WILEY | Hindawi

Research Article

Fault Diagnosis Algorithm of Gearbox Based on NAPSO-VMD Self-Adaptive Noise Reduction and Dual-Sensor Feature Fusion

Yangshuo Liu [1], Jianshe Kang [1], Chiming Guo [1], Yunjie Bai,[2] and Shuo Meng[1]

[1]*Army Engineering University of PLA, Shijiazhuang 050003, China*
[2]*66029 Unit of the Chinese People's Liberation Army, Xilinguolemeng 011200, China*

Correspondence should be addressed to Chiming Guo; guochiming@nudt.edu.cn

Condition monitoring plays a very important role in equipment fault diagnosis technology. However, existing monitoring methods often collect equipment fault signals from a single dimension, resulting in a major lack of fault information. To improve the problem, we built a gearbox preset fault test bench and constructed a dual-sensor acquisition system to realize the multiple dimensions of vibration signal acquisition in the horizontal and vertical directions of the gearbox. At the same time, given the poor adaptability of most current signal preprocessing methods, the improved nonlinear adaptive inertial weight particle swarm optimization algorithm (NAPSO) and variational modal decomposition (VMD) are combined to optimize the key parameters in VMD with the maximum correlation kurtosis convolution (MCKD) as the fitness function. Further, after extracting fault features from the intrinsic mode functions (IMFS) decomposed by VMD, the single-layer sparse autoencoder network (SAE) and the double-layer stacked sparse autoencoder network (SSAE) with different structures are used to realize an effective fusion of multidimensional information and deep feature extraction. Finally, the hybrid fault diagnosis of gearboxes is realized by using the random forest algorithm (RF) as the classifier. The experimental results show that the accuracy of the method proposed in this paper can reach 96.0%, and the accuracy can be improved by 3.0% and 4.0%, respectively, when compared with a single horizontal or vertical sensor signal input.

1. Introduction

The health management of equipment plays an important role and significance in maintaining the reliability of the equipment. Among them, fault diagnosis technology, as a key technology for monitoring and judging equipment status, has long received extensive attention from industrial managers and researchers [1–3]. At present, the research on fault diagnosis mode based on the combination of equipment fault signal data-driven and deep learning has reached a relatively mature stage. Wang et al. [4] used the synchronous extraction transformation to convert the gearbox vibration signal into a time-frequency map of uniform size and finally learned the optimal classification strategy automatically under the framework of deep reinforcement learning. Y. Wang and S. Wang [5] used variable-parameter time-frequency analysis to expand the number of bearing abnormal samples to address the imbalance between normal and

abnormal samples and then rebalanced the dataset to train a convolutional neural network (CNN) to classify bearing health. Deng et al. [6] first performed a continuous wavelet transform (CWT) on the vibration signal of the rolling bearing to obtain a time-frequency map and then input the time-frequency map as a feature map to the BP neural network and CNN to compare the feature extraction and classification effects of different networks. Bai et al. [7] used a multiscale crop fusion data enhancement method to enhance the rolling bearing fault signal before converting it into a short-time Fourier time-frequency image; following these steps, a multichannel deep network (DNN) was then used to fuse multisensor-derived image data for feature extraction and failure mode classification.

Although the above fault diagnosis algorithms based on deep learning have achieved good classification results, they only consider the single-dimensional signal of the equipment. When a device fails, it will generate fault excitations

in multiple directions. Only collecting signals from a single dimension of the device will undoubtedly result in the lack of important device failure information. In this regard, this paper builds a gearbox preset fault test bench and builds a dual-sensor acquisition system to realize the vibration signal acquisition in the horizontal and vertical directions of the gearbox, to ensure that there are sufficient multidimensional data for monitoring and judging the equipment operation status. Further, although deep learning methods can autonomously learn deep features from the raw signals of the device, they cannot effectively integrate the information of multisource signals. In addition, the signal preprocessing analysis technology can describe the fault state of the equipment from the perspective of the fault mechanism, which has its irreplaceable advantages. In this regard, Shao et al. [8] proposed a stacked wavelet autoencoder structure with a Morlet wavelet function for multisensory data fusion and a flexible weighted assignment of fusion strategies, which realized the effective extraction and fusion of multichannel sensor fault features. Aiming at the problem that the fault detection results of the vibration signal of a single sensor may be unreliable and unstable, Liu et al. [9] proposed an intelligent multisensor data fusion method based on the ant colony optimization algorithm and correlation vector machine (RVM) for gearbox faults detection. The method removes a large amount of redundant information in the features and improves the accuracy of fault diagnosis; Cheng et al. [10] combined wavelet correlation scale entropy (WCFSE) and Dempster-Shafer (DS) evidence theory algorithm and proposed a new method for gear fault diagnosis under strong noise conditions based on multisensor information fusion. The introduction of multisensor fusion technology eliminates the uncertainty and defects of single sensor identification and improves the accuracy of fault diagnosis. On the other hand, shallow classifiers often show better accuracy and stability than deep networks in discriminating fault mechanism features. Han et al. [11] compared the discriminative accuracy of random forest (RF), extreme learning machine (ELM), probabilistic neural network (PNN), and support vector machine (SVM) on different time-frequency domain features. Experimental results show that RF outperforms comparative classifiers in terms of recognition accuracy, stability, and robustness to features, especially when the training set is small; Yang et al. [12] proposed a fault diagnosis scheme based on a hierarchically structured ELM. Representation learning based on multilayer ELM covers functions such as data preprocessing, feature extraction, and dimensionality reduction; each ELM layer is processed independently for its corresponding role, thereby realizing data representation feature learning and fault classification. Compared with deep learning (DL) based on backpropagation, iterative fine-tuning of parameters is omitted and the efficiency of diagnosis is improved; Cerrada et al. [13] established a robust multilevel fault diagnosis system for gearboxes. The optimal set of conditional parameters in time, frequency, and time-frequency domains is extracted from vibration signals, and data classification is performed in a supervised environment using genetic algorithms and RF-based classifiers. High-efficiency fault diagnosis under variable working conditions of the gearbox is real-

ized. It can be seen from the above research that the method of combining signal preprocessing and the shallow classifier can accurately and stably realize the fault diagnosis of equipment.

However, there are still some problems that can be improved. On the one hand, various types of fault signals of equipment will produce nonlinear fluctuations with changes in the external environment such as working conditions and operating time. However, in most cases, different types of signals have not been adaptively analyzed and preprocessed, and the differences between the data have not been further highlighted. In this regard, Wang et al. [14] first used ensemble empirical mode decomposition (EEMD) to decompose the signal and then used the kurtosis spectral entropy as the objective function to search for the filter length of multipoint optimal minimum entropy deconvolution adjusted (MOMEDA) in a grid format, which is used for searching complex fault pulse signal in strong noise environment; a comparative study of adaptive algorithms derived from empirical mode decomposition (EMD), empirical wavelet transform (EWT), variational mode decomposition (VMD), and Vold-Kalman filter order tracking was conducted by Liu et al. [15], and the improvement and application of these methods in dynamic analysis and fault diagnosis of the mechanical transmission system are elaborated; Ni et al. [16] proposed a fault information-guided VMD (FIVMD) method for extracting the weak bearing repetitive transient. Two nested statistical models based on the fault cyclic information, incorporated with the statistical threshold at a specific significance level, are used to approximately determine the mode number of VMD; then, the ratio of fault characteristic amplitude (RFCA) is defined and utilized to identify the optimal bandwidth control parameter of VMD, which enables the decomposed BLIMF to be sensitive to the bearing fault signature and less affected by the abnormal impulses and vibrations from other components. Cheng et al. [17] proposed an adaptive weighted symplectic geometric distribution (AWSGD) method, which can adaptively adjust the algorithm parameters according to periodic kurtosis (CK) and periodic impact intensity (PII) to effectively denoise gear signals. The above self-adaptive algorithm can perform self-adaptive preprocessing on the equipment fault signal according to different sensitive indicators, which can effectively improve the accuracy and effect of the diagnosis algorithm. However, based on the standard of adaptive preprocessing, the adjustment of the parameters of the above algorithm is still carried out using empirical values or grid traversal search, which is less efficient. In this regard, a method combining the population optimization algorithm with the adaptive optimization objective is proposed to efficiently process the effective fault information in the signal. In addition, the selection of signal adaptive optimization criteria is also very important. Feng et al. [18] used the vibration cyclostationarity index (CS) to track the wear evolution of the equipment and studied the relationship between the tribological characteristics of the two wear phenomena, fatigue pitting and abrasive wear, and gearmesh-modulated second-order cyclostationary (CS2) properties of the vibration signal. As a result, the mechanism recognition of the gearbox wear is implemented. Fan et al.

[19] proposed a sparse representation-based transient feature extraction technique for gearbox fault diagnosis. The method can effectively identify the transient pulse time and period based on wavelet, to extract transient features. The effectiveness of the proposed method is verified by simulation signals and actual gearbox vibration signals. Since the fault pulses in the vibration signal of rotating machinery often appear in a periodic form, how to extract and separate the fault pulse characteristics is a problem that needs to be focused on. Traditional time-domain features such as kurtosis can more sensitively detect pulse components in signals, but they cannot distinguish noise components from periodic components; although the minimum entropy deconvolution method (MED) can make the signal after filtering through the inverse filter, it prefers to deconvolve a single pulse or group of pulses rather than the expected periodic pulses that recur during the fault period [20]. To this end, Zhang et al. [21] proposed a new method called maximum correlation kurtosis deconvolution (MCKD). This method optimizes the correlation kurtosis of the signal by finding the optimal filter, which can more effectively extract the periodic pulse component of the signal and restrain the noise influence of the signal compared with the MED method. Therefore, this paper combines the good noise reduction performance of variational modal decomposition for mechanical equipment vibration signals and the optimization performance of a typical swarm optimization algorithm—particle swarm optimization (PSO)—and takes the MCKD value of the signal as the fitness function, so as to adaptively optimize the key parameters in VMD. The decomposed modal component (IMF) matrix retains the periodic fault pulse components in the equipment vibration signal. Further in terms of feature extraction, the time-frequency domain signal analysis method can accurately describe and cover the fault information of the signal from the perspective of the fault mechanism [22, 23]; therefore, this paper extracts the MCKD value, 14 common time-domain features, and wavelet packet decompose energy scale features from the IMF matrix decomposed by VMD to construct the feature matrix of the signal [24]. At the same time, to improve the global optimization ability of the PSO algorithm, the inertia weight in the PSO algorithm is nonlinearly improved (NAPSO).

On the other hand, there is redundant information between multisensor signals, and the accuracy and effect of equipment fault diagnosis can be improved through effective information fusion. In this regard, Xie et al. [25] used principal component analysis (PCA) to fuse and convert multisensor signal features into RGB images, and then, the image samples are input into a convolutional neural network (CNN) with residuals for further extraction of deep features; Li et al. [26] proposed an adaptive channel weighted neural network to study the importance of different sensor signals in the feature fusion method while maximizing the mining of the deep fault feature information of each sensor and finally realized the condition monitoring of the gearbox transmission system and the helicopter transmission system; Cao and Yunusa-Kaltungo [27] proposed a gearbox fault classification framework for the automatic fusion of multisensor data, generating features through coherent composite spectros-

copy (CCS) and using PCA for data dimensionality reduction. The final diagnosis results were obtained from artificial neural network training feature samples. Although the above information fusion method has played a role in reducing the dimension of the data and extracting features, it still cannot achieve the effect of deep fusion and high-dimensional fault feature extraction for the effective information of the sensor; in addition, combined with the actual situation of this paper, the signal decomposed by adaptive VMD will have different feature dimensions due to the different decomposition layers. Therefore, aligning the feature dimensions of the fault data of different channels is also a key step in the diagnosis process; and while aligning the features, maintaining the original components and information of the signal is also an aspect that needs to be considered. Sparse autoencoder (SAE) is one of the important structures in deep networks. It adds sparsity constraints to traditional autoencoders and has powerful data reconstruction and feature reexpression capabilities, so it is widely used in various fields of production and life (pattern recognition, target detection, and natural language processing) [28–30]. Therefore, this paper uses a single-layer SAE to align data of different dimensions; at the same time, for the problem that the signals collected by multiple sensors may contain redundant information, this paper builds a two-layer stacked sparse autoencoder. In this way, the features of the dual sensors are fused to achieve the purpose of data compression and feature extraction. Finally, the random forest algorithm (RF), as an emerging and highly flexible machine learning algorithm, integrates the decision value of each decision tree into the final classification result through the idea of ensemble learning [31, 32]. Therefore, the fused feature samples are input into the RF classifier to realize the hybrid fault diagnosis under the variable working conditions of the gearbox. Experiments show that the fault diagnosis algorithm proposed in this paper can obtain a good classification effect for the mixed faults of the gearbox under variable working conditions, and the classification accuracy is improved compared with the single sensor signal input.

In summary, a hybrid fault diagnosis method for gearboxes based on NAPSO-VMD adaptive noise reduction of vibration signals and multitype sparse autoencoder dual-channel sensor feature alignment and fusion is proposed in this article. Experiments and diagnostic algorithms are verified on the gearbox preset fault test bench, and the experimental results demonstrate the effectiveness of the proposed dual-channel sensor data fusion diagnostic algorithm. The work in the rest of this paper is described as follows. Section 2 introduces the theory and the process of the fault diagnosis algorithm. Section 3 provides the experiment and data preparation process for the mixed failure of the gearbox under variable operating conditions. Section 4 analyzes and discusses the experimental results. Section 5 discusses the effectiveness of the diagnosis algorithm. Finally, the conclusions of the paper will be summarized in Section 6.

2. Methodology

2.1. Adaptive Noise Reduction of Signals. When using VMD decomposition and noise reduction, the noise reduction effect of the signal is more sensitive to the decomposition

layers K and the penalty factor alpha [33]; in addition, a unified decomposition parameter cannot be used for the vibration data of different types of gearbox faults. In this paper, the improved adaptive weight particle swarm optimization algorithm is used to optimize the parameters of the signal-to-VMD with the maximum correlation kurtosis deconvolution value of the signal as the optimization objective, so that each type of fault data can obtain appropriate decomposition parameters and adaptive noise reduction.

2.1.1. Maximum Correlation Kurtosis Deconvolution (MCKD). The maximum correlation kurtosis deconvolution is a new method that is extremely sensitive to the expected periodic pulses that recur during the fault cycle, and its purpose is to deconvolve the periodic fault pulses from the vibration signal of the equipment. The following is the principle and algorithm flow of MCKD [34], assuming that the input signal and the system are in the form of convolution, which is

$$x(n) = h(n) * y(n) + e(n). \tag{1}$$

In the case of ignoring noise, the purpose is to find an optimal filter $f = [f_1 f_2 \cdots f_L]^T$ to solve $y(n)$, that is,

$$y(n) = f(n) * x(n) = \sum_{k=1}^{L} f_k x_{n-k+1}. \tag{2}$$

Among them, $x(n)$ and $y(n)$ are the input signal and output signal, respectively, and f, L, and N are the filter coefficient, filter length, and sampling times, respectively. Then, the concept of correlation kurtosis (CK) is introduced, which is defined as

$$\mathrm{CK}_M(T_s) = \frac{\sum_{n=1}^{N} \left(\prod_{m=0}^{M} y_{n-mT_s} \right)^2}{\left(\sum_{n=1}^{N} y_n^2 \right)^{M+1}}. \tag{3}$$

Among them, M is the shift number, and T_s is the sampling point corresponding to the iterative update period, which can be expressed as

$$T_s = f_s \cdot T, \tag{4}$$

where f_s is the sampling frequency and T is the epoch of fault feature detection. The entire MCKD algorithm adopts the objective function method; that is, a finite impulse response filter is selected iteratively to maximize the correlation kurtosis value of the filtered signal, and its expression is

$$\max_f \mathrm{CK}_M(T_s) = \max_f \frac{\sum_{n=1}^{N} \left(\prod_{m=0}^{M} y_{n-mT_s} \right)^2}{\left(\sum_{n=1}^{N} y_n^2 \right)^{M+1}}. \tag{5}$$

Among them, the general value range of M is 1 to 7, and if it is greater than 7, the precision will be reduced. Then, take the derivative of the objective function:

$$\frac{d}{df_k} \mathrm{CK}_M(T_s) = 0 \, k = 1, 2, \cdots, L. \tag{6}$$

Finally, the expression form of the matrix is obtained by derivation as

$$f = \frac{\|y^2\|}{(M+1)\|\beta\|^2} \left(X_0 X_0{}^T \right)^{-1} \sum_{m=0}^{M} X_{mT} \alpha_m, \tag{7}$$

where $X_0 X_0{}^T$ is the Toeplitz autocorrelation matrix of the original signal $x(n)$, assuming that $\left(X_0 X_0{}^T \right)^{-1}$ exists; the superscript T denotes the transpose operation; in addition,

$$X_r = \begin{bmatrix} x_{1-r} & x_{1-r} & x_{1-r} & \cdots & x_{1-r} \\ 0 & x_{1-r} & x_{1-r} & \cdots & x_{1-r} \\ 0 & 0 & x_{1-r} & \cdots & x_{1-r} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & x_{1-r} \end{bmatrix}_{LbyN}, \tag{8}$$

$$r = mT_s, m = 1, 2, \cdots, M,$$

$$\alpha_m = \begin{bmatrix} y^{-1}{}_{1-mT_s} \left(y_1{}^2 y^2{}_{1-T_s} \cdots y^2{}_{1-MT_s} \right) \\ y^{-1}{}_{2-mT_s} \left(y_2{}^2 y^2{}_{2-T_s} \cdots y^2{}_{2-MT_s} \right) \\ \cdots \\ y^{-1}{}_{N-mT_s} \left(y_N{}^2 y^2{}_{N-T_s} \cdots y^2{}_{N-MT_s} \right) \end{bmatrix}, \tag{9}$$

$$\beta = \begin{bmatrix} y_1 y_{1-T_s} \cdots y_{1-MT_s} \\ y_2 y_{2-T_s} \cdots y_{2-MT_s} \\ \cdots \\ y_N y_{N-T_s} \cdots y_{N-MT_s} \end{bmatrix}.$$

2.1.2. Adaptive Variational Mode Decomposition. Particle swarm optimization (PSO) is a swarm evolutionary computing technology derived from the study of predation behavior of bird flocks; its basic idea is to design a massless particle with only two properties—speed and position—to simulate the birds in the flock. The optimal solution is found through cooperation and information sharing among particle individuals [35]. For an optimization problem with n dimensions $x_1, x_2 \cdots, x_n$, x_i and v_i are used to represent the position and velocity of the i-th particle, respectively, and the expressions are as follows:

$$\begin{aligned} x_i &= \left(x_i{}^1, x_i{}^2 \cdots, x_i{}^n \right), \\ v_i &= \left(v_i{}^1, v_i{}^2 \cdots, v_i{}^n \right). \end{aligned} \tag{10}$$

The particles in the population find the optimal parameter value corresponding to the objective function by iteratively updating the position and velocity. At the time

TABLE 1: Function information.

Function	Expression	Parameter value range	Theoretical optimal solution (minimum value)
Goldstein and Price	$f_1(x) = \left\{\left[1 + (x_1 + x_2 + 1)^2 \left(\begin{array}{c}19 - 14x_1 + 3x_1^2 \\ -14x_2 + 6x_1x_2 + 3x_2^2\end{array}\right)\right] \times \left[30 + (2x_1 - 3x_2)^2 \times \left(\begin{array}{c}18 - 32x_1 + 12x_1^2 \\ +48x_2 - 36x_1x_2 + 27x_2^2\end{array}\right)\right]\right\}$	[-2, 2]	3
Branin	$f_2(x) = \left\{\left(x_2 - (5/4\pi^2)x_1^2 + (5/\pi)x_1 - 6\right)^2 + 10(1 - 1/8\pi)\cos(x_1) + 10\right\}$	[-5, 15]	0.397887
Schaffer F6	$f_3(x) = \sin^2\sqrt{x_1^2 + x_2^2} - 0.5/\left[1 + 0.001(x_1^2 + x_2^2)\right]^2 - 0.5$	[-100, 100]	-1

TABLE 2: Average fitness of different inertia weights.

| Inertia weight	Index	Test function		
		f_1	f_2	f_3
Linear weight	BE	6.49010^{-7}	3.773×10^{-7}	1.072×10^{-3}
	WE	27.00	3.215	3.725×10^{-2}
	AE	0.271	4.249×10^{-1}	1.127×10^{-2}
Log weight	BE	1.510×10^{-6}	5.275×10^{-7}	3.860×10^{-3}
	WE	0.301	2.308	4.595×10^{-2}
	AE	4.866×10^{-2}	2.311×10^{-2}	1.622×10^{-2}
Square weight	BE	1.151×10^{-6}	3.686×10^{-7}	3.301×10^{-5}
	WE	3.957×10^{-2}	6.171×10^{-4}	9.716×10^{-3}
	AE	2.342×10^{-3}	1.880×10^{-4}	7.199×10^{-3}
Exponential weight	BE	1.337×10^{-8}	2.295×10^{-7}	0
	WE	0.144	5.217×10^{-4}	0
	AE	1.759×10^{-3}	1.009×10^{-4}	0
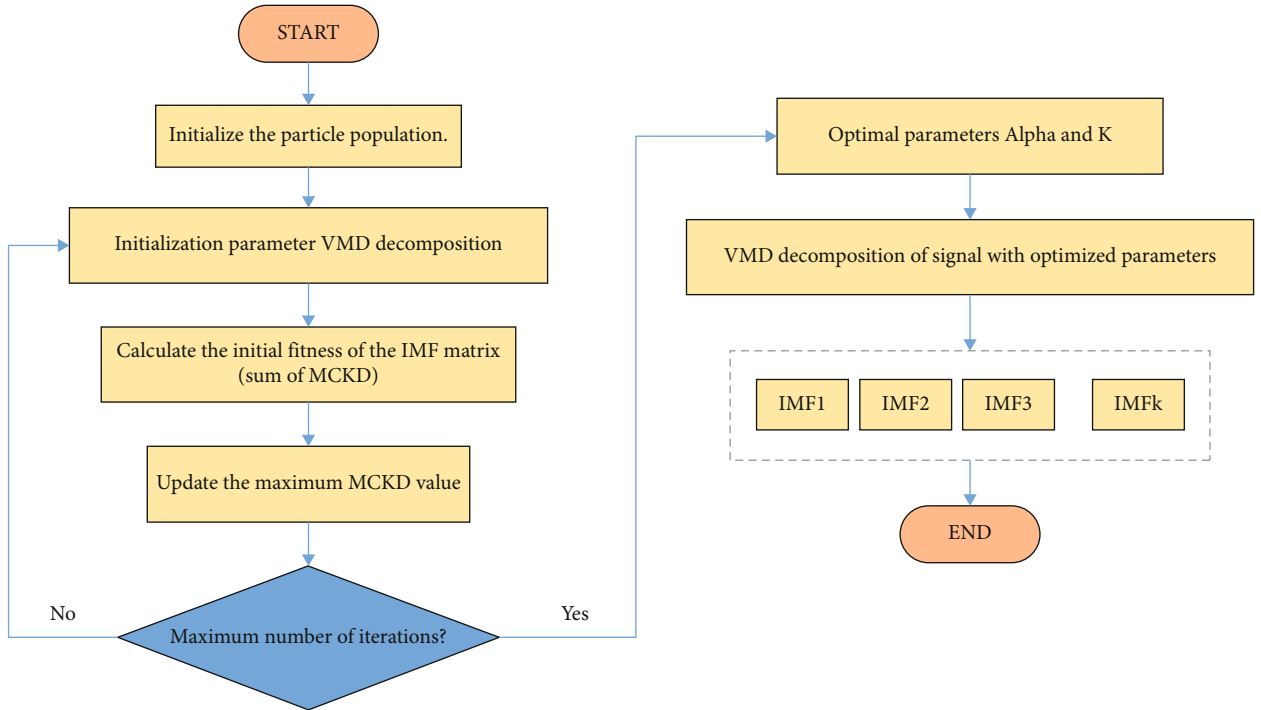
FIGURE 1: Parameter alpha and K value optimization process.

$t + \Delta t$, the particle velocity and position update formula are as follows:

$$v_i(t + \Delta t) = \omega v_i(t) + c_1 r_1(\overline{x_i} - x_i) + c_2 r_2(\overline{x_g} - x_i), \quad (11)$$

$$x_i(t + \Delta t) = x_i(t) + v_i(t + \Delta t)\Delta t, \quad \Delta t = 1. \quad (12)$$

t in the formula represents the current number of iterations, c_1, c_2 is a positive acceleration factor, usually taken as 2;

r_1, r_2 is a random number between [0,1]; $\overline{x_i}, \overline{x_g}$ represents the historical optimal position of the i-th particle and the entire population, respectively; ω is the inertia weight, which changes linearly with the increase of the number of iterations to update the speed of the particle [36]. The basic expression of ω is

$$\omega = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times a = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times \frac{t}{T}. \quad (13)$$
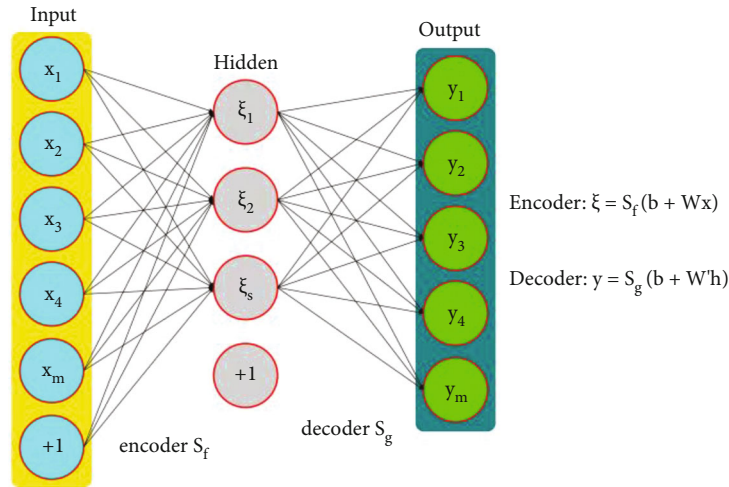
Encoder: $\xi = S_f(b + Wx)$

Decoder: $y = S_g(b + W'h)$

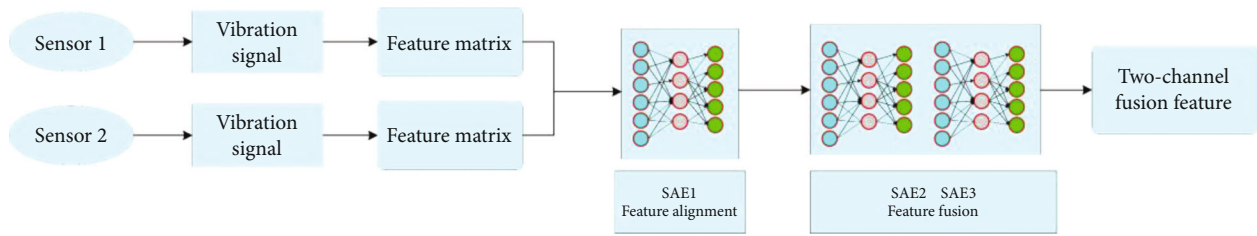FIGURE 2: Structure of SAE.

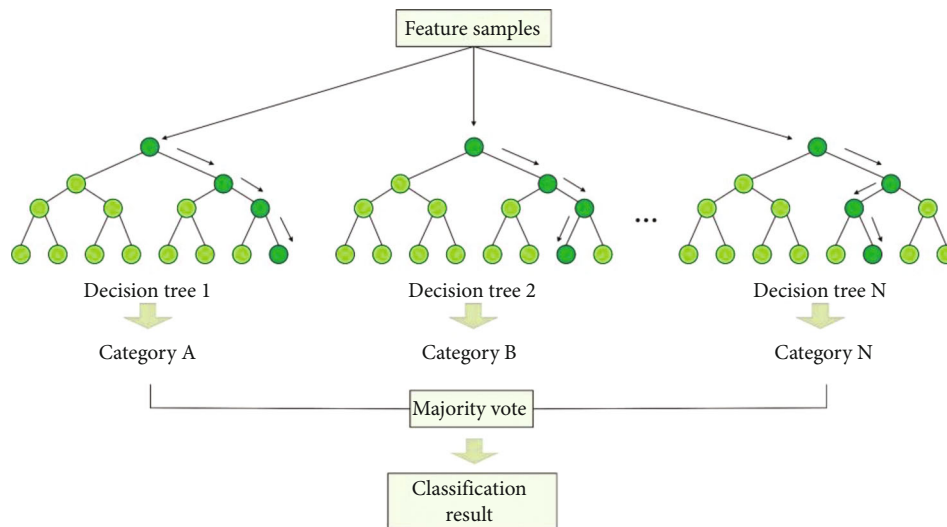
FIGURE 3: Feature fusion process.

FIGURE 4: Schematic diagram of random forest algorithm.

In the formula, a is the weight decay method, and linear decay is the most commonly used method of the standard particle swarm algorithm. $\omega_{max}, \omega_{min}$ are the maximum and minimum values of the inertia weight, usually taken as 0.9 and 0.4, and T is the maximum number of iterations. In recent years, with the deepening and diversification of research problems, the two biggest problems faced by the PSO algorithm are the speed of convergence and falling into a local minimum. Given

this situation, some scholars have improved PSO from the aspects of evolution state evaluation, elite learning strategy, and system adaptive parameters. Among them, the method based on system adaptive parameters increases the diversity of particle swarms and accelerates the convergence speed by giving the inertial weight ω adaptiveness or controlling the acceleration factor [37]; in this paper, by changing the original linear inertia weight ω to exponential decay weight ω_1, it is

ensured that inertia weight has a larger value in the exploration stage, which is beneficial to the global search of the algorithm; in the development stage, a smaller value is beneficial to the local search of the algorithm and reduces fluctuations. At the same time, to compare with the weight values of other attenuation methods, this paper also lists two nonlinear weights besides linear weights: logarithmic attenuation weight ω_2 and square attenuation weight ω_3. Equation (14) lists the expressions of the four inertia weights.

$$
\begin{cases}
\omega = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times \dfrac{t}{T}, \\[2mm]
\omega_1 = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times \dfrac{e^{t/T} - 1}{e^t - 1}, \\[2mm]
\omega_2 = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times \dfrac{\ln \sqrt{T + 1 - t}}{\ln \sqrt{T + 1}}, \\[2mm]
\omega_3 = \omega_{\max} - (\omega_{\max} - \omega_{\min}) \times \dfrac{t^2}{T^2}.
\end{cases}
\tag{14}
$$

To obtain more accurate results, this paper compares the optimization performance of different inertia weights through simulation experiments. The performance test function of the intelligent optimization algorithm used is shown in Table 1 [38, 39]. These test functions are complex and nonlinear, including multiple peaks and valleys, and there are a large number of local extreme points, which can be used to detect the global search performance, local optimization performance, and stability of the algorithm.

Taking the function value as the fitness of the PSO algorithm, the number of particles is set to 100, the maximum number of iterations is 500, and each test function is optimized 100 times. The performance of each optimization algorithm is measured by indicators such as the minimum error $BE = \min(F_i) - F_{\mathrm{to}}$, the maximum error $\mathrm{WE} = \max(F_i) - F_{\mathrm{to}}$, and the average error $AE = \mathrm{mean}(F_i) - F_{\mathrm{to}}$. The smaller the minimum error, the better the local optimization ability of the optimization algorithm. The larger the maximum error, the worse the antipremature ability of the optimization algorithm (ability to resist premature convergence). The smaller the average error, the more stable the optimization performance of the optimization algorithm. In short, the smaller the value of these three indicators, the better the performance of the optimization algorithm, and the average fitness is shown in Table 2.

It can be seen from the overall test results that the three indicators of the nonlinear weight are smaller than those of the linear weight. Then comparing the test results of different nonlinear weights, it can be found that the square weight is not as effective as the other two types of weights in other test functions except that the maximum error WE corresponding to the test function f_1 is smaller; in addition, the indicators of exponential weight are further reduced as a whole compared with the logarithmic weight. Therefore, among the four types of inertia weights, the exponential weight has the best comprehensive optimization performance. In this paper, the exponential weight is used to replace the linear weight of the PSO algorithm to obtain a
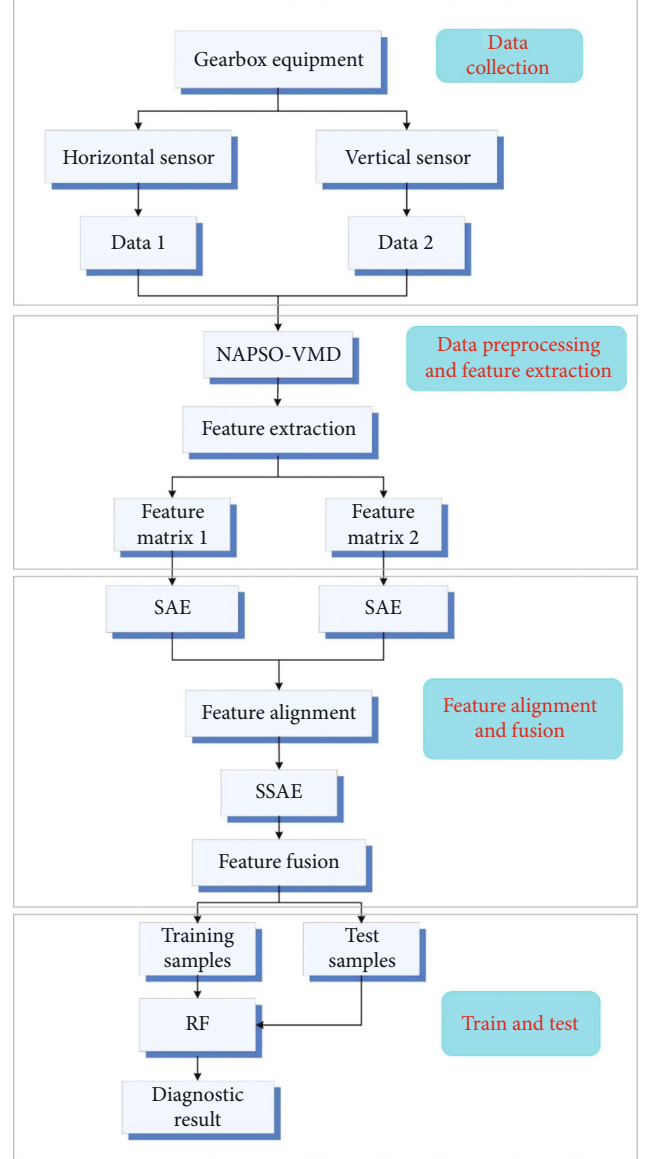

FIGURE 5: Diagnostic strategy flowchart.

nonlinear adaptive inertia weight particle swarm optimization (NAPSO). Further, the MCKD value of the signal is used as the fitness function to optimize the parameters in the VMD. The specific steps are shown in Figure 1.

2.2. Applications of Multitype Sparse Autoencoders.
In this section, a single-layer SAE and a two-layer stacked sparse autoencoder (SSAE) are used to align and fuse the data of dual-channel sensors, using the powerful data depth reconstruction and feature extraction capabilities of the sparse autoencoder (SAE).

2.2.1. Autoencoder (AE).
Autoencoder (AE), as one of the most typical unsupervised learning deep networks, its main role is to reconstruct the input data and learn its features and make the output data as equal to the input as possible. Its structure mainly includes input, hidden, and output
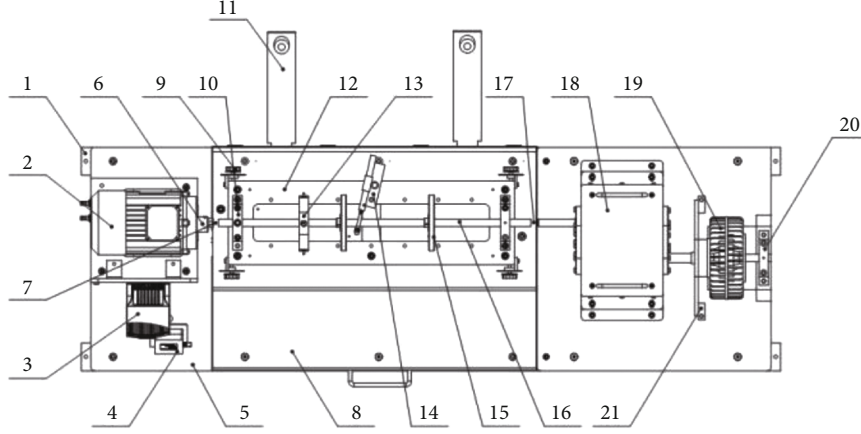
Figure 6: Structure diagram of the experimental platform.

layers (Figure 2); and its network learning includes encoding and decoding two processes. The data is encoded and reconstructed in the hidden layer to obtain a deep representation of the features [40–42].

2.2.2. Sparse Autoencoder (SAE). SAE is the same as AE in structure, including output layer, hidden layer, and output layer. The improvement is that SAE increases the sparsity limit based on AE. By constraining part of the network activation state in the hidden layer, the neuron nodes related to the input data are made active. On this basis, KL dispersion is introduced to measure the similarity between the average activation output of a hidden layer node and the sparse ρ we set. Therefore, SAE can learn more effective feature expressions and improve the efficiency of feature extraction. The training is achieved by minimizing the error as a loss function. Determine whether the error converges by calculating the output of SAE; the SAE networks are trained until the error converges [43, 44].

2.2.3. SAE-Based Feature Dimension Alignment. Based on the good data reconstruction ability of SAE, this paper uses a single-layer SAE as a feature dimension alignment tool. Firstly, the MCKD value feature, 14 commonly used time-domain features, and 8 energy scale features after 3-layer wavelet packet decomposition are extracted from the dual-channel samples after adaptive noise reduction, a total of 23 category features. Among them, time-domain features include dimensional features (mean, root mean square, root square amplitude, absolute mean, skewness, kurtosis, variance, maximum value, minimum value, and peak-to-peak value) and dimensionless features (shape factor, crest factor, margin factor, and impulse factor) [45, 46]; the "db3" wavelet basis is used in the wavelet packet decomposition [47]. In this way, after further feature extraction, a $K \times 23$-dimensional feature vector can finally be obtained.

However, since different decomposition level K values may be obtained, this paper uses the above SAE to align features of different data in each channel, so that the feature dimensions of data in the same channel are kept consistent.
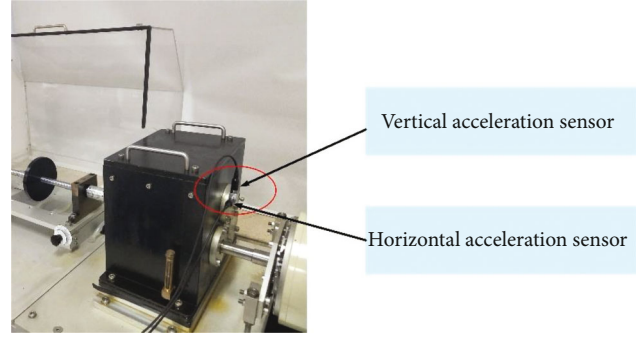

Figure 7: Two-channel sensor installation diagram.

2.2.4. SSAE-Based Deep Feature Fusion. In this section, two layers of SAE are stacked to form an SSAE network, which is used as a deep feature fusion network (DFFN). In the network, the output features of the hidden layer of the first SAE are input to the second SAE, and the deep feature extraction and fusion of multichannel data are realized through multiple unsupervised training of SSAE. For example, in Section 2.2.3, the data features of the horizontal channel sensor are unified to $K_1 \times 23$, and the data features of the vertical channel sensor are unified to $K_2 \times 23$; then, the features of the two channels are concatenated into a $(K_1 + K_2) \times 23$ vector and input to the DFFN network for feature fusion. The concatenated features are first input to the first layer of SAE in the DFFN network, and the output features of the hidden layer are used as the input of the second layer of SAE, and the final output of the second layer of SAE hidden layer is the dual-channel fusion feature. The process of feature fusion is shown in Figure 3.

2.3. Random Forest Algorithm (RF). Random forest is a machine learning algorithm based on the idea of ensemble learning. The basic unit of a random forest is a decision tree, and a random forest is formed by integrating multiple decision trees. For classification problems, first, use each decision tree in the random forest to classify the sample, each decision tree will randomly select some features of the sample, and make classification decisions based on them;
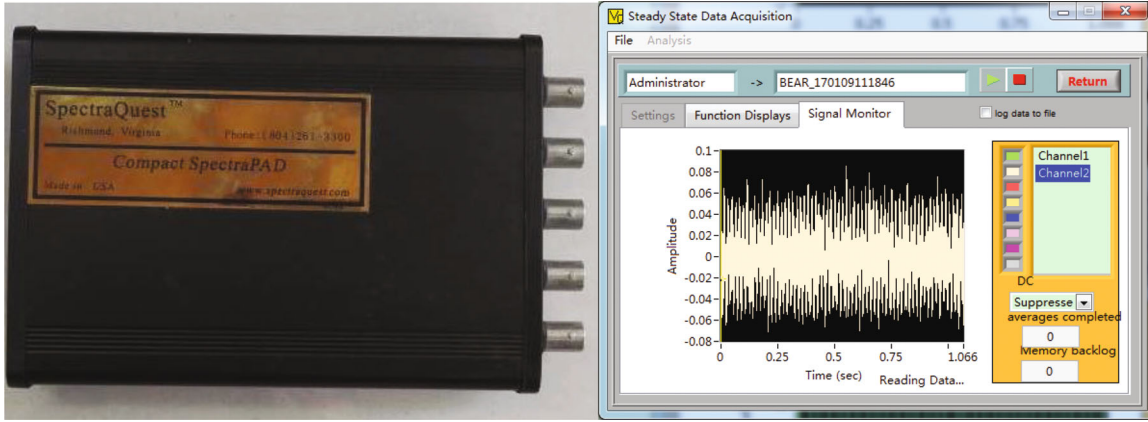
FIGURE 8: Data acquisition board and acquisition software.

these classification decisions are finally integrated to form the classification result. Figure 4 visually shows how random forests work [13, 48].

The random forest has two main elements: "random" and "forest." "Forest" is easy to understand, that is, multiple decision trees constitute a forest, which reflects the integration idea of a random forest. And "random" refers to the random selection of samples and the random selection of features to be selected. In recent years, random forests have been widely used in various fields because random forests have the following advantages [49]:

(1) It can run effectively on large datasets

(2) It can process input samples with high-dimensional features without dimensionality reduction

(3) Since each tree selects some samples and some features, overfitting can be avoided to a certain extent, and the performance is stable

In Section 2.2, the two-channel sensor fault samples have large dimensions after feature extraction and deep fusion, so random forest is selected as the classifier of the fault diagnosis method. The specific implementation process of random forest is as follows [50, 51]:

(1) Build a decision tree. Assume that the number of samples to be classified is N, and the number of features of the samples is U; when the decision tree analyzes and decides the samples, it needs to randomly extract u sample features, where u is much smaller than U, and then determines the decision result of a node on the decision tree according to the attributes of the selected features. After that, all samples are randomly sampled with a replacement for N times (that is, bootstrap sampling), thus forming a training sample set. The samples that are not drawn from the sample population are used as the test set to evaluate the error of the decision tree. So far, for all nodes in the decision tree, repeat the aforementioned analysis and decision-making process; that is, each node randomly selects u features. According to these
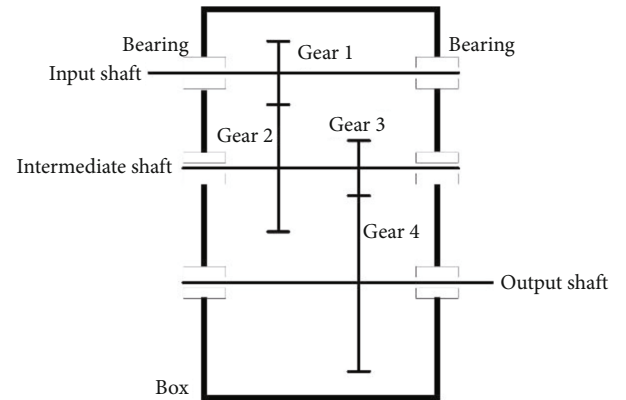

FIGURE 9: Internal structure diagram of experimental gearbox.

TABLE 3: The parameters of gears of the experimental gearbox.

Gear shaft	Gear number	Modulus	Number of teeth	Hub diameter (mm)
Input shaft	Gear 1	1.5	41	45
Intermediate shaft	Gear 2	1.5	79	60
Intermediate shaft	Gear 3	1.5	36	45
Output shaft	Gear 4	1.5	90	60

u features, calculate the best splitting method; in addition, each decision tree will grow completely without being pruned

(2) Select important features. Among the randomly selected features in a tree, the importance of each feature is different, so it is necessary to select more important features. Arbitrarily take one of u features of a tree and randomly change the feature value, compare the error rate of the test set before and after the change, and the test set is the remaining samples after sampling. Then, calculate the error rate of the features of a tree, then calculate the importance of
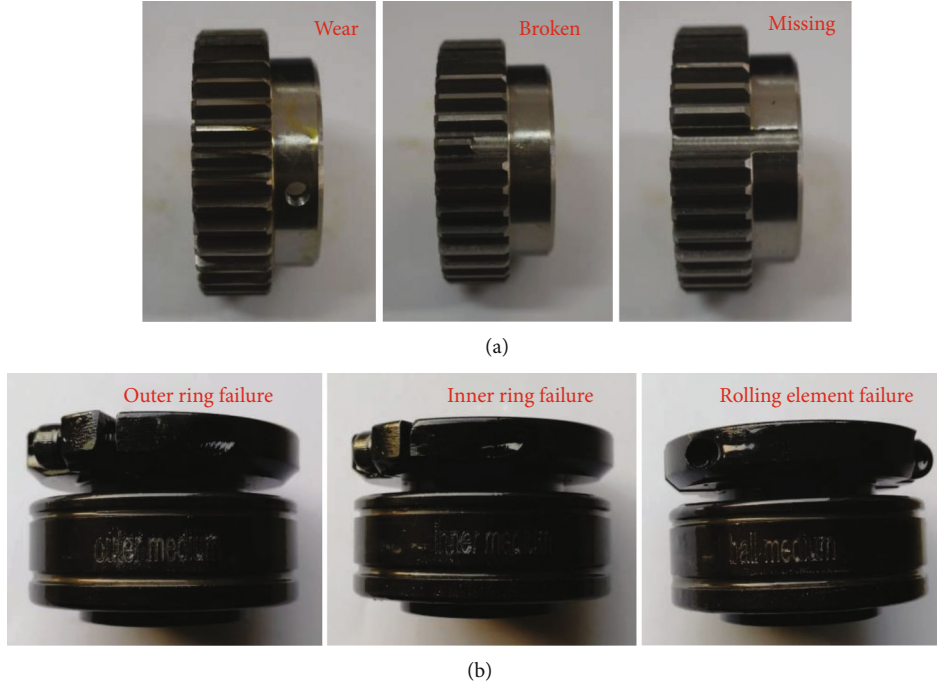
(a)

(b)

FIGURE 10: Preset failure experiments of (a) gears and (b) bearing.

the features in all trees in their respective trees, and finally calculate the mean of the importance of this feature value in multiple trees. This gives the importance of all features in the forest. Sort all the features according to their importance, remove some features with low importance in the forest, and get a new feature set

(3) Follow the above steps for multiple iterations, gradually remove features with relatively low importance, and generate a new forest each time until the number of remaining features is u. After each iteration of the forest, the out-of-bag error rate is also calculated to evaluate the performance of the forest, and the forest with the smallest out-of-bag error rate is selected as the final random forest model. The out-of-bag error rate is mainly calculated in the following ways: first, a sample is used as a decision tree of the test set to classify the sample, and the final result is obtained through the voting classification of multiple decision trees. Then, perform the first step on all samples, and finally, calculate the ratio of the number of misclassified samples to the total number of samples as the out-of-bag error rate

2.4. Fault Diagnosis Algorithm Flow. The dual-channel fusion feature samples obtained in Section 2.2 are input into the random forest classifier for training, to realize the fault diagnosis of the mixed fault of the gearbox. So far, the complete diagnosis process of this paper is shown in Figure 5.

(1) Install two accelerometers in different directions on the gearbox test bench to perform multichannel acquisition of vibration signals

(2) Perform adaptive VMD on the dual-channel fault vibration signal of the gearbox

(3) Fault features are extracted from the denoised signal; then use single-layer sparse autoencoder and double-layer stacked sparse autoencoder to align and fuse the features, respectively

(4) Use the RF classifier to train and test the feature samples to realize the fault diagnosis of the gearbox

3. Data Preparation

To research the hybrid fault diagnosis method of the gearbox, the preset hybrid fault experiment is carried out by relying on the simulation fault test bench of the general components of the mechanical transmission system. The simulation failure test bench for the general components of the mechanical transmission system is composed of gearboxes, bearings, motors, magnetic powder brakes, and sensors. The specific structure of the test bench is shown in Figure 6. The components corresponding to the numbers in the figure are as follows: 1, test bench foundation lifting ear; 2, motor; 3, motor frequency conversion controller; 4, speed sensor display table; 5, test bench foundation; 6, speed sensor; 7, coupling; 8, safety cover; 9, centering adjustment code disc (2 axial and 4 horizontal); 10, split rolling (sliding) bearing seat (2); 11, safety cover support; 12, rotor system base; 13, intermediate positioning device (for sliding bearings); 14, mechanical friction device; 15, rotors with evenly distributed screw holes (2); 16, rotor shaft (20 mm); 17, rotor shaft and gearbox connection coupling; 18, parallel gearbox (two-stage); 19, magnetic powder brake; and 20, gearbox support bearing seat (including rolling bearings).

TABLE 4: The parameters of bearings of the experimental gearbox.

Bearing pitch diameter D (mm)	Rolling body diameter d (mm)	Number of rolling elements Z/piece	Contact angle α (°)
38.5	7.9	9	0

TABLE 5: Details of 12 test working conditions.

Condition number	Rotating speed	Load current
1	1200 rpm (constant)	0.1 A
2	1500 rpm (constant)	0.1 A
3	1800 rpm (constant)	0.1 A
4	600 rpm~1200 rpm (even speed change)	0.1 A
5	1200 rpm (constant)	0.2 A
6	1500 rpm (constant)	0.2 A
7	1800 rpm (constant)	0.2 A
8	600 rpm~1200 rpm (even speed change)	0.2 A
9	1200 rpm (constant)	0.3 A
10	1500 rpm (constant)	0.3 A
11	1800 rpm (constant)	0.3 A
12	600 rpm~1200 rpm (even speed change)	0.3 A

TABLE 6: Data category and operating condition information for fault diagnosis.

Fault type	Condition number
F1	4
F2	9
F3	5
F4	5
F5	4
F6	6
F7	10
F8	8
F9	2
F10	6

TABLE 7: Optimization results of parameters.

Type	Channel	Alpha value	K value	MCKD
F1	Horizontal	9016	6	2.88e-04
	Vertical	9082	5	5.62e-04
F2	Horizontal	9892	4	3.66e-05
	Vertical	9003	4	2.56e-05
F3	Horizontal	7629	3	3.41e-07
	Vertical	7682	4	2.57e-04
F4	Horizontal	9973	4	4.04e-04
	Vertical	9928	4	5.71e-05
F5	Horizontal	7839	4	3.86e-04
	Vertical	9374	3	3.44e-06
F6	Horizontal	7914	4	6.09e-05
	Vertical	7762	4	6.07e-04
F7	Horizontal	8538	4	6.33e-05
	Vertical	9714	5	2.81e-04
F8	Horizontal	8167	4	3.94e-05
	Vertical	8831	5	3.14e-04
F9	Horizontal	9416	5	3.05e-04
	Vertical	8801	5	1.99e-05
F10	Horizontal	8983	6	3.00e-06
	Vertical	9786	4	1.56e-06

The power of the test bench is provided by the motor, the magnetic powder brake is used to provide variable load, and the gearbox used in the experiment is a two-stage parallel shaft gearbox To realize the dual-channel vibration data acquisition, two acceleration vibration sensors are installed in the vertical and horizontal directions of the bearing housing of the preset fault bearing in the gearbox, respectively (Figure 7); the data acquisition board and data acquisition software interface are shown in Figure 8.

In this paper, a preset fault experiment is carried out on the gears and bearings in the gearbox. The gearbox used in the experiment is a two-stage parallel shaft gearbox. The internal structure diagram is shown in Figure 9, and the specific parameters of the gear are shown in Table 3. The preset faulty gear is the pinion of the intermediate shaft (gear 3 in Table 3). The preset faults of gears are set as gear wear, gear broken teeth, and gear missing teeth, as shown in Figure 10(a). Among them, the gear wear fault is that the upper and lower tooth surfaces of the gear teeth are ground equidistantly inward by 0.2 mm; the gear broken tooth fault is the removal of half of one of the gear teeth; the missing tooth fault is that one of the gear teeth is completely removed. The rolling bearing model of the gearbox used in the experiment is ER-16K, and its specific parameters are shown in Table 4. The bearing with the preset fault in the experiment is the bearing of the intermediate shaft close to the gear 3. The preset faults of the bearing mainly include outer ring fault, inner ring fault, and rolling element fault. The faulty bearing is shown in Figure 10(b). The outer ring

fault is to machine a 0.5 mm deep groove at the center of the outer ring; the inner ring fault is to machine a 0.5 mm deep groove at the center of the inner ring; the rolling element fault is to machine a groove on one of the rolling elements 0.5 mm deep groove.

Assembling the failure parts described above into the gearbox constitutes different mixed failure modes. A total of 9 mixed fault states and normal state data were collected in the experiment, of which 9 mixed fault states include broken gear teeth and bearing rolling element fault (hereinafter referred to as F1 for simplification, other fault forms are
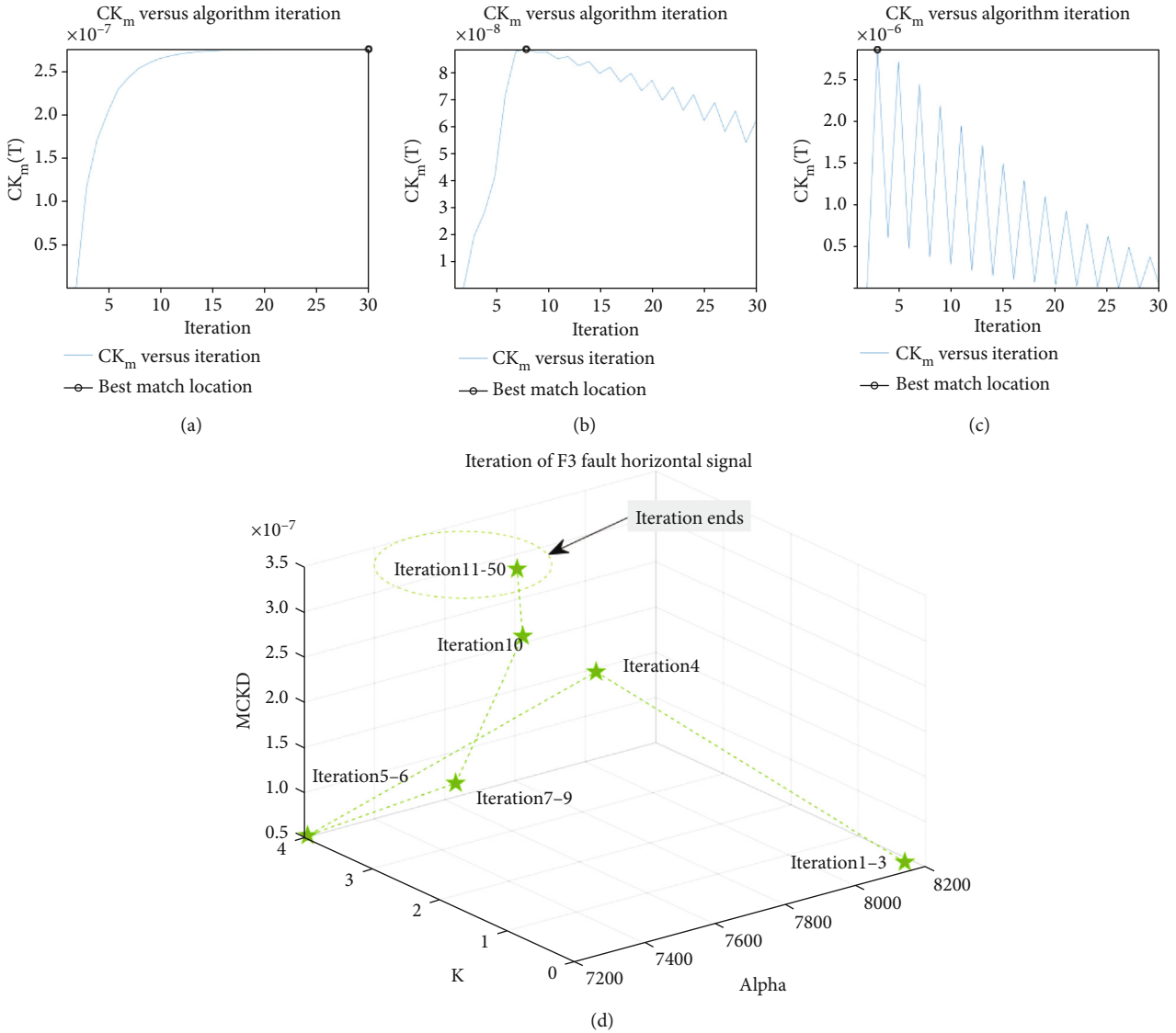
FIGURE 11: MCKD calculation of F3 fault horizontal channel signal (a–c) and iterative diagram of NAPSO algorithm (d).

simplified in this way), gear broken teeth and bearing inner ring fault (F2), gear broken teeth and bearing outer ring fault (F3), gear wear and bearing rolling element fault (F4), gear wear and bearing inner ring fault (F5), gear wear and bearing outer ring fault (F6), gear missing teeth and bearing rolling element fault (F7), gear missing teeth and bearing inner ring fault (F8), and gear missing teeth and bearing outer ring fault (F9), and the data in the normal state is expressed as F10. To more comprehensively diagnose different faults under the mixed working conditions of shaft teeth, mixed fault experiments under different working conditions were carried out. The specific working conditions are set as shown in Table 5, in which the speed of 600 rpm~1200 rpm (uniform speed change) means that the speed is increased uniformly from 600 rpm to 1200 rpm within 10 seconds, and changing the current simulates different loads.

In this paper, the data under a certain working condition is randomly selected from the 10 types of experimental data to realize the fault diagnosis under the mixed working con-

dition of shaft teeth. Table 6 lists the working condition information of the selected data. The sampling frequency of the vibration signal is 20.48 kHz, the sampling length of each group of data is 49.6 s, 5000 points are taken as a sample, and each type of data of the two channels is taken 60 samples.

After the data samples are collected, adaptive VMD noise reduction is performed on them, the number of particles is set to 100, and the number of iterations is set to 50. Table 7 lists the average value of 10 calculations. It can be seen from the MCKD calculation of the two channels of F3 type data and the iterative process of the NAPSO algorithm shown in Figures 11 and 12 that after updating the iterative filter many times, each IMF component of the signal has obtained the MCKD value. This means that after the adaptive VMD decomposition, the signal retains a sufficient number of periodic fault pulses, which is conducive to feature extraction and fault classification. In terms of NAPSO iteration, on the one hand, with each fitness update, the
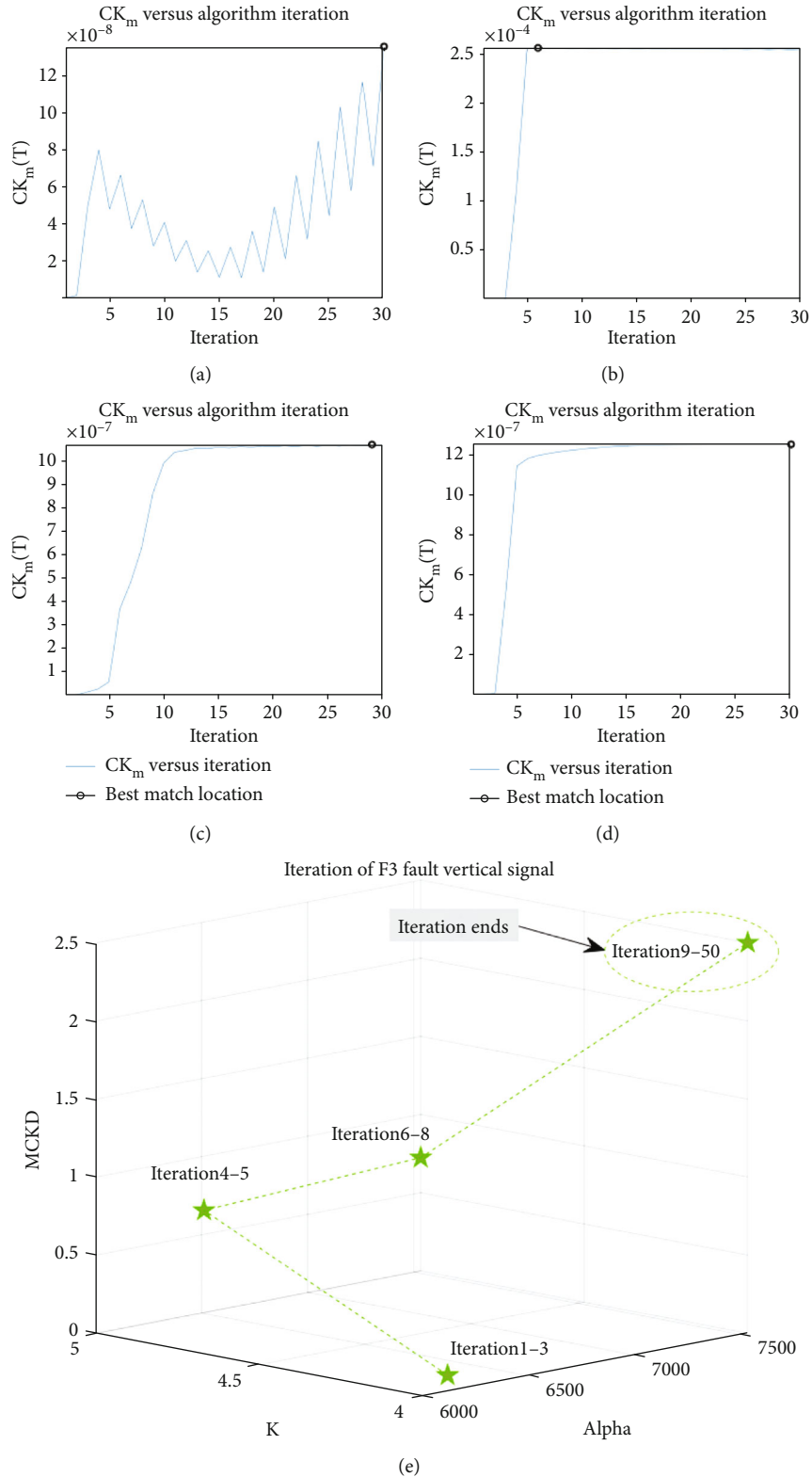
(a)

(b)

(c)

(d)

(e)

FIGURE 12: MCKD calculation of F3 fault vertical channel signal (a–d) and iterative diagram of NAPSO algorithm (e).

MCKD of the signal is iteratively computed; on the other hand, the horizontal signal and vertical signal reach the maximum MCKD value at the 11th and 9th iterations, respectively, and in the subsequent iterations, the MCKD value remains around the maximum value (iterative results overlap).

After extracting features and unifying dimensions, the feature dimension of the horizontal channel is unified to

TABLE 8: Parameter settings of sparse autoencoders with different functions.

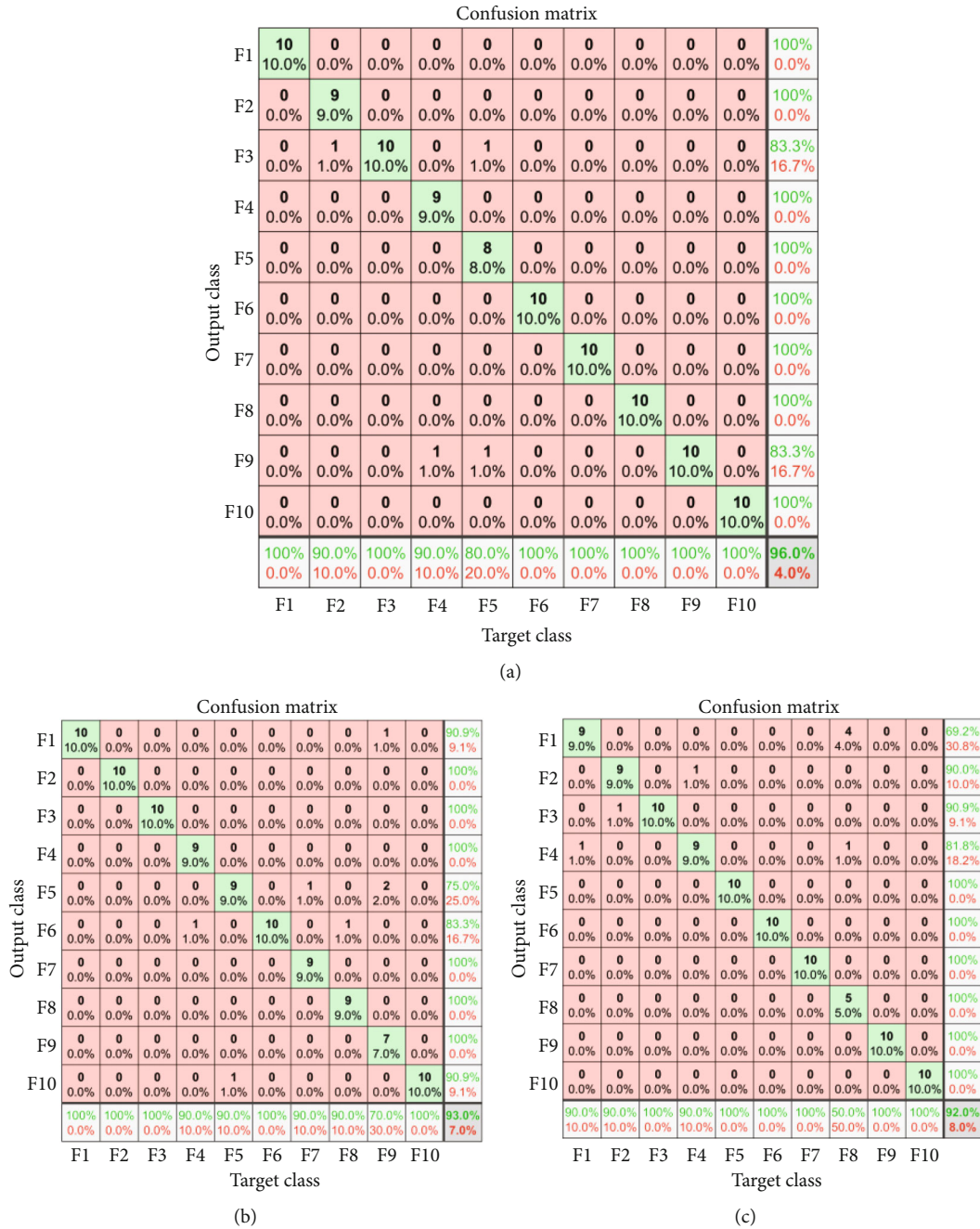| Parameter	SAE1	SAE2	SAE3
Hidden layer node number	69/92	100	30
L2 regularization weight decay coefficient	0.002	0.001	0.001
Sparse penalty weight	0.003	0.002	0.002
Sparsity parameter	0.0001	0.0001	0.0001
Decoder transfer function	Purelin	Purelin	Purelin
The maximum number of iterations	100	100	100

FIGURE 13: Confusion matrix of fault diagnosis results for different channels: (a) fusion features, (b) horizontal data features, and (c) vertical data features.

3×23 dimensions, and the feature dimension of the vertical channel is unified to 4×23 dimensions; that is, the hidden layer node number of the single-layer sparse autoencoder is set to 69 and 92, respectively, and they can be connected in series to form 7×23-dimensional feature samples; then, the series feature samples are normalized by Z-score and input into the DFFN network for feature fusion, and hidden layer node number of the two-layer sparse autoencoder in the DFFN network is 100 and 30, respectively. The final fault data can get a total of 600 30-dimensional feature samples of 10 types. Among them, the settings of the parameter of the three-layer sparse autoencoder are shown in Table 8 [52].

4. Fault Diagnosis Using Random Forest Classifiers

First, the samples are divided into two parts according to the ratio of $5 : 1$, the training set, and the test set. The number of decision trees of the RF classifier is set to 1000, and the number of features randomly selected for each decision tree is set to 3; after training the training set with the RF classifier, the test set was input into the model, and the diagnostic accuracy was 96.00% (Figure 13(a)). At the same time, the feature matrices of the horizontal and vertical channels are input into RF for training and testing, and the diagnostic accuracy of the horizontal data is 93.00% (Figure 13(b)), and the diagnostic accuracy of the vertical data is 92.00% (Figure 13(c)). The diagnostic accuracy was 3% and 4% lower than that of fusion features, respectively. The out-of-bag error rates of the three types of features in RF are shown in Figure 14. It can be seen from the figure that the out-of-bag error rate of the fusion feature is significantly lower than that of the other two-directional sensor data features. The above experimental results demonstrate the effectiveness of the dual-channel feature fusion method in this paper.

5. Model Analysis

5.1. Effectiveness Analysis of Dual-Channel Fusion Features. To observe the fusion feature matrix extracted in this paper more intuitively, the data visualization method t-distributed stochastic neighborhood embedding (t-SNE) is used to visualize the fusion feature matrix and the feature matrix of the sensor data in the vertical and horizontal directions, and the results are shown in Figure 15. It can be seen from the figure that the fused feature scattergram can completely distinguish different fault states, while the data feature scattergram collected from the sensors in the vertical and horizontal directions cannot clearly distinguish some fault states. Therefore, the t-SNE analysis results intuitively show that the fused features perform better than the unfused features in distinguishing different mixed fault states. It also further illustrates the effectiveness of the feature extraction method in this paper.

5.2. Parameter Sensitivity Analysis of Sparse Autoencoder. The data reconstruction ability of SAE is closely related to the L2 regularization weight decay coefficient, sparse penalty weight, sparsity parameter, and other parameter values.
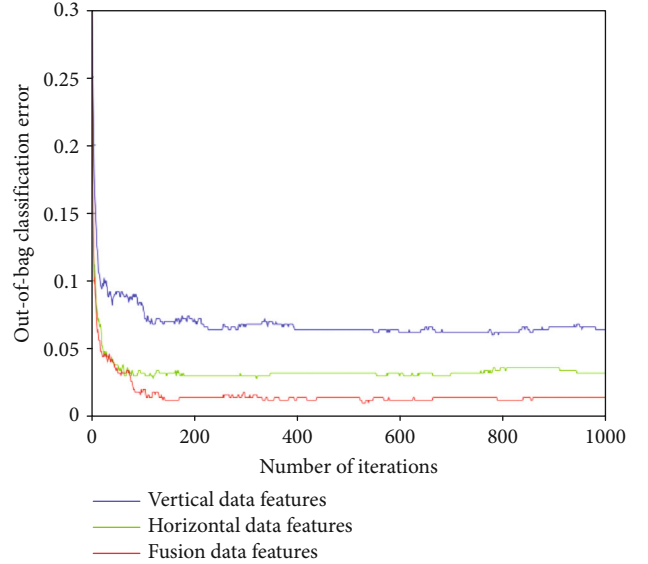

Figure 14: RF out-of-bag classification error for different channel data.

Therefore, to analyze the robustness of the diagnosis algorithm to the above parameters, this paper adjusts the above parameters of the sparse autoencoder in the DFNN network and calculates the final fault accuracy as shown in Tables 9–11 (when adjusting each parameter, the rest of the parameters remain unchanged).

It can be seen from Table 9 that when the L2 regularization weight decay coefficient is 3e-01 and 3e-02, the diagnostic accuracy is very low; when it is 1e-03 or even smaller, the diagnostic accuracy is basically consistent with the original accuracy, both of which are about 96.00%, indicating that the value of the L2 regularization weight decay coefficient should not be too large. Observing Tables 10 and 11, it is found that the diagnostic results corresponding to sparse penalty weight and sparsity parameters of different orders of magnitude are not much different, indicating that the diagnostic method in this paper is not sensitive to the changes in these two parameters. Based on the above analysis, the method in this paper has better robustness to sparse penalty weight and sparsity parameters, while the L2 decay coefficient should not be set too large, otherwise, the diagnosis effect will be affected.

5.3. Performance Comparison Analysis. To further prove the effectiveness of the fault diagnosis algorithm in this paper, this section first adopts different adaptive preprocessing methods to analyze the signal, including only VMD decomposition is performed on the signal; only empirical mode decomposition (EMD) is performed on the signal [53]; the particle swarm optimization algorithm is combined with VMD to decompose the signal, and the adaptive decomposition standard is the signal power spectrum entropy widely used in current research. It is noted as PSO-VMD-DE [54]. The particle swarm optimization algorithm is still combined with VMD, but the adaptive decomposition standard is consistent with the method in this paper, which is the MCKD
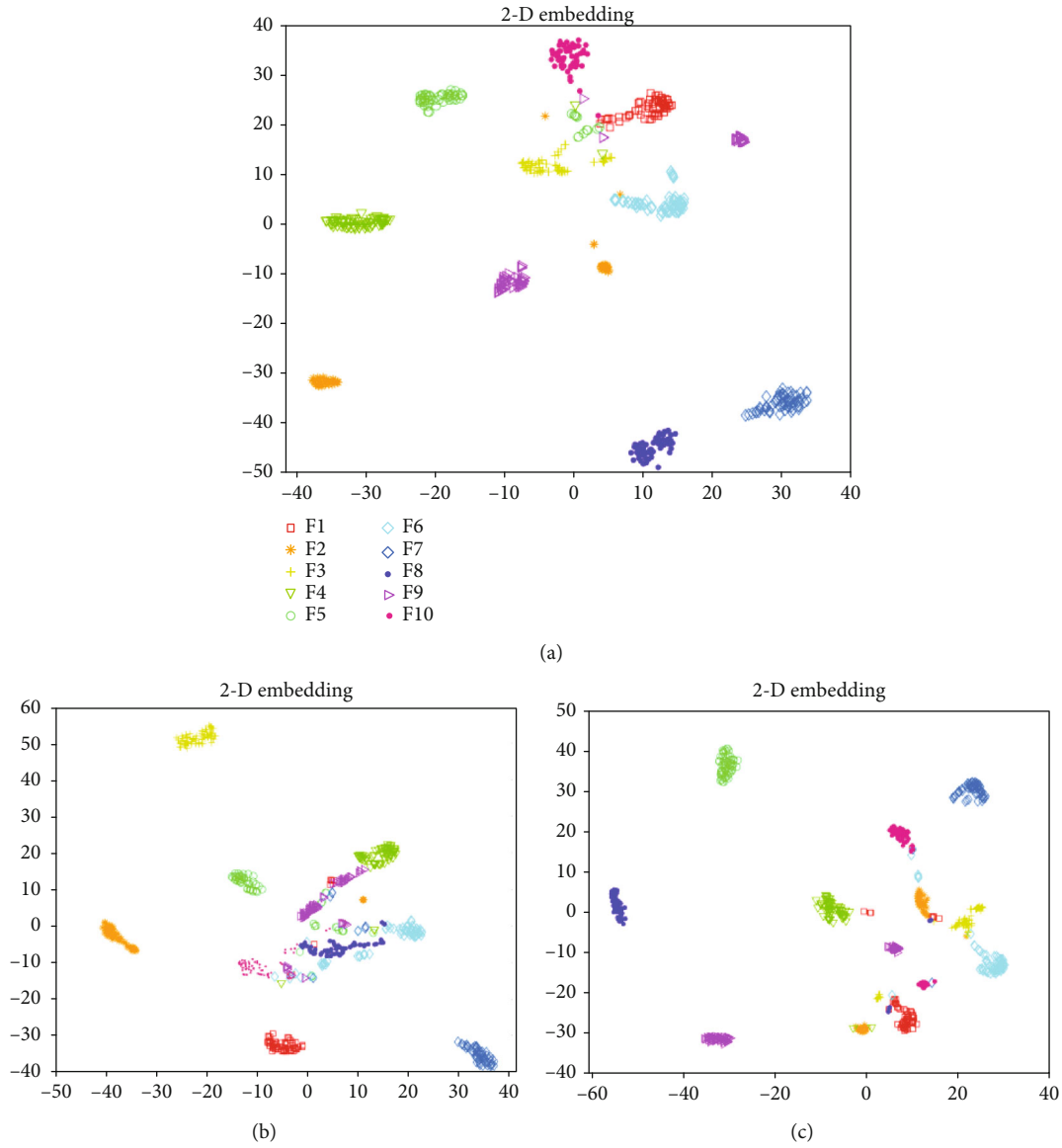
FIGURE 15: Feature visualization processing results: (a) fusion features, (b) horizontal data features, and (c) vertical data features.

value of the signal, which is recorded as PSO-VMD-MCKD; this method is denoted as NAPSO-VMD-MCKD. In addition, when the modal decomposition of the signal is performed directly, the parameter values such as the decomposition layers are taken as empirical values [55, 56]. The feature extraction and fusion steps are the same as those in Section 2.2, the classifier still uses RF, and the parameter settings are the same as the previous ones. The mean values of the multiple experimental results are shown in Table 12.

It can be seen from the above results that the VMD method has a better diagnosis effect than the EMD method because VMD effectively solves the modal aliasing problem in signal decomposition to a certain extent. However, since the signal is not decomposed adaptively, the diagnostic effect of the VMD method is lower than that of the PSO-VMD-DE and PSO-VMD-MCKD methods; the diagnostic effect of the PSO-VMD-MCKD method is higher than that of the PSO-VMD-DE method, which shows that in the fault diagnosis

of the gearbox, compared with the power spectrum entropy, it is better to perform adaptive decomposition based on the MCKD value of the signal. It retains the fault information of the signal, thereby improving the accuracy of fault diagnosis. Finally, because the NAPSO algorithm has a stronger optimization ability than PSO, the NAPSO-VMD-MCKD method obtains more adaptive decomposition parameter values than the PSO-VMD-MCKD method, and thus, the diagnosis effect is better. The above experimental results demonstrate the effectiveness of the NAPSO-VMD-MCKD signal adaptive decomposition method in this paper. And on the whole, no matter which adaptive processing method is used, the accuracy of dual-channel data is higher than that of single-channel data. The validity of the establishment of the dual-channel sensor information acquisition system in this paper has been proved again.

On the other hand, this section adopts several typical feature fusion methods to fuse the channel feature data

TABLE 9: Fault diagnosis accuracy of methods with different L2 regularization weight decay coefficient.

L2 regularization weight decay coefficient	3e-01	3e-02	1e-03	1e-05	3e-06
Diagnostic accuracy (%)	55.23	62.64	96.00	95.33	96.00

TABLE 10: Fault diagnosis accuracy of methods with different sparse penalty weight.

Sparse penalty weight	2e-01	2e-02	2e-03	1e-05	2e-06
Diagnostic accuracy (%)	95.33	96.00	96.00	95.33	95.33

TABLE 11: Fault diagnosis accuracy of methods with different sparsity parameters.

Sparsity parameter	1e-01	1e-02	1e-04	1e-05	1e-06
Diagnostic accuracy (%)	96.00	96.00	96.00	95.33	96.00

TABLE 12: Fault diagnosis accuracy of the different adaptive preprocessing methods.

Method	VMD	EMD	PSO-VMD-DE	PSO-VMD-MCKD	NAPSO-VMD-MCKD
Horizontal data accuracy (%)	78.23	73.20	88.00	90.03	93.00
Vertical data accuracy (%)	74.00	73.00	89.30	90.20	92.00
Dual-channel data accuracy (%)	89.35	85.30	91.32	94.23	96.00

TABLE 13: Fault diagnosis accuracy of different feature fusion methods.

Method	KPCA	FA	LDA	MDS	MT-SAE
Horizontal data accuracy (%)	88.86	82.02	75.00	85.92	93.00
Vertical data accuracy (%)	89.02	83.00	76.23	86.50	92.00
Dual-channel data accuracy (%)	93.57	87.33	81.00	91.33	96.00

before alignment in Section 3 to verify the effectiveness of feature alignment and feature fusion methods based on multitype SAE. The methods mainly include kernel principal component analysis (KPCA), factor analysis (FA), linear discriminant analysis (LDA), and multidimensional scaling (MDS) [57–60]. The kernel function used in KPCA is the Gaussian kernel function. The method in this paper is denoted as multitype sparse autoencoder (MT-SAE). To maintain the principle of invariance, the classifier still uses RF, and the parameter settings are the same as the previous ones. The specific diagnosis results are shown in Table 13. It can be seen from the table that the KPCA method is higher than FA, LDA, and MDS. This is because the nonlinear mapping component is introduced in KPCA, which is more suitable for dimension reduction and fusion of nonstationary signal features. The accuracy of the MT-SAE method is still higher than that of KPCA because the data feature alignment and feature fusion based on MT-SAE can reexpress the data features while minimizing the data reconstruction error. The feature information of the data is nonlinearly reconstructed, and the deep-level features of the data will be further extracted, fully retaining the individual character-

istics of the data, and the differences between different types of data will be effectively highlighted. Therefore, the effectiveness and superiority of the MT-SAE data feature fusion method have been proved.

Finally, different fault diagnosis methods are used to diagnose the fault of the gearbox to verify the superiority of the diagnosis algorithm proposed in this paper. These methods include inputting the raw vibration signal into a one-dimensional convolutional network (1D-CNN) and inputting the feature samples used in this paper into the SVM and Softmax classifiers, respectively [61–63]. In the training process of SVM, its two key parameters, the penalty parameter C and the kernel function parameter g, have been optimized by the NAPSO algorithm, and the objective function is the accuracy of fault diagnosis. The experimental results are shown in Table 14. It can be concluded that SVM cannot solve the multiclassification problem well, and the diagnosis result of dual-channel data only reaches 84.67%. In addition, although the diagnostic accuracy of 1D-CNN and Softmax classifiers can reach more than 90%, the accuracy of the two is still lower than that of RF classifiers because they have not been trained with sufficient data.

TABLE 14: Fault diagnosis accuracy of different classifiers.

Method	1D-CNN	SVM	Softmax	RF
Horizontal data accuracy (%)	85.36	79.32	89.00	93.00
Vertical data accuracy (%)	86.01	80.54	90.11	92.00
Dual-channel data accuracy (%)	91.22	84.67	93.21	96.00

The above experimental results demonstrate the effectiveness of the proposed algorithm for the gearbox hybrid fault classification problem.

In summary, the algorithm proposed in this paper can not only adaptively decompose and denoise various types of fault signals of the gearbox but also effectively and deeply fuse the data features of the dual-channel sensor. Finally, the hybrid fault diagnosis under variable working conditions of the gearbox is realized with high precision.

6. Conclusion

The condition monitoring is of great significance to the healthy and smooth operation of the equipment. A novel fault diagnosis algorithm of gearbox based on NAPSO-VMD self-adaptive noise reduction and dual-sensor feature fusion is proposed to solve the problem that adaptive noise reduction of the vibration signal of the equipment and the single-channel information cannot completely cover the signal fault information. Through the experimental verification of the gearbox, the following conclusions are obtained.

(1) By combining the nonlinear adaptive weight particle swarm algorithm and the variational mode decomposition method, the deconvolution of the maximum correlation spectrum kurtosis of the signal is used as the fitness function to perform adaptive VMD decomposition and noise reduction of the signal. The decomposed IMF matrix retains the periodic fault pulse components in the signal, which is very beneficial to the extraction and discrimination of fault features. In addition, after the inertia weight in particle swarm optimization is improved to exponential nonlinear weight, the optimization ability of the algorithm has been greatly improved

(2) In this paper, MCKD and common time-frequency domain features are extracted for the dual-channel signal, and the feature matrix can describe the fault feature performance from the original point of view of the signal. And when the time-frequency features of the signal are reconstructed and fused by multi-layer sparse autoencoders and then output, the fault features are deeply expressed, which is very beneficial to the diagnosis of fault data

(3) In the process of method verification, on the one hand, this paper compares and analyzes the fault diagnosis results of fusion features and single sensor signal features, which proves that fusion features can cover and describe the fault features of signals more

comprehensively and have higher diagnostic accuracy; on the other hand, the influence of the sensitivity parameters of SAE in the DFFN network on the diagnosis results is analyzed, and it is concluded that the robustness of the diagnostic algorithm in this paper to sparse penalty weight and sparsity parameter is better, but the diagnostic accuracy will be reduced if L2 regularization weight decay coefficient is set too large

(4) Sparse autoencoders have very powerful data feature reexpression capabilities. In this paper, two different sparse autoencoders, single-layer SAE and double-stacked SAE, are used to achieve the alignment and fusion of two-channel data features, respectively. Verified by RF classification experiments and compared with other common feature fusion methods, it is proved that the feature fusion method of DFFN in this paper is higher than other methods, and the effectiveness of the RF algorithm in classifying high-dimensional and multiclass samples has also been proved. To sum up, the fault diagnosis algorithm proposed in this paper can well realize the mixed fault diagnosis of gearbox equipment under variable working conditions

Effective fault diagnosis is of great significance to the smooth operation of equipment. In future work, further research will be done on the effective standard of adaptive preprocessing standard of the signal, the improvement of the optimization ability of the group optimization algorithm, the efficient fusion of multichannel signal features, and the improvement of the learning performance of the classifier.

Data Availability

The experimental data of this study can be obtained by contacting the corresponding author by email.

Conflicts of Interest

The authors of this article declare that there are no known competing financial interests or personal relationships that could influence the work of this article.

Acknowledgments

References

[1] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Computing and Applications*, vol. 32, no. 10, pp. 6111–6124, 2020.

[2] Z. Liu, K. Ding, H. Lin, G. He, C. Du, and Z. Chen, "A novel impact feature extraction method based on EMD and sparse decomposition for gear local fault diagnosis," *Machines*, vol. 10, no. 4, article 242, 2022.

[3] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Few-shot transfer learning for intelligent fault diagnosis of machine," *Measurement: Journal of the International Measurement Confederation*, vol. 166, article 108202, 2020.

[4] H. Wang, J. Xu, C. Sun, R. Yan, and X. Chen, "Intelligent fault diagnosis for planetary gearbox using time-frequency representation and deep reinforcement learning," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 2, pp. 985–998, 2022.

[5] Y. Wang and S. Wang, "Soft sensor for VFA concentration in anaerobic digestion process for treating kitchen waste based on SSAE-KELM," *IEEE Access*, vol. 9, pp. 36466–36474, 2021.

[6] H. Deng, W. X. Zhang, and Z. F. Liang, "Application of BP neural network and convolutional neural network (CNN) in bearing fault diagnosis," *IOP Conference Series: Materials Science and Engineering*, vol. 1043, no. 4, 2021.

[7] R. Bai, Q. Xu, Z. Meng, L. Cao, K. Xing, and F. Fan, "Rolling bearing fault diagnosis based on multi-channel convolution neural network and multi-scale clipping fusion data augmentation," *Measurement: Journal of the International Measurement Confederation*, vol. 184, article 109885, 2021.

[8] H. Shao, J. Lin, L. Zhang, D. Galar, and U. Kumar, "A novel approach of multisensory fusion to collaborative fault diagnosis in maintenance," *Information Fusion*, vol. 74, pp. 65–76, 2021.

[9] Z. Liu, W. Guo, Z. Tang, and Y. Chen, "Multi-sensor data fusion using a relevance vector machine based on an ant colony for gearbox fault detection," *Sensors (Switzerland)*, vol. 15, no. 9, pp. 21857–21875, 2015.

[10] G. Cheng, X. H. Chen, X. L. Shan, H. G. Liu, and C. F. Zhou, "A new method of gear fault diagnosis in strong noise based on multi-sensor information fusion," *Journal of Vibration and Control*, vol. 22, no. 6, pp. 1504–1515, 2016.

[11] T. Han, D. Jiang, Q. Zhao, L. Wang, and K. Yin, "Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery," *Transactions of the Institute of Measurement and Control*, vol. 40, no. 8, pp. 2681–2693, 2018.

[12] Z. X. Yang, X. B. Wang, and J. H. Zhong, "Representational learning for fault diagnosis of wind turbine equipment: a multi-layered extreme learning machines approach," *Energies*, vol. 9, no. 6, article 379, 2016.

[13] M. Cerrada, G. Zurita, D. Cabrera, R. V. Sánchez, M. Artés, and C. Li, "Fault diagnosis in spur gears based on genetic algorithm and random forest," *Mechanical Systems and Signal Processing*, vol. 70, pp. 87–103, 2016.

[14] Z. Wang, W. Du, J. Wang et al., "Research and application of improved adaptive MOMEDA fault diagnosis method," *Measurement*, vol. 140, pp. 63–75, 2019.

[15] T. Liu, Z. Luo, J. Huang, and S. Yan, "A comparative study of four kinds of adaptive decomposition algorithms and their applications," *Sensors*, vol. 18, no. 7, article 2120, 2018.

[16] Q. Ni, J. C. Ji, K. Feng, and B. Halkon, "A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis," *Mechanical Systems and Signal Processing*, vol. 164, article 108216, 2022.

[17] J. Cheng, Y. Yang, N. Hu, Z. Cheng, and J. Cheng, "A noise reduction method based on adaptive weighted symplectic geometry decomposition and its application in early gear fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 149, article 107351, 2021.

[18] K. Feng, W. A. Smith, P. Borghesani, R. B. Randall, and Z. Peng, "Use of cyclostationary properties of vibration signals to identify gear wear mechanisms and track wear evolution," *Mechanical Systems and Signal Processing*, vol. 150, article 107258, 2021.

[19] W. Fan, G. Cai, Z. K. Zhu, C. Shen, W. Huang, and L. Shang, "Sparse representation of transients in wavelet basis and its application in gearbox fault feature extraction," *Mechanical Systems and Signal Processing*, vol. 56, pp. 230–245, 2015.

[20] Y. Miao, M. Zhao, J. Lin, and Y. Lei, "Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 92, pp. 173–195, 2017.

[21] X. Zhang, J. Zhao, X. Ni, F. Sun, and H. Ge, "Fault diagnosis for gearbox based on EMD-MOMEDA," *International Journal of Systems Assurance Engineering and Management*, vol. 10, no. 4, pp. 836–847, 2019.

[22] S. Zhou, M. Xiao, P. Bartos, M. Filip, and G. Geng, "Remaining useful life prediction and fault diagnosis of rolling bearings based on short-time Fourier transform and convolutional neural network," *Shock and Vibration*, vol. 2020, Article ID 8857307, 14 pages, 2020.

[23] Y. Yao, H. Wang, S. Li et al., "End-to-end convolutional neural network model for gear fault diagnosis based on sound signals," *Applied Sciences (Switzerland)*, vol. 8, no. 9, article 1584, 2018.

[24] Q. Hu, A. Qin, Q. Zhang, J. He, and G. Sun, "Fault diagnosis based on weighted extreme learning machine with wavelet packet decomposition and KPCA," *IEEE Sensors Journal*, vol. 18, no. 20, pp. 8472–8483, 2018.

[25] T. Xie, X. Huang, and S. -K. Choi, "Intelligent mechanical fault diagnosis using multisensor fusion and convolution neural network," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 3213–3223, 2022.

[26] T. Li, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Adaptive channel weighted CNN with multisensor fusion for condition monitoring of helicopter transmission system," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8364–8373, 2020.

[27] R. Cao and A. Yunusa-Kaltungo, "An automated data fusion-based gear faults classification framework in rotating machines," *Sensors*, vol. 21, no. 9, article 2957, 2021.

[28] P. Yang, C. Wen, H. Geng, and P. Liu, "Intelligent fault diagnosis method for blade damage of quad-rotor UAV based on stacked pruning sparse denoising autoencoder and convolutional neural network," *Machines*, vol. 9, no. 12, article 360, 2021.

[29] S. Ullah and D. H. Kim, "Benchmarking Jetson platform for 3D point-cloud and hyper-spectral image classification," in *2020 IEEE International conference on big data and smart computing (BigComp)*, pp. 477–482, Busan, Korea (South), February 2020.

[30] G. Zhang, S. Zhao, W. Li, Q. Du, Q. Ran, and R. Tao, "HTD-Net: a deep convolutional neural network for target detection in hyperspectral imagery," *Remote Sensing*, vol. 12, no. 9, article 1489, 2020.

[31] S. Liu, X. Qian, H. Wan, Z. Ye, S. Wu, and X. Ren, "NPC three-level inverter open-circuit fault diagnosis based on adaptive electrical period partition and random forest," *Journal of Sensors*, vol. 2020, Article ID 9206579, 18 pages, 2020.

[32] T. Han and D. Jiang, "Rolling bearing fault diagnostic method based on VMD-AR model and random forest classifier," *Shock and Vibration*, vol. 2016, Article ID 5132046, 11 pages, 2016.

[33] X. Zhang, Q. Miao, H. Zhang, and L. Wang, "A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery," *Mechanical Systems and Signal Processing*, vol. 108, pp. 58–72, 2018.

[34] Y. He, H. Wang, H. Xue, and T. Zhang, "Research on unknown fault diagnosis of rolling bearings based on parameter-adaptive maximum correlation kurtosis deconvolution," *Review of Scientific Instruments*, vol. 92, no. 5, article 055103, 2021.

[35] H. Chen, Y. Xiong, S. Li, Z. Song, Z. Hu, and F. Liu, "Multisensor data driven with PARAFAC-IPSO-PNN for identification of mechanical nonstationary multi-fault mode," *Machines*, vol. 10, no. 2, article 155, 2022.

[36] M. Ye, X. Yan, and M. Jia, "Rolling bearing fault diagnosis based on VMD-MPE and PSO-SVM," *Entropy*, vol. 23, no. 6, article 762, 2021.

[37] M. Zellagui, A. Lasmari, S. Settoul, C. Z. El-Bayeh, and R. Chenni, "Assessment integration of hybrid PV-DSTAT-COM-BES-DG system in EDS under uncertainties using chaotic adaptive inertia weight PSO algorithms," in *2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, pp. 1–8, Bucharest, Romania, March 2021.

[38] D. Połap and M. Woźniak, "Red fox optimization algorithm," *Expert Systems with Applications*, vol. 166, article 114107, 2021.

[39] J. Xiong, D. Yu, S. Liu, L. Shu, X. Wang, and Z. Liu, "A review of plant phenotypic image recognition technology based on deep learning," *Electronics*, vol. 10, no. 1, pp. 1–19, 2021.

[40] Y. Mao, D. Xiao, J. Cheng et al., "Multigrades classification model of magnesite ore based on SAE and ELM," *Journal of Sensors*, vol. 2017, Article ID 9846181, 9 pages, 2017.

[41] X. Wang, Z. Chu, B. Han, J. Wang, G. Zhang, and X. Jiang, "A novel data augmentation method for intelligent fault diagnosis under speed fluctuation condition," *IEEE Access*, vol. 8, pp. 143383–143396, 2020.

[42] M. Sun, H. Wang, P. Liu, S. Huang, P. Wang, and J. Meng, "Stack autoencoder transfer learning algorithm for bearing fault diagnosis based on class separation and domain fusion," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 3, pp. 3047–3058, 2022.

[43] S. Pang, X. Yang, X. Zhang, and Y. Sun, "Fault diagnosis of rotating machinery components with deep ELM ensemble induced by real-valued output-based diversity metric," *Mechanical Systems and Signal Processing*, vol. 159, article 107821, 2021.

[44] Z. Dong, Y. Liu, J. Kang, and S. Zhang, "A novel generative method for machine fault diagnosis," *Journal of Sensors*, vol. 2022, Article ID 5420478, 11 pages, 2022.

[45] Y. Li, X. Chen, J. Yu, and X. Yang, "A fusion frequency feature extraction method for underwater acoustic signal based on variational mode decomposition, duffing chaotic oscillator and a kind of permutation entropy," *Electronics (Switzerland)*, vol. 8, no. 1, article 61, 2019.

[46] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018.

[47] X. Sui, K. Wan, and Y. Zhang, "Pattern recognition of SEMG based on wavelet packet transform and improved SVM," *Optik*, vol. 176, pp. 228–235, 2019.

[48] S. Ma, M. Chen, J. Wu, Y. Wang, B. Jia, and Y. Jiang, "Intelligent fault diagnosis of HVCB with feature space optimization-based random forest," *Sensors*, vol. 18, no. 4, article 1221, 2018.

[49] J. Tian, L. Liu, F. Zhang, Y. Ai, R. Wang, and C. Fei, "Multi-domain entropy-random forest method for the fusion diagnosis of inter-shaft bearing faults with acoustic emission signals," *Entropy*, vol. 22, no. 1, article 57, 2020.

[50] N. Liu, B. Liu, and C. Xi, "Fault diagnosis method of rolling bearing based on the multiple features of LMD and random forest," *IOP Conference Series: Materials Science and Engineering*, vol. 892, no. 1, 2020.

[51] R. H. Lin, Z. X. Pei, Z. Z. Ye, C. C. Guo, and B. D. Wu, "Hydrogen fuel cell diagnostics using random forest and enhanced feature selection," *International Journal of Hydrogen Energy*, vol. 45, no. 17, pp. 10523–10535, 2020.

[52] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, and X. Chen, "Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2416–2425, 2019.

[53] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1–2, pp. 108–126, 2013.

[54] F. Zhang, W. Sun, H. Wang, and T. Xu, "Fault diagnosis of a wind turbine gearbox based on improved variational mode algorithm and information entropy," *Entropy*, vol. 23, no. 7, article 794, 2021.

[55] S. Mohanty, K. Kumar Gupta, and K. Solomon Raju, "Comparative study between VMD and EMD in bearing fault diagnosis," in *2014 9th International Conference on Industrial and Information Systems (ICIIS)*, pp. 1–6, Gwalior, India, December 2014.

[56] X. Li, Z. Ma, D. Kang, and X. Li, "Fault diagnosis for rolling bearing based on VMD-FRFT," *Measurement*, vol. 155, article 107554, 2020.

[57] Y. Sun, W. Qin, Z. Zhuang, and H. Xu, "An adaptive fault detection and root-cause analysis scheme for complex industrial processes using moving window KPCA and information geometric causal inference," *Journal of Intelligent Manufacturing*, vol. 32, no. 7, pp. 2007–2021, 2021.

[58] R. Yan, F. Shen, C. Sun, and X. Chen, "Knowledge transfer for rotary machine fault diagnosis," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8374–8393, 2020.

[59] J. Liu, Y. Hu, B. Wu, and Y. Wang, "An improved fault diagnosis approach for FDM process with acoustic emission," *Journal of Manufacturing Processes*, vol. 35, pp. 570–579, 2018.

[60] Y. Qu, Y. Zhang, M. He, D. He, C. Jiao, and Z. Zhou, "Gear pitting fault diagnosis using disentangled features from

unsupervised deep learning," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 233, no. 5, pp. 719–730, 2019.

[61] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier," *Journal of Signal Processing Systems*, vol. 91, no. 2, pp. 179–189, 2019.

[62] L. M. R. Baccarini, V. V. R. Silva, B. R. de Menezes, and W. M. Caminhas, "SVM practical industrial application for mechanical faults diagnostic," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6980–6984, 2011.

[63] S. Tao, T. Zhang, J. Yang, X. Wang, and W. Lu, "Bearing fault diagnosis method based on stacked autoencoder and softmax regression," in *2015 34th Chinese Control Conference (CCC)*, pp. 6331–6335, Hangzhou, China, July 2015.