Research on a Bearing Fault Enhancement Diagnosis Method with Convolutional Neural Network Based on Adaptive Stochastic Resonance

Chen Wang 1,2, Zijian Qiao 3, Zhangjun Huang 1,2, Junchen Xu 1,2, Shitong Fang 1,2, Cailiang Zhang 4, Jinjun Liu 5, Ronghua Zhu 4,* and Zhihui Lai 1,2,*

1 Shenzhen Key Laboratory of High Performance Nontraditional Manufacturing, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China
2 Guangdong Key Laboratory of Electromagnetic Control and Intelligent Robots, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China
3 School of Mechanical Engineering and Mechanics, Ningbo University, Ningbo 315211, China
4 Ocean College, Zhejiang University, Zhoushan 316021, China
5 School of Control and Mechanical Engineering, Tianjin Chengjia University, Tianjin 300384, China
* Correspondence: zhu.richard@zju.edu.cn (R.Z.); laizh@szu.edu.cn (Z.L.)

Abstract: As a powerful feature extraction tool, a convolutional neural network (CNN) has strong adaptability for big data applications such as bearing fault diagnosis, whereas the classification performance is limited when the quality of raw signals is poor. In this paper, stochastic resonance (SR), which provides an advanced feature enhancement approach for weak signals with strong background noise, is introduced as a data pre-processing method for the CNN to improve its classification performance. First, a multiparameter adjusting bistable Duffing system that can achieve SR under large-parameter weak signals is introduced. A hybrid optimization algorithm (HOA) combining the genetic algorithm (GA) and the simulated annealing (SA) is proposed to adaptively obtain the optimized parameters and output signal-to-noise ratio (SNR) of the Duffing system. Therefore, the data optimization based on the multiparameter-adjusting SR of Duffing system can be realized. An SR-based mapping method is further proposed to convert the outputs of the Duffing system into grey images, which can be further processed by a normal CNN with batch normalization (BN) layers and dropout layers. After verifying the feasibility of the HOA in multiparameter optimization of the Duffing system, the bearing fault data set from the CWRU bearing data center was processed by the proposed fault enhancement classification and identification method. The research showed that the weak features of the bearing signals could be enhanced significantly through the adaptive multiparameter optimization of SR, and classification accuracies for 10 categories of bearing signals could achieve 100% and those for 20 categories could achieve more than 96.9%, which is better than other methods. The influences of the population number on the classification accuracies and calculation time were further studied, and the feature map and network visualization are presented. It was demonstrated that the proposed method can realize high-performance fault enhancement classification and identification.

Keywords: bearing fault diagnosis; stochastic resonance; optimization algorithm; convolutional neural network

1. Introduction

Modern mechanical equipment develops towards complexity and automation; hence, minor malfunctions of parts may bring serious chain reactions. Rolling bearings, as a type of important component of rotating machinery, are widely used in wind turbines, compressors, high-speed railways, and other modern mechanical equipment. Bearing failure may cause the failure of the whole equipment, resulting in significant economic loss.
and even casualties. Therefore, it is necessary to monitor the working condition of rolling bearings and diagnose bearing faults in time to prevent potential accidents, which can be realized by intelligent fault diagnosis methods widely studied in recent years [1,2].

The framework for an intelligent fault diagnosis method can be divided into three parts [3]: signal acquisition, fault feature extraction, and fault classification. The objects of the signal acquisition include many kinds of signals, such as vibration signals [4], acoustic signals [5] and electrical signals [6], among which the vibration signals are most focused as they contain rich essential information of mechanical faults. Common signal processing methods for feature extraction features include short Fourier transform (STFT) [7], wavelet transform (WT) [8], empirical mode decomposition (EMD) [9], principal component analysis (PCA) [10] and stochastic resonance (SR) [11,12], etc. Fault classification is used to determine the health conditions of mechanical equipment. Common fault classification methods include deep learning, support vector machine (SVM) [13], and the K-nearest neighbor (KNN) algorithm [14]. The selections of the fault feature extraction method and the fault classification method are key to establishing a vibration-signal-based fault diagnosis method.

In practical applications, the fault features in the vibration signals are always weak when the fault is in its early stage or the faulty part operates in a terrible working environment. The weak features are difficult to extract through a normal feature extraction method, thus limiting the weak fault diagnosis performance of the corresponding intelligent fault diagnosis methods. Therefore, a feature extraction method with the capability of weak-signal enhancement should be selected in these cases. The stochastic resonance (SR)-based methods are a type of weak feature extraction method that can obtain a higher signal-to-noise ratio (SNR) compared to other traditional feature extraction methods [15]. Stochastic resonance is a type of commonly seen nonlinear phenomenon, which was first proposed by Benzi in the 1980s to describe the periodicity associated with the ice age of the earth in climatology [16]; it occurs under the cooperation of input weak signal, noise and a nonlinear system, which enables the weak signal to increase its amplitude by absorbing energy from noise. Hence, this interesting phenomenon is widely applied in physics and engineering research [17], such as energy harvesting [18] and weak-signal detection in bearing fault diagnosis [19,20]. Previous research shows that the weak-signal enhancement performance of the SR system is significantly influenced by the nonlinear system, which involves monostable systems [21,22], bistable systems [23–25], tri-stable systems [26] and multi-stable systems [27]. Among these nonlinear systems, bistable systems such as the Duffing system are widely studied due to their advanced weak-signal detection performance [28].

However, SR systems need to meet the small-parameter conditions [29] (the frequency and amplitude of weak-signal should be within small-parameter ranges) due to the adiabatic approximation theory, which cannot be satisfied for most practical vibration signals. To solve this issue, a multiparameter adjusting SR system has been proposed [30] by introducing an amplitude transformation coefficient and a frequency transformation coefficient, which are used to transform the amplitude and frequency of the signals to appropriate small-parameter ranges. Therefore, SR for larger-parameter signals can be achieved. In order to realize the SR output of the multiparameter adjusting system, the optimal parameters should be obtained, which can be realized by adaptive multiparameter optimization methods [31], such as the particle swarm optimization (PSO) [32], genetic algorithm (GA) [33] and simulated annealing (SA) algorithm [34], etc.; however, these algorithms all have their limitations in achieving optimal SR outputs. For example, the GA has a powerful global search capability but it is easy to trap in the local optimum [35], while SA can escape from the local optimum but has a slow optimization speed [36]. Therefore, the combination of some of these algorithms can improve their optimization performance, which has been studied in some literature [37].

As for the fault classification method, deep learning technology has achieved great success in many fields in recent years due to its powerful feature extraction and classification ca-
The deep learning technologies include deep neural network (DNN) [40], recurrent neural network (RNN) [41] and convolutional neural network (CNN) [42], etc. They can extract features from input automatically by building multiple neural layers and make predictions accordingly, which can also be used in many application fields including the fault diagnosis of mechanical equipment. In the past years, deep learning has been used as an advanced classification tool, which can effectively classify the signals obtained from traditional feature extraction methods [43]. In this paper, the CNN, which is one of the main types of deep learning technology and has been applied widely in fault classification, was selected as the fault classification method. However, it was noted that its classification accuracy was not high enough when the quality of the raw data was poor [44], such as the weak fault signals. Therefore, it is necessary to pre-process the raw data using a weak-fault feature extraction method to extract useful features, which can be further classified by the CNN.

In this work, a fault enhancement classification method combining the adaptive SR, which utilizes a hybrid optimization method combining the SA and GA as the weak-fault feature extraction method and a normal CNN as the fault classification method, is proposed for high-performance bearing weak-fault diagnosis. This paper is organized as follows. The multiparameter adjusting bistable Duffing system, the hybrid optimization method and a mapping method are introduced in Section 2. In Section 3, the normal CNN is presented by appropriately selecting its parameters. A series of simulations and experiments are conducted in Section 4 to verify the proposed signal pre-processing and fault diagnosis methods. Conclusions are drawn in Section 5.

2. Data Preprocessing Based on Adaptive Multiparameter Optimization of SR

In this section, the classical bistable Duffing system we investigated previously is introduced as a data preprocessing model for the further classification algorithm. A hybrid intelligent optimization algorithm combining genetic algorithm (GA) and simulated annealing (SA) was used to achieve stochastic resonance (SR) in this system and obtain the optimal parameters. To make the SR outputs able to be processed by the CNN for classification, a mapping method based on a noise intensity sequence is further proposed to convert the time-domain output of the Duffing system into an image that can be further processed in classification.

2.1. Introduction of the Bistable Duffing System That Can Achieve SR

The bistable Duffing system, which is a typical form of nonlinear system that can achieve SR, can be described as [20]:

$$\ddot{x}(t) + k\dot{x}(t) - ax(t) + bx^3(t) = s(t) + n(t)$$  \hspace{1cm} (1)

where $k$ denotes the damping ratio; $a$ and $b$ are system parameters deciding the potential function of the system; $s(t) = A \cos(2\pi f_0 t)$ indicates a harmonic characteristic signal with amplitude $A$ and frequency $f_0$; $n(t) = \sqrt{2D}\xi(t)$ indicates a Gaussian white noise with noise intensity $D$, where $\xi(t)$ is a zero-mean and unit-variance Gaussian white noise. In this system, $sn(t) = s(t) + n(t)$ is defined as the input signal, and $x(t)$ is the output signal, which can be obtained by solving Equation (1) numerically.

In Equation (1), the term of $dU(x)/dx = -ax + bx^3$ can be understood as the tangential force of a bistable potential field given by $U(x) = -ax^2/2 + bx^4/4$, which has two stable equilibrium points at $x = \pm \sqrt{a/b}$ and one unstable equilibrium point at $x = 0$, as shown in Figure 1. This shows that a potential barrier with a height of $\Delta U = a^2/(4b)$ separates two symmetrical potential wells, showing why the Duffing system is bistable. Moreover, the output $x(t)$ of Equation (1) can be understood as the trajectory of a unit-mass Brownian particle moving in the potential field $U(x)$, which is suffered from the damping force $-k\dot{x}(t)$ and the external excitation $sn(t)$ as well. Stochastic resonance indicates an optimal matching result between the signal, noise, and a nonlinear system. When SR occurs, the particle can get energy from noise and cross the barrier regularly even though the amplitude...
of the signal is relatively low, thus enhancing the weak features of the weak signal. Hence, SR provides a feasible way to extract the features of the input signal from the enhanced output signal, especially under weak-signal conditions.

![Bistable potential function without input.](image)

Figure 1. Bistable potential function without input.

Previous research results show that due to the adiabatic approximation theory, the bistable Duffing system can only achieve SR under small-parameter conditions, i.e., the amplitude, frequency and noise intensity should be small [30]; however, most of the practical engineering signals do not satisfy this small-parameter requirement. Hence, to enhance the signal features of such large-parameter signals, an improved multiparameter-adjusting SR model based on the bistable Duffing system was proposed by the authors previously [45]. By introducing two adjusted parameters $\varepsilon$ and $R$, this model can be written as:

$$\ddot{x}(t') + kx(t') - ax(t') + bx(t')^3 = \varepsilon \cdot sn(t')$$

where $\varepsilon$ is the amplitude transform coefficient used to transform the amplitude of the input signal to an appropriate range, and $R$ is the scale-transformation coefficient used to transform the time scale of the input signal from $t$ to $t' = R t$. The scale transformation can be simply realized by applying a time step of $\Delta t' = R / f_s$ instead of $\Delta t = 1 / f_s$ in numerical calculation, where $f_s$ denotes the sampling frequency of the system. Therefore, the frequency of the characteristic signal ($f_0$) can be regarded as $f_0 / R$ in the numerical calculation. The large frequency of the input signal can be compressed accordingly in the numerical calculation by setting an appropriate value of $R$.

The output signal-to-noise ratio (SNR) of the Duffing system can be regarded as an objective optimization function to decide whether the system achieves SR. The output SNR of the system is defined as:

$$\begin{align*}
\text{SNR}_{\text{in}} &= 10 \log_{10} \left( \frac{\text{Am}^2_{\text{in}}}{\sum \left| \text{SN}(f) \right|^2 - \text{Am}^2_{\text{in}}} \right) \\
\text{SNR}_{\text{out}} &= 10 \log_{10} \left( \frac{\text{Am}^2_{\text{out}}}{\sum \left| \text{X}(f) \right|^2 - \text{Am}^2_{\text{out}}} \right)
\end{align*}$$

where $\text{SN}(f)$ represents the single-side spectrum of input $sn(t)$, and $X(f)$ represents the single-side spectrum of the system output $x(t)$. Moreover, $\text{Am}_{\text{in}} = \text{SN}(f_0)$ and $\text{Am}_{\text{out}} = X(f_0)$ indicates the amplitudes of the system input signal and output signal at the characteristic frequency $f = f_0$.

2.2. Hybrid Optimization Algorithm Combining GA and SA

To achieve SR adaptively in a bistable Duffing system under an input signal with fixed parameters, an optimization algorithm is needed to obtain a group of appropriate system parameters $(a, b)$, damping ratio $(k)$ and adjusted parameters ($\varepsilon, R$) that match the fixed input signal. Among various optimization algorithms, the genetic algorithm (GA) is an effective intelligent optimization algorithm when the objective optimization function is not
differentiable, and it can obtain a local optimization value greater than 90% of the global optimization one in a short time [35]. In the GA, every individual represents a solution, and its principle is to obtain the optimal population by selecting the parents to do cross and mutation according to the fitness function.

However, the optimal results obtained from the GA are local optimization solutions, which could become better for the objective optimization function (such as the output SNR of the Duffing system). This can be obtained by adopting the simulated annealing (SA), which is another important optimization algorithm proposed by Metropolis et al. in 1953 based on the solid annealing process in physics. The SA can accept a solution worse than the current one with a certain probability, resulting in a capacity of jumping out of the local optimal solution and reaching the global optimal solution. The probability of accepting a new solution in the Metropolis criterion in this paper is defined as:

$$ P = \begin{cases} 1, & E_{t+1} > E_t \\ \frac{1}{e^{\frac{E_t - E_{t+1}}{kT}}}, & E_{t+1} \leq E_t \end{cases} $$

(4)

where $E_{t+1}$ and $E_t$ are the new condition and temporary optimal condition, respectively, where $T$ is the current updated temperature, and $k$ is a Boltzmann constant set as 1 in this work [34]. The new condition is undoubtedly accepted as the updated temporary optimal condition when $E_{t+1} > E_t$ while when $E_{t+1} \leq E_t$, the new condition can be also accepted if the acceptance probability $P$ is greater than a random number between $[0, 1]$, thus finally obtaining a satisfactory optimization result. The main disadvantage of the SA is its slow optimization speed, which is not satisfied in practical conditions such as the adaptive multiparameter optimization of SR for big data.

Therefore, a hybrid optimization algorithm (HOA) based on both the GA and the SA combining their advantages was utilized in this work for high-performance multiparameter optimization of SR. In the HOA, the Metropolis criterion of the SA was added in the parents’ selection of cross and mutate stages based on the GA framework. Hence, the HOA can get a better optimization solution through the capacity of the SA that jumps out of local optimization in a short time. As a result, we can get an acceptable SNR for the Duffing system in a short time using the HOA.

2.3. Data Optimization Based on Multiparameter-Adjusting SR of Duffing System

According to previous analyses, the proposed HOA provides an effective approach to achieving multiparameter-adjusting SR in a Duffing system, thus improving the quality of the input raw signal, and enhancing its SNR. The relevant data optimization method is presented in this subsection.

It is noted that the fourth-order Runge–Kutta algorithm is adopted in this work to solve the Duffing system. The HOA used in this paper adopts a binary encoding format, and each parameter consists of 15-bit binary numbers to guarantee sufficient resolution. Moreover, the value of $R$ is pre-set as an appropriate value to ensure that the calculation results will not overflow. Hence, the optimization parameter dimension of the Duffing system is 4 ($k$, $a$, $b$, and $\varepsilon$) in the optimization, and the flowchart of the optimization process to achieve multiparameter-adjusting SR of the Duffing system using the hybrid optimization algorithm is shown in Figure 2.

In order to use roulette to select the crossed parents and mutation parents of the HOA (see Figure 2), a fitness function $F$ is defined:

$$ F = E_{j}^{t} - E_{j} + \delta $$

(5)

where $j = 1, 2, 3, \ldots, L$, and $L$ is the maximum number of iterations; $E_j$ denotes the minimum value of the SNR$_{out}$ in the whole population in the $j$th iteration, and $E_{j}^{t}$ denotes the value of the SNR$_{out}$ of the $i$th individual in the $j$th iteration. It is noted that $\delta$ with a quite small value is used in Equation (5) to avoid the fitness function equaling to 0.
(\(\delta = 0.001\) is set in this work), thus, the optimization parameters in terms of minimum \(\text{SNR}_{\text{out}}\) can be abandoned. Hence, the value of the fitness function \(F\) is always positive. Other parameters in the HOA were set as follows: population number \(PN = 50\), cross probability \(P_c = 0.9\), mutation probability \(P_m = 0.9\), initial temperature \(T = 10\), minimum temperature \(T_{\text{min}} = 0.001\) and update weight \(\Delta = 0.9\). It is necessary to mention that \(X_1\) and \(X_2\) represent the random numbers in the cross and mutate stages of GA.

![Flowchart](image)

**Figure 2.** Flowchart of the optimization process to achieve multiparameter-adjusting SR of the Duffing system using the hybrid optimization algorithm.

### 2.4. SR-Based Mapping Method with a Noise Intensity Sequence

When using the SR-based methods, engineers are required to have extensive experience, hence, it takes a lot of time and manpower to find out the characteristic frequency of the bearing fault vibration signal. This can be solved by using an intelligent diagnosis
method proposed in this work that combines SR with a neural network classifier. For this purpose, the SR output of the Duffing system should be converted into a grey image, which can be further used for feature extraction and fault classification by the neural network. The process of the mapping method is as follows.

First, \( M \) continuous time domain points are intercepted from a raw signal to form a new signal \( s_1(t) \), which is always a noisy signal. To produce more feature information in one image, a sequence of noise \( n_k(t) \), whose noise densities are given by

\[
D_k = 0.04 \times k
\]

is further added to the signal \( s_1(t) \). Therefore, a sequence of input signals can be obtained:

\[
s_{1k}(t) = s_1(t) + n_k(t)
\]

where \( k = 1, 2, 3, \ldots, m \), where \( m \) is the pre-set number of the input signals.

Next, by inputting \( s_{1k}(t) \) into the Duffing system of Equation (2), an optimal parameter sets \( P_k = [k, a_k, b_k, \varepsilon_k] \) and the output signal \( x_k(t) \) can be obtained based on the proposed data optimization method. Therefore, the matrix of the output signals \( G = [x_1(t), x_2(t), \ldots, x_m(t)] \) can be further converted into a visual grey matrix \( G_{\text{gray}} \) according to:

\[
G_{\text{gray}}(i, j) = \frac{G(i, j) - g_{\text{min}}}{g_{\text{max}} - g_{\text{min}}}
\]

where \( i = 1, 2, \ldots, M \) and \( j = 1, 2, \ldots, m \); \( g_{\text{max}} \) and \( g_{\text{min}} \) are the maximum and minimum values of \( G(i, j) \). It is noted that \( M = 512 \) and \( m = 128 \) are pre-set in this work.

Hence, for each detected signal, a grey image can be obtained by adding a group of noises with different intensities into the input signal and then being processed by the adaptive multiparameter adjusting Duffing system. More detected signals can produce more grey images, which can be further used by the convolutional neural network (CNN) for feature extraction and fault diagnosis.

3. Construction of the CNN

In this section, basic knowledge of a CNN is briefly introduced. The network architecture used for fault classification was obtained by modifying the parameters of the conventional visual geometry group (VGG) net architecture to satisfy the resolution of grey images obtained by the proposed mapping method. It is noted that compared to the conventional VGG, the batch normalization (BN) modules are added in the convolution layer and the dropout modules are added in the full connected (FC) layer to enhance the generalization of the VGG in this work.

3.1. Brief Introduction of CNN

The architecture of CNN is briefly introduced in this subsection. A CNN consists of some filter stage and one classification stage [46]. The filter stage contains convolutional layers, activation layers, BN layers, and pooling layers.

The convolutional layer convolves the input local region with kernel filters, and the following activation layer generates a feature map. The kernel that extracts the local features keeps the same in each filter, thus reducing the complication of CNN. The convolution process is described as follows:

\[
y_{l+1}^j = x_l^j \ast K_{l+1}^j + b_{l+1}^j
\]

where \( \ast \) is a convolutional operator; \( K_{l+1}^j \) and \( b_{l+1}^j \) represent the weight and bias of the \( j \)th kernel filter from layer \( l \) to layer \( l + 1 \); \( x_l^j \) denotes the \( j \)th local region of layer \( l \), \( y_{l+1}^j \) denotes the output of layer \( l \) calculated by convolution. Moreover, the padding method is used in convolution to make full use of all the features of the grey images.
In the activation layer, a nonlinear activation function of Rectified Linear Unit (ReLU) is widely used to improve the expression ability of the whole network, which means the functions that can be expressed are more abundant. The ReLU can prevent the occurrence of overfitting by making the output of some neurons to be zero, thus resulting in the sparsity of networks.

A BN layer is further designed to speed up training and convergence of the network and reduce the shift of internal covariance. The pooling layer generally adopts the max-pooling layer, which enhances the generalization of the model by reducing the parameters while retaining the main features.

Moreover, the classification stage is composed of several FC layers. The FC layer is used for enhancing the generation of the model after convolution, and the number of neurons in the output layer denotes the types of bearing health conditions.

3.2. Architecture of the Proposed CNN Model

The whole architecture of the CNN used in this work is shown in Figure 3, which includes convolutional layers, ReLU layers, BN layers, max-pooling layers, and FC layers. The number of convolutional layers depends on the size of the grey image produced by the proposed mapping method. Small convolutional kernels make the networks deeper, which helps to improve the generalization ability of the network, and the size of the convolutional kernel is set as 5 × 5 accordingly. The BN is implemented after the convolutional layers to accelerate the training process, and the ReLU is utilized in the next layer to prevent the occurrence of overfitting. Max-pooling is used to reduce the parameters of the networks, and the kernel size is set as 2 × 2. The classification stage includes three FC layers for classification, and the output layer has ten outputs, which represent ten different bearing health conditions.

Figure 3. The architecture of the CNN model used in this work.
In the process of training, the number of iterations was set as 300, and an Adam optimizer was utilized to minimize the loss function, where the learning rate was set as $Lr = 0.001$ initially. After every 100 iterations, the learning rate reduced 10 times to get more accurate optimal solutions.

4. Verification of the Proposed Method

The proposed method, which can be used for fault classification and fault identification in practical applications, is verified in this section. It is necessary to point out that the computer used for numerical simulations had a CPU of Intel(R) Core (TM) i5-10400 and RAM with 16.00 GB as its main configurations.

For practical noisy fault signals, the characteristic frequency $f_0$ of the fault signal is always unknown in advance. Therefore, the characteristic frequency $f_0$ should be pre-estimated according to the specific working environment before fault diagnosis, and the objective function $\text{SNR}_{\text{out}}$ for optimization and $\text{SNR}_{\text{in}}$ for comparison are redefined as follows:

\[
\text{SNR}_{\text{out}} = 10 \log_{10}\left(\frac{\max\{|X(f)|_{f_1 \leq f \leq f_h}\}^2}{\sum |X(f)|^2 - \max\{|X(f)|_{f_1 \leq f \leq f_h}\}^2}\right) \tag{10}
\]

\[
\text{SNR}_{\text{in}} = 10 \log_{10}\left(\frac{\max\{|\text{SN}(f)|_{f_1 \leq f \leq f_h}\}^2}{\sum |\text{SN}(f)|^2 - \max\{|\text{SN}(f)|_{f_1 \leq f \leq f_h}\}^2}\right) \tag{11}
\]

where $f_1 = f_0 - \Delta f$ and $f_h = f_0 + \Delta f$; where $\Delta f = f_s / (RM)$ is the adjusted frequency resolution after scale transformation. Consequently, several spectral lines around the pre-estimated frequency are involved to avoid the characteristic spectral line being missing.

4.1. Verification of the HOA

In this subsection, the advantage of the HOA is verified by processing a simulated signal with different injected noises. The signal was set as a pure harmonic signal with amplitude $A = 0.1$ and frequency $f_0 = 40$ Hz; the sampling frequency $f_s = 20000$ Hz and the sampling points $N = 2000$. This signal was injected with noises of different intensities $D_k$ ranging from 0.04 to 5.12, and the obtained noisy signals were then input into the multiparameter-adjusting Duffing system shown in Equation (2). Both the HOA and the conventional GA were used to find the optimal $\text{SNR}_{\text{out}}$ for each signal.

Before optimizations, the value of $R$ should be determined, and the range of each adjustable parameter ($k$, $a$, $b$, $\varepsilon$) should be selected as well. The value of $R$, which is the scale transformation parameter, was fixed as $R = 2000$ according to the large frequency domain and the large sampling frequency $f_s$ to ensure that the calculation results will not overflow in the numerical simulation. Moreover, $\varepsilon$ is the amplitude transformation coefficient, whose selected range should be determined according to the amplitude of the input signal. Based on our previous research [30], the value of $\varepsilon A$ should be between 0.001 and 0.1, where $A = \sum_{f_1 < f < f_2} A(f) / n$ with $A(f)$ the single side spectrum of the input signal and $n$ the number of spectral lines during the frequency range of $[f_1, f_2]$ which is the manual-selected frequency range involving $f_0$. In our simulation, $f_1 = 35$ Hz and $f_2 = 45$ Hz were set. For the given signals with different noises, the values of $A$ are within $[0.0056, 0.2416]$, hence, the range of $[0.2, 0.4]$ was selected for $\varepsilon$ to guarantee $0.001 < \varepsilon A < 0.1$. Moreover, the selected ranges of the other three parameters were set artificially as $a \in [0.1, 5]$, $b \in [0.1, 5]$ and $k \in [0.1, 5]$ in optimization. Under different noise intensities, the optimal SNRout of the Duffing system obtained from both the HOA and the GA with a population number of 500 are shown in Figure 4. The SNR of input signals with different noise intensities are plotted in this figure as well.
4.2. Application for Practical Bearing Fault Data Classification

4.2.1. Introduction of the Used Bearing Fault Data Set

In this subsection, the vibration signals of the rotating bearing from the bearing data center of Case Western Reserve University (CWRU) were processed by the proposed method, thus verifying its feasibility in bearing fault data classification and fault diagnosis. The test rig is shown in Figure 5, which contained a motor with a load of up to 3 hp, a torque transducer or encoder, and a dynamometer.

Figure 4. SNR of the optimization results based on HOA and GA with a population number of 500 and SNR of the input signal against noise intensity.

Figure 4 shows that the SNR_{in} of input signal presents a decreasing trend as the noise intensity increases, while both the HOA and the GA can achieve a relatively high SNR_{out} regardless of the noise intensity, demonstrating the feasibility of the multiparameter optimization algorithms in achieving SR in the Duffing system. Moreover, in most cases (80 of 128) the optimal SNR_{out} obtained from the HOA is larger than that obtained from the GA. The advantage of the HOA is it can be also quantitatively concluded that the average value of the optimal SNR_{out} obtained from the HOA (−0.8616 dB) is larger than that obtained from the GA (−1.3808 dB). This result indicates that the HOA has a higher possibility to obtain a better local optimization result compared to the GA.

Besides, it is convenient to set a large population number to obtain a local optimization close to the global optimization with a large SNR_{out}, but it takes a lot of extra time. Whereas, in practical analysis, the time for fault diagnosis is relatively short. Therefore, a smaller population number should be used in practical engineering to achieve acceptable time cost and SNR_{out}. Its influence on the classification results will be studied in Section 4.2.4. The population number was set as 50 in the following of this section.

4.2. Application for Practical Bearing Fault Data Classification

4.2.1. Introduction of the Used Bearing Fault Data Set

In this subsection, the vibration signals of the rotating bearing from the bearing data center of Case Western Reserve University (CWRU) were processed by the proposed method, thus verifying its feasibility in bearing fault data classification and fault diagnosis. The test rig is shown in Figure 5, which contained a motor with a load of up to 3 hp, a torque transducer or encoder, and a dynamometer.

Figure 5. CWRU bearing test rig.
In the test rig, the test bearings, which were deep groove ball bearings of type 6205-2RS JEM SKF, were used to support the motor shaft. The bearing details are listed in Table 1. In the test, motor bearings were seeded with faults using electro-discharge machining. The diameters of the faults ranged from 0.007 to 0.04 inches, and the faults were separately located at the inner ring, outer ring and rolling element. Faulty bearings were installed onto the test motor, and the vibration data was recorded under the load of 0 to 3 hp (the motor speed was 1797 to 1720 rpm). Therefore, the bearings contained different faults with different health conditions, producing a variety of vibration signals of faulty bearings when they operated.

Table 1. Drive end bearing details and fault frequency.

| Position on Rig | Inside Diameter (inch) | Outside Diameter (inch) | Thickness (inch) | Ball Diameter (inch) | Defect Frequencies (Multiple of Running Speed in Hz) |
|----------------|------------------------|-------------------------|------------------|---------------------|-----------------------------------------------|
|                |                        |                         |                  |                     | Inner Ring /srev | Outer Ring /srev | Rolling Element /srev |
| Drive end      | 0.9843                 | 2.0472                  | 0.5906           | 0.3126              | 5.4152          | 3.5848          | 4.7135              |

The bearing data used in the experiment was sampled at the end of the drive with a sampling frequency of 12,000 Hz. As the location of the fault relative to the bearing load area affected the vibration response of the whole motor system, the bearing data at 6 o’clock, 3 o’clock and 12 o’clock directions of the bearing load area were listed, respectively. In this work, the bearing data at 6 o’clock was used for verification.

However, the bearing data set only contains one bearing signal of each fault type, which is not enough for data training. To obtain more fault signals to make the classification results more generalized, each bearing signal was expanded to 400 samples, as shown in Figure 6. The first 512 time domain points of each bearing signal are intercepted from each bearing signal, thus forming a new signal indicated as \( s_1(t) \). Next, the 257th to 768th time domain points of the bearing signal are intercepted to form a new signal indicated as \( s_2(t) \). More new signals can be further obtained using the same method. In this work, each bearing signal was expanded to 400 new signals. Moreover, to reduce the amount of computation and save time cost, the parameter set optimized by \( s_1(t) \) was used for other expanded new signals.

Figure 6. The signal expanding method used in this work.

The details of the datasets are shown in Table 2. The datasets contain the signal of the 10 different healthy condition categories under four different loads of 1, 2, 3 and 4 hp, which are represented as datasets A, B, C and D, respectively. Each bearing signal was expanded to 400 samples, among which 320 samples were training samples and 80 samples were testing samples.
Table 2. Details of the bearing datasets.

| Category labels | Inner Race | Outer Race | Ball | Normal |
|-----------------|------------|------------|------|--------|
| Fault diameters (inches) | 0.007 | 0.014 | 0.021 | 0.007 | 0.014 | 0.021 | 0.007 | 0.014 | 0.021 | 0 |
| Dataset A: Load 1 | Train | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 |
|                  | Test   | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  |
| Dataset B: Load 2 | Train | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 |
|                  | Test   | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  |
| Dataset C: Load 3 | Train | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 |
|                  | Test   | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  |
| Dataset D: Load 4 | Train | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 | 320 |
|                  | Test   | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  | 80  |

4.2.2. Optimization Results

In this subsection, the samples of dataset A are taken as examples to be processed by adaptive optimization of SR. As we mentioned at the beginning of Section 4, the characteristic frequencies of the faulty bearings should be estimated first to calculate the SNR_{out}. When the system rotated at a constant speed, the characteristic frequency of the bearings can be calculated according to Table 1.

In the optimization processes, the scale transformation parameter was set as $R = 2000$, thus, the frequency resolution $\Delta f$ at sampling frequency of 12,000 Hz can be obtained as $\sim 0.0117$ Hz. The optimization ranges of other adjustable parameters are pre-set, as shown in Table 3. Note that the characteristic frequency of the normal bearing signals cannot be estimated. As a result, the parameters of the Duffing system with normal bearing signals input cannot be optimized, whereas the normal bearing signals are also processed by the Duffing system with a group of fixed parameters ($R = 2000, k = 0.5, a = b = 1, \varepsilon = 10$) to maintain the uniformity compared to other 9 fault bearing signals.

Table 3. Optimization ranges of each adjustable parameter of dataset A.

| Fault Position | Fault Diameter (inch) | $k$ | $a$ | $b$ | $\varepsilon$ |
|----------------|----------------------|-----|-----|-----|------------|
| inner race     | 0.007                | [0.1, 5] |
|                | 0.014                | [0.1, 5] |
|                | 0.021                | [0.15, 10] |
| ball           | 0.007                | [0.1,15] | [0.1,10] | [0.1,10] | [0.5, 15] |
|                | 0.014                | 0.021 | [0.15, 10] |
| outer race     | 0.007                | [0.1, 10] |
|                | 0.014                | [0.1, 10] |
|                | 0.021                | [0.1, 5] |

Hence, the SNR_{out} of the Duffing system with different input signals can be optimized according to the HOA. For example, the values of the optimized SNR_{out} for $s_1(t)$ of the faulty signals of dataset A against the noise intensity of the injected noise are shown in Figure 7. It is necessary to mention that the SNR_{in} curves are not drawn because signal numbers are too small to find precise feature frequency, but the SNR_{in} can be obtained by using Equation (11), which are less than $-30$ dB. Figure 7 shows that all the SNR_{out} are more than $-20$ dB. It can be obtained that the optimized SNR_{out} has significant enhancements compared to the SNR_{in}, demonstrating the feasibility and effectiveness of the HOA in enhancing the weak features of practical signals.
Figure 7. The optimized SNR\textsubscript{out} of the faulty bearing signals with faults of (a) inner ring, (b) rolling element and (c) outer ring against the intensity of injected noise.

4.2.3. Accuracies of Classification

The classification accuracies of the bearing signals are presented in this subsection. It is noted that the batch size has a significant influence on the coverage speed and classification result. In this work, the batch size was set as 32 to obtain the highest classification accuracy with a relatively high convergence speed. Through simulations, an accuracy of 100% can be obtained in Table 4 for the 10 categories of all datasets presented in Table 2, which is higher than that obtained from other classification methods including SVM, Multilayer Perceptron (MLP), and DNN [47], showing that the proposed method has a good performance in fault classification and feature extraction. Moreover, Figure 8 shows the confusion matrix of dataset A, which clearly shows that each label was classified well.

Table 4. Accuracies for the signals with 10 conditions using optimization method.

| Dataset | A     | B     | C     | D     |
|---------|-------|-------|-------|-------|
| Accuracy| 100%  | 100%  | 100%  | 100%  |

However, an accuracy of 100% is difficult to achieve in practical engineering as the number of fault categories is far more than 10. To study the classification performance of the proposed method for more fault categories, the datasets A, B, C and D presented in Table 2 were combined in pairs to increase the health conditions to 20. The new datasets, which include 20 categories of bearing signals, were further processed by the proposed method, and the classification accuracies of the testing data are shown in Table 5. An accuracy of more than 96.9% was achieved. As comparisons, the raw signals, which were not processed by the multiparameter adjusting Duffing system before classification, were also processed using the proposed CNN and the traditional CNN; the optimal signals were also processed using the traditional CNN. The classification accuracies of the testing data are shown in Table 5. One can see that the classification accuracies are enhanced by either pre-processing...
the raw signals or using our CNN architecture, and the classifications accuracies achieved the highest when both methods were adopted. Therefore, both the optimization method and our CNN architecture play important roles in enhancing classification accuracies, showing that the proposed method has the capability to realize high-performance fault classification and feature extraction results better than conventional methods. It is noted that as the accuracy is different for every training and inference, each classification accuracy presented in Table 5 was obtained by taking the average of the accuracy results of five simulations. To observe which healthy conditions were difficult to classify, Figure 9 shows the confusion matrix of datasets A and B using optimal signals with our CNN architecture. Only two normal signals under different loads had misclassification, which means that two normal signals contain similar information and features and it was difficult to classify them correctly.

Figure 8. The confusion matrix of dataset A (10 healthy conditions) with 100% accuracy using optimal signals with our CNN architecture.

Table 5. Accuracies for the signals with 20 healthy conditions using three different classification methods.

| Datasets A and B | Optimal Signals with | Optimal Signals with | Raw Signals with | Raw Signals with |
|------------------|----------------------|----------------------|------------------|------------------|
|                  | Our CNN Architecture | Traditional CNN Architecture (without BN and Dropout Layers) | Our CNN Architecture | Traditional CNN Architecture (without BN and Dropout Layers) |
| Datasets A and B | 99.16%               | 98.06%               | 98.44%           | 94.31%           |
| Datasets B and C | 96.90%               | 95.94%               | 95.44%           | 85.44%           |
| Datasets C and D | 97.15%               | 95.96%               | 95.94%           | 90.44%           |

4.2.4. Influence of the Population Number on Classification Accuracies and Calculation Time

In addition to the classification accuracies, the calculation time is another important index to evaluate the performance of a classification method. Both indexes are affected by the population number, which is studied in the subsection.

The datasets shown in Table 4 were re-processed using the proposed method with different population numbers. The accuracies and calculation time against the population number are shown in Figure 10. It can be seen from Figure 10a that when the population number increased from 1 to 50, the classification accuracies had slight increases of less than 1%, meaning that the population number had a relatively low influence on the classification accuracies. Figure 10b shows that increasing the population number significantly increases
the calculation time. When the population number increased from 1 to 50, the cost time for the optimization process increased from 30 s to 6500 s. Therefore, it is possible to get an acceptable high-accuracy classification result in a short time using the proposed method.

![Confusion Matrix](image)

**Figure 9.** The confusion matrix of datasets A and B (20 healthy conditions) using optimal signals with our CNN architecture.

![Accuracy and Calculation Time](image)

**Figure 10.** (a) The classification accuracies and (b) the calculation time against the population number.

### 4.2.5. Visualizations of Feature Maps and Networks

Generally, CNN is an efficient tool to extract features, but it is hard to understand how CNN processes grey images. In this subsection, the feature maps and networks are plotted for a better understanding of the powerful feature extraction and classification capacities of CNN.

For dataset A, Figure 11 shows the feature distributions of some representative layers in the CNN visualized by the t-distributed stochastic neighbor embedding (t-SNE) [48]. The features of the input signals of the CNN, which are the output signals of the Duffing system, are not obvious. The features are continuously separated by each convolutional layer, and the classification result of each fault type becomes obvious after the fourth convolutional layer. Moreover, the features in the fully connected layer are easier to be divided and an accuracy of 100% can be obtained.
A novel intelligent fault diagnosis method combining the adaptive multiparameter stochastic resonance (SR) as a weak-fault feature extraction method and the normal convolutional neural network (CNN) as a fault classification method is proposed in this paper for bearing fault enhancement diagnosis. A multiparameter bistable Duffing system that can realize SR for large-parameter signals is introduced as a weak-fault feature extraction model, whose optimal output signal-to-noise ratio (SNR) can be achieved adaptively using a proposed hybrid optimization algorithm (HOA) combining the genetic algorithm (GA) and simulated annealing (SA) with high optimization speed. Therefore, poor-quality raw data can be pre-processed by the adaptive multiparameter-adjusting SR of the Duffing system. A mapping method that can convert the SR outputs of the Duffing system into grey images is proposed. The obtained grey images are further processed by a normal CNN with batch normalization (BN) and dropout layers, thus achieving fault enhancement classification. The research shows that a relatively high output SNR can be obtained from the adaptive multiparameter-adjusting Duffing system using the proposed HOA that has a better optimization performance compared to the GA. Therefore, weak features of raw data can be significantly enhanced. The bearing fault dataset from the CWRU bearing data center was pre-processed by the feature extraction method and was further processed by the proposed CNN model. Classification accuracies of 100% were achieved for 10 categories of bearing signals and those of more than 96.9% were achieved for 20 categories of bearing signals, which is better than other methods. The influences of the population number on the classification accuracies and calculation time were further studied, indicating that the proposed method can realize a rather good classification result in a short time. The feature map and network visualization are presented by t-SNE to show how the features were identified by the CNN. Hence, it was demonstrated that the proposed method can realize high-performance fault enhancement classification and identification. Moreover, to make full use of the advantage of SR in detecting weak signals, the abundant information of SR output, such as the frequency domain output, can be further combined with other appropriate intelligent methods for fault classification or prediction in future work.

Figure 11. Feature visualization via t-SNE: feature representations of (a) input images, (b–e) images obtained through four convolutional layers, and (f) images obtained through the last fully-connected layer.

5. Conclusions
Author Contributions: Conceptualization, C.W. and Z.L.; methodology, C.W., Z.Q. and Z.L.; software, C.W. and J.X.; validation, Z.Q. and C.Z.; formal analysis, Z.H.; investigation, Z.Q.; writing—original draft preparation, C.W.; writing—review and editing, S.F. and Z.L.; supervision, Z.L.; project administration, R.Z.; funding acquisition, Z.Q., J.L. and Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant Nos. 51905349, U2013603, 52205569), Guangdong Basic and Applied Basic Research Foundation (Grant Nos. 2022A151010126, 2020A151011509), Zhejiang Provincial Natural Science Foundation of China (LQ22E050003), Research Program of Tianjin Education Commission (Grant No. 2019KJ100), Ningbo Science and Technology Major Project (2020Z110, 2022Z057, 2022Z002), and Natural Science Foundation of Shenzhen University (Grant No. 860-000002110264).

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Qiao, Z.; Shu, X. Coupled neurons with multi-objective optimization benefit incipient fault identification of machinery. Chaos Solitons Fractals 2021, 145, 110813. [CrossRef]
2. Gong, W.; Chen, H.; Zhang, Z.; Zhang, M.; Wang, R.; Guan, C.; Wang, Q. A Novel Deep Learning Method for Intelligent Fault Diagnosis of Rotating Machinery Based on Improved CNN-SVM and Multichannel Data Fusion. Sensors 2019, 19, 1693. [CrossRef] [PubMed]
3. Lei, Y.; Jia, F.; Lin, J.; Xing, S.; Ding, S.X. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data. IEEE Trans. Ind. Electron. 2016, 63, 3137–3147. [CrossRef]
4. Lei, Y.; Li, N.; Lin, J.; He, Z. Two new features for condition monitoring and fault diagnosis of planetary gearboxes. J. Vib. Control 2013, 21, 755–764. [CrossRef]
5. Glowacz, A.; Glowacz, W.; Glowacz, Z.; Kozik, J. Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals. Measurement 2018, 113, 1–9. [CrossRef]
6. Ottewill, J.; Orkisz, M. Condition monitoring of gearboxes using synchronously averaged electric motor signals. Mech. Syst. Signal Process. 2013, 38, 482–498. [CrossRef]
7. Zhang, M.; Jiang, Z.; Feng, K. Research on variational mode decomposition in rolling bearings fault diagnosis of the multi-stage centrifugal pump. Mech. Syst. Signal Process. 2017, 93, 460–493. [CrossRef]
8. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. Signal Process. 2014, 96, 1–15. [CrossRef]
9. Loutridis, S. Damage detection in gear systems using empirical mode decomposition. Eng. Struct. 2004, 26, 1833–1841. [CrossRef]
10. Sun, W.; Chen, J.; Li, J. Decision tree and PCA-based fault diagnosis of rotating machinery. Mech. Syst. Signal Process. 2007, 21, 1300–1317. [CrossRef]
11. Lei, Y.; Han, D.; Lin, J.; He, Z. Planetary gearbox fault diagnosis using an adaptive stochastic resonance method. Mech. Syst. Signal Process. 2013, 38, 113–124. [CrossRef]
12. Yang, K.; Tong, W.; Huang, Z.; Qiu, T.; Lai, Z. Signal-to-noise ratio improvement of the signal immersed in the strong background noise using a bistable circuit with tunable potential-well depth. Mech. Syst. Signal Process. 2022, 177, 109201. [CrossRef]
13. Santos, P.; Villa, L.F.; Reñones, A.; Bustillo, A.; Maudes, J. An SVM-Based Solution for Fault Detection in Wind Turbines. Sensors 2015, 15, 5627–5648. [CrossRef] [PubMed]
14. Pandya, D.; Upadhyay, S.; Harsha, S. Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN. Expert Syst. Appl. 2013, 40, 4137–4145. [CrossRef]
15. Zhou, P.; Lu, S.; Liu, F.; Liu, Y.; Li, G.; Zhao, J. Novel synthetic index-based adaptive stochastic resonance method and its application in bearing fault diagnosis. J. Sound Vib. 2017, 391, 194–210. [CrossRef]
16. Benzi, R.; Sutera, A.; Vulpiani, A. The mechanism of stochastic resonance. Rev. Mod. Phys. 1998, 70, 223. [CrossRef]
17. Duan, L.; Ren, Y.; Duan, F. Adaptive stochastic resonance based convolutional neural network for image classification. Chaos Solitons Fractals 2022, 162, 112429. [CrossRef]
18. Zhou, S.; Cao, J.; Inman, D.J.; Lin, J.; Liu, S.; Wang, Z. Broadband tristable energy harvester: Modeling and experiment verification. Appl. Energy 2014, 133, 33–39. [CrossRef]
19. Qiao, Z.; Lei, Y.; Li, N. Applications of stochastic resonance to machinery fault detection: A review and tutorial. Mech. Syst. Signal Process. 2019, 122, 502–536. [CrossRef]
20. Lu, S.; He, Q.; Wang, J. A review of stochastic resonance in rotating machine fault detection. Mech. Syst. Signal Process. 2018, 116, 230–260. [CrossRef]
21. Agudov, N.V.; Krichigin, A.V.; Valenti, D.; Spagnolo, B. Stochastic resonance in a trapping overdamped monostable system. Phys. Rev. E 2010, 81, 051123. [CrossRef] [PubMed]
22. Yao, M.; Xu, W.; Ning, L. Stochastic resonance in a bias monostable system driven by a periodic rectangular signal and uncorrelated noises. Nonlinear Dyn. 2011, 67, 329–333. [CrossRef]
23. Qin, Y.; Tao, Y.; He, Y.; Tang, B. Adaptive bistable stochastic resonance and its application in mechanical fault feature extraction. *J. Sound Vib.* 2014, 333, 7386–7400. [CrossRef]

24. Li, J.; Zhang, J.; Li, M.; Zhang, Y. A novel adaptive stochastic resonance method based on coupled bistable systems and its application in rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* 2019, 114, 128–145. [CrossRef]

25. Gammaitoni, L.; Marchesi, F.; Menichella-Saetta, E.; Santucci, S. Stochastic resonance in bistable systems. *Phys. Rev. Lett.* 1989, 62, 349–352. [CrossRef]

26. Lai, Z.H.; Liu, J.S.; Zhang, H.T.; Zhang, C.L.; Zhang, J.W.; Duan, D.Z. Multi-parameter-adjusting stochastic resonance in a standard tri-stable system and its application in incipient fault diagnosis. *Nonlinear Dyn.* 2019, 96, 2069–2085. [CrossRef]

27. Han, D.; Su, X.; Shi, P. Stochastic resonance in multi-stable system driven by Lévy noise. *Chin. J. Phys.* 2018, 56, 1559–1569. [CrossRef]

28. Qiao, Z.; Elhattab, A.; Shu, X.; He, C. A second-order stochastic resonance method enhanced by fractional-order derivative for mechanical fault detection. *Nonlinear Dyn.* 2021, 106, 707–723. [CrossRef]

29. Jin-Jun, L.; Yong-Gang, L.; Zhi-Hui, L.; Dan, T. Stochastic resonance based on frequency information exchange. *Acta Phys. Sin.* 2016, 65, 220501. [CrossRef]

30. Lai, Z.-H.; Leng, Y.-G. Generalized Parameter-Adjusted Stochastic Resonance of Duffing Oscillator and Its Application to Weak-Signal Detection. *Sensors* 2015, 15, 21327–21349. [CrossRef]

31. Lai, Z.H.; Wang, S.B.; Zhang, G.Q.; Zhang, C.L.; Zhang, J.W. Rolling Bearing Fault Diagnosis Based on Adaptive Multi-parameter-Adjusting Bistable Stochastic Resonance. *Shock Vib.* 2020, 2020, 6096024. [CrossRef]

32. Tong, L.; Li, X.; Hu, J.; Ren, L. A PSO Optimization Scale-Transformation Stochastic-Resonance Algorithm with Stability Mutation Operator. *IEEE Access* 2018, 6, 1167–1176. [CrossRef]

33. Xia, P.; Xu, H.; Lei, M.; Ma, Z. An improved stochastic resonance method with arbitrary stable-state matching in underdamped nonlinear systems with a periodic potential for incipient bearing fault diagnosis. *Meas. Sci. Technol.* 2018, 29, 085002. [CrossRef]

34. Li, Z.; Wei, W.; Hu, K.; Chen, H.; Wang, Y.; Liu, Q.; Liu, S. Simulated Annealing Wrapped Generic Ensemble Fault Diagnostic Strategy for VRF System. *Energy Build.* 2020, 224, 110281. [CrossRef]

35. Katoch, S.; Chauhan, S.S.; Kumar, V. A review on genetic algorithm: Past, present, and future. *Multimed. Tools Appl.* 2021, 80, 8091–8126. [CrossRef] [PubMed]

36. Bertsimas, D.; Tsitsiklis, J. Simulated annealing. *Stat. Sci.* 1993, 8, 10–15. [CrossRef]

37. Xiong, J.; Liu, X.; Zhu, X.; Zhu, H.; Li, H.; Zhang, Q. Semi-Supervised Fuzzy C-Means Clustering Optimized by Simulated Annealing and Genetic Algorithm for Fault Diagnosis of Bearings. *IEEE Access* 2020, 8, 181976–181987. [CrossRef]

38. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. 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]

39. Guo, X.; Chen, L.; Shen, C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement* 2016, 93, 490–502. [CrossRef]

40. Zhou, F.; Sun, T.; Hu, X.; Wang, T.; Wen, C. A sparse denoising deep neural network for improving fault diagnosis performance. *Signal Image Video Process.* 2021, 15, 1889–1898. [CrossRef]

41. Shenfield, A.; Howarth, M. A Novel Deep Learning Model for the Detection and Identification of Rolling Element-Bearing Faults. *Sensors* 2020, 20, 5112. [CrossRef] [PubMed]

42. Li, H.; Huang, J.; Ji, S. Bearing Fault Diagnosis with a Feature Fusion Method Based on an Ensemble Convolutional Neural Network and Deep Neural Network. *Sensors* 2019, 19, 2034. [CrossRef] [PubMed]

43. Guo, S.; Yang, T.; Gao, W.; Zhang, C. A Novel Fault Diagnosis Method for Rotating Machinery Based on a Convolutional Neural Network. *Sensors* 2018, 18, 1429. [CrossRef] [PubMed]

44. Jing, L.; Zhao, M.; Li, P.; Xu, X. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* 2021, 111, 1–10. [CrossRef]

45. Lai, Z.H.; Leng, Y.G. Weak-signal detection based on the stochastic resonance of bistable duffing oscillator and its application in incipient fault diagnosis. *Mech. Syst. Signal Process.* 2016, 81, 60–74. [CrossRef]

46. Lu, W.; Liang, B.; Cheng, Y.; Meng, D.; Yang, J.; Zhang, T. Deep Model Based Domain Adaptation for Fault Diagnosis. *IEEE Trans. Ind. Electron.* 2017, 64, 2296–2305. [CrossRef]

47. Zhang, W.; Peng, G.; Li, C.; Chen, Y.; Zhang, Z. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals. *Sensors* 2017, 17, 425. [CrossRef]

48. Chatzimparmpas, A.; Martins, R.M.; Kerren, A. T-visne: Interactive assessment and interpretation of tsne projections. *IEEE Trans. Vis. Comput. Graph.* 2020, 26, 2696–2714. [CrossRef]