

## Research Article

# A Novel Cuckoo Search Optimized Deep Auto-Encoder Network-Based Fault Diagnosis Method for Rolling Bearing

Jinyu Tong <sup>1,2</sup>, Jin Luo,<sup>2</sup> Haiyang Pan,<sup>2</sup> Jinde Zheng,<sup>2</sup> and Qing Zhang <sup>1</sup>

<sup>1</sup>College of Mechanics and Materials, Hohai University, Nanjing 211100, China

<sup>2</sup>School of Mechanical Engineering, Anhui University of Technology, Maanshan, Anhui 243002, China

Correspondence should be addressed to Qing Zhang; lxzhangqing@hhu.edu.cn

Received 16 July 2020; Revised 16 August 2020; Accepted 9 September 2020; Published 24 September 2020

Academic Editor: Changqing Shen

Copyright © 2020 Jinyu Tong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To enhance the performance of deep auto-encoder (AE) under complex working conditions, a novel deep auto-encoder network method for rolling bearing fault diagnosis is proposed in this paper. First, multiscale analysis is adopted to extract the multiscale features from the raw vibration signals of rolling bearing. Second, the sparse penalty term and contractive penalty term are used simultaneously to regularize the loss function of auto-encoder to enhance the feature learning ability of networks. Finally, the cuckoo search algorithm (CS) is used to find the optimal hyperparameters automatically. The proposed method is applied to the experimental data analysis. The results indicate that the proposed method could more effectively distinguish fault categories and severities of rolling bearings under different working conditions than other methods.

## 1. Introduction

Rolling bearings are widely used in rotary machines, which play an important role in determining the running states of equipment. Under complex working conditions, rolling bearings will inevitably obtain various faults. This may affect the operating performance of the whole machine and even lead to enormous losses and serious casualties. Therefore, it has important practical significance for the state monitoring and fault diagnosis of rolling bearings [1].

At present, the vibration analysis method has been widely used in fault diagnosis of rolling bearings, which generally contains three steps: (1) feature extraction, (2) feature selection, and (3) pattern recognition [2]. For (1), the effectiveness of feature extraction is related to the accuracy of fault diagnosis. The common feature extraction method mainly includes time-domain methods, frequency-domain methods, and time-frequency methods. In time domain, the most popular methods are the statistical analysis [3], such as root mean square, root amplitude, and maximum peak value. The frequency-domain methods mainly analyze the characteristic frequency [4], such as frequency standard deviation and root mean square frequency. In the time-

frequency domains, the famous methods include short-time Fourier transformation (STFT) [5], wavelet transformation (WT) [6], and empirical mode decomposition (EMD) [7]. For (2), feature selection can reduce the input dimension of classifier and improve classification efficiency. The existing feature selection methods include decision tree [6], random forest [8], and distance-based feature selection method [9]. For (3), the selected features are classified by Bayesian classification [10], support vector machines (SVM) [11], artificial neural network (ANN) [12], and other machine learning algorithms.

The shallow structure of machine learning lacks powerful representation capabilities and has difficulty in effectively learning the complex nonlinear relationships in mechanical fault diagnosis problems [13]. As a breakthrough in the field of machine learning, deep learning has received widespread attention from all fields [14, 15]. It can automatically mine the representative information hidden in the raw data and directly establish an accurate mapping relationship between data and the operating state of the equipment [16].

AE is a kind of deep learning model, which can automatically learn features from samples and is widely used

in the field of mechanical fault diagnosis [17]. Jia et al. [18] built a deep neural network based on AE and applied it to bearing fault diagnosis. Shao et al. [19] proposed a new deep feature fusion method for rotating machinery fault diagnosis, which has achieved satisfactory diagnostic results in bearing fault diagnosis and gear fault experiment analysis. Meng et al. [20] improved the performance of the denoising auto-encoder (DAE) and successfully identified rolling bearing faults. Sun et al. [21] successfully used a deep neural network based on sparse auto-encoder (SAE) for fault classification of induction motors. In addition, Shen et al. [22] proposed a fault diagnosis method for rotating machinery based on contractive auto-encoder (CAE). Zhang et al. [23] conducted mechanical fault diagnosis using the ensemble deep contractive auto-encoder (EDCAE) under noisy environment. The deep structure can automatically extract more representative and discriminating high-level features from the input and obtain satisfactory fault diagnosis results.

However, the effective characteristics of the signal are often submerged in industrial background noise. This leads to the performance decline of deep AE [24]. Moreover, the selection of hyperparameters in deep auto-encoder networks has always been a challenge. It is difficult to find the most suitable parameters. Therefore, it is necessary to design a novel deep AE.

In order to further enhance the performance of deep auto-encoder and select hyperparameters adaptively, a novel CS optimized deep auto-encoder network-based fault diagnosis method for rolling bearing is proposed in this paper. The proposed method was applied to the experimental data analysis of rolling bearing under different working conditions. The results show that the proposed method is more effective and robust than other methods. The main contributions of this paper can be summarized as follows:

- (1) The multiscale features of vibration signals are extracted as the input of the network, which improves test accuracy and reduces training time
- (2) The sparse penalty term and contractive penalty term are simultaneously used to regularize the loss function of the AE, which enhance the feature learning ability and robust of deep AE
- (3) CS algorithm is used to optimize hyperparameters of the new AE network automatically, which makes the new auto-encoder network more suitable for the signal characteristics

The rest of the paper is organized as follows. The basic AE is described in Section 2. The proposed method is described in Section 3. In Section 4, the experimental diagnosis results for rolling bearing are analyzed and discussed. Finally, conclusions are given in Section 5.

## 2. Basic AE

Basic AE is a three-layer unsupervised neural network [25], and it can learn useful features automatically by nonlinear

dimension reduction. The structure of the basic AE is shown in Figure 1. It consists of two steps: encode and decode. The encoding procedure maps the input data to hidden features vector, and the decoding procedure maps the hidden features vector back to reconstructed vector.

Given the unlabeled sample data  $\mathbf{x}^d = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$  ( $d = 1, 2, \dots, D$ ), the encoding procedure is defined as

$$\mathbf{h}_j(\mathbf{x}^d) = s(\mathbf{W}^{(1)}\mathbf{x}^d + \mathbf{b}^{(1)}). \quad (1)$$

The decoding procedure is defined as

$$\hat{\mathbf{x}}^d = s(\mathbf{W}^{(2)}\mathbf{h}_j(\mathbf{x}^d) + \mathbf{b}^{(2)}), \quad (2)$$

where  $s$  is the sigmoid function,  $\mathbf{W}^{(1)}$  and  $\mathbf{W}^{(2)}$  are the weight matrices,  $\mathbf{b}^{(1)}$  and  $\mathbf{b}^{(2)}$  are the bias matrices,  $\mathbf{h}_j(\mathbf{x}^d)$  is the hidden layer features vector and  $\hat{\mathbf{x}}^d$  is the output layer features vector,  $D$  is the number of training samples, and  $d$  stands for the  $d^{\text{th}}$  sample.

AE minimizes the reconstruction error by optimizing parameters  $\mathbf{W}$  and  $\mathbf{b}$ . The loss function of AE is usually defined by the mean square error (MSE) as

$$L(\mathbf{x}^d, \hat{\mathbf{x}}^d) = \frac{1}{2} \|\mathbf{x}^d - \hat{\mathbf{x}}^d\|^2, \quad (3)$$

where  $\|\cdot\|$  represents the norm.

Thus, the total loss function of  $M$  samples is defined as

$$J(\mathbf{W}, \mathbf{b}) = \left[ \frac{1}{M} \sum_{d=1}^M L(\mathbf{x}^d, \hat{\mathbf{x}}^d) \right]. \quad (4)$$

## 3. The Proposed Method

**3.1. DCSAE Network.** If the encoder and decoder are given too large capacity, the AE will perform the copy task without capturing any useful information about data distribution. The regular term constraints are added to the AE loss function that can encourage the model to learn useful features without restricting the capacity of the model [26].

Sparse penalty is a common regular term constraint, which is added to the loss function of AE, can reduce the dimension of input data effectively, and can speed up network training; the sparse penalty items is defined as [27]

$$J_{\text{sparse}}(\mathbf{W}, \mathbf{b}) = \sum_{j=1}^{s_2} KL(\rho \|\hat{\rho}_j), \quad (5)$$

$$\hat{\rho}_j = \frac{1}{M} \sum_{i=1}^M (h_{(i)}^d),$$

$$KL(\rho \|\hat{\rho}_j) = \sum_{j=1}^s \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j},$$

where  $j$  is the number of each unit in the hidden layer,  $s_2$  is the number of units in the second layer,  $\rho$  is the sparse parameter, which is a small value artificially given, and  $\hat{\rho}_j$  is the average activation value of the  $i^{\text{th}}$  hidden layer.

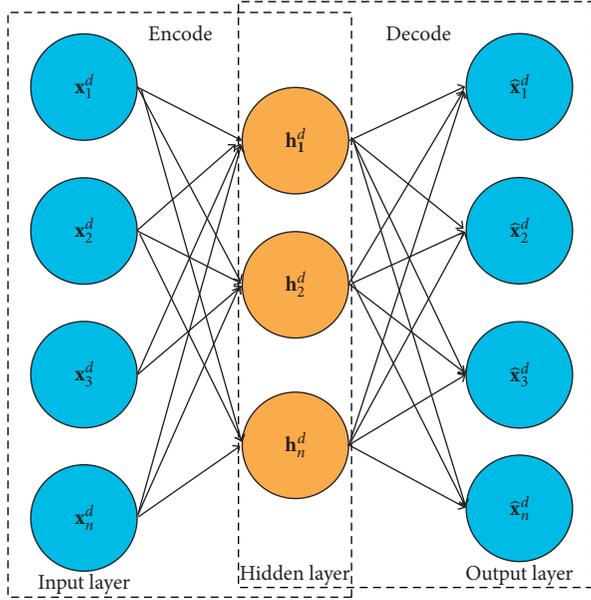


FIGURE 1: Structure of the basic AE.

Contractive penalty is added to the AE, which can learn the representative and robust hidden features from noisy environment. The contractive penalty term is defined as [28]

$$\|J_f(\mathbf{x}^d)\|_F^2 = \sum_{i=1}^n \sum_{j=1}^e \left( \frac{\partial \mathbf{h}_j(\mathbf{x}^d)}{\partial x_i} \right)^2. \quad (6)$$

The element of the  $J_f(\mathbf{x}^d)_{ji}$  in the  $j^{\text{th}}$  row and  $i^{\text{th}}$  column is defined as

$$J_f(\mathbf{x}^d)_{ji} = \frac{\partial \mathbf{h}_j(\mathbf{x}^d)}{\partial x_i}. \quad (7)$$

When the activation function of AE is the sigmoid function, equation (7) can be calculated by

$$J_f(\mathbf{x}^d)_{ji} = \mathbf{h}_j(\mathbf{x}^d)(1 - \mathbf{h}_j(\mathbf{x}^d))\mathbf{W}_{ji}^{(3)}, \quad (8)$$

where  $\mathbf{W}_{ji}^{(3)}$  is the weight matrix of the unit in layer  $i$  and the unit in layer  $j$ .

Based on equations (8) and (6), it can be further written as

$$\|J_f(\mathbf{x}^d)\|_F^2 = \sum_{j=1}^e (\mathbf{h}_j(\mathbf{x}^d)(1 - \mathbf{h}_j(\mathbf{x}^d)))^2 \sum_{i=1}^n \mathbf{W}_{ji}^{(3)2}. \quad (9)$$

Considering the advantages of sparse penalty and contractive penalty, in this paper, the sparse penalty term and contractive penalty term are added to the loss function of AE at the same time, and a contractive sparse autoencoder (CSAE) can be obtained. CSAE loss function can be defined as

$$J_{\text{CSAE}}(\mathbf{W}, \mathbf{b}) = \frac{1}{M} \sum_{d=1}^M [L(\mathbf{x}^d, \hat{\mathbf{x}}^d)] + \beta \sum_{j=1}^{s_2} KL(\rho \|\hat{\rho}_j) + \lambda \|J_f(\mathbf{x}^d)\|_F^2, \quad (10)$$

where  $\beta$  is the sparse penalty factor used to control the relativity between the sparse penalty and the reconstruction error and  $\lambda$  is the contractive penalty parameter used to control the relativity between the contractive penalty term and the reconstruction error.

Two CSAEs and one softmax classifier can be stacked to form a DCSAE network with several hidden layers, as shown in Figure 2. The detailed process of DCSAE network is as follows:

Step 1: the input data and the first hidden layer are made up of the first CSAE for feature learning

Step 2: the hidden layer features of the first CSAE are used as the input of the second CSAE; repeat this process until all hidden layers are pretrained

Step 3: update the parameters by minimizing the CSAE loss function

Step 4: the hidden features of the last CSAE are used as the input of softmax classifier for supervised training

Step 5: the backpropagation (BP) algorithm is used to fine tune the entire network to obtain a well-trained DCSAE

Moreover, the batch normalization (BN) layer is added to the DCSAE that is beneficial to the training network weight matrix.

**3.2. Parameter Optimization.** The sparse parameter  $\rho$ , the sparse penalty factor  $\beta$ , and the contractive penalty parameter  $\lambda$  in equation (10) are key hyperparameters of the DCSAE. Setting the parameters manually is hard to ensure that the set of parameter is optimal.

The CS algorithm is a simple and effective optimization approach [29], the path can be randomly searched during the optimization process, and the optimal solution is easily obtained by setting fewer parameters [30]. Thus, the CS algorithm is used to select hyperparameters of DCSAE. Process of the CS algorithm is as follows:

Step 1: set the following cuckoo parameters number of nests, population size, probability of discovery, and number of iterations, set fitness function, and initialize nest position.

Step 2: calculate the value corresponding to each nest according to the fitness function and select the current best nest.

Step 3: retain the optimal value of the previous generation and its corresponding nest position, and then, update other bird nest positions and states.

Step 4: compare the current value with the retained previous generation optimal value; replace if better, otherwise keep it unchanged.

Step 5: after updating the nest position, a new probability is randomly generated. If the new probability is less than the set discovery probability, the nest position remains unchanged; otherwise, the nest position is randomly changed. Finally, keep the best nest position.

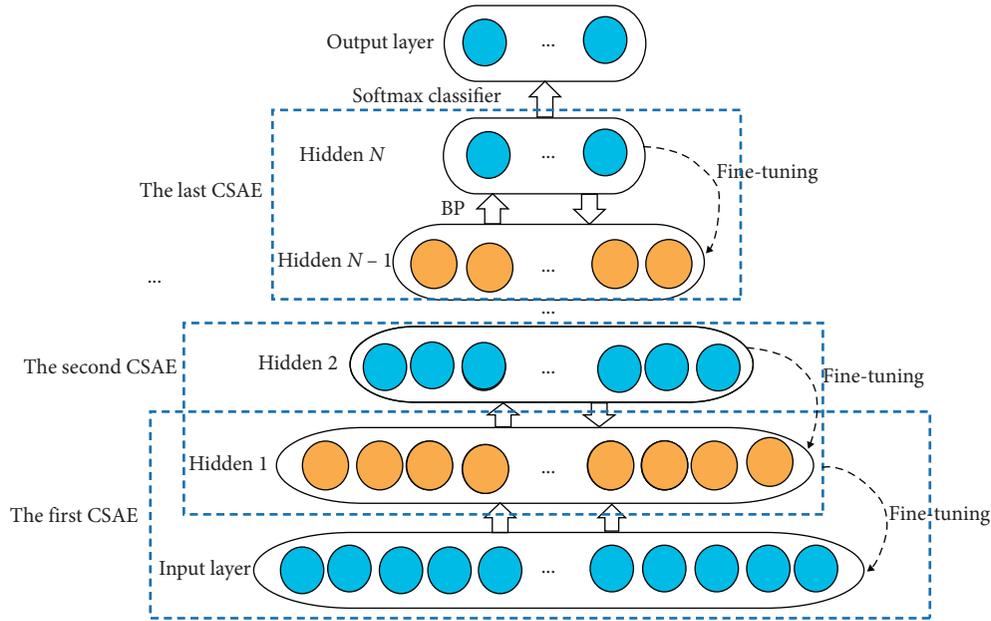


FIGURE 2: The architecture of DCSAE network.

Step 6: if the number of iterations does not reach the maximum or the error does not reach the minimum, then return to step 2; otherwise, the global best nest position is the output.

Step 7: the global optimal nest position is output as the optimization parameters and the optimization is finished.

**3.3. Procedure of the Proposed Method.** A new fault diagnosis method for rolling bearing is proposed based on CS and DCSAE in this paper. The flowchart of the proposed method is given in Figure 3.

The procedure of the proposed method is as follows:

Step 1: multiscale features are extracted from the rolling bearing vibration signals to constitute feature set

Step 2: multiscale features are divided into training samples and testing samples according to a certain proportion as the input of network

Step 3: the CS algorithm is used to optimize the key hyperparameters of DCSAE

Step 4: training samples are used to train the DCSAE network

Step 5: testing samples are used to verify the trained DCSAE; the fault diagnosis results are obtained

## 4. Experimental Verification

**4.1. Rolling Bearing Fault Test Rig.** The rolling bearing experiment data are provided by the Anhui University of Technology (AHUT). The test rig is shown in Figure 4. The test rolling bearings are 6206-2RS1 SKF deep groove ball bearings, and the single point faults are seeded by using wire cutting technology with different depths of 0.2 mm, 0.3 mm,

and 0.4 mm. The vibration signals from four kinds of rolling bearings, normal bearing (NO), bearings with inner race (IR), outer race (OR), and rolling ball (RB) faults, were collected. The motor speeds are 150 r/min, 300 r/min, 900 r/min, and 1500 r/min with the load of 0 kN, 5 kN, and 7 kN. The sampling frequency is 10.24 kHz, and the data length is 204800.

### 4.2. Experimental Verification

**4.2.1. Dataset Introduction of Experiment 1.** In this experiment, rolling bearing dataset is collected under the motor speed of 900 r/min and the load of 0 kN. The operation conditions are listed in Table 1. Each condition consists of 200 samples, and each sample is a measured vibration signal consisting of 1024 data points. The random 180 samples of each fault are chosen as training samples, and the rest 20 samples are chosen as testing samples. The time-domain vibration signals corresponding to seven types of faults are shown in Figure 5.

**4.2.2. Dataset Introduction of Experiment 2.** In this experiment, rolling bearing data of the same fault degree under the different loads were selected to verify the recognition ability of the DCSAE. The rolling bearing data set is collected under the motor speed of 900 r/min and the loads of 5 kN and 7 kN. The operation conditions are listed in Table 2. Each condition also consists of 200 samples, and each sample also consists of 1024 data points. The random 180 samples of each fault are chosen as training samples, and the rest 20 samples are chosen as testing samples.

**4.3. Multiscale Feature Extraction.** It is common to perform multiscale analysis on vibration signals to extract multiscale

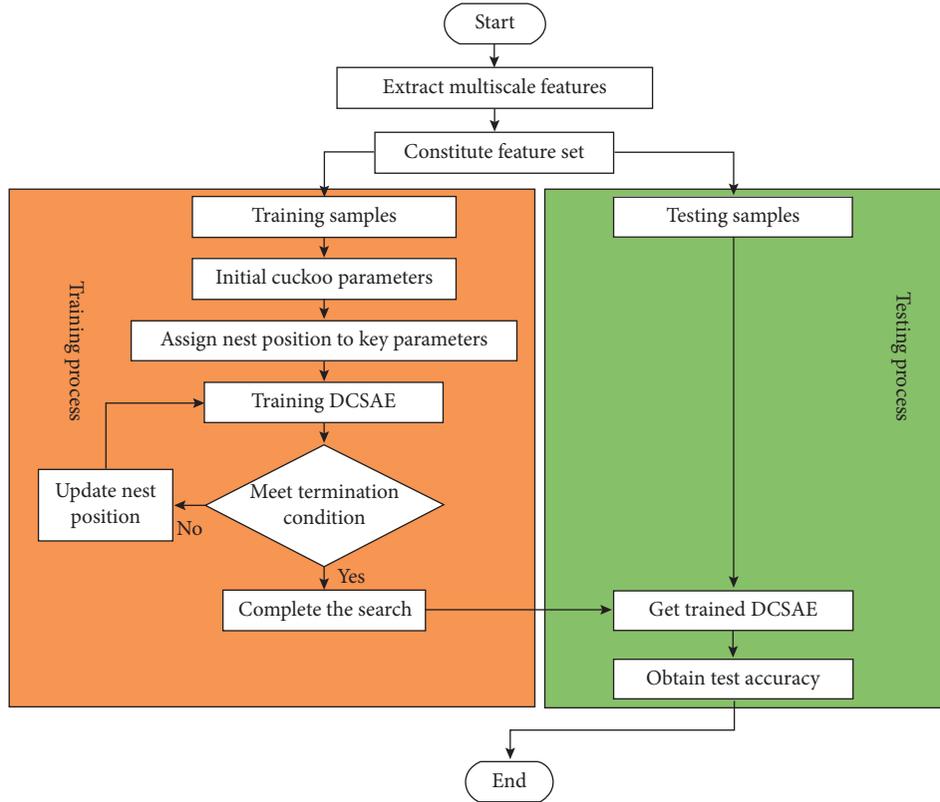


FIGURE 3: The flowchart of the proposed method.

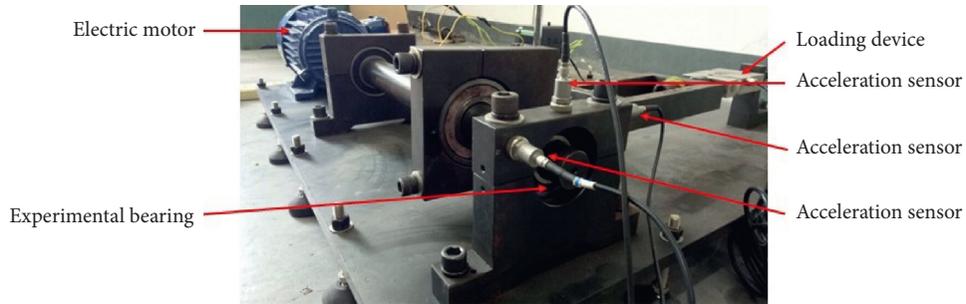


FIGURE 4: The diagram of the rolling bearing test rig.

TABLE 1: Description of the operation conditions in experiment 1.

Fault types	Fault depth	Sizes of training samples	Sizes of testing samples	Label
Inner race fault 1 (IR1-1)	0.3	180	20	1
Inner race fault 2 (IR1-2)	0.4	180	20	2
Outer race fault 1 (OR1-1)	0.2	180	20	3
Outer race fault 2 (OR1-2)	0.3	180	20	4
Rolling ball fault 1 (RB1-1)	0.2	180	20	5
Rolling ball fault 2 (RB1-2)	0.4	180	20	6
Normal (NO1-1)	0	180	20	7

features of the signals to mine fault information more fully. Multiscale analysis based on signal coarse-graining is widely used. It mainly contains two steps. First, the coarse-grained sequence of raw signals at different scales is calculated, and then, the eigen values of the coarse-grained sequence of each scale are calculated to obtain the multiscale features [31].

For a certain length of sample, the coarse-grained sequence of different scales is given by [32].

$$x_{\tau}(j) = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} u(i), \quad (11)$$

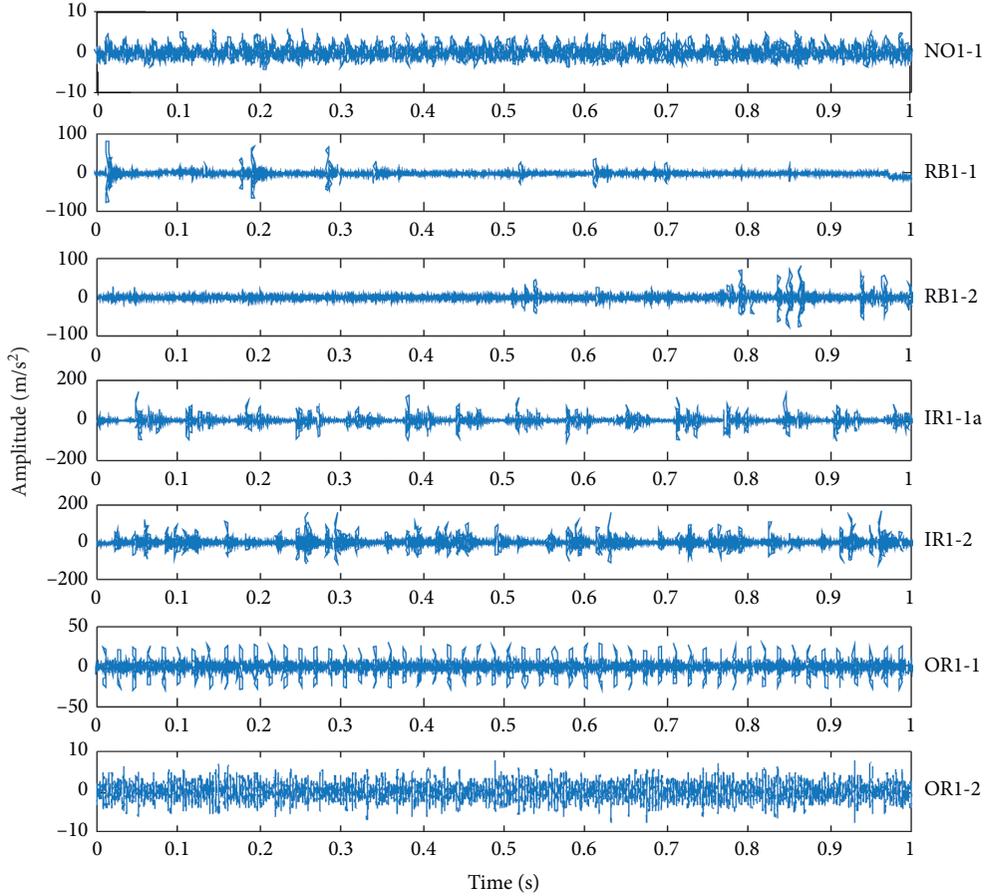


FIGURE 5: The time-domain signals of seven operation conditions.

TABLE 2: Description of the operation conditions in experiment 2.

Fault types	Fault depth (mm)	Load (kN)	Sizes of training sample	Sizes of testing samples	Label
Inner race fault 1 (IR2-1)	0.4	5	180	20	1
Inner race fault 2 (IR2-2)	0.4	7	180	20	2
Outer race fault 1 (OR2-1)	0.3	5	180	20	3
Outer race fault 2 (OR2-2)	0.3	7	180	20	4
Rolling ball fault 1 (RB2-1)	0.2	5	180	20	5
Rolling ball fault2 (RB2-2)	0.2	7	180	20	6
Normal 1 (NO2-1)	0	5	180	20	7
Normal 2 (NO2-2)	0	7	180	20	8

where  $x_\tau(j)$  is the coarse-grained sequence,  $\tau$  is the scale,  $u(i)$  is the sample, and  $L$  is the sample length,  $1 \leq j \leq (L/\tau)$ .

To illustrate the effectiveness of multiscale features as network input, raw bearing vibration signals, single-scale features, and multiscale features of bearing vibration signals are used as input for DCSAE. In multiscale features extraction based on the data of experiment 1, each sample consisting of 1024 data points and 17 types of time-domain and frequency-domain features in the first 5 scales of the raw rolling bearing vibration signals are extracted to form a new features set with a dimension of  $1 * 85$ . The extracted features [33] are as follows:

- (1) Time-domain features: mean, root mean square, root amplitude, average amplitude, maximum peak value, standard deviation, skewness index, kurtosis index,

peak value index, margin index, waveform index, and pulse index

- (2) Frequency-domain features: center of gravity frequency, frequency standard deviation, root mean square frequency, skewness frequency, and kurtosis frequency

The formulas of partial time-domain features and frequency-domain features are shown in Table 3.

The basic parameters of the CS algorithm are set as follows: the number of nests is 20, the probability of discovery is 0.25, the number of iterations is 20, and the upper and lower bounds of parameter optimization are [0.001, 1]. Optimal hyperparameters  $[\rho \beta \lambda]$  are [0.15 0.27 0.13] through the CS algorithm. The deep structure DCSAE is

TABLE 3: The formula of the partial extracted features.

Statistical features in time domain					
Skewness index	$c_w = (((1/N) \sum_{i=1}^N (x_i - \bar{x})^3) / x_r^3)$	Kurtosis value	$K_f = (x_r / \beta)$	Peak value index	$C_f = (x_{\max} / \bar{x}_x)$
Margin index	$L_f = (x_{\max} / x_r)$	Waveform index	$S_f = (\bar{x}_x /  \bar{x} )$	Pulse index	$I_f = (x_{\max} /  \bar{x} )$
Note: $x_r =  (1/N) \sum_{i=1}^N \sqrt{ x_i - \bar{x} } $ , $\beta = [(1/N) \sum_{i=1}^N (x_i - \bar{x})^2]^2 = \sigma^4$					
Statistical features in frequency domain					
Center of gravity frequency	$f_c = (\int_0^{+\infty} f S(f) df / \int_0^{+\infty} S(f) df)$	Frequency standard deviation	$f_{SD} = \sqrt{(\int_0^{+\infty} (f - f_c)^2 S(f) df / \int_0^{+\infty} S(f) df)}$		
Root mean square frequency, $f_{RMS} = ((\int_0^{+\infty} \sqrt{f - f_p} S(f) df) / \sqrt{\sigma} N)$					
Skewness frequency	$f_s = ((\int_0^{+\infty} (f - f_c)^3 S(f) df) / f_{SD}^3 \cdot N)$	Kurtosis frequency	$f_k = (\int_0^{+\infty} (f - f_c)^4 S(f) df) / \sigma^4 \cdot N)$		
Note: $S(f)$ is the power spectral density, $f_p = (\int_0^{+\infty} f \cdot S(f) df) / \int_0^{+\infty} S(f) df$ , $\sigma = \sqrt{(\int_0^{+\infty} (f - f_p)^2 S(f) df) / N}$					

selected as [85 50 30 7] by experimentation, the learning rate is 0.01, and the number of iterations is 200. The comparison results of different inputs for DCSAE are shown in Figure 6. The test accuracy and training time are shown in Table 4.

In Figure 6 and Table 4, compared with the other groups, the training time is the longest and the average test accuracy is the lowest in each trial when raw data are used as input for DCSAE; this is because raw data of rolling bearing are high-dimension data, and they contain a lot of redundant information which is not conducive to the weight training in the network and will lead to long training time. The training time of single-scale feature as the input for DCSAE is the shortest, but the recognition rate is far lower than that of multiscale feature.

Multiscale feature as the input for network can fully mine fault information; therefore, multiscale features are used as input for all training networks of experimental verification in Section 4.4.

#### 4.4. Results and Analysis

**4.4.1. Results and Analysis of Experiment 1.** To verify the effectiveness of the proposed method, it is compared with SVM, SAE, stacked sparse auto-encoder (SSAE), and stacked contractive auto-encoder (SCAE).

In this experiment, the basic parameters of the CS are set as given in Section 4.2, and the optimal hyperparameters  $[\rho \beta \lambda]$  are [0.15 0.27 0.13] through the CS algorithm. The deep structure DCSAE is selected as [85 50 30 7] by experimentation, the learning rate is 0.01, and the number of iterations is 200. The kernel function of SVM is RBF function, and the penalty factor and kernel function parameters are obtained by the CS, which are 0.64 and 0.017, respectively. The structure of SAE is [85 30 85], the learning rate is 0.01, and the hidden layer features are input to the softmax for classification. The structures of SSAE and SCAE are the same as that of the DCSAE, the learning rate is 0.01, the number of iterations is

200, and the sparse parameters  $\rho$  and  $\beta$  in SSAE and the contractive penalty parameters  $\lambda$  in SCAE are also obtained by the CS as well, which are 0.08, 0.18, and 0.69, respectively.

Ten trials are carried out for the five methods to eliminate the influence of accidental errors, and the average result of the ten trials was used as the evaluation index. The comparison results of the five methods are shown in Figure 7 and Table 5. The following can be seen from Figure 7 and Table 5:

- (1) Comparison between SSAE, SCAE, and DSCAE: the average test accuracies of SSAE and SCAE are 96.43% and 95%, respectively; they are lower than that of DSCAE, which is 98.57%. The standard deviations of SSAE and SCAE are smaller than that of DSCAE. It is indicated that the sparse penalty term and the contractive penalty term are applied to the loss function of the stacked AE at the same time, which can obtain more satisfactory diagnosis results.
- (2) Comparison between SVM, SAE, and other three deep networks: the average test accuracies of SVM and SAE are 52.14% and 70.71%, respectively; they are much lower than that of other three deep networks as SSAE, SCAE, and DSCAE. This illustrates that deep learning methods have better performance than shallow models in dealing with large samples, and the reason is that the deep learning method can mine more useful features from fault data.

The confusion matrix of the DCSAE for experiment 1 is shown in Figure 8. The columns stand for the true label, and the rows stand for the predicted label; the chart bar in the right shows the correspondence between color and numbers (from 0 to 1). DCSAE obtains the best result of 100% on the IR1-1, IR1-2, OR1-1, OR1-2, and NO1-1 in Figure 8. The only misclassification occurred on RB fault samples, and about 10% of RB1-1 samples were misclassified as RB 1-2.

A manifold learning method is used to visualize the data samples to further analyze the feature extraction capability

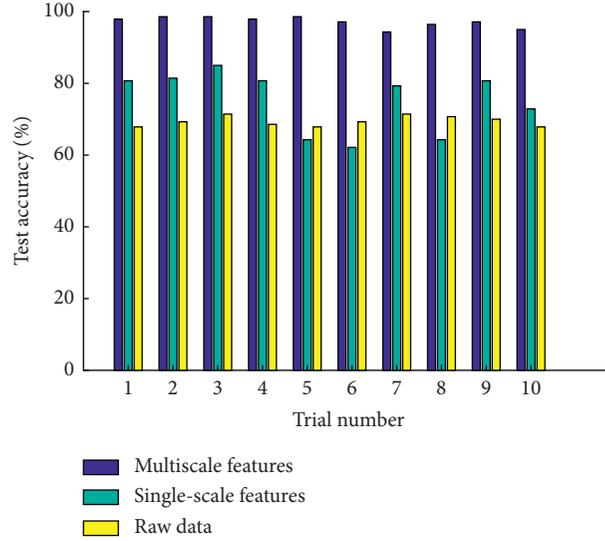


FIGURE 6: Comparison chart of different inputs for DCSAE.

TABLE 4: Training time and average test accuracy of different inputs for DCSAE.

Input	Training time (s)	Average test accuracy (%)
Raw data	136.97	68.43
Single-scale features	2.67	75.14
Multiscale features	5.83	98.57

TABLE 5: Diagnosis results of five different methods.

Method	Average test accuracy (%)	Standard deviation
SVM	52.14 (73/140)	0.1588
SAE	70.71 (99/140)	0.0714
SSAE	96.43 (135/140)	0.0232
SCAE	95 (133/140)	0.0357
DCSAE	98.57 (138/140)	0.0132

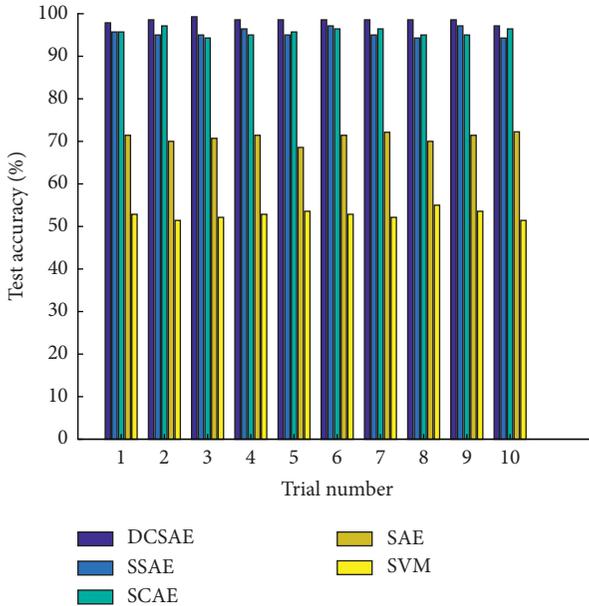


FIGURE 7: Diagnosis results of the 10 trials using the five methods.

of DCSAE. t-distribution stochastic neighbor embedding (t-SNE) [34] is an embedding model that can map data in high-dimensional space to low-dimensional space and retain the local characteristics of the data set. It is mainly used for dimension reduction and visualization of high-dimensional data. Therefore, t-SNE is adopted to extract the raw data and

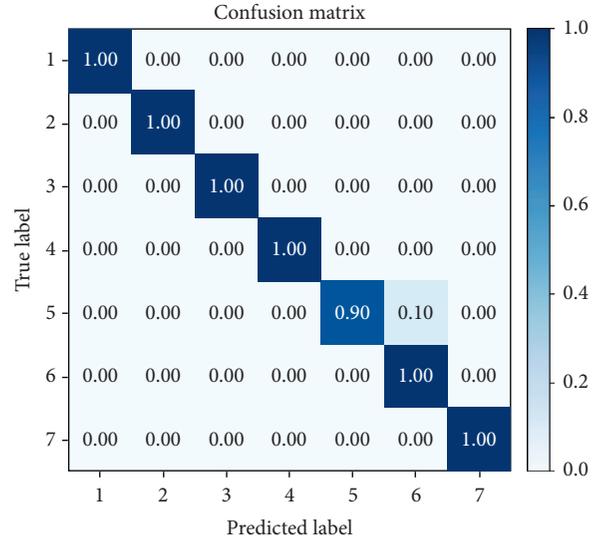


FIGURE 8: Confusion matrix of the DCSAE for experiment 1.

the first three components of the last layer of DCSAE and draw the scatter plots in Figures 9 and 10. As shown in Figure 10, fault samples of the same category converge in a center and fault samples of different categories can be easily distinguished. In conjunction with Figures 8 and 10, it can be seen that a small number of samples overlap on the RB1-1 and RB1-2. This leads to the misclassification of the samples with the RB fault condition.

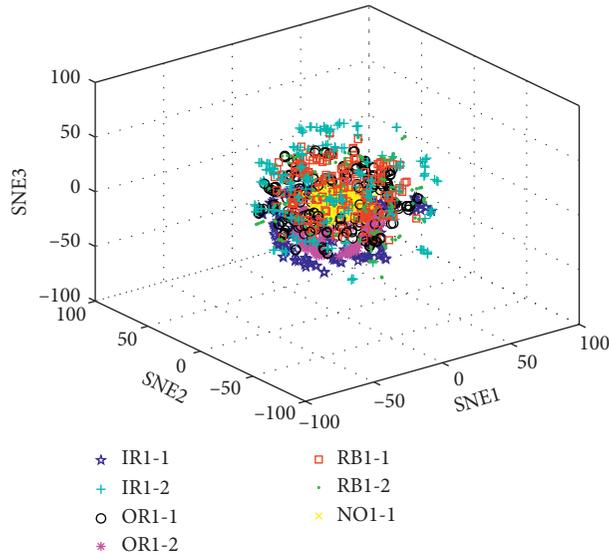


FIGURE 9: Visualization map of raw data in experiment 1.

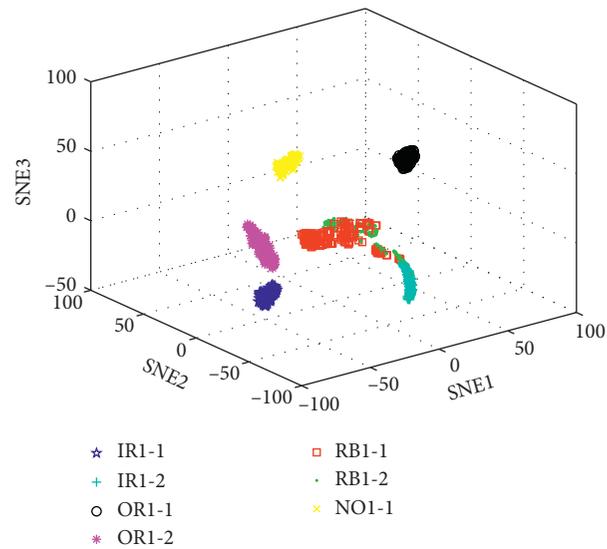


FIGURE 10: Feature visualization map of DCSAE in experiment 1.

In general, DCSAE has achieved a high accuracy to distinguish the different categories under the same motor speed and load. The average test accuracy is higher than that of existing deep learning methods and traditional machine learning methods.

**4.4.2. Results and Analysis of Experiment 2.** In this section, DCSAE is compared with other three deep AE networks: stacked AE, SSAE, and SCAE. Architecture of the four deep AE networks are [85 50 30 8] by experimentation, the learning rate is 0.01, and the number of iterations is 200. Multiscale features of the time domain and frequency domain were extracted as the input of different networks. The optimal hyperparameters  $[\rho \beta \lambda]$  are [0.26 0.56 0.5] that were obtained through the CS algorithm.

The result of the 10 trials is shown in Figure 11, and the average test accuracy and standard deviation of the specific 10 trials are shown in Table 6. As shown in Figure 11 and Table 6, compared with other three deep AE networks, the average test accuracy of DCSAE is  $91.25 \pm 0.0092$  which has the highest test accuracy and the lowest standard deviation. It indicates that DCSAE can obtain better diagnosis accuracy when dealing with fault data under different conditions.

The confusion matrix of the DCSAE for experiment 2 is shown in Figure 12. From the result, about 5% of NO2-1 are misclassified as NO2-2, about 25% of RB2-1 are misclassified as RB 2-2, and about 35% of RB2-2 are misclassified as RB 2-1; all of other categories are classified correctly while processing bearing fault data with different loads.

Similar to the above process, t-SNE is used to extract the raw data and the first three components of the last

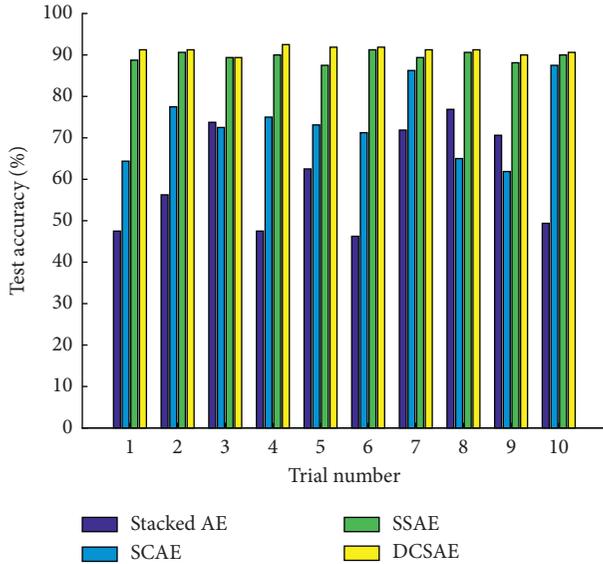


FIGURE 11: Diagnosis results of the 10 trials using the four methods.

TABLE 6: Diagnosis results of four different methods.

Method	Average test accuracy (%)	Standard deviation
Stacked AE	60.63 (97/160)	0.123
SSAE	89.38 (143/160)	0.0118
SCAE	73.75 (118/140)	0.0865
DCSAE	91.25 (146/160)	0.0092

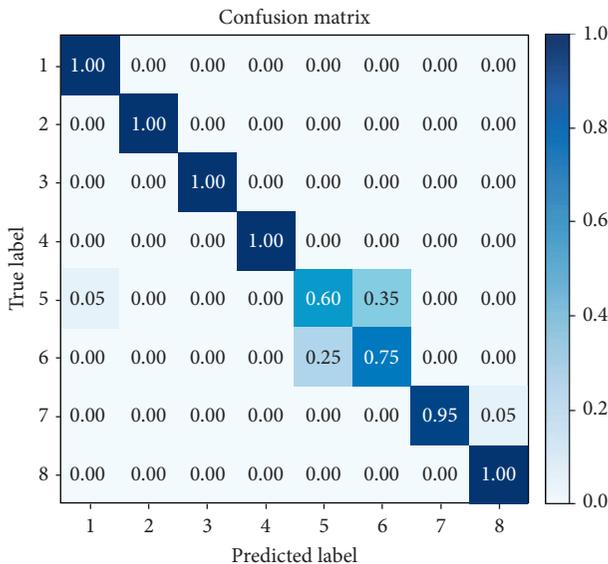


FIGURE 12: Confusion matrix of DCSAE for experiment 2.

layer of DCSAE in experiment 2 and draw the scatter plots in Figures 13 and 14. As shown in Figure 14, fault samples of different categories under different loads can be distinguished. In conjunction with Figures 12 and 14, it can be seen that a large number of samples overlap on the

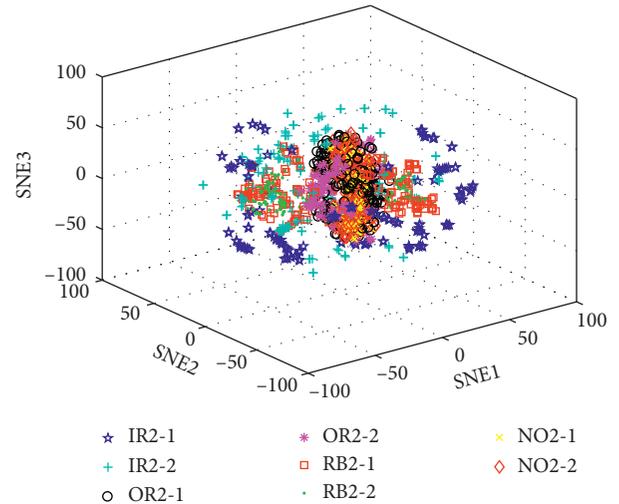


FIGURE 13: Visualization map of raw data in experiment 2.

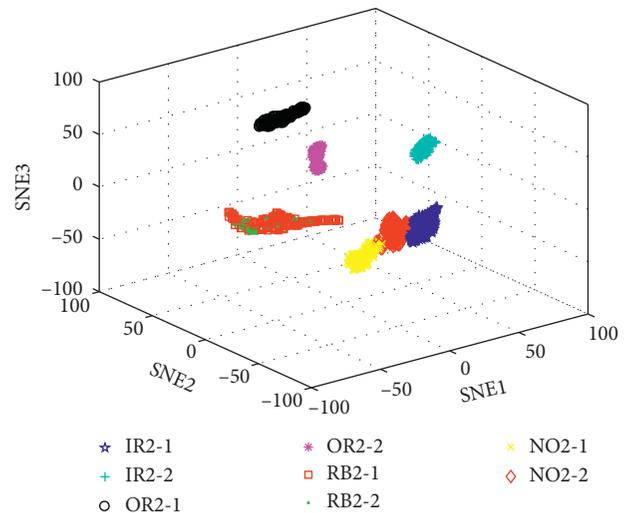


FIGURE 14: Feature visualization map of DCSAE in experiment 2.

RB2-1 and RB2-2, and it is difficult for the softmax classifier to achieve a high-precision classification effect. This leads to the misclassification of the samples with the RB fault condition. Overall, DCSAE has robustness and generalization ability to distinguish the different categories with the different loads.

## 5. Conclusion

In this paper, a novel deep AE network is proposed for rolling bearing fault diagnosis. In the proposed network, the sparse penalty term and the contractive penalty term are adopted to regularize the novel deep AE loss function to enhance the ability of feature learning. Moreover, the multiscale features of rolling bearing vibration signal are extracted as the input of DCSAE to reduce the training time and data dimension. Furthermore, the CS algorithm is used to optimize the key hyperparameters of the deep auto-encoder automatically. Effectiveness of the proposed method is

verified by two different experiments of rolling bearings fault diagnosis. The results show that the proposed method is more effective and robust for fault diagnosis than other methods. It is a worthy research topic to apply the deep learning method to other mechanical fault diagnosis fields. The authors would continue to study this field in the future work.

### Data Availability

The data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 6/12 months after publication of this article, will be considered by the corresponding author.

### Conflicts of Interest

The authors declare that they have no conflicts of interest concerning the publication of this manuscript.

### Acknowledgments

This work was supported by the National Key Technologies Research & Development Program of China (no. 2017YFC0805100), the National Natural Science Foundation of China (nos. 51975004, 11932006, and U1934206), the Natural Science Foundation of Anhui Province, China (no. 2008085QE215), and the Key Program of Natural Science Research of Higher Education in Anhui Province of China (nos. KJ2019A053 and KJ2019A092).

### References

- [1] A. Hussein and T. Irina, "A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling," *Journal of Sound and Vibration*, vol. 369, pp. 246–265, 2016.
- [2] B. Samantha and K. R. Al-Balushi, "Artificial neural network based fault diagnostic of rolling element bearing using time domain features," *Mechanical Systems and Signal Processing*, vol. 17, no. 2, pp. 317–328, 2003.
- [3] Y. Lei, Z. He, Y. Zi et al., "Application of an intelligent classification method to mechanical fault diagnosis," *Expert Systems With Applications*, vol. 36, no. 6, pp. 9941–9948, 2009.
- [4] J. Shi and M. Liang, "Intelligent bearing fault signature extraction via iterative oscillatory behavior based signal decomposition (IOBSD)," *Expert Systems with Applications*, vol. 45, pp. 40–55, 2016.
- [5] K. Pichler, E. Lughofer, M. Pichler, T. Buchegger, E. P. Klement, and M. Huschenbett, "Fault detection in reciprocating compressor valves under varying load conditions," *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 104–119, 2016.
- [6] V. Muralidharan and V. Sugumaran, "Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump," *Measurement*, vol. 46, no. 1, pp. 353–359, 2013.
- [7] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1–2, pp. 108–126, 2013.
- [8] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mechanical Systems and Signal Processing*, vol. 76–77, pp. 283–293, 2016.
- [9] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Sparse bayesian classification of EEG for brain-computer interface," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2256–2267, 2016.
- [10] L. Meng, W. Miao, and C. Wang, "Research on SVM classification performance in rolling bearing diagnosis," *International Conference on Intelligent Computation Technology and Automation*, vol. 3, pp. 132–135, 2010.
- [11] L. Jinshan, C. Doub, and Q. Wang, "Comparisons of MFDFA, EMD and WT by neural network, mahalanobis distance and SVM in fault diagnosis of gearboxes," *Computer Modeling in Engineering & Sciences*, vol. 52, no. 2, pp. 11–15, 2018.
- [12] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, pp. 16–27, 2015.
- [13] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.
- [14] C. Shang, F. Yang, D. Huang, and W. Lyu, "Data-driven soft sensor development based on deep learning technique," *Journal of Process Control*, vol. 24, no. 3, pp. 223–233, 2014.
- [15] Y. Wang and X. Yan, "Novel ensemble modeling method for enhancing subset diversity using clustering indicator vector based on stacked autoencoder," *Computer Modeling in Engineering & Sciences*, vol. 121, no. 1, pp. 123–144, 2019.
- [16] H. Jiang, H. Shao, and X. Li, "Intelligent fault diagnosis method for aircraft based on deep learning," *Journal of Mechanical Engineering*, vol. 55, no. 7, pp. 27–34, 2019.
- [17] S.-Z. Su, Z.-H. Liu, S.-P. Xu, S.-Z. Li, and R. Ji, "Sparse auto-encoder based feature learning for human body detection in depth image," *Signal Processing*, vol. 112, pp. 43–52, 2015.
- [18] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619–628, 2018.
- [19] H. Shao, H. Jiang, F. Wang, and H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowledge-Based Systems*, vol. 119, pp. 200–220, 2017.
- [20] Z. Meng, X. Zhan, J. Li, and Z. Pan, "An enhancement denoising autoencoder for rolling bearing fault diagnosis," *Measurement*, vol. 130, pp. 448–454, 2018.
- [21] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.
- [22] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.
- [23] Y. Zhang, X. Li, L. Gao, W. Chen, and P. Li, "Ensemble deep contractive auto-encoders for intelligent fault diagnosis of machines under noisy environment," *Knowledge Based Systems*, vol. 196, Article ID 105764, 2020.

- [24] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: properties and applications in non-Gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.
- [25] Y. Wang, H. Yao, X. Sun et al., "Research on the expressive ability of auto-encoder in deep learning," *Computer Science*, vol. 42, no. 9, pp. 56–60, 2015.
- [26] W. Liu, T. Ma, D. Tao, and J. You, "HSAE: a hessian regularized sparse auto-encoders," *Neurocomputing*, vol. 187, pp. 59–65, 2016.
- [27] D. Yu, K. Yao, H. Su, G. Li, and F. Seide, "KL-divergence regularized deep neural network adaptation for improved large vocabulary speech recognition," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7893–7897, Vancouver, Canada, May 2013.
- [28] S. Rifai, P. Vincent, X. Muller et al., "Contractive auto-encoders: explicit invariance during feature extraction," *International Conference on Machine Learning*, pp. 833–840, Honolulu, HI, USA, December 2011.
- [29] X. Liu and M. Ma, "Application of cuckoo search algorithm in multi-threshold image segmentation," *Computer Engineering*, vol. 39, no. 7, pp. 274–278, 2013.
- [30] P. Korosec, J. Silc, B. Filipic et al., "The differential ant-stigmergy algorithm," *Information Sciences*, vol. 192, no. 1, pp. 82–97, 2012.
- [31] C. Li, J. Zheng, H. Pan, J. Tong, and Y. Zhang, "Refined composite multivariate multiscale dispersion entropy and its application to fault diagnosis of rolling bearing," *IEEE Access*, vol. 7, pp. 47663–47673, 2019.
- [32] C. Wang, "Multi-scale models and related analysis methods," *Complex Systems and Complexity Science*, vol. 1, no. 1, pp. 9–19, 2004.
- [33] V. Srinivasan, C. Eswaran, and N. Sriraam, "Artificial neural network based epileptic detection using time-domain and frequency-domain features," *Journal of Medical Systems*, vol. 29, no. 6, pp. 647–660, 2005.
- [34] V. D. M. Laurens and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.