Research Article

A Parameter-Optimized DBN Using GOA and Its Application in Fault Diagnosis of Gearbox

Jingbo Gai,[1] **Junxian Shen** ⓘ**,**[1] **He Wang,**[1] **and Yifan Hu**[2]

[1]*College of Aerospace and Civil Engineering, Harbin Engineering University, Harbin 150001, Heilongjiang Province, China*
[2]*School of Electrical Engineering and Automation, Harbifn Institute of Technology, Harbin 150001, Heilongjiang Province, China*

Correspondence should be addressed to Junxian Shen; shenjunxian@hrbeu.edu.cn

Aiming at the problems of poor self-adaptive ability in traditional feature extraction methods and weak generalization ability in single classifier under big data, an internal parameter-optimized Deep Belief Network (DBN) method based on grasshopper optimization algorithm (GOA) is proposed. First, the minimum Root Mean Square Error (RMSE) in the network training is taken as the fitness function, in which GOA is used to search for the optimal parameter combination of DBN. After that the learning rate and the number of batch learning in DBN which have great influence on the training error would be properly selected. At the same time, the optimal structure distribution of DBN is given through comparison. Then, FFT and linear normalization are introduced to process the original vibration signal of the gearbox, preprocess the data from multiple sensors and construct the input samples for DBN. Finally, combining with deep learning featured by powerful self-adaptive feature extraction and nonlinear mapping capabilities, the obtained samples are input into DBN for training, and the fault diagnosis model for gearbox based on DBN would be established. After several tests with the remaining samples, the diagnosis rate of the model could reach over 99.5%, which is far better than the traditional fault diagnosis method based on feature extraction and pattern recognition. The experimental results show that this method could effectively improve the self-adaptive feature extraction ability of the model as well as its accuracy of fault diagnosis, which has better generalization performance.

1. Introduction

As a key part of mechanical transmission system, the gearbox is widely used in wind turbine generators, coal mining, and military equipment. When operating, the gearbox is exposed to alternating load, and key parts such as gears and transmission shafts are prone to failure. If the fault is not diagnosed in time and the equipment keeps running, minor faults may turn into serious faults, resulting in machine shutdown, production stagnation, and even casualties [1, 2]. Therefore, real-time state monitoring and fault diagnosis of the gearbox are necessary measures to ensure the safe operation of these equipments [3, 4].

The fault diagnosis process for the gearbox generally includes four steps: data collection, feature extraction, feature fusion, and pattern recognition. Among them, feature extraction is the most critical step, which directly determines the performance of fault diagnosis. Sun et al. [5] proposed a fault diagnosis method for the planetary gearbox based on parameter optimized VMD, determining the parameters of mode number and center frequency adaptively according to the extreme value of power spectral density. Such method can effectively extract fault feature frequency, making accurate diagnosis for crack faults in gears under strong background noises and subtle fault signals. Isham et al. [6] decomposed the vibration signal of the gearbox by VMD, then extracted the time-domain, frequency-domain, and time-frequency-domain features of each IMF component to construct the eigen matrix of signal, and finally trained ELM to establish a fault diagnosis model to complete the intelligent diagnosis of the gearbox in the wind turbine. Zhang et al. [7] took advantage of GWO algorithm to search for the optimized parameters in TVF-EMD matching with the input signal, eliminating the influence of parameter selection on

the decomposition results. Then, the fault characteristics of rotating machinery were extracted by analyzing the IMF component with the maximum weighted kurtosis index. The abovementioned methods are effective for simulation signals and certain specific fault signals, but they need abundant knowledge in signal processing and rich experience in expert diagnosis. In complex industrial test sites, with huge amounts of data, fault information is often complex and changeable, and it may also contain internal and external excitation as well as the coupling of multiple faults. It is unrealistic to just rely on professional technicians and diagnostic experts for manual analysis. At present, in health monitoring, with the increase of measuring points, sampling frequency, and time length of data collection, a larger amount of data has been acquired by the monitoring system. These massive data makes the traditional fault diagnosis methods fall into a bottleneck in real-time monitoring efficiency, fault diagnosis accuracy, and self-adaptive analysis capability. Therefore, exploiting information from the big data to efficiently and accurately identify the health status has become a new problem in the health monitoring of equipment [8].

With the development of machine learning, fault diagnosis methods based on machine learning models have become a research hotspot, such as BP Neural Network, Support Vector Machines (SVM), and Extreme Learning Machines (ELM). However, in case of high-dimensional big data, when applying the shallow learning model for gearbox fault diagnosis, there is lack of diagnosis and generalization ability in fault diagnosis, the accuracy of which relies on the extraction quality of fault features among big data [9]. As a new method in the field of machine learning, deep learning is increasingly applied in fault diagnosis due to its powerful modeling and characterization capabilities. Different from the traditional fault diagnosis methods of feature extraction and pattern recognition, deep learning integrates them into the deep neural network to carry out the feature extraction of signals in the hidden layer and the recognition of state patterns in the output layer. Lei et al. [8] used denoising autoencoder (DAE) as an unsupervised algorithm in the pretraining stage and BP algorithm as a supervised algorithm in the fine-tuning stage to build a deep neural network, achieving adaptive extraction of fault characteristics and accurate identification of health conditions of different faults in the gearbox under various working conditions and a large number of samples. Jin et al. [9] introduced the multiobjective optimization algorithm to optimize multiple Stack Denoising Automatic Encoders (SDAE) and extracted the diverse fault features of the planetary gearbox. Lei et al. [10] proposed a two-stage learning method for machine intelligence diagnosis, learning the characteristics of signals directly with unsupervised two-layer neural network and then adopting softmax regression to classify the health status. After that the method was successfully verified with relevant data sets. Deep learning avoids the dependence on a large number of signal processing technologies and diagnostic experience, directly extracts fault features self-adaptively from signals in frequency domain, integrates feature extraction and pattern recognition methods in traditional

fault diagnosis, and achieves self-adaptive extraction of fault features as well as intelligent diagnosis of health conditions under big data.

Deep learning opens a new way for intelligent fault diagnosis. Wen [11] used DBN with different structures to establish the fault diagnosis model for bearing, evaluated the models through multiple indexes of performance, and selected the network structure with the best diagnostic performance. Through experiments, Zhang [12] deeply analyzed the influence of number of nodes in the hidden layer, learning rate, and number of iterations on feature extraction ability of DBN and determined how the main parameters should be set. The abovementioned methods have achieved certain effects in DBN network construction and self-adaptive fault feature extraction. However, in the process of parameter selection for DBN, network parameters are still modified according to experience. At this time, the diagnosis model has disadvantages of insufficient stability and high randomness of diagnosis. Based on this, this paper designed a new fault diagnosis method based on DBN to give a set of optimal diagnosis scheme. First, parameter optimization is carried out through GOA to reduce the influence of manual parameter setting on training results. Then, the influence of optimal network structure distribution and parameter optimization on feature extraction capability of the hidden layer is analyzed. Finally, the preprocessed data is input into the network for training, and a fault diagnosis model for the gearbox based on DBN is constructed. Through experiments, it has been proved that the method proposed in this paper can effectively improve DBN's self-adaptive fault feature extraction ability and identification accuracy effectively solving the shortcomings in traditional methods under big data.

2. The Parameter-Optimized DBN Method

In this section, some related algorithms which include DBN, GOA, and the parameter determining criterion are introduced. Based on these algorithms, a new parameter-optimized DBN method is proposed.

2.1. Brief Overview of DBN. DBN is a probability generation network composed of several Restricted Boltzmann Machines (RBMs) [13, 14]. The network consists of a visible layer, a hidden layer, and an output layer. The visible layer and the hidden layer are connected by weights, and each neuron itself has an offset to represent its own weight. The output layer and the previously hidden layer form a BP neural network which is mainly used to adjust the initial parameters of the hidden layer to achieve supervised training of the entire network. In the DBN learning process, the data is input from the bottom layer and then through the various hidden layers to complete the training process. The learning process can be divided into two parts: pretraining and fine tuning. Figure 1 shows a DBN structure with n layers hidden.

2.1.1. Pretraining. Pretraining uses an unsupervised greedy layer-by-layer approach to initialize the connection weights and offsets between the RBM layers. Then, each layer of RBM
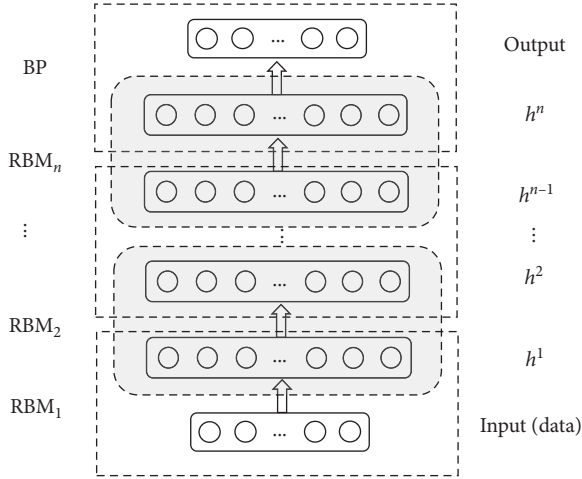
FIGURE 1: The structure of DBN.

is trained separately from bottom to top [15]. Suppose RBM is an energy-generated Bernoulli model, given the energy of state (v, h) [16]:

$$\begin{cases} E(v, h; \theta) = -\sum_{i=1}^{V} \sum_{j=1}^{H} \omega_{ij} v_i h_j - \sum_{i=1}^{V} b_i v_i - \sum_{j=1}^{H} a_j h_j, \\ \theta = \{w, a, b\}. \end{cases} \quad (1)$$

In this formula, θ is a parameter of the RBM and ω_{ij} is a connection weight between the visible layer node and the hidden layer node. V and H are the number of visual units and hidden units. v_i and h_j are the node states of the visible layer and the hidden layer. b_i and b_j are the offsets of the visible layer and the hidden layer. In order to maintain sparseness, the visible layer offset b_j can be initialized to $lb(\hat{p}_i/(1 - \hat{p}_i))$, where $\hat{p}_i$ is the probability of $v_i = 1$. The hidden layer offset a_j is initialized to a large positive number and ω_{ij} is initialized to a smaller random number. At this time, the joint probability of the model is as follows:

$$\begin{cases} P(v, h) = \frac{1}{Z} e^{-E(v,h)}, \\ Z = \sum_{v,h} e^{-E(v,h)}. \end{cases} \quad (2)$$

In this formula, Z is a normalization factor. Since there is no connection between the peer nodes, the probability of the visible layer unit v_i and the hidden layer unit h_j is independent:

$$\begin{cases} p(v_i = 1 \mid h; \theta) = \sigma\left(\sum_{j=1}^{H} \omega_{ij} h_j + b_j\right), \\ p(h_j = 1 \mid h; \theta) = \sigma\left(\sum_{i=1}^{H} \omega_{ij} v_i + a_j\right), \\ \sigma(x) = \frac{1}{(1 + \exp(x))}. \end{cases} \quad (3)$$

In the formula, $\sigma(x)$ is a Sigmoid function. Find the edge distribution of $p(v, h; \theta)$ to h:

$$p(v; \theta) = \frac{1}{Z} \sum_h \exp(-E(v, h; \theta)). \quad (4)$$

θ can be obtained by solving the maximum log-likelihood estimation function on the training set, and the RBM parameter update criterion is obtained by the contrast divergence method [17]:

$$\begin{cases} \Delta b_i = \varepsilon\left(\langle v_i \rangle_{\text{data}} - \langle v_i \rangle_k\right), \\ \Delta a_j = \varepsilon\left(\langle h_j \rangle_{\text{data}} - \langle h_j \rangle_k\right), \\ \Delta \omega_{ij} = \varepsilon\left(\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_k\right), \end{cases} \quad (5)$$

where ε is the learning efficiency and $\langle \cdot \rangle_{\text{data}}$ and $\langle \cdot \rangle_k$ are the expected values of the distribution defined by the current model and the reconstructed model.

2.1.2. Fine Tuning. Since pretraining is unsupervised learning, the initial values of the parameters obtained through pretraining are not optimal parameters. At this stage, the BP neural network is combined with the label to fine tune the parameters for the problem of large output error. The BP neural network is set up at the output layer of DBN and supervised training is performed from top to bottom. According to formula (20), the connection parameters between each layer are optimized to make the best classification ability of DBN. For the complex characteristics of early fault signal, DBN is able to establish a deep model by simulating the deep tissue structure of the brain, which can more effectively characterize the complex mapping relationship between vibration signal and running state of the gearbox.

2.2. Parameter Determining Criterion: Minimum Root Mean Square Error (RMSE). It is necessary to evaluate the network error in the training process. RMSE is the square root of the difference between the reconstructed visual layer state vector and the original data input vector after one Gibbs sampling of RBM with the training sample as the initial state. The specific definition is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \left(x_i^j - y_i^j\right)^2}{m \times n}}. \quad (6)$$

In equation (6), x_i^j means the state vector of the visual layer; y_i^j means the input vector of original data; and m and n, respectively, represent the number of nodes and samples in the visible layer.

The smaller the RMSE, the better the training effect. Through the observation error, the training situation of the model can be judged, and the parameters such as iteration times, learning rate, and number of batch learning can be adjusted to achieve better training effect. Therefore, RMSE is an excellent choice as a fitness function in the optimization process.

2.3. Grasshopper Optimization Algorithm. Grasshopper Optimization Algorithm (GOA) imitates the swarm foraging behavior of grasshoppers in nature and shows excellent performance in dealing with multiobjective optimization problems [18]. The network formed by grasshopper populations connects all the individuals so that all grasshoppers keep in step, and one individual can determine the direction of predation through others in the group. Since the location of the target is unknown, the position of grasshopper with the best fitness is considered to be the closest to the target. Then, the grasshoppers will move in the same direction as the target in the network. With the position update of grasshoppers, in order to achieve a balance between global search and local search, the appropriate range area would decline self-adaptively until finally grasshoppers get together and approach the optimal solution [19, 20]:

$$X_i^d = c\left(\sum_{j=1,i\neq j}^{N} c\,\frac{\mathrm{ub}_d - \mathrm{lb}_d}{2} s\left(\left|x_j^d - x_i^d\right|\right)\frac{x_j - x_i}{d_{ij}}\right) + \widehat{T}_d. \tag{7}$$

In equation (7), N is the population size; ub_d and lb_d represent the upper and lower bounds of the dth dimension, respectively; $\widehat{T}_d$ represents the current iterative optimal solution; and

$$c = c_{\max} - l\frac{c_{\max} - c_{\min}}{L}. \tag{8}$$

In this equation, $c_{\max}$ is the maximum value of c; $c_{\min}$ is the minimum value of c; l represents the current number of iterations; and L represents the maximum number of iterations.

In order to make each grasshopper move towards the optimal solution during each search, it is assumed that the optimal fitness value among individuals in the current search process is the target value. GOA starts optimization with a random initial set of solutions and updates position according to formula (7), where the update of factor c depends on formula (8). The best location of target is updated after each iteration until the termination condition is met and the location and fitness value of the optimal individual are returned.

2.4. Proposed Method. As shown in Figure 2, the optimization steps of GOA for DBN parameters are as follows:

(1) Set all parameters of GOA and initialize the population

(2) Take the DBN training RMSE value as fitness function, evaluate individual fitness value fit(i) according to the learning rate and the number of batch learning, and then mark the optimal individual

(3) Judge whether the current iteration times have reached the termination condition; if so, end iteration and output the result; if not, continue to the next step

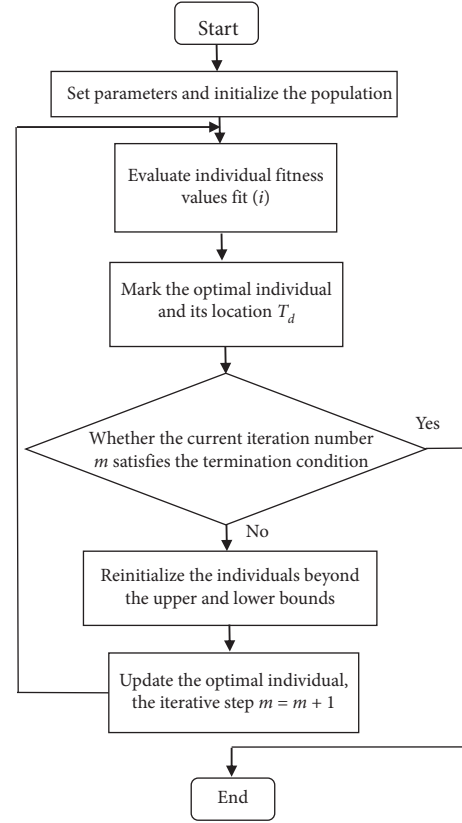(4) Update the position of each individual and reinitialize the individuals beyond the upper and lower bounds

Figure 2: The flow chart of improved DBN based on GOA.

(5) Update the optimal individual and start a new iteration: $m = m + 1$

3. The Construction of Gearbox Fault Diagnosis Model Based on Parameter-Optimized DBN

Combined with the characteristics of big data from equipment monitoring and the advantages of deep learning, a fault diagnosis method for the gearbox based on parameter-optimized DBN is proposed. This method achieves the organic combination of unsupervised learning and supervised learning and is capable for self-adaptive extraction of fault features under big data as well as the identification of equipment running state. Also, it is superior to traditional methods which are with poor self-adaptive ability in feature extraction as well as insufficient generalization performance of shallow network in fault identification. The method flow chart is shown in Figure 3. The specific steps are as follows:

(1) The vibration signal of the gearbox is preprocessed by FFT and linear normalization.

(2) The vibration signals of multiple sensors are preprocessed and formed into eigenvector $A = [x_1, x_2, \ldots, x_n]^T$.

(3) Minimize the training RMSE in DBN by searching for the optimal combination of DBN parameters with GOA, and the optimal structure distribution of the network is given by comparison.
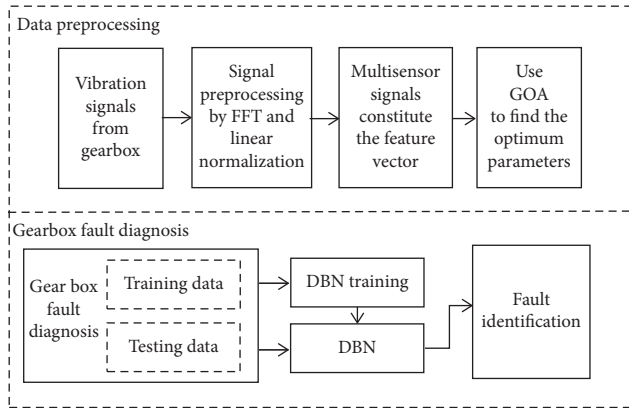
FIGURE 3: The flow chart of the proposed method.

(4) Input the standard samples of different states of the gearbox into the optimized DBN. After establishing the fault diagnosis model, the test samples of the gearbox in different states would be diagnosed.

4. Case Study

4.1. Experiment Setup.
In this paper, the transmission system of the gearbox is taken as the research object to verify the effectiveness of the proposed method by monitoring and diagnosing its running state. The test bed of drive system in the gearbox is shown in Figure 4(a). It is composed of a drive motor, gearbox, and magnetic powder brake. The schematic diagram of the test rig and accelerometer layout is illustrated in Figure 4(b). Different fault states of the gearbox are implanted by wire electrical-discharge machining. The number of teeth between pinion and gearwheel, pressure angle, and tooth width are 55/75, 20°, and 20 mm.

The four running states of the gearbox are simulated on the test bed, and samples are collected under five different working conditions (880 rpm no load, 1500 rpm no load, 880 rpm 0.2 A, 880 rpm 0.1 A, and 880 rpm 0.05 A) in each state with the sampling frequency of 5120 Hz. As shown in Table 1, 500 sample groups are obtained from each running state, each containing 512 points. In conclusion, the data set for all running states contains 2000 samples, which simulate the gearbox running states under various working conditions and with various faults. During the training and testing of the network, 50% samples are randomly selected for training and the other 50% for testing. Table 1 illustrates 4 running states of normal, pitting, snaggletooth, and abrasion, as well as their corresponding status labels.

4.2. Fault Diagnosis Using Parameter-Optimized DBN

4.2.1. Data Preprocessing.
According to the procedure of the proposed method, the vibration signal of the gearbox is preprocessed. The FFT spectrum of different running states are given in Figure 5.

Each signal corresponds to a superposition of several components in frequency domain and can be decomposed by frequency-domain analysis. In order to make the signal more concise and more convenient to represent, each group of samples would go through FFT transformation, obtaining 1024 points. In view of the symmetry of the spectrum, half of the data points are taken for the eigenvector, so as to reduce the dimension of signal feature. The data from sensors in different measurement positions are superimposed to increase the information included in the eigenvector about space and angle. In order to reduce the influence of noises and abnormal samples on the network training, the obtained eigenvectors are normalized linearly to reduce the training time and to speed up the convergence.

4.2.2. Determination for Optimal Parameter Combinations and Network Structure of DBN.
Considering that no formula or theory is known in setting the number of neuron nodes in each hidden layer, many experiments and relevant knowledge are required then. In this paper, three types of hidden layer structures would be analyzed: smooth type (200-200-200), increasing type (100-200-400), and decreasing type (400-200-100). In order to determine both the optimal parameter combination and the optimal network structure at a time, GOA is applied. After searching for the optimal parameter combination of the network with different structures under the same training conditions, the structure whose RMSE converges to the minimum is considered to be the best.

First, the optimal parameters of learning rate and batch extraction in DBN are searched by GOA, with the search range of [0, 1] and [1, 100], respectively. According to Zhang's suggestions [18], parameter setting of GOA are shown in Table 2.

After parameter setting of the optimization algorithm, the parameter search of different network structures is started. As shown in Figure 6, in order to explain the parameter search process in detail, the optimization curve under the network structure of 400-200-100 is given, where the RMSE converges to the minimum value of about 0.0074. Also, the iteration begins to converge after 31 times of calculation, indicating that the algorithm has strong global optimization ability and fast convergence speed, making it suitable for searching optimal parameter combination of DBN. In this case, the optimal combination of parameters obtained by descending type (400-200-100) is [0.1711, 25]. Meanwhile, the same operation is taken in smooth type (200-200-200) and increasing type (100-200-400), with the result of [0.3498, 20] and [0.6356, 16], respectively.

In order to determine the optimal network structure, the error curves of three network types with corresponding optimal parameter combination are given. It is indicated in Figure 7(a) that the RMSE of decreasing network (400-200-100) converges faster with smaller value. Figure 7(b) shows the convergence in later period (after 50 iterations), and RMSE of the decreasing type is significantly smaller than that of other types. Therefore, the decreasing type is taken as the best structure in this paper.

In addition, the DBN model achieves a good training effect and tends to a stable state at the 100th iteration. Although the increase in the number of iterations is beneficial to improve the effectiveness of fault recognition, the
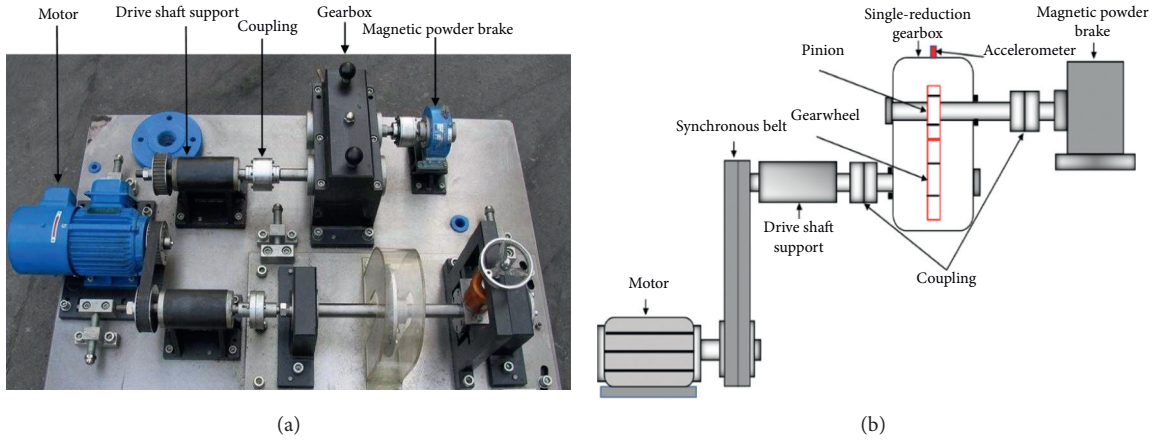
(a)   (b)

FIGURE 4: Test bed of the transmission system: (a) overview and (b) diagram.

TABLE 1: Four running states of the gearbox system.

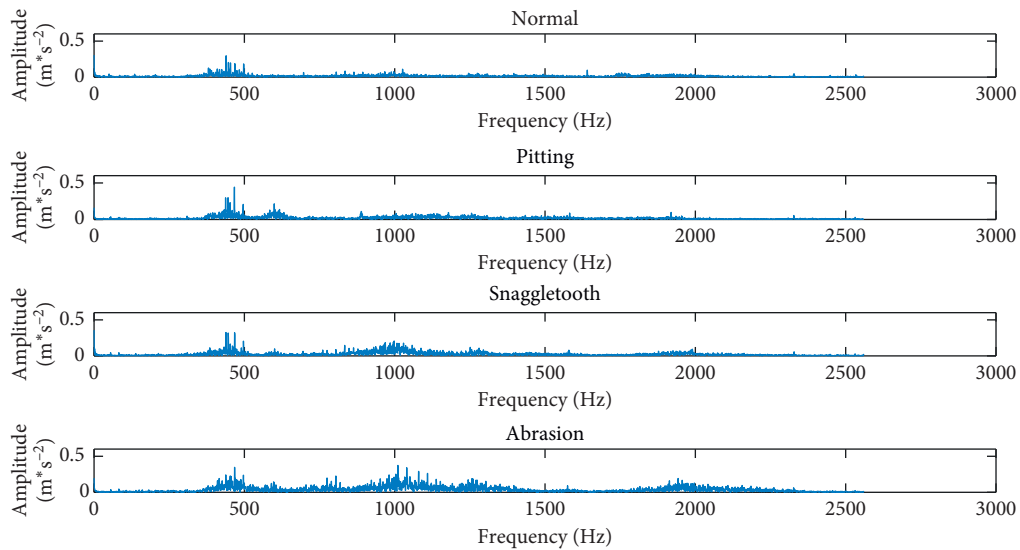| Running state	Sample groups	Working conditions	Status labels
Normal	500	5	1000
Pitting	500	5	0100
Snaggletooth	500	5	0010
Abrasion	500	5	0001

FIGURE 5: FFT spectrum of different running states.

TABLE 2: Parameter setting of GOA.

Number of search agents	Maximum number of iterations	Variable dimension	Range of learning rate	Range of batch number
10	100	2	[0, 1]	[1, 100]

calculation time required would also increase greatly. Considering the recognition effect and calculation cost comprehensively, the number of iterations is set to 100.

The number of nodes in the input layer depends on the sample dimension (2560 dimensions), and the number of nodes in the output layer is determined by the running state (4 states). In this paper, the decreasing structure type with minimum RMSE is applied in the hidden layer. The finally determined parameters of the structure in the DBN model are shown in Table 3.
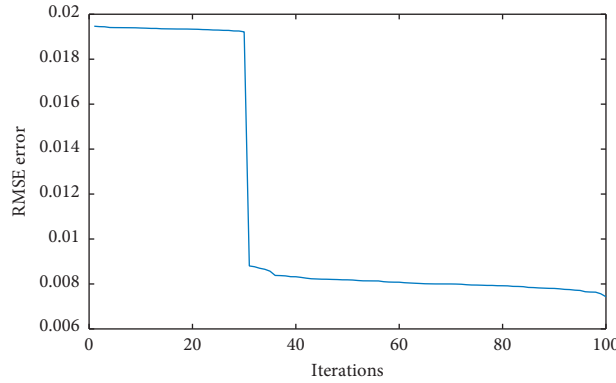
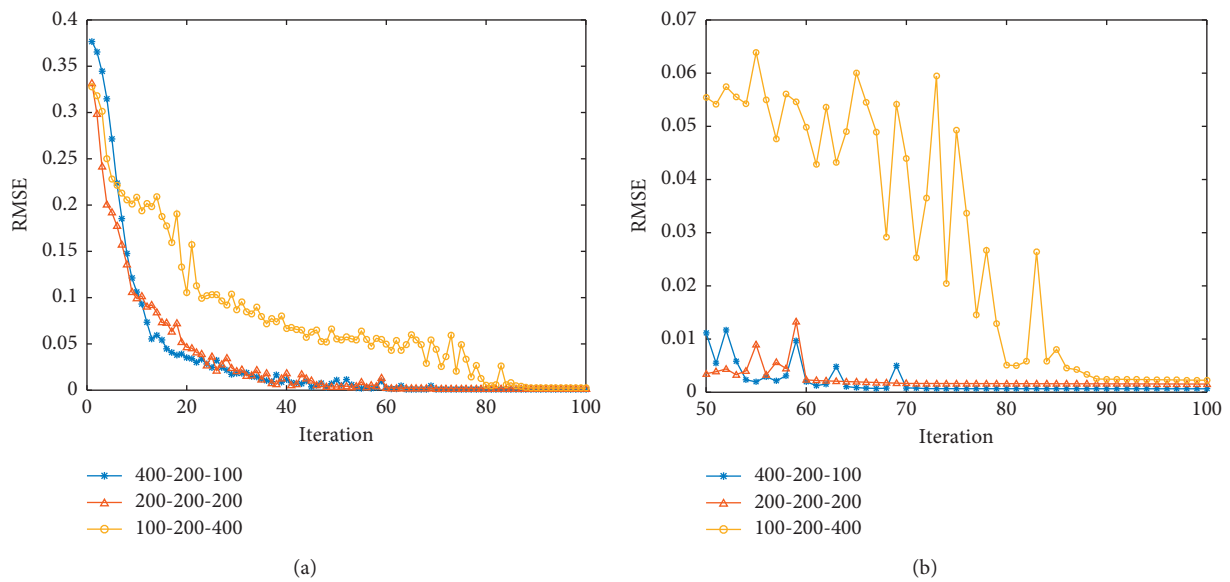FIGURE 6: The variation of RMSE with the number of iterations.

(a)

(b)

FIGURE 7: RMSE curve during fine tuning: (a) global and (b) detail.

TABLE 3: DBN structure parameter setting.

Input layer	RBM1	RBM2	RBM3	Output layer
2560	400	200	100	4

TABLE 4: Learning parameter setting.

Batch Size	Learning rate	Momentum value	Iteration
25	0.7716	0.05	100

According to the parameter combination obtained after optimization, the setting of learning parameters in DBN is shown in Table 4.

4.2.3. DBN Hidden Layer Feature Extraction Capability Analysis. In order to verify that optimized DBN is more capable for feature extraction, the extraction capability of hidden layers before and after optimization is compared. According to the advice given by Hinton et al. [21], the learning rate and the number of batch extraction in DBN selected by experience (viewed as DBN before optimization) are [0.1, 10]. With the same sample and network structure for training, the node values of the third hidden layer are output, and its sparsity is taken as the evaluation index of feature extraction capability.

As illustrated in Figure 8, the features extracted by DBN after optimization are with more sparsity than that of DBN before optimization. Such sparse features can effectively express the essential features of data and can improve the generalization ability of fault features. Table 5 lists the changes of parameter combination, RMSE, and comprehensive distance value during iterations. The comprehensive distance within and between classes, obtained by dividing the distance between classes and that within classes, is an essential criterion for the separability of samples in different classes and the aggregation in the same classes. Under a specified feature, the longer the distance between classes, the more separable the samples in different states. Similarly, the shorter the distance within classes, the more concentrated the samples in same states. Therefore, the increase of the
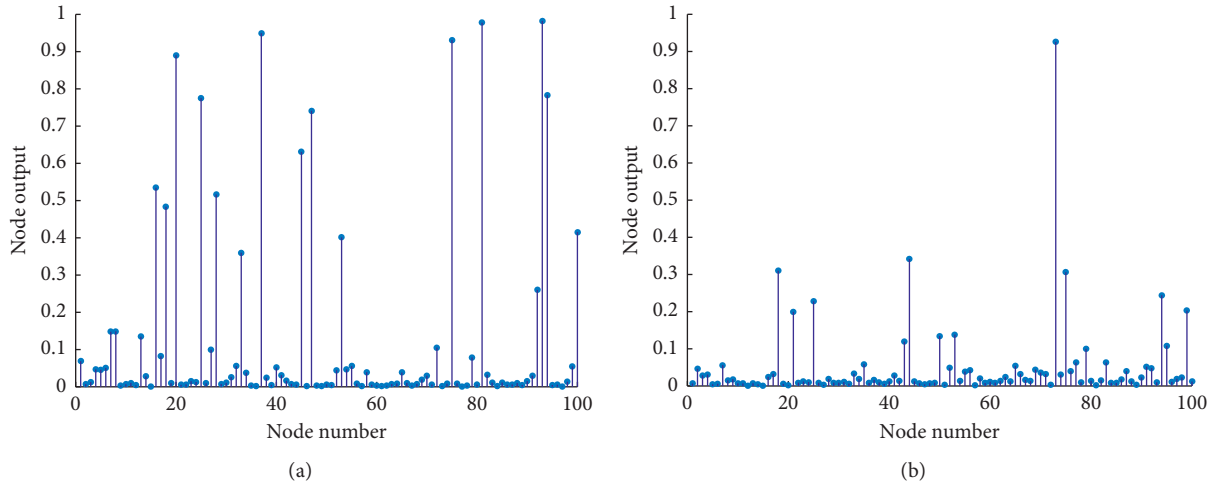
(a)

(b)

FIGURE 8: Comparison of fault feature sparsity between networks: (a) before optimization and (b) after optimization.

TABLE 5: Iteration process.

Number of iterations	Batch size	Learning rate	Comprehensive distance	RMSE
1	20	0.3166	2.9238	0.01945
2	20	0.1067	2.9233	0.01945
3	25	0.1067	2.9250	0.01944
4	25	0.1067	2.9267	0.01941
		. . .		
28	25	0.1068	2.9676	0.01926
29	25	0.1068	2.9793	0.01925
30	25	0.1068	2.9837	0.01922
31	25	0.1712	5.0172	0.00880
		. . .		
97	25	0.1712	5.1849	0.007640
98	25	0.1712	5.1862	0.007635
99	25	0.1711	5.1893	0.007563
100	25	0.1711	5.1982	0.007431

comprehensive distance between and within classes is capable for expressing the improvement in feature extraction ability of the network. As shown in Table 5, RMSE gradually decreases with the iteration, and the comprehensive distance also appears in an upward trend. At the 31st iteration, RMSE decreases significantly and the comprehensive distance increases significantly, until finally they reach stabilization. This fact indicates that, with the iteration of parameters, the feature extraction ability of DBN is improving, which has a direct impact on the reduction of RMSE.

The proposed method is capable for extracting fault features self-adaptively from the spectrum of running states. In order to further verify the feature extraction ability of the proposed method, the first three principal components of these features are extracted by KPCA and visualized. Then, optimized DBN, DBN set by experience, shallow probability network, and traditional feature extraction are compared, respectively. The shallow probability network adopts a single hidden layer Probability Neural Network (PNN), which follows Bayesian law of prior probability and Bayesian rules of decision to simplify the network training and carry out the nonlinear mapping between original data and features.

Traditional feature extraction method is to extract 20 common characteristics in time domain, frequency domain, and time-frequency domain from vibration signals of the gearbox: mean value, standard deviation, peak value, RMS, root amplitude, margin index, kurtosis index, waveform index, pulse index, peak index, mean frequency, center frequency, RMS frequency, standard deviation frequency, kurtosis frequency, and the first 5 orders of energy entropy in IMF components from EMD.

Figure 9(a) is the scatter diagram of principal elements for feature extraction in the proposed method, indicating that the samples in the same state cluster completely in its own space, while those in different states separate effectively without overlapping. Figure 9(b) is associated with DBN set by experience. In the scatter diagram of the first three principal elements, little overlap appears among pitting, snaggletooth, and abrasion, which would have an adverse impact on the accuracy of fault diagnosis. At the same time, the significance of parameter optimization is verified as well, which significantly affects the ability of feature extraction in the network. Figure 9(c) is a shallow probability network. Compared with deep probability network such as DBN, it is
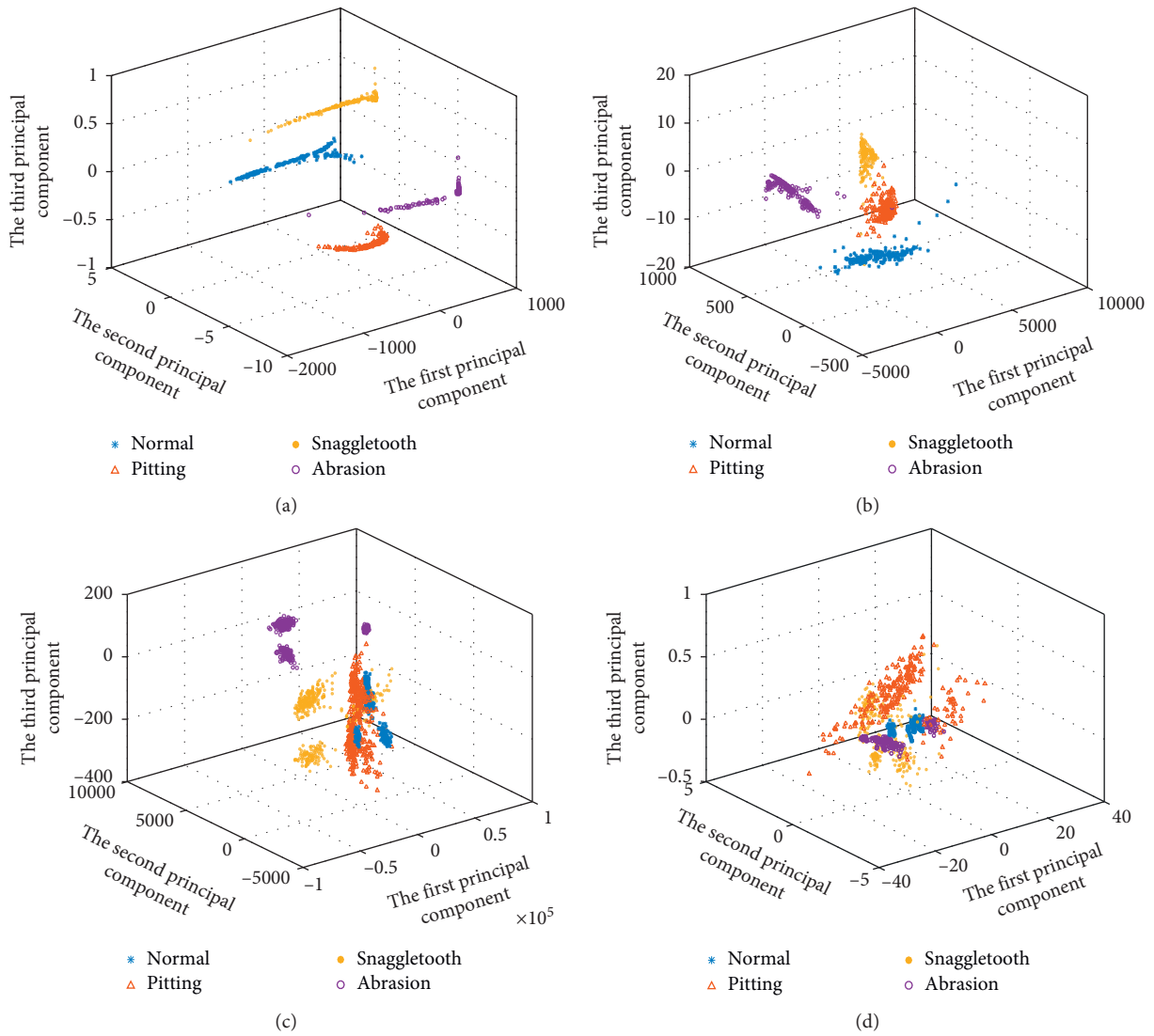
FIGURE 9: Visual scatter diagram of feature extraction. (a) The proposed method. (b) DBN set by experience. (c) Shallow probability network. (d) Traditional feature extraction method.
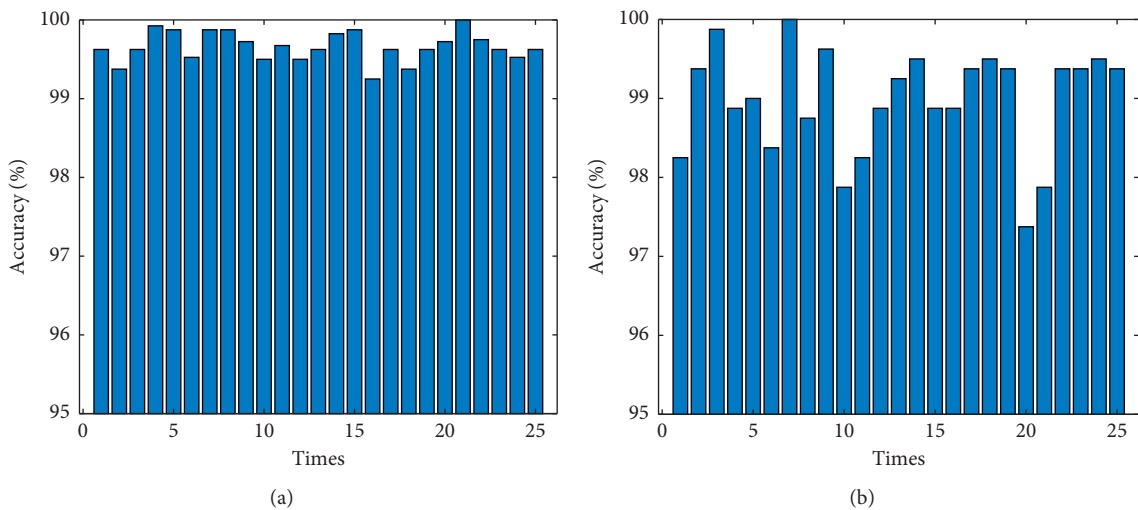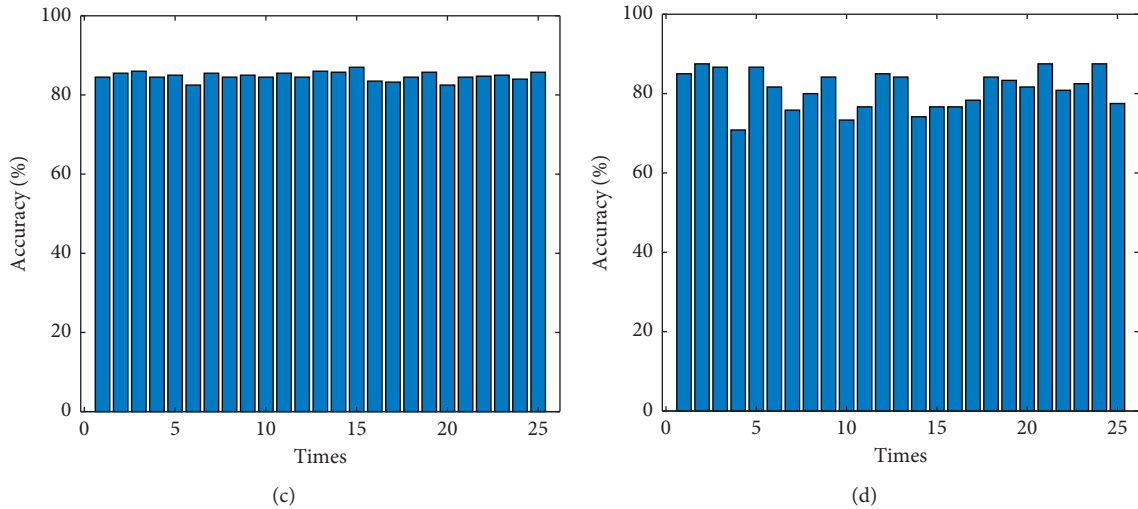

FIGURE 10: Continued.

FIGURE 10: Accuracy of different methods. (a) The proposed method. (b) DBN set by experience. (c) Shallow probability network. (d) Traditional fault diagnosis method.

discovered that the deep probability network is more capable for feature extraction, while serious overlapping exists among different states under shallow probability network. Figure 9(d) is associated with the traditional feature extraction method. By observing the scatter diagram, the distances between different states are too close with aliasing phenomenon, which is also a main reason for the poor diagnosis effect in traditional fault diagnosis.

4.2.4. Comparative Analysis with Other Methods. In order to verify the advantages in diagnosis accuracy, the diagnosis rates of the proposed method is compared with DBN set by experience, shallow probability network, and traditional feature extraction combined with ELM. 250 groups (the remaining 50% samples) from each of the four running states in the gearbox are randomly selected. In order to eliminate the errors and to verify the fault identification ability and stability of the model, the test is repeated for 25 times. The test results are as follows.

As illustrated in Figure 10(a), the accuracy of the fault diagnosis model established by the proposed method is higher than 99.5% among 25 random sampling tests, and the average diagnosis rate can reach 99.66%, indicating that the proposed method is characterized by the high diagnosis rate and stability for the gearbox under multiple working conditions. Figure 10(b) is the diagnosis rate of the DBN model with empirically selected parameters. The average diagnosis rate is 98.89%, slightly lower than the optimized DBN. Figure 10(c) shows the diagnosis rate of shallow probability network, with an average diagnosis rate of 84.79%. Compared with the shallow network, the deep network is more suitable for big data and self-adaptive fault diagnosis under complex working conditions. Figure 10(d) shows the diagnosis rate of traditional feature extraction combined with ELM, with an average diagnosis rate of 80.93%. Compared with the deep network model, traditional fault diagnosis methods lack in self-adaptive fault feature extraction, monitoring diagnosis accuracy, and generalization performance.

5. Conclusion

(1) A parameter-optimized DBN method was proposed to improve the feature extraction ability and fault diagnosis accuracy, in which the minimum RMSE in the network is considered as the fitness function, and the newly proposed GOA is properly employed to search for the optimal parameter combination.

(2) The parameter-optimized DBN method can self-adaptively extract fault information contained in the signal spectrum of the gearbox, avoiding the dependence on a large number of signal processing methods, and diagnosis experience, which has more advantages in fault diagnosis ability and generalization performance.

(3) A novel integrated fault diagnosis model based on FFT, linear normalization, and the optimized DBN is established, which provides a set of new intelligent fault diagnosis procedure. Through experimental analysis, this method is superior to shallow layer networks and traditional methods based on the combination of feature extraction and pattern recognition, which greatly contributes to the new era of intelligent fault diagnosis mode under "big data."

Data Availability

The data used in this manuscript are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jingbo Gai provided the main idea of the study; Junxian Shen analyzed the experiment and completed the paper; He

References

[1] C. Li, R.-V. Sanchez, G. Zurita et al., "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mechanical Systems and Signal Processing*, vol. 76-77, pp. 283–293, 2016.

[2] J. Gai and Y. Hu, "Research on fault diagnosis based on singular value decomposition and fuzzy neural network," *Shock and Vibration*, vol. 2018, Article ID 8218657, 7 pages, 2018.

[3] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[4] Y. G. Lei and Z. J. He, "Advances in applications of hybrid intelligent fault diagnosis and prognosis technique," *Journal of Vibration and Shock*, vol. 30, no. 9, pp. 129–135, 2011.

[5] C. F. Sun, Y. R. Wang, Y. Shen et al., "Fault diagnosis of planetary gearbox based on adaptive parameter variational mode decomposition," *Journal of Aerospace Power*, vol. 33, no. 11, pp. 205–214, 2018.

[6] M. F. Isham, M. S. Leong, M. H. Lim et al., "Intelligent wind turbine gearbox diagnosis using VMDEA and ELM," *Wind Energy*, vol. 22, no. 6, pp. 813–833, 2019.

[7] X. Zhang, Q. Z. Liu, and L. Wang, "An optimized time varying filtering based empirical mode decomposition method with grey wolf optimizer for machinery fault diagnosis," *Journal of Sound and Vibration*, vol. 418, pp. 55–78, 2018.

[8] Y. G. Lei, F. Jia, X. Zhou et al., "A deep learning-based method for machinery health monitoring with big data," *Journal of Mechanical Engineering*, vol. 51, no. 21, pp. 55–62, 2015.

[9] Q. Jin, Y. R. Wang, and J. Wang, "Planetary gearbox fault diagnosis based on multiple feature extraction and information fusion combined with deep learning," *Journal of Mechanical Engineering*, vol. 30, no. 2, pp. 74–82, 2019.

[10] Y. Lei, F. Jia, J. Lin et al., "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.

[11] H. Q. Wen, *A Fault Diagnosis Method Based on Deep Belief Network*, Beijing Jiaotong University, Beijing, China, 2018.

[12] S. Q. Zhang, *Research on Fault Diagnosis Technology Based on Deep Learning*, Harbin Institute of Technology, Harbin, China, 2018.

[13] J. Hu, Z. L. Zhang, and Y. Wang, "An intelligent fault diagnosis system for process plant using a functional HAZOP and DBN integrated methodology," *Engineering Applications of Artificial Intelligence*, vol. 45, pp. 119–135, 2015.

[14] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: a review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[15] B. Schölkopf, J. Platt, and T. Hofmann, "Greedy layer-wise training of deep networks," *Advances in Neural Information Processing Systems*, vol. 19, pp. 153–160, 2007.

[16] J. Ruan, P. R. Ukrainetz, and R. Burton, "Frequency domain modelling and identification of 2D digital servo valve," *International Journal of Fluid Power*, vol. 1, no. 2, pp. 49–59, 2000.

[17] C. J. Alonso-Gonzalez, N. Moya, and G. Biswas, "Dynamic bayesian network factors from possible conflicts for continuous system diagnosis," in *Advances in Artificial Intelligence*, Springer, Berlin, Germany, 2011.

[18] Y. F. Zhang, *A Research Based on Hybrid Methods GOA and LSSVM for Electric Load Forecasting*, Lanzhou University, Lanzhou, China, 2018.

[19] S. Saremi, A. S. Mirjalili, and A. Lewis, "Grasshopper optimisation algorithm: theory and application," *Advances in Engineering Software*, vol. 105, pp. 30–47, 2017.

[20] A. Saxena, "A comprehensive study of chaos embedded bridging mechanisms and crossover operators for grasshopper optimisation algorithm," *Expert Systems with Applications*, vol. 132, pp. 166–188, 2019.

[21] 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.