1 Wavelet Packet Decomposition and Reconstruction

In the actual situation, the rolling bearing vibration signal is usually carried with many noise signals and the noise signal is very easy to cover the weak characteristic signal hidden in the vibration signal, which causes that the useful information can not be obtained. The analysis method of wavelet packet is to decompose the original signal with noise, then the characteristic signal containing energy information will be collected in some frequency bands, while the energy information of the noise signal is evenly dispersed in each frequency band. Therefore, we can separate the noise signal from the characteristic signal by setting an appropriate threshold to make the wavelet coefficient of the noise signal zero, as a result, the goal of reducing noise is achieved. The specific process of using Wavelet Packet Transform is as follows.

- (1) First, the Wavelet Packet Transform is used to decompose the collected original vibration signal. We first select the appropriate wavelet function to determine the number n of decomposition layers of the wavelet packet, that is, the n -layer decomposition of the vibration signal is as follows.
- (2) Calculate the best tree for the wavelet packet. We derive the best tree of wavelet packets based on an entropy criterion given by the wavelet packet.
- (3) The threshold of the Wavelet Packet Decomposition coefficient is quantized. The appropriate threshold and its calculation method are selected, and then the threshold quantization is used to process the Wavelet Packet Decomposition coefficient.
- (4) Decomposition signal wavelet packet reconstruction. Based on the decomposition coefficients of the wavelet packet and the quantized coefficients, the wavelet packet coefficients is reconstructed.

3.2 Wavelet Packet Energy Feature Extraction

The sum of the squares of the individual subspace signals is used to represent the energy value of the wavelet packet. Once the rolling bearing fails, the energy value in each frequency band of its vibration signal will be greatly disturbed. The type of fault is different, and the amount

of energy reflected in each frequency band of the vibration signal is also different. Therefore, the type of fault can be determined based on the distribution of energy values in each frequency band of the vibration signal. The wavelet packet energy value after the fault signal is decomposed by the wavelet packet is taken as a feature, and then the pattern recognition of the fault is performed according to the feature. The main steps of extracting features of vibration signals are as follows.

- (1) Wavelet Packet Decomposition is performed. The same wavelet base is selected from the noise-reduced signals, and the noise-reduced signal is decomposed by 'N' layer. The component characteristics of the 2^N bands of the decomposed N_{th} layer (j = 0, 1, 2, ..., 2^N - 1) is picked, where 'j' is the number of nodes.
- (2) The signals of 2^N frequency band components are selected for reconstruction. S_{Nj} represents the reconstructed signal of X_{Nj}, the total signal S can be expressed as

$$S = \sum_{j=0}^{2^N-1} S_{Nj}. \tag{6}$$

- (3) To calculate the sum of energy in each frequency range of the reconstructed signal, let the energy corresponding to the reconstructed signal s_{Nj}(j = 0, 1, 2, ..., 2^N - 1) of each band component be: E_{Nj}, (j = 0, 1, 2, ..., 2^N - 1), so

$$E_{Nj} = \int |S_{Nj}(t)|^2 dt = \sum_{k=1}^n |X_{jk}|^2, \tag{7}$$

where X_{jk}, (j = 0, 1, 2, ..., 2^N - 1, k = 0, 1, 2, ..., n) represents the magnitude of each discrete point in the reconstructed signal S_{Nj}.

- (4) Constructing the eigenvector of the energy value, the energy value in each frequency band component of the N_{th} layer is thought of as the eigenvector T:

$$T = [E_{N0}, E_{N1}, E_{N2}, \dots, E_{N2^N-1}]. \tag{8}$$

- (5) The energy value is normalized in order to facilitate subsequent data processing:

$$E = \sqrt{\sum_{j=0}^{2^N-1} |E_{Nj}|^2}, \tag{9}$$

$$T' = T/E, \tag{10}$$

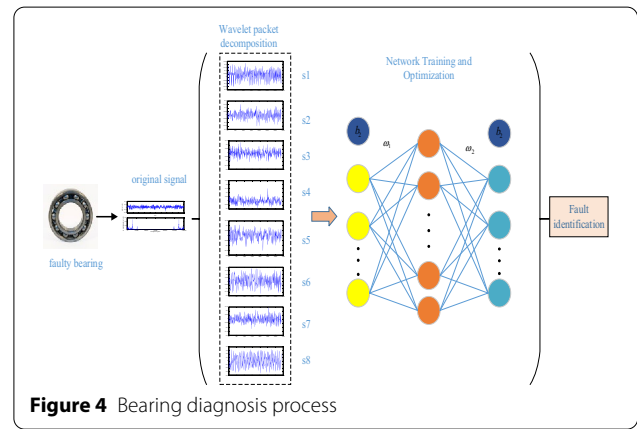
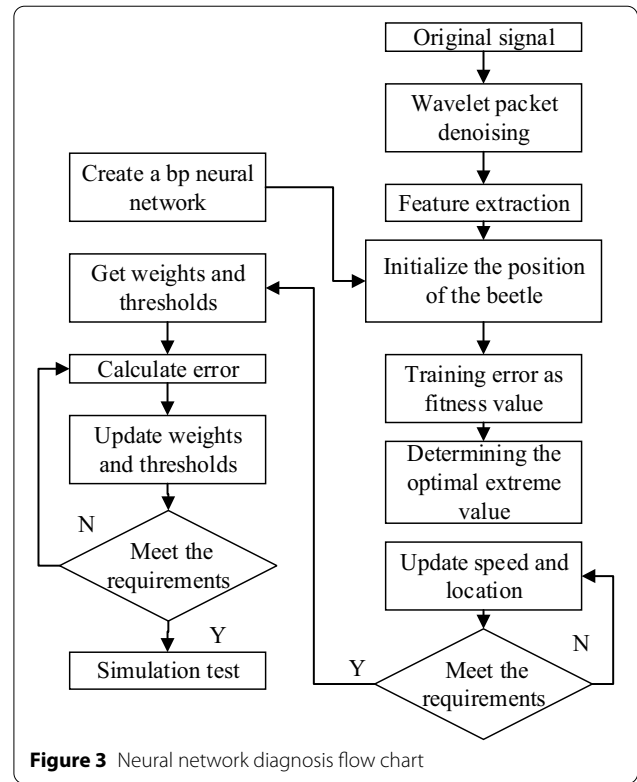
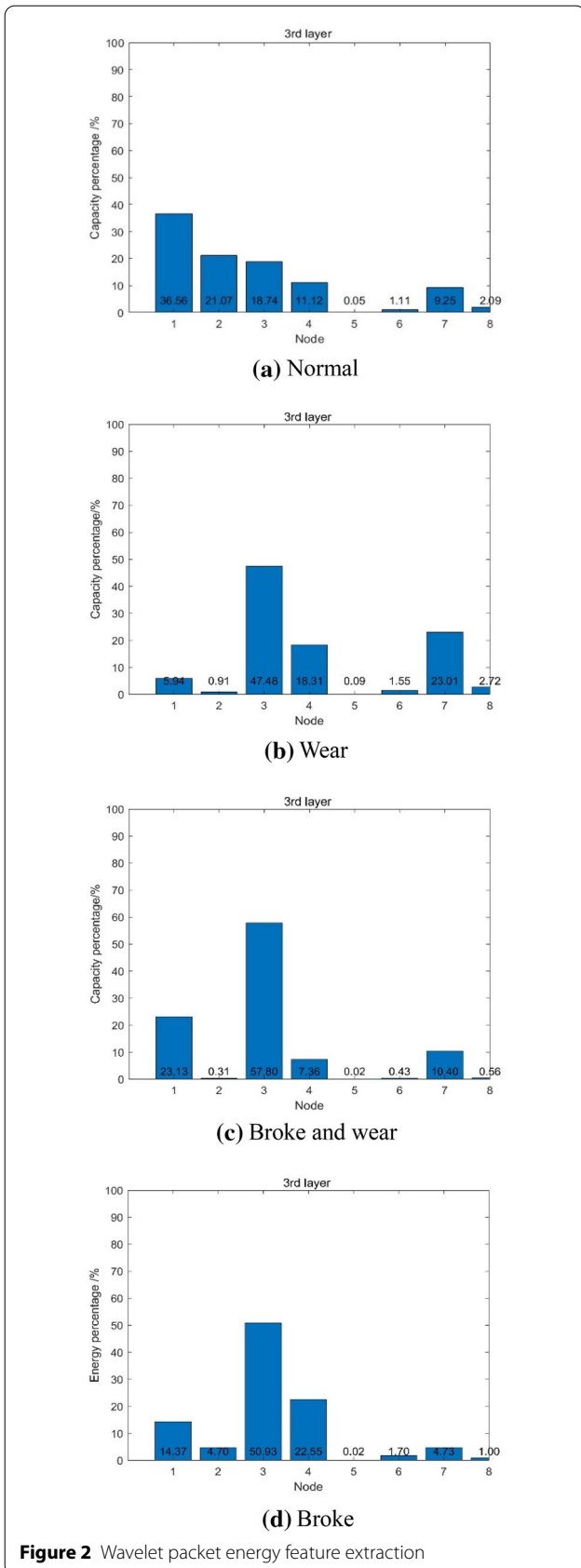
where the vector T' is a feature vector after the normalization process.

As is shown Figure 2, the wavelet packet energy value feature is extracted for the noise-reduced vibration signal. The input speed of the rolling bearing is about 1800 r/min, and its energy value is mostly in the range of 0–2500 Hz. Four Wavelet Packet Decomposition layers are selected, and the first 8 wavelet packet components from the low frequency to the high frequency in the fourth layer are selected to extract the energy values.

3.3 BP Neural Network Optimized by Beetle Algorithm

BP Neural Network which is suitable for the identification of complex patterns, such as multiple symptoms and multiple faults, has strong self-learning, self-adaptation and associative memory capabilities. The BP Neural Network is a single-point search algorithm that does not have the global search ability, on account of using the error function gradient descent method (Figure 3).

Therefore, in the process of learning, BP Neural Network has the disadvantages of poor network performance, poor robustness, slow convergence speed and easy to fall into local extreme. The principle of particle swarm optimization algorithm is relatively simple, and easy to understand. At the same time, the information of speed and position can be updated continuously to correct the search direction, until the global optimal solution is found. The particle swarm optimization algorithm can use the mutual cooperation between individuals and the way of competing to achieve the purpose of global search, which greatly reduces the probability of falling into the local optimal solution. The BP Neural Network optimized by Particle Swarm Algorithm [23, 24] combines the advantages of the two above, and then fully utilizes the strong global search ability of the particle swarm optimization algorithm to optimize the initial weight and threshold of the neural network. Similar to the particle swarm optimization algorithm for BP Neural Network, Beetle Algorithm [25, 26] can also be combined with BP Neural Network (Figure 4). First, the global search capabilities of the Beetle Algorithm are used to optimize the initial weights and thresholds of the neural network. Then the BP algorithm is used to update the initial weights and thresholds, and the final neural network is obtained. Compared with the Particle Swarm Algorithm, the Beetle Algorithm is much simpler, which greatly



reduces the amount of calculation and greatly improves the efficiency of classification [27, 28].

4 Experimental Demonstrations

The raw data used in the fault simulation are from the rolling bearing failure comprehensive test bench of Case Western Reserve University in the United States, which is shown in Figure 5. Acceleration sensors are used to collect vibration acceleration signals of the rolling bearing under different working conditions. The

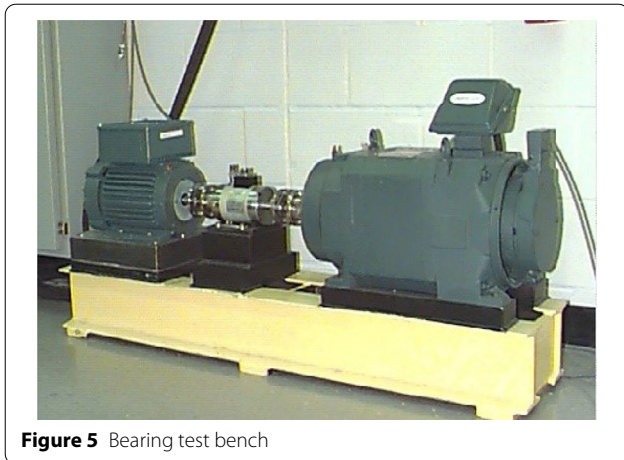


Figure 5 Bearing test bench

test platform mainly includes a 2-horsepower motor, a torque sensor, a power meter and some electronic control equipment. In the experiment, an acceleration sensor was used to collect vibration signals, and the sensor was placed on the motor housing by using a magnetic base. The acceleration sensor is mounted on the drive end of the motor housing. The speed of the motor is 1797 r/min. Vibration signals were collected via a 16-channel DAT recorder and processed later in the MATLAB environment. The sampling frequency of

the digital signal is 12000 S/s, the sampling time is set to 20 s.

4.1 Case 1: Fault Diagnosis of Bearing

4.1.1 Description of Test and Training Samples

The first set of experiments is the simplest experiment. A total of 240 sets of feature vectors are used as training samples: T(r1): 60 sets of feature vectors of normal signals, T(r2): 60 sets of inner ring fault signals, T(r3): 60sets of outer ring fault signals and T(r4): 60 sets of rolling element fault signals. The fault diameter are all 0.021 inches. The test sample set contains T(e1): 60 sets of normal signal feature vectors, T(e2): 60 sets of inner ring fault signal, T(e3): 60 sets of outer ring fault signal and T(e4): 60 sets of rolling element fault signal. Part of the data is shown in Table 1.

4.1.2 Visualization of Characteristic Data

In order to display the extracted energy features, the K-means clustering method is used to cluster the samples, and the generated 2D elements are drawn as a scatter diagram in Figure 6. Different colors represent different categories. It can be seen from the figure that the wavelet packet energy features can produce well clustering performance. Among them, the third type of data points are scattered around the center point in the figure,

Table 1 Feature data of bearing with a fault diameter of 0.021 inches

Type	Feature vector								Label
None	0.9200	0.0502	0.0822	0.3798	0.0002	0.0016	0.0501	0.0623	1
	0.8565	0.0574	0.1251	0.4863	0.0003	0.0021	0.0682	0.0790	1
	0.8074	0.0553	0.1337	0.5581	0.0003	0.0024	0.0852	0.0915	1
	0.8307	0.0727	0.1106	0.5297	0.0003	0.0023	0.0664	0.0849	1
	0.8940	0.0533	0.1024	0.4221	0.0002	0.0018	0.0666	0.0678	1
IRF	0.0558	0.0007	0.0371	0.0079	0.5043	0.2548	0.0210	0.1469	2
	0.0503	0.0005	0.0328	0.0064	0.5386	0.2533	0.0255	0.1125	2
	0.0598	0.0005	0.0179	0.0043	0.5699	0.2225	0.0122	0.0940	2
	0.0729	0.0009	0.0374	0.0056	0.6289	0.1591	0.0234	0.0852	2
	0.0598	0.0006	0.0234	0.0043	0.6555	0.2315	0.0162	0.1041	2
ORF	0.0042	0.0399	0.0199	0.0164	0.5395	0.3213	0.0075	0.0249	3
	0.0035	0.0273	0.0173	0.0091	0.4426	0.1459	0.0048	0.0208	3
	0.0032	0.0298	0.0213	0.0124	0.5038	0.3959	0.0077	0.0333	3
	0.0052	0.0434	0.0146	0.0169	0.5439	0.3265	0.0066	0.0270	3
REF	0.0038	0.0326	0.0251	0.0146	0.5768	0.2530	0.0047	0.0351	3
	1.0000	0.0002	0.0003	0.0006	0.0002	0.0001	0.0004	0.0002	4
	0.9999	0.0004	0.0052	0.0005	0.0023	0.0090	0.0030	0.0045	4
	0.9989	0.0009	0.0005	0.0001	0.0173	0.0102	0.0004	0.0010	4
	0.9999	0.0008	0.0018	0.0005	0.0029	0.0021	0.0004	0.0039	4
	0.9976	0.0020	0.0003	0.0001	0.0265	0.0203	0.0008	0.0052	4

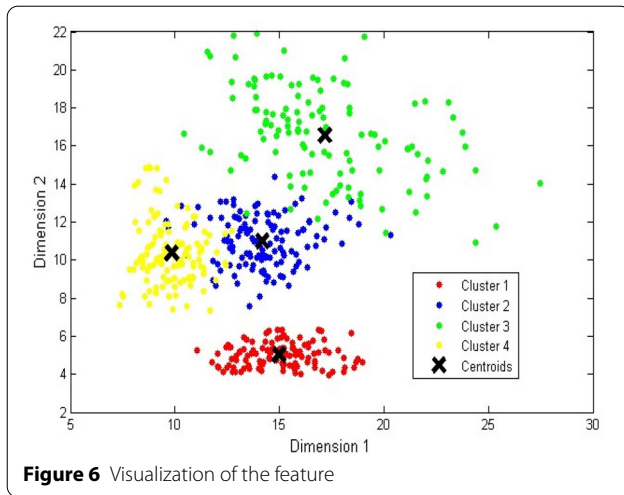


Figure 6 Visualization of the feature

Table 2 Test results of each sample set (case 1)

Sample	T1	T2	T3	T4
Accuracy (%)	100	100	100	100
Sample	T(1,2,3)	T(1,3,4)	T(2,3,4)	All
Accuracy (%)	99.44	96.5	96.2	95

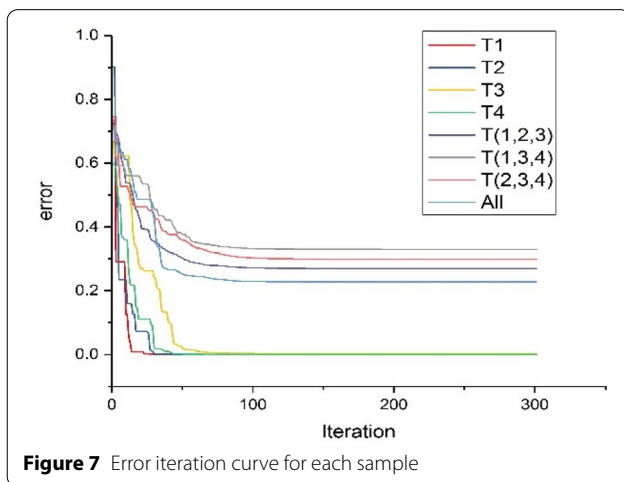


Figure 7 Error iteration curve for each sample

4.1.3 Diagnosis Results of Model

In order to test the performance and effect of the BAS-BP model, different sample sets were set up for experiments. The results is shown in Table 2. The model can diagnose a single sample set with a probability of 100%. However, after the data set is mixed, the diagnosis result has dropped significantly. When mixing the three data sets,

the test results were 99.44%, 96.5% and 96.2%. It can be concluded that there is a coupling effect between sample set 3 (broken and wear) and sample set 4 (broken tooth), which affects the fault diagnosis result. Overall, the diagnostic efficiency of the model can reach 95%, indicating that the model has high diagnostic performance. Figure 7 shows the error iteration curve of each data set training.

4.2 Case 2: Fault Diagnosis of Inner Ring

4.2.1 Description of Test and Training Samples—Inner ring

In the second experiment, a total of 240 sets of feature vectors are used as training samples which contains S(r1): 60 sets of inner ring fault signals with a fault diameter of 0.028 inches, S(r2): 60 sets of inner ring fault signals with a fault diameter of 0.021 inches, S(r3): 60 sets of inner ring fault signals with a fault diameter of 0.007 inches and S(r4): 60 sets of inner ring fault signals for 0.014 inches. The test sample contains S(e1): 60 sets of inner ring fault signals with a fault diameter 0.028 inches, S(e2): 60 sets of inner ring fault signals with a fault diameter 0.021 inches, S(e3): 60 sets of inner ring fault signals with a fault diameter 0.007 inches, and S(e4): 60 sets of inner ring fault signals with a fault diameter 0.014 inches. Part of the data is shown in Table 3.

4.2.2 Visualization of Characteristic Data

These data sets mainly describe the bearing inner ring faults of different sizes, so the coupling interference between the data sets is relatively large, which may greatly interfere with the diagnostic effect of the model. The data set is visualized through K-means clustering, and the data point aggregation is shown in the Figure 8. It can be seen from the Figure 8 that the aggregation effect of each data set is not very satisfactory. It can be analyzed that the diagnostic efficiency for this data set will have a certain drop.

4.2.3 Diagnosis Results of Model

Similarly, set up different sample sets to test model performance. The results are shown in Table 4. The diagnosis result of a single data set is still very accurate, while the diagnosis effect of the mixed data set has been reduced to a certain extent. Even if there is coupling interference between data sets 2 and 3, the overall diagnosis rate can still reach 89.17%, which shows that the model has a certain degree of anti-interference. The error iteration curve is shown in Figure 9.

4.3 Diagnosis Performance and Comparison

In order to further study the model performance, compare the constructed model with the network model optimized by stochastic gradient descent method and

Table 3 Date of the inner ring with different fault diameter

Size	Feature vector								Label
0.028	0.7866	0.0826	0.1359	0.5830	0.0004	0.0025	0.0810	0.0963	1
	0.8242	0.0579	0.1213	0.5380	0.0003	0.0023	0.0733	0.0873	1
	0.8644	0.0603	0.1015	0.4777	0.0003	0.0020	0.0666	0.0781	1
	0.8548	0.0479	0.1063	0.4943	0.0003	0.0020	0.0643	0.0839	1
	0.8529	0.0527	0.1158	0.4957	0.0003	0.0021	0.0634	0.0805	1
0.021	0.0614	0.0008	0.0463	0.0106	0.6735	0.2350	0.0276	0.1436	2
	0.0511	0.0006	0.0414	0.0084	0.6371	0.3552	0.0335	0.0934	2
	0.0273	0.0005	0.0340	0.0084	0.5865	0.2859	0.0249	0.1138	2
	0.0708	0.0239	0.0392	0.0119	0.3932	0.2744	0.0207	0.1207	2
	0.0322	0.0006	0.0324	0.0086	0.6983	0.1620	0.0176	0.1328	2
0.007	0.6262	0.0036	0.2665	0.0342	0.4402	0.2612	0.1061	0.1796	3
	0.4359	0.0040	0.3048	0.0360	0.3652	0.3484	0.1364	0.2477	3
	0.6254	0.004	0.2634	0.0315	0.3427	0.2489	0.1134	0.2325	3
	0.7946	0.0023	0.2096	0.0165	0.2783	0.2476	0.0997	0.1676	3
	0.7778	0.0018	0.1933	0.0208	0.3001	0.2547	0.0813	0.1465	3
0.014	0.2655	0.0007	0.0117	0.0050	0.6392	0.2963	0.0117	0.0588	4
	0.2881	0.0003	0.0075	0.0027	0.6747	0.4394	0.0060	0.0379	4
	0.2420	0.0001	0.0057	0.0014	0.7332	0.4243	0.0020	0.0475	4
	0.2663	0.0003	0.0098	0.0033	0.4430	0.4645	0.0058	0.0994	4
	0.2171	0.0003	0.0085	0.0040	0.4532	0.4578	0.0069	0.1054	4

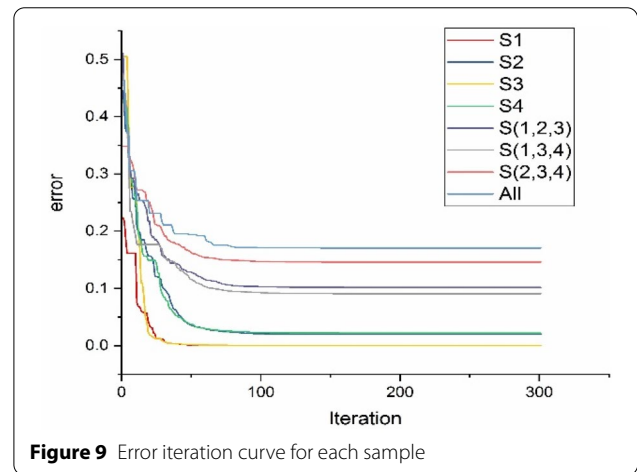
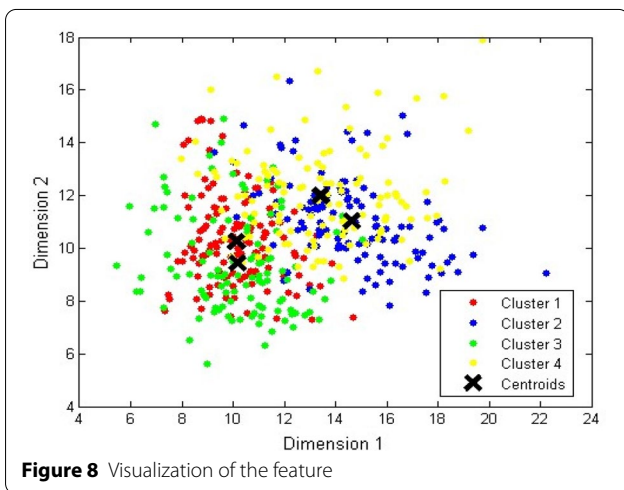
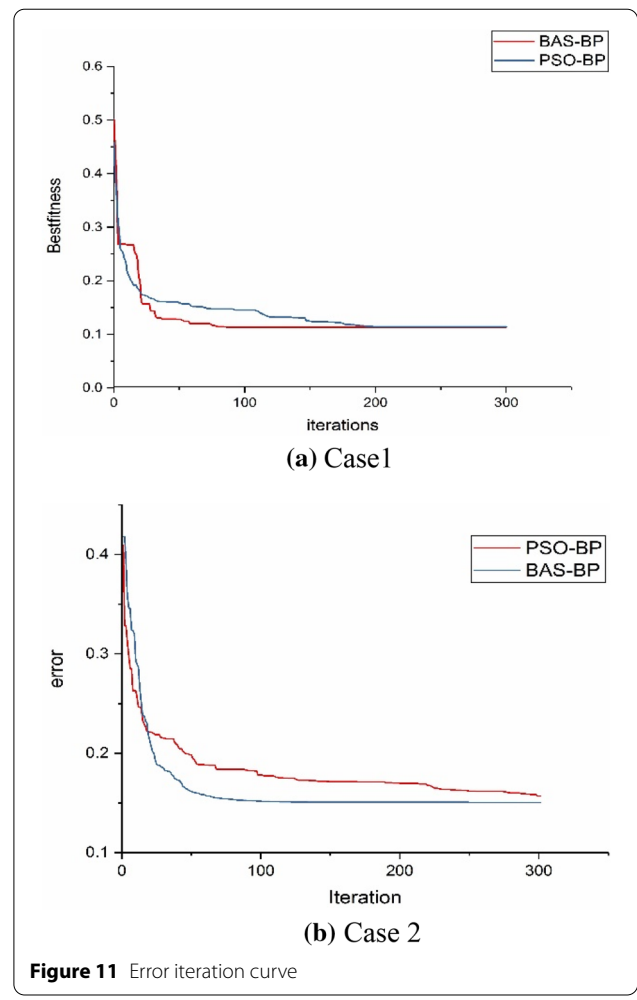
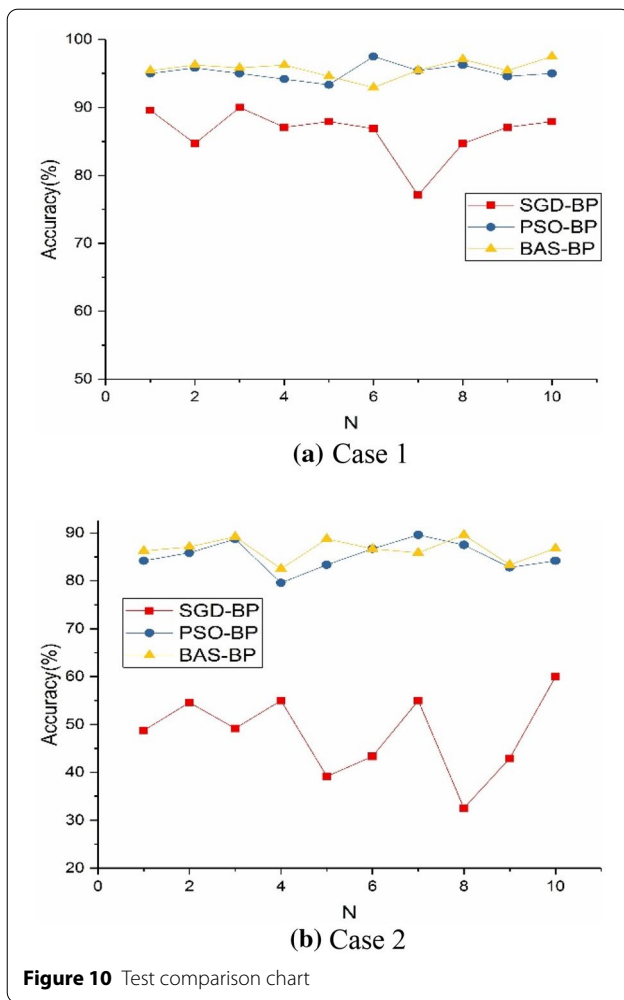


Table 4 Test results of each sample set (case 2)

Sample	S1	S2	S3	S4
Accuracy (%)	100	100	100	100
Sample	S(1,2,3)	S(1,3,4)	S(2,3,4)	All
Accuracy (%)	89.4	96.11	88.9	89.17

Particle Swarm Algorithm. The comparison is made from the diagnosis rate, update speed and the number of iterations to reach the minimum fitness value. The result is shown in Figures 10, 11.

The overall diagnosis rate of the BAS-BP model is significantly higher than the stochastic gradient descent method and Particle Swarm Algorithm. However, the stochastic gradient descent algorithm randomly selects a certain number of samples for training in each iteration, resulting in unsatisfactory training results and easy to fall into local extreme values, which greatly affects the



performance of the model. In the second sample set test, its diagnostic performance almost failed. On the contrary, the model optimized by the BAS algorithm and the Particle Swarm Algorithm has a certain degree of anti-coupling. In the case of aliasing of samples in the sample set, the diagnostic level of the model can still be maintained above 85%. Comparing the error iteration curves of the two algorithms, the declining speed and convergence speed of the BAS curve are significantly faster than the PSO curve, and in the actual training process, the PSO takes significantly longer than the BAS algorithm. Therefore, it can be concluded that the BAS-BP model bearing fault diagnosis has anti-coupling properties, and can quickly and accurately realize fault identification.

5 Conclusions

- (1) A fault diagnosis model based on BP Neural Network optimized by Beetle Algorithm is proposed to identify bearing failure status. The Wave Packet

Energy feature are obtained by Wave Packet Transform learned from different frequency bands of signals and fed into the supervised model for health condition recognition.

- (2) A visual study on the extracted features has been carried out. The study shows that the wavelet packet energy feature can accurately distinguish different working conditions and can be used to evaluate the health status of bearings.
- (3) This article implemented two experiments to test the performance of the BAS-BP model, including diagnostic rate and anti-interference ability. The results show that the BAS-BP model can improve the problem of BP easily falling into local extreme values and identify bearing faults in a shorter time with higher accuracy. At the same time, the coupling interference between data can be avoided to a certain extent.

Acknowledgements

Not applicable.

Authors' Contributions

MX is responsible for the conception and writing of the thesis. WZ participated in the writing of the paper, data processing and analysis. KW is responsible for the design and calculation of experiments. YZ participated in data processing and analysis. YY participated in the writing and retouching of the thesis. All authors read and approved the final manuscript.

Authors' Information

Maohua Xiao, PhD, professor at *Department of Mechanical Engineering, Nanjing Agricultural University*. His research interests include high-end agricultural machinery equipment health maintenance and intelligent manufacturing technology and equipment.

Wei Zhang, born in 1995, is currently a master candidate at *College of Engineering, Nanjing Agricultural University, China*.

Yue Zhu is a lecturer at *Department of Mechanical Engineering, Nanjing Agricultural University, China*. He mainly teaches mechanical vibration technology and theoretical mechanics. His research interests mainly include vibration theory and its applications. He obtained his PhD degree from *Nanjing University of Aeronautics and Astronautics* majoring in mechanical design and theory.

Yilidaer Yiliyas is an associate professor at *College of Mechanical and Electrical Engineering, Xinjiang Agricultural University, China*. His research direction is mainly vibration theory and its applications. He obtained his PhD degree from *Nanjing Agricultural University* majoring in agricultural mechanization engineering.

Funding

Supported by Agricultural Science and Technology Independent Innovation Fund of Jiangsu Province of China (Grant No. CX(19)3081) and Key Research and Development Program of Jiangsu Province of China (Grant No. BE2018127).

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Author Details

¹College of Engineering, Nanjing Agricultural University, Nanjing 210031, China. ²College of Mechanical and Electrical Engineering, Xinjiang Agricultural University, Urumqi 830052, China.

Received: 23 March 2020 Revised: 26 April 2021 Accepted: 10 November 2021

Published online: 04 December 2021

References

- [1] Y G Lei, J Lin, Z J He, et al. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, 2011, 25(5): 1738-1749.
- [2] M M Tahir, A Q Khan, N Iqbal, et al. Enhancing fault classification accuracy of ball bearing using central tendency based time domain features. *IEEE Access*, 2017, 5(99): 72-83.
- [3] Y Li, M Xu, Y Wei, et al. Rotating machine fault diagnosis based on intrinsic characteristic-scale decomposition. *Mechanism and Machine Theory*, 2015, 94: 9-27.
- [4] S R Carpenter, W A Brock. Early warnings of regime shifts in spatial dynamics using the discrete Fourier transform. *Ecosphere*, 2010, 1(5): 1-15.
- [5] Z Feng, D Zhang, M J Zuo. Planetary gearbox fault diagnosis via joint amplitude and frequency demodulation analysis based on variational mode decomposition. *Applied Sciences*, 2017, 7(8): 775.
- [6] H Shao, H Jiang, F Wang, et al. Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet. *ISA Transactions*, 2017, 69: 187-201.
- [7] F Jia, Y Lei, J Lin, et al. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, 2016, 72-73: 303-315.
- [8] X A Yan, M P Jia. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neurocomputing*, 2018, 313: 47-64.
- [9] P K Kankar, S C Sharma, S P Harsha. Fault diagnosis of ball bearings using continuous wavelet transform. *Applied Soft Computing*, 2011, 11(2): 2300-2312.
- [10] H T Zhou, J Chen, G M Dong, et al. Bearing fault recognition method based on neighbourhood component analysis and coupled hidden Markov model. *Mechanical Systems and Signal Processing*, 2016, 66-67: 568-581.
- [11] J Calvo-Zaragoza, A H Toselli, E Vidal. Hybrid hidden Markov models and artificial neural networks for handwritten music recognition in mensural notation. *Pattern Analysis and Applications*, 2019, 22(4): 1573-1584.
- [12] X Chen, Y P Chen, Z Y Long, et al. SVM-HMM-based bearing fault diagnosis method. *Journal of Wuhan University of Technology (Information and Management Engineering Edition)*, 2016, 38(2): 267-270.
- [13] S Khan, T Yairi. A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing*, 2018, 107(1): 241-265.
- [14] X P Pu, K X Chen, J W Xu, et al. Fault diagnosis of bearings based on neural network Breaking method. *Journal of University of Science and Technology Beijing*, 1993, 12(15): 610-613.
- [15] T Wu, S C Yuan, X Meng, et al. A neural network method for vibration diagnosis of rolling bearings. *Mechanical Design & Manufacturing*, 2010, 1: 198-200.
- [16] Y Jiang, Z X Li, Y Tang. LVQ neural network in rolling bearing fault diagnosis applied research. *Mechanical Science and Technology*, 2011, 30(3): 408-411.
- [17] Z Z Zhao, Q S Xu, M P Jia. Improved shuffled frog leaping algorithm-based BP Neural Network and its application in bearing early fault diagnosis. *Neural Computing and Applications*, 2016, 27(2): 375-385.
- [18] A P Markopoulos, S Georgiopoulos, D E Manolakos. On the use of back propagation and radial basis function neural networks in surface roughness prediction. *Journal of Industrial Engineering International*, 2016, 12(3): 389-400.
- [19] J B Ali, N Fnaiech, L Saidi, et al. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Applied Acoustics*, 2015, 89: 16-27.
- [20] Q Long, Y Q Liu, Y P Yang. Fault diagnosis method for wind turbine gearbox based on particle swarm optimization BP Neural Network. *Journal of Dynamics*, 2012, 33(1): 120-125.
- [21] I Zaidi, M Chtourou, M Djemel. Robust neural control of discrete time uncertain nonlinear systems using sliding mode back propagation training algorithm. *International Journal of Automation and Computing*, 2019, 16(2): 213-225.
- [22] Y Q Fan, J P Shao, G T Sun. Optimized PID controller based on Beetle Antennae Search Algorithm for electro-hydraulic position servo control system. *Sensors*, 2019, 19(12): 2727.
- [23] C Ren, N An, J Z Wang, et al. Optimal parameters selection for BP Neural Network based on particle swarm optimization: A case study of wind speed forecasting. *Knowledge-Based Systems*, 2014, 56: 226-239.
- [24] A M Montes de Oca, T Stutzle, M Birattari. Frankenstein's PSO: A composite particle swarm optimization algorithm. *IEEE Transactions on Evolutionary Computation*, 2009, 13(5): 1120-1132.
- [25] P Poupart, N A Vlassis, J Hoey. *An analytic solution to discrete Bayesian reinforcement learning*. *Machine Learning, Proceedings of the Twenty-Third International Conference (ICML 2006)*, Pittsburgh, Pennsylvania, USA, 2006: 697-704.
- [26] N A Kallioras, N D Lagaros, D N Avtzis. Pity Beetle Algorithm—A new metaheuristic inspired by the behavior of bark beetles. *Advances in Engineering Software*, 2018, 121: 147-166.
- [27] X M Xu, L Zhang, Y P Jiang, et al. Active control on path following and lateral stability for truck-trailer combinations. *Arabian Journal for Science and Engineering*, 2019, 44(2): 1365-1377.
- [28] X M Xu, Y P Jiang, H P Lee. Multi-objective optimal design of sandwich panels using a genetic algorithm. *Engineering Optimization*, 2017, 49(10): 1665-1684.