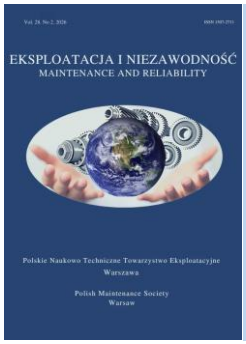


Eksploatacja i Niezawodnosc – Maintenance and Reliability

Volume 28 (2026), Issue 2

journal homepage: http://www.ein.org.pl

Article citation info:
Jiang F, Wang H, Su J, Zhang Z, Experimental Research on Vibration Feature Recognition for Bearing Reliability Based on Improved Convolutional Neural Networks, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2026: 28(2) http:/10.17531/ein/211279

Experimental Research on Vibration Feature Recognition for Bearing Reliability Based on Improved Convolutional Neural Networks

Indexed by: Web of Science Group

Fang Jiang[a], Huiliang Wang[a,*], Jianxin Su[a], Zhijie Zhang[a]

[a] Henan University of Science and Technology, China

Highlights

- A fault recognition method based on an improved convolutional neural network (PCNN) is proposed, and a model combining the zebra optimization algorithm (ZOA), PCNN, and attention mechanism (AT) is established.
- Comparative validation between the ZOA-PCNN-AT model and the traditional CNN model demonstrates that the proposed model can recognize different fault types with fewer iterations, achieves higher recognition accuracy, and enhances the reliability of fault identification.
- The Empirical Mode Decomposition (EMD) algorithm is upgraded to the Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) algorithm, yielding denoised signals with improved signal-to-noise ratios.

Abstract

Among various faults in computer numerical control (CNC) machine tools, motor spindle bearing faults make up the largest proportion. At the heart of the electric spindle, the bearing's performance has a direct impact on the service life, machining accuracy, and efficiency of the entire electric spindle. Once a bearing fails, it usually causes strong vibrations, and the vibration characteristics of damage in different parts are different. In the past network model application, Convolutional Neural Networks (CNN) have better performance in the recognition of vibration characteristics of electric spindle bearings, but due to their structural limitations, the actual recognition is still misjudged, which affects the efficient operation of machine tools and product quality. To enhance the reliability of bearing fault identification, this paper selects motorized spindle bearings as the main research object. First, it analyzes the identification process and issues of traditional CNN regarding vibration characteristics. Subsequently, it innovatively proposes a Parallel Convolutional Neural Network-Attention Mechanism (ZOA-PCNN-AT) model improved by the Zebra Optimization Algorithm (ZOA). This model incorporates an attention mechanism to improve the capacity to capture important features and optimizes the parallel convolutional neural network's structure using the ZOA.

Keywords
vibration signals, convolutional neural networks, zebra optimization algorithm, fault identification

1. Introduction

As the core functional component of high-speed precision machining equipment, the operating condition of electric spindle bearings directly affects machining accuracy and equipment reliability. When the equipment is operating, different vibrations are signals through which the equipment conveys faults to people. Various methods need to be used to capture and analyze the vibration signals. Fault diagnosis based on vibration signals in various fields has also laid the foundation for the fault diagnosis of the electric spindle. Wang et al. [1] sought to address the challenge of identifying the earliest fault features in rolling element bearing vibration signals. They used empirical mode decomposition to decompose the signals with added noise. He et al. [2] merged vibration analysis with deep learning. This structure has proven to be highly effective for signal processing in bearing fault diagnosis. With the increasing demands of intelligent manufacturing for machining efficiency,

(*) Corresponding author.
E-mail addresses: F. Jiang (ORCID: 0009-0002-0620-2204) 1023436837@qq.com, H. Wang (ORCID: 0000-0001-8456-4345) whl-ly@163.com. J. Su (ORCID: 0000-0001-8264-0762) jianxinsu@126.com. Z. Zhang (ORCID:0009-0003-0775-020X) 1184528366@qq.com

electric spindle bearings are subjected to complex working environments characterized by high speed, heavy loads, and variable operating conditions, resulting in vibration signals that exhibit strong non-stationary and multi-modulation characteristics. This poses significant challenges to traditional fault diagnosis methods. Zhang et al. [3] employed the super-resolution sparse representation approach to enhance the Intrinsic Mode Function (IMF) components of vibration signals processed by Ensemble Empirical Mode Decomposition (EEMD), achieving effective denoising and fine-grained processing of abnormal vibration signals. Fang et al. [4] applied the Continuous Wavelet Transform (CWT) and established a vibration model that simplifies the robot into a forced vibration system under harmonic excitation, and effectively identified the fault characteristics of belt looseness. Liu et al. [5] extracted typical frequency features from the wellhead vibration data of normal and faulty electrical submersible pump (ESP) through wavelet analysis, and applied them to efficiently diagnose impeller tip screening faults in oilfields. Vibration signal-based feature recognition technology has become a crucial means for bearing health monitoring due to its high information-carrying capacity. However, existing methods still exhibit notable shortcomings in feature decoupling capability and recognition accuracy under complex operating conditions.

Recently, deep learning techniques have demonstrated prominent advantages in mechanical fault diagnosis. Among them, convolutional neural networks, which leverage their local feature extraction capability, have achieved notable success in vibration signal processing applications. Shi et al. [6] addressed the issue of low accuracy in asynchronous motor fault diagnosis by constructing a stacked fault diagnosis model that integrates stacked autoencoders and CNN. Tang et al. [7] employed a reliable multi-scale quadratic attention embedding CNN to diagnose bearing faults. Guo et al. [8] integrated the convolutional block attention module with AlexNet and synchronized it with the Anomaly Transformer model to enhance the accuracy of fault diagnosis. However, traditional CNN models face three key challenges when dealing with vibration signals from motorized spindle bearings: First, the multi-scale characteristics of raw vibration signals are difficult to capture using single-structured convolutional kernels effectively; second, parameter optimization efficiency is low during the training of deep networks; and third, the lack of an attention mechanism for critical fault features results in insufficient identification capability for weak faults. These issues significantly constrain the practical application value of diagnostic models in industrial scenarios.

To address the aforementioned technical bottlenecks, this study proposes a ZOA-PCNN-AT model. The solution involves: constructing a multi-branch parallel convolutional architecture to achieve multi-scale feature extraction of vibration signals; introducing a channel-spatial dual attention mechanism to enhance weight allocation for critical fault features; and employing a novel zebra optimization algorithm to overcome network parameter optimization challenges. The ZOA has demonstrated significant advantages in optimizing model parameters and enhancing algorithm performance. In the research directions of fault diagnosis and optimization, numerous scholars have also actively explored the potential applications. Jiang [9] and others utilized the ZOA to optimize the core parameters of the support vector machine (SVM), achieving a better fault classification effect. Liang et al. [10] utilized an enhanced ZOA to optimize the deep hybrid kernel extreme learning machine model, demonstrating promising performance in wood processing applications. Liang et al. [11] focused on the issue of trajectory planning for obstacle avoidance in unmanned aerial vehicles (UAVs) and employed the ZOA to enhance the convergence speed. In this study, the ZOA is applied to optimize the parallel convolutional neural network (PCNN) and integrate it with the attention mechanism (AT) to improve the capability of capturing salient features. This study designed a comparative experimental protocol: first, a full life-cycle test platform for motorized spindle bearings was established to acquire vibration data under multiple operating conditions; second, time-frequency analysis was performed to construct the feature dataset; finally, comparative verification of recognition accuracy and generalization capability was conducted under different operating conditions with traditional CNN, ZOA-CNN and other models. The experimental results indicate that the proposed method attains an average recognition accuracy of 99.82%, which is more than 9% points higher than that of traditional methods.

2. Experimental fundamental theory

A. ZOA Zebra Optimization Algorithm

The ZOA was initially introduced by Eva Trajkovski et al. in 2022 as an optimization algorithm inspired by the behavioral patterns of the African zebra. ZOA optimizes parameters by simulating the foraging and defense behaviors of the zebras against predators' attacks, etc. [12], and has advantages such as fast convergence speed. Currently, it has been applied in multiple fields, including power, automotive, and robotics. The calculation process is as follows:

1) Population initialization

In ZOA, a zebra individual serves as the unknown solution, represented by a vector whose elements represent the solution vector components. The plane where the zebra is located represents the problem search space. The location of a zebra individual affects the value of the decision variables. Firstly, a population matrix is established, as shown in Formula 1：

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N\times m} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \end{bmatrix}_{N\times m} \quad (1)$$

where X represent the zebra population; X_i is the $i-th$ zebra; and $x_{i,j}$ denotes the value of the $j-th$ decision variable raised by the $i-th$ zebra, where N is the count of population members (zebras) and m represents the dimension of the decision variable vector. The objective function is assessed according to the magnitudes of the problem parameters corresponding to each zebra, and the results are converted into a vector form denoted as F, as shown in Formula 2:

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N\times 1} \quad (2)$$

where F_i represents the objective function value obtained by the $i-th$ zebra. By comparing it, the optimal candidate solution is identified, thereby updating the positions of the zebras.

2) Foraging behavior

Within the population, there is a type of zebra that feeds on the upper canopy of trees and is called a pioneer zebra (denoted as *PZ*). They provide survival conditions for other low-growing and food-dependent species [13]. Pioneer zebras, as the optimal members, guide other members to their respective positions. As specified in Formulas 3 and 4, zebra positions in the foraging phase are updated through the proposed mechanism.

$$x_{i,j}^{new,P1} = x_{i,j} + r\cdot\left(PZ_j - I\cdot x_{i,j}\right) \quad (3)$$

$$x_i = \begin{cases} x_i^{new,P1}, F_i^{new,P1} < F_i \\ X_i, else \end{cases} \quad (4)$$

where $x_i^{new,P1}$ represents the new state of the $i-th$ zebra, $x_{i,j}^{new,P1}$ is its $j-th$ dimensional value, $F_i^{new,P1}$ is its objective function value, PZ_i is its $j-th$ dimension, r is a random number in the interval [0,1], *I=round(1+rand)*, and $I \in \{1,2\}$.

3) Defensive behavior

Assume that zebras have an equal probability of choosing these two behaviors:

① Large predators attack zebras; zebras choose to flee.

② Small predators attack zebras; zebras choose their attack strategies.

In the mathematical model, the first case can be regarded as S1 and the second case as S2, as shown in Formula 5. During position updating, the update is performed by comparing the objective function values at the zebra's new position, as mathematically expressed by Formula 6:

$$x_{i,j}^{new,P2} = \begin{cases} S_1: x_{i,j} + R\cdot(2r-1)\cdot\left(1-\frac{1}{T}\right)\cdot x_{i,j}, P_s \le 0.5; \\ S_2: x_{i,j} + r\cdot\left(AZ_j - I\cdot x_{i,j}\right), else, \end{cases} \quad (5)$$

$$X_i = \begin{cases} X_i^{new,P2}, F_i^{new,P2} \le F_i; \\ X_i, else, \end{cases} \quad (6)$$

where t stands for iteration contour, T indicates the maximum number of iterations, (R=0.01), P_s is the probability that two strategies occur randomly ($P_s \in [0,1]$). AZ shows the state of the attacked zebra, and AZ_i is its first-dimensional value.

The flowchart is shown in Figure 1.

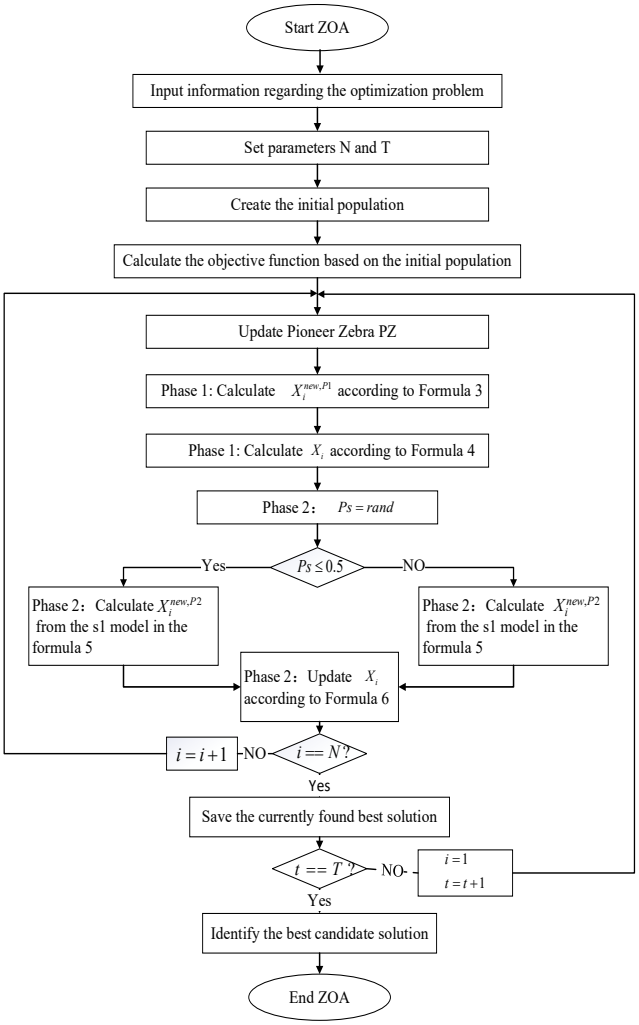


Figure 1. Flowchart of ZOA optimization process.

According to the above process, in CNN, different parameter ranges are set to optimize the hyperparameters in the network model. This enables the adaptive adjustment of the size of each parameter, thereby seeking the optimal solution within the model.

B. Attention Mechanism

Important information is filtered out of the input and given greater weights by the attention mechanism, a computational model that mimics human visual or cognitive processes [14]. Specifically, the attention mechanism weights the input data, reducing attention to irrelevant or noisy information.

Combining CNN with AT can enhance model performance and interpretability. AT mainly consists of transformation layers and attention layers.

1) The transformation layer

It performs a linear transformation to map the input sequences $X \in R^{n_x \times d_x}$ and $Y \in R^{n_y \times d_y}$ to three different vectors Q (query), K (key) and V (value), where d and n represent the dimension and length of the input sequences [15]. The vector calculation is shown in Formula 7:

$$Q = XW^Q, \quad K = YW^K, \quad V = YW^V \tag{7}$$

where $W^K \in R^{d_y \times d_k}$, and $W^V \in R^{d_y \times d_v}$ are linear matrices, d_k represents the dimensions of Q and K, and d_v represents the dimension of V. K and V are projected from Y and Q is projected from X respectively. This way is called the AT.

2) The attention layer

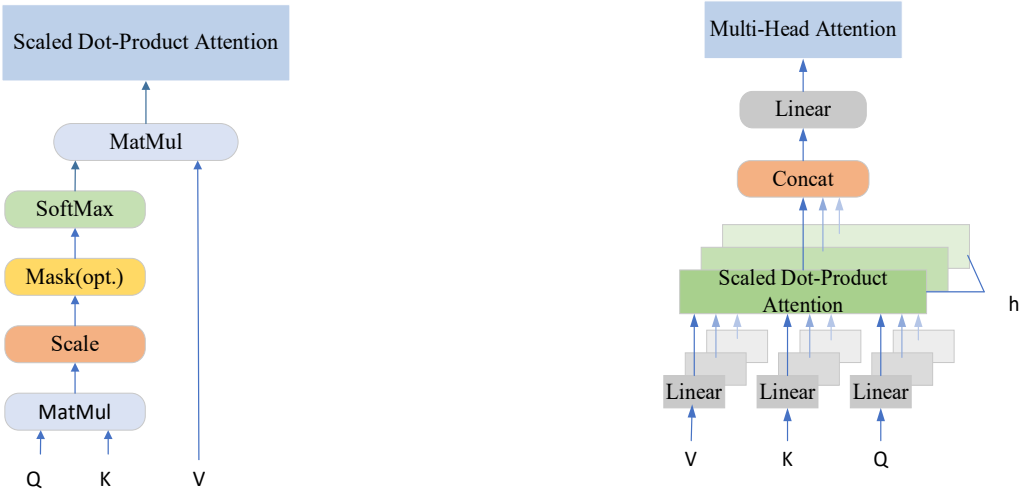
It aggregates the vector Q and the corresponding K, and then aggregates the result with the vector V again to update the vector. Its mathematical expression is:

$$Attention(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{8}$$

The multi-head AT is essentially composed of multiple independent single-head attentions that operate in parallel for

computation [16]. To capture relationships across diverse subspaces and fuse representations from various subspaces.

Attention [17] architecture is shown in Figure 2.



(a) Single-head attention (b) Multi-headed attention

Figure 2. Attention structure diagram.

C. Improve the CNN network

On the basis of the traditional CNN model, an additional channel is added and improved to form the ZOA-PCNN-AT model. Firstly, the wavelet time-frequency map [18] is sent into two channels respectively. Each channel contains two convolution kernels and pooling kernels, with their sizes differing between channels. Two channels are used to learn different weight values, and the learned feature maps of the two channels are concatenated. The fully connected layer is improved to an eight-head attention mechanism to strengthen the key features. The hyperparameters, such as the learning rate and the sizes of the convolution kernels in the two channels are optimized using the ZOA algorithm. Finally, the Softmax function is used to determine different categories [19]. The model structure is shown in Figure 3.

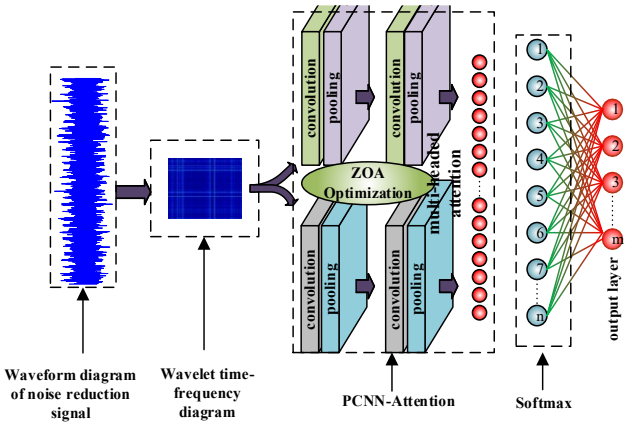


Figure 3. The structure diagram of the ZOA-PCNN-AT model.

3. Vibration identification experiment preparation

A. Experimental Design

In this section, the validity of the model is verified by collecting the vibration signals of the motorized spindle under different fault modes. Firstly, the experimental platform is set up. The experimental equipment used in this chapter consists of the main body of the CNC grinding machine, M+P VibPilot measuring instrument, motorized spindle, acceleration sensor, signal processing software, and computer. The motorized spindle model studied in this experiment is HGE-MD230Z3, and the rolling bearing type used in the front-end bearing group is SKF7305. The M+P VibPilot measuring instrument used in the experiment supports the highest 8-channel signal acquisition with synchronous sampling and 8 analog input channels. The selected sensor model is ZKX-3A-450BM-G05E, with a sensitivity range of and a sampling frequency of 2048Hz. The experimental device is shown in Figure 4.

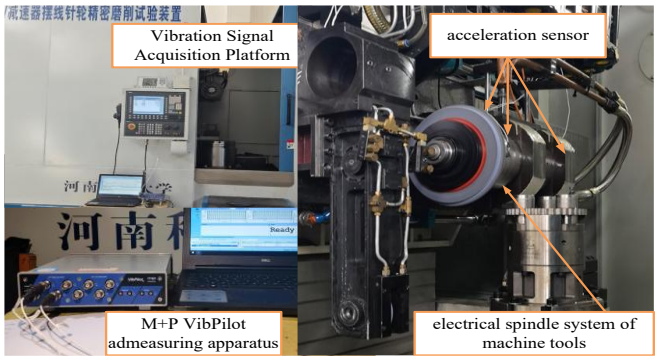


Figure 4. Experimental facility.

Arrange the experimental equipment and select appropriate measurement points on the electric spindle to install acceleration sensors. The sensor measurement points are arranged along the axial and radial directions of the electric spindle. It measures the vibration acceleration of the electric spindle in all three axes of a three-dimensional Cartesian coordinate system during operation, as shown in Figure 5.

The experimental center adopted electrical discharge machining [20] to artificially induce four distinct vibration patterns in rolling bearings and carried out tests across varying load conditions.

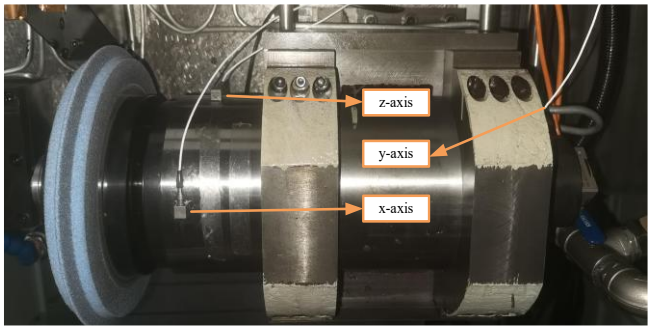


Figure 5. Sensor measurement point layout.

Tests were conducted under both normal and faulty conditions for the inner ring, outer ring, and rollers. In this section, vibration data sets from the drive end bearing under various conditions, including a rotational speed of 1800 rpm, 12 kHz sampling frequency, and a faulty damage diameter of 7 mils, are selected for experimental verification.

According to Shannon's sampling theorem [21], the above data set is resampled twice. Each sample point is set to 1024, and 100 samples of each signal type are collected. The waveform diagrams of different signals after the second sampling are shown in Figures 6 and 7.

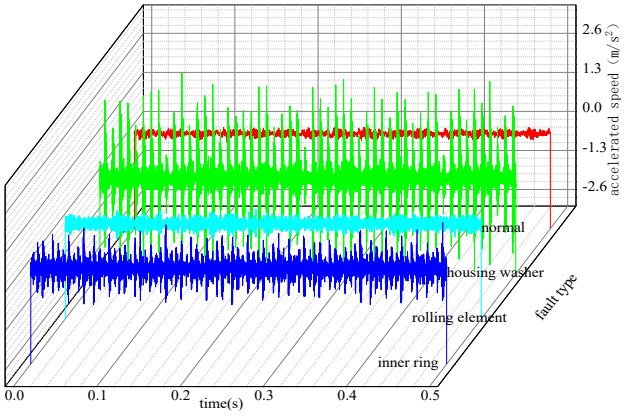


Figure 6. Bearing time domain waveform.

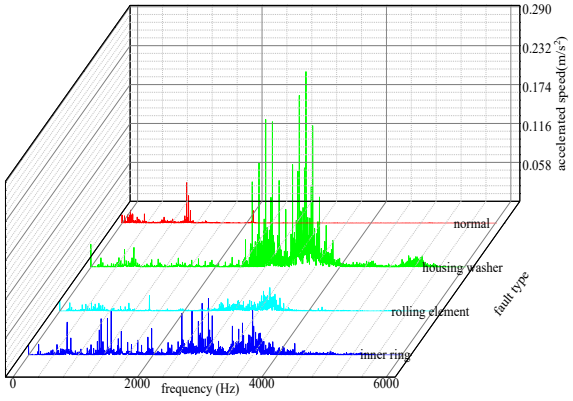


Figure 7. Bearing frequency domain waveform diagram.

The traditional CNN structure established in this section is that the pooling layers adopt the maximum pooling method, the convolutional, pooling, and fully connected layers use the ReLU activation function [22], and the output layer uses the Softmax function for feature recognition to handle multi-classification situations. The optimizer employs the Adam algorithm to dynamically adapt the learning rate for every model parameter, and the loss function is a cross-entropy loss function [23]， shown in Table 1.

Table 1. CNN Structure.

Serial number	Network layer	Size and step length	Number of channels
1	Convolutional 1	3*3，1*1	16
2	Pooling 1	2*2，2*1	16
3	Convolutional 2	3*3，1*1	32
4	Pooling 2	2*2，2*1	32
5	Convolutional 3	3*3，1*1	64
6	Pooling 3	2*2，2*1	64
7	FC	128	1

B. Data Preparation and Preprocessing

The denoising algorithm based on the wavelet packet-improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) first performs wavelet packet decomposition on the original signal and processes components

with higher energy proportions [24]. The EMD algorithm is upgraded to the ICEEMDAN. By applying wavelet threshold denoising reconstruction to intrinsic mode function (IMF) signals with higher sample entropy values, denoised signals with improved signal-to-noise ratios (SNR) are obtained. Unlike traditional EMD, ICEEMDAN integrates adaptive noise and a complete ensemble strategy, enhancing the stability and accuracy of the decomposition process. During decomposition, ICEEMDAN incorporates IMF components derived directly from the original signal, avoiding residual noise and spurious modes caused by random Gaussian white noise injection in improved methods like EEMD. The ICEEMDAN algorithm workflow is as follows:

Introduce adaptive noise to the original signal $s(t)$:

$$s^{(i)}(t) = s(t) + \beta_0 E_1\left(\omega^{(i)}(t)\right) \tag{9}$$

Decompose the obtained signal $s^{(i)}(t)$ using the EMD method, calculate the local mean $N\left[s^{(i)}(t)\right]$ of each component (as shown in Formula 10), and obtain the first residual R_1:

$$R_1 = N\left[s^{(i)}(t)\right] \tag{10}$$

Compute the first IMF component:

$$\widetilde{IMF_1} = s(t) - R_1 \tag{11}$$

Continue adding adaptive noise and decompose to compute the second residual:

$$R_2 = R_1 + \beta_1 E\left(w^{(i)}(t)\right) \tag{12}$$

Obtain the second IMF component:

$$\widetilde{IMF_2} = R_1 - R_2 \tag{13}$$

Repeat the process to compute the $n-th$ residual and mode component. The $n-th$ residual can be expressed as:

$$R_n = N\left\{R_{n-1} + \beta_{n-1} E\left[w^{(i)}(t)\right]\right\} \tag{14}$$

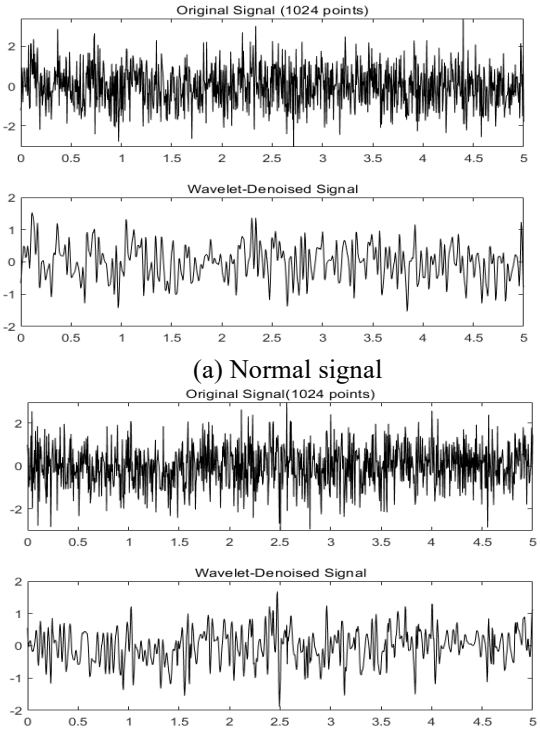
The $n-th$ IMF component is:

$$\widetilde{IMF_n} = R_{n-1} - R_n \tag{15}$$

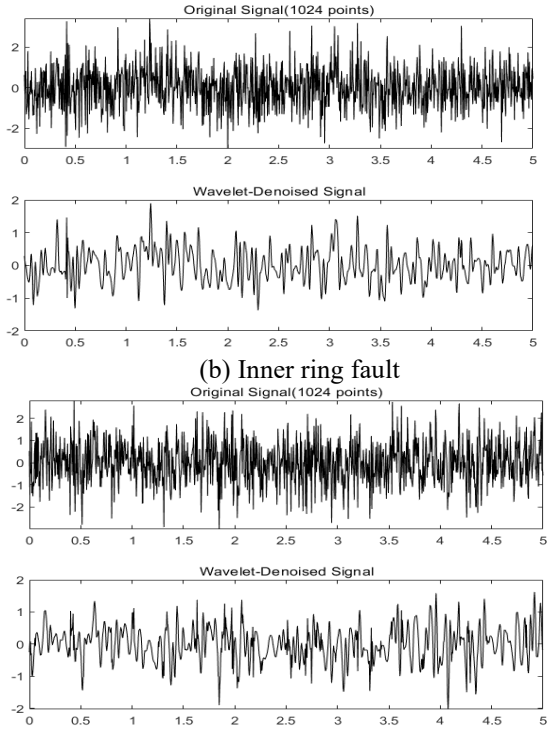
After ICEEMDAN decomposition, the original signal can be reconstructed as:

$$s(t) = \sum_{n=1}^{n} \widetilde{IMF_n} + R_n \tag{16}$$

Using this method to preprocess the four collected vibration signals, the results of filtering are shown in Figure 8.



(a) Normal signal



(b) Inner ring fault

(c) Outer ring fault

(d) Rolling element fault

Figure 8. The result diagrams of the four types of signals after filtering.

C. Two-dimensional image conversion based on CWT

CNN can recognize the same patterns appearing at different positions in the image. This paper employs the wavelet time-frequency transform to map the denoised signal into a two-dimensional image and investigates its impact on the recognition accuracy of CNN.

Wavelet transform illustrates the original vibration signal in both the time and frequency domains by translating and scaling the mother wavelet. The mathematical principle of the mother wavelet can be expressed as:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}), \quad a,b \in R, a > 0 \tag{17}$$

where $\psi_{a,b}(t)$ represents the continuous wavelet function basis, a represents the scaling factor that changes the shape of the wavelet, and its function is to affect the scaling ratio of the wavelet function. b represents the translation factor of the wavelet shift.

Since the measured signal is energy-limited, its allowable condition is that the amplitude satisfies the condition:

$$C_\Psi = \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|^2} d(\omega) < +\infty \tag{18}$$

Continuous Wavelet Transform (CWT) decomposes the signal into different frequencies [25], and analyzes the local variations of the signal at each frequency. For any A, there is:

$$W_s(a,b) = \langle f, \psi_{a,b} \rangle = \frac{1}{\sqrt{a}} \int f(t)\psi^*(\frac{t-b}{a})dt \tag{19}$$

where $W_s(a,b)$ represents wavelet transform coefficients, $\langle f, \psi_{a,b} \rangle$ represents inner product, $f(t)$ represents the original signal, and ψ^* represents complex conjugate.

The Complex Morlet wavelet, as the complex form of the Morlet wavelet, compared to other wavelets, it exhibits excellent resolution, and its adaptive performance is also more outstanding. Therefore, we choose it as the wavelet basis of CWT. The expression of the Complex Morlet wavelet is shown in Formula 20:

$$\psi_0(t) = \frac{1}{\sqrt{\pi f_b}} exp(2i\pi f_c t)\, exp(\frac{-t^2}{f_b}) \tag{20}$$

where f_b is the bandwidth, and f_c is the center frequency of the wavelet. Compared with the Gram angle field and other two-dimensional image conversion methods, CWT can analyze signals in different frequency domains based on the different scale and translation characteristics of the wavelet function, and has better feature recognition performance [26].

The process of wavelet time-frequency transformation is as follows:

1) Determine the relationship between scale and frequency

The relationship can be determined by Formula 21:

$$F_a = f_c \times \frac{f_s}{a} \tag{21}$$

where f_s is the sampling frequency. According to the sampling theorem, to make the frequency range of the scale diagram $\left(0, \frac{f_s}{2}\right)$, the scale range should be $(2 \times f_c, \infty)$.

2) Determine the scale sequence

To ensure the frequency sequence constitutes an arithmetic sequence, the scale sequence is represented in the following form:

$$c/totalscal, c/(totalscal-1), \ldots, c/4, c/2, c \tag{22}$$

where $totalscal$ represents the length of the scale sequence used in the wavelet transform, and c represents a constant. For a given scale $\frac{c}{totalscal}$, the corresponding frequency is $\frac{f_s}{2}$, and there is:

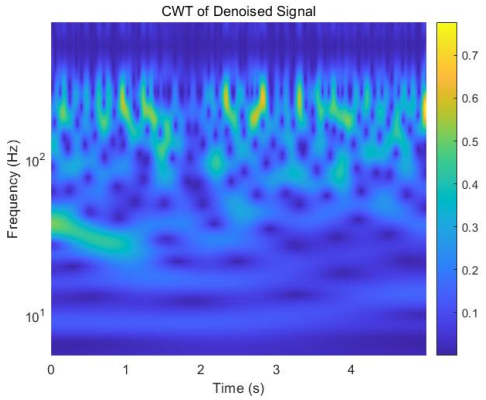
$$c = 2 \times f_c \times totalscal \tag{23}$$

According to Formulas 21 and 23, the corresponding scale sequences can be obtained.

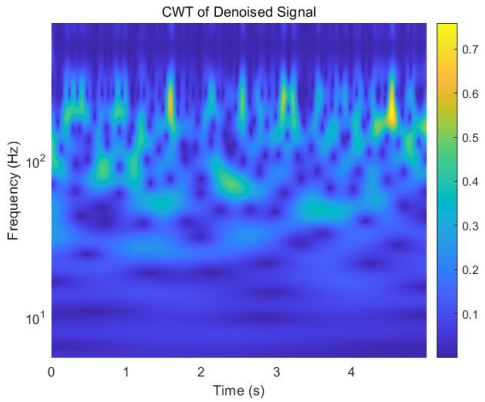
3) Draw the time-frequency diagram

After determining the scale and wavelet basis, the actual frequency sequence can be calculated using Formulas 19 and 23. Combining this with the time sequence, a time-frequency diagram can be drawn.

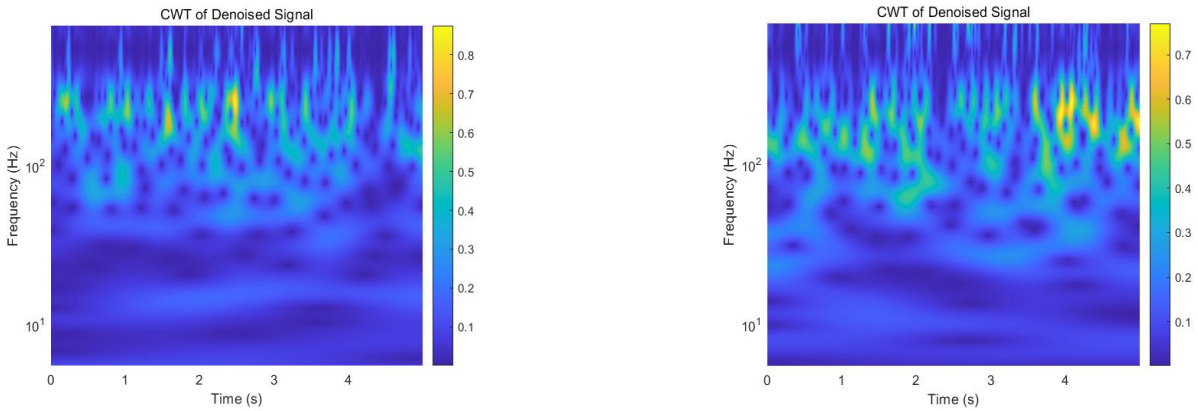
The four types of signals after noise reduction are converted into two-dimensional images using CWT. Finally, the time-frequency diagrams of the bearing in different states are obtained, as shown in Figure 9



(a) Normal



(b) Inner ring fault



(c) Outer ring fault (d) Rolling element fault

Figure 9. Time-frequency diagrams of bearing under different conditions.

D. Data set partitioning

Record the signals after the two-dimensional image conversion, divide them into different fault types, and attach corresponding labels to them as the sample set for training. The specific configuration is shown in Table 2.

Table 2. Specific details of data and labels.

Signal type	Training set (%)	Test set (%)	Sample label
Normal	80	20	1
Inner ring fault	80	20	2
Outer ring fault	80	20	3
Rolling element fault	80	20	4
Sum total	320	80	

4. Bearing vibration feature recognition based on traditional CNN

A. T-SNE Visualization

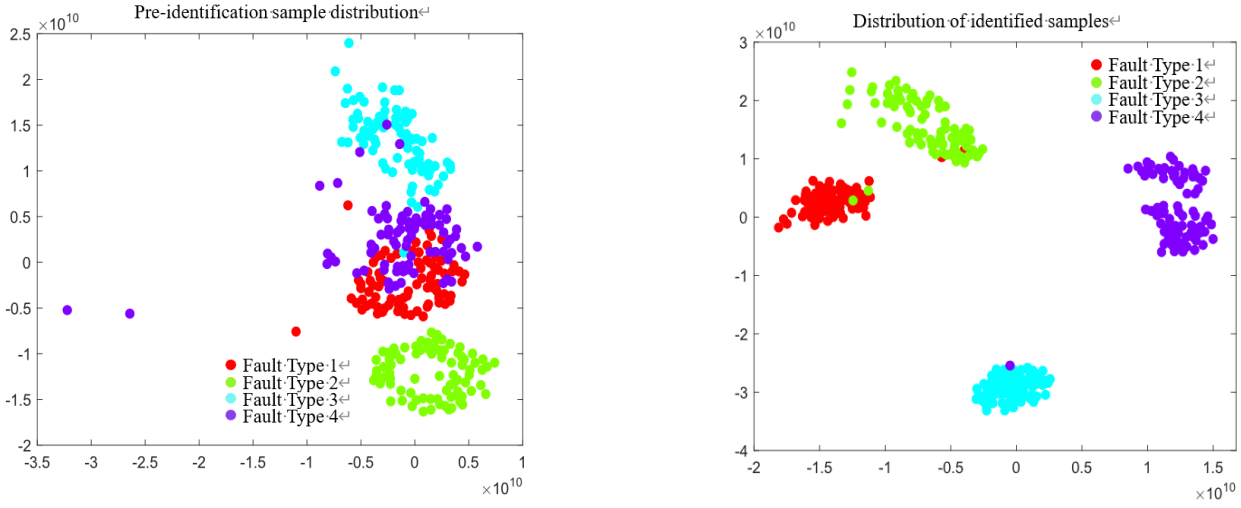
t-SNE (t-distributed Stochastic Neighbor Embedding) is a machine learning algorithm used for data dimensionality reduction and visualization [27]. It uses gradient descent to minimize the cost function and adjusts data points in low-dimensional space to maximize their relationships in high-dimensional space.

Visualizing the classification results obtained from training a traditional CNN, it is evident that the model still has misclassification cases for samples, as shown in Figure 10.

B. Forecast the Outcome

The model was used to make predictions on the test set, as shown in Figure 11. The training accuracy and loss are shown in Figure 12.

In the figure, the model has reached convergence when the training iteration is about 300 times, and the accuracy of the test set prediction can reach 91.25%.



(a) The sample distribution before identification (b) The sample distribution after identification

Figure 10. T-SNE visualization diagram.

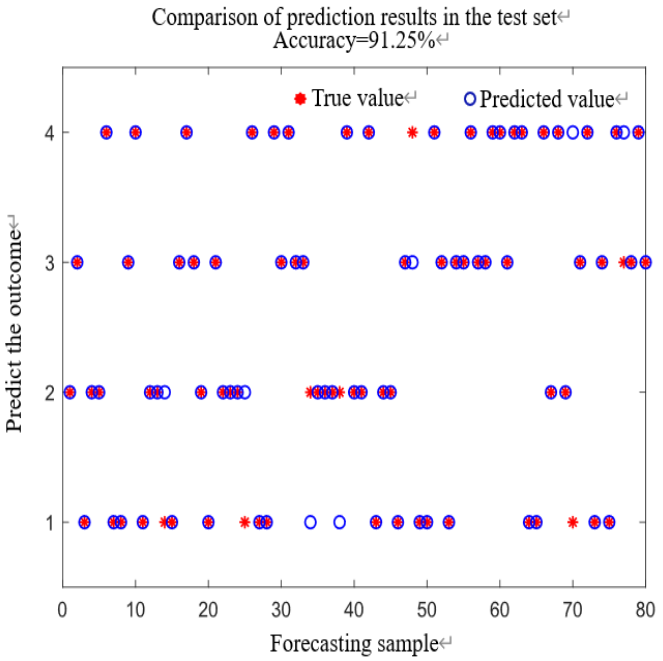


Figure 11. Forecast result diagram.

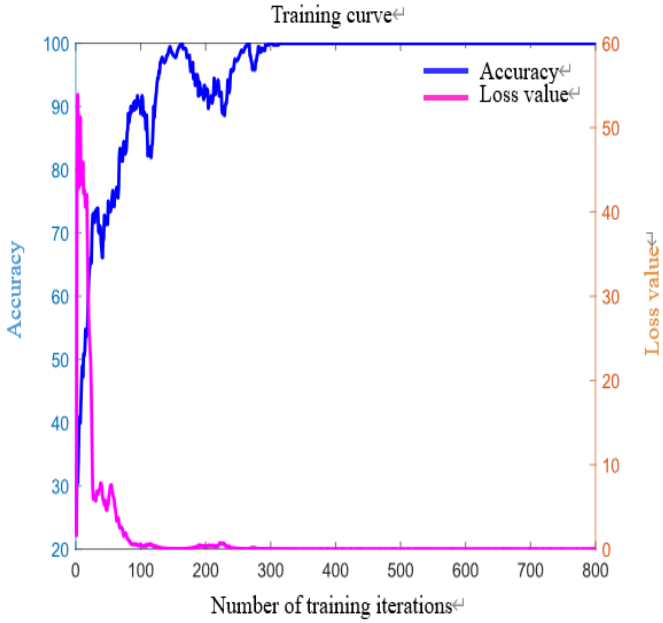


Figure 12. Accuracy and loss curves.

C. Confusion matrix

The Confusion Matrix is a table used to evaluate the performance of classification models in the field of machine learning. Using the actual classes and predicted classes [28], the confusion matrix divides the classification results into four different categories: true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

The calculation rules of the confusion matrix are shown in Table 3.

Table 3. Calculation rules.

Actual category/forecast category	Prediction is a positive example	Prediction is a negative example
The actual situation is a positive example	TP	FN
The actual example is negative	FP	TN

According to the calculation rules of the confusion matrix, the confusion matrix of the training outcomes can be obtained, as shown in Figure 13.

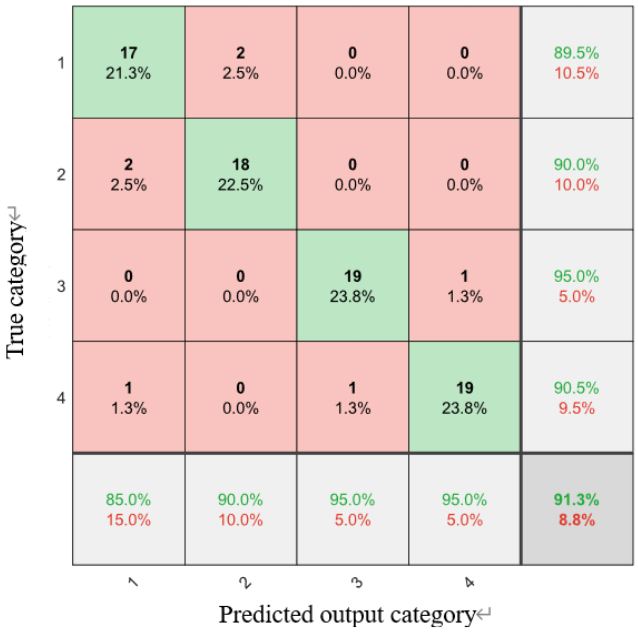


Figure 13. Forecast result diagram.

Through the confusion matrix, multiple evaluation metrics can be calculated, and these metrics are used to assess the performance of the classification model [29].

1) Accuracy

To measure the model's overall prediction accuracy, this metric is defined as the proportion of correctly predicted samples to the total number of samples.

$$\text{accuracy rate} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{24}$$

2) Precision

To assess the model's effectiveness in correctly predicting positive instances, this metric is defined as the rate of true positives among all predicted positive samples.

$$\text{precision rate} = \frac{TP}{(TP+FP)} \tag{25}$$

3) Recall

To assess the model's effectiveness in identifying positive

instances, this metric is defined as the rate of true positives among all actual positive examples.

$$\text{recall rate} = \frac{TP}{(TP+FN)} \tag{26}$$

4) F1 score

To evaluate the model's ability to balance precision and recall, this metric is defined as the harmonic mean of precision and recall, making it particularly suitable for imbalanced datasets.

$$F1 = 2 \times \frac{(\text{precision rate} \times \text{recall rate})}{(\text{precision rate} + \text{recall rate})} \tag{27}$$

D. Experimental Result

According to the above formulas, the indicators of the experimental results can be obtained. The recall rates, precisions, and F1 scores of different categories are shown in Table 4.

Table 4. Calculation results of different indicators.

Categories/ Indicators	Recall rate (%)	Accurate rate (%)	F1-score mode (%)
1	85	89.4737	87.1795
2	90	90	90
3	95	95	95
4	95	90.4762	92.6829

Considering the rigor of the experiment, the above method was used to conduct ten repeated independent experiments. The accuracies of the prediction are shown in Table 5.

Table 5. The results of 10 experiments were predicted.

Number of experiments	Accuracy rate（%）
1	91.25
2	91.3
3	91.56
4	92.02
5	91.01
6	90.95
7	91.21
8	91.63
9	91.37
10	91.26

The table shows that the traditional CNN model achieves an average recognition accuracy of 91.4%. This suggests that the model demonstrates relatively strong judgment performance for various types of fault signals.

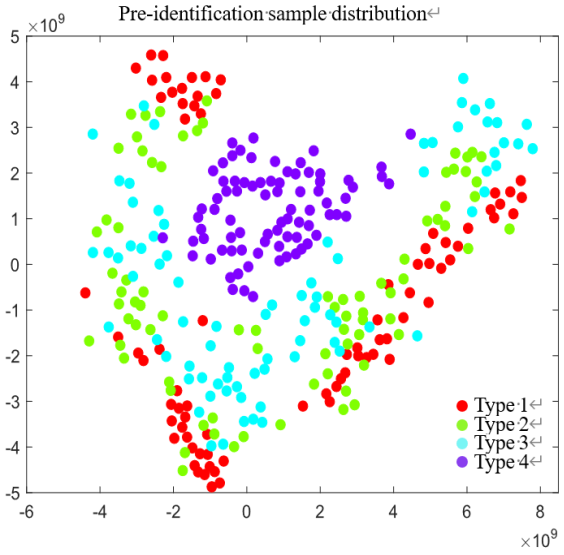
This shows that it is feasible to convert vibration signals into two-dimensional images for feature recognition. However, due to the simple structure of the traditional CNN model, this paper will improve the CNN structure to enhance the performance of the model.

5. Bearing vibration feature recognition based on improved CNN

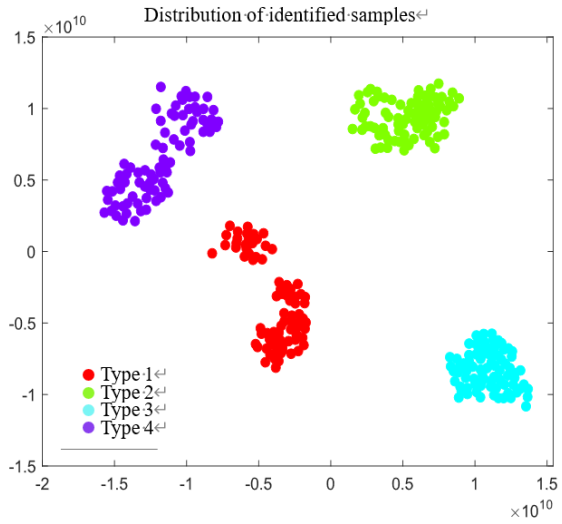
Building upon the experimental data presented in the earlier subsection, the traditional CNN network model has relatively better performance in the recognition of the vibration characteristics of the electric spindle bearing. However, due to the problem of the model structure, there are still cases of incorrect recognition. To address this issue, in this section, the CNN structure will be improved to enhance the recognition performance of the CNN network. Through experiments and comparison with the traditional model, the proposed method's generalization capability will be verified.

The improved CNN model was trained using the same datasets as in the previous section, with the maximum number of iterations of ZOA set to 10, the optimization range of the learning rate set from 0.001 to 0.1, the size range of the convolution kernel of branch 1 set from 1 to 5, the size range of the convolution kernel of branch 2 set from 2 to 6, and the optimization range of batch size set from 10 to 50. After optimizing the hyperparameters through ZOA, the optimal learning rate of the model is 0.0048492, the convolution kernel size of branch 1 is $1 * 1$, the convolution kernel size of branch 2 is $4 * 4$, and the batch size is 30.

The experimental results are visualized by t-SNE, as shown in Figure 14.



(a) The sample distribution before identification



(b) The sample distribution after identification

Figure 14. T-SNE visualization diagram.

Through t-SNE visualization, the distribution of samples before and after dataset recognition can be more intuitively observed. Compared with the original samples, the improved model can achieve aggregation of the same category and distinction of different categories, and no outliers appear in the data after classification. Therefore, the improved CNN has better visualization effects

The prediction result graph of the training set is a tool used to visualize the model's performance on the training data. By displaying the model's prediction results on the training data as charts or graphs, the fitting of the model to the training data can be more clearly observed. After model training, the prediction effect of the improved CNN model on the training set is shown in Figure 15.

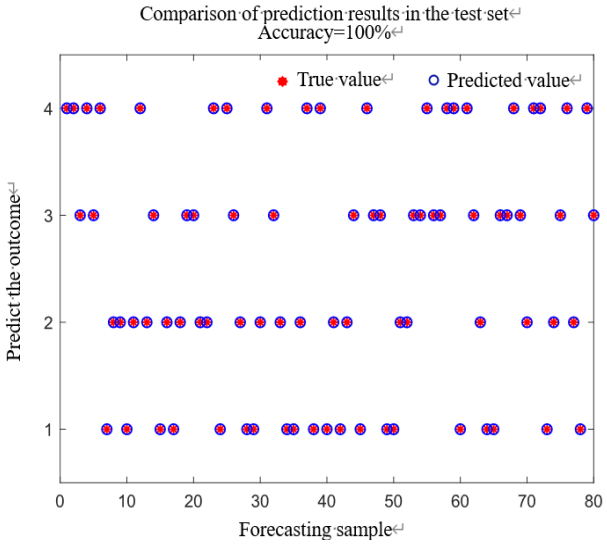


Figure 15. Predictive rendering of the training set.

The fitness curve [30] usually refers to the curve plotted based on the fitness of different candidate solutions in an optimization problem. It can show the performance of different solutions in the search space and help analyze and compare the quality of different candidate solutions. Judging from the shape and trend of the fitness curve, one can determine the distribution of candidate solutions and whether there are local or global optimal solutions, etc.

The fitness curve of ZOA during training is shown in Figure 16.

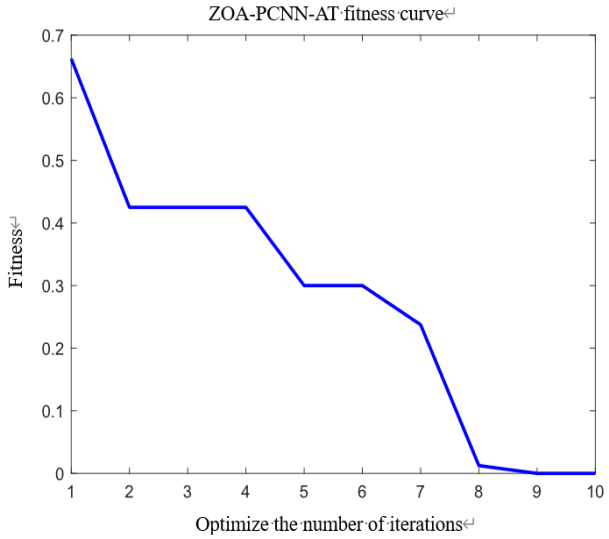


Figure 16. Fitness curve of ZOA.

As is evident from the above graph, with the increase in optimization iterations, the fitness curve monotonically declines. This indicates that during the training of the model, with the increment of iterations, the target function values (fitness) of the candidate solutions found by the optimization algorithm keep decreasing. This suggests that the algorithm is moving towards a better solution.

In Figure 17, the ZOA-PCNN-AT model has reached convergence when the training iterations are around 200. Compared with the traditional CNN model, the ZOA-PCNN-AT model can identify different fault types with fewer iterations and has a better prediction effect.

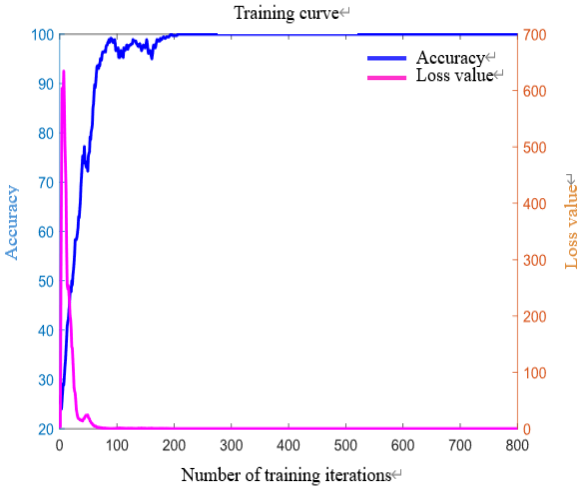


Figure 17. CWRU data set model training curve.

By the training results, the confusion matrix of the ZOA-PCNN-AT model on the CWRU dataset can be obtained.

as shown in Figure 18.

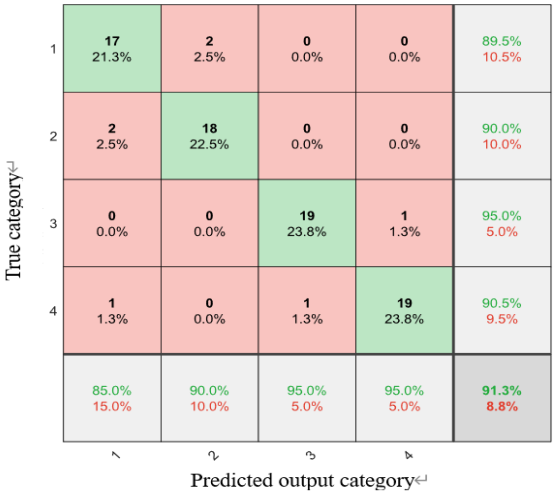


Figure 18. CWRU data set confusion matrix.

To ensure the accuracy of the experimental outcomes, the above-mentioned method was used for ten independent experiments with repetition. The accuracy is shown in Table 6.

Table 6. The results of 10 experiments were predicted.

Number of experiments	Accuracy rate（%）
1	100
2	99.5
3	99.8
4	100
5	99.7
6	99.5
7	100
8	100
9	99.9
10	99.8

From the above table, the average recognition accuracy of the model can reach 99.82%, which indicates that this model has better judgment performance for different types of fault signals. Compared with the traditional CNN model, the ZOA-PCNN-AT model has improved its diagnostic performance by 9.21%.

To further evaluate the model's performance, ablation experiments on the network structure were carried out [31]. By gradually deleting different component modules in the network to study their impact on performance. After each ablation, ten independent experiments were repeated, and the average recognition rate of training was recorded and is shown in Table 7.

Table 7. Table of accuracy comparison in ablation experiments.

Construction	Training set recognition rate	Test set recognition rate
ZOA-PCNN-AT	100%	99.82%
CNN	81.65%	81.21%
ZOA-CNN	82.43%	81.74%
ZOA-CNN-AT	99.03%	98.16%
PCNN-AT	96.41%	94.33%
ZOA-PCNN	93.62%	90.78%

As shown in Table 7, the improved ZOA-PCNN-AT network structure proposed in this chapter shows notably better recognition results than the other five methods. The average recognition accuracy can reach up to 99.82%. Through ablation experiments, it can be observed that different component modules have a notable effect on improving the performance.

6. Bearing vibration feature recognition based on improved CNN under service conditions

A. Vibration Signal Acquisition

After noise reduction, signal samples are collected under different states. Because of the scarce training samples, data augmentation is performed on the collected vibration data to reduce the occurrence of training overfitting. In this paper, a random overlapping sampling method of signals is adopted for dataset enhancement; that is, each segment of the signal is randomly sampled with continuous data points of the same length as the original signal. The number of sampling points is 1,024, and the sampling frequency is 512 Hz. 100 samples are

collected for each rotational speed, as shown in Figure 19.

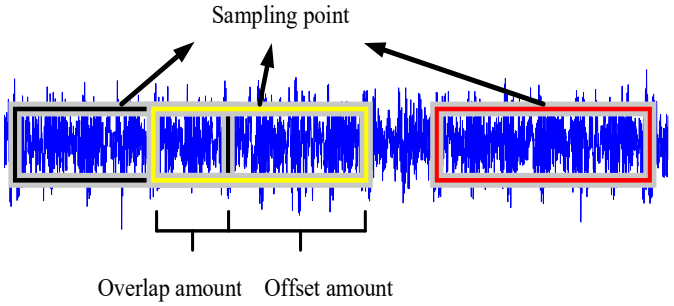


Figure 19. Data set extension.

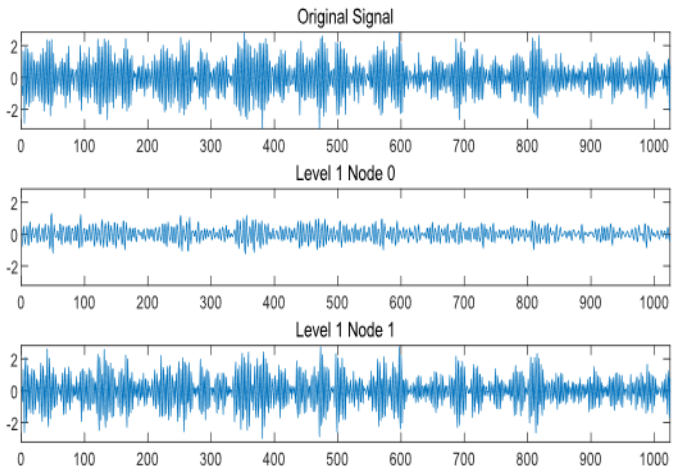
Fault signals of various types shall be divided according to the same criteria, as shown in Table 8.

Table 8. Specific details of the data and labels.

Signal type	Training set	Test set	Sample label
Normal state at 1000 rpm	80	20	1
Normal state at 3000 rpm	80	20	2
Pull stud wear at 1000 rpm	80	20	3
Pull stud wear at 3000 rpm	80	20	4
Bearing wear at 1000 rpm	80	20	5
Bearing wear at 3000 rpm	80	20	6
Sum total	480	120	600

B. Vibration Signal Preprocessing

The vibration signals of the electric spindle under 6 different states are preprocessed. Taking the bearing wear signal at a rotational speed of 3000 rpm as an example. Figure 20 shows the frequency spectra of each node after wavelet packet decomposition, and the energy proportion, as shown in Figure 21.



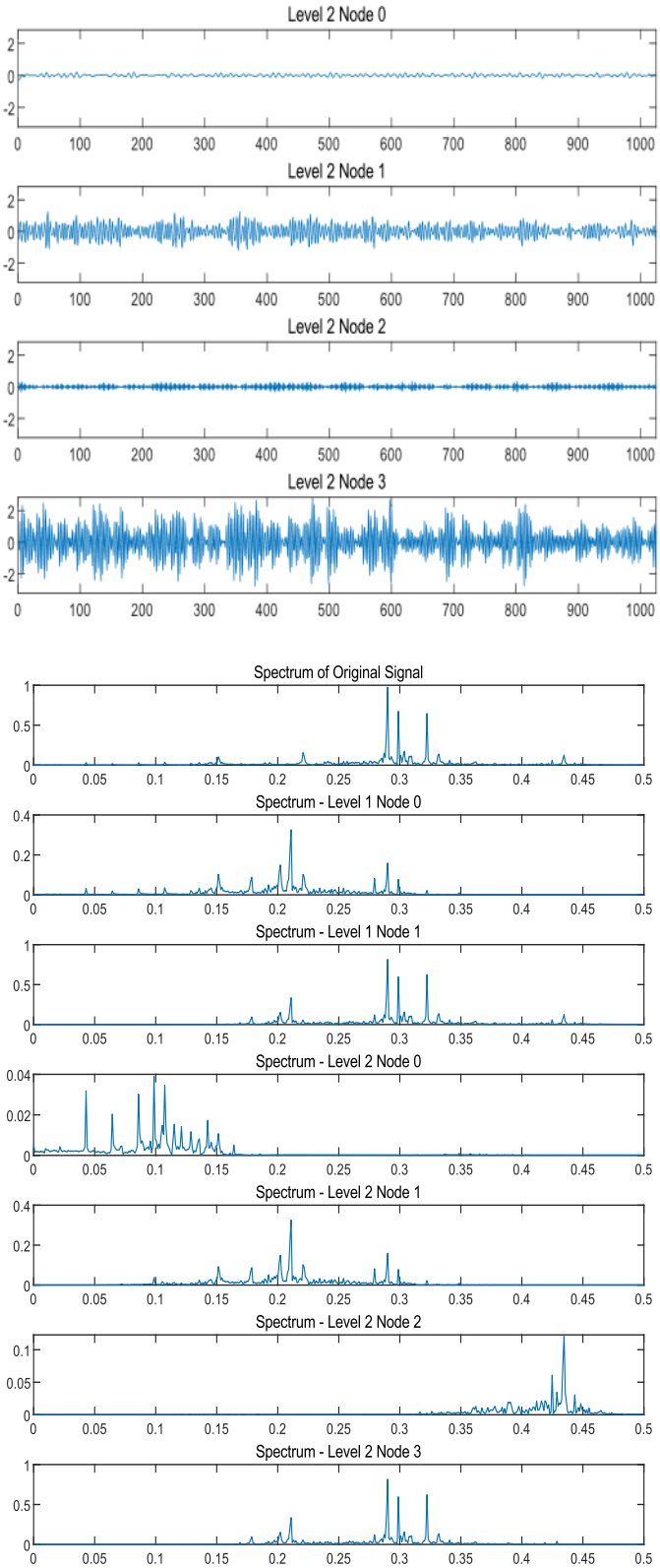


Figure 20. Bearing wear signal at 3000 rpm wavelet packet decomposition node spectrum diagram.

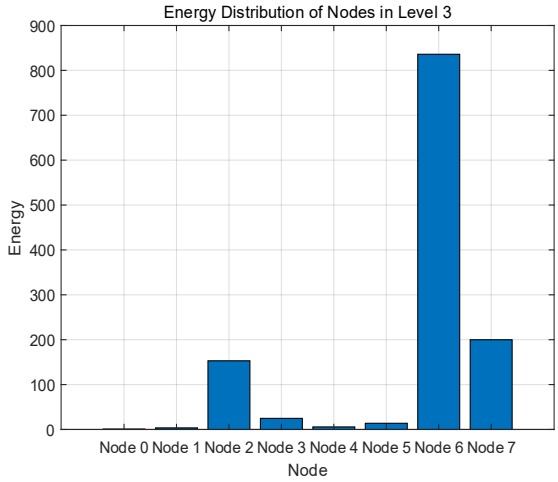


Figure 21. Frequency distribution of energy.

In Figure 21, the signal energy corresponding to node [3,6] accounts for the highest proportion, while the remaining signal energy proportions are less than 10%. Therefore, the frequency distribution of the original signal is concentrated in this node, while other nodes represent the frequency distribution of noise signals. Therefore, the signal of this node is used as the component for wavelet packet reconstruction.

Due to the high energy proportion of nodes [3,6], ICEEMDAN decomposition was performed on them to obtain IMF components in different frequency bands, and the results are shown in Figure 22.

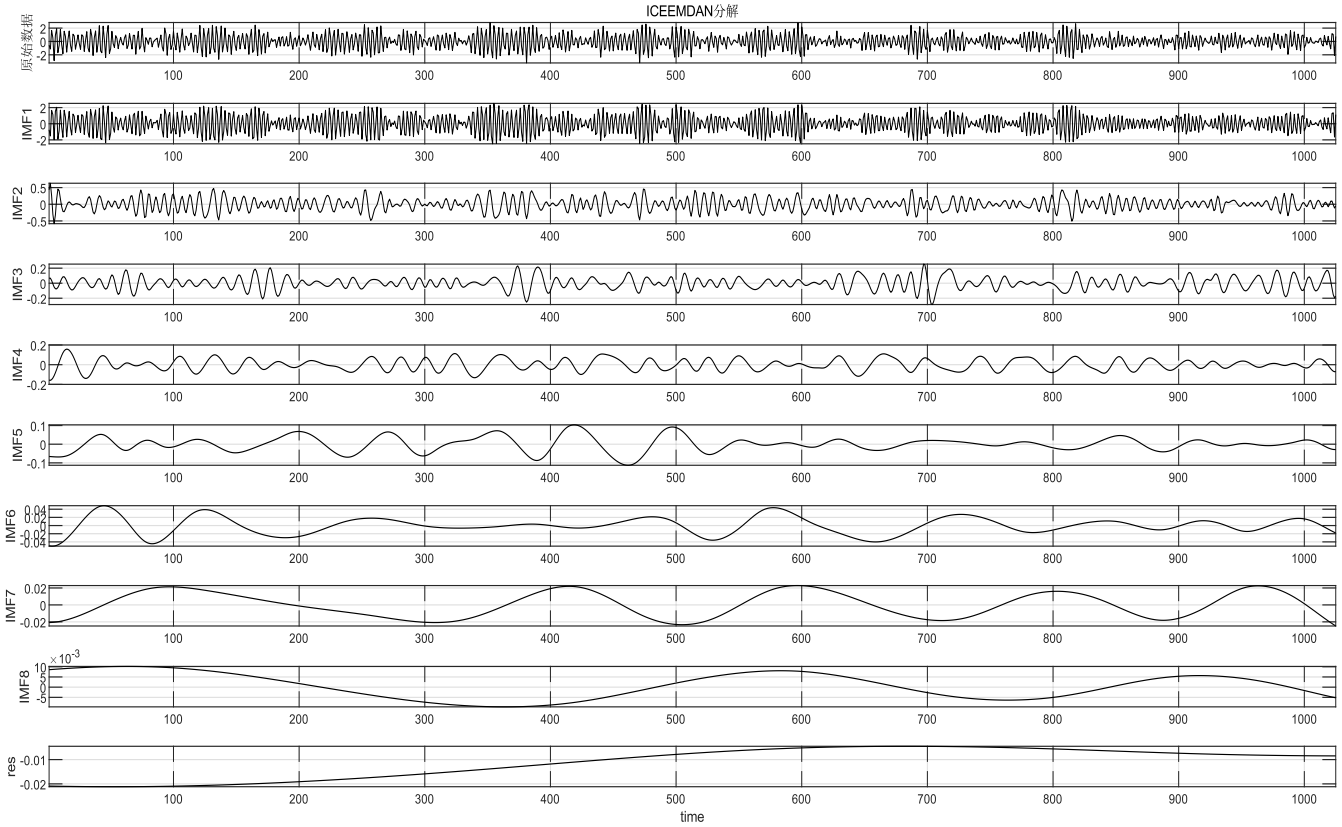


Figure 22. ICEEMDAN decomposition result of the bearing wear signal at 3000 rpm.

The sample entropy of all IMFs is calculated, as shown in Table 9.

Table 9. Sample entropy values of different IMFs for bearing wear signals at 3000 rpm.

IMF serial number	Sample entropy value
IMF1	1.2961
IMF2	0.99029
IMF3	0.66611
IMF4	0.59559
IMF5	0.36808
IMF6	0.26979
IMF7	0.13193
IMF8	0.04704
Average IMF	0.5456

Filtering is performed according to the magnitude of their sample entropy, and the comparison of results before and after filtering is shown in Figure 23.

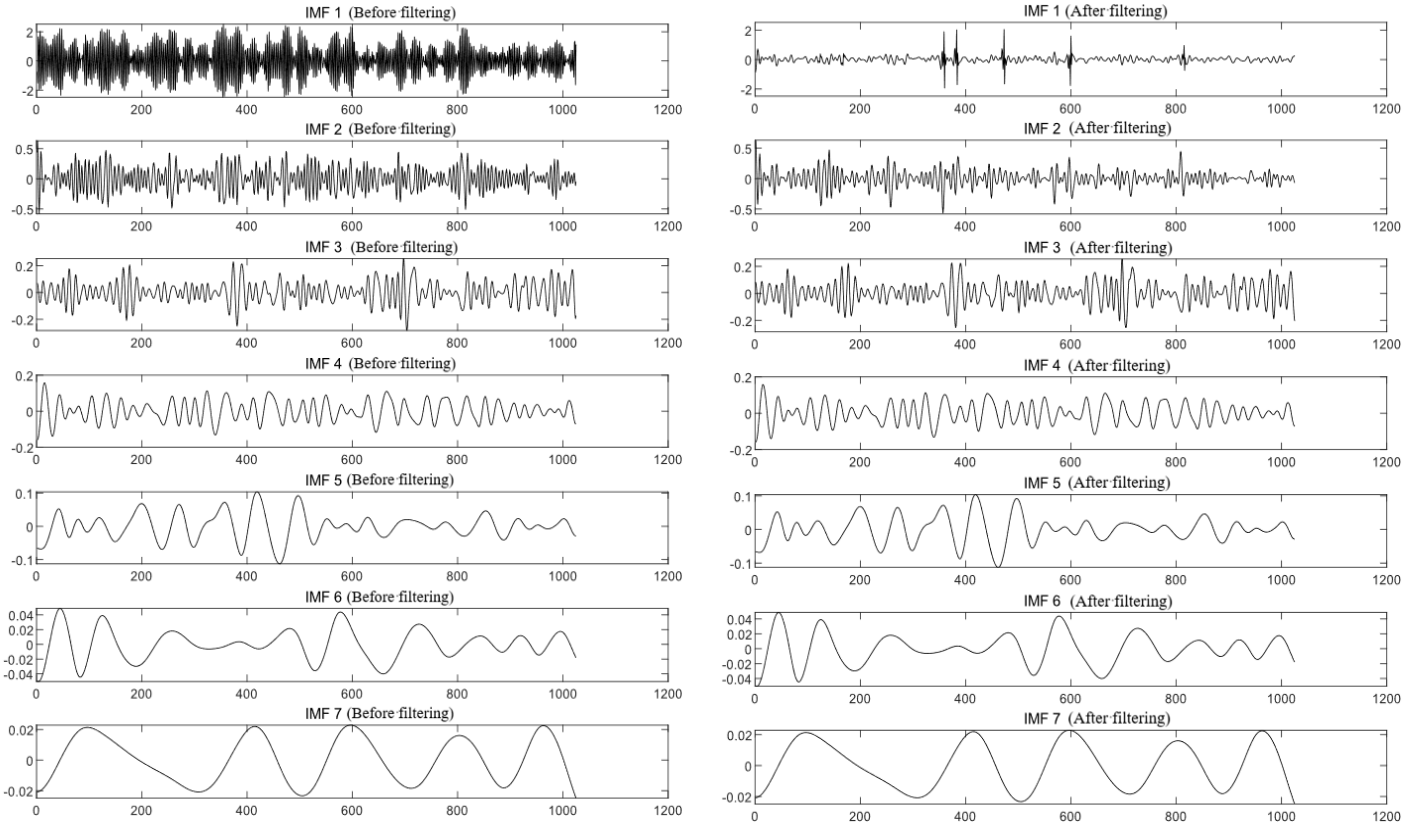


Figure 23. Comparison of bearing wear signals before and after IMF filtering at 3000 rpm.

The filtered IMFs are reconstructed to obtain the denoised vibration signal of the electric spindle bearing wear at 3000 rpm, as shown in Figure 24:

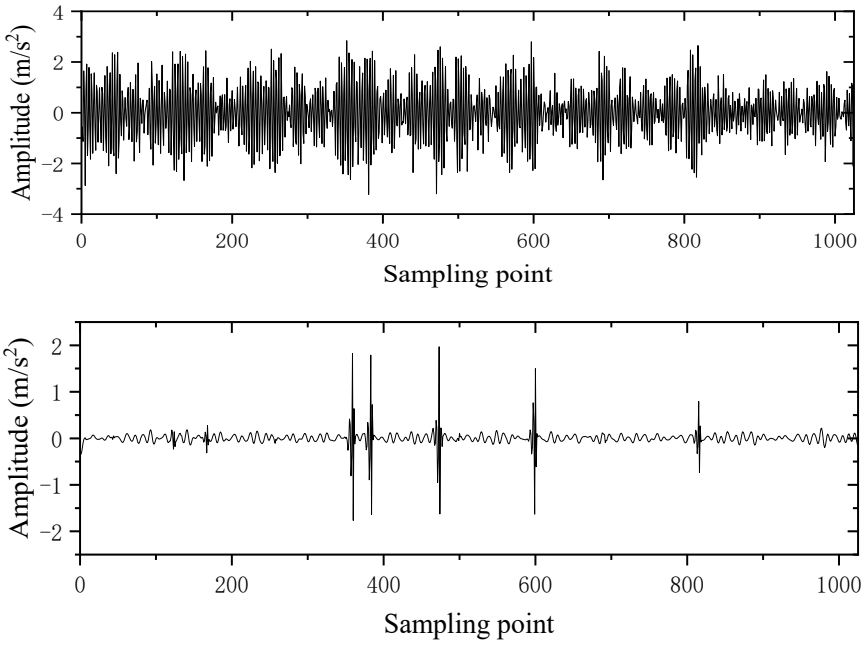
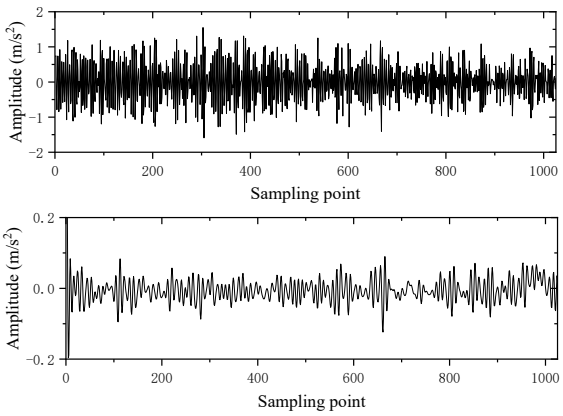
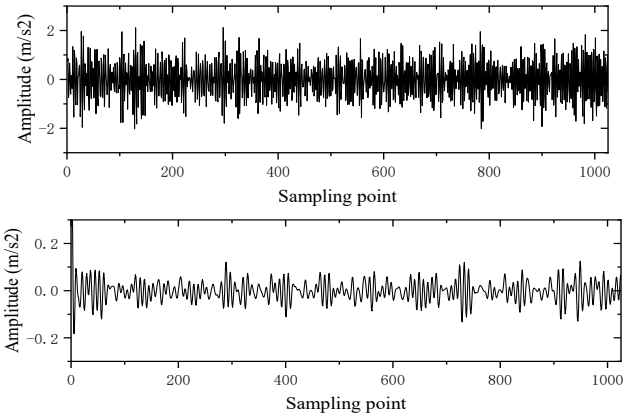


Figure 24. Comparison between the original bearing wear and the filtered signal at 3000 rpm.

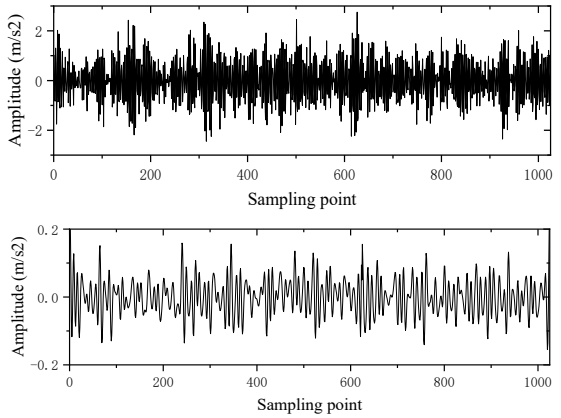
The vibration signals of the electric spindle under fault and normal states at different rotational speeds are processed using the above method. The comparison results between the original signals and the denoised signals under different states are shown in Figure 25.



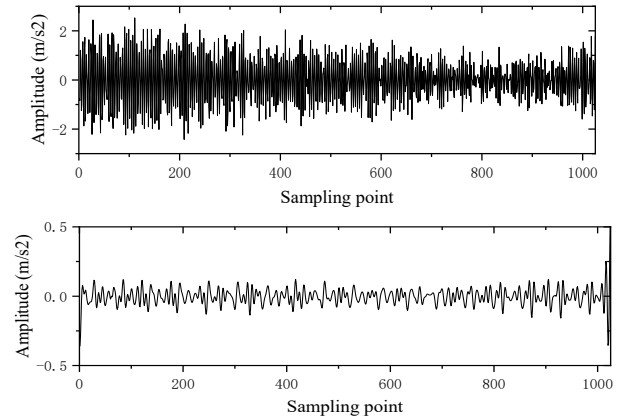
(a) Normal state at 1000 rpm



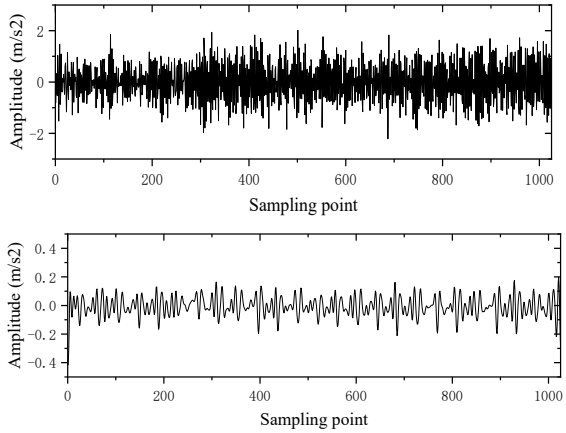
(b) Normal state at 3000 rpm



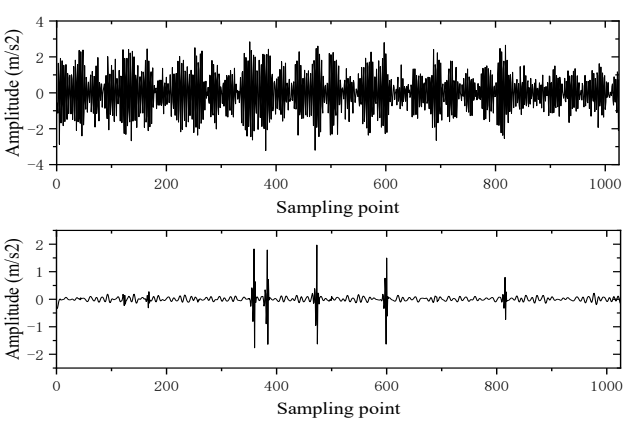
(c) Pull stud wear at 1000 rpm



(d) Pull stud wear at 3000 rpm



(e) Bearing wear at 1000 rpm



(f) Bearing wear at 3000 rpm

Figure 25. Comparison Chart of Original Signal and Filtered Signal.

C. Feature Recognition

The obtained signals are converted into two-dimensional images through CWT and subjected to compression processing. The processed sample set is then input into the PCNN network for training. ZOA's maximum iteration count is set to 10, with the optimization interval for the learning rate defined as 0.001 to 0.1. The convolution kernel size interval for Branch 1 is set as 1 to 5, and that for Branch 2 is 2 to 6. The optimization interval for batch size is specified as 10 to 50. Through model training, the fitness curve of the PCNN is obtained as shown in Figure 26. Obviously, with the increase of iteration times, the fitness of candidate solutions found by the optimization algorithm continuously decreases, indicating that the algorithm is progressing toward better solutions.

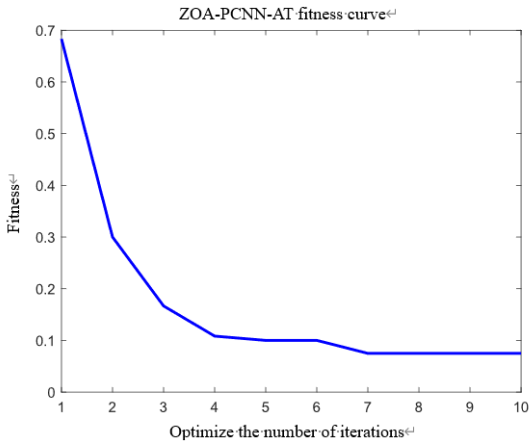
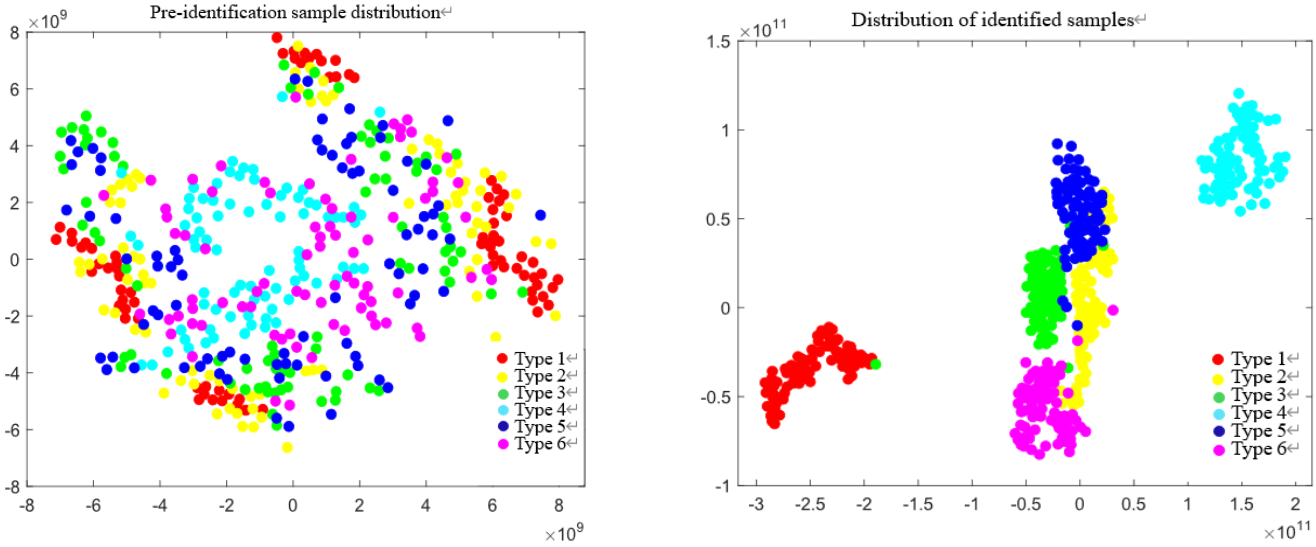


Figure 26. The fitness curve of CNN was improved.

When the iteration count is approximately 7, the model's fitness starts to converge, demonstrating the effectiveness of the ZOA optimization algorithm in hyperparameter optimization. Through ZOA hyperparameter optimization for the model, a learning rate of 0.0052619, a convolution kernel size of 1×1 for Branch 1, a convolution kernel size of 2×2 for Branch 2, and a batch size of 20. The sample classification diagram after t-SNE dimensionality reduction is shown in Figure 27.



(a) The sample distribution before identification

(b) The sample distribution after identification

Figure 27. t-SNE visualization schematic.

As shown in Figure 27, the model can intuitively identify the vibration conditions of the electric spindle under different states after dimensionality reduction by t-SNE. Figure 28 shows the prediction outcomes for the test set.

According to the figure, the model achieves a prediction accuracy of 92.5% for the electric spindle under different states.

Based on the calculation of experimental results, the confusion matrix is shown in Figure 29.

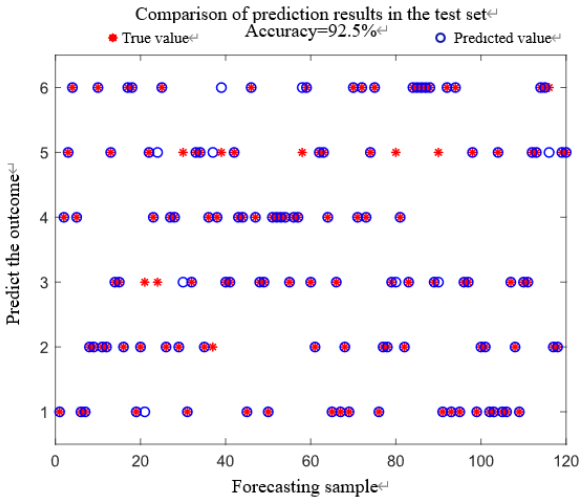


Figure 28. Test set prediction results.

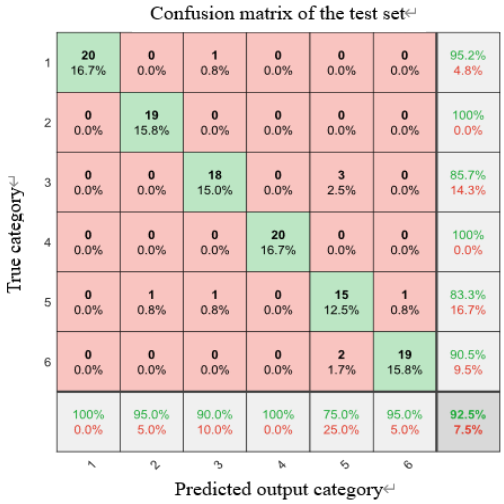


Figure 29. Confusion Matrix of Test Set.

The calculation results of different indicators are shown in Table 10.

Table 10. Different metric calculation results.

Categories/ Indicators	Recall rate(%)	Accurate rate(%)	F1-score mode (%)
1	100	95.2381	97.561
2	95	100	97.436
3	90	85.7142	87.805
4	100	100	100
5	85	83.3333	88.947
6	95	90.4761	92.683

As can be seen from the above table, all calculated values exceed 80%, indicating that the model performs well and the training results are relatively reliable.

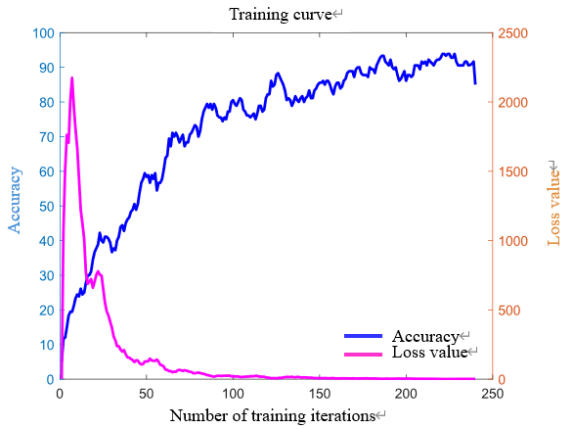


Figure 30. Training curve of the model.

As shown in Figure 30, the model tends to converge when the training set is iterated approximately 225 times, with the accuracy reaching 93%. The model is effective in identifying the vibration conditions of the electro-spindle in various operating conditions.

Considering the rigor of the experiment, ten repeated independent experiments were conducted using the above method, as shown in Table 11.

Table 11. Results of ten experimental predictions.

Number of experiments	Accuracy rate（%）
1	92.5
2	95.3
3	96.1
4	94.2
5	94.6
6	93.5
7	94.4
8	94.6
9	94.5
10	95.1

An analysis of the table content reveals that the ZOA-PCNN-AT model has an average recognition accuracy of 95% for the vibration characteristics of electric spindles under service conditions, indicating the effectiveness of the proposed method.

D. Analysis of Model Recognition Performance Under Strong Noise Background

To ensure the integrity of the experiment, the collected original signals were subjected to noise addition processing, with Gaussian white noise added to simulate the recognition performance of the model for different working conditions of the electric spindle in a strong noise environment. The signal-to-noise ratio (SNR) index was used to evaluate the degree of signal interference by noise: a smaller SNR indicates stronger interference of noise on the original signal, making the vibration characteristics more difficult to identify.

The SNR range after noise addition was set from -10 dB to 10 dB for testing the proposed model. Ten experiments were conducted under each SNR condition, and the results were recorded. To ensure the reliability of the experimental results, the recognition accuracy of the model for vibration characteristics under that SNR condition was determined by averaging the 10 experimental results. To intuitively observe the performance level of the model, ablation of the model structure was performed. The recognition effects of CNN, PCNN-AT, ZOA-PCNN, and ZOA-CNN-AT models on the vibration characteristics of the electric spindle in a noise environment were tested as references.

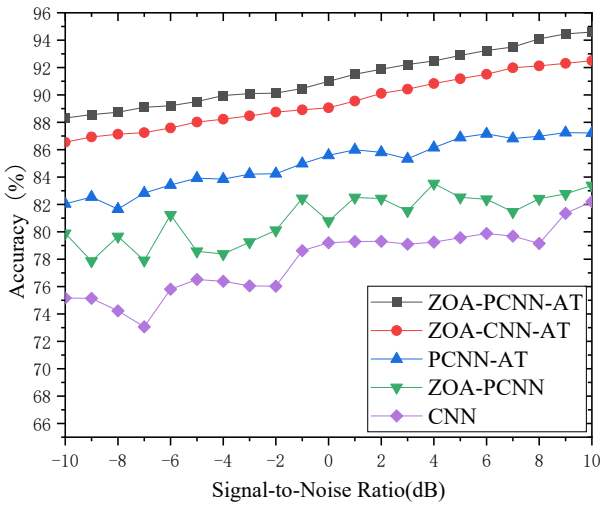


Figure 31. The accuracy of models in a noisy environment.

In Figure 31, the recognition accuracy of vibration characteristics by different models gradually improves. Among them, the model used in this chapter exhibits the highest recognition accuracy, reaching approximately 94.5% when the SNR is 10 dB. The CNN model shows the poorest recognition performance, with a maximum accuracy of only about 80%.

Therefore, the model still demonstrates good recognition performance in noisy environments.

7. Conclusion

This paper presents a study on recognizing bearing vibration features using an enhanced CNN, informed by a literature review and field investigations. First, the EMD algorithm is upgraded to the ICEEMDAN. By performing wavelet threshold denoising reconstruction on IMF signals with higher sample entropy values, denoised signals with improved signal-to-noise ratios are obtained. Subsequently, a ZOA module is incorporated into the traditional CNN architecture, where the Zebra Optimization Algorithm (ZOA) adaptively optimizes hyperparameters such as learning rate and convolutional kernel size. This process yields an optimal hyperparameter combination, including a best learning rate of 0.0048. Finally, by conducting ablation experiments, it can be concluded that compared with the traditional CNN model, the improved model enhances diagnostic performance by 9.21%, and can reach approximately 94.5% when the signal-to-noise ratio is 10 dB. This significantly improves the reliability of CNC machine tool fault diagnosis.

Acknowledgement

This work was partially supported by the National Natural Science Foundation of China (Grant No. 52375052), in part by the Provincial Science and Technology Research and Development Program Joint Fund under Grant 222103810040, and in part by the 2023 Key Scientific Research of Universities in Henan Province Project, under Grant 23A460017.

References

1. Wang H, Deng S, Yang J, et al. A fault diagnosis method for rolling element bearing (REB) based on reducing REB foundation vibration and noise-assisted vibration signal analysis[J]. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2019, 233 (7): 2574-2587. https://doi.org/10.1177/0954406218791209.
2. He M, He D. A new hybrid deep signal processing approach for bearing fault diagnosis using vibration signals[J]. Neurocomputing, 2020, 396(prepublish): 542-555. https://doi.org/10.1016/j.neucom.2018.12.088.
3. Zhang J, Selvakumar S R, Deng N, et al. Bearing Fault Vibration Signal Feature Extraction and Recognition Method Based on EEMD Superresolution Sparse Decomposition[J]. Shock and Vibration, 2022. https://doi.org/10.1155/2022/9985131.
4. Fang S, Hu B, Liu P S. The Analysis on Fire Smoke Exhausting Robot Vibration Based on Wavelet Transform[J]. Advanced Materials Research, 2011, 1442 (328-330): 1887-1891. https://doi.org/10.4028/www.scientific.net/AMR.328-330.1887.
5. Liu F G, Tao Y F, Li G X. Research on the Impeller Tip Screenout Fault Feature Extraction of the Submersible Pump Units Based on Wavelet Analysis[J]. Advanced Materials Research, 2011, 1250 (225-226): 835-838. https://doi.org/10.4028/www.scientific.net/AMR.225-226.835.
6. Shi J, Li J, Wang T. A Convolutional Neural Network Model-Based Approach for Multi-Fault Diagnosis of Asynchronous Motors[C]//2023 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia). IEEE, 2023: 883-888. https://doi.org/10.1109/ICPSAsia58343.2023.10294534.
7. Tang Y, Zhang C, Wu J, et al. Deep learning-based bearing fault diagnosis using a trusted multi-scale quadratic attention-embedded convolutional neural network[J]. IEEE Transactions on Instrumentation and Measurement, 2024. https://doi.org/10.1109/TIM.2024.3374311.
8. Guo J, Yang Y, Li H, et al. A parallel deep neural network for intelligent fault diagnosis of drilling pumps[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108071. https://doi.org/10.1016/j.engappai.2024.108071.
9. Jiang R, Hu P, Cheng S, et al. Rolling bearing fault diagnosis based on VMD optimized by zebra algorithm[C]//2023 10th International Forum on Electrical Engineering and Automation (IFEEA). IEEE, 2023: 1145-1150. https://doi.org/10.1109/IFEEA60725.2023.10429138.
10. Liang J, Wang W, Qu Z, et al. Prediction of Color Change in Heat-Treated Wood Based on Improved Zebra Algorithm Optimized Deep

Hybrid Kernel Extreme Learning Machine Model (IZOA-DHKELM)[J]. Forests, 2025, 16 (2): 253. https://doi.org/10.3390/f16020253.

11. Liang Q, Niu M, Xu X, et al. UAV Obstacle Avoidance Trajectory Planning Based on Improved Zebra Optimization Algorithm[C]//2024 43rd Chinese Control Conference (CCC). IEEE, 2024: 3174-3179. https://doi.org/10.23919/CCC63176.2024.10661711.
12. Deepanraj B, Saravanan A, Senthilkumar N, et al. Drilling characteristics optimization of polymer composite fortified with eggshells using box-Behnken design and zebra optimization algorithm[J]. Results in Engineering, 2025, 25104102-104102. https://doi.org/10.1016/j.rineng.2025.104102.
13. Punia P, Raj A, Kumar P. Enhanced zebra optimization algorithm for reliability redundancy allocation and engineering optimization problems[J]. Cluster Computing, 2025, 28 (4): 267-267. https://doi.org/10.1007/s10586-024-04931-4.
14. Liu H, Yu S, Wang X. Mega-constellation satellite maneuver forecast via network with attention mechanism[J]. Advances in Space Research, 2025, 75 (6): 4942-4962. https://doi.org/10.1016/j.asr.2024.12.058.
15. Sen S, Pandit K A, Chakraborty P, et al. Performance evaluation of efficient interpretable CNN-transformer model for redshift prediction[J]. Signal, Image and Video Processing, 2025, 19 (4): 326-326. https://doi.org/10.1007/s11760-025-03855-9.
16. Yang M, Song M, Guo Y, et al. Prediction of shield tunneling-induced ground settlement using LSTM architecture enhanced by multi-head self-attention mechanism[J]. Tunnelling and Underground Space Technology incorporating Trenchless Technology Research, 2025, 161106536-106536. https://doi.org/10.1016/j.tust.2025.106536.
17. Ding S, Chen R, Liu H, et al. ASG-HOMGAT: a high-order multi-head graph attention network with adaptive small graph structure for rolling bearing fault diagnosis[J]. Measurement Science and Technology, 2024, 35 (6). https://doi.org/10.1088/1361-6501/AD3293.
18. Zhang H, Wu Q, Tang W, et al. Acoustic Signal-Based Defect Identification for Directed Energy Deposition-Arc Using Wavelet Time–Frequency Diagrams[J]. Sensors, 2024, 24 (13): 4397-4397. https://doi.org/10.3390/S24134397.
19. Zhang Q, Zhang X, Yang J, et al. Introducing Euclidean distance optimization into Softmax loss under neural collapse[J]. Pattern Recognition, 2025, 162111400-111400. https://doi.org/10.1016/J.PATCOG.2025.111400.
20. Xiangzhi W, Chaojiang L, Hun G, et al. Electrical discharge machining of polycrystalline diamond: A review[J]. Proceedings of the Institution of Mechanical Engineers, 2023, 237 (11): 1611-1627. https://doi.org/10.1177/09544054221136511.
21. Koçillari L, Lorenz M G, Engel M N, et al. Sampling bias corrections for accurate neural measures of redundant, unique, and synergistic information[J]. bioRxiv: the preprint server for biology, 2024. https://doi.org/10.1101/2024.06.04.597303.
22. Layton W O, Peng S, Steinmetz T S. ReLU, Sparseness, and the Encoding of Optic Flow in Neural Networks[J]. Sensors, 2024, 24 (23): 7453-7453. https://doi.org/10.3390/S24237453.
23. Zhao W, Zhu X, Shi H, et al. Global Cross-Entropy Loss for Deep Face Recognition[J]. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society, 2025, PP. https://doi.org/10.1109/TIP.2025.3546481.
24. Wang Q, Xu F, Ma T. Wavelet packet decomposition with motif patterns for rolling bearing fault diagnosis under variable working loads[J]. Journal of Vibration and Control, 2025, 31 (3-4): 516-527. https://doi.org/10.1177/10775463241229512.
25. 25. Assri E N, Jallal A M, Chabaa S, et al. Enhancing building energy consumption prediction using LSTM, Kalman filter, and continuous wavelet transform[J]. Scientific African, 2025, 27e02560-e02560. https://doi.org/10.1016/J.SCIAF.2025.E02560.
26. Ping Y, Bin R L, Jie C, et al. Bearing fault diagnosis method for unbalance data based on Gramian angular field[J]. Journal of Intelligent & Fuzzy Systems, 2024, 47 (1-2): 45-54. https://doi.org/10.3233/JIFS-233797.
27. Adrien B, Valentin D, Benoît F. DT-SNE: t-SNE discrete visualizations as decision tree structures[J]. Neurocomputing, 2023, 529101-112. https://doi.org/10.1016/J.NEUCOM.2023.01.073.
28. Mandal K A, Dehuri S, Sarma D K P. Analysis of machine learning approaches for predictive modeling in heart disease detection systems[J]. Biomedical Signal Processing and Control, 2025, 106107723-107723. https://doi.org/10.1016/J.BSPC.2025.107723.
29. Zeng Q, Wang M, Zhang Y, et al. Machine learning-assisted development of TMDs-type gas-sensitive materials for dissolved gases in oil-immersed transformer oils[J]. Materials Today Chemistry, 2025, 44102583-102583. https://doi.org/10.1016/J.MTCHEM.2025.102583.
30. DevarasiddappaD, ChandrasekaranM, RavikumarM, et al. Modified teaching learning based optimization for maximization of MRR in wire-cut EDM of Ti6Al4V alloy for sustainable production[J]. AIP Conference Proceedings, 2019, 2128 (1): 040007. https://doi.org/10.1063/1.5117969.
31. Ayetiran F E, Özgöbek Ö. An inter-modal attention-based deep learning framework using unified modality for multimodal fake news, hate

speech and offensive language detection[J]. Information Systems, 2024, 123102378. https://doi.org/10.1016/J.IS.2024.102378.