



*Research article*

## **A rolling bearing fault detection method based on compressed sensing and a neural network**

**Lu Lu<sup>1</sup>, Jiyou Fei<sup>2</sup>, Ling Yu<sup>3</sup> and Yu Yuan<sup>2,4,\*</sup>**

<sup>1</sup> School of Mechanical Engineering, Dalian Jiaotong University, Dalian 116028, China

<sup>2</sup> College of Locomotive and Rolling Stock Engineering, Dalian Jiaotong University, Dalian 116028, China

<sup>3</sup> China Household Electric Appliance Research Institute, Beijing 100037, China

<sup>4</sup> Traction Power State Key Laboratory of Southwest Jiaotong University, Chengdu 610031, China

\* **Corresponding author:** E-mail: [yuan@djtu.edu.cn](mailto:yuan@djtu.edu.cn).

**Abstract:** The high sampling frequency of traditional Nyquist sampling theory not only puts greater requirements on the sampling equipment, but also generates a large amount of data, which increases the difficulty of information transmission and storage. To this end, this paper proposes a rolling bearing fault signal detection method based on compressed sensing combined with a neural network. Based on the theory of compressed sensing, the observations obtained from compression sampling are divided into two sets of data. Given the one set of data, the predictive ability of the nonlinear time series through the neural network can predict the second set of observed values. The predicted observations are used to reconstruct the signal, thereby reducing the amount of data to be stored and transmitted and realizing secondary compression of the signal.

**Keywords:** compressed sensing; neural network; fault detection; feature extraction

---

### **1. Introduction**

A mechanical vibration signal is inextricably related to its operating state. To obtain the key information contained in the signal, the signal is usually collected by a method following the classic Nyquist sampling theorem [1–3]. Most acquisition chips currently used are also based on the Nyquist

sampling theorem, to uniformly sample signals at equal intervals. The theorem requires that the sampling frequency be at least twice the highest frequency of the signal being tested, and in practical applications this frequency can often be five to 10 times as high [4,5]. The purpose of this high frequency is to avoid signal aliasing and other phenomena, to obtain complete and accurate signal information. However, this method not only requires high sampling equipment, but also generates a large amount of data with greater redundancy, which increases the difficulty of data transmission and analysis [6–8]. To solve the problem of excessive data volumes, people often compress the data, such as the JPEG format in image compression, the MP3 format in audio compression, and the ZIP format in file encoding compression, extracting only important information from the data and discarding a great deal of redundant data [9–11]. Although this method can achieve information compression, the compression process is carried out after the acquisition and requires a complex algorithm to complete, which not only wastes sampled data, but also consumes a large amount of front-end hardware resources. The method thus consists of distortion compression, and the data set itself will be damaged by predictability.

To resolve the problems of the traditional sampling theorem, Donoho et al. [1] formally proposed an entirely new method of signal acquisition and processing based on sparse representation and signal approximation theory, that is, compressed sensing [12–14]. Compressed sensing can compress the original signal while collecting it [15,16]. The sampling frequency is related to the structure of the signal and the information is no longer determined by the signal bandwidth [17–19]. Part of the noise and redundant information in the signal can also be eliminated, which can greatly reduce the amount of data collected and reduce pressure on the acquisition end.

When the theory of compressed sensing was introduced, Donoho and others strictly proved its mathematical correctness. Therefore, in view of its significant advantages, compressed sensing theory has since quickly attracted the attention of many researchers. In terms of practical applications, Rice University has successfully developed a single-pixel compression digital camera [20–22]. The observation process is equivalent to a single-pixel video stream. The camera can take high-quality images with pixel values much lower than that of the original image [23,24]. Self-adaptive ability, however, is not available in traditional equipment. Similar devices include a coded aperture camera developed by the Massachusetts Institute of Technology, a hyperspectral imager developed by Yale University, and a DNA microarray sensor developed by Illinois State University [25–27].

Compressed sensing technology is currently being used in the fields of image and video compression, nuclear magnetic resonance imaging, compressed sensing radar, data communications, and so forth [28–30]. Since certain breakthroughs and given more in-depth research on its theoretical framework, compressed sensing has also been applied in sub-Nyquist sampling systems [31–33] and compressed sensing networks [34–36], among other fields [36,37]. However, there have been fewer applications of compressed sensing in fault diagnosis of mechanical equipment, which is the research topic of this paper. The use of compressed sensing theory to collect and detect mechanical vibration signals can reduce the sampling frequency and number of samples, with a high probability of detecting mechanical failure information, thus reducing pressure in terms of information transmission, storage, and processing [38,39]. In addition, some optimization algorithms are also used to optimize the compressed sensing networks [40–42].

This paper uses compressed sensing technology to compress and collect the fault signals of rolling bearings. Through the analysis and selection of the sparse representation algorithm, measurement matrix, and reconstruction algorithm and the comparative analysis of the impact of each parameter value, the original signal is recovered from a small amount of collected data with high accuracy. This paper also introduces a neural network model into the compressed sensing process, the neural network's predictive ability is used to predict compressed sensing observation values for nonlinear time series to achieve further compression. After testing and selecting the number of neurons and transfer functions for the neural network, it is possible for the entire set of observation values to be accurately predicted, with high probability, under the condition of greatly compressing the observation values. Finally, it can accurately identify fault information bearing signal through the reconstruction of the spectrum.

## 2. Compressed sensing framework

The core idea of compressed sensing theory is the assumption that the signal can be sparsely represented under a transform basis technology. A linear random observation method is used to project the signal onto a low-dimensional space to obtain a small number of observations containing most of the information of the original signal. The nonlinear optimization algorithm then restores the reconstructed signal, which is similar to the original signal.

A real signal  $\mathbf{x} \in \mathbf{R}^{N \times 1}$  can be linearly expanded under a set of  $N$ -dimensional orthogonal bases  $\{\boldsymbol{\psi}_i\}_{i=1}^N$ , and the vector  $\{\boldsymbol{\psi}_i\}_{i=1}^N$  is formed as a column vector to build an  $N \times N$  wike matrix  $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \dots, \boldsymbol{\psi}_N] \in \mathbf{R}^{N \times N}$ . The signal  $\mathbf{x}$  can then be expressed as

$$\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\theta} \quad (1)$$

where the coefficient vector  $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_N] \in \mathbf{R}^{N \times N}$  and  $\boldsymbol{\theta}_i = \langle \mathbf{x}, \boldsymbol{\psi}_i \rangle = \boldsymbol{\psi}_i^T \mathbf{x}$ , which shows that  $\boldsymbol{\theta}$  is the representation of the time domain signal  $\mathbf{x}$  in the  $\boldsymbol{\Psi}$  domain. If there are only  $K$  elements in  $\boldsymbol{\theta}$  that are nonzero or that have large coefficients and  $K \ll N$ , then  $\boldsymbol{\theta}$  is said to be  $K$ -sparse, that is, the signal  $\mathbf{x}$  is a sparse signal under the orthogonal basis  $\boldsymbol{\Psi}$ .

Given an  $M \times N$  ( $M < N$ ) dimensional measurement matrix  $\boldsymbol{\Phi}$  that is not related to the orthogonal basis  $\boldsymbol{\Psi}$ , the dimension reduction is carried out on the signal  $\mathbf{x}$  to obtain the observation vector  $\mathbf{y} \in \mathbf{R}^{M \times 1}$ ; that is,

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} \quad (2)$$

where the observation vector  $\mathbf{y}$  contains the main information of the original signal  $\mathbf{x}$ , and the number of elements in  $\mathbf{y}$  is much lower than in the original signal  $\mathbf{x}$ ; that is, the amount of data is reduced, and the original value can be reconstructed from the observation value  $\mathbf{y}$  by solving linear equations. Because  $M < N$ , the number of unknowns in the system of equations is greater than the number of equations. This is an underdetermined problem, but, according to the above, the signal  $\mathbf{x}$  can be expressed sparsely, so the observation process can be expressed as

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta} \quad (3)$$

By reconstructing  $\boldsymbol{\theta}$ , the original signal  $\mathbf{x}$  can be reconstructed indirectly. Although it is an underdetermined problem, because  $\boldsymbol{\theta}$  is sparse, it greatly reduces the number of unknowns and

reduces the size of the understanding space, making it possible to reconstruct the original signal. The following optimization problems need to be solved:

$$\theta^{(1)} = \arg \min \|\theta\|_1 \quad \text{s.t.} \quad \mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \theta \quad (4)$$

According to the process of compressed sensing, the core problems can be summarized into the following three points:

(1) Design an appropriate measurement matrix  $\Phi$  to observe the original signal  $\mathbf{x}$  to obtain the measurement value  $\mathbf{y} = \Phi \mathbf{x}$ .

(2) According to a priori knowledge, select the sparse basis  $\Psi$  that enables the original signal  $\mathbf{x}$  to be represented sparsely under the basis.

(3) Use an appropriate reconstruction algorithm to find the optimal solution  $\hat{\theta}$ , and finally calculate the reconstructed signal  $\hat{\mathbf{x}} = \Psi \hat{\theta}$ .

The sparsity of the signal is an important factor that affects the efficiency and accuracy of compressed sensing. The traditional orthogonal base dictionary sometimes cannot guarantee the signal's sufficient sparsity. Therefore, the method of using the optimized learning algorithm to construct the redundant dictionary has received more attention. This paper uses the most representative K-SVD algorithm, widely used because of its relatively good sparse decomposition effect on various signals [43,44].

When applying the K-SVD algorithm, it is necessary to determine the original signal  $\mathbf{x}$ , the initial dictionary atom length  $n$ , the number of atoms  $K$ , the sample set atom number  $N$ , the linearly combined atom number  $L$  in the sparse representation, and the number of iterations  $J$ . The proper parameter value combination in the K-SVD algorithm is the key to ensuring that the sparse coefficient has sufficient sparsity and attenuation. However, there is currently no uniform selection standard, which is usually selected based on the experience of researchers or repeated experiments.

This paper uses faulty rolling bearing data collected by the Case Western Reserve University Bearing Data Center for the original signal, to test combinations of various parameter values in the K-SVD algorithm. The original data were taken from 6205-2RS JEM SKF deep groove ball bearings. Inner ring fault signals were analyzed. The sampling frequency was 12 kHz, the motor speed was 1797 rpm, and the fault point was the inner ring score.

The atomic length  $n$  is chosen to be 512, because the parameter values to be determined are not unique and are related to each other. This study uses single-factor analysis to pre-set the initial values of each parameter. Each round of the test fixes three parameter values and adjusts a fourth. The order of the parameters adjustments is  $K$ ,  $N$ ,  $L$ , and  $J$ .

From experience, when parameter  $K$  is tested, the initial values of the other parameters are set to  $N = 1100$ ,  $L = 10$ , and  $J = 10$ . To ensure a complete dictionary, the value range of  $K$  in the test is 600–950. The test results are shown in Table 1.

**Table 1.** Reconstruction errors for different values of  $K$ .

Value	700	750	800	850	900	950
Relative Error	0.432	0.425	0.425	0.417	0.412	0.438

It is not difficult that the relative error of reconstruction has a tendency to decrease with the increase of  $K$ . In this paper, there is  $K = 900$ .

Table 2 is a comparison chart of the reconstruction errors for different values of  $N$ . In the test,  $K$  takes the value of 900, selected in the previous test, and the values of  $L$  and  $J$ , both equal to 10, remain unchanged. Because of the limitation on the value of  $K$ , the dictionary needs to guarantee sufficient training. Therefore,  $N = 950, 1000, \dots, 1200$  is used for testing.

**Table 2.** Reconstruction errors for different values of  $N$ .

Value	950	1000	1050	1100	1150	1200
Relative Error	0.424	0.418	0.414	0.413	0.413	0.414

Through the test results, one can find that the relative error of reconstruction is relatively stable, and the reconstruction accuracy when  $N = 1100$  is slightly better than for the other values.

The value of the parameter  $L$  is generally small. In this paper,  $L = 2, 4, \dots, 12$  is used for the test. The test results are shown in Table 3.

**Table 3.** Reconstruction errors for different values of  $L$ .

Value	2	4	6	8	10	12
Relative Error	0.425	0.447	0.509	0.409	0.418	0.422

It is not difficult to see that, when the value of  $L$  is small, the error fluctuation is large, so this paper uses  $L = 8$  to ensure that the error is stable.

The number of iterations  $J$  is similar to the parameter  $L$ . The values in the relevant literature are usually around eight. In this paper,  $J = 2, 4, \dots, 12$  is used for testing. The test results are shown in Table 4.

**Table 4.** Reconstruction errors for different values of  $J$ .

Value	2	4	6	8	10	12
Relative Error	0.421	0.408	0.401	0.404	0.408	0.413

To reduce the dictionary training time, the number of iterations  $J$  also needs to take on a smaller value. According to the reconstruction error curve,  $J = 6$  is more suitable.

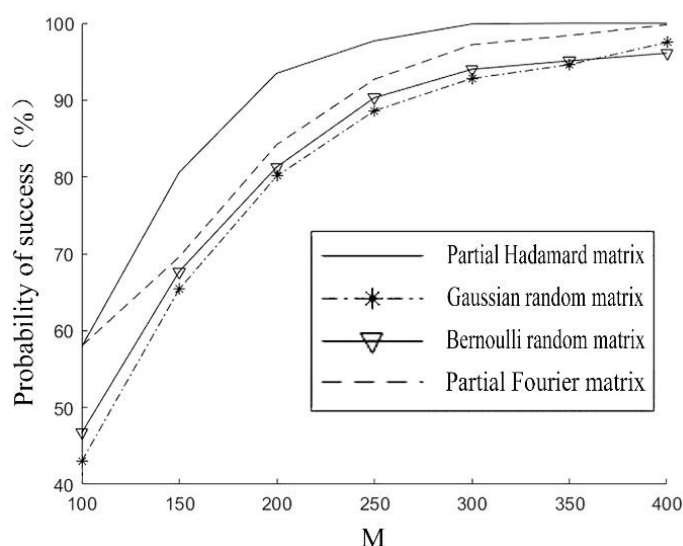
In the follow-up experiments conducted, the values of the parameters of the  $K$ -SVD algorithm are given.

The measurement matrix is the key to the dimensionality reduction projection of the signal, and it is divided into a random matrix and a deterministic matrix. The former matrix is used in this paper. Common random measurement matrices used currently include Gaussian random matrices, Bernoulli random matrices, and partial Hadamard matrices. The performance of each matrix is different, depending on the signals of different properties. The performance of each measurement matrix in

different applications is quite different, with no uniform selection standard. Therefore, this paper compares and analyzes the reconstruction performance of the above-mentioned common random measurement matrices under the conditions.

The test selects 512 continuous data observations of the vibration signal of a faulty rolling bearing as the original signal. It uses each measurement matrix to perform compression sensing and then compares the accuracy of each set of fault information detected by the matrix. In other words, the frequency corresponding to the highest spectral peak of the reconstructed signal spectrum is required to be consistent with the original signal fault frequency, and the amplitude ratio of the secondary spectral peak to the highest spectral peak of the reconstructed signal is less than 0.9.

The research shows that the frequency characteristics corresponding to each harmonic component in the Fourier domain appear as two spectral lines. Therefore, to facilitate the experiment, the signal sparsity  $K = 2$  is used, which is equivalent to detecting only one harmonic component at a time. Under different values for the number of data  $M$ , each measurement matrix performs 1000 compressed sensing processes to calculate the probability of the conditions being satisfied. The test results are shown in Figure 1.



**Figure 1.** Measurement matrix failure detection probability.

As can be seen from Figure 1, the success rate of the failure information extraction of the partial Hadamard matrix is significantly higher than that of the other matrices. Therefore, this paper selects the partial Hadamard matrix as the measurement matrix of compressed sensing.

The essence of signal reconstruction is to solve the optimization problem of equation (4) and find a solution that is as sparse as possible in the equation. At present, there are two main types of methods, namely, the base tracking method and the greedy algorithm. The latter is used for analysis in this paper. According to the principle of compressed sensing, for a signal with a large dimension, the amount of computation for the reconstructed signal to solve the L1 norm minimization problem is large, and the computation time required will increase. The iterative method using the greedy algorithm can reduce computational complexity, can shorten the signal recovery time, and is easy to implement in hardware.

The most commonly used iterative method is the orthogonal matching pursuit algorithm [45–47]. The subspace pursuit algorithm selected in this paper is an improvement over the traditional orthogonal matching pursuit algorithm in terms of efficiency and stability [48–50].

This paper adds neural network prediction theory to the compressed sensing process. First, the signal is collected by the compressed sensing method to obtain a small amount of observation values, which are divided into several consecutive data groups  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$ , where  $\mathbf{X}_i = [x_i, x_{i+1}, \dots, x_{i+m-1}]$ . Then,  $\mathbf{X}_1$  is used as the learning sample of the neural network, and the next time data corresponding to each group of samples are used as the expected output, that is, the expected output value corresponds to the input signal. For example, the expected output value corresponding to the input signal  $\mathbf{X}_1$  is  $x_{m+1}$ . The neural network is thus trained, until the trained network is obtained. Then the observation value is divided into two parts according to the set ratio, and only the first part of the data is transmitted and stored, which is equivalent to greatly reducing the amount of information data transmitted and stored. When the signal needs to be recovered, the last  $m$  data observations of the previous observations stored are used as input, which ensures that the number of input data points is equal to the number of elements of  $\mathbf{X}_i$ . The predicted value at the next moment can then be obtained through the trained neural network. According to the principle of rolling prediction, after repeated iterations, the predicted values of all later observations can be obtained. Finally, the standard first part of the observation values and the predicted subsequent observation values are combined to obtain the complete observation values, and the original signal can be restored by using the reconstruction algorithm.

Although the application of multilayer perception neural network has been quite extensive, there are still no unified guidelines or established formulas for the specific design of its structure in different fields, which is usually selected based on the designer's experience or the experiment.

The total number of layers of the neural network mainly depends on the number of hidden layers, which is generally one to two. In this paper, the predicted value of compressed sensing is input into the network as the original signal. Because of the high level of randomness of the signal, the use of multiple hidden layers is suitable to ensure the accuracy of the prediction. Therefore, this paper uses a double hidden layer structure.

The transfer function, an important part of the neural network, is a continuous differentiable function. Its main function is to increase the nonlinearity of the network. In this paper, the tan-sigmoid function is selected as the transfer function between the input layer and the first hidden layer, and between the first and second hidden layers. The transfer function between the second hidden layer and the output layer uses a linear transfer function.

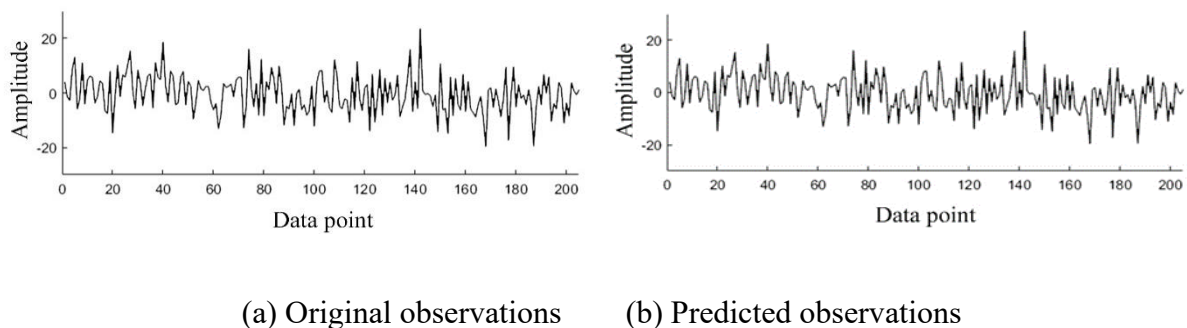
The training algorithm of the neural network is a method for solving the local minimum. The more common algorithms include the gradient descent method, the conjugate gradient method, the Newton method, and the Levenberg–Marquardt method. In this paper, after several experiments using the above training algorithms, the network training is found to reach the upper limit of 20,000 iterations when using the gradient descent or the conjugate gradient method, but it is still difficult for the training error to satisfy the requirements. When the Newton method is used, although the stopping condition can be satisfied after several thousand iterations, the training error is still large and does not meet the error requirements in practical applications. Therefore, this paper uses the Levenberg–Marquardt algorithm, with greater speed and higher accuracy for training the network.

The specific values of the neural network parameters are given later.

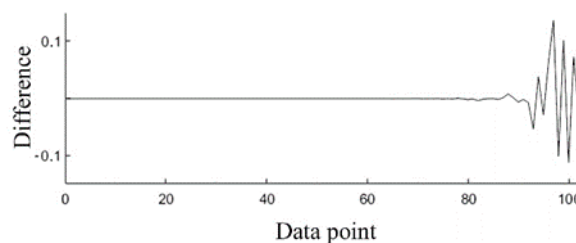
### 3. Simulation test

This article uses the bearing signal described earlier as the original signal. Given the number of original data  $n = 512$ , the sparseness of the signal is 110, and the K-SVD algorithm is used to sparsely represent the signal. Then, a partial Hadamard matrix is used to compress and observe the signal. The test uses the compressed data number  $M = 205$  as an example. That is, the compressed data are compressed by about 60%. Then, the observations obtained are input into the neural network for learning. This experiment intends to compress the observed value by about 50%, that is, the first 103 data are fixed values, and the last 102 data are predicted. After testing, the numbers of neurons in the input layer, first and second hidden layers, and output layer of the neural network are 14, 11, 5, and 1, respectively. This combination of numbers can guarantee a higher prediction success rate with fewer neurons. Figure 2 shows the results of a comparison between the predicted and actual observed values. The first 103 data points in panels (a) and (b) are exactly same and are fixed values. The 104th to 205th data points in panel (b) are the predicted values for the corresponding data in panel (a). The error between the predicted and actual values is shown in Figure 3.

It can be seen error of the first 70 data points of the prediction is almost zero, and the prediction error of the subsequent data increases gradually with the rolling iteration process. The error fluctuation of the 90th to 102th data points is more obvious. Nevertheless, the overall forecast error is still small compared to the observed amplitude, which is basically maintained within a range of  $\pm 0.15$ , accounting for only about 0.4% of the total amplitude. It is difficult to distinguish the difference between the two from Figure 2.



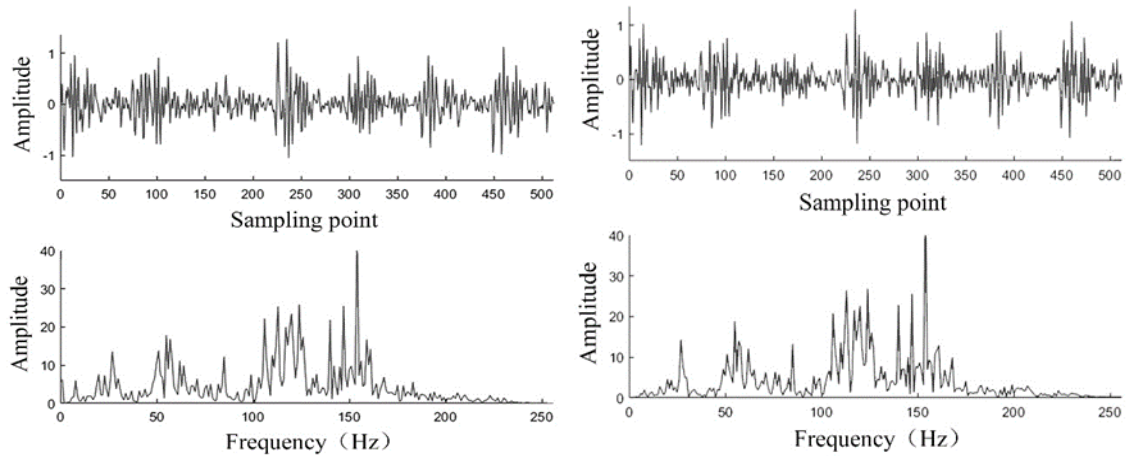
**Figure 2.** Original and predicted values.



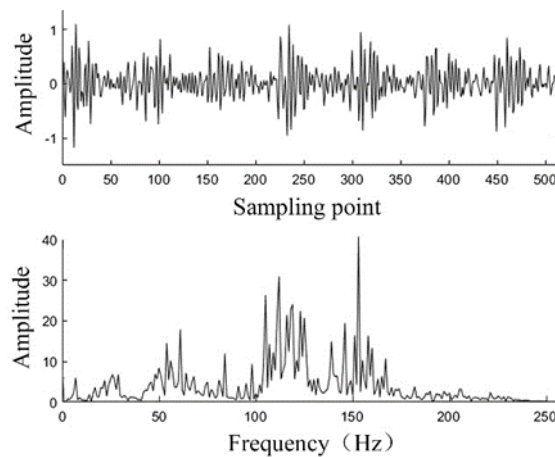
**Figure 3.** Predicted value error curve.



Finally, all the predicted observations are reconstructed using the subspace pursuit algorithm, and MATLAB software is used to generate the reconstructed signals and their spectrograms, as shown in Figure 4.



(a) Original signal and its spectrum (b) Reconstructed signal and its spectrum



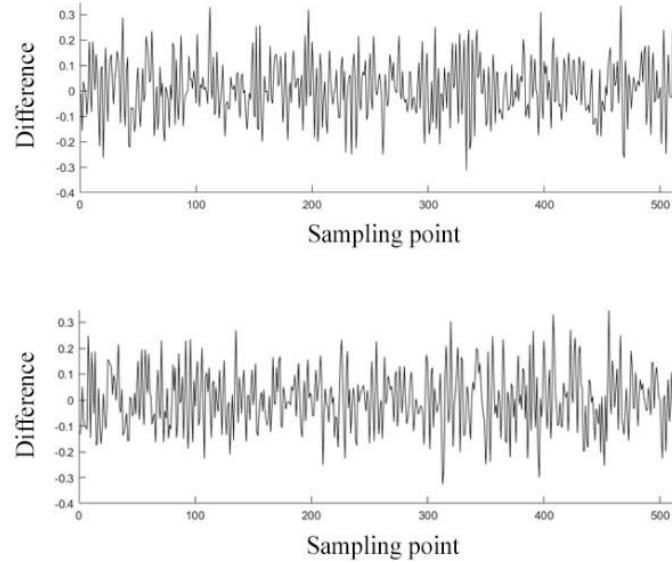
(b) Reconstructed signal and its spectrum for the traditional method

**Figure 4.** Original and reconstructed signals of the inner ring fault.

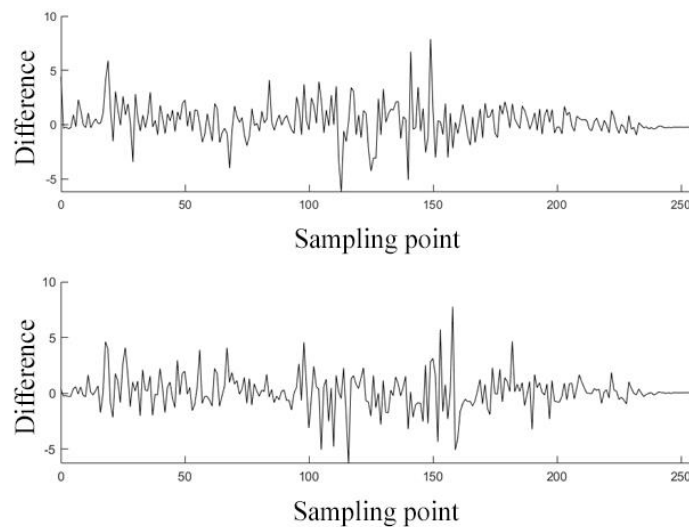
It is not difficult to determine that the reconstructed signal obtained with the method in this paper can simulate the original signal well, with only a certain error at some sampling points. The error for the traditional method is more obvious, for example, the signal near the 160th sampling point is seriously distorted. Although, given the experimental compression ratio, the proposed method can accurately detect the fault frequency of the bearing about 154 Hz through their respective spectra, the traditional method is already exhibiting obvious errors.

The error curve in Figure 5 is the difference between the original signal and its frequency spectrum. In panels (a) and (b), the upper curve is the result of the comparison of the method in this

paper, and the lower curve is the result of the traditional method. Although the signal error is not obvious, the spectral error comparison shows that the reconstruction error of this method is slightly better than that of the traditional method.



(a) Error curve between the reconstructed and original signals of each method



(b) Error curve between the reconstructed and original spectra of each method

**Figure 5.** Error curves of the signal and spectrum of each method.

To quantitatively compare the reconstruction accuracy of the neural network prediction method used in this paper with the traditional method, the concept of a matching rate (MR) is introduced as a measure of the reconstruction accuracy of this paper. The calculation formula of the MR is

$$MR = 1 - \frac{\|\hat{x} - x\|_2}{\|\hat{x} + x\|_2} \quad (5)$$

where  $x$  represents the original signal and  $\hat{x}$  represents the reconstructed signal. The value of the MR is between zero and one, where the closer the value is to one, the better the reconstruction effect, and the closer to zero, the greater the reconstruction error. This method bypasses the disadvantage of the traditional error measurement standard, where the error size cannot be determined because there is no fixed value range. According to the test, for the application instances in this paper, as long as  $MR \geq 0.6$  is ensured, the fault information can be accurately identified through the reconstructed signal spectrum.

Table 5 presents the average values of the reconstructed signal MR for the prediction and traditional methods for multiple experiments under different compression amounts.

**Table 5.** MR values for the prediction and traditional methods for different compression ratios.

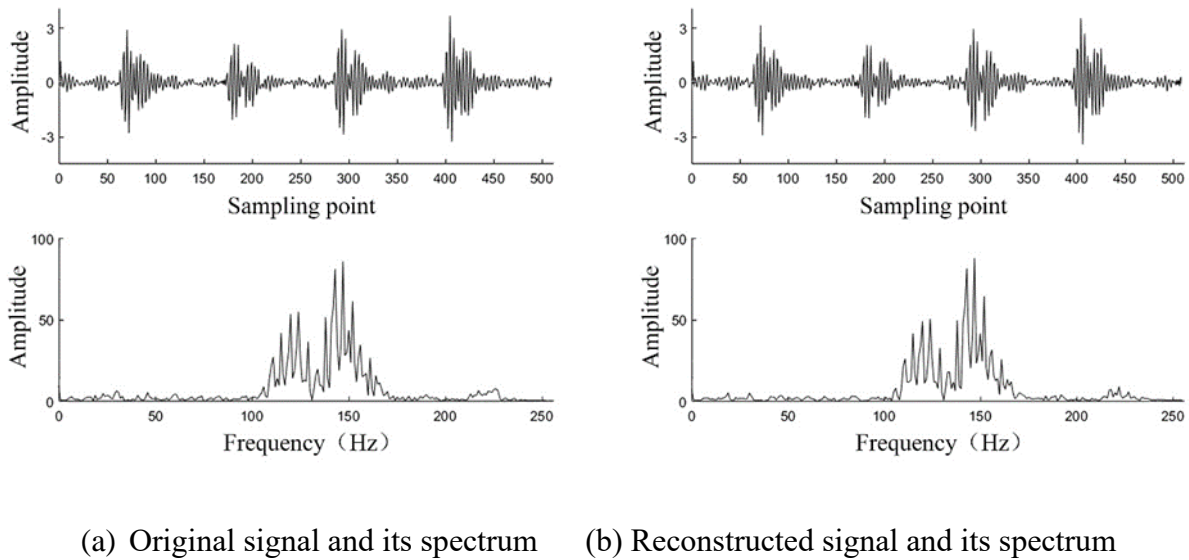
Compression ratio	MR of our method	MR of the traditional K-SVD method
0.72	0.8284	0.8111
0.76	0.8318	0.7943
0.80	0.8277	0.7822
0.84	0.8331	0.7695
0.88	0.8296	0.7349
0.92	0.8315	0.7344

It can be seen from Table 5 that, when the total compression is 72–92%, the MR of the prediction method in this paper is significantly higher than that of the traditional method, which is basically stable at around 0.83. The MR of the traditional method decreases significantly with the increase in compression. At a compression of 92%, the gap between the traditional and prediction methods is quite obvious. In the experiment, when using the traditional method to compress the signal by a more compression amount, although there is only a difference of 0.1 between the final calculated match value and the predicted value, the traditional method is found to be unable to guarantee the accuracy of the extraction of the bearing fault information performance, and the success rate of the compressed sensing process is significantly reduced. At 92% compression, the success rate is lower than 50%.

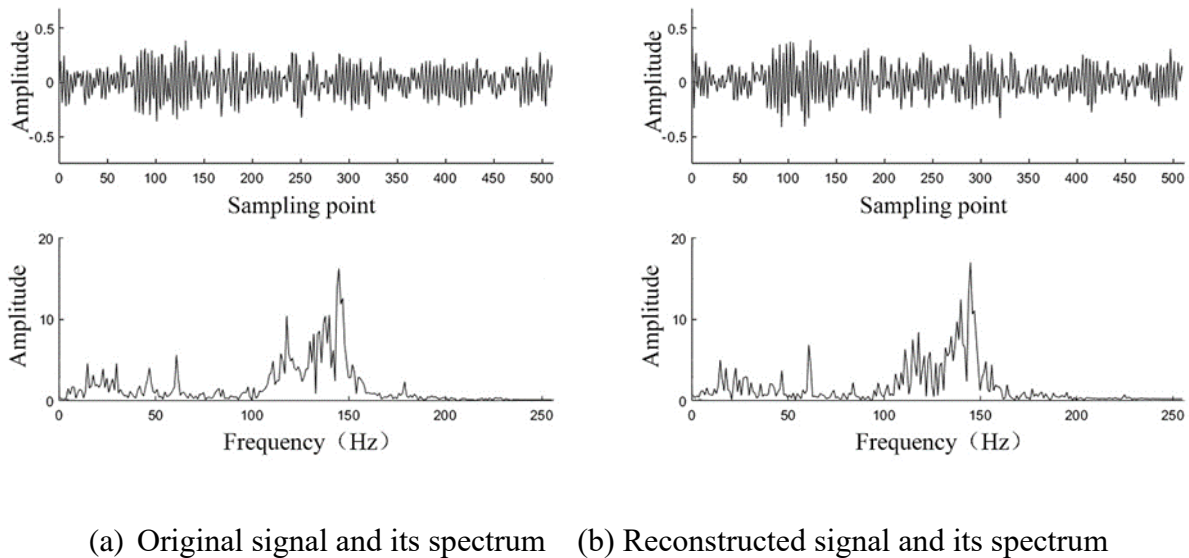
In addition to inner ring failure, the bearing can undergo outer ring failure and rolling element failure. To verify the applicability of this method to the bearing's outer ring and rolling element fault signals, the original signal is replaced by, first, the outer ring vibration signal and, then, the rolling element vibration signal, to test separately. The number of original signal data points  $n = 512$ , using the method described above to perform compression observation, observation value prediction, and signal reconstruction. The signal reconstruction results are shown in Figures 6 and 7.

Figure 6 shows that the reconstructed signal can simulate the original signal well. After calculation, the MR of the signal at this time is about 0.91. In Figure 7, the reconstructed rolling element fault signal contains significant error. The MR of the reconstructed signal at this time is about 0.75. Although there is a big difference between the original and reconstructed signals, the change

trend of the two is basically the same. The amplitude and position of the main spectrum peak in the spectrum are basically the same, which can be used to identify the failure frequency of the bearing.



**Figure 6.** Original and reconstructed signals of outer ring fault.



**Figure 7.** Original and reconstructed signals of roller fault.

#### 4. Performance analysis of the compressed sensing of a damaged signal

In practical applications, the signal is often disturbed due to factors such as the environment and equipment. Among these disturbances, the noise interference at the acquisition end, which refers to the noise contained in the original vibration signal, which have affect to a certain extent on the observation values that need to be predicted through the compressed sensing process. Neural network predictions have higher signal structure requirements. To verify the performance of our signal processing method

in this paper when the original signal is disturbed by noise, noise is artificially added to the original signal for testing, and its impact on the original signal noise is analyzed by comparing the reconstructed signal spectrum and the MR. The level of noise is measured by the signal-to-noise ratio (SNR).

The original signal without noise interference used in the experiment is the same as above, but Gaussian white noise is added to the original signal before compressive sensing sampling. In this paper, the same experiment is carried out for a different SNR. At the same time, the MRs of each reconstructed signal with a noise-free interference signal and a noise-containing signal are respectively calculated. The results are shown in Table 6. In the experiment, the MR corresponding to each SNR value is the average value of the MR of the reconstructed signal under the condition that the neural network successfully predicts it five times.

**Table 6.** Reconstructed signal MRs under different SNRs.

SNR (dB)	MR for a pure signal	MR for a noisy signal
50	0.833	0.834
30	0.834	0.834
25	0.809	0.806
20	0.763	0.761
15	0.682	0.677
10	0.592	0.600
8	0.514	0.587
5	0.475	0.559

It is not difficult to see that, when the SNR is at least 30 dB, the MR tends to be stable, and a stage is reached where no noise interference is added in the previous test, indicating that the noise interference on the signal is basically negligible at that time. When the SNR is no more than 25 dB, the signal MR decreases significantly as the SNR decreases. When the SNR equals 5 dB, the reconstructed signal has difficulty simulating the original signal, and the fault frequency cannot be accurately judged by its frequency spectrum. In addition, when the SNR is at least 15 dB, the MR of the reconstructed signal relative to the undisturbed original signal is slightly greater than that of the noisy signal. The reason for this phenomenon is that the compressed sensing process itself has a certain noise reduction capability. If the signal structure contains less interference, weak noise interference can be eliminated through its own calculation process.

In addition to the original signal being disturbed by noise, the observations can also be disturbed. A major feature of compressed sensing is that a small amount of observations after compression are used for the transmission of information, and the noise interference of observations is the impact of noise during the transmission of analog signals.

The test method involves mixing Gaussian white noise into the observation value before the observation value is input into the neural network for training. Then, the neural network is trained using observations with noise, and the latter part of the data with noise observations is predicted by the neural network. Finally, the predicted observations are used to reconstruct the signal to obtain fault frequency information through the spectrum. In this paper, the same experiment is conducted under different SNRs, and the MR of the reconstructed signal is calculated at the same time. The results are

shown in Table 7. In the experiment, the MRs corresponding to each value of the SNR are the average MR values in the case of five successful predictions by the neural network.

**Table 7.** Reconstructed signal MRs under different SNRs.

SNR (dB)	MR
60	0.838
55	0.835
50	0.825
45	0.785
40	0.735
35	0.600
30	0.514

It can be seen from Table 7 that, when the SNR is no more than 50 dB, as the SNR decreases, the signal MR decreases significantly. When the SNR is at least 50 dB, the signal MR tends to be stable and is close to the matching degree without noise interference in the previous section.

The experiment finds that, when the observed SNR is less than 35 dB, the neural network prediction rate is significantly reduced, the reconstructed signal cannot accurately simulate the original signal, and it is difficult to judge the fault frequency through its spectrum. The test results indicate that, when the compressed sensing observations are disturbed by noise, the original signal must be reconstructed with higher accuracy by ensuring that the SNR is no greater than 35 dB in this method.

In addition to noise interference, data loss is also an inevitable phenomenon. To simulate data loss, the loss mechanism before processing the data must be analyzed. At present, the loss mechanism is usually divided into three: nonrandom loss, random loss, and complete random loss. However, there is no good analysis and processing method for the nonrandom loss or complete random loss of data. Therefore, this study simulates the random loss of data.

In the compressed sensing process, data loss can be split into data loss at the collecting end and data loss at the observation value. However, because a neural network is needed to predict the sequence of observations, this method has strict requirements on the integrity of the sequence data. If data are lost from the observations, it is difficult to for the neural network to make accurate predictions. Therefore, in the analysis of the impact of data loss, only the loss of original data at the acquisition end is considered.

The original signal with no data loss that is selected for the experiment is the same as previously, and, based on this, part of the data are randomly set to zero to simulate data loss. In this paper, the same experiment is conducted for different data loss ratios. In the experiment, the signal with no data loss is segmented into groups of 9 to 2 data. In each data segment, one data point is randomly selected set to zero. The corresponding data loss ratio is about 11.1–50.0%. Table 8 shows the MRs of the reconstructed signal with the data loss ratios to the original signal with data loss when the other parameter conditions remain unchanged. The MR is the average value of 10 successful predictions by the neural network.

**Table 8.** Reconstructed signal MRs for different data loss ratios.

Loss ratio (%)	MR
11.1	0.732
12.5	0.717
14.3	0.712
16.7	0.694
20.0	0.681
25.0	0.638
33.3	0.603
50.0	0.523

It can be seen from Table 8 that, in the case of loss of data from the original signal, the MR of the reconstructed signal is significantly lower than in the case of no data loss in the previous experiment. As the proportion of data loss increases, the MR of the reconstructed signal gradually decreases. The experiment finds that the loss of data randomness destroys the structure of the signal to a certain extent, so the information contained in the observations obtained through compressed sensing is disturbed and, eventually, the neural network prediction rate under the same conditions is significantly reduced. In addition, when the proportion of data loss is high, the reconstructed signal has difficulty simulating the original signal, and its frequency spectrum will occasionally appear, since the original fault frequency spectrum peak is lost according to the actual data loss.

From the results, one can conclude that, in the case of random loss of the original data, as long as the loss ratio is less than 30%, the reconstructed signal obtained by the method in this paper is guaranteed to simulate the original signal with greater accuracy. The bearing's frequency can be used to accurately determine its fault frequency.

## 5. Conclusion

This paper proposes a method for extracting rolling bearing fault information by combining compressed sensing and a neural network. Four combinations of the main numerical parameters of the K-SVD algorithm are selected for use in sparse signal representation. The reconstructed performances of several common measurement matrices and greedy algorithms are compared and analyzed, and the measurement matrix and signal reconstruction algorithms are suitable for the signal feature extraction method are selected. A neural network is used to predict the latter part of the observation data obtained by compression sampling. Further compression of the observation value is thereby achieved, and the original signal is reconstructed from the prediction value. The feasibility of the method is verified by simulation experiments. For the reconstruction of the bearing inner ring, outer ring, and rolling element signals, the MRs can be guaranteed to be 0.83, 0.77, and 0.91, respectively. This method is obviously superior to the traditional method. Additionally, considering that the signal will inevitably be disturbed in practical applications, two situations are simulated in which the signal is disturbed by noise and data are randomly lost. The minimum SNR and the maximum data loss ratio to ensure the accuracy of the

extracted fault information are obtained: the original signal's SNR must be at least 10 dB, the observed SNR value must be at least 35 dB, and the original signal data random loss rate must be less than 30%.

## Acknowledgments

This work was supported in part by the Natural Science Foundation of Liaoning Province under Grants 2019ZD0112 and 2019ZD0099, and in part by the National Natural Science Foundation of China under grants 51475065, 51605068, and 51879027; the Traction Power State Key Laboratory of Southwest Jiaotong University under grant TPL2002; and the Liaoning BaiQianWan Talents Program.

## Conflict of interest

The authors declare no conflicts of interest.

## References

1. D. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory*, **52** (2006), 1289–1306.
2. H. Zhao, H. Liu, J. Xu, W. Deng, Performance prediction using high-order differential mathematical morphology gradient spectrum entropy and extreme learning machine, *IEEE Trans. Instrum. Meas.*, **69** (2020), 4165–4172.
3. T. Li, J. Shi, X. Li, J. Wu, F. Pan, Image encryption based on pixel-level diffusion with dynamic filtering and DNA-level permutation with 3D Latin cubes, *Entropy*, **21** (2019), 1–21.
4. M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, et al, Single-pixel imaging via compressive sampling, *IEEE Signal Process. Mag.*, **25** (2008), 83–91.
5. Y. Xu, H. Chen, J. Luo, Q. Zhang, S. Jiao, X. Zhang, Enhanced Moth-flame optimizer with mutation strategy for global optimization, *Inf. Sci.*, **492** (2019), 181–203.
6. H. Zhao, J. Zheng, J. Xu, W. Deng, Fault diagnosis method based on principal component analysis and broad learning system, *IEEE Access*, **7** (2019), 99263–99272.
7. E. Candes, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory*, **52**, (2006), 489–509.
8. D. Takhar, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. Kelly, et al., A new compressive imaging camera architecture using optical domain compression, *Proc. SPIE*, (2006), 6065.
9. R. Chen, S. K. Guo, X. Z. Wang, T. Zhang, Fusion of multi-RSMOTE with fuzzy integral to classify bug reports with an imbalanced distribution, *IEEE Trans. Fuzzy Syst.*, **27** (2019), 2406–2420.
10. H. Zhao, S. Zuo, M. Hou, W. Liu, L. Yu, A novel adaptive signal processing method based on enhanced empirical wavelet transform technology, *Sensors*, **18** (2018), 1–17.
11. Y. Liu, X. Wang, Z. Zhai, R. Chen, Y. Jiang, Timely daily activity recognition from headmost sensor events, *ISA Trans.*, **94** (2019), 379–390.
12. H. Zhao, J. Zheng, W. Deng, Y. Song, Semi-supervised broad learning system based on manifold regularization and broad network, *IEEE Trans. Circuits Syst. I*, **67** (2020), 983–994.
13. T. Li, Z. Qian, T. He, Short-term load forecasting with improved CEEMDAN and GWO-based multiple kernel ELM, *Complexity*, **2020** (2020), 1–20.



14. J. Yu, Y. Xu, G. Yu, L. Liu, Fault severity identification of roller bearings using flow graph and Non-naive Bayesian inference, *P. I. Mech. Eng. C-J. Mec.*, **233** (2019), 5161–5171.
15. Z. Wang, Z. Ji, X. Wang T., Wu, W. Huang, A new parallel DNA algorithm to solve the task scheduling problem based on inspired computational model, *BioSystems*, **162** (2017), 59–65.
16. H. Zhao, D. Li, W. Deng, Research on vibration suppression method of alternating current motor based on fractional order control strategy, *Proc. Inst. Mech. Eng. E J. Process*, **231** (2017), 786–799.
17. M. Lustig, D. Donoho, L. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, *Magn. Reson. Med.*, **58** (2007), 1182–1195.
18. R. Baraniuk, P. Steeghs, Compressive radar imaging, *IEEE Radar Conf.*, (2007), 128–133.
19. W. Deng, H. Liu, J. Xu, H. M. Zaho, An improved quantum-inspired differential evolution algorithm for deep belief network, *IEEE Trans. Instrum. Meas.*, **2020** (2020), 1–8.
20. Z. Ji, Z. Wang, X. Deng, W. Huang, T. Wu, A new parallel algorithm to solve one classic water resources optimal allocation problem based on inspired computational model, *Desalin. Water Treat.*, **160** (2019), 214–218.
21. W. Deng, J. Xu, Y. Song, H. M. Zaho, An effective improved co-evolution ant colony optimization algorithm with multi-strategies and its application, *Int. J. Bio-Inspired Comput.*, **2020** (2020), 1–10.
22. H. Shao, J. Cheng, H. Jiang, Y. Yang, Z. Wu, Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing, *Knowledge-Based Syst.*, **188** (2020), 1–14.
23. H. Chen, Q. Zhang, J. Luo, Y. Xu, X. Zhang, An enhanced bacterial foraging optimization and its application for training kernel extreme learning machine, *Appl. Soft Comput.*, **86** (2020), 1–24
24. I. Maravic, M. Vetterli, K. Ramchandran, High resolution acquisition methods for wideband communication systems, *Proc. Acoust. Speech Signal Process.*, **4** (2003), IV-133-6.
25. K. Gedalyahu, Y. Eldar, Time-delay estimation from low-rate samples: A union of subspaces approach, *IEEE Trans. Signal Process.*, **58** (2010), 3017–3031.
26. M. Davenport, C. Hegde, M. Duarte, R. Baraniuk, Joint manifolds for data fusion, *IEEE Trans. Image Process.*, **19** (2010), 2580–2594.
27. W. Deng, J. Xu, H. Zhao, An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem, *IEEE Access*, **7** (2019), 20281–20292.
28. Y. Liu, Y. Mu, K. Chen, Y. Li, J. Guo, Daily activity feature selection in smart homes based on Pearson correlation coefficient, *Neural Process. Lett.*, **51** (2020), 1771–1787.
29. Z. He, H. Shao, X. Zhang, J. Cheng, Y. Yang, Improved deep transfer auto-encoder for fault diagnosis of gearbox under variable working conditions with small training samples, *IEEE Access*, **7** (2019), 115368–115377.
30. J. Luo, H. Chen, A. A. Heidari, Y. Xu, Q. Zhang, C. Li, Multi-strategy boosted mutative whale-inspired optimization approaches, *Appl. Math. Model.*, **73** (2019), 109–123.
31. H. Chen, F. Miao, X. Shen, Hyperspectral remote sensing image classification with CNN based on quantum genetic-optimized sparse representation, *IEEE Access*, **8** (2020), 99900–99909.
32. J. Yu, M. Bai, G. Wang, X. Shi, Fault diagnosis of planetary gearbox with incomplete information using assignment reduction and flexible naive Bayesian classifier, *J. Mech. Sci. Technol.*, **32** (2018), 37–47.

33. F. R. Sun, Y. D. Yao, G. Z. Li, W. Liu, Simulation of real gas mixture transport through aqueous nanopores during the depressurization process considering stress sensitivity, *J. Petrol. Sci. Eng.*, **178** (2019), 829–837.
34. W. Deng, H. M. Zhao, L. Zou, G. Y. Li, X. H. Yang, D. Q. Wu, A novel collaborative optimization algorithm in solving complex optimization problems, *Soft Comput.*, **21** (2017), 4387–4398.
35. J. Zheng, Y. Yuan, L. Zou, W. Deng, C. Guo, H. Zhao, Study on a novel fault diagnosis method based on VMD and BLM, *Symmetry*, **11** (2019), 747.
36. W. Deng, W. Li, X. Yang, A novel hybrid optimization algorithm of computational intelligence techniques for highway passenger volume prediction, *Expert Syst. Appl.*, **38** (2011), 4198–4205.
37. J. Yu, Y. He, Planetary gearbox fault diagnosis based on data-driven valued characteristic multigranulation model with incomplete diagnostic information, *J. Sound Vib.*, **429** (2018), 63–77.
38. H. L. Fu, M. M. Wang, P. Li, S. Jiang, M. Cao, Tracing knowledge development trajectories of the internet of things domain: A main path analysis, *IEEE T. Ind. Inform.*, **15** (2019), 6531–6540.
39. W. Deng, R. Yao, H.M. Zhao, X. H. Yang, G. Y. Li, A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm, *Soft Computing*, **23** (2019), 2445–2462.
40. Z. Wang, X. Ren, Z. Ji, W. Huang, T. Wu, A novel bio-heuristic computing algorithm to solve the capacitated vehicle routing problem based on Adleman–Lipton model, *Biosystems*, **184** (2019), 103997.
41. H. L. Fu, M. M. Wang, P. Li, S. Jiang, M. Cao, Tracing knowledge development trajectories of the internet of things domain: A main path analysis, *IEEE T. Ind. Inform.*, **15** (2019), 6531–6540.
42. J. Yu, Y. Xu, K. Liu, Planetary gear fault diagnosis using stacked denoising autoencoder and gated recurrent unit neural network under noisy environment and time-varying rotational speed conditions, *Meas. Sci. Technol.*, **30** (2019), 095003.
43. M. Aharon, M. Elad, A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, signal processing, *IEEE Trans. Inf. Theory*, **54** (2006), 4311–4322.
44. M. Davenport, M. Wakin, Analysis of orthogonal matching pursuit using the restricted isometry property, *IEEE Trans. Inf. Theory*, **56** (2011), 4395–4401.
45. S. Lu, Q. He, J. Wang, A review of stochastic resonance in rotating machine fault detection, *Mech. Syst. Signal Pr.*, **116** (2019), 230–260.
46. W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, *IEEE Trans. Inf. Theory*, **55** (2009), 2230–2249.
47. H. shao, M. Xia, G. Han, Y. Zhang, J. Wan, Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer CNN and thermal images, *IEEE T Ind. Inform.*, 2020, Doi: 10.1109/TII.2020.3005965.
48. J. Zheng, Z. Dong, H. Pan, Q. Ni, J. Zhang, Composite multi-scale weighted permutation entropy and extreme learning machine based intelligent fault diagnosis for rolling bearing, *Measurement*, **143** (2019), 69–80.
49. S. Lu, P. Zheng, Y. Liu, Z. Cao, H. Yang, Q. Wang, Sound-aided vibration weak signal enhancement for bearing fault detection by using adaptive stochastic resonance, *J Sound Vib.*, **449** (2019), 18–29.
50. J. Zheng, H. Pan, Q. Liu, K. Ding, Refined time-shift multiscale normalised dispersion entropy and its application to fault diagnosis of rolling bearing, *Phys. A*, **545** (2020), 123641.



AIMS Press

©2020 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License

(<http://creativecommons.org/licenses/by/4.0>)