Article

Application of Multi-Scale Convolutional Neural Networks and Extreme Learning Machines in Mechanical Fault Diagnosis

Wei Zhang [1], Junxia Li [1,*], Shuai Huang [1], Qihang Wu [1], Shaowei Liu [1] and Bin Li [2]

[1] College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China; 18734114327@163.com (W.Z.); 19935112407@163.com (S.H.); tyutwqh@163.com (Q.W.); 15530080650@163.com (S.L.)

[2] Technical Department of Taiyuan Satellite Launch Center, Taiyuan 030027, China; msepaper@163.com

* Correspondence: lijunxia@tyut.edu.cn

Abstract: Extracting fault features in mechanical fault diagnosis is challenging and leads to low diagnosis accuracy. A novel fault diagnosis method using multi-scale convolutional neural networks (MSCNN) and extreme learning machines is presented in this research, which was conducted in three stages: First, the collected vibration signals were transformed into images using the continuous wavelet transform. Subsequently, an MSCNN was designed to extract all detailed features of the original images. The final feature maps were obtained by fusing multiple feature layers. The parameters in the network were randomly generated and remained unchanged, which could effectively accelerate the calculation. Finally, an extreme learning machine was used to classify faults based on the fused feature maps, and the potential relationship between the fault and labels was established. The effectiveness of the proposed method was confirmed. This method performs better in mechanical fault diagnosis and classification than existing methods.

1. Introduction

Rolling bearings are extensively used in manufacturing, hydropower, transportation, and other industries [1]. Rolling bearings are susceptible to various failures when operating at high speeds and heavy loads, affecting the performance and efficiency of the entire mechanical system and eventually causing economic losses, environmental pollution, casualties, and other serious effects [2]. Therefore, the early and accurate detection of rolling bearing faults is essential to ensure efficient production and avoid catastrophic accidents in industrial applications.

Recently, as a result of industrial automation, data-driven fault diagnosis technology, which can discover the potential relationship between collected data and bearing status, has attracted increasing attention [3]. Traditional fault diagnosis technology includes two main steps: feature extraction and fault classification [4]. The collected fault signals were first analyzed using signal analysis techniques and statistical calculations to identify features related to the machine's operation state. The obtained features were then input into the pattern recognition algorithm for fault diagnosis [5]. For example, Wang et al. [6] extracted multiscale features using generalized composite multiscale weighted permutation entropy technology. They used a support vector machine (SVM) optimized using the marine predators algorithm to diagnose and identify bearing faults according to the extracted features. Vashishtha et al. [7] extracted fault features using a filter-based relief technique and then realized intelligent recognition using an extreme learning machine (ELM). They also verified that ELM is superior to SVM in terms of fault diagnosis. Owing to its few training parameters and short running time, it is a promising model for fast fault diagnosis.

Traditional fault diagnosis methods extract features mainly through engineering experience, which significantly affects their diagnostic performance [8]. Many data features

are often extracted to improve fault diagnosis accuracy, easily leading to data redundancy and calculation waste. To realize adaptive feature extraction, convolutional neural networks (CNNs) [9], inspired by deep learning, automatically extract fault features by alternately using convolution and pooling operations, and they have been extensively used in the field of mechanical fault diagnosis [10]. CNNs have been shown to perform fault diagnosis more effectively than conventional techniques using shallow structures, such as SVM and artificial neural networks [11]. Sun et al. [12] combined a symmetrical dot pattern image with a CNN and realized an accurate rolling mechanical diagnosis based on an optimal CNN. Khodja et al. [13] used a CNN and vibration spectrum imaging to classify bearing faults, which exhibited excellent performance in terms of accuracy and robustness.

Traditional CNNs have some flaws, such as information loss during feature extraction, which reduces detection accuracy. Their generalization ability and accuracy are greatly improved by increasing the depth of the network structure. However, blindly increasing the network depth causes a waste of computing resources and overfitting [14]. Therefore, extracting features from multiple convolution kernels of different sizes makes the network wider. The realization of multiscale feature extraction can provide a solution to this problem [15]. For example, Deng et al. [16] designed a novel multi-scale feature fusion block to fuse sensitive fault features. At the same time, CNNs have shortcomings such as many parameters and a long running time. Some scholars have found that CNNs with randomly generated parameters exhibit good feature extraction abilities [17]. Pinto and Cox [18] extracted features by randomly generating a large number of complex, nonlinear, and multilayered CNNs combined with standard machine learning techniques, and they achieved good results in the visual system. Jarrett et al. [19] adopted a CNN with random weights, which could perform the object recognition task well without training, thus avoiding a time-consuming learning process. The feature extraction ability of multi-scale convolutional neural networks (MSCNN) with randomly generated parameters is yet to be studied. This topic is studied in this paper.

Aiming at the problems of the feature extraction of the bearing fault not being sufficient and the time required to adjust the parameters of the deep learning network, we proposed a novel mechanical fault diagnosis method based on an MSCNN-ELM. It includes three consecutive steps: signal preprocessing, feature extraction based on the MSCNN, and fault classification based on the ELM. The main contributions of this study are as follows:

(1) A new fault diagnosis model composed of an MSCNN and ELM was proposed. The MSCNN was used to extract multiscale features from images obtained using the continuous wavelet transform (CWT). The ELM was used as a classifier to determine the potential relationship between the fault features and the labels.

(2) The parameters of the MSCNN in the model were randomly generated with a Gaussian probability distribution, which reduced the calculation time for adjusting the parameters compared with a deep neural network. The best parameters were found using the grid search method to improve the fault diagnosis accuracy.

(3) The self-made idler dataset and bearing dataset of Case Western Reserve University (CWRU) were used in the fault diagnosis experiments, and the effectiveness of the proposed method was confirmed. Using accuracy and running time as the evaluation indicators, the fault diagnosis performance of the MSCNN-ELM was thoroughly examined in comparison to other methods.

The following summarizes the structure of the paper: Section 2 introduces the basic principles of the MSCNN and the ELM. Section 3 introduces the algorithm flow chart based on the MSCNN-ELM. Two other experimental datasets are described to further substantiate the advantages of the MSCNN-ELM in Section 4. Section 5 summarizes the conclusions and prospects of this research.

2. Basic Theory

2.1. MSCNN

The MSCNN is mainly composed of a multi-scale convolution layer (MSCL), pooling layer, full connection layer, and other structures [20], as shown in Figure 1. In an MSCL, the input image is convolved by multiple parallel convolution kernels of different scales. The features are fully expressed from different feature-receptive fields to adaptively extract the advanced features in the image. The MSCL and pool layers appear alternately, and the obtained feature maps are concatenated, input to the full connection layer, and finally classified using a multi-classifier.
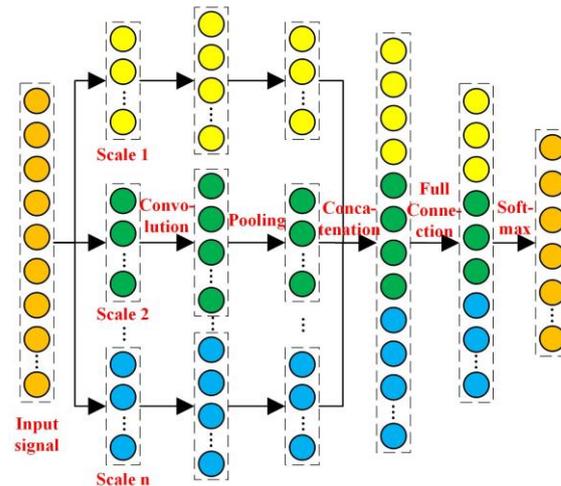

Figure 1. The structure of an MSCNN.

2.1.1. MSCL

In an MSCL, the input picture is convolved by a plurality of convolution kernels with different scales, and the feature map is obtained using an activation function [21]. The expression for the convolution process is as follows:

$$y_j^l = f\left(\sum_{i=1}^{n} x_i^{l-1} * \omega_{ij}^l + d_j^l\right),\tag{1}$$

where x_i^{l-1} denotes the feature map, and ω and d represent the weight and bias of the corresponding convolution kernel, respectively. l and j represent the number of layers and the number of scales, respectively.

2.1.2. Pooling Layer

Based on retaining essential features, a pooling layer changes the size of a feature map through a downsampling operation to reduce the feature dimension, improve the operation speed, and avoid overfitting in the network model. The mathematical expression is as follows:

$$y_i = f[\beta_i down(x) + b_i],\tag{2}$$

where $down(\ )$ indicates the downsampling function, β_i and b_i indicate the weight and the bias of the i-th feature, respectively, and x represents the output of the previous layer (convolution layer) [22].

2.1.3. Full Connection Layer and Softmax Classifier

A full connection layer can integrate local information with a classification in a pooling layer. The collected fault features were classified using the Softmax multi-classifier. Assuming there are the k classification problems, the output of the Softmax multi-classifier can be calculated as follows:

$$O = \begin{bmatrix} P(y = 1|x; \beta_1, b_1) \\ P(y = 2|x; \beta_2, b_2) \\ \cdots \\ P(y = k|x; \beta_k, b_k) \end{bmatrix}. \tag{3}$$

2.2. ELM

An ELM is primarily composed of an input layer, a hidden layer, and an output layer [23], as shown in Figure 2. The training speed was improved by randomly generating weights and bias and keeping them unchanged during the training process.
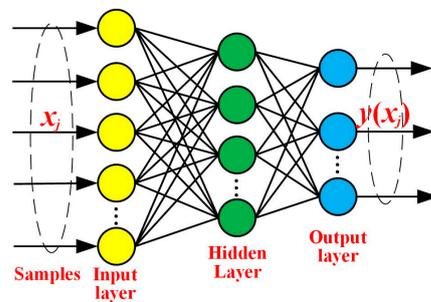

Figure 2. The structure of an ELM.

For N different training samples, the output result of an ELM is attained when the hidden layer has L nodes:

$$y(x_j) = \sum_{i=1}^{L} \beta_i h(\omega_i \times x_i + b_i), \; j = 1, 2, \cdots, N, \tag{4}$$

where h denotes the activation function, and ω_i and b_i represent the weight and bias of the hidden layer neuron, respectively. They were generated randomly and obeyed a continuous probability distribution. β_i represents the weight of the output layer.

Formula (4) can be simplified as follows:

$$Y = H\beta, \tag{5}$$

where the weight β is solved by finding the minimum sum of the prediction error loss functions. The objective function z is represented as follows:

$$z = \min_{\beta \in R^{L \times c}} \frac{1}{2}\|\beta\|^2 + \frac{C}{2}\|H\beta - Y\|^2, \tag{6}$$

where the first term represents the weight, the second term represents the training error, Y is the target matrix, and c is the class number [24].

The solution of the objective function is obtained by transforming it into a least-squares optimization problem. When the gradient of the objective function to β is zero, the following can be obtained [25]:

$$\beta^+ - CH^T(Y - H\beta^+) = 0, \tag{7}$$

where the optimal solution of the output weight β^+ can be calculated using matrix inversion. Based on the relationship between the sample number N and the hidden node L, β^+ exists in the following two situations:

$$\beta^+ = \begin{cases} \left(H^T H + \frac{I_L}{C}\right)^{-1} H^T Y, N \geq L; \\ H^T \left(H^T H + \frac{I_N}{C}\right)^{-1} Y, N \leq L. \end{cases} \tag{8}$$

3. The Proposed Method

3.1. Design of the MSCNN-ELM

If the convolution kernel of a CNN is too large, the model's detection effect on local features will be reduced; in contrast, if the convolution kernel is too small, the network cannot obtain global information well [26]. Inspired by the Inception Network and SqueezeNet [27], we designed an MSCNN to extract fault feature information to solve this problem entirely. The MSCNN extracts the scale information through three parallel 1×1, 3×3, and 5×5 convolution kernels. The dimensions of the channel number and parameter variables were reduced using a 1×1 convolution kernel. Detailed fault features can be captured, and the loss of detailed feature information can be reduced using the 3×3 and 5×5 convolution kernels, which are more sensitive to all detailed features of the image. The network depth and nonlinear feature extraction ability of the model were enhanced by using multiple small convolution kernels instead of large convolution kernels [28]. Finally, the obtained feature information was fused to form a new feature map, which was input into the strong classifier ELM to build the mapping relationship between the features and the labels. The MSCNN-ELM model is shown in Figure 3, considering the feature extraction capabilities of the MSCNN and the classification capabilities of the ELM.
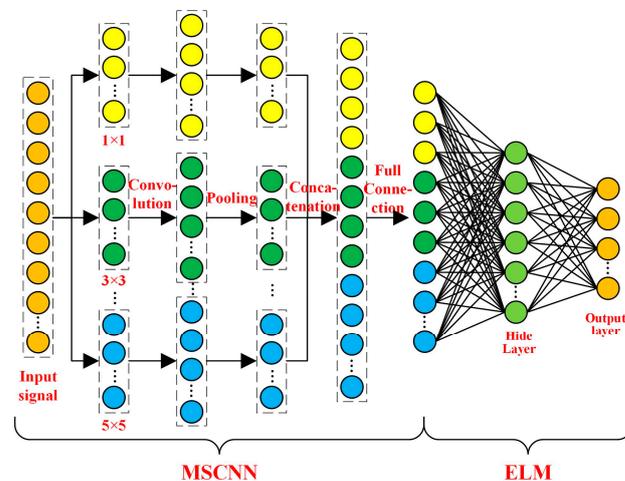

Figure 3. The structure of the proposed MSCNN-ELM.

3.2. Fault Diagnosis Procedure Flow Chart Based on the MSCNN-ELM

As shown in Figure 4, the framework for fault diagnosis based on the MSCNN-ELM is proposed, and it includes four main steps:

Step 1: Fault-diagnosis experiments were performed on different test rigs, and the original vibration signals were collected.

Step 2: The vibration signal was converted into a two-dimensional image by the CWT, which was compressed to pixels 28×28 in size. Samples were made according to a certain proportion.

Step 3: The parameters of the MSCNN were randomly generated and remained unchanged during the training process, and the best parameter combination in the ELM was selected using the grid search method. The MSCNN-ELM model was built layer by

layer. The fault features of the training set were obtained using the MSCNN, and then they were input into the ELM to train the model.

Step 4: The accuracy of the test set was calculated based on the trained MSCNN-ELM, the fault identification results were recorded, and the model's performance was evaluated from all aspects.
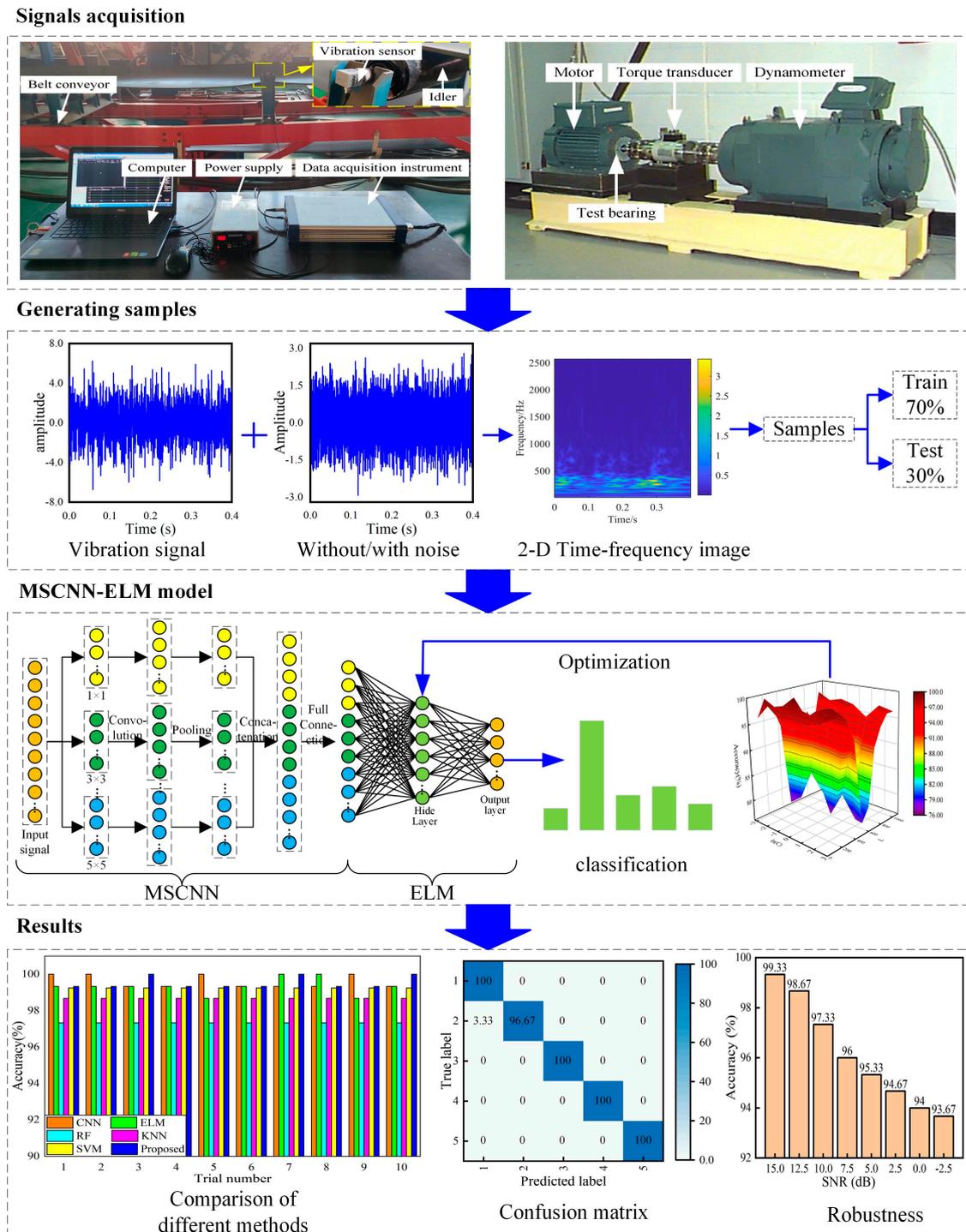

Figure 4. The proposed framework for fault diagnosis based on the MSCNN-ELM.

4. Experimental Cases

To reflect the good performance of the MSCNN-CNN, we conducted fault diagnosis experiments using the self-made idler dataset and bearing dataset from the CWRU. The experimental results showed that the model had good noise robustness and fault diagnosis ability, and its advantages were verified by comparison with those of other existing models. The experiments were implemented in a Python 3.10 environment and carried out using an AMD Ryzen 5 CPU @ 3.30 ghz, an NVIDIA GPU with 16 GB memory, and an x64-based processor.

4.1. Fault Diagnosis Based on the Self-Made Idler Dataset
4.1.1. Data Description

As an essential part of the belt conveyor, the idler gradually deteriorates in performance and eventually fails owing to coal dust intrusion, material impact, improper installation, and other factors during service [29]. To solve this problem, an idler test rig for fault diagnosis was developed. The test rig consisted of a belt conveyor, vibration sensor, data acquisition instrument, power supply, computer, and fault idlers [30], as shown in Figure 5. Four common fault types of the idler were simulated: the inner and outer race faults were formed by machining 2 mm cracks on the inner and outer races of the bearing with electro-discharge machining; the idler rotated intermittently by invading the bearing with foreign matter (gravel and dust); and the idler could not rotate because it was improperly installed. When the conveyor belt was running at 1 m/s, the running data in the states of health, inner-race fault (IF), outer-race fault (OF), intermittent rotation (IR), and non-rotation (NR) were collected using a data acquisition instrument. The original signals were processed using an overlapping sampling technique to obtain more samples, and the step size of the sliding segment was 1024, while the sampling rate was 5120 Hz. Two hundred samples in each state were collected, and the labels of the five operation states were indicated from 1 to 5, respectively. The dataset information is presented in Table 1.
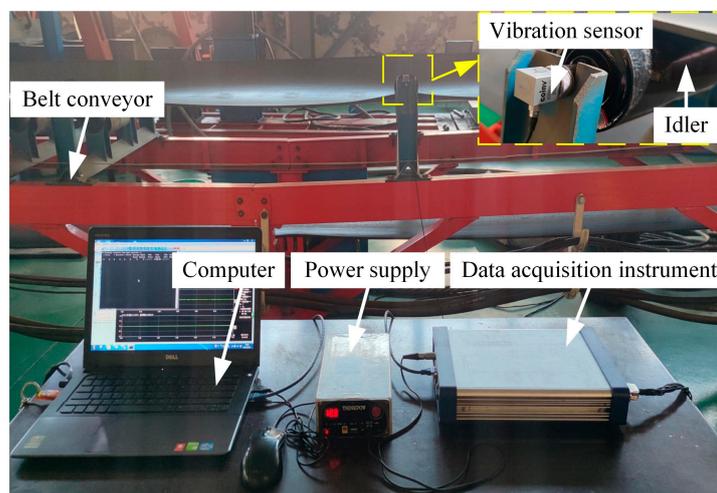

Figure 5. The idler fault diagnosis test rig.

Table 1. Sample description of the idler dataset.

Bearing Conditions	Number of Samples	Class Label
Health	200	1
IF	200	2
OF	200	3
IR	200	4
NR	200	5

A CWT has good locality in the time and frequency domains and can extract the macro and micro characteristics of signals from different scale spaces [31]. It is often more sensitive to mutation signals, easily captures mutation points, and accurately detects signal mutations. Therefore, the original vibration signals were preprocessed using the complex Morlet wavelet function to obtain the impact characteristics.

The wavelet scalograms of the idler in the five operating states are shown in Figure 6a–e. There are significant differences in the energy distributions of some of the fault types, and the energy distributions of some faults are similar, which makes it difficult to distinguish them. Therefore, it is necessary to use the MSCNN to further extract fault features to accurately identify many faults.
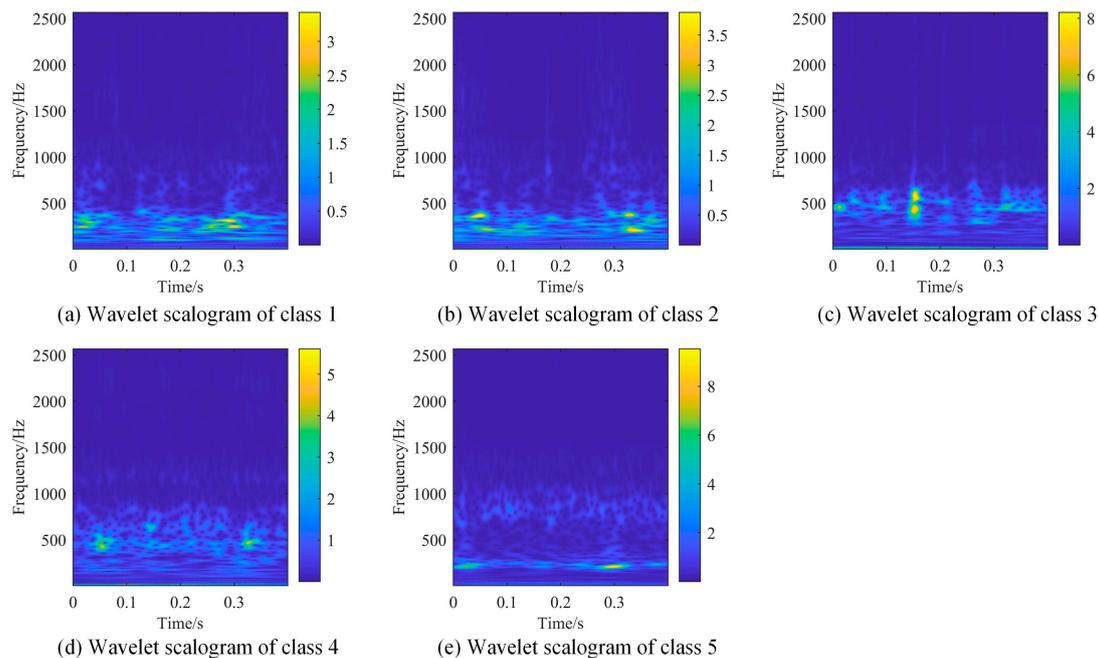
(a) Wavelet scalogram of class 1

(b) Wavelet scalogram of class 2

(c) Wavelet scalogram of class 3

(d) Wavelet scalogram of class 4

(e) Wavelet scalogram of class 5

Figure 6. Wavelet scalograms of the idler in five operating states.

4.1.2. Parameter Optimization

In the proposed MSCNN-ELM model, based on the ideas of the Inception Network and SqueezeNet [27], we determined the size and number of the convolution kernels. However, some key parameters still needed to be determined, such as the hidden node L (L is the number of neurons in the hidden layer) and penalty factor C in the ELM. Currently, heuristic optimization algorithms, bionic algorithms, and swarm intelligence algorithms are often used to optimize the parameters. These methods may fall into the optimal local solution during optimization and require assistance in obtaining the optimal solution. Although computationally intensive, the grid search method can search for all parameter combinations. Therefore, this study adopted the simplest grid search method, taking the test accuracy as the objective function, to determine the parameter combination corresponding to the maximum test accuracy. First, the range of L is set to (100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000), and the range of C is set to (10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, and 10^{3}).

The results of optimizing the parameters using the grid search method are shown in Figure 7. When L is a constant, there is an irregular change between parameter C and the test accuracy. When C is a constant, the precision first decreases and then increases with an increase in L. When L is in the range of 500–700, the test accuracy is low; therefore, it should be avoided. When the values of the parameters L and C are 300 and 0.1, the maximum test accuracy is 100%.
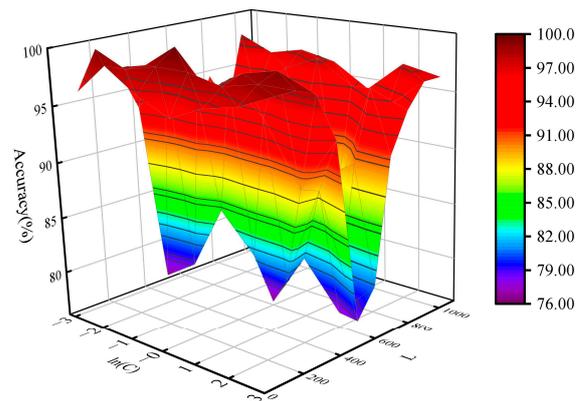
Figure 7. Effects of the parameters *L* and *C* on the test accuracy.

4.1.3. The Efficiency of the MSCNN

We built a new network model (called a CNN-ELM) to highlight the feature extraction capability of the MSCNN by comparing it with the MSCNN-ELM model. The CNN-ELM model was introduced as follows: The input image was convolved by three 4×4 convolution kernels, and then the dimension was reduced by a 2×2 maximum pool kernel. Finally, the obtained features were combined into a column vector through the full connection layer and input into the ELM for fault classification. The parameters of the CNN in the two structures were randomly generated using a Gaussian distribution.

The hidden nodes *L* and penalty factor *C* in the CNN-ELM were also selected by the grid search method, and the optimal parameters were (200, 0.01). Ten experiments were conducted using the same dataset to reduce the random error. The test accuracy and calculation time (excluding the preprocessing time) are listed in Table 2. The training accuracy of both methods reached 100%, indicating that a good fault diagnosis model was established through training. The test accuracy of both methods was very high. Compared with the CNN-ELM, the test accuracy of the MSCNN-ELM was slightly higher. Both methods have low calculation costs and can meet the demand for idler fault diagnosis in coal mines.

Table 2. Comparison of the accuracy and calculation time between the CNN-ELM and the MSCNN-ELM.

Model	Training Accuracy (%)	Testing Accuracy (%)	Training Time (s)	Testing Time (s)
CNN-ELM	100 ± 0	99.40 ± 0.60	1.15	0.08
MSCNN-ELM	100 ± 0	99.53 ± 0.47	1.12	0.08

4.1.4. Result Comparison with Other Methods

Machine learning techniques, including ELM, random forest (RF), *k*-nearest neighbors (KNN), SVM, and CNN, were tested in the same scenarios as the proposed method to further highlight its superiority [32]. The parameters of the other algorithms were chosen in the manner described below.

(1) CNN: Two convolution and pooling layers were constructed. The first convolution layer had four convolution kernels with a size of three, whereas the parameters of the second were eight and three, respectively. There were 200 neurons in this dense layer. Taking the cross entropy as the objective function, the Aam optimizer was selected to update iteratively 100 times, and Softmax was used for classification.

(2) ELM: The hidden node *L* and penalty coefficient *C* were set to 200 and 0.001, respectively.

(3) RF: The number of classifiers (n_estimators) and random number (random_state) were set to 100 and 0, respectively.

(4) KNN: The value of parameter *K* in the range (1, 2, . . . , 10) was set to 2.

(5) SVM: Radial-basis SVM was used for classification. The penalty coefficient *C* was 10 in the range of (10, 20, . . . , 100), and the kernel parameter *g* was 0.001 in the range of $(10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3)$.

All algorithms were tested ten times to reduce the effect of random initialization on the fault diagnosis results, and the average value was used as the final result, as shown in Figure 8. The test accuracy of all algorithms was greater than 97%, which meets the requirements for fault diagnosis. In particular, the CNN and MSCNN-ELM have higher test accuracies than the other methods. To quantitatively evaluate the benefits and drawbacks of all the methods, they were compared in terms of accuracy and calculation time, as shown in Table 3. The training accuracy of all the algorithms reached 100%, and all algorithms achieved good training results. The test accuracy of the ELM alone was 99.32%, whereas that of the MSCNN combined with the ELM was 99.53%, which improved the classification performance. Compared with deep learning methods, machine learning methods consume less time. The CNN updates 100 times iteratively, which requires the most time, and the training time is 3.25 s. The ELM and MSCNN-ELM consume less time—0.07 s and 0.08 s, respectively. The computation time of the MSCNN-ELM method is much less than that of the CNN method, mainly because the weights and bias in the MSCNN are randomly generated and remain unchanged during the training process. The ELM classifier can provide faster convergence speeds and a superior classification performance.
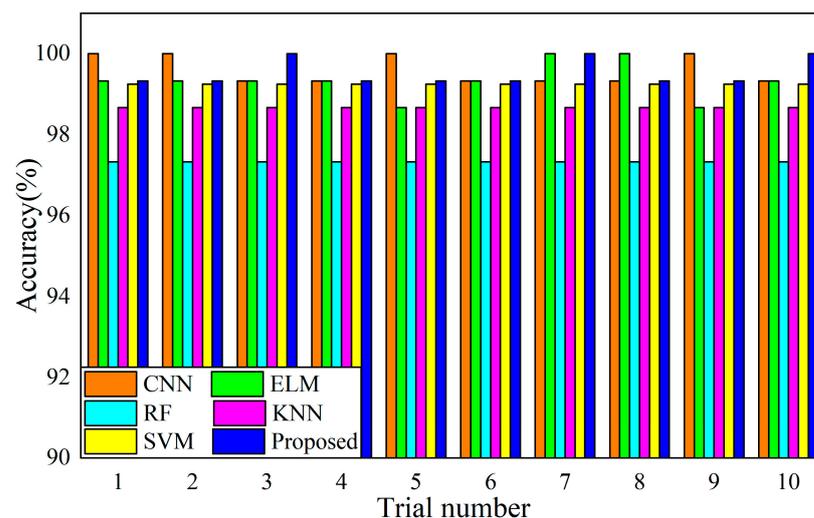

Figure 8. Comparison of test accuracy on an idler dataset by the different methods.

Table 3. Comparison of the accuracy and calculation times of the different methods.

Model	Training Accuracy (%)	Testing Accuracy (%)	Training Time (s)	Testing Time (s)
CNN	100 ± 0	99.60 ± 0.40	3.25	0.23
ELM	100 ± 0	99.32 ± 0.68	1.02	0.07
RF	100 ± 0	97.33 ± 0	1.13	0.08
KNN	100 ± 0	98.67 ± 0	1.10	0.08
SVM	100 ± 0	99.25 ± 0	1.08	0.11
MSCNN-ELM	100 ± 0	99.53 ± 0.47	1.12	0.08

Finally, the confusion matrix for the idler dataset obtained using this method is shown in Figure 9. Many of the samples were correctly predicted and classified. Some idlers with inner-ring failure are considered healthy. This is mainly because the inner-race fault impact is weak, and its fault features are submerged by the periodic harmonic generated by the rotating shaft, mistaking the inner-race fault for a healthy state.
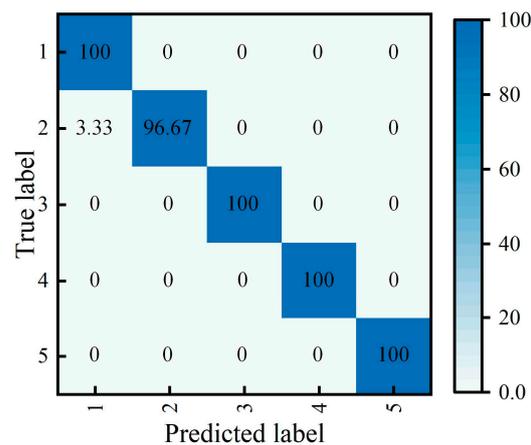
Figure 9. Confusion matrix for the idler dataset.

4.1.5. The Effect of Noise on the Model

Considering the harsh working environment of a coal mine, the collected idler running signals inevitably interfere with various noises, setting a higher standard for the robustness of the MSCNN-ELM method. Therefore, noise with various signal-to-noise ratios (SNR) ranging from 15 to −2.5 dB with an interval of −2.5 dB were added to the original vibration signal to simulate noise interference in industrial production [33]. The processed signals were converted into datasets and input into the MSCNN-ELM model to test the anti-interference performance of the model. It should be noted that the parameters of the MSCNN-ELM remained unchanged in different experiments.

Figure 10 shows the test accuracy of the MSCNN-ELM under different noise interference conditions. In the initial stage of noise increase, even if these images are slightly disturbed by noise, the MSCNN-ELM model can still learn relevant fault features from the images, correctly classify the faults, and achieve high test accuracy. With a further increase in noise interference, some fault features in the signal are submerged by noise, which leads to a failure of the model to fully extract the correct fault features and a decline in the test accuracy. When the input interference noise is −2.5 dB, the MSCNN-ELM model can still achieve 93.67% accuracy and has reliable fault diagnosis performance, which can meet the demand for fault diagnosis of idlers in coal mines.
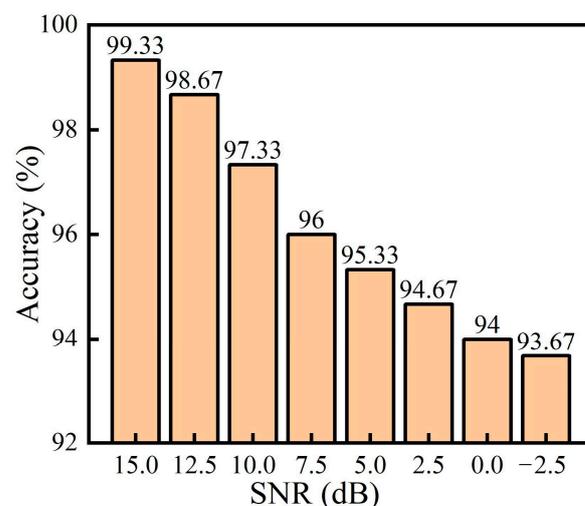

Figure 10. Test accuracy under different noise interferences.

4.2. Fault Diagnosis Based on the Bearing Dataset of the CWRU

4.2.1. Data Description

The bearing data provided by the CWRU Bearing Data Center were used for the experiments to verify the performance of the MSCNN-ELM method in the field of mechanical fault diagnosis (data source: https://engineering.case.edu/bearingdata center (accessed on 5 January 2023)) [34]. The test rig consisted of a motor, torque transducer, dynamometer, and test bearings [35], as shown in Figure 11. Single-point faults with different wear levels were machined on SKF6205 bearings using electro-discharge machining. Under various load conditions (0, 1, 2, and 3 hp), the operating data of the bearings under the conditions of health, inner-race fault (IF), outer-race fault (OF), and ball fault (BF) with different wear degrees were collected. The original data were divided into samples of 2000 points in length. The sliding step was set at 1000, and 400 samples were collected in each state; thus, 4000 samples in 10 states were obtained. The dataset information is presented in Table 4.
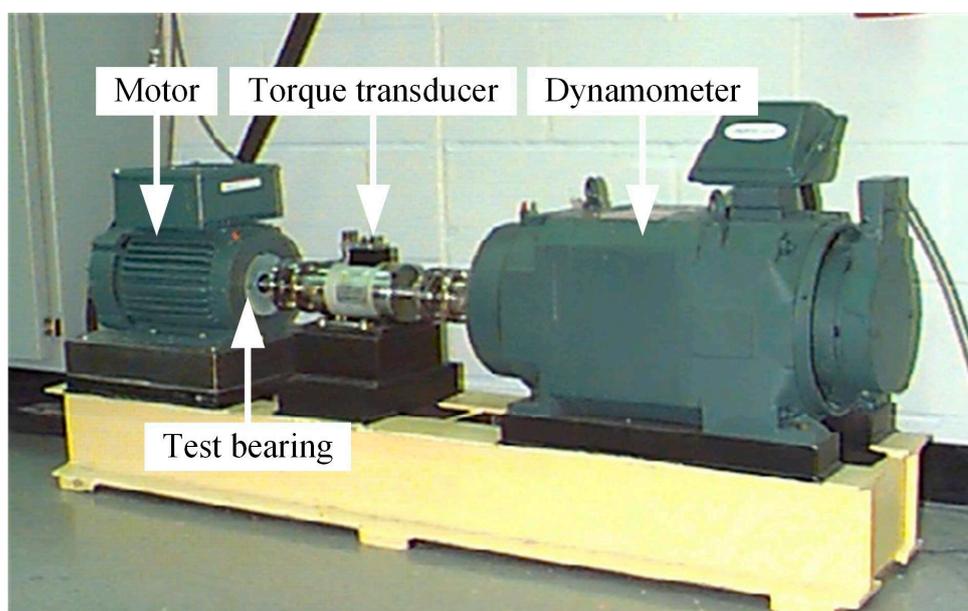

Figure 11. The bearing test rig of the CWRU.

Table 4. Sample descriptions of the bearing dataset of the CWRU.

Load (hp)	Bearing Conditions	Fault Diameters (inch)	Number of Samples	Class Label
	Health	0	400	1
	BF	0.007	400	2
	IF	0.007	400	3
	OF	0.007	400	4
	BF	0.014	400	5
0, 1, 2, and 3	IF	0.014	400	6
	OF	0.014	400	7
	BF	0.021	400	8
	IF	0.021	400	9
	OF	0.021	400	10

The CWT was performed on the collected data samples, and the wavelet-scale diagrams are shown in Figure 12a–j. The energy distributions of the different faults are significantly different, which is conducive to accurate fault diagnosis.
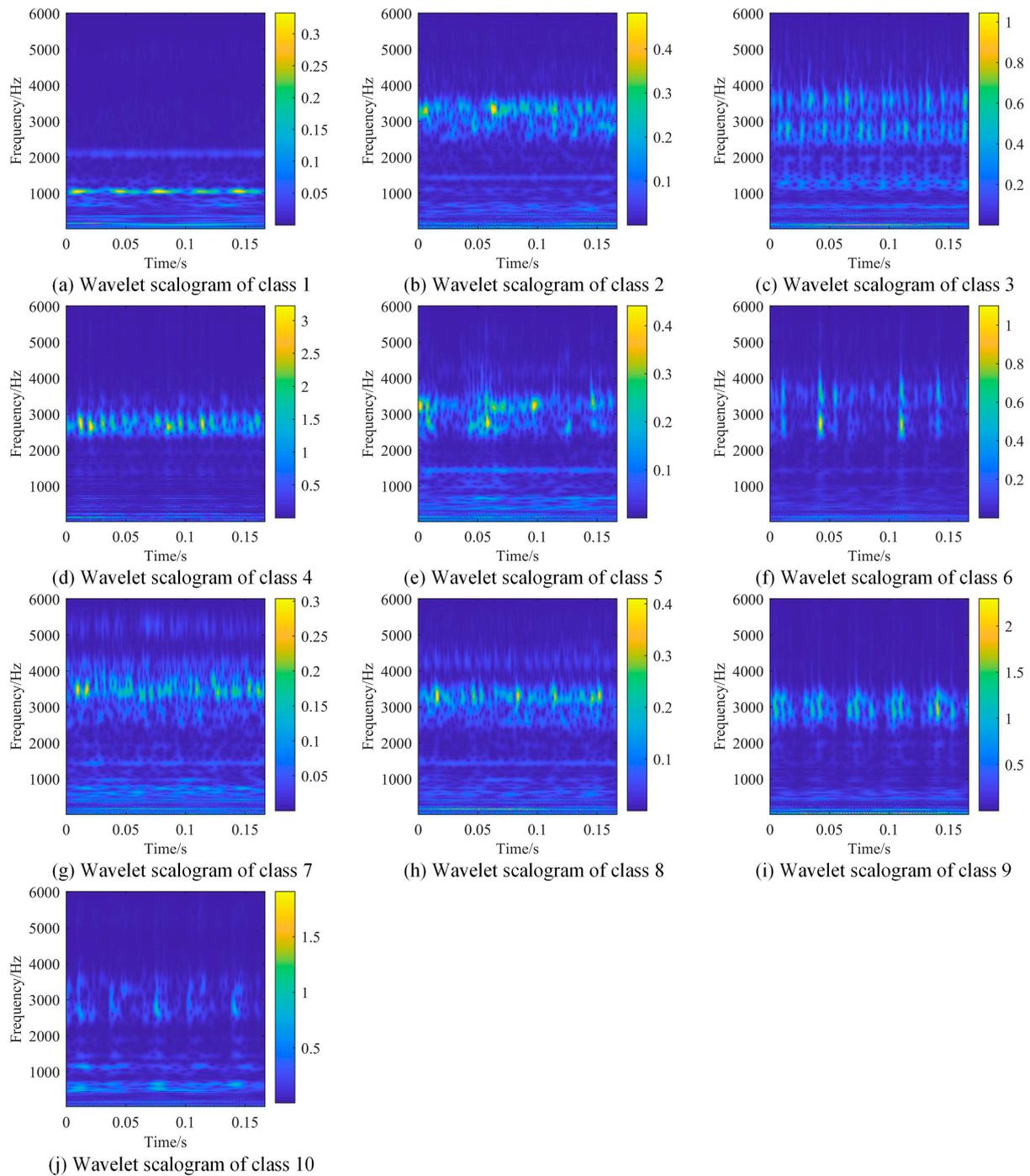
Figure 12. Wavelet scalograms of the bearing dataset for ten operating states.

4.2.2. The Efficiency of the MSCNN

The comparative results between the CNN-ELM and MSCNN-CNN models are presented in Table 5. The two models are similar in terms of computing time and testing accuracy. The testing accuracies of the MSCNN-ELM and the CNN-ELM are 99.43% and 99.37%, respectively. The MSCNN-ELM model has a higher testing accuracy and smaller standard deviation, which indicates that the MSCNN has a better feature extraction capability.

Table 5. Comparison of the accuracy and calculation times between the CNN-ELM and the MSCNN-ELM.

Model	Training Accuracy (%)	Testing Accuracy (%)	Training Time (s)	Testing Time (s)
CNN-ELM	99.85 ± 0.11	99.37 ± 0.21	9.37	0.67
MSCNN-ELM	99.88 ± 0.13	99.43 ± 0.15	9.45	0.68

4.2.3. Results Comparison with Other Methods

The experiment was repeated ten times using the same dataset, and the average testing accuracy and training accuracy were taken as the final results, as shown in Figure 13. The accuracy and calculation costs (excluding the preprocessing time) of the other methods are listed in Table 6. The results show that, based on good training, each model achieves good testing accuracy, and the overall accuracy is above 98%. The training and testing accuracies of the RF, KNN, and SVM remained unchanged in many experiments, which shows that these models have excellent stability. The training times for the ELM, RF, KNN, and SVM are 8.97 s, 9.98 s, 6.16 s, and 10.08 s, respectively, and the testing times are 0.64 s, 0.71 s, 0.65 s, and 0.72 s, respectively. These methods require less time to train the models and accurately identify faults. The training time and testing time of the CNN are 28.55 s and 2.04 s, respectively, and the calculation time is approximately three times that of the shallow learning method. There are many parameters in the CNN, such as weights and bias, that need to be determined through 100 iterations. The training and testing times of the MSCNN-ELM model are 9.45 s and 0.67 s, respectively, which are similar to the calculation costs of the shallow learning method. Its testing accuracy is 99.43%, which is higher than that of the shallow learning model, and it saves considerable time compared to the CNN.
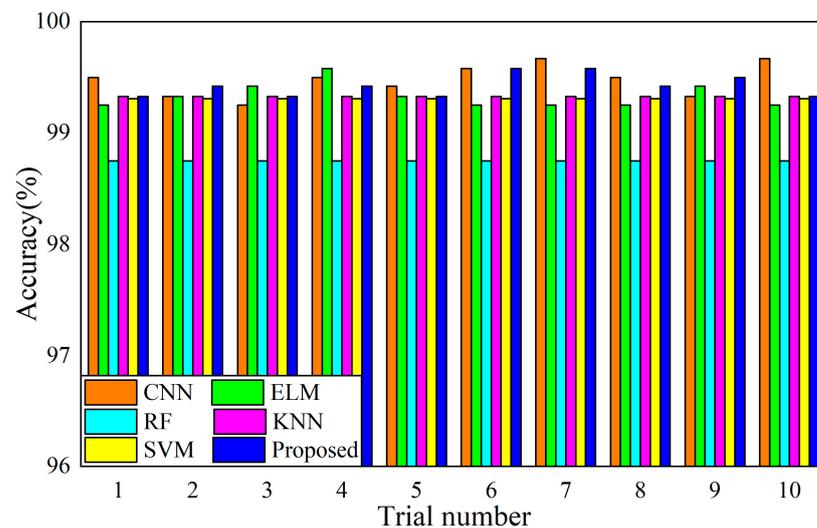

Figure 13. Comparison of test accuracy on the bearing dataset for the different methods.

Table 6. Comparison of the accuracy and calculation times of the different methods.

Model	Training Accuracy (%)	Testing Accuracy (%)	Training Time (s)	Testing Time (s)
CNN	99.80 ± 0.13	99.47 ± 0.20	28.55	2.04
ELM	99.99 ± 0.01	99.33 ± 0.22	8.97	0.64
RF	100 ± 0	98.75 ± 0	9.98	0.71
KNN	99.82 ± 0	99.33 ± 0	9.16	0.65
SVM	100 ± 0	99.31 ± 0	10.08	0.72
MSCNN-ELM	99.88 ± 0.13	99.43 ± 0.15	9.45	0.67

The confusion matrix for the final test is shown in Figure 14. The results show that misclassification occurs mainly in samples 2, 3, 5, and 9, and a few inner-race faults are mistaken for rolling-element faults. The energy distributions of these two faults are similar, as shown in Figure 12, which makes it difficult to distinguish faults.
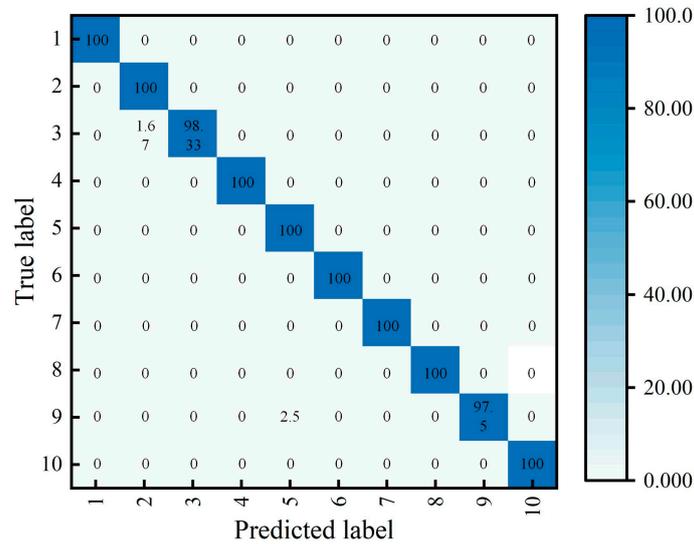

Figure 14. Confusion matrix on the bearing dataset of the CWRU.

4.2.4. The Effect of Noise on the Model

The fault diagnosis results of the MSCNN-ELM model under different noise levels are shown in Figure 15. With a gradual increase in noise interference, the model's fault diagnosis performance gradually decreased. Faults were accurately classified in the initial stage of an increase in noise. When SNR = 10 dB, this method can accurately identify most of the faults. Below that, the model's fault diagnosis performance drops sharply because the fault features are masked by noise interference. Finally, it could still accurately identify samples 4, 5, and 10.
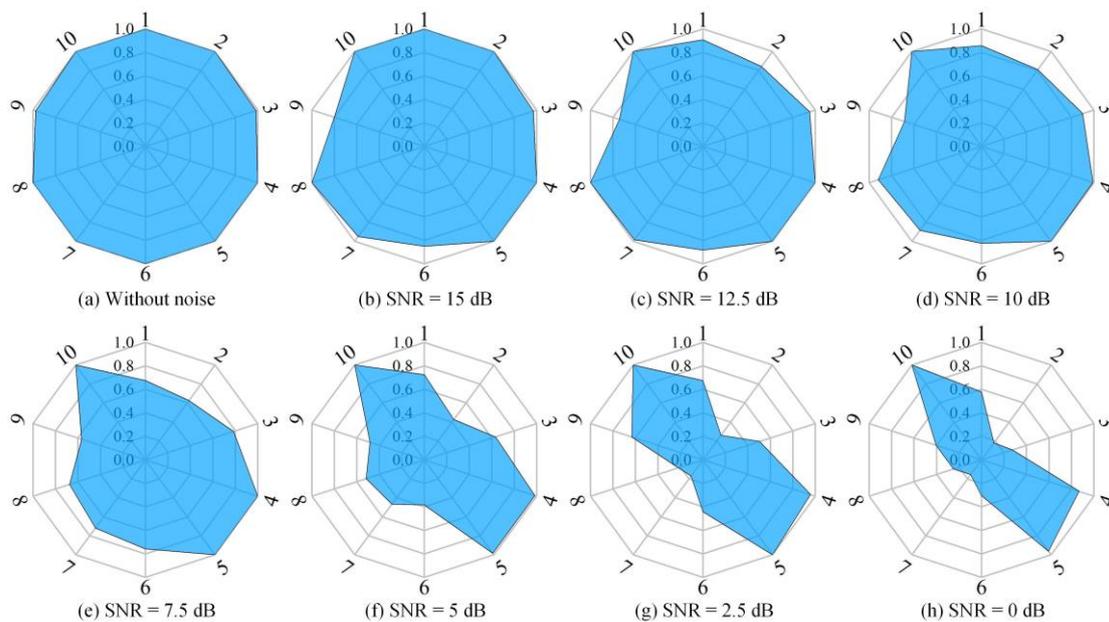

Figure 15. Fault diagnosis results under different noise levels.

5. Conclusions

A novel method of mechanical fault diagnosis based on the MSCNN-ELM is proposed, which effectively solves the problems of difficult fault feature extraction and low diagnostic accuracy in industrial production. We used numerous convolution layers with various branches to extract features and verified that their feature extraction ability was better than that of a single-scale model. The parameters in the MSCNN are randomly generated and follow a Gaussian probability distribution, which speeds up the calculation. The extracted final features were input into the strong classifier ELM to achieve fault classification. We employed a grid search algorithm to optimize the parameters of the ELM. The validity of the MSCNN-ELM was verified using the self-made idler dataset and the bearing dataset of the CWRU. The MSCNN-ELM has a high diagnostic accuracy and low running cost. The model exhibited outstanding fault identification ability and robustness, even under challenging working conditions.

In the future, we will consider combining transfer learning with the MSCNN-ELM to achieve fault diagnosis under variable operating conditions and ensure that the MSCNN-ELM model maintains a high fault diagnosis accuracy under complex operating conditions.

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

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

References

1. Zhao, B.; Zhang, X.; Li, H.; Yang, Z. Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions. *Knowl.-Based Syst.* **2020**, *199*, 105971. [CrossRef]
2. Lei, Y.; Lin, J.; He, Z.; Zuo, M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2013**, *35*, 108–126. [CrossRef]
3. Cerrada, M.; Sánchez, R.-V.; Li, C.; Pacheco, F.; Cabrera, D.; Valente de Oliveira, J.; Vásquez, R.E. A review on data-driven fault severity assessment in rolling bearings. *Mech. Syst. Signal Process.* **2018**, *99*, 169–196. [CrossRef]
4. Zhang, T.; Chen, J.; Li, F.; Zhang, K.; Lv, H.; He, S.; Xu, E. Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions. *ISA Trans.* **2022**, *119*, 152–171. [PubMed]
5. Xu, G.; Liu, M.; Jiang, Z.; Shen, W.; Huang, C. Online Fault Diagnosis Method Based on Transfer Convolutional Neural Networks. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 509–520. [CrossRef]
6. Wang, Z.; Yao, L.; Chen, G.; Ding, J. Modified multiscale weighted permutation entropy and optimized support vector machine method for rolling bearing fault diagnosis with complex signals. *ISA Trans.* **2021**, *114*, 470–484. [CrossRef]
7. Vashishtha, G.; Chauhan, S.; Singh, M.; Kumar, R. Bearing defect identification by swarm decomposition considering permutation entropy measure and opposition-based slime mould algorithm. *Measurement* **2021**, *178*, 109389. [CrossRef]
8. Tang, S.; Yuan, S.; Zhu, Y. Deep Learning-Based Intelligent Fault Diagnosis Methods toward Rotating Machinery. *IEEE Access* **2020**, *8*, 9335–9346. [CrossRef]
9. Han, T.; Liu, C.; Wu, L.; Sarkar, S.; Jiang, D. An adaptive spatiotemporal feature learning approach for fault diagnosis in complex systems. *Mech. Syst. Signal Process.* **2019**, *117*, 170–187. [CrossRef]
10. Onan, A. Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 2098–2117. [CrossRef]

11. Wang, H.; Li, S.; Song, L.; Cui, L. A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals. *Comput. Ind.* **2019**, *105*, 182–190. [CrossRef]

12. Sun, Y.; Li, S. Bearing fault diagnosis based on optimal convolution neural network. *Measurement* **2022**, *190*, 110702. [CrossRef]

13. Youcef Khodja, A.; Guersi, N.; Saadi, M.N.; Boutasseta, N. Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks. *Int. J. Adv. Manuf. Technol.* **2019**, *106*, 1737–1751. [CrossRef]

14. Cheng, Y.; Lin, M.; Wu, J.; Zhu, H.; Shao, X. Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network. *Knowl.-Based Syst.* **2021**, *216*, 106796. [CrossRef]

15. Jiang, G.; He, H.; Yan, J.; Xie, P. Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox. *IEEE Trans. Ind. Electron.* **2019**, *66*, 3196–3207. [CrossRef]

16. Deng, F.; Ding, H.; Yang, S.; Hao, R. An improved deep residual network with multiscale feature fusion for rotating machinery fault diagnosis. *Meas. Sci. Technol.* **2020**, *32*, 024002. [CrossRef]

17. Huang, G.-B.; Bai, Z.; Kasun, L.L.C.; Vong, C.M. Local Receptive Fields Based Extreme Learning Machine. *IEEE Comput. Intell. Mag.* **2015**, *10*, 18–29. [CrossRef]

18. Cox, D.; Pinto, N. Beyond simple features: A large-scale feature search approach to unconstrained face recognition. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Santa Barbara, CA, USA, 21–25 March 2011; pp. 8–15.

19. Jarrett, K.; Kavukcuoglu, K.; Ranzato, M.; LeCun, Y. What Is the Best Multi-Stage Architecture for Object Recognition? In Proceedings of the IEEE 12th International Conference on Computer Vision (ICCV), Kyoto, Japan, 29 September–2 October 2009; pp. 2146–2153.

20. Zhu, J.; Chen, N.; Peng, W. Estimation of Bearing Remaining Useful Life Based on Multiscale Convolutional Neural Network. *IEEE Trans. Ind. Electron.* **2019**, *66*, 3208–3216. [CrossRef]

21. Jin, Y.; Qin, C.; Zhang, Z.; Tao, J.; Liu, C. A multi-scale convolutional neural network for bearing compound fault diagnosis under various noise conditions. *Sci. China Technol. Sci.* **2022**, *65*, 2551–2563. [CrossRef]

22. Pang, Y.; Jia, L.; Zhang, X.; Liu, Z.; Li, D. Design and implementation of automatic fault diagnosis system for wind turbine. *Comput. Electr. Eng.* **2020**, *87*, 106754. [CrossRef]

23. Wang, D.; Chen, Y.; Shen, C.; Zhong, J.; Peng, Z.; Li, C. Fully interpretable neural network for locating resonance frequency bands for machine condition monitoring. *Mech. Syst. Signal Process.* **2022**, *168*, 108673. [CrossRef]

24. Chen, Z.; Gryllias, K.; Li, W. Mechanical fault diagnosis using Convolutional Neural Networks and Extreme Learning Machine. *Mech. Syst. Signal Process.* **2019**, *133*, 106272. [CrossRef]

25. Li, X.; Yang, Y.; Hu, N.; Cheng, Z.; Cheng, J. Discriminative manifold random vector functional link neural network for rolling bearing fault diagnosis. *Knowl.-Based Syst.* **2021**, *211*, 106507. [CrossRef]

26. Han, S.; Shao, H.; Cheng, J.; Yang, X.; Cai, B. Convformer-NSE: A Novel End-to-End Gearbox Fault Diagnosis Framework under Heavy Noise Using Joint Global and Local Information. *IEEE/ASME Trans. Mechatron.* **2022**, *28*, 340–349. [CrossRef]

27. Zhong, H.; Lv, Y.; Yuan, R.; Yang, D. Bearing fault diagnosis using transfer learning and self-attention ensemble lightweight convolutional neural network. *Neurocomputing* **2022**, *501*, 765–777. [CrossRef]

28. Wang, Y.; Ding, X.; Zeng, Q.; Wang, L.; Shao, Y. Intelligent Rolling Bearing Fault Diagnosis via Vision ConvNet. *IEEE Sens. J.* **2021**, *21*, 6600–6609. [CrossRef]

29. Liu, Y.; Miao, C.; Li, X.; Ji, J.; Meng, D. Research on the fault analysis method of belt conveyor idlers based on sound and thermal infrared image features. *Measurement* **2021**, *186*, 110177. [CrossRef]

30. Zhang, W.; Li, J.; Li, T.; Ge, S.; Wu, L. Research on feature extraction and separation of mechanical multiple faults based on adaptive variational mode decomposition and comprehensive impact coefficient. *Meas. Sci. Technol.* **2022**, *34*, 025110. [CrossRef]

31. Wang, J.; He, Q.; Kong, F. Automatic fault diagnosis of rotating machines by time-scale manifold ridge analysis. *Mech. Syst. Signal Process.* **2013**, *40*, 237–256. [CrossRef]

32. Wang, Z.; Zhang, Q.; Xiong, J.; Xiao, M.; Sun, G.; He, J. Fault Diagnosis of a Rolling Bearing Using Wavelet Packet Denoising and Random Forests. *IEEE Sens. J.* **2017**, *17*, 5581–5588. [CrossRef]

33. Zhang, P.; Wen, G.; Dong, S.; Lin, H.; Huang, X.; Tian, X.; Chen, X. A Novel Multiscale Lightweight Fault Diagnosis Model Based on the Idea of Adversarial Learning. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3518415. [CrossRef]

34. Liu, H.; Wang, Y.; Li, F.; Wang, X.; Liu, C.; Pecht, M.G. Perceptual Vibration Hashing by Sub-Band Coding: An Edge Computing Method for Condition Monitoring. *IEEE Access* **2019**, *7*, 129644–129658. [CrossRef]

35. Liang, L.; Liu, F.; Li, M.; He, K.; Xu, G. Feature selection for machine fault diagnosis using clustering of non-negation matrix factorization. *Measurement* **2016**, *94*, 295–305. [CrossRef]