SCIPEDIA

Bearing Fault Diagnosis Based on AVMD and HPO-DBN

Xinyu Sha[1] and Fucai Qian[1,2,*]

[1] School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, 710048, China

[2] Autonomous Systems and Intelligent Control International Joint Research Centre, Xi'an Technological University, Xi'an, 710021, China

INFORMATION

Keywords:
Variational mode decomposition
DBN
hunter-prey optimisation algorithm
fault diagnosis

DOI: 10.23967/j.rimni.2025.10.71862

SCIPEDIA Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

Revista Internacional Métodos numéricos para cálculo y diseño en ingeniería RIMNI

In cooperation with CIMNE

Bearing Fault Diagnosis Based on AVMD and HPO-DBN

Xinyu Sha[1] **and Fucai Qian**[1,2,*]

[1]School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, 710048, China

[2]Autonomous Systems and Intelligent Control International Joint Research Centre, Xi'an Technological University, Xi'an, 710021, China

ABSTRACT

To overcome difficulties such as non-stationary vibrations, high-dimensional feature redundancy, and mode selection issues that may arise during signal decomposition in bearing fault diagnosis. We propose an adaptive method called Adaptive Variational Mode Decomposition (AVMD) for extracting time-frequency domain characteristics from the bearing vibration displacement signals to the maximum extent possible. Next, the ReliefF algorithm is employed to select desired features, and an autoencoder is used to reduce the selected features dimensionally. Furthermore, because the Hunter-Prey Optimisation (HPO) algorithm can balance multiple objectives during the search process by utilising the concepts of hunter and prey to generate a better solution set, incorporating this algorithm into the Deep Belief Network (DBN) establishes an HPO-DBN fault diagnosis model. Subsequently, we validate the proposed method using both public datasets and field compressor data. Moreover, we compare the results with those obtained from the Support Vector Machine (SVM). The findings indicate that this approach enhances the bearing fault identification rate, thus supporting predictive maintenance of bearings.

OPEN ACCESS

Received: 13/08/2025

Accepted: 16/10/2025

Published: 23/01/2026

DOI
10.23967/j.rimni.2025.10.71862

Keywords:
Variational mode decomposition
DBN
hunter-prey optimisation algorithm
fault diagnosis

1 Introduction

Rolling bearings are sophisticated mechanical components that transform sliding friction between a rotating shaft and its housing into rolling friction, thereby reducing frictional losses [1]. The failure of rolling bearings, a crucial component in contemporary industrial production, can have a profound impact on equipment safety and the efficiency of industrial manufacturing. Damaged rolling bearings or those with reduced reliability often constitute the primary causes of rotating machinery failures, accounting for approximately 30% of total failures [2].

In 2000, a major safety accident occurred in the Three Gorges Project in the Yangtze River, resulting in three fatalities and 31 injuries. The accident investigation report indicated that the cause of the accident was the incorrect use of bearing models, leading to bearing damage [3]. In 2005, a bearing failure in a large industrial fan in Miami, Florida, caused the entire fan to suddenly lose control, detach from its base, and collapse [4]. In 2010, a centrifugal blower bearing failure occurred

*Correspondence: Fucai Qian (fcqian@xaut.edu.cn).

in a chemical plant in Zhuhai, Guangdong Province, China, triggering the high-speed rotation of the blower. Because the blower cannot be shut down, the blower casing eventually ruptured, releasing high-temperature gas and chemicals, resulting in multiple personnel injuries [5]. In 2014, an engine bearing failure at an aircraft engine maintenance centre in Frankfurt, Germany, caused the engine to abruptly come to a halt during taxiing, leading to an emergency situation [6]. During a drill in 2018, the inaugural vessel of the United States Navy's new generation Ford-class nuclear-powered aircraft carrier encountered severe mechanical issues. These problems included a major failure in the main propeller bearing. Consequently, the entire exercise had to be suspended, leading to substantial losses in the military-industrial sector's economy [7].

The above examples provide clear evidence regarding the importance of conducting further research on bearing fault diagnosis. Such research is crucial for ensuring the secure operation of diverse mechanical equipment.

In the year 2014, Wang and colleagues described the Fault Characteristic Order (FCO) Analysis method for diagnosing rolling element bearing faults in the journal Mechanical Systems and Signal Processing. However, they did not employ advanced signal processing methods to extract faults more effectively. In the presented work, the AVMD-based feature extraction method is integrated with HPO-DBN to enhance the diagnostic accuracy and robustness through varying operation conditions [8]. Li and Wang suggested the Filtered-Waveform Variational Mode Decomposition (VMD) to extract the rolling bearing fault features in Entropy in 2025. Although VMD successfully improved the separation of fault features, it did not incorporate deep learning models for real-time fault detection. The provided work demonstrated a feature extraction method that employed VMD, but the proposed method incorporated an HPO-DBN to yield a much more favourable sensitivity and accuracy, and it adapted well to complexitiesassociated with fault signatures [9]. In 2009, Liu et al. introduced Empirical Mode Decomposition (EMD) as a method for diagnosing bearing faults [10]. However, they did not take into consideration the issue of mode mixing that may occur during its practical application. In 2014, Dragomiretskiy introduced the Variational Mode Decomposition (VMD) method in the IEEE [11]. Since then, VMD has been extensively utilised by numerous scholars in the field of bearing fault diagnosis. In the domain of automation, the VMD technique has demonstrated sensitivity to noise and the requirement for manual parameter adjustment. To address these issues, Salakutdinov et al. achieved this by optimising and configuring VMD parameters using neural networks. However, the actual results were not satisfactory compared to the expected values [12]. The combination of data fusion and deep learning methods, as proposed by Dinesh Kumar Reddy Basani, has been successfully implemented in this project using AVMD and HPO-DBN for bearing fault diagnosis. This implementation improves both the accuracy of fault detection and real-time monitoring of the bearing system. In addition, the combination is potentially more efficient in terms of computational effort, providing a deeper and more accurate analysis of bearing conditions, unlike Panga's system [13]. Yin et al. transformed one-dimensional time series vibration signals into three different time-frequency domain representations. These representations were fed into the diagnostic network [14]. However, during the network computation process, it was discovered that the high correlation between features resulting from their lack of selection greatly reduced the computing speed.

Furthermore, the simple network structure frequently led to infinite looping. In 2015, Jie and co-authors introduced a method utilising Deep Belief Networks (DBN) to enhance bearing fault diagnosis. They made modifications to the network structure, aiming to address the challenge of infinite loops. However, they did not take into consideration the parameter selection issue of the DBN network [15]. In 2024, Naresh Kumar Reddy Panga's work utilises the Discrete Wavelet Transform (DWT) for feature extraction in ECG signals in IoT health monitoring systems. However, it does

not consider any limitations in signal processing that relate to non-stationary signals. Variational Mode Decomposition (VMD) and Hybrid Particle Optimised Deep Belief Networks (HPO-DBN) are adopted in the proposed work of a bearing fault diagnosis approach, which furthers the signal processing and fault detection. The proposed method leads to greater accuracy and real-time monitoring of bearing conditions [16].

This paper introduces an innovative approach to diagnosing faults in mechanical equipment, employing a combination of diverse features. First, the vibration signals are subjected to spectral analysis and time-frequency analysis. Second, the feature extraction process integrates the application of the Adaptive Variational Mode Decomposition (AVMD) method on various domains, including the time domain, frequency domain (consisting of frequency spectrum and envelope spectrum), and time-frequency domain [17]. Next, the ReliefF algorithm is utilised for feature selection, and the selected features are fused using an autoencoder. Lastly, the application of the Hunter-Prey Optimisation (HPO) algorithm empowers the Deep Belief Network to autonomously identify the optimal parameters, thereby enhancing the precision of fault diagnosis outcomes [18].

The main contributions of this article are:

(1) By employing adaptive variational mode decomposition (AVMD) for feature extraction, several advantages can be achieved, including the elimination of manual parameter adjustment, mitigation of noise impact on decomposition results, and enhancement of feature extraction stability.

(2) By incorporating the hunter-prey optimisation algorithm (HPO), the deep belief network (DBN) can be optimised, effectively solving the problem of using manual parameter setting.

Below is the structural diagram of the bearing fault diagnosis system, which is based on AVMD and HPO-DBN (Fig. 1).

2 Feature Extraction

For the diagnosis of bearing faults, it is imperative to collect the original signals of the bearings. Subsequently, crucial features that reflect bearing faults from multiple perspectives should be extracted from these original signals.

2.1 Time Domain Feature

The initial signal illustrates the movement of the bearing and is comprised of elements not linked to the fault, components associated with the fault patterns, significant noise, and any other intervention. This signal can be mathematically expressed as follows: $x(n) = \theta(n) + \alpha(n) + \varepsilon(n)$. Where $\theta(n)$ is the normal component, $\alpha(n)$ is the fault component when the fault occurs, and $\varepsilon(n)$ is the noise.

In order to reduce the dependency on prior knowledge, M statistical feature is extracted from the original signal. Some of these characteristics are:

Mean: If no fault occurs in $x(n)$, $\alpha(n) = 0$. Because noise exists and noise $\varepsilon(n)$ fluctuates around 0. If $x(n)$ is averaged over a period of time, the noise can be cancelled, and the normal component of the original signal can be obtained. But when a fault occurs, the mean value will change, so by analysing the change of the mean value, the type and degree of bearing fault can be detected.

$$f_1 = \frac{1}{N}\sum_{n=1}^{N} x(n) \tag{1}$$

where $x(n)$ is the original signal and N is the number of samples.

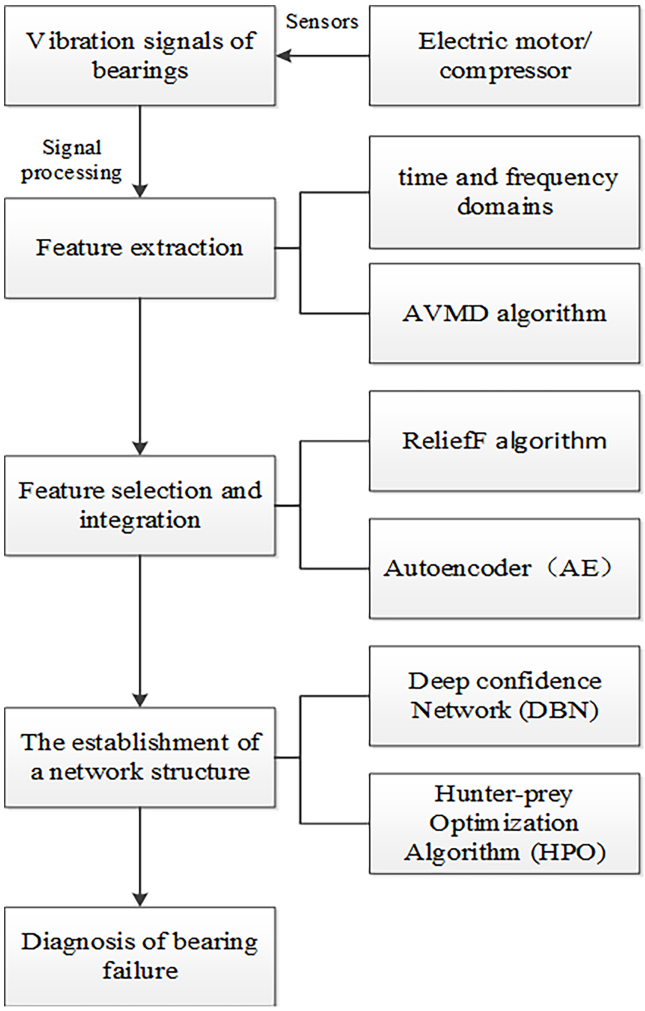

Figure 1: System architecture

Maximum amplitude: An increase in bearing signal amplitude indicates the occurrence of bearing failure. Different types of bearing failures, such as rolling element damage, raceway wear, eccentricity, and stuck conditions, can be identified by monitoring the variations in the maximum amplitude [19]. The maximum amplitude can also provide information about the severity of the bearing failure. When the failure gradually develops or intensifies, the maximum amplitude tends to increase gradually, and the moment of the failure can also be determined.

$$f_2 = \max(|x(n)|) \tag{2}$$

Standard deviation: Standard deviation can be used to detect whether the bearing has abnormal vibration. When the bearing fails, the vibration signal often presents irregularity or unstable changes. An abnormal vibration signal will cause the standard deviation of the vibration signal to increase. Bearing faults can be detected and diagnosed by monitoring the change of standard deviation [20]. At the same time, the change of the standard deviation of the abnormal vibration signal can provide

information about the fault type. For example, faults such as rolling element damage and raceway wear may increase the amplitude variation of the vibration signal, thus increasing the standard deviation.

$$f_3 = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x(n)-f_1)^2} \tag{3}$$

Square root amplitude: The square root amplitude is used to measure the fluctuation size of the data set, which can provide an estimate of the energy of the vibration signal and provide important information in the bearing fault diagnosis [21]. For example, failures such as rolling element damage and raceway wear may increase the amplitude variation of the vibration signal, thus increasing the square root amplitude. By monitoring the change of square root amplitude, bearing faults can be detected and diagnosed.

$$f_4 = \left(\frac{1}{N}\sum_{n=1}^{N}\sqrt{|x(n)|}\right)^2 \tag{4}$$

Skewness: Skewness describes how far and in what direction the data distribution deviates from the mean. When the bearing fails, the vibration signal may have an atypical skew distribution, and different types of bearing faults may lead to different skew directions of the vibration signal distribution [22,23]. For example, faults such as rolling body damage and raceway wear may skew the vibration signal distribution to the right, that is, with positive skew. By analysing the variation mode of skewness, the type of bearing fault can be preliminarily determined [24].

$$f_5 = \frac{1}{N-1}\sum_{n=1}^{N}\left(\frac{x(n)-f_1}{f_4}\right)^3 \tag{5}$$

Kurtosis: Kurtosis describes the peak characteristics of the data distribution, i.e., the degree of concentration of values in the data set and the shape of the tail distribution. When bearings fail, vibration signals may exhibit atypical peak characteristics. For example, faults such as rolling body damage and raceway wear may cause the peak state of vibration signal distribution to be steeper, that is, with higher positive kurtosis [25].

$$f_6 = \frac{1}{N-1}\sum_{n=1}^{N}\left(\frac{x(n)-f_1}{f_4}\right)^4 \tag{6}$$

2.2 Frequency Domain Features

For bearing vibration signal feature extraction, fault features that are not obvious in the time domain can be obtained through spectrum analysis of vibration displacement signals and extraction of frequency domain features [26]. Some of the frequency domain features are:

The calculation of the average frequency of a vibration signal enables the comprehensive representation of the signal's frequency domain characteristics. By conducting this calculation, the influence of noise on the original signal is effectively eliminated, facilitating the extraction of precise and accurate signals [27].

$$f_1 = \frac{1}{N_{fft}}\sum_{j=1}^{N_{fft}}X(j) \tag{7}$$

In the frequency domain after transformation, the original signal, denoted as $X(j)$, corresponds to the transformed sampling count represented by N_{fft}.

Centre frequency: The centre frequency is determined by calculating the weighted average of each frequency component on the spectrum to determine the concentration location of the signal's main energy [28]. The occurrence of bearing failure often leads to alterations in the vibration signal spectrum, manifesting as the emergence of supplementary frequency peaks or bands. These changes can be used to identify bearing faults by comparing the centre frequency of vibration signals with the centre frequency under normal operation [29].

$$f_2 = \sum_{j=1}^{N_{fft}} [f(j)X(j)] / \sum_{j=1}^{N_{fft}} X(j) \tag{8}$$

RMS frequency: RMS frequency is a statistic that measures the frequency distribution of a signal, which reflects the balance of energy or amplitude of the signal at different frequencies [30]. The more serious the bearing fault, the more obvious the change in energy or amplitude of the fault frequency component, and the larger the amplitude of the root mean square frequency value. For example, rolling element failure usually causes an increase in the RMS of frequency in the lower frequency range; Raceway failure usually causes an increase in the RMS of frequency in the high frequency range [31].

$$f_3 = \sqrt{\sum_{j=1}^{N_{fft}} [f(j)X(j)] / \sum_{j=1}^{N_{fft}} X(j)} \tag{9}$$

Frequency standard deviation: The frequency standard deviation characterises the extent of dispersion among frequency components with respect to the frequency mean. It specifically highlights the variability range observed within the signal's frequency distribution. When a bearing fails, the distribution of the frequency component often changes, resulting in an increase in the frequency standard deviation. By monitoring the change of frequency standard deviation, the presence of bearing faults can be found in time [32].

$$f_4 = \sqrt{\sum_{j=1}^{N_{fft}} \{[f(j) - f_3]^2 X(j)\} / \sum_{j=1}^{N_{fft}} X(j)} \tag{10}$$

2.3 Time-Frequency Domain Features

As the vibration signal from bearings typically exhibits non-stationary behaviour, its statistical properties undergo temporal variations. By extracting time-frequency domain features, one can effectively capture the signal's temporal and spectral changes, thereby obtaining fault characteristics during the evolving process [33].

The time-frequency domain provides the following characteristic features:

Entropy feature: The entropy feature describes the complexity and amount of information of the signal spectrum; it is also used to describe the randomness and periodicity of the signal. The vibration signal caused by a bearing fault is often drowned out by noise, which makes feature extraction difficult. Entropy features are robust to noise reduction and fault feature enhancement, and can extract nonlinear vibration features from fault signals, thus enhancing bearing fault diagnosis performance. For example, when the bearing works normally, the vibration signal generally has an obvious periodicity. When the bearing fails, the vibration signal may lose periodicity due to the

influence of friction, fracture or loosening. Entropy features can quantify the disorder or aperiodicity of signals, providing a powerful means to detect bearing faults [34].

$$F = -\sum_{i=1}^{N} p_i \ln p_i \tag{11}$$

where the probability density function of each component is denoted as p_i.

2.4 Feature Extraction

For feature extraction in the time-frequency domain, an improved variational mode decomposition (AVMD) method is proposed. The AVMD method builds upon enhancements made to the variational mode decomposition (VMD) technique. The VMD method is introduced below.

The signal is decomposed by the VMD method into k eigenmode functions, and the original signal is formed by the superposition of all these components [35]. The expression for the intrinsic mode function is as follows:

$$v_k(t) = A_k(t)\cos(\phi_k(t)) \tag{12}$$

In this context, the phase function $\phi_k(t)$ exhibits a non-monotonic decrease.

For each $v_k(t)$, the Hilbert transform yields the analytical signal.

$$\left[\delta(t) + \frac{j}{\pi t}\right] * v_k(t) \tag{13}$$

The equation involves the Dirac function represented by $\delta(t)$ and the convolution denoted by "$*$".

Next, the eigenmode function's spectral diagram is shifted to the corresponding frequency baseband, denoted by the Fourier transform:

$$[(\delta(t) + \frac{j}{\pi t}) * v_k(t)]e^{-j\omega_k t} \tag{14}$$

where, $e^{-j\omega_k t}$ is the central frequency index.

Then the centre frequency of the eigenmode function is estimated, and the corresponding eigenmode function is reconstructed. This paper constructs the following optimisation problems:

$$\begin{aligned} &\min_{\{v_2\},\{\alpha_2\}} \left\{ \sum_k \left\| \frac{\partial\left[\left(\delta(t)+\frac{j}{\pi t}\right) * v_k(t)\right] e^{-j\alpha_2 t}}{\partial t} \right\|_{[2}^2 \right\} \\ &\text{st.} \sum_{k=1}^{K} v_k = X(j) \end{aligned} \tag{15}$$

For the given problem, the objective is to minimise the total centre frequency bandwidth of each modal component by considering the 2-norm of the spectral term as the objective function. This minimisation allows for the decomposition of the original signal into optimal individual eigenmode functions. Additionally, it is necessary to ensure that the total energy of the original signal equals the sum of the energies of the modal components.

To solve the variational equation mentioned above, the constrained variational problem is converted into an unconstrained variational problem. This transformation incorporates the augmented

Lagrange function, which combines the quadratic penalty term and the Lagrange multiplier method:

$$L(\{v_k\},\{\omega_k\},\lambda)=\alpha\sum_k||\partial_t[\left(\delta(t)+\frac{j}{\pi t}\right)*v_k(t)]e^{-j\omega_k t}||_2^2+||f(t)-\sum_k v_k(t)||_2+\{\lambda(t),f(t)-\sum_k v_k(t)\} \quad (16)$$

where α is the penalty function, λ is the Lagrange multiplier.

Then iterate:

$$\hat{v}_k^{n+1}(\omega)=\frac{f(\omega)-\sum_{i=1,i<k}^{K}\hat{v}_i^{n+1}(\omega)-\sum_{i=1,i<k}^{K}\hat{v}_i^{n}(\omega)+\frac{\hat{\lambda}^n(\omega)}{2}}{1+2\alpha(\omega-\omega_k^n)^2} \quad (17)$$

$$\omega_k^{n+1}=\frac{\int_0^\infty \omega|\hat{v}_k^{n+1}(\omega)|^2d\omega}{\int_0^\infty|\hat{v}_k^{n+1}(\omega)|^2d\omega} \quad (18)$$

$$\hat{\lambda}^{n+1}(\omega)=\lambda^n(\omega)+\gamma\left[\hat{f}(\omega)-\sum_{k=1}^{K}\hat{v}^{n+1}(\omega)\right] \quad (19)$$

γ stands for noise tolerance. Finally, each eigenmode function is obtained.

The previously described method is a classical approach known as Variational Mode Decomposition (VMD). However, its effectiveness relies on manually selecting the number of eigenmode function components for the final decomposition. To ensure a complete decomposition of the original signal and maximise the signal's energy ratio, it is necessary to carefully choose this value. To address these limitations and capture the complete characteristics of all signals, this paper proposes an adaptive mode decomposition approach called Energy Ratio-based Adaptive Mode Decomposition (AVMD).

The following formulas are used to calculate the energy of the original signal and each intrinsic mode function:

$$E=\sqrt{\frac{\sum_{n=1}^{N}x^2(n)}{N}} \quad (20)$$

In this context, the signal energy E is calculated based on the signal sequence $x(n)$ and the number of sampling points N.

The expression for the energy ratio is as follows:

$$\delta=\frac{\sum_{m=1}^{K}E_m}{E_x} \quad (21)$$

where the total energy of the signal E_x and the energy of each intrinsic mode function E_m are defined.

The optimal value for k, which maximises the energy ratio for the first time, can be observed from the provided formula. The AVMD algorithm can be expressed using the following steps:

Step1. Input the original signal and compute the energy ratio when $k=2$.

Step2. Set $k=k+1$, calculate the energy sum of each eigenmode function and calculate the energy ratio.

Step3. Check if the current energy ratio surpasses the energy ratio from the previous iteration. Suppose it doesn't, consider the energy ratio from the previous iteration as the peak value. Once the peak value of the energy ratio is attained, output the corresponding k value. Others, return to Step 2.

Step4. Use the output value at this time for VMD decomposition of the original signal.

3 Feature Selection

3.1 Feature Dimension Reduction

It is challenging to depict the equipment's fault condition solely based on a single characteristic when it comes to bearings. Instead, the combination of multiple features selected using a feature selection algorithm is very necessary in order to express bearing fault states effectively.

The ReliefF algorithm is used for feature selection in this paper. The ReliefF algorithm determines the sensitivity of features by calculating the relative size of the intra-class spacing of the same type of features and the inter-class spacing of different types of features [36]. The algorithm is as follows:

Step1. Initialise the weight matrix vector $W = 0$;

Step2. Obtain a sample R from the feature vector, and select k nearest neighbour samples H_j and M_j from samples of the same or different categories as R; $j(1, 2, \ldots, k)$;

Step3. Update the weight A_j of each feature in R according to the following formula:

$$W(A_{j+1}) = W(A_j) - \sum_{j=1}^{k} diff(A_j, R, H_j)/k + \sum_{j=1}^{k} diff(A_j, R, M_j)/k \tag{22}$$

$$diff(A_j, R, H_j) = \frac{|value(A_j, R) - value(A_j, H_j)}{\max(A_j) - \min(A_j)} \tag{23}$$

$$diff(A_j, R, M_j) = \frac{|value(A_j, R) - value(A_j, M_j)}{\max(A_j) - \min(A_j)} \tag{24}$$

In the above formula, $value(A_j, B)$ is the value of the j eigenvalue A_j of sample B;

Step4. Cycle step 1–3 m times to calculate the weight of each sample iteratively.

The overall structure of the feature selection algorithm based on AVMD and ReliefF is shown below (Fig. 2):

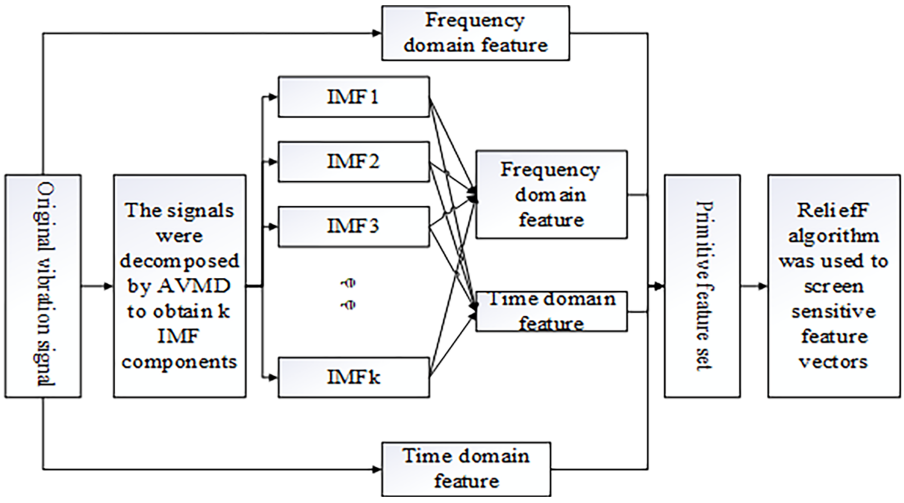

Figure 2: Structure diagram of the feature selection algorithm

3.2 Feature Fusion

For the extracted features, it is necessary to reduce the feature redundancy. Since most rotating machinery is a nonlinear system, it is necessary to adopt a nonlinear feature fusion method for feature fusion. In this paper, an automatic encoder is used to carry out feature fusion [37].

Before using the autoencoder, all feature data needs to be normalised, as follows:

$$z = \frac{x - u}{s} \tag{25}$$

The equation above defines u as the sample mean and s as the sample standard deviation.

The autoencoder consists of an encoder and a decoder. This paper uses a trained autoencoder and retrograde feature fusion. The recursive structure incorporates the residual by comparing the input and output of the feedforward network.

The output layer and input layer are related in the form of $\hat{x}_i \approx, x_i$ in this study. To quantify the reconstruction loss, the mean square error is employed, with the objective function aiming to minimise this loss. The following equation presents the objective function for minimising the reconstruction loss:

$$J(\hat{x}, x) = \frac{1}{m}\sum_{i=1}^{m}(\hat{x} - x_i)^2 \tag{26}$$

In the formula mentioned above, the symbol m represents the count of neurons in the hidden layer.

The dimensionality of high-dimensional features can thus be reduced by back-propagation and gradient descent using trained autoencoders.

4 Fault Diagnosis Model Based on HPO-DBN

A Deep Belief Network (DBN) consists of multiple Restricted Boltzmann Machines (RBMs) and a top-level classifier. The DBN architecture comprises multiple hidden layers. The input to the network is the sample data that satisfies the given criteria, and the corresponding label for the input data is placed at the top. Initially, the input samples are sequentially allocated to the visible neurons in the first layer. Through deep learning, the weights between layers and the biases of each neuron gradually stabilise, enabling the input data samples to be maximally fitted [38].

Upon completion of training, the network undergoes testing by running test data. The test data is automatically analysed, and classification into the most suitable category is performed. The training process of a DBN consists of two steps: unsupervised training utilising RBM, followed by supervised fine-tuning [39].

4.1 Unsupervised Training and Supervised Fine-Tuning Based on RBM

The most important part of this step is to train the RBM, which is done in each step. To initiate the training process, the training data is input to the neurons in the visible layer. The data is then mapped to the hidden layer using a greedy learning approach. Subsequently, the hidden layer data is transmitted back to the visible layer through the CD-k algorithm (where "k" represents the number of samples). Finally, the reconstruction error is computed to facilitate adjustments to the internal weight and bias [40].

Forward unsupervised training based on RBM can only approximate the internal parameter vector of each RBM; it cannot find the optimal solution of the entire model. Therefore, supervised fine-tuning is needed to find the optimal solution from a large set of solutions.

Supervised fine-tuning in the backward direction falls under the umbrella of supervised learning. Typically, the classifier for the final output layer employs the BP algorithm. Upon defining the network architecture, it becomes crucial to determine the appropriate number of nodes in the hidden layer, as well as the number of neurons and the learning rate. The DBN model, which meets the desired criteria, can be obtained through a specific number of iterations. Refer to the accompanying figure for the illustration of the DBN network structure (Fig. 3):

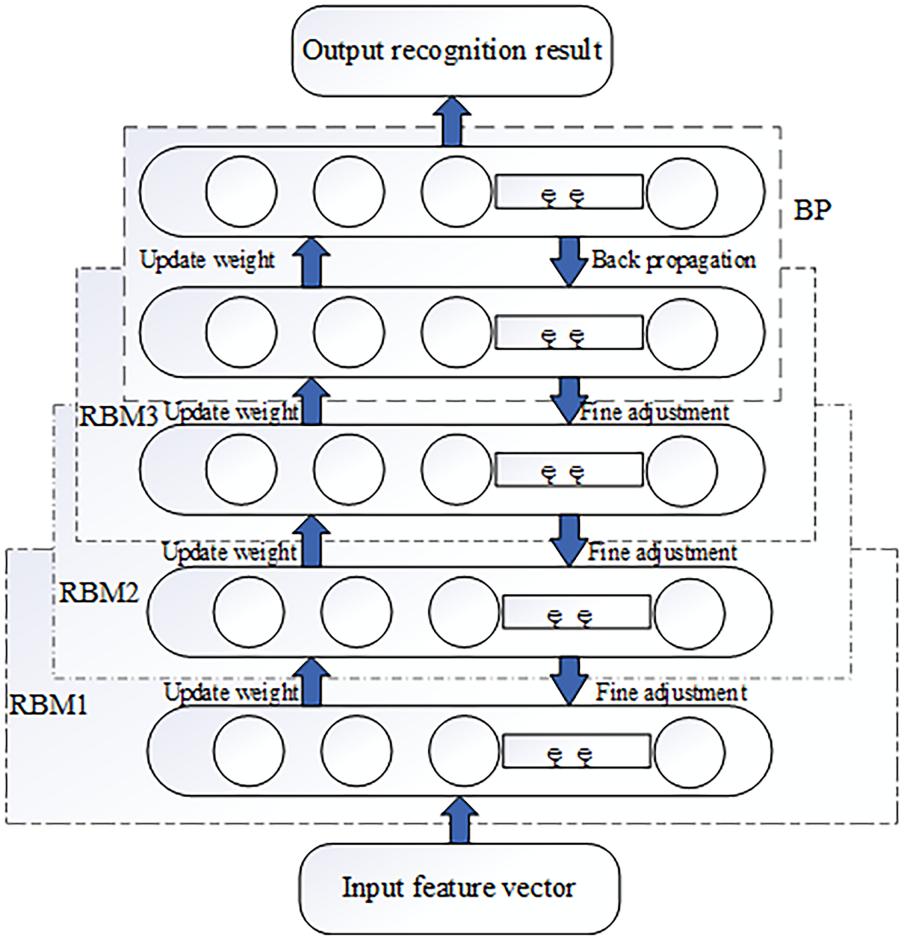

Figure 3: DBN network structure diagram

4.2 Parameter Selection and Algorithm

When using DBN, it is necessary to set the related parameters. The DBN structure design involves two key aspects: the quantity of network layers and the size of hidden layers in terms of node count. Expanding the number of hidden layers can narrow the gap between nodes across layers, thereby enhancing classification accuracy. However, it may also lead to prolonged training time. Additionally, the inappropriate configuration of relevant parameters can significantly impact the classification

outcome [41]. The network structure utilised in this study comprises four layers, including an input layer, an output layer, and two hidden layers.

Optimising the number of nodes in the hidden layer is crucial for achieving desirable training outcomes. Currently, there is no universally applicable approach for determining the optimal number of nodes in this layer. In this study, we propose a methodology utilising the Hunter-Prey Optimiser to optimise the selection of hidden layer nodes.

The algorithm employed in this study is the Hunter-Prey Optimiser (HPO), which draws inspiration from the behavioural patterns observed in predatory animals, such as lions, leopards, and wolves, as well as their prey, including deer and antelope [42]. The core principle of this population-based optimisation algorithm can be summarised as follows:

(1) Initialize the population

To begin with, the initial population randomness is defined as $\vec{x} = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$, and the objective function is evaluated as $\vec{O} = \{O_1, O_2, \ldots, O_n\}$ for all individuals within the population. The subsequent procedures involve the utilisation of prescribed rules and strategies to regulate and direct the population's exploration in the search space, thereby facilitating an iterative search process. Within each iteration, the algorithm rules dictate the update of each member's position within the population, which is then assessed by the objective function. The population initialisation follows the subsequent rules:

$$\begin{cases} x_i = rand(1, d). * (u - l) + l \\ u = [u_1, u_2, \ldots, u_d] \\ l = [l_1, l_2, \ldots, l_n] \end{cases} \tag{27}$$

The formula mentioned above involves several variables, each serving a distinct role in the computation. Specifically, x_i represents the hunting position, l corresponds to the minimum search range, u denotes the maximum search range, d represents the number of dimensions in the problem variable, and *rand* is a random number utilised within the algorithm.

(2) Predator search

Once the initial population has been generated and individual members identified, the objective function is used to calculate the fitness value associated with each solution. This fitness function serves as a determinant of solution quality, distinguishing between favourable and unfavourable outcomes. Based on these results, subsequent search procedures are repeated iteratively to identify the optimal position. The search mechanism is as follows:

$$x_{i,j}(t+1) = x_{i,j}(t) + 0.5[(2CZP_{pos(j)} - x_{i,j}(t)) + (2(1-C)Z\gamma(j) - x_{i,j}(t))] \tag{28}$$

The formula provided above encompasses various variables, each playing a distinct role in the computation process. Specifically, $x_{i,j}(t)$ denotes the current position of the predator, $x_{i,j}(t+1)$ represents the subsequent position of the predator, P_{pos} signifies the position of the prey, γ denotes the average of all positions, and Z pertains to the adaptive parameter. Then Z,P is as follows:

$$P = R_1 < C; IDX = (P == 0) \tag{29}$$

$$Z = R_2 \otimes IDX + R_3 \otimes (\sim IDX) \tag{30}$$

Within the mentioned formula, the random vectors R_1 and R_3 are defined within the interval $[0, 1]$. The variable P represents the index value determined by comparing R_1 and C. Additionally, the random number R_2 is constrained to the range $[0, 1]$. The index value IDX corresponds to vector

R_1, which satisfies the condition ($P = 0$). Moreover, parameter C serves as a balancing factor that governs the trade-off between exploration and development. During the iterative process, the value of f decreases to 0.02. The subsequent calculation for C is as follows:

$$C = 1 - i\left(\frac{0.98}{M}\right) \tag{31}$$

Within the aforementioned formula, the variable i represents the current iteration number, while M signifies the maximum iteration number. γ in the above formula is calculated as follows:

$$\gamma = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{32}$$

The following equation describes the Euclidean distance:

$$D_i = \left(\sum_{j=1}^{d}(x_{i,j} - \gamma_j)^2\right)^{\frac{1}{2}} \tag{33}$$

In the hunting scenario, when the hunter successfully captures its prey, the captured prey will perish, necessitating the hunter to relocate to a new prey location for the subsequent hunt. To address this challenge, the algorithm incorporates a decrement mechanism outlined below:

$$k = round(C \times N) \tag{34}$$

Within the formula, the variable N denotes the number of individuals in the search population. At the algorithm's onset, the value of k is initialised to be equal to N. In order to select and capture the prey, the search individual farthest from the average position γ is chosen. The global best position, which offers the prey a higher chance of survival, corresponds to the safest location. Consequently, the prey's position is updated according to the following procedure:

$$x_{i,j}(t+1) = T_{pos(j)} + CZ\cos(2\pi R_4) \times [T_{pos(j)} - x_{i,j}(t)] \tag{35}$$

Within the provided formula, the variable $x_{i,j}(t)$ represents the current position of the prey. Following that, $x_{i,j}(t+1)$ denotes the position of the prey in the next iteration, and $T_{pos(j)}$ signifies the global optimal position. Additionally, Z stands for the adaptive parameter, whereas R_4 is the random number belonging to the interval [0, 1]. By utilising the cosine function and its input parameters, the position of the subsequent prey can achieve global optimality across various radii and angles. The selection process between predator and prey operates according to the following criteria:

$$x_i(t+1) = \begin{cases} x_i(t) + 0.5\left[\left(2CZP_{\_pos(j)} - x_i(t)\right) + (2(1-C)Z\gamma(j) - x_i(t))\right] & \text{if } R_5 < \beta \\ x_i(t+1) = T_{pos} + CZ\cos\left(2\pi R_4\right) \times \left(T_{po} - x_i(t)\right) & \text{else} \end{cases} \tag{36}$$

Within the formula provided above, we assign the variable R_5 to a randomly generated number falling within the interval [0, 1]. Meanwhile, β represents a regulating parameter, which we set to 0.1 for the purposes of this experiment. If the value of R_5 is smaller than β, we consider the search population to portray a hunter, thus necessitating an update in the search's next position using the aforementioned formula.

Below outlines the sequential procedures involved in the intelligent optimisation algorithm employing a hunter-prey approach:

Step 1: Initialise population initialisation, set random variables.

Step 2: Start the predator search and set the search mechanism.

Step 3: Set the search mechanism parameters according to the search mechanism.

Step 4: The fitness value is computed, and the optimal position of each iteration is recorded.

Step 5: The balance parameter C, which determines the trade-off between exploration and exploitation, is updated using a specific formula. Additionally, the number of populations N is searched, along with the adaptive parameter Z.

Step 6: Update the position according to the value of R_4.

Step 7: Calculate the fitness value.

Step 8: Checking if the stop condition has been satisfied. If the condition is met, the optimal solution is output. Otherwise, the process proceeds to Step 5. In combination with Genetic Algorithm (GA) and Particle Swarm Optimisation (PSO), HPO offers a better balance between exploration and exploitation, leveraging its predator–prey mechanism. While GA depends primarily on crossover and on mutation, PSO depends on velocity updates toward the global best and toward the personal best. HPO tends to maintain diversity in populations longer while minimising premature convergence, which makes it more effective in optimising the parameters of deep neural networks for a fault detection task.

Step 9: End the algorithm and output the optimal weight.

According to the above algorithm steps, the algorithm flow is shown as follows (Fig. 4).

4.3 Fault Diagnosis Model

To determine the optimal number of nodes in the hidden layer, HPO is employed. This technique enables the identification of the most suitable number of nodes in each RBM layer. Consequently, the intricate features encapsulated within DBN can be effectively mapped to distinct feature spaces. By reducing the feature vectors without compromising recognition accuracy, the objective is to achieve the highest possible recognition accuracy.

The HPO-DBN structure diagram is as follows (Fig. 5).

Suppose the input $x = \{x_1, x_2, \ldots, x_m\}$ is an m-dimensional training sample, the first hidden layer is $h_1 = \{h_{11}, h_{12}, \ldots, h_{1a}\}$, and the second hidden layer is $h_2 = \{h_{21}, h_{22}, \ldots, h_{2b}\}$.

The optimisation of the number of hidden-layer nodes in the DBN network involves the utilisation of the HPO algorithm, which consists of three steps. To evaluate the distinguished characteristics of the HPO algorithm, it was compared to several other commonly used optimisation techniques, such as grid search, random search, PSO, and genetic algorithms. The comparison focused on two aspects, convergence speed and output quality. HPO outperformed the evaluated techniques in both speed and quality of results, most especially with higher population sizes. The main areas for studying hyperparameters were: Hidden nodes: [20, 200], Learning rates: [1e−3, 1e−1]. Also, we evaluated sensitivity to population size and found that a population size equal to 30 gave the best trade-off between computational cost and model quality. Convergence curves also bore out that HPO converged significantly faster than the other optimisation techniques.

Step 1: Initialisation of the HPO algorithm and the DBN network parameters. The HPO algorithm is applied to assign initial values to the node count in the hidden layer.

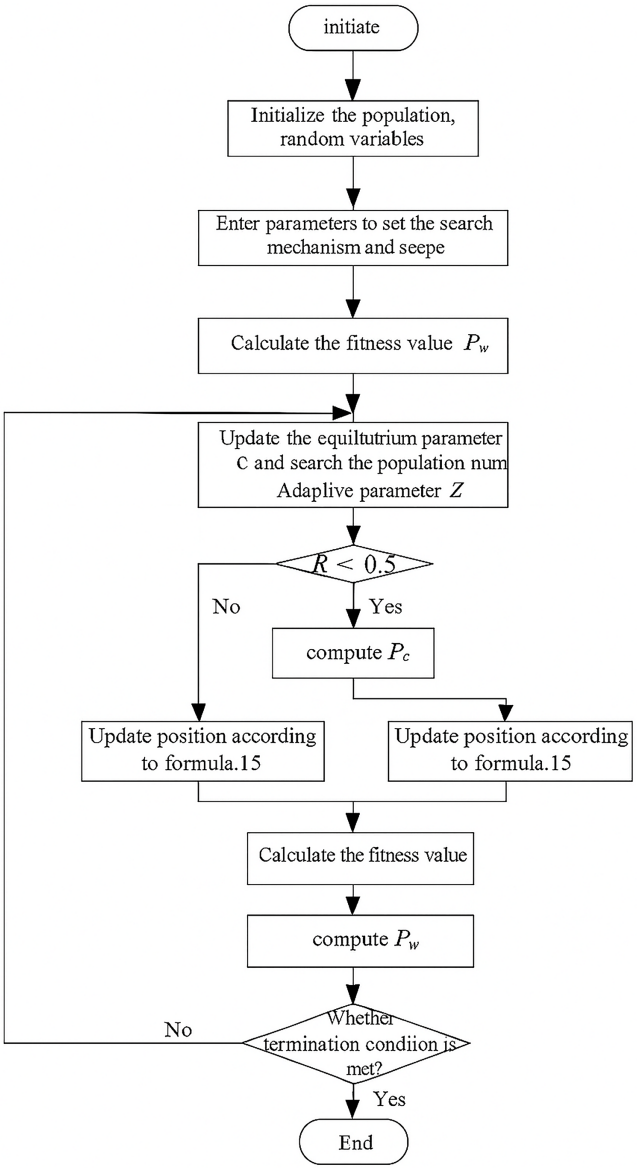

Figure 4: Flowchart of the hunter-prey intelligent optimisation algorithm

Step 2: Iteratively updating the node count in the hidden layer based on predefined rules within the HPO algorithm. The termination criteria for the algorithm include reaching an error rate of 0 or reaching the maximum number of iterations. These steps are repeated until the optimal number of hidden layer nodes is determined. In contrast, the optimised DBN Parameters using Genetic Algorithms exhibited slower convergence speed, lower accuracy gains, and were not as effective. Under the same conditions, the predator–prey search strategy of HPO retained population diversity and avoided premature convergence, allowing it to reach optimal solutions more efficiently than GA.

Step 3: Utilising the obtained number of hidden layer nodes as input for the DBN network. The effectiveness and accuracy of the network are validated and measured using the test set. The

hierarchical structure being learned with the Deep Belief Network can also be understood using the principle of fault diagnosis. Each lower layer operates like a detector for basic patterns exhibited in the signals, such as changes in amplitude or peaks of frequency, and more abstract and complex combinations are depicted in the deeper layers that relate to certain fault categories. This represents a more interpretable hierarchy of features that relate directly back to the bearing condition states, which were learned. When training a Deep Belief Network (DBN), the energy function is minimised using a technique called contrastive divergence, which reduces the error in observing reconstructed data from a model. Minimising the energy function reduces the reconstruction error, leading the network to learn useful hidden representations for classification. The Hunter-Prey Optimisation (HPO) algorithm has some unique characteristics that can enhance tuning of deep models. HPO utilises a predator-prey update strategy, in combination with an adaptable balance factor f (Eq. (31)) to accommodate exploration and exploitation, which makes it different from conventional optimisers. The dual-agent framework also prevents premature convergence, while the population size maintains diversity. These reasons partially underlie why HPO is readily suited to execute optimisation of deep neural networks to perform fault detection, where there is a need for global search and convergence stability for high-dimensional, nonlinear phenomena.

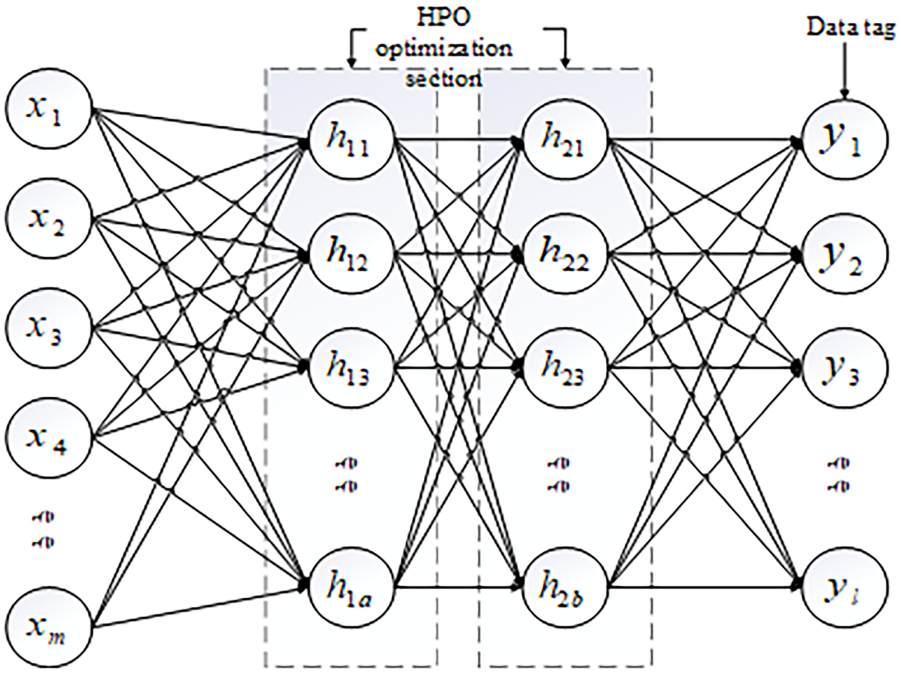

Figure 5: HPO-DBN diagnostic model

5 Simulation Verification

To assess the proposed methods presented in this research paper, the following protocol is implemented. Firstly, the validity and effectiveness of the HPO-DBN diagnostic model, based on AVMD, are verified using the standard dataset provided by Case Western Reserve University (CWRU). Subsequently, the model is employed to diagnose faults in motor vibration data obtained from the National Pipeline Company of China.

5.1 Data Set

The experimental platform diagram, depicted below, Fig. 6, illustrates the source of the vibration data, which is derived from the publicly available dataset provided by Case Western Reserve University.

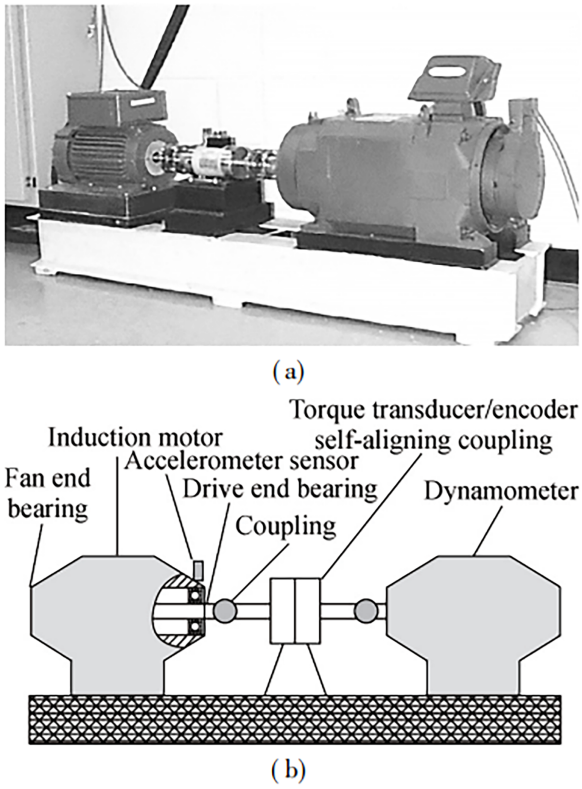

Figure 6: Schematic diagram of the experimental platform. (**a**) Test bench (**b**) schematic diagram

The experimental setup comprises essential components, including a 2 HP motor, a torque sensor/autoencoder, and a power testing unit. Vibration data was acquired at a sampling frequency of 48 kHz, capturing various conditions such as normal state, inner ring fault, outer ring fault, and rolling element fault. To support vibration data acquisition, it is important that both the mounting of the sensor and the filtering of the signals are done correctly. In this study, accelerometers were rigidly mounted to the bearing housing to ensure that the acceleration signal was persistent. Anti-aliasing filters were configured for frequencies greater than half the sampling rate of 48 kHz prior to sampling in order to minimise high-frequency noise. Improper mounting or filtering could create additional artefacts that could degrade the quality of AVMD decomposition and reduce the accuracy of later fault diagnosis.

To validate the efficacy of the approach, a total of 2048 data points were grouped, with each state containing 60 sample groups. Among these groups, 40 were allocated as the training set, while the remaining 20 groups were designated as the test set. To investigate the necessity of each core module, we performed ablation studies removing AVMD, ReliefF, and the autoencoder in separate experiments. The results were that performance dropped significantly in each case where any core module was removed, demonstrating that they were all important to the effectiveness of the model overall.

5.2 Signal Transformation and Feature Extraction

After transforming the signal into the time-frequency domain, features are extracted from three distinct domains: the time domain, the frequency domain, and the domain of time-frequency domain. The contact of different types of bearings, including ball bearings and roller bearings, has different modes of vibration as a result of their contact mechanics. Rolling element type has a role in determining the level of amplitude distribution of vibration signals and how they are modulated. By using an adaptive mode of decomposition, AVMD effectively captures the unique design characteristics of various bearing types, allowing for accurate feature extraction. The time-domain waveform, frequency spectrum, envelope spectrum, and intrinsic mode function (IMF) components obtained after AVMD transformation (with the K value selected by AVMD) are illustrated in Figs. 7 to 12.

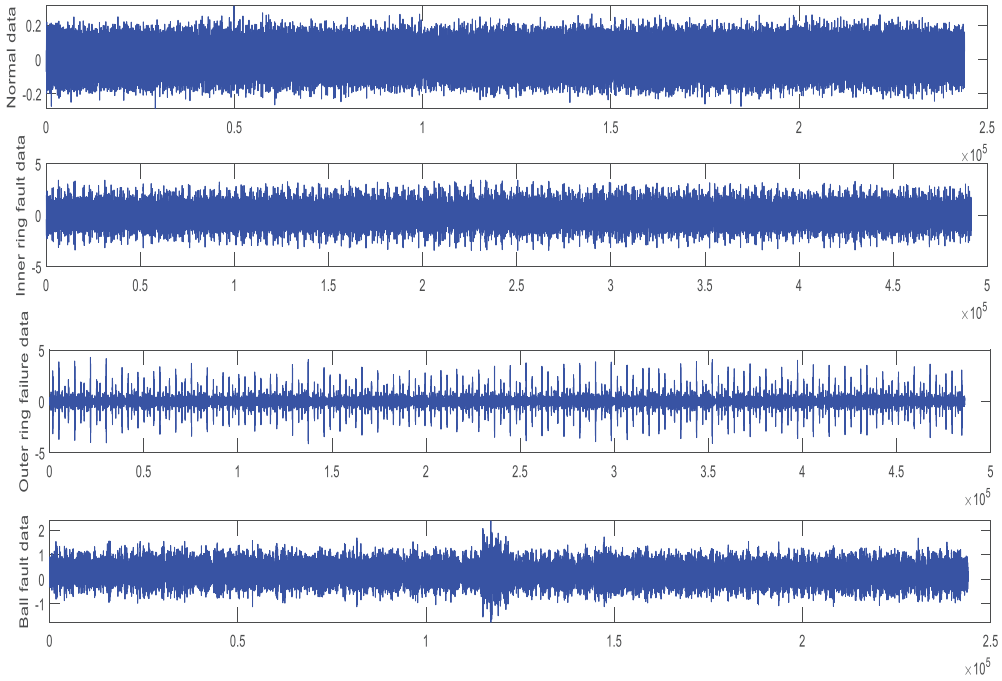

Figure 7: Diagram of vibration signal in the time domain

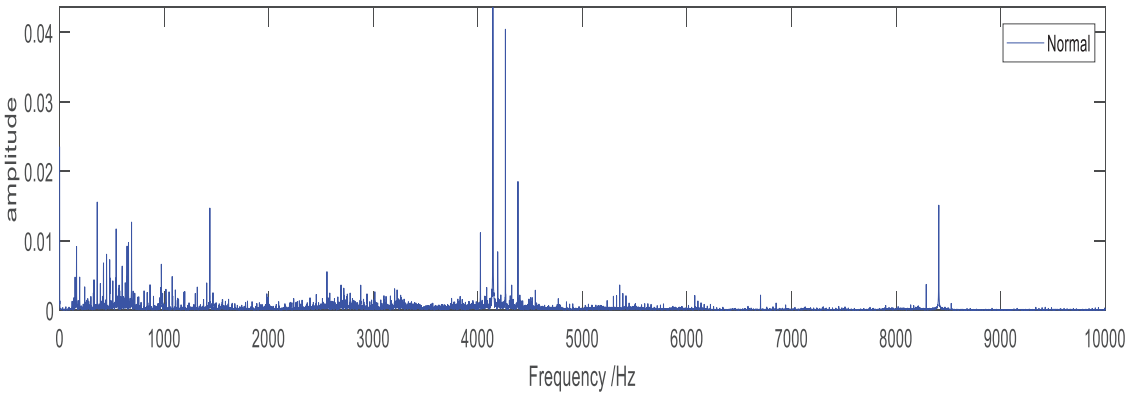

Figure 8: (Continued)

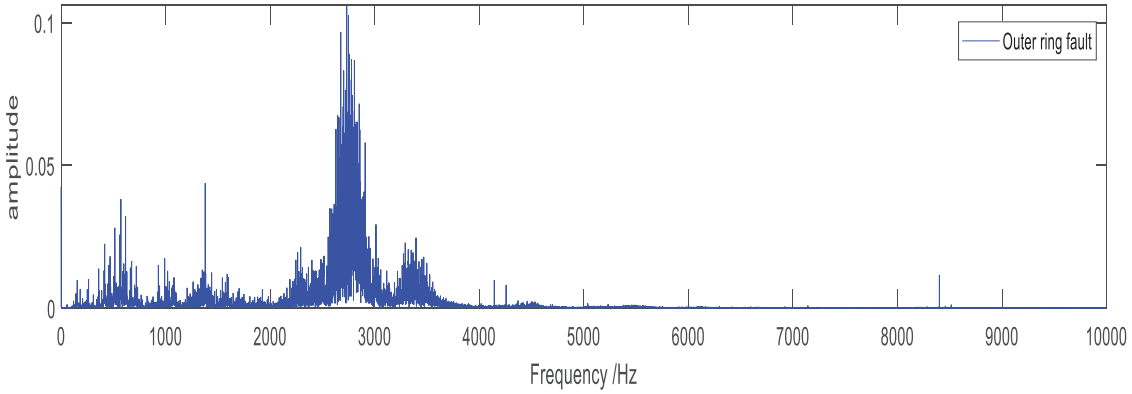

Figure 8: Diagram showcasing the spectrum of a vibration signal

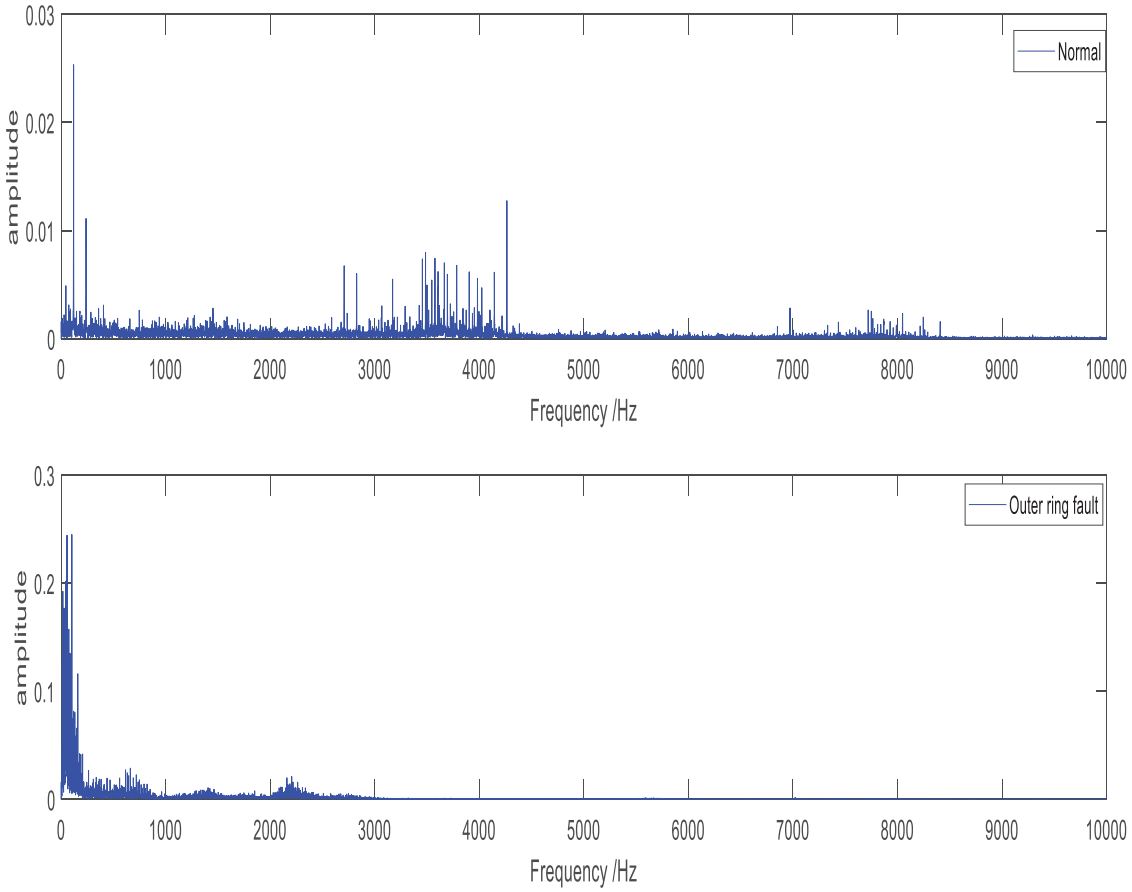

Figure 9: The vibration signal's envelope spectrum

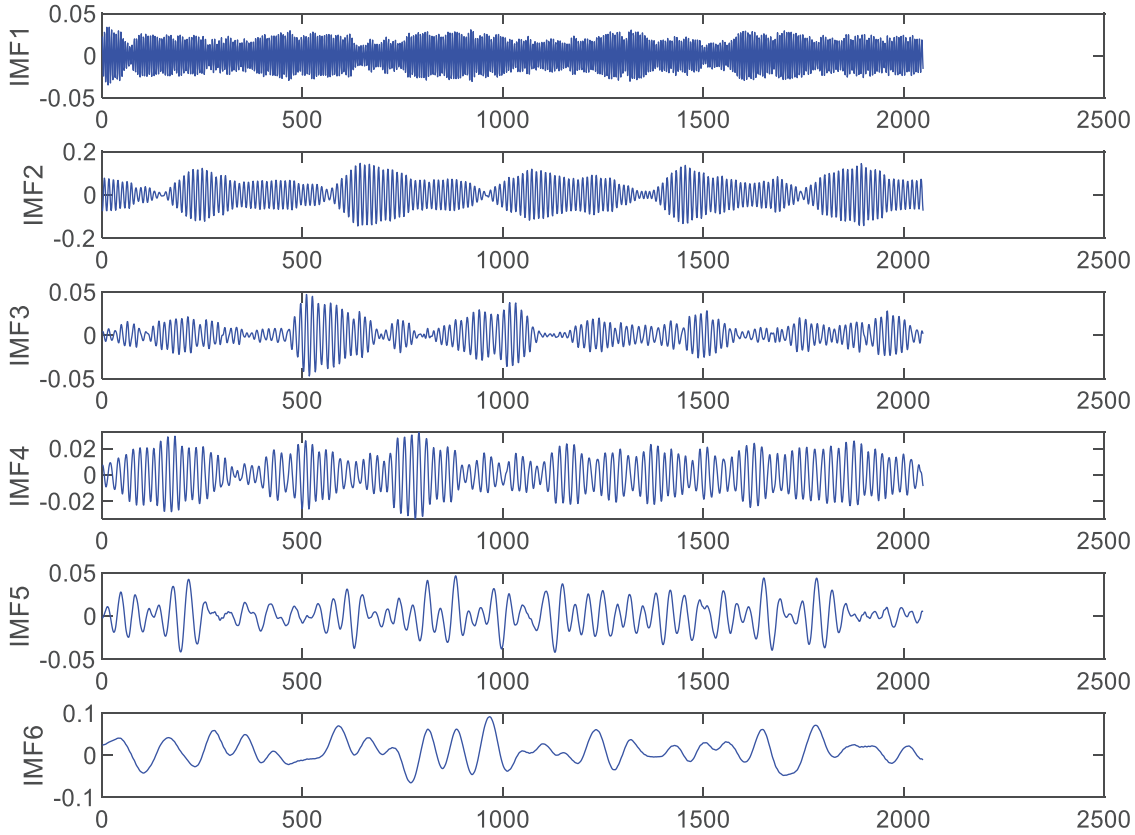

Figure 10: IMF after AVMD decomposition

The statistical features in the time domain are followed by statistical features and four frequency-related features in the frequency domain. Additionally, the time-frequency domain encompasses statistical features, entropy features, and a fractal box dimension. These time-frequency domain features are obtained from IMF signals decomposed by the AVMD method. AVMD was selected for use because it does not involve the manual selection of a number of modes, as is required in standard VMD. The adaptive aspect also supports more reliable decompositions with reduced operator bias. Empirical studies demonstrated that AVMD produced slightly better accuracy than fixed-K VMD, affirming its practicality. AVMD avoids mode mixing and instability, in particular when compared to Empirical Mode Decomposition (EMD). It adaptively selects the number of modes and is more resistant to noise. Moreover, it can achieve consistent signal separation under different conditions. AVMD adaptively breaks down the vibration signal into intrinsic mode functions that emphasise fault-related patterns. These IMFs clarify features, leading to more accurate rolling bearing fault diagnosis. The dimensions of each feature are presented in Table 1 below.

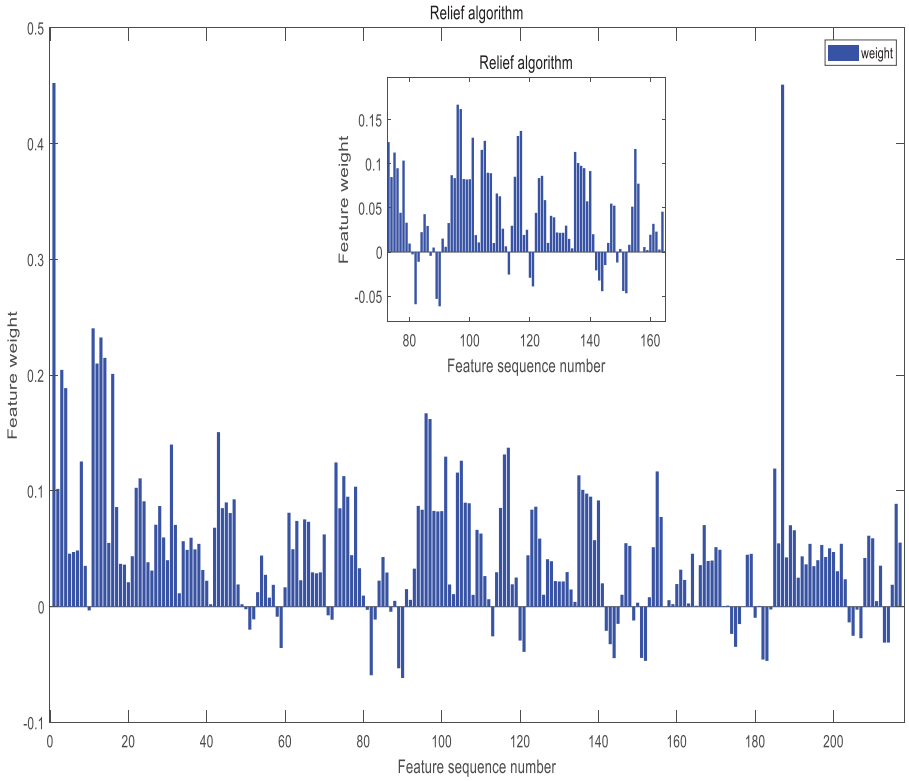

Figure 11: Weight diagram of the ReliefF algorithm

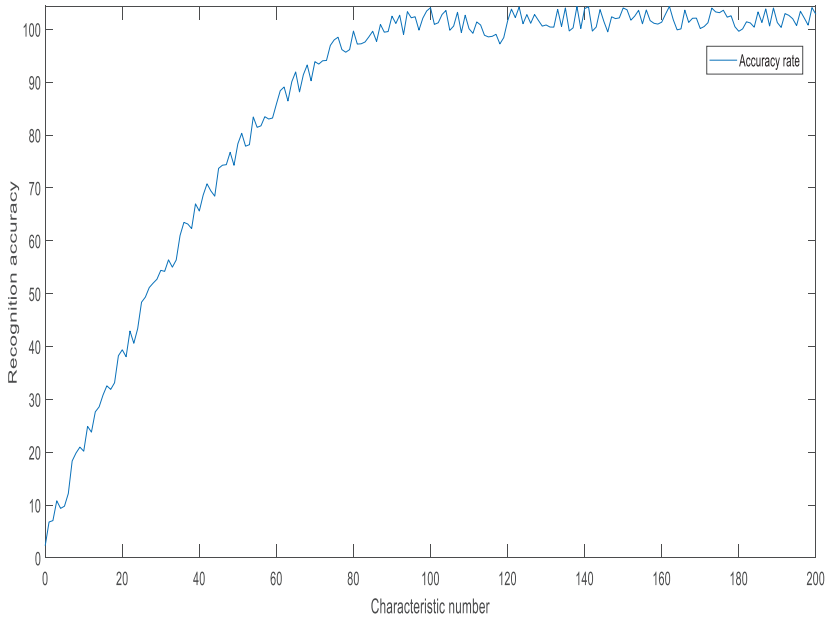

Figure 12: The correlation between the number of features and the accuracy of recognition

Table 1: Feature selection

Time domain feature	Frequency domain feature	Time-frequency domain features
43 statistical characteristics	86 statistical characteristics 4 frequency domain features	86 statistical characteristics 1 entropy feature Dimension of 1 parting box

As can be seen from the above table, each sample needs to extract a feature with dimension 221. Initially, we extracted a wide array of features (221 dimensions) to fulfil the goal of covering all possible operational and fault conditions. We were able to characterise the data in high-dimensional space, but then applied dimensionality reduction (ReliefF and Autoencoder) to maximise relevant features and improve accuracy and efficiency. By starting with a larger conformation of feature space, we ensured that no important features were ignored early on, but still benefited from the reduction process. Although multi-dimensional features can reflect equipment conditions in an all-round way, some features do not change with the occurrence of faults. Therefore, feature selection and dimensionality reduction of these features are needed to ensure the fault diagnosis effectiveness.

5.3 Feature Selection and Dimensionality Reduction

The weight of each feature is calculated using the ReliefF algorithm. The figure below presents the weights of the aforementioned features.

The figure above reveals that certain features possess negative weights, while others demonstrate significantly low weights. ReliefF was used because it measures feature relevance with respect to class distinction. For supervised fault diagnosis, this is somewhat preferable to PCA, which can only measure variance without considering the class. Mutual Information methods can provide supervision, but they have a heavier computational burden and can be sensitive to discretisation; ReliefF provides a good and balanced methodology for a supervised process that is efficient and discriminative. Consequently, the HPO-DBN model is employed to compare recognition rates based on different feature quantities. By selecting varying numbers of features for recognition under identical parameter configurations, the corresponding accuracy rates are depicted in the figure below.

The figure above indicates that the model attains the highest recognition rate when 120 features are selected. Therefore, the ReliefF algorithm can reduce the dimensionality of features to eliminate some unimportant features, improve the recognition accuracy and improve the operation speed greatly. However, it is also necessary to use the autoencoder to fuse the selected features to further reduce the redundancy of the selected features.

The autoencoder adopted in this study includes a hidden layer and employs L1 regularization to prevent overfitting. The specific parameters used for the autoencoder are detailed in Table 2.

Table 2: Autoencoder parameters

Parameter name	value
Hidden layers	1
Nodes in each layer	120-30-120

(Continued)

Table 2 (continued)

Parameter name	value
Regularised term	L1
Cost function	Cross entropy
Batch size	32

The original vibration time domain data undergoes a transformation into the frequency domain. Subsequently, the time domain, frequency domain, and frequency domain characteristics are obtained through AVMD decomposition of the data. The ReliefF algorithm is used to select these features, and a more important feature input encoder is extracted for feature fusion, thus reducing the feature dimension and improving the accuracy and speed of operation. The high-dimensional features fused by the autoencoder are used as input. The autoencoder was used because, unlike PCA (which can only perform linear projection), it allows for nonlinear compression while preserving the relevant structure of the data. Furthermore, while t-SNE is primarily a visualisation method and does not offer a parametric mapping to deal with unseen samples, it prevents it from being practical for integration into classifiers. Thus, the autoencoder provides compact and discriminative features that can conveniently be inserted into DBN training.

5.4 HPO-DBN Classification

To classify the data using DBN, it is crucial to configure the number of nodes in the network's hidden layer and determine the initial parameters. The HPO termination condition is set as error 0 or when the predetermined number of iterations is reached. Details of the specific parameters can be found in Table 3 below.

Table 3: HPO-DBN parameter table

Parameter description	Value
Number of RBM layers	2
RBM iterations per layer	150
Number of neurons comprising the initial hidden layer	100
The second hidden layer's neuron count	50
learning rate	0.1
momentum factor	0.2
activation function	SoftMax
Number of HPO iterations	100

In order to highlight the superior recognition accuracy of the To solve the problem of low recognition rate in bearing fault diagnosis, the HPO-DBN algorithm, we conducted a comparison between the results achieved using SVM and those obtained using the HPO-DBN method. The simulation results are presented in the figure below.

Figs. 13–15 illustrate the diagnostic results for the test set using confusion matrices for a more quantitative measure of performance. The confusion matrices summarise the number of true positives, false positives, true negatives, and false negatives for each of the fault categories, providing a clear insight to the reader as to how effective the model was in differentiating the four faults. Use of these confusion matrices allows for a more intuitive set of results and allows the reader to observe the accuracy of the model in classifying faults and, therefore, inform the user of its merits and opportunities for future work in fault diagnosis. As shown in Figs. 13 and 14, the horizontal coordinate is the sample serial number, a total of 160 samples, and the vertical coordinate is 4 label types, namely normal bearings, bearings with inner ring failure, bearings with rolling element failure, and bearings with outer ring failure. The classification accuracy of the HPO-DBN multi-classification model is 99.8%, indicating that each prediction label corresponds to the test label. The model exhibits high accuracy, as evidenced by the root-mean-square error of the prediction, which is 0.001. The square correlation coefficient is 0.998, indicating that the four maximum segmentation hyperplanes of the HPO-DBN quadripartite classification model have a good linearity.

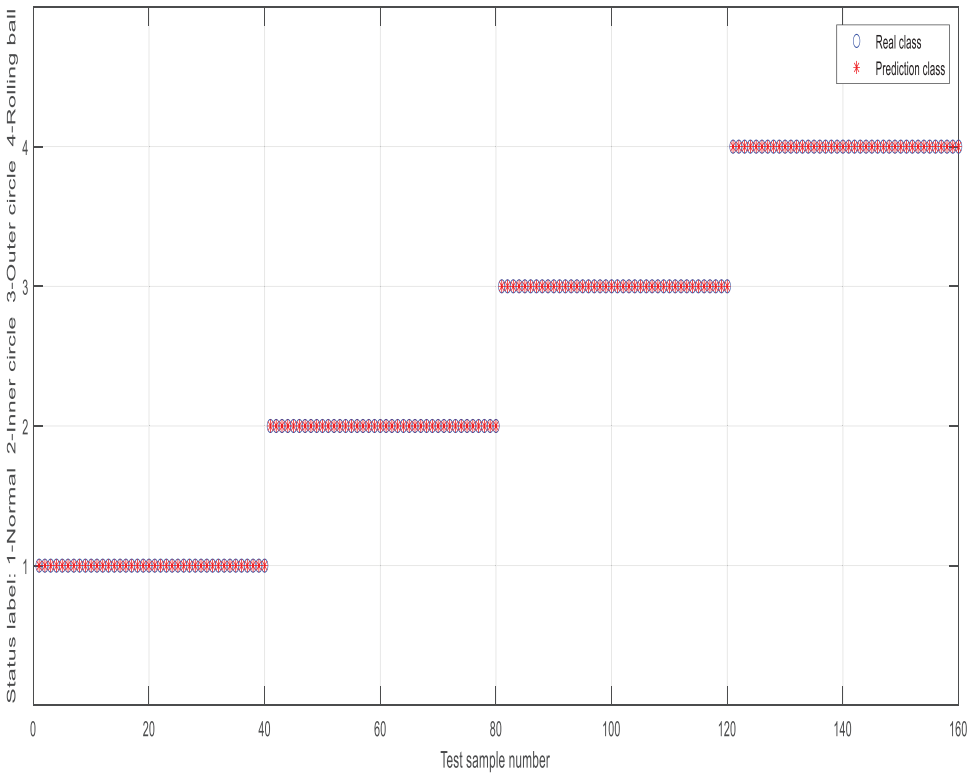

Figure 13: DBN test results

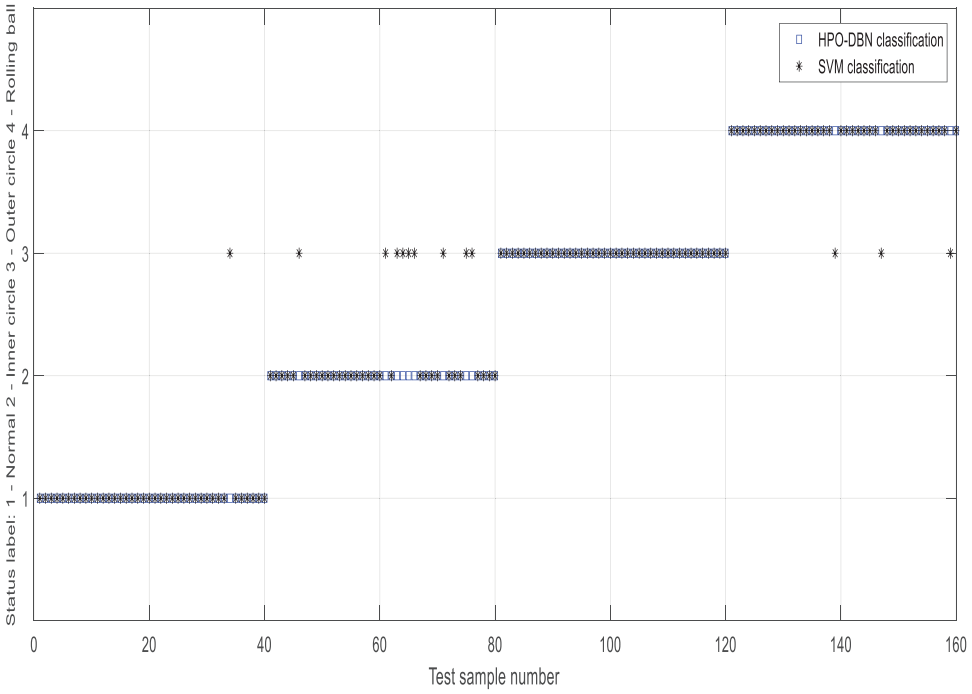

Figure 14: Comparison of HPO-DBN and SVM results

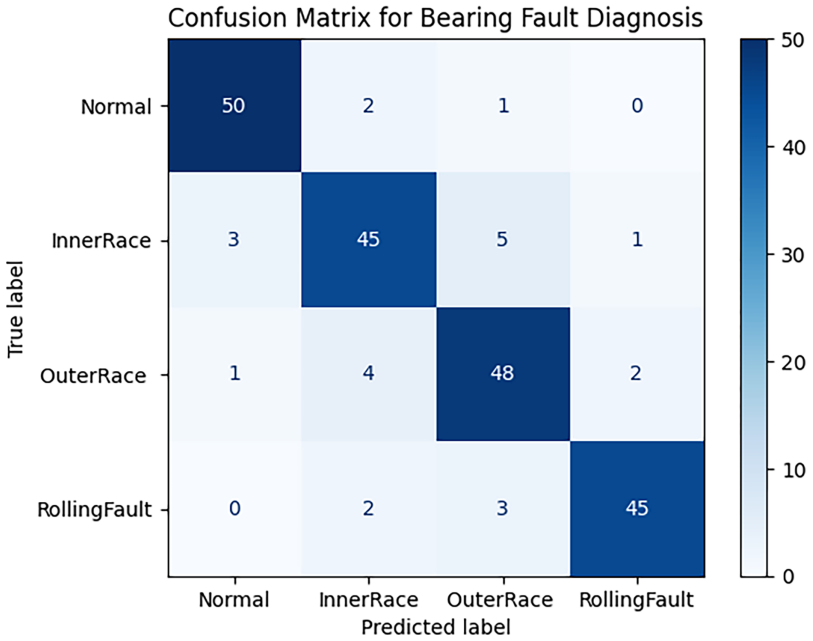

Figure 15: Confusion matrix

Fig. 14 illustrates that the fault diagnosis accuracy achieved with the HPO-DBN method exceeds that of SVM by nearly 10%, with SVM only attaining a 90% accuracy rate. Therefore, the experiment verifies the validity of the HPO-DBN diagnostic model based on AVMD. To enhance the validation

of the HPO-DBN model, experiments were performed that compared HPO-DBN against recently advanced methods and techniques like CNN and models that utilise Transfer Learning, which are typically developed for bearing fault diagnosis using the same dataset. HPO-DBN showed great performance when compared to the more advanced models, again, producing higher accuracy and convergence performance. The HPO-DBN obtained almost 10% higher accuracy on the CWRU dataset than the SVM, 99.8% vs. 90%. The use of ReliefF and feature reduction with an autoencoder resulted in lower computational cost, producing a model that is both more accurate and efficient.

5.5 Validation of Measured Data

According to the vibration data of motor bearing x and y axis of No. 1 Nanchang-Shanghai Branch of West-East Gas transmission in Fuzhou Compressed Gas Station from April 2022 to April 2023, provided by the State Pipeline Network Company, the motor is a synchronous motor produced by Shanghai Electric, the specific model is TZW1000-2, and the measuring point code is VIA181X/Y/Y. Rated power 12,000 KW, rated speed 4800 rpm, rated current 182 A, design efficiency 0.975, rated voltage 6000 V, power factor is 1.

The trained model mentioned above was applied to diagnose faults in the measured data, and the diagnostic results are illustrated in Figs. 16–18.

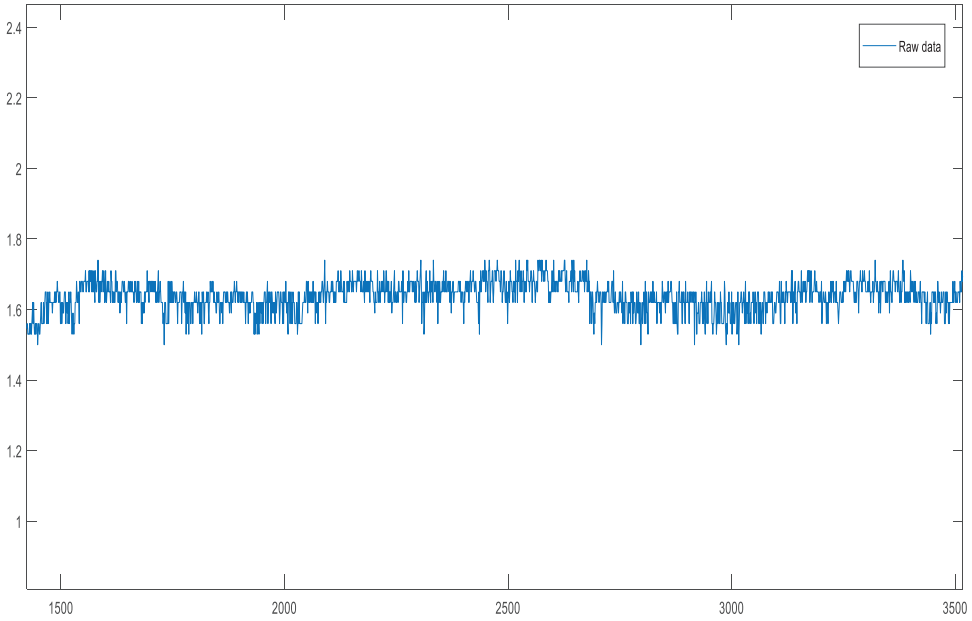

Figure 16: Time domain signal diagram of measured data

Based on the figure presented above, the trained HPO-DBN network demonstrates proficient fault type identification. The accuracy rates for the individual samples are as follows: 97.5% for the first sample, 87.50% for the second sample, 100.00% for the third sample, and 92.50% for the fourth sample. The overall accuracy rate is 95.88%, accompanied by a mean square error (MSE) of 0.0010068. These results lead to the conclusion that this method enables a relatively accurate and efficient diagnosis of the measured data. Despite the high accuracy, some false positives and false negatives were seen and observed, specifically between inner-race and rolling-element faults. In a practical sense, false positives may mean unnecessary inspections, while false negatives could delay required maintenance.

Both, however, were low rates, and overall reliability was adequate to justify a predictive maintenance system. As bearing vibration signals vary according to a variable load and speed, this framework integrates Adaptive Variational Mode Decomposition (AVMD) with HPODBN to achieve diagnostic stability. The AVMD will adapt the number of modes K through energy ratio maximisation so that signals can be adaptively decomposed in the presence of changes in load. After this, the HPODBN will utilise corresponding features generated from the signal decomposition to learn nonlinear mappings, producing a robust classification approach. In a field compressor dataset (April 2022–April 2023) with variation in load and speed, the method maintained an overall 95.88% accuracy, confirming the application can produce reliable results under varying operating conditions.

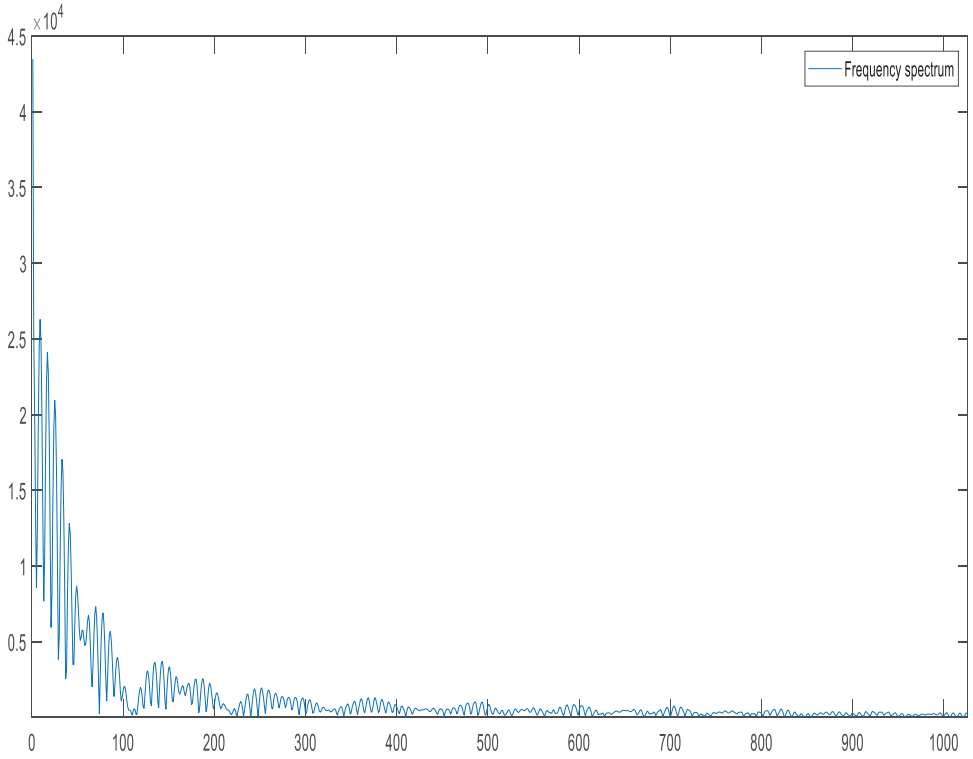

Figure 17: Spectrum diagram of measured data

The fault categories in the field data are inner ring failure, outer ring failure, and rolling element failure. The data set comprises a total of 2048 data points, and there are 60 sample groups for each fault type. There are a total of 40 data groups designated for training and 20 groups designated for testing. Each of the samples is labelled by the researcher based on visual and vibration characteristics of the data. The parameters used for data collection and the distribution of fault types are detailed in Table 4 below.

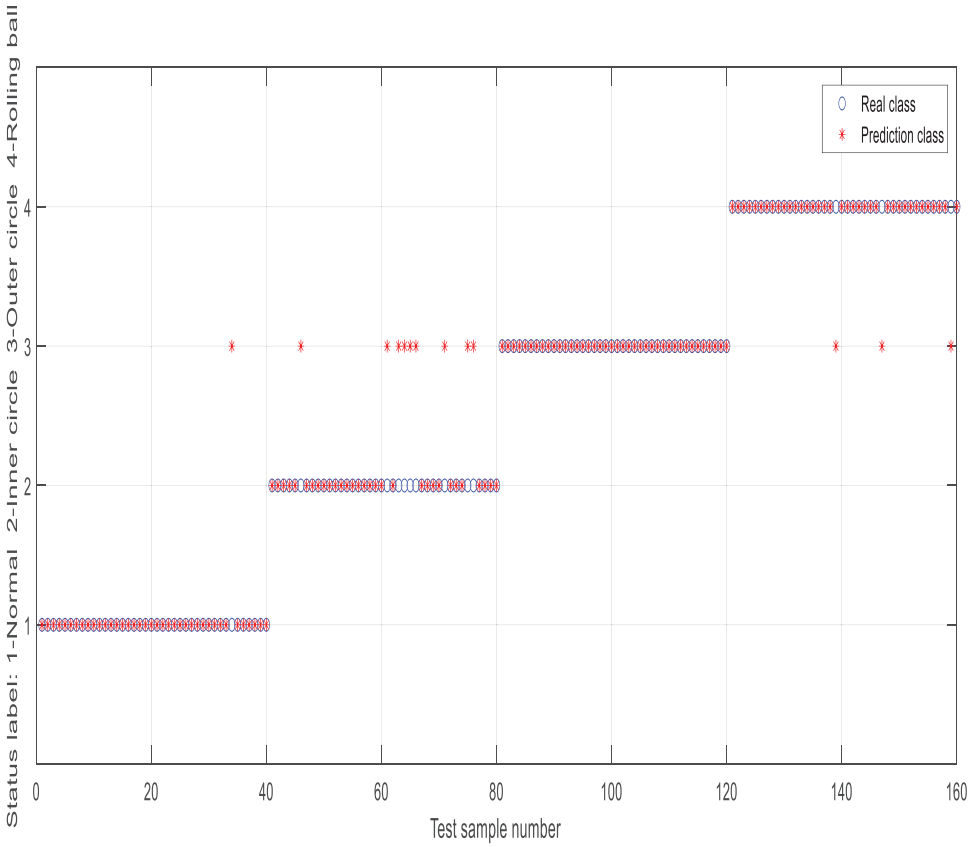

Figure 18: Fault diagnosis results of measured data

Table 4: Data collection parameters and fault distribution

Fault type	Sample size	Labeling method
Normal operation	60	Expert inspection + trend analysis
Inner ring failure	60	Envelope spectrum analysis
Outer ring failure	60	Envelope spectrum analysis
Rolling element	60	Spectrum band analysis

6 Conclusion

This paper presents the AVMD algorithm, which extracts multi-angle features to effectively represent the actual operational condition of bearings. By utilising the proposed algorithm, features in the time-frequency domain can be extracted. Furthermore, feature selection and fusion techniques enhance the distinction between normal and faulty states. Moreover, the application of the HPO algorithm optimises the node count in the hidden layer of the DBN network. This optimisation leads to the creation of the HPO-DBN fault diagnosis model, which guarantees network stability, enhanced speed, and improved accuracy in fault diagnosis. The method suggested is useful for predictive maintenance because it can accurately detect early-stage bearing faults. Predictive maintenance

promotes the confident scheduling of maintenance to avoid severe failures that increase downtime and maintenance costs. Compared to reactive maintenance, which only responds to breakdowns, predictive maintenance, through the use of AVMD and HPO-DBN, provides time-proven reliability and safety while improving long-term operational efficiency of industrial systems.

In follow-up research, the number of network iterations can be further studied, and the method can also be applied to gearbox fault diagnosis and compressor health management.

Acknowledgement: This research was supported by the National Natural Science Foundation of China and the Key R&D projects in Shaanxi Province.

Funding Statement: This research was supported by the National Natural Science Foundation of China (grant numbers 62073259 and 61973094) and the Key R&D Projects in Shaanxi Province (grant number 2023-YBGY-380).

Author Contributions: The authors confirm contribution to the paper as follows: Xinyu Sha is responsible for designing the framework, Fucai Qian is analysing the performance, validating the results, and writing the article. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author upon request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Du X, Jia W, Yu P, Shi Y, Cheng S. A remaining useful life prediction method based on time-frequency images of the mechanical vibration signals. Measurement. 2022;202(2):111782. doi:10.1016/j.measurement.2022.111782.
2. Shen Y, Tang B, Li B, Tan Q, Wu Y. Remaining useful life prediction of rolling bearing based on multi-head attention embedded Bi-LSTM network. Measurement. 2022;202:111803. doi:10.1016/j.measurement.2022.111803.
3. Cheng H, Kong X, Wang Q, Ma H, Yang S, Chen G. Deep transfer learning based on dynamic domain adaptation for remaining useful life prediction under different working conditions. J Intell Manuf. 2023;34(2):587–613. doi:10.1007/s10845-021-01814-y.
4. Xu Q, Fan Z, Jia W, Jiang C. Quantile regression neural network-based fault detection scheme for wind turbines with application to monitoring a bearing. Wind Energy. 2019;22(10):1390–401. doi:10.1002/we.2375.
5. Zhu J, Chen N, Peng W. Estimation of bearing remaining useful life based on multiscale convolutional neural network. IEEE Trans Ind Electron. 2019;66(4):3208–16. doi:10.1109/tie.2018.2844856.
6. Zeng F, Li Y, Jiang Y, Song G. An online transfer learning-based remaining useful life prediction method of ball bearings. Measurement. 2021;176(7):109201. doi:10.1016/j.measurement.2021.109201.
7. Lu W, Mao H, Lin F, Chen Z, Fu H, Xu Y. Recognition of rolling bearing running state based on genetic algorithm and convolutional neural network. Adv Mech Eng. 2022;14(4):168781322210956. doi:10.1177/16878132221095635.
8. Wang T, Liang M, Li J, Cheng W. Rolling element bearing fault diagnosis via fault characteristic order (FCO) analysis. Mech Syst Signal Process. 2014;45(1):139–53. doi:10.1016/j.ymssp.2013.11.011.

9. Li N, Wang H. Variable filtered-waveform variational mode decomposition and its application in rolling bearing fault feature extraction. Entropy. 2025;27(3):277. doi:10.3390/e27030277.
10. Liu S, Fan L. An adaptive prediction approach for rolling bearing remaining useful life based on multistage model with three-source variability. Reliab Eng Syst Saf. 2022;218:108182. doi:10.1016/j.ress.2021.108182.
11. Zosso D, Dragomiretskiy K, Bertozzi AL, Weiss PS. Two-dimensional compact variational mode decomposition. J Math Imag Vis. 2017;58(2):294–320. doi:10.1007/s10851-017-0710-z.
12. Gupta M, Wadhvani R, Rasool A. A real-time adaptive model for bearing fault classification and remaining useful life estimation using deep neural network. Knowl Based Syst. 2023;259:110070. doi:10.1016/j.knosys.2022.110070.
13. Basani DKR, Gudivaka BR, Gudivaka RL, Gudivaka RK. Enhanced fault diagnosis in IoT: uniting data fusion with deep multi-scale fusion neural network. Internet Things. 2024;2022(20):101361. doi:10.1016/j.iot.2024.101361.
14. Yin C, Wang Y, Ma G, Wang Y, Sun Y, He Y. Weak fault feature extraction of rolling bearings based on improved ensemble noise-reconstructed EMD and adaptive threshold denoising. Mech Syst Signal Process. 2022;171(3):108834. doi:10.1016/j.ymssp.2022.108834.
15. Tao J, Liu Y, Yang D, Tang F, Liu C. Fault diagnosis of rolling bearing using deep belief networks. In: Proceedings of the 2015 International Symposium on Material, Energy and Environment Engineering; 2015 Nov 28–29; Changsha, China. Huizen, The Netherlands: Atlantis Press; 2015. p. 566–9. doi:10.2991/ism3e-15.2015.136.
16. Naresh KRP. Applying discrete wavelet transform for ECG signal analysis in IoT health monitoring systems. Int J Inf Technol Comput Eng. 2022;10(4):2347–3657.
17. Padmavathi V, Suresh C, Kumar RL, Punitha P. Adaptive AI-enhanced signal processing framework for optimized communication in the Internet of underwater things (IOUT). Int J Patt Recogn Artif Intell. 2025;39(7):2558002. doi:10.1142/s0218001425580029.
18. Aggarwal K, Khoa BT, Daya Sagar KV, Agrawal R, Dhingra M, Dhingra J, et al. Marketing information system based on unsupervised visual data to manage transportation industry using signal processing. Expert Syst. 2025;42(1):e13384. doi:10.1111/exsy.13384.
19. Wang H, Du W, Li H, Li Z, Hu J. Weak fault feature extraction of rolling element bearing based on variational mode extraction and multi-objective information fusion band-pass filter. J Vibroeng. 2022;24(1):30–45. doi:10.21595/jve.2021.22067.
20. Jagtap HP, Bewoor AK, Kumar R. Failure analysis of induced draft fan used in a thermal power plant using coordinated condition monitoring approach: a case study. Eng Fail Anal. 2020;111(2):104442. doi:10.1016/j.engfailanal.2020.104442.
21. Saraza-Canflanca P, Martin-Martinez J, Castro-Lopez R, Roca E, Rodriguez R, Fernandez FV, et al. Statistical characterization of time-dependent variability defects using the maximum current fluctuation. IEEE Trans Electron Devices. 2021;68(8):4039–44. doi:10.1109/ted.2021.3086448.
22. Suh S, Lukowicz P, Lee YO. Generalized multiscale feature extraction for remaining useful life prediction of bearings with generative adversarial networks. Knowl Based Syst. 2022;237(12):107866. doi:10.1016/j.knosys.2021.107866.
23. ghgjLiu S, Guo Z, Chen L. Effects of radar absorption materials application on the stealth performance of the inlet based on risk theory. Rev Int De Métodos Numéricos Para Cálculo Y Diseño En Ing. 2024;40(3):42.
24. Ang L, Chen X, Zhao Y, Yu J, Zou H. Research on an early fault detection method for a new distribution system based on automatic arc power. Rev Int De Métodos Numéricos Para Cálculo Y Diseño En Ing. 2024;40(2):15.
25. Rathore MS, Harsha SP. An attention-based stacked BiLSTM framework for predicting remaining useful life of rolling bearings. Appl Soft Comput. 2022;131(7):109765. doi:10.1016/j.asoc.2022.109765.

26. Hu T, Guo Y, Gu L, Zhou Y, Zhang Z, Zhou Z. Remaining useful life prediction of bearings under different working conditions using a deep feature disentanglement based transfer learning method. Reliab Eng Syst Saf. 2022;219:108265. doi:10.1016/j.ress.2021.108265.
27. Jiang L, Li X, Wu L, Li Y. Bearing fault diagnosis method based on a multi-head graph attention network. Meas Sci Technol. 2022;33(7):075012. doi:10.1088/1361-6501/ac56f1.
28. Han T, Pang J, Tan ACC. Remaining useful life prediction of bearing based on stacked autoencoder and recurrent neural network. J Manuf Syst. 2021;61(5):576–91. doi:10.1016/j.jmsy.2021.10.011.
29. Su X, Liu H, Tao L, Lu C, Suo M. An end-to-end framework for remaining useful life prediction of rolling bearing based on feature pre-extraction mechanism and deep adaptive transformer model. Comput Ind Eng. 2021;161(1):107531. doi:10.1016/j.cie.2021.107531.
30. Ding N, Li H, Yin Z, Jiang F. A novel method for journal bearing degradation evaluation and remaining useful life prediction under different working conditions. Measurement. 2021;177:109273. doi:10.1016/j.measurement.2021.109273.
31. Deng Y, Yang X, Huang Y, Pan T, Zhu H. Calculation method of intermediate bearing displacement value for multisupported shafting based on neural network. J Ship Res. 2021;65(4):286–92. doi:10.5957/josr.02200007.
32. Li X, Kong X, Zhang J, Hu Z, Shi C. A study on fault diagnosis of bearing pitting under different speed condition based on an improved inception capsule network. Measurement. 2021;181(2):109656. doi:10.1016/j.measurement.2021.109656.
33. Wang Y, Zhao J, Yang C, Xu D, Ge J. Remaining useful life prediction of rolling bearings based on Pearson correlation-KPCA multi-feature fusion. Measurement. 2022;201:111572. doi:10.1016/j.measurement.2022.111572.
34. Zhang T, Liu S, Wei Y, Zhang H. A novel feature adaptive extraction method based on deep learning for bearing fault diagnosis. Measurement. 2021;185:110030. doi:10.1016/j.measurement.2021.110030.
35. Wu Y, Liu L, Qian S. A small sample bearing fault diagnosis method based on variational mode decomposition, autocorrelation function, and convolutional neural network. Int J Adv Manuf Technol. 2023;124(11):3887–98. doi:10.1007/s00170-021-08126-8.
36. Wang S, Wang Q, Xiao Y, Liu W, Shang M. Research on rotor system fault diagnosis method based on vibration signal feature vector transfer learning. Eng Fail Anal. 2022;139:106424. doi:10.1016/j.engfailanal.2022.106424.
37. Zou L, Zhuang KJ, Zhou A, Hu J. Bayesian optimization and channel-fusion-based convolutional autoencoder network for fault diagnosis of rotating machinery. Eng Struct. 2023;280:115708. doi:10.1016/j.engstruct.2023.115708.
38. Zhao H, Yang X, Chen B, Chen H, Deng W. Bearing fault diagnosis using transfer learning and optimized deep belief network. Meas Sci Technol. 2022;33(6):065009. doi:10.1088/1361-6501/ac543a.
39. Niu G, Wang X, Golda M, Mastro S, Zhang B. An optimized adaptive PReLU-DBN for rolling element bearing fault diagnosis. Neurocomputing. 2021;445:26–34. doi:10.1016/j.neucom.2021.02.078.
40. Mushtaq S, Manjurul Islam MM, Sohaib M. Deep learning aided data-driven fault diagnosis of rotatory machine: a comprehensive review. Energies. 2021;14(16):5150. doi:10.3390/en14165150.
41. Wang H, Huang H, Yu S, Gu W. Size and location diagnosis of rolling bearing faults: an approach of kernel principal component analysis and deep belief network. Int J Comput Intell Syst. 2021;14(1):1672. doi:10.2991/ijcis.d.210518.002.
42. Wen L, Wang Y, Li X. A new automatic convolutional neural network based on deep reinforcement learning for fault diagnosis. Front Mech Eng. 2022;17(2):17. doi:10.1007/s11465-022-0673-7.