


Article

A Novel Intelligent Fault Diagnosis Method for Rolling Bearing Based on Integrated Weight Strategy Features Learning

Jun He ¹ , Ming Ouyang ^{1,*}, Chen Yong ¹, Danfeng Chen ¹, Jing Guo ¹ and Yan Zhou ²

¹ College of Automation, Foshan University, Foshan City 528000, Guangdong Province, China; hejun_723@fosu.edu.cn (J.H.); chen_y@fosu.edu.cn (C.Y.); cdf2017@fosu.edu.cn (D.C.); guojing@fosu.edu.cn (J.G.)

² College of Computer Science, Foshan University, Foshan City 528000, Guangdong Province, China; zhouyan791266@163.com

* Correspondence: ouymouym@outlook.com; Tel.: +86-134-3311-0891

Received: 2 February 2020; Accepted: 19 March 2020; Published: 23 March 2020



Abstract: Intelligent methods have long been researched in fault diagnosis. Traditionally, feature extraction and fault classification are separated, and this process is not completely intelligent. In addition, most traditional intelligent methods use an individual model, which cannot extract the discriminate features when the machines work in a complex condition. To overcome the shortcomings of traditional intelligent fault diagnosis methods, in this paper, an intelligent bearing fault diagnosis method based on ensemble sparse auto-encoders was proposed. Three different sparse auto-encoders were used as the main architecture. To improve the robustness and stability, a novel weight strategy based on distance metric and standard deviation metric was employed to assign the weights of three sparse auto-encodes. Softmax classifier is used to classify the fault types of integrated features. The effectiveness of the proposed method is validated with extensive experiments, and comparisons with the related methods and researches on the widely-used motor bearing dataset verify the superiority of the proposed method. The results show that the testing accuracy and the standard deviation are 99.71% and 0.05%.

Keywords: stacked auto-encoder; weighting strategy; deep learning; bearing fault diagnosis

1. Introduction

With the upgrading of industrial capacity, the connection between machine equipment is increasingly inseparable. Once unexpected faults happen in a machine, it may indirectly effect the reliability of other connected machineries [1]. These failures will cause heavy economic loss, and even more seriously, they could be life-threatening [2]. Therefore, the automatic, accurate, and timely recognition of the health conditions of machine equipment is highly necessary [3,4].

In the past few years, intelligent fault diagnosis methods have attracted great attentions and widely adopted in the condition monitoring systems [5,6]. Generally, intelligent fault diagnosis methods can be divided three main steps: (1) signals acquisition; (2) feature extraction and selection; (3) fault classification [7,8]. After a literature review, it can be found that a tremendous amount of researches have focused on how to extract discriminative features from collected signals based on abundant signal processing technologies [9,10], such as time-domain [11,12], frequency-domain [13], time-frequency-domain statistics analytical methods [14], or other waveform transform methods [15,16]. To classify the extracted features, a few artificial intelligence methods (ANN, SVM, etc.) are applied. For instance, Fu et al. [17] proposed a novel hybrid approach coupling variational mode decomposition and SVM to identify rolling bearing fault types. Ali et al. [15] used empirical mode decomposition

to extract 10 time-domain statistical features and an artificial neural network is used to identify the health conditions of rolling bearing. He et al. [18] proposed an ensemble error minimized learning machine method to recognize rolling bearing faults, empirical mode decomposition technology is adopted to extract the ensemble time-domain features. However, although these traditional intelligent methods did work and achieved an accurate diagnosis result, they still have two deficiencies: (1) the features are usually manually extracted depending on prior knowledge and diagnostic expertise, which accorded to a specific fault type and probably unsuitable for other faults [19,20]; (2) In real industries, the collected signals are usually exposed to environmental noises, which cause the signals to be complex and non-stationary, and signal processing technologies need to be employed to filter the collected signals to obtain the effective features [3,21]. Consequently, there is an urgent need to develop new intelligent fault diagnosis methods to accomplish fault diagnosis tasks automatically.

As an emerging research field, deep learning has a powerful ability to extract the representative features from the collected signals, which makes it has the potential to overcome the shortcomings of the traditional intelligent diagnosis methods [22,23]. The advantage of deep learning is that can automatically learn discriminative features and classified faults, which removes the requirements of manual feature extraction and prior knowledge from the diagnosis model. After more than ten years of development, deep learning has been gradually applied to the field of fault diagnosis. For example, Liu et al. [24] presented a fault diagnosis method for rolling bearings based on convolution neural network (CNN) in which the step k is used to discretize the vibration signal and the discrete sequence as the input data of CNN. Jia et al. [25] used the normalized sparse AEs to constitute local connection network, and the model can learn to avoid similar, repeated features and overcome the problem of feature change. Shao et al. [26] proposed an improved convolution deep belief network method based on compressed sensing technology, this method used compressed data as the input of the model and obtained less time consumption of the fault diagnosis. A novel cross-domain fault diagnosis method was proposed by Li et al. [27] whereby multiple deep generative neural networks were employed to generate corresponding-domain fake samples, and faults in different domains could be discriminated well. Long et al. [28] used a competitive swarm optimizer and a local search algorithm to optimize the weights of echo state networks for decreasing the affect caused by random selection of input weights and reservoir weights. Although the above researches are successfully applied in fault diagnosis, there still exist shortcomings in that these intelligent diagnosis methods based on deep learning mainly focus on the research of the individual learning model. Due to complexity of the collected vibration data and even there are exiting the imbalance between different data [29], the generalization can seldom perform well consistently when used individual deep learning model. This problem derives from the limitation of individual deep learning models for the fault diagnosis of complicate mechanical equipment [30]. Ensemble learning is another method of machine learning that can effectively deal with this problem, ensemble learning uses several models and an integration strategy to maximize the strengths of individual models and achieve better results than an individual model [31,32]. Among them, the integration strategy plays an important role in the ensemble learning, and directly affect the accuracy of the diagnosis results. Therefore, it is meaningful to study ensemble learning models.

In this paper, a novel ensemble learning method based on multiple stacks sparse AEs is proposed for bearing intelligent fault diagnosis. The proposed method is mainly included three steps: Firstly, three stack sparse AEs with different weights are used to extract the representative features from the raw vibration signals. Secondly, a feature integrated strategy based on distance and standard deviation metrics is designed to fine tune the extracted features, which improves the robustness and stability of the diagnosis result. Finally, the softmax classifier is used to classify the fault types based on the integrated features. Experimental results show that the proposed method can get rid of the dependence of manual design algorithm to extract features, and overcome the limitations of an individual deep learning model, which is superior compared with other similar intelligent diagnosis methods. In brief, the contributions of this paper are summarized as follows:

Suppose a n -dimensional unlabeled training sample is $x = \{x_1, x_2, \dots, x_n\} \in \mathfrak{R}^{1 \times n}$, the training process of AE is a representation that transform the input sample x into a hidden layer vector a_1 , the vector a_1 can be denoted as $a_1 = \{a^1, a^2, \dots, a^s\} \in \mathfrak{R}^{1 \times s}$, the calculation procedures is as follows:

$$a_1 = f(w_1^{(1)}x + b_1^{(1)}) \quad (1)$$

where $w_1^{(1)}$ is the weight matrix, $b_1^{(1)}$ and $f(\cdot)$ are the offset vector and the activation function, respectively. Sigmoid [36,37] as the activation function used to train AE given as follows:

$$f(z) = 1/(1 + e^{-z}) \quad (2)$$

Then, the hidden vector a_1 will be decoded and reconstructed as the vector \hat{x} by the Equation (3), the vector \hat{x} can be denoted as $\hat{x} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n\} \in \mathfrak{R}^{1 \times n}$. Equation (3) gives as follows:

$$\hat{x} = f(w_1^{(2)}a_1 + b_1^{(2)}) \quad (3)$$

where the $w_1^{(2)}$ and $b_1^{(2)}$ are the parameters of hidden layer to output layer. This works as the Equation (1).

The aims of training process is to obtain the approximation optimal value of parameter w and b through minimized the reconstruction errors.

For a sample set $\{x^m\}_{m=1}^M$ with M samples, its reconstruction cost function can be expressed as follow:

$$J_1(w, b) = \frac{1}{M} \sum_{m=1}^m L(x^m, \hat{x}^m) \quad (4)$$

where $L(x^m, \hat{x}^m)$ is the reconstruction error square, which is given as Equation (5)

$$L(x^m, \hat{x}^m) = \|x^m - \hat{x}^m\|^2 \quad (5)$$

2.2. Sparse Auto-Encoder

In the training process of AE, training samples usually contain a lot of redundant information, which means that the training samples only contain a small amount of useful information, and the hidden neurons are not all activated to represent the information of input data, especially when the dimension of input data is less than the number of hidden neurons. Therefore, for each AE, a sparse constraint is adopted to limit the number of activated neurons in the hidden layer [37,38]. Kullback–Leibler (KL) divergence, as a constraint condition usually used in AE training, can be expressed as follows:

$$KL(\rho \parallel \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \quad (6)$$

where ρ and $\hat{\rho}_j$ are the sparse factor and average activated number of j th hidden neurons, respectively, and $KL(\rho \parallel \hat{\rho}_j)$ denotes the discrepancy ρ and $\hat{\rho}_j$.

To sum up the above conclusions, it can minimize the cost function of each AE to get optimal pre-training parameters w and b . So, the cost function can be rewritten as follows:

$$J(W, b) = \frac{1}{M} \sum_{m=1}^M L(x^m, \hat{x}^m) + \beta \cdot \sum_{j=1}^s KL(\rho \parallel \hat{\rho}_j) \quad (7)$$

where β is the dilution penalty factor.

2.3. Softmax Classifier

Softmax classifier is a linear classifier that commonly used in multi-classification tasks, whose output is the probability value of each class [39]. Given a training sample set $\{x^m\}_{m=1}^M$ and $x^m \in \mathcal{X}^{1 \times n}$, its corresponding sample label set is $\{y^m\}_{m=1}^M$ with $y^m \in \{1, 2, \dots, K\}$. For each given sample x^m , softmax classifier will compute the values $p(y^m = k|x^m)$, which is the probability of each class. Therefore, for each different input sample, the output is always a K dimension vector of probability, and the position of the maximum probability determines the class of the sample, which can be expressed by the following hypothetical functions

$$h_{\theta}(x^m) = \begin{bmatrix} p(y^m = 1|x^m; \theta) \\ p(y^m = 2|x^m; \theta) \\ \vdots \\ p(y^m = K|x^m; \theta) \end{bmatrix} = \frac{1}{\sum_{k=1}^K e^{\theta_k^T x^m}} \begin{bmatrix} e^{\theta_1^T x^m} \\ e^{\theta_2^T x^m} \\ \vdots \\ e^{\theta_K^T x^m} \end{bmatrix} \quad (8)$$

where $\theta = [\theta_1, \theta_2, \dots, \theta_K]^T$ is the parameter of Softmax classifier, h_{θ} is the normalized probability. The optimization of model parameters can be achieved by minimizing the cost function $J(\theta)$.

$$J(\theta) = -\frac{1}{M} \left[\sum_{m=1}^M \sum_{k=1}^K I\{y^m = k\} \log \frac{e^{\theta_k^T x^m}}{\sum_{k=1}^K e^{\theta_k^T x^m}} \right] \quad (9)$$

where $I\{\cdot\}$ is an indicator function, when the condition is true, the function return 1 otherwise return 0.

3. Proposed Fault Diagnosis Method

In this section, the proposed bearing fault diagnosis method is presented. First, three different sparse auto-encoders are constructed and used to extract the features from the raw vibration signal in Section 3.1. The weight strategy is described in Section 3.2. In Section 3.3, the feature integration is introduced. Softmax classifier is used to classify the health condition of the integrated features, and the detailed process is shown in Figure 2.

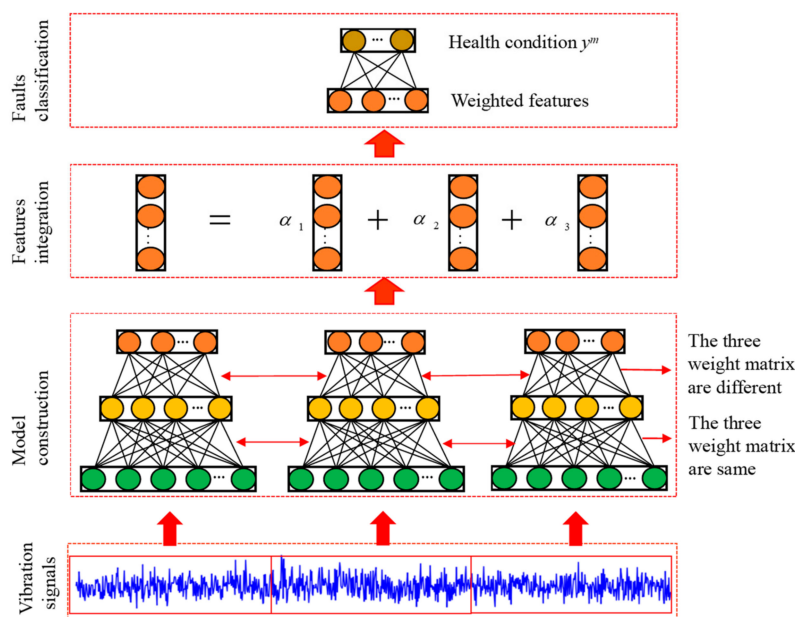


Figure 2. Illustration of the proposed.

3.1. Ensemble Auto-Encoders Construction

In order to construct three different sparse auto-encoder models, we divide the original vibration signal into three segments, and each segment uses a SAE to extract features. Assuming the input dimension of SAE is N_{in} , when training individual SAE, we randomly select N_t training samples from the data set, which are obtained by overlapping sampling method. Each training sample consists of three N_{in} segments, which means that there are three segment samples that can be used to training in each N_t . The details of training process are shown in the Figure 3.

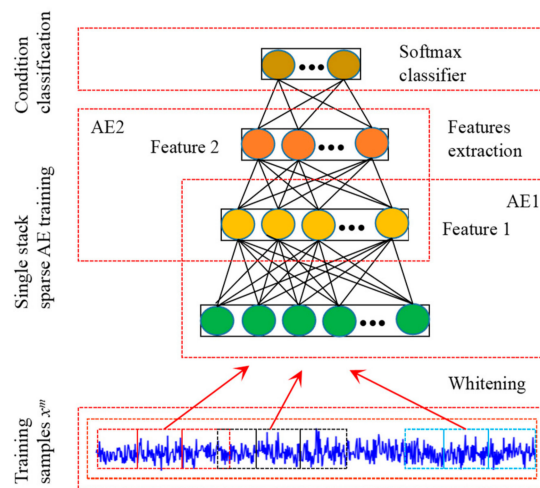


Figure 3. Illustration of the SAE training process.

When the training of individual SAE is completed, removing softmax classifier, and reproduce the parameters of the feature extraction part to other two SAEs. Then, keeping the weight $w_1^{(1)}$ unchanged, and add a small variety to the weight $w_2^{(1)}$ of other two SAEs. This not only can extract representative features from raw data and increase the diversity between features, but it is also beneficial when the input samples are similar. Therefore, the proposed model is very concise, greatly reducing training time and increasing the practicality. In addition, in order to improve the robustness of the model, Gaussian white noise is added to the training samples.

3.2. Weighting Strategy

A common ensemble strategy is voting method, which has been widely applied in different ensemble learning models [40,41]. The voting method includes majority voting and weighting voting [31,32]. The majority voting is used to directly calculate the average value of features, and the advantages of this method are convenience and intelligibility. When machines work in a stable environment or without noise interference, majority voting can get good results for the mechanical fault diagnosis. Different with the majority voting, weighting voting assigns different weight for each feature. Obviously, the majority voting is a special case of the weighting voting. When the working environment of the machine changes or the signal contains a lot of noise, the weighting voting has better performance than the majority voting. There are other ensemble methods for integration features, such as the learning method which outputs features to form a new data set, and learning with a new model [42,43].

In this paper, in order to improve the robustness and stability of the proposed method, we select the weight voting method to design an integration strategy. As show in Figure 4, assuming that the output features of the three SAEs are $\{a_n^1, a_n^2, a_n^3\}$, their distances to the expectation \bar{a}_n are d_1, d_2, d_3 , respectively. When the distance is larger, it means that the feature deviates from the category, and the lower accuracy will be obtained for fault diagnosis. Therefore, we select the distance metric of the three features to the expectation to measure the weight, such as, the larger the distance, the smaller

the weight, and vice versa. Suppose the weights of the three SAEs based on distance metric are $\alpha' = [\alpha'_1, \alpha'_2, \alpha'_3]^T$, mathematically, it can be written as:

$$\alpha' = \frac{1}{\sum_{i=1}^{N_s} \|a_n^i - \bar{a}_n\|_2} \begin{bmatrix} \sum_{i=2}^{N_s} \|a_n^i - \bar{a}_n\|_2 \\ \sum_{i=1, i \neq 2}^{N_s} \|a_n^i - \bar{a}_n\|_2 \\ \sum_{i=1, i \neq 3}^{N_s} \|a_n^i - \bar{a}_n\|_2 \end{bmatrix} \quad (10)$$

where N_s denotes the number of input segments; n is the n th hidden layer; $\|\cdot\|_2$ is the Euclidean distance; denominator $\sum_{i=1}^{N_s} \|a_n^i - \bar{a}_n\|_2$ is to normalize the weight distribution.

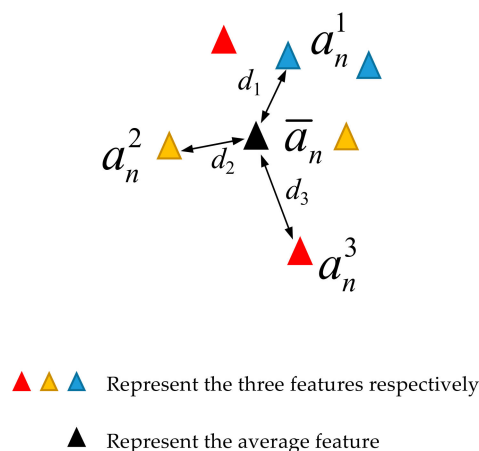


Figure 4. Illustration of the weight selection.

It should be noticed that, although the distance metric can constrain the feature deviation from the average feature on the same faulty category, when the feature itself has a large deviation, the distance metric may not have a good function. Based on this, we introduce the second weight measure condition: standard deviation metric. Standard deviation can reflect the degree of data fluctuation, the larger the standard deviation gets, the greater the data fluctuation is. So, the standard deviation can reflect the stability of features. Suppose that the weights of the three SAEs based on standard deviation metric are $\alpha'' = [\alpha''_1, \alpha''_2, \alpha''_3]^T$, they are defined by:

$$\alpha'' = \frac{1}{\sum_{i=1}^{N_s} \rho_i} \begin{bmatrix} \sum_{i=2}^{N_s} \rho_i \\ \sum_{i=1, i \neq 2}^{N_s} \rho_i \\ \sum_{i=1, i \neq 3}^{N_s} \rho_i \end{bmatrix} \quad (11)$$

where ρ denotes the standard deviation of each feature; Denominator $\sum_{i=1}^{N_s} \rho_i$ is to normalize the weight distribution.

Now we have two feature-related weight vectors, distance metric weight and standard deviation metric weight. To implement an excellent integration strategy, we assume the target weight $\alpha = [\alpha_1, \alpha_2, \alpha_3]^T$, it is defined as follow:

$$\alpha = \lambda\alpha' + \gamma\alpha'' \quad (12)$$

where λ and γ are two hyper-parameters by user-specifying, which the limits of the values are between 0 and 1 and their sum is 1. In the proposed method, the two hyper-parameters will be studied in detail for the effect of diagnosis performance.

3.3. Feature Integration

After the above analysis, the weight vector $\alpha = [\alpha_1, \alpha_2, \alpha_3]^T$ can be determined for each sample x^m . Meanwhile, three feature vectors $[f_1, f_2, f_3]$ are extracted from the input sample x^m by the three SAEs. The final object features f^m are aggregated using the weight strategy, which is written as follows:

$$f^m = \alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3 \quad (13)$$

This weighted strategy is beneficial that can decrease the influence of the random features caused by ambient noise and interference. Also, the weighted way enhances the discriminative features that these features are complementary and improves the stability due to having the weight constraint term. The detailed process of the proposed method given as Figure 5.

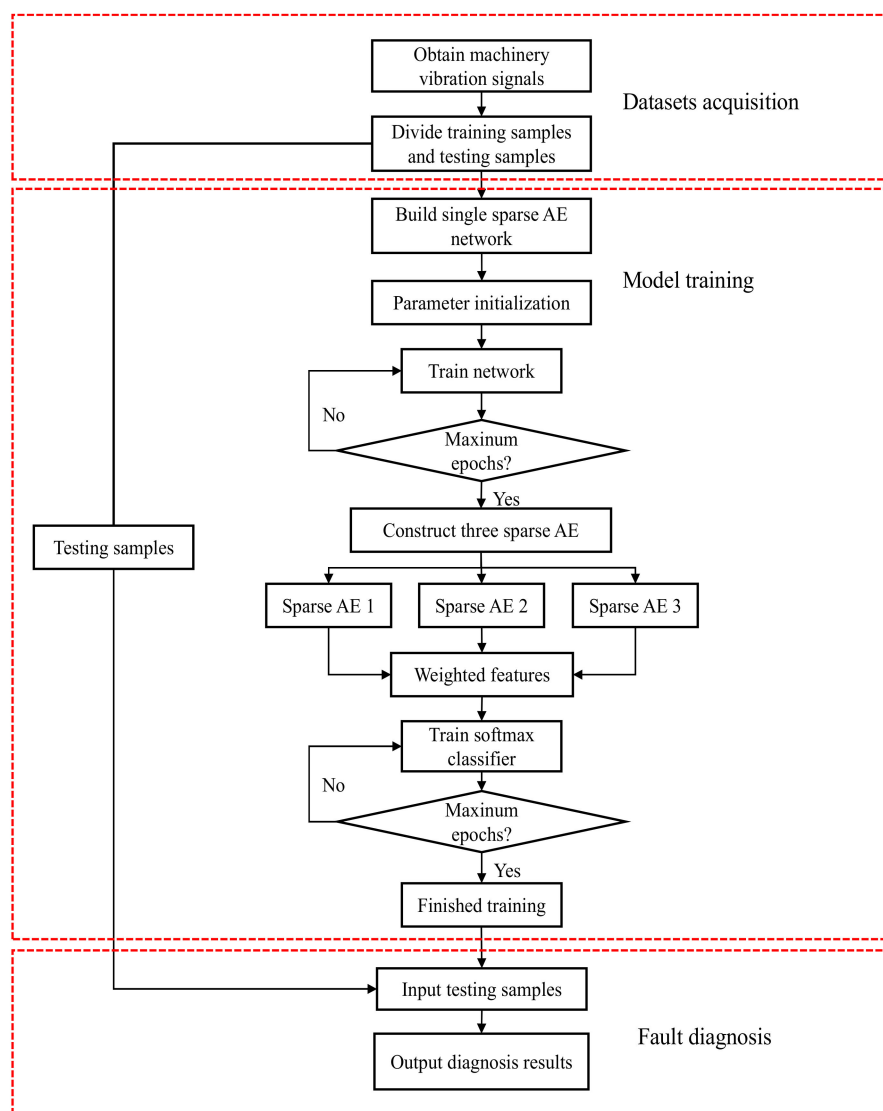


Figure 5. Flow chat of the proposed bearing fault diagnosis method.

4. Experiment and Analysis

4.1. Dataset Description

The bearing dataset provided by Case Western Reserve University [44] is analyzed in this section. As shown in Figure 6, the test rig mainly consists of a 2-horsepower (hp) motor, a torque converter/encoder, a dynamometer and a control circuit. The vibration data were collected from the drive end of a motor under four different conditions: normal condition, inner race fault (IF), roller fault (RF), and outer race fault (OF). Single point faults were introduced on the motor with fault diameters of 0.18 mm, 0.36 mm, and 0.54 mm, respectively. The bearing data were all collected under four load conditions (0, 1, 2, and 3 hp) with the sampling frequency of 12 kHz.

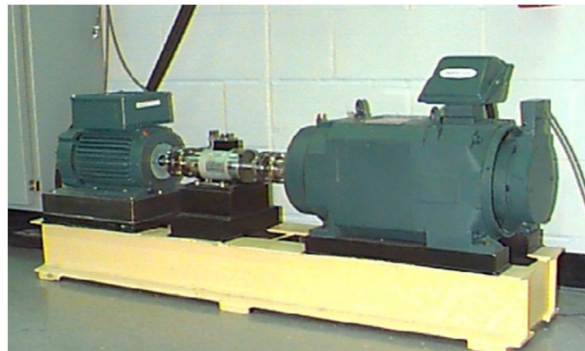


Figure 6. Bearing platform used for experiment.

These vibration data compose the motor bearing dataset, which is used to verify the effectiveness of the proposed method. These data contain ten bearing health conditions under four loads, where the same health condition under different loads is defined as one class. The details of the experimental condition are summarized in Table 1. In this experiment, the first 120,000 points of the vibration data are selected as the preprocessed data under each condition. These preprocessed data are divided into training set and test set.

Table 1. Bearing data information used to experiment in this proposed.

Fault Type	Fault Size (mm)	Load (hp)	Label
Normal	0.0	0,1,2,3	1
IF	0.18	0,1,2,3	2
IF	0.36	0,1,2,3	3
IF	0.53	0,1,2,3	4
RF	0.18	0,1,2,3	5
RF	0.36	0,1,2,3	6
RF	0.53	0,1,2,3	7
OF	0.18	0,1,2,3	8
OF	0.36	0,1,2,3	9
OF	0.53	0,1,2,3	10

4.2. Compare Studies

In order to verify the superiority of the proposed method, three methods were selected to compare with the proposed method, namely, Support Vector Machine (SVM), Back-Propagation Neural Network (BPNN) with two hidden layers, and the individual stack sparse AE with two hidden layers. They are widely used in fault diagnosis of rotating machinery. The input data is raw vibration data, and the comparison of diagnosis performance of the three methods under 20 experiments is shown in Figure 7.

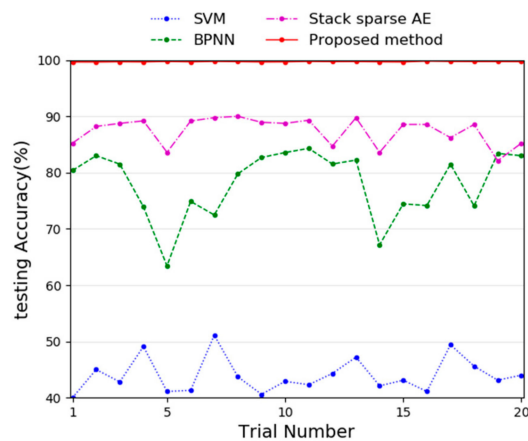


Figure 7. Comparison of different method with 20 experiments.

In the Figure 7, the results show that the proposed method has the highest diagnostic accuracy and the smallest fluctuations. Compared with the proposed method, the individual stack sparse AE has smaller diagnosis accuracy and greater diagnosis fluctuation, which indicates that the proposed ensemble method has better diagnosis performance than an individual stack sparse AE. Of course, the individual stack sparse AE is better than the other two diagnosis methods. Since the BPNN is not pre-trained like the individual stack sparse AE, it is under fitting, this proves that AE can reduce the number of training samples. SVM has minimal diagnosis accuracy, because SVM is not suitable for processing high-dimensional data, usually, it needs to preprocess the original vibration data and transform them into statistical features. The specific diagnosis results are summarized in Table 2.

Table 2. Experimental results of average accuracy and standard deviation of various methods.

Method	Average Accuracy	Standard Deviation
SVM	43.99%	3.09%
BPNN	78.07%	5.91%
SAE	87.40%	2.44%
Our method	99.71%	0.05%

From the perspective of the average accuracy in Table 2, the proposed method shows the best marks with the highest average accuracy (99.71%), while the SVM has the worst diagnosis performance (43.99%), and the individual stack sparse AE only gets intermediate testing accuracy (87.40%). In addition, from the perspective of stability, the proposed method has the optimum performance, with the smallest standard deviation (0.05%), and the BPNN with two hidden layer has the worst performance, with the largest standard deviation (5.91%).

For further proving the superiority of the proposed method, we also compared with other similar studies used the same dataset. In [45], a method adopting 15 stack sparse AEs to extract bearing features was proposed. The 15 stack sparse AEs use different activation function, and the extracted features are integrated with an accuracy threshold. This proposed method classifies the health conditions of 12 motor bearings at 0 hp, and finally obtained an average test accuracy of 97.18% and a standard deviation of 0.11%. Sun et al. [46] proposed a method based on compressed sensing theory. Their method combined with stack sparse AEs to extract features from the compressed data which were used to represent seven bearing health conditions under the load 2. The fault recognition rate of this method is 97.47% and the standard deviation is 0.43% in the bearing database. Lei et al. [47] proposed a bearing diagnosis method to integrate 12 sparse filter networks. The method used a simple average weighted combination strategy to process 12 local features that extracted from raw vibration data and white Gaussian noise is added during training. The method achieved 99.66% diagnosis accuracy and 0.19%

standard deviation under 10 fault types and 4 different loads. In method [48], a dynamic weighted average method is designed to aggregate these learned features. This method used three different deep auto-encoder to extract the features, and the accuracy of k-fold cross-validation is used as a metric to assign the weights of the three deep auto-encoders. They obtained the accuracy of 99.69% and standard deviation of 0.24%. Comparing with the above methods, the proposed method in this paper achieved the highest accuracy of fault identification and the smallest standard deviation. The results of the above comparison are displayed in Table 3.

Table 3. Performance comparison with various studies.

Method	Load(hp)	No. of Health Condition	Testing Accuracy	Standard Deviation
[45]	0	12	97.18%	0.11%
[46]	2	7	97.41%	0.43%
[47]	0,1,2,3	10	99.66%	0.19%
[48]	0,1,2,3	10	99.69%	0.24%
Proposed	0,1,2,3	10	99.71%	0.05%

4.3. Visualization of Learned Representation

In this section, to qualitatively illustrate the effectiveness of the proposed fault diagnosis method, we visualize the features using four methods. The other three methods are sparse AEs with two hidden layers, the proposed model without whitening method, and the weight average method, respectively. The visual features are extracted from testing sample by the four methods, and the experiment conducted under the condition of noise for a better visual comparison of the result.

A technique called 't-SNE' is used to map the extracted features into a two-dimensional space to achieve visualization of high-dimensional data [49]. This technique has two processes, firstly, the principal component analysis (PCA) is used to reduce the dimension of the features to 50. Then, a technology called 't-SNE' is used to represent the 50-dimensional data as two-dimensional planar data.

Figure 8 is the feature visualization of individual sparse AEs with two hidden layers. It can be seen from the figure that the individual sparse AE method performs aggregation poorly on different fault types. this method cannot correctly diagnose the bearing fault, and only 82% of the test accuracy is obtained.

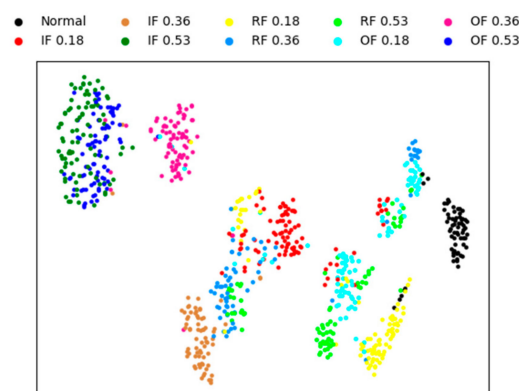


Figure 8. Visualization of sparse AE for the learned features.

Figure 9 is the feature visualization of the proposed method without whitening. Comparing with Figure 8, it can be noticed that most of the testing data are clustered in their own category and different types of faults are scattered in different regions. In Figure 9, there are only less intersection between the different fault classes, and the mainly error of fault diagnosis is concentrated in IF 0.18, that is mean that the proposed method cannot completely classify IF 0.18. The distances between different

classes are far away, which also shows that the proposed method is robust. The final test accuracy of this method is 96.93%.

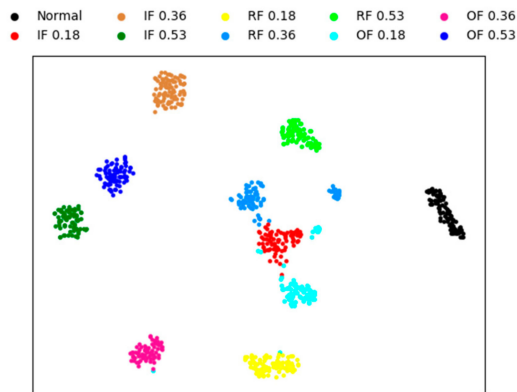


Figure 9. Visualization of without whitening for the learned features.

Figures 10 and 11 shows the feature visualization of the weight average method and our method, respectively. Figures 10 and 11 are very similar, which verifies that the features extracted by the three SAEs are similar, weighting strategy only a fine-tuning operation. Although the difference is not great, the proposed method has better performance than the weight average method, with the fault identification accuracy of 97.78% and 98.23% respectively obtained.

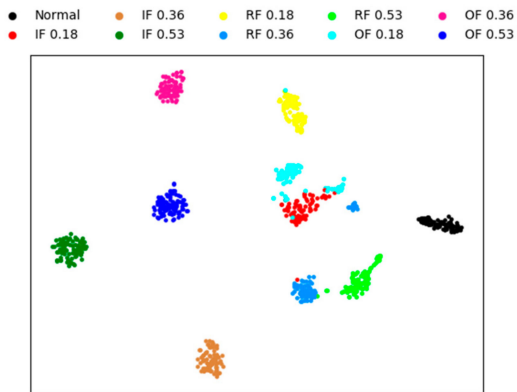


Figure 10. Visualization of weight average for the learned features.

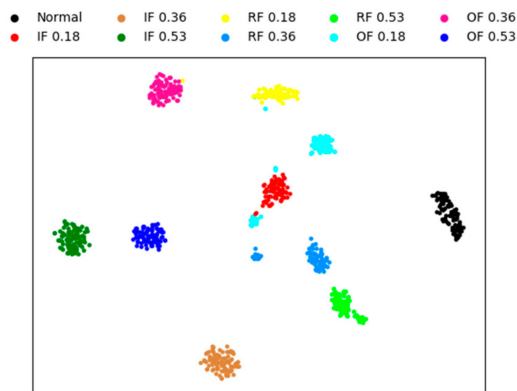


Figure 11. Visualization of proposed method for the learned features.

4.4. Parameters Selection of the Proposed Method

There are several key parameters need to determine in the proposed method, such as: the input dimension of SAE, the number of hidden layer neurons and sparse parameters ρ , etc. Next, we will

respectively investigate the selection of these parameters. In addition, in order to reduce the influence of the randomness, 20 trials are repeated for each experiment. The environment of all experiments are 4G RAM and python 3.6.

First, we investigate the selection of the input dimension. We select a certain number of samples to train the proposed method, where 40,000 samples are sampled from the bearing dataset, and the rest samples are used for testing. For each trial of different input dimension, we always keep other parameters unchanged. The diagnosis results are displayed in Figure 12, wherein the positive error bars show the standard deviations and the point of time are the average time. It can be seen that when the input dimensions are increasing from 100 to 300, the accuracies are going higher, and when the input dimension is 300, the standard deviation is the smallest. When the input dimension is greater than 300, the average test accuracy only decreases slightly, but the time consumption is growing linearly. Therefore, considering the results from the experiment, we choose 300 as the input dimension.

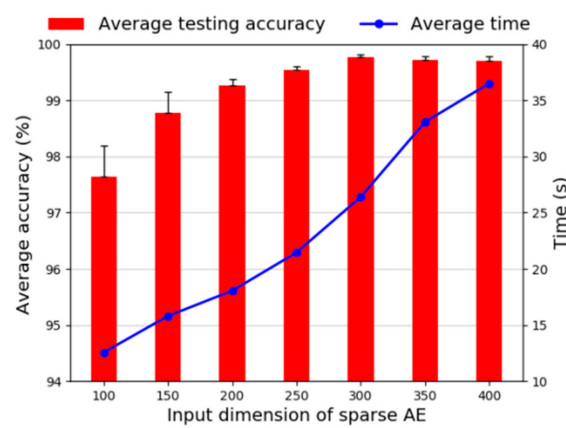


Figure 12. Diagnosis accuracy of various input dimension.

Next we investigate the number of the first hidden layer neurons. As shown Figure 13, the fault recognition accuracy increases gradually, and standard deviation is also reduced as the number of neurons in the first hidden layer increases from 50 to 200. When the number of neurons is greater than 200, the accuracy is stable and corresponding standard deviations are higher. The average time is also increasing. So, we choose 200 as the number of the first hidden layer neurons.

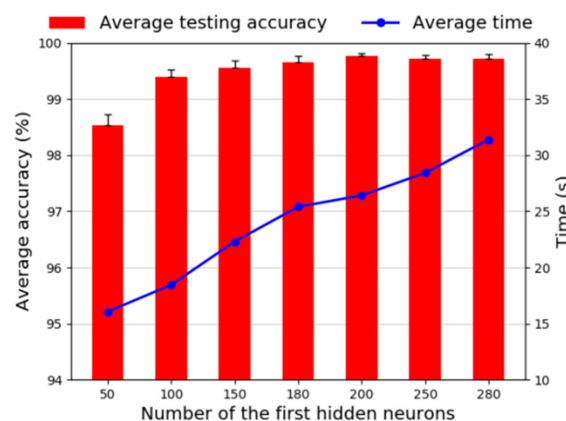


Figure 13. Diagnosis accuracy of neurons in the first layer.

Then, we investigate the number of the second hidden layer neurons. Generally, the number of neurons in the second hidden layer is less than the first hidden layer. Therefore, the number of neurons we studied was between 40 and 200, and the diagnosis results are shown in Figure 14. It can be seen that the accuracy varies only slightly in the whole neural unit interval. When the number of

neurons is 100, the average testing accuracy is highest and standard deviation is smallest. Although time consumption has increased, the increasing values are far from acceptable. Therefore, we choose 100 as the number of the second hidden layer neurons.

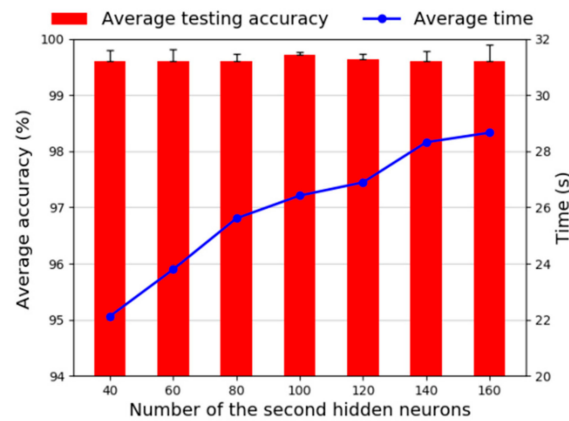


Figure 14. Diagnosis accuracy of neurons number in the second layer.

Afterward we investigate the selection of sparse parameter ρ . The sparse parameter plays an important role in the process of achieving high accuracy. In general, it is a small value close to zero. According to the general experiment results, the selection of sparse parameter varying from 0.05 to 0.5 is studied. Figure 15 shows the average diagnosis accuracy with different sparse parameters. It can be seen from the figure that, as the value of sparse factor is 0.15, the highest average test accuracy and smallest standard deviation are obtained. Therefore, 0.15 is chosen as the value of the sparse parameter.

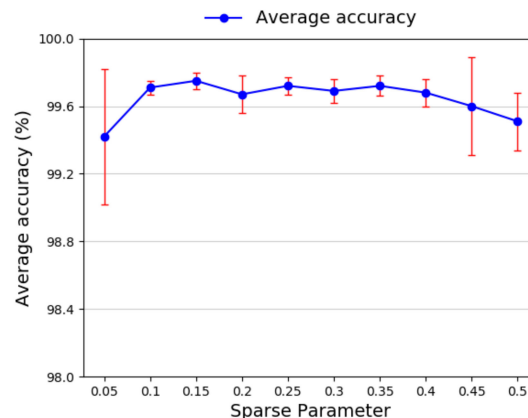


Figure 15. Diagnosis accuracy of different sparse parameters.

Finally, we investigate the selection of parameters λ and γ . These two parameters are the hyper-parameters of distance metric and standard metric. The diagnosis results are shown in Figure 16, where the two hyper-parameters λ and γ are selected as $[0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1]$ and $[1,0.9,0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1,0]$, respectively. It can be known that when the model is determined, the average diagnosis accuracy slightly changes as the two parameters correspond to different values, which means that the two hyper-parameters only have the function of fine-tuning. The highest accuracy can be obtained when the parameter λ is between 0.4 and 0.5. Therefore, it is reasonable that λ chooses 0.4 or 0.5. Furthermore, when the parameter γ is greater than λ , the average accuracy is higher, which indicates that the proposed method prefers to choose the standard deviation metric and the distance metric has a smaller impact.

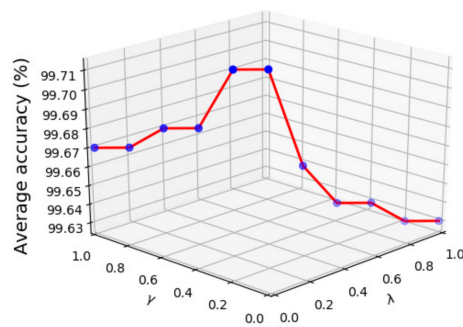


Figure 16. The relationship between average accuracy and parameters λ , γ .

In summary, the detailed parameters of the proposed method are as follows, the input dimension of sparse AE is 300, the number of two hidden layer neurons is 200, 100, respectively, and the value of sparsely parameter ρ is 0.15, the two hyper-parameters λ γ are 0.5. A wider selection of these parameters in the proposed method are listed in Table 4.

Table 4. Key parameters of the proposed method.

Parameters Description	Value
The dimension of Sparse AE	300
The number of the hidden layers	2
The number of the first hidden neurons	200
The number of the second hidden neurons	100
Learning rate	0.007
Sparse parameter	0.15
Sparse penalty factor	2
Batch size	100
Hyper-parameters (λ , γ)	0.5

4.5. Effect of Segments and Training Samples

The proposed model involves different number of segments and training samples, i.e., a different number of segments for training input and the percentage of training samples for training the proposed model will both significantly impact the diagnosis accuracy and time consumption of the proposed method. Therefore, we study the effect of different number of segments and training samples.

(1) Effect of segments: Different segments determine the structure and diagnosis performance of the model. In this study, signals with different segments will be used as the input for the proposed framework. In order to quantitatively evaluate the effect of the input segments on the classification performance, different segments ranging from one to four are studied. Figure 17 shows the diagnosis accuracy and training time choosing various segments. It is easily observed that when the segment number goes from one to four, the superior diagnosis performance is obtained. The results indicate that the more segments are used, the proposed model can achieve better and more stable performance, because these extracted features from different segments are rich and complementary, it is helpful for classification. Furthermore, a significant accuracy increase and standard deviation decrease from one segment to two segments can be noticed. More segments can achieve better diagnosis performance, however, in reality, it does not mean that more segments are always beneficial, from the comprehensive consideration of the model complexity and computational cost, choosing three segments are reasonable. Table 5 lists the diagnosis performance of different segments corresponding to Figure 17. This result validates that the proposed method can extract more discriminative and stable features from raw vibration signals.

(2) Effect of training samples: In general, as more samples are used to train the model, the higher accuracy can be achieved. The diagnosis results using different percentage of training samples are

shown in Figure 18. It can be seen that when the training samples goes larger, the average test accuracy is higher, and the standard deviation is smaller. However, the time consumption is increasing linearly. It means that the selection of training samples is a trade-off between the diagnosis accuracy and the time consumption. The same is true for Figures 12–14 and Figure 17. In Figure 18, when the proportion of training samples is 40%, the average test accuracy is 99.71% and the standard deviation is only 0.05%, which means that our proposed method achieves very high diagnosis accuracy and has good stability.

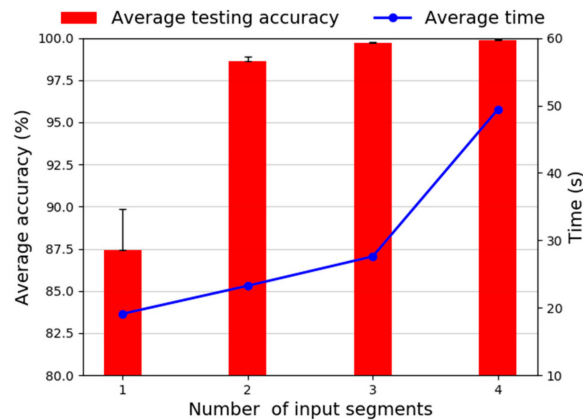


Figure 17. Diagnosis performance with different segments from 1 to 4.

Table 5. Diagnosis accuracy and time consuming for different segments.

Segments	Average Accuracy	Standard Deviation	Training Time (s)	Testing Time (s)
1	87.40%	2.44%	19.11	0.28
2	98.62%	0.23%	23.27	0.36
3	99.71%	0.05%	27.62	0.44
4	99.88%	0.04%	49.44	0.50

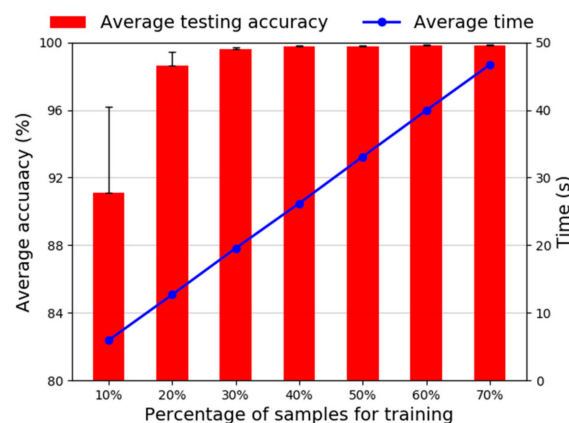


Figure 18. Diagnosis accuracy of different percentage of training samples.

4.6. Robustness Against Environmental Noises

In the actual industrial production process, noise is everywhere. The raw vibration signals are collected often contain a lot of noise, which has complex variability. For all possible noise, we can't get all the label samples corresponding to noises. So, in this section, we will study the effect of noise on diagnosis performance by adding Gaussian white noise. The robustness of the proposed method against environmental noise is verified by adding noise to the test data based on the original

experiments. Specifically, the noise data is generated by adding Gaussian white noise with different signal-to-noise ratio (SNR) to the test data. The signal-to-noise ratio is defined as

$$SNR = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \quad (14)$$

where P_{signal} and P_{noise} represent the power of the original signal and added noise, respectively, the unit of SNR is dB. In this study, we evaluate the proposed method adding noisy signals with different SNR ranging from 0 dB to 8 dB. The results are shown in Figure 19.

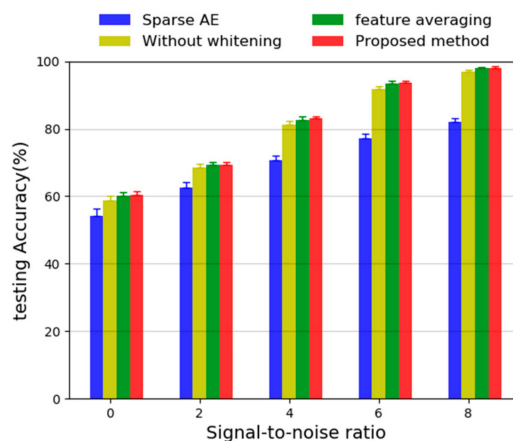


Figure 19. Diagnosis results with environmental noises by different methods.

It can be seen from the figure that when the SNR increases from 0 dB to 8 dB, the test accuracy of the four methods is increasing. Among them, at each SNR, the proposed method has the highest accuracy and the smallest deviation, the second is the method that using feature averaging method and next is the method without adding Gaussian white noise to the training data, the individual sparse AE with two hidden layer gets worst performance. Compared with the individual sparse AE, it is obvious that the proposed method has better anti-noise performance. In addition, it can be noticed that the proposed method is only slightly better (it is about 0.5%) than the feature average method; it infers that the input data has the same distribution, also, the feature average method can be considered as a sample mean filter, thus random noises will be filtered to some extent. Maybe, when the vibration signals are collected by multiple sensors or the distribution of input data is different, the proposed method may achieve better performance.

5. Concluding Remarks

In this paper, a novel bearing fault diagnosis method based on ensemble stack sparse auto-encoder was proposed. A common bearing data set is used, and a large number of experiments are carried out to verify the effectiveness of the proposed method. This paper studies the selection of several key parameters and the influence of segments and training samples on the diagnosis performance. By a comparison with other methods and related studies using the same data set, the superiority of the proposed method is proved. Additionally, the robustness of the proposed method against environmental noises is demonstrated under different levels of noise.

Future research will be extended to other complex models and other fault diagnosis problems such as using a CNN model and remaining useful-life prediction for rolling bearings. In addition, although the proposed method in this paper obtained a high accuracy of fault recognition, it did not achieve satisfactory results in a noisy environment, and there is still much room for improvement. This is one of the future research directions on how to improve the anti-noise performance of the model.

Author Contributions: Writing—original draft, J.H.; data curation, M.O.; supervision, C.Y.; funding acquisition, D.C., J.G. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The research is supported by National Natural Science Foundation of China (Grant No.61703104, 61803087, 61972091), by Guangdong Natural Science Fund (Grant No. 2017A030310580, 2017A030313388), by Key Project of University of Guangdong Province (Grant No. 2019KZDXM007).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lei, Y.; Li, N.; Guo, L.; Li, N.; Yan, T.; Lin, J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 799–834. [[CrossRef](#)]
2. Lu, C.; Wang, Z.-Y.; Qin, W.-L.; Ma, J. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Process.* **2017**, *130*, 377–388. [[CrossRef](#)]
3. Li, X.; Zhang, W.; Ding, Q. A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning. *Neurocomputing* **2018**, *310*, 77–95. [[CrossRef](#)]
4. Xu, Y.; Sun, Y.; Wan, J.; Liu, X.; Song, Z. Industrial big data for fault diagnosis: Taxonomy, review, and applications. *IEEE Access* **2017**, *5*, 17368–17380. [[CrossRef](#)]
5. Gao, Z.; Cecati, C.; Ding, S.X. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3757–3767. [[CrossRef](#)]
6. Gao, Z.; Ding, S.X.; Cecati, C. Real-time fault diagnosis and fault-tolerant control. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3752–3756. [[CrossRef](#)]
7. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **2016**, *72*, 303–315. [[CrossRef](#)]
8. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [[CrossRef](#)]
9. Feng, Z.; Ma, H.; Zuo, M.J. Vibration signal models for fault diagnosis of planet bearings. *J. Sound Vib.* **2016**, *370*, 372–393. [[CrossRef](#)]
10. Rai, A.; Upadhyay, S. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **2016**, *96*, 289–306. [[CrossRef](#)]
11. Bustos, A.; Rubio, H.; Castejón, C.; García-Prada, J.C.J. EMD-based methodology for the identification of a high-speed train running in a gear operating state. *Sensors* **2018**, *18*, 793. [[CrossRef](#)] [[PubMed](#)]
12. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15. [[CrossRef](#)]
13. Bustos, A.; Rubio, H.; Castejon, C.; Garcia-Prada, J.C. Condition monitoring of critical mechanical elements through Graphical Representation of State Configurations and Chromogram of Bands of Frequency. *Measurement* **2019**, *135*, 71–82. [[CrossRef](#)]
14. Li, X.; Yang, Y.; Pan, H.; Cheng, J.; Cheng, J.J.M.; Theory, M. Non-parallel least squares support matrix machine for rolling bearing fault diagnosis. *Mech. Mach. Theory* **2020**, *145*, 103676. [[CrossRef](#)]
15. Ali, J.B.; Fnaiech, N.; Saidi, L.; Chebel-Morello, B.; Fnaiech, F. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl. Acoust.* **2015**, *89*, 16–27.
16. Lee, J.; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L.; Siegel, D. Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Mech. Syst. Signal Process.* **2014**, *42*, 314–334. [[CrossRef](#)]
17. Fu, W.; Shao, K.; Tan, J.; Wang, K.J. Fault diagnosis for rolling bearings based on composite multiscale fine-sorted dispersion entropy and SVM with hybrid mutation SCA-HHO algorithm optimization. *IEEE Access* **2020**, *8*, 13086–13104. [[CrossRef](#)]
18. He, C.; Liu, C.; Wu, T.; Xu, Y.; Wu, Y.; Chen, T. Medical rolling bearing fault prognostics based on improved extreme learning machine. *J. Comb. Optim.* **2019**, 1–22. [[CrossRef](#)]

19. Kang, M.; Kim, J.; Kim, J.-M.; Tan, A.C.; Kim, E.Y.; Choi, B.-K. Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis. *IEEE Trans. Power Electron.* **2014**, *30*, 2786–2797. [[CrossRef](#)]
20. Cao, H.; Fan, F.; Zhou, K.; He, Z. Wheel-bearing fault diagnosis of trains using empirical wavelet transform. *Measurement* **2016**, *82*, 439–449. [[CrossRef](#)]
21. Huang, S.; Tan, K.K.; Lee, T.H. Fault diagnosis and fault-tolerant control in linear drives using the Kalman filter. *IEEE Trans. Ind. Electron.* **2012**, *59*, 4285–4292. [[CrossRef](#)]
22. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Trans. Ind. Electron.* **2017**, *65*, 5990–5998. [[CrossRef](#)]
23. Khan, S.; Yairi, T. A review on the application of deep learning in system health management. *Mech. Syst. Signal Process.* **2018**, *107*, 241–265. [[CrossRef](#)]
24. Liu, R.; Meng, G.; Yang, B.; Sun, C.; Chen, X. Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine. *IEEE Trans. Ind. Inform.* **2016**, *13*, 1310–1320. [[CrossRef](#)]
25. Jia, F.; Lei, Y.; Guo, L.; Lin, J.; Xing, S. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. *Neurocomputing* **2018**, *272*, 619–628. [[CrossRef](#)]
26. Shao, H.; Jiang, H.; Zhang, H.; Duan, W.; Liang, T.; Wu, S. Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing. *Mech. Syst. Signal Process.* **2018**, *100*, 743–765. [[CrossRef](#)]
27. Li, X.; Zhang, W.; Ding, Q. Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks. *IEEE Trans. Ind. Electron.* **2018**, *66*, 5525–5534. [[CrossRef](#)]
28. Long, J.; Zhang, S.; Li, C. Evolving Deep Echo State Networks for Intelligent Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2019**. [[CrossRef](#)]
29. Mao, W.; He, L.; Yan, Y.; Wang, J. Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine. *Mech. Syst. Signal Process.* **2017**, *83*, 450–473. [[CrossRef](#)]
30. Ma, S.; Chu, F. Ensemble deep learning-based fault diagnosis of rotor bearing systems. *Comput. Ind.* **2019**, *105*, 143–152. [[CrossRef](#)]
31. Rokach, L. Ensemble-based classifiers. *Artif. Intell. Rev.* **2010**, *33*, 1–39. [[CrossRef](#)]
32. Polikar, R. Ensemble based systems in decision making. *IEEE Circuits Syst. Mag.* **2006**, *6*, 21–45. [[CrossRef](#)]
33. Qi, Y.; Shen, C.; Wang, D.; Shi, J.; Jiang, X.; Zhu, Z. Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery. *IEEE Access* **2017**, *5*, 15066–15079. [[CrossRef](#)]
34. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [[CrossRef](#)] [[PubMed](#)]
35. Hinton, G.E.; Osindero, S.; Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)]
36. Makhzani, A.; Frey, B. K-sparse autoencoders. *arXiv* **2013**, arXiv:1312.5663.
37. Ng, A. Sparse autoencoder. *CS294A Lect. Notes* **2011**, *72*, 1–19.
38. Erhan, D.; Bengio, Y.; Courville, A.; Manzagol, P.-A.; Vincent, P.; Bengio, S. Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.* **2010**, *11*, 625–660.
39. Tao, S.; Zhang, T.; Yang, J.; Wang, X.; Lu, W. Bearing fault diagnosis method based on stacked autoencoder and softmax regression. In Proceedings of the 2015 34th Chinese Control Conference (CCC), Hangzhou, China, 28–30 July 2015; pp. 6331–6335.
40. Zheng, J.; Pan, H.; Cheng, J. Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mech. Syst. Signal Process.* **2017**, *85*, 746–759. [[CrossRef](#)]
41. Li, S.; Liu, G.; Tang, X.; Lu, J.; Hu, J. An ensemble deep convolutional neural network model with improved DS evidence fusion for bearing fault diagnosis. *Sensors* **2017**, *17*, 1729. [[CrossRef](#)]
42. Zhou, F.; Hu, P.; Yang, S.; Wen, C. A Multimodal Feature Fusion-Based Deep Learning Method for Online Fault Diagnosis of Rotating Machinery. *Sensors* **2018**, *18*, 3521. [[CrossRef](#)]
43. Jiang, G.; He, H.; Yan, J.; Xie, P. Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. *IEEE Trans. Ind. Electron.* **2018**, *66*, 3196–3207. [[CrossRef](#)]
44. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64*, 100–131. [[CrossRef](#)]

45. Shao, H.; Jiang, H.; Lin, Y.; Li, X. A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. *Mech. Syst. Signal Process.* **2018**, *102*, 278–297. [[CrossRef](#)]
46. Sun, J.; Yan, C.; Wen, J. Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. *IEEE Trans. Instrum. Meas.* **2017**, *67*, 185–195. [[CrossRef](#)]
47. Lei, Y.; Jia, F.; Lin, J.; Xing, S.; Ding, S.X. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3137–3147. [[CrossRef](#)]
48. Zhang, Y.; Li, X.; Gao, L.; Chen, W.; Li, P. Intelligent fault diagnosis of rotating machinery using a new ensemble deep auto-encoder method. *Measurement* **2020**, *151*, 107232. [[CrossRef](#)]
49. Maaten, L.V.d.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).