Research Article

Intelligent Fault Severity Detection of Rotating Machines Based on VMD-WVD and Parameter-Optimized DBN

Ning Jia, Yao Cheng, Youyuan Tian, and Feiyu Yang

College of Mechanical Engineering, Chongqing University of Technology, Chongqing 400054, China

Correspondence should be addressed to Yao Cheng; chy_cqit@cqut.edu.cn

Received 16 November 2021; Revised 12 February 2022; Accepted 17 February 2022; Published 21 March 2022

Academic Editor: Selçuk Erkaya

Copyright © 2022 Ning Jia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An intelligent fault severity detection method based on variational mode decomposition- (VMD-) Wigner-Ville distribution (WVD) and sparrow search algorithm- (SSA-) optimized deep belief network (DBN) is suggested to address the problem that typical fault diagnostic algorithms are inappropriate for extremely comparable vibration signals when the samples are insufficient. VMD is used to process the original vibration signal to obtain the band intrinsic mode functions (BIMFs) with different frequencies. WVD produces the two-dimensional spectrum of the key BIMF with the highest variance contribution rate. The input sample of DBN is composed of a characteristic matrix formed by the two-dimensional spectrum of multiple fault signals. DBN's learning rate and batch size are both tuned globally by SSA, which has a significant influence on network error. The fitness function in the parameter optimization process is the network’s root mean square error (RMSE). Finally, the input samples are loaded into a DBN that has the best structure for detecting severity. Experiments show that, based on VMD-WVD and SSA-DBN, accuracy of the fault severity detection model for rotating machines, which has good generalization ability and robustness, can reach 98%. Compared with BPNN, the traditional DBN, VMD-DBN, VMD-PSO-DBN, and other methods, the proposed algorithm has strong adaptive feature extraction ability and generalization of application.

1. Introduction

As an important part of transmission instruments, rotating machines are widely used in various transmission systems in electric, wind power, military, industrial machinery production, etc. Because of their long working time, high working frequency, and bad working environment, rotating machines are often accompanied by faults such as wear, noise, and unbalanced load.

Fault severity detection of rotating machines is divided into three parts: data acquisition, feature extraction, and pattern recognition. Among them, feature extraction is the most important part, affecting pattern recognition accuracy’s upper and lower limits. Various one-dimensional or two-dimensional signal feature extraction technologies are applied to vibration signal feature extraction, such as Fourier transform (FFT), empirical mode decomposition (EMD), integrated empirical mode decomposition (IEMD), continuous wavelet transforms (CWT), Hilbert envelope spectrum (HES), and variational mode decomposition (VMD). Huang et al. [1] proposed the empirical mode decomposition (EMD) signal time-frequency processing method in 1998. Although this method can process non-linear and nonstationary fault signals, the inherent problems of mode mixing and endpoint effect have not been solved, which affects the accuracy of subsequent classification and identification. Dragomiretskiy et al. [2] proposed variational modal decomposition (VMD) in 2014. This method can overcome the inherent problems of EMD and IEMD and is very helpful to improve the accuracy of fault classification. Qin et al. [3] decomposed the vibration signal into multiple modal functions using VMD to obtain the time-series eigenvalues of each component. Iterative random forest (IRF) is used for fault classification, and the fault classification results are compared with other classifiers. The results show that the fault diagnosis accuracy of VMD-IRF is higher than that of VMD-C5.0, VMD-SVM, and VMD-RF. An et al. [4] used parameter-optimized VMD to obtain IMF and
constructed multiobjective evaluation function to select the best IMF. The results show that the improved VMD is helpful to improve the accuracy of signal decomposition. Zhang et al. [5] decomposed the signal into multiple components using VMD, estimated the fractal dimension using the least square method, and finally carried out fault diagnosis according to the fractal dimension of the two scales. The results show that the two-scale fractal dimension extracted by VMD can express the fractal characteristics of vibration signals and further carry out fault diagnosis.

With the development of deep learning, deep learning models are often used for optimization, diagnosis, and prediction. Common network models include deep belief network (DBN), convolutional neural network (CNN), recurrent neural network (RNN), sparse autoencoder (SAE), long-term and short-term memory (LSTM), support vector machine (SVM), etc. After signal feature processing, the accuracy of classification and recognition using the deep learning network model is higher than that of other machine learning algorithms. Chen et al. [6] integrated CNN with LSTM to realize the end-to-end fault classification model. Su et al. [7] used the DBN feature index to replace the feature index extracted by entropy and used the GWO-SVM method for fault recognition. However, compared with the methods of feature extraction using the DBN hidden layer and pattern recognition using the output layer, the accuracy of the method proposed by Su needs to be improved. Zhang et al. [8] used the LE algorithm to extract key features, characterize the sample characteristics, and carry out classification diagnosis through DBN, but the DBN network parameters were selected by experience. They did not consider the coordination between classification accuracy and training efficiency.

Various optimization algorithms are used to strengthen further the coordination between pattern recognition accuracy and training efficiency to solve the global optimal solutions of important parameters such as the number of neural network layers, learning rate, and batch size. Gai et al. [9] regarded the RMSE in the DBN network as the objective function and used the grasshopper optimization algorithm (GOA) to find the optimal combination of the parameter. The analysis shows that the diagnosis rate of the hyperparametric optimized model is much higher than that of the empirical model. Li et al. [10] used the shuffled frog leaping algorithms (SLFA) to optimize the CNN structure and directly input the original vibration signal into the CNN with the optimal structure. The training process is simple and efficient. Zheng et al. [11] used the improved whale optimization algorithm (IWOA) to optimize the search area of weights and obtain the optimal global solution of LSTM and then used IWOA-LSTM for fault diagnosis, which is more accurate than LSTM and WOA-LSTM.

This paper proposes a fault severity detection method based on variational mode decomposition (VMD) and the parameter-optimized DBN using a sparrow search algorithm (SSA). This method overcomes the bottleneck problem of enhancing DBN’s recognition accuracy by combining the powerful global optimization and local search capabilities of SSA with the greedy learning advantage of DBN.

Furthermore, the algorithm proposed in this paper can fully extract the effective fault feature information under the background of high noise, which applies to a variety of working conditions, and provides a certain algorithm idea and scientific verification for the fault diagnosis of rotating parts in multiple fields and equipment, based on the verification of a variety of multi-working condition data sets.

2. Methodology

In this section, some related algorithms, including VMD algorithm, WVD algorithm, DBN algorithm, and sparrow search algorithm, for hyperparametric optimization, were introduced. Based on these algorithms, a new rotating machinery fault severity detection method is proposed.

2.1. VMD Algorithm. The VMD algorithm is a vibration signal processing method that decomposes complex signals into a series of AM-FM signals around a fixed center frequency and limited bandwidth [12]. The AM-FM signal expression is as follows:

$$u_k(t) = A_k(t) \cos(\phi_k(t)).$$  \hspace{1cm} (1)

In this formula, $A_k(t)$ is the instantaneous amplitude of the $u_k(t)$, and $A_k(t) \geq 0$, $\phi_k(t)$ is the phase of $u_k(t)$, and $\cos(\phi_k(t))$ is the instantaneous angular frequency parameter.

$$\omega_k(t) = \frac{d\phi_k(t)}{dt} \geq 0,$$ \hspace{1cm} (2)

where $\omega_k(t)$ is the instantaneous angular frequency of $u_k(t)$. Each phase corresponds to a unique instantaneous amplitude and instantaneous angular frequency. The change of $\phi_k(t)$ can determine the change of $A_k(t)$ and $\omega_k(t)$. Therefore, within the range of interval $[t - \delta(t), t + \delta(t)]$ ($\delta(t) = 2\pi/\phi_k(t)$), $u_k(t)$ can be perceived as the harmonic signal of amplitude $A_k(t)$ and $\omega_k(t)$.

VMD transforms the decomposition process into the construction and solution process of variational problems. The construction process of variational problems is as follows:

(i) Obtaining the unilateral spectrum of the original vibration signal by Hilbert transform.

$$\left(\delta(t) + \frac{j}{\pi t}\right) \ast u_k(t)$$  \hspace{1cm} (3)

(ii) Estimating the center frequency of each analytical signal and then modulating the spectrum of each modal to the corresponding estimated fundamental frequency band.

$$\left(\delta(t) + \frac{j}{\pi t}\right) \ast u_k(t) e^{-j\omega t}.$$  \hspace{1cm} (4)

(iii) Estimating the bandwidth of each BIMF.

The expression of variational construction process is as follows:
\[
\begin{align*}
&\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[ \left( \delta (t) + \frac{j}{\pi t} \right) \ast u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \\
&\text{s.t.} \quad \sum_k u_k = f,
\end{align*}
\] (5)

where \( \{u_k\} = \{u_1, u_2, \ldots, u_k\} \) is a set of K BIMFs, \( \{\omega_k\} = \{\omega_1, \omega_2, \ldots, \omega_1\} \) is the set of centroid frequencies of each BIMF, \( \ast \) is a convolution symbol, \( \partial_t \) is a gradient operator symbol, and \( j \) is an imaginary unit.

The solution process of variational problem is to transform constrained problems into unconstrained problems [13, 14]. The method is to introduce quadratic multiplication parameters and Lagrange operator. The augmented Lagrange function is as follows:

\[
L(\{u_k\}, \{\omega_k\}, \lambda) = a \sum_k \left\| \partial_t \left[ \left( \delta (t) + \frac{j}{\pi t} \right) \ast u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle.
\] (6)

Suppose there is a d-dimensional search space with \( n \) sparrows. In the whole sparrow group, they are divided into producers and predators according to the tasks they undertake in the group. The producer is responsible for foraging and guiding the whole group to more food places. Predators constantly monitor the producers and take the opportunity to plunder food. In the group, the identity of producer and predator changes dynamically. Every sparrow has the opportunity to become a producer as long as it can find more abundant food. However, the ratio of producers to predators remains unchanged. If high-yield predators become producers, backward producers will become predators. In addition, biological groups will have natural enemies. Sparrows with vigilance account for 10%–20% of the total number of sparrows, who can find the enemy and give an alarm. When the warning value exceeds the safety threshold, the whole group will give up food in time and go to a safer place [18,19]. The location information of the sparrow group can be represented by array \( X \):

\[
X = \begin{bmatrix}
x_{1,1} & x_{1,2} & \cdots & \cdots & x_{1,d} \\
x_{2,1} & x_{2,2} & \cdots & \cdots & x_{2,d} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
x_{n,1} & x_{n,2} & \cdots & \cdots & x_{n,d}
\end{bmatrix}.
\] (8)

Each sparrow has an initial fitness, and the fitness of the sparrow population is stored in the vector \( F_x \):

\[
F_x = \begin{bmatrix}
f\left( \begin{bmatrix}
x_{1,1} & x_{1,2} & \cdots & \cdots & x_{1,d}
\end{bmatrix} \right) \\
f\left( \begin{bmatrix}
x_{2,1} & x_{2,2} & \cdots & \cdots & x_{2,d}
\end{bmatrix} \right) \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
f\left( \begin{bmatrix}
x_{n,1} & x_{n,2} & \cdots & \cdots & x_{n,d}
\end{bmatrix} \right)
\end{bmatrix}.
\] (9)

In the population, producers with higher fitness and stronger foraging ability account for 10%–20% of the total population. Therefore, producers get a larger range of foraging. In each iteration, the location of the producer update principle is based on formula (10). \( i \) and \( j \) in formula (10)
represent the position coordinates of the sparrow. The formula can be better understood according to the pseudo-code in Table 1.

\[
X_{i,j}^{t+1} = \begin{cases} 
X_{i,j}^t \cdot \exp(-i/\alpha \cdot \text{iter}_\text{max}) & \text{if } (R_2 < ST) \\
X_{i,j}^t + Q \cdot L & \text{if } (R_2 \geq ST), 
\end{cases}
\]

where \(t = 1, 2, \ldots, \text{iter}_\text{max}\) is the number of iterations for optimization, \(\text{iter}_\text{max}\) is a fixed value representing the maximum number of iterations, \(X_{i,j}^t\) represents the value of the \(i\)-th \(j\)-dimensional sparrow in the \(t\)-th iteration, \(\alpha\) and \(Q\) are random numbers \((\alpha \in (0, 1), Q \sim N(\mu, \sigma^2))\), and \(L\) represents a matrix of size \(1 \times d\). The elements of the \(L\) matrix are all 1. \(R_2\) indicates indicating warning value, and \(ST\) indicates safety threshold. When \(R_2 \geq ST\), the sparrow group needs to move to a safe area immediately to avoid intruders [18].

The predator updates the position according to formula (11), and the sparrow with alert ability updates the position according to formula (12).

\[
X_{i,j}^{t+1} = \begin{cases} 
Q \cdot \exp \left( \frac{X_{\text{worst}}^t - X_{i,j}^t}{\beta} \right), & \text{if } (i > \frac{n}{2}), \\
X_{P}^{t+1} + \left| X_{i,j}^t - X_{P}^{t+1} \right| \cdot A \cdot L, & \text{otherwise}, 
\end{cases}
\]

where \(X_P\) represents the optimal position in the producer and \(X_{\text{worst}}\) denotes the worst position in the current range. When \(i > \frac{n}{2}\), it showed that the \(i\)-th predator had poor fitness and was likely to be hungry. \(A\) indicates a matrix of \(1 \times d\) for which each element inside is randomly assigned 1 or -1.

For simplification, when \(f_i = f_g\), it showed that the sparrow found the enemy and approached other sparrows to avoid being attacked by the enemy. \(\beta\) is a step control parameter. \(K\) is a random number in the interval [-1, 1].

The basic architecture of SSA can be summarized as pseudo-code shown in Table 1.

### 2.4. Deep Belief Network

A deep belief network (DBN) is a probability generation model. The greater the probability, the lower the classification error rate. Structurally, DBN is composed of several layers of neurons. The constituent elements are \(n\) layers of restricted Boltzmann machines (RBM) and one back propagation neural network (BPNN) layer. The structure diagram of DBN is shown in Figure 1. The data samples are input from the visible layer and trained by multiple hidden layers. The hidden layer realizes data feature extraction. Finally, the network output layer obtains the diagnosis results [19, 20].

In the process of DBN training, forward learning and reverse fine-tuning is performed on each layer of the network in turn. The multilayer RBM training follows the unsupervised greedy learning algorithm. The hidden layer of each trained RBM is the visible layer of the next RBM to be trained. The layer-by-layer depth feature extraction of the original signal is realized to obtain the high-level non-concrete feature expression. BPNN dominates the reverse fine-tuning process. The error after supervision training is reversed to fine-tune the parameters of RBM layer-by-layer.
to obtain the optimal global solution of the whole DBN network [21].

2.4.1. Forward Learning. An RBM contains 2 layers of neurons: the visible layer and the hidden layer. The visible layer is employed to look at a feature of the information, and also the hidden layer is employed for data extraction. A vector often draws every layer of neurons, and also, the range of neurons corresponds to the dimension of the vector. As shown in Figure 1, the hidden layer contains \( N \) neurons, and the visible layer contains \( M \) neurons. Suppose the visible layer state vector is \( v = (v_1, v_2, \ldots, v_M)^T \). The state vector of the hidden layer is \( h = (h_1, h_2, \ldots, h_N)^T \). The visible layer bias vector is \( a = (a_1, a_2, \ldots, a_M)^T \). The offset vector of the hidden layer is \( b = (b_1, b_2, \ldots, b_N)^T \). The weight matrix between the visible layer and hidden layer is \( W = (\omega_{ij}) \in \mathbb{R}^{M \times N} \). The energy function of RBM is shown in the following formula:

\[
E_{\theta}(v, h) = -\sum_{i=1}^{M} a_i v_i - \sum_{j=1}^{N} b_j h_j - \sum_{i=1}^{M} \sum_{j=1}^{N} v_i h_j \omega_{ij},
\]

\[\theta = (W, a, b)\].

Using the energy function, the joint probability distribution of state \((V, H)\) can be obtained as

\[
P_{\theta}(v, h) = e^{-E_{\theta}(v, h)} Z_{\theta},
\]

(14)

where \( Z_{\theta} = \sum_{v,h} e^{-E_{\theta}(v, h)} \) is the normalization factor.

The structural diagram of RBM is a typical bipartite diagram. There is no link between neurons in each layer. The neurons in each layer are not connected internally, and the neurons in adjacent layers are connected with each other. Once the worth of the visible cell state is decided, the dynamic state of each hidden layer cell is freelance of every alternative, and therefore the dynamic chance of the \( j \)-th hidden cell is

\[
P(h_j = 1 | v, \theta) = \omega \left( b_j + \sum_i v_i W_{ij} \right),
\]

\[
\omega(x) = \text{sigmoid}(x) = \frac{1}{1 + \exp(-x)}.
\]

(15)

When the value of the hidden cell state is determined, the dynamic state of each visible layer cell is independent of each other, and the dynamic probability of the \( i \)-th hidden cell is

\[
P(v_i = 1 | h, \theta) = \omega \left( a_i + \sum_j h_j W_{ij} \right),
\]

\[
\omega(x) = \text{sigmoid}(x) = \frac{1}{1 + \exp(-x)}
\]

(16)

The random gradient rise method is usually used to maximize the log-likelihood function to obtain the parameters. In addition, the contrast divergence method is used to update the parameter set \( \theta \).

\[
\Delta W_{ij} = \varepsilon \left( \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{recon}} \right),
\]

\[
\Delta a_i = \varepsilon \left( \langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{recon}} \right),
\]

\[
\Delta b_j = \varepsilon \left( \langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{recon}} \right),
\]

(17)

where \( \varepsilon \) is the ratio of the learning rate to the batch size.
2.4.2. Fine-Tuning. The weight value and the offset value obtained by RBM forward learning are used as the initial parameters of the reverse fine-tuning process. The prediction data trained by BP neural network is used as the guide parameter, and the root mean square error (RMSE) between the prediction data and the input data of the visible layer is used as the evaluation index. The RBM parameters are updated layer-by-layer by using the gradient descent method. The smaller the RMSE, the higher the network convergence efficiency and the stronger the feature extraction ability. In other words, the accuracy of network classification and recognition has reached a higher state [22]. The evaluation index (RMSE) is defined as follows:

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=1}^{m} (x_{ij} - y_{ij})^2}{m \times n}}
\]  

(18)

In (18), \(x_{ij}\) represents the input data of the visible layer and \(y_{ij}\) represents forecast data and \(m\) and \(n\) represent the number of features and samples in the visible layer, respectively.

2.5. The Parameter-Optimized DBN Method Based on SSA. As shown in Figure 2, the optimization steps of SSA for DBN parameters are as follows:

**Step 1.** Set all parameters of SSA, including the number of sparrows and groups who perceive the danger, as well as the alarm value, and initialize the utmost iterations and, therefore, the search scope.

**Step 2.** The RMSE of DBN training data is taken as the fitness function of SSA with the ability of super parameter optimization of neural network. The target variables of SSA are the learning rate \(a\) and the batch size \(\beta\) of DBN. Estimated individual fitness value \(fit(i)\) of sparrow based on vector \(X[a, \beta]\).

**Step 3.** By comparing the individual estimated fitness of sparrows, the initial optimal position and the worst position can be obtained. The location of the sparrow with the highest fitness is the current global optimal position \(X_{\text{best}}\). And the location of the sparrow with the lowest fitness is the global worst position \(X_{\text{worst}}\).

**Step 4.** Determine whether the current number of optimization iterations has met the termination requirements; if so, the iteration is ended, and the global optimal result \(X_{\text{best}}[a^*, \beta^*]\) is output, the optimal parameter combination of DBN used for training data [23]; if not, proceed to the next step.

**Step 5.** Renovate the location of individual sparrows and their fitness. Reinitialize sparrows outside the search range.

**Step 6.** The current optimal solution \(X_{\text{best}}[a^*, \beta^*]\) is used as the parameter value of the DBN training part in the next iteration. Renew iterations \(t = t + 1\). The following iteration only needs to repeat Steps 2 ~ 4.

3. The Construction of Rotating Machinery Fault Severity Detection Model Based on Parameter-Optimized DBN

The traditional methods of rotating machinery fault signal feature extraction and fault diagnosis are mostly based on pre-knowledge reserve and experience. Combined with the advantages of artificial intelligence technology and the characteristics of big data from multisensor monitoring, a rotating machinery fault severity detection model based on VMD-WVD and SSA optimized DBN is proposed. Firstly, the vibration signal is discretized into a series of band intrinsic mode functions (BIMFs) with different frequencies by variational mode decomposition (VMD). Variational mode decomposition has good sparsity, which can separate the fault information hidden in strong noise to improve the accuracy of fault diagnosis. The variance contribution rate of each BIMF is compared. The BIMF with the largest square difference contribution rate is converted from a one-dimensional time-domain signal to a two-dimensional characteristic spectrum through Wigner-Ville distribution (WVD) to obtain the two-dimensional characteristic matrix of fault state as the input sample of the deep confidence network. DBN parameter combination is optimized by using the global optimization ability of SSA. The learning rate and batch size of DBN, which greatly impact the network error, are taken as the optimization objectives of the sparrow search algorithm. The root means square error (RMSE) of

![Figure 2: The flow chart of parameter-optimized DBN based on SSA.](image-url)
the network is taken as the fitness function to obtain the best network structure of DBN [22]. DBN extracts nonlinear key features through unsupervised feature extraction and supervised parameter fine-tuning. Based on this, SSA is introduced to optimize the learning rate and batch size processing quantity of DBN to strengthen further the coordination between the accuracy of pattern recognition and training efficiency. A fault severity detection model for rotating machines based on parameter optimization DBN is established.

The flow chart of this method is shown in Figure 3. The concrete steps are as follows:

1. The fault signals under different conditions and the normal vibration signals of rotating machinery are used as the original signals. VMD and WVD are used to preprocess the data to obtain a two-dimensional characteristic matrix $\lambda$. Singular value decomposition (SVD) is used to decrease the dimension of rows and columns of the two-dimensional characteristic matrix $\lambda$ to meet the dimension requirements of the deep belief network input samples.

\[
\lambda = \begin{bmatrix}
\lambda_{1,1} & \lambda_{1,2} & \cdots & \lambda_{1,n} \\
\lambda_{2,1} & \lambda_{2,2} & \cdots & \lambda_{2,m} \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_{n,1} & \lambda_{n,2} & \cdots & \lambda_{n,m}
\end{bmatrix}.
\]

(19)

2. Training samples and testing samples are constructed, respectively. After marking, they are combined as input samples.

3. Search the parameter combination of DBN through SSA. Refer to Section 2.5 for the detailed processing process of this algorithm.

4. The fault severity detection model of rotating machines is established by using DBN with optimal structure. Training accuracy and testing accuracy can be obtained [24, 25].

4. Case Study

4.1. Description of Experimental Data. The multi-condition bearing data from the bearing center of Western Reserve University’s electrical laboratory is utilized as the experimental data for comparison testing in order to assess the efficacy and robustness of the created fault detection model. The test bench of the bearing data acquisition system is shown in Figure 4(a). A driving motor, load motor, acceleration sensor, torque sensor, and several bearings are all part of the bearing test bench. The schematic diagram of the test rig and accelerometer layout is illustrated in Figure 4(b). The acceleration sensor records data under four distinct loads (load 0 = 0 HP/1797 RPM, load 1 = 1 HP/1772 RPM, load 2 = 2 HP/1750 RPM, and load 3 = 3 HP/1730 RPM), with a sample frequency of 12 kHz. Inner ring faults, outer ring faults, and rolling element faults are among the fault types [26, 27]. The flaw in the rolling element is located inside the bearing and is not apparent. Each fault type contains three fault sizes, namely, 0.007, 0.014, and 0.021 inches.

This paper adopts fault signals with a fault diameter of 0.014 inches, a motor load of 3 HP, and a motor speed of 1730 rpm. The normal signal contains 485643 data points, and the three fault signals each contain 122136 data points. The window function containing 1000 points is employed to divide the samples of the collected original vibration signal. The training data of each bearing state contains 100 samples. The testing data for each bearing condition contains 50 samples. Four types of bearing states constitute input data, as shown in Table 2. Figure 5 is a time-domain diagram of vibration signals in 4 bearing states in load 3.

As shown in Figure 5, under normal conditions, the vibration signal amplitude of the rolling bearing is very small. In the fault state, the amplitude of the vibration signal increases suddenly, which is approximately periodic. The time-domain waveform of the rolling element fault signal is similar to that of the outer ring fault signal, making it difficult to distinguish specific fault conditions. Therefore, it is necessary to extract the characteristics of vibration signals [28].

4.2. Vibration Signal Feature Extraction. VMD and WVD are used for feature extraction of vibration signals. The bearing state signal in load 3 is the experimental data. Figure 6 is the time-domain diagram of BIMF obtained by variational modal decomposition of the outer ring fault signal in load 3, and Figure 7 is the frequency domain diagram obtained by Fourier transform of each BIMF.

To select the key BIMF, the variance contribution rate of each BIMF to the vibration signal under each fault state is calculated. The average value of the variance contribution rate of each BIMF component is obtained, respectively, as shown in Table 3. The fourth BIMF has the largest average variance contribution. The average variance contribution rate of each BIMF represents the contribution of BIMF to the characteristics of the initial vibration signal [29]. The larger the average variance contribution rate is, the closer the feature of the BIMF is to the feature of the original signal without noise. Calculation shows that the fourth BIMF has a great impact on the vibration signal, so the WVD is carried out for the fourth BIMF. In the outer ring fault state, the three-dimensional spectrum diagram after WVD transformation for the fourth BIMF is shown in Figure 8.

The feature matrix after feature extraction is operated by kernel principal component analysis (KPCA) to obtain the visual scatter diagram of vibration signal feature extraction, as shown in Figure 9(a). The coordinate system in Figure 9(a) has a smaller order of magnitude, which indicates that the feature information has good aggregation. After feature extraction, various types of feature information are distinguished. Figure 9(b) shows that normal state features and various fault state features are mixed before feature extraction, which is difficult to distinguish [30].

4.3. DBN Parameter Optimization Using SSA. This paper uses DBN with a network structure of 100–100. Parameter optimization is also based on this structure. SSA is used to...
search the optimal global solution of two parameters: learning rate and batch size of DBN. The search range of learning rate is [0.01 1], and the search range of batch size is [1 90]. The parameter settings of SSA are shown in Table 4. Figure 10 shows the change curve of RMSE of the testing data with the number of iterations. After 5 iterations, RMSE
starts to converge, showing that the algorithm has a quick convergence speed. After the 10th iteration, the RMSE of the testing data converges to 0.095. After the 15th iteration, RMSE converges stably at 0.085, which shows that the optimization algorithm using SSA to optimize DBN has excellent global optimization ability and strong robustness.

Finally, the optimal combination of learning rate and batch size is [0.1673,4].

The structural parameters of DBN are shown in Table 5. Input samples and fault categories determine the number of nodes in the input and output layers. The setting of the DBN parameter combination obtained after SSA global optimization is shown in Table 6.
Table 3: Variance contribution rate of BIMF.

<table>
<thead>
<tr>
<th>Fault classification</th>
<th>BIMF1</th>
<th>BIMF2</th>
<th>BIMF3</th>
<th>BIMF4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>47.08</td>
<td>10.15</td>
<td>23.58</td>
<td>19.18</td>
</tr>
<tr>
<td>Inner race fault</td>
<td>2.94</td>
<td>16.72</td>
<td>24.04</td>
<td>56.30</td>
</tr>
<tr>
<td>Ball fault</td>
<td>16.92</td>
<td>19.00</td>
<td>17.57</td>
<td>46.52</td>
</tr>
<tr>
<td>Outer race fault</td>
<td>18.54</td>
<td>6.01</td>
<td>45.95</td>
<td>29.49</td>
</tr>
<tr>
<td>Average</td>
<td>21.37</td>
<td>12.97</td>
<td>27.78</td>
<td>37.87</td>
</tr>
</tbody>
</table>

Figure 8: Three-dimensional spectrum of BIMF.

Figure 9: Visual scatter diagram of feature extraction. (a) The proposed method. (b) Traditional method.

Table 4: Parameter setting of SSA.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maximum iterations</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Sparrow population size</td>
<td>10</td>
</tr>
<tr>
<td>SSA</td>
<td>Number of producers</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Number of sparrows who perceive the danger</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Alarm value</td>
<td>0.8</td>
</tr>
</tbody>
</table>
4.4. Analysis and Comparison of Fault Severity Detection Model Results

4.4.1. Diagnosis Accuracy Test and Multimodel Comparison. After time-frequency analysis, the fault of bearing data is diagnosed using a deep belief network with an optimal structure. To begin, the network is trained and diagnosed using the training set. The defect diagnostic model based on VMD-WVD and SSA-DBN began to converge stably in this region around the 200th iteration, as shown in Figure 11, suggesting that the model is well trained. The error converges to 0.0201 after 500 iterations, and the average correctness of the training data is 98 percent. The trained model is then utilized to diagnose the test data issue, with a diagnostic accuracy of 95.5 percent. Ten repeated tests are carried out to decrease the impact of random error on the accuracy of the defect diagnostic model, and the accuracy histogram is displayed in Figure 12. The recognition rates for ten diagnoses are 98 percent, 97.7 percent, 98.2 percent, 97.8 percent, 97.67 percent, 98 percent, 97.7 percent, 97.98 percent, 98 percent, and 98 percent, respectively, with an average accuracy of 98 percent, indicating that the model is robust and accurate for rolling bearing fault classification.

To deeply verify the effectiveness of the diagnostic model established in this paper, this model is contrasted with BPNN, the traditional DBN (non-opt DBN), VMD-DBN, and VMD-PSO-DBN. On each model, conduct 10 equal sample and equal amount experiments, with the results displayed in Figure 13. The average diagnostic accuracy and associated standard deviation of 10 experiments are determined to objectively compare the diagnostic accuracy of the five approaches, as shown in Table 7.

It can be seen from the diagnosis results that although the network structure of shallow neural network BPNN is simple and the requirements for operating equipment are not high, the fault diagnosis accuracy is the lowest, and the

---

**Table 5: The structural parameters of DBN.**

<table>
<thead>
<tr>
<th>Input layer</th>
<th>RBM (1)</th>
<th>RBM (2)</th>
<th>Output layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>100</td>
<td>100</td>
<td>4</td>
</tr>
</tbody>
</table>

**Table 6: The setting of DBN parameter combination.**

<table>
<thead>
<tr>
<th>Batch size</th>
<th>Learning rate</th>
<th>Momentum value</th>
<th>Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.1673</td>
<td>0.01</td>
<td>100</td>
</tr>
</tbody>
</table>

---

![Figure 10: The variation of RMSE with the number of iterations.](image)

![Figure 11: Error variation curve of training data.](image)

![Figure 12: Accuracy histogram for test data.](image)

![Figure 13: Diagnostic accuracy comparison.](image)
results of many experiments are very different. The standard deviation of accuracy is as high as 2.195, and the diagnosis accuracy is unstable. The accuracy of fault detection using the standard DBN (non-opt DBN) is also increased since DBN has more layers and a more acceptable structure than BPNN. However, the results of non-opt DBN experiments are still highly variable, and the diagnostic effect ranks in the middle of the other methods. The use of the time-frequency analysis method to extract the features of the original data aids in fault diagnostic accuracy. As a result, when compared to earlier approaches, the diagnostic accuracy and stability of VMD-DBN have increased, although the diagnosis accuracy
still has room for development. The advanced search method is utilized to optimize DBN to break past the bottleneck of the VMD-DBN process. The particle swarm optimization algorithm has a tremendous memory capacity and can do global optimization. PSO is utilized to improve the super parameters of DBN and construct the VMD-PSO-DBN fault diagnostic model. Many experiments suggest that diagnostic accuracy has improved from the beginning. However, the
4.4.2. Generalization Capability Verification. To test the
generalization ability of the proposed model, it is used to
Southeast University’s defect diagnostic of induction motor ML bearing data set. To retrieve the original
signal, choose a working state with a rotation frequency of
20 Hz and a load configuration of 2 in ML bearing data,
and split 1019200 sampling points for each kind of data.
VMD and WVD are used to create the two-dimensional
data characteristic matrix. Singular value decomposition
is used to lower the matrix’s dimension, and the data is	hen normalized to create the network training and test sets. Normal, ball fault, inner fault, outer fault, and
combination fault are the five states in the ML data set.
Table 8 describes the sample sizes and status labels for the
five bearing states. BPNN, Non-opt DBN, PSO-DBN, and
SSA-DBN are used for fault diagnosis of new network
samples, respectively. The categorization ability is
measured using the confusion matrix. Figure 14 depicts
the findings of the fault severity detection confusion matrix.

As shown in Figure 14, BPNN has the lowest classi-
fication ability of all approaches for bearing state, with a
classification ability of less than 50% for bearing condition
4 (inner ring fault). PSO-DBN has a lot greater fault
classification ability than non-opt DBN and BPNN. Its
performance disparity with SSA-DBN is mostly seen in
the categorization and identification of state 3 (combined
fault). As can be shown, SSA-DBN virtually perfectly
classifies motor bearings in five states into the right
category, with the combined defect with the lowest
identification rate over 90%. The model has an identifi-
cation rate of more than 99 percent for a single fault
condition. This demonstrates that the VMD-WVD-SSA-
DBN fault detection model developed in this research
satisfies the technical criteria for multi-working condition
fault detection, has strong generalization, and is an ideal
to bear failure diagnosis in industrial equipment.

5. Conclusion

To handle the problem of bearing failure detection under
high noise and varied operating circumstances, a rolling
bearing fault diagnostic model based on VMD-WVD and
SSA-DBN is presented in this research. A range of multi-
working condition bearing data sets are used in the ex-
perimental analysis procedure. The algorithms given in
this work have a high diagnostic accuracy, which proves
the method’s viability and generalization, as well as
providing a particular notion and scientific proof for the
defect diagnosis algorithm of rotating components in the
industrial production process. The key conclusions are as
follows:

1. A fault diagnostic model is developed using VMD,
WVD, and optimized DBN, as well as a novel in-
telligent fault diagnosis approach. The model ex-
hibits a greater fault identification accuracy and a
lower precision standard deviation when compared to
classic intelligent diagnosis approaches. It can
extract feature information buried in noise and is
excellent for mining complicated and variable
characteristics hidden in huge data. It also aids in
improving pattern recognition accuracy.

2. To prevent the impact of empirical DBN on diag-
nostic accuracy, SSA is utilized to improve the
learning rate and batch learning periods of DBN
worldwide. The DBN optimized by SSA has a greater
fault diagnostic accuracy and stability than VMD-
DBN and VMD-PSO-DBN.

3. The follow-up investigation is focused on real-time
defect diagnosis using the new vibration signal detected
by the sensor to implement the sensor-diagnostic
mechanism interconnection. The next stage in the
research is to apply this methodology to a
broader range of fault diagnosis, classification, and
prediction problems, such as motor or gear failure
analysis, acoustic signal-based detection and analysis,
and so on.

Data Availability

The data used in the experimental part are from the bearing
data sets of Case Western Reserve University and Southeast
University. Readers can search on the Internet. The other
data used in this manuscript are available from the corre-
sponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was jointly supported by Chongqing Research
Fund (cstc2016jcyjA0497), Chongqing University of Tech-
nology Research Fund (cglycx 20202088), and Chongqing
Graduate Innovation Research Fund (CYS21464).
References


