electronics

MDPI

Article

Bearing Fault Diagnosis Method Based on Osprey–Cauchy–Sparrow Search Algorithm-Variational Mode Decomposition and Convolutional Neural Network-Bidirectional Long Short-Term Memory

Zhiyuan Xiong [1], Haochen Jiang [2], Da Wang [2], Xu Wu [2,*] and Kenan Wu [3]

[1] Department of Mechanical Engineering, Monash University, Melbourne 3800, Australia; xiongzhiyuan98@163.com
[2] Information Science and Technology College, Dalian Maritime University, Linghai Road 1, Dalian 116026, China; 13998966658@163.com (H.J.); wd666@dlmu.edu.cn (D.W.)
[3] Key Laboratory of Chemical Lasers, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Zhongshan Road 457, Dalian 116023, China; wukn@dicp.ac.cn
* Correspondence: xuwu@dlmu.edu.cn

Abstract: To solve the problem of the low diagnosis rate of early weak faults of rolling bearings, a novel bearing fault diagnosis method based on Variational Mode Decomposition (VMD) and convolutional neural network (CNN)−Bidirectional Long Short-Term Memory (BiLSTM) was proposed. Based on the basic Sparrow Search Algorithm, the tent chaotic mapping, the Osprey Optimization Algorithm, and the Cauchy mutation were used to enhance the global search ability of the algorithm. To improve the accuracy of fault diagnosis, the BiLSTM layer is introduced into CNN to preserve the global and local features to the maximum extent. The experimental results show that VMD avoids the end effect problem in Empirical Mode Decomposition (EMD). The accuracy rate of the diagnosis model based on CNN-BILSTM reached 97.6667%, which was higher than that of the common diagnosis model.

Keywords: fault diagnosis; variational mode decomposition; Osprey optimization algorithm; Cauchy mutation; CNN-BiLSTM

check for updates

Citation: Xiong, Z.; Jiang, H.; Wang, D.; Wu, X.; Wu, K. Bearing Fault Diagnosis Method Based on Osprey–Cauchy–Sparrow Search Algorithm-Variational Mode Decomposition and Convolutional Neural Network-Bidirectional Long Short-Term Memory. *Electronics* **2024**, *13*, 4853. https://doi.org/10.3390/electronics13234853

Academic Editor: Davide Astolfi

Received: 3 November 2024
Revised: 3 December 2024
Accepted: 5 December 2024
Published: 9 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Rolling bearings are crucial components of mechanical equipment and play a vital role in ensuring the normal operation of the machinery. Statistical data indicate that approximately 30% of mechanical failures are attributed to bearing issues [1]. Once bearing failures occur, they can lead to abnormal operation of equipment, potentially causing economic and safety concerns [2]. Therefore, researching bearing fault diagnosis holds significant practical significance as it aids in preventing equipment failures, reducing economic losses, and ensuring equipment safety. However, vibration signals from rolling bearings typically exhibit non-linear and non-stationary characteristics, accompanied by complex noise signals, which increases the difficulty of extracting fault features from them [3]. There are four stages for bearing fault detection, which are the shallow surface damage stage: at this stage, the bearing only shows shallow surface damage, with friction and slight impact, and the vibration frequency is above 20 KHz, which is difficult to hear.

Mild surface or shallow damage stage: During this stage, the amplitude of ultra-high frequency vibration increases, causing resonance and presenting the fault frequency in the spectrum.

Mild injury development stage: The injury further develops, and the amplitude of ultra-high frequency vibration continues to increase.

Severe damage stage: The damage is more extensive, which may lead to the damage of other components, and ultra-high frequency vibration may decrease. Addressing this challenge will provide a more accurate means of bearing fault diagnosis.

The Empirical Mode Decomposition (EMD) method can recursively decompose vibration signals into Intrinsic Mode Functions (IMFs). However, it suffers from mode mixing and end effect issues [4,5], which reduces its capability for feature extraction in bearing fault diagnosis. Although Ensemble Empirical Mode Decomposition (EEMD) has addressed some of these issues, it increases computational complexity and cannot eliminate white noise in decomposition, leading to increased reconstruction errors [6,7]. In contrast, the Variational Mode Decomposition (VMD) algorithm proposed by Dragomiretskiy et al. effectively eliminates mode mixing and exhibits good robustness to noise [8,9]. However, the VMD algorithm requires manual setting of parameters such as the number of modes and penalty factor, both of which significantly influence the effectiveness of feature extraction. Overall, compared to wavelet decomposition methods, the VMD algorithm has advantages in handling vibration signals but requires careful parameter tuning to achieve optimal results.

Bearing fault diagnosis based on vibration signals requires the identification of faults from analyzed and processed data [10]. Ref. [11] employed a one-dimensional deep residual neural network (DRNN) for bearing fault diagnosis, which to some extent enhanced the generalization ability of the dataset. However, DRNN's receptive field is limited, making it difficult to capture global information. Ref. [12] utilized Long Short-Term Memory (LSTM) neural networks for bearing fault diagnosis, but LSTM networks restrict the model's parallel capability, resulting in longer diagnostic times. Ref. [13] established an anomaly detection model based on CNN-LSTM, which can detect early faults. Ref. [14] proposes a feature-adaptive method called IES, which determines the information spectrum band from SCoh through the Candidate Fault Frequency Optimization Map (IESCFFOgram) for bearing fault diagnosis. However, the unoptimized complex network structure is challenging to train and may even suffer from degradation issues.

To fully extract features from bearing vibration signals and improve fault diagnosis accuracy, based on the aforementioned research findings, this paper proposes a bearing fault diagnosis method based on optimized VMD and CNN-BiLSTM (Bidirectional Long Short-Term Memory). By introducing Tent mapping, Cauchy mutation, and the Osprey Optimization Algorithm to enhance the Sparrow Search Algorithm, the global exploration capability improves and the algorithm's ability to escape local optima becomes enhanced. The improved Sparrow Algorithm is used to optimize VMD for decomposing bearing fault signals, effectively avoiding mode mixing and end effects issues [15]. Simultaneously, introducing a BiLSTM layer into CNN enables the network structure to preserve both global and local features to the maximum extent, thereby improving fault diagnosis accuracy.

2. OCSSA

2.1. SSA

The Sparrow Search Algorithm (SSA) is an optimization algorithm based on the foraging behavior of sparrows, proposed by Jiankai Xue et al. in 2020 [16]. This algorithm simulates two different types of sparrows: leaders and followers. When the population encounters predators, some individuals transmit danger signals through sound to ensure that the entire population can avoid danger promptly. These alert sparrows are referred to as vigilant individuals. These collaborative characteristics make SSA a biologically inspired iterative optimization algorithm that can be effectively applied to optimization problems.

However, it should be noted that the algorithm may fall into local optima during the convergence process due to a decrease in population diversity, which is considered one of its main shortcomings. To improve the global search capability of the Sparrow algorithm, this paper introduces tent chaotic mapping, the Osprey Optimization Algorithm, and Gaussian mutation to improve it.

2.2. Tent Chaotic Mapping

One notable feature of the Sparrow Search Algorithm is the preservation of the random distribution of elements during the initialization phase [17]. During this phase, the positions and behaviors of sparrow individuals are influenced by randomness. However, this randomness may lead to an uneven distribution of the sparrow population in the search space, which could affect the subsequent convergence speed of the algorithm. Because the initial uneven distribution may result in some regions being overly explored while others are less explored. Therefore, this paper replaces the conventional random number generator in the algorithm with chaotic mapping to initialize the population. Chaotic mapping exhibits a high dependence on initial values, meaning that two nearby initial values will yield completely different random number sequences. In the current optimization algorithm application scenario, the following three different types of chaotic mappings are mainly utilized: Tent mapping, Logistic mapping, and Circle mapping. Their function graphs are depicted in Figure 1.

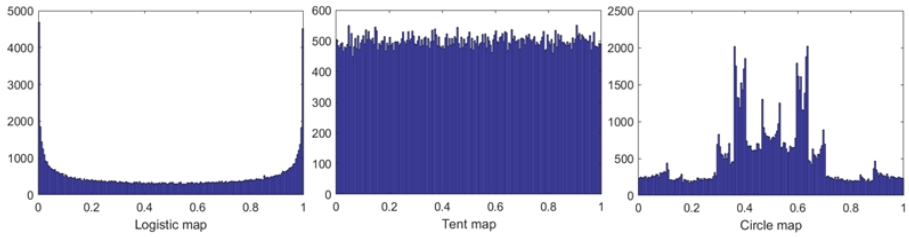


Figure 1. Chaos mapping function.

From the graph, it can be observed that Tent mapping stands out for its characteristic uniform traversal. Compared to other chaotic mappings, it demonstrates superior traversal performance. This makes Tent mapping an excellent generator of chaotic sequences, particularly suitable for optimization algorithms and related fields. Its expression is as follows:

$$z_{i+1} = \begin{cases} 2z_i & 0 \le z \le \frac{1}{2} \\ 2(1 - z_i) & \frac{1}{2} \le z \le 1 \end{cases} \quad (1)$$

2.3. Osprey Optimization Algorithm

The Osprey Optimization Algorithm (OOA) is an optimization algorithm proposed in 2023, based on the hunting behavior of ospreys. This algorithm establishes a mathematical model by simulating the hunting behavior of ospreys, featuring strong optimization capability and fast convergence speed.

Comparing the two algorithms, it can be found that both are cluster-based metaheuristic algorithms, and the Osprey Optimization Algorithm and Sparrow Optimization Algorithm have many similarities.

However, compared to the Sparrow Search Algorithm, the Osprey Optimization Algorithm is based on the hunting behavior of ospreys. Each position update is relatively independent, which accelerates the convergence speed of the algorithm and reduces the possibility of falling into local optima traps.

Therefore, to compensate for the excessive reliance on the previous generation sparrow position update method in the leader position update of the sparrow optimization algorithm, this paper adopts the global exploration strategy of the Osprey Optimization Algorithm in the first stage to replace the leader position update formula of the original sparrow algorithm. While improving the convergence speed of the algorithm, the running time will not suddenly increase.

The new leader position update formula is as follows:

$$x_{i,j}^{P1} = x_{i,j} + r_{i,j} \cdot \left(SF_{i,j} - I_{i,j} \cdot x_{i,j} \right) \quad (2)$$

$x_{i,j}^{P1}$ is the new position of the i-th fish eagle in the j-th dimension during the first stage; F_i^{P1} represents the corresponding fitness values; $SF_{i,j}$ is a random number, whose scope ranges between 0 and 1; and $I_{i,j}$ is a random number; whose scope ranges between 1 and 2.

2.4. Cauchy Mutation Strategy

In sparrow foraging strategies, it is observed that followers tend to seek food around the best leaders and sometimes even compete to become new leaders. This strategy helps to avoid local optimization problems; however, to further improve the global optimization ability, the Cauchy mutation strategy is introduced to optimize the update formula of the follower. The new follower positions are updated as follows:

$$X_{i,j}^{t+1} = X_{\text{best}}(t) + cauthy(0,1) \oplus X_{\text{best}}(t) \tag{3}$$

In the equation $cauthy(0,1)$ represents the standard Cauchy distribution function, while $\oplus$ denotes the multiplication operation.

The comparison between the standard Cauchy distribution and the normal distribution is shown in Figure 2.

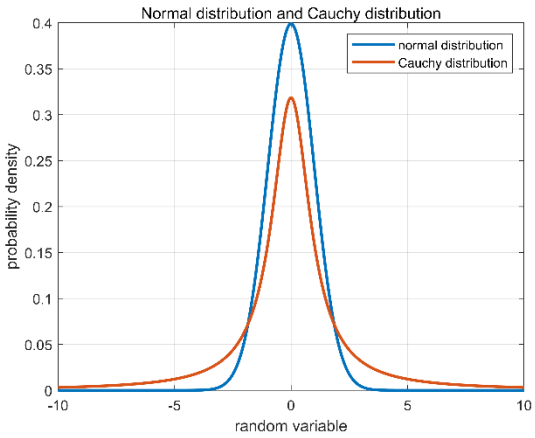


Figure 2. Standard normal distribution and Cauchy distribution function.

The graph shows that the Cauchy distribution exhibits a continuous nature and shares similarities with the conventional normal distribution in terms of shape. However, its shape is more elongated, and numerically, it approaches zero more closely near the origin, with a relatively slower decrease towards zero. This characteristic makes the Cauchy distribution particularly suitable for introducing larger perturbations when incorporated into optimization algorithms, aiding in exploring solutions more extensively within the search space. In the Sparrow Search Algorithm, by introducing the mutation strategy based on the Cauchy distribution, the positions of elements participating in the algorithm are refreshed. This helps increase the breadth of exploration, thereby enhancing the algorithm's potential to overcome local optima challenges and achieve a more comprehensive search effect during optimization.

In summary, building upon the traditional Sparrow Search Algorithm, this paper proposes the Osprey–Cauchy–Sparrow Search Algorithm (OCSSA), which integrates the Osprey Optimization Algorithm and Cauchy mutation. By enhancing the optimization capability of the Sparrow Search Algorithm, the complexity of the algorithm has not increased, and there is no significant increase in runtime. This approach enables better optimization of parameters for the Variational Mode Decomposition (VMD) algorithm.

2.5. OCSSA Performance Testing

In this section, comparative experiments are conducted between the Sparrow Search Algorithm (SSA) and the Osprey–Cauchy–Sparrow Search Algorithm (OCSSA) to evaluate their performance. In the experiments, the parameters of the algorithms are set, including a

population size of 50 and a maximum iteration of 1000. Four different types of benchmark functions are chosen to comprehensively assess the algorithm's performance, with the benchmark function indicators listed in Table 1. The main metrics for testing include the mean and standard deviation, which help provide a comprehensive understanding of the algorithm's performance under different conditions [18]. The test results are presented in Table 2.

Table 1. Four benchmark functions.

Number	Function	Dimension	Search Interval	Theoretical Extreme
F1	Sphere	30	[−100,100]	0
F2	Schwefell 1.2	30	[−100,100]	0
F3	Rastrigin	30	[−5.12,5.12]	0
F4	Ackley	30	[−32,32]	0

Table 2. Test results for four benchmark functions.

Function	Index	SSA	OCSSA
F1	mean	6.43×10^{-130}	0.00
	std	0.00	0.00
F2	mean	8.89×10^{-69}	0.00
	std	0.00	0.00
F3	mean	3.36×10^{-6}	0.00
	std	1.03×10^{-5}	0.00
F4	mean	8.88×10^{-16}	3.39×10^{-16}
	std	0.00	0.00

From the data in Table 2, it can be observed that in the comparative experiments on specified search space functions, the OCSSA demonstrates stronger optimization capability compared to SSA. OCSSA can approach the theoretical optimum more closely, while SSA can only find a value that is close to the theoretical optimum under the same conditions. Specifically, for the given functions, OCSSA is closer to the theoretical optimum compared to SSA, with a std value of 0, indicating that the optimization process of OCSSA is more stable. This result indicates that the OCSSA exhibits better performance in finding function optima compared to the traditional SSA.

To further validate the performance of the OCSSA, a comparison was made among five algorithms, including the basic Sparrow Search Algorithm (SSA), Dung Beetle Optimization Algorithm (DBO) [17], Subtractive Averaging-Based Optimization (SABO) algorithm [19], and Gorilla Troop Optimization Algorithm (GTO) [18], in addition to the OCSSA. The comparison was conducted on the four benchmark functions listed in Table 1.

To comprehensively compare the performance of different algorithms, we have used graphical methods and plotted convergence curves. We set the population size for all five algorithms as 30, the dimension as dim = 20, and limited the experiments to complete within a maximum of 1000 iterations. We repeated the experiments independently for 100 rounds to obtain the data representation of the convergence process. By exploring the graphs below, we observe that the curves stop decreasing after the number of iterations increases, indicating that all algorithms, to some extent, achieve convergence under the given experimental conditions. Particularly, the OCSSA demonstrates effective optimization capability towards the theoretical optimum of 0 during this process. This suggests that OCSSA has a competitive advantage in problem convergence performance compared to other algorithms under the given experimental conditions.

In Figures 3 and 4, from the convergence curves of F1 and F2, it can be observed that OCSSA exhibits a significant advantage in convergence speeds compared to other algorithms. The curve rapidly decreases from the beginning, requiring fewer iterations during the convergence process. At the same time, it does not lag in convergence accuracy.

This demonstrates that our improvement strategy has made our algorithm's performance more outstanding.

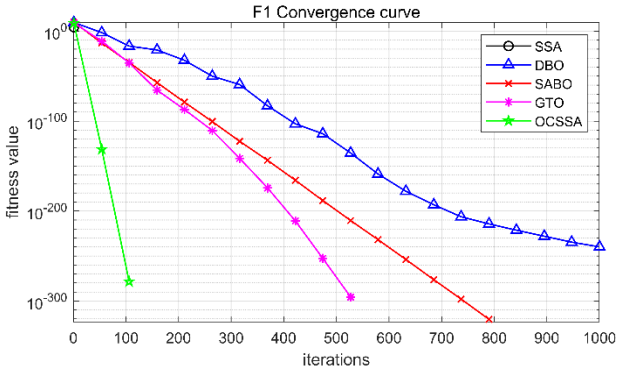


Figure 3. F1 convergence curves.

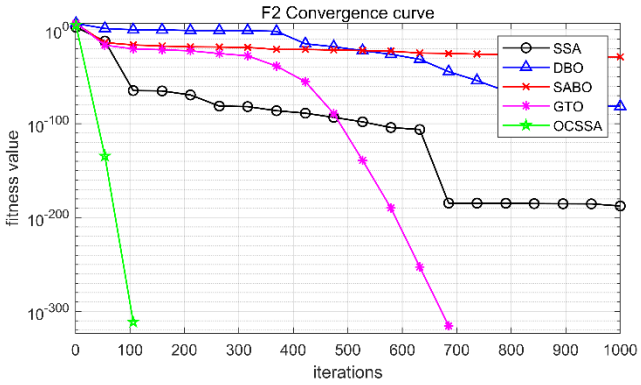


Figure 4. F2 convergence curves.

In Figures 5 and 6, the multi-modal convergence curves of F3 and F4 demonstrate that while the curves of other algorithms tend to flatten and stagnate, with lower optimization accuracy, our OCSSA shows stronger optimization accuracy and speed. This further confirms that introducing the Osprey Optimization Algorithm and Cauchy mutation strategy can significantly enhance the algorithm's ability to escape local optima and conduct rapid searches.

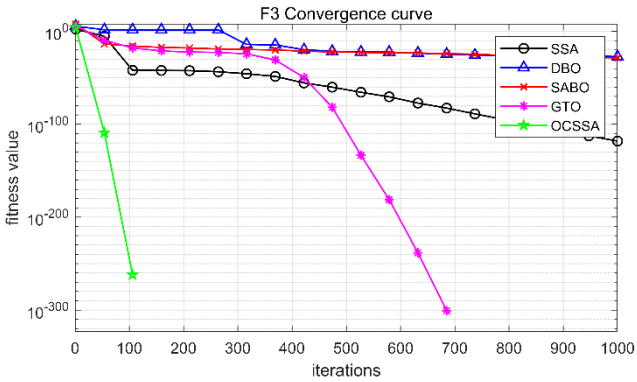


Figure 5. F3 convergence curves.

In conclusion, OCSSA demonstrates a significant improvement in optimization capability compared to SSA across the four benchmark functions tested. Its outstanding stability and robustness indicate its adaptability under different conditions. Additionally, OCSSA exhibits faster convergence speed and stronger global search capabilities compared to SSA, effectively avoiding falling into local optima. These experimental results fully

validate the feasibility and superiority of the OCSSA, making it an effective tool for solving optimization problems.

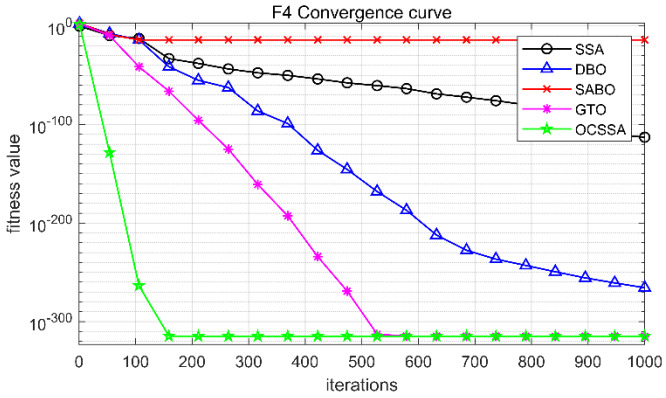


Figure 6. F4 convergence curves.

3. Fault Diagnosis Method for Rolling Bearings

3.1. Fault Diagnosis Ideas for Rolling Bearings

Firstly, a detailed discussion was conducted on the characteristic frequency of rolling bearings. Subsequently, the Variational Mode Decomposition (VMD) algorithm and convolutional neural network (CNN) model in deep learning were adopted, combined with existing knowledge to form the technical process of bearing from data collection to fault diagnosis, as shown in Figure 7.

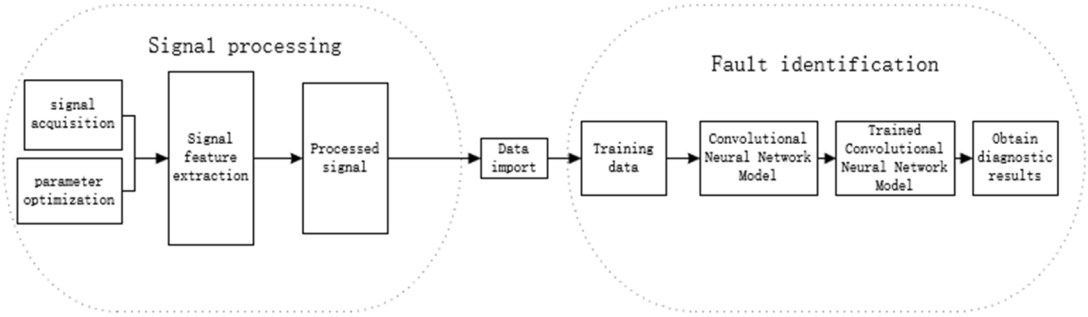


Figure 7. A technical approach for identifying bearing fault problems.

3.2. VMD Algorithm

Traditional VMD [20] decomposes signals into two main steps:

The first step is to construct a variational model, transforming the signal decomposition problem into a constrained optimization problem:

$$\begin{cases} \min\limits_{\{u_k\},\{\omega_k\}} \left\{ \sum\limits_{k=1}^{K} \left\| \delta_t \left[\left(\delta(t) + \dfrac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ s.t. \sum\limits_{k=1}^{K} u_k(t) = f(t) \end{cases} \tag{4}$$

The equation $\{u_k\} = \{u_1, \ldots, u_K\}$ represents the K-th mode component of the signal and the central frequency of each mode component.

The second step involves introducing the Lagrange operator λ and penalty factor α into Equations (3) and (4) to construct the augmented Lagrangian function, with an aim to obtain the optimal solution to the variational problem. This transformation converts the constrained problem of the variational model into an unconstrained problem, resulting in the following equation:

$$L(\{u_k\},\{\omega_k\},\lambda) = \alpha\sum_i \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\| + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \quad (5)$$

This equation represents the Lagrange transformation.

The core idea of the VMD algorithm is to decompose the input signal into a specified number of intrinsic mode functions (IMFs), each IMF having an independent central frequency. This decomposition process results in each IMF exhibiting sparsity characteristics in the frequency domain, indicating that different components are mainly concentrated in certain frequency ranges.

However, the VMD algorithm also has its drawbacks. Refs. [20,21] have demonstrated that the decomposition results of the VMD algorithm are directly related to the selection of parameters $[K, \alpha]$. Since there is no clear rule for selecting the values of these two parameters, manual selection may not yield the optimal decomposition results. Therefore, the OCSSA is needed to optimize the VMD parameters and globally search for the optimal decomposition results.

3.3. OCSSA Optimized VMD Algorithm

To optimize VMD, this paper adopts the OCSSA and pays special attention to the selection of the objective function. In the early stages of bearing failure, background noise often accounts for a significant proportion. As time progresses, bearing defects worsen, and periodic impact signals related to the defects become more pronounced. Therefore, this paper proposes envelope entropy E_p, where a lower value indicates the presence of significant regular impact signals inside the bearing, while a higher value indicates weaker impact signals. Thus, the magnitude reflects the sparsity of the bearing vibration signal, effectively assessing the bearing's fault status. The definition of envelope entropy is as follows:

$$\begin{cases} E_p = -\sum\limits_{j=1}^{N} p_j \lg p_j \\ p_j = \dfrac{a(j)}{\sum\limits_{j=1}^{N} a(j)} \end{cases} \quad (6)$$

This equation represents the Hilbert envelope signal of the original signal and represents the normalized result.

After decomposing the signal into components using the VMD algorithm, this paper takes the minimum value of the envelope entropy of these components as the fitness function for the OCSSA, aiming to optimize the global minimum value.

The flowchart of the OCSSA-VMD algorithm is shown in Figure 8, with the following steps:

(1) Setting the basic parameters of the algorithm involves specifying the upper and lower bounds of variables, such as $[K, \alpha]$.

(2) The population size and the number of iterations directly affect the speed and accuracy of the algorithm. Setting these parameters too large may lead to a decrease in algorithm speed and an increase in running time. Conversely, setting them too small may cause the algorithm to converge prematurely, making it difficult to find the optimal solution. To strike a balance between speed and accuracy, the paper sets the population size to 20 and the number of iterations to 30.

(3) Calculate the initial fitness value for each individual in the sparrow population, which corresponds to the envelope entropy of the IMF components mentioned earlier. Select the minimum envelope entropy as the optimal fitness value.

(4) By evaluating the performance of each individual, the initial positions of leaders, followers, and food sources within the population are preliminarily determined. This evaluation process relies on identifying the best performers and using their positions to locate the initial food sources, where "best" refers to individuals with the minimum

envelope entropy value. The division into leaders and followers is based on sorting the individuals by their fitness values, with the top 50% of individuals designated as leaders and the remaining individuals designated as followers.

(5) Update the position information by calculating the new fitness values for each individual during each iteration period and then updating their positions accordingly.

Update location information, calculate new fitness values for individuals during each iteration, and update them.

(6) Repeat step 5 until the iteration count is reached.

(7) Terminate iterations and output the optimal parameter combination.

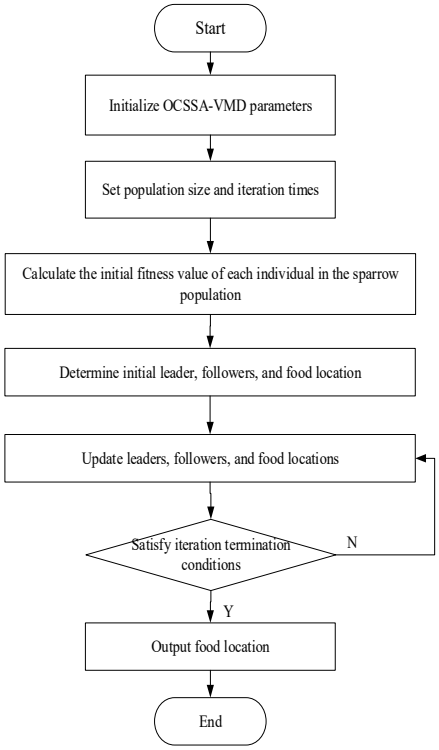


Figure 8. OCSSA optimizes the VMD algorithm flowchart.

3.4. Fault Diagnosis Model and Optimization

This article focuses on using a method that combines convolutional neural network layers and BiLSTM layers to identify the operating condition of rolling bearings. This method does not rely on the complexity of deep network structures. Based on this, we have implemented a series of improvement measures within the framework of the existing LeNet-5 model to improve the accuracy of recognition. The number of kernels in the convolutional layer is 64 and 32, with step sizes of and, respectively. The number of kernels is 32, and small step sizes are used to deeply mine fault features. To accelerate model training and effectively avoid overfitting, a Batch Normalization (BN) layer was added after each convolutional layer to normalize the feature parameters. Connecting a fully connected layer and a Dropout layer after the convolutional layer structure plays a role in linearly reshaping features and suppressing random elements to enhance the model's ability to cope with new data. Through these optimizations, the performance of the rolling bearing operating state recognition model has been improved.

3.4.1. BiLSTM Model Optimization CNN

The basic structure of a Convolutional Neural Network (CNN) includes the input layer, convolutional layers, activation functions, pooling layers, and fully connected layers [21]. Traditional CNNs operate on input data in the convolutional layers; however, they suffer from a limitation where learned features correspond only to limited regions of the input

data, making it difficult to capture global information. To address this issue, as shown in Figure 9, Bidirectional Long Short-Term Memory (BiLSTM) is introduced, which is positioned between the pooling layers and the fully connected layers, aiming to capture long-range dependencies in the vibration signals of the bearings. Each layer of the network serves a distinct function: the input layer receives vibration signals; the convolutional layers extract spatial features from the data; activation functions introduce non-linearity and sparsity to the neural network; pooling layers reduce overfitting by reducing the dimensionality of the feature vectors outputted by the convolutional layers; BiLSTM layers deeply extract features from the original sequences; and the fully connected layers connect all the features and feed the output values to the classifier. This network architecture is designed to leverage deep learning models efficiently for the classification of vibration signals from bearings.

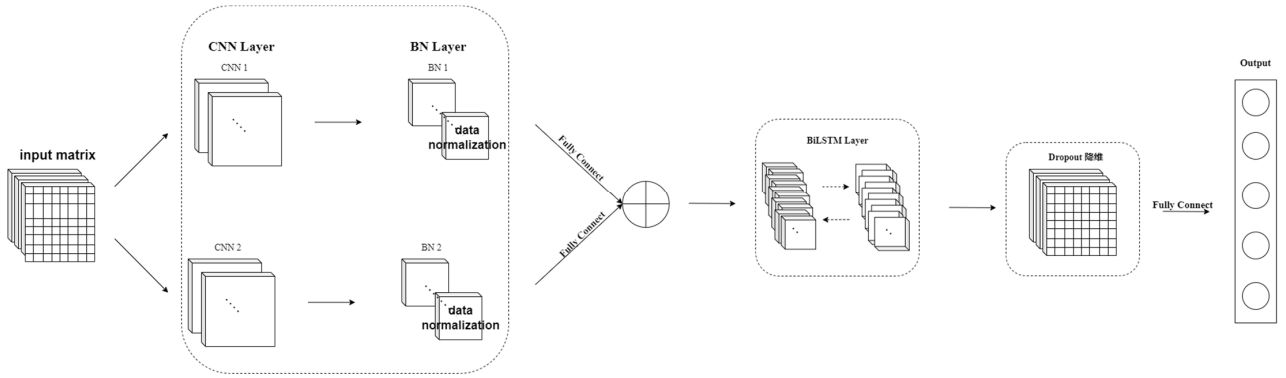


Figure 9. CNN-BiLSTM model structure.

3.4.2. Model Structural Parameters

In the network design, the convolutional layers use "same" padding to ensure that the spatial dimensions of the input and output remain the same. For pooling operations, max pooling is chosen to extract the most significant features from each pooling window. During the training process of the model, stochastic optimization algorithms are employed to effectively adjust network parameters and improve performance. The parameters related to the model structure are listed in Table 3.

Table 3. Structural parameters of the CNN BiLSTM model.

Number	Network Layer	Kernel Size/Step Size	Nuclear Quantity	Output Size	Zero Complement
1	network layer 1	$2 \times 2/2 \times 1$	64	$5 \times 6 \times 64$	yes
2	BN layer 1	-	64	$5 \times 6 \times 64$	-
3	network layer 2	$2 \times 2/2 \times 1$	32	$3 \times 6 \times 32$	yes
4	BN layer 2	-	32	$3 \times 6 \times 32$	-
5	BiLSTM layer	-	200	200	-
6	fully connected layer	100	1	100×1	-
7	Dropout layer	-	-	100×1	-

4. Verification of Fault Diagnosis Methods

4.1. Data Source Analysis

In this study, publicly available data from the rolling bearing fault simulation test rig at Case Western Reserve University, USA, was used to test the CNN-BiLSTM model. The test rig consists of two bearings, located at the drive end and the fan end, respectively. The primary focus of the research was on the test bearing SKF6205 located at the drive end, with a rotational speed of 1797 r/min. The data were sampled at a frequency of 12 kHz; the specific parameters of the bearing are shown in Table 4.

Table 4. The 6205-2RS JEM SKF deep groove ball-bearing dimensional information.

Internal Diameter/mm	External Diameter/mm	Thickness/mm	Rolling Element Diameter/mm	Pitch/mm
25	52	15	7.94	39.04

The fault type of the bearing is single-point damage caused by electrical discharge machining, with damage sizes of 0.007, 0.014, and 0.021 inches, respectively.

The data consist of four types: normal signals, inner race fault signals, outer race fault signals, and rolling element fault signals, comprising a total of ten signal sets in the dataset. Each signal set contains 300 samples, with each sample consisting of 2048 sampling points.

4.2. Analysis of Vibration Signals of Real Rolling Bearings Based on VMD

In this experiment, the faulty bearings were classified into three main types of faults: inner ring fault, outer ring fault, and rolling element fault. The placement of the faulty bearing is located closer to the motor side to minimize the vibration caused by the universal joint and improve the quality of signal acquisition. In engineering surveying applications, sensors are generally fixed on static parts such as mechanical casings or bearing seats to capture bearing vibration data. Therefore, to better meet the needs of real-life scenarios, we chose to position the sensor points collecting vibration data at the vertical top of the bearing seat of the tested bearing on the experimental platform.

This article uses the rolling bearing test bench experimental system mentioned above to collect bearing vibration signals. The sampling frequency is set to 12 kHz. The experiment includes normal bearings and the three types of faulty bearings mentioned earlier, among which the faulty bearings are manually machined with damage diameters of 0.5 mm, 0.7 mm, and 1 mm, covering a total of nine different damage states. The specific dataset partitioning is shown in Table 5.

Table 5. Real data experimental dataset.

Dataset		A		B		C		D	
Speed (RPM)		5000		7000		9000		ALL	
Position	size/mm	train	test	train	test	train	test	train	test
No		270	30	270	30	270	30	270	30
Rolling element	0.5	270	30	270	30	270	30	270	30
	0.7	270	30	270	30	270	30	270	30
	1.0	270	30	270	30	270	30	270	30
Inner circle	0.5	270	30	270	30	270	30	270	30
	0.7	270	30	270	30	270	30	270	30
	1.0	270	30	270	30	270	30	270	30
Outer circle	0.5	270	30	270	30	270	30	270	30
	0.7	270	30	270	30	270	30	270	30
	1.0	270	30	270	30	270	30	270	30

The time-domain signal chart of each part of the motor bearing under fault conditions is shown in Figure 10, which shows the time amplitude changes of signals under normal and abnormal conditions. Although there are certain differences in the time-domain waveforms of different faults, these differences are not sufficient for accurate bearing condition judgment. Therefore, FFT technology is used to perform frequency-domain conversion on vibration signals for a more detailed analysis. Figure 11 shows the spectral distribution of normal and various types of faulty bearing signals after FFT conversion.

In Figure 11, the frequency-domain image of the bearing vibration signal obtained through the Fourier transform is presented. Normal bearings and faulty bearings exhibit different frequency-domain features and amplitudes in the image. To quickly and effectively

diagnose bearing faults, it is necessary to establish a convolutional neural network model, as there are identifiable fault features in frequency-domain images, and convolutional neural networks can improve the speed and accuracy of bearing fault diagnosis in this regard.

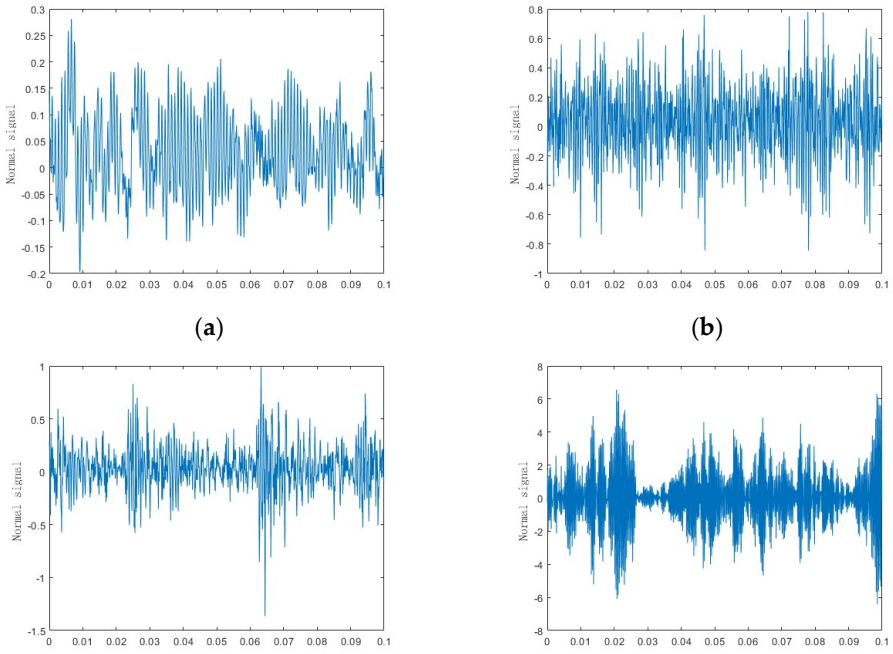


Figure 10. The time-domain diagram of four states of bearings (**a**–**d**) are normal signal, inner circle malfunction, outer ring fault, and rolling element malfunction, respectively.

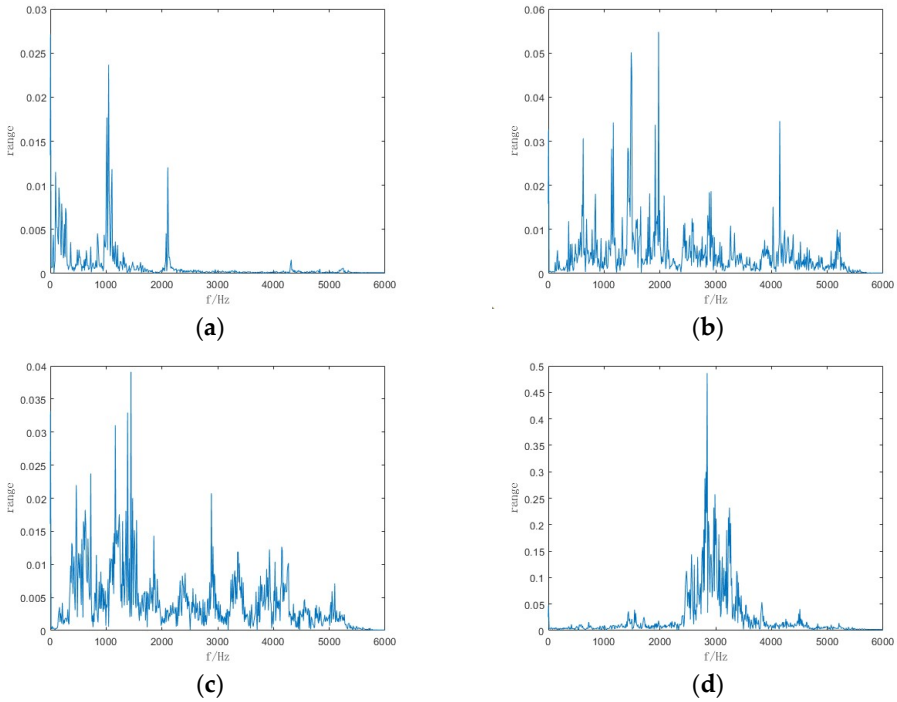
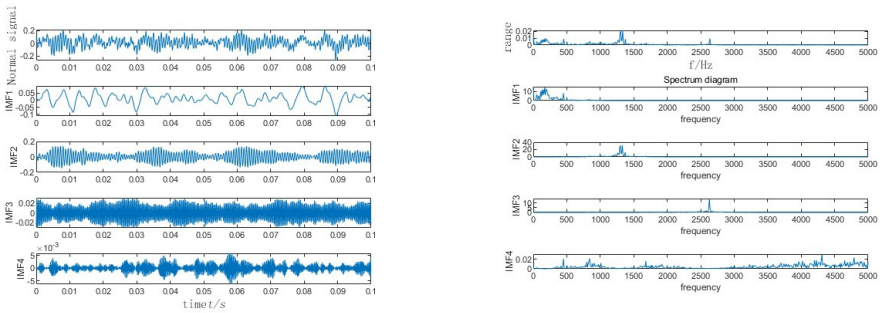


Figure 11. The frequency-domain diagram of four states of bearings (**a**–**d**) are normal signal, inner circle malfunction, outer ring fault, and rolling element malfunction, respectively.

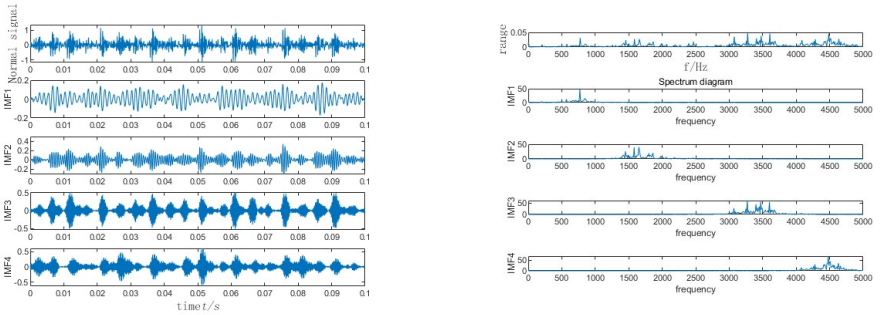
Subsequently, through the VMD algorithm, the vibration signals of the bearing in different states were successfully decomposed into multiple intrinsic mode components (IMFs).

In Figures 12–15, the performance of each IMF component in the time–frequency domain is thoroughly illustrated, clearly revealing the time–frequency characteristics of the signal. Through these detailed illustrations, we can intuitively identify the key frequencies of the vibration signal.



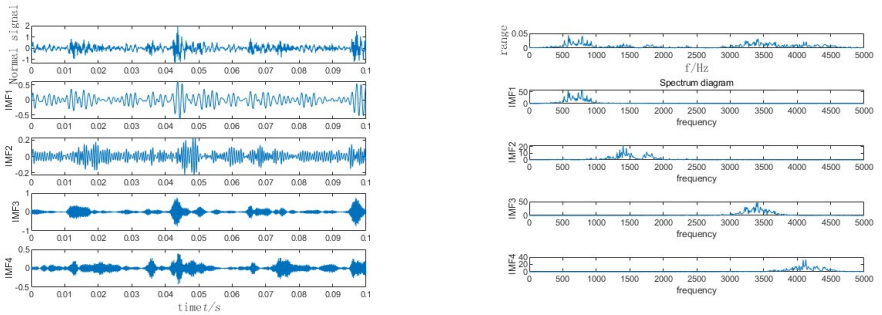
(**a**) (**b**)

Figure 12. (**a**,**b**) Normal state of VMD.



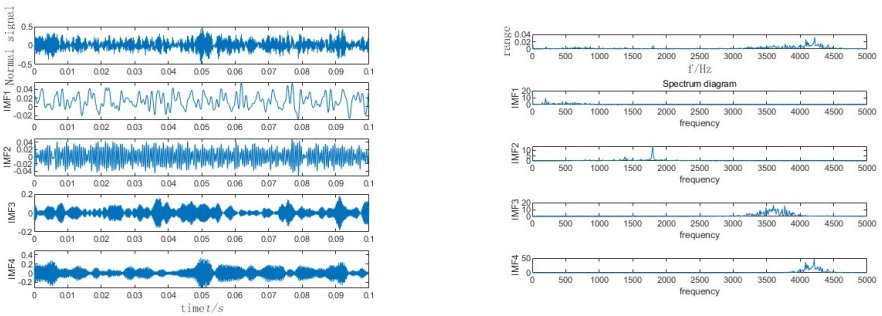
(**a**) (**b**)

Figure 13. (**a**,**b**) VMD of the inner ring fault.



(**a**) (**b**)

Figure 14. (**a**,**b**) VMD of the outer ring fault.



(**a**) (**b**)

Figure 15. (**a**,**b**) Rolling element fault of VMD.

4.3. Performance Comparison of Fault Diagnosis Models

After decomposing each sample signal into several IMF components using VMD, the optimal IMF component is extracted based on the criterion of minimum envelope entropy. Subsequently, nine indicators are computed for each IMF component, including mean, variance, peak value, kurtosis, effective value, peak factor, pulse factor, waveform factor, and margin factor. These nine indicators are then used to construct the feature vector for each sample.

After extracting the feature vectors, each state consists of 400 samples, totaling 10 states, resulting in a matrix of size 4000 × 9 columns. Subsequently, the data are labeled according to the categories in Table 4. The numbers 1 through 10 represent different types of faults. The data in each state are divided into training and testing sets in a ratio of 7:1. Then, the labeled data are input into the fault diagnosis model for fault identification.

The objective of this section is to evaluate the performance differences among the newly developed CNN-BiLSTM diagnostic model and other existing techniques, including CNN-LSTM, BiLSTM, and a single CNN model. The specific test results are presented in Figure 16, which demonstrates that the use of a standalone CNN method performs poorly in terms of fault diagnosis accuracy, while both the BiLSTM and CNN-LSTM models achieve significantly higher accuracy compared to CNN alone. Particularly noteworthy is the superior performance of the CNN-BiLSTM model, which outperforms all other models with the highest accuracy of 96.6667%. In this comparative analysis, the CNN-BiLSTM model demonstrates outstanding performance, leading with the highest accuracy among the models and significantly enhancing the reliability of fault diagnosis.

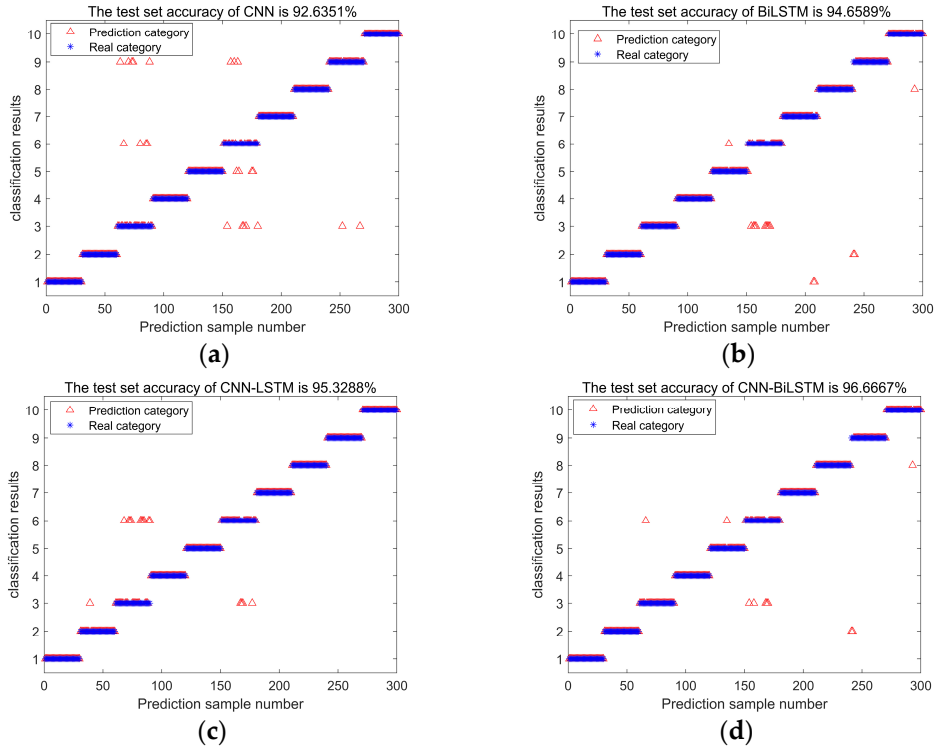


Figure 16. Experimental results of the fault diagnosis model (**a**–**d**) are CNN, BiLSTM, CNN-LSTM, and CNN-BiLSTM, respectively.

A comprehensive evaluation of the training duration and fault detection accuracy for these four models is presented in Table 6. The data in Table 6 show that the CNN model exhibits a relatively low accuracy only with the shortest training time. Both the BiLSTM and CNN-LSTM models achieve some success in improving diagnostic performance, with similar accuracy rates, although the CNN-LSTM model's training time is slightly longer than that of BiLSTM. However, the CNN-BiLSTM model stands out with the best

diagnostic performance, achieving an accuracy rate as high as 97.6667%. It is worth noting that due to the increased complexity of the network model caused by BiLSTM in extracting bidirectional time series features, the training efficiency is reduced. Therefore, considering both accuracy and training time, the CNN-BiLSTM model demonstrates the most outstanding performance in fault diagnosis.

Table 6. Dataset overview.

Damage Location		No.	Rolling Element			Inner Circle			Outer Circle			Load
Labeled damage diameter		1	2	3	4	5	6	7	8	9	10	
(inch)		0	0.007	0.014	0.021	0.007	0.014	0.021	0.007	0.014	0.021	
A	train	450	450	450	450	450	450	450	450	450	450	no
	test	50	50	50	50	50	50	50	50	50	50	
B	train	450	450	450	450	450	450	450	450	450	450	1
	test	50	50	50	50	50	50	50	50	50	50	
C	train	450	450	450	450	450	450	450	450	450	450	2
	test	50	50	50	50	50	50	50	50	50	50	
D	train	450	450	450	450	450	450	450	450	450	450	3
	test	50	50	50	50	50	50	50	50	50	50	

The comprehensive evaluation is based on the training duration and fault detection accuracy of these four models, and the specific comparison results are shown in Table 7. From the data presented in Table 7, it can be seen that the CNN model exhibits relatively low accuracy in model comparison and is characterized by the shortest training time. The BiLSTM and CNN-LSTM models have achieved certain success in improving diagnostic performance, with similar accuracy. However, the training time of the CNN-LSTM model is slightly longer than that of BiLSTM. However, the ICNN BiLSTM model stood out and demonstrated the best diagnostic performance. It is worth noting that BiLSTM increases the complexity of the network model during bidirectional time series feature extraction, resulting in a decrease in training efficiency. Therefore, considering both accuracy and training time, the ICNN BiLSTM model performs the best in fault diagnosis.

Table 7. Evaluation results of various fault diagnosis models.

Algorithm	Accuracy/%	Training Time/s
CNN	92.6351	10
BiLSTM	94.6589	11
CNN-LSTM	95.3288	13
CNN-BiLSTM	96.6667	14

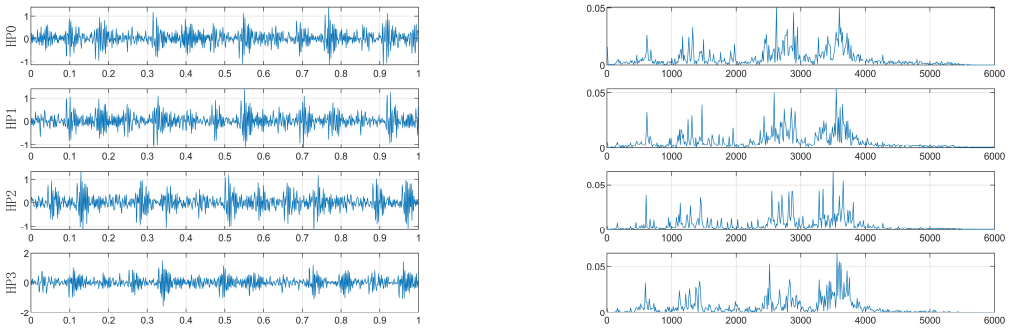
5. Analysis of Fault Diagnosis Results of Rolling Bearings Under Variable Load Conditions

5.1. Description of Variable Load Problem

In actual industrial operations, the load borne by equipment is often fluctuating, which results in corresponding changes in the signals captured by sensors. Therefore, the focus of this section is to collect operational data on bearings under different load conditions, aiming to demonstrate that the model can effectively apply to fault diagnosis when bearing load changes. Through this series of experiments, we will gain a more comprehensive understanding of the performance of the model under complex operating conditions, providing reliable references for practical applications.

Figure 17a depicts the vibration signal caused by inner ring faults under various load conditions and Figure 17b shows the frequency corresponding to the vibration signal. Through these graphs, it can be observed that due to different load conditions, vibration

signals differ in multiple dimensions, such as feature quantity, vibration intensity, periodic fluctuations, and phases, making it difficult to identify manually, especially in the process of feature extraction, which may lead to incorrect classification. This makes it particularly crucial to adopt convolutional neural network models with stronger feature extraction capabilities when facing complex conditions, which is expected to significantly improve the accuracy of fault detection.



(**a**) (**b**)

Figure 17. Time-frequency-domain waveform diagram of inner ring faulty bearing under different load states.

Therefore, training fault detection models based on vibration data of bearings under certain workload conditions is not only a major academic advancement but also opens up broad prospects in practical applications. This section focuses on the analysis of diagnostic results under different loads, specifically, the data corresponding to the same fault and different loads at Western Reserve University, identified as datasets A, B, C, and D in Table 6. By training the model on these four datasets and using the remaining datasets to test the efficiency of the model, we further explored the adaptability of the model. The experimental details for the corresponding variable load conditions are shown in Table 8.

Table 8. Description of the variable load experiment.

Experiment: Using Source Domain Signals for Training and Diagnosing Target Domain Signals				
Field type Signal description	Source domain dataset Labeled training data	Target domain dataset Untamed test data		
Dataset	Dataset A	Dataset B	Dataset C	Dataset D
	Dataset B	Dataset A	Dataset C	Dataset D
	Dataset C	Dataset A	Dataset B	Dataset D
	Dataset D	Dataset A	Dataset B	Dataset C

5.2. Analysis of Diagnostic Results Under Variable Load Conditions

As shown in Figure 18, The average diagnostic rate of the CNN (1) algorithm is the worst, all below 90%. The BiLSTM (2) algorithm performs well, with a diagnostic rate of around 90%. After comparing the structures of CNN-LSTM (3) and CNN BiLSTM (4), it was found that both algorithms can achieve a fault detection rate of approximately 95%, with the latter method showing a slight advantage. Especially when dataset B is used as the training set and tested on dataset C, this new diagnostic method improves recognition efficiency by about 10% compared to the ordinary BiLSTM (2) algorithm. This comparison confirms that our proposed fault diagnosis model, ICNN BiLSTM, has better adaptability than traditional fault diagnosis techniques and other popular deep learning methods when processing data under different workload conditions.

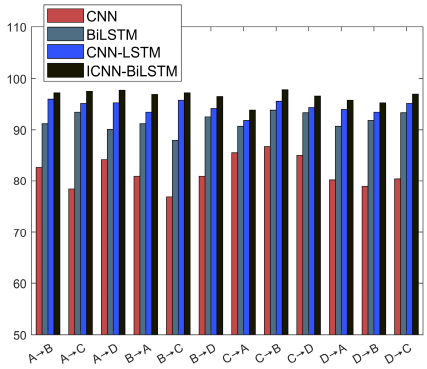


Figure 18. Accuracy rates of various algorithms under different load conditions.

5.3. Simulation Analysis of the OCSSA-VMD Algorithm

To further analyze the algorithm performance, analysis validation was performed on simulated EMD signals and VMDs set using this parameter. The results indicate that VMD processing optimized by OCSSA can significantly improve signal analysis performance.

This article initially created a simulation signal to test the effectiveness of the proposed algorithm. This simulated signal combines three components from different frequency domains, with their frequencies set to f_1 = 3 Hz, f_2 = 29 Hz, and f_3 = 352 Hz. In order to be closer to practical application scenarios, we also added a layer of Gaussian white noise. The composition formula of the signal is as follows:

$$\mathrm{F} = \frac{1}{2}\cos(2\pi f_1 t) + \frac{1}{4}\cos(2\pi f_2 t) + \frac{1}{8}\cos(2\pi f_3 t) + \xi \tag{7}$$

Figure 19 vividly illustrates the decomposition effect of simulated signals using the EMD algorithm. The time-domain and frequency-domain graphs on both sides clearly present the time–frequency characteristics of each IMF component. The original simulation signal is located at the top of the image, and the following layers are the specific IMF components. However, in analyzing the decomposition results generated by EMD, some components exhibit modal aliasing issues. Specifically, the frequency information displayed by the first modal function (IMF1) and the third modal function (IMF3) are relatively clear and distinct; however, the fourth and fifth modal functions (IMF4 and IMF5) show an overlap between frequency components. In addition, the second mode function (IMF2) fails to clearly show its frequency information.

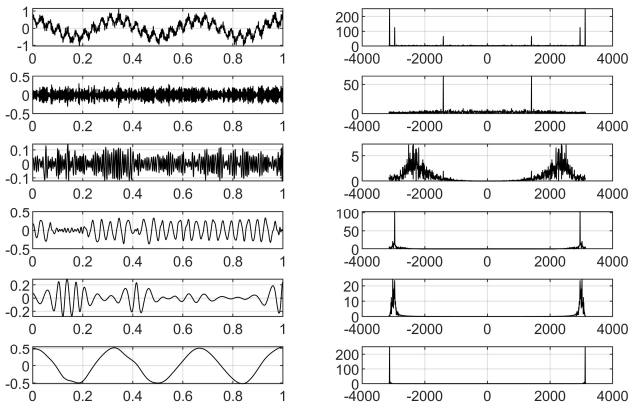


Figure 19. EMD diagram of the simulated signal.

The VMD algorithm was used to decompose the same signal, and the result is shown in Figure 20. The top part is the original simulation signal, followed by the IMF component. The VMD algorithm sets the decomposition level to 3 and the penalty factor to 1500. Compared with the EMD algorithm, it can be found that there is a significant decrease in

the modal aliasing phenomenon in the frequency-domain graph. Moreover, low, medium, and high frequencies are clearly displayed in the IMF component. This indicates that the VMD algorithm has stronger signal processing capabilities compared to the END algorithm.

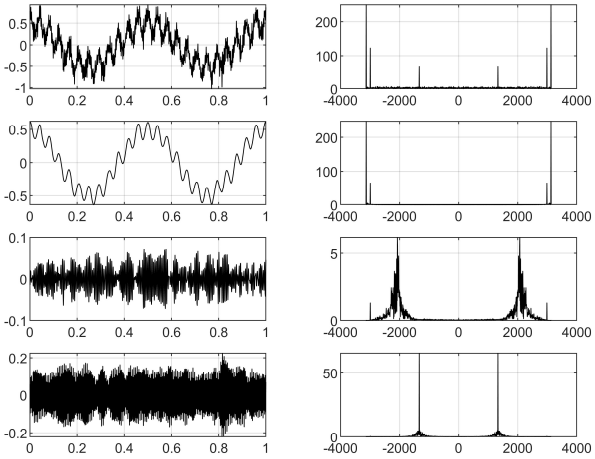


Figure 20. VMD diagram of the simulated signal.

Manually setting decomposition levels and penalty parameters in the previously applied VMD algorithm may have an impact on the quality of signal decomposition. This article uses the OCSSA to improve the core parameters of the VMD algorithm and achieve the optimal combination ratio search. This improvement strategy has enabled the VMD algorithm to be efficiently applied in vibration data processing, and the final confirmed $[K, \alpha]$ optimal ratio corresponds to a value of $[4, 1500]$. The actual decomposition effect is shown in Figure 21.

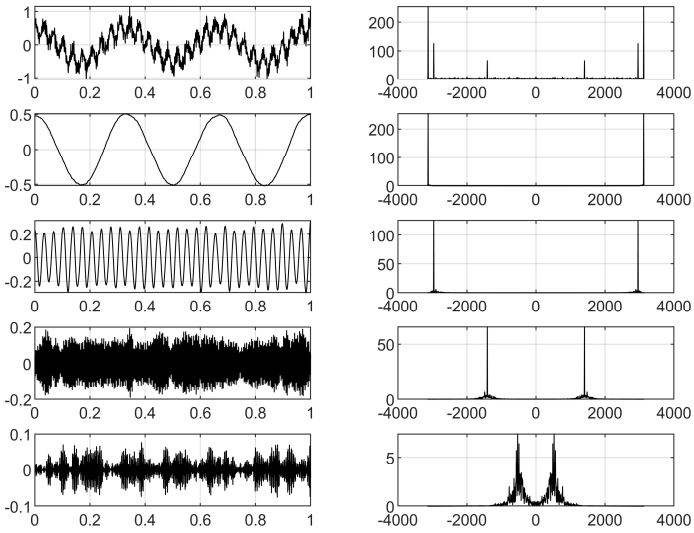


Figure 21. VMD diagram for parameter K = 4.

According to the data analysis in Figure 21, setting the decomposition level to 4 and adjusting the penalty factor to 1500, the simulated signal showed excellent decomposition performance on the IMF component, effectively avoiding mode mixing and demonstrating robustness and good decomposition ability. The optimal parameter combination obtained by optimizing the key parameter K value and penalty factor of VMD through the OCSSA performs well in signal decomposition. Overall, through the parameter optimization of the OCSSA, VMD exhibits better performance in signal decomposition, especially in avoiding mode aliasing, with significant advantages.

6. Conclusions

(1) The traditional Sparrow Search Algorithm (SSA) is enhanced by incorporating the Tent chaotic map, the Osprey Optimization Algorithm (OOA), and the Cauchy mutation strategy. This enhancement results in the development of the Osprey–Cauchy–Sparrow Search Algorithm (OCSSA), which is utilized to optimize the parameters of the Variational Mode Decomposition (VMD) algorithm.

(2) A feature extraction method based on OCSSA-VMD for bearing vibration signals is proposed. Experimental results demonstrate that OCSSA-VMD does not suffer from mode mixing issues or false components in the decomposition of bearing fault signals. This method is highly beneficial for extracting fault information from the original data.

(3) Based on the Sparrow Search Algorithm, this article combines Tent mapping, Osprey Optimization Algorithm, and the Cauchy mutation technique to form the OCSSA optimization algorithm, which improves the efficiency and accuracy of optimization. The VMD algorithm optimized by OCSSA parameters shows significant improvements in both time and frequency domains when processing bearing fault signals, making the characteristic frequencies of the signals more prominent. This article also confirms the effectiveness of the proposed algorithm in reducing signal mode overlap and improving interference resistance by comparing it with EMD and VMD algorithms with different configurations.

(4) The CNN-BiLSTM fault diagnosis model is employed to recognize the input feature vector dataset. Comparative experiments show that the accuracies of the CNN, BiLSTM, CNN-LSTM, and CNN-BiLSTM fault diagnosis models are 92.64%, 94.66%, 95.33%, and 96.67%, respectively.

The VMD and convolutional neural network models developed in this article are initially limited to the detection of single bearing faults, which makes their application scope relatively narrow. In the future, its application can be gradually expanded by considering the use of federated learning and transfer learning to cover more mechanical components and composite fault diagnosis.

Author Contributions: Z.X.: initial draft writing, algorithm research and design, data analysis, and manuscript analysis. H.J.: data collection, data-assisted analysis, and software-assisted design. D.W.: data-assisted analysis, chart creation, literature search, and manuscript revision. X.W.: research and design guidance, algorithm rationality review, fund management, and finalization. K.W.: researching ideas, algorithm recommendations, and project management. 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 numbers 61875197 and 61205139] and the Youth Innovation Promotion Association of the Chinese Academy of Sciences [grant number 2016168].

Data Availability Statement: We did not create any new data.

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

References

1. Li, Z.; Jiang, Y.; Hu, C.; Peng, Z. Recent progress on decoupling diagnosis of hybrid failures in gear transmission systems using vibration sensor signal: A review. *Measurement* **2016**, *90*, 4–19. [CrossRef]
2. He, D.; Wang, X.; Li, S.; Lin, J.; Zhao, M. Identification of multiple faults in rotating machinery based on minimum entropy deconvolution combined with spectral kurtosis. *Mech. Syst. Signal Process.* **2016**, *81*, 235–249. [CrossRef]
3. Randall, R.B.; Antoni, J. Rolling element bearing diagnostics—A tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520. [CrossRef]
4. Wang, Y.; Liang, M. Identification of multiple transient faults based on the adaptive spectral kurtosis method. *J. Sound. Vib.* **2012**, *331*, 470–486. [CrossRef]
5. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc.* **1998**, *454*, 903–993. [CrossRef]
6. 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]
7. Lei, Y.; He, Z.; Zi, Y. Application of the EEMD method to rotor fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2009**, *23*, 1327–1338. [CrossRef]

8. Li, G.; Tang, G.; Luo, G.; Wang, H. Underdetermined blind separation of bearing faults in hyperplane space with variational mode decomposition. *Mech. Syst. Signal Process.* **2019**, *120*, 83–97. [CrossRef]
9. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process. A Publ. IEEE Signal Process. Soc.* **2014**, *62*, 531–544. [CrossRef]
10. Yang, K.; Wang, G.; Dong, Y.; Zhang, Q.; Sang, L. Early chatter identification based on an optimized variational mode decomposition. *Mech. Syst. Signal Process.* **2019**, *115*, 238–254. [CrossRef]
11. Zhan, Y.; Mechefske, C.K. Robust detection of gearbox deterioration using compromised autoregressive modeling and Kolmogorov-Smirnov test statistic-Part I: Compromised autoregressive modeling with the aid of hypothesis tests and simulation analysis. *Mech. Syst. Signal Process.* **2007**, *21*, 1953–1982. [CrossRef]
12. Pan, J.; Qu, L.; Peng, K. Deep residual neural-network-based robot joint fault diagnosis method. *Sci. Rep.* **2022**, *12*, 17158. [CrossRef] [PubMed] [PubMed Central]
13. Wang, R.; Feng, Z.; Huang, S.; Fang, X.; Wang, J. Research on Voltage Waveform Fault Detection of Miniature Vibration Motor Based on Improved WP-LSTM. *Micromachines* **2020**, *11*, 753. [CrossRef] [PubMed] [PubMed Central]
14. Fu, G.; Wei, Q.; Yang, Y. Bearing fault diagnosis with parallel CNN and LSTM. *Math. Biosci. Eng.* **2024**, *21*, 2385–2406. [CrossRef] [PubMed]
15. Cheng, Y.; Wang, S.; Chen, B.; Mei, G.; Zhang, W.; Peng, H.; Tian, G. An improved envelope spectrum via candidate fault frequency optimization-gram for bearing fault diagnosis. *J. Sound Vib.* **2022**, *523*, 116746. [CrossRef]
16. Lian, J.; Liu, Z.; Wang, H.; Dong, X. Adaptive variational mode decomposition method for signal processing based on mode characteristic. *Mech. Syst. Signal Process.* **2018**, *107*, 53–77. [CrossRef]
17. Xue, J.; Shen, B. A novel swarm intelligence optimization approach: Sparrow search algorithm. *Syst. Sci. Control. Eng.* **2020**, *8*, 22–34. [CrossRef]
18. Dehghani, M.; Trojovský, P. Osprey optimization algorithm: A new bio-inspired metaheuristic algorithm for solving engineering optimization problems. *Front. Mech. Eng.* **2023**, *8*, 1126450. [CrossRef]
19. Wang, J.; Guo, S. Application of Adaptive VMD Algorithm in Fault Diagnosis of Rolling Bearings. *Mech. Electr. Eng. Technol.* **2020**, *49*, 161–164.
20. Miao, Y.; Zhao, M.; Lin, J. Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition. *ISA Trans.* **2019**, *84*, 82–95. [CrossRef] [PubMed]
21. Dibaj, A.; Hassannejad, R.; Ettefagh, M.M.; Ehghaghi, M.B. Incipient fault diagnosis of bearings based on parameter-optimized VMD and envelope spectrum weighted kurtosis index with a new sensitivity assessment threshold. *ISA Trans.* **2021**, *114*, 413–433. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.