Article

Localization of Rolling Element Faults Using Improved Binary Particle Swarm Optimization Algorithm for Feature Selection Task

Chun-Yao Lee * and Guang-Lin Zhuo

Department of Electrical Engineering, Chung Yuan Christian University, No. 200, Zhongbei Road, Zhongli District, Taoyuan City 320, Taiwan; s10528245@cycu.org.tw
* Correspondence: cyl@cycu.edu.tw; Tel.: +886-3-2654827

Abstract: The accurate localization of the rolling element failure is very important to ensure the reliability of rotating machinery. This paper proposes an efficient and anti-noise fault diagnosis model for rolling elements. The proposed model is composed of feature extraction, feature selection and fault classification. Feature extraction is composed of signal processing and signal noise reduction. Signal processing is carried out by local mean decomposition (LMD), and signal noise reduction is performed by product function (PF) selection and wavelet packet decomposition (WPD). Through the steps of signal noise reduction, high-frequency noise can be effectively removed, and the fault information hidden under the noise can be extracted. To further improve the effectiveness of the diagnostic model, an improved binary particle swarm optimization (IBPSO) is proposed to find the most important features from the feature space. In IBPSO, cycling time-varying inertia weight is introduced to balance exploitation and exploration and improve the capability to escape from local solutions, and crossover and mutation operations are also introduced to improve exploration and exploitation capabilities, respectively. The main contributions of this research are briefly described as follows: (1) The feature extraction process applied in this research can effectively remove noise and establish a high-accuracy feature set. (2) The proposed feature selection algorithm has higher accuracy than the other state-of-the-art feature selection algorithms. (3) In a strong noise environment, the proposed rolling element fault diagnosis model is compared with the state-of-the-art fault diagnosis model in terms of classification accuracy. Experimental results show that the model can maintain high classification accuracy in a strong noise environment. Therefore, it can be proved that the fault diagnosis model proposed in this paper can be effectively applied to the fault diagnosis of rotating machinery.

Keywords: rolling element; fault diagnosis model; local mean decomposition; wavelet packet decomposition; binary particle swarm optimization

1. Introduction

With the improvement of industrial automation levels, the development of rotating machinery is more precise than ever. Therefore, the monitoring and fault diagnosis methods of rotating machinery have always been the field that researchers are committed to developing [1]. M. Van and H.J. Kang [2] proposed a bearing fault diagnosis model. The model combines a new feature extraction technology based on non-local mean denoising and empirical mode decomposition (EMD), and a two-stage feature selection technology based on hybrid distance evaluation technology (DET) and particle swarm optimization (PSO). The model proved its effectiveness in bearing failure experiments. F. Alvarez-Gonzalez et al. [3] proposed an online statistical analysis method based on Hilbert–Huang transform (HHT) to detect permanent magnet synchronous motor (PMSM) stator short-circuit faults and proved reliable fault detection through simulation results. S. Haroun et al. [4] proposed multiple feature extraction techniques to detect stator winding faults of induction motors.

First, the three-phase stator current is analyzed using Park transform, zero-crossing time signal and envelope. Then, the time domain and frequency domain statistical features are extracted from the analysis results. Experimental results show that the proposed method can detect stator winding faults and identify fault phases under various faulty cases and different load variations. The above results prove that fault diagnosis can greatly improve system reliability, reduce maintenance costs and even avoid major production losses caused by failures. In addition, in the statistics of the Electric Power Research Institute (EPRI), rolling element failures accounted for 41% of all rotating machinery failures, the highest proportion of failure types [5]. Moreover, even early rolling element fault can quickly develop into serious fault [6]. Therefore, this study focuses on constructing an efficient rolling element fault diagnostic model.

In recent years, with the advent of accelerometers, it is easy to measure vibration signals and generally provide a wide frequency range, so fault diagnosis models based on vibration signals have been widely proposed. Z. Wang et al. proposed an efficient and robust hybrid model, using wavelet packet decomposition (WPD) and mutual dimension-less indexing to extract the best features, and random forest for classification [7]. Y. Shao et al. proposed a rolling element fault diagnosis model based on the principle of coherent demodulation. By extracting the feature frequencies of different fault types, the fault type can be accurately classified [8]. Z. Huo et al. proposed a rolling element fault diagnosis model, which can be effectively applied to multi-speed environments. The model uses particle swarm optimization and quasi-Newton minimization algorithm to optimize the parameters of the continuous wavelet transform (WT) model. Then, it performs feature extraction in the 3-D feature space and the k-nearest neighbor (k-NN) classifier for fault classification [9]. S. Wei et al. proposed a time-varying envelope filtering (TVEF) to extract the features of rolling element faults. Using the instantaneous frequency and instantaneous amplitude extracted by this method to reconstruct the high-resolution time-frequency distribution can more accurately extract the fault features [10].

The fault diagnosis model based on vibration signals is usually divided into three stages: feature extraction, feature selection and fault classification. Among them, feature extraction and fault classification are key stages. The signal processing technology in feature extraction [9] is an important step in reducing the dimensionality of vibration signals and extracting key fault messages. Due to the complex working environment of rotating machinery, the measured vibration signal contains non-stationary components and noise. Signal processing techniques based only on the time domain or frequency domain may not be effective. Therefore, some time-frequency analysis signal processing techniques such as fast Fourier transform (FFT), short-time Fourier transform (STFT) and continuous wavelet transform (CWT) are widely used. The classification of signal analysis results by neural network (NN) or machine learning (ML) is the final stage. Feature selection [11] is an option of the model, and its function is to solve the diagnosis performance degradation caused by redundant or irrelevant fault features.

However, the above-mentioned time-frequency analysis signal processing techniques still have their own limitations. For example, FFT and STFT affect the decomposition performance due to the use of fixed-length windows [12]. CWT solves the problems of FFT and STFT with an adjustable window size [13]. However, once the decomposition scale of CWT is defined, CWT can only decompose signals in the defined frequency band, which makes CWT non-adaptive [14]. Based on the above analysis, adaptive signal processing technology may be able to analyze vibration signals more effectively [15]. Local mean decomposition (LMD) [16] is a new adaptive signal processing technology and has many advantages that can be applied to rolling element fault diagnosis. First, the decomposition process of LMD does not need to use Hilbert transformation (HT), so it will not encounter negative frequencies [17]. Secondly, LMD can decompose and demodulate signals at the same time. Third, LMD can decompose the amplitude and frequency modulation characteristics of the vibration signal when the rolling element fails [18,19].

However, in the actual working environment, the features of rolling element faults are usually masked by noise or other rotating machinery components' disturbing vibrations [20,21], causing LMD to decompose redundant or irrelevant product function (PF) components. Therefore, this study used a denoising technique combining PF selection and WPD. First, the PF selection method removes redundant or irrelevant PF components and selects the most valuable PF components for further denoising. Then, WPD denoising technology can effectively remove noise and present fault information in wavelet packet coefficients [22,23]. Finally, the fault features are extracted from these wavelet packet coefficients.

Although the hard work in the feature extraction stage extracts the fault features in the original signal, there may still be redundant or irrelevant fault features in the feature set, resulting in a decrease in diagnostic performance [11]. Therefore, feature selection is applied to prevent overfitting and improve model performance [11]. Feature selection can be divided into filter methods and wrapper methods. The filter method mainly uses correlation coefficient (CC) or univariate mutual information (MI) to calculate the linear intensity between each input and the target, and sorts according to their intensity and removes irrelevant features [24]. The wrapper method combined with a specific classifier for accurate evaluation can usually achieve better performance than the filter method [11,24]. Therefore, some optimization algorithms such as binary particle swarm optimization (BPSO) [25], genetic algorithm (GA) [26] and binary chicken swarm optimization (BCSO) [27] are widely used in feature selection. However, the above algorithms generally have many defects, such as premature convergence [28,29] and falling into local optima [28–30]. Although there is no optimization algorithm that can guarantee the best feature subset, PSO has successfully solved many nonlinear optimization problems in the engineering field due to its excellent computational efficiency and simple operation [31,32]. Therefore, PSO is still an optimization algorithm that many researchers are dedicated to researching [33–35]. Therefore, this study proposes an improved binary particle swarm optimization (IBPSO) as the feature selection task of the fault diagnosis model. In this study, three mechanisms are proposed to improve the performance of PSO. First, cycling time-varying inertia weights are introduced to balance exploration and exploitation and enhance the capability to avoid local optima. Considering the crossover and mutation mechanism can improve the exploration and exploitation capabilities of PSO and solve the problem of premature convergence of PSO.

Fault classification is another important stage that constitutes a rolling element fault diagnosis model. Researchers have widely used NN and ML in the fault diagnosis of rolling elements [36,37]. Traditional NN such as multilayers perceptron (MLP) has the problem of a complex structure and difficult training process [38]. ML has the advantage of being simple and easy to implement, and the classification results are better, especially the support vector machine (SVM) algorithm, which has many papers to prove its classification efficiency and anti-noise capability [39,40]. In recent years, a new type of NN, fully connected neural network (FCNN), achieves powerful performance through a new way of connecting neurons. FCNN has the following advantages: (1) The complexity of FCNN is like that of traditional single hidden layer NN, but the performance is very powerful [38]. (2) The addition of too many neurons in the traditional single hidden layer NN leads to overfitting and poor generalization of the model [41]. FCNN needs fewer neurons to achieve powerful performance and good generalization [42]. (3) In [42], the author proved that FCNN has excellent anti-noise capability and can complete classification under low signal-to-noise ratio (SNR). (4) In [38], the author showed the excellent performance of FCNN, and its classification performance is better than SVM. Therefore, in this study, we adjusted the number of layers and the number of neurons and compared the performance of five FCNNs to establish the most robust rolling element fault diagnosis model. The advantages of the above feature extraction, feature selection and fault classification motivate us to propose a robust rolling element diagnosis model with both classification accuracy and anti-noise capability.

The organization of this paper is as follows: Section 2 introduces the basic methods of the proposed model, including feature extraction process, binary particle swarm algorithm and fully connected neural network. Section 3 introduces the detailed description of the improved binary particle swarm algorithm and the flow of the rolling element fault diagnosis model. Section 4 discusses the experimental results of the University of California, Irvine (UCI) feature selection dataset and Case Western Reserve University (CWRU) rolling element failure dataset. Section 5 evaluates the diagnostic model and future work. Finally, Section 6 explains the conclusion.

2. Methodology

In this section, three important stages in fault diagnosis are presented: feature extraction, feature selection and classification.

2.1. Feature Extraction

In actual cases, the fault features are usually masked by a lot of noise (e.g., background noise and Gaussian noise). Therefore, the most important step in the fault diagnosis model is the denoising and extraction of fault features. In the feature extraction process of this study, first, local mean decomposition (LMD) decomposes the vibration signal into a set of product function (PF) components. Second, it uses PF selection to select the PF that contains the most fault information. Then, wavelet packet decomposition (WPD) is used for analysis further of fault information and denoising. Finally, the potential features of bearing fault features are extracted.

2.1.1. Local Mean Decomposition

The LMD is an effective technique that usually adopts to decompose non-stationary signals [16]. LMD can decompose non-stationary signals into simple PFs. Each PF is the product of the envelope signal and the pure frequency modulated (FM) signal. Then, the time-frequency distribution of the original signal can be easily derived. The process of LMD decomposing the vibration signal $v(t)$ is shown as following a loop.

Step 1. Calculate all local extrema $(e_1, e_2, \dots, e_i, \dots)$ from $v(t)$.

Step 2. All the local mean value m_i and local envelope value a_i can be determined by two successive local extrema e_i and e_{i+1}.

$$m_i = \frac{(e_i + e_{i+1})}{2} \tag{1}$$

$$a_i = \frac{|e_i - e_{i+1}|}{2}. \tag{2}$$

Step 3. Apply the moving averaging (MA) method to smooth the local mean function $m_{11}(t)$ and local envelope function $a_{11}(t)$. $m_{11}(t)$ and $a_{11}(t)$ are the straight lines extending between the successive local extrema of the signal.

Step 4. The residue signal $r_{11}(t)$ can be subtracted by the local mean function $m_{11}(t)$ from the original signal $v(t)$.

$$r_{11}(t) = v(t) - m_{11}(t) \tag{3}$$

Step 5. The FM signal $f_{11}(t)$ can be obtained from $r_{11}(t)$ and $a_{11}(t)$:

$$f_{11}(t) = \frac{r_{11}(t)}{a_{11}(t)}. \tag{4}$$

Step 6. Take $f_{11}(t)$ as a new signal and Step 1 to 5 will be repeated until $f_{11}(t)$ is a purely FM signal. Then, go to Step 7.

Step 7. The instantaneous envelope function $a_1(t)$ is obtained by all local envelope functions produced during iteration.

$$a_1(t) = a_{11}(t)a_{12}(t)\ldots a_{1n}(t) = \prod_{q=1}^{n} a_{1q}(t) \tag{5}$$

Then, the first product function PF_1 is generated by multiplying the instantaneous envelope function and the final FM signal:

$$PF_1 = a_1(t)f_{1n}(t). \tag{6}$$

Step 8. The new signal u_1 is treated as $v(t)$ minus PF_1. Then, the above process is repeated from Step 1 to 7 until $u_i(t) = u_{i-1}(t) - PF_i(t)$ is a monotonic function. Finally, $v(t)$ is decomposed into a set of PF components.

$$v(t) = \sum_{i=1}^{p} PF_i(t) + u_p(t). \tag{7}$$

2.1.2. Product Function Selection

The second step is to remove redundant components and extract the components that contain most of the fault information. Some parameters are used to extract those effective components, such as kurtosis [23], root mean square (RMS) [43] and correlation coefficient (CC) [44]. Kurtosis is very sensitive to early or weak failures, but when the failures become more severe, kurtosis cannot maintain an increasing trend [43]. RMS is not sensitive to early failure [45]. CC is used to evaluate the similarity of components, but it is not sensitive to early failures [17]. Therefore, a single parameter cannot effectively select components. This study uses weights to balance the statistical values of the above parameters and effectively select the PF component.

2.1.3. Wavelet Packet Decomposition

The third step is to use WPD to further analyze and remove high-frequency noise to extract the fault features that are hidden under the noise. WPD is a powerful noise reduction tool due to its high resolution at both high and low frequencies [22].

The wavelet packet coefficients (WPC) of a signal can be described as follows:

$$WPC_{j,n,k}(t) = \int x(t)WPT_{j,k}^{n}(t)dt \tag{8}$$

where $x(t)$ is the signal, $WPT_{j,k}^{n}(t)$ denotes the wavelet packet function, j and k are scale and translation values, respectively, and $n = 1, 2, \ldots, 2^j$ is the oscillation parameter.

2.2. Binary Particle Swarm Optimization for Feature Selection

The traditional particle swarm optimization (PSO) is based on the concept of swarm exploration of the search space to find solutions. Each particle moves in the search space at a certain velocity and updates the velocity according to its own search experience and the group's search experience in each iteration, as shown in Equation (9) [25]. In Equation (9), each particle will update its velocity according to the best position it has searched (personal best) and the best position searched by the group (global best).

$$v_{i+1} = wv_i + ac_1 Rand(pb_i - p_i) + ac_2 Rand(gb - p_i) \tag{9}$$

where i represents the current iteration; w denotes the inertia weight; ac_1 and ac_2 denote the acceleration coefficients to control the strength of exploration and exploitation; v_i and p_i are current velocity and current position of particle, respectively; and pb_i and gb represent the personal best position and global best position, respectively.

In binary particle swarm optimization (BPSO) [25] for feature selection, each particle is converted from a continuous position to a position in the binary search space through a transfer function. The converted position is represented as a bit string (i.e., feature subset), which means that the corresponding feature is selected (1) or not selected (0). The sigmoid function is used as a transfer function in this study. The position of each particle is updated based on the velocity of each particle. The steps to transfer from a continuous position to a binary position are described as follows:

$$sig(v_{i+1}) = \frac{1}{1 + e^{-v_{i+1}}} \tag{10}$$

$$p_{i+1} = \begin{cases} 1 & if \quad Rand < sig(v_{i+1}) \\ 0 & else \end{cases} \tag{11}$$

3. Proposed Rolling Element Fault Diagnosis Model

This section details the proposed rolling element fault diagnosis model based on three main stages: feature extraction, improved binary particle swarm optimization for feature selection and classifier.

3.1. Feature Extraction Process

In the feature extraction process, first, the motor signal is decomposed into PF components using LMD. Second, the PF selection is used to select the effective PF components which contain the most failure information. In this study, we choose two effective PF components. Third, WPD is used to denoise the effective PF components and extract failure information further. In this study, a two-level WPD is used to decompose the effective PF components into four wavelet packet coefficients. Finally, eight statistical fault features such as max value, min value, root mean square, mean square error, standard deviation, kurtosis, crest factor and clearance factor are extracted from each WPC and extract a total of 64 ($2 \times 4 \times 8$) features. Figure 1 illustrates the detail of feature extraction process.
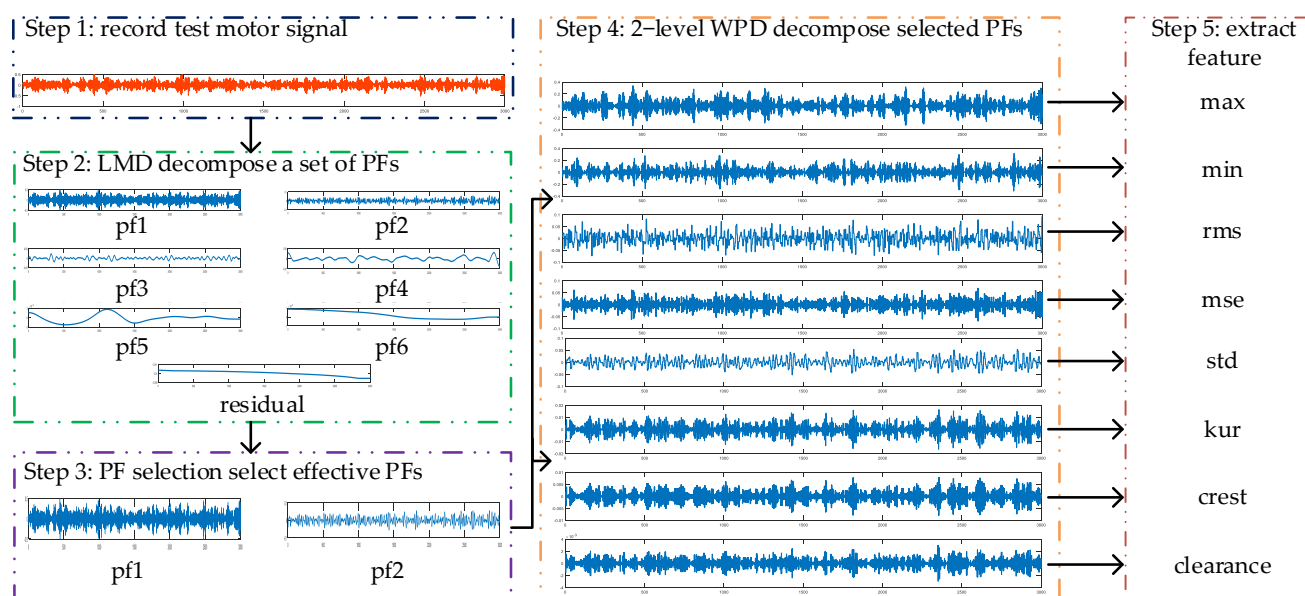

Figure 1. Feature extraction process.

3.2. Improved Binary Particle Swarm Optimization

In this subsection, two mechanisms and one parameter are introduced to enhance the BPSO. First, cycling time-varying inertia weight can not only balance exploration and exploitation but also enhance the capability to escape the local optima. Second, the

position update formula based on crossover operation can select high-potential solutions to generate better solutions and improve the exploration capability. Finally, the position update formula based on mutation operation can improve the capability to escape the local optima without increasing the computational cost.

3.2.1. Cycling Time-Varying Inertia Weight

In BPSO, the inertia weight is an important parameter responsible for balancing exploitation and exploration in the algorithm. The inertia reduction of the weight value can make the algorithm transition from exploration to exploitation smoothly. However, BPSO still has the problem of premature convergence, and the algorithm cannot continue to find a better solution in the later stages of the iteration. In this study, cycling time-varying inertia weight is applied to not only balance exploitation and exploration but also improve the algorithm's capability to escape local solutions [46]. The cycling time-varying inertia weight will cycle the number of cycles set by the user, so that the weight value linearly decreases from the maximum weight value (2) to the minimum weight value (0), and linearly increases from the minimum weight value (0) to the maximum weight value (2) until the number of cycles set by the user is met. The cycling time-varying inertia weight defined in Equation (12) is calculated as follows:

$$w = \left| 2 - \frac{(t \bmod T/C)}{T/(4C)} \right| \tag{12}$$

where C is user defined cycling time and t and T denote the current iteration and maximum number of iterations, respectively.

3.2.2. Improve the Exploration Capability Based on Crossover Operation

In genetic algorithms, crossover operations can explore new areas in the search space, avoid premature convergence and further improve convergence accuracy. In this study, the three-point crossover operation is combined with BPSO. During the iteration process, when the best solution of the current population matches the best solution of the population obtained in the previous iteration, it may indicate that the overall evolution is slow, and a three-point crossover operation will be performed to create a high-potential solution. The three-point crossover operation formula is shown in Equation (13). First, a pair of solutions are randomly selected from the individual best solutions of each particle. Next, the three-point crossover operation randomly selects three positions to cut the pair of solutions. Then, the pair of solutions swaps segments to create two new solutions. Finally, one of the new solutions is randomly selected as the current solution.

$$p_i = Crossover(pb_{rand1}, pb_{rand2}) \tag{13}$$

where *Crossover* is the three-point crossover operation applied in this study, and pb_{rand1} and pb_{rand2} are randomly selected particles from the individual best solutions.

3.2.3. Escape the Local Trap Based on Mutation Operation

Introducing the mutation operation in the genetic algorithm, this operation enables the solution to move slightly in the search space, increases the capability of escaping from the local solution and increases the diversity of the population further. In this study, the mutation probability *MR* is introduced to specify whether to change the position of each particle in the specified dimension. The mutation operation formula is shown in Equation (14). If *Rand*(*i*) (*i* = 1, 2, . . . , *n*) < *MR*, one dimension of the *i*th particle is selected randomly, and its value is mutated.

$$p_{i+1} = \begin{cases} 1 - p_i & if \quad Rand(i) < MR \\ p_i & else \end{cases} \tag{14}$$

where p_i is current position of *i*th particle.

3.3. Rolling Element Fault Diagnosis Model

The proposed rolling element fault diagnosis model can be divided into three stages, including feature extraction, feature selection and classification.

In the feature extraction stage, as shown in Section 3.1, the measured signal is decomposed into a set of PF components through LMD. Next, the most important PF components are selected. Then, wavelet packet decomposition is used to extract further the fault information and denoise the most important PF components. Finally, eight statistical fault features are extracted through each wavelet packet coefficient.

In the feature selection stage, as shown in Section 3.2, the improved binary particle swarm optimization (IBPSO) optimizes the fault feature set obtained in the feature extraction stage to remove redundant features. This stage can improve the classification accuracy of the rolling element fault diagnosis model. The process of IBPSO is shown in Figure 2.
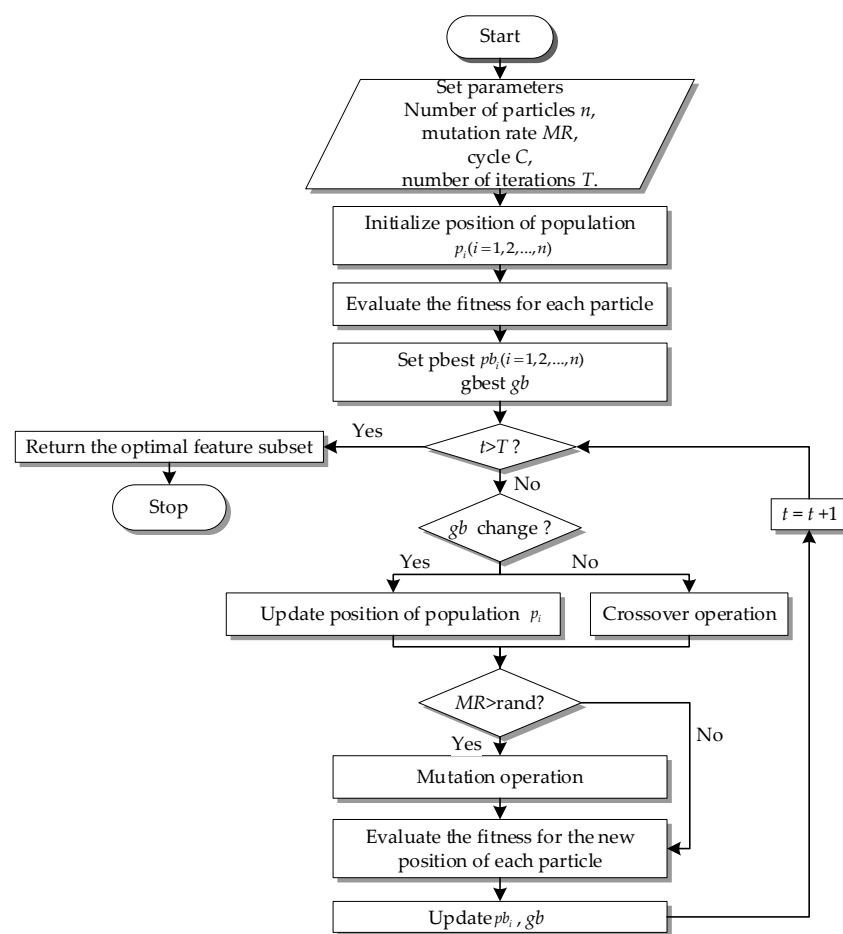
Figure 2. The procedure of IBPSO.

In the fault classification stage, the best feature subset obtained in the feature selection stage is classified through the fully connected neural network. In this study, the hyperparameters of the fully connected neural network will be adjusted to obtain the best classification accuracy to obtain the best rolling element fault diagnosis model.

4. Results

In this study, the results can be divided into two experiments, namely, the University of California, Irvine (UCI) feature selection dataset [47] and the Case Western Reserve University (CWRU) bearing failure dataset [48], and the experiments were simulated on Intel(R) Core (TM) i7-3930K CPU @ 3.20 GHz 3.60 GHz, 24 GB RAM and MATLAB platform.

4.1. Experiment 1: UCI Feature Selection Datasets

4.1.1. Experiment Setup and Parameter Setting

To evaluate the effectiveness of IBPSO in the field of feature selection, nine UCI feature selection datasets were used in this experiment, including BreastCancer, Wine, CongressEW, SpectEW, BreastEW, Ionosphere, krvskp, WaveformEW and Sonar. Table 1 describes the nine UCI feature selection data sets used in this experiment. In this experiment, the results of feature selection by IBPSO for nine UCI feature selection data sets are shown and compared with three basic feature selection algorithms, including BPSO, GA and BCSO. Finally, IBPSO is compared with other state-of-the-art feature selection models.

Table 1. Description of UCI benchmark datasets used in this case study.

Datasets	Features	Instances	Classes
BreastCancer	10	699	2
Wine	13	178	3
CongressEW	16	435	2
SpectEW	22	267	2
BreastEW	30	569	2
Ionosphere	34	351	2
krvskp	36	3196	2
WaveformEW	40	5000	3
Sonar	60	208	2

In this experiment, the k-nearest neighbors (k-NN) classifier is used as the classifier of the wrapper feature selection model. Table 2 is the parameter setting of this experiment. T is regarded as the convergence criterion. Each model performed 30 independent feature selections on all datasets and collected experimental results, including the average classification accuracy, the average number of selected features, the standard deviation of the classification error and the average computational time. Based on the requirements of a high-accuracy fault diagnosis model, minimizing the classification error is the primary task of the proposed feature selection model, so the classification error is defined as the fitness value in this experiment.

Table 2. Parameter setting for this experiment.

Parameter	Value
Number of nearest neighbor of k-NN classifier	1
k-fold cross-validation	10
Number of solutions	10
Maximum number of iterations T	100
Independent runs	30
Inertia weight w in BPSO	[0.9, 0.4]
Acceleration c_1 and c_2 in BPSO	2.05
Crossover rate in GA	0.8
Mutation rate in GA and IBPSO	0.01
Rooster parameter in BCSO	0.15
Hen parameter in BCSO	0.7
Mother parameter in BCSO	0.5
C in IBPSO	4

4.1.2. Comparison with Basic Feature Selection Algorithm

This subsubsection shows the convergence curves of feature selection algorithms including IBPSO, BPSO, GA, and BCSO for each dataset, as shown in Figure 3. Observing from the convergence curve in Figure 3, IBPSO has lower classification errors than other feature selection algorithms in all datasets when it reaches the defined convergence criterion. Especially in the BreastCancer, WaveformEW and Sonar datasets, IBPSO has a relatively

poor initial population, but IBPSO can continue to find better solutions and eventually achieve the best classification error.
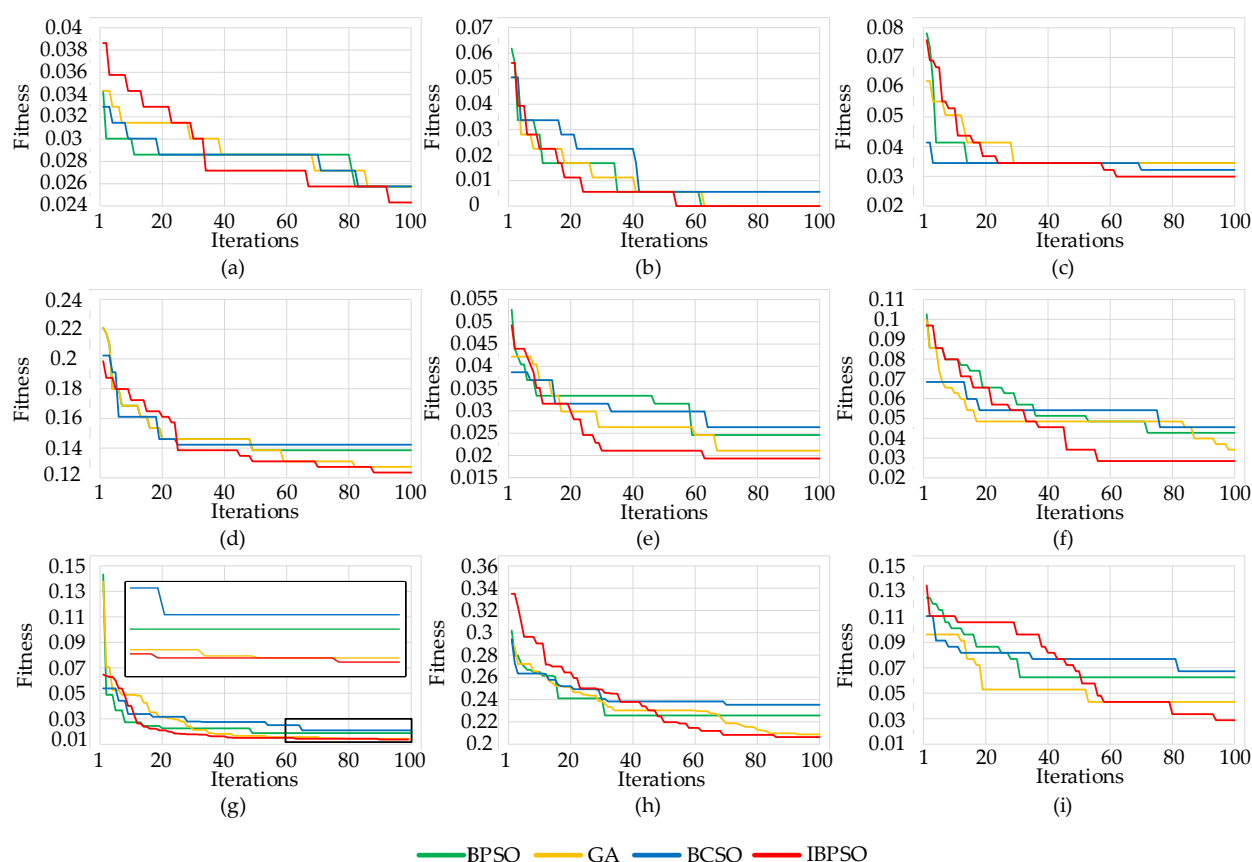

Figure 3. Comparison between IBPSO based on convergence curves of UCI benchmark datasets. (**a**) BreastCancer, (**b**) Wine, (**c**) CongressEW, (**d**) SpectEW, (**e**) BreastEW, (**f**) Ionosphere, (**g**) krvskp, (**h**) WaveformEW, (**i**) Sonar.

Table 3 shows the results of the average classification accuracy and the average number of selected features. IBPSO has the best average classification accuracy in all datasets. GA has the second-ranked average classification accuracy in all datasets. BPSO has competitive results of average classification accuracy with GA in datasets with smaller dimensions, such as BreastCancer, Wine, and CongressEW. This result shows that although BPSO has good exploitation capabilities, high-dimensional datasets such as WaveformEW and Sonar, GA with better exploration capabilities perform better than BPSO. In summary, the proposed hybrid algorithm IBPSO has good exploitation and exploration capabilities and achieves the best classification accuracy in this experiment.

In addition, because the results of the algorithm are different each run, this experiment uses the standard deviation to analyze the stability of the algorithm. Table 4 shows the standard deviation of the classification error of each algorithm. It can be seen from Table 4 that IBPSO has the best standard deviation of classification error in all datasets, showing that IBPSO has stronger stability than other algorithms. Table 5 shows the average calculation time of each algorithm. It can be clearly seen from Table 5 that GA is the worst computational cost algorithm. There is no significant difference in computational time between BPSO, BCSO and IBPSO. From this result, IBPSO can achieve higher classification accuracy without consuming extra computational costs.

In summary, case study 1 uses the public feature selection datasets to verify the superiority of IBPSO. The results in Tables 3 and 4 show that based on the average classification accuracy and stability, the proposed method is better than the three comparison algorithms (BPSO, GA, BCSO). This shows that IBPSO enhances its exploitation and exploration capa-

bilities during the iteration through its more efficient cycling time-varying inertia weights and crossover operators than traditional BPSO. Although IBPSO has no advantage in the average number of selected features, accurate classification of rolling element failures is the priority criterion of this study.

Table 3. Comparison between IBPSO based on classification accuracy and selected features in this experiment.

Datasets	BPSO		GA		BCSO		IBPSO	
	Avg Acc (%)	Avg No.F	Avg Acc (%)	Avg No.F	Avg Acc (%)	Avg No.F	Avg Acc (%)	Avg No.F
BreastCancer	97.19	5.83	97.24	6.06	97.1	**5.56**	**97.28**	6.13
Wine	98.72	**6.36**	98.82	6.63	98.2	6.46	**99.27**	6.63
CongressEW	96.16	3.93	96.13	4.2	95.95	4.03	**96.3**	**3.86**
SpectEW	84.49	**12.4**	85.39	12.53	83.79	12.76	**85.73**	12.56
BreastEW	96.89	**14.7**	97.38	15.96	96.81	16	**97.49**	15.13
Ionosphere	94.93	**14.63**	95.21	15.8	94.2	14.93	**95.43**	15.1
krvskp	97.79	20.73	98.29	20.96	97.19	**20.46**	**98.45**	20.5
WaveformEW	75.76	18.5	77.91	17.63	74.96	18.96	**78.16**	**16.36**
Sonar	91.95	**30.13**	93.57	31.2	91.49	31.9	**94.04**	32.7

Note: the algorithms that achieve better results are bold.

Table 4. Comparison between IBPSO based on standard deviation.

Datasets	BPSO	GA	BCSO	IBPSO
	Std	Std	Std	Std
BreastCancer	1.79×10^{-3}	1.38×10^{-3}	2.12×10^{-3}	$\mathbf{1.31 \times 10^{-3}}$
Wine	8.1×10^{-3}	1.08×10^{-2}	8.37×10^{-3}	$\mathbf{5.89 \times 10^{-3}}$
CongressEW	3.74×10^{-3}	4.14×10^{-3}	3.65×10^{-3}	$\mathbf{2.2 \times 10^{-3}}$
SpectEW	8.89×10^{-3}	9.38×10^{-3}	8.34×10^{-3}	$\mathbf{7.59 \times 10^{-3}}$
BreastEW	3.38×10^{-3}	3.61×10^{-3}	2.97×10^{-3}	$\mathbf{2.7 \times 10^{-3}}$
Ionosphere	7.86×10^{-3}	1.05×10^{-2}	6.36×10^{-3}	$\mathbf{5.33 \times 10^{-3}}$
krvskp	2.45×10^{-3}	1.81×10^{-3}	4.3×10^{-3}	$\mathbf{1.03 \times 10^{-3}}$
WaveformEW	7.56×10^{-3}	8.1×10^{-3}	9.36×10^{-3}	$\mathbf{6.7 \times 10^{-3}}$
Sonar	7.78×10^{-3}	9.74×10^{-3}	8.58×10^{-3}	$\mathbf{6.17 \times 10^{-3}}$

Note: the algorithm that achieves better results is bold.

Table 5. Comparison between IBPSO based on average computational time.

Datasets	BPSO	GA	BCSO	IBPSO
	Sec	Sec	Sec	Sec
BreastCancer	**108.45**	176.43	109.33	110.28
Wine	98.05	152.56	**97**	98.37
CongressEW	140.07	216.17	**138.33**	140.46
SpectEW	105.35	153.96	**98.51**	103.99
BreastEW	138.88	237.47	139.17	**138.08**
Ionosphere	140.7	217.08	**134.29**	138.78
krvskp	**200.22**	322.99	205.42	229.92
WaveformEW	235.62	352.55	237.4	**215.45**
Sonar	153.02	213.39	135.72	**132.75**

Note: the algorithms that achieve better results are bold.

4.1.3. Comparison with State-of-the-Art Models

Table 6 shows the classification results that reference the same UCI dataset to evaluate the effectiveness of the proposed algorithm in the field of feature selection. This study cites four state-of-the-art feature selection models, and their brief descriptions are as follows: the continuous symbiotic organism search algorithm uses adaptive S-shaped transfer function to convert into a binary symbiosis organism search algorithm, named BSOS [49]; the basic PSO introduces two dynamic correction coefficient and spiral-shaped mechanisms to improve the position update formula of PSO, and uses the logic diagram

sequence to enhance the diversity, named HPSO-SSM [50]; a binary butterfly optimization algorithm based on sigmoid transfer function can better converge to the optimal solution, named s-bBOA [51]; the grasshopper optimization algorithm combined with the mutation operator with linearly decreasing mutation rate enhances the exploration stage, named BGOA-M [52].

In Table 6, BSOS achieves the least number of selected features in BreastEW and Sonar, which is significantly reduced compared to other algorithms. This result is because the author emphasizes fewer selected features to achieve acceptable accuracy. HPSO-SSM achieves better performance in low-dimensional datasets, such as BreastCancer, Wine and CongressEW. Especially in BreastCancer and Wine, HPSO-SSM achieves the highest classification accuracy and the least number of selected features. This result is because the spiral-shaped position update mechanism introduced by HPSO-SSM can exploitation the search area more intensively, provide more solutions and maximize the exploitation capability of the PSO algorithm. However, in high-dimensional datasets, too much exploitation may lead to local optima. s-bBOA generally performs well in the selected datasets. This result is because the adaptive mechanism in s-bBOA plays a role in balancing exploration and exploitation. s-bBOA achieves the second highest classification accuracy among SpectEW and BreastEW. However, in datasets with many local solutions, such as CongressEW, Ionosphere, WaveformEW and Sonar, the classification accuracy of s-bBOA is not well and cannot escape the local solution well. BGOA-M has better classification accuracy in low-dimensional datasets, especially in CongressEW. BGOA-M achieves the highest classification accuracy. However, in higher-dimensional datasets, such as WaveformEW and Sonar, the classification accuracy of BGOA-M is poor. This result shows that the exploration capability of mutation operators with a linearly decreasing mutation rate is not enough to solve high-dimensional datasets. The proposed algorithm IBPSO performs best in classification accuracy, especially in medium and high-dimensional datasets including SpectEW, BreastEW, Ionosphere, krvskp, WaveformEW and Sonar, achieving the highest classification accuracy. In addition, to evaluate the stability of the algorithm, Table 6 also compares the standard deviation of the classification accuracy of the algorithm. Based on the standard deviation of classification accuracy, BGOA-M and IBPSO performed best, achieving the best or second-ranked standard deviation of the classification accuracy in all datasets.

Table 6. Comparison between IBPSO and state-of-the-art models.

Datasets	Algorithms	Avg Acc (%)	std	Avg No.F
BreastCancer	HPSO-SSM [50]	**98.03**	2.25×10^{-3}	**4**
	s-bBOA [51]	96.86	6×10^{-3}	5.6
	BGOA_M [52]	97.43	**0**	5
	IBPSO	97.28	1.31×10^{-3}	6.13
Wine	HPSO-SSM [50]	**99.38**	8.91×10^{-3}	**4.43**
	s-bBOA [51]	98.43	5.6×10^{-3}	6.2
	BGOA_M [52]	98.88	**0**	4.4
	IBPSO	99.27	5.89×10^{-3}	6.63
CongressEW	HPSO-SSM [50]	96.64	8.13×10^{-3}	**2.97**
	s-bBOA [51]	95.93	2×10^{-2}	6.4
	BGOA_M [52]	**97.64**	$\mathbf{1.6 \times 10^{-3}}$	5
	IBPSO	96.3	2.2×10^{-3}	3.87
SpectEW	HPSO-SSM [50]	79.92	2.61×10^{-2}	**8.43**
	s-bBOA [51]	84.63	1×10^{-2}	10.8
	BGOA_M [52]	82.61	7.6×10^{-3}	9.96
	IBPSO	**85.73**	$\mathbf{7.59 \times 10^{-3}}$	12.57

Table 6. *Cont.*

Datasets	Algorithms	Avg Acc (%)	std	Avg No.F
BreastEW	BSOS [49]	94.73	9×10^{-3}	**5**
	HPSO-SSM [50]	94.89	6.87×10^{-3}	6.76
	s-bBOA [51]	97.09	3×10^{-3}	16.8
	BGOA_M [52]	96.97	4×10^{-3}	12.5
	IBPSO	**97.49**	$\mathbf{2.7 \times 10^{-3}}$	15.13
Ionosphere	BSOS [49]	90	1.2×10^{-2}	8
	HPSO-SSM [50]	92.57	1.62×10^{-2}	**7.1**
	s-bBOA [51]	90.7	1×10^{-2}	16.2
	BGOA_M [52]	94.58	7.3×10^{-3}	11.46
	IBPSO	**95.43**	$\mathbf{5.33 \times 10^{-3}}$	15.1
krvskp	HPSO-SSM [50]	96.37	7.11×10^{-3}	18.27
	s-bBOA [51]	96.6	3×10^{-3}	**17.6**
	BGOA_M [52]	97.36	3×10^{-3}	17.73
	IBPSO	**98.45**	$\mathbf{1.03 \times 10^{-3}}$	20.5
WaveformEW	s-bBOA [51]	74.29	$\mathbf{1 \times 10^{-3}}$	25
	BGOA_M [52]	75.11	6.4×10^{-3}	20.9
	IBPSO	**78.17**	6.7×10^{-3}	**16.37**
Sonar	BSOS [49]	90.47	1.4×10^{-2}	**19**
	s-bBOA [51]	93.62	$\mathbf{1 \times 10^{-3}}$	32.8
	BGOA_M [52]	91.47	1.09×10^{-2}	26.8
	IBPSO	**94.04**	6.17×10^{-3}	32.7

Note: the algorithms that achieve better results are bold.

4.2. Experiment 2: CWRU Bearing Dataset

In this experiment, a bearing failure dataset was used to evaluate the robustness of the diagnostic model. In this experiment, fully connected neural network was used and hyperparameters were adjusted to set a total of five fully connected neural networks to select a robust fault diagnosis model. In the experiment, white Gaussian noise was also added to simulate various noise environments, making the experimental results more in line with the real working environment. Finally, we compared the robustness of the proposed bearing fault diagnosis model with the state-of-the-art model.

4.2.1. Experiment Setup and Parameter Setting

This experiment used a bearing failure dataset from Case Western Reserve University (CWRU) [48]. The source of the vibration signal is an accelerometer set at the driving end of the test motor, collected at a sampling frequency of 12 kHz. The test motor was tested under a load of 2 horsepower and the speed was 1750 rpm. Bearing defects were completed using electrical discharge machining. This experiment used 10 different rolling bearing faults, including normal bearings, inner ring faults (0.007-, 0.014- and 0.021-inch fault depth), outer ring faults (0.007-, 0.014- and 0.021-inch fault depth) and ball defects (0.007-, 0.014- and 0.021-inch fault depth). Table 7 shows the description of the dataset. Table 8 shows the detailed information of the fully connected neural network used in this experiment. The fully connected neural network set 5-fold cross-validation to effectively measure the generalization ability of the model to new data. A total of three kind of layer sizes of single fully connected layer neural networks were set up, and the layer sizes of the FCNN-D and FCNN-E fully connected neural networks were both 10. In addition, setting the penalty term lower can solve overfitting, and standardizing can converge data to optimize neural network performance.

Table 7. Description of CWRU bearing dataset used in this case study.

Fault Type	Fault Depth (Inches)	Dataset (2Hp) Category Label
Healthy	-	1
Inner	0.007	2
	0.014	3
	0.021	4
Outer	0.007	5
	0.014	6
	0.021	7
Ball	0.007	8
	0.014	9
	0.021	10

Table 8. Parameter setting of five classifiers.

Parameter	FCNN-A	FCNN-B	FCNN-C	FCNN-D	FCNN-E
k-fold cross-validation	5	5	5	5	5
Layer sizes	10	25	100	[10, 10]	[10, 10, 10]
Activation function	ReLU	ReLU	ReLU	ReLU	ReLU
Maximum number of training iterations	1000	1000	1000	1000	1000
Regularization penalty term	0	0	0	0	0
Standardize data	true	true	true	true	true

4.2.2. Comparison in Feature Selection Stage

In this study, not only the UCI feature selection datasets, but also the bearing failure dataset were used to evaluate the effectiveness of the proposed feature selection algorithm. In addition, in this experiment, BPSO, GA and BCSO were also used for comparison. Figure 4 shows the convergence curves of the four algorithms. It can be seen from Figure 4 that the proposed algorithm has excellent performance, converging to the best solution at the 29th iteration. In contrast, BPSO, GA and BCSO converged at the 59th, 64th and 77th iterations, respectively, and did not converge to the optimal solution. According to Table 9, the proposed algorithm achieves the best performance in the average fitness value, the average selected feature number and the standard deviation of the average fitness value. Especially in the result of average fitness value, the proposed algorithm has a significant improvement compared with the compared algorithms. Table 10 shows the detailed information of the best feature subsets obtained by each algorithm. According to Table 10, the proposed algorithm achieves the optimal feature subset with the least number of selected features. The above results show that the proposed algorithm has better exploitation and exploration capabilities, and better converges to the best solution.

Table 9. Comparison between IBPSO based on fitness value and number of selected features in this case study.

Dataset	BPSO			GA			BCSO			IBPSO		
	Avg Fitness	Avg No.F	Std	Avg Fitness	Avg No.F	Std	Avg Fitness	Avg No.F	Std	Avg Fitness	Avg No.F	Std
2 Hp	0.021	25.73	6.4×10^{-3}	0.0207	29.67	1.4×10^{-2}	0.0339	29.9	9.7×10^{-3}	**0.0067**	**25**	$\mathbf{6.2 \times 10^{-3}}$

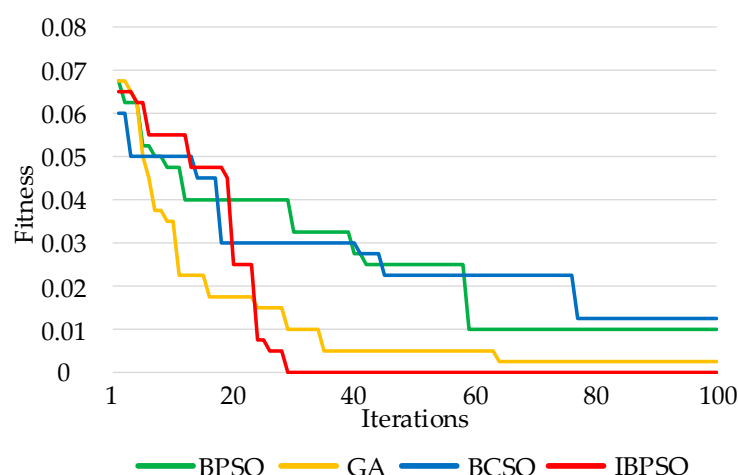Note: the algorithm that achieves better results is bold.

Figure 4. Comparison between IBPSO based on convergence curves of CWRU bearing dataset.

Table 10. Details of obtained optimal feature subset.

Dataset	Algorithm	No.F	Feature Indicators (F)
2 Hp	BPSO	26	2, 7, 8, 9, 10, 11, 12, 17, 18, 19, 20, 23, 24, 33, 37, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 56.
	GA	25	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 14, 16, 17, 18, 19, 21, 36, 37, 38, 40, 43, 45, 49, 52, 56.
	BCSO	21	1, 5, 6, 7, 9, 10, 12, 13, 15, 17, 18, 19, 20, 36, 37, 41, 43, 44, 48, 50, 56.
	IBPSO	16	1, 2, 3, 5, 7, 10, 14, 15, 16, 18, 19, 20, 33, 41, 45, 47.

4.2.3. Comparison in Classification Stage

In the classification stage, five fully connected neural networks were used to classify the optimal feature subset based on Table 10 and the original feature set obtained in the feature extraction stage. In addition, this experiment also evaluated the performance of the proposed fault diagnosis model in a noisy environment. In this experiment, Gaussian white noise was added to the original vibration signal to simulate the real environment. Equation (13) defines the signal to noise ratio (SNR).

$$SNR = 10 \log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \tag{15}$$

where P_{signal} and P_{noise} are the power of signal and the power of noise, respectively. In this experiment, the SNR values were set to 20, 15, 10, 5 and 0 dB. In order to make the experimental results fair, the average classification accuracy rate is obtained after 50 trainings. Tables 11–15 show the classification results of each fully connected neural network under different noise levels. It can be seen from each table that after feature selection eliminates redundant features, the performance of the fault diagnosis model can be further improved, and the feature subset obtained by the proposed feature selection algorithm has the best classification accuracy. Based on classification accuracy, the three types of single fully connected layer neural networks have better performance, while FCNN-D was fourth and FCNN-E has the worst performance. The FCNN-A is the worst among the three types of single fully connected layer neural networks. There are competing results between FCNN-B and FCNN-D. The classification accuracy under each noise level is very similar, but the FCNN-B is still slightly higher, except that when SNR = 0 dB, the two classifiers achieve the same classification accuracy, which is 96.56%. In summary, according to the above analysis, the FCNN-B is highly efficient in this experiment and has high robustness for bearing fault detection.

Table 11. Classification results using FCNN-A.

Dataset	Algorithm	No.F	Avg Acc (%)					
			∞ dB	20 dB	15 dB	10 dB	5 dB	0 dB
2 Hp	Without FS	64	94.44	93.87	93.65	92.52	90.35	85.22
	BPSO	26	95.88	95.51	95.2	94.82	93.08	89.83
	GA	25	97.01	96.89	96.59	96.41	95.64	93.21
	BCSO	21	96.55	96.5	96.36	96.01	95.22	92.03
	IBPSO	16	**97.4**	**97.24**	**97.18**	**96.93**	**96.29**	**96.08**

Note: the algorithm that achieves better results is bold.

Table 12. Classification results using FCNN-B.

Dataset	Algorithm	No.F	Avg Acc (%)					
			∞ dB	20 dB	15 dB	10 dB	5 dB	0 dB
2 Hp	Without FS	64	95.18	94.16	93.55	92.3	89.76	84.11
	BPSO	26	95.88	95.75	95.56	95.1	92.98	89.65
	GA	25	97.42	97.16	96.96	96.41	95.85	93.64
	BCSO	21	96.94	96.74	96.68	96.52	95.72	92.38
	IBPSO	16	**98.05**	**97.53**	**97.49**	**97.27**	**96.63**	**96.56**

Note: the algorithm that achieves better results is bold.

Table 13. Classification results using FCNN-C.

Dataset	Algorithm	No.F	Avg Acc (%)					
			∞ dB	20 dB	15 dB	10 dB	5 dB	0 dB
2 Hp	Without FS	64	94.3	93.88	93.5	92.71	89.5	82.22
	BPSO	26	96.18	96.11	95.07	94.98	92.68	89.54
	GA	25	97.3	97.13	96.85	96.72	96.03	93.5
	BCSO	21	96.89	96.75	96.59	96.49	95.48	91.78
	IBPSO	16	**97.97**	**97.43**	**97.27**	**97.24**	**96.64**	**96.56**

Note: the algorithm that achieves better results is bold.

Table 14. Classification results using FCNN-D.

Dataset	Algorithm	No.F	Avg Acc (%)					
			∞ dB	20 dB	15 dB	10 dB	5 dB	0 dB
2 Hp	Without FS	64	92.08	91.67	91.46	90.62	88.85	86.1
	BPSO	26	94.97	94.83	94.29	93.81	93.12	89.82
	GA	25	95.79	95.64	95.01	94.91	94.88	92.97
	BCSO	21	95.59	95.5	95.27	94.58	93.52	91.13
	IBPSO	16	**96.75**	**96.31**	**96.01**	**95.9**	**95.42**	**95.37**

Note: the algorithm that achieves better results is bold.

Table 15. Classification results using FCNN-E.

Dataset	Algorithm	No.F	Avg Acc (%)					
			∞ dB	20 dB	15 dB	10 dB	5 dB	0 dB
2 Hp	Without FS	64	91.61	91.56	91.02	90.36	88.85	85.07
	BPSO	26	94.65	93.73	93.65	92.8	90.71	88.38
	GA	25	94.95	94.42	94.3	93.74	93.14	91.63
	BCSO	21	94.16	94.04	93.78	92.97	91.57	88.55
	IBPSO	16	**95.17**	**95.13**	**94.62**	**94.09**	**93.5**	**92.58**

Note: the model that achieves better results is bold.

4.2.4. Comparison with State-of-the-Art Bearing Fault Diagnosis Models

Tables 16 and 17 show the comparison for the classification accuracy of the proposed fault diagnosis model and the most advanced bearing diagnosis model. Among them, Tables 16 and 17 are the classification accuracy rate (SNR = ∞ dB) under the normal environment and the classification accuracy rate under the noise environment, respectively. The diagnostic models in the table all use the same data set (CWRU bearing fault dataset) for fair comparison, and the classification results are all cited from their papers. Some brief descriptions of state-of-the-art fault diagnosis models are as follows: C. Grover and N. Turk [53] proposed a fault diagnosis model that combines Hjorth parameters to extract features from intrinsic mode functions (IMFs) and rule-based machine learning. M. Zhao et al. [54] proposed a new supervised dimensionality reduction method to extract features and improved the iterative trace ratio method to solve the trace ratio problem in linear discriminant analysis for fault classification. X. Zhanga and J. Zhoub [55] proposed a bearing fault diagnosis model combining ensemble empirical mode decomposition (EEMD) and optimized support vector machine by inter-cluster distance (ICD). X. Zhang et al. [41] proposed a hybrid fault diagnosis model, which uses permutation entropy (PE) to extract feature vectors from intrinsic mode functions and supports vector machines optimized by ICD for classification.

Table 16. Comparison with published in literature for bearing fault diagnosis models in normal condition.

Diagnosis Model	Dataset (2 Hp)
Ref. [53]	93.82
Ref. [54]	95.8
Ref. [55]	99.33
Ref. [41]	**100**
Proposed work	98.05

Note: the algorithm that achieves better results is bold.

Table 17. Comparison with published in literature for bear.ng fault diagnosis models in noisy condition.

Diagnosis Model	Avg Acc (%)									
	20 dB	15 dB	10 dB	8 dB	6 dB	5 dB	4 dB	3 dB	2 dB	0 dB
Ref. [56]	-	-	95.4	85	76.8	-	68.4	-	66.6	64.3
Ref. [57]	-	-	-	-	**99.61**	-	97.33	-	**95.71**	91.33
Ref. [58]	-	-	99.33	-	-	-	-	98.63	-	96.35
Ref. [59]	-	**99.64**	-	-	-	98.57	-	95	-	-
Proposed work	97.53	97.49	97.27	-	-	96.63	-	-	-	**96.56**

Note: the models that achieve better results are bold.

Some brief descriptions of state-of-the-art fault diagnosis models which are tested under the noise environment are as follows: Y. Zhang et al. [56] integrated 15 ensembles deep

shrinkage autoencoders (EDCAE) and combined these EDCAEs through a combination strategy. The proposed fault diagnosis method can effectively diagnose the fault even in the noise environment. H. Wenyi et al. [57] used filters of different scales to obtain the diversity resolution expression of the signal in the frequency domain, and enhanced the classification information of the input, and proposed an improved CNN called multi-scale cascaded convolutional neural network (MC-CNN). S. Ma et al. [58] proposed an end-to-end deep learning model based on wavelet packet transform. The experimental results show that this model is highly efficient and has excellent anti-noise ability. X. Yu et al. [59] proposed a combination of window marginal spectrum clustering (WMSC) and Hilbert–Huang transform (HHT) feature extraction technology and SVM, named HHT-WMSC-SVM.

The comparison results of classification accuracy under normal conditions are shown in Table 16. Ref. [41] achieves 100% classification accuracy, Ref. [55] achieves 99.33% classification accuracy and ranks second, and the proposed model achieves 98.05% classification accuracy and ranks third. The above results show that SVM has excellent classification performance under normal condition. However, Table 17 shows the classification results of each model in a strong noise environment, and the proposed model shows excellent anti-noise capability. When the SNR value is reduced from 20 dB to 0 dB, the classification accuracy of the proposed model is only reduced by 0.97%. In contrast, when the SNR value from 10 dB to 0 dB, the classification accuracy was reduced by 31.1% in Ref. [56]. When the SNR value is reduced from 6 dB to 0 dB, the classification accuracy is reduced by 8.28% in Ref. [57]. When the SNR value is reduced from 10 dB to 0 dB, the classification accuracy is reduced by 2.98% in Ref. [58]. When the SNR value is reduced from 15 dB to 0 dB, the classification accuracy is reduced by 4.64% in Ref. [59]. In addition, when the SNR value is 0 dB, the proposed model achieves a classification accuracy of 96.56%, which is better than the 96.35% classification accuracy of the deep learning model proposed by Ref. [57]. In summary, the above analysis results show that the proposed fault diagnosis model achieves high accuracy and excellent noise immunity.

5. Discussion

Based on the reason for the highest incidence of rolling element failures, this research proposes an efficient rolling element failure diagnosis model. To verify the effectiveness of the proposed model, two public data sets were used, namely, UCI feature selection dataset and CWRU bearing fault dataset for fair comparison with other state-of-the-art methods. Based on the experimental results in the fourth subsection, the main contributions of this research can be divided into the following two points.

(1) The proposed fault diagnosis model can be applied to a strong noise environment: Although local mean decomposition can effectively deal with non-stationary signals, it can extract the time-frequency domain information from the signal. However, this method is sensitive to noise and generates redundant PF components. Therefore, this study proposes a feature extraction technique that combines local mean decomposition with PF selection and WPD. Three fault features are introduced in PF selection, because these features perform well in early faults or severe faults, so the weight value is considered to balance the contribution of each fault feature. Benefitting from the wavelet packet denoising technology that maintains high resolution at both low and high frequencies, high-frequency noise is removed and fault information is further extracted.

In addition, feature selection technology can improve the performance of fault diagnosis models further, remove redundant features, improve classification accuracy and reduce computational costs. Based on the above criteria, the proposed feature selection algorithm IBPSO performs best among the comparison algorithms (BPSO, GA, BCSO). Table 12 shows the classification results of the best feature subsets obtained by each feature selection algorithm. It can be seen from Table 12 that IBPSO removes 75% of the redundant features and achieves the best classification accuracy under each noise level. Moreover, after increasing the noise level, there was a reduction in the classification accuracy of the diagnostic model. Based on the classification results of the original feature set, the

classification accuracy rate drops from 95.18% to 84.11% when the SNR value rises from ∞ dB to 0 dB. Based on the classification result of the best feature subset obtained by IBPSO, the classification accuracy rate drops from 98.05% to 96.56% when the SNR value rises from ∞ dB to 0 dB. Based on the above analysis results, IBPSO can select the most important features and significantly improve the performance of the fault diagnosis model.

(2) Determine the appropriate number and size of fully connected layers: In neural networks, the most difficult task is to determine the appropriate number of fully connected layers and the number of neurons. A deeper number of layers may cause overfitting problems, increase the difficulty of training and make it hard for the model to converge. Using too few neurons will result in underfitting. Conversely, using too many neurons will also lead to overfitting. When the amount of information contained in the training set is not enough to train all the neurons in the fully connected layer, it will lead to overfitting. Therefore, it is important to select the appropriate number of fully connected layers and neurons. In this study, five types of layer numbers and sizes are set, including FCNN-A (layer sizes: 10), FCNN-B (layer sizes: 25), FCNN-C (layer sizes: 100), FCNN-D (layer sizes: [10, 10]) and FCNN-E (layer sizes: [10, 10, 10]). Tables 11–15 show the classification results of each classifier. The experimental results from Tables 14 and 15 show that the classification accuracy of deeper layers is worse than that of single layer, and the model is difficult to converge. Therefore, the fault diagnosis model is more suitable for using a single fully connected layer neural network. Based on the experimental results from Tables 11–13, the performance of the FCNN-B is better than narrow neural network. This result shows that increasing the number of neurons can improve classification performance. However, increasing the number of too many neurons does not help improve the classification performance. In Tables 12 and 13, the classification results of FCNN-B and FCNN-C are almost the same. This result shows that too many neurons are a waste of computational cost and do not help to improve the classification performance.

In addition to the above advantages, the proposed model still has the following shortcomings.

(1) The types of features selected by the fault diagnosis model are highly dependent on the knowledge of the engineer, and the quality of the features determines the accuracy of the fault diagnosis model. This also affects the versatility of the fault diagnosis model. Engineers choose appropriate features for different types of faults based on prior knowledge. Therefore, automatic feature extraction technology should be considered in the future.

(2) The computational time of the algorithm. In Table 5, the computational time of the proposed algorithm performs poorly in low-dimensional or less local optimum datasets. However, in high-dimensional or more local optimum datasets, the proposed algorithm performs better than other algorithms. This result shows that although the crossover operator mechanism has powerful exploration capabilities, the low-dimensional or less local optimum datasets usually require strong exploitation capabilities to approach the global optimal. Therefore, it is necessary to further study the algorithm to reduce the computational time.

(3) Optimization of the computational complexity of the fault diagnosis model. This study only discusses the influence of the number of fully connected neural network layers and the number of neurons on the classification accuracy of fault diagnosis model but does not mention the computational complexity of the diagnosis model. Some methods may reduce the computational complexity of a fully connected neural network, such as sparsity [60] or improvement of neural network architecture and parameters.

6. Conclusions

This paper proposes an improved binary particle swarm optimization algorithm for the feature selection task of rolling element fault classification model, which is accurate and robust. The proposed model uses effective and anti-noise feature extraction technology. The process is to decompose the vibration signal by LMD and combine the noise reduc-

tion technology of PF selection and wavelet packet decomposition. The improved binary particle swarm optimization can remove redundant features and improve classification accuracy. The CWRU bearing fault dataset was used to evaluate the proposed diagnostic model. The IBPSO removes 75% of the redundant features from the original feature set (64 features) and selects the 16 most important features. By adjusting the hyperparameters, five fully connected neural networks are applied. The FCNN-B achieves the best classification results, reaching a classification accuracy of 98.05%. In addition, the proposed diagnostic model was compared for anti-noise capability, and the result is better than the state-of-the-art diagnostic model. When the SNR value is reduced from 20 dB to 0 dB, the classification accuracy of the proposed model is only reduced by 0.97%. However, as mentioned in the discussion section, the proposed model still has the following limitations: (1) the selection of feature types depends on prior knowledge, which affects the classification accuracy and generalization of the fault diagnosis model, and (2) computational complexity. Therefore, it is necessary to further study the automatic feature extraction technology and the improvement of neural network architecture and parameters.

Author Contributions: Methodology, C.-Y.L. and G.-L.Z.; visualization, C.-Y.L. and G.-L.Z.; software, C.-Y.L.; data curation, G.-L.Z.; writing—original draft preparation, G.-L.Z.; writing—review and editing, C.-Y.L. 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.

References

1. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal. Process.* **2020**, *138*, 106587. [CrossRef]
2. Van, M.; Kang, H.J. Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection. *Sci. Meas. Technol.* **2015**, *9*, 671–680. [CrossRef]
3. Alvarez-Gonzalez, F.; Griffo, A.; Wang, B. Permanent magnet synchronous machine stator windings fault detection by Hilbert–Huang transform. *J. Eng.* **2019**, *2019*, 3505–3509. [CrossRef]
4. Haroun, S.; Seghir, A.N.; Touati, S. Multiple features extraction and selection for detection and classification of stator winding faults. *IET Electr. Power Appl.* **2018**, *12*, 339–346. [CrossRef]
5. Kumar, S.; Mukherjee, D.; Guchhait, P.K.; Banerjee, R.; Srivastava, A.K.; Vishwakarma, D.N.; Saket, R.K. A comprehensive review of condition based prognostic maintenance (CBPM) for induction motor. *IEEE Access* **2019**, *7*, 90690–90704. [CrossRef]
6. Hamadache, M.; Lee, D.; Veluvolu, K.C. Rotor speed-based bearing fault diagnosis (RSB-BFD) under variable speed and constant load. *IEEE Trans. Ind. Electron.* **2015**, *62*, 6486–6495. [CrossRef]
7. Wang, Z.; Zhang, Q.; Xiong, J.; Xiao, M.; Sun, G.; He, J. Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests. *IEEE Sens. J.* **2017**, *17*, 5581–5588. [CrossRef]
8. Shao, Y.; Kang, R.; Liu, J. Rolling Bearing Fault Diagnosis Based on the Coherent Demodulation Model. *IEEE Access* **2020**, *8*, 207659–207671. [CrossRef]
9. Huo, Z.; Zhang, Y.; Francq, P.; Shu, L.; Huang, J. Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures. *IEEE Access* **2017**, *5*, 19442–19456. [CrossRef]
10. Wei, S.; Wang, D.; Wang, H.; Peng, Z. Time-Varying Envelope Filtering for Exhibiting Space Bearing Cage Fault Features. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3504313. [CrossRef]
11. Kang, M.; Islam, M.R.; Kim, J.; Kim, J.-M.; Pecht, M. A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3299–3310. [CrossRef]
12. Yiyuan, G.; Dejie, Y.; Haojiang, W. Fault diagnosis of rolling bearings using weighted horizontal visibility graph and graph Fourier transform. *Measurement* **2020**, *149*, 107036.
13. Gao, M.; Yu, G.; Wang, T. Impulsive gear fault diagnosis using adaptive morlet wavelet filter based on alpha-stable distribution and kurtogram. *IEEE Access* **2019**, *7*, 72283–72296. [CrossRef]
14. Ye, X.; Hu, Y.; Shen, J.; Feng, R.; Zhai, G. An improved empirical mode decomposition based on adaptive weighted rational quartic spline for rolling bearing fault diagnosis. *IEEE Access* **2020**, *8*, 123813–123827. [CrossRef]

15. Minhas, A.S.; Singh, G.; Singh, J.; Kankar, P.K.; Singh, S. A novel method to classify bearing faults by integrating standard deviation to refined composite multi-scale fuzzy entropy. *Measurement* **2020**, *154*, 107441. [CrossRef]
16. Smith, J.S. The local mean decomposition and its application to EEG perception data. *J. R. Soc. Inter.* **2005**, *2*, 443–454. [CrossRef]
17. Yu, J.; Lv, J. Weak fault feature extraction of rolling bearings using local mean decomposition-based multilayer hybrid denoising. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 3148–3159. [CrossRef]
18. McFadden, P. Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration. *J. Vib. Acoust. Stress Reliab. Des.* **1986**, *108*, 165–170. [CrossRef]
19. Radcliff, G.A. Condition Monitoring of Rolling Element Bearings Using the Enveloping Technique. In *IMechE Paper, Solid Mechanics and Machine Systems Group Seminar*; IMechE: London, UK, 1990; pp. 55–65.
20. Khanam, S.; Dutt, J.K.; Tandon, N. Extracting rolling element bearing faults from noisy vibration signal using Kalman filter. *J. Vib. Acoust. Trans. ASME* **2014**, *136*, 11. [CrossRef]
21. Wang, Y.; Liang, M. An adaptive SK technique and its application for fault detection of rolling element bearings. *Mech. Syst. Signal. Process.* **2011**, *25*, 1750–1764. [CrossRef]
22. Wang, Y.; Xu, G.; Liang, L.; Jiang, K. Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis. *Mech. Syst. Signal. Process.* **2015**, *54–55*, 259–276. [CrossRef]
23. Sun, J.; Xiao, Q.; Wen, J.; Wang, F. Natural gas pipeline small leakage feature extraction and recognition based on LMD envelope spectrum entropy and SVM. *Measurement* **2014**, *55*, 434–443. [CrossRef]
24. Souza, F.A.; Araújo, R.; Mendes, J. Review of soft sensor methods for regression applications. *Chemom. Intell. Lab. Syst.* **2016**, *152*, 69–79. [CrossRef]
25. Kennedy, J.; Eberhart, R.C. A Discrete Binary Version of the Particle Swarm Algorithm. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Orlando, FL, USA, 12–15 October 1997; pp. 4104–4108.
26. Holland, J.H. Genetic algorithms. *Sci. Am.* **1992**, *267*, 66–72. [CrossRef]
27. Hafez, A.I.; Zawbaa, H.M.; Emary, E.; Mahmoud, H.A.; Hassanien, A.E. An Innovative Approach for Feature Selection Based on Chicken Swarm Optimization. In Proceedings of the 2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR), Fukuoka, Japan, 13–15 November 2015; pp. 19–24.
28. Karim, A.A.; Isa, N.A.M.; Lim, W.H. Modified particle swarm optimization with effective guides. *IEEE Access* **2020**, *8*, 188699–188725. [CrossRef]
29. Liang, X.; Kou, D.; Wen, L. An improved chicken swarm optimization algorithm and its application in robot path planning. *IEEE Access* **2020**, *8*, 49543–49550. [CrossRef]
30. Ooi, C.S.; Lim, M.H.; Leong, M.S. Self-Tune Linear Adaptive-Genetic Algorithm for Feature Selection. *IEEE Access* **2019**, *7*, 138211–138232. [CrossRef]
31. Chang, W.D.; Shih, S.P. PID controller design of nonlinear systems using an improved particle swarm optimization approach. *Commun. Nonlinear Sci. Numer. Simul.* **2010**, *15*, 3632–3639. [CrossRef]
32. Cacciola, M.; Calcagno, S.; Morabito, F.C.; Versaci, M. Swarm optimization for imaging of corrosion by impedance measurements in Eddy current test. *IEEE Trans. Magn.* **2007**, *43*, 1853–1856. [CrossRef]
33. Xue, B.; Zhang, M.; Browne, W.N.; Yao, X. A Survey on Evolutionary Computation Approaches to Feature Selection. *IEEE Trans. Evol. Comput.* **2016**, *20*, 606–626. [CrossRef]
34. Lee, J.H.; Kim, J.W.; Song, J.Y.; Kim, Y.J.; Jung, S.Y. A novel memetic algorithm using modified particle swarm optimization and mesh adaptive DIRECT search for PMSM design. *IEEE Trans. Magn.* **2016**, *52*, 7001604. [CrossRef]
35. Lee, C.-Y.; Le, T.-A. Intelligence bearing fault diagnosis model using multiple feature extraction and binary particle swarm optimization with extended memory. *IEEE Access* **2020**, *8*, 198343–198356. [CrossRef]
36. Long, J.; Zhang, S.; Li, C. Evolving deep echo state networks for intelligent fault diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 4928–4937. [CrossRef]
37. Lee, C.-Y.; Le, T.-A. An Enhanced Binary Particle Swarm Optimization for Optimal Feature Selection in Bearing Fault Diagnosis of Electrical Machines. *IEEE Access* **2021**, *9*, 102671–102686. [CrossRef]
38. Deshpande, G.; Wang, P.; Rangaprakash, D.; Wilamowski, B. Fully Connected Cascade Artificial Neural Network Architecture for Attention Deficit Hyperactivity Disorder Classification from Functional Magnetic Resonance Imaging Data. *IEEE Trans. Cybern.* **2015**, *45*, 2668–2679. [CrossRef] [PubMed]
39. Zhang, W.; Li, C.H.; Peng, G.L.; Chen, Y.H.; Zhang, Z.J. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal. Process.* **2018**, *100*, 439–453. [CrossRef]
40. Lee, C.-Y.; Zhuo, G.-L. Effective Rotor Fault Diagnosis Model Using Multilayer Signal Analysis and Hybrid Genetic Binary Chicken Swarm Optimization. *Symmetry* **2021**, *13*, 487. [CrossRef]
41. Zhang, X.; Liang, Y.; Zhou, J.; Zang, Y. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement* **2015**, *69*, 164–179. [CrossRef]
42. Hussain, S.; Mokhtar, M.; Howe, J.M. Sensor Failure Detection, Identification, and Accommodation Using Fully Connected Cascade Neural Network. *IEEE Trans. Ind. Electron.* **2014**, *62*, 1683–1692. [CrossRef]
43. Meng, L.; Xiang, J.; Wang, Y.; Jiang, Y.; Gao, H. A hybrid fault diagnosis method using morphological filter–translation invariant wavelet and improved ensemble empirical mode decomposition. *Mech. Syst. Signal. Process.* **2015**, *50–51*, 101–115. [CrossRef]

44. Lei, Y.; Li, N.; Lin, J. A new method based on stochastic process models for machine remaining useful life prediction. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 2671–2684. [CrossRef]

45. Xu, F.; Song, X.; Tsui, K.; Yang, F.; Huang, Z. Bearing performance degradation assessment based on ensemble empirical mode decomposition and affinity propagation clustering. *IEEE Access* **2019**, *7*, 54623–54637. [CrossRef]

46. Askari, Q.; Saeed, M.; Younas, I. Heap-based optimizer inspired by corporate rank hierarchy for global optimization. *Expert Syst. Appl.* **2020**, *161*, 113702. [CrossRef]

47. UCI Machine Learning Repository. Available online: http://archive.ics.uci.edu/ml (accessed on 22 May 2021).

48. Case Western Reserve University Bearing Data Center. Available online: http://csegroups.case.edu/bearingdatacenter/home (accessed on 12 June 2021).

49. Han, C.; Zhou, G.; Zhou, Y. Binary Symbiotic Organism Search Algorithm for Feature Selection and Analysis. *IEEE Access* **2019**, *7*, 166833–166859. [CrossRef]

50. Chen, K.; Zhou, F.; Yuan, X. Hybrid particle swarm optimization with spiral-shaped mechanism for feature selection. *Expert Syst. Appl.* **2019**, *128*, 140–156. [CrossRef]

51. Arora, S.; Anand, P. Binary butterfly optimization approaches for feature selection. *Expert Syst. Appl.* **2019**, *116*, 147–160. [CrossRef]

52. Mafarja, M.; Aljarah, I.; Faris, H.; Hammouri, A.I.; Al-Zoubi, A.M.; Mirjalili, S. Binary grasshopper optimisation algorithm approaches for feature selection problems. *Expert Syst. Appl.* **2019**, *117*, 267–286. [CrossRef]

53. Grover, C.; Turk, N. Rolling element bearing fault diagnosis using empirical mode decomposition and hjorth parameters. *Procedia Comput. Sci.* **2020**, *167*, 1484–1494. [CrossRef]

54. Zhao, M.; Jin, X.; Zhang, Z.; Li, B. Fault diagnosis of rolling element bearings via discriminative subspace learning: Visualization and classification. *Expert Syst. Appl.* **2014**, *41*, 3391–3401. [CrossRef]

55. Zhang, X.; Zhou, J. Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines. *Mech. Syst. Signal. Process.* **2013**, *41*, 127–140. [CrossRef]

56. Zhang, Y.; Li, X.; Gao, L.; Chen, W.; Li, P. Ensemble deep contractive auto-encoders for intelligent fault diagnosis of machines under noisy environment. *Knowl. Based Syst.* **2020**, *48*, 34–50. [CrossRef]

57. Huang, W.; Cheng, J.; Yang, Y.; Guo, G. An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis. *Neurocomputing* **2019**, *359*, 77–92. [CrossRef]

58. Ma, S.; Cai, W.; Liu, W.; Shang, Z.; Liu, G. A Lighted Deep Convolutional Neural Network Based Fault Diagnosis of Rotating Machinery. *Sensors* **2019**, *19*, 2381. [CrossRef]

59. Yu, X.; Ding, E.; Chen, C.; Liu, X.; Li, L. A Novel Characteristic Frequency Bands Extraction Method for Automatic Bearing Fault Diagnosis Based on Hilbert Huang Transform. *Sensors* **2015**, *15*, 27869–27893. [CrossRef] [PubMed]

60. Wang, Y.K.; Zhang, F.; Zhang, S.W. A new methodology for identifying arc fault by sparse representation and neural network. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 2526–2537. [CrossRef]