Article

Fault Diagnosis of Rotating Machinery Bearings Based on Improved DCNN and WOA-DELM

Lijun Wang [1,*], Dongzhi Ping [1], Chengguang Wang [2], Shitong Jiang [1], Jie Shen [1] and Jianyong Zhang [3,*]

1 School of Mechanical Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450045, China; 201721704@stu.ncwu.edu.cn (D.P.); duoheleshui@163.com (S.J.); shenjie@ncwu.edu.cn (J.S.)
2 School of Management and Economics, North China University of Water Resources and Electric Power, Zhengzhou 450045, China; 201709502@stu.ncwu.edu.cn
3 School of Computing, Engineering & Digital Technologies, Teesside University, Middlesbrough TS1 3BA, UK
* Correspondence: wanglijun@ncwu.edu.cn (L.W.); j.zhang@tees.ac.uk (J.Z.)

Abstract: A bearing is a critical component in the transmission of rotating machinery. However, due to prolonged exposure to heavy loads and high-speed environments, rolling bearings are highly susceptible to faults, Hence, it is crucial to enhance bearing fault diagnosis to ensure safe and reliable operation of rotating machinery. In order to achieve this, a rotating machinery fault diagnosis method based on a deep convolutional neural network (DCNN) and Whale Optimization Algorithm (WOA) optimized Deep Extreme Learning Machine (DELM) is proposed in this paper. DCNN is a combination of the Efficient Channel Attention Net (ECA-Net) and Bi-directional Long Short-Term Memory (BiLSTM). In this method, firstly, a DCNN classification network is constructed. The ECA-Net and BiLSTM are brought into the deep convolutional neural network to extract critical features. Next, the WOA is used to optimize the weight of the initial input layer of DELM to build the WOA-DELM classifier model. Finally, the features extracted by the Improved DCNN (IDCNN) are sent to the WOA-DELM model for bearing fault diagnosis. The diagnostic capability of the proposed IDCNN-WOA-DELM method was evaluated through multiple-condition fault diagnosis experiments using the CWRU-bearing dataset with various settings, and comparative tests against other methods were conducted as well. The results indicate that the proposed method demonstrates good diagnostic performance.

Keywords: rotating machinery; convolutional neural network; fault diagnosis; Efficient Channel Attention Module; Bi-directional Long Short-Term Memory; DELM

1. Introduction

With industrial modernization, the direction of rotating machinery development has been toward large-scale, intelligent, high-precision, and high-efficiency. Rotating machinery usually continuously works at high speed under heavy loads. Rolling bearings are used to convert sliding friction between the shaft and shaft seat into rolling friction, as a result, rolling bearings have become one of the most prone to failure parts in rotating machinery and equipment. According to relevant statistics, 40% of motor failures come from bearings [1], and therefore, in order to ensure the reliable and safe operations of rotating machinery, the fault diagnosis of rolling bearings is of great importance.

The signals used in fault diagnosis methods are usually vibration signals, acoustic signals, current signals, speed signals, temperature signals, etc. The bearing fault diagnosis relies on various sensors. One of the effective methods is based on the signals from vibration sensors [2,3]. Many mechanical failures, such as local defects in rotating machinery. The vibration signal is manifested as a series of pulse events [4]. The fault diagnosis process is generally divided into two stages: feature extraction and fault classification. The time–frequency analysis can be used to extract the information contained in the signal in the time and frequency domains, and commonly the Short-Time Fourier Transform (STFT) [5], Fast

Fourier Transform (FFT) [6], Wavelet Transform (WT) [7], Variational Modal Decomposition (VMD) [8], and Ensemble Empirical Modal Decomposition (EEMD) [9], etc., are used in the feature extraction. Hou et al. [10] used EEMD to decompose the vibration signal to obtain the intrinsic modal components. Combining the permutation entropy eigenvectors of each modal component, the Linear Discriminant Analysis (LDA) method is used to process the entropy eigenvectors as the input of the clustering algorithm, which has the advantage of better intra-class clustering compactness, but the EEMD relies too much on expert experience in the decomposition process. He et al. [11] introduced the hybrid impact index (SII) as a new metric to evaluate the fault components in the VMD method and the optimal parameters of VMD were selected using an artificial bee colony algorithm. The models that can be used for fault classification include the support vector machine (SVM) [12], K-Nearest Neighbor (K-NN) [13], Artificial Neural Network (ANN) [14], etc.; Deng et al. [15] optimized the least squares support vector machine (LS-SVM) parameters using the Particle Swarm Optimization algorithm (PSO) to improve the classification accuracy. Lu et al. [16] proposed a case-based reconstruction algorithm to adaptively locate the nearest neighbors of each test sample, which can achieve the classification of bearing faults using both parameters and cases.

In recent years, research on deep learning has received more and more attention from scholars, and its applications in object recognition, image segmentation, speech recognition, machine health detection, and medical health diagnosis have become more widespread [17–19]. Traditional machine learning architectures are simple and difficult to automatically extract the information carried by features in higher-order samples, and only through feature engineering that relies on expert experience may more desirable classification results be obtained. Therefore, deep learning methods are more often used for bearing fault diagnosis of rotating machinery in production practice, and the commonly used deep learning methods are convolutional neural networks (CNN) [20], Deep Belief Network (DBN) [21], and Generative Adversarial Network (GAN) [22]. Jiang et al. [23] improved the feature learning capability through a layered learning structure for convolution and pooling layers by taking the multiscale characteristics of gearbox vibration signals into consideration. Gong et al. [24] proposed an improved convolutional neural network support vector machine, in which the raw data from multiple sensors were directly input to the CNN-Softmax model, and the extracted feature vectors were input to the support vector machine for fault classification. The results were better than those achieved using the SVM and K-nearest neighbor methods. Deng et al. [25] proposed a Multi-Swarm Intelligence Quantum Differential Evolution (MSIQDE) algorithm to optimize the DBN parameters to avoid premature convergence and improve the global search capability, and the experimental results showed that higher classification accuracy was achieved using the MSIQDE-DBN than other comparative methods. Zhang et al. [26] used CNN to extract features from data and then combined them with the Long Short-Term Memory (LSTM) neural network to process time series data. The Sparrow Search Algorithm (SSA) is used to optimize the parameters of the Long Short-Term Memory (LSTM) neural network and improve the accuracy and feasibility of fault diagnosis. Chen et al. [27] proposed a fault diagnosis method combining a convolutional neural network (CNN) and Extreme Learning Machine (ELM), in which, firstly, the original vibration signal was processed using the continuous wavelet transform, then advanced features were extracted with the CNN. The classification performance was improved by using the ELM as the classifier, the proposed method was able to detect different fault types and the classification accuracy was higher than other methods. Chen et al. [28] proposed a novel fault diagnosis method for rolling bearings based on hierarchical refined composite multiscale fluctuation-based dispersion entropy (HRCMFDE) and PSO-ELM. This method solves the problem of missing high-frequency signals in the process of coarse-grained decomposition and improves the anti-interference ability and computational efficiency. The extracted feature vectors can effectively describe the fault information. Finally, the PSO-ELM classifier is used to classify the fault characteristics. Experimental results show that this method has high recognition

accuracy and good load migration effect. Zhou et al. [29] proposed a new adversarial Generative Adversarial Network (GAN) generator and per-discriminator, which was improved by extracting fault features through an Auto-Encoder (AE) instead depending on fault samples., and the generator training was enhanced from the original statistical overlap to the fault features and diagnostic result errors guided model. The experimental results proved the effectiveness of Zhou's method. Mao et al. [30] used the spectral data obtained through the original signal processing as the input to the GAN, which generated synthetic samples with fewer fault classes according to the real samples. Such synthetic samples were used as the training set, and the experiment results demonstrated that the improved generalization ability was achieved.

Based on the investigation of the above research, a rotating machinery fault diagnosis method based on an improved deep convolutional neural network (DCNN) and Whale Optimization Algorithm (WOA) optimized Deep Extreme Learning Machine (DELM) is proposed in this paper. The proposed method enhances the ability of the DCNN networks to extract important features while leveraging the excellent generalization ability of the WOA-DELM model. It addresses the issues related to poor feature extraction and low diagnostic accuracy in traditional convolutional neural networks due to feature masking caused by background noise under varying operational conditions. The main contributions of this paper are as follows:

1. An improved DCNN (IDCNN) classification network with Bi-directional Long Short-Term Memory (BiLSTM) and Efficient Channel Attention Net (ECA-Net) is constructed. The BiLSTM added to DCNN can extract the deep features of the data based on the timing information. The ECA-Net in the DCNN is introduced to weight different features. Therefore, IDCNN can make expressive features play a greater role.
2. The initial weights of the first input layer of DELM are optimized by WOA to effectively improve the overall stability of DELM. Using WOA-DELM as a classifier the classification accuracy as well as the generalization performance are improved.
3. The IDCNN-WOA-DELM fault diagnosis method proposed in this paper is experimentally validated using multiple bearing data sets and its effectiveness and generalization capability for bearing fault diagnosis have been verified by comparing it with other network models.

2. Theoretical Derivation

2.1. DCNN

The deep convolutional neural network is a class of the feedforward neural network with convolutional computation and deep structure, which is one of the representative algorithms of deep learning. The convolutional layer and pooling layer are feature extraction layers, which can be set alternately. The structure of the classical DCNN network is shown in Figure 1.
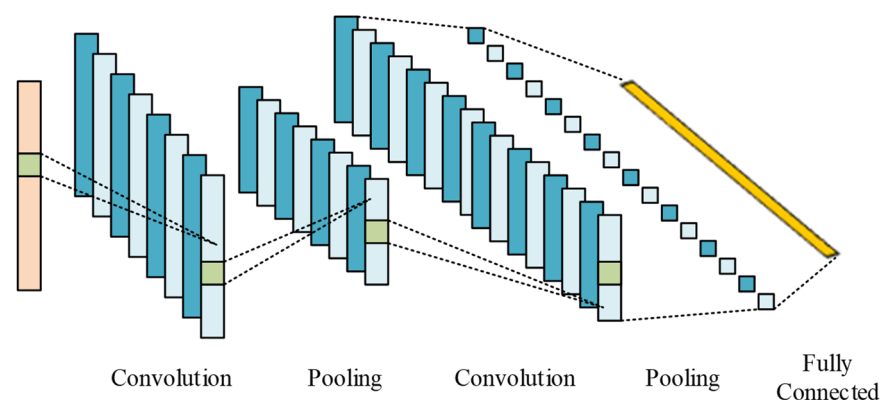
Figure 1. One dimensional-DCNN structure.

In the convolutional layer, a certain size of convolutional kernels is used to convolve local regions of the input features. Each kernel convolution represents a feature map, multiple feature surfaces are output by the nonlinear activation function, and the same input feature surface and the same feature output surface use the same convolutional kernel, so as to achieve a weight-sharing network structure. The mathematical model of the convolutional layer is expressed as follows:

$$x_n^l = f\left(\sum_{i=1}^{M} x_n^{l-1} \times k_{ni}^l + b_n^l\right) \tag{1}$$

In this equation, x_n^l is the nth feature mapping of the lth layer; $f(\cdot)$ is the activation function; M is the number of input feature mappings; x_n^{l-1} is the number of nth feature mappings of the $l - 1$th layer; k_{ni}^l is the trainable convolution kernel; and b_n^l is the bias.

The pooling layer is a downsampling layer, and the size of the input matrix will change in this layer, but not the depth of the matrix. The pooling layer can be used to reduce the number of nodes for the fully connected layer, which, in the neural network training parameters optimization, has a certain role to play. The pooling layer has no parameter, so there is no need for weight updates. The mathematical model of the pooling layer is expressed as follows:

$$x_n^l = f\left(\beta_n^l down\left(x_n^{l-1} + b_n^j\right)\right) \tag{2}$$

In this formula down() represents the subsampling function.

The fully connected layer is a traditional feedforward neural network in which the neurons of the fully connected layer are connected to the neurons of the upper layer, and the features that have been convolved and pooled can continue to be expressed nonlinearly, and the inputs of the fully connected layer are all one-dimensional feature vectors. The mathematical model expression of the fully connected layer is as follows:

$$a^l = \sigma\left(W^l a^{l-1} + b^l\right) \tag{3}$$

a^l is the output of the fully connected layer, W^l is the weight factor, l is the network layer sequence number, b^l is the bias, a^{l-1} is the unfolded one-dimensional vector, $\sigma(\cdot)$ is the activation function, and the classification task usually uses the Softmax function.

2.2. DELM

ELM, proposed by Huang et al. [31], is a new training model for Single-hidden Layer Feedforward Networks without iterative tuning its network structure of the extreme learning machine consists of an input layer, an implicit layer, and an output layer. Since it contains only one implicit layer, its generalization ability is better than the classical neural network model, and in its training process, the learning parameters in the hidden nodes are randomly chosen, which are not required to be adjusted. The output weights are obtained through generalized inverse operation, and only the number of hidden nodes needs to be determined, which is no longer propagated backward during the training process. Compared with traditional deep learning models, the training speed of the ELM is significantly improved, and the model has better generalization ability.

The mathematical model of the ELM output with a sample set of x samples and l number of hidden layers is expressed as follows:

$$Y_j = \sum_{i=1}^{l} \beta_i g\left(w_i x_j + b_i\right), j = 1, 2, \cdots, N \tag{4}$$

In the formula, l is the number of hidden nodes; w_i and β_i are the input weight and output weight between the nodes of the ith hidden layer and the output layer, respectively;

b_i is the threshold of the ith hidden neuron. $g(\cdot)$ is the activation function. N is the arbitrary number of samples.

By setting $w_i \cdot x_j$ to be the inner product of w_i and x_j, the above equation can be written as follows:

$$H\beta = T \tag{5}$$

where H is the output matrix of the hidden layer, and T is the desired output.

In order to make the error between the output and the desired output close to 0, let the network cost function $\|Y - T\|$, approach minimal. Based on the ELM theory, the learning parameters of the hidden nodes can be generated randomly without considering the input data, and hence the above equation can be changed into a linear function, and the output weights can be determined by the least squares method. The mathematical model is as follows:

$$\hat{\beta} = H^+ T \tag{6}$$

H^+ is the Moore-Penrose generalized inverse matrix of H.

The basic unit in the DELM is the ELM-AE, which is a combination of the ELM and an Auto-Encoder (AE), which can be seen as connecting multiple ELMs into an instrument. The structure of DELM is shown in Figure 2. This allows a more comprehensive extraction of mapping relationships between data, exhibiting better performance for processing high-dimensional and nonlinear data. The mathematical model of DELM is as follows:

$$Y_j = \sum_{j=1}^{L} \sum_{k=0}^{Z} \beta_j^k g_{jk} \left(\sum_{p=1}^{n} w_P^{ih} x_i + b_P \right), j = 1, 2, \cdots, Q \tag{7}$$

where L is the number of hidden layer neurons; Z is the number of derived neurons corresponding to the hidden neuron, β_j^k is the weight vector between the jth hidden layer neuron and the output layer; g_{jk} is the kth order derivative of the implicit layer neuron activation function of the jth implicit layer neuron activation function; n is the number of input layer neurons; w_j^{ih} is the weight vector between the input layer and the jth hidden layer neuron, and b_j is the jth hidden layer node bias. Q is the number of training data sets.
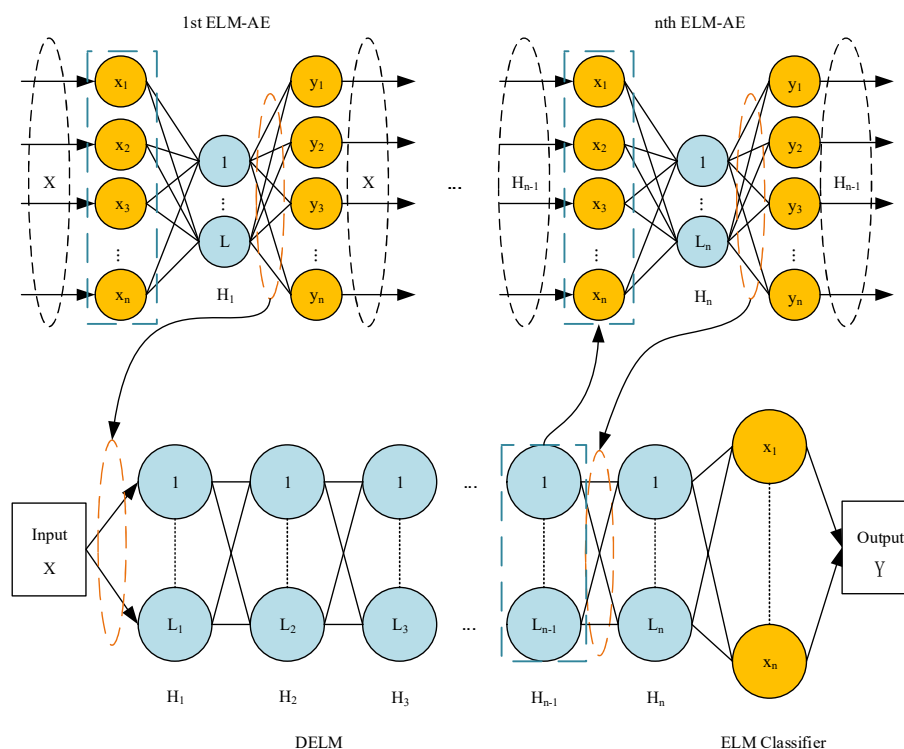

Figure 2. DELM structure.

The weight of the DELM input layer is an orthogonal random matrix randomly generated in the first ELM-AE pre-training stage. In DELM, the least square method can only adjust the weight parameter of the output layer. The weight of the hidden layer must be obtained through iteration. The input weight of each ELM-AE in DELM will affect the final DELM effect.

2.3. ECA-Net

ECA-Net is proposed by Wang et al. [32], which is a local cross-channel interaction network without dimensionality reduction. It can be used to reduce model complexity while maintaining performance. The attention mechanism is a method to optimize deep learning models by simulating the attention mechanism of the human brain, in the ECA-Net, mainly the SE-net module is improved so that the ECA-Net can adaptively select the size of one-dimensional convolution kernels. Only a few parameters need to be added into the model, but obvious performance gains have been achieved. The structure of ECA-Net is shown in Figure 3.

The ECA module is implemented by fast one-dimensional convolution of size K, where K represents the coverage of local cross-channel interactions. K is related to the channel dimension C, and the larger the channel dimension C, the stronger the long-term interaction. The mapping between K and C is shown below:

$$C = \Phi(K) \tag{8}$$

When the channel dimension is given, the size of K can be determined adaptively.

$$C = \Phi(K) = \left| \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{9}$$

where $|t|_{odd}$ represents the most recent odd number; $\gamma = 2$; $b = 1$.
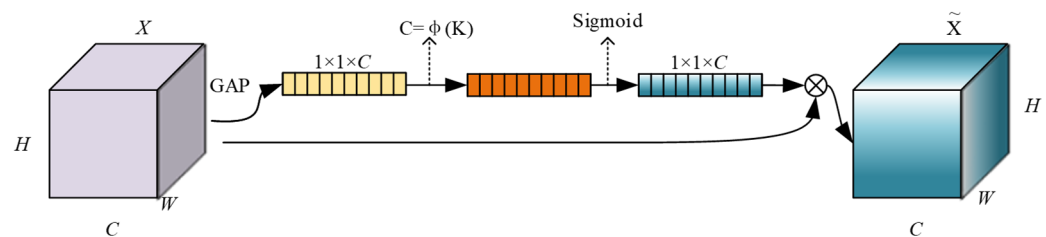
Figure 3. Structure of ECA-Net.

2.4. BiLSTM

LSTM was developed by Hochreiterand Schmidhuber from Recurrent Neural Network (RNN) [33] in 1998. The structure of LSTM is shown in Figure 4. LSTM has a special gating mechanism, so it can learn long-term dependencies between two sequences and has long-term memory. The key to LSTM is the transmission of information from C_{t-1} to C_t and the selective retention of desired features in the process. LSTM unit includes a forget gate f_t, input gate i_t and output gate o_t. The mathematical expression of the gates is shown below.

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{10}$$

$$f_t = \sigma\left(w_f[h_{t-1}, x_t] + b_f\right) \tag{11}$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{12}$$

where σ is the activation function, w_x is the weight of the corresponding gate, h_{t-1} is the

output of the previous LSTM unit, x_t is the input at the current time, and b_x is the bias of the corresponding gate.
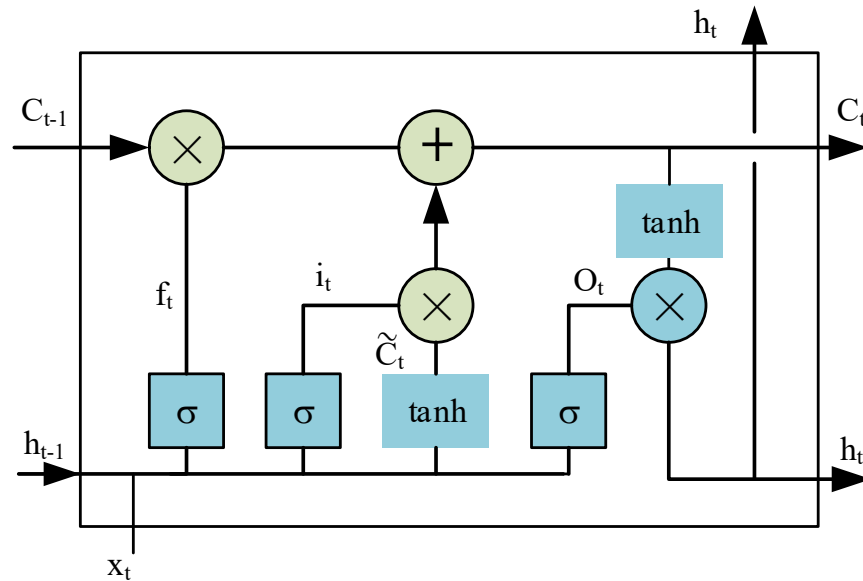
Figure 4. Structure of LSTM.

The formula for calculating the cell state C_t and hidden layer state h_t of LSTM unit is as follows:

$$C_t = f_t * C_{t-1} + i_t * \tanh(w_c[h_{t-1}, x_t + b_c]) \tag{13}$$

$$h_t = o_t * \tanh(C_t) \tag{14}$$

where w_c is the weight coefficient matrix of the current input cell state.

BiLSTM is composed of a forward LSTM and a reverse LSTM. BILSTM makes use of known time series and reverse position series, and deepens the feature extraction of the original sequence through forward and back propagation bi-directional operation. The final output of the BiLSTM neural network is the sum of LSTM output results propagated forward and back. The structure of BiLSTM is shown in Figure 5.
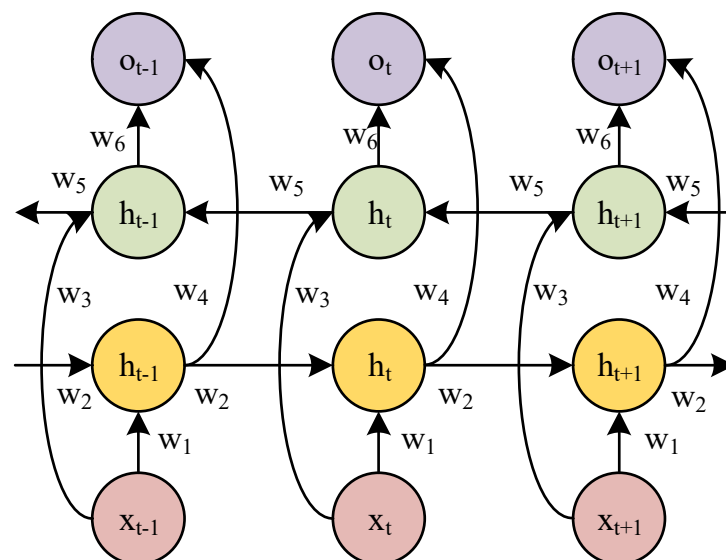
Figure 5. Structure of BiLSTM.

Forward calculation is performed from time 1 to time t in the forward layer to obtain and save the output $\overrightarrow{h_t}$ of the forward hidden layer at each time. The backward calculation is performed at the backward layer along time t to time 1 to obtain and save the output $\overleftarrow{h_t}$ of the backward hidden layer at each time. Finally, the final output is obtained by combining the output results O_t of the corresponding moments of the forward layer and backward layer at each moment. The calculation formula is as follows:

$$\overrightarrow{h_t} = f(w_1 x_t + w_2 \overrightarrow{h}_{t-1}) \tag{15}$$

$$\overleftarrow{h_t} = f(w_3 x_t + w_5 \overleftarrow{h}_{t+1}) \tag{16}$$

$$O_t = g(w_4 \overrightarrow{h_t} + w_6 \overleftarrow{h_t}) \tag{17}$$

2.5. WOA

The whale optimization algorithm (WOA) algorithm is a novel natural heuristic optimization algorithm proposed by Mirjalili et al. [34], and its main idea is to mimic the unique behaviors and predation strategy adopted by humpback whales when exploring prey, shrinking the encirclement, and spiral prey location updating.

At the exploration stage, whales randomly look for prey according to each other's location during the search process, but the prey location is generally unknown, so whales need to update the position according to their own position, and the update equation is as follows:

$$D = C \cdot X^*(t) - X(t) \tag{18}$$

$$X(t+1) = X^*(t) - A \cdot D \tag{19}$$

The current position of a whale is denoted by $X(t)$, while X^* refers to the position of a randomly selected whale. The distance between the current individual and the randomly selected individual whales is represented by D. A is the coefficient vector, which is mathematically expressed as follows:

$$A = 2\overrightarrow{a}r_1 - \overrightarrow{a} \tag{20}$$

$$C = 2r_2 \tag{21}$$

where r_1 and r_2 are random vectors from 0 to 1. $\overrightarrow{a}$ is a vector that decreases linearly from 2 to 0.

$$\overrightarrow{a} = 2 - \frac{2t}{T_{\max}} \tag{22}$$

where $T_{\max}$ is the maximum number of iterations.

There are two predation modes The first mode is to narrow down the search. It can be described using the following mathematical expression:

$$D = C \cdot X_{best}(t) - X(t) \tag{23}$$

$$X(t+1) = X_{best}(t) - A \cdot D \tag{24}$$

At this point, all individuals move towards the position with the best fitness value, thus forming a contraction surround.

The second mode uses a spiral equation to update the whale's position based on the prey's location, and its mathematical expression is shown below:

$$X(t+1) = D_{best} \cdot e^{bl} \cdot \cos(2\pi l) + X(t) \tag{25}$$

$$D = C \cdot X_{best}(t) - X(t) \tag{26}$$

In Formulas (25) and (26), t is the current number of iterations, X is the current coordinate vector of the whale, X_{best} is the location of the prey found by the whale, b is the logarithmic spiral shape constant, l is a random number between -1 and 1, and D is the distance between the humpback whale and the prey.

The above two processes are synchronized in the actual predation process, and therefore it requires setting the probability control to determine the strategy to choose between the two methods, which can be explained with Equation (27).

$$X(t+1) = \begin{cases} X_{best} - AD, p < 0.5 \\ X(t) + D_{best} \cdot e^{bl} \cdot cos(2\pi l) \cdot p \geq 0.5 \end{cases} \tag{27}$$

The real behavior of humpback whales is simulated by assigning equal probabilities of 50% to each of the two methods. The search is judged to be over when the number of iterations is maximized.

3. The Proposed Bearing Fault Diagnosis Method Based on IDCNN Feature Extraction and WOA-DELM

In rotating machinery, bearing fault diagnosis under variable conditions is often a challenging task due to the high noise environment, heavy load, and high speed. Such a task requires sufficient expertise and abundant experience. To address this issue, a fault diagnosis method based on a combination of IDCNN and WOA-optimized DELM is proposed in this paper. This method aims to improve the feature extraction capability of the convolutional neural network and to leverage the excellent classification capability and stable global dynamic search capability of the WOA-DELM. The overall approach is presented in the following roadmap in Figure 6.

As shown in Figure 6, the specific troubleshooting process is as follows:

Step 1: The bearing vibration signal is collected from the experimental platform, and the collected one-dimensional vibration signal is cut into small segments of 1×2048, and the training set and testing set are selected from the small segments to simulate the working conditions under variable working conditions.

Step 2: The IDCNN model is trained using the training set and the deep convolutional neural network combined with BiLSTM and ECA-Net is used to mine the deep features of the data so that the features with expressive power in the samples can play a greater role. The extracted fault features are then fed into the WOA-DELM model for training.

Step 3: The testing set samples are imported into the optimal IDCNN network, and the results of the fully connected layer are output as a new testing set into the trained WOA-DELM classifier. Then the diagnostic results and various evaluation indexes are combined to illustrate the effectiveness of the model. The traditional feedforward neural network using gradient descent iterative algorithm to adjust the weight parameters will lead to slow training speed, poor generalization ability, and more training parameters, which affect the effectiveness of the feedforward neural network. Therefore, using the WOA-DELM classifier instead of the commonly used Softmax classifier can effectively improve the accuracy, efficiency, and generalization ability of classification. The fault features extracted from the IDCNN network are input into the DELM classifier for classification. The features input to the DELM classifier is trained using the WOA to find the optimal weights and bias parameters for the DELM classifier to improve the diagnostic performance of the model.
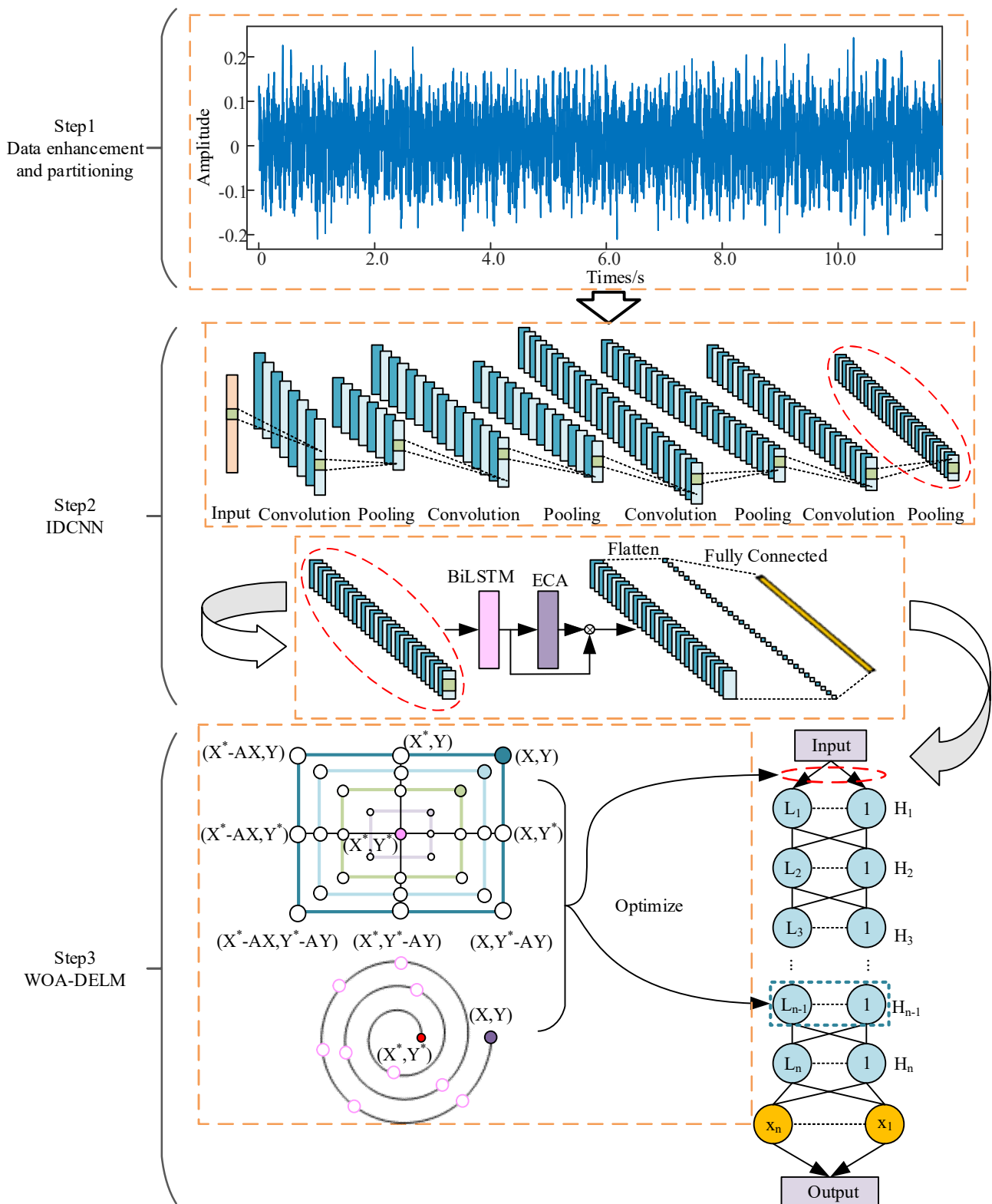
Figure 6. Overall technical roadmap for the proposed methodology.

WOA–DELM Classifier

WOA is used to optimize the original randomly generated weights of the first input layer in the DELM model. The optimization of DELM using the WOA is shown in Figure 7. The specific optimization steps are as follows:

Step 1: Set the number of ELM-AEs as 180, the activation function is Sigmoid, the input weights are w_i, and the hidden layer bias is b_i.

Step 2: Initialization operation of the WOA parameters. The population size is set to 80 and the number of iterations is set to 20.

Step 3: Initialize the individual whale with the randomly generated x, x from DELM as the initial position vector.

Step 4: Set the fitness function, which in this paper is set as the error rate of the training set, and calculate the individual fitness values in the initialized population to obtain the optimal individual.

Step 5: After randomly generating p values, $|A|$ and p are used jointly to decide how to determine the formula for the location update. When $|A| \geq 1$, Formula (19) is chosen. When $|A| < 1$, Formula (27) is chosen in combination with the probability p.

Step 6: Recalculate the fitness values and find a better solution.

Step 7: Examine whether the WOA algorithm meets the termination condition, and outputs the optimal result if the termination condition is met, otherwise repeat the above steps according to the set number of iterations.

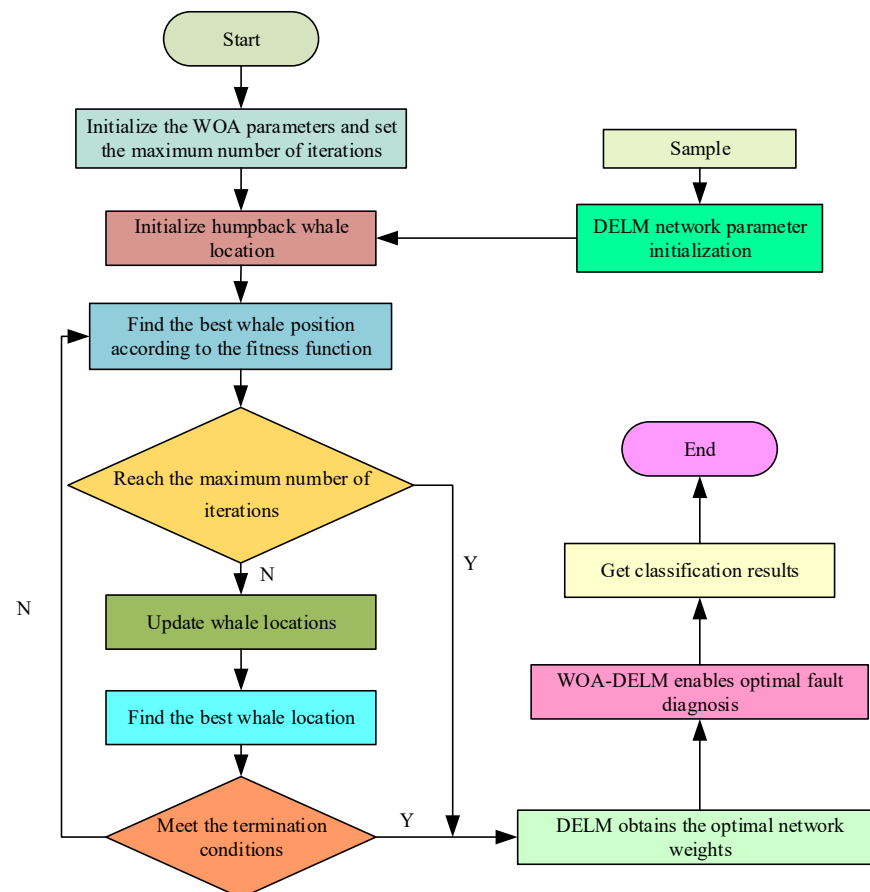Step 8: Enter the search parameters in the WOA-DELM model to start fault diagnosis.

Figure 7. Flow chart of WOA-DELM.

4. Experiments and Analysis

This paper uses a bearing experimental dataset to validate the proposed method. The experimental data are a publicly available dataset from Case Western Reserve University.

The computer hardware environment was configured with Windows 11 operating system, CPU i5-12400F@2.5 GHz, and GPU Nvidia GeForce RTX 3060. The program deep learning framework was built and run in Tensorflow using Python 3.7 and the WOA-DELM in Matlab R2018b.

4.1. Data Description and Processing

The CWRU (Case Western Reserve University) dataset, which is a commonly used dataset for bearing fault diagnosis, consists of a motor, torque sensor, and dynamometer in its experimental platform. The experimental platform of CWRU bearing is shown in Figure 8. It encompasses four different types of bearing failures, including inner ring damage, outer ring damage, ball failure, and normal operation, with failure diameters of 0.007 inches, 0.014 inches, and 0.028 inches, respectively. Furthermore, outer ring damage is placed in three positions: 3 o'clock, 6 o'clock, and 12 o'clock. The drive end (DE) features two sampling frequencies of 12 KHz and 48 KHz, while the fan end (FE) has only a 12 KHz sampling frequency. Each fault type contains different bearing operating states, comprising four different loads of 0, 1, 2, and 3 hp, as well as four different speeds of 1797, 1772, 1750, and 1730 rpm. The CWRU (Case Western Reserve University) dataset, which is a commonly used dataset for bearing fault diagnosis, consists of a motor, torque sensor, and dynamometer in its experimental platform. It encompasses four different types of bearing failures, including inner ring damage, outer ring damage, ball failure, and normal operation, with failure diameters of 0.007 inches, 0.014 inches, and 0.028 inches, respectively. Furthermore, outer ring damage is placed in three positions: 3 o'clock, 6 o'clock, and 12 o'clock. The drive end (DE) features two sampling frequencies of 12 KHz and 48 KHz, while the fan end (FE) has only a 12 KHz sampling frequency. Each fault type contains different bearing operating states, comprising four different loads of 0, 1, 2, and 3 hp, as well as four different speeds of 1797, 1772, 1750, and 1730 rpm.
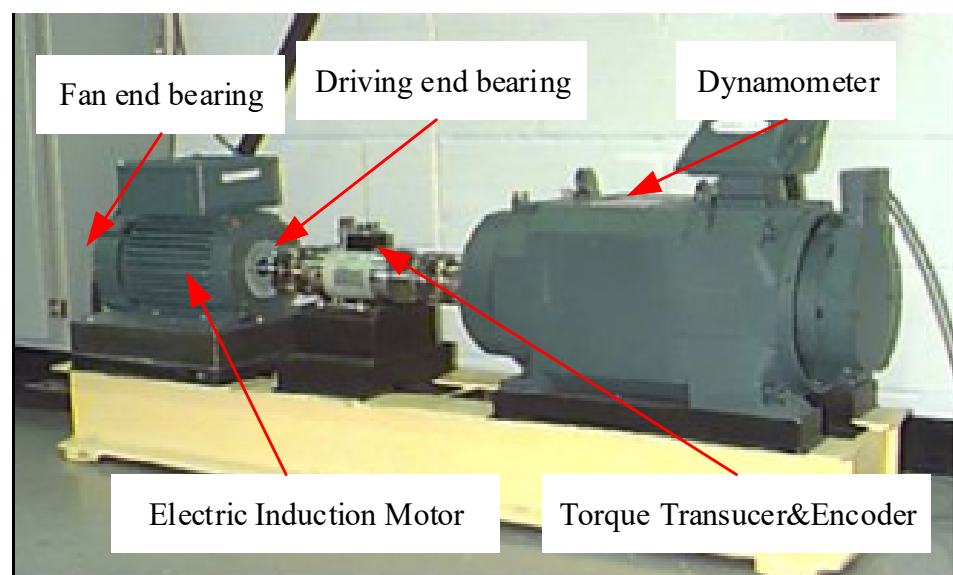
Figure 8. The experimental platform of CWRU bearing.

The data used in experiments 1 and 2 are sampled at 12 KHz frequency at the DE end, with speeds of 1797 rpm, 1772 rpm, 1750 rpm, and 1730 rpm. Different speeds correspond to 0 hp, 1 hp, 2 hp, and 3 hp under four kinds of load in different fault positions, including inner ring, ball, and outer ring fault states with a fault size of 0.014, as well as the normal state. The above data are divided into data sets A, B, C, and D according to the speed and load, with each data set including 4 different working conditions. Each working condition has 200 samples. The original vibration signal of each health state contains 121,048 points. In order to avoid overfitting due to the small amount of data, the data are enhanced by overlapping sampling. Each sample contains 2048 points, and each health condition yields 200 samples, totaling 800 samples for each data set. The samples of each data set are provided in Table 1.

Table 1. Sample of data set for experiments 1 and 2.

Name	Motor Speed/rpm	Motor Load/(hp)	Fault Diameter (Inches)	Status	Label	Sample
A	1797	0	0.014	Normal IF BF OF	0, 1, 2, 3	800
B	1772	1	0.014	Normal IF BF OF	0, 1, 2, 3	800
C	1750	2	0.014	Normal IF BF OF	0, 1, 2, 3	800
D	1730	3	0.014	Normal IF BF OF	0, 1, 2, 3	800

The data used in Experiment 3 is also sampled at 12 KHz frequency at the DE end. Experiment 3 was carried out under four conditions which are 1797 rpm/0 hp, 1772 rpm/1 hp, 1750 rpm/2 hp, and 1730 rpm/3 hp. In Experiment 3, three fault types of inner race fault, outer race fault, and ball fault were selected. Each fault type included 0.07 inches, 0.014 inches, and 0.021 inches fault sizes, respectively. Therefore, the data set used in Experiment 3 contains nine different fault conditions and one normal condition. After overlapping sampling, each fault type includes 300 samples The data set used in Experiment 3 is shown in Table 2.

Table 2. Sample of data set for Experiment 3.

Label	Status	Fault Diameter (Inches)	Number of Training Samples	Number of Testing Samples
0	Normal	-	300	128
1		0.007	300	128
2	Ball fault	0.014	300	128
3		0.021	300	128
4		0.007	300	128
5	Inner race fault	0.014	300	128
6		0.021	300	128
7		0.007	300	128
8	Outer race fault	0.014	300	128
9		0.021	300	128

This study comprises three experiments, denoted Experiment 1, Experiment 2, and Experiment 3, respectively. Experiment 1 involves the use of data from two distinct working conditions as the training set, and data from a third, distinct working condition as the testing set. The sample numbers of the training set and testing set are 1600 and 686, respectively. The training set and the testing set used in Experiment 2 are, respectively, from two different working conditions. The sample numbers of the training set and testing set are 800 and 343, respectively. Since the situation of 0 load rarely occurs during the actual operation of bearings, Experiment 2 does not consider the situation of 0 hp.

4.2. Model Parameter Setting

The structure of the IDCNN feature extraction is shown in Table 3. There are 4 convolutional layers and 4 maximum pooling layers in the model. The sizes of the four convolutional filters are 16, 32, 64, and 32. The sizes of the convolutional layers are 64, 3, 3, and 3. The step sizes are 16, 1, 1, and 1, respectively. The four maximum pooling layers are behind the four convolutional layers. The kernel size and step size are both 2. The Bilstm is set after the fourth maximum pooling layer to extract the bi-directional deep features of the positive and negative extraction time series data. The ECA-Net is set after the BiLSTM layer, and this feature extraction layer multiplies the features after the convolutional pooling layer with its feature weight matrix after the ECA attention mechanism completes to achieve feature weighting. Except for the activation function of the ECA module, which uses Sigmoid, the rest of the activation functions use the ReLU function. The parameters of the WOA-DELM classifier are set as follows: the sum of the error rates of the training and testing sets is

used as the fitness function during the training iterations, and the fixed parameters are calculated at the end of the optimization in the final accuracy. The hidden layer of ELM-AE is 180, while the number of populations is 80, and the number of iterations is 20; moreover, the activation function is chosen as Sigmoid.

Table 3. Structural parameters of IDCNN.

Name	Kernal Size	Stride	Input Size	Output Size	Activation Function
Covn1	64	16	2048×1	128×16	ReLU
Max pool1	2	2	128×16	64×16	-
Covn2	3	1	64×16	64×32	ReLU
Max pool2	2	2	64×32	32×32	-
Covn3	3	1	32×32	32×64	ReLU
Max pool3	3	2	32×64	16×64	-
Covn4	3	1	16×64	16×64	ReLU
Max pool4	3	2	16×64	8×32	-
BiLSTM	-	-	8×32	8×32	
ECA	-	-	8×32	8×32	Sigmoid
Flatten	-	-	8×32	256	-
Fc	-	-	256	128	ReLU

4.3. Analysis of Experiment 1

To evaluate the classification effect, the accuracy and F1 Score are usually used as criteria, which are calculated based on multiple experiment results so that the probability of wrong conclusions can be reduced when dealing with data containing random errors.

The following basic concepts are included in the evaluation metrics: TP (True Positives) indicates that positive cases are classified as positive cases; FP (False Positives) means that negative cases are classified as positive cases; FN (false Negatives) depicts that positive cases are classified as negative cases; TN (True Negatives) represents that negative cases are classified as negative cases.

Accuracy is the ratio of the number of correctly classified samples to the total number of samples. The formula is as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{28}$$

F1 Score is the summed average of precision and recall. The formula is as follows:

$$F1 = 2 \frac{\frac{TP}{TP+FP} \times \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \tag{29}$$

In Experiment 1, each data set in Table 1 is used. Both IDCNN and 1D-DCNN use Softmax as the classifier The network structure of the above two methods is the same as the network structure selected in this paper. Ten experiments are used during the experiments to reduce random errors.

In Experiment 1, the data from two conditions were used as the training set, and the data from the other conditions as the testing set. The average accuracy of ten experiments is shown in Figure 9. It can be seen from the experimental results that in 12 groups of experiments, the lowest average accuracy reached 96.65%, and the highest accuracy reached 99.85%. The proposed method shows high accuracy and stability.

Accuracy of two conditions training set and one condition testing set (%)
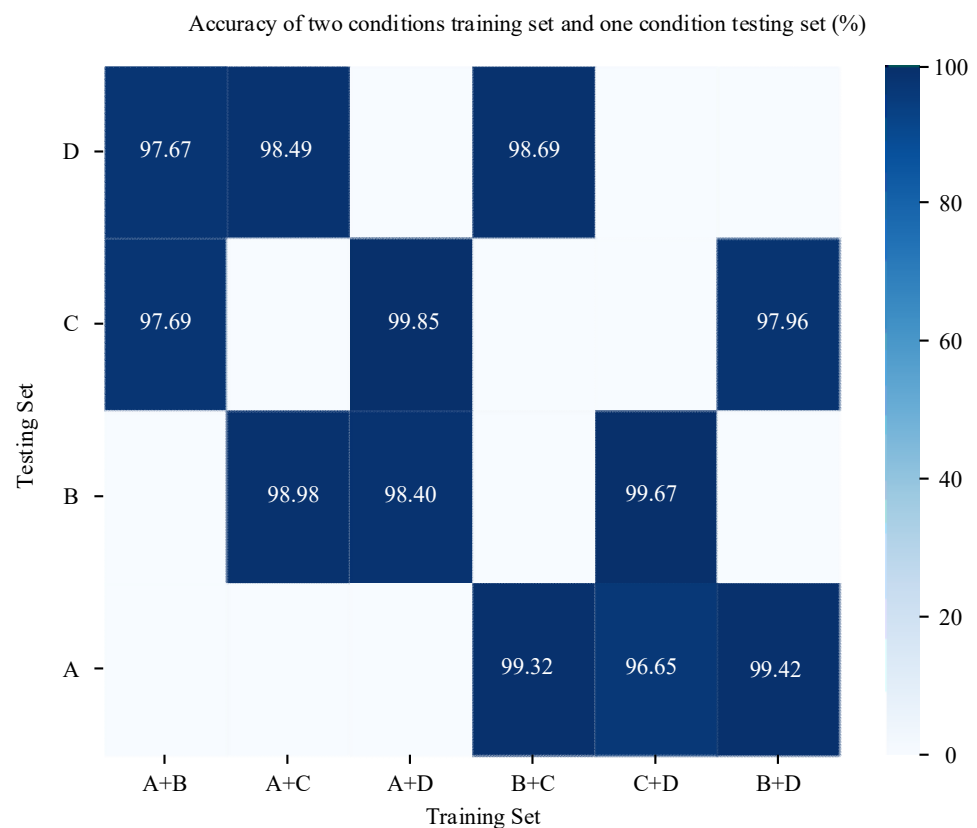

Figure 9. The accuracy of fault diagnosis when the data of two working conditions is used as the training set and the data of another different working condition is used as the testing set.

Four of the twelve groups were selected. They are numbered Tasks 1–4. The other two methods were used to further verify the effectiveness of the proposed method. The sample situations of Tasks 1–4 are shown in Table 4.

Table 4. Sample selection for Tasks 1–4.

Task Number	Train	Sample	Test	Sample
1	A, B	800, 800	C	686
2	A, C	800, 800	D	686
3	B, C	800, 800	A	686
4	C, D	800, 800	B	686

One experiment was randomly selected from each of the four tasks, and its accuracy and F1 Score are shown in Figure 10. In the four tasks, all the indicators were higher than 97%, among which the accuracy and F1 Score of Task 4 reached 100%. The result of Task 1 was relatively poor, but the accuracy and F1 Score still reached 97.96%. The average accuracy and average F1 Score of the ten experiments are shown in Table 5. The average accuracy and average F1 Score of the proposed method in four tasks are significantly higher than those of the other two methods, which proves the effectiveness of the proposed method in fault classification and feature extraction. The training time and testing time of the three methods are shown in Table 6. The total time of the proposed method is longer than that of the other two methods, but it is within the acceptable range. This proves that the modified method can improve diagnostic performance even with a small increase in training and testing time.
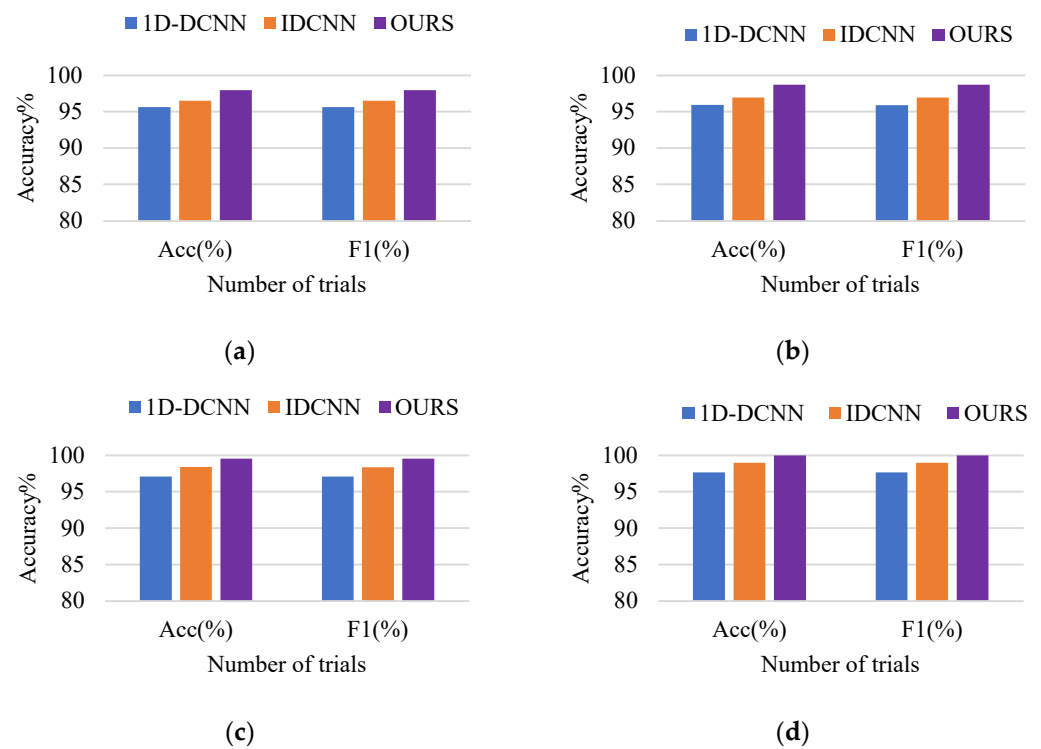
Figure 10. Results of trials; (**a**) results for Task 1; (**b**) results for Task 2; (**c**) results for Task 3; (**d**) results for Task 4.

Table 5. Average evaluation index for different methods in Task 1, 2, 3, 4.

Task Number	1D-DCNN Average Accuracy/F1 Score	IDCNN Average Accuracy/F1 Score	OURS Average Accuracy/F1 Score
1	95.44%/95.44%	96.32%/96.33%	97.69%/97.68%
2	95.68%/95.68%	96.46%/96.47%	98.49%/98.49%
3	96.85%/96.85%	98.19%/98.20%	99.32%/99.30%
4	97.42%/97.42%	98.80%/98.80%	99.67%/99.66%

Table 6. Average time for different methods in Task 1, 2, 3, 4.

Method	Time	Task 1	Task 2	Task 3	Task 4
1D-DCNN	Average Training Time (s)	12.42	11.89	12.30	12.04
	Average Testing Time (s)	0.36	0.45	0.39	0.37
IDCNN	Average Training Time (s)	20.87	20.42	20.36	20.21
	Average Testing Time (s)	1.04	1.03	1.01	1.04
OURS	Average Training Time (s)	20.99	20.17	20.55	20.18
	Average Testing Time (s)	10.32	10.45	10.14	10.28

The convergence curves are shown in Figure 11. The WOA-DELM classifier achieves the optimal parameter values within 20 iterations for the task types in Table 5. The proposed method has been compared with the 1D-DCNN and IDCNN methods to show its advantages in terms of classification efficiency for rotating machine bearing faults, which reflects the excellent parameter-finding ability of the WOA-DELM classifier.
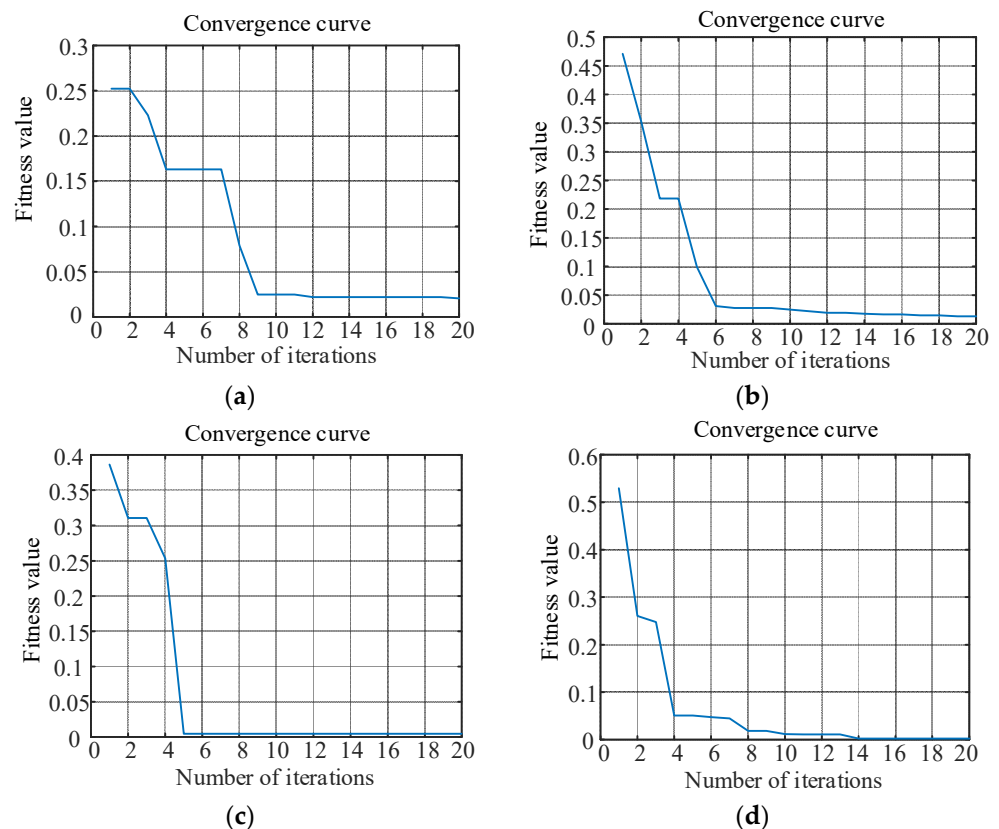
Figure 11. Convergence curves for the four tasks: (**a**) convergence curve for Task 1; (**b**) convergence curve for Task 2; (**c**) convergence curve for Task 3; (**d**) convergence curve for Task 4.

In order to demonstrate the better feature extraction ability of the improved convolutional neural network, the testing set samples in the input model for the four tasks and the output samples of the fully connected layer are visualized and analyzed using the t-SNE method. Figure 12 shows the data distribution of the input testing set samples of one of the four task types and the distribution of the testing set samples of the output of the fully connected layer. It can be seen that In the four tasks, there is still some confusion about individual features but the results obtained using the proposed method in Task 2 show good diagnostic and generalization capabilities.

4.4. Analysis of Experiment 2

In Experiment 2, each data set in Table 1 has been used. A choice of 800 samples were selected from each of the four data sets as the training set, and 343 samples are randomly selected from one of the remaining data sets as the testing set. The rest of the experiment conditions, parameter settings, and evaluation criteria were the same as those used in Experiment 1. For the same reason to avoid random results, 10 experiments were conducted.

In Experiment 2, the data of one working condition are selected as the training set, and one of the remaining data sets is selected as the testing set. The average accuracy of 10 experiments is shown in Figure 13. It can be seen from the experimental results that in the 6 groups of experiments, the lowest average accuracy reached 97.73%, and the highest accuracy reached 99.42%. Under the condition that the training set is reduced, the average accuracy of the diagnosis of the same fault type still has good and stable performance.

Three of the six groups were selected. They are numbered Tasks 5–7. The other two methods are still used to compare with the proposed method. The sample situations of Tasks 5–7 are shown in Table 7.
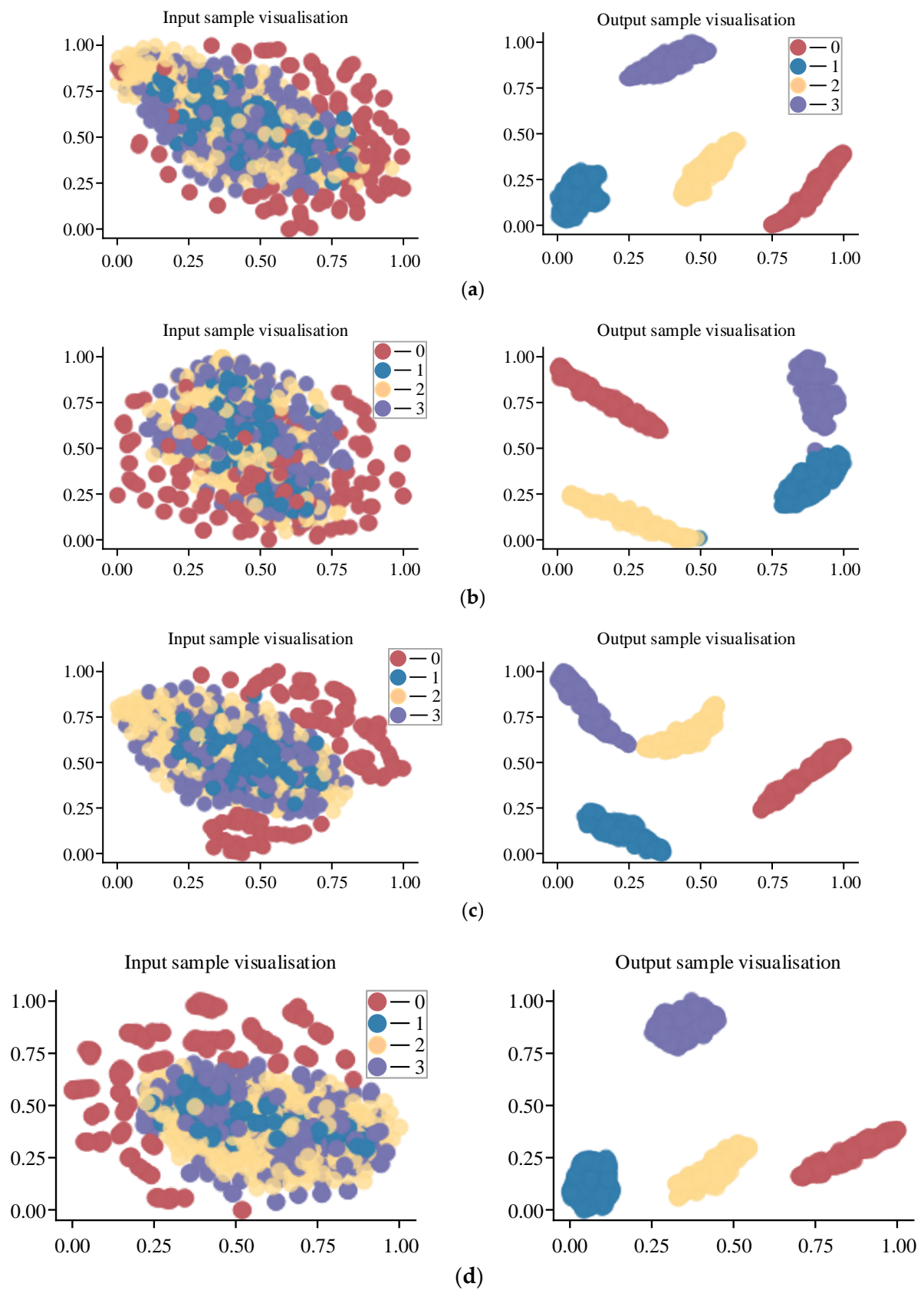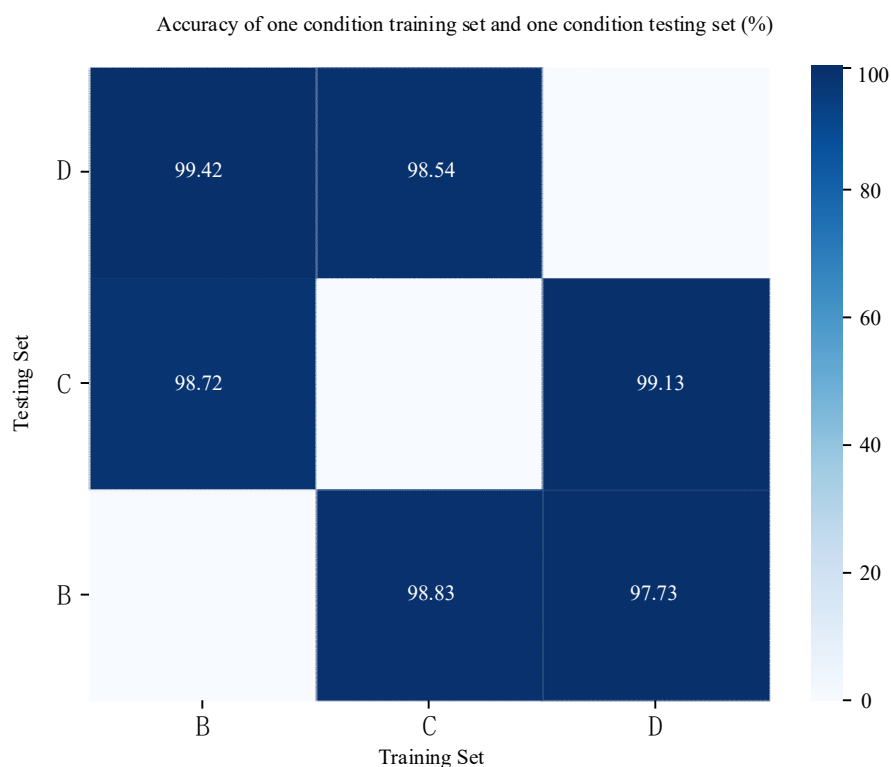
Figure 12. Feature visualization: (**a**) Task 1; (**b**) Task 2; (**c**) Task 3; (**d**) Task 4.

Accuracy of one condition training set and one condition testing set (%)

Figure 13. The accuracy of fault diagnosis when the data of one working condition are used as the training set and the data of another different working condition are used as the testing set.

Table 7. Sample selection for Tasks 5–7.

Task Number	Train	Sample	Test	Sample
5	B	800	C	343
6	C	800	D	343
7	D	800	B	343

One experiment was randomly selected from each of the four tasks, and its accuracy and F1 Score are shown in Figure 14. The accuracy and F1 Score of the proposed method in the three tasks are all over 97.37%, which is better than the other two methods. In Task 5, accuracy and F1 Score reach 98.83%, which proves the excellent fault diagnosis capability of the proposed method under variable working conditions. The average accuracy and average F1 Score of the ten experiments are shown in Table 8. The average evaluation index obtained from ten experiments in different tasks can again prove the generalization and stability of the proposed method. The training time and testing time of the three methods are shown in Table 9. The average time spent on different tasks decreases as the number of training and testing sets decreases. Although the time used is higher than the other two methods, it is still within the acceptable range.

Figure 15 shows the confusion matrix of the method proposed in this paper for one of the experiments for each of the three tasks, particularly included are the results of the classification for all four fault states. It can be seen from this figure that the accuracy reached 98.83% in Task 5, and in Task 6, three samples of fault type 3 were incorrectly diagnosed as fault type 1, whilst eight samples of fault type 3 were incorrectly diagnosed as fault type 2. The accuracy for label 1 is 91.2% and that for label 3 is 98.8%. From the signal frequency analysis, it is known that when the outer and inner ring faults occur, the low-frequency component almost disappears, and for the rolling body fault, the vibration signal is dominated by low-frequency components, that is the reason that the samples of IF

and OF in Task 6 cause confusion. In Task 7, except for the accuracy of all types but label 1 reaches 100%, the accuracy for label 1 is 92.9%.
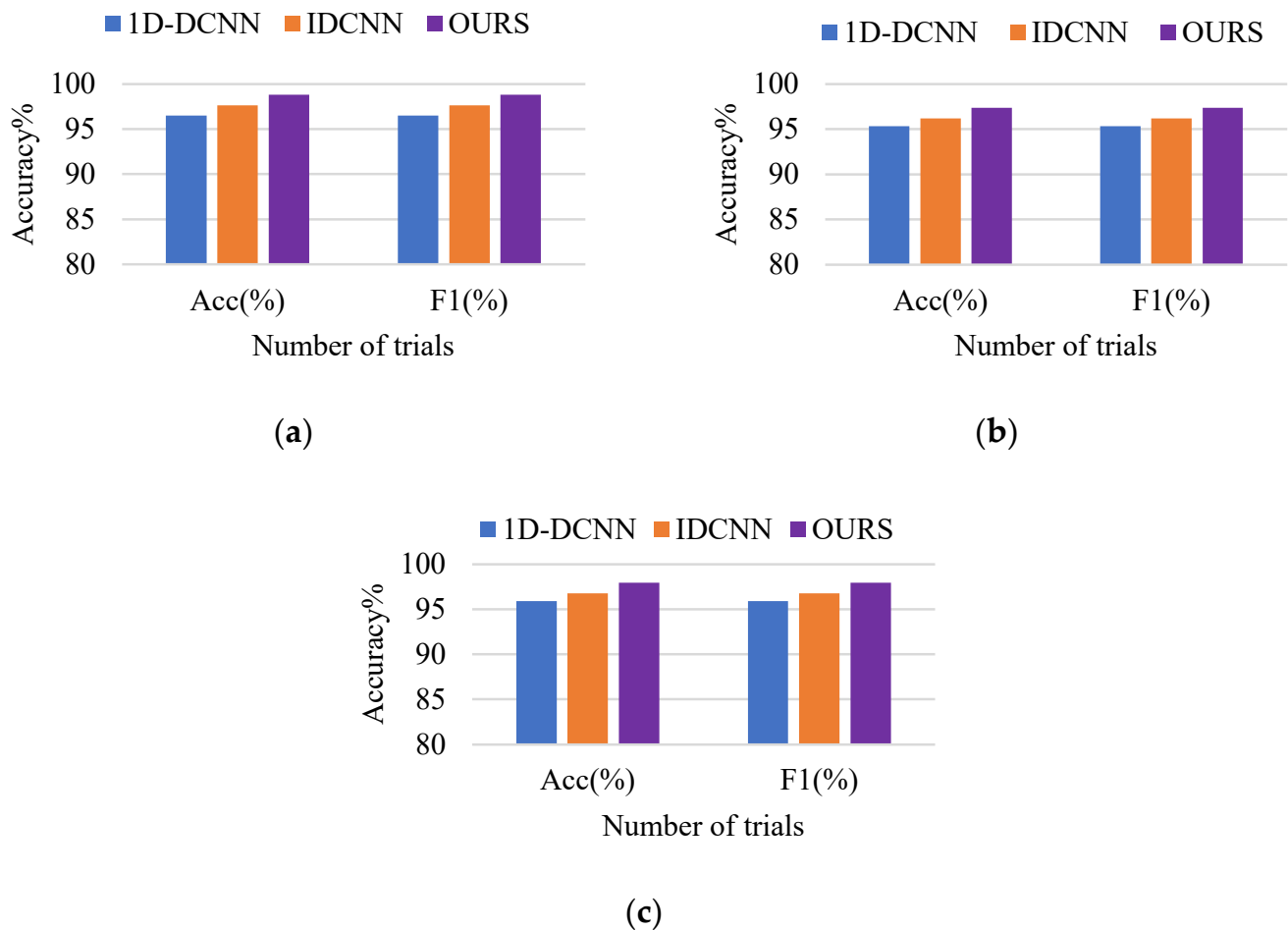

(**a**)

(**b**)

(**c**)

Figure 14. Results of trials; (**a**) Task 5; (**b**) Task 6; (**c**) Task 7.

Table 8. Average evaluation index for different methods in Task 5, 6, 7.

Task Number	1D-CNN Average Accuracy/F1 Score	IDCNN Average Accuracy/F1 Score	OURS Average Accuracy/F1 Score
5	96.33%/96.33%	97.32%/97.33%	98.72%/98.72%
6	96.12%/96.11%	97.11%/97.11%	98.54%/98.54%
7	95.60%/95.60%	96.68%/96.68%	97.73%/97.73%

Table 9. Average time for different methods in Task 5, 6, 7.

| Method	Time	Task 5	Task 6	Task 7
1D-DCNN	Average Training Time (s)	6.42	6.54	6.47
	Average Testing Time (s)	0.43	0.44	0.42
IDCNN	Average Training Time (s)	11.41	11.59	11.62
	Average Testing Time (s)	0.98	0.96	0.98
OURS	Average Training Time (s)	11.47	11.35	11.34
	Average Testing Time (s)	8.47	8.25	8.46
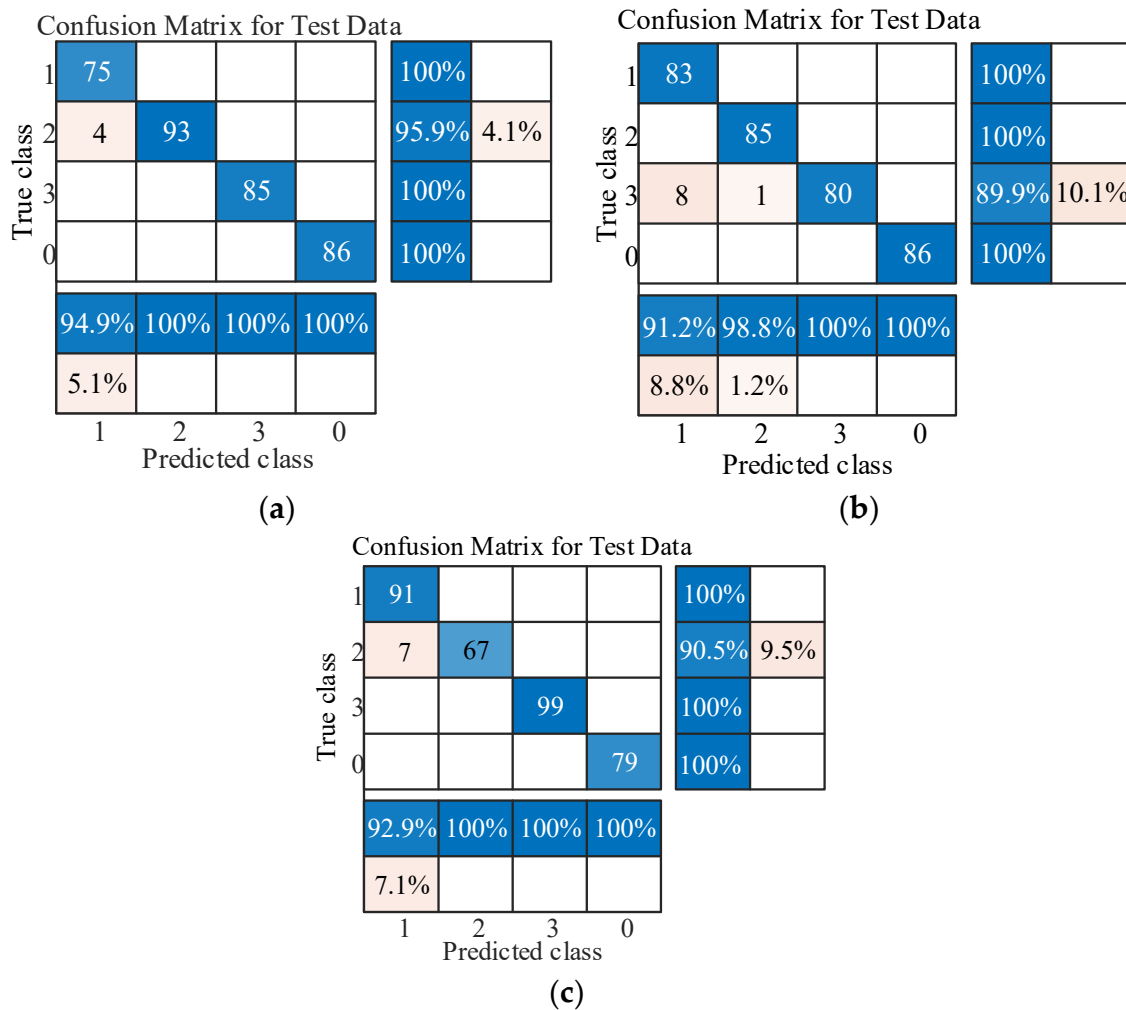
Figure 15. Confusion matrix for the three tasks: (**a**) Task 5; (**b**) Task 6; (**c**) Task 7.

4.5. Analysis of Experiment 3

In Experiment 3, the data in Table 1 were used. A total of 3000 training set samples and 1280 testing set samples. Other experimental conditions, parameter settings and evaluation criteria are the same as in Experiment 1. By the same token, we conducted 10 experiments to avoid random results. When the working condition is 1772 rpm/1 hp, the fault status identification result is shown in Figure 16. Furthermore, the average recognition accuracy reached 100% under this condition.

In order to verify the effectiveness of the proposed method, 1D-DCNN, IDCNN, and the proposed method in this paper were, respectively, used for experiments at 1730 rpm/3 hp. Figure 17 shows the confusion matrix of the results of any of the ten experiments. It can be seen from the experimental results that the accuracy of the proposed method reaches 100% under non-variable working conditions. In addition, compared with the IDCNN method, the accuracy of the proposed method is also improved, indicating that the construction of the WOA-DELM classifier plays a certain role in improving the accuracy of fault types.

In order to verify the generalization of the method proposed in this paper. The proposed method was used to carry out experiments under four conditions, and the average accuracy and F1 Score of ten experiments were obtained. The experimental results are shown in Table 10. According to the results, it can be seen that the average accuracy and F1 Score of the proposed method reach 100% in the three working conditions. It shows that the proposed method has a high accuracy of fault identification.
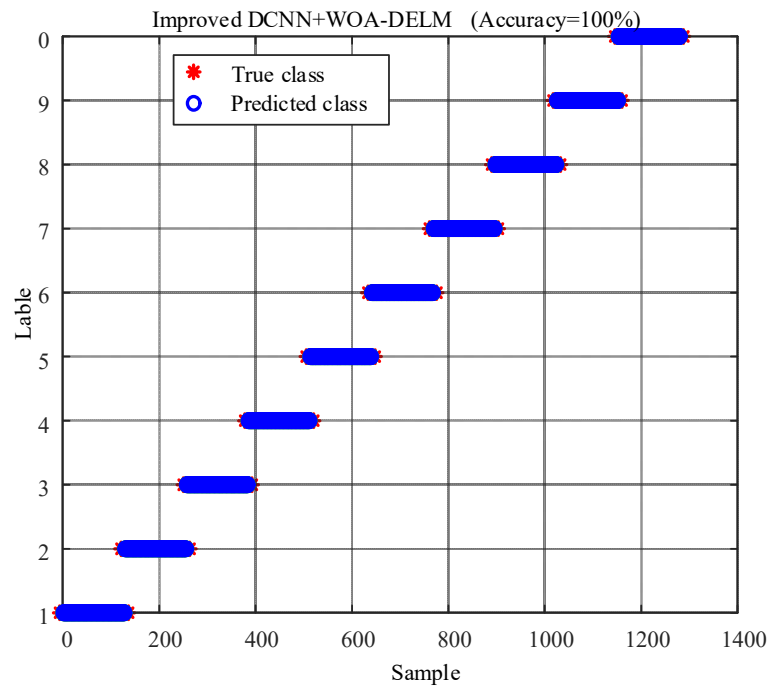
Figure 16. The fault identification results of the proposed method are obtained when the working condition is 1772 rpm/1 hp.
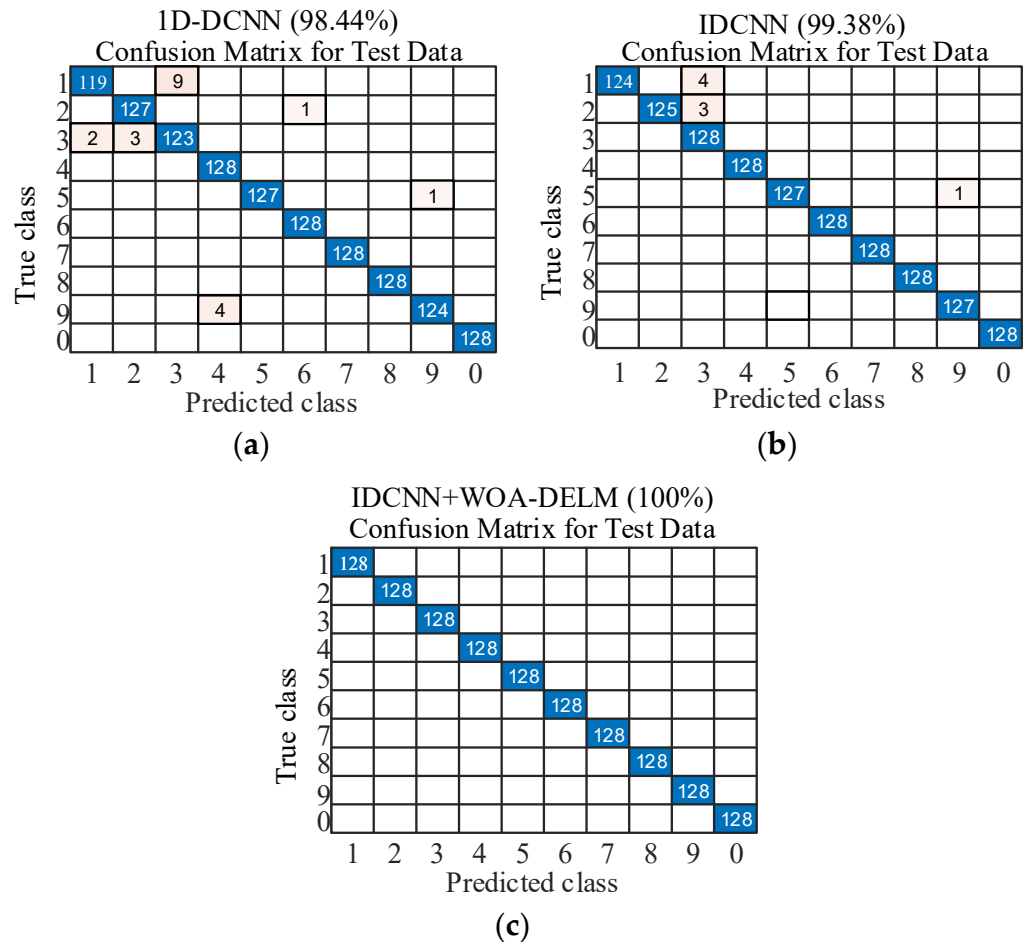
Figure 17. Results of fault classification by different methods at 1730 rpm/3 Hp: (**a**) 1D-DCNN; (**b**) IDCNN; (**c**) IDCNN-DELM.

Table 10. The fault identification results of the proposed method are obtained under different load and speed conditions.

Speed/Load	Status	Number of Training Samples	Number of Testing Samples	Average Accuracy	Average F1 Score
1797 rpm/0 hp	Normal IF BF OF	300, 900, 900, 900	128, 384, 384, 384	100%	100%
1772 rpm/1 hp	Normal IF BF OF	300, 900, 900, 900	128, 384, 384, 384	100%	100%
1750 rpm/2 hp	Normal IF BF OF	300, 900, 900, 900	128, 384, 384, 384	99.77%	99.77%
1730 rpm/3 hp	Normal IF BF OF	300, 900, 900, 900	128, 384, 384, 384	100%	100%

Table 11 shows the accuracy and average accuracy of the three methods in ten experiments under different working conditions. According to the results, it can be seen that the proposed method has good performance in four different working conditions.

Table 11. Fault identification accuracy of different methods under different working conditions.

Method	1797 rpm/0 hp	1772 rpm/1 hp	1750 rpm/2 hp	1730 rpm/3 hp	Average Accuracy
1D-DCNN	98.44%	98.42%	98.45%	98.46%	98.44%
IDCNN	99.32%	99.29%	99.30%	100%	99.47%
OURS	100%	100%	99.77%	100%	99.94%

Eight commonly used fault diagnosis methods were used for Experiment 2, in which the training set and testing set were randomly selected from the corresponding data set and compared with the proposed method in this paper, and 10 experiments were conducted to find the average accuracy in order to reduce random error. The results are shown in Table 12. Apparently, under the variable working conditions, the accuracy and F1 Score of the six methods are lower than that of the proposed method.

Table 12. Fault diagnosis results of other methods.

Model	Train	Test	Accuracy	F1 Score
	B	C	71.5%	-
SVM	C	D	71.7%	-
	D	B	66.9%	-
	B	C	97.3%	97.3%
IDCNN	C	D	97.1%	97.1%
	D	B	96.7%	96.7%
	B	C	84.7%	-
Multilayer Perceptron (MLP)	C	D	80.6%	-
	D	B	82.6%	-
	B	C	77.9%	-
Deep Neural Networks (DNN)	C	D	74.1%	-
	D	B	78.9%	-
	B	C	42.3%	-
Decision Tree (DT)	C	D	41.5%	-
	D	B	46.7%	-

Table 12. *Cont.*

Model	Train	Test	Accuracy	F1 Score
DCNN-BiLSTM	B	C	96.7%	96.7%
	C	D	96.9%	96.9%
	D	B	97.3%	97.4%
CEEMDAN-PSO-SVM	B	C	78.6%	-
	C	D	79.2%	-
	D	B	78.3%	-
HRCMFDE-PSO-ELM [28]	B	C	92.0%	-
	C	D	98.5%	-
	D	B	84.8%	-
OURS	B	C	98.7%	98.7%
	C	D	98.5%	98.5%
	D	B	97.7%	97.7%

5. Conclusions

A bearing fault diagnosis method based on the IDCNN-WOA-DELM is proposed and applied to rotating machinery fault diagnosis under variable load and varying speed operating conditions. The main findings from this research are summarized as follows:

1. By adding BiLSTM to DCNN, the deep features of timing signals can be extracted bi-directional. By using the adaptive one-dimensional convolutional kernel size determination in the ECA-Net, A certain level of performance improvement can be achieved with a small number of added parameters, and by weighting the effective features, the effective features can play even greater roles.

2. The application of the IDCNN in bearing fault features extraction under varying working conditions can reduce the dependence on expert experience and enhances the stability and integrity of the algorithm. This is demonstrated by visualizing the test samples and the fully connected layer output of the IDCNN model using t-SNE.

3. The initial weights of the DELM exhibit randomness, whereas the WOA algorithm possesses excellent global search capability. By leveraging the contraction surround mechanism and spiral advance mechanism in the WOA algorithm, the randomly generated initial weights in the DELM can be optimized, and the WOA-DELM classifier is constructed. The experiment results indicate that the WOA-DELM classifier can significantly enhance the classification performance of bearing faults under varying working conditions, demonstrating the outstanding performances of the WOA algorithm and the improved WOA-DELM classifier.

4. Various experiments were conducted using the CWRU bearing dataset to simulate bearing fault diagnosis under varying working conditions. The training and testing sets were selected randomly across multiple experiments. In Experiment 1 and Experiment 2, a total of 18 different experiments were carried out. In the 12 groups of experiments in Experiment 1, the average accuracy of the testing set reached 98.65%. In the six experiments in Experiment 2, the average accuracy of the testing set was more than 97.69%. Among the 4 groups of experiments in Experiment 3, the average accuracy of the testing set in 3 groups reached 100%. The averaged experiment results prove that the proposed method can provide better fault classification performance and generalization ability. Furthermore, the comparison with the SVM, IDCNN, MLP, DNN, DT, DCNN-BiLSTM, and other methods confirms the effectiveness of the proposed method.

In practical production, bearing faults may conform to complex patterns, leading to more intricate vibration signals. Therefore, addressing fault analysis under compound

fault conditions in the following investigation will be a critical step forward. At the same time, in order to deploy deep learning models in mobile and embedded devices, the work related to lightweight models is also of great significance.

References

1. Gai, J.B.; Shen, J.X.; Hu, Y.F.; Wang, H. An integrated method based on hybrid grey wolf optimizer improved variational mode decomposition and deep neural network for fault diagnosis of rolling bearing. *Measurement* **2020**, *162*, 107901. [CrossRef]
2. Nishat Toma, R.; Kim, J.-M. Bearing Fault Classification of Induction Motors Using Discrete Wavelet Transform and Ensemble Machine Learning Algorithms. *Appl. Sci.* **2020**, *10*, 5251. [CrossRef]
3. Tang, S.N.; Yuan, S.Q.; Zhu, Y. Deep Learning-Based Intelligent Fault Diagnosis Methods Toward Rotating Machinery. *IEEE Access* **2020**, *8*, 9335–9346. [CrossRef]
4. Wang, Y.X.; Xiang, J.W.; Markert, R.; Liang, M. Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications. *Mech. Syst. Signal Process.* **2016**, *66-67*, 679–698. [CrossRef]
5. Mustafa, D.; Yicheng, Z.; Minjie, G.; Jonas, H.; Jürgen, F. Motor current based misalignment diagnosis on linear axes with short-time Fourier transform (STFT). *Procedia CIRP* **2022**, *106*, 239–243. [CrossRef]
6. Gowid, S.; Dixon, R.; Ghani, S.; Shokry, A. Robustness analysis of the FFT-based segmentation, feature selection and machine fault identification algorithm. *Insight* **2019**, *61*, 271–278. [CrossRef]
7. Sifuzzaman, M.; Islam, M.R.; Ali, M. Application of wavelet transform and its advantages compared to Fourier transform. *Int. J. Mag. Eng. Technol. Manag. Res.* **2009**, *3*, 1078–1083.
8. Miao, Y.; Zhao, M.; Lin, J. Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition. *ISA Trans.* **2019**, *84*, 82–95. [CrossRef]
9. Amarouayache, I.I.E.; Saadi, M.N.; Guersi, N.; Boutasseta, N. Bearing fault diagnostics using EEMD processing and convolutional neural network methods. *Int. J. Adv. Manuf. Technol.* **2020**, *107*, 4077–4095. [CrossRef]
10. Hou, J.; Wu, Y.; Gong, H.; Ahmad, A.S.; Liu, L. A Novel Intelligent Method for Bearing Fault Diagnosis Based on EEMD Permutation Entropy and GG Clustering. *Appl. Sci.* **2020**, *10*, 386. [CrossRef]
11. He, X.Z.; Zhou, X.Q.; Yu, W.N.; Hou, Y.X.; Mechefske, C.K. Adaptive variational mode decomposition and its application to multi-fault detection using mechanical vibration signals. *ISA Trans.* **2021**, *111*, 360–375. [CrossRef] [PubMed]
12. Wang, Z.Y.; Yao, L.G.; Cai, Y.W. Rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine. *Measurement* **2020**, *156*, 107574. [CrossRef]
13. Zhang, S.C.; Li, X.L.; Zong, M.; Zhu, X.F.; Wang, R.L. Efficient kNN Classification With Different Numbers of Nearest Neighbors. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 1774–1785. [CrossRef]
14. Kanai, R.A.; Desavale, R.G.; Chavan, S.P. Experimental-Based Fault Diagnosis of Rolling Bearings Using Artificial Neural Network. *J. Tribol.-Trans. ASME* **2016**, *138*, 031103. [CrossRef]
15. Deng, W.; Yao, R.; Zhao, H.M.; Yang, X.H.; Li, G.Y. A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm. *Soft Comput.* **2019**, *23*, 2445–2462. [CrossRef]
16. Lu, J.; Qian, W.; Li, S.; Cui, R. Enhanced K-Nearest Neighbor for Intelligent Fault Diagnosis of Rotating Machinery. *Appl. Sci.* **2021**, *11*, 919. [CrossRef]
17. Hannun, A.Y.; Rajpurkar, P.; Haghpanahi, M.; Tison, G.H.; Bourn, C.; Turakhia, M.P.; Ng, A.Y. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **2019**, *25*, 65–69. [CrossRef]
18. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.; van Ginneken, B.; Sanchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef]

19. Zhao, R.; Yan, R.Q.; Chen, Z.H.; Mao, K.Z.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [CrossRef]
20. Kiranyaz, S.; Avci, O.; Abdeljaber, O.; Ince, T.; Gabbouj, M.; Inman, D.J. 1D convolutional neural networks and applications: A survey. *Mech. Syst. Signal Process.* **2021**, *151*, 107398. [CrossRef]
21. Zhang, C.; Tan, K.C.; Li, H.Z.; Hong, G.S. A Cost-Sensitive Deep Belief Network for Imbalanced Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 109–122. [CrossRef] [PubMed]
22. Shao, S.Y.; Wang, P.; Yan, R.Q. Generative adversarial networks for data augmentation in machine fault diagnosis. *Comput. Ind.* **2019**, *106*, 85–93. [CrossRef]
23. Jiang, G.Q.; He, H.B.; Yan, J.; Xie, P. Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox. *IEEE Trans. Ind. Electron.* **2019**, *66*, 3196–3207. [CrossRef]
24. Gong, W.; Chen, H.; Zhang, Z.; Zhang, M.; Wang, R.; Guan, C.; Wang, Q. A Novel Deep Learning Method for Intelligent Fault Diagnosis of Rotating Machinery Based on Improved CNN-SVM and Multichannel Data Fusion. *Sensors* **2019**, *19*, 1693. [CrossRef] [PubMed]
25. Deng, W.; Liu, H.L.; Xu, J.J.; Zhao, H.M.; Song, Y.J. An Improved Quantum-Inspired Differential Evolution Algorithm for Deep Belief Network. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 7319–7327. [CrossRef]
26. Zhang, C.; Chen, P.; Jiang, F.; Xie, J.; Yu, T. Fault Diagnosis of Nuclear Power Plant Based on Sparrow Search Algorithm Optimized CNN-LSTM Neural Network. *Energies* **2023**, *16*, 2934. [CrossRef]
27. Chen, Z.Y.; Gryllias, K.; Li, W.H. Mechanical fault diagnosis using Convolutional Neural Networks and Extreme Learning Machine. *Mech. Syst. Signal Process.* **2019**, *133*, 106272. [CrossRef]
28. Chen, Y.; Yuan, Z.; Chen, J.; Sun, K. A Novel Fault Diagnosis Method for Rolling Bearing Based on Hierarchical Refined Composite Multiscale Fluctuation-Based Dispersion Entropy and PSO-ELM. *Entropy* **2022**, *24*, 1517. [CrossRef]
29. Zhou, F.N.; Yang, S.; Fujita, H.; Chen, D.M.; Wen, C.L. Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. *Knowl.-Based Syst.* **2020**, *187*, 104837. [CrossRef]
30. Mao, W.T.; Liu, Y.M.; Ding, L.; Li, Y. Imbalanced Fault Diagnosis of Rolling Bearing Based on Generative Adversarial Network: A Comparative Study. *IEEE Access* **2019**, *7*, 9515–9530. [CrossRef]
31. Huang, G.; Huang, G.-B.; Song, S.; You, K. Trends in extreme learning machines: A review. *Neural Netw.* **2015**, *61*, 32–48. [CrossRef] [PubMed]
32. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.
33. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. [CrossRef] [PubMed]
34. Mirjalili, S.; Lewis, A. The Whale Optimization Algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [CrossRef]