

Received September 25, 2018, accepted October 16, 2018, date of publication October 23, 2018, date of current version November 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2877447

# An End-to-End Model Based on Improved Adaptive Deep Belief Network and Its Application to Bearing Fault Diagnosis

JIAQI XIE<sup>1</sup>, GUIFU DU<sup>1</sup>, CHANGQING SHEN<sup>1,2</sup>, NAN CHEN<sup>2</sup>, (Member, IEEE), LIANG CHEN<sup>1</sup>, AND ZHONGKUI ZHU<sup>1</sup>

<sup>1</sup>School of Rail Transportation, Soochow University, Suzhou 215000, China

<sup>2</sup>Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore 119077

Corresponding author: Guifu Du (gfd@suda.edu.cn).

This work was supported in part by the National Natural Science Foundation of China under Grant 51875375 and Grant 51505311, in part by the Suzhou Prospective Research Program under Grant SYG201802, and in part by the Natural Science Foundation of Jiangsu Province under Grant BK 20150339.

**ABSTRACT** Effective machinery prognostics and health management play a crucial role in ensuring the safe and continuous operation of equipment, and satisfactory characteristics' expression of machine health status plays a key role in the ability to diagnose faults with high accuracy. At present, most methods based on signal processing and the shallow learning model rely on artificial feature extraction to identify the machine fault type. In practical applications, however, meaningful health management requires correct recognition of not only the health type but also the fault degree, if any occurs. Such recognition is useful for determining the priority level of mechanical maintenance and minimizing economic losses. Deep learning techniques, such as deep belief network (DBN), have demonstrated great potential in exploring characteristic information from machine status signals. In this paper, an end-to-end fault diagnosis model based on an adaptive DBN optimized by the Nesterov moment (NM) is proposed to extract deep representative features from rotating machinery and recognize bearing fault types and degrees simultaneously. Frequency-domain signals are inputted into the model for feature learning, and NM is introduced to the training process of the DBN model. Individual adaptive learning rate algorithms are then applied to optimize parameter updating. The performance of the proposed method is validated using a self-made bearing fault test platform, and the model is shown to achieve satisfactory convergence and a testing accuracy higher than those obtained from standard DBN and support vector machine.

**INDEX TERMS** Bearing, health management, deep belief network, end-to-end model.

## I. INTRODUCTION

Rotating machinery is widely used in the fields of manufacturing and aerospace, among others. Given developments in science and technology, machines are becoming increasingly complicated, intelligent, and integrated, and their working conditions are becoming more complex and dynamic. Rolling bearings, which usually operate at heavy loads and high speeds, are key components of many machines. The health status of a rolling bearing directly affects the performance, quality, and reliability of equipment in various manufacturing process. Depending on the type and size of the machine, bearing issues account for approximately 40% (large machines) to 90% (small machines) of all machine faults [1]. Therefore, to improve the reliability and ensure the safe operation

of these machines, effectively diagnosing bearings is essential [2]; such diagnosis can also help reduce operating and maintenance costs.

Numerous researchers have proposed various bearing fault diagnosis methods from the signal-processing aspect, such as empirical mode decomposition [3], wavelet packet transform [4], and morphological filter [5]. Wang *et al.* [6] proposed a new multi-speed fault diagnostic approach by using self-adaptive wavelet transform components generated from bearing vibration signals to identify four health states of rolling bearings. Zhang *et al.* [7] proposed a combination of the improved ensemble empirical mode decomposition (EEMD) and spectral kurtosis methods to address the problem of poor rolling bearing signals under the high background

noise of high-speed printing presses. Cui *et al.* [8] used a search algorithm to extract the morphological characteristics of signals of different gear fault types to recognize the damage pattern of gear signals. Yao *et al.* [9] used harmonic wavelet entropy to convert vibration signals into time-scale representations and achieved fault diagnosis of train bearings through demodulation and the envelope spectrum. A novel double-dictionary matching pursuit for fault extent evaluation of rolling bearings based on the Lempel–Ziv complexity index has been proposed to achieve satisfactory quantitative rolling bearing fault diagnosis [10]. Overall, these methods consistently achieve good results; however, they also require special expertise to enable effective signal processing. When strong background noise signals of complicated machines are encountered, health management through signal processing-based methods becomes difficult, and the extracted features are often unclear, thereby prohibiting accurate manual analysis and evaluation.

Several intelligent fault diagnosis methods, such as support vector machine (SVM) [11] and artificial neural network [12], have recently been proposed. However, traditional intelligent systems with shallow architectures depend on distinctive characteristics, which are often extracted manually. The representative ability of extracted features greatly affects the training and testing of fault diagnosis models. Thus, a self-learning model that can automatically extract features containing effective health information from complex signals is urgently needed to assist machine condition monitoring. As a breakthrough in neural networks, deep learning [13] combines multiple nonlinear learning layers to simulate the multi-layer abstract learning mechanism of the human brain; this approach thus possesses the capacity to process raw data in different domains and avoids the complexity and uncertainty often observed in traditional feature extraction processes. Some typical deep-learning networks include stacked auto-encoder [14], convolutional neural network (CNN) [15], and deep belief network (DBN) [16]. These methods have been successfully applied to speech recognition [17], natural language processing [18], and computer vision [19]. DBN has also been introduced to the field of mechanical fault diagnosis. Tran *et al.* [20], for example, successfully used DBN to identify defects in the valves of a reciprocating compressor. Guo *et al.* [21] proposed a two-layer hierarchical diagnosis network based on a CNN to classify bearing fault types and diagnose fault sizes. Gan *et al.* [22] employed a two-layer hierarchical network based on DBN to diagnose bearing faults, and Jia *et al.* [23] utilized a DNN based on auto-encoders for bearing fault diagnosis.

Deep learning shows strong ability in feature learning but is challenged in the field of mechanical equipment fault diagnosis. On the one hand, some studies are still based on manual preprocessing of characteristics prior to deep-model training, which does not make full use of the advantages of deep learning. On the other hand, during the training of deep-learning models, overfitting or missing of optimal values may

easily occur, and some measures or strategies must be developed to achieve successful model construction. Moreover, most of the existing research identifies fault types but rarely attempts to determine their degree despite the consensus that recognition of both fault type and degree will be beneficial to machine health management.

This paper addresses these issues and develops an end-to-end fault diagnosis model based on an adaptive DBN optimized by the Nesterov moment (NM) to extract deep representative features from rotating machinery and recognize bearing fault types and degrees simultaneously. First, frequency-domain signals are prepared for inputting into the deep-learning model for feature learning. Then, NM is introduced to the training process of DBN. Next, an independent adaptive learning rate is developed to optimize the parameter update process. Finally, the characteristics learned by DBN are used to identify fault types and degrees. Compared with the standard DBN method, NM can roughly judge where the parameters are proceeding and decline in advance validly to avoid missing the optimal point, thereby avoiding the problem of overfitting and local convergence. The development of an independent adaptive learning rate improves the convergence speed, and the proposed method enables simultaneous identification of fault types and degrees using one model. For verification, bearing signals from different health conditions are collected through a self-made bearing fault test bench for analysis, and the performance of the proposed method demonstrates satisfactory convergence. In fact, the testing accuracy of the proposed method is higher than those of standard DBN and SVM.

The rest of this paper is organized as follows. Section 2 explains the proposed adaptive DBN optimized by NM, Section 3 introduces an end-to-end model based on the improved DBN for recognizing both bearing fault type and degree, and Section 4 validates the performance and superiority of the proposed method compared with other methods through experimental analysis. Finally, conclusions are given in Section 5.

## II. ADAPTIVE DEEP BELIEF NETWORK OPTIMIZED BY THE NESTEROV MOMENT

### A. RESTRICTED BOLTZMANN MACHINES AND DEEP BELIEF NETWORK

A restricted Boltzmann machine (RBM; Figure 1) is a special form of the Markov random field proposed by Smolensky. RBM consists of visible and hidden layers containing visible  $v = \{v_1, v_2, \dots, v_i, \dots, v_n\}$  and hidden  $h = \{h_1, h_2, \dots, h_j, \dots, h_m\}$  neurons, respectively.

The visible layer is used to represent the input data, where each visible neuron is only associated with  $m$  hidden neurons and independent of other visible neurons. The hidden layer is understood as the intrinsic expression of the data. Each hidden neuron is only affected by  $n$  visible neurons and independent of other hidden neurons. As a stochastic network, RBM is mainly described by probability distribution

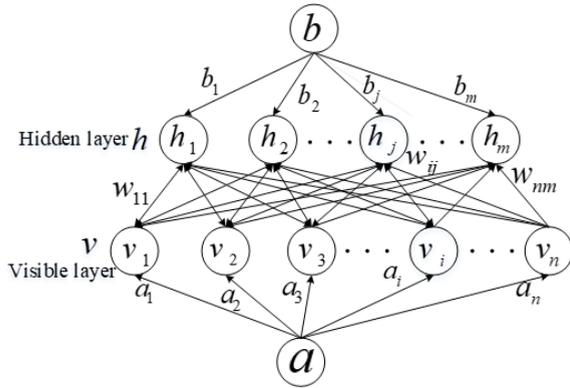


FIGURE 1. Schematic architecture of RBM.

and energy functions. Since the state of the neuron value of the RBM is random, according to Bayesian network theory, knowledge of the joint probability distribution, edge probability distribution, and conditional probability of visual neurons and hidden neurons is necessary to describe the whole network.

RBM neurons belong to the Boolean data type, which describes only two states: state 1 indicates that neurons are activated, and state 0 indicates that neurons are suppressed. Then, for a defined set of neuronal states  $(v, h) \in \{0, 1\}$ , the energy function of RBM is defined as:

$$E(v, h; \theta) = - \sum_{i=1}^n \sum_{j=1}^m w_{ij} v_i h_j - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j \quad (1)$$

where  $\theta = \{W, a, b\}$ ,  $v_i$  represents the state of the  $i^{th}$  neuron in the visible layer,  $h_j$  represents the state of the  $j^{th}$  hidden neuron,  $a_i$  represents the bias of the  $i^{th}$  neuron in the visible layer,  $b_j$  represents the bias of the  $j^{th}$  neuron in the hidden layer, and  $w_{ij}$  indicates the connection weight between the  $i^{th}$  neuron in the visible layer and the  $j^{th}$  neuron in the hidden layer.

The joint probability of visible neurons and hidden neurons is defined as

$$P(v, h; \theta) = \frac{1}{Z(\theta)} \exp(-E(v, h; \theta)) \quad (2)$$

which is the Gibbs distribution of the RBM network,  $Z(\theta)$  is the normalization factor, and the sum of the joint probabilities is guaranteed to be 1,  $Z(\theta)$  is the normalization factor and defined as:

$$Z(\theta) = \sum_v \sum_h \exp(-E(v, h; \theta)) \quad (3)$$

The edge probabilities and conditional probabilities of visible and hidden neurons can be obtained from the joint probabilities:

$$P(v; \theta) = \frac{\sum_h \exp(-E(v, h; \theta))}{\sum_v \sum_h \exp(-E(v, h; \theta))}$$

$$P(v | h; \theta) = \frac{\sum_h \exp(-E(v, h; \theta))}{\sum_v \exp(-E(v, h; \theta))} \quad (4)$$

$$P(h; \theta) = \frac{\sum_v \exp(-E(v, h; \theta))}{\sum_v \sum_h \exp(-E(v, h; \theta))}$$

$$P(h | v; \theta) = \frac{\exp(-E(v, h; \theta))}{\sum_h \exp(-E(v, h; \theta))} \quad (5)$$

After providing a set of visible neuron states, the probability that the  $i^{th}$  hidden neuron is activated is:

$$p(h_j = 1 | v; \theta) = \sigma(\sum_{i=1}^n w_{ij} v_i + b_j) \quad (6)$$

where  $\sigma$  represents the sigmoid function  $\sigma = 1 / (1 + e^{-x})$ . Similarly, after providing a set of hidden neuron states, the probability that the  $i^{th}$  visible neuron is activated is:

$$p(v_i = 1 | h; \theta) = \sigma(\sum_{j=1}^m w_{ij} h_j + a_i) \quad (7)$$

RBM training involves searching for parameters that maximize the edge probability of the visible neurons in the Gibbs distribution represented by RBM, that is, maximizing the logarithmic likelihood function:

$$\theta^* = \arg \max_{\theta} L(\theta) = \arg \max_{\theta} \sum_v \ln P(v; \theta) \quad (8)$$

Based on log-likelihood function, parameters are calculated by the following update rules:

$$\frac{\partial \ln p(v; \theta)}{\partial w_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{\text{mod el}}$$

$$\frac{\partial \ln p(v; \theta)}{\partial a_i} = \langle v_i \rangle_{data} - \langle v_i \rangle_{\text{mod el}}$$

$$\frac{\partial \ln p(v; \theta)}{\partial b_j} = \langle h_j \rangle_{data} - \langle h_j \rangle_{\text{mod el}} \quad (9)$$

where  $\langle \cdot \rangle_{data}$  represents the training data distribution expectation and  $\langle \cdot \rangle_{\text{mod el}}$  represents the expectation related to the distribution defined by the model. The sample of  $\langle \cdot \rangle_{data}$  can be easily obtained, whereas the sample of  $\langle \cdot \rangle_{\text{mod el}}$  cannot. Therefore, contrastive divergence is used to approximate the gradient by one full-step of Gibbs sampling. The updated rule can be modified as follows:

$$\Delta w_{ij} = \eta_w (\langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^1)$$

$$\Delta a = \eta_a (\langle v_i \rangle^0 - \langle v_i \rangle^1)$$

$$\Delta b = \eta_b (\langle h_j \rangle^0 - \langle h_j \rangle^1) \quad (10)$$

where  $\eta_w$ ,  $\eta_a$ , and  $\eta_b$  represent the learning rates of the weight  $w_{ij}$ , the visible layer deviation value  $a$ , and the hidden layer deviation value  $b$ , and have values ranging from 0 to 1.

In 2006, Hinton *et al.* [24] proposed DBN, which is a stack of several RBMs. In DBN, data enter from the visible layer of the underlying RBM, and the features are extracted bottom-up through the stacked RBMs. Finally, the deep features of the original data are obtained, and a classification layer is added to the top layer to classify the extracted features.

Figure 2 shows a stacked DBN model with three RBMs. The first RBM is formed by layer1 (as the visible layer) and layer2 (as the hidden layer), while the second RBM is formed

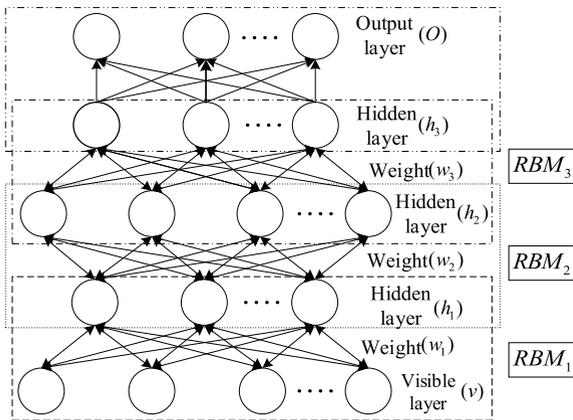


FIGURE 2. Schematic architecture of DBN.

by layer2 (as the visible layer) and layer3 (as the hidden layer). The third RBM is formed by layer3 (as the visible layer) and layer4 (as the hidden layer). Finally, the third layer of the hidden layer  $h_3$  and the classification layer  $O$  constitute the top classifier. Deep features from the third RBM are fed into a softmax classifier to conduct the classification task and one-hot encoding is adopted to binarize the labels. The number of neurons in the visible layer is determined by the dimensions of the input data, and the number of neurons in the classification layer is determined by the number of categories of input data.

**B. NESTEROV MOMENT FOR DEEP BELIEF NETWORK LEARNING**

Stochastic gradient descent (SGD) is often used to determine the gradient descent of likelihood functions. However, the objective functions encountered in practice tend to show various gradients in different dimensions. Figure 3 shows that the function is steeper in the longitudinal direction than in the lateral direction, forming a gully. In this case, since the change in gradient is large in the  $y$  direction and small in  $x$  the direction, the SGD oscillates in the  $y$  direction, slowly moves toward the  $x$  direction, and gradually approaches an optimal value, resulting in a slow training speed of the RBM.

The momentum method [25] can effectively speed up the gradient descent of SGD through the gully and suppress oscillations. The momentum method is based on the SGD and accomplished by multiplying the gradient of the last update by a momentum factor  $\gamma$  and adding the gradient calculated at in this case as the gradient of this update.

$$\begin{aligned}
 v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta_{t-1}) \\
 \theta_t &= \theta_{t-1} - v_t
 \end{aligned}
 \tag{11}$$

In Equation (11), the momentum factor  $\gamma$  is usually set to 0.9. Since the objective function oscillates back and forth in the  $y$  direction, the two successive gradients converge. If they have oscillating values in the  $y$  direction, the gradients in this direction cancel each other out and suppress the sway

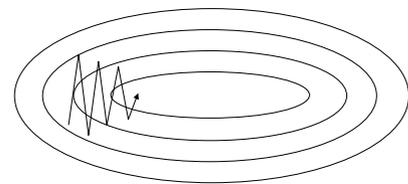


FIGURE 3. Mini-batch stochastic gradient descent along gullies.

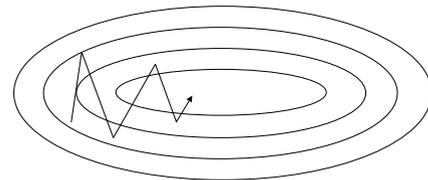


FIGURE 4. Mini-batch stochastic gradient descent along gullies with a momentum.

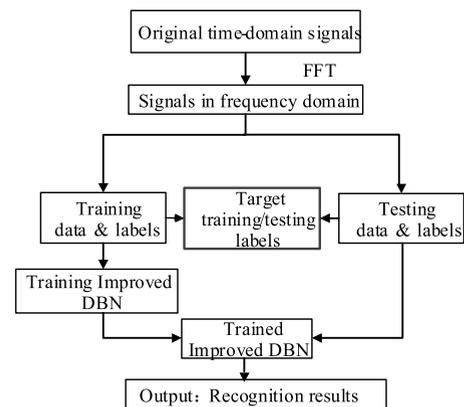


FIGURE 5. General procedure of the improved DBN method.

after addition. The gradient in the  $x$  direction is maintained. Thus, when the two gradients are added, the gradient descent in the  $x$  direction is accelerated and the optimal point is quickly approached, as illustrated in Figure 4.

The use of the momentum method in SGD causes accumulation of gradients in the same direction and cancellation of gradients in different directions, resulting in acceleration of approaching the optimal point. However, the gradient descent of the momentum method is random; it cannot judge where the parameters are heading and it decreases rapidly, causing the parameter to fall rapidly near the optimal point and miss it. NM [26] can effectively solve this problem by calculating the gradient of  $J(\theta - \gamma v_{t-1})$  to predict the position of the next drop of the parameter and decelerating before the parameter reaches the optimal point to avoid missing it.

$$\begin{aligned}
 v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta_{t-1} - \gamma v_{t-1}) \\
 \theta_t &= \theta_{t-1} - v_t
 \end{aligned}
 \tag{12}$$

Unfortunately, the NM method is sometimes excessively conservative. As the speed of the parameter approaching the optimal point slows down, if the parameter is not near

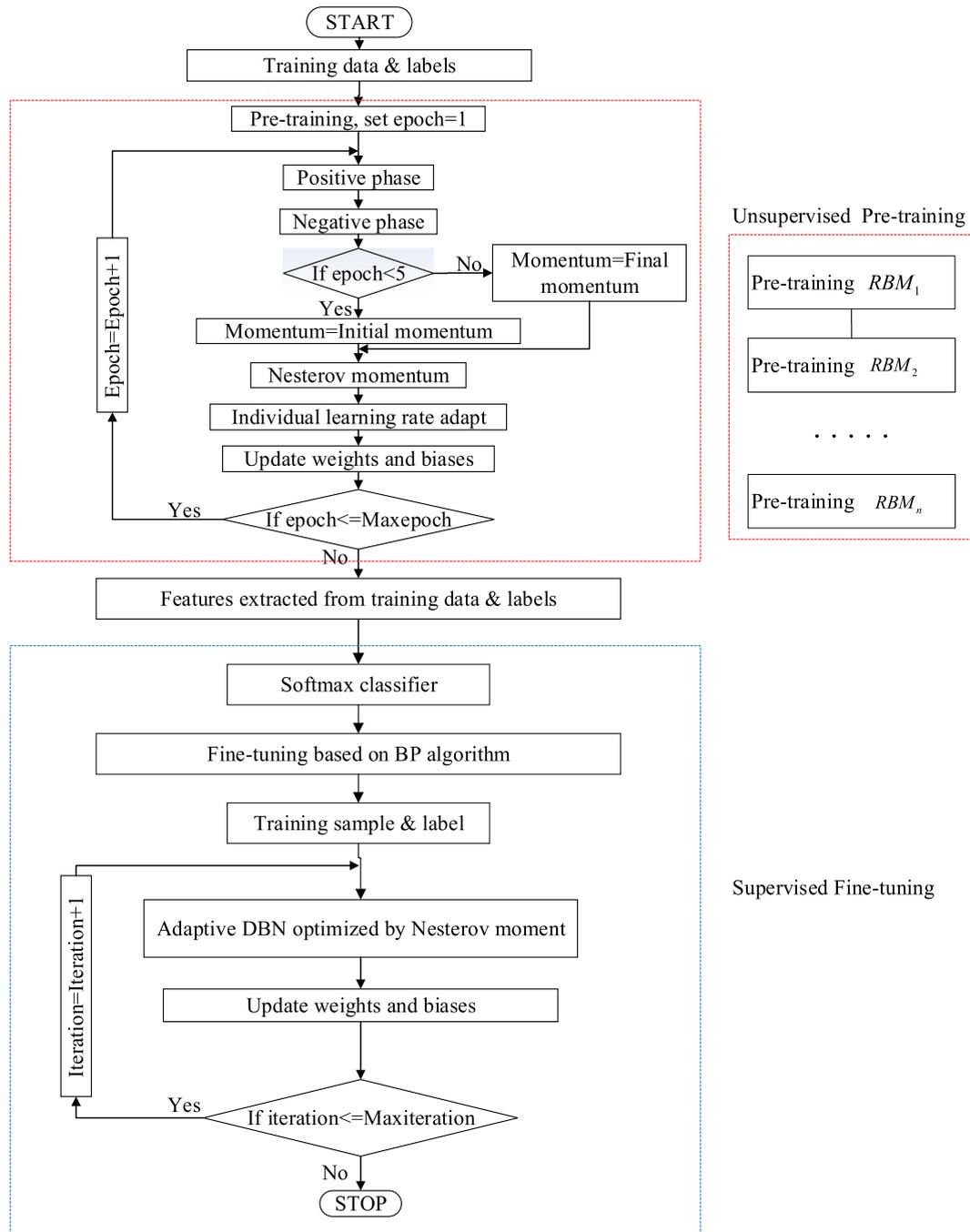


FIGURE 6. Training process of adaptive DBN optimized by the NM method.

TABLE 1. Specifications of the test bearing.

Outer diameter (mm)	Inner diameter (mm)	Pitch diameter (mm)	Roller number	Roller diameter (mm)	Contact angle (°)
52	25	38.5	9	7.938	0

the optimal point, the NM method still reduces the updated gradient, leading to a decline in speed well before the parameter approaches the optimal point; in addition, the speed of

model training is slow [27]. Hence, the individual adaptive learning rate, which is developed to avoid this problem, is introduced.

TABLE 2. List of bearing health conditions.

Bearing health condition No.	Fault type	Fault degree (mm)	Bearing health condition No.	Fault type	Fault degree (mm)
Condition 1	Normal	/	Condition 9		0.4
Condition 2		0.2	Condition 10	Inner fault	0.5
Condition 3		0.3	Condition 11		0.6
Condition 4	Outer fault	0.4	Condition 12		0.2
Condition 5		0.5	Condition 13		0.3
Condition 6		0.6	Condition 14	Roller fault	0.4
Condition 7	Inner fault	0.2	Condition 15		0.5
Condition 8		0.3	Condition 16		0.6

TABLE 3. Description of the dataset for the end-to-end fault diagnosis model.

Bearing health condition No.	Training samples	Testing samples	Identification label
Condition 1	75	25	1
Condition 2	75	25	2
Condition 3	75	25	3
Condition 4	75	25	4
Condition 5	75	25	5
Condition 6	75	25	6
Condition 7	75	25	7
Condition 8	75	25	8
Condition 9	75	25	9
Condition 10	75	25	10
Condition 11	75	25	11
Condition 12	75	25	12
Condition 13	75	25	13
Condition 14	75	25	14
Condition 15	75	25	15
Condition 16	75	25	16

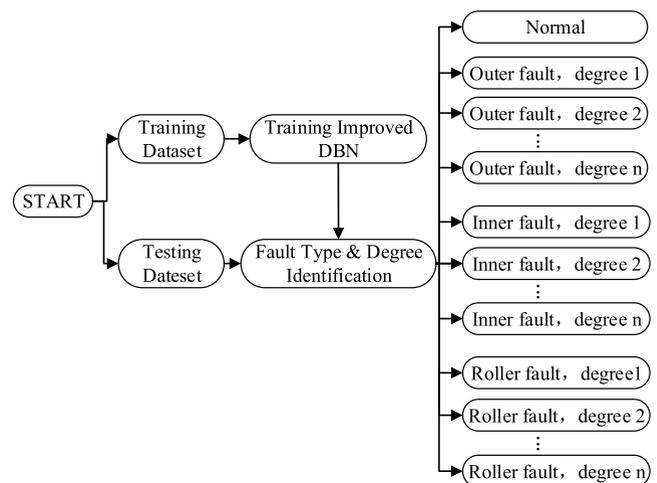


FIGURE 7. End-to-end model based on adaptive DBN optimized by the NM method.

C. INDIVIDUAL ADAPTIVE LEARNING RATE FOR UPDATING DEEP BELIEF NETWORK PARAMETERS

To obtain the ideal classification effect and training speed, an individual adaptive learning rate was used to train the RBM model. The connection weight  $w_{ij}$  update equations are:

$$h_{ij}(k) = \begin{cases} h_{ij}(k-1) + \alpha \text{if}(\text{grad}_{ij}^k \text{grad}_{ij}^{k-1}) > 0 \\ h_{ij}(k-1) \times (1 - \alpha) \text{if}(\text{grad}_{ij}^k \text{grad}_{ij}^{k-1}) < 0 \end{cases} \quad (13)$$

$$\Delta w_{ij}^k = \eta \cdot h_{ij}^k \cdot \nabla_{\theta} J(\theta_{t-1} - \gamma v_{t-1}) \quad (14)$$

where  $\text{grad}_{ij}^k$  indicates the gradient of the weight  $w_{ij}$  after training  $k$  times and  $h_{ij}(0)$  is set to 1. If the current direction of the weight gradient is consistent with the previous gradient direction,  $h_{ij}(t)$  increases by  $\alpha$  times; by contrast, if the current direction of the weight gradient is opposite the previous gradient direction,  $h_{ij}(t)$  is reduced by  $1 - \alpha$  times. The value of this parameter should be small, such as 0.1, to ensure that the learning rate does not grow too fast and the model does not miss the optimal solution when the learning rate increases by  $\alpha$  multiples. The limitation of  $h_{ij}(t)$ , which remains within the range of [0.01, 100], can effectively prevent the vanishing gradient problem.

Since the basis of judgment of the independent adaptive learning rate is different from that in the NM method,

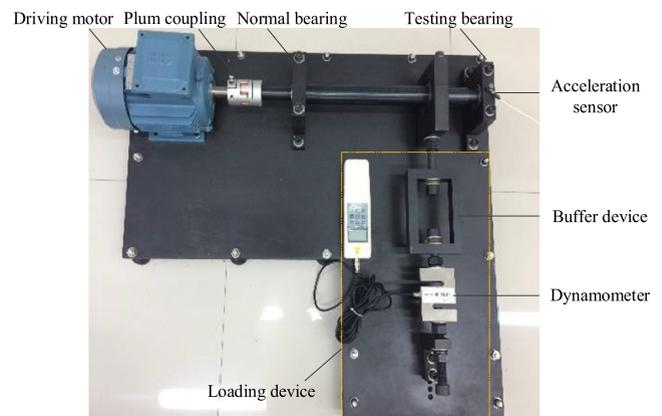


FIGURE 8. Bearing experimental fault platform.

the former can effectively avoid the problem that the model training speed slows down during the training process of the NM method. When the NM method judges incorrectly, the independent adaptive learning rate compensates the gradient update by increasing the descending step size to ensure that the parameters can obtain faster training speeds without exceeding the optimal point.

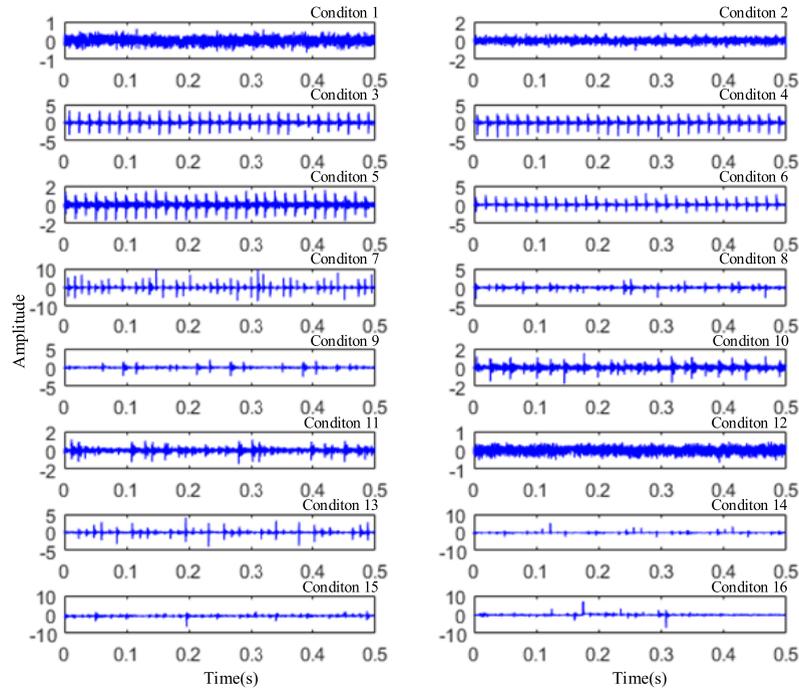


FIGURE 9. Time-domain signals under different bearing health conditions.

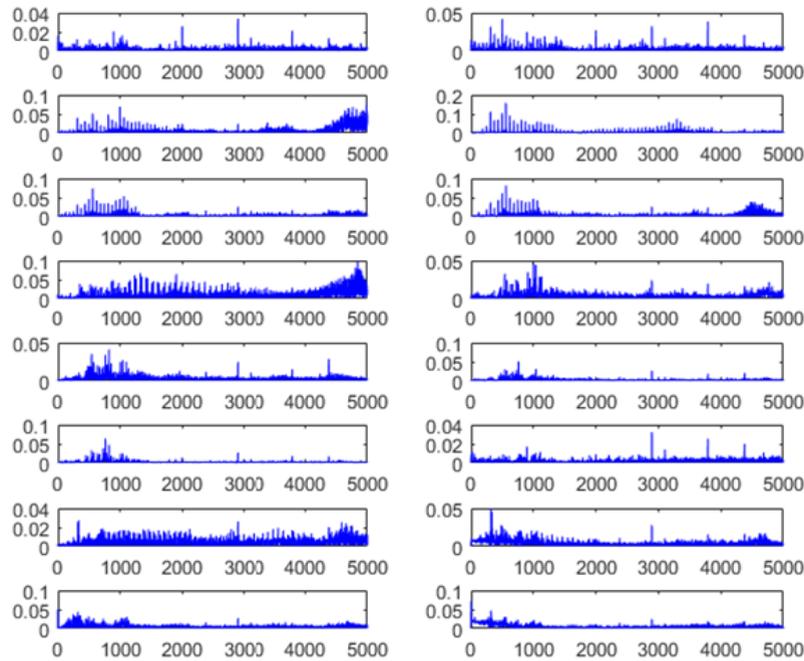


FIGURE 10. Frequency-domain signals under different bearing health conditions.

**III. END-TO-END FAULT DIAGNOSIS MODEL BASED ON ADAPTIVE DEEP BELIEF NETWORK OPTIMIZED BY THE NESTEROV MOMENT**

The DBN classification procedure is divided into two major processes, namely, pre-training and fine-tuning. In the

pre-training process, each RBM network is individually trained using a greedy layer-by-layer algorithm. The specific steps are as follows. The data are inputted into DBN’s visible layer to train the first RBM, and updating of the RBM parameters through forward and reverse propagation is

TABLE 4. Results obtained by the end-to-end fault diagnosis model.

Fault type	Condition number	Experiment number										Fault type recognition accuracy	
		1	2	3	4	5	6	7	8	9	10		
Normal	1	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Outer fault	2	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	3	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	
	4	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	
	5	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	
	6	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	
Inner fault	7	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	8	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	
	9	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	
	10	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	
Roller fault	11	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	12	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	
	13	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	
	14	92%	92%	92%	92%	92%	92%	92%	92%	92%	92%	92%	
	15	96%	92%	92%	96%	96%	100%	96%	96%	96%	96%	100%	
16	88%	96%	88%	96%	96%	96%	96%	96%	96%	96%	96%		
Average accuracy for each condition(%)		98.75	99.00	98.50	99.00	99.00	99.25	99.00	99.00	99.00	99.25		

performed until the maximum number of cycles is reached. The layer characteristics of the data are first learned from the first RBM output in the hidden layer. Then, fixing the parameters of the first RBM, these features are inputted into the second RBM. The above steps are repeated to train the parameters of the second RBM. Fine-tuning is applied to improve the performance of the DBN model by utilizing the back-propagating (BP) algorithm [28]. After pre-training of the Softmax classifier, the tagged data are used to begin input from the bottom layer, and BP [29] is used for global fine-tuning.

Combining the independent adaptive learning rate and the NM method based on standard DBN, the adaptive DBN optimized by the NM method, that is, the improved DBN, is established, as presented in Figure 5. Figure 6 illustrates training process of the improved DBN, which pre-trains each RBM to complete initialization of the end-to-end model through greedy unsupervised learning and unlabeled training sample sets.

This paper builds an end-to-end fault diagnosis model based on the improved DBN to identify fault types and degrees. First, the NM method and the independent adaptive learning rate are developed to improve the DBN in terms of model optimization. Then, an end-to-end fault diagnosis model is established based on the improved DBN. Signals under different health conditions are analyzed, as presented in Figure 7. Hence, the end-to-end model can directly recognize various fault types and degrees.

#### IV. EXPERIMENTAL VERIFICATION

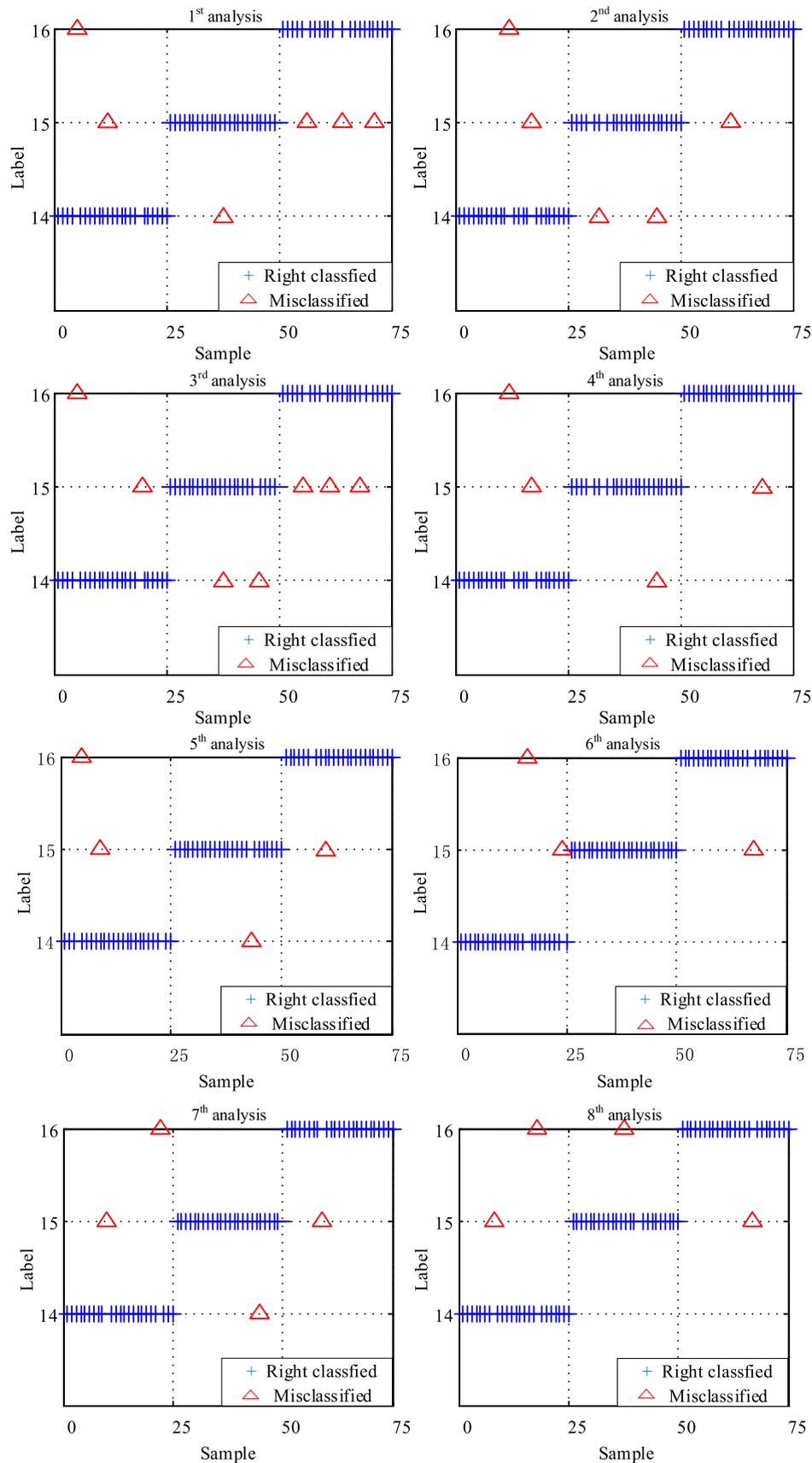
##### A. BRIEF INTRODUCTION OF THE DATASET

Figure 8 shows the experimental platform, which includes a drive motor, normal bearing, test bearing, shaft, bolt-nut loading system, accelerometer, and NI PXIe-1082 data acquisition system. The plum coupling transmits power to the shaft, where the test bearing (6205-2RS SKF) is installed. The bolt-nut loading system includes loading bolts, nuts, cushioning devices, and the SGSF-20K dynamometer. By tightening the nut above the bolt, the nut is squeezed against the support to create an adjustable radial load. As shown by the red block in Figure 8, a PCB 352C33 accelerometer was installed on the bearing block at 12 o'clock. Here, the sampling frequency is 10 KHz.

Table 1 shows the geometric parameters of the test bearing. A total of 16 health conditions, including four fault types (normal, outer race fault, inner race fault, and roller fault) and five fault degrees (0.2, 0.3, 0.4, 0.5, and 0.6 mm), are collected from the self-made bearing test platform and analyzed, as shown in Table 2. The waveforms of signals are presented in Figures 9 and 10.

For analysis, 3/4 of the total samples are randomly selected as training data, and the remaining samples are used for testing. Detailed information of the dataset is summarized in Table 3.

In the end-to-end fault diagnosis model, the number of visible layer neurons is 624, and the number of neurons in the first hidden layer is 500. In the second hidden layer, the number of neurons is 300. The connection weights of each



**FIGURE 11.** Ten repeated analysis results by the end-to-end fault diagnosis model for 0.4-0.6 mm roller faults.

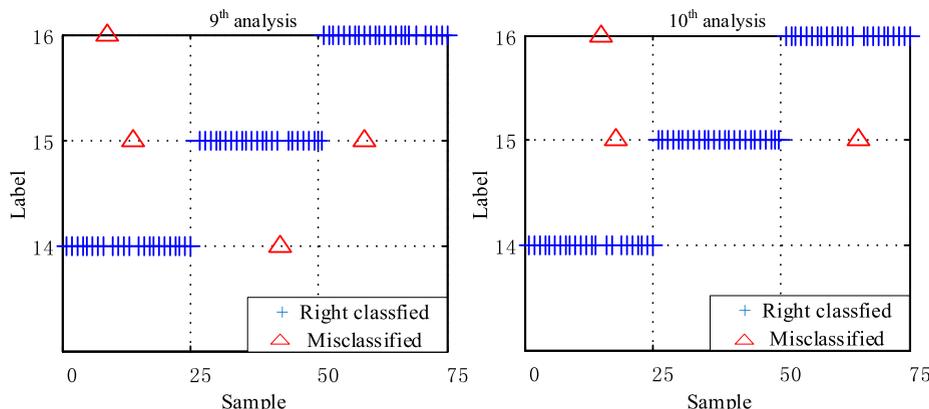


FIGURE 11. Ten repeated analysis results by the end-to-end fault diagnosis model for 0.4-0.6 mm roller faults.

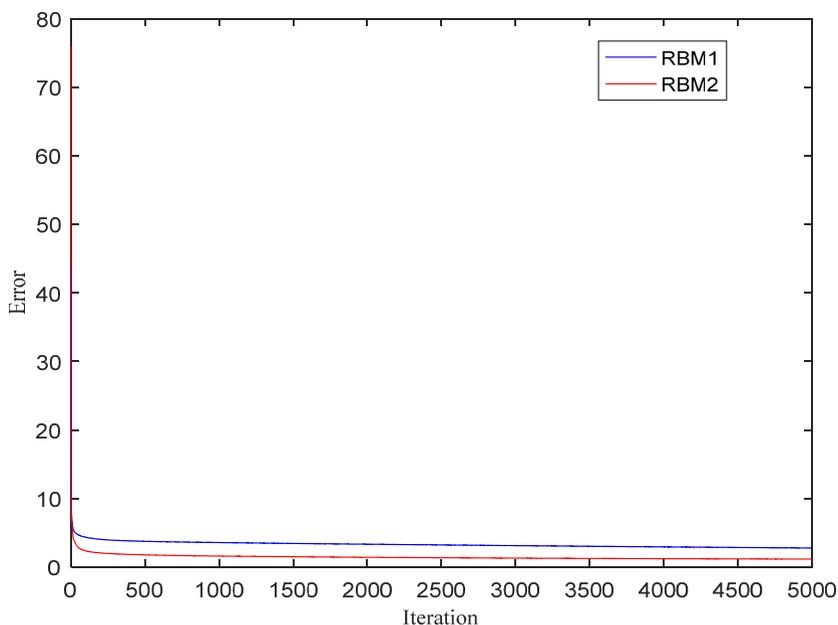


FIGURE 12. Training processes of RBM1 and RBM2 in the fine-tuning iterations.

layer are initialized by  $N(0, 0.01)$ . In the pre-training process, the learning rate of the first RBM is initialized to 0.1, and the learning rate of the second RBM is initialized to 0.0025. The weight decay term  $L2$  is added to penalize the large weight term, and the attenuation coefficient is 0.0001. The number of pre-training iterations is set to 5000, and the number of iterations is fine-tuned to 5000. To reduce the impact of random factors, the experiment was repeated 10 times, and the average results are considered the final recognition results.

**B. EXPERIMENTAL RESULTS AND ANALYSIS**

Table 4 reveals the results obtained by the end-to-end fault diagnosis model. The average testing accuracy of the model is 98.98%. For the first 13 bearing-health conditions, all of the recognition results of the model are achieved with the expected accuracy of 100%; for bearing-health conditions

Nos. 14, 15, and 16, the average accuracies for roller fault degree recognition are 92.0%, 96.0%, and 94.4%, respectively. Compared with those of inner and outer race faults, dynamic characteristics are more complicated when the surface of the roller suffers from defects; thus, the representative ability of the features learned from complex dynamic responses is relatively weak. Figure 11 reveals the specific results of 10 repeated analyses of bearing health conditions Nos. 14 to No. 16, which feature different roller fault degrees. Incorrect recognition only occurs in these three conditions. For other conditions, the recognition accuracy of the end-to-end fault diagnosis model is 100%.

Comparative analysis between the proposed method and standard DBN and SVM was conducted. The same frequency-domain signals were used as inputs to these models for recognition, and the average results of the comparative

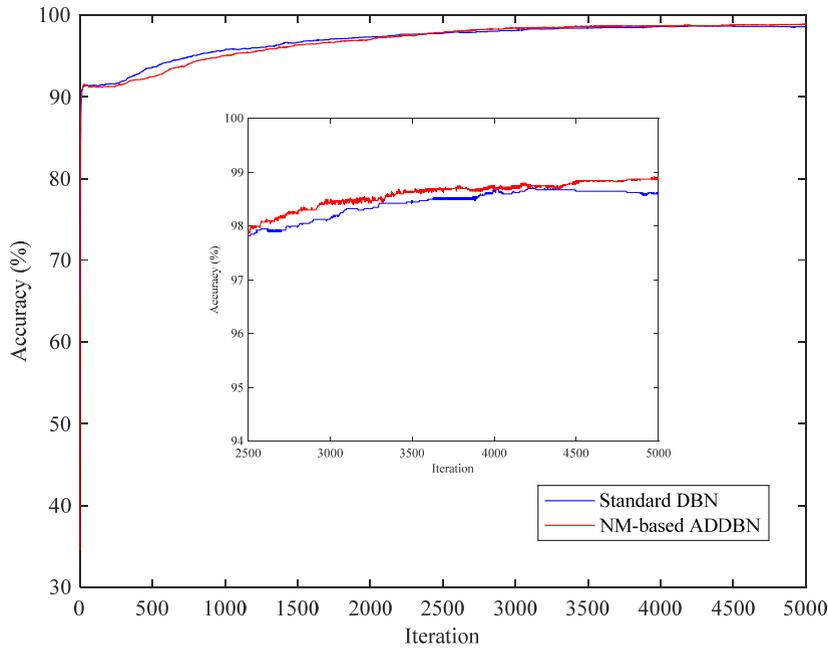


FIGURE 13. Trends of testing accuracy of two models in the fine-tuning iterations.

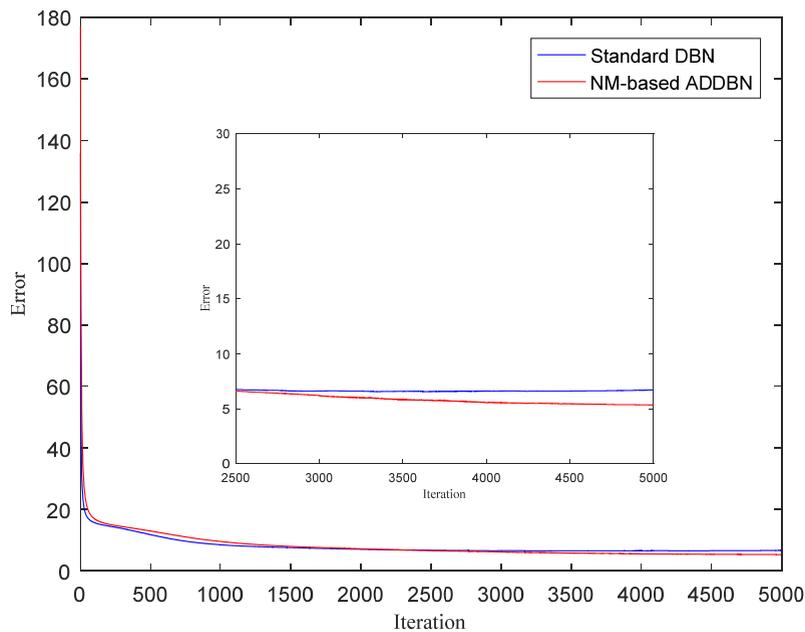


FIGURE 14. Trends of error for the testing samples in each fine-tuning iteration.

methods are expressed in Table 5. The first and second methods revealed testing accuracies of 64.25% and 98.67%, while the proposed method achieved an identification result of 98.98%. In general, the proposed end-to-end fault diagnosis model based on improved DBN performs better than either standard DBN or SVM. Figure 12 depicts the training process of RBM1 and RBM2 in fine-tuning iterations and shows satisfactory convergence trends. As shown in Figure 13, while the

fault-identification accuracy of the proposed method is lower than that of the standard DBN at the beginning of the analysis, it gradually and steadily increases. After 2500 iterations, the accuracy of the improved DBN gradually surpasses that of the standard DBN and shows the highest values among the methods tested. Similarly, in Figure 14, although the error of the proposed method is higher than that of the standard DBN at the beginning of analysis, it steadily decreases and

TABLE 5. Average testing accuracy using different methods.

Methods	The initial learning rate		Average training accuracy of fault type(%)	Average training accuracy of fault degree(%)	Average testing accuracy of fault type(%)	Average testing accuracy of fault degree(%)
	1 <sup>st</sup> RBM	2 <sup>nd</sup> RBM				
SVM+FFT	-	-	98	98	64.25	64.25
Standard DBN+FFT	0.1	0.0025	100	98.57	100	98.72
Improved DBN+FFT	0.1	0.0025	100	98.57	100	98.98

gradually becomes lower than that of the standard DBN after 2500 iterations. These improvements may be attributed to the ability of the NM method to calculate the position of the next drop of the gradient prediction parameter of  $J(\theta - \gamma v_{t-1})$ , thereby ensuring that the parameter is always close to the optimal point at the appropriate speed to avoid missing it. At the same time, the independent adaptive learning rate can compensate the gradient descent to a certain extent, effectively avoiding the problem that the training speed of the NM method is too slow during the training process. Taking the results together, the end-to-end model based on the improved DBN achieves ideal results and exhibits better performance than conventional methods.

## V. CONCLUSION

In this study, a novel end-to-end fault diagnosis method is proposed for the fault diagnosis of rotating machinery. The frequency-domain signals are inputted into the model for feature learning. To ensure that parameters decline at the appropriate speed to avoid missing the optimal point and improve training performance, NM is introduced to the training process of DBN. An individual adaptive learning rate algorithm is developed to select a suitable step length with which to accelerate the descent to optimize the process of parameter updating. Finally, the characteristics learned by DBN are used to identify various fault types and degrees. The proposed method can automatically extract valid fault features and avert from the extraction of complex features from frequency-domain signals manually. Verification using a self-made bearing fault testing platform indicated that the proposed method demonstrates excellent recognition performance on the defective bearing dataset and achieves higher accuracy than standard DBN and SVM.

## ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grants No. 51875375, 51505311), the Suzhou Prospective Research Program (Grant No. SYG201802) and Natural Science Foundation of Jiangsu Province (Grant No. BK 20150339).

## REFERENCES

- [1] F. Immovilli, C. Bianchini, M. Cocconcelli, A. Bellini, and R. Rubini, "Bearing fault model for induction motor with externally induced vibration," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3408–3418, Aug. 2013.
- [2] L. Cui, J. Huang, H. Zhai, and F. Zhang, "Research on the meshing stiffness and vibration response of fault gears under an angle-changing crack based on the universal equation of gear profile," *Mech. Mach. Theory*, vol. 105, pp. 554–567, Nov. 2016.
- [3] Y. Imaouchen, M. Kedadouché, R. Alkama, and M. Thomas, "A frequency-weighted energy operator and complementary ensemble empirical mode decomposition for bearing fault detection," *Mech. Syst. Signal Process.*, vol. 82, pp. 103–116, Jan. 2017.
- [4] Y. Wang, G. Xu, L. Liang, and K. Jiang, "Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis," *Mech. Syst. Signal Process.*, vols. 54–55, pp. 259–276, Mar. 2015.
- [5] Y. Li, X. Liang, and M. J. Zuo, "Diagonal slice spectrum assisted optimal scale morphological filter for rolling element bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 85, pp. 146–161, Feb. 2017.
- [6] D. Wang, K.-L. Tsui, and Q. Miao, "Prognostics and health management: A review of vibration based bearing and gear health indicators," *IEEE Access*, vol. 6, pp. 665–676, 2018.
- [7] J. Zhang, J. Wu, Y. Wang, K. Li, L. Ma, and F. Zhao, "Fault diagnosis in printing press roller bearing based on spectrum kurtosis and improved EEMD," *Appl. Sci. Graphic Commun. Packag.*, no. 477, pp. 545–552, 2018.
- [8] L. Cui, J. Huang, and F. Zhang, "Quantitative and localization diagnosis of a defective ball bearing based on vertical-horizontal synchronization signal analysis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 11, pp. 8695–8706, Nov. 2017.
- [9] D. Yao et al., "Harmonic wavelet envelope method applied in railway bearing fault diagnosis," *J. Eng. Sci. Technol. Rev.*, vol. 6, no. 2, pp. 24–28, 2013.
- [10] L. Cui, X. Gong, J. Zhang, and H. Wang, "Double-dictionary matching pursuit for fault extent evaluation of rolling bearing based on the Lempel-Ziv complexity," *J. Sound Vib.*, vol. 385, pp. 372–388, Dec. 2016.
- [11] B.-C. Zhang, R. Xu, X.-J. Yin, and Z. Gao, "Research on fault diagnosis for rail vehicle compartment of LED lighting system of analog circuit based on WP-EE and BP neural network," in *Proc. Control Decis. Conf.*, May 2016, pp. 2989–2993.
- [12] T. Khouldia, A. E. Hadjadj, K. Bouacha, and D. O. Abdeslam, "Multi-objective optimization of ANN fault diagnosis model for rotating machinery using grey rational analysis in Taguchi method," *Int. J. Adv. Manuf. Technol.*, vol. 89, nos. 9–12, pp. 3009–3020, 2017.
- [13] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
- [14] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3235–3243, Jul. 2018.
- [15] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [16] T. A. Ngo, Z. Lu, and G. Carneiro, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Med. Image Anal.*, vol. 35, pp. 159–171, Jan. 2017.
- [17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

- [18] A. Kumar et al., "Ask me anything: Dynamic memory networks for natural language processing," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1378–1387.
- [19] Y. Jia et al., "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.
- [20] V. T. Tran, F. Althobiani, and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4113–4122, 2014.
- [21] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, Nov. 2016.
- [22] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mech. Syst. Signal Process.*, vols. 72–73, no. 2, pp. 92–104, 2016.
- [23] F. Jia, Y. G. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016.
- [24] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [25] D. A. T. D. do Rio Vaz, A. L. A. Mesquita, J. R. P. Vaz, C. J. C. Blanco, and J. T. Pinho, "An extension of the blade element momentum method applied to diffuser augmented wind turbines," *Energy Convers. Manage.*, vol. 87, pp. 1116–1123, Nov. 2014.
- [26] G. Yang, L. Liu, Z. Jiang, J. Guo, and T. Wang, "Incoherent beam combining based on the momentum SPGD algorithm," *Opt. Laser Technol.*, vol. 101, pp. 372–378, May 2018.
- [27] T. Tieleman and G. Hinton, "Divide the gradient by a running average of its recent magnitude," COUR-SERA, Neural Netw. Machine Learn., Tech. Rep. Accessed: Apr. 21, 2017. [Online]. Available: <https://zh.coursera.org/learn/neural-networks/lecture/YQHki/rmspr-op-divide-the-gradient-by-a-running-average-of-its-recent-magnitude>
- [28] Z. Zhao, Q. Xu, and M. Jia, "Improved shuffled frog leaping algorithm-based BP neural network and its application in bearing early fault diagnosis," *Neural Comput. Appl.*, vol. 27, no. 2, pp. 375–385, 2016.
- [29] W. Y. Liu, B. P. Tang, J. G. Han, X. N. Lu, N. N. Hu, and Z. Z. He, "The structure healthy condition monitoring and fault diagnosis methods in wind turbines: A review," *Renew. Sustain. Energy Rev.*, vol. 44, pp. 466–472, Apr. 2015.



**JIAQI XIE** received the B.S. degree in vehicle engineering from the Nanjing Institute of Technology in 2017. She is currently pursuing the M.S. degree in measurement techniques and instruments with Soochow University. Her research interests include machine fault diagnosis.



**GUIFU DU** was born in Shandong, China. He received the B.E. and Ph.D. degrees from the China University of Mining and Technology, Xuzhou, China, in 2012 and 2017, respectively. In 2017, he joined the School of Rail Transportation, Soochow University, Suzhou, China, as a Lecturer. His current research interests include fault diagnosis of dc traction power system and railway electrification.



**CHANGQING SHEN** received the B.S. and Ph.D. degrees in instrument science and technology from the University of Science and Technology of China in 2009 and 2014, respectively, and the Ph.D. degree (joint program) in systems engineering and engineering management from the City University of Hong Kong in 2014. He is currently an Associate Professor with the School of Rail Transportation, Soochow University, China. His research interests include signal processing and artificial intelligence based machinery fault diagnosis and prognosis.



**NAN CHEN** (M'11) received the B.S. degree in automation from Tsinghua University, Beijing, China, in 2006, the M.S. degree in computer science in 2009, and the M.S. degree in statistics and the Ph.D. degree in industrial engineering from University of Wisconsin–Madison, Madison, WI, USA, in 2010. He is currently an Associate Professor with the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore. His research interests include statistical modeling and surveillance of engineering systems, simulation modeling design, condition monitoring, and degradation modeling.



**LIANG CHEN** received the B.S. and Ph.D. degrees in control science and engineering from Zhejiang University in 2004 and 2009, respectively, and the Ph.D. degree (joint program) in control science and engineering from the Berlin Institute of Technology in 2009. He is currently an Associate Professor with the School of Mechanical and Electric Engineering, Soochow University, China. His research interests include artificial intelligence-based machinery process control and performance evaluation.



**ZHONGKUI ZHU** received the B.S. degree in automobile and tractor (automobile) and the M.S. degree in vehicle engineering from Hefei Polytechnic University in 1997 and 2002, respectively, and the Ph.D. degree in instrument science and technology from the University of Science and Technology of China in 2005. He is currently a Professor with the School of Rail Transportation, Soochow University. His research interests include vehicle system dynamics and control, vibration measurement, and signal processing.

• • •