lubricants

MDPI

Article

Intelligent Workshop Bearing Fault Diagnosis Method Based on Improved Convolutional Neural Network

Xuan Su [1,2], Jitai Han [1,*], Chen Chen [3,*], Jingyu Lu [3], Weimin Ma [1] and Xuesong Dai [1]

[1] School of Automation, Wuxi University, Wuxi 214105, China
[2] Wuxi Key Laboratory of Intelligent Manufacturing Technology for Core Components of High-End Equipment, Wuxi 214105, China
[3] School of Intelligent Manufacturing, Jiangnan University, Wuxi 214122, China
* Correspondence: 860085@cwxu.edu.cn (J.H.); 8202401021@jiangnan.edu.cn (C.C.)

Abstract

A bearing intelligent fault diagnosis method based on an improved convolutional neural network is proposed to address the problems of high noise, difficult fault feature extraction, and low fault diagnosis recognition rate in rolling bearing vibration signals collected under complex working conditions. Firstly, in the data preprocessing stage, the wavelet denoising method is used to preprocess the data to obtain higher-quality signals. Then, the convolutional neural network LeNet-5 model was improved through batch normalization, Dropout, and L2 regularization methods. The wavelet denoised signal was input into the optimized LeNet-5 model to achieve more accurate fault diagnosis output for rolling bearings. Finally, to demonstrate the generalization ability of the model, this paper uses publicly available rolling bearing data from a university as the dataset and conducts experimental verification of the model using MATLAB-2023b software under different loads. The experimental results show that the improved neural network model has a fault diagnosis accuracy of 94.27%%, which is 17.84% higher than the traditional neural network model in terms of accuracy. Moreover, for different loads, the improved convolutional neural network model still maintains good fault diagnosis accuracy.

Keywords: rolling bearings; fault diagnosis; improving convolutional neural networks



Received: 3 November 2025
Revised: 19 November 2025
Accepted: 27 November 2025
Published: 30 November 2025

Citation: Su, X.; Han, J.; Chen, C.; Lu, J.; Ma, W.; Dai, X. Intelligent Workshop Bearing Fault Diagnosis Method Based on Improved Convolutional Neural Network. *Lubricants* **2025**, *13*, 521. https://doi.org/10.3390/lubricants13120521

Copyright: © 2025 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

Currently, as the Made in China 2025 process continues to accelerate, China is emerging as a manufacturing powerhouse. Rolling bearings, as an essential basic component, are widely used in various mechanical devices in the manufacturing industry. During long-term operation of the equipment, rolling bearings are susceptible to various external factors, which can cause faults in the inner ring, rolling elements, or outer ring, thereby affecting the normal operation of the equipment. Therefore, in-depth exploration of intelligent diagnostic methods for rolling bearing faults is crucial for improving the operational efficiency of mechanical equipment and ensuring production safety [1].

In the field of mechanical operation monitoring, vibration signals contain rich and comprehensive information about the operating status of equipment [2]. However, such signals often exhibit complex nonlinear and non-stationary characteristics, which pose significant challenges for signal analysis and fault diagnosis. Especially in the early stages of bearing operation, due to the extremely small defects on the bearing raceway or rolling element surface, the fault characteristic signals generated are often weak. Based on case

studies of CWRU and XJTU bearings, Zhou et al. [3] found that diagnostic accuracy dropped significantly and prediction uncertainty surged by more than 10% under strong noise interference, confirming that noise is a key factor leading to performance gaps. Under the interference of a strong noise environment, these key fault features are easily masked by noise, making it difficult to effectively extract fault features with diagnostic value from vibration signals [4]. From the perspective of time-domain analysis, typical time-domain statistical features provide an intuitive and effective way to describe the dynamic characteristics of vibration signals [5,6]. Sahu P K et al. [7] Improved denoising techniques based on Complete Ensemble Empirical Mode Decomposition (CEEMD) to enhance the early fault detection performance of bearings under strong noise. Wavelet analysis method, as the most commonly used tool in the field of signal processing, is widely applied in the data processing stages of various fields [8,9]. Although EMD is widely used for adaptive decomposition of vibration signals, it was not adopted in the final framework due to the problems of mode mixing and reconstruction instability under strong noise conditions. However, it has an important reference value as a background technology.

In recent years, Convolutional Neural Networks (CNNs) have been widely used in the field of deep learning due to their powerful feature learning ability and ability to process complex data, which has greatly promoted the research and development of fault diagnosis methods for rolling bearings. However, the application of this method in the field of fault diagnosis presents challenges in terms of computational efficiency, effectiveness of initial feature extraction, and difficulty in capturing time-domain features. In terms of innovative fault diagnosis methods based on convolutional neural networks, numerous scholars have conducted in-depth research and achieved a series of results. Yu et al. [10] proposed a Convolutional Neural Network (CNN-GAM) equipped with a global attention module, which enhances feature representation, improves computational efficiency, and optimizes attention mechanisms. Sabyasachi et al. [11] proposed an ensemble parameter learning method and applied it as a structured function of a filter bank to a CNN architecture, solving the problem of "insufficient initial feature extraction or dependence on manual design" in model limitations. Dai et al. [12] introduced the fast Fourier transform method into neural networks to capture specific fault information accurately. Li et al. [13] used wide kernel convolutional layers to preprocess the original signal, achieving the goal of data dimensionality reduction and feature channel expansion. He [14] proposed a CNN-LSTM diagnostic method driven by fused wavelet time-frequency graph features, which solves the problems of "insufficient information in a single domain" and "ignoring signal temporal dynamics" in complex and variable operating conditions challenges. Chen et al. [15] proposed a wind turbine bearing fault diagnosis model based on efficient cross space multi-scale CNN transformer parallel connection, which solves the problem of multi-scale feature extraction in model limitations.

Significant progress has also been made in the application of graph models in the field of fault diagnosis in recent years. The graph model can effectively extract spatiotemporal features from one-dimensional time-domain signals by converting them into the spectral domain [16,17]. However, the application of this method in the field of fault diagnosis presents challenges in effectively joint extraction of spatiotemporal features, coupling and hiding of fault information in the spatiotemporal dimension, and feature extraction of non-stationary signals. Specifically, Sun et al. [18] adopted a two-stage framework for bearing fault diagnosis, using a graph model to identify the operating state of the bearing during the detection stage, improving the robustness of diagnosis under complex and variable operating conditions. Wang et al. [19] proposed a bearing fault detection and diagnosis method based on spatiotemporal graphs, which deeply explored the ability of graph models to extract fault information hidden in spatial form and temporal dynamics,

and solved the problem of coupling and hiding fault information in the spatiotemporal dimension under complex and variable operating conditions. To extract the time-frequency distribution of the original signal, Tao et al. [20] used the short-time Fourier transform to transform the original signal from the time domain to the time-frequency domain, solving the problem of feature extraction of non-stationary signals under noise interference.

However, in practical applications, fault diagnosis of rolling bearings poses many challenges. External factors such as changes in operating conditions and environmental fluctuations may cause significant nonlinear fluctuations in the characteristic curve. This fluctuation leads to unclear, weak, and even complete disappearance of initial symptoms of the fault, thereby increasing the risk of misjudgment and missed detection in the fault detection process. Entropy, as an indicator for measuring the uncertainty or complexity of data sequences [21], can effectively describe nonlinear dynamic characteristics and has important application value in fault diagnosis. For example, sample entropy [22], fuzzy entropy [23], etc., distinguish the fault state of bearings by estimating the complexity of time-domain signals. However, there are challenges in applying this method to the field of fault diagnosis in terms of feature evaluation indicators and model structure design. Wang et al. [24] considered frequency distribution and its amplitude changes in the entropy calculation process, and proposed a cumulative spectral distribution entropy for rotating machinery fault diagnosis. The entropy measure was extended to the frequency domain to solve the problem of difficult quantification and identification of fault features in the presence of noise interference and weak background features. Wang et al. [25] proposed a multi-branch convolutional network, where each branch focuses on a different feature subspace and is fused to obtain diverse feature representations. Chen et al. [26] proposed a multi-scale convolutional neural network with feature alignment to address the issue of data distribution differences in fault diagnosis of rolling bearings, such as changes in working conditions and equipment. The network extracts features of different scales from each branch and fuses them into a global feature representation. He et al. [27] proposed a Multi-scale Mixed Convolutional Neural Network (MSMCNN) for fault diagnosis of industrial robot harmonic reducers under complex working conditions, aiming to improve the model's ability to extract fault features and diagnostic performance of harmonic reducers. Choudhary et al. [28] used a multi-input convolutional neural network to fuse vibration and acoustic signals, achieving fault diagnosis under different operating conditions and solving the problem of limited fault diagnosis under a single data source.; Li et al. [29] proposed a Transformer based on variational attention to establish causal relationships between signal patterns and fault types, improving the ability to focus on key features; Wu et al. [30] combined GAF-MAT with Transformer to reduce the influence of inter-sample time shift and improve the diagnostic accuracy of the model; Yao et al. [31] studied the acoustic signals of planetary gearboxes and combined Fourier decomposition with energy, time-frequency kurtosis, and random forest to achieve a balance between high-precision diagnosis and computational efficiency under limited sample conditions.

This article focuses on the significant demand for intelligent operation and maintenance of rotating machinery in the context of Industry 4.0. Taking rolling bearing fault diagnosis as the starting point, a bearing intelligent fault diagnosis method based on an improved convolutional neural network is proposed. Three aspects of research are systematically carried out: firstly, in the data preprocessing stage, a nonlinear mapping relationship between time-frequency domain features and fault modes is established through the joint optimization of wavelet threshold denoising and empirical mode decomposition; Secondly, at the level of model optimization, an innovative multi-dimensional regularization collaborative mechanism is proposed, which includes batch normalization, Dropout random deactivation, and L2 weight decay; Thirdly, in the verification stage, data preprocessing

experiments were designed for analysis, network structure modification impact analysis, and different load experiments were conducted to verify the generalization ability of the experimental model, providing theoretical support for rotating machinery fault prediction and health management. The relevant methodology can be extended to intelligent diagnosis scenarios for complex equipment such as gearboxes and turbines.

2. Data Preprocessing

In the field of fault diagnosis, existing research mostly focuses on algorithm validation under standardized experimental datasets. Although these methods demonstrate excellent diagnostic performance under ideal conditions, the equipment operation data in real industrial scenarios often accompanies significant noise interference. Therefore, it is necessary to study signal preprocessing methods for the raw data collected in actual environments to improve the engineering application value of fault diagnosis methods. Signal denoising is a commonly used signal preprocessing method that aims to reduce or eliminate noise as much as possible while preserving the authenticity of the original signal, in order to obtain higher-quality signals and build a robust data foundation for subsequent feature extraction and pattern recognition.

2.1. Wavelet Transform

The wavelet transform is a time-frequency analysis method that decomposes signals into components of different scales. Let $\psi(t) \in L^2(R)$ satisfy the condition:

$$\int_R |\psi(\omega)|^2 |\omega|^{-1} d\omega < +\infty \tag{1}$$

In the equation, $\psi(\omega)$ is the Fourier transform of $\psi(t)$, which is a wavelet mother function or fundamental wavelet. By scaling and translating the wavelet function, we can obtain:

$$\psi_{\alpha,\tau}(t) = \alpha^{-\frac{1}{2}} \psi\left(\frac{t-\tau}{\alpha}\right) \alpha > 0, \tau \in R \tag{2}$$

In the formula, α is the scaling factor and τ is the ranslation factor. $\psi_{\alpha,\tau}(t)$ is a wavelet basis function that depends on α and τ, and they are a sequence of functions obtained by stretching and translating a set of mother functions $\psi(t)$ [2].

If there exists a signal $x(t) \in L^2(R)$, and it is the inner product with the wavelet basis function, we can obtain:

$$WT_\chi(\alpha,\tau) = \langle \chi, \psi_{\alpha,\tau} \rangle = \alpha^{-\frac{1}{2}} \int_R \chi(t) \psi_{\alpha,\tau}\left(\frac{t-\tau}{\alpha}\right) dt \tag{3}$$

Equation (3) is the formula for the continuous wavelet transform of $x(t)$, known as the Continuous Wavelet Transform (CWT).

2.2. Wavelet Denoising

Wavelet denoising is a signal processing technique that uses wavelet transform technology to eliminate interference noise in the original signal. Assuming that the signal x (t) is contaminated by noise and becomes s (n), its basic noise model is as follows:

$$s(t) = \chi(t) + \sigma e(t) \tag{4}$$

In the equation, $e(t)$ is the noise of the signal, and σ is the noise intensity of the signal.

In signal denoising, this study uses db4 as the mother wavelet to decompose the vibration signal into three layers, effectively representing both temporal abrupt changes

and frequency domain details. To address the differences in noise distribution across sub-bands at each scale, the WDCBM method is used to calculate layered thresholds, ensuring that the threshold values reflect the statistical characteristics of wavelet coefficients at different scales. Simultaneously, a soft threshold is used to nonlinearly compress the wavelet coefficients, avoiding the coefficient discontinuity issues that may arise from a hard threshold. Finally, inverse wavelet reconstruction significantly reduces high-frequency random noise and effectively enhances fault characteristic components, improving the stability of subsequent feature extraction and the quality of time-frequency analysis.

3. GADF Rolling Bearing Fault Feature Transformation

The Gram angle field undertakes the task of time series encoding in polar coordinate systems. Given the data collected by the sensor $X = \{x_1, x_2, \ldots, x_n\}$, by scaling X so that all values fall within the interval $[-1, 1]$, the calculation formula is shown in Equation (5) below.

$$\widetilde{\chi}_{-1}^{i} = \frac{(\chi_i - \max(X)) + (\chi_i - \min(X))}{\max(X) - \min(X)} \tag{5}$$

Convert the calculation result of the above formula into polar coordinates, and then encode it in polar coordinates. Encode the scaled X value using the inverse cosine formula so that its value falls between $[0, \pi]$, and its corresponding timestamp is encoded as the radius, as shown in (6) below.

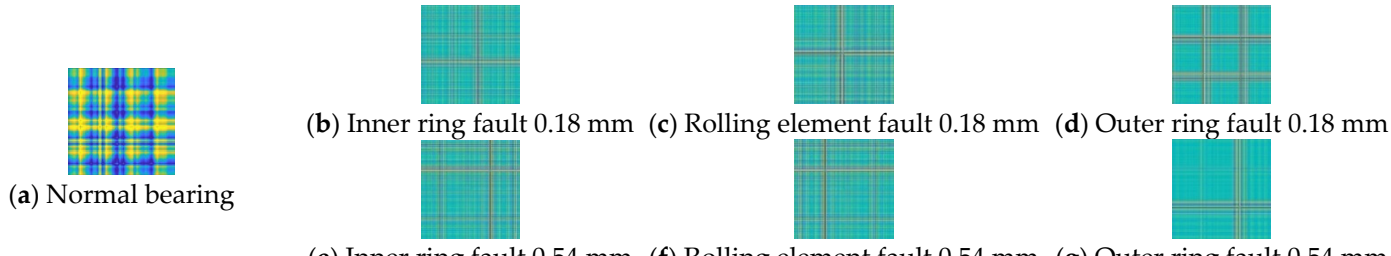
$$\begin{cases} \vartheta = \arccos(\widetilde{\chi}_i), -1 \leq \widetilde{\chi}_i < 1, \widetilde{\chi}_i \in \widetilde{X} \\ r = \frac{t_i}{N}, t_i \in R \end{cases} \tag{6}$$

In the equation, t_i is the timestamp, and N is the constant factor of the regularized polar coordinate system generated space.

The GADF calculation formula is shown in Equation (7) below.

$$GADF = \left[\sin\left(\theta_i - \theta_j\right)\right] \tag{7}$$

This paper employs GADF to effectively represent the phase and periodic evolution characteristics of time series in two-dimensional space, thereby enhancing the adaptability of feature representation to deep models. The temporal resolution and window length are set based on the sampling frequency and the fault feature period to ensure sufficient coverage of key information. Although polar coordinate mapping introduces finite amplitude perturbations, it maintains good preservation of the fault-related phase structure and does not affect subsequent classification and discrimination performance. The GADF encoding method converts the image into a two-dimensional image, and the fault feature map is shown in Figure 1.



(**a**) Normal bearing (**b**) Inner ring fault 0.18 mm (**c**) Rolling element fault 0.18 mm (**d**) Outer ring fault 0.18 mm

(**e**) Inner ring fault 0.54 mm (**f**) Rolling element fault 0.54 mm (**g**) Outer ring fault 0.54 mm

Figure 1. Fault characteristic diagram.

4. Design of Convolutional Neural Network

LeNet-5 has one input layer responsible for inputting feature images, two convolutional layers responsible for extracting image features, two pooling layers responsible for reducing the dimensionality of feature maps, two fully connected layers, and one output layer responsible for mapping features to corresponding categories to complete classification operations. After undergoing multiple fine adjustments in convolutional and pooling layers, image data can be transformed into informative feature data.

4.1. Classic LeNet-5 Network Model

The structure of the LeNet-5 model proposed by LeCunundefined is shown in Figure 2 [32], which consists of two convolutional and pooling combination layers, two fully connected layers, and one output layer.

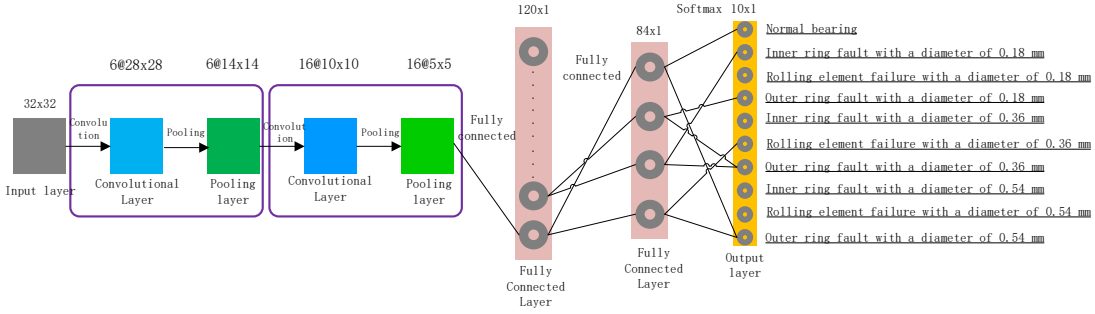


Figure 2. Structure diagram of classic convolutional neural network LeNet-5. (1: Normal bearing; 2: Inner ring fault with a diameter of 0.18 mm; 3: Rolling element failure with a diameter of 0.18 mm; 4: Outer ring fault with a diameter of 0.18 mm; 5: Inner ring fault with a diameter of 0.36 mm; 6: Rolling element failure with a diameter of 0.36 mm; 7: Outer ring fault with a diameter of 0.36 mm; 8: Inner ring fault with a diameter of 0.54 mm; 9: Rolling element failure with a diameter of 0.54 mm; 10: Outer ring fault with a diameter of 0.54 mm.)

4.1.1. Convolutional Layer

The convolutional layer constitutes the key part of a convolutional neural network, which performs sliding convolution operations on the input and kernel according to a set step size, thereby extracting local features of the input and generating feature maps. Due to the varying degrees of detail and edge contours present in images at different scales, a multi-resolution model was employed for processing. By activating the function, the output data of the convolution operation can be obtained. Usually, the mathematical description of convolutional layers can be presented through Formula (8).

$$\chi_j^l = f\left(\sum_{i \in M_j} x_j^{l-1} * k_{ij}^l + b_j^l\right) \tag{8}$$

In the formula, * represents a convolution operation, M_j represents the selection of input mapping, l is the l-th layer in the network, k is the kernel matrix of size s $\times$ s, and f is the nonlinear activation function.

4.1.2. Pooling Layer

In the application of convolutional neural networks, the pooling operation mainly calculates values that can represent specific points in a particular region of the feature map. Max-pooling and average-pooling are common downsampling methods. In contrast, max-pooling, by selecting the feature value with the highest response in a local sub-region, effectively preserves abrupt changes, impulses, and other high-energy anomalies in the

fault signal, regardless of their specific location within the sub-region. Average-pooling, on the other hand, smooths out local differences, potentially weakening diagnostically significant peak features. Given that mechanical fault features typically exhibit local energy spikes, this paper employs max-pooling to better preserve key discriminative information.

$$y_{k_{ij}} = \max_{(p,q)\in R_{ij}} x_{k_{pq}} \tag{9}$$

In the formula, $y_{k_{ij}}$ represents the maximum pooling output value related to the k-th feature map in the matrix region R_{ij}, and $x_{k_{pq}}$ represents the element located at (p, q) in the matrix region.

4.1.3. Activation Function

The activation function is a specific function executed on neurons that maps the input neurons to the output function. In convolutional neural networks, the ReLU function is often used as an activation function. For input feature vectors or feature maps, the system can reset factors that are less than zero, while keeping the remaining factors in their unchanged state. The ReLu function expression is as follows:

$$f(x) = \begin{cases} 0, x < 0 \\ x, x \geq 0 \end{cases} \tag{10}$$

4.1.4. Fully Connected Layer

The fully connected layer tightly connects all neurons in the front and back layers, thereby transforming the two-dimensional image of the previous layer into a one-dimensional vector. The output layer uses the Softmax function as the activation function. The Softmax function takes a vector of any real value and compresses it between 0 and 1. The Softmax function is defined as follows:

$$\sigma(z)_j = \frac{e^{e_j}}{\sum_{K=1}^{K} e^{z_k}}, \text{for } j = 1, \cdots, K \tag{11}$$

4.2. Improving the LeNet-5 Network Model

This article optimizes the network structure based on the classic LeNet-5 model structure, and the optimized network model is shown in Figure 3. Improving the model can enhance the accuracy of fault diagnosis and recognition by adjusting the size and number of convolution kernels, as well as optimizing the algorithm.

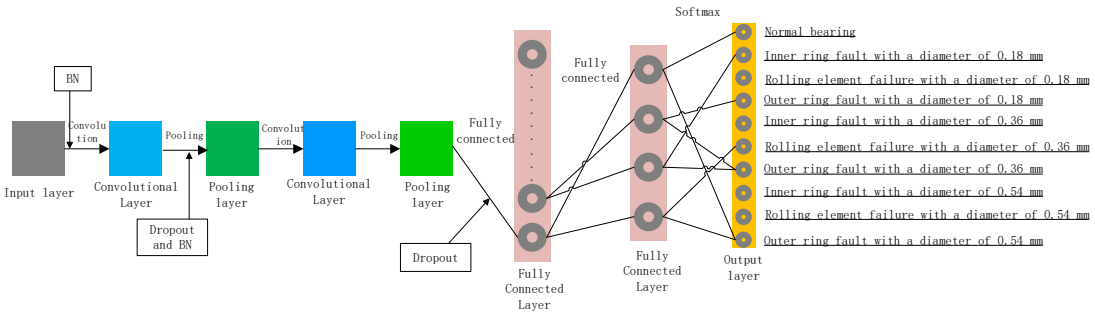


Figure 3. Improved LeNet-5 Network Structure. (1: Normal bearing; 2: Inner ring fault with a diameter of 0.18 mm; 3: Rolling element failure with a diameter of 0.18 mm; 4: Outer ring fault with a diameter of 0.18 mm; 5: Inner ring fault with a diameter of 0.36 mm; 6: Rolling element failure with a diameter of 0.36 mm; 7: Outer ring fault with a diameter of 0.36 mm; 8: Inner ring fault with a diameter of 0.54 mm; 9: Rolling element failure with a diameter of 0.54 mm; 10: Outer ring fault with a diameter of 0.54 mm.)

4.2.1. Network Structure Optimization

This article made the following three network structure optimizations when constructing the neural network structure of CNN.

Add batch normalization program segment: The function of BN is to forcibly pull the input values of any neuron in each layer of the neural network back to the standard normal distribution with a mean of 0 and a variance of 1. This essentially forces the gradually deviating distribution back to the relatively standard distribution. Preliminary experiments comparing different positions of BN before and after the activation function revealed that placing BN before the activation function resulted in more stable training and better performance, while placing it after the activation function led to unstable feature distribution. Therefore, this study placed BN before the activation function. The specific steps of BN are shown in Figure 4.

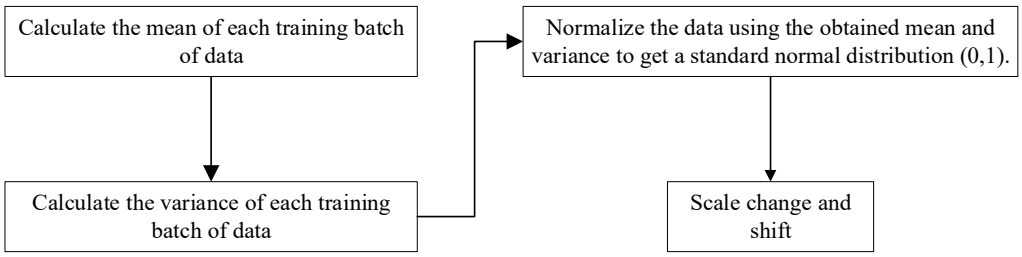


Figure 4. BN operation steps diagram.

Add a discard layer: Dropout is a method used in deep learning, and the purpose of regularization is to reduce the risk of overfitting in the model. In the training phase of the neural network, the Dropout layer randomly resets the output of some neurons to zero, simplifying the computational burden and improving the model's generalization performance, effectively preventing overfitting. Based on preliminary experimental results, after comparing different settings such as 10%, 20%, 30%, and 50%, it was found that too low a ratio resulted in insufficient regularization, while too high a ratio could easily lead to the loss of key features. Therefore, this study sets the Dropout ratio to 20%, which effectively achieves the best balance between preventing model overfitting and preserving its expressive power. The specific implementation of this method is shown in Figure 5. During network training, 20% of neurons are randomly selected and their outputs are set to zero.

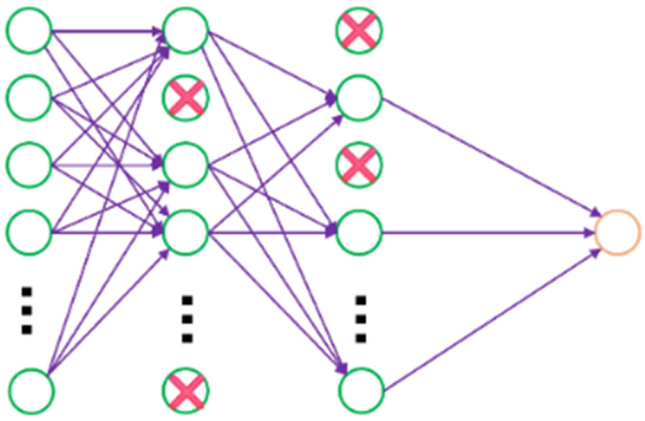


Figure 5. Schematic diagram of adding Dropout operation.

Add L2 Regularization to the training parameter settings: The L2 regularization method reduces the weight by reducing the weight to avoid excessive weight, thereby

reducing weight. Even if the weights are smoothed, it simplifies the model to some extent, speeds up training, and avoids overfitting of the model.

4.2.2. Improving Fault Diagnosis Methods

The specific steps of the improved fault diagnosis method are as follows, with the flowchart shown in Figure 6 and the algorithm pseudocode shown in Table 1.

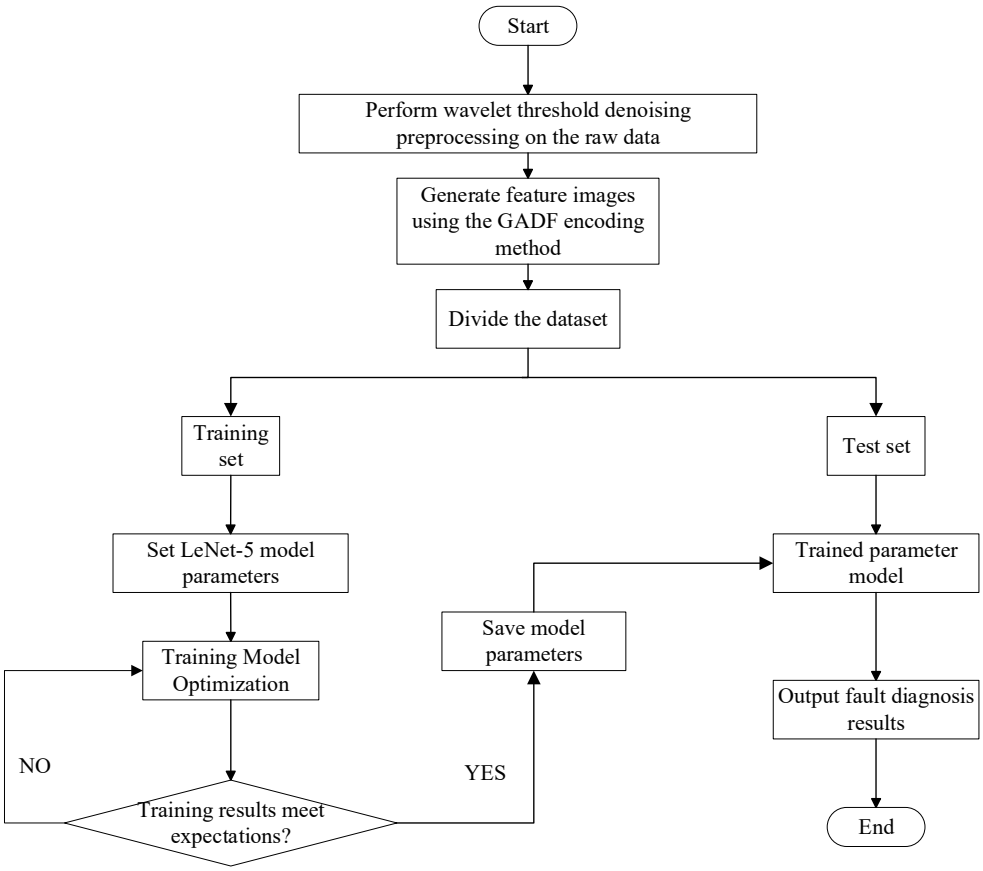


Figure 6. Fault diagnosis flowchart.

Table 1. Optimization of LeNet-5 Model Algorithm Pseudo code.

Optimize LeNet-5 Model

Input: Feature image set (a + b)
Output: Confusion Matrix (C)
Algorithm process:
While running LeNet-5 fault classification model do
  Set Train (a); Test (b)
  For $C > 90\%$
  For $C \leq 90\%$
  If the training parameters have L2 regularization, then
  Add Batch Normalization between Convolutional Layer 1 and Pooling Layer 1
  Set Dropout Layer (0.2) before the fully connected layer 1
  Else
  Add L2 regularization to the training parameters
  Add Batch Normalization between Convolution Layer 1 and Pooling Layer 1
  Add Dropout Layer (0.2) before both fully connected layers
  End if
  End for
  Save training parameters
  End for
End while

Step 1: Collect one-dimensional temporal vibration signals of rolling bearings during operation for subsequent experiments.

Step 2: Denoise the collected vibration signals using a selected wavelet threshold and perform appropriate sample segmentation to generate a new experimental dataset.

Step 3: Use the one-dimensional dataset denoised by wavelet as input and transform it into a two-dimensional feature image using the GADF image encoding method.

Step 4: Divide the encoded feature images into training and testing sets according to the set ratio.

Step 5: Build a deep learning network system based on the traditional LeNet-5 architecture and set experimental parameters.

Step 6: Optimize the traditional LeNet-5 model using three different methods and adjust the model parameters.

Step 7: Use the experimentally partitioned training set to train the optimized model to see if it achieves the expected accuracy. If it does not achieve the expected accuracy, continue with Step 6. Otherwise, save the model parameters and complete the training of the model.

Step 8: After completing the training optimization of the model, import the preset test set into the model for convolutional neural network fault diagnosis, and output the fault diagnosis accuracy and confusion matrix as experimental results.

5. Experiment and Analysis

5.1. Data Denoising Experiment and Analysis

This article adds Gaussian white noise with different signal-to-noise ratios to the original data to simulate noise interference in real industrial scenarios. The signal-to-noise ratio is considered a criterion for evaluating whether the noise reduction effect meets expectations, but the original signal often overlaps with the noise and is difficult to distinguish. In order to more accurately evaluate the similarity between the denoising effect and the original signal, this paper adopts two different evaluation criteria.

(1) The energy ratio of the denoising result to the original signal is shown in Equation (12).

$$per = \frac{\sqrt{\sum_i \chi d_i^2}}{\sqrt{\sum_i S_i^2}} \tag{12}$$

In the equation, s_i is the original signal. The energy ratio reveals the ratio between the energy of the denoised signal and the original signal. The better the noise reduction effect, the less energy is required, so a lower energy ratio means a better noise reduction effect.

(2) The standard deviation between the denoised signal and the original signal is shown in Equation (13).

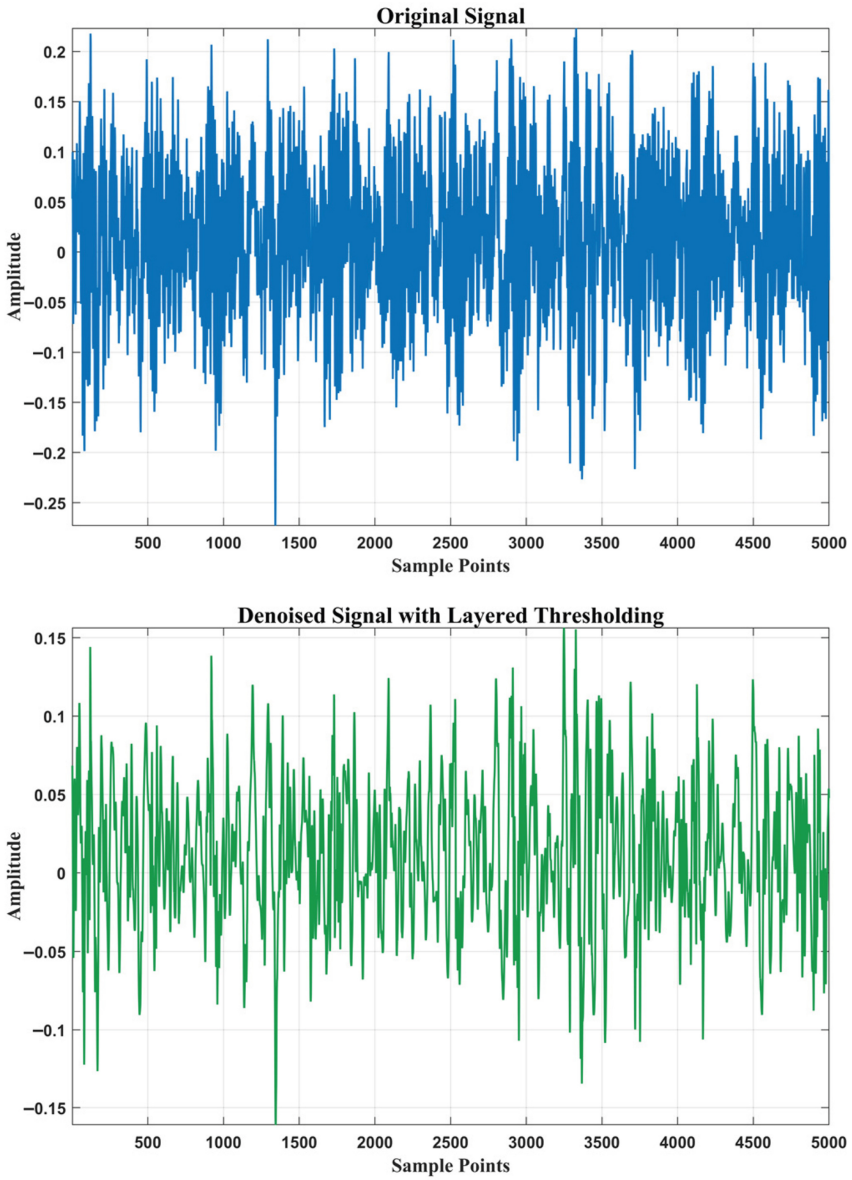
$$e = \sqrt{\sum_i (\chi d_i - S_i)^2} \tag{13}$$

In the equation, xd_i is the denoised result signal. The standard deviation reveals the degree of deviation between the denoised signal and the original signal. The smaller the standard deviation, the more obvious the noise reduction effect.

In order to achieve noise reduction in vibration signals and obtain effective experimental data, this paper proposes wavelet denoising processing based on the original collected vibration signals, which is implemented using MATLAB. The experimental results are shown in Figure 7. From the comparison of soft and hard threshold denoising images in Figure 7, it can be seen that, through wavelet denoising technology, the smoothness of the signal is significantly enhanced, and the abrupt details of the original signal are almost

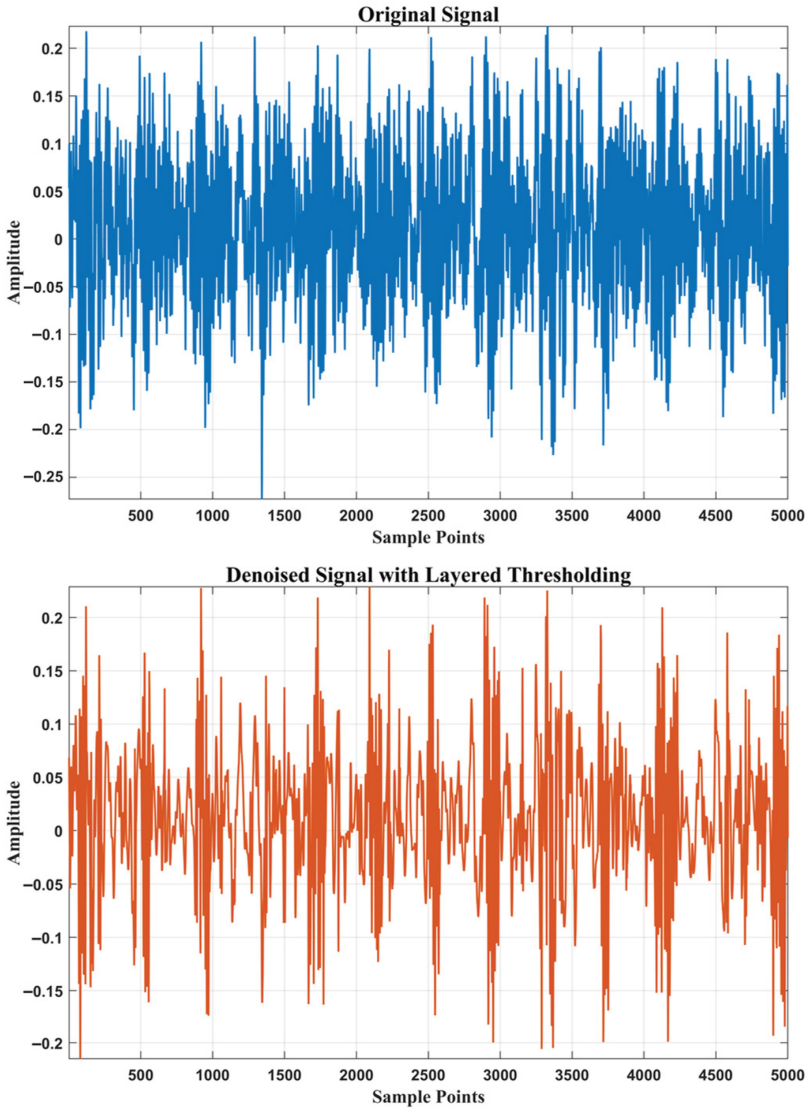
eliminated, effectively maintaining the peaks and abrupt points of the useful signal in the original signal.

This paper tests the impact of wavelet threshold selection on noise reduction using the CWRU rolling bearing dataset number 97.mat (i.e., normal bearing data). A 3-level decomposition using the db4 wavelet is employed, and the layer thresholds are calculated using wdcbm. Simultaneously, an LVD adaptive strategy based on local variance estimation is combined to dynamically adjust the threshold according to the noise level. Through soft thresholding and inverse wavelet reconstruction, high-frequency noise is effectively suppressed while preserving impact characteristics. The experimental results are shown in Table 2. From the comparison of soft and hard threshold verification results in Table 2, it can be seen that the standard deviation of the soft threshold is higher, while the signal energy corresponding to the soft threshold is lower than that corresponding to the hard threshold. Based on the above, it can be concluded that the noise reduction effect of the soft threshold is better than that of the hard threshold, and the denoised signal retains the overall signal trend well. Therefore, this paper selects soft thresholding.



(**a**) Comparison between the original signal image and the soft threshold denoising processed image

Figure 7. *Cont.*



(**b**) Comparison between the original signal image and the hard threshold denoising processed image

Figure 7. Comparison of Soft and Hard Threshold Denoising Images.

Table 2. Comparison of Soft and Hard Threshold Verification Data.

	Energy Ratio	Standard Deviation
Soft threshold denoising	0.6075	18.5402
Hard threshold denoising	0.7804	19.3949

5.2. Case Introduction

This article used the rolling bearing data from a certain university's bearing data center website as the experimental dataset for the experiment. Taking SKF's 6205-2RS deep groove ball bearing as an example, the driving end bearing data with a sampling frequency of 12 kHz, motor load of 0, and speed of 1797 r/min were selected for verification. In the data processing stage, the input dataset is divided into ten categories of fault labels based on the inner ring, rolling element, and outer ring of the rolling bearing. The experimental sample structure of the rolling bearing is shown in Table 3. The experimental results of different algorithms with different dataset partitioning ratios are shown in Table 4. As the accuracy deviation between the training and testing sets is small when the dataset is partitioned by

7:3, this ratio is selected for the partitioning of the training and testing sets. The training parameters and network structure parameters are shown in Tables 5 and 6.

Table 3. Sample Structure of Rolling Bearings with a Load of 0 kW.

Fault diameter/mm	0	0.18	0.18	0.18	0.36	0.36	0.36	0.54	0.54	0.54
	Normal	inner circle	rolling	outer circle	inner circle	rolling	outer circle	inner circle	rolling	outer circle
Tags	1	2	3	4	5	6	7	8	9	10

Table 4. Comparison of different algorithms combined with different datasets for partitioning.

Algorithm	Algorithm Training Set: Test Set = 7:3		Training Set: Test Set = 8:2	
	Training Accuracy/%	Testing Accuracy/%	Training Accuracy/%	Testing Accuracy/%
GADF-CNN	88.65	88.42	88.27	87.36
GADF-ResNet	87.72	87.53	87.94	86.89
GADF-BP	86.58	86.01	87.13	86.24

Table 5. Training Parameter Setting Table.

Name	Small Batch Size	Maximum Training Times	Initial Learning Rate	Verification Frequency
Value	128	150	0.001	10

Table 6. Structural Parameters of LeNet-5.

Name	Remarks
he input image size	32 × 32, with 1 channel
Convolutional layer 1	kernel size of 5 × 5 and a total of 8 kernels
Maximum pooling layer 1	adopts the maximum pooling method, with a pooling area of 2 × 2 and a stride of 2
Convolutional layer 2	kernel size of 5 × 5 and 16 kernels
Maximum pooling layer 2	adopts the maximum pooling method, with a pooling area of 2 × 2 and a stride of 2
Fully connected layer 1	connects 120 neurons
Fully connected layer 2	connects 84 neurons

5.3. Experimental Optimizer Selection

Optimizer as a method of adjusting the parameters of neural network models aims to minimize the loss function to the greatest extent possible. This experiment compares and tests four commonly used optimizers, namely SGD, SGDM, Adam, and RMsprop, as shown in Figure 8. Among them, SGD and SDGM are gradient descent optimizers. The problem with SGD is that it updates frequently and has a large variance, so it cannot converge quickly to the local optimal solution, while SGDM has a slower training speed. Adam and RMsprop are both adaptive optimizers, and the most important feature is that they do not require adjusting the learning rate value. However, RMsprop as an optimizer requires manual setting of initial learning rate and decay coefficient, making parameter tuning more complex. Adam was selected as the experimental optimizer for this experiment due to its high computational efficiency and the ability to naturally adjust accuracy.

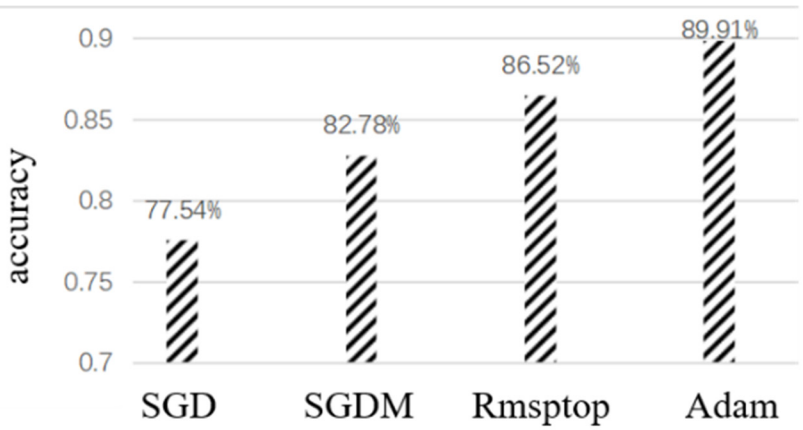


Figure 8. Comparison of Optimizer Accuracy.

5.4. Analysis of Experimental Results

5.4.1. Input Dimension Sensitivity Analysis

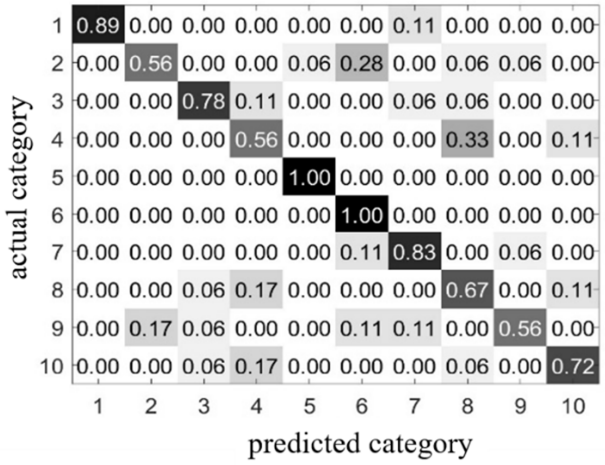
To evaluate the impact of input image size on fault diagnosis performance, this study compared and analyzed the input image resolution by adjusting only the input image resolution on the basic LeNet-5 model structure without any optimization. The experiments used the GADF two-dimensional feature map generated in this paper as input, setting three sizes: 28 × 28, 32 × 32, and 48 × 48, respectively, to analyze the sensitivity of different input dimensions to the model's recognition performance. The experimental results are shown in Table 7 below. At the 28 × 28 size, due to image pixel compression, some fault feature information was lost, and key details were weakened, thus affecting the model's recognition performance. As the input image size increased, the model's recognition performance improved, but the training time also increased. After comprehensively weighing accuracy and computation time, this study chose 32 × 32 as the input dimension to balance model performance and training efficiency, thus serving as the standard input size for subsequent fault diagnosis experiments.

Table 7. Comparison of results with different input dimensions.

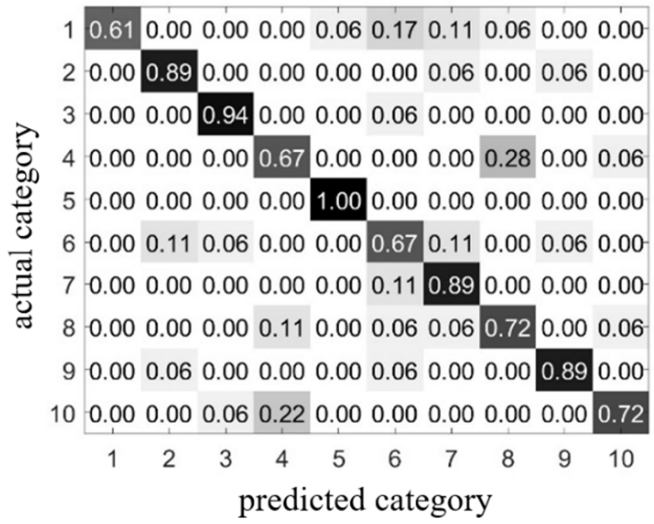
Input Dimensions	Recognition Accuracy (Mean ± Standard Deviation)/%	Average Training Time (Min)
28 × 28	81.28 ± 0.59	7.32
32 × 32	83.14 ± 0.45	9.07
48 × 48	85.02 ± 0.63	14.53

5.4.2. Analysis of the Impact of Noise Reduction Preprocessing

In order to further verify the necessity of adding data preprocessing before image processing, this paper takes the above experimental data as an example and uses the experimental signals that have undergone denoising preprocessing and those that have not. The experimental signals are encoded into feature images using GADF image encoding method, and then put into the LeNet-5 model that has not undergone any modification for fault diagnosis. The experimental results are shown in Figure 9. According to the analysis in Figure 9, the confusion matrix output by the model of the experimental data without denoising treatment has a more scattered distribution of accuracy points and lower accuracy compared to the denoised data. This demonstrates the feasibility of using wavelet denoising scheme in the data processing stage proposed in this paper.



(**a**) Undenoised model confusion matrix



(**b**) Model confusion matrix after denoising

Figure 9. Comparison of the Confusion Matrix of Noise Reduction Models.

5.4.3. Analysis of the Impact of Network Structure Transformation

This article is based on the optimization and modification of the traditional neural network LeNet-5, which does not add BN, Dropout, or L2 regularization. To eliminate the influence of random initialization on the experimental results, all experiments were repeated under 10 different random seeds, and the results are presented as mean ± standard deviation. Firstly, the dataset in this article is combined and preprocessed for noise reduction. The experimental results run in the unmodified neural network model are used as a reference. The experimental results are shown in Figure 10, which compares and demonstrates the experimental results of adding BN, Dropout, and L2 regularization.

Adding a BN layer to the neural network model and placing it before the activation function have a better effect, making the data standardized and the learning convergence speed faster, which can greatly accelerate the training speed. Figure 11a shows the experimental results with only BN added without Dropout and L2 regularization. From Figure 11, it can be seen that the fault diagnosis accuracy after adding only BN is 86.95% compared to 80.00% for the traditional neural network model without BN. This shows a significant improvement in the accuracy of fault diagnosis, and the point with the highest accuracy during the training process also meets the above conclusion.

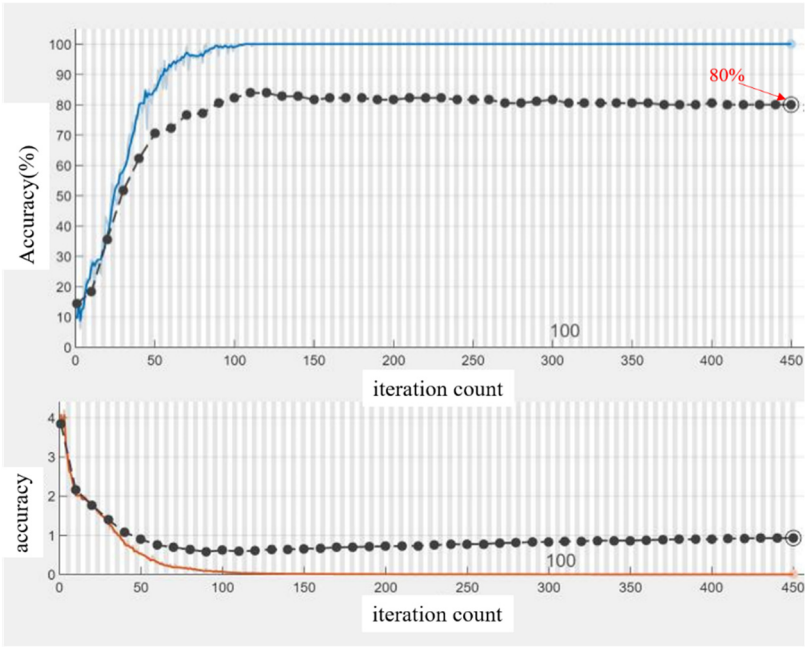
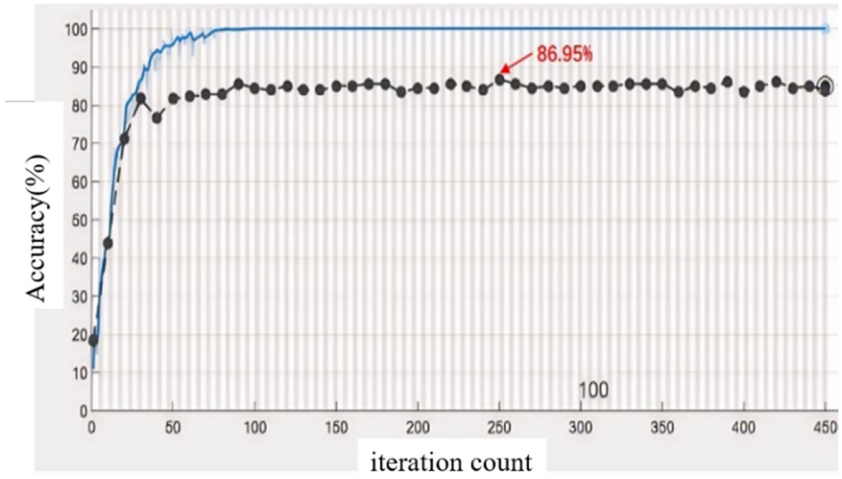


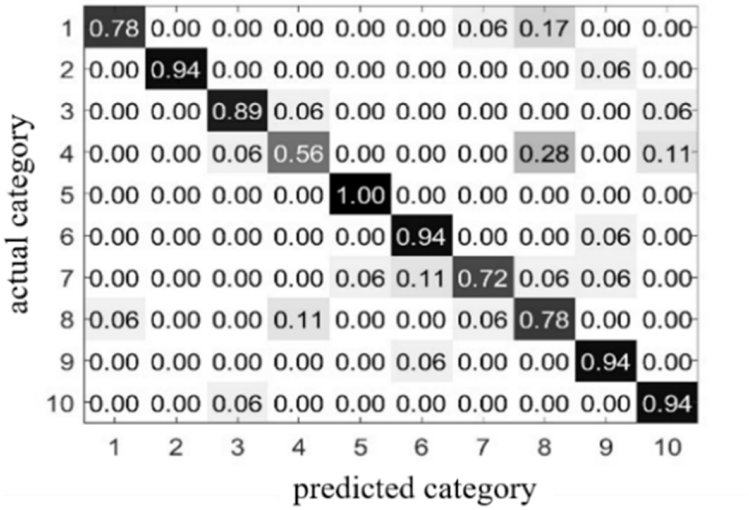
Figure 10. Experimental results of the unoptimized network model.

Adding Dropout layers to neural network models and placing them after fully connected layers with a high number of parameters allow hidden layer neurons to be deactivated with a certain probability. Random deactivation reduces the number of network parameters involved in calculating each gradient update, thereby reducing model capacity and preventing overfitting. Figure 12a shows the experimental results with only Dropout added without BN and L2 regularization. From Figure 12, it can be seen that, only after adding Dropout, the accuracy of fault diagnosis is 90.03%, which is significantly improved compared to the 80.00% accuracy trained by traditional neural networks without any optimization layers. The accuracy of fault diagnosis at the highest point also conforms to the above conclusion. It should be noted that in practice, the commonly used Dropout ratio is between 0.2 and 0.5, which means randomly setting 20% to 50% of the parameters to zero. When the ratio set in this article is 0.2, the experimental accuracy is good.



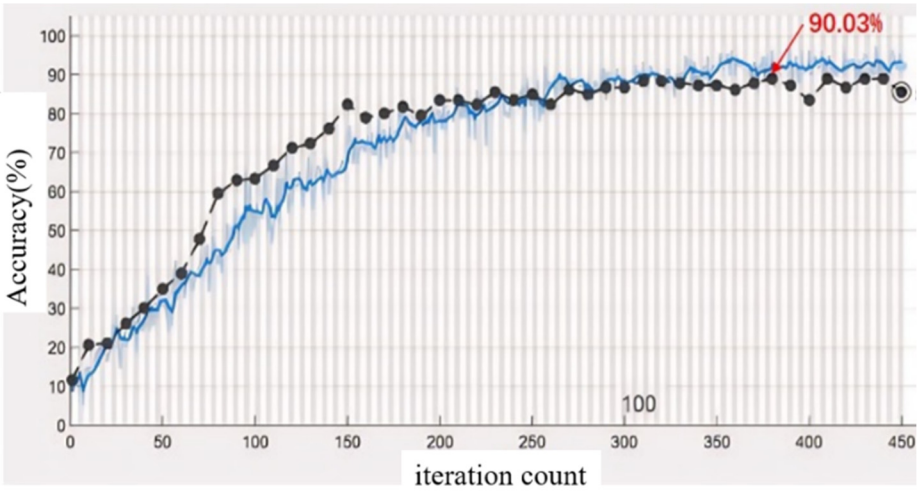
(**a**) The modified network validates the model

Figure 11. *Cont.*

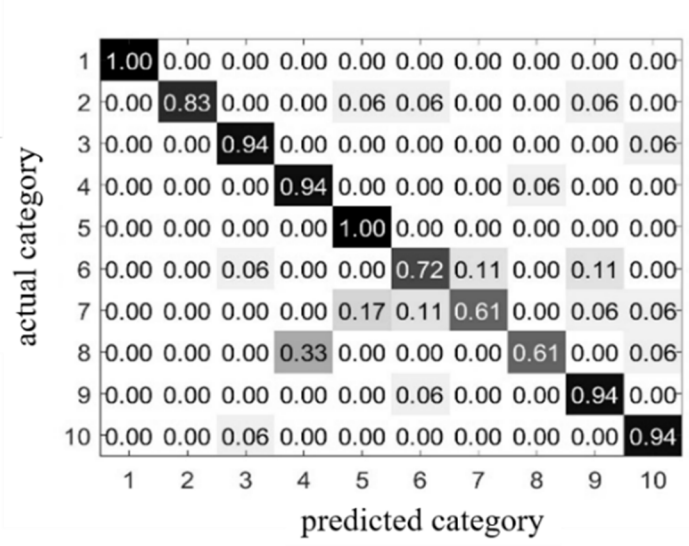


(**b**) The confusion matrix of the modified network model

Figure 11. Comparison of BN's impact on diagnostic results.



(**a**) Validation of the Model by the Modified Network

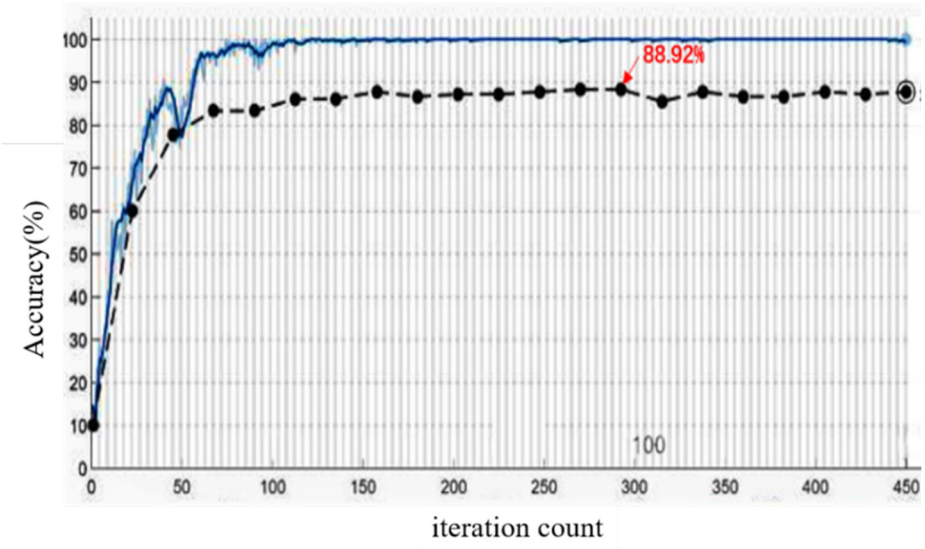


(**b**) The confusion matrix of the modified network model

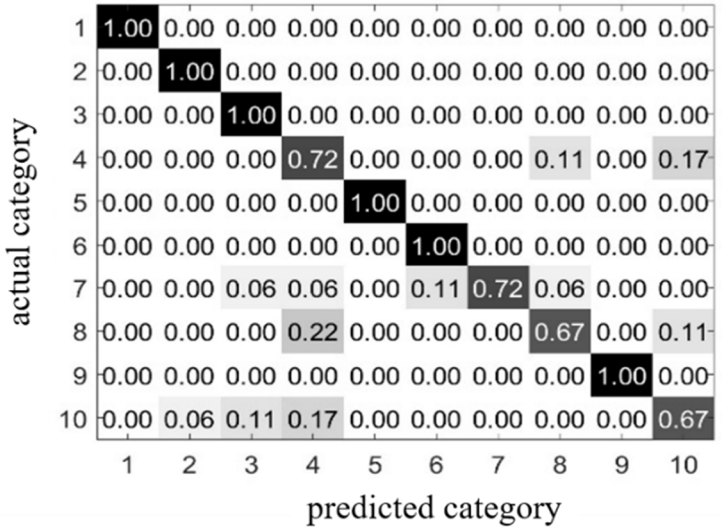
Figure 12. Comparison of Dropout's impact on diagnostic results.

Adding L2 regularization layers to neural network models is the most common type of regularization, which often helps to avoid overfitting or reduce network errors. Figure 13a shows the experimental results with only L2 regularization added without BN or Dropout. As shown in Figure 13, the fault diagnosis accuracy increased significantly to 88.92% only after adding L2 regularization, compared to the 80.00% accuracy achieved by traditional neural networks without any optimization layers. This further confirms the necessity of incorporating L2 regularization.

Finally, this article demonstrates the effectiveness of adding three model optimization methods, BN, Dropout, and L2 regularization, simultaneously in traditional neural network models, and increases the maximum training times to 200 to verify the training ability of the model. The experimental results are shown in Figure 14. According to Figure 14a, the fault accuracy of the optimized neural network model is 94.27%, which is 17.84% higher than the traditional neural network model without any modification, demonstrating the rationality and efficiency of the experimental scheme proposed in this paper.

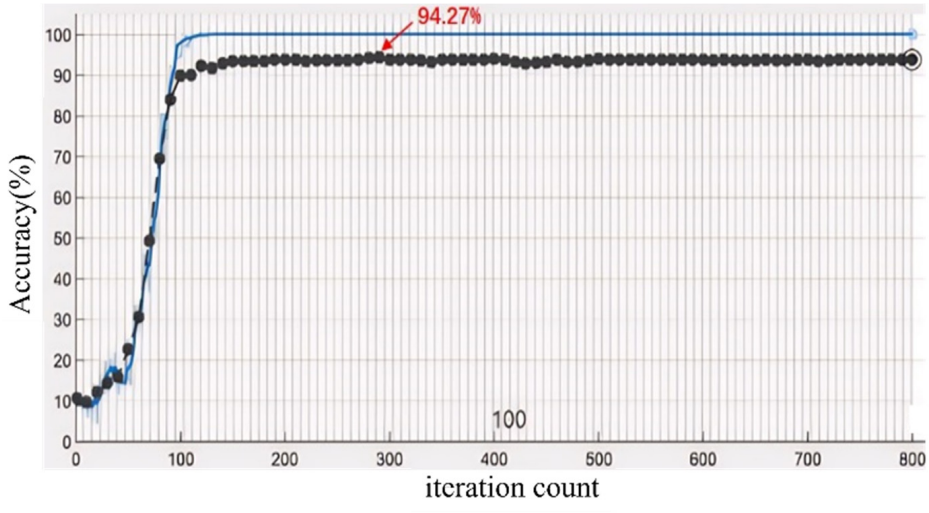


(**a**) Validation of the Model by the Modified Network

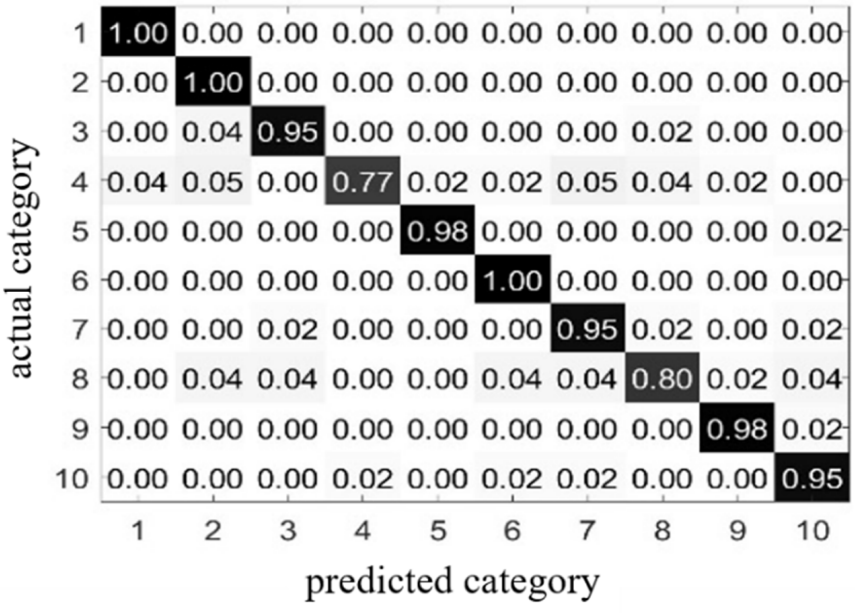


(**b**) The confusion matrix of the modified network model

Figure 13. Comparison of the impact of L2 regularization on diagnostic results.



(**a**) Model accuracy verification



(**b**) Confusion matrix

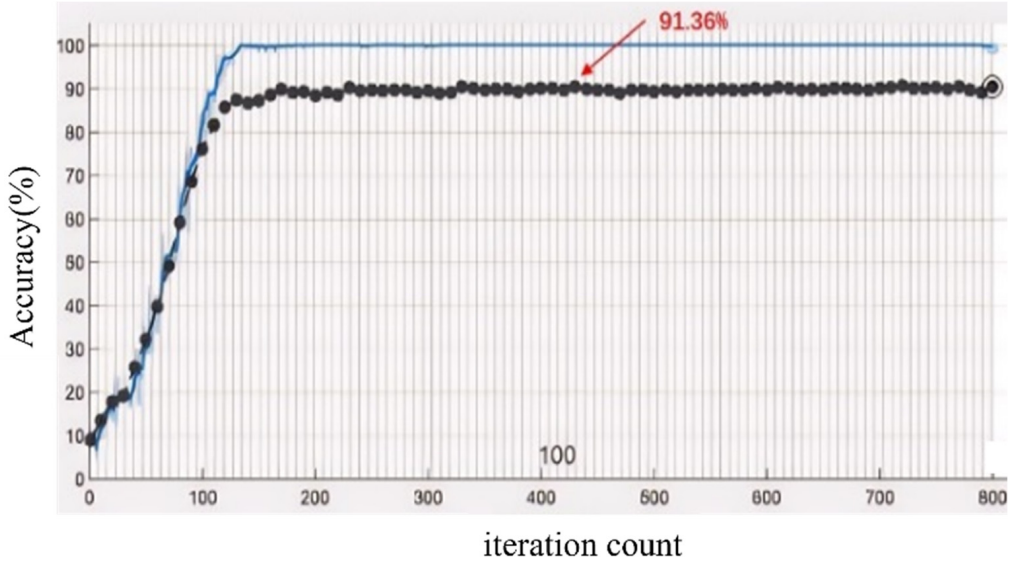
Figure 14. Experimental results of the optimized model.

5.4.4. Model Performance Analysis Under Load Changes

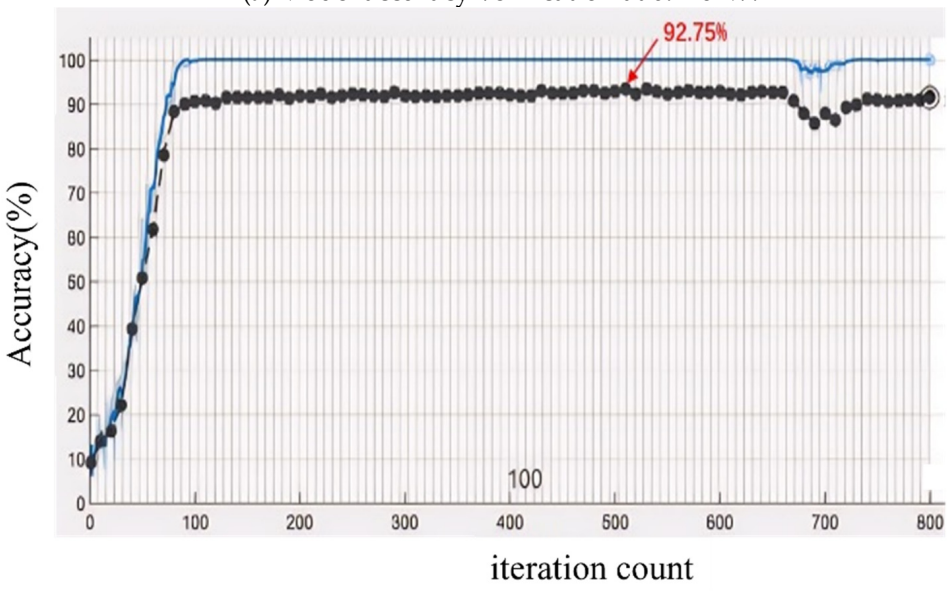
Based on the data from the CWRU Bearing Data Center, select bearing data under two different loads for experimental verification. Select datasets with loads of 0.746 kW and 2.237 kW, respectively. The dataset in the experiment has undergone denoising preprocessing, and the selected neural network model is the optimized LeNet-5 model. The experimental results are shown in Figure 15. As shown in Figure 15, the model and parameters designed in this experiment maintain high accuracy in fault diagnosis and recognition under different loads, with a fault diagnosis accuracy of 91.36% under a load of 0.746 kW and 92.75% under a load of 2.237 kW. The feasibility and rationality of the experimental ideas proposed in this article have been further verified, and the experimental comparison results are shown in Table 8.

Table 8. Comparison of Experiments before and after Optimization of Different Load Combination Models.

Load/kW	GADF-CNN (Unoptimized) Recognition Accuracy (Mean ± Standard Deviation)/%	GADF-CNN (Optimized) Recognition Accuracy (Mean ± Standard Deviation)/%
0	80.00 ± 1.21	93.69 ± 0.54
0.746	78.26 ± 1.35	90.45 ± 0.68
2.237	79.31 ± 1.19	91.53 ± 0.72

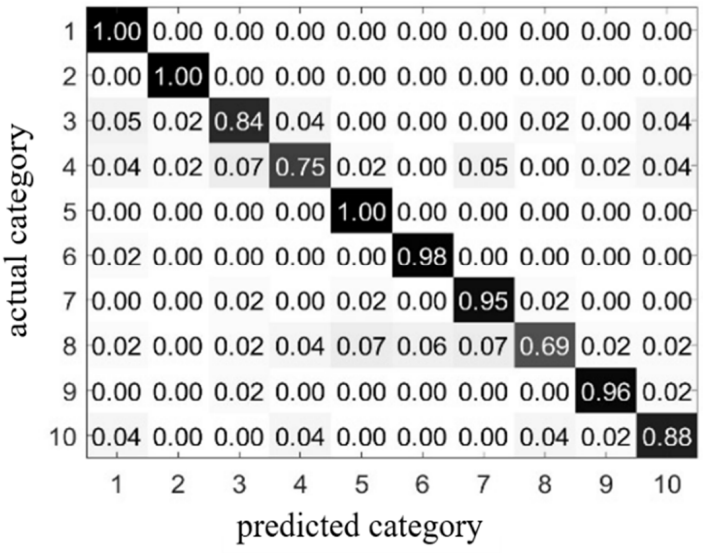


(**a**) Model accuracy verification at 0.746 kW

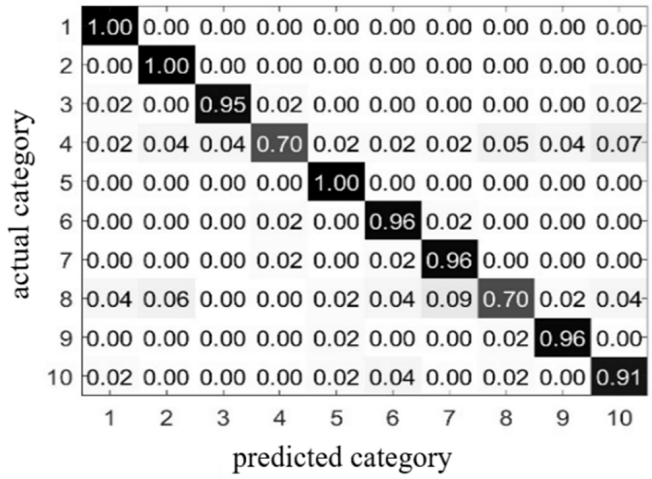


(**b**) Model accuracy verification at 2.237 kW

Figure 15. *Cont.*



(**c**) Confusion matrix at 0.746 kW



(**d**) Confusion matrix at 2.237 kW

Figure 15. Experimental results under different loads.

6. Conclusions

To address the problems of high noise levels, difficulty in fault feature extraction, and low identification accuracy in rolling bearing vibration signals under complex operating conditions, this paper proposes a fault diagnosis method based on wavelet denoising and an improved convolutional neural network. The specific conclusions are as follows:

(1) This article proposes the experimental idea of first performing signal denoising preprocessing to achieve more effective signal-to-noise separation and extract more effective experimental data. The experimental results indicate the necessity of using noise reduction processing in the data processing stage.

(2) The inevitability of optimizing neural network models. After comparing the optimization and transformation methods of different network structures, it was found that Batch Normalization, Dropout, and L2 Regularization, when optimized separately, all have a positive effect on improving the accuracy of neural network models. The experiment shows that by combining the three optimization methods together, the accuracy of the final fault diagnosis model reached 93.69%.

(3) In order to verify the generalization ability of the model in the face of new data, this study conducted experimental verification on different training and testing sets under three loading conditions. Experimental data show that, even in the case of significant fluctuations in load, the model can still maintain a relatively high level of fault detection accuracy.

Author Contributions: Methodology, X.S.; Software, J.L.; Resources, J.H.; Data curation, J.L.; Writing—original draft, X.S.; Writing—review & editing, J.H. and C.C.; Visualization, W.M.; Funding acquisition, X.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Jiangsu Provincial Basic Research Program Natural Science Foundation—Youth Fund Project (BK20230173), Basic Research Program of Jiangsu (BK20240316) and General Project of Basic Science (NATURAL SCIENCE) Research in Colleges and Universities of Jiangsu Province (23kjb460031).

Data Availability Statement: Data is contained within the article. The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

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

References

1. Wang, D.; Tsui, K. Statistical modeling of bearing degradation signals. *IEEE Trans. Reliab.* **2017**, *66*, 1331–1344. [CrossRef]
2. Zhou, H.; Huang, X.; Wen, G.; Lei, Z.; Dong, S.; Chen, X. Construction of health indicators for condition monitoring of rotating machinery: A review of the research. *Expert Syst. Appl.* **2022**, *203*, 117297. [CrossRef]
3. Zhou, H.; Chen, W.; Liu, J.; Chen, L.; Xia, M. Trustworthy and intelligent fault diagnosis with effective denoising and evidential stacked GRU neural network. *J. Intell. Manuf.* **2023**, *35*, 3523–3542. [CrossRef]
4. Luo, J.; Wen, G.; Lei, Z.; Su, Y.; Chen, X. Weak signal enhancement for rolling bearing fault diagnosis based on adaptive optimized VMD and SR under strong noise background. *Meas. Sci. Technol.* **2023**, *34*, 064001. [CrossRef]
5. Liu, Y.; He, B.; Liu, F.; Zhao, Y. Feature fusion using kernel joint approximate diagonalization of eigen-matrices for rolling bearing fault identification. *J. Sound Vib.* **2016**, *385*, 389–401. [CrossRef]
6. Miao, Y.; Zhao, M.; Lin, J. Improvement of kurtosis-guided-grams via Gini index for bearing fault feature identification. *Meas. Sci. Technol.* **2017**, *28*, 125001. [CrossRef]
7. Sahu, P.; Rai, R. Fault diagnosis of rolling bearing based on an improved denoising technique using complete ensemble empirical mode decomposition and adaptive thresholding method. *J. Vib. Eng. Technol.* **2023**, *11*, 513–535. [CrossRef]
8. Rai, A.; Sanjay, S. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **2016**, *96*, 289–306. [CrossRef]
9. Sun, Y.; Li, S.; Wang, X. Bearing fault diagnosis based on EMD and improved chebyshev distance in SDP image. *Measurement* **2021**, *176*, 109100. [CrossRef]
10. Yu, L.; Du, X.; Yang, D.; Liu, Y. Improved swin Transformer integrated with GADF method for bearing fault detection. *Eng. Res. Express* **2025**, *7*, 035420. [CrossRef]
11. Biswas, S.; Mamun, A.; Islam, S.M.; Bappy, M. Interpretable CNN models for computationally efficient bearing fault diagnosis using learnable Gaussian/Sinc filters. *Manuf. Lett.* **2025**, *44*, 110–120. [CrossRef]
12. Dai, M.; Jo, H.; Kim, M.; Ban, S.W. MSFF-Net: Multi-sensor frequency-domain feature fusion network with lightweight 1D CNN for bearing fault diagnosis. *Sensors* **2025**, *25*, 4348. [CrossRef]
13. Li, H.; Wang, G.; Shi, N.; Li, Y.; Hao, W.; Pang, C. A lightweight multi-angle feature fusion CNN for bearing fault diagnosis. *Electronics* **2025**, *14*, 2774. [CrossRef]
14. He, X. Research on bearing fault diagnosis based on the fusion of CNN and LSTM algorithms. *J. Phys. Conf. Ser.* **2025**, *3057*, 012054. [CrossRef]
15. Chen, Q.; Zhang, F.; Wang, Y.; Yu, Q.; Lang, G.; Zeng, L. Bearing fault diagnosis based on efficient cross space multiscale CNN transformer parallelism. *Sci. Rep.* **2025**, *15*, 12344. [CrossRef]
16. Lu, G.; Liu, J.; Yan, P. Graph-based structural change detection for rotating machinery monitoring. *Mech. Syst. Signal Process.* **2018**, *99*, 73–82. [CrossRef]
17. Li, K.; Zhang, H.; Lu, G. Graph entropy-based early change detection in dynamical bearing degradation process. *IEEE Internet Things J.* **2024**, *11*, 23186–23195. [CrossRef]

18. Sun, W.; Zhou, Y.; Chen, B.; Feng, W.; Chen, L. A two-stage method for bearing fault detection using graph similarity evaluation. *Measurement* **2020**, *165*, 108138. [CrossRef]
19. Wang, T.; Liu, Z.; Lu, G.; Liu, J. Temporal-spatio graph based spectrum analysis for bearing fault detection and diagnosis. *IEEE Trans. Ind. Electron.* **2020**, *68*, 2598–2607. [CrossRef]
20. Tao, H.; Wang, P.; Chen, Y.; Stojanovic, V.; Yang, H. An unsupervised fault diagnosis method for rolling bearing using STFT and generative neural networks. *J. Franklin Inst.* **2020**, *357*, 7286–7307. [CrossRef]
21. Li, A.; Pan, Y. Structural information and dynamical complexity of networks. *IEEE Trans. Inf. Theory* **2016**, *62*, 3290–3339. [CrossRef]
22. Noman, K.; Li, Y.; Wang, S. Continuous health monitoring of rolling element bearing based on nonlinear oscillatory sample entropy. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–14. [CrossRef]
23. Wang, Y.; Wong, D. Investigations on sample entropy and fuzzy entropy for machine condition monitoring: Revisited. *Meas. Sci. Technol.* **2023**, *34*, 125104. [CrossRef]
24. Wang, S.; Li, Y.; Noman, K.; Wang, D.; Feng, K.; Liu, Z.; Deng, Z. Cumulative spectrum distribution entropy for rotating machinery fault diagnosis. *Mech. Syst. Signal Process.* **2024**, *206*, 110905. [CrossRef]
25. Wang, G.; Zhang, M.; Xiang, L.; Hu, Z.; Li, W.; Gao, J. A multibranch convolutional transfer learning diagnostic method for bearings under diverse working conditions and devices. *Measurement* **2021**, *182*, 109627. [CrossRef]
26. Chen, J.; Huang, R.; Zhao, K.; Wang, W.; Liu, L.; Li, W. Multiscale convolutional neural network with feature alignment for bearing fault diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [CrossRef]
27. He, Y.; Chen, J.; Zhou, X.; Huang, S. In-situ fault diagnosis for the harmonic reducer of industrial robots via multiscale mixed-convolutional neural networks. *J. Manuf. Syst.* **2023**, *66*, 233–247. [CrossRef]
28. Choudhary, A.; Mishra, R.K.; Fatima, S.; Panigrahi, B.K. Multi-input CNN based vibro-acoustic fusion for accurate fault diagnosis of induction motor. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105872. [CrossRef]
29. Li, T.; Zhou, Z.; Li, S.; Sun, C.; Yan, R.; Chen, X. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. *Mech. Syst. Signal Process.* **2022**, *168*, 108653. [CrossRef]
30. Wu, R.; Liu, C.; Han, T.; Yao, J.; Jiang, D. A planetary gearbox fault diagnosis method based on time-series imaging feature fusion and a transformer model. *Meas. Sci. Technol.* **2022**, *34*, 024006. [CrossRef]
31. Yao, J.; Liu, C.; Song, K.; Feng, C.; Jiang, D. Fault diagnosis of planetary gearbox based on acoustic signals. *Appl. Acoust.* **2021**, *181*, 108151. [CrossRef]
32. Wan, L.; Chen, Y.; Li, H.; Li, C. Rolling-Element Bearing Fault Diagnosis Using Improved LeNet-5 Network. *Sensors* **2020**, *20*, 1693. [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.