Article

Improved Variational Mode Decomposition and One-Dimensional CNN Network with Parametric Rectified Linear Unit (PReLU) Approach for Rolling Bearing Fault Diagnosis

Xiaofeng Wang, Xiuyan Liu *, Jinlong Wang, Xiaoyun Xiong, Suhuan Bi and Zhaopeng Deng

School of Information and Control Engineering, Qingdao University of Technology, No. 777 Jialingjiang East Rd., Qingdao 266525, China

* Correspondence: liuxiuyan@qut.edu.cn

Abstract: As a critical component of rotating machinery, rolling bearings are essential for the safe and efficient operation of machinery. Sudden faults of rolling bearings can lead to unscheduled downtime and substantial economic costs. Therefore, diagnosing and identifying the equipment status is essential for ensuring the operation and decreasing the additional maintenance costs of the machines. However, extracting the features from the early bearing fault signals is challenging under background noise interference. With the purpose of solving the above problem, we propose an integrated rolling bearing fault diagnosis model based on the improved grey wolf optimized variational modal decomposition (IGVMD) and an improved 1DCNN with a parametric rectified linear unit (PReLU). Firstly, an improved grey wolf optimizer (IGWO) with the fitness function, the minimum envelope entropy, is designed for adaptively searching the optimal parameter values of the VMD model. The performance of the basic grey wolf optimizer (GWO) algorithm by introducing three improvement strategies, the non-linear convergence factor adjustment strategy, the grey wolf adaptive position update strategy, and the Levy flight strategy in the IGWO algorithm, is improved. Then, an improved 1DCNN model with the PReLU activation function is proposed, which extracts the bearing fault features, and a grid search to optimize the model parameters of the 1DCNN is introduced. Finally, the effectiveness of the proposed model is demonstrated well by employing two experimental datasets. The preliminary comparative results of the average identification accuracy in the proposed method in two datasets are 99.98% and 99.50%, respectively, suggesting that this proposed method has a relatively higher recognition accuracy and application values.

Keywords: bearing fault diagnosis; grey wolf optimization (GWO); variational modal decomposition (VMD); convolutional neural network (CNN); parametric rectified linear unit (PReLU)

1. Introduction

As an important part of rotating equipment, rolling bearings have been applied in modern industrial areas such as shipbuilding, vehicles, and aerospace. The failure of rolling bearings can cause massive time and economic losses, especially for large plant equipment. Therefore, the early fault monitoring and diagnosis of rolling bearings are vital for the reliability and safety of industry equipment [1,2]. However, it is hard to diagnose faults of rolling bearings under complex non-linear and nonstationary vibration signals. It is an outstanding challenge to successfully extract vulnerable fault features under very noisy conditions.

Generally, the traditional time–frequency fault diagnosis feature extractors such as the classic Fourier transform (FFT), wavelet transforms (WT), and empirical mode decomposition (EMD) are commonly used to extract features and diagnose faults. However, the vibration signals of faults are nonstationary and non-linear characteristics, so the analysis

method of a single time–frequency domain cannot adequately extract the nonstationary features from the fault data of rotating machinery [3]. Huang et al. [4] designed the EMD method as an adaptive analysis method employed extensively for mechanical fault detection and diagnosis. The EMD method can effectively reduce low-frequency interference by highlighting the high-frequency resonant part. However, EMD uses the recursive decomposition method to separate components with similar frequencies, which can easily cause the modal confusion and endpoint effects [5]. The errors in the EMD process are passed in layers, which are prone to false components and endpoint effects that affect the accuracy of fault feature extraction. To overcome these EMD disadvantages, many scholars have tried to adapt the EMD method [6,7]. These improved EMD methods have some inhibitory effects on the problems of mode mixing but cannot eliminate the effects of mode mixing. In 2014, Dragomiretskiy et al. [8] proposed a variational mode decomposition (VMD) method which has attracted increasing attention in bearing fault diagnosis in industrial engineering. As a non-recursive signal decomposition method, the VMD method, characterized by noise robustness and high accuracy, has an outstanding advantage in conquering the issues of modal confusion and endpoint effects compared with EMD [9]. Yang et al. [10] utilized VMD in the feature extraction of bearing signals, and the experimental results verified that VMD had good results in bearing fault detection. Although the method of the VMD has a better decomposition ability than the EMD, it also has some disadvantages in that the modal decomposition fraction K and the learning rate α must be set in advance by the people's prior experience. In addition, the accuracy of decomposition relies heavily on these parameters in the decomposition process of nonstationary signals. An improperly chosen size of the modal number K can cause over- or under-decomposition phenomena, and the quadratic penalty term α can affect the bandwidth of the modal components in the VMD solution. With the development of intelligence optimization algorithms such as particle swarm optimization (PSO), genetic, swarm, whale, and grey wolf (GWO), researchers have used these algorithms to optimize and solve the VMD parameters. Qing et al. [11] proposed a fault information guided variational decomposition method (FIVMD) which determines parameters α and K based on statistical thresholds of IGGCS\GGS, IGGCS\GGS, and RFCA, effectively extracting repeated transients of weak bearings but lacking adaptability. Deqiang et al. [12] proposed a VMD mode that adaptively obtained the globally optimal mode number K through a sparrow-inspired algorithm but omitted the impact of penalty factor α on signal decomposition. The GWO algorithm [13] is a heuristic method with strong global optimization searchability, which is more advantageous than other intelligent algorithms in finding the best parameter combination. Zhao et al. [14] successfully applied GWO to fault detection and verified the feasibility of GWO in the field of fault detection. However, the GWO algorithm still has some weaknesses, such as premature convergence of the algorithm, poor solution accuracy, and the tendency to fall into local optimality on some complex problems [15,16]. Therefore, many scholars have conducted relevant studies to improve the performance of GWO [17–19]. It has been demonstrated through existing research that there is significant room for performance improvement in the grey wolf algorithm. This study focuses on the following aspects to overcome these weaknesses. Firstly, a new non-linear convergence factor is designed to maintain a proper balance between exploration and exploitation. Secondly, Levy flight [20] is introduced to enhance the exploration capability of the algorithm by increasing the population diversity. Finally, an adaptively tuned position update strategy is utilized to speed up the convergence of the algorithm. This article proposes an improved grey wolf optimization algorithm (IGWO) based on the above strategies. Because the signal with a periodic pulse feature of the fault feature has strong signal sparsity, its envelope entropy is small, and conversely, its envelope entropy is larger. Therefore, the minimum envelope entropy [21] is introduced as the fitness function of the optimization algorithm, and the method of optimizing VMD parameters based on IGWO (IGVMD) is designed to extract bearing fault features.

The essence of bearing fault diagnosis is actually a classification problem. Many classification algorithms such as shallow artificial neural networks (ANNs), support vector

machines (SVMs), and convolutional neural networks (CNNs) have been proposed for classification problems in engineering applications. Compared with traditional machine learning algorithms, deep learning algorithms [22] have greater advantages in dealing with massive, high-dimensional, and noisy data, and can automatically and efficiently extract fault characteristics from massive data. Convolutional neural networks (CNNs) have achieved better results for mechanical faults [23]. Other commonly used methods, such as CNN-based VGG [24], ResNet [25], and other network architectures, have achieved great success in computer vision and successfully applied to fault diagnosis. Lu et al. [26] proposed a fault diagnosis method by converting original vibration signals into two-dimensional feature maps and feeding them into a CNN for training. Chen et al. [27] combined the continuous wavelet transform (CWT) method, the CNN model, and an extreme learning machine (ELM) classifier to extract high-level fault features. Jiang et al. [28] developed a novel multiscale convolutional neural network (MSCNN) to extract multiscale features of gearboxes and simultaneously recognize fault conditions. However, important features such as periodic or short-time pulse features are easily lost when they are transformed into two-dimensional feature maps. To avoid the abovementioned problems, Ince et al. [29] proposed the one-dimensional CNN method, which can directly extract the best features from vibration data. Compared with the two-dimensional CNN, it has higher recognition accuracy and faster convergence speed. Peng et al. [30] realized fault identification of rolling bearings through a one-dimensional CNN and had a strong generalization ability. Inspired by Ince and Peng et al., one-dimensional convolution is used to improve structures such as deep convolutional generative adversarial networks [31], residual neural networks [32], and long-short memory networks [33]. Habbouche et al. [34] applied the VMD method as a depression filter to eliminate the dominant modes and then input the features into a one-dimensional CNN model for fault diagnosis. However, the one-dimensional CNN still suffers from the shortcomings of a long training time and slow convergence in fault diagnosis, leading to low efficiency in fault identification.

A new framework that integrates grey wolf optimized variational modal decomposition (IGVMD) and a parametric rectified linear unit (PReLU) activation layer in a 1DCNN model is proposed in this paper for bearing fault diagnosis to overcome the above problems. Firstly, the grey wolf algorithm is improved based on three adjustment strategies, and then the improved GWO method is used to optimize the parameters of VMD (IGVMD). The objective of the IGVMD is to extract features of fault signals from a noisy background. The fault signal is decomposed by the IGVMD algorithm to obtain K modal components, and the valid modal components are obtained by correlation coefficient threshold screening. Then, a PReLU-based 1DCNN is introduced to promote the efficiency and accuracy of model training. IGVMD is used to preprocess the fault signals to effectively extract the fault features, and then the extracted fault features are used as the input of 1DCNN to train the network for accurate classification of the fault signals. The main contributions of this paper are summarized as follows:

(1) An improved grey wolf optimization algorithm is designed to improve the global search performance of the GWO by introducing three improved strategies. We propose a hybrid diagnosis framework for rolling bearings with integrated signal denoising and dynamic training. The improved grey wolf algorithm adaptive optimization variational modal method is used for effectively extracting weak fault signals under a noise background.

(2) A one-dimensional convolutional neural network is designed by introducing grid search and proposing a PReLU-1DCNN with a dynamically trained activation layer and adaptive learning rate. This can achieve high accuracy and a fast convergence rate by solving the problems of mean shift and negative interval death of neurons.

The arrangement of this paper is as follows. Section 2 briefly describes the relevant theories of the grey wolf algorithm, variational modal decomposition, and one-dimensional convolutional neural networks. Section 3 introduces the fault diagnosis framework based on IGVMD–1DCNN. In Section 4, the experimental dataset collected in the Electrical

Engineering Laboratory of Case Western Reserve University (CWRU) and the gearbox at Southeast University is used for experiments to assess the effectiveness of the proposed method, and the results are analyzed and discussed in detail. Finally, our conclusions and prospects are shown in Section 5.

2. Methodology

2.1. Introduction of Variational Modal Decomposition

The VMD algorithm is a decomposition method for vibration signal decomposition, which is completely non-recursive and adaptive [8]. Its basic process is the following: first, the sequence signal is decomposed into K modal components (IMF) so that the sum of the estimated frequency bandwidth of each mode is minimized, and the sum of the IMFs is the original decomposed signal; second, the bandwidth of each IMF component is estimated; lastly, the central frequency corresponding ω_k to each IMF component is extracted. This mainly includes two parts: variational problem construction and variational problem solving. The steps of variable fraction problem construction are as follows:

(1) The original signal is decomposed to a series of intrinsic mode functions by VMD algorithm, each of which can be regarded as the AM-modulated signal. The symbol is u_k, and its expression is:

$$u_k(t) = A_k(t) \cos[\varphi_k(t)] \tag{1}$$

(2) The original signal is obtained via a one-sided spectrum by being Hilbert-transformed. Then, the frequency spectrum of IMFs is modulated to the corresponding baseband by mixing the predicted central frequency spectrum with the analyzed signal. In addition, the method to estimate bandwidth is by calculating the L^2 parametric square of the signal gradient. The constrained variational model of the VMD algorithm is [8]:

$$\begin{cases} \min\left\{ \sum_K ||\partial_t\left[\left(\partial(t) + \frac{j}{\pi t}\right) * u_k(t)\right]e^{-j\omega_k t}||_2^2 \right\} \\ s.t. \ \sum_k u_k = f \end{cases} \tag{2}$$

Then, the steps of variational problem solving are as follows:

(3) The Lagrangian operator $\lambda(t)$ and the quadratic penalty factor α are introduced to the method to turn the constrained variational problem into an unconstrained variational problem. The extended Lagrangian expression is:

$$L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_k ||\partial_t\left[\left(\partial(t) + \frac{j}{\pi t}\right)u_k(t)\right]e^{-j\omega_k t}||_2^2 + ||f(t) - \sum_k u_k(t)||_2^2 + \langle\lambda(t), f(t) - \sum_k u_k(t)\rangle \tag{3}$$

(4) The iterative solution by alternating multiplier method (ADMM) obtains each modal component u_k and its central frequency ω_k.

The iterative update formulae for u_k, ω_k, and λ_k are as follows [8]:

$$u_k^{n+1}(\omega) = \frac{f(\omega) - \sum_{i \neq k} u_i(\omega) + \frac{\lambda(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{4}$$

$$\omega_k^{n+1}(\omega) = \frac{\int_0^\infty \omega |u_k(\omega)|^2 d\omega}{\int_0^\infty |u_k(\omega)|^2 d\omega} \tag{5}$$

$$\lambda^{n+1}(\omega) = \lambda^n(\omega) + \tau\left(f(\omega) - \sum_k u_k^{n+1}(\omega)\right) \tag{6}$$

when the iteration termination condition $\sum_k (||u_k^{n+1} - u_k^n||_2^2 / ||u_k^n||_2^2) < \varepsilon$ is satisfied, the variational modal decomposition ends, and k modal components of finite bandwidth are obtained.

2.2. Grey Wolf Optimizer

The grey wolf algorithm as a heuristic algorithm that mainly imitates the hunting process of grey wolf packs. The grey wolf pack is divided into wolves α, β, δ, and ω by the strict social hierarchy from high to low. In hunting, the whole wolf pack is led by wolf α, then wolf β and δ are responsible for the hunt, and finally, wolf ω tracks and seizes the prey [13]. The whole process is mainly divided into the three following parts:

(1) In the process of wolf hunting, the prey is first surrounded, and the mathematical model of its behavior is as follows [13]:

$$\overrightarrow{D} = \left| \overrightarrow{C} \cdot \overrightarrow{X_p(t)} - \overrightarrow{X(t)} \right| \tag{7}$$

$$\overrightarrow{X(t+1)} = \overrightarrow{X_p(t)} - \overrightarrow{A} \cdot \overrightarrow{D} \tag{8}$$

$$\overrightarrow{A} = 2 \overrightarrow{a} \cdot \overrightarrow{r_1} - \overrightarrow{a} \tag{9}$$

$$\overrightarrow{C} = 2 \cdot \overrightarrow{r_2} \tag{10}$$

where $\overrightarrow{X_p}$ represents the prey location and $\overrightarrow{X}$ represents the location of the individual grey wolf; $\overrightarrow{A}$ and $\overrightarrow{C}$ are coefficient vectors; t is the number of iterations; $\overrightarrow{r_1}$ and $\overrightarrow{r_2}$ are random numbers in the range of [0, 1]; a in $a = 2 - 2t/T$ represents the convergence factor; and the maximal number of iterations is T.

(2) The grey wolf pack rounds up the prey, projecting the location by wolves α, β, and δ leading wolf ω. The location update is completed with the following process equation of the grey wolf:

$$\overrightarrow{D_\alpha} = \left| \overrightarrow{C_1} \cdot \overrightarrow{X_\alpha(t)} - \overrightarrow{X_\omega(t)} \right| \tag{11}$$

$$\overrightarrow{D_\beta} = \left| \overrightarrow{C_2} \cdot \overrightarrow{X_\beta(t)} - \overrightarrow{X_\omega(t)} \right| \tag{12}$$

$$\overrightarrow{D_\delta} = \left| \overrightarrow{C_3} \cdot \overrightarrow{X_\delta(t)} - \overrightarrow{X_\omega(t)} \right| \tag{13}$$

$$\overrightarrow{X_1} = \overrightarrow{X_\alpha} - \overrightarrow{A_1} \cdot \overrightarrow{D_\alpha} \tag{14}$$

$$\overrightarrow{X_2} = \overrightarrow{X_\beta} - \overrightarrow{A_2} \cdot \overrightarrow{D_\beta} \tag{15}$$

$$\overrightarrow{X_3} = \overrightarrow{X_\delta} - \overrightarrow{A_3} \cdot \overrightarrow{D_\delta} \tag{16}$$

$$\overrightarrow{X(t+1)} = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3} \tag{17}$$

In the above equations, $\overrightarrow{X_\alpha}, \overrightarrow{X_\beta}, \overrightarrow{X_\delta}, \overrightarrow{X_\omega}$ represent the current positions of wolves α, β, δ, and ω, respectively; $\overrightarrow{D_\alpha}, \overrightarrow{D_\beta}, \overrightarrow{D_\delta}$ represent the distances between grey wolf ω and wolves α, β, and δ before the point, respectively; $\overrightarrow{A_1}, \overrightarrow{A_2}, \overrightarrow{A_3}, \overrightarrow{C_1}, \overrightarrow{C_2}, \overrightarrow{C_3}$ correspond to the coefficient vectors of wolves α, β, and δ, respectively; and $\overrightarrow{X_1}, \overrightarrow{X_2}, \overrightarrow{X_3}$ correspond to the vector positions of wolves α, β, and δ, respectively.

Prey is captured through an attack by the grey wolf pack. This attack behavior is mainly achieved by decreasing the value of a in Equation (9) by 2 to 0. When $\left| \overrightarrow{A} \right| \leq 1$, the grey wolf pack attacks the prey and conducts a local search; when $\left| \overrightarrow{A} \right| > 1$, the grey

wolf pack disperses and conducts all searches; lastly, the optimal solution of the grey wolf algorithm is output.

2.3. One-Dimensional Convolutional Neural Network

The LeNet-5 model proposed by LeCun in 1998 was the first convolutional neural network (CNN), and its structure mainly contained input and output layers, convolutional, pooling, and fully connected layers. CNNs have strong self-learning, parallel-processing, and fault-tolerance capabilities, are characterized by a low risk of overfitting and high model training efficiency compared with traditional neural networks, and they have been widely used in image processing. The main difference between one-dimensional convolutional neural networks (1DCNN) and CNNs is the difference in input data dimension, so 1DCNNs have a greater advantage in one-dimensional signal processing. Due to the rolling bearing vibration signal being a one-dimensional time-series signal, the 1DCNN structure is used in this paper for model training. Figure 1 shows the basic structure of the 1D convolutional neural network.
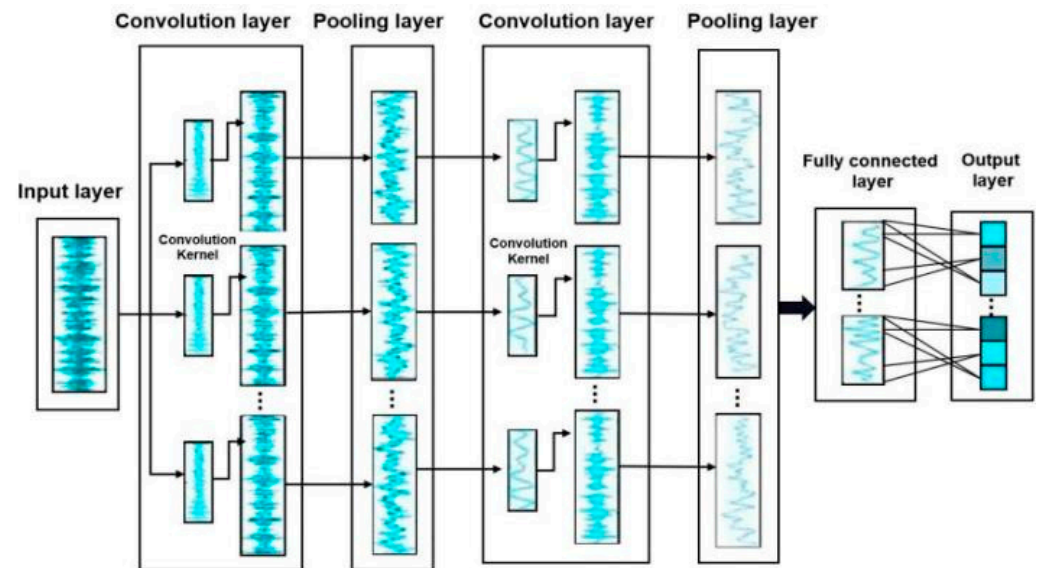
Figure 1. The 1DCNN structure.

The role of the one-dimensional convolutional layer is to perform the feature extraction of the input signal [29]. The formula for the convolutional operation performed by current convolutional layer l is shown below [29]:

$$x_j^l = f\left[\sum_{i \in M}\left(x_i^{l-1} * k_{ij}^l\right) + b_j^l\right] \tag{18}$$

In the above equation, x_j^l is the output mapping of the *j*-th neuron of the current (*l*) layer; x_i^{l-1} is the output mapping on the previous layer, which is also the input of the current (*l*) layer; *k* is the convolutional kernel; *j* is the number of convolutional kernels; *M* is the number of channels of input x_i^{l-1}; *b* is the deviation of the corresponding kernel; * is the convolutional operator; *f* is the activation function, generally the ReLU function. The activation function is PReLU in this paper to solve the problem of zero gradients of the ReLU activation function, the formula of which is as follows:

$$a_j^l = f\left(x_j^l\right) = \max\left(cx_j^l, x_j^l\right) \tag{19}$$

Parameter c is learnable and generally takes a number between 0 and 1. In this paper, parameter c is set to 0.25. The role of the activation layer is to non-linearly map the output of the previous layer, and the ReLU activation function has strong convergence

to overcome the problem of gradient disappearance. However, when the value of the input activation function is always negative, then the gradient through this point in the backpropagation process is a constant of 0 that cannot update the bias parameters and the corresponding weight. Therefore, the ReLU activation function has neuronal decay, which can be effectively solved by the PReLU activation function.

The function of the pooling layer is to downsample the increased amount of feature sequence data extracted from the convolutional layer to prevent the occurrence of overfitting by reducing the quantity of data computation in the neural network. This usually uses maximal pooling for downsampling; the formula of maximal pooling is as follows:

$$y_i^l = maxpooling\left(x_j^l, \text{scale}, \text{stride}\right) \tag{20}$$

In the above equation, y_i^l represents the output of the i-th neuron of the current (l) layer; $maxpooling(x)$ is the downsampling function, which serves to take the maximal value of a certain range; scale is the scale of pooling; and stride is the step of pooling.

The dropout layer further prevents the occurrence of the overfitting phenomenon. Using the dropout function can reduce the sensitivity of the neural network to small data changes and effectively improve the accuracy of data processing.

The fully connected layer plays the role of a classifier in the whole convolutional neural network architecture, and its role is to classify extracted features from the convolutional and pooling layers with the following equation:

$$z^{l+1}(j) = f\left(\sum_{i=1}^{m}\sum_{i=1}^{n} W_{itj}^l a_i^l(t) + b_j^l\right) \tag{21}$$

In the above equation, $z^{l+1}(j)$ represents the logits value of the j-th neuron in layer $(l+1)$; $a_i^l(t)$ is the output value of the t-th neuron of the i-th feature in layer l; W_{itj}^l represents the weight between the t-th neuron of the i-th feature in layer l and the j-th neuron in the next layer; b is the deviation; f denotes the activation function, and here represents the PReLU activation function.

The output layer usually uses a softmax classifier to output the classification labels.

3. Proposed IGVMD–1DCNN Fault Diagnosis Model

3.1. Improved Grey Wolf Optimizer (IGWO)

3.1.1. Non-Linear Convergence Factor Adjustment Strategy

The linear convergence factor a is the core parameter in the search process of the GWO algorithm, which decreases linearly with the number of iterations in the interval [0, 2] to achieve the global search time in the early stage and the local mining in the later stage [13]. In the actual search process, the global search time in the early stage should be longer than that of the local search in the later stage, but this linear adjustment strategy makes the global search time and local search time of the algorithm the same. For this reason, this paper designs a non-linear convergence factor for balancing the global search time and local search time, and the formula for the non-linear convergence factor is as follows:

$$a = 1 + \sin\left(\frac{\pi}{2} + \left(\frac{t}{T}\right)^3 \pi\right) \tag{22}$$

The trends of the two convergence factors as the number of iterations increases are shown in Figure 2.
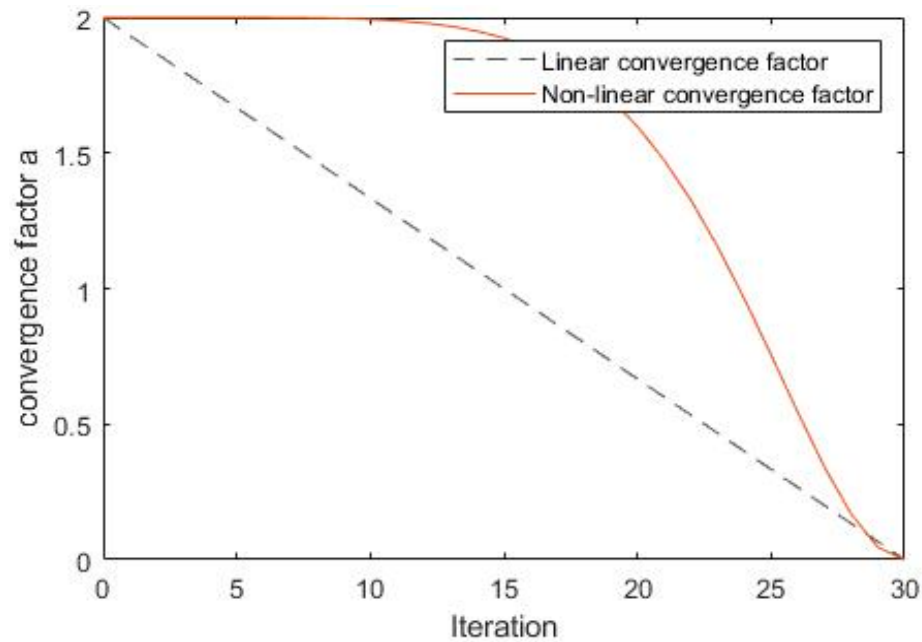
Figure 2. Two convergence factors.

3.1.2. Introduction of Levy Flight Strategy

In the GWO algorithm, all grey wolves converge under the leadership of alpha wolves, which makes the algorithm prone to local optimality during the iterative search process. The random wandering performance of Levy flight [20] during the iteration is used to improve the variety of the population, thus avoiding premature convergence and enhancing the global search capability of the algorithm. The new position update formula is as follows [20]:

$$
\begin{aligned}
\overrightarrow{X(t+1)} &= \overrightarrow{X(t)} + a*\text{Levy} \\
\text{Levy} &= \frac{u}{|v|^{1/z}} \\
u &\sim N(0, \phi^2) \quad v \sim N(0,1) \\
\phi &= \left\{ \frac{\tau(1+z)*\sin(\pi z/2)}{\tau((1+z)/2)*z*2^{(z-1)/2}} \right\}^2
\end{aligned}
\tag{23}
$$

where * is the multiplication of elements, *a* = 1, *z* = 1.5.

3.1.3. Adaptive Position Adjustment Strategy

Based on the strict hierarchy of the grey wolf population, the position update of ω wolves should follow the decreasing leadership weights of α, β, and δ wolves. However, the ω wolf position update shown in Equation (17) of the standard GWO takes an equal contribution rate calculation for α, β, and δ wolves, which cannot reflect its hierarchy. The above problems will affect the exploration search efficiency of the GWO algorithm. Thus, a new position update formula is introduced for this paper to improve the convergence speed of the grey wolf algorithm as follows:

$$
\overrightarrow{X(t+1)} = \frac{w_1 \overrightarrow{X_1} + w_2 \overrightarrow{X_2} + w_3 \overrightarrow{X_3}}{3}
\tag{24}
$$

In the update of the grey wolf position, led by α, β, and δ wolves, α has the best fitness value, β has the second-best fitness function value, and δ has the third-best fitness value. The dynamic weighting formula is as follows:

$$
\begin{aligned}
w_1 &= \frac{f_\alpha}{f_\alpha + f_\beta + f_\delta} \\
w_2 &= \frac{f_\beta}{f_\alpha + f_\beta + f_\delta} \\
w_3 &= \frac{f_\delta}{f_\alpha + f_\beta + f_\delta}
\end{aligned}
\tag{25}
$$

In the above equation, f_α, f_β, and f_δ represent the fitness values for α, β, and δ wolves, respectively.

3.2. Improved Variational Mode Decomposition with Grey Wolf Optimizer (IGVMD)

The goal of IGVMD optimization is to find the optimal parameter combination of VMD under which an IMF component obtained by VMD decomposition should have the minimum enveloping entropy. The algorithm procedure (Figure 3) is as follows:
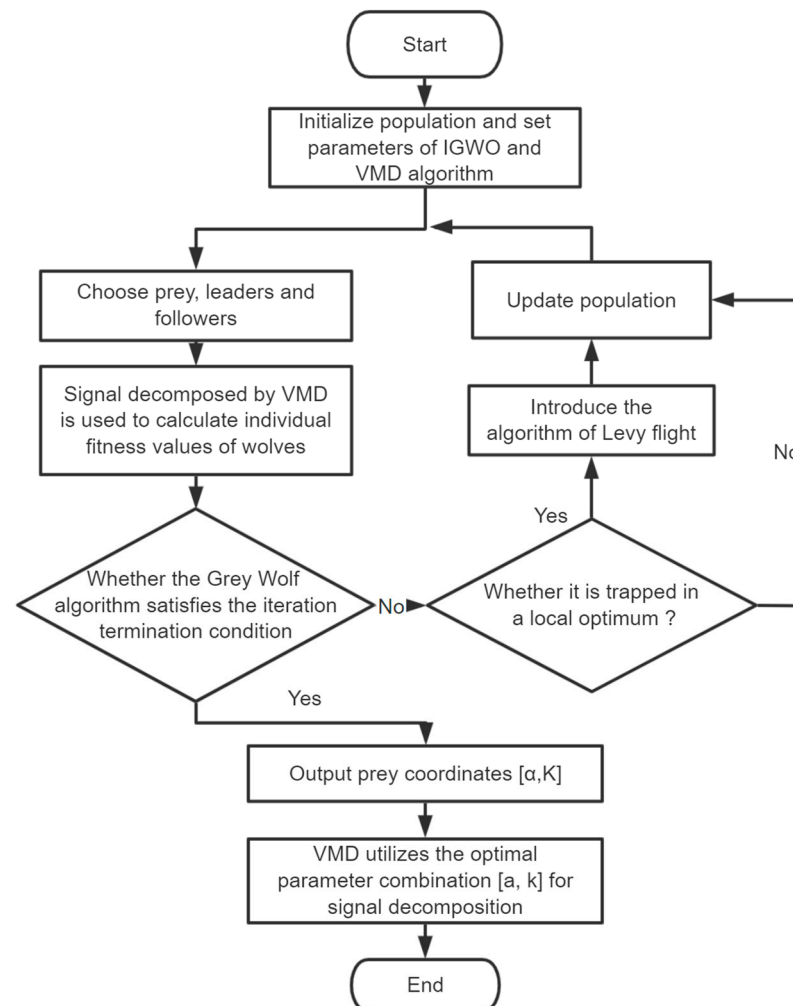
Figure 3. IGVMD algorithm flow chart.

(1) Set the parameters of the IGWO and VMD algorithms and initialize the population.
(2) Take the local envelope entropy of the modal component as the fitness function. Then, the position of the individual population by the parameter combination $[K, \alpha]$ is estimated. Take the place of the grey wolf individual with the minimum fitness value as the position of the current prey.

(3)　Calculate the fitness function value of the VMD decomposition corresponding to the current position of each grey wolf.

(4)　According to Equations (11)–(16) and Equations (24) and (25), update the spatial position of the grey wolf.

(5)　Calculate the fitness function values of individual grey wolves under VMD decomposition after the grey wolf position update and compare them with the fitness function values of current prey. If the fitness function value of the updated grey wolf is smaller than the prey, take the position of the grey wolf as the new prey position.

(6)　Determine whether it has trapped in a local optimum, and when the fitness value of the prey does not change in 5 consecutive iterations, introduce a levy flight strategy for a grey wolf position update.

(7)　Repeat steps 3 to 6 until the iteration termination condition is met, and output the prey coordinates, which is the best combination of parameters $[K,\alpha]$. Then, the parameters are utilized to decompose the fault signal by VMD.

3.3. IGVMD–1DCNN Model

A one-dimensional convolutional neural network generally contains input, convolutional, pooling, fully connected, and output layers.

The 1DCNN network used in this paper has 13 layers, including input, convolutional, pooling, dropout, fully connected, softmax, and output layers, among which the dropout layer is mainly used to randomly remove some hidden layer units to further avoid the phenomenon of overfitting of the model. The dropout layer probability in this experimental network is 0.5. A batch normalization (BN) layer is added between each convolutional and activation layer, making the input data distribution of each layer relatively stable and accelerating the learning speed of the model. The parameter settings of the 1D convolutional neural network are shown in Table 1.

Table 1. The 1DCNN parameters.

Network Layer Type	Convolution Kernel and Number of Channels	Activation Functions	Output Data Length and Number of Channels
Input	-	-	1024, 1
Conv_1	55, 27	PReLU	1024, 27
Pooling_1	16, 27	-	64, 27
Conv_2	55, 27	PReLU	64, 27
Dropout_2	-	-	-
Conv_3	55, 27	PReLU	64, 27
Pooling_3	16, 27		4, 27
Conv_4	55, 27	PReLU	4, 27
Dropout_4	-	-	-
Conv_5	55, 27	PReLU	4, 27
Flatten	-	-	108
Softmax	-	-	10, 1

The network structure of fault identification detection in this paper is shown in Figure 4 and is mainly divided into four stages: fault signal preprocessing, signal input, fault feature intelligent extraction, and fault feature identification stages. In the signal processing stage, the improved grey wolf optimized variational modal function algorithm proposed in this paper extracts the characteristics of the fault signal from the noise background. The input signal in the signal input stage is a one-dimensional vibration signal of length 1024 [35,36]. The input signal is convolved for intelligent feature extraction. Then, PReLU learns negative parameter features and solves the zero-gradient problem, downscales by maximal pooling, further downscales by the dropout layer, and executes a fault diagnosis by softmax for fault classification. The output layer has ten neurons to classify the ten health states of the bearing. Figure 5 shows a flowchart of the rolling bearing fault detection system.
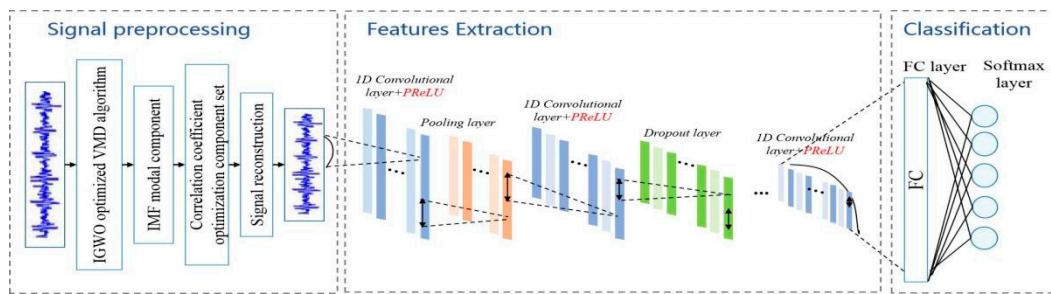
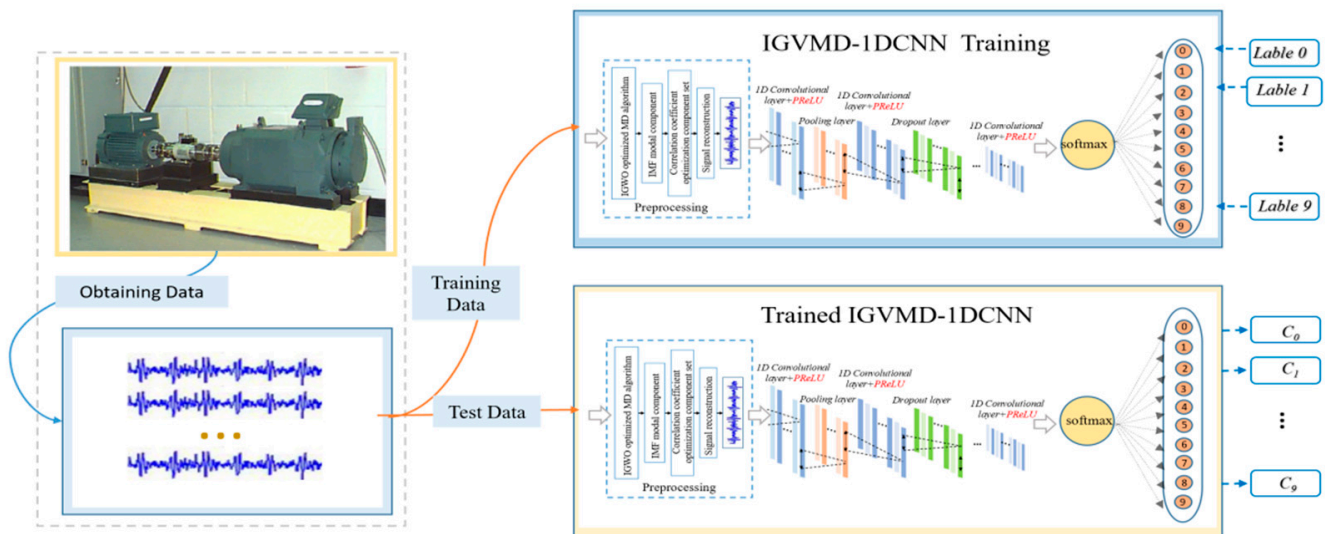**Figure 4.** The 1DCNN structure in the experiment.

Figure 5. The flow chart of rolling bearing fault diagnosis.

3.4. Optimization of IGVMD–1DCNN Model Parameters

To improve the performance of the IGVMD–1DCNN model, this study uses a grid search algorithm [37] to optimize the hyperparameters in the model and select an optimal set of hyperparameters for the model. This study avoids repeated manual tests and improves the model training efficiency. Grid search is used to optimize the model batch_size, iteration number, and learning rate. Too large of a batch_size tends to not converge, and too small tends to cause memory overflow. Too large of an iteration number tends to cause overfitting, while too small of an iteration number will cause underfitting. The size of the learning rate directly affects the convergence speed of the model and whether it can eventually converge. The parameter range of batch_size is defined as 16, 32, 64, and 128; the parameter range of iteration number is defined as 10, 30, 50, and 100; and the parameter range of learning rate is 0.1, 0.01, and 0.001. The average accuracy of the test set in the 4-fold cross-validation is set as the evaluation index of the model, and the optimal combination of parameters is selected to construct the IGVMD–1DCNN fault diagnosis model. In addition, the fault diagnosis model is created on Python 3.8 and performed on a GTX3090 GPU. The optimal parameters of the IGVMD–1DCNN are obtained by grid search, and the experimental settings and results are shown in Table 2.

Table 2. The experimental setup and results.

The Serial Number	Batch_Size	Epochs	Learning Rate
0	16	10	0.1
1	32	30	0.01
2	64	50	0.001
3	128	100	-
The optimal results	32	50	0.001

3.5. Fault Diagnosis Process of IGVMD–1DCNN Model

The flowchart of the IGVMD–1DCNN model is shown in Figure 6, and the steps of its model processing process are as follows:
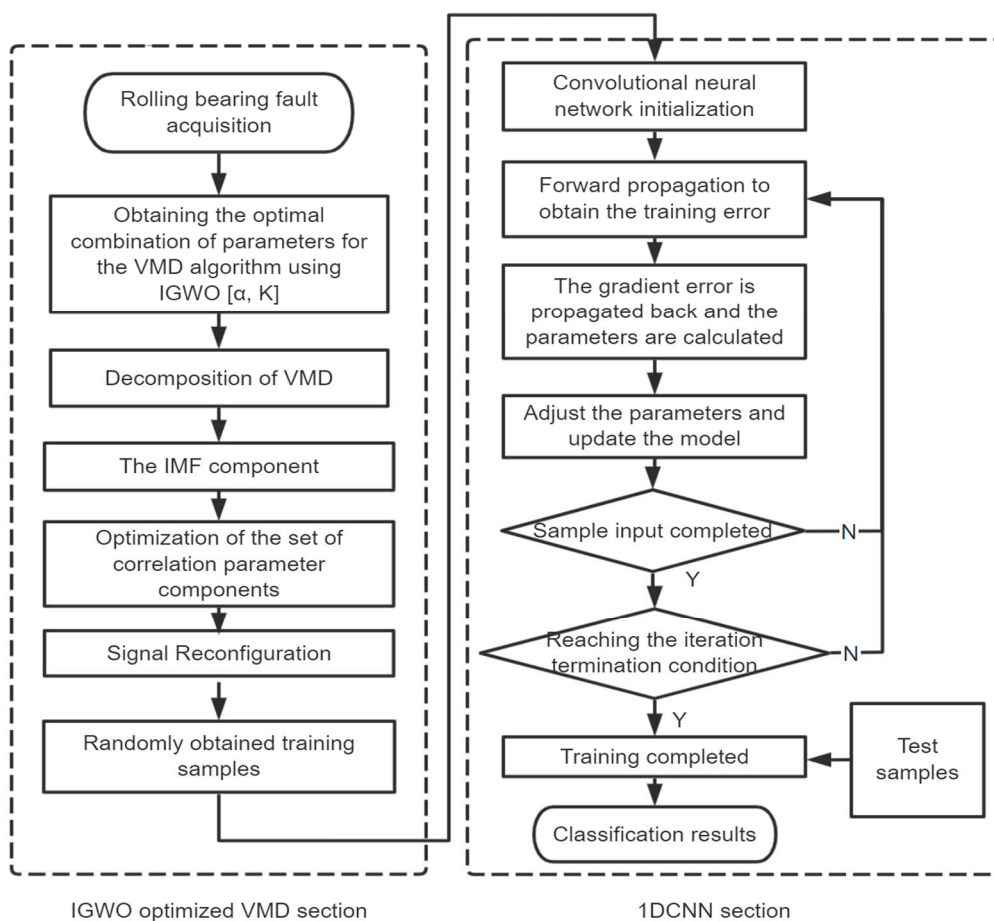

Figure 6. The flow chart of IGVMD–1DCNN model.

Step 1. Signal acquisition of rolling bearing data as signal input.

Step 2. The input bearing signal is preprocessed by using the optimized variational modal decomposition method of the improved grey wolf algorithm. The optimal combination of parameters [K,α] is obtained by continuous iterative search, and K modal components IMF are obtained.

Step 3. Effective modal component IMF is selected for vibration signal reconstruction by correlation coefficients to complete the effective extraction of features.

Step 4. Reconstructed signal X is divided into training data and test data. The training data are divided into training set Xtrain and validation set Xvalidation, which are used as the input signals of 1DCNN

Step 5. Combining the grid search and the 1D convolutional neural network model, the optimal combination of parameters is determined, and the 1DCNN fault diagnosis model is established.

Step 6. Training sample Xtrain is used as input to train the 1DCNN model. Model parameters are continuously updated by inputting training samples through forward and backward propagation.

Step 7. Validation set Xvalidation participates in model training and continually adjusts the model parameters.

Step 8. Whether the model meets the iteration termination condition N is determined. If yes, training is stopped for the next operation, otherwise, step 7 is repeated.

Step 9. Test sample Xtest is input into the trained 1DCNN model for classification and identification detection of fault signals.

4. Experimental Results and Analysis

In this section, we demonstrate the effectiveness of the proposed IGVMD–1DCNN through experiments on different datasets. Compared with EMD and traditional VMD, the proposed IGVMD has a better feature extraction ability by using the bearing vibration signal under noise conditions. Then, compared with the neural network with the ReLU activation unit, the higher training efficiency of the 1DCNN model with the PReLU activation unit designed in this paper is verified. Furthermore, we compared it with SVM, gcForest, LeNet-5, and Resnet.

4.1. Experimental Data

To ensure the reliability of the experimental data, open dataset rolling bearing experimental data from the Electrical Engineering Laboratory of Case Western Reserve University, USA were used [38]. The experimental platform consists of an electric motor, a torque sensor/converter, a power test meter, and an electronic controller. The experimental platform equipment [38] is shown in Figure 7.
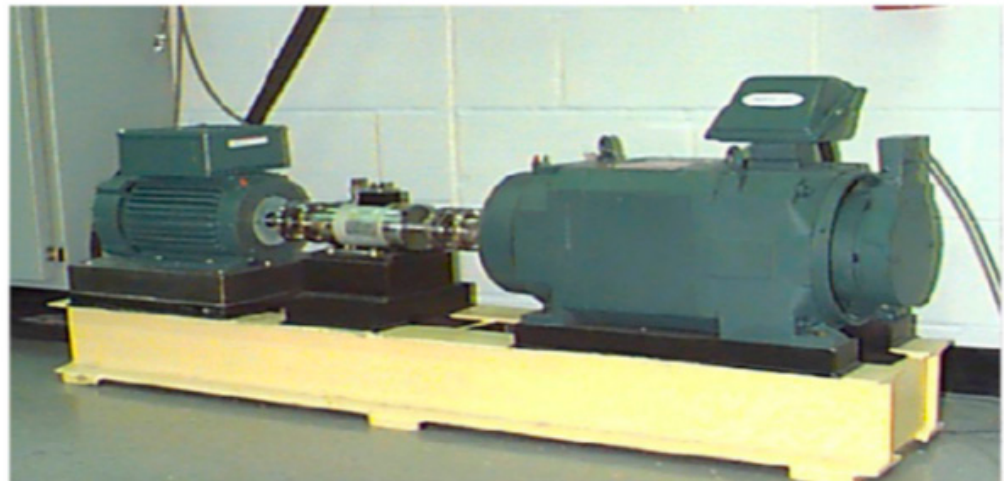

Figure 7. Equipment of experimental platform.

Collected bearing failure data for the experiments were single-point failure data manufactured by EDM single-point machining, and a total of 10 data categories under 0HP functional model are used here. Bearing type SKF6205 drive end data are used for the test data with sampling frequency of 12 KHz, the sample uses damage diameter of 0.1178 mm, 0.3556 mm, a 0.5334 mm inner ring, outer ring, and roller body failure data, and a class of normal data. According to the resampling method [37], the original data of ten states are divided. Each sample has a length of 1024 [35,36] and a sampling step size of 512. In this way, a total of 4500 data samples are obtained. A total of 4500 samples were generated from 10 types of data, 90% for the training set and 10% for the test set. The specific sample

information in the experiment is shown in Table 3, and the time and frequency domain waveforms of each type of bearing operation state are shown in Figure 8.

Table 3. Experimental sample information.

Sample Types	Fault Width (in)	Motor Speed	Length of Sample Length	Label
normal	0	1797 (rpm)	1024	0
IR007	0.007	1797 (rpm)	1024	1
B007	0.007	1797 (rpm)	1024	2
OR007@6	0.007	1797 (rpm)	1024	3
IR014	0.014	1797 (rpm)	1024	4
B014	0.014	1797 (rpm)	1024	5
OR014@6	0.014	1797 (rpm)	1024	6
IR021	0.021	1797 (rpm)	1024	7
B021	0.021	1797 (rpm)	1024	8
OR021@6	0.021	1797 (rpm)	1024	9

(**a**) Fault 1

(**b**) Fault 2

(**c**) Fault 3

(**d**) Fault 4
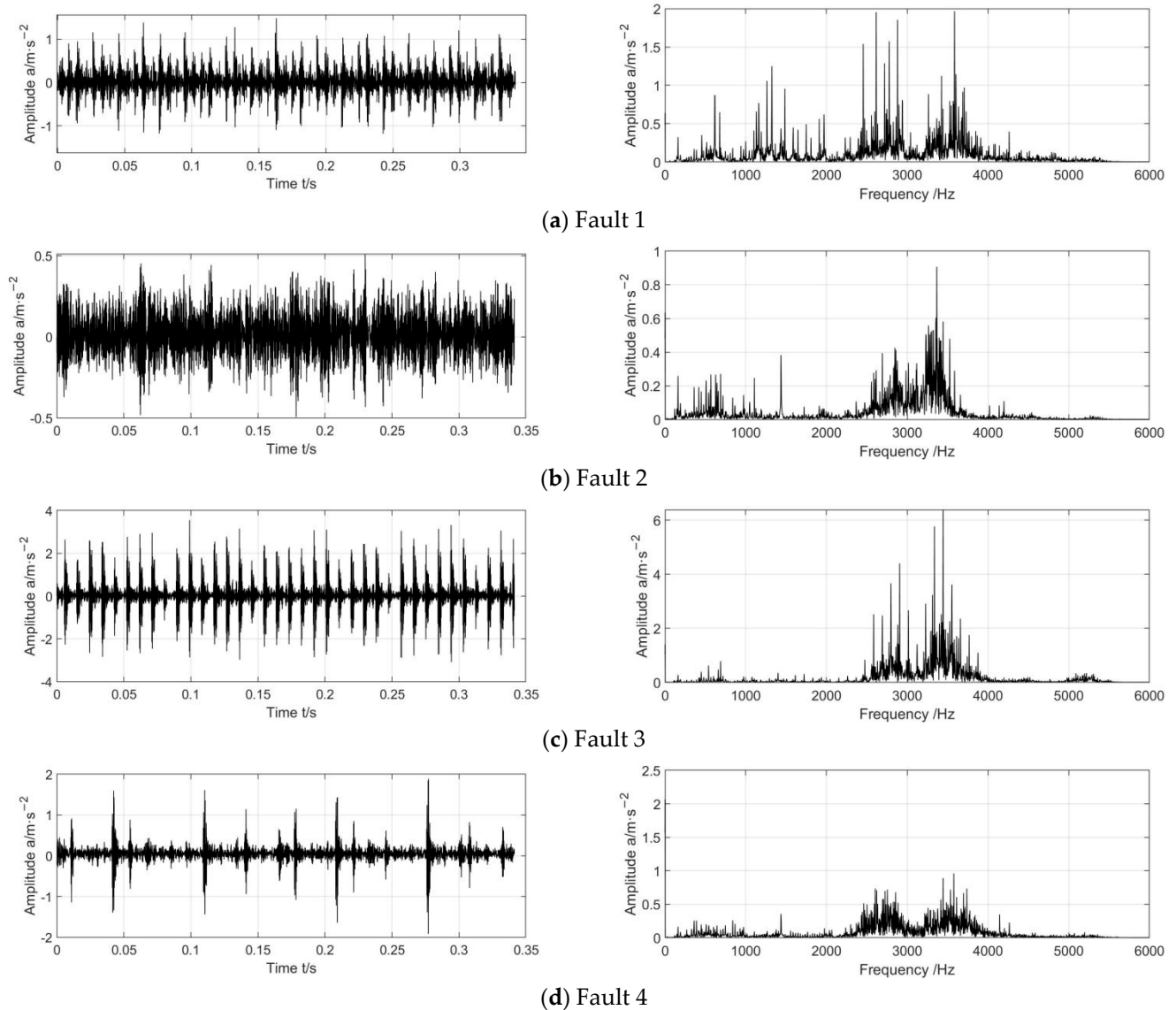
Figure 8. *Cont.*
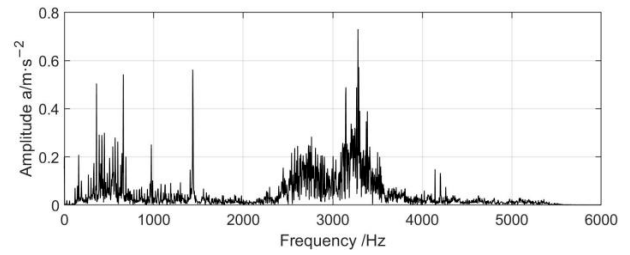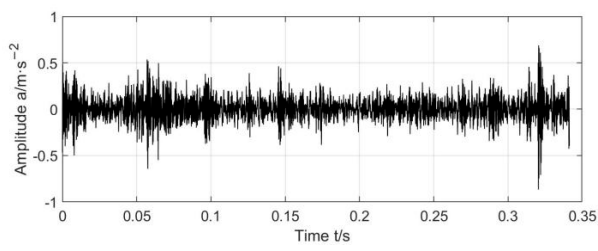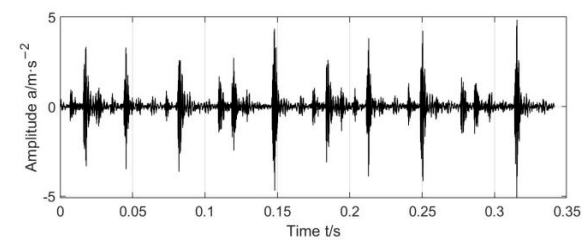
(**e**) Fault 5
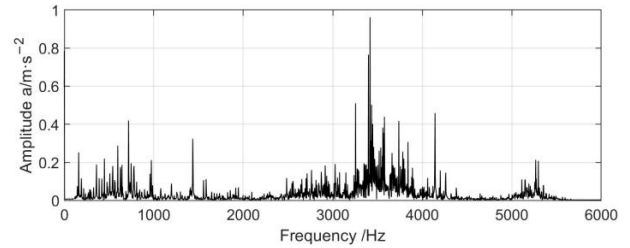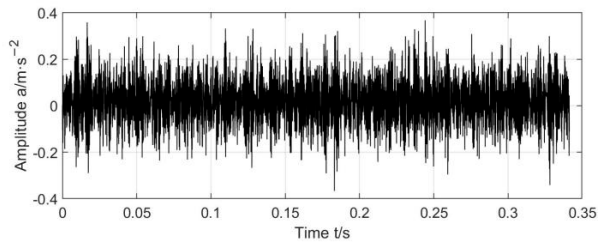
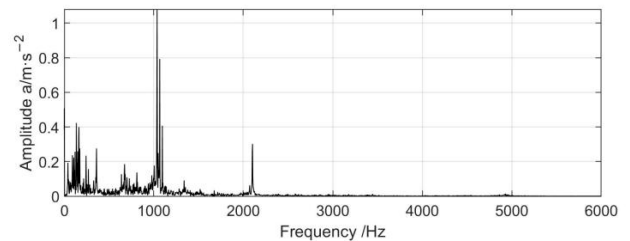(**f**) Fault 6
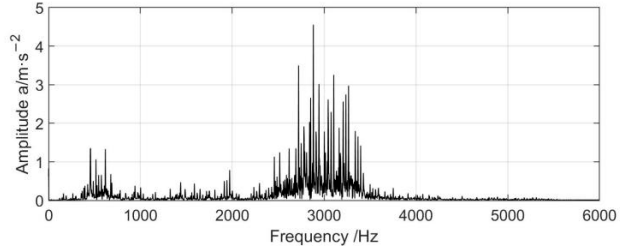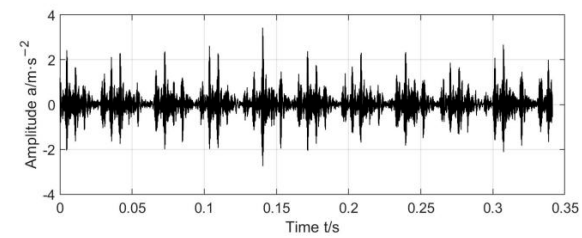
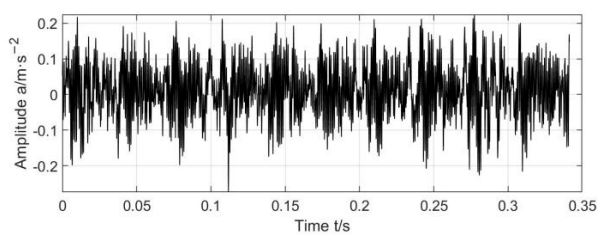

(**g**) Fault 7
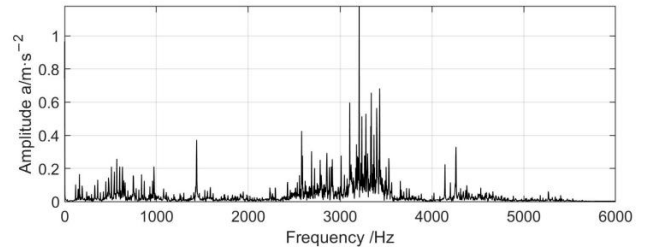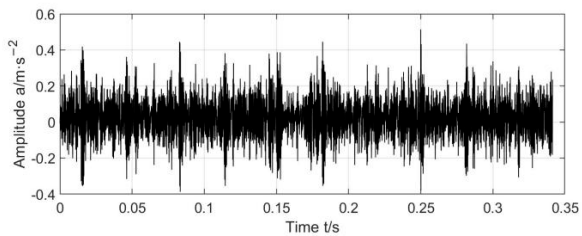
(**h**) Fault 8
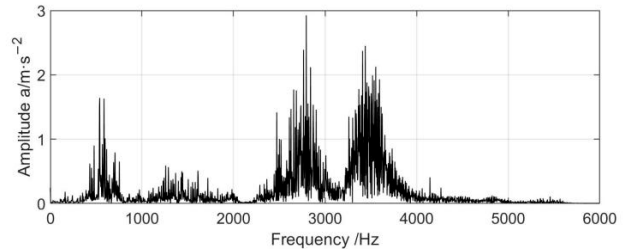
(**i**) Fault 9

(**j**) Normal

Figure 8. Time and frequency domain waveforms of bearing operating conditions.

4.2. IGVMD Decomposition

　　To address the problem of premature convergence and falling into the local optimum of the GWO algorithm, three adjustment strategies are introduced in the GWO algorithm to obtain the IGVMD algorithm. The number of IGVMD algorithm population was set to 30, the maximum number of iterations was 30, the range of *K* values was set as [1,15], and the range of α values was set as [500, 4000] [11]. In the process of optimizing the VMD model, the fitness convergence curves obtained using four different optimization algorithms are shown in Figure 9. GWO represents the standard GWO; IGWO-1 represents the GWO algorithm that introduces a non-linear convergence factor adjustment strategy on top of GWO; and IGWO-2 represents the GWO algorithm that introduces a levy flight adjustment strategy on top of IGWO-1.
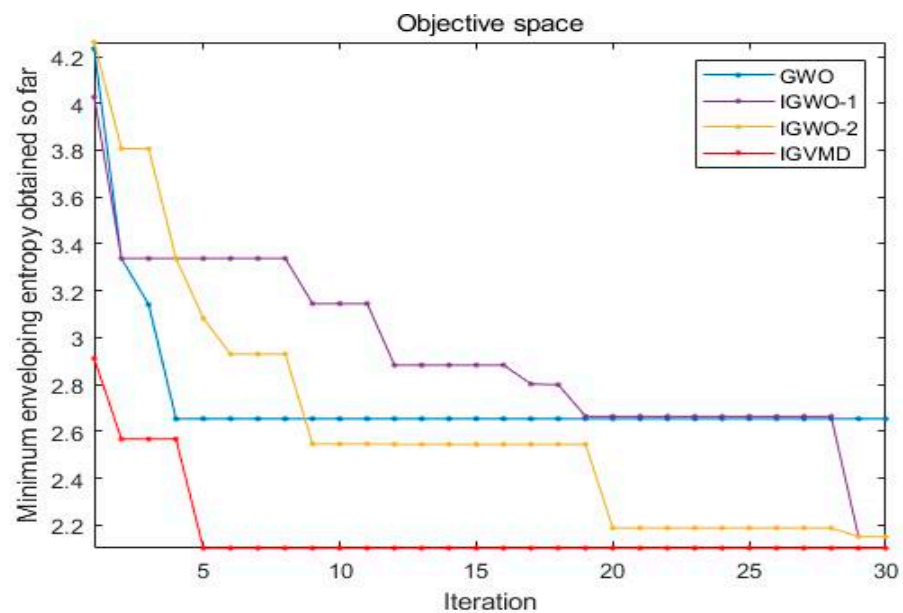

Figure 9. Fitness value curves for each algorithm.

　　The curves in Figure 9 show that the fitness value of GWO converges too early; compared with GWO, IGWO-1 and IGWO-2 algorithms have the ability to jump out of the local optimum but converge more slowly; IGVMD not only has a good ability to avoid falling into the local optimum but also has faster convergence performance compared with the first three optimization algorithms. It is verified that the improved strategy proposed in this paper can effectively improve the global search performance of GWO and has a strong search capability in optimizing this model problem.

　　Taking Fault 1 as an example, optimal [*K*,α] combinations are obtained by the IGVMD algorithm: *K* = 4, α = 791, and fault signals are decomposed into four IMF components. Time and frequency domain waveforms of each modal component obtained after the signal decomposition are shown in Figure 10.

　　For comparison, the Fault 1 signal is decomposed using EMD and conventional VMD. EMD is decomposed into 11 IMF components, and the first 6 IMF components are analyzed for ease of observation. The time–frequency diagrams of IMF1–IMF6 are shown in Figure 11. According to the spectrum after EMD, it is obvious that IMF5 and IMF6 showed the modal mixing phenomenon. A traditional VMD reference [38] empirically sets [K = 4, α = 2000]. Figure 12 shows the IMF component time–frequency diagram of traditional VMD, which is carefully compared with the improved VMD, and the IMF component period pulse pattern of the improved VMD, are more obvious.
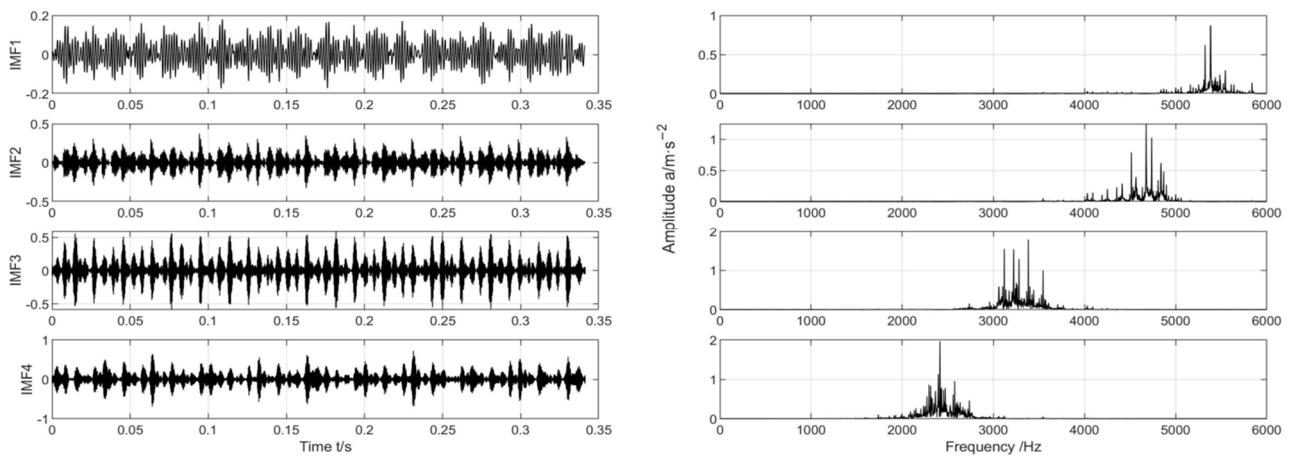
Figure 10. Time domain and frequency domain diagrams of IMF components decomposed by IGVMD.
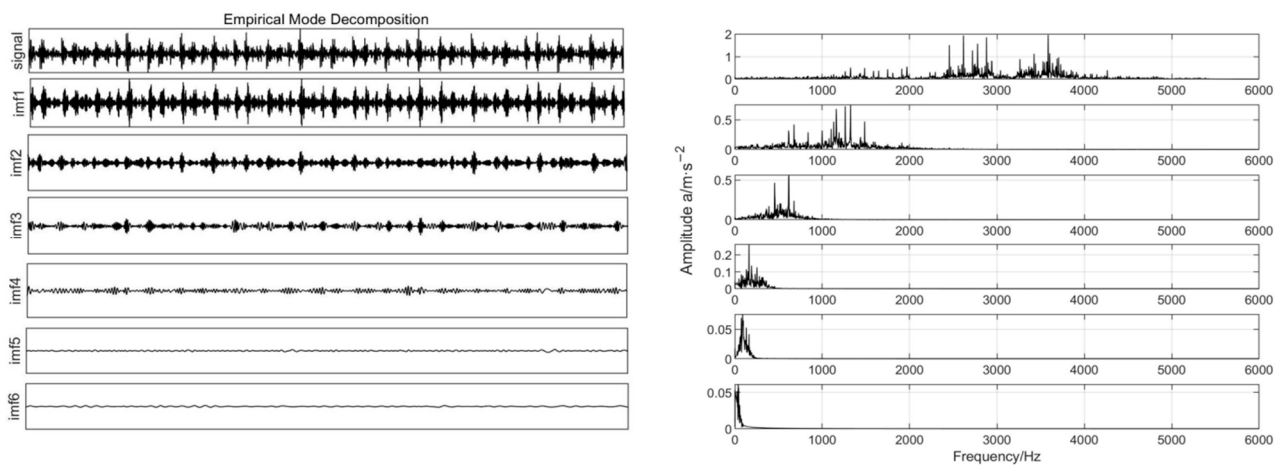

Figure 11. Time domain and frequency domain diagrams of IMF components decomposed by EMD.
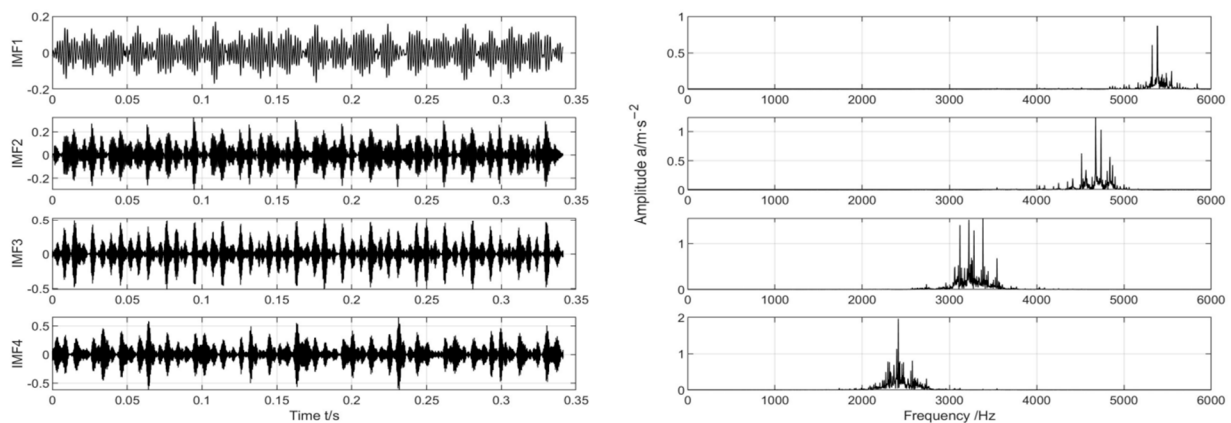
Figure 12. Time domain and frequency domain diagrams of IMF components decomposed by traditional VMD.

Experimental results showed that, compared with the EMD method and the traditional VMD method, the IGVMD method effectively avoids modal confusion and endpoint effects generated by EMD; this solves the problem of adaptive VMD solution and avoids the influence of the traditional VMD method on the experimental results due to improper signal decomposition caused by human factors.

4.3. Filtering Modal Components

The IMF component of the bearing signal obtained by IGVMD contains noise signals, and to extract effective fault features, this paper uses the correlations to filter the IMF component for signal denoising. The correlation mainly reflects the correlation degree between two signals and is commonly used in the analysis of vibration signals [39]. Its algorithm is expressed as follows:

$$\rho XY = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}, |\rho XY| \leq 1 \tag{26}$$

where $Corr(X,Y)$ represents the covariance, and $\sqrt{D(X)}$ and $\sqrt{D(Y)}$ are variances of signals X and Y, respectively. The closer the correlation coefficient is to 0, the lower the correlation of the signal; the closer the absolute value of the correlation coefficient is to 1, the higher the correlation of the signal.

In this paper, by selecting the modal components with a correlation threshold greater than 0.3 with the original signal for signal reconstruction, signal noise reduction is effectively achieved, and the impulse characteristics of the fault are retained to a great extent. To verify the extraction effect of the weak signal under the background noise of the improved VMD, the reconstructed signal of the improved VMD was analyzed and compared with the reconstructed signal of the EMD and the conventional VMD in the envelope spectrum, and the envelope spectrum of the reconstructed signal of each method is shown in Figure 13. Figure 13a clearly shows that, although fault frequency and twofold frequency exist in the envelope signal after EMD, the other multiples are disturbed by noise and the fault characteristics are not obvious; Figure 13b,c clearly show that, in the analysis of traditional VMD and improved VMD envelope spectrum, the feature frequencies and their multiples in the improved VMD envelope had more obvious peaks. The experiment shows that the improved VMD can more accurately decompose the signal, which is more conducive to extracting early faint fault features in the noise background.
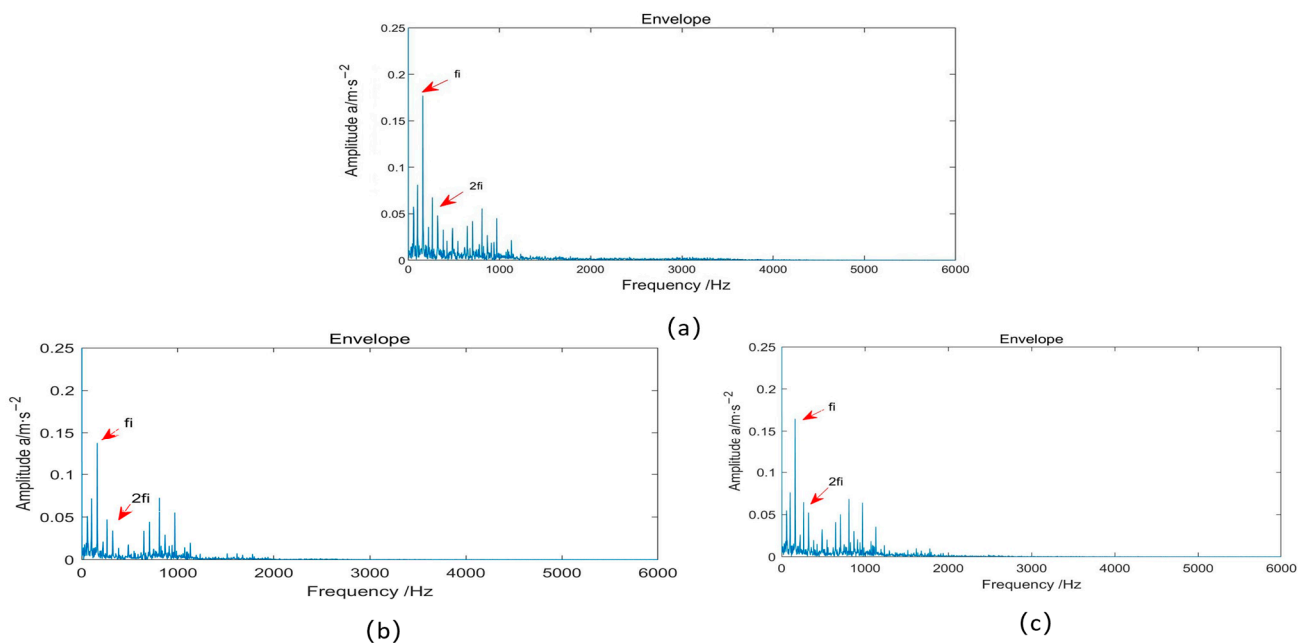

Figure 13. Envelope spectrum analysis of reconstructed signals by different methods: (**a**) EMD; (**b**) traditional VMD; (**c**) improved VMD.

4.4. Diagnostic Results of Model

In the 1D convolutional neural network used in this paper, batch processing and the number of iterations of the model training set are set to 32 and 50 by grid search; learning

rates are set to 0.001. The cross-entropy loss function and Adam optimization algorithm are used. In bearing fault diagnosis, the accuracy rate and loss rate are the most important indices to evaluate the model, so this paper uses these two indices to make a comprehensive evaluation of the model diagnosis effect. The results are shown in Figure 14. Figure 14a,b show that the accuracy rate and loss rate of the model fluctuated significantly in the first six iterations because the model learned fault features quickly. After that, the accuracy rate and loss rate of the model tended to be flat and finally reached a stable state. The accuracy of the validation set is higher than that of the training set, so the results show that the proposed model has not been fitted.
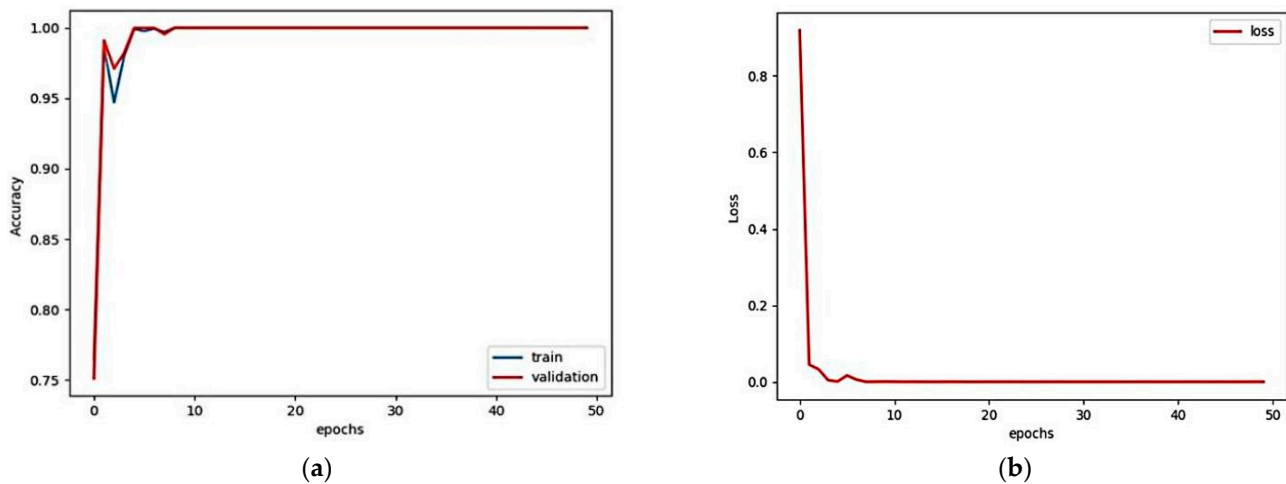

Figure 14. Experimental results. (**a**) The change in accuracy. (**b**) The variation in loss rate.

To further analyze the proposed method, three evaluation metrics were applied: *precision*, *recall*, and *F1-score*. *Precision* is the probability of actual positive samples from all predicted positive samples. *Recall* is the probability of being predicted as a positive sample from the actual positive samples, and the *F1-score* is the weighted average of precision and recall [40]. The precision, recall, and *F1-score* for each bearing category are defined as follows [40]:

$$
\begin{aligned}
precision &= \frac{TP}{(TP+FP)} \\
recall &= \frac{TP}{(TP+FN)} \\
F1\_score &= 2\frac{precision*recall}{(precision+recall)}
\end{aligned}
\tag{27}
$$

In Equation (27), *TP* represents the number of positive samples predicted correctly, FP represents the number of negative samples predicted incorrectly, *TN* represents the number of negative samples predicted correctly, and *FN* represents the number of positive samples predicted incorrectly.

The calculation results of the *precision*, *recall*, and *F1-score* of the method in this paper are shown in Table 4. From Table 4, it can be seen that except for a few samples in fault category 2, fault category 5, and fault category 8 that are not identified, all other types of samples are effectively identified, and the experiment proves that the proposed method in this paper has a strong fault diagnosis capability.

In addition, the impact of emd, traditional vmd, and improved vmd on the detection accuracy was evaluated. As shown in Table 5, the results of the study clearly show that the improved VMD method outperforms the other two methods in terms of accuracy for the same dataset with a signal-to-noise ratio of −6 dB under 0 HP functioning modes.

Table 4. The diagnosis result evaluation of improved 1DCNN.

| Type	Precision (%)	Recall (%)	F1-Score (%)
0	100	100	100
1	100	100	100
2	100	100	99.9
3	100	100	100
4	100	100	100
5	100	99.77	100
6	100	100	100
7	100	100	100
8	99.77	100	99.89
9	100	100	100
Avg	99.98	99.98	99.98

Table 5. The diagnosis result evaluation of different methods.

Method	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
EMD	96.28	96.10	96.10	95.33
VMD	98.36	98.56	98.41	98.22
IGVMD	99.59	99.60	99.59	99.56

4.5. Discussion

4.5.1. Performance of Parameterized Linear Elements

PReLU can effectively diagnose bearing faults by solving the equal displacement and negative interval death problems of neurons. To evaluate the influence of PReLU on the model performance in IGVMD–1DCNN, an IGVMD–1DCNN model with the same parameters was tested with regard to the size and number of convolution kernels, the learning rate, etc., except for the difference of a PReLU activation unit and ReLU activation unit. The same signals were input, and the validation accuracy and loss of the two models were recorded. Figures 15 and 16 show the comparison results of accuracy and loss changes, respectively, in the training process of the two algorithms. These figures show that the model using the PReLU activation function had higher accuracy, faster loss convergence, and a more stable training process. It was proven that PReLU makes the model have better learning ability, and it enhances the fitting of the 1DCNN network for fault diagnosis without increasing the amount of computation.
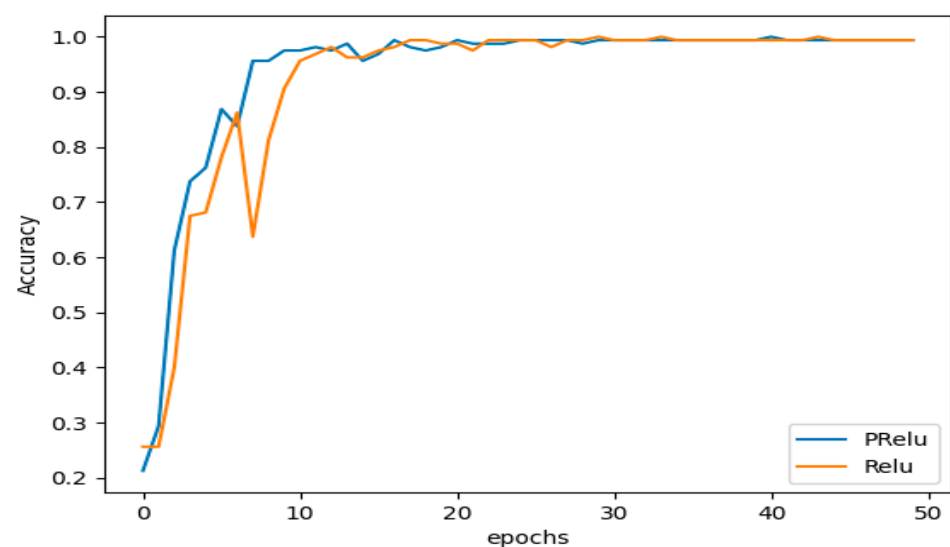
Figure 15. Variation in accuracy with iteration number under different activation functions.
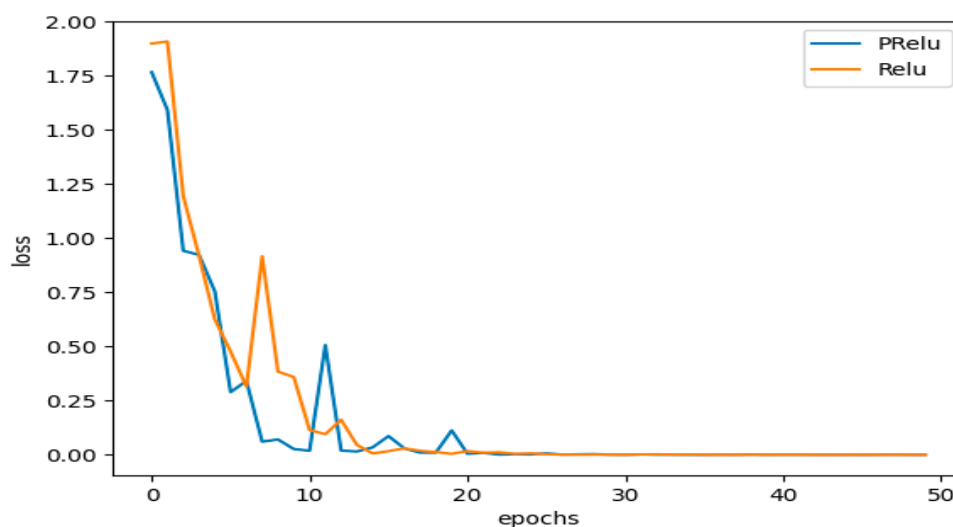
Figure 16. Variation in loss with iteration number under different activation functions.

4.5.2. Influence of Data Preprocessing on the Model

The IGVMD method can eliminate the influence of background noise on the model through decomposition of the original signal and screening of effective information. A group of comparison tests was set, the same data and parameter settings were input, and the number of iterations to was set to 50. The IGVMD–1DCNN and 1DCNN models were compared, and the accuracy and loss of the two models on the validation set were recorded. The accuracy rate and loss changes of the two algorithms in the training process are shown in Figures 17 and 18, respectively.
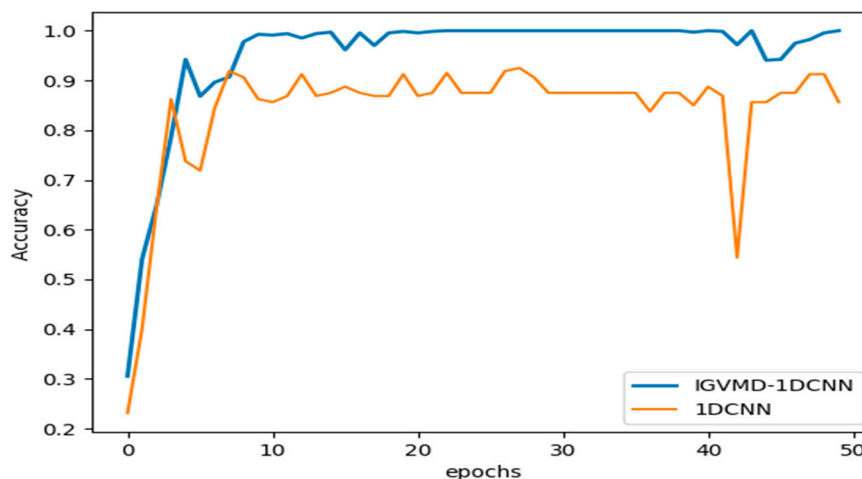

Figure 17. Variation in accuracy with iteration number.

Figure 17 clearly shows that after the 23rd iteration, the training accuracy of the 1DCNN method before pretreatment reached 91.50%, and the training accuracy of the improved model in this paper reached 100%, indicating that the model in this paper is more efficient at fault identification. In addition, as shown in Figure 18, the loss of the model proposed in this paper decreased faster, at only 0.00003, indicating that the improved model in this paper has a better fitting effect, and the training process is more stable.

A confusion matrix is used to further observe the recognition of different types of faults by these two models, and the training model is used to classify the faults of 200 randomly divided test sets. Figures 19 and 20 represent the confusion matrices for the recognition of real and predicted samples of the test set, respectively, under the two models. Figure 19 shows that the recognition rate of the 1DCNN method before IGVMD pretreatment in Faults 2, 5, and 8 did not reach 100%; its recognition rates were 95.24%, 95.92%, and 97.78%,

respectively; the overall classification recognition rate was 98.89%. Experimental result analysis showed that the interference of strong noise gave the traditional 1DCNN model a low recognition rate of bearing rolling element fault features. The confusion matrix identified by our method is shown in Figure 20, and the recognition rate reached 100%. The experiment showed that this method can accurately identify bearing signals faster by extracting effective fault features as input signals.
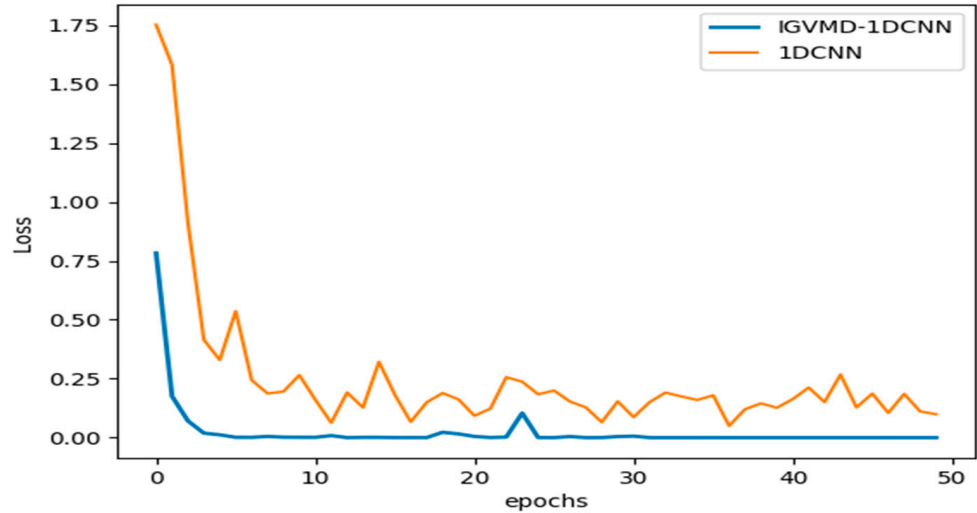
Figure 18. Variation in loss with iteration number.
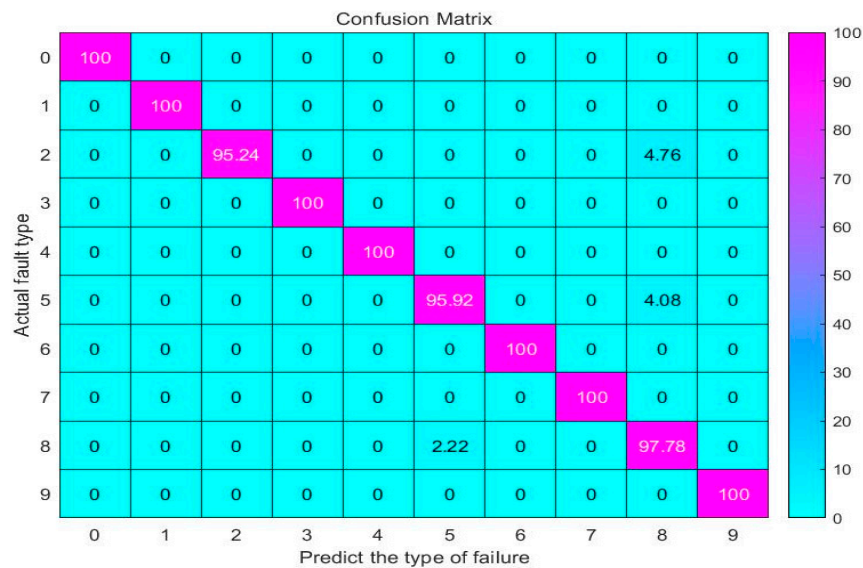
Figure 19. The confusion matrix of 1DCNN.

The excellent diagnostic performance of this research algorithm was demonstrated in fault diagnosis experiments on the above 0HP functional mode dataset. To further validate the algorithms in this study, diagnostic experiments were conducted for 1HP, 2HP, and 3HP functional mode fault signals with equal size datasets. The dataset in each functional mode was subjected to 10 iterations of the experiment, and the four evaluation metrics obtained from the experiment are shown in Table 6. The fault recognition rates under the three loads were 99.91%, 99.84%, and 99.33%, respectively, and the results show that the IGVMD–1DCNN model can effectively identify faults under different loads.

Figure 20. The confusion matrix of IGVMD–1DCNN.

Table 6. The diagnosis result evaluation of different functioning mode.

Load	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
1HP	99.91	99.93	99.92	99.91
2HP	99.83	99.87	99.84	99.84
3HP	99.61	99.27	99.18	99.33

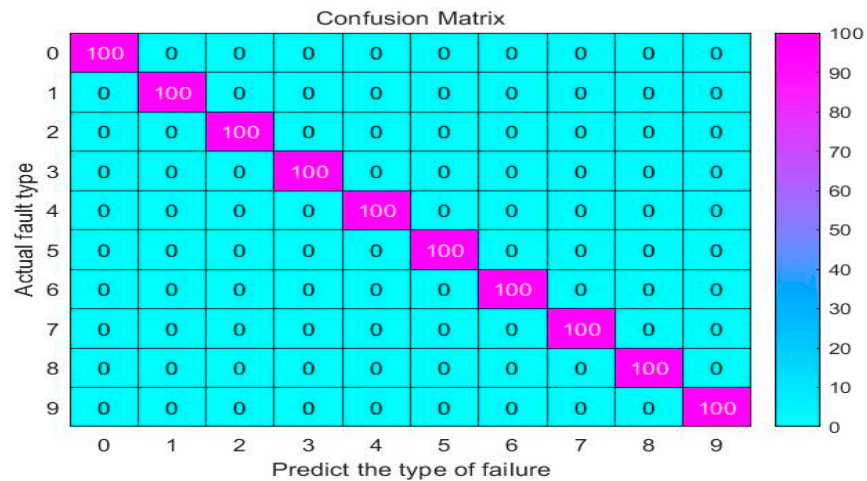### 4.5.3. Datasets Divided by Different Proportions

To verify the feature extraction ability of the model, a set of comparative experiments was set up, and the training data and test data were divided on the dataset with ratios of 1:1, 4:1, 7:3, 3:2, and 9:1. To ensure the reliability of the experimental results, the average results after 10 tests were adopted, and the results are shown in Table 7.

Table 7. Comparison results of different scales of dividing the dataset.

The Serial Number	Ratio of Training Set to Test Set	Diagnostic Accuracy %
1	1:1	99.73
2	3:2	99.83
3	7:3	99.87
4	4:1	99.94
5	9:1	99.98

It can be seen from Table 7 that the higher the proportion of training data, the better the diagnosis result of the model. When the training data and test data are divided by 9:1, the diagnostic rate of the model is 99.98%. The experimental results show that the greater the number of training samples, the more specific fault features included in the samples, and the better the model diagnostic performance training. It is proven that this paper can achieve good results by dividing training samples and test samples in a ratio of 9:1.

The t-SNE visualization technique was used to map the high-dimensional features of the whole connection layer into two-dimensional features for visualization to demonstrate the feature extraction capability of the proposed method. As shown in Figure 21a, the data in the input layer are in a disordered state, and the features of the fully connected layer of the proposed method after processing are shown in Figure 21d. It can be seen that the proposed method gathers fault samples of the same type together well and effectively separates fault samples of different types. The results show that the proposed method has a better fault feature identification capability.
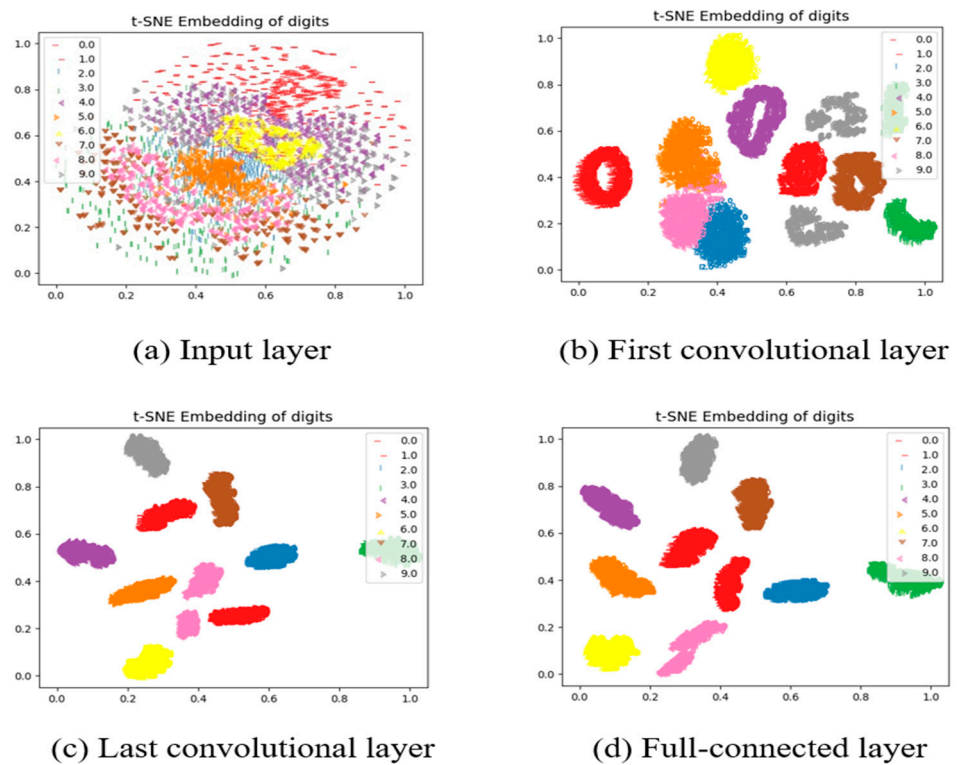
(a) Input layer

(b) First convolutional layer

(c) Last convolutional layer

(d) Full-connected layer

Figure 21. Results of t-SNE visualization analysis.

4.5.4. Comparison with Other Methods

To verify the effectiveness of the proposed method, the support vector machine (SVM), LeNet-5, Resnet [41], and multigrained cascade forest (gcForest) [42] methods were selected to conduct comparison group tests for verification. SVM adopts one-to-many classification, the Gaussian kernel function is selected as the kernel function, the penalty parameter is 30, and gamma is 0.01. LeNet-5 adopts a classical structure. The two-dimensional convolution kernel was improved to one-dimensional convolution to directly process one-dimensional vibration data, two-layer convolution, and three-layer full connection, and the full connection parameter is 120. Resnet was modified to directly process the input 1D vibration signal. The network structure of Resnet consists of five residual blocks, two pooling layers, and a fully connected layer. For gcForest, the number of estimators in each cascade layer n_estimators is set to 20, the number of trees in each cascade layer n_trees is set to 100, and the maximum cascade layer max_layers is set to 20.

SVM is a traditional machine learning method, LeNet-5 is a classical convolutional neural network model, Resnet is a recent CNN architecture, and gcForest is a new tree-based model. Under the same input signal setting, this paper compares these methods to ensure the reliability of the proposed method. To test the generalization ability of the model, the training data are divided into training sets and validation sets by quadrifold cross-validation for model evaluation. The cross-validation results of different models through multiple experiments are shown in Figure 22. To eliminate random errors, repeated validation and testing are carried out. The experimental results of different types of fault data under each model adopted the average results of 10 experiments, and the model performance evaluation results were finally obtained, as shown in Table 8. The table shows that the proposed method has the highest classification accuracy and the best stability.
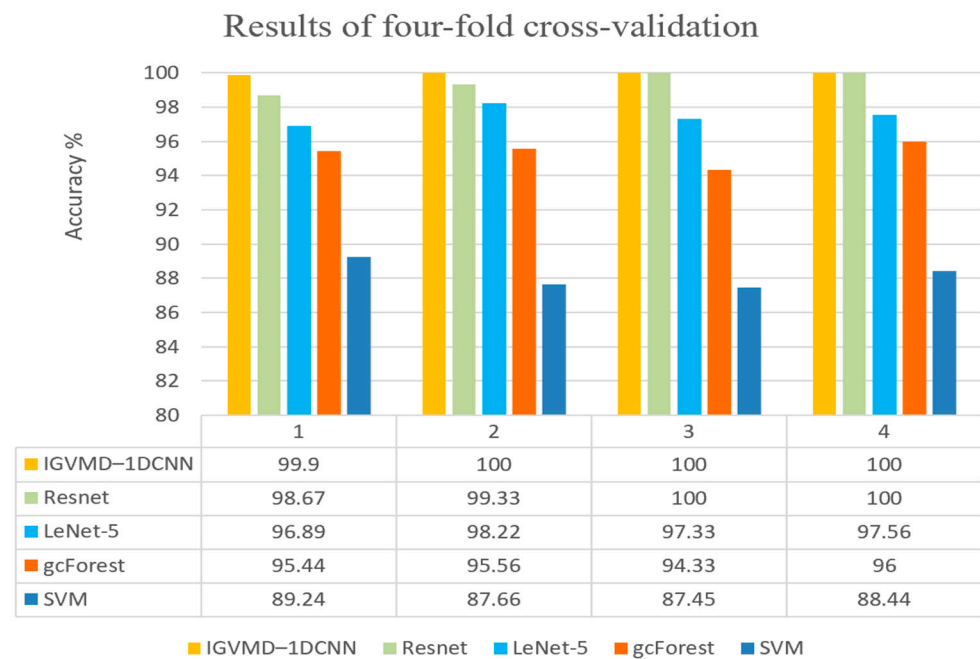
Figure 22. The cross-validation results of different models with CWRU.

Table 8. Comparison results with the different models using the CWRU dataset.

The Serial Number	Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
1	SVM	83.50	88.80	88.70	88.89
2	gcForest	96.70	96.60	96.47	96.44
3	LeNet-5	98.48	98.43	98.41	98.22
4	Resnet	99.58	99.58	99.58	99.56
5	IGVMD–1DCNN	99.98	99.98	99.98	99.98

The experimental results show that the accuracy of our method reached 99.98% on the test set, which is higher than the existing SVM, LeNet-5, gcForest, and Resnet methods, and the loss function value is 0.00003, which is very small. This indicates that the actual output of the proposed method is closest to the predicted value and has the best recognition effect compared with the other methods.

4.5.5. Performance under Different Noise Conditions

To verify the reliability and superiority of the method, the effect of different noise intensities on the performance of the diagnostic model is discussed in a comparison with four models from SVM, gcForest, LeNet-5, and Resnet. Gaussian white noise with different signal-to-noise ratios (SNRs) is added to the original signal to construct different composite noise signals and simulate different noise environments. The SNR is defined as follows:

$$\text{SNR(dB)} = 10 \log_{10} \left(\frac{P_{\text{singal}}}{P_{\text{noise}}} \right) \tag{28}$$

where P_{singal} denotes the signal power and P_{noise} denotes the noise power.

The training and test samples in the dataset under a 0 HP load were selected. Gaussian white noise with SNRs of −6, 0, and 6 dB were added to the test samples. The training strategy was the same for all 4 models, and each model is trained 10 times (50 epochs each) under the original sample to obtain 10 training models. Ten training models of each model were tested with noise test samples of different SNRs. The final results were averaged, and the experimental results are shown in Figure 23. When the SNR varied from −6 to 6 dB, only the IGVMD–1DCNN model consistently maintained more than 99% accuracy, which was significantly better than the other four models. In addition, the SNR = −6 dB indicates

that the noise power is much greater than the original signal power, indicating that the model is effective in detecting very weak fault signals. The IGVMD–1DCNN model has good performance in strong noise scenes, which indicates that the model has good noise robustness performance.
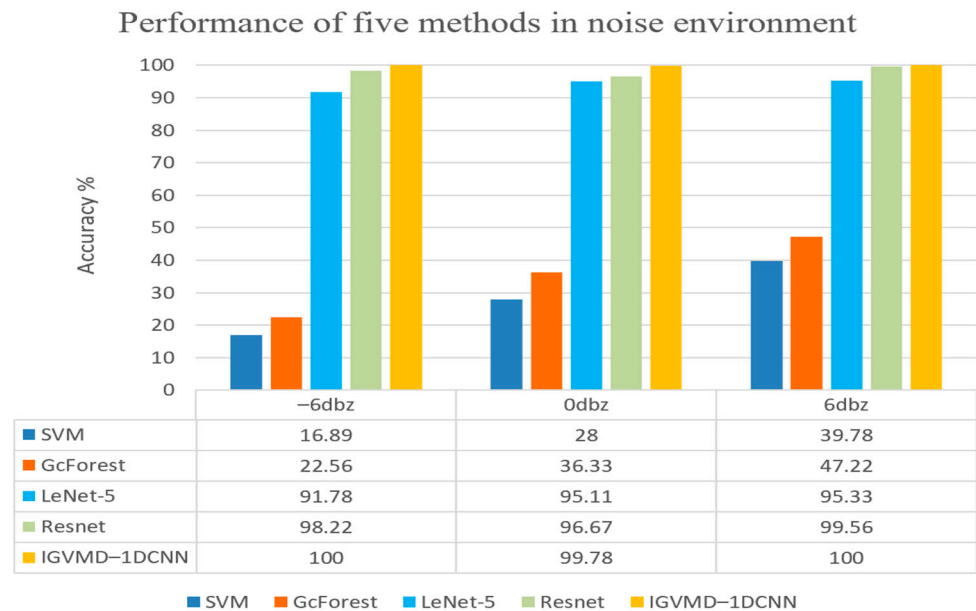

Figure 23. The performance of five methods in noise environment.

4.5.6. Comparison with Other Bearing Datasets

To further verify the effectiveness of the proposed method, bearing data from a gearbox at Southeast University were used for experimental verification [43].

The data were taken from the Drivetrain Dynamic Simulator (DDS), which collects a total of eight channels of data in the rotating speed-load configuration set to 20 Hz-0 V and 30 Hz-2 V. This experiment used channel 2 data at a speed-load setting of 20 Hz-0 V. The sampling rate is 20 Hz and number of data points is 1024. There are five types of the bearing data: normal, inner ring fault, outer ring fault, rolling element fault, and combined failure of inner and outer rings. The 500 samples of each fault are selected, and the training data and test data are divided by the ratio of 9:1.

The model structure used for this experiment was similar to that used for the Case Western Reserve University bearing dataset, and the specific model structure parameters are listed in Table 1. The results of the fourfold cross-validation of the different methods through multiple experiments are shown in Figure 24. To avoid the contingency of experimental results, the average results of 10 experiments are used, while the experimental results of the proposed method are compared with the SVM, LeNet-5, Resnet, and gcForest methods, and the results are shown in Table 9.

Table 9. Comparison results with the different models using the gearbox dataset.

The Serial Number	Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
1	SVM	36.08	39.11	37.53	37.60
2	gcForest	92.83	92.15	92.49	90.80
3	LeNet-5	84.83	85.16	84.99	84.80
4	Resnet	95.38	95.41	95.37	94.80
5	IGVMD–1DCNN	99.54	99.49	99.51	99.50

Results of four-fold cross-validation

	1	2	3	4
■ IGVMD–1DCNN	98.22	99.64	99.82	99.64
■ Resnet	95.2	94.8	94.8	94
■ LeNet-5	83.48	79.04	86.3	82.56
■ gcForest	90.76	91.12	90.39	91.1
■ SVM	35.88	36.23	35.05	32.92

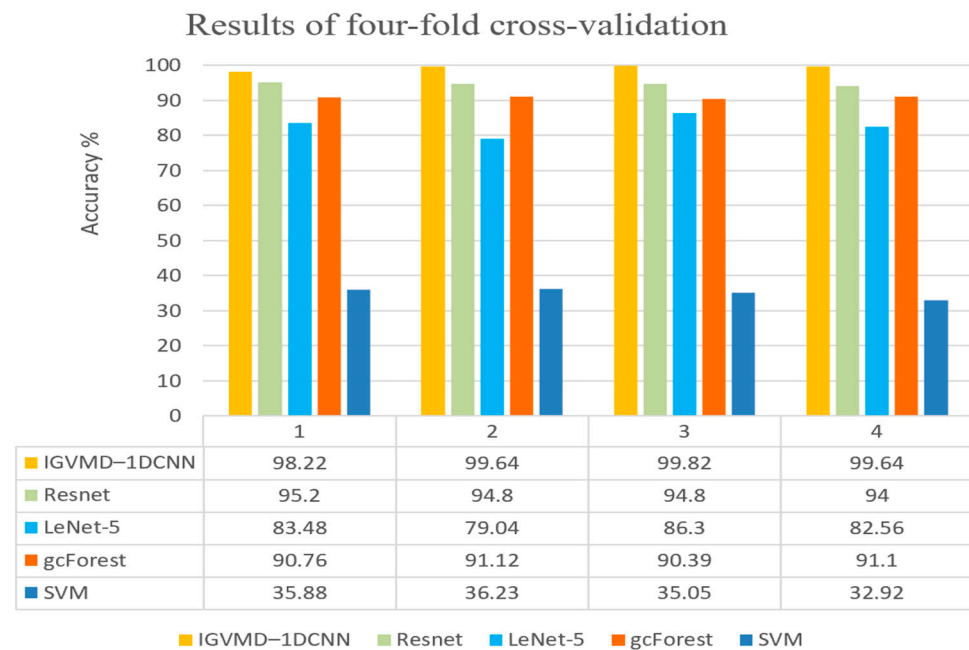■ IGVMD–1DCNN ■ Resnet ■ LeNet-5 ■ gcForest ■ SVM

Figure 24. The cross-validation results of different models.

Figure 24 and Table 9 indicate that compared with other methods, the proposed method significantly improved the accuracy and stability of the model. SVM is a shallow model that has limited feature learning ability, and the learned features do not have good classification characteristics. The gcForest deep model has a better diagnostic effect but has large memory consumption and low operation efficiency. Compared with LeNet-5, the depth of the model proposed is greater, and the feature extraction and diagnosis ability are stronger. The accuracy is improved compared with Resnet because this model eliminates the interference of strong noise. The fault identification rate of the proposed method is 99.50%, and it has the highest recognition accuracy. Therefore, the method proposed can be used for bearing fault diagnoses of different types with good generalization ability.

5. Conclusions

In this paper, the improved grey wolf optimized variational modalities method was used as a new method for fault feature extraction of rolling bearings, and a 1DCNN model with adaptive PReLU activation layers was suggested for identifying the fault conditions of the working machine. First, the improved grey wolf optimized variational model was applied to optimize the VMD penalty parameters to decrease the effect of background noise on the input data features. A PReLU activation layer was added to increase the training process of the 1DCNN network. The accuracy of fault classification and recognition was validated by experimental data and the concluding remarks are given as follows.

(1) First, the objective function, the minimal envelope entropy was used for the decomposition signal on the improved grey wolf optimized variational modal model. The comparison with EMD and traditional VMD verifies that the method in this paper is more accurate for early faint signal feature extraction, which can provide suitable conditions for subsequent rolling bearing fault classification detection.

(2) IMF components were screened by correlation coefficients to avoid the loss of fault feature signals compared to the random screening of modal components, so that fault features of rolling bearing signals under background noise conditions could be extracted effectively.

(3) Compared with traditional methods such as SVM, LeNet-5, Resnet, and gcForest, the IGVMD–1DCNN method has a better fault feature extraction effect and higher classification accuracy, which verifies that the IGVMD–1DCNN method as a new structured deep learning algorithm has better performance. The IGVMD–1DCNN

model reduces the dependence on the inspector's professional knowledge and practical testing experience through adaptive feature extraction and the complexity of the diagnosis process. It provides a practical method on fault diagnosis and bearing condition monitoring.

The fault types diagnosed in this paper focused on single working condition faults. Future work will investigate various operating conditions for compound fault types and build a more efficient and robust framework to accommodate different types of bearing fault diagnosis.

Author Contributions: Methodology, X.L., J.W. and X.W.; writing—original draft preparation, X.W.; writing—review and editing, X.L. and X.W.; formal analysis, S.B. and X.X.; funding acquisition, X.L. and Z.D.; supervision, X.X. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest and the manuscript is approved by all authors for publication.

Abbreviations

The following abbreviations are used in this manuscript:

VMD	Variational modal decomposition
GWO	Grey wolf optimization algorithm
IGVMD	Improved grey wolf optimized variational modal decomposition
1DCNN	One-dimensional convolutional neural network
SVM	Support vector machine
gcForest	Multi-grained cascade forest
ReLU	Rectified linear unit
PReLU	Parametric rectified linear unit

The list of symbols:

symbols

u_k	each modal component	a	the convergence factor
ω_k	central frequency	T	the maximal number of iterations
$\overrightarrow{X}$	location of the individual grey wolf	ρXY	correlation coefficient
a	convergence factor	$Corr(X, Y)$	the covariance between signals
K	The number of modal component	k	convolutional kernel
α	the quadratic penalty term	$*$	convolutional operator
$\overrightarrow{X}$	the location of the individual grey wolf	f	activation function
$\overrightarrow{A}, \overrightarrow{C}$	coefficient vectors	$maxpooling(x)$	the downsampling function

References

1. Jiang, F.; Zhu, Z.; Li, W. An improved VMD with empirical mode decomposition and its application in incipient fault detection of rolling bearing. *IEEE Access* **2018**, *6*, 44483–44493. [CrossRef]
2. Li, H.; Liu, T.; Wu, X.; Chen, Q. Research on bearing fault feature extraction based on singular value decomposition and optimized frequency band entropy. *Mech. Syst. Signal. Process.* **2019**, *118*, 477–502. [CrossRef]
3. Ding, X.; Li, Q.; Lin, L.; He, Q.; Shao, Y. Fast time-frequency manifold learning and its reconstruction for transient feature extraction in rotating machinery fault diagnosis. *Measurement* **2019**, *141*, 380–395. [CrossRef]

4. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. R. Soc. Lond. A Mat.* **1998**, *454*, 903–995. [CrossRef]

5. Li, S.; Wang, H.; Song, L.; Wang, P.; Cui, L.; Lin, T. An adaptive data fusion strategy for fault diagnosis based on the convolutional neural network. *Measurement* **2020**, *165*, 108122. [CrossRef]

6. Ji, J.; Qu, J.; Chai, Y.; Zhou, Y.; Tang, Q.; Ren, H. An algorithm for sensor fault diagnosis with EEMD-SVM. *Trans. Inst. Meas. Control.* **2018**, *40*, 1746–1756. [CrossRef]

7. Lu, Y.; Xie, R.; Liang, S.Y. CEEMD-assisted kernel support vector machines for bearing diagnosis. *Int. J. Adv. Manuf. Tech.* **2020**, *106*, 3063–3070. [CrossRef]

8. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE. Trans. Signal. Process.* **2013**, *62*, 531–544. [CrossRef]

9. Tang, G.J.; Wang, X.L. Parameter optimized variational mode decomposition method with application to incipient fault diagnosis of rolling bearing. *J. Xi'an Jiaotong Univ.* **2015**, *49*, 73–81.

10. Yang, W.; Peng, Z.; Wei, K.; Shi, P.; Tian, W. Superiorities of variational mode decomposition over empirical mode decomposition particularly in time-frequency feature extraction and windturbine condition monitoring. *IET Renew. Power Gen.* **2017**, *11*, 443–452. [CrossRef]

11. Ni, Q.; Ji, J.C.; Feng, K.; Halkon, B. A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis. *Mech. Syst. Signal. Process.* **2022**, *164*, 108216. [CrossRef]

12. He, D.; Liu, C.; Jin, Z.; Ma, R.; Chen, Y.; Shan, S. Fault diagnosis of flywheel bearing based on parameter optimization variational mode decomposition energy entropy and deep learning. *Energy* **2022**, *239*, 122108. [CrossRef]

13. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey wolf optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]

14. Zhao, H.; Guo, S.; Gao, D. Fault feature extraction of bearing faults based on singular value decomposition and variational modal decomposition. *J. Vib. Shock.* **2016**, *35*, 183–188.

15. Rodriguez, L.; Castillo, O.; Soria, J. Grey wolf optimizer with dynamic adaptation of parameters using fuzzy logic. In Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, Canada, 24–29 July 2016; pp. 3116–3123.

16. Xing, H.; Zhou, X.; Wang, X.; Luo, S.; Dai, P.; Li, K.; Yang, H. An integer encoding grey wolf optimizer for virtual network function placement. *Appl. Soft Comput.* **2019**, *76*, 575–594. [CrossRef]

17. Yang, Y.; Yang, B.; Wang, S.; Jin, T.; Li, S. An enhanced multi-objective grey wolf optimizer for service composition in cloud manufacturing. *Appl. Soft Comput.* **2020**, *87*, 106003. [CrossRef]

18. Wang, Y.; Zhang, T.; Cai, Z.; Zhao, J.; Wu, K. Multi-UAV coordination control by chaotic grey wolf optimization based distributed MPC with event-triggered strategy. *Chin. J. Aeronaut.* **2020**, *33*, 2877–2897. [CrossRef]

19. Miao, D.; Chen, W.; Zhao, W.; Demsas, T. Parameter estimation of PEM fuel cells employing the hybrid grey wolf optimization method. *Energy* **2020**, *193*, 571–582. [CrossRef]

20. Mantegna, R.N. Fast, accurate algorithm for numerical simulation of Levy stable stochastic processes. *Phys. Rev. E* **1994**, *49*, 4677–4683. [CrossRef]

21. Dibaj, A.; Hassannejad, R.; Ettefagh, M.M.; Ehghaghi, M.B. Incipient fault diagnosis of bearings based on parameter-optimized VMD and envelope spectrum weighted kurtosis index with a new sensitivity assessment threshold. *ISA. Trans.* **2021**, *114*, 413–433. [CrossRef]

22. Emary, E.; Zawbaa, H.M.; Grosan, C.; Hassenian, A.E. Feature subset selection approach by gray-wolf optimization. In Proceedings of the Afro-European Conference for Industrial Advancement, Addis Ababa, Ethiopia, 17–19 November 2014; Springer: Cham, Switzerland, 2014; pp. 1–13.

23. Chen, Z.; Gryllias, K.; Li, W. Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network. *IEEE Trans. Ind. Inform.* **2019**, *16*, 339–349. [CrossRef]

24. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NE, USA, 26 June–1 July 2016; pp. 770–778.

26. Lu, C.; Wang, Z.; Zhou, B. Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. *Adv. Eng. Informat.* **2017**, *32*, 139–151. [CrossRef]

27. Chen, Z.; Gryllias, K.; Li, W. Mechanical fault diagnosis using convolutional neural networks and extreme learning machine. *Mech. Syst. Signal Process.* **2019**, *133*, 106272. [CrossRef]

28. Jiang, G.; He, H.; Yan, J.; Xie, P. Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. *IEEE Trans. Ind. Electron.* **2018**, *66*, 3196–3207. [CrossRef]

29. Ince, T.; Kiranyaz, S.; Eren, L.; Askar, M.; Gabbouj, M. Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Trans. Ind. Electron.* **2016**, *63*, 7067–7075. [CrossRef]

30. Peng, D.; Liu, Z.; Wang, H.; Qin, Y.; Jia, L. A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains. *IEEE Access* **2018**, *7*, 10278–10293. [CrossRef]

31. Jiang, W.; Wang, C.; Zou, J.; Zhang, S. Application of Deep Learning in Fault Diagnosis of Rotating Machinery. *Processes* **2021**, *9*, 919. [CrossRef]

32. Li, X.; Li, J.; Zhao, C.; Qu, Y.; He, D. Gear pitting fault diagnosis with mixed operating conditions based on adaptive 1D separable convolution with residual connection. *Mech. Syst. Signal. Process.* **2020**, *142*, 106740. [CrossRef]

33. Hao, S.; Ge, F.X.; Li, Y.; Jiang, J. Multisensor bearing fault diagnosis based on one-dimensional convolutional long short-term memory networks. *Measurement* **2020**, *159*, 107802. [CrossRef]

34. Habbouche, H.; Amirat, Y.; Benkedjouh, T.; Benbouzid, M. Bearing Fault Event-Triggered Diagnosis using a Variational Mode Decomposition-based Machine Learning Approach. *IEEE Trans. Energy Conver.* **2021**, *37*, 466–474. [CrossRef]

35. Ding, Y.; Jia, M. A multi-scale convolutional auto-encoder and its application in fault diagnosis of rolling bearings. *J. Southeast Univ.* **2019**, *35*, 417–423.

36. Ma, S.; Liu, W.; Cai, W.; Shang, Z.; Liu, G. Lightweight deep residual CNN for fault diagnosis of rotating machinery based on depthwise separable convolutions. *IEEE Access* **2019**, *7*, 57023–57036. [CrossRef]

37. Yao, L.; Fang, Z.; Xiao, Y.; Hou, J.; Fu, Z. An intelligent fault diagnosis method for lithium battery systems based on grid search support vector machine. *Energy* **2021**, *214*, 118866. [CrossRef]

38. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal. Process.* **2015**, *64–65*, 100–121. [CrossRef]

39. Jiang, X.Y.; Li, S. BAS: Beetle antennae search algorithm for optimization problems. *Int. J. Control* **2018**, *1*, 1599–1603. [CrossRef]

40. Ouadine, A.Y.; Mjahed, M.; Ayad, H.; El Kari, A. Aircraft Air Compressor Bearing Diagnosis Using Discriminant Analysis and Cooperative Genetic Algorithm and Neural Network Approaches. *Appl. Sci.* **2018**, *8*, 2243. [CrossRef]

41. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 630–645.

42. Zhou, Z.H.; Feng, J. Deep forest: Towards an alternative to deep neural networks. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 3553–3559.

43. Shao, S.; McAleer, S.; Yan, R.; Baldi, P. Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2446–2455. [CrossRef]