



*Research article*

## **Intelligent fault diagnosis algorithm of rolling bearing based on optimization algorithm fusion convolutional neural network**

**Qiushi Wang<sup>2,4</sup>, Zhicheng Sun<sup>2,4,\*</sup>, Yueming Zhu<sup>2,4</sup>, Chunhe Song<sup>1,2,3</sup> and Dong Li<sup>1,2,3</sup>**

<sup>1</sup> State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China No.114 Nanta Street, Shenyang, Liaoning Province, China

<sup>2</sup> Key Laboratory of Networked Control Systems, Chinese Academy of Sciences, Shenyang 110016, China

<sup>3</sup> Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China

<sup>4</sup> Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

\* **Correspondence:** Email: [sunzhicheng@sia.cn](mailto:sunzhicheng@sia.cn).

**Abstract:** As an essential component of mechanical equipment, the fault diagnosis of rolling bearings may not only guarantee the systematic operation of the equipment, but also minimize any financial losses caused by equipment shutdowns. Fault diagnosis algorithms based on convolutional neural networks (CNN) have been widely used. However, traditional CNNs have limited feature representation capabilities, thereby making it challenging to determine their hyperparameters. This paper proposes a fault diagnosis method that combines a 1D-CNN with an attention mechanism and hyperparameter optimization to overcome the aforementioned limitations; this method improves the search speed for optimal hyperparameters of CNN models, improves the diagnostic accuracy, and enhances the representation of fault feature information in CNNs. First, the 1D-CNN is improved by combining it with an attention mechanism to enhance the fault feature information. Second, a swarm intelligence algorithm based on Differential Evolution (DE) and Grey Wolf Optimization (GWO) is proposed, which not only improves the convergence accuracy, but also increases the search efficiency. Finally, the improved 1D-CNN alongside hyperparameters optimization are used to diagnose the faults of rolling bearings. By using the Case Western Reserve University (CWRU) and Jiangnan University (JNU) datasets, when compared to other common diagnosis models, the results demonstrate the usefulness and dependability of the DE-GWO-CNN algorithm in fault diagnosis applications by demonstrating the increased diagnostic accuracy and superior anti-noise capabilities of the proposed method. The fault diagnosis methodology presented in this paper can accurately identify faults and provide dependable fault classification, thereby assisting technicians in promptly resolving faults and minimizing equipment failures and operational instabilities.

---

**Keywords:** fault diagnosis; DE-GWO; rolling bearing; attention mechanism; 1DCNN

---

## 1. Introduction

To ensure the safety and stable operation of industrial production within the industrial field, equipment fault diagnosis is extremely important. Because of the challenging atmosphere and harsh working circumstances, rolling bearings within mechanical equipment are often damaged [1]. According to statistics, rolling bearings are responsible for around 30% of mechanical faults in rotating equipment and about 20% of mechanical faults in gearboxes [2]. The rolling bearings' condition monitoring and fault diagnosis technology serve as an important tool in understanding the performance status of bearings and identifying potential faults in time [3]; in addition, it can effectively improve the operation management and maintenance efficiency of mechanical equipment, thus significantly improving the economic performance of enterprises [4].

Nowadays, gathering temperature or pressure data of the equipment for fault diagnosis does not yield accurate findings due to the influence of the factory's actual production process; alternatively, the equipment is directly connected to the vibration signal, which can therefore represent the device's operational status [5]. Thus, there is a high importance placed on research into fault diagnosis algorithms for vibration signals. In recent years, there has been a significant increase in computer performance due to advances in science and technology. Additionally, experts and academics have used machine learning and deep learning algorithms to diagnose faults using vibration signals [6]. Common fault diagnosis algorithms include support vector machine [7], k-nearest neighbor [8], random forest [9], convolution neural network (CNN) [10], long short-term memory (LSTM)[11] and transfer learning(TL) [12]. Wu [13] used the enhanced quantum-inspired differential evolution (MSIQDE) method to optimize deep confidence networks and facilitate fault diagnosis. Yuan [14] effectively utilized hypergraph algorithms to reduce the dimensionality of fault data features and implemented KNN classifiers for precise fault classification. Qing [15] proposed a novel Physical Information Residual Network (PIResNet) for the fault diagnosis of rolling bearings. Feng [16] created a digital replica of the gearbox's structure and employed a transfer learning algorithm to acquire knowledge on faults, which was then used to evaluate surface faults in an actual gearbox. Yuan [17] used the local-global standard hypergraph embedding (LGSHE) method to reduce the dimensionality of fault information and improve the accuracy of fault classification. Among them, CNN has attracted much attention because of its powerful feature extraction ability. Wang [18] employed a wavelet analysis to transform time-frequency picture data from one-dimensional (1D) vibration data into feature extraction and defect classification using a deep convolution neural network. Qiu [19] built a fault diagnosis model based on the CNN model by integrating auxiliary classifier generative adversarial network (ACGAN) and a self-attention mechanism. Gao [20] proposed a fault diagnosis method by combining maximum correlated kurtosis deconvolution (MCKD) and CNN. Hoang [21] transformed the 1D vibration signal into a binary graph and used CNN for fault diagnosis; however, this method lost the time series information. Wang [22] used improved Markov transformation field (MTF) to convert 1D vibration data into two-dimensional (2D) image data, and then used CNN to diagnose faults of rolling bearings. To achieve fault diagnosis, Zhao [23] proposed a new dimensionality reduction method and applied it to 1D-CNN with an adaptive activation function.

However, in the actual use of CNN for fault diagnosis, hyperparameters are often selected based on experience, which cannot allow CNN to play an optimal role, as the hyperparameter optimization problem for CNN has multi-modal characteristics [24]. Therefore, many scholars use intelligent optimization algorithms to optimize the hyperparameters of CNN. Sun [25] took the diagnosis accuracy and stability of the model as the optimization objective and used Differential Evolution (DE) to optimize the hyperparameters of CNN; however, the convergence speed was not satisfactory. Li [26] proposed a fault diagnosis method by combining the symplectic frog leapfrog algorithm (SFLA) and CNN, the SFLA is used to optimize the network structure and improve the feature extraction ability of CNN, thus improving the fault diagnosis accuracy of the model. Wang [27] combined auxiliary classifier generative adversarial network (VMD) with CNN for fault diagnosis by first optimizing the parameters of VMD and processing the data by using the improved grey wolf optimization algorithm; then, the processed data was inputted into the CNN, which optimized the model parameters through a grid search for fault diagnosis. However, the global search capability was not satisfactory.

Although CNN is an effective method in the field of rolling bearing fault diagnosis, there are still some challenges in practical applications such as quickly finding the most suitable hyperparameters and improving the feature representation ability of CNN models. To this end, this paper compares other intelligent optimization algorithms, and finds that though the DE algorithm has excellent global optimization ability [28], the iterative process of the grey wolf optimization (GWO) algorithm is quicker [29]. Therefore, this paper proposes a fault diagnosis algorithm that combines the two optimization algorithms to optimize the hyperparameters of the 1D convolutional neural network (1D-CNN) model and adds an attention mechanism after the 1D-CNN to highlight more useful fault feature information. Then, the algorithm uses the optimized 1D-CNN to build a fault diagnosis model to diagnose rolling bearing faults. The contribution of this paper can be summarized as follows:

1. A novel deep learning method based on 1D-CNN is proposed, which effectively enhances the fault feature information by adding an attention mechanism layer, prevents overfitting of the network by adding a dropout layer, and efficiently improves the accuracy of the classification.

2. A new swarm intelligence optimization algorithm based on DE-GWO is proposed, which not only enhances the global search ability, but also improves the convergence speed.

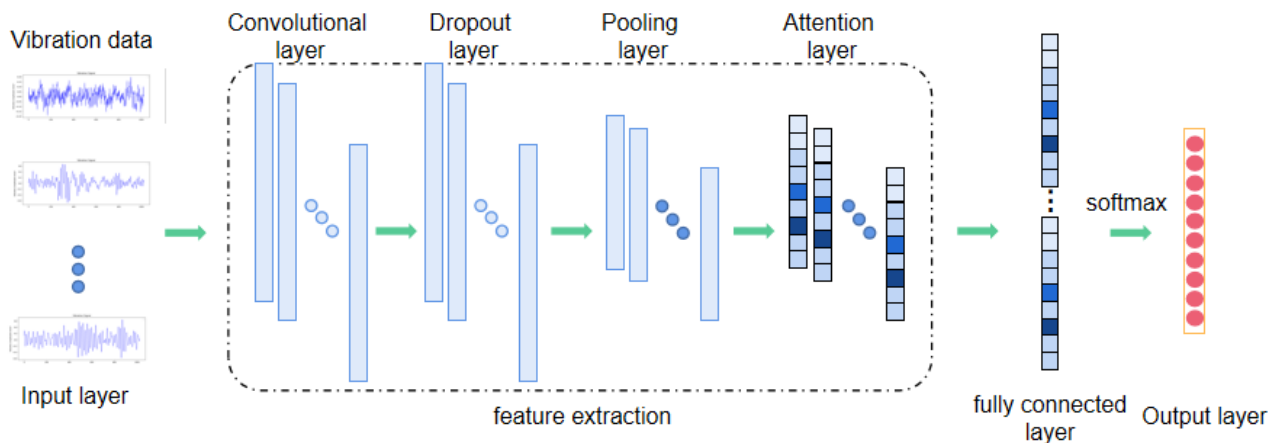
3. A fault diagnosis method for rolling bears based on DE-GWO-CNN is proposed, which efficiently identifies the best combination of six important hyperparameters for the improved 1D-CNN algorithm and effectively improves the fault diagnosis accuracy while ensuring fast convergence.

The remaining of this paper is displayed as follows. Section 2 represents the method used in this paper. Section 3 shows the training and testing datasets. The results and comparison are shown in Section 4 and the conclusion is drawn in Section 5.

## 2. Methods

The structure of the fault diagnosis method proposed in this paper is shown in Figure 1.

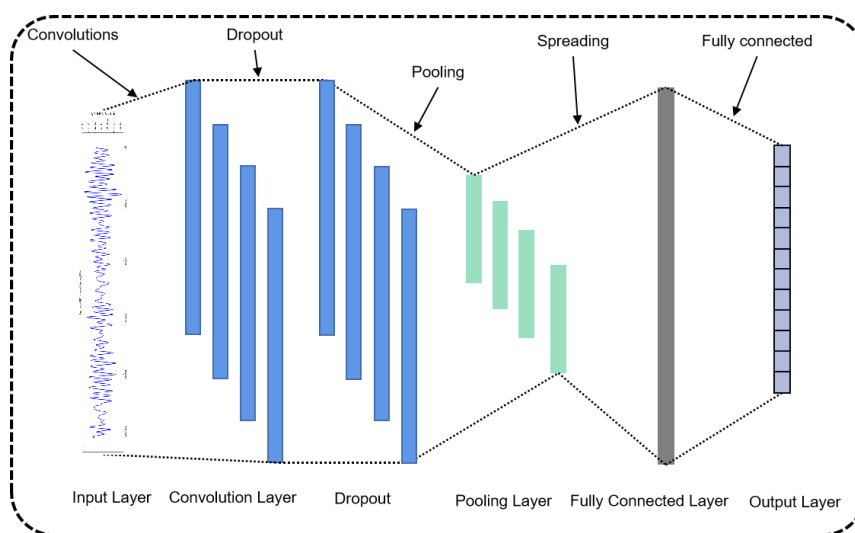
As seen from Figure 1, the first step is constructing the datasets involves normalizing fault signals and tags and dividing them into a training set and a test set. Next, we utilize the DE-GWO algorithm proposed in this paper to optimize the hyperparameters of the CNN method. Subsequently, we train the fault detection model using both the optimized CNN and the training data. Finally, we evaluate the trained model using a test set to ensure its accuracy and effectiveness.



**Figure 1.** The structure of the fault diagnosis method.

### 2.1. DCNN algorithm

The main components of a 1D-CNN, which consists of a convolution layer, a pooling layer, and a fully connected layer, are similar to those of a typical feed-forward neural network. However, 1D-CNN has the added benefits of reducing model complexity, avoiding difficult feature extraction procedures, and reducing the number of weights required. The convolution, dropout, and pooling layers are utilized for feature extraction from the original signal, while the full connection layer creates a mapping relationship between the retrieved features and labels to realize the classification function. The structure of the 1D-CNN can be visualized in Figure 2.



**Figure 2.** The structure of the 1DCNN.

The convolution layer applies a convolution operation to the input's original 1D data using convolution kernels, and then uses the activation function to make nonlinear changes to obtain a series of feature maps. The following formula [30] can be used to describe the convolution process:

$$y_{i,j,k} = f\left(\sum_{i=l}^s x_{i,k} * w_{j,i} + b_i\right), \quad (1)$$

where  $y_{i,j,k}$  is the result of the convolution operation,  $f$  represents the activation function, which is Relu in this paper,  $x_{i,k}$  represents the input data,  $*$  represents the convolution operation,  $w_{j,i}$  represents weights and  $b_i$  represents bias.

The dropout is used to prevent an overfitting problem during the training of the CNN model, which utilizes random sampling of weights based on a certain probability.

The purpose of the pooling operation is to improve the training speed of the convolution neural network and further avoid overfitting by reducing the dimension of the data. Popular pooling techniques include the maximum and average pooling methods; this paper adopts the maximum pooling method [30]:

$$X = \max_{(i-1)l+1 \leq t \leq il} X^{l-1}(t), \quad (2)$$

where  $X$  represents the feature map after dimension reduction and  $l$  represents the length of the pooling area.

Finally, the extracted features are inputted into the full connection layer, the output probability is identified through the softmax function; then, the classification results are obtained. In this paper, before the full connection layer, a dropout layer is inserted to disregard some neurons and prevent overfitting during the training model.

Although 1D-CNN has a strong feature extraction ability, the selection of hyperparameters in the training process of the model will also have a great impact on the training results. At present, hyperparameters are generally selected through experience, though the effect is often general. This paper optimizes the hyperparameters of 1D-CNN by using the proposed DE-GWO algorithm.

The hyperparameters that have a great impact on the training results of 1D-CNN are the number of convolution kernels, the size of the convolution kernel, the dropout rate, the size of the pooling layer, the batch size and the learning rate. As a result, this paper mainly uses DE-GWO to optimize these parameters.

## 2.2. Attention mechanism

The attention mechanism can highlight the fault features with important information and suppress invalid features through adaptive weighting of different signal segments. The attention mechanism in this paper is added between the pooling layer and the fully connected layer. According to the attention mechanism, features that have a significant impact on the results will be given a greater weight. The structure is shown in Figure 3.

The input features are automatically extracted through convolution. The attention weight of each channel of the feature is obtained by adding an attention layer after the feature map. Then, the output feature map is produced using the dot product of the acquired attention weight and the original feature. The attention mechanism can be described as follows [31]:

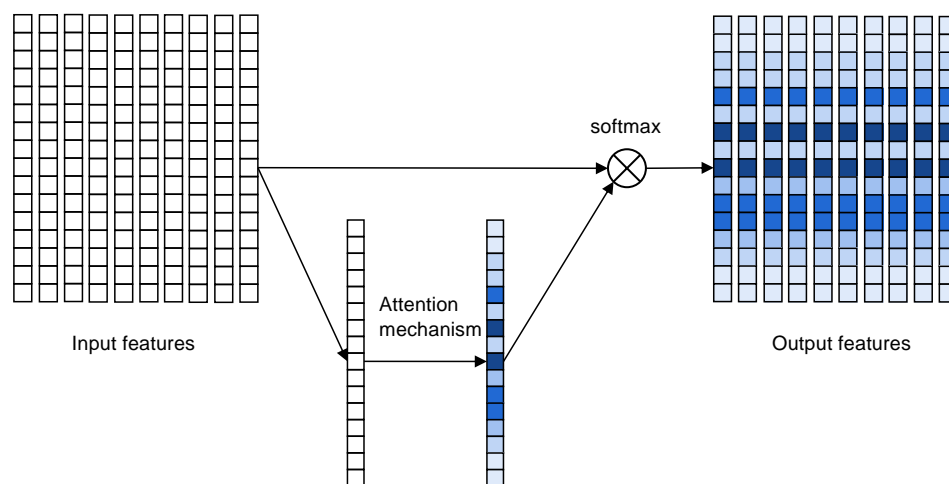
$$H' = M_c(H) \otimes H, \quad (3)$$

where  $H \in n \times m$  is the original feature map after the convolution operation,  $M_c \in 1 \times n$  is the attention weights for the channel dimensions,  $H'$  is the feature map after the attention mechanism,  $\otimes$  is the dot product, and  $M_c(H)$  is described as follows [31]:

$$M_c(H) = \sigma(\text{Conv}([H_{avg}^C; H_{max}^C])), \quad (4)$$

where  $H_{avg}^C$  and  $H_{max}^C$  are the global average pooling and global max pooling, respectively,  $\text{Conv}()$  is the convolution operation and  $\sigma$  is the activation function.

In this paper, the attention mechanism is added after the maxpooling layer to enhance the fault feature information of 1D-CNN.



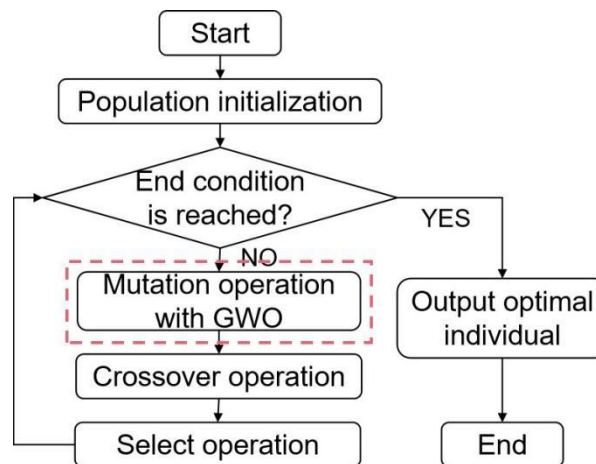
**Figure 3.** The structure of the attention mechanism.

### 2.3. The DE-GWO method

In 1995, R. Storn and K. Price proposed the DE algorithm as a population-based optimization method. Compared with other optimization algorithms, the advantages of DE mainly lie in its controllability and few control parameters to be adjusted. Its attributes include a straightforward structure, straightforward realization, quick convergence, and high resilience; however, the convergence speed becomes slower during the latter part of the algorithm and sometimes even falls into the local best. In 2014, Mirjalili proposed the GWO algorithm as a population-based optimization method. It benefits from having a strong convergence performance, a straightforward structure, few adjustable parameters, and ease of implementation. When solving problems, it performs well in terms of convergence speed and accuracy. However, when faced with difficult issues, it easily converges early, and the convergence accuracy is not good. Narayan Nahak and Ranjan Kumar Mallick [32] combined the two algorithms into one by considering the final population of DE as the initial population of GWO, which took the advantages of both DE and GWO.

The traditional DE algorithm has a strong global search ability, though its convergence rate is not satisfactory. This paper effectively improves the rate of convergence and accuracy of the algorithm by integrating the GWO algorithm into the mutation process of DE, which makes the

population evolution more directional. Figure 4 depicts the method's organizational structure.



**Figure 4.** The structure of the DE-GWO.

The DE-GWO algorithm's operational approaches are as follows.

#### Step 1. Population initialization

The standard DE algorithm uses real number coding. First, generate an initial population with a scale of  $[N, D]$ .  $N$  represents the number of individuals in the population, and  $D$  represents the number of decision variables. The initial population can be expressed as follows:

$$x = \{x_{1,D}, x_{2,D}, \dots, x_{N,D}\}, \quad (5)$$

where  $x$  represents the initial population and  $x_{i,D}$  represents an individual in the population.

#### Step 2. Mutation operation with GWO

The mutation operator used by the traditional mutation strategy is shown as follows [33]:

$$V_{i,g} = X_{r0,g} + F(X_{r1,g} - X_{r2,g}), i \neq r0 \neq r1 \neq r2, \quad (6)$$

where  $V_{i,g}$  is a mutation individual,  $X_{r0,g}$ ,  $X_{r1,g}$  and  $X_{r2,g}$  are individuals randomly selected from the population which are different from each other, and  $F$  is the scaling factor, which is usually a float data type set between 0 and 2.

However, in this paper, the mutation operation is completed by GWO. The following are the precise steps of the operation. First, make a fitness calculation for each person in the population. The individuals of the first three optimal solutions are called the head wolf, represented by  $\alpha$ ,  $\beta$  and  $\delta$ , respectively, and the other individuals are grey wolves. Grey wolves update their hunting positions by simulating the hunting process according to the positions of the three head wolves. The mathematical formula for this process of hunting is as follows [34]:

$$D_\alpha = |C_1 X_\alpha(t) - X(t)|, \quad (7)$$

$$D_\beta = |C_2 X_\beta(t) - X(t)|, \quad (8)$$

$$D_\delta = |C_3 X_\delta(t) - X(t)|, \quad (9)$$

$$X_1 = X_\alpha(t) - A_1 D_\alpha, \quad (10)$$

$$X_2 = X_\beta(t) - A_2 D_\beta, \quad (11)$$

$$X_3 = X_\delta(t) - A_3 D_\delta, \quad (12)$$

where  $X$  represents an individual in the population,  $C$  represents the wobble factor, which is usually a float data type set randomly between 0 and 2,  $D$  represents the distance between the head wolves and the grey wolves, and  $A$  represents the convergence factor, which can decide whether the grey wolves move towards the head wolves' position. The new position of the individual in the population is expressed as follows:

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3}, \quad (13)$$

where  $X(t+1)$  represents an individual in the next generation population.

Finally, after all grey wolves have updated their positions, the new population is the mutation population of the DE algorithm.

#### Step 3. Crossover operation

The population variety may be further increased using the crossover approach. In this paper, the binomial crossover is the employed crossover technique [33]:

$$U_{j,i,g} = \begin{cases} V_{j,i,g}, & \text{randj}(0,1) \leq Cr \text{ or } j = jrand \\ X_{j,i,g}, & \text{others} \end{cases}, \quad (14)$$

where  $U_{j,i,g}$  represents a crossover individual and  $Cr$  is the crossover operator, which is often a float data type set at random between 0 and 1. The crossover individual is the mutation individual if  $\text{randj}(0,1) \leq Cr$  or  $j = jrand$ , and the original individual for the others.

#### Step 4. Select operation

The selection strategy is the last step of the DE algorithm. The experimental vector obtained through the crossover and the mutation is compared with the original vector. In this paper, to solve the minimum optimization problem and to enter the next generation, individuals with smaller fitness functions are chosen. It can be stated as follows [33]:

$$X_{i,g+1} = \begin{cases} U_{i,g}, & f(U_{i,g}) \leq f(X_{i,g}) \\ X_{i,g}, & \text{others} \end{cases}, \quad (15)$$

where  $X_{i,g+1}$  is the next generation,  $U_{i,g}$ , is the individual after the crossover operation,  $X_{i,g}$  is the parent individual before the mutation and crossover operations, and  $f$  stands for the fitness function.



### 3. Data

#### 3.1. CWRU bearing datasets

The Rolling Bearing Data Center at CWRU provided the defect diagnosis datasets for this study. This paper employs a 48k drive end bearing fault data, a 3 hp motor load, and a 1,730 rpm motor speed. Table 1 displays a description of the data.

**Table 1.** The data description of CWRU.

Fault Type	Fault Size	Fault Position	Training Data sets	Testing Data sets	Label
Inner Race Fault	0.1778	—	320	80	0
Ball Fault	0.1778	—	320	80	1
Outer Race Fault	0.1778	Centered@6:00	320	80	2
Inner Race Fault	0.3556	—	320	80	3
Ball Fault	0.3556	—	320	80	4
Outer Race Fault	0.3556	Centered@6:00	320	80	5
Inner Race Fault	0.5334	—	320	80	6
Ball Fault	0.5334	—	320	80	7
Outer Race Fault	0.5334	Centered@6:00	320	80	8
Normal	—	—	320	80	9

The experimental data consist of 10 types, including 9 fault types and 1 normal type. At the hour mark, the inner race, ball, and outside race are where the fault is placed; the corresponding fault diameters are 0.1778 mm, 0.3556 mm, and 0.5334 mm, respectively. Each type has 400 samples, of which 320 serve as training data sets and 80 serve as test data sets; each sample includes 1024 vibration data. To solve the problem that classifiers are not good at processing attribute data, the one-hot coding is used to replace the real number encoding.

#### 3.2. JNU bearing datasets

The JNU datasets, obtained from Jiangnan University in China [35], is a comprehensive collection of bearing data. The datasets were generated using a centrifugal fan system test bed specifically designed for fault diagnosis. The test bed utilized a Mitsubishi SB-JR induction motor, with the rotor supported by two bearings, one of which was intentionally faulty. To capture the vibration signals, accelerometers were strategically placed in the vertical direction of the bearings. The datasets consider four distinct health states: normal condition (N), inner ring failure (IF), outer ring failure (OF), and rolling body failure (BF). The vibration acceleration signals were meticulously collected at three different speeds - 600, 800, and 1000 rpm - with a sampling frequency of 50 kHz, ensuring rich and diverse datasets for various analytical purposes. In this paper, four operating states of the tachometer 600 are used to carry out troubleshooting experiments. Table 2 displays a description of the data.

The experimental data in this study is comprised of four distinct types, with three fault types and one normal type. The faults are deliberately placed in the inner ring, outer ring, and rolling body of the bearing. For each type, there are 480 available samples, with 400 samples designated as training data sets and 80 samples as test data sets. Each sample contains 1,024 data points of vibration information, thereby providing ample data for analysis and model training.

This article opted not to process the data, and instead conducted an analysis and research using

the original datasets. The objective was to preserve the salient features of the original signal with the utmost fidelity, thus enabling a better comprehension of the intrinsic properties and characteristics of the data.

The test was conducted on a computer with an i7-12700 CPU, featuring a main frequency of 3.6 GHz and 32 GB of memory. The programming software used was Python 3.9.7, with the TensorFlow2.0 environment developed by Google. The framework utilized was Keras, and the model was sequential.

**Table 2.** The data description of JNU.

Fault Type	Training Data sets	Testing Data sets	Label
Inner Ring Failure (IF)	400	80	0
Outer Ring Failure (OF)	400	80	1
Rolling Body Failure (BF)	400	80	2
Normal	400	80	3

## 4. Results and discussion

### 4.1. The performance of the DE-GWO method

To verify the optimization ability of the proposed algorithm under unimodal and multimodal functions, four standard test functions [36,37] are selected and used for simulation experiments, where the functions F1–F2 are unimodal, whereas F3–F4 are multimodal, and the proposed algorithm's convergence results are contrasted with those of other optimization algorithms and the optimization performance is analyzed. The expressions, search intervals and theoretical optimal values of the four standard test functions are listed in Table 3.

**Table 3.** The standard test functions.

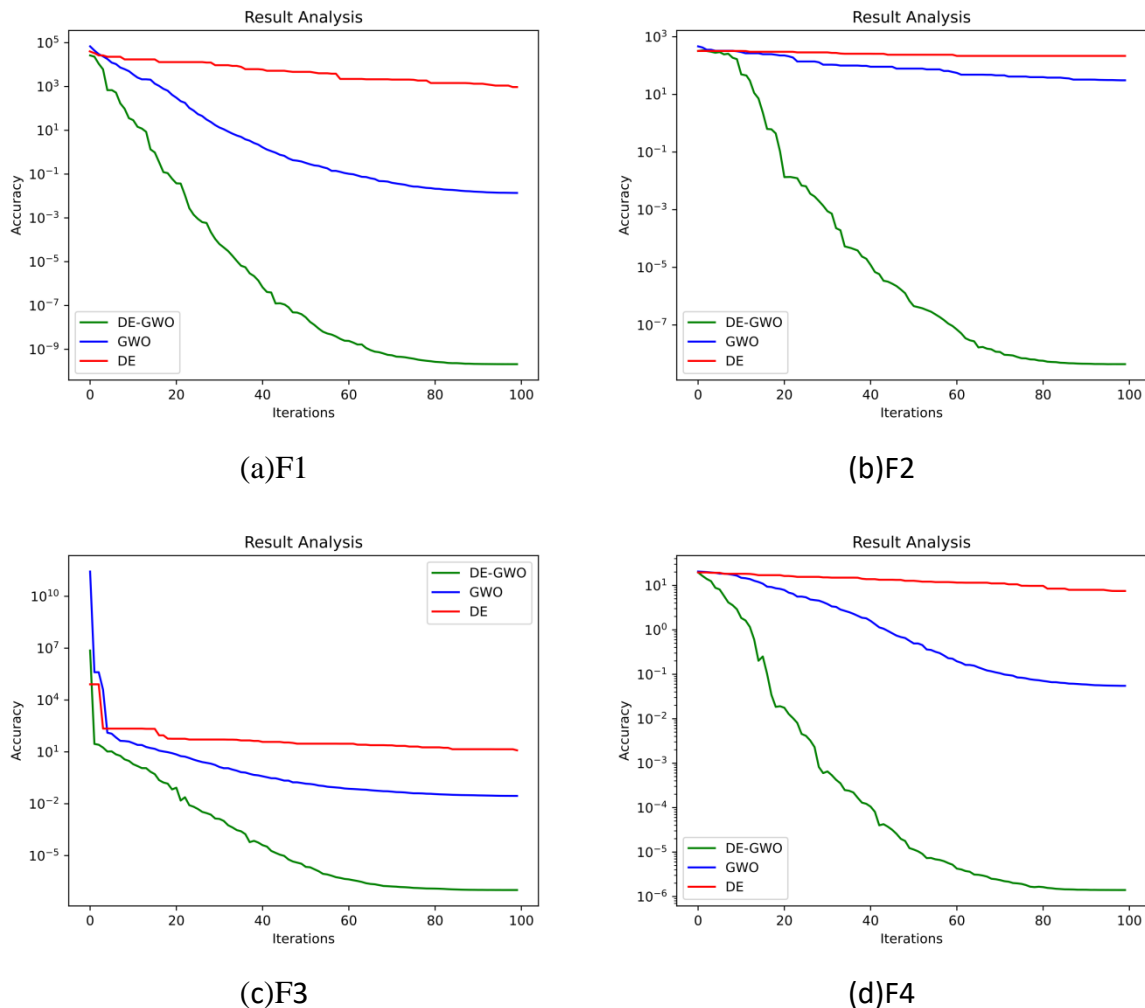
Function	Function Expression	d	S	$f_{\min}$
F1	$F_1 = \sum_{i=1}^d x_i^2$	30	[-100,100]	0
F2	$F_2 = \sum_{i=1}^d  x_i  + \prod_{i=1}^d  x_i $	30	[-10,10]	0
F3	$F_3 = \sum_{i=1}^d [x_i^2 - 10 \cos(2\pi x_i) + 10]$	30	[-5.12,5.12]	0
F4	$F_4 = -20 \exp\left(-0.2 \sqrt{\frac{1}{d} \sum_{i=1}^d x_i^2}\right) - \exp\left(\frac{1}{d} \sum_{i=1}^d \cos 2\pi x_i\right) + 20 + e$	30	[-32,32]	0

To compare the proposed DE-GWO method, the DE method, and the GWO method's optimum performances, the three algorithms were used to solve the four standard test functions listed in Table 1. For the fairness of comparison, we set the population number to 30, the maximum number of iterations was 100, the crossover rates of DE-GWO and DE were both 0.5, and the mutation rate of DE was 0.5. For each test function, the three algorithms were independently run 20 times under the dimensions  $d = 30$ . The solution results are shown in Table 4.

The table shows that the proposed DE-GWO algorithm performs better in terms of convergence accuracy and robustness compared to either the single DE or single GWO algorithms, whether in unimodal or multimodal functions. This supports the efficacy of the enhancement technique suggested in this research. Figure 5 displays the convergence curves of the methodologies for the four test functions.

**Table 4.** The solution results of the test functions.

Test function	d	DE-GWO		DE		GWO	
		mean	SD	mean	SD	mean	SD
F1	30	2.42E-10	2.40E-10	1.27E+3	3.65E+2	1.82E-2	9.32E-2
F2	30	2.95E-7	1.76E-7	1.77E+1	1.60E+0	2.49E-2	8.42E-3
F3	30	3.24E-9	1.62E-9	2.12E+2	1.33E+1	3.18E+1	6.49E+0
F4	30	4.31E-6	2.10E-6	8.94E+0	7.39E-1	2.3E-2	5.6E-3

**Figure 5.** The convergence curves of the three algorithms for the four test functions.

#### 4.2. 1DCNN optimized by DE-GWO

The DE-GWO proposed in this paper is used to calculate the result in order to choose the hyperparameters of 1D-CNN with the highest fitness; then, the attention mechanism is added in the fault diagnosis model.

This paper chooses six hyperparameters that need to be optimized using DE-GWO in the 1D-CNN, as they have a great impact on the algorithm performance through numerous experiments, namely the number of convolution kernels, convolution step, dropout rate, pooling step, learning rate and batch size. An appropriate number of convolution kernels can avoid both underfitting and overfitting. The convolution step can affect the training speed and the ability of extract features. The

dropout rate can effectively reduce the risk of overfitting of the model. The pooling step can control the complexity of the model. The learning rate can influence the convergence speed. The batch size has a significant impact on the convergence rate.

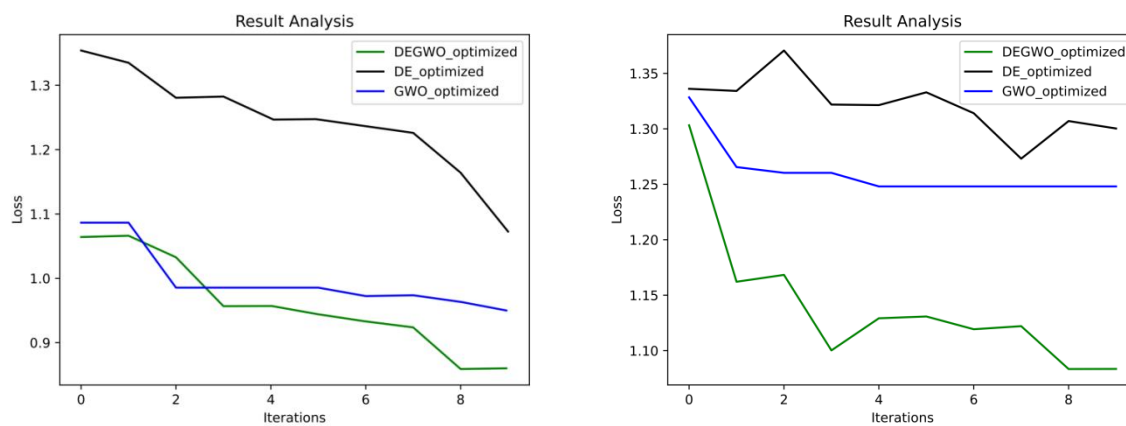
Before using the optimization algorithm, the value range of these six hyperparameters should be determined. According to experience, this paper sets the range of the number of convolution cores as 1 to 100, the convolution step size as 1 to 100, the dropout rate as 0.1 to 0.9, the pooling step size as 1 to 100, the batch size as 10 to 100, and the learning rate as 0 to 1.

The parameters of the DE-GWO algorithm are as follows: the maximum evolution times is 10, the number of populations is 10, the dimension is 6 and the cross probability is 0.5.

This paper uses rolling bearing data to train the model; the fitness function of the DE-GWO algorithm is as follows:

$$f(x_{i,g}) = A_{x_{i,g}}, \quad (16)$$

where  $x_{i,g}$  represents the  $i$ th individual of the  $g$  generation and  $A_{x_{i,g}}$  represents the loss result of one iteration of the 1D-CNN. Figure 6 shows the convergence curves of the three optimized algorithms for 1D-CNN.



(a) The convergence curves using CWRU datasets

(b) The convergence curves using JNU datasets

**Figure 6.** The convergence curves of the three optimized algorithms for 1D-CNN.

After the optimization of the DE-GWO, DE and GWO, the fitness hyperparameters of the 1D-CNN using different algorithms and normal 1D-CNN with hyperparameters selected by experience are shown in Table 5.

**Table 5(a).** The hyper parameters of 1D-CNN using different algorithms for CWRU datasets.

Algorithm	KC	KL	PL	Dropout	Learning Rate	Batch Size
DE-GWO	10	64	9	0.26524	0.00423	42
DE	17	54	18	0.38586	0.00402	17
GWO	42	40	36	0.26008	0.00516	33
normal	8	8	3	0.2	0.004	40

**Table 6(b).** The hyper parameters of 1D-CNN using different algorithms for JNU datasets.

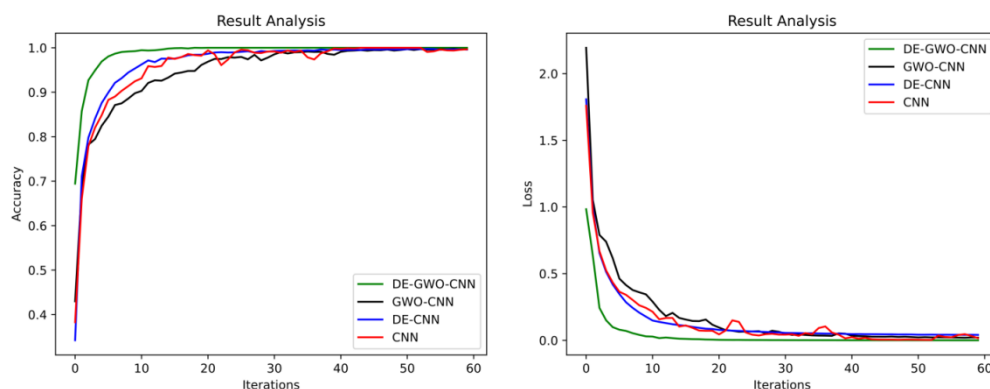
Algorithm	KC	KL	PL	Dropout	Learning Rate	Batch Size
DE-GWO	14	90	19	0.16168	0.00659	32
GWO	8	72	40	0.38763	0.00727	25
DE	58	50	22	0.38547	0.00519	12
normal	8	8	3	0.2	0.004	40

The padding of the model is the "same", the activation of the convolution layer is relu and the classifier is softmax. To enhance fault feature information of 1D-CNN, the attention mechanism is added after the maxpooling layer.

### 4.3. Fault diagnosis of rolling bearings based on DE-GWO-CNN

#### 4.3.1. Fault diagnosis experiments using the CWRU datasets

Taking categorical\_crossentropy as the loss function to update model parameters, 80% of the data set is used as the training data set of the algorithm and the remaining 20% as the test data set. Then, DE-GWO-CNN with attention mechanism, DE-CNN, GWO-CNN and 1D-CNN are used to diagnose rolling bearing faults. The iteration number of the algorithm is set as 60. The fault diagnosis accuracy and loss function of the four algorithms are shown in Figure 7.



(a) The accuracy of the fault diagnosis methods

(b) The loss of the fault diagnosis methods

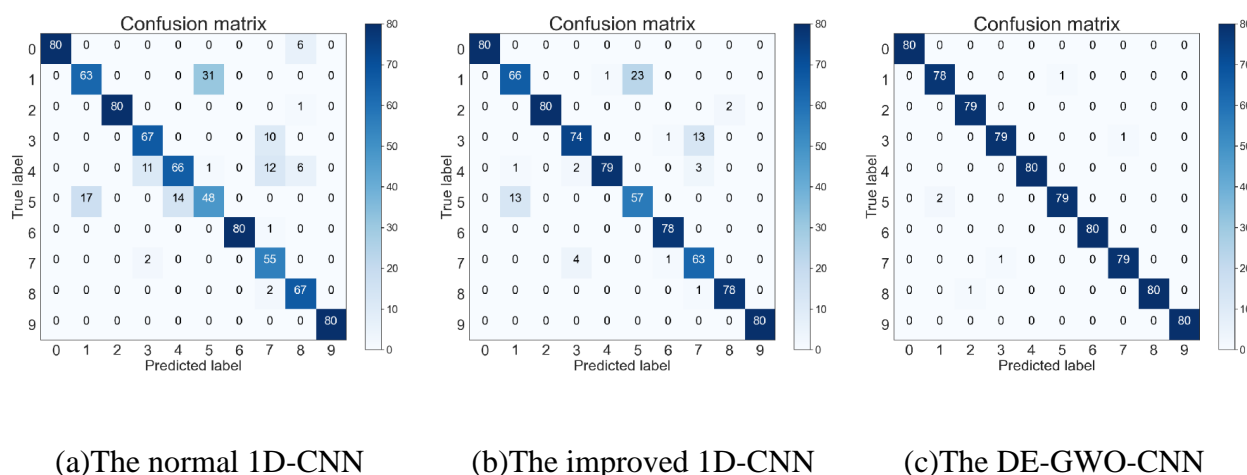
**Figure 7.** The accuracy and loss of the fault diagnosis methods.

The four methods in the training set can all achieve 100% accuracy after 40 iterations; however, the approach suggested in this study has the fastest convergence speed, a smoother curve, and superior stability. Table 6 displays the various algorithms' fault diagnostic accuracy and validation time for the test data set. Though not statistically different from other algorithms, the proposed fault diagnosis method does not have the shortest validation time. However, its accuracy is the highest among these algorithms, indicating that it is effective for diagnosing faults in rolling bearings.

To further clarify the diagnostic superiority of the algorithm proposed in this paper, the confusion matrix of the normal 1D-CNN, the improved 1D-CNN and the DE-GWO-CNN are compared with each other. The results are shown in Figure 8, which express the accuracy of different fault diagnosis benchmarks for different fault types.

**Table 7.** The accuracy and calculation time of fault diagnosis in the test data set.

Fault diagnosis model	DE-GWO-CNN	DE-CNN	GWO-CNN	1D-CNN
Accuracy	99.25%	97.25%	92.875%	85.75%
Recall	99.253%	96.966%	93.011%	85.623%
F1	99.251%	97.108%	92.893%	85.686%
Validation time (s/epoch)	0.36	0.68	0.29	0.27

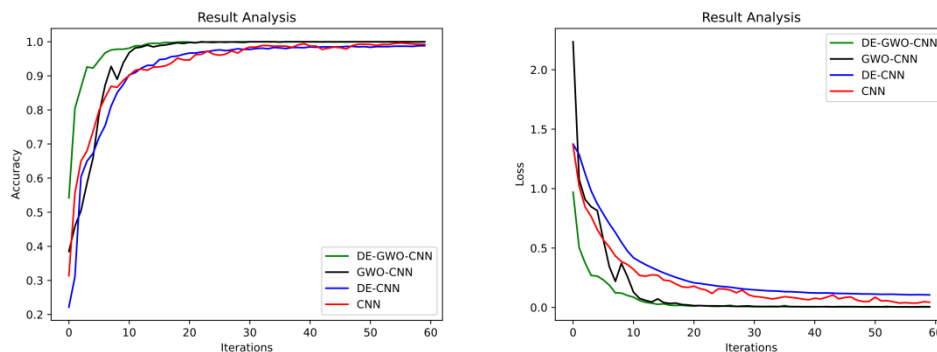
**Figure 8.** The confusion matrix of the different algorithms.

The figure indicates that the proposed algorithm achieves a diagnostic accuracy of 100% for states one, five, seven, nine and ten, while the lowest diagnostic accuracy for the remaining five states is 97.5%. Compared to the normal 1D-CNN and the improved 1D-CNN, the proposed method achieves the highest accuracy, indicating that the fault diagnosis model presented in this paper has a superior diagnostic ability.

#### 4.3.2 Fault diagnosis experiments using the JNU datasets

Taking categorical\_crossentropy as the loss function to update model parameters, 75% of the data set is used as the training data set of the algorithm and the remaining 25% as the test data set. Then, DE-GWO-CNN with attention mechanism, DE-CNN, GWO-CNN and 1D-CNN are used to diagnose rolling bearing faults. The iteration number of the algorithm is set as 60. The fault diagnosis accuracy and loss function of the four algorithms are shown in Figure 9.

The GWO-CNN and DE-GWO-CNN algorithms achieved 100% accuracy after just 30 iterations in the training set, outperforming other methods that required more iterations. The proposed approach in this study exhibits the fastest convergence speed, a smoother curve, and superior stability. Table 7 presents the fault diagnostic accuracy and validation time for various algorithms on the test datasets. While the proposed fault diagnosis method does not have the shortest validation time, the difference is not significant compared to other algorithms. More importantly, the method achieves the highest accuracy among these algorithms, demonstrating its effectiveness in diagnosing faults in rolling bearings. This superior accuracy justifies the use of the proposed method for fault diagnosis in rolling bearing.



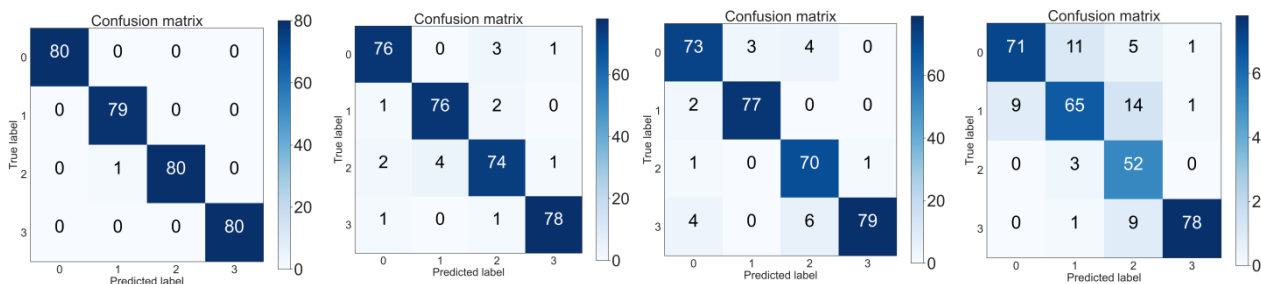
(a)The accuracy of the fault diagnosis methods (b)The loss of the fault diagnosis methods

**Figure 9.** The accuracy and loss of the fault diagnosis methods using JNU datasets.

**Table 8.** The accuracy and calculation time of fault diagnosis in the test data set.

Fault diagnosis model	DE-GWO-CNN	GWO-CNN	DE-CNN	1D-CNN
Accuracy	99.69%	95.00%	92.43%	83.13%
Precision	99.69%	95.00%	92.43%	83.13%
Recall	100%	95.02%	93.68%	84.22%
F1	99.84%	95.01%	93.05%	83.67%
Validation time (s/epoch)	0.36	0.46	0.28	0.22

To further demonstrate the diagnostic superiority of the algorithm proposed in this paper, a comparison of confusion matrices is conducted between the normal 1D-CNN, the DE-CNN, the GWO-CNN and the DE-GWO-CNN. As shown in Figure 10, the results illustrate the accuracy of different fault diagnosis methods for various fault types. This comparison effectively highlights the improved performance of the proposed algorithm.



(a)The DE-GWO-CNN (b)The GWO-CNN (c)The DE-CNN (d)The normal 1D-CNN

**Figure 10.** The confusion matrix of the different algorithms.

Using Figure 10, the average recognition accuracy for each operating condition of different algorithms is calculated and the results are listed in Table 8.

According to the table, the proposed algorithm demonstrates a diagnostic accuracy of 100% for states one, three, and four, with a slightly lower accuracy of 98.75% for the remaining states. In comparison to other algorithms, the proposed method exhibits the highest accuracy. These results indicate that the fault diagnosis model presented in this paper possesses superior diagnostic

capabilities.

**Table 9.** The accuracy for each operation condition of different algorithms.

Fault diagnosis model	DE-GWO-CNN	GWO-CNN	DE-CNN	1D-CNN
IF	100%	95.00%	91.25%	88.75%
OF	98.75%	95.00%	96.25%	81.25%
BF	100%	92.50%	87.50%	65.00%
Normal	100%	97.50%	98.75%	97.50%

#### 4.3.3 Comparison with other methods

To demonstrate the superiority of the proposed fault diagnosis algorithm, a quantitative comparison was made between the algorithm proposed in this paper and DE-GWO-CNN without an attention mechanism, as well as three classic models: GoogLeNet, LeNet-5, and AlexNet. All models were trained on the CWRU datasets but with different hyperparameters. The test results are presented in Table 9.

**Table 10.** Comparison between algorithms.

Fault diagnosis model	Accuracy
The proposed method	99.25%
DE-GWO-CNN without attention mechanism	94.50%
GoogLeNet	93.75%
LeNet-5	91.87%
AlexNet	96.13%

According to the table, when compared to DE-GWO-CNN without attention mechanism, GoogLeNet, LeNet-5, and AlexNet, the proposed algorithm demonstrates a superior accuracy in fault diagnosis. Specifically, the proposed method achieves 4.75%, 5.5%, 7.48%, and 2.92% higher accuracy than these respective models. This suggests that the proposed algorithm may offer a more effective solution for fault diagnosis.

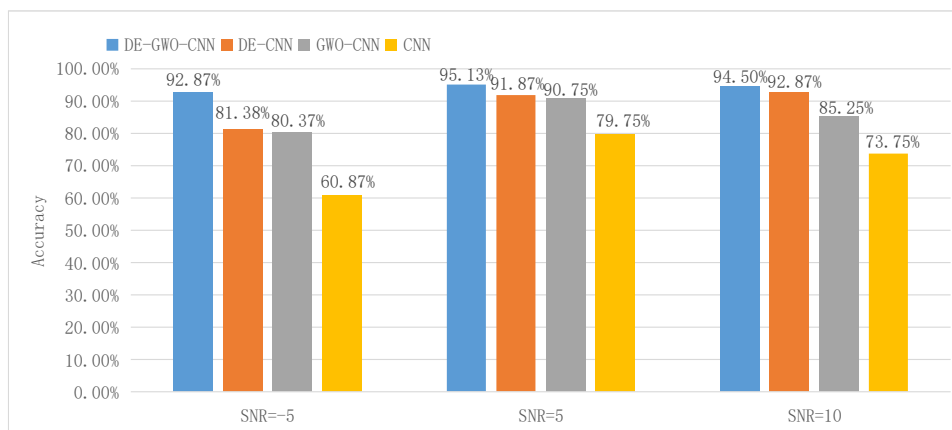
Finally, to verify the anti-noise ability of the algorithm proposed in this paper, Gaussian white noise is added to the original rolling bearing vibration data, and the signal-noise ratio (SNR) is used to measure the noise, which is shown as follows [38]:

$$SNR = 10 \log_{10} \frac{P_{signal}}{P_{noise}} \quad (17)$$

where  $P_{signal}$  represents the power of the signal and  $P_{noise}$  represents the power of the noise.

This paper sets up three sets of noise experiments, with signal-to-noise ratios of -5 dB, 5 dB, and 10 dB. The four algorithms used for fault diagnosis are DE-GWO-CNN, DE-CNN, GWO-CNN, and normal 1D-CNN. The results are presented in Figure 11.





**Figure 11.** Diagnostic accuracy of each algorithm under noise interference.

The experimental findings clearly demonstrate that the suggested algorithm outperforms the other three algorithms in terms of fault diagnostic performance for these three types of Gaussian white noise. Moreover, the results also highlight the program's exceptional anti-noise capability and its potential for higher engineering application value.

## 5. Conclusions

While the 1D-CNN algorithm is effective in feature extraction, selecting optimal hyperparameters can be challenging and its feature representation ability may not always be satisfactory. To address these issues, we propose a novel approach that utilizes the DE-GWO intelligent optimization algorithm to optimize the hyperparameters of 1D-CNN. By finding the most suitable hyperparameters, our method enhances the ability of the 1D-CNN model to extract features from bearing vibration signals and improve the fault classification accuracy. Additionally, we enhance the feature representation ability of 1D-CNN by incorporating an attention mechanism. Comparative tests demonstrate that our proposed algorithm improves the accuracy and convergence speed of fault diagnosis. Notably, the algorithm exhibits high anti-noise performance, as it achieves a relatively higher accuracy even when the signal-to-noise ratio is set to -5 dB, 5 dB, and 10 dB.

Furthermore, although the diagnostic accuracy of the rolling bearing fault diagnosis model proposed in this paper is reasonably high, there is still room for improvement in terms of its transferability. Therefore, our future research efforts will be focused on enhancing the transferability of the fault diagnosis model, with the aim of facilitating its broader application in real-world industrial settings.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

This research was supported by Key Research and Development Project of Liaoning Province (Grant No. 2020JH2/10300040); Applied Basic Research Project of Liaoning Province, China (Grant No. 2023JH2/101300183).

## Conflict of interest

All authors declare that no conflict of interest exists.

## References

1. C. Liu, X. Zhang, Z. Bao, Z. He, M. Gao, W. Song, A novel deep transfer learning method for intelligent fault diagnosis based on variational mode decomposition and efficient channel attention, *Entropy*, **24** (2022),1087. <https://doi.org/10.3390/e24081087>
2. Z. Hou, J. Zeng, Condition monitoring technology for bearing ring groove grinding, *J. Phys. Conf. Ser.*, **1213** (2019), 052036. <https://doi.org/10.1088/1742-6596/1213/5/052036>
3. X. Gao, H. Wei, T. Li, G. Yang, A rolling bearing fault diagnosis method based on LSSVM, *Adv. Mechan. Eng.*, **12** (2020), 1–10. <https://doi.org/10.1177/1687814019899561>
4. Z. Yan, G. Liu, J. Wang, H. Bao, Z. Zhang, X. Zhang, et al., A new universal domain adaptive method for diagnosing unknown bearing faults, *Entropy*, **23** (2021), 1052. <https://doi.org/10.3390/e23081052>
5. J. Tang, J. Hu, J. Qing, T. Kang, Rolling bearing fault monitoring for sparse time-frequency representation and feature detection strategy, *Entropy*, **24** (2022), 1822. <https://doi.org/10.3390/e24121822>
6. Y. Zhang, L. Duan, M. L. Duan, A new feature extraction approach using improved symbolic aggregate approximation for machinery intelligent diagnosis, *Measurement*, **133** (2019), 468–478. <https://doi.org/10.1016/j.measurement.2018.10.045>
7. Z. Wang, X. He, B. Yang, N. Li, Subdomain adaptation transfer learning network for fault diagnosis of roller bearings, *IEEE Trans. Ind. Electron.*, **69** (2022), 8430–8439. <https://doi.org/10.1109/TIE.2021.3108726>
8. W. Li, Y. Cao, L. Li, S. Hou, An orthogonal wavelet transform-based k-nearest neighbor algorithm to detect faults in bearings, *Shock Vibr.*, **2022** (2022),1–13. <https://doi.org/10.1155/2022/5242106>
9. D. Wu, C. Jemmings, J. Terpenning, A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests, *J. Manuf. Sci. Eng. Trans. ASME*, **139** (2017), 179–187. <https://doi.org/10.1115/1.4036350>
10. S. Guo, B. Zhang, T. Yang, D. Lyu, W. Gao, Multitask convolutional neural network with information fusion for bearing fault diagnosis and localization, *IEEE Transact. Industr. Electron.*, **67** (2020), 8005–8015. <https://doi.org/10.1109/TIE.2019.2942548>
11. H. Pan, X. He, S. Tang, F. Meng, An improved bearing fault diagnosis method using one-dimensional CNN and LSTM, *J. Mechan. Eng.*, **64** (2018), 443–452. <https://doi.org/10.5545/sv-jme.2018.5249>
12. Z. Meng, W. Cao, D. Sun, Q. Li, W. Ma, F. Fan. Research on fault diagnosis method of MS-CNN rolling bearing based on local central moment discrepancy, *Adv. Eng. Inform.*, **54** (2022), 101797. <https://doi.org/10.1016/j.aei.2022.101797>
13. W. Deng, H. Liu, J. Xu, H. Zhao, Y. Song, An improved quantum-inspired differential evolution algorithm for deep belief network, *IEEE Transact. Instrument. Measur.*, **69** (2020), 7319–7327. <https://doi.org/10.1109/TIM.2020.2983233>
14. J. Yuan, R. Zhao, T. He, P. Chen, K. Wei, Z. Xing, Fault diagnosis of rotor based on Semi-supervised Multi-Graph Joint Embedding, *ISA Trans.*, **131** (2022), 516–532, <https://doi.org/10.1016/j.isatra.2022.05.006>

15. Q. Ni, J.C. Ji, H. Benjamin, K. Feng, K. A. Nandi, Physics-Informed Residual Network (PIResNet) for rolling element bearing fault diagnostics, *Mechan. Syst. Signal Process.*, **200** (2023), 110544. <https://doi.org/10.1016/j.ymsp.2023.110544>
16. K. Feng, J. C. Ji, Y. Zhang, Q. Ni, Z. Liu, B. Michael, Digital twin-driven intelligent assessment of gear surface degradation, *Mechan. Syst. Signal Process.*, **186** (2023), 109896. <https://doi.org/10.1016/j.ymsp.2022.109896>
17. J. Yuan, R. Zhao, P. Chen, T. He, K. Wei, Dimensionality reduction using local-global standard hypergraph embedding for rotor fault diagnosis, *Meas. Sci. Technol.*, **34** (2023), 034006. <https://doi.org/10.1088/1361-6501/acab1e>
18. P. Wang, R. Yan, R. Gao, Virtualization and deep recognition for system fault classification, *J. Manuf. Syst.*, **44** (2017), 310–316. <https://doi.org/10.1016/j.jmsy.2017.04.012>
19. J. Qiu, H. Tao, L. Cheng, L. Shen, Bearing fault diagnosis based on self-attention mechanism ACGAN, *Inform. Control.*, **51** (2022), 753–762. <https://doi.org/10.13976/j.cnki.xk.2022.2002>
20. S. Gao, S. Shi, Y. Zhang, Rolling bearing compound fault diagnosis based on parameter optimization MCKD and convolutional neural network, *IEEE Trans. Instrum. Meas.*, **71** (2022), 1–8. <https://doi.org/10.1109/TIM.2022.3158379>
21. D. Hoang, H. Kang, Rolling element bearing fault diagnosis using convolutional neural network and vibration image, *Cogn. Syst. Res.*, **53** (2018), 42–50. <https://doi.org/10.1016/j.cogsys.2018.03.002>
22. M. Wang, W. Wang, X. Zhang, Iu HH-C, A new fault diagnosis of rolling bearing based on markov transition field and CNN, *Entropy*, **24** (2022), 751. <https://doi.org/10.3390/e24060751>
23. X. Zhao, K. Ji, M. Zhang, H. Huang, F. Wang, Y. Liu, et al., Intelligent fault diagnosis of gearbox under variable working conditions with adaptive intraclass and interclass convolutional neural network, *IEEE Transact. Neural Networks Learn. Syst.*, **34** (2023), 6339–6353. <https://doi.org/10.1109/TNNLS.2021.3135877>
24. W. Deng, X. Zhang, Y. Zhou, Y. Liu, X. Zhou, H. Chen, et al., An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems, *Inform. Sci.*, **585** (2022), 441–453. <https://doi.org/10.1016/j.ins.2021.11.052>
25. Q. Sun, Y. Li, Research on fault diagnosis of rolling bearings based on de algorithm optimization of CNN, *Noise Vibr. Control*, **42** (2022), 165–171. <https://doi.org/10.3969/j.issn.1006-1355.2022.04.027>
26. Y. Li, J. Ma, L. Jiang, Fault diagnosis of rolling bearing based on an improved convolutional neural network using SFLA, *J. Vibr. Shock*, **39** (2020), 187–193. <https://doi.org/10.13465/j.cnki.jvs.2020.24.026>
27. X. Wang, X. Liu, J. Wang, X. Xiong, S. Bi, Z. Deng, Improved variational mode decomposition and one-dimensional CNN network with parametric rectified linear unit (PReLU) approach for rolling bearing fault diagnosis, *Appl. Sci.*, **12** (2022), 9324. <https://doi.org/10.3390/app12189324>
28. G. Ning, D. Cao, Y. Zhou, Improved differential evolution algorithm for solving 0-1 programming problems, *J. Syst. Sci. Math. Sci.*, **39** (2019), 120–132. <https://doi.org/10.12341/jssms13549>
29. X. Zhang, X. Wang, Comprehensive review of grey wolf optimization algorithm, *Computer Sci.*, **46** (2019), 30–38. <https://doi.org/10.11896/j.issn.1002-137X.2019.03.004>
30. J. Jiao, M. Zhao, J. Lin, K. Liang, A comprehensive review on convolutional neural network in machine fault diagnosis, *Neuro. Comput.*, **471** (2020), 36–63. <https://doi.org/10.1016/j.neucom.2020.07.088>
31. G. Brauwers, F. Frasincar, A general survey on attention mechanisms in deep learning, *IEEE*

- Transact. Knowl. Data Eng.*, **35** (2023), 3279–3298.  
<https://doi.org/10.1109/TKDE.2021.3126456>
32. N. Nahak, R. K. Mallick, Damping of power system oscillations by a novel DE-GWO optimized dual UPFC controller, *Eng. Sci. Technol. Int. J.*, **20** (2017), 1275–1284.  
<https://doi.org/10.1016/j.jestch.2017.09.001>
33. M. Yu, D. Wang, Model-based health monitoring for a vehicle steering system with multiple faults of unknown types, *IEEE Transact. Industr. Electron.*, **61** (2014), 3574–3586.  
<https://doi.org/10.1109/TIE.2013.2281159>
34. Z. Jin, D. He, Z. Wei, Intelligent fault diagnosis of train axle box bearing based on parameter optimization VMD and improved DBN, *Eng. Appl. Artif. Intell.*, **110** (2022), 104713.  
<https://doi.org/10.1016/j.engappai.2022.104713>
35. K. Li, X. Ping, H. Wang, P. Chen, Y. Cao, Sequential fuzzy diagnosis method for motor roller bearing in variable operating conditions based on vibration analysis, *Sensors*, **13** (2013), 8013–8041. <https://doi.org/10.3390/s130608013>
36. Q. Gu, X. Li, C. Lu, Hybrid genetic greywolf algorithm for high dimensional complex function optimization, *Control Decision*, **35** (2020), 1191–1198.  
<https://doi.org/10.13195/j.kzyjc.2018.1253>
37. Z. Teng, J. Lv, L. Guo, An improved hybrid greywolf optimization algorithm, *Soft Computing*, **23** (2019), 6617–6631. <https://doi.org/10.1007/s00500-018-3310-y>
38. R. Sun, J. Yang, D. Yao, J. Wang, A new method of wheelset bearing fault diagnosis, *Entropy*, **24** (2022), 1381. <https://doi.org/10.3390/e24101381>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)