

## Research Article

# A Denoising Autoencoder-Based Bearing Fault Diagnosis System for Time-Domain Vibration Signals

Yi Gu , Jiawei Cao, Xin Song, and Jian Yao

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214122, China

Correspondence should be addressed to Yi Gu; 8202101437@jiangnan.edu.cn

Received 6 April 2021; Revised 27 April 2021; Accepted 4 May 2021; Published 15 May 2021

Academic Editor: Shan Zhong

Copyright © 2021 Yi Gu 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.

The condition monitoring of rotating machinery is always a focus of intelligent fault diagnosis. In view of the traditional methods' excessive dependence on prior knowledge to manually extract features, their limited capacity to learn complex nonlinear relations in fault signals and the mixing of the collected signals with environmental noise in the course of the work of rotating machines, this article proposes a novel approach for detecting the bearing fault, which is based on deep learning. To effectively detect, locate, and identify faults in rolling bearings, a stacked noise reduction autoencoder is utilized for abstracting characteristic from the original vibration of signals, and then, the characteristic is provided as input for backpropagation (BP) network classifier. The results output by this classifier represent different fault categories. Experimental results obtained on rolling bearing datasets show that this method can be used to effectively diagnose bearing faults based on original time-domain signals.

## 1. Introduction

Rolling bearings are the modules that have frequent usage in mechanical equipment and are also among the most vulnerable components. When a fault occurs, it often results in enormous losses; consequently, bearing fault diagnosis technology is receiving increasing attention. Statistics show that around 30% of machinery failures arise from bearings [1, 2]. To comprehensively detect the health status of rotating machinery, a condition monitoring system is used for data collection. After a long period of operation, such a machine will have generated a large amount of real-time data. We can obtain the features that are extracted from the monitored data through constructing feature engineering. These data can be used to establish a classifier that different bearing faults can be identified [3, 4]. Therefore, detecting bearing faults is essentially the task of classification [5–7], and the health status of bearings can be identified through such artificial intelligence techniques as deep learning models [8–10], support vector machines (SVMs) [11], neural networks (NNs) [12, 13], extreme learning machines (ELMs) [14], and autoencoders (AEs) [15–17]. In particular, in recent

years, deep learning technology has been widely used in this field and achieved good results [18, 19].

Most of the common methods used to detect bearing faults at present are supervised algorithms where labor label cost is pretty high. Because of the powerful function of hidden layers [20–22], the algorithm for training the classifier that detects bearing faults is NNs commonly. However, high-performance neural network classifiers also rely on numerous good-quality labeled data. But overfitting of NNs may occur when the samples used for training is noisy, or the amount is not enough, or the test distribution cannot be covered. Thus, the ability of NNs to generalize becomes poor particularly for complicated classification problems. Diagnosing bearing faults is challenging because the subject to be solved is a sophisticated mechanical instrument [23, 24]. Hence, lots of approaches have been put forward to figure this issue out, like senior techniques of processing the signals that are used for analyzing the vibration of signal to enable the extraction of helpful bearing fault features. But these advanced techniques require practitioners to have a deep knowledge reserve of vibration signal and the whole mechanical systems. Moreover, because it is difficult to obtain the

professional knowledge related to this, these methods are not universal, meaning that they are not smart enough compared to machine learning. As an alternative and with the power of deep learning, deep neural networks (DNNs) are proposed to achieve unsupervised learning of characteristic. In recent years, considering the impressive achievements in the area of image recognition, deep learning has become the most popular and promising research methods. Among them, deep AE (autoencoder) network structures, as one of the representatives of unsupervised learning for detecting bearing faults, have become a common solution. In addition, as a well-known unsupervised feature learning method, denoising self-encoding is widely used for the realization of various tasks because it can be used to learn more robust feature expressions for input signals, that is, it has a strong generalization ability [25–27].

## 2. Brief Introduction to DNNs

DNN is usually a nonlinear converter with multilayer structures. In general, it contains an input layer, an output layer, and several layers in the middle. The parameter values of each layer in DNN can be learned by the training mechanism of the neural network, as Hinton et al. said [10]. By comparison with smaller NNs, DNNs can express more information. With the power of deep networks, we propose to use a SAE-based (stacked autoencoder) method to automatically classify bearing faults through feature selection without any manual intervention. This method consists of three steps. Step one: use unsupervised learning to train the entire model gradually. Step two: use supervised learning to fine-tune this network with the training data which contain true labels. Step three: use the network to perform feature extraction and classification of the bearing data after completing the training of this network.

A DAE (denoising autoencoder) [25] is a variant of an AE obtained by adding denoising technology to an AE. Traditionally, an AE comprises an encoder and a decoder. Besides, AE is also one of the representative methods of unsupervised learning. The principle that AE can achieve unsupervised feature extraction is through continuously minimizing the deviation between the input and the reconstructed data. For the sake of enhancing the generalization performance of the entire model, we add some noise to the input layer. This approach also effectively avoids overfitting in the training of the network. Usually, a DAE is a symmetrical neural network with an odd number of layers, and its goal is to make the output and input data as consistent as possible. The purpose of the encoder of DAE is to condense the input data without losing important information as much as possible and get a concise representation of the input data. The diagram of a DAE is as seen in Figure 1.

$$f_{\{W^1, b^1\}}(x) = s_f(W^1 x + b^1), \quad (1)$$

$$g_{\{W^2, b^2\}}(x) = s_g(W^2 x + b^2). \quad (2)$$

In Figure 1, the encoder is made up by an input layer and

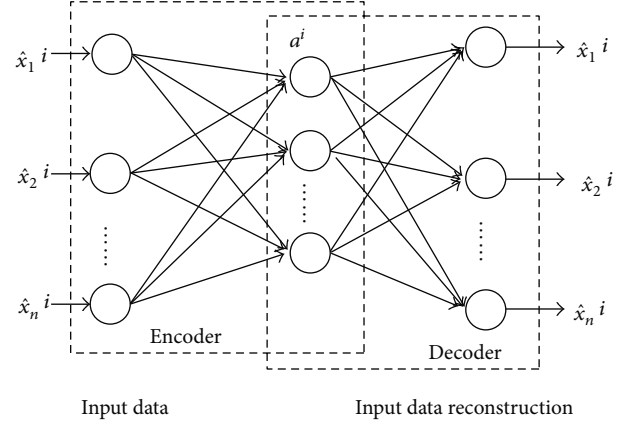


FIGURE 1: Structure graph of an AE.

parts of the layers near the input, and the decoder is composed by an output layer and parts of the layers near the output. The process that the encoder performs processing can be expressed by Equation (1), and Equation (2) represents the processing performed by the decoder, where  $s_f$  and  $s_g$ , respectively, denote as the activation functions of the coding process and the decoding process. For these activation functions, the nonlinear sigmoid function is commonly used. Suppose a dataset is represented as  $X = \{x^i\}_{i=1}^M$ , where  $M$  represents the size of  $X$ . The expression of the sample  $x^i$  in  $X$  obtained by the encoder is shown as  $a^i$ . Furthermore,  $a^i$  is also a concise expression of the input.

$$a^i = f_{\{W^1, b^1\}}(x^i). \quad (3)$$

Suppose the reconstructed error is represented as  $x \wedge^i$ . The connection between  $x \wedge^i$  and  $a^i$  is expressed as

$$x \wedge^i = g_{\{W^2, b^2\}}(a^i). \quad (4)$$

The parameters of DAE are updated iteratively during the training phase. The objective function of DAE is as follows:

$$J(W, b) = \|x - g(f(\tilde{x}))\|_2^2, \quad (5)$$

$$\tilde{x} = \text{dropout}(x).$$

In Equation (5),  $\tilde{x}$  represents the input  $x$  after adding some noise. In DAE, it is implemented through dropout technology. Therefore, DAE must be able to recover  $x$  from this pollution rather than just using the data with noise. By this means, the inner distribution of the generated data structure can be implicitly captured by the encoding function  $f(\bullet)$  and the decoding function  $g(\bullet)$  of DAE in common, thus enabling the extraction of robust features. DAE-based automatic detecting for bearing faults method.

We propose a novel automatic detecting for bearing faults method in this article. This method is based on DAE. The breakdown characteristic from primitive vibration signals of rotating equipment like bearings can be mined by this

method. By adaptively selecting these representative features, this method can automatically classify the condition category, which is the health level, of the machinery on these breakdown features. We measure the primitive signals of vibration in the time domain. The analysis of converted spectral signals of time domain and frequency domain is opposite, where frequency domain applies the fast Fourier transform (FFT). The temporal waveform determines the analysis in the time domain, and such a time-domain signal provides relatively vague information on the health of the rotating machinery. Therefore, it is very challenging to classify these original time-domain signals directly.

The proposed method contains four steps as seen in Figure 2.

Step one: acquire time-domain data from rotating equipment with various states of health. Assume that the training dataset composed of these time-domain data is denoted as  $\{x^i, d^i\}_{i=1}^M$ , where  $x^i$  represents the  $i$ -th sample in the training-phase,  $d^i$  represents the state label of health related to  $x^i$ , and  $M$  represents the amount of time-domain data.

Step two: construct a DNN which consists of several hidden layers. The amount of neurons in the first layer is identical to the dimensionality of the time-domain data  $x^i$ . Next, pretrain the DNN on each layer through SAE with the unlabeled dataset  $x = \{x^i\}_{i=1}^M$ . In the pretraining, the required number of AE is the same as that of the amount of hidden layers in DNN. The specific method of pretraining layer by layer is that firstly take the first three layers of DNN, where it consists of an input and two hidden layers, as the three-layer structure of AE1. The correspondence between above two is that the input layer of DNN is also the input layer of AE1, the hidden layer 1 of DNN is the hidden layer of AE1 and the hidden layer 2 of DNN is the output layer of AE1. The dataset of training AE1 is the unlabeled dataset  $x$ , which is also regarded as the output of AE1. After completing the training for AE1, use the parameters  $\{W^1, b^1\}$  of AE1 to finish the initialization of the parameters of hidden layer 1 of DNN. Let  $a^1$  be the encoding vector which is from time-domain data of the time rotating machinery by AE1. Then, take  $a^1$  as the input data and objective output of AE2 for training and finish the initialization of the parameters of hidden layer 2 of DNN via  $\{W^2, b^2\}$ , and obtain  $a^2$  accordingly. Finally, repeat the above steps sequentially till AEN is obtained and the time-domain data is encoded as  $a^N$ . By using the above method, we complete the pretraining of the whole of the middle layers of DNN.

Step three: the size of the last layer of DNN, which performs classification, should be consistent with the amount of different states of health that machine may have. Then, by making the reconstruction value from the original time-domain data as close to the given health status label as possible, the parameters of DNN are fine-tuned. The method to optimize the entire network is the backpropagation (BP) algorithms.

Step four: apply the trained DNN to faults diagnose of rotating equipment.

In the approach of this paper, DNN learns multiple nonlinear transformations through a pretraining process to dis-

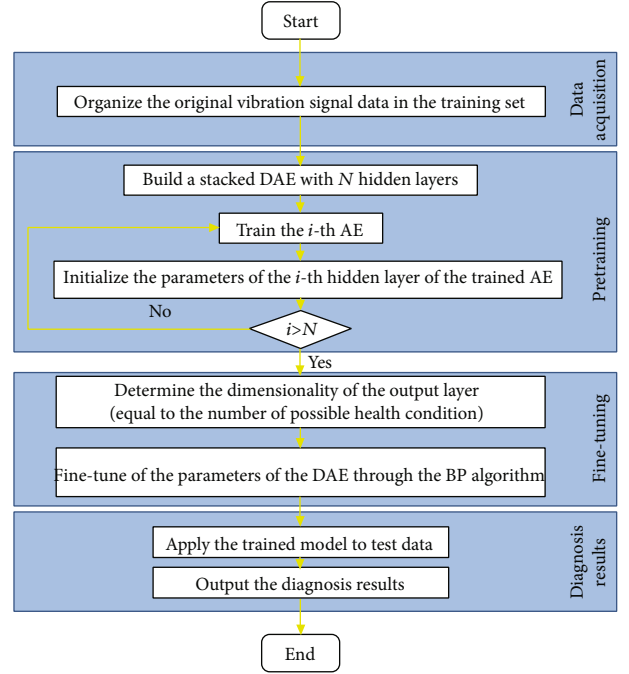


FIGURE 2: Framework of the method in this paper.

cover the main changes in the time-domain signal characteristics. In this way, DAE can extract more robust features. After that, the fine-tuning process can help DNN mine discriminative information in the vibration signal data. By these means, the well-trained DNN is capable of automatically capturing the fundamental features from the time-domain data and set up a complex nonlinear relationship between the bearing status labels and time-domain data. Thus, this method can be used to realize effective attribute selection and rapid smart detecting of rotating bearing breakdown features.

### 3. Fault Diagnosis Experiment

Rolling bearings are the core parts in rotating machinery. The performance, service life, and reliability of this type of machine usually depend on the health of these components. However, because of their harsh working environment, these components are susceptible to various damages, resulting in malfunctions and causing serious economic damage. Several diagnostic situations of rolling bearings will be considered to verify the proposed method in this part.

**3.1. Introduction to the Training Data.** The raw data of rotating bearing were gathered from a system of machine which is driven by a motor. The approach of acquisition is to utilize a specific frequency to complete the sampling, where the frequency is 12 kHz. In general, there are four ways of loading in these bearing data. Thus, the data were obtained under the following four experimental conditions: (a) normal state (N), (b) outer race fault (OF), (c) inner race fault (IF), and (d) roller fault (RF). Meanwhile, when constructing the raw data, the diameter of the failed rotating bearing is also considered into the driving termination engine. In this method,

TABLE 1: Descriptions of the training data.

Datasets	Load (hp)	The amount of samples	Fault type	Fault size	Classification label
A B C D E	0 1 2 3 0-3	600/600/600/1800	N	0''	1
		600/600/600/1800	RF	0.007''	2
		600/600/600/1800	RF	0.014''	3
		600/600/600/1800	RF	0.021''	4
		600/600/600/1800	IF	0.007''	5
		600/600/600/1800	IF	0.014''	6
		600/600/600/1800	IF	0.021''	7
		600/600/600/1800	OF	0.007''	8
		600/600/600/1800	OF	0.014''	9
		600/600/600/1800	OF	0.021''	10

TABLE 2: Diagnosis results obtained on different bearing datasets.

Dataset	Metric	ELM	SVM	BPNN1 (500)	BPNN2 (500-500)	BPNN3 (600-200-100)	SAE+BP	Proposed method
A	Acc-mean	11.23	77.07	83.97	85.96	78.01	86.32	92.29
	Acc-std	0.82	0.57	0.37	0.56	0.49	0.39	0.19
B	Acc-mean	11.6	74.47	85.77	88.58	79.87	87.14	93.32
	Acc-std	0.86	0.61	0.47	0.36	0.39	0.30	0.33
C	Acc-mean	11.70	72.69	89.28	89.24	77.57	88.92	93.76
	Acc-std	0.86	0.50	0.37	0.41	0.30	0.28	0.25
D	Acc-mean	11.37	72.45	87.67	89.57	81.13	88.70	94.06
	Acc-std	0.52	0.00	0.43	0.34	0.35	0.43	0.21
E	Acc-mean	13.10	93.34	96.11	97.84	93.55	97.79	99.04
	Acc-std	1.19	0.28	0.13	0.10	0.13	0.12	0.05

there are three sizes of the fault: 0.007 inches, 0.014 inches, and 0.021 inches. In this article, five datasets, respectively, named A, B, C, D, and E, are used for checking the classification capability of this method. The detailed description of these five datasets is seen in Table 1. Furthermore, the health states of bearing data are divided into ten different types of labels, named 1 to 10. Therefore, A, B, C, and D in the datasets cover these condition type labels with the loading of 0, 1, 2, and 3 hp, respectively. Each state of health involves 600 signals, and there are 2048 data points for every single signal.

**3.2. Experimental Setup.** The designed one side of DNN consists of four layers. The size of the first layer of DNN is identical to the feature dimension of the training data, which is 2048. As for the size of the hidden layers in the middle of the encoder, we uniformly set them to 500. The size of the last layer of the encoder depends on the categories of possible bearing states of health. Specifically, the shape of the entire DNN is 2048-500-500-10-500-500-2048, where the structure of encoder and decoder is symmetric. We use rectified linear units (ReLU) in all encoder/decoder pairs except the first (which needs to reconstruct the input data, which may have both positive and negative values) and the last (to ensure that the final encoding retains complete information). For these experiments, during greedy layerwise pretraining, the initial-

ization of the weights is set to a random number with a standard deviation of 0.01 derived from a zero-mean Gaussian distribution. Each layer was pretrained for 500 iterations with a dropout rate of 20% and a learning rate of 0.001. We select 50% of the data at random to pretrain the network, and at the same time, these samples are also utilized to optimize the entire network through fine-tuning the parameters. After that, the rest of the data (another 50% data) are used for testing the classification capability. The fine-tuning process was performed by removing the decoder from the pretrained DAE and adding a soft-max layer for BP network, and we set the maximum value of fine-tuning epochs to 500. Table 2 shows the diagnostic results of ten tests performed on each dataset in A to E using the proposed method. We are aware from the experimental results of these test sets that the accuracy of diagnose for bearing fault data is greater than 92%, and some even reach to 99%, which signifies that our method is capable of well distinguishing the status of health of ten types of bearings.

**3.3. Comparison with Traditional Methods.** Backpropagation neural networks (BPNNs) and SVMs have been widely used in fault detection for rotating equipment. In this part, we compare our method with a multiple layer architecture network, that is, BPNN, whose structure is 2048-500-500-10,

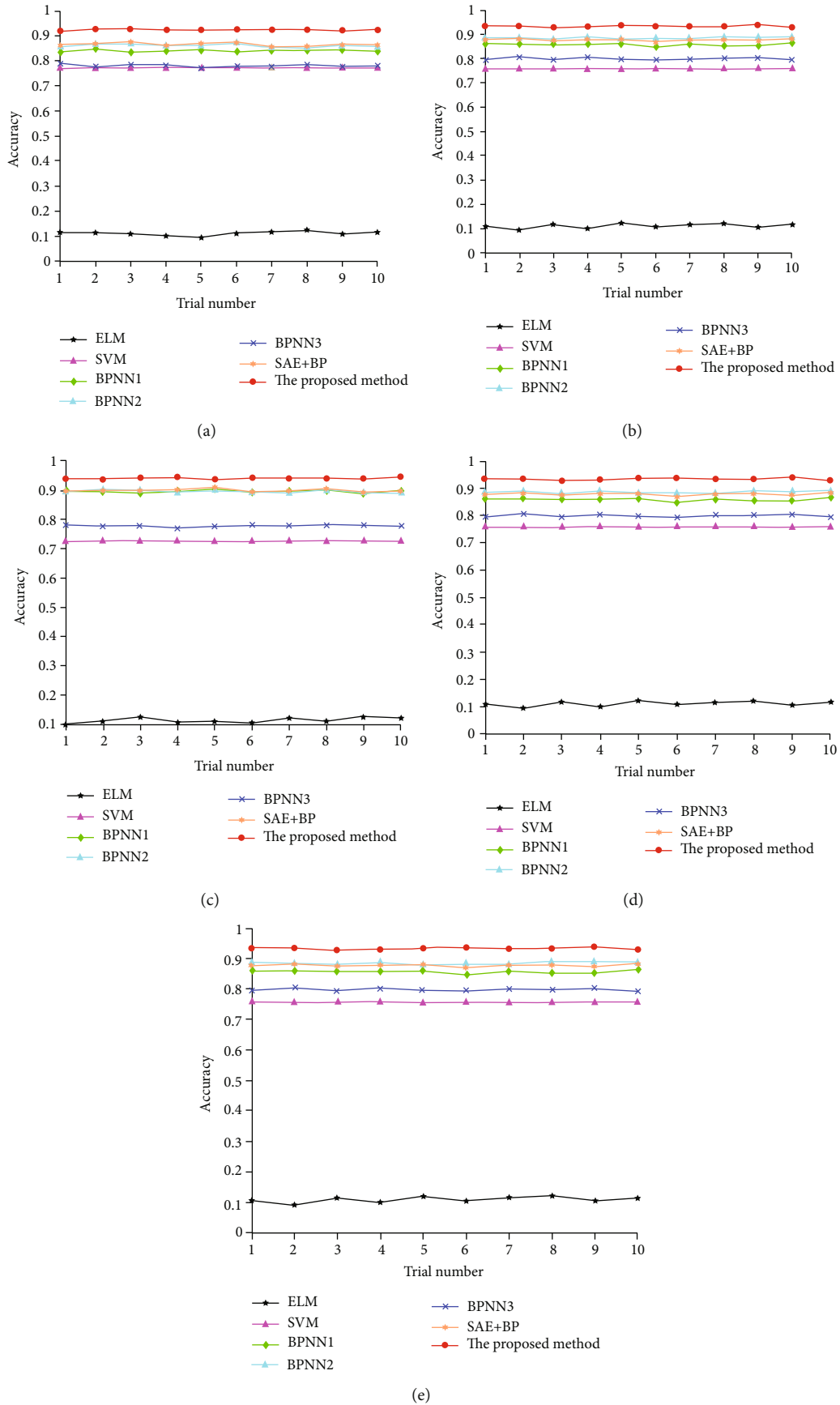


FIGURE 3: Diagnosis results from 10 trials on different bearing datasets: (a) A, (b) B, (c) C, (d) D, and (e) E.



another multi-hidden-layer BPNN used in a previous study, whose network structure is 2048-600-200-100-10, a neural network with only one hidden layer, whose network structure is 2048-500-10, an SVM-based model, an ELM, and a traditional SAE-based model with the same architecture used in the proposed method. A comparison of the diagnostic accuracy of the different methods across 10 experiments is seen in Figure 3.

The test results demonstrate that the method with the most accurate detection and the lowest volatility is our proposed method, and they also show a greater advantage in stability. Compared with the proposed method, the SVM has relatively stable classification performance and slightly greater volatility. What is more, compared with other methods, its diagnostic performance is relatively stable. Compared with our method or SVM, only one hidden layer BPNN shows worse recognition precision and greater volatility. In addition, the performance of multilayer BPNN2 is only better than that of smaller BPNN1, but the diagnostic results of the multi-hidden-layer BPNN3, with the deepest structure, are poorer. This shows that the classification accuracy of a BPNN cannot always be improved simply by increasing the number of network layers; instead, a reasonable network structure must be selected. In addition, we can see that the results obtained from the traditional SAE-based method are similar to the results of BPNN1, which has the same hidden layer structure. This indicates that even if a DNN is used for pretraining, it is not necessarily possible to effectively extract the useful features of the original time-domain data. In addition, ELM achieves the worst results in all experiments, indicating that this classification method cannot process the original time-domain data effectively. We calculated the mean diagnostic accuracies and standard deviations of more than 10 tests for quantitative comparisons. Table 2 details the experimental results.

It is worth noting that, in each trial, the same training dataset is used for checking the performance of each comparison approach. From the perspective of average classification accuracy, our proposed method can achieve the best diagnostic accuracy, while the ELM has the worst diagnostic accuracy. Moreover, our method also has the best performance in terms of classification stability. The results indicate that our method in this article is able to effectively and stably identify the type of bearing failure and the severity of damage to rotating machinery. In comparison, the results obtained by BPNN in the table are relatively poor in both average accuracy and standard deviation, which shows that our method is better than BPNN. In particular, dataset E contains a large number of samples representing the states of health of bearing data with ten different loadings, and the detecting precision of the method proposed in this paper on this dataset indicates that this approach can detect bearing breakdowns independently from load fluctuations.

The above results show that, in comparison with the conventional diagnosis approaches, our proposed approach of combining unsupervised DAE pretraining with the supervised BP algorithm has significant advantages. In this method, the DAE is trained with the reconstruction errors generated from multidimensional temporal data to explore

the relationship between time-domain data. Then, the decoder is replaced with a soft-max layer for health condition monitoring and fault diagnosis. Therefore, compared with the traditional recognition methods based on time-domain signal analysis, the proposed method can obtain higher recognition accuracy by means of a reasonable network structure and the introduction of dropout technology in the pretraining process. The characteristic structure of a time-domain vibration signal is fuzzy, and its regularity is poor. Therefore, the method of combining traditional time-domain features with an SVM classifier results in a low average diagnostic accuracy. However, a DAE is capable of deep learning from the original time-domain data, mining hidden information and identifying differences. Therefore, the DAE-based classification method proposed in this paper can obtain high recognition accuracy in repeated trials and maintain stable performance. Because of the challenges presented by the noise in the original time-domain signals, although the traditional SAE-based method also relies on a deep structure for network pretraining, good classification performance cannot be obtained. In addition, because the structure of a BPNN is relatively shallow and its feature learning ability is limited, its diagnostic accuracy is not high. The BP algorithm based on a multi-hidden-layer BPNN results in unstable network performance and poor generalization ability.

#### 4. Conclusions

We come up with a smart method in this article which is on the basis of DAE to diagnose faults in rotating machinery. The feasibility of this approach for breakdown detecting of rolling bearings is verified on five test datasets. These datasets contain a large number of samples that represent the health of different bearings under various operating conditions. It demonstrates that based on the diagnosis outcome obtained from these datasets this method can adaptively mine fault features from time-domain data, so as to successfully solve a variety of detecting problems and efficiently implement the classification of the condition of the machine. Throughout the introduction of data dropout, the generalization ability and anti-interference ability of the model are improved. Since the fault features are automatically extracted, compared with traditional BPNN-based methods, our proposed method can avoid the dependence on artificial label and without the expert-level understanding of mechanical signal and diagnosis so that it can be easily used in practical applications.

#### Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

#### Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61772241 and in part by the 2018 Six Talent Peaks Project in Jiangsu Province under Grant XYDXX-127.

## References

- [1] T. Junbo, L. Weining, A. Juneng, and W. Xueqian, "Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder," in *the 27th Chinese Control and Decision Conference (2015 CCDC)*, Qingdao, China, 2015.
- [2] T. D. Bui and Guangyi Chen, "Translation-invariant denoising using multiwavelets," *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3414–3420, 1998.
- [3] Y. Bengio and O. Delalleau, "On the expressive power of deep architectures," in *Algorithmic Learning Theory. ALT 2011*, J. Kivinen, C. Szepesvári, E. Ukkonen, and T. Zeugmann, Eds., vol. 6925 of Lecture Notes in Computer Science, pp. 18–36, Springer, Berlin, Heidelberg, 2011.
- [4] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [5] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Transactions*, vol. 77, pp. 167–178, 2018.
- [6] F. Pacheco, M. Cerrada, R. V. Sánchez, D. Cabrera, C. Li, and J. Valente de Oliveira, "Attribute clustering using rough set theory for feature selection in fault severity classification of rotating machinery," *Expert Systems with Applications*, vol. 71, pp. 69–86, 2017.
- [7] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep auto-encoder feature learning method for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 95, no. 95, pp. 187–204, 2017.
- [8] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Transactions on Industry Applications*, vol. 53, no. 3, pp. 3057–3065, 2017.
- [9] Y. Wang, M. Liu, and Z. Bao, "Deep learning neural network for power system fault diagnosis," in *2016 35th Chinese Control Conference (CCC)*, pp. 6678–6683, Chengdu, China, 2016.
- [10] G. E. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [11] P. Chen, L. Yuan, Y. He, and S. Luo, "An improved SVM classifier based on double chains quantum genetic algorithm and its application in analogue circuit diagnosis," *Neurocomputing*, vol. 211, pp. 202–211, 2016.
- [12] B. Bessam, A. Menacer, M. Boumehraz, and H. Cherif, "Detection of broken rotor bar faults in induction motor at low load using neural network," *ISA Transactions*, vol. 64, pp. 241–246, 2016.
- [13] M. A. Atoui, S. Verron, and A. Kobi, "Fault detection with conditional Gaussian network," *Engineering Applications of Artificial Intelligence*, vol. 45, pp. 473–481, 2015.
- [14] M. Luo, C. Li, X. Zhang, R. Li, and X. An, "Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings," *ISA Transactions*, vol. 65, pp. 556–566, 2016.
- [15] C. Lu, Z. Wang, W. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol. 130, no. 130, pp. 377–388, 2017.
- [16] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 72–73, pp. 303–315, 2016.
- [17] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 1, pp. 185–195, 2018.
- [18] M. Sohaib and J. Kim, "Reliable fault diagnosis of rotary machine bearings using a stacked sparse autoencoder-based deep neural network," *Shock and Vibration*, vol. 2018, Article ID 2919637, 11 pages, 2018.
- [19] 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.
- [20] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—reviews, methodology and applications," *Mechanical Systems and Signal Processing*, vol. 42, no. 1–2, pp. 314–334, 2014.
- [21] K. Worden, W. J. Staszewski, and J. Hensman, "Natural computing for mechanical systems research: a tutorial overview," *Mechanical Systems and Signal Processing*, vol. 25, no. 1, pp. 4–111, 2011.
- [22] B. Samanta, "Artificial neural networks and genetic algorithms for gear fault detection," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1273–1282, 2004.
- [23] B. Samanta and C. Nataraj, "Use of particle swarm optimization for machinery fault detection," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 2, pp. 308–316, 2009.
- [24] V. T. Tran, B. Yang, M. Oh, and A. C. C. Tan, "Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1840–1849, 2009.
- [25] P. Vincent, H. Larochelle, I. Lajoie, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.
- [26] Q. V. Le, M. Ranzato, R. Monga et al., "Building high-level features using large scale unsupervised learning," in *International conference on machine learning*, pp. 507–514, Vancouver, BC, Canada, 2012.
- [27] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *International conference on machine learning*, pp. 478–487, New York, USA, 2016.