IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 5, 2020, accepted June 27, 2020, date of publication July 9, 2020, date of current version July 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3008208

An Intelligent Fault Diagnosis Method of Variable Condition Gearbox Based on Improved DBN Combined With WPEE and MPE

DONGYING HAN[1,2], XIAOCI GUO[3], AND PEIMING SHI[3]

[1]School of Vehicles and Energy, Yanshan University, Qinhuangdao 066004, China
[2]School of Mechanical Engineering, Yanshan University, Qinhuangdao 066004, China
[3]School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China

Corresponding author: Peiming Shi (spm@ysu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61973262 and Grant 51875500, in part by the Natural Science Foundation of Hebei Province under Grant E2019203146, and in part by the Project of introducing overseas talents in Hebei Province under Grant C20190516.

ABSTRACT Gear transmission is one of the most commonly used transmission methods in mechanical equipment. By analyzing the vibration data of gearbox, an improved deep belief network (DBN) algorithm for gear fault diagnosis based on wavelet packet energy entropy (WPEE) and multiscale permutation entropy (MPE) is proposed. Firstly, the vibration data of gearbox with various fault types under multiple working conditions are collected. Secondly, the energy entropy of wavelet packet and the entropy distribution of multiscale permutation are calculated respectively to form a combined feature matrix. Then, the improved threshold adaptive DBN is used to further extract the fault signal features, and finally the deep layer features are classified. By analyzing the vibration data of multi-platform gearbox, a high and stable diagnostic accuracy is obtained.

INDEX TERMS Gear fault, wavelet packet energy entropy, multiscale permutation entropy, deep belief network.

I. INTRODUCTION

Gearbox is a vital component of rotating machinery [1]. The complexity of operation environment makes gear fault occur frequently, which is very easy to cause equipment failure [2], [3]. The fault signal detection of gearbox under multiple working conditions is of significant practical meanings to monitor the occurrence of serious faults and guarantee the normal operation of mechanical equipment [4]–[6]. Because the change of gear operation condition will cause the change of vibration signal characteristics, and it is difficult to extract the gear fault characteristics under variable load effectively by traditional methods [7]–[10]. Therefore, in order to get rid of the dependence on expert experience and the tedious steps of feature extraction, it is very meaningful to find a simple intelligent diagnosis algorithm, which can realize the vibration state detection of multi working condition gearbox.

In recent years, deep learning develops rapidly [11]–[16], more and more people apply it to the field of gearbox fault diagnosis, and complete diagnosis and classification based on data-driven [17]–[23]. Wang et al [24] presented an intelligent fault diagnosis method combining generation countermeasure network and stack de-noising automatic encoder (SDAE) to identify fault modes of planetary gearbox from frequency spectrum. Han et al [25] fused the multistage wavelet packet coefficients of planetary gearbox with a dynamic integrated convolutional neural network (CNN) to diagnose faults.

Li et al [26] used the DBN to realize the fault diagnosis of the gearbox and bearing with time frequency characteristics. Chen et al [27] applied the CNN to sort out the health conditions of the gearbox with time frequency features. Shao et al [28] constructed an optimized DBN for the bearing diagnosis faults, and 18 time domain characteristics are enhanced. Guo et al [29] proposed a layered deep CNN to monitor the health states of bearings. Tang [30] studied the hybrid domain of wind turbine gearbox vibration and Shannon wavelet support vector machine to diagnose gearbox faults. Lei *et al.* [31] calculated 10 statistical parameters describing bearing state, and input these characteristics into neural network for fault classification. Chen et al [32]

The associate editor coordinating the review of this manuscript and approving it for publication was You Yang.

combined with empirical mode decomposition and Teager Kaiser energy operator to extract and select features at the same time, and used DBN to realize fault classification of planetary gearbox. Zhao et al [33] established a deep residual network to diagnose planetary gearbox faults based on wavelet packet coefficients in different frequency bands.

In this paper, an intelligent fault diagnosis method based on improved DBN model combined with WPEE and MPE is proposed. In view of the complex operation environment of gearbox, in order to get rid of the extreme dependence of traditional time-frequency feature diagnosis methods on artificial experience, the combination of wavelet packet energy entropy and multiscale permutation entropy are selected. To make the iterations of deep belief network set adaptively, the improved deep belief network is used for diagnosis and classification, so it gets rid of the problem. By analyzing the vibration signal of the fault gear with variable load and speed, the accuracy of fault diagnosis is greatly improved. In the result analysis of two examples, this paper compares the proposed method with some other existing classification algorithms, and optimizes the parameters such as diagnosis time, accuracy and stability.

II. THEORETICAL BACKGROUND AND THRESHOLD ADAPTIVE IMPROVED DBN METHOD

A. DBN MODEL

The DBN was proposed by Hinton in 2006 [28], [34], which is a probability generation model, as shown in Fig.1. It can be used as a classifier when there is supervised learning. DBN consists of several restricted Boltzmann machines (RBM) and softmax regression layers, as shown in Fig.2.

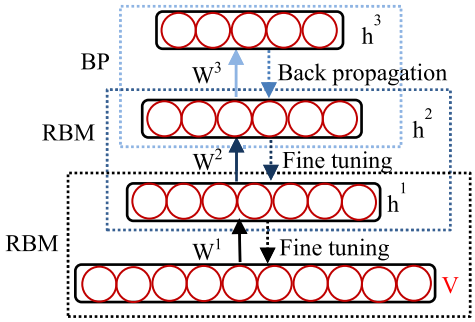


FIGURE 1. DBN structure diagram.

RBM is a randomly generated neural network structure, which is intrinsically an undirected graph model composed of one layer of visible neurons and one layer of hidden neurons. It only has a connection between hidden layer and visible layer neurons but no connection between visible layer neurons and hidden layer neurons, as shown in Fig. 2. Moreover, hidden layer neurons generally take binary and obey Bernoulli distribution. The vibration signal data are real values, which can be selected as the input of the visible layer neurons.

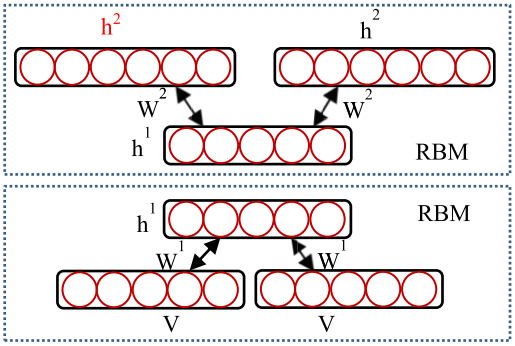


FIGURE 2. RBM structure diagram.

The DBN model mainly consists of two steps: unsupervised pre training based on RBM and supervised reverse optimization.

Step 1: unsupervised pre training based on RBM. Using the CD-K algorithm (contrast divergence algorithm) to initialize the weights is as follows. Hinton found that when K is 1, it can get a better learning effect.

1) Random initialization weights $\{W, a, b\}$, where X, W and a are the input sample data, the weight vector and the bias vector of the visible layer respectively, and b is the bias vector of the hidden layer.

$$X = v = \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ v_M \end{bmatrix}, h = \begin{bmatrix} h_1 \\ h_2 \\ \cdots \\ h_N \end{bmatrix},$$
$$W = \begin{bmatrix} W_{1,1} & W_{2,1} & \cdots & W_{M,1} \\ W_{1,2} & W_{2,2} & & W_{M,2} \\ \cdots & \cdots & \cdots & \cdots \\ W_{1,N} & W_{2,N} & \cdots & W_{M,N} \end{bmatrix},$$
$$a = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_M \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_N \end{bmatrix} \quad (1)$$

where, M and N are the number of neurons in the visible layer and the hidden layer, respectively. Initialize W based on the normal distribution N (0, 0.01), and initialize a_i according to $a_i = \log \frac{p_i}{1-a_1}$, where p_i is the proportion of the samples whose the i-th sample is in the active state (i.e. the value is 1), and b is initialized to 0. The calculation of the activation value h of the hidden element is:

$$h = (W \cdot X + b)$$
$$= \begin{bmatrix} W_{1,1} \cdot v_1 + W_{2,1} \cdot v_2 + \ldots + W_{M,1} \cdot v_M \\ W_{1,2} \cdot v_1 + W_{2,2} \cdot v_2 + \ldots + W_{M,2} \cdot v_M \\ \cdots \quad \cdots \quad \cdots \\ W_{1,N} \cdot v_1 + W_{2,N} \cdot v_2 + \ldots + W_{M,N} \cdot v_M \end{bmatrix}$$
$$+ \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_N \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ \cdots \\ h_N \end{bmatrix} \quad (2)$$

The calculation of visual layer neuron value v can be expressed as follows:

$$v = \left(W^T \cdot h + a\right) = \begin{bmatrix} W_{1,1} \cdot h_1 + W_{1,2} \cdot h_2 + \ldots + W_{1,N} \cdot h_N \\ W_{2,1} \cdot h_1 + W_{2,2} \cdot h_2 + \ldots + W_{2,N} \cdot h_N \\ \cdots \quad \cdots \quad \cdots \\ W_{M,1} \cdot h_1 + W_{M,2} \cdot h_2 + \ldots + W_{M,N} \cdot h_N \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_M \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ v_M \end{bmatrix} \tag{3}$$

2) Assign X to $v^{(0)}$, and calculate the probability that it turns on the hidden element:

$$p\left(h_j^{(0)} = 1 \mid v^{(0)}\right) = \sigma\left(W_j \cdot v^{(0)} + b_j\right) \tag{4}$$

The superscript in equation (4) represents the vector, and the subscript represents the dimension of the vector.

3) Perform one-step Gibbs sampling based on the obtained probability distribution, and the corresponding value is selected for each unit of the hidden layer from {0, 1}, which satisfies $h^{(0)} \sim p(h^{(0)}|v^{(0)})$. The process is as follows:

First generate a random number r_j on [0, 1], then determine the value of h_j as follows:

$$h_j = \begin{cases} 1, & \text{if } p\left(h_j^{(0)} = 1 \mid v^{(0)}\right) > r_j \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

4) Reconstruct the visible layer with $h^{(0)}$, first calculate the probability density, and then perform Gibbs sampling. The visual layer neuron is a Bernoulli visual layer neuron, so it satisfies:

$$p\left(v_i^{(1)} = 1 \mid h^{(0)}\right) = \sigma(W_i^T h^{(0)} + a_i) \tag{6}$$

5) Based on the obtained probability distribution, perform another Gibbs sampling to select the corresponding values from the {0, 1} neurons in the visible layer for sampling reconstruction, that is, first generate {0, 1} random number, then determine the value:

$$v_i = \begin{cases} 1, & \text{if } p\left(v_i^{(1)} = 1 \mid h^{(0)}\right) > r_j \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

6) Compute the probability of the neuron being turned on using the reconstructed explicit element according to the following formula:

$$p\left(h_j^{(1)} = 1 \mid v^{(1)}\right) = \sigma\left(W_j \cdot v^{(1)} + b_j\right) \tag{8}$$

7) Update the weights and offsets according to the following formula to get new weights and offsets:

$$\begin{aligned} \mathrm{W} &\leftarrow \mathrm{W} + \lambda[\mathrm{p}\left(h^{(0)} = 1 \mid v^{(0)}\right) v^{(0)T} \\ &\quad -\mathrm{p}\left(h^{(1)} = 1 \mid v^{(1)}\right) v^{(1)\mathrm{T}}] \\ \mathrm{b} &\leftarrow \mathrm{b} + \lambda[\mathrm{p}\left(h^{(0)} = 1 \mid v^{(0)}\right) \\ &\quad -\mathrm{p}\left(h^{(1)} = 1 \mid v^{(1)}\right)] \\ \mathrm{a} &\leftarrow \mathrm{a} + \lambda\left[v^{(0)} - v^{(1)}\right] \end{aligned} \tag{9}$$

where λ represents the learning rate.

This step completes the supervised pre-training process on a basis of RBM.

Step 2: supervised reverse parameter adjustment. In this step, a certain output is obtained with the forward propagation algorithm, and then the weight value and bias value are updated by using the back propagation algorithm.

1) Get output values in forward propagation. First, the $\{W, b\}$ determined in *Step 1* is used to determine whether the hidden element is open or not. Then the activation value can be expressed as:

$$h^{(l)} = W^{(l)} \cdot v + b^{(l)} \tag{10}$$

where l represents the layer index of neural network. The W and b can be expressed as:

$$W = \begin{bmatrix} W_{1,1} & W_{2,1} \ldots & W_{M,1} \\ W_{1,2} & W_{2,2} \quad \ldots & W_{M,2} \\ \cdots & \cdots \quad \ldots & \cdots \\ W_{1,N} & W_{2,N} \ldots & W_{M,N} \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_N \end{bmatrix} \tag{11}$$

where $W_{i,j}$ refers to the weight between the i-th visible layer neuron and the j-th hidden layer neuron.

The activation values of each hidden layer neuron are calculated in turn. The sigmoid function is used as the activation function to complete the standardization, as follows:

$$\sigma(h_j)^{(l)} = \frac{1}{1 + e^{-h_j}} \tag{12}$$

Finally, the activation value $h^{(l)}$ and output of the output layer $\hat{X}$ can be obtained:

$$\begin{cases} h^{(l)} = W^{(l)} \cdot h^{(l-1)} + b^{(l)} \\ \hat{X} = f\left(h^{(l)}\right) \end{cases} \tag{13}$$

where $f(\cdot)$ represents activation function of the output layer, and the activation functions used in this model are sigmoid functions.

2) Update weight and offset values in back propagation. Firstly, update all network parameters with backward propagation algorithm following the principle of minimum mean square error, and the expression of cost function is as follows:

$$E = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{X}_i\left(W^{(l)}, b^{(l)}\right) - X_i\right)^2 \tag{14}$$

where E represents the mean square error of learning, $\hat{X}_i$ and X_i are the output of output layer and ideal output respectively, and i represents the sample index. $\left(W^{(l)}, b^{(l)}\right)$ represent the weights to be learned and the parameters of the offsets in the l layer.

Finally, the gradient descent method is selected to update the weight and bias parameters, as shown below:

$$\left(W^{(l)}, b^{(l)}\right) \leftarrow \left(W^{(l)}, b^{(l)}\right) - \lambda \cdot \frac{\partial E}{\partial \left(W^{(l)}, b^{(l)}\right)} \quad (15)$$

where λ is the learning rate.

The final DBN model can be obtained by selecting the appropriate number of hidden layers, layer neurons and learning rate, and then iterating a certain number of times for training.

B. EARLY STOPPING BASED ON ITERATIVE ERROR THRESHOLD

In view of the over fitting phenomenon of DBN model due to over training, a method of early stopping based on iterative error threshold is proposed to prevent over fitting in this paper. In references [35] and [36], Kamada and Carlson *et al.* found that hidden layer bias b plays a key role in RBM training convergence, and has the following relationship:

$$F\left(\left\{a^k, b, W^k\right\}\right) \leq F\left(\theta^K\right) + \langle \nabla_b F(\theta^k), b - b^k \rangle + \frac{n}{2}\left\|b - b^k\right\|_\infty^2 \quad (16)$$

which is derived from the Lipschitz continuous (please see [28] for details), where, $\left\{a^k, b, W^k\right\}$ represent the bias vector of visible layer, the bias vector of hidden layer and weight respectively.

Through observation, it is found that the optimal model training result can be achieved when the relationship between cumulative iteration error and hidden layer bias b satisfies the formula (17):

$$\varepsilon \leq \frac{n}{2} \times \sum_{i=1}^{n} b_i^2 \quad (17)$$

where ε is the iteration error of single iterative training for a single RBM according to n training samples, and b is the bias vector of the hidden layer.

Therefore, this paper proposes to determine the stop time by observing the cumulative error ε of multiple iterations, and proposes to use the hidden layer bias b to judge whether the threshold of cumulative error ε is reached and stop the iteration.

Therefore, it is proposed to use the hidden layer bias b to judge whether the accumulated error ε threshold is reached and stop the iteration. The implementation process is shown in Fig. 3. Among them, the meaning of v_1, v_2, b are mentioned in formula (1).

The main implementation steps of iterative optimization algorithm on the basis of loss threshold can be described as:

(1) Parameter setting before RBM pre training: set the initial value i of the number of iterations and the maximum value T of the number of iterations according to the experience value, so as to ensure that the model is optimal after T cycles.

(2) Start RBM pre training. After *Step 1* in Section 2.1 of Chapter 2, the single training iteration error is obtained according to formula (18) below.

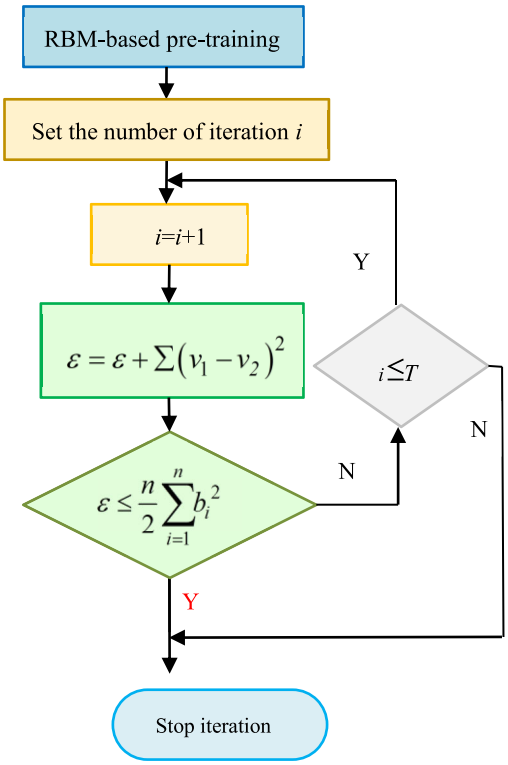


FIGURE 3. Variation of iteration error and threshold for each layer during layer-by-layer training.

$$\varepsilon = \sum (v_1 - v_2)^2 \quad (18)$$

(3) Determine whether $\varepsilon \leq \frac{n}{2} \times \sum_{i=1}^{n} b_i^2$ is satisfied. If the result is true, the model training is considered to be optimal and the iteration is stopped. Otherwise, continue the iteration according to the new iteration error, where the iteration of ε is realized according to the following formula:

$$\varepsilon = \varepsilon + \sum (v_1 - v_2)^2 \quad (19)$$

Repeat *Step 2* until the number of iterations i reaches the preset maximum number T, RBM pre training stops.

Fig. 4 can be obtained by data verification. It is found that the model stops iteration when ε satisfies formula (17), which not only guarantees the sufficiency of model training, but also greatly shortens the training time. At the same time, the setting of training times t also achieves the adaptive effect. The number of iterations corresponding to the intersection of the two lines is the number of iterations stopped during RBM pre training.

Taking the gear vibration signals collected on the QPZZ-II platform as an example, when classifying gear diagnosis signals of 6 states, on the premise that the optimal diagnosis results can be achieved, an

optimization algorithm based on the ε threshold and a traditional algorithm are used. Perform classification experiments

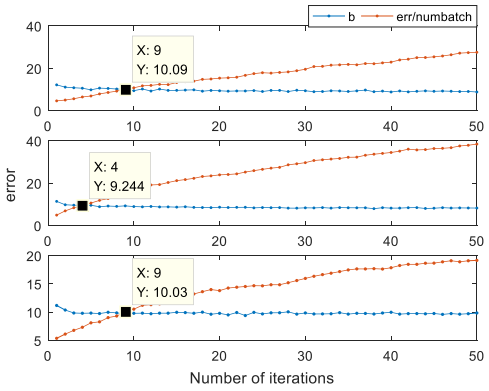


FIGURE 4. Variation of iteration error and threshold for each layer during layer-by-layer training.

on the signals. The diagnostic time and diagnostic accuracy are recorded in the tables below:

From Tables 1 and Table 2, it can be found that when an optimization algorithm based on the ε threshold is used to ensure higher diagnostic accuracy. Compared with traditional algorithms, the diagnosis time is also greatly reduced, and the ideal effect of preventing overfitting is achieved.

TABLE 1. Comparison 1 of experimental results under different methods.

Traditional Method | Article Method	MPE+DBN	WPEE+DBN	WPEE-MPE+DBN
Time(s)	54|41	31|22	41|31
Accuracy (%)	90.37|94.44	98.89|98.93	100|100

TABLE 2. Comparison 2 of experimental results under different methods.

Traditional Method | Article Method	MPE+DBN	WPEE+DBN	WPEE-MPE+DBN
Time(s)	142|140	94|40	113|58
Accuracy (%)	68.39|70.83	97.65|98.70	100|99.97

C. WAVELET PACKET ENERGY ENTROPY

When the high frequency band of wavelet transform is further decomposed, the signal can be analyzed more precisely, and the frequency band can be divided in the full band range, further improving the frequency resolution [37], [38]. Therefore, the wavelet packet decomposition based on the orthogonal wavelet basis function can not only decompose the signal in the low-frequency and high-frequency parts at the same time but also select the resolution of the signal in different frequency bands adaptively. The signals in each decomposition frequency band are independent of each other without redundancy or omission. The wavelet packet decomposition obeys the energy conservation criterion [39], [40]. The expression orthogonal wavelet packet decomposition is as follows:

$$\begin{cases} c_{2n}(t) = 2^{1/2} \sum_{k \in Z} h(k)\, c_n(2t-k) \\ c_{2n+1}(t) = 2^{1/2} \sum_{k \in Z} g(k)\, c_n(2t-k) \end{cases} \tag{20}$$

where $h(k)$ and $g(k)$ represent high pass filter and low pass filter respectively, $g(\mathrm{k}) = (-1)^k h(1-k)$ and the two coefficients are orthogonal.

When $n = 0$, $c_0(t)$ degenerates to scale function $\varphi(t)$ and $c_1(t)$ degenerates to wavelet basis function $\psi(t)$. Therefore, the function system $\{c_n(t)\}$ is named the orthogonal wavelet packet.

If the data length of the original signal $x(\mathrm{t})$ is N, the data length of the discrete signal $x^{k,m}(i)$ in the decomposition band is reduced to $2^{-k}N$, and its energy can be expressed as:

$$\mathrm{E}\left(x^{k,m}(i)\right) = \frac{1}{2^{-k}(N-1)} \sum\nolimits_{i=1}^{2^{-k}N} \left(x^{k,m}(i)\right)^2 \tag{21}$$

where k represents the number of decomposition, $m = 0, 1, 2, \ldots, 2^k - 1$, indicating the position sequence number of the decomposition band. The relative energy of the m-band decomposition signal can be expressed as:

$$E_m = \frac{E(x^{k,m}(i))}{E(x(t))} \tag{22}$$

In the formula (22), $E(x(\mathrm{t}))$ is the sum of total energy. According to the principle of conservation of energy, $\sum_{m=0}^{2^{-k}} E_m = 1$.

WPEE is to take the normalized energy feature of wavelet packet as the probability distribution of vibration signal, carry out information entropy operation and extract the vibration signal feature [41]–[43]. The advantage of this method is that one-dimensional data of energy entropy is used to replace the number of eigenvectors that may be multi-dimensional, which makes the algorithm simpler and faster when it is used for state monitoring and prediction. In the fault feature extraction of bearing, the method of obtaining the wavelet packet energy entropy can be described as:

(1) The fault data of rolling bearing is decomposed by wavelet packet in layer i, and the signals of 2^i frequency bands are obtained, which are arranged in order from low frequency band to high frequency band.

(2) After the coefficients of each band are reconstructed, the band energy is computed; the total energy of each layer is obtained by accumulation. The total energy of layer I was normalized.

(3) According to formula (22), the information entropy is calculated, and the eigenvectors of each state of vibration signal are obtained:

$$E_p = \{E_m, m = 1, 2, \ldots, M\} \tag{23}$$

where, p represents the serial number of signal acquisition, $p = 1, 2, 3, \ldots, p$.

(4) Definition of WPEE: due to different wavelet packet energy distribution of different faults, each fault state can be represented by different H_E.

D. MULTISCALE PERMUTATION ENTROPY

Multiscale permutation entropy is defined as permutation entropy at different scales [44]–[46]. This method can be described as:

(1) There is a time-series signal $\{x(i), i = 1, 2, \ldots, N\}$, and coarse-grained processing is performed to obtain a coarse-grained sequence $\{y_j^{\mathrm{T}}\}$.

y_j^{T} is expressed as:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i \quad j = 1, 2, \ldots, [N/\tau] \qquad (24)$$

where $[N/\tau]$ means rounding N/τ and it is a scale factor, $\tau = 1, 2, \ldots$. Obviously, the coarse-grained sequence is the original sequence at $\tau = 1$; the original sequence is coarse-grained into a coarse-grained sequence with a length of $[N/\tau]$ at $\tau > 1$.

(2) The permutation entropy of each coarse-grained sequence is calculated and plotted as a function of scale factor. The process is called multiscale permutation entropy analysis.

Generally, the maximum value of scale factor is more than 10 [47], [48]. To keep entropy independent of the length of coarse-grained sequence $[N/\tau]$, $\tau = 12$ is selected in this paper. If the value of m is too small, PE value will decrease with the increase of scale factor, but the larger m is, the more time-consuming the calculation is. In this paper, $m = 4$ is selected.

III. FAULT DIAGNOSIS METHOD

In view of the dependence of traditional feature extraction methods on expert experience and the long time-consuming analysis of vibration data, a DBN fault diagnosis method based on WPEE-MPE is proposed. The implementation process is shown in Fig.5.

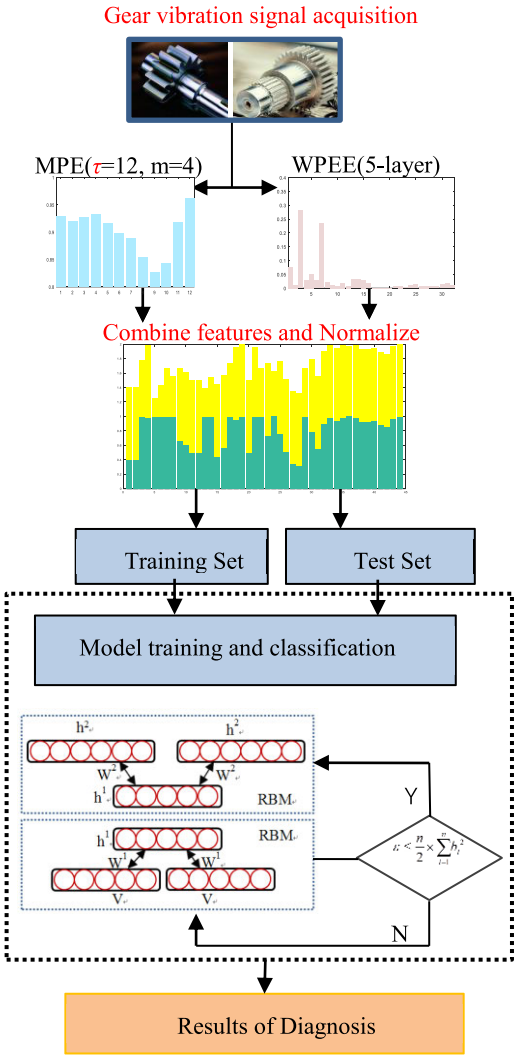


FIGURE 5. Flow chart of the proposed method.

IV. EXAMPLES OF DIAGNOSIS

To demonstrate the effectiveness of this method in gear fault diagnosis, this paper performs fault diagnosis experiments on the rotary machinery vibration analysis fault diagnosis test platform system named QPZZ-II and a multi-stage gear transmission system test bench. The device is shown in Fig.6 and Fig.7.

Fig. 6 is a QPZZ-II rotating machinery fault diagnosis test platform, which can simulate single gear fault such as large gear pitting, large gear broken teeth, pinion wear, and compound gear fault such as large gear broken tooth compound small gear wear and Gear pitting compound pinion wear fault. Fig.7 is a fixed-shaft variable-speed gearbox. The test bench can simulate single fault of various gearboxes, such as missing gear teeth, gear cutting, and root cracks. This paper collects vibration data through these two platforms, and uses

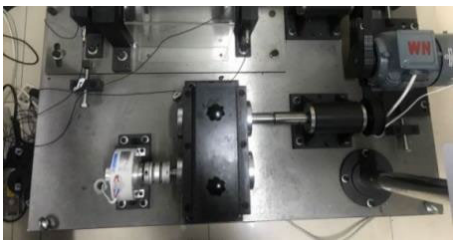


FIGURE 6. QPZZ-II test platform.

the method proposed in this paper to diagnose and classify them.

A. GEARBOX DATA OF QPZZ-II TEST PLATFORM

The types of gear fault collected on the QPZZ-II test platform include normal status signals, three single fault signals, and two composite fault signals, which are the gear pitting fault signal, the gear breakage fault signal, the pinion wear fault

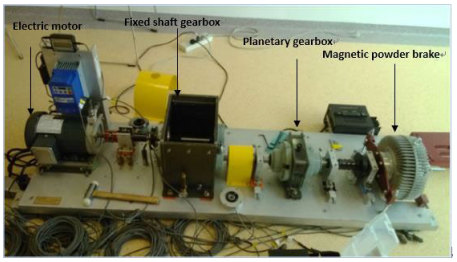


FIGURE 7. Fixed-shaft variable speed gearbox.

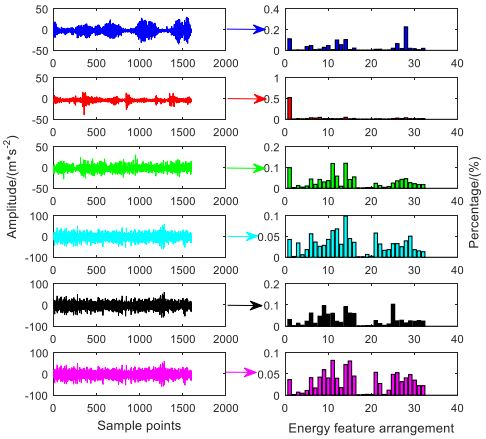


FIGURE 8. WPEE feature distribution of the original signal.

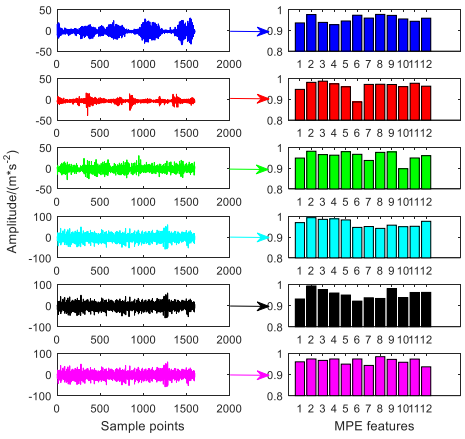


FIGURE 9. MPE feature distribution of the original signal.

signal, and the gear broken tooth compound pinion wear fault signal, large gear pitting compound pinion wear fault signal, a total of 6 states of gearbox vibration signal data. For these six signals, the wavelet packet energy entropy and multi-scale permutation entropy distribution are obtained, as shown in Fig.8 and Fig.9, and they are combined to form a feature matrix as shown in Fig.10. Normalize the data according to formula (25), and use them as training samples and test samples.

$$y = \frac{(y_{max} - y_{min}) \times (x - x_{min})}{x_{max} - x_{min}} + y_{min} \tag{25}$$

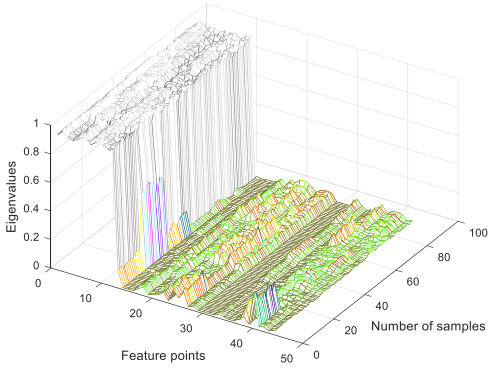


FIGURE 10. WPEE-MPE-based combined feature distribution.

The training samples are input to the DBN neural network, and the threshold is set according to formula (17) to prevent over fitting of a single RBM during training. Where ε represents the cumulative error of a single iteration training of a single RBM according to n training samples, and b represents the bias of the hidden layer. Comparison results of this method and the other three methods are shown in Table 3 and Fig. 11.

TABLE 3. Comparison results of this method and the other three methods.

Method	Results of 15 diagnoses			Number of samples
	Accuracy (100%)	Time (s)	Standard deviation	
MPE+SVM	90.52	56	0.1725	160
MPE+DBN	90.37.	41	0.1522	240
WPEE+DBN	98.07	25	0.0098	240
WPEE+MPE +DBN	100	32	0	160

In Fig. 11, the method 1 is to select MPE for feature extraction and SVM for fault classification; the method 2 is to select MPE for feature extraction and DBN neural network for fault classification; the method 3 is to select WPEE for feature extraction and DBN for fault classification; the last method is to select WPEE-MPE for feature extraction and improved DBN for fault classification. From Table 3 and Fig. 11, it can be found that for example 1 data, when six types of faults are classified, the method proposed in this paper can get 100% diagnosis results, and achieve stable and reliable diagnosis results at a lower time cost.

B. MEDIUM SPEED GEARBOX DATA

Fig. 7 is a fixed-shaft variable-speed gearbox. This test bench can simulate several single fault of the gearbox under various operating conditions, such as missing gear teeth, gear cutting, cracks on the roots, etc. The vibration signals in the four gear states collected under single working conditions and multiple working conditions with variable speed and variable load were analyzed to demonstrate the effectiveness of the proposed method. First, by calculating the wavelet packet energy entropy distribution of the vibration signal and

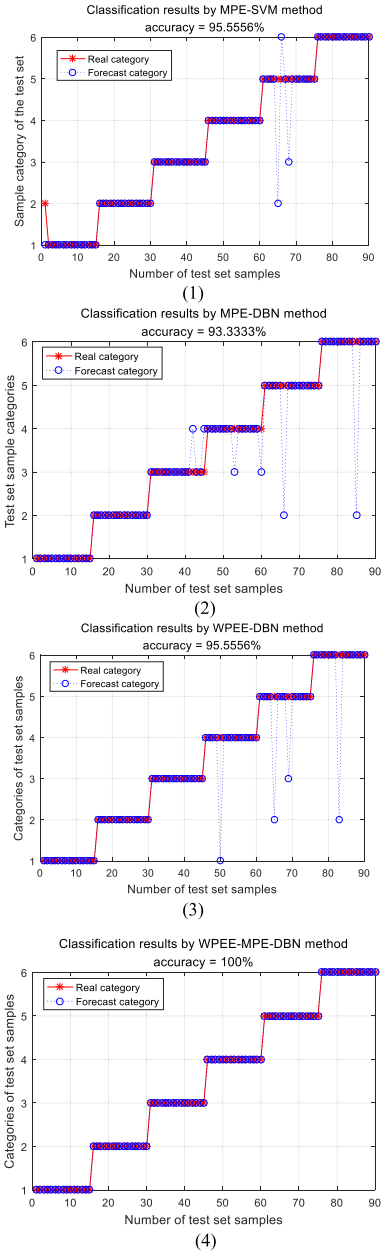


FIGURE 11. Classification results of multiple methods.

the multiscale permutation entropy distribution, the feature matrix is data normalized according to equation (25) to form a feature matrix. Then determine training and test samples; adjust the parameters of the DBN neural network; input training samples into it and prevent overfitting during RBM training according to formula (17) and further generate a model. Finally, using the trained model to classify the test samples.

First, classify the vibration data under a single operating condition. The signal takes vibration data with a speed of 880 r/min under no load as the signal sample trains the model through the above steps, and then inputs the test sample into the model to obtain the classification result shown in Fig.12. Among them, the first label represents a vibration signal in a normal state, and the second label, the third label, and the last label respectively represent vibration signals when a gear is missing a gear, a gear is cut, and a tooth root is cracked. It can be found that because the working conditions are single and the signal distribution is relatively simple, when the proposed method is applied to classification, not only can a high accuracy rate be obtained, but the diagnosis results are very stable, and the standard deviation of the accuracy of multiple diagnosis is 0.

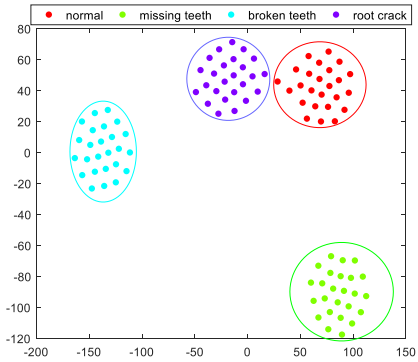
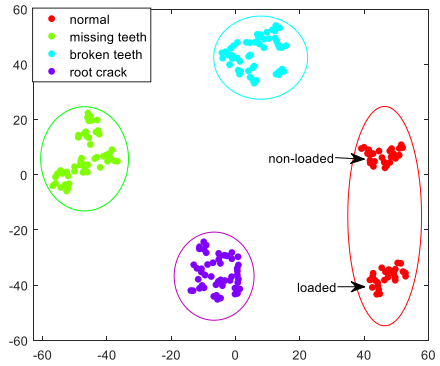


FIGURE 12. Classification of single working conditions.

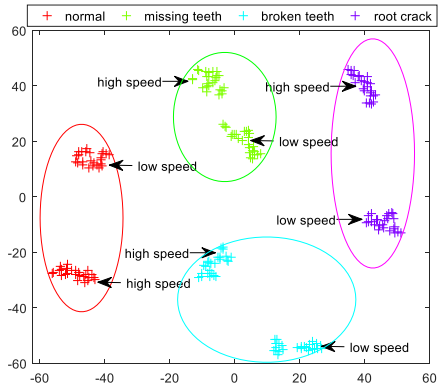
When classifying complex data of variable load, although the signals are signals of the same fault type, but the operating conditions are different, even if the vibration data of the same fault type, the distribution will be dispersed, as shown in Fig.13(1) and Fig.13(2). This situation is more obvious under the condition of variable speed, as shown in Fig.13(2). Experiments show that the proposed model can realize high diagnostic accuracy and stable diagnosis results, no matter under single operating conditions or multiple operating conditions with variable speed and variable load.

Fig.14 is shown the distribution of the number of iteration stops after optimization of 15 repeated experiments under variable speed and variable load conditions, it can be seen that after threshold-based adaptation is used to prevent over fitting, in repeated experiments with variable speed, the solid blue line indicates the change in the number of iteration stops during training after threshold adaptation is used. The maximum number of stops is 69. Therefore, it can be considered that when the adaptive method is not used, the model must be trained at least 69 epochs to train the model to the optimal. But in some trainings, it only takes far less than 69 epochs to get the optimal model. Too many epochs will not only waste a lot of time, but also cause over fitting, making the classification results inaccurate. This is the significance of the threshold-based adaptive prevention of over fitting mentioned in this article.

Comparison of the accuracy and standard deviation of multi-method and multi-condition diagnosis are shown in Fig.15 and in Table 4, in order to highlight the standard deviation difference, the standard deviation of all methods is expressed as 100 times the original standard deviation. The



(1) The classification results when the rotation speed is 90% and the loading is 0 and 20% respectively



(2) Classification results when the load is 0 and the speed is 90% and 50% respectively

FIGURE 13. Classification results of gear fault under variable load condition.

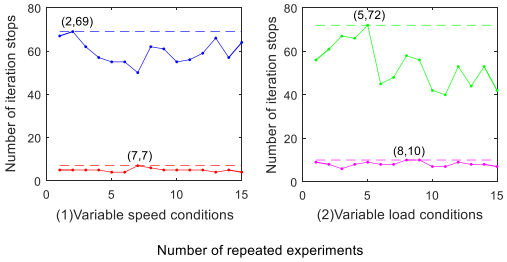


FIGURE 14. Distribution of the number of iteration stops after optimization of 15 repeated experiments under variable speed and variable load conditions.

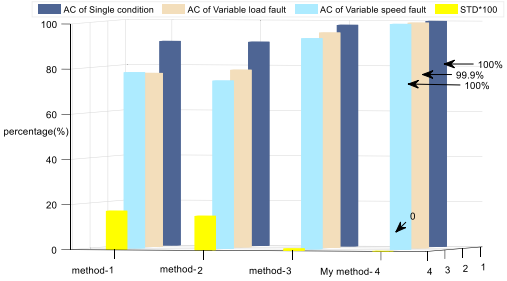


FIGURE 15. Comparison of the accuracy and standard deviation of multi-method and multi-condition diagnosis.

method proposed in this article is Method 4. Through comparison, it is found that the standard deviation of the accuracy of the 15 diagnostic tests in a single working condition is 0,

TABLE 4. Comparison of accuracy and standard deviation of multi-method and multi-condition diagnosis.

Method	15 diagnostic accuracy (%)			Standard deviation in a single working condition
	Single working condition	Variable load	Variable speed	
MPE+SVM	90.52	77	78	0.1725
MPE+DBN	90.37.	78.63	74.50	0.1522
WPEE+DBN	98.07	95.33	93.47	0.0098
WPEE+MPE +DBN	100	99.90	100	0

that is, in 15 consecutive diagnoses; the model classifies all samples correctly, so the model is considered to have high application reliability.

V. CONCLUSION

The gearbox has a complicated operating environment. To improve the effectiveness and ease of feature extraction of gear vibration signals, a feature extraction method based on WPEE and MPE is proposed, and the iteration number is adaptive through the method of preventing overfitting based on the loss threshold. Set up the trained model to achieve the best convergence effect and complete the fault classification based on DBN. The validity and stability of the proposed method are confirmed by two sets of variable-mode vibration data collected on the QPZZ-II rotating machinery vibration analysis fault diagnosis test platform system and a company's multi-stage gear transmission system test bench. The analysis results show that this method not only simplifies the feature extraction step of the gearbox vibration signal, not only can accurately diagnose and classify the fault signal under varying conditions, but also greatly reduces the time spent on diagnosis.

REFERENCES

[1] Y. Lei, J. Lin, M. J. Zuo, and Z. He, "Condition monitoring and fault diagnosis of planetary gearboxes: A review," *Measurement*, vol. 48, pp. 292–305, Feb. 2014.
[2] P. Shi, W. Zhang, D. Han, and M. Li, "Stochastic resonance in a high-order time-delayed feedback tristable dynamic system and its application," *Chaos, Solitons Fractals*, vol. 128, pp. 155–166, Nov. 2019.
[3] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.
[4] F. Liu, C. Shen, Q. He, A. Zhang, Y. Liu, and F. Kong, "Wayside bearing fault diagnosis based on a data-driven Doppler effect eliminator and transient model analysis," *Sensors*, vol. 14, no. 5, pp. 8096–8125, May 2014.
[5] S. Lu, R. Yan, Y. Liu, and Q. Wang, "Tacholess speed estimation in order tracking: A review with application to rotating machine fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 7, pp. 2315–2332, Jul. 2019.
[6] S. Lu, P. Zheng, Y. Liu, Z. Cao, H. Yang, and Q. Wang, "Sound-aided vibration weak signal enhancement for bearing fault detection by using adaptive stochastic resonance," *J. Sound Vib.*, vol. 449, pp. 18–29, Jun. 2019.
[7] D. Li, S. Liu, and H. Zhang, "Negative selection algorithm with constant detectors for anomaly detection," *Appl. Soft Comput.*, vol. 36, pp. 618–632, Nov. 2015.
[8] B. Hu, C. Guo, J. Wu, J. Tang, J. Zhang, and Y. Wang, "An adaptive periodical stochastic resonance method based on the grey wolf optimizer algorithm and its application in rolling bearing fault diagnosis," *J. Vib. Acoust.*, vol. 141, no. 4, Aug. 2019, Art. no. 041016.

[9] Z. Qiao, Y. Lei, and N. Li, "Applications of stochastic resonance to machinery fault detection: A review and tutorial," *Mech. Syst. Signal Process.*, vol. 122, pp. 502–536, May 2019.
[10] Z. Qiao, Y. Lei, J. Lin, and F. Jia, "An adaptive unsaturated bistable stochastic resonance method and its application in mechanical fault diagnosis," *Mech. Syst. Signal Process.*, vol. 84, pp. 731–746, Feb. 2017.
[11] Z. Du, X. Chen, H. Zhang, and R. Yan, "Sparse feature identification based on union of redundant dictionary for wind turbine gearbox fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6594–6605, Oct. 2015.
[12] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mech. Syst. Signal Process.*, vols. 76–77, pp. 283–293, Aug. 2016.
[13] D. Zappalá, P. J. Tavner, S. Sheng, and C. J. Crabtree, "Side-band algorithm for automatic wind turbine gearbox fault detection and diagnosis," *IET Renew. Power Gener.*, vol. 8, no. 4, pp. 380–389, May 2014.
[14] S. Tang, S. Yuan, and Y. Zhu, "Deep learning-based intelligent fault diagnosis methods toward rotating machinery," *IEEE Access*, vol. 8, pp. 9335–9346, 2020.
[15] J. Singh, M. Azamfar, A. Ainapure, and J. Lee, "Deep learning-based cross-domain adaptation for gearbox fault diagnosis under variable speed conditions," *Meas. Sci. Technol.*, vol. 31, no. 5, May 2020, Art. no. 055601.
[16] S. Haidong, C. Junsheng, J. Hongkai, Y. Yu, and W. Zhantao, "Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing," *Knowl.-Based Syst.*, vol. 188, Jan. 2020, Art. no. 105022.
[17] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[18] 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," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016.
[19] C. Sankavaram, A. Kodali, K. R. Pattipati, and S. Singh, "Incremental classifiers for data-driven fault diagnosis applied to automotive systems," *IEEE Access*, vol. 3, pp. 407–419, 2015.
[20] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.
[21] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, Jan. 2017.
[22] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017.
[23] Q. He, J. Zhao, G. Jiang, and P. Xie, "An unsupervised multiview sparse filtering approach for current-based wind turbine gearbox fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 8, pp. 5569–5578, Aug. 2020.
[24] Z. Wang, J. Wang, and Y. Wang, "An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition," *Neurocomputing*, vol. 310, pp. 213–222, Oct. 2018.
[25] Y. Han, B. Tang, and L. Deng, "Multi-level wavelet packet fusion in dynamic ensemble convolutional neural network for fault diagnosis," *Measurement*, vol. 127, pp. 246–255, Oct. 2018.
[26] C. Li, R.-V. Sánchez, G. Zurita, M. Cerrada, and D. Cabrera, "Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning," *Sensors*, vol. 16, no. 6, p. 895, Jun. 2016.
[27] Z. Chen, C. Li, and R.-V. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," *Shock Vib.*, vol. 2015, pp. 1–10, Oct. 2015.
[28] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, Nov. 2015, Art. no. 115002.
[29] 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.
[30] B. Tang, T. Song, F. Li, and L. Deng, "Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine," *Renew. Energy*, vol. 62, pp. 1–9, Feb. 2014.
[31] Y. Lei, Z. He, and Y. Zi, "EEMD method and WNN for fault diagnosis of locomotive roller bearings," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7334–7341, Jun. 2011.
[32] H. Chen, J. Wang, B. Tang, K. Xiao, and J. Li, "An integrated approach to planetary gearbox fault diagnosis using deep belief networks," *Meas. Sci. Technol.*, vol. 28, no. 2, Feb. 2017, Art. no. 025010.
[33] M. Zhao, M. Kang, B. Tang, and M. Pecht, "Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4290–4300, May 2018.
[34] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, and S. Van Hoecke, "Convolutional neural network based fault detection for rotating machinery," *J. Sound Vib.*, vol. 377, pp. 331–345, Sep. 2016.
[35] S. Kamada, T. Ichimura, A. Hara, and K. J. Mackin, "Adaptive structure learning method of deep belief network using neuron generation–annihilation and layer generation," *Neural Comput. Appl.*, vol. 31, no. 11, pp. 8035–8049, Nov. 2019.
[36] D. Carlson, V. Cevher, and L. Carin, "Stochastic spectral descent for restricted Boltzmann machines," in *Proc. 18th Int. Conf. Artif. Intell. Statist.*, 2015, pp. 111–119.
[37] P. Shi, X. Guo, D. Han, and R. Fu, "A sparse auto-encoder method based on compressed sensing and wavelet packet energy entropy for rolling bearing intelligent fault diagnosis," *J. Mech. Sci. Technol.*, vol. 34, no. 4, pp. 1445–1458, Apr. 2020.
[38] G. Hu, F. Zhu, and Z. Ren, "Power quality disturbance identification using wavelet packet energy entropy and weighted support vector machines," *Expert Syst. Appl.*, vol. 35, nos. 1–2, pp. 143–149, Jul. 2008.
[39] R. Yan and R. X. Gao, "Base wavelet selection for bearing vibration signal analysis," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 7, no. 4, pp. 411–426, Jul. 2009.
[40] Z. Liu, Z. Han, Y. Zhang, and Q. Zhang, "Multiwavelet packet entropy and its application in transmission line fault recognition and classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 11, pp. 2043–2052, Nov. 2014.
[41] Q. He, "Vibration signal classification by wavelet packet energy flow manifold learning," *J. Sound Vib.*, vol. 332, no. 7, pp. 1881–1894, Apr. 2013.
[42] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network," *Mech. Syst. Signal Process.*, vol. 27, pp. 696–711, Feb. 2012.
[43] H. Göksu, "BCI oriented EEG analysis using log energy entropy of wavelet packets," *Biomed. Signal Process. Control*, vol. 44, pp. 101–109, Jul. 2018.
[44] J. Li, P. Shang, and X. Zhang, "Financial time series analysis based on fractional and multiscale permutation entropy," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 78, Nov. 2019, Art. no. 104880.
[45] C. Li and P. Shang, "Multiscale tsallis permutation entropy analysis for complex physiological time series," *Phys. A, Stat. Mech. Appl.*, vol. 523, pp. 10–20, Jun. 2019.
[46] Y. Ye, Y. Zhang, Q. Wang, Z. Wang, Z. Teng, and H. Zhang, "Fault diagnosis of high-speed train suspension systems using multiscale permutation entropy and linear local tangent space alignment," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106565.
[47] M. El Sayed Hussein Jomaa, P. Van Bogaert, N. Jrad, N. E. Kadish, N. Japaridze, M. Siniatchkin, M. A. Colominas, and A. Humeau-Heurtier, "Multivariate improved weighted multiscale permutation entropy and its application on EEG data," *Biomed. Signal Process. Control*, vol. 52, pp. 420–428, Jul. 2019.
[48] J. Zheng, H. Pan, S. Yang, and J. Cheng, "Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 99, pp. 229–243, Jan. 2018.

DONGYING HAN received the Ph.D. degree from the Mechanical Engineering Institute, Yanshan University, Qinhuangdao, China, in 2008. She is currently an Associate Professor with the Institute of Vehicles and Energy, Yanshan University. Her current research interests include fault diagnosis and signal processing.

XIAOCI GUO received the B.S. degree from the Measurement and Control Technology and Communication Engineering Institute, Yanshan University, Hebei, China, in 2017, where she is currently pursuing the M.S. degree in fault signal detection and intelligent fault diagnosis.

PEIMING SHI received the Ph.D. degree from the Information Science and Engineering Institute, Yanshan University, Qinhuangdao, China, in 2009. He is currently a Professor with the Institute of Electrical Engineering, Yanshan University. His current research interests include intelligent fault diagnosis and signal processing.