KSME

Springer

Journal of Mechanical Science and Technology 34 (4) 2020

Original Article
DOI 10.1007/s12206-020-0306-1

Keywords:
· Intelligent diagnosis
· Rolling bearing
· Compressed sensing
· Sparse auto-encoder
· Wavelet packet energy entropy

Correspondence to:
Peiming Shi
spm@ysu.edu.cn

Citation:
Shi, P, Guo, X., Han, D., Fu, R. (2020). A sparse auto-encoder method based on compressed sensing and wavelet packet energy entropy for rolling bearing intelligent fault diagnosis. Journal of Mechanical Science and Technology 34 (4) (2020) 1445~1458.
http://doi.org/10.1007/s12206-020-0306-1

Received September 3rd, 2019
Revised December 14th, 2019
Accepted February 4th, 2020

† Recommended by Editor
No-cheol Park

A sparse auto-encoder method based on compressed sensing and wavelet packet energy entropy for rolling bearing intelligent fault diagnosis

Peiming Shi[1], Xiaoci Guo[1], Dongying Han[2] and Rongrong Fu[1]

[1]School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China, [2]School of Vehicles and Energy, Yanshan University, Qinhuangdao 066004, China

Abstract Improving diagnostic efficiency and shortening diagnostic time is important for improving the reliability and safety of rotating machinery, and has received more and more attention. When using intelligent diagnostic methods to diagnose bearing faults, the increasingly complex working conditions and the huge amount of data make it a great challenge to diagnose fault quickly and effectively. In this paper, a novel fault diagnosis method based on sparse auto-encoder (SAE), combined with compression sensing (CS) and wavelet packet energy entropy (WPEE) for feature dimension reduction is proposed. Firstly, vibration signals of each fault type are projected linearly through compressed sensing to obtain compressed signals, which are merged into a low-dimensional compressed signal matrix of multiple fault types. Secondly, the WPEE of low-dimensional compressed signal matrix of multi-fault type is determined, and the eigenvector matrix of bearing fault diagnosis is formed, which greatly reduces the dimension of the eigenvector matrix. Finally, SAE are constructed by adding sparse penalty to auto-encoder (AE) for high-level feature learning and bearing fault classification, and it not only further learns the high-level features of data, but also reduces the feature dimension. Compared with traditional feature extraction methods and the standard deep learning method, the proposed method not only guarantees high accuracy, but also greatly reduces the diagnosis time.

1. Introduction

The vibration process of the bearing is not stable, and the vibration signal is mostly nonlinear, and the noise is complex and changeable in modern industry, and it is more and more difficult to identify bearing fault accurately and quickly by using traditional diagnosis methods. Therefore, it is very important to identify and diagnose bearing faults quickly and accurately with appropriate intelligent diagnosis methods [1-7].

In recent years, artificial intelligence and some signal processing technologies have developed rapidly [8-17], and many intelligent diagnosis methods for rotating machinery faults, including support vector machine [9], artificial immunity [10, 11], blind source separation [12, 13], principal component analysis (PCA) [11], evidence theory [15, 16] and information fusion [17], have been paid more and more attention. Generally, rotating mechanical intelligent diagnosis consists of two parts: feature extraction and fault classification [18]. Widodo et al. [19] found that using support vector machine for feature extraction can achieve high accuracy in fault classification. Lei et al. [20] proposed to use improved distance evaluation technology to extract optimal features and to use adaptive neural fuzzy inference system (ANFIS) to classify faults. Xu et al. [21] used improved fuzzy ARTMAP (IFAM) to identify different fault types after extracting the optimal features. He et al. [22] used the correlation vector machine (RVM) method to establish an intelligent diagnostic model for multi-stage bearings. Wang et al. [23] used possibility theory and the probabilistic neural network (PNN) to identify faults. Li et al. [24, 25] proposed

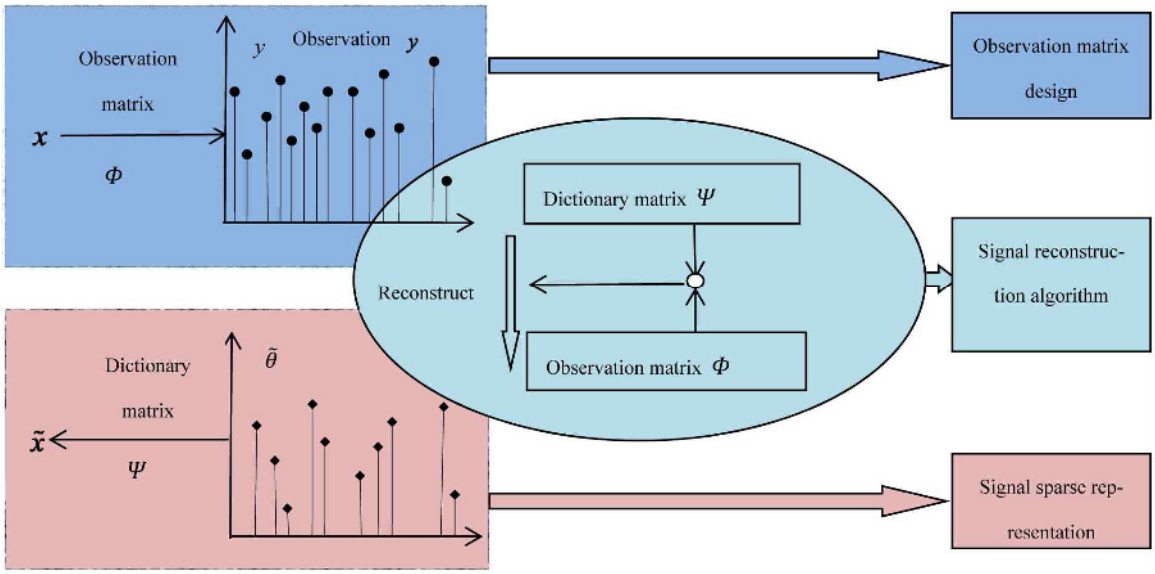


Fig. 1. Key issues in the study of compressed sensing.

combining non-dimensional symptom parameters (NSPs) with ant colony optimization (ACO) to classify early bearing faults. Combining empirical mode decomposition (EMD), improved rule matching strategy (IRMS) and fault decision table (FDT), Dou et al. [26] proposed an easy-to-transplant fault diagnosis method. Based on the symbolic aggregation approximation (SAX) framework, Georgoulas et al. [27] proposed a new method for intelligent fault diagnosis of bearings which can detect bearing faults by the vibration signals and the resulting feature vectors. Zhang et al. [28] proposed the synchronization feature selection and parameter optimization of vibration signals based on SVM and ACO. Jia et al. [29] found that models can be built by deep neural networks (DNN) to extract useful information from raw data.

Although there are many kinds of intelligent diagnostic methods for bearings [30, 31], there are still some problems in the application process. Firstly, the engineering experience of a diagnostic expert is critical to the precise selection of features. Secondly, since the mechanical structures are usually complicated and related to each other, bearing fault types range from single faults to composite faults. Therefore, fault feature extraction can be more effective after mastering the intelligent diagnosis method. Similarly, fault classification algorithms need to be mastered and integrated with feature extraction algorithms to diagnose faults intelligently and accurately. Thirdly, if the size of the sensitive feature matrix is large, it will take a long time to diagnose fault of rotating bearing with intelligent diagnosis algorithm. Therefore, it is urgent to find a suitable dimension reduction algorithm for sensitive feature space [32].

In recent years, only some research on mechanical feature detection based on compressed sensing (CS) has been carried out, and has shown great advantages under experimental conditions. This is a booming research field, which is still in its early stage [33]. For fault diagnosis of rotating machinery, direct spectrum reconstruction in compressed measurement domain is proposed [34]. However, because there is interference in the actual project, it is not feasible to extract the fault features directly from the spectrum. Wang et al. [35] introduced CS theory to solve the problem of time-frequency feature extraction in noise observation. However, it cannot restore the characteristic waveform. Sun et al. [36] proposed an intelligent diagnosis method which based on the CS-DNN, however, the dimension of compressed data is still very high, and the time consumed in classification and diagnosis using DNN is still very long.

Based on the above theories and practical experiments, an intelligent diagnosis method based on CS dimension reduction and wavelet packet energy entropy (WPEE) feature extraction is proposed. The bearing vibration signal was analyzed by this method [37]. The method greatly reduces the complexity of the neural network after the second compression of the data. The high accuracy rate also proves that the useful information is not lost after the data is compressed, and it also proves the method has high reliability [38].

The remainder of this paper is structured as follows. In Sec. 2, CS dimensional reduction theory is briefly introduced. Feature extraction and classification based on WPEE-SAE is given in Sec. 3. The proposed intelligent fault diagnosis method for bearing fault is presented in Sec. 4. In Sec. 5, the experimental results and comparison are analyzed and discussed. In Sec. 6, conclusion is given.

2. Compressed sensing dimensional reduction theory

2.1 Compressed sensing

CS is a new information acquisition theory proposed by Donoho [39] and Candes et al. [40]. CS mainly includes three

aspects: Signal sparse representation, coding measurement and signal reconstruction [41, 42]. The Key issues in the study of compressed sensing can be represented by Fig. 1.

For example, a discrete-time signal *x* of a certain dimension, the signal length is *N*, expressed as *x*(*n*), *n* = 1, 2,...*N*, which can be regarded as n-column vector in R^N. If the transform coefficient on an orthogonal basis of $x \in R^N$ are sparse or compressible, and Ψ can be expanded into a standard orthogonal base $\Psi = [\psi_1, \psi_2, \ldots, \psi_N]$, as in Eq. (1):

$$x = \sum_{i=1}^{N} a_i \psi_i = \Psi a. \tag{1}$$

where a is the coefficient of the sequence x and it satisfies $a_i = \langle x, \psi_i \rangle = \psi_i^T x$. This is the sparse representation of the signal [43, 44].

The signal *x* is linearly projected on the observation matrix $\Phi \epsilon R^{M \times N} (M \ll N)$ to obtain a linear measurement value $y \in R^M$, and the projection process can be expressed by the Eq. (2):

$$y = \Phi x = \Phi \Psi a = \Theta a. \tag{2}$$

where $\Theta = \Phi\Psi$ is the $M \times N$ matrix, which is called the perceptual matrix. The next step is signal reconstruction. If the perceptual matrix Θ satisfies the RIP criterion or Φ and Ψ are uncorrelated, the recovery of the sparse signal minimizes the l_1 norm instead of l_0, as in Eq. (3):

$$\min_a \|a\|_{l_1} s.t. y = \Phi\Psi a. \tag{3}$$

Signal reconstruction can be achieved by solving the optimal norm problem below, as in Eq. (4):

$$\min_a \|a\|_{l_1} s.t. \|y - \Phi\Psi a\| < \varepsilon. \tag{4}$$

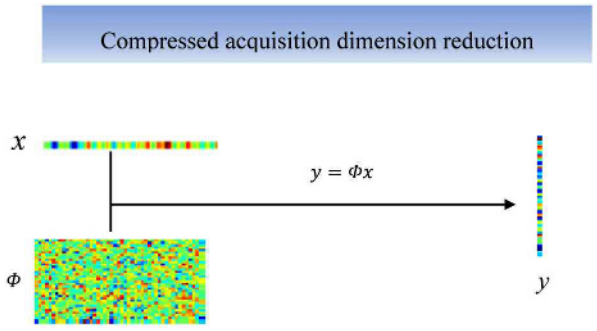


Fig. 2. Schematic diagram of compressed acquisition and dimension reduction.

2.2 Preliminary feature extraction and dimensionality reduction using linear projection of compressed sensing

Suppose *x* is a one-dimensional discrete-time signal of length *N*, expressed as *x*(*n*), *n* = 1, 2,...*N*, which can be regarded as n-column vector in R^N. If the transform coefficient on an orthogonal basis of $x \in R$ is sparse or compressible, and Ψ can be expanded into a standard orthogonal base $\Psi = [\psi_1, \psi_2, \ldots, \psi_N]$ [45], as in Eq. (5):

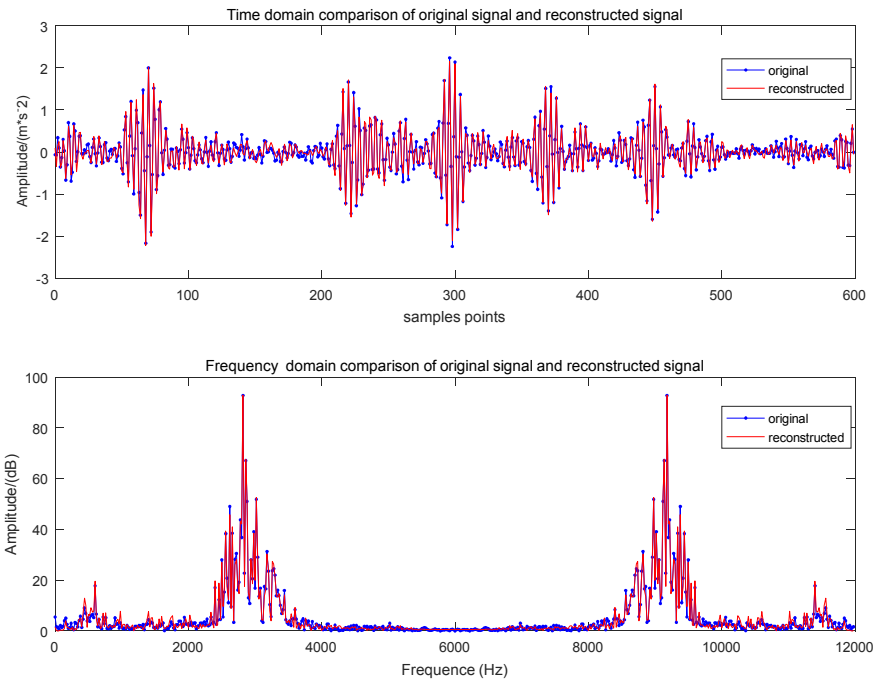


Fig. 3. Time-frequency domain comparison of the original signal and the reconstructed signal.

$$x = \sum_{i=1}^{N} a_i \psi_i = \Psi a. \tag{5}$$

where *a* is the coefficient of the sequence *x* and it satisfies $a_i = \langle x, \psi_i \rangle = \psi_i^T x$. This is the sparse representation of the signal.

The signal x is linearly projected on the observation matrix $\Phi \epsilon R^{M \times N} (M \ll N)$ to obtain a linear measurement value $y \in R^M$, and the projection process can be expressed by the Eq. (6):

$$y = \Phi x \tag{6}$$

This process is a linear projection of the signal. And y is a low-dimensional observation obtained by compression of the original signal. Fig. 2 is a schematic diagram of the linear projection process of the signal.

CS can completely reconstruct high-dimensional signals with low-dimensional random projection under certain conditions. Therefore, under the condition of sparsity and irrelevance, CS can completely reconstruct the original signal using a small amount of observation data.

When the one-dimensional signal x is compressed, the data dimension is greatly reduced, and a low-dimensional observation value is obtained. When the original signal needs to be transmitted and feature extraction, the low-dimensional data greatly reduces the calculation amount and calculation time of the project. When using low-dimensional observations to reconstruct the original signal, it is found that its time domain and frequency are perfectly restored.

3. Feature extraction and classification based on WPEE-SAE

3.1 Wavelet packet energy entropy theory

In the bearing fault diagnosis, when the bearing structure element faults, the energy of each frequency component changes accordingly [46, 47]. Therefore, the change in the energy value of one or several frequency components can be used to characterize the corresponding fault type.

The feature extraction method based on WPEE [47-49] can be performed as following:

(1) The j-layer wavelet packet decomposition is performed on the compressed low-dimensional observation S of a certain fault type. The three-layer wavelet packet decomposition tree structure is shown in Fig. 4, where $W_1^0, W_1^1, W_2^0, W_2^1, \ldots W_3^7$ are the characteristic semaphores of all the frequency components of the first three layers, respectively. Through experiments, it is found that the higher diagnostic accuracy can be obtained when the energy envelope of the wavelet packet of 5 layers is decomposed.

(2) Reconstruct the wavelet packet decomposition coefficients. The characteristic signal sequence of the 32 frequency bands is decomposed by the 1) step to reconstruct 32 wavelet packet reconstructed signals.

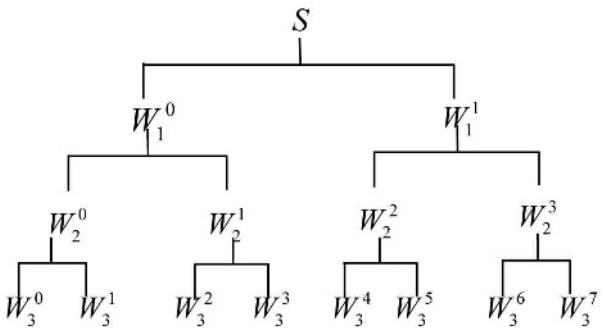


Fig. 4. Three-layer wavelet packet decomposition tree.

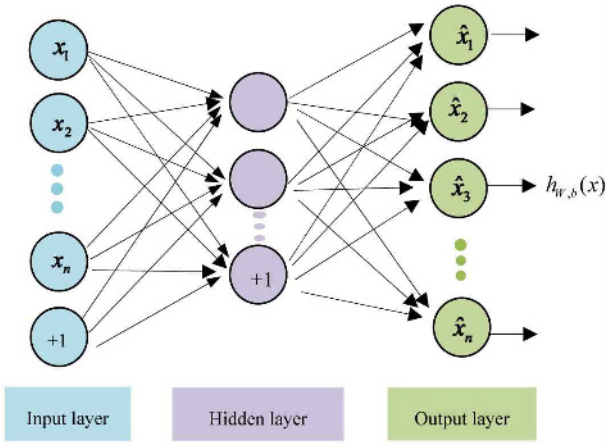


Fig. 5. AE neural network.

(3) The j-layer wavelet packet decomposition sequence of the signal is S_{jk} $(k = 0 \to 2^j - 1)$. Assuming $E_1, E_2, \ldots, E_k$ is the energy spectrum of the signal on *k* scales [50], then, the energy division of the signal is formed on the scale domain E, and the sum of the power E_i (*i* = 1,2,…*k*) of each point is equal to the total signal power E. Then the proportion of energy at each scale can be expressed as follows: $\varepsilon_{jk}(i) = \frac{E_i}{E}$, and $\sum_k \varepsilon_{jk} = 1$. According to the definition of WPEE:

$$H_{jk} = -\sum_{i=1}^{N} \varepsilon_{jk}(i) lg \varepsilon_{jk}(i) \tag{7}$$

where *N* is the original signal length and H_{jk} is the energy entropy of *k*-th wavelet packet of j-level signal [18].

Thereby, the eigenvectors of the 5-layer WPEE decomposition of the low-dimensional observations can be obtained for subsequent classification detection.

3.2 Sparse auto-encoder (SAE) classification theory

3.2.1 Auto-encoder

Suppose there is a training sample set $x^{(i)} \in Re^n$. The AE

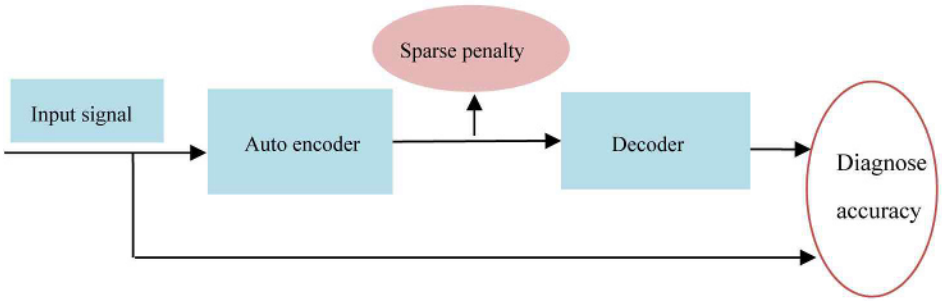


Fig. 6. SAE model.

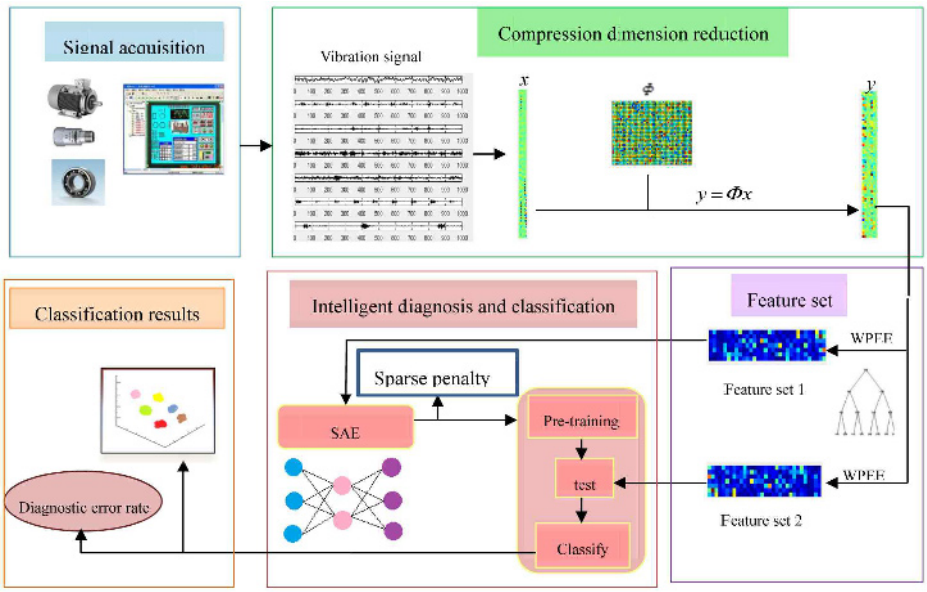


Fig. 7. Schematic diagram of the diagnostic process.

neural network is an unsupervised learning algorithm which uses a back-propagation algorithm and makes the target value equal to the input value, such as $y^{(i)} = x^{(i)}$ [51].

The AE neural network attempts to realize the function of $h_{W,b}(x) \approx x$. That is to say, it tries to approximate an identity function [52]. When the AE neural network is given some restrictions, such as restricting the number of hidden neurons, some interesting relationships can be found [53].

3.2.2 Sparse auto-encoder neural network

The above discussion of AE is based on the assumption that there are few hidden neurons. But even if there are many hidden neurons, we still discover the structure in the input data by applying some other constraints to the AE neural network [54-60].

Adding the regularity limit of L_1 to the AE (L_1 mainly restricts most of the nodes in each layer to be 0, only a few are not 0), and then obtains the SAE. As shown in Fig. 6, it is actually limiting the expression code obtained each time as sparse as possible. Similar to the human brain, an input only stimulates certain neurons, and most other neurons are suppressed, so sparse expressions tend to be more effective than others.

Pre-training of shallow network (only one hidden layer) using supervised training methods usually allows the parameters to converge to a reasonable range. The AE uses a semi-supervised machine learning method of layer-by-layer greedy training, that is, unsupervised learning is performed using AE in the stage of extracting new features from the original sample, but after learning a certain degree of features, the supervised learning method (such as softmax classifier) is used for training and classification. To sum up, it is the semi-supervised learning of layer-by-layer greedy training.

4. Intelligent fault diagnosis method for bearing fault based on CS dimension reduction and WPEE-SAE

Since a large amount of data is generated when bearing vibration data is collected [61-63], an intelligent diagnosis method of bearing fault based on CS dimension reduction and WPEE-SAE is proposed in this paper. The intelligent method proposed in this paper can be divided into following steps,

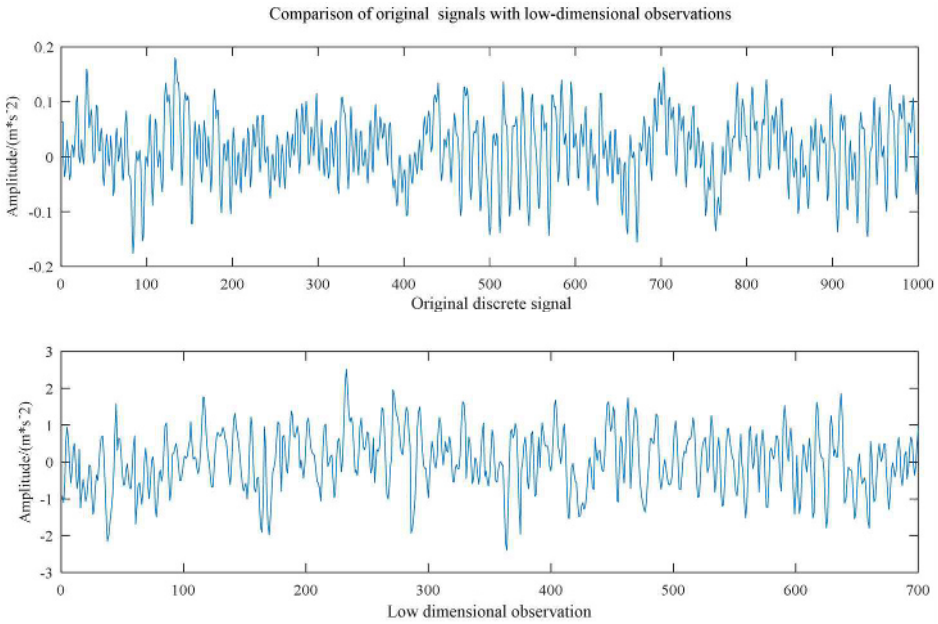


Fig. 8. Comparison of original signals with low-dimensional observations.

namely data dimension reduction, feature extraction, feature vector construction and diagnostic classification. The overall process of this method can be shown in Fig. 7. Original signal is linearly projected to obtain low-dimensional observation *y* whose dimension is $M \times 1$, where $M = N \times c$ and *c* represent the compression ratio. Fig. 8 shows the original discrete signal and the low- dimensional observations.

Step 1: Data reduction and feature extraction using linear projection of compressed sensing: The signal *x* is a one-dimensional vibration signal of $N \times 1$, and the dimension reduction part is compressed and acquired by the CS algorithm. According to Eq. (2), a random Gauss matrix is used as the observation matrix, and the high-dimensional where *N* is the original signal length and H_{jk} is the *k*-th wavelet packet energy entropy of the *j*-th layer of the signal. The WPEE decomposition vector of the *j*-layer of the low-dimensional observations constitutes the feature vector *Y* for subsequent classification detection.

Step 2: Construct a feature vector matrix: Take the first vibration state letter *t* samples of the bearing, each sample consists of *N* signal points, and sequentially perform steps 1 to obtain the WPEE feature set Y_i of the *i*-th vibration state signal. The feature sets $Y_1, Y_2, Y_3, \ldots, Y_i$ of the *i* kinds of vibration state signals are sequentially obtained, and these Y_i constitute TR. Another *t* samples are taken, and according to this step, the WPEE feature sets $YY_1, YY_2, \ldots, YY_i$ of the *i* kinds of vibration state signals are obtained, and these YY_i constitute *TE*. *TR* and *TE* form the feature vector and satisfy $TR \in R^{p \times q}$, $TE \in R^{p \times q}$, *p* and *q* to represent $t \times i$ samples and 2^j feature vector lengths, respectively.

Step 3: Diagnostic classification: The *TR* and *TE* obtained from step 2 will enter the AE neural network as input to adjust the neural network parameters. The training sample *TR* is pre-trained by the AE to realize the deep neural network weight initialization. Then, the BP algorithm is used to achieve the overall fine-tuning of the AE that completes the initialization, and the test samples are tested and classified to obtain fault identification and classification. Because the feature vector dimension obtained in step 2 is only $p \times q$, using low-dimensional observations directly as input into the AE deep neural network, the network complexity is greatly reduced. The diagnostic time required is greatly reduced, so the diagnostic efficiency can be significantly improved.

5. Experimental analysis and comparison

5.1 The first set of experimental data

5.1.1 Basic parameter of test platform

The experiment selected the experimental platform data of Case Western Reserve University in the United States, and Fig. 10 is the experimental platform. A single point of failure was placed on the bearings using EDM technology with fault diameters of 0.007 inches, 0.014 inches, 0.021 inches, 0.028 inches, and 0.040 inches. Data on normal bearings, single-point drive and fan end faults were collected. For the drive end bearings, data was collected using frequencies of 12 kHz and 48 kHz, respectively. For fan-end bearings, data was collected using only a frequency of 12 kHz. The main idea of this method is shown in Fig. 7. The first selected data is shown in Fig. 11.

5.1.2 Low-dimensional compressed signal matrix based on linear projection of compressed sensing

Take the normal signal at the fan-end at 12 kHz, and the inner ring fault signal, the rolling ball fault signal, and the outer ring fault signal with a fault diameter of 0.007 inches and 0.021

Table 1. Changes in the average diagnostic accuracy rate of 15 times with the learning rate changing.

Learning rate	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Accuracy rate (%)	93.95	94.79	96.04	97.97	97.08	94.58	95.21	92.08	85.42	77.92
Time (s)	41	41	41	41	41	41	42	42	42	42

Table 2. Changes in the average diagnostic accuracy rate of 15 times with the number of hidden layer nodes changing.

Hidden layer nodes	20	40	60	80	90	100	110	120	130	140
Accuracy rate (%)	93.73	94.12	95.23	97.43	97.82	98.54	97.67	96.82	96.0	95.24
Time (s)	33	34	36	38	39	41	41	42	42	43

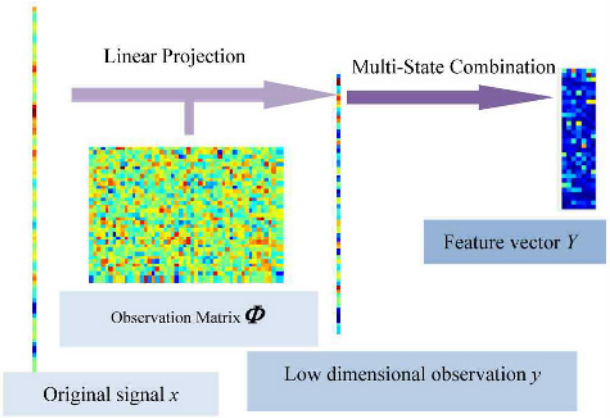


Fig. 9. Schematic diagram of feature extraction.

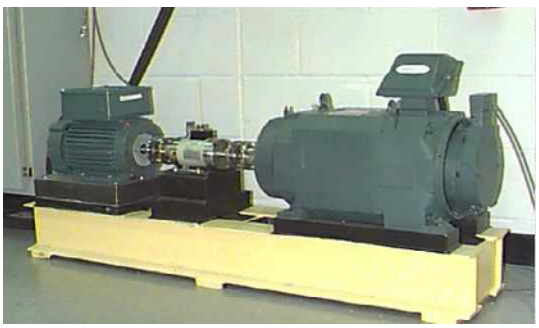


Fig. 10. Experimental platform of Case Western Reserve University.

inches as the experimental signals [64]. Twenty-four samples were selected for each signal, and 2400 data points were selected for each sample to form training group and test group sample data sources. The sample data sources of training group and testing group are 2400×84 data sets.

Set the sample compression ratio (*c*) to 70 %, take the Gaussian random matrix as the observation matrix, and obtain the low-dimensional observation sets *Y* and *YY* of the training group and the test group using linear projection of compressed sensing. The dimensions are both 1680×84, which is used as the training group sample data source and test group sample data source, reducing the workload of data transmission.

5.1.3 Extracting feature sequences using WPEE

In the previous step, the low-dimensional observation sets Y and YY of the training group and the test group with dimensions of 1680×84 were obtained.

Compared with the original data, the dimension of data is much reduced by compressed low-dimensional observations. However the larger the amount of data is, the longer the time it takes when classifiers are needed to classify faults. Therefore, it is very important to further reduce the data dimension of feature parameters.

However, the contradiction is that the data dimension is reduced, and the difficulty of containing useful information of the data becomes more and more difficult. Therefore, it is more important to find a method that can extract useful information of lengthy data. The method uses the WPEE function to extract useful information from the low-dimensional data sets Y and YY. Using the Eq. (6) the WPEE decomposition of the low-dimensional observation set Y and YY is performed, and the feature set with the dimension of only 84×32 is obtained, which are *TR* and *TE*. Using this ultra-low dimensional feature set, the fault signal can be identified and classified faster.

5.1.4 Diagnosis and classification of bearing faults using SAE

In the previous step, we obtained the feature set with only 84×32 dimensions, namely *TR* and *TE*, and then input it into the SAE neural network to identify and classify the existing signal set. The SAE neural network layer selects 4 layers, which are input layer, output layer and two hidden layers. The 5-layer wavelet energy entropy distribution selected by this method is used as the eigenvalue, so the number of nodes in the input layer is 32, the number of hidden layer nodes and output layer nodes are 100, 100 and 7, respectively. So the SAE model is 32-100-100-7. After a series of experimental comparisons, the coefficients are configured as shown in Table 4.

We can get that in Tables 1-3: When the learning rate and the number of hidden layer nodes are changed, the average value of the 15 diagnostic results and the diagnosis time are recorded, and it can be seen that the diagnostic accuracy is greatly different.

Table 3. Changes in diagnostic accuracy rate, diagnostic time, and diagnostic standard deviation with diagnostic times changing from 5 to 50.

Diagnostic times	5	10	15	20	25	30	35	40	45	50
Accuracy rate (%)	92.5	94.53	94.79	96.13	96.15	96.87	96.99	97.97	97.65	97.25
Accuracy standard deviation	0.06	0.05	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.01
Time	21	31	42	53	63	73	82	90	102	110

Table 4. Parameter setting of SAE neural network.

Parameter name	Parameter setting	Parameter name	Parameter setting
Activation function	Sigmoid	L2 norm constraint term	0
Learning rate	2	Non sparse penalty	0.05
Weight momentum factor	0.5	Sparse target value	0.05
Dropout fraction	0.005	Number of network nodes	32-100-100-7

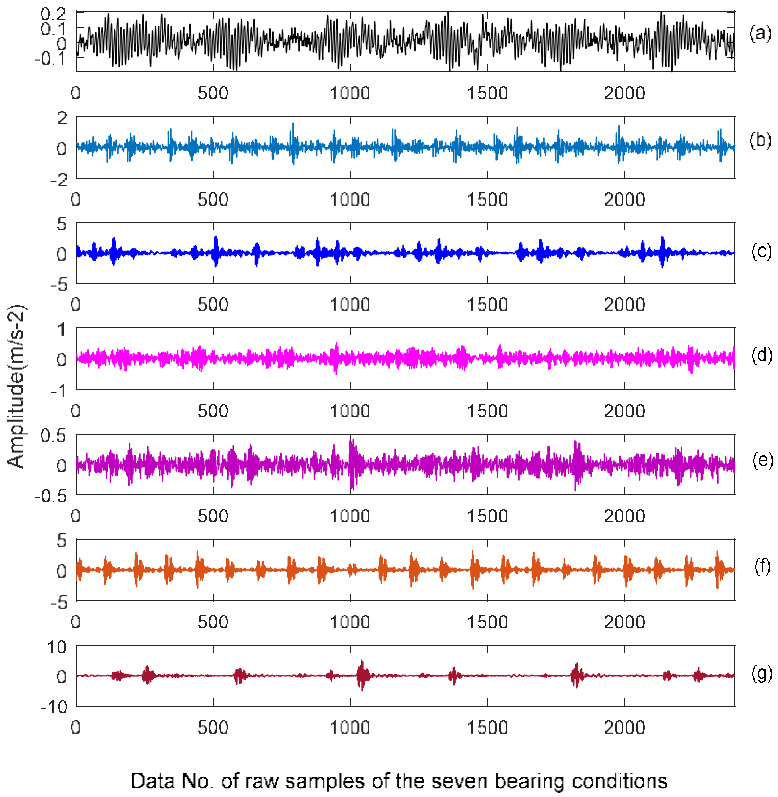
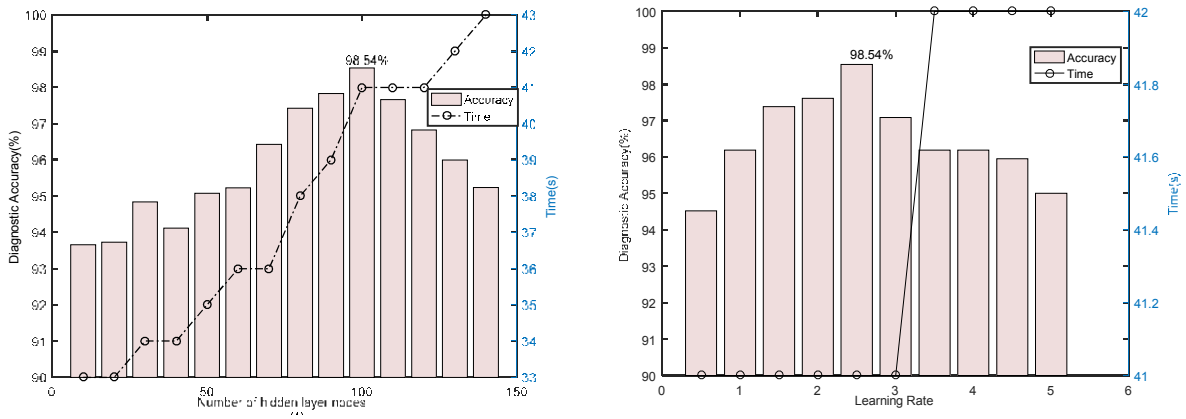


Fig. 11. Raw samples of the seven bearing conditions: (a) Normal condition; (b) slightly inner race fault; (c) severe inner race fault; (d) slightly ball fault; (e) severe ball fault; (f) slightly outer race fault; (g) severe outer race fault.

It can be seen in Figs. 12 and 13 that the diagnostic accuracy is the highest when the learning rate is 2, and the experimental accuracy gradually decreases when the learning rate is more than 2. Even when the learning rate is 5, the accuracy is only 77.92 %. So the learning rate is set to 2. When the number of hidden layer nodes is changed, the change of diagnostic rate is obvious, and the regularity is not so obvious. However, the diagnostic accuracy is the highest when the number of hidden layer nodes is set to 100. When the number of diagnoses is increasing, the accuracy of the diagnosis is gradually increased, and the standard deviation of the accuracy is gradually reduced. The parameters can be set as follows in Table 4.

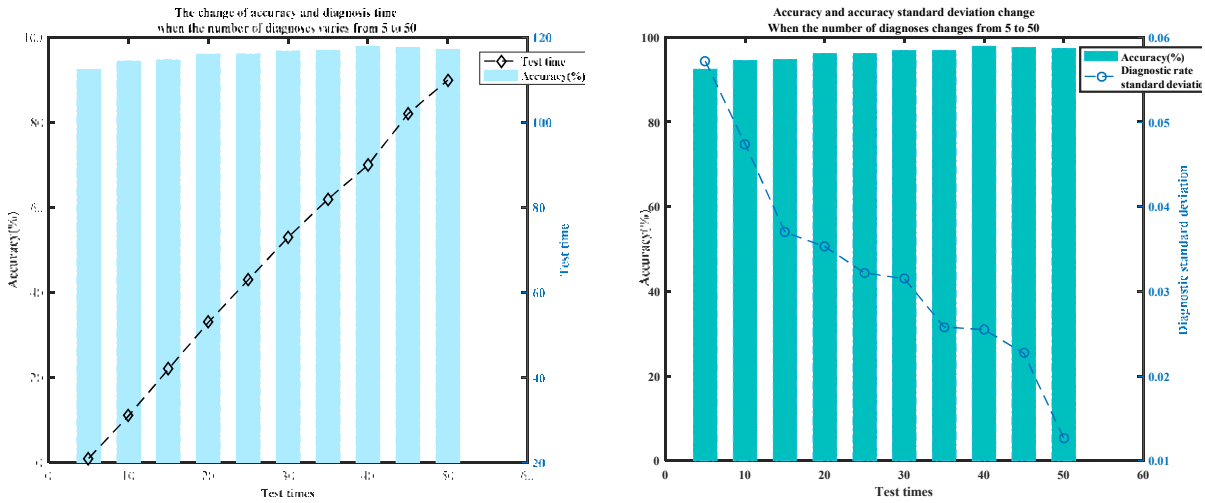
5.1.5 The diagnostic results of this method were compared with those of other methods

By using the energy entropy function of wavelet packet to process the low-dimensional observations, the ultra-low-dimensional feature set is obtained. Without affecting the accuracy of diagnosis, the classification and detection tasks of the AE are reduced, the diagnosis time is greatly reduced, and the diagnosis efficiency is greatly improved. The original data node



(a) (b)

Fig. 12. (a) Changes in diagnostic accuracy and diagnostic time with the number of hidden layer nodes changing from 10 to 140; (b) changes in diagnostic accuracy and diagnostic time with the learning rate changing from 0.5 to 5.



(a) (b)

Fig. 13. (a) Changes in accuracy and time with the number of tests changing from 5 to 50; (b) changes in accuracy and accuracy standard deviation with the number of diagnoses changing from 5 to 50.

settings need to be processed are 720-200-60-7, because the construction of the neural network is very complex, the time required and equipment requirements will be high. However, the WPEE will concentrate the feature information to the feature set of 32×84. When the data is input to the SAE neural network, the training and testing time can be greatly reduced, but at the same time, the fault detection is not reduced. In addition, compared with the previous method of calculating and integrating feature vectors by using various parameters (harness factor, variance, waveform factor, etc.), this method only needs to extract the WPEE of low-dimensional observations to accurately identify the fault type. Compared with the PSO-SVM algorithm, the fault diagnosis rate of the proposed method can reach 98.5417 %, and the diagnosis rate of a few cases can be as high as 100 %. From this it can be concluded that the method can be applied to bearing fault diagnosis.

Contrast 1: Directly input the bearing signals of various fault types into the neural network. Without compression sensing, the diagnostic time will be extremely long and there will be large fluctuations, and the diagnostic accuracy cannot be guaranteed.

Contrast 2: The low-dimensional observations obtained by using CS are directly input into the neural network, and the low-dimensional data is not processed by the WPEE. Although the average diagnostic rate of the experiment is 97 %, the diagnostic time is up to 170 s.

Contrast 3: Using low-dimensional observations for feature extraction, extracting parameters (kurtness factor, variance, waveform factor, etc.) and then using sensitive feature parameters, the PSO-SVM algorithm can only achieve 95.83 %.

In this paper, when the method is used, the average diagnostic rate of the experiment can reach 98.5417 %, and the diagnosis time is only 41 s. It not only does not need to extract multiple sensitive feature parameters, but also reduces the

Table 5. Comparison of three existing methods with this method.

Experimental method	Number of samples	Sample contains data points	Time spent on 15 diagnoses (s)	Accuracy standard deviation	Accuracy
1	2800	2400	416.7	0.87	96.75 %
2	2800	2400	170	0.43	97.47 %
3	100	4096	*	1.52	95.83 %
CS+WPEE-SAE	168	2400	41	0.0269	98.54 %

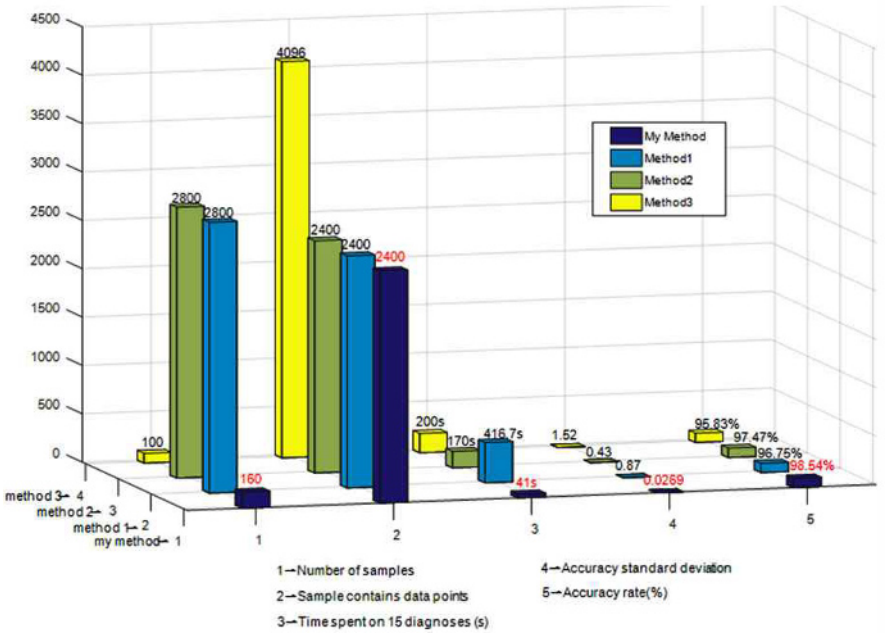


Fig. 14. The variation of five observation parameters with different methods.

complexity of the method, it can also use the relatively intelligent method to diagnose and analyze, improve the operability and facilitate the modular application.

The advantages of this method are now compared in Table 5 and Fig. 14. In Table 5, since the method 3 cannot directly obtain the diagnosis result based on the acquired signal by the program design, the diagnosis time is not counted. It is easy to find from Table 5 and Fig. 13 that there are many samples required for the diagnosis of method 1 and method 2. The method 1 takes too long to diagnose the process, while method 2 shortens the diagnosis time, it requires a large number of samples. Although method 3 requires only a small amount of samples, it requires a large amount of analysis to extract sensitive characteristic parameters, and the accuracy rate can only reach 95.83 %. In contrast, in addition to extracting useful signal features from finite samples, the use of WPEE can also simplify the time and complexity of neural network training and testing, and obtain stable diagnostic results and high accuracy rate. This will save a lot of time in the actual diagnosis of the fault. Since the fault can be accurately identified and classified when there are only a small number of samples, the fault can be found more timely and the loss caused by the bearing failure can be reduced.

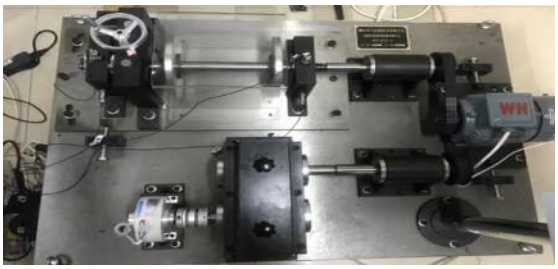


Fig. 15. The experimental platform and rolling bearing.

5.2 The second set of experimental data

The second set of data is also used to verify the validity of the method. This set of data was collected on the QPZZ-II rotating machinery vibration analysis experimental platform, where the sampling frequency is 3 kHz and the bearing speed is 1500 r/min. Bearing vibration data of normal state, inner-race fault, outer-race fault, and rolling ball fault are collected to verify the effectiveness of the method proposed in this paper as show in Fig. 15.

Step 1: Similar to Sec. 5.1.2, the raw vibration signals of four states are shown in Fig. 16. Take 2400 points of each state

Table 6. Parameter setting of SAE neural network.

Parameter name	Parameter setting	Parameter name	Parameter setting
Activation function	Sigmoid	L2 norm constraint term	0
Learning rate	2	Numepochs of pre training	20
Weight momentum factor	0.5	Numepochs of fine tuning	20
Dropout fraction	0.001	Number of network nodes	32-20-20-4

Table 7. Changes in diagnostic accuracy rate, diagnostic time, and diagnostic standard deviation with diagnostic times changing from 5 to 50.

Diagnostic times	5	10	15	20	25	30	35	40	45	50	55
Accuracy rate (%)	100	100	98.33	100	99	99.17	98.05	99.38	98.04	99	99.48
Accuracy standard deviation	0	0	0.064	0	0.05	0.045	0.07	0.039	0.065	0.049	0.033
Time (s)	16	20	25	31	37	42	44	50	54	60	64

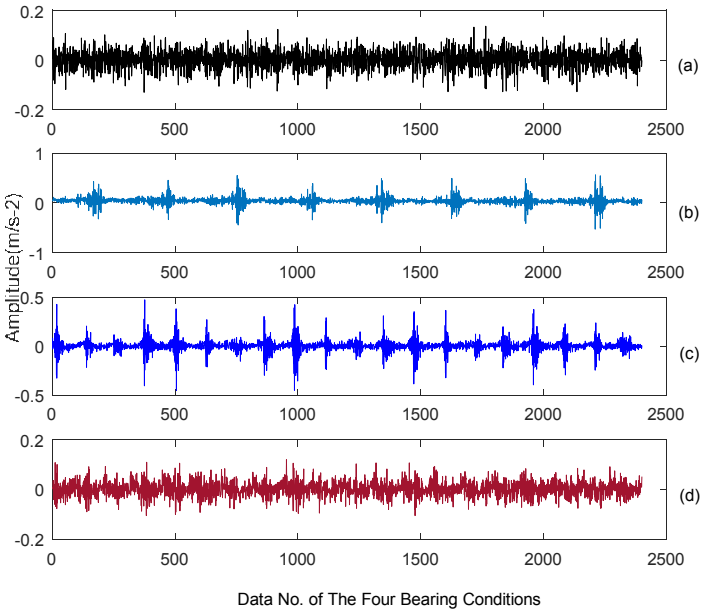


Fig. 16. Raw samples of the seven bearing conditions: (a) Normal condition; (b) inner-race fault; (c) rolling ball fault; (d) outer-race fault.

vibration data as a sample, and take 15 samples from each of the four states to form a 2400×60 training set.

Step 2: Based on the compression sensing algorithm, the sample compression rate (c) is set to 70 %, and the Gauss random matrix is used as the observation matrix. The original vibration signal is linearly projected to obtain a low-dimensional observation matrix, which is recorded as y. similarly, based on the vibration data different from step1, another set of low-dimensional observation matrices can be obtained, which is recorded as *YY*. After compression, the dimension of the training set is 1680×60.

Step 3: Feature sequence extraction based on WPEE. Using the Eq. (6), the WPEE decomposition of the low-dimensional observation set *Y* and *YY* is performed, and the feature set with the dimension of only 60×32 is obtained, which are *TR* and *TE*. Using this ultra-low dimensional feature set, the fault signal can be identified and classified faster.

Step 4: Diagnosis and classification using SAE. The *TR* obtained from step 3 is used as training sample to adjust each parameter, and the result of parameter adjustment is shown in Table 6. The model is built by using the parameters after fine adjustment, *TR* is used as training sample to input into SAE neural network for unsupervised pre training, and then the supervised fine adjustment process is carried out according to the label. *TE* is input into the trained model as a test sample. The changes of accuracy, standard deviation of multiple diagnosis accuracy and diagnosis time with the number of repeated diagnosis are shown in Fig. 17 and Table 7.

It can be seen from Table 7 and Fig. 16 that because the features of the example are obvious, only a relatively simple neural network structure is needed to get a high diagnostic accuracy. It also takes only 25 seconds for 15 repetitions. It can be

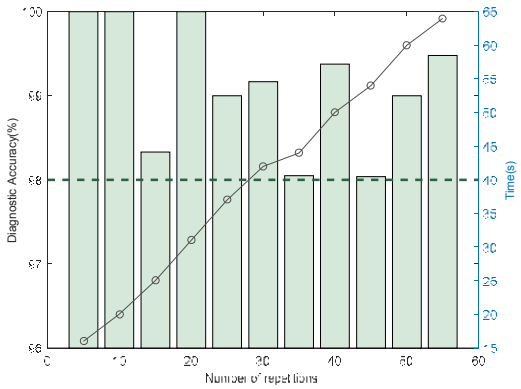


Fig. 17. Diagnostic accuracy and diagnostic time from 5 to 55 times with repeated training.

found from Table 7 that in repeated training, the value of standard deviation is generally lower, that is, the diagnosis result is more stable. Therefore, it can be verified that the proposed SAE neural network intelligent diagnosis method based on CS and WPEE can get stable and accurate diagnosis results on the premise of only consuming a short diagnosis time, and the method is effective.

6. Conclusion

In the existing fault diagnosis methods for rotary bearings, most of them need to analyze and process a large amount of data, and it takes a long time to transfer a large amount of data, but by compressing the data and only processing the compressed low-dimensional data, the fault diagnosis will become faster and more efficient. In this paper, the low-dimensional observations containing key information of the signal are obtained by using CS, and then the ultra-low-dimensional feature sets are obtained by WPEE. Finally, the SAE neural network is used to classify them. The acquisition of compressed observations greatly reduces the computational complexity of subsequent analysis of WPEE and sparse classification. The WPEE was used to analyze the signal in the diagnosis of circuit faults. However, this paper finds that it can be used in the bearing fault signal, and the feature set dimension obtained by WPEE processing is very low. At this time, the feature set also contains enough information to classify the faults by using the SAE neural network combined with the softmax classifier, thereby reducing the diagnostic complexity of the classification, shortening the diagnosis time, and ensuring high diagnostic accuracy. The application of technology to the industry provides a theoretical basis.

Acknowledgments

The studies were funded by the National Natural Science Foundation of China [Grant numbers 61973262 and 51875500], Natural Science Foundation of Hebei Province (Grant number E2019203146), and Hebei Province Graduate Innovation Funding Project [Grant number 2019000629].

References

[1] Y. G. Lei, F. Jia, J. Lin, S. B. Xing and S. X. Ding, An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, *IEEE Transactions on Industrial Electronics,* 63 (5) (2016) 137-3147.

[2] Y. Wang, J. Xiang, Q. Mo and S. He, Compressed sparse time-frequency feature representation via compressive sensing and its applications in fault diagnosis, *Measurement,* 68 (2015) 70-81.

[3] H. Shao, H. Jiang, F. Wang and Y. Wang, Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet, *ISA Transactions,* 69 (2017) 187-201.

[4] X. Yan, M. Jia, W. Zhang and L. Zhu, Fault diagnosis of rolling element bearing using a new optimal scale morphology analysis method, *ISA Transactions,* 73 (2018) 165-180.

[5] V. T. Thang, N. A. Tuan and N. V. Tiep, Evaluation of grinding wheel wear in wet profile grinding for the groove of the ball bearing's inner ring by pneumatic probes, *Journal of Mechanical Science and Technology*, 32 (3) (2018) 1297-1305.

[6] N. Upadhyay and P. K. Kankar, Diagnosis of bearing defects using tunable Q-wavelet transform, *Journal of Mechanical Science and Technology,* 32 (2) (2018) 549-558.

[7] D. Y. Dou, J. Jiang, Y. Wang and Y. Zhang, A rule-based classifier ensemble for fault diagnosis of rotating machinery, *Journal of Mechanical Science and Technology,* 32 (6) (2018) 2509-2515.

[8] R. R. Fu, H. Wang, M. M. Han, D. Y. Han and J. D. Sun, Scaling analysis of phase fluctuations of brain networks in dynamic constrained object manipulation, *International Journal of Neural Systems,* 32 (1) (2018) 91-99.

[9] B. Pang, G. Tang, C. Zhou and T. Tian, Rotor fault diagnosis based on characteristic frequency band energy entropy and support vector machine, *Entropy,* 20 (12) (2018) 932.

[10] G. C. Silva, R. M. Palhares and W. M. Caminhas, Immune inspired fault detection and diagnosis: A fuzzy-based approach of the negative selection algorithm and participatory clustering, *Expert Systems with Applications,* 39 (16) (2012) 12474-12486.

[11] N. Bayar, S. Darmoul, S. Hajri-Gabouj and H. Pierreval, Fault detection, diagnosis and recovery using artificial immune systems: A review, *Engineering Applications of Artificial Intelligence,* 46 (2015) 43-57.

[12] T. Benkedjouh, N. Zerhouni and S. Rechak, Tool wear condition monitoring based on continuous wavelet transform and blind source separation, *International Journal of Advanced Manufacturing Technology,* 97 (2018) 3311-3323.

[13] H. C. Sun, L. Fang and J. Z. Guo, A fault feature extraction method of rotating shaft with multiple weak faults based on underdetermined blind source signal, *Measurement Science and Technology,* 29 (2018).

[14] Y. Song, J. Liu, N. Chu, P. Wu and D. Wu, A novel demodulation method for rotating machinery based on time-frequency analysis and principal component analysis, *Journal of Sound*

and Vibration, 442 (2019) 645-656.

[15] H. P. Zhang and Y. Deng, Engine fault diagnosis based on sensor data fusion considering information quality and evidence theory, *Advances in Mechanical Engineering,* 10 (11) (2018).

[16] J. Q. Liu, A. F. Chen and N. Zhao, An intelligent fault diagnosis method for bogie bearings of metro vehicles based on weighted improved D-S evidence theory, *Energies,* 11 (2018) 232.

[17] F. Y. Xiao, Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy, *Information Fusion,* 46 (2019) 23-32.

[18] J. He, M. Cao, W. Wang and X. Zhu, Partial discharge pattern recognition algorithm based on sparse self - coding and extreme learning machine, *High Voltage Apparatus,* 11 (2018).

[19] A. Widodo and B. S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, *Mech. Syst. Signal Process,* 21 (6) (2007) 2560-2574.

[20] Y. G. Lei, Z. J. He and Y. Y. Zi, A new approach to intelligent fault diagnosis of rotating machinery, *Expert Systems With Applications,* 35 (4) (2008) 1593-1600.

[21] Z. B. Xu, J. B. Xuan, T. L. Shi, B. Wu and Y. M. Hu, A novel fault diagnosis method of bearing based on improved fuzzy ARTMAP and modified distance discriminant technique, *Expert Systems With Applications,* 36 (9) (2009) 11801-11807.

[22] C. He, C. Liu, Y. Li and J. Yuan, Intelligent fault diagnosis of rotating machinery based on multiple relevance vector machines with variance radial basis function kernel, *Proceedings of The Institution of Mechanical Engineers Part C-Journal of Mechanical Engineering Science*, 225 (2011) 1718-1729.

[23] H. Q. Wang and P. Chen, Intelligent diagnosis method for rolling element bearing faults using possibility theory and neural network, *Computers & Industrial Engineering,* 60 (4) (2011) 511-518.

[24] K. Li, P. Chen and H. Wang, Intelligent diagnosis method for rotating machinery using wavelet transform and ant colony optimization, *IEEE Sensors Journal*, 12 (2012) 2474-2484.

[25] K. Li, P. Chen and S. M. Wang, An intelligent diagnosis method for rotating machinery using least squares mapping and a fuzzy neural network, *Sensors,* 12 (5) (2012) 5919-5939.

[26] D. Y. Dou, J. G. Yang, J. T. Liu and Y. K. Zhao, A rule-based intelligent method for fault diagnosis of rotating machinery, *Knowledge-Based Systems,* 36 (2012) 1-8.

[27] G. Georgoulas, P. Karvelis, T. Loutas and C. D. Stylios, Rolling element bearings diagnostics using the symbolic aggregate approximation, *Mech. Syst. Signal Process,* 60-61 (2015) 229-242.

[28] X. L. Zhang, W. Chen, B. J. Wang and X. F. Chen, Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization, *Neurocomputing,* 167 (2015) 260-279.

[29] F. Jia, Y. G. Lei, J. Lin, X. Zhou and N. Lu, Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, *Mech. Syst. Signal Process,* 72-73 (2016) 303-315.

[30] H. D. Shao, H. K. Jiang, H. Z. Zhang, W. J. Duan, T. C. Liang and S. P. Wu, Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing, *Mech. Syst. Signal Process,* 100 (2018) 743-765.

[31] H. Shao, H. Jiang, H. Zhang and T. Liang, Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network, *IEEE Transactions on Industrial Electronics,* 99 (2017) 2727-2736.

[32] I. Attoui, N. Fergani, N. Boutasseta, B. Oudjani and A. Deliou, A new time-frequency method for identification and classification of ball bearing faults, *Journal of Sound and Vibration,* 397 (2017) 241-265.

[33] Z. H. Du, X. F. Chen, H. Zhang and B. Y. Yang, Compressed-sensing-based periodic impulsive feature detection for wind turbine systems, *IEEE Transactions on Industrial Informatics* (2017) 2933-2945.

[34] Z. Du, X. Chen, H. Zhang, H. Miao, Y. Guo and B. Yang, Feature identification with compressive measurements for machine fault diagnosis, *IEEE Transactions on Instrumentation and Measurement,* 65 (2016) 977-987.

[35] Y. Wang, J. Xiang, Q. Mo and S. He, Compressed sparse time-frequency feature representation via compressive sensing and its applications in fault diagnosis, *Measurement,* 68 (2015) 70-81.

[36] J. Sun, C. Yan and J. Wen, Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning, *IEEE Transactions on Instrumentation and Measurement,* 99 (2017) 185-195.

[37] P. Ma, H. Zhang, W. Fan, C. Wang and X. Zhang, A novel bearing fault diagnosis method based on 2D image representation and transfer learning-convolutional neural network, *Measurement Science and Technology,* 30 (2019) 055402.

[38] H. Shao, H. Jiang and W. Duan, Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing, *Mechanical Systems and Signal Processing,* 100 (2018) 743-765.

[39] D. L. Donoho, Compressed sensing, *IEEE Transactions on Information Theory,* 52 (4) (2006) 1289-1306.

[40] E. J. Candes and T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies, *IEEE Transactions on Information Theory,* 52 (12) (2006) 5406-5425.

[41] G. Tang, Q. Yang, H. Q. Wang, G. G. Luo and J. W. Ma, Sparse classification of rotating machinery faults based on compressive sensing strategy, *Mechatronics,* 31 (2015) 22-29.

[42] D. Takhar et al., A compressed sensing camera: New theory and an implementation using digital micromirrors, *Proc. of Computational Imaging IV* (2006).

[43] K. Zhu, X. Lin, K. Li and L. Jiang, Compressive sensing and sparse decomposition in precision machining process monitoring: From theory to applications, *Mechatronics,* 31 (2015) 3-15.

[44] C. Liu, X. Wu, J. Mao and X. Liu, Acoustic emission signal processing for rolling bearing running state assessment using compressive sensing, *Mech. Syst. Signal Process,* 91 (2017) 395-406.

[45] X. Wang, Z. Zhao, Y. Xia and H. Zhang, Compressed sensing for efficient random routing in multi-hop wireless sensor networks, *International Journal of Communication Networks and Distributed Systems,* 7 (3/4) (2011) 275-292.

[46] T. Han, C. Liu, L. J. Wu, S. Sarkar and D. X. Jiang, An adaptive spatiotemporal feature learning approach for fault diagnosis in complex systems, *Mech. Syst. Signal Process,* 117 (2019) 170-187.

[47] C. Wang, M. Gan and C. Zhu, A supervised sparsity-based wavelet feature for bearing fault diagnosis, *Journal of Intelligent Manufacturing,* 30 (2019) 229-239.

[48] Z. Y. He, Y. M. Cai and Q. Q. Qian, Study of wavelet entropy theory and its application in electric power system fault detection, *Proceedings of the Csee,* 5 (2005).

[49] M. Y. Yang and Y. K. Yang, A study of transient-based protection using wavelet energy entropy for power system ehv transmission line, *2010 International Conference on Wavelet Analysis and Pattern Recognition* (2010) 283-288.

[50] S. Li and Z. Liu, Application of improved wavelet packet energy entropy and GA-SVM in rolling bearing fault diagnosis, *2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)* (2018).

[51] H. Ghodrati and A. Hamza, Nonrigid 3D shape retrieval using deep auto-encoders, *Proceedings of the 2017 International Conference on Communications, Signal Processing, and Systems*, 47 (2017) 44-61.

[52] K. Han, Designing extreme learning machine network structure based on tolerance rough set, *International Journal of Intelligent Information Technologies,* 13 (2017) 38-55.

[53] Z. Zhang, Y. Wang and K. Wang, Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network, *Journal of Intelligent Manufacturing,* 24 (6) (2013) 1213-1227.

[54] J. Liu, A. Chen and N. Zhao, An intelligent fault diagnosis method for bogie bearings of metro vehicles based on weighted improved D-S evidence theory, *Energies,* 11 (2018) 232.

[55] C. Hou, Y. Li and Y. Cao, Analysis on vibration and acoustic joint mechanical fault diagnosis of high voltage vacuum circuit based on wavelet packet energy relative entropy, *2016 27th International Symposium on Discharges and Electrical Insulation in Vacuum (ISDEIV)* (2016).

[56] L. Wen, L. Gao and X. Li, A new deep transfer learning based on sparse auto encoder for fault diagnosis, *IEEE Transactions on Systems, Man, and Cybernetics : Systems,* 49 (2019) 136-144.

[57] H. Liu, J. Zhou, Y. Zheng, W. Jiang and Y. Zhang, Fault diagnosis of rolling bearings with recurrent neural network-based auto encoders, *ISA Transactions,* 77 (2018) 167-178.

[58] C. Wang, M. Gan and C. A. Zhu, Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory, *Journal of Intelligent Manufacturing,* 29 (2018) 937-951.

[59] Z. Zhang, Y. Wang and K. Wang, Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network, *Journal of Intelligent Manufacturing,* 24 (6) (2013) 1213-1227.

[60] C. Wang, M. Gan and C. Zhu, Intelligent fault diagnosis of rolling element bearings using sparse wavelet energy based on overcomplete DWT and basis pursuit, *Journal of Intelligent Manufacturing,* 28 (6) (2017) 1377-1391.

[61] Y. Qiu, W. Zhou, N. Yu and P. Du, Denoising sparse autoencoder based ictal EEG classification, *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 99 (2018).

[62] G. Helbing and M. Ritter, Deep learning for fault detection in wind turbines, *Renewable and Sustainable Energy Reviews,* 98 (2018) 189-198.

[63] H. O. A. Ahmed, M. L. D. Wong and A. K. Nandi, Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features, *Mech. Syst. Signal Process,* 99 (2018) 459-477.

[64] J. Sun, C. Yan and J. Wen, Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning, *IEEE Transactions on Instrumentation and Measurement,* 67 (2018) 185-195.

Peiming Shi received Ph.D. degree in information science and engineering institute from Yanshan University, Qinhuangdao, China, in 2009. Now he is a Professor in Institute of Electrical Engineering of Yanshan University. His current research interests include fault diagnosis and signal processing.