applied sciences

MDPI

Article

Adaptive Multi-Channel Residual Shrinkage Networks for the Diagnosis of Multi-Fault Gearbox

Wenxian Chen [1], Kuangchi Sun [2], Xinxin Li [1,*], Yanan Xiao [1], Jiangshu Xiang [1] and Hanling Mao [1]

[1] School of Mechanical Engineering, Guangxi University, Nanning 530004, China
[2] College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400030, China
* Correspondence: lixinx@gxu.edu.cn

Abstract: Intelligent fault diagnosis is a hot research topic in machinery and equipment health monitoring. However, most intelligent fault diagnosis models have good performance in single fault mode, but poor performance in multiple fault modes. In real industrial scenarios, the interference of noise also makes it difficult for intelligent diagnostic models to extract fault features. To solve these problems, an adaptive multi-channel residual shrinkage network (AMC-RSN) is proposed in this paper. First, a channel attention mechanism module is constructed in the residual block and a soft thresholding function is introduced for noise reduction. Then, an adaptive multi-channel network is constructed to fuse the feature information of each channel in order to extract as many features as possible. Finally, the Meta-ACON activation function is used before the fully connected layer to decide whether to activate the neurons by the model outputs. The method was implemented in gearbox fault diagnosis, and the experimental results show that AMC-RSN has better diagnostic results than other networks under various faults and strong noises.

Keywords: deep learning; multi-fault; fault diagnosis; residual network; adaptive weights; gearbox

check for updates

Citation: Chen, W.; Sun, K.; Li, X.; Xiao, Y.; Xiang, J.; Mao, H. Adaptive Multi-Channel Residual Shrinkage Networks for the Diagnosis of Multi-Fault Gearbox. *Appl. Sci.* **2023**, *13*, 1714. https://doi.org/10.3390/app13031714

Academic Editor: Abílio Manuel Pinho de Jesus

Received: 28 December 2022
Revised: 13 January 2023
Accepted: 23 January 2023
Published: 29 January 2023

Copyright: © 2023 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

Mechanical troubleshooting is vital for the reliable operation of modern industrial systems. Gearboxes are the main transmission components of industrial equipment that have been widely used in aircraft, wind turbines, automobiles, and other equipment [1,2], whereas gearboxes often operate in a complex and harsh working environment for a long time, and their components often work continuously at high speed with heavy load leading to a high failure rate and coupling faults [3]. Therefore, the study of fault diagnosis methods for gearboxes should be a concern that can prolong the use of machinery and equipment, reduce economic losses, and avoid accidents.

The most effective and direct fault detection method for gearboxes is vibration signal analysis. The existing system diagnostic methods can be divided into two types, namely traditional signal analysis methods and machine learning methods [4]. Most of the traditional signal analysis methods are based on physical and mathematical principles to extract and detect the fault-related features in the original signal [5]. However, the working environment of gearboxes is complex, so the vibration signal usually contains information about the characteristics of other components, such as the rotation of shafts and bearings, and the meshing of gears. Moreover, fault coupling of different components may also occur under long-term operation. These factors add difficulties to gearbox fault diagnosis. In addition, the vibration signal of the component associated with the fault is thoroughly weak and can be susceptible to being swamped by other vibrating components in the early stages [6]. Therefore, traditional signal-based analysis methods are difficult to identify the vibration components associated with a fault.

On the other hand, intelligent fault diagnosis has come a long way with the development of machine learning theory, and prevalent machine learning methods include artificial

neural networks (ANN) [7], support vector machines (SVM) [8], and k-means method [9], etc. The intelligent diagnosis steps of machine learning are divided into signal acquisition, feature extraction, and machine learning model classification. However, the difficulty of machine learning fault diagnosis lies in finding an effective feature [10].

Recently, with the development of artificial intelligence, deep learning (DL) methods have become a practical tool for fault diagnosis based on vibration signals [11]. Deep learning methods are machine learning methods with multi-level nonlinear transformations. However, in real industrial scenarios, vibration signals collected by the gearbox contain a lot of noise, and excessive noise will drown out the signal that characterizes the fault. In this case, a general DL mode can make a rough judgment. Therefore, minimizing the impact of noise on diagnostic models is the focus of current research. Several researchers have studied intelligent fault diagnosis in environments with strong noise and variable loads. Su et al. [12] proposed a hierarchical branching CNN (HB-CNN) method with excellent robustness in bearing fault diagnosis, Zhang et al. [13] proposed a training interference convolutional neural networks (TICNN) method that shows excellent performance in a noisy environment, and Mo et al. [14] proposed a new method of integrating a learnable variational kernel into a one-dimensional CNN. To further improve the diagnostic performance of the network model, many researchers have given the model a stronger ability to extract features. Yu et al. [15] proposed a broad convolutional neural network (BCNN) with certain incremental learning abilities. Li et al. [16] constructed a CNN-BGM model with a mixture of neural networks and Bayesian Gaussians. Jiao et al. [17] proposed a deeply coupled dense convolutional network (CDCN).

Neural networks are generally optimized by a backpropagation algorithm which updates parameters by the gradient. Some networks obtain better performance by increasing the number of layers of the network, which tends to the disappearance of the gradient, and the underlying parameters cannot be updated. Deep residual networks (ResNet), a popular derivative of convolutional networks, use identity shortcuts to alleviate the difficulty of parameter optimization. In residual networks, optimization of the underlying parameters can be solved because gradients can be imported with identity shortcuts [18,19]. Chen et al. [20] proposed a dual-path mixed-domain residual threshold network (DP-MRTN) to improve the diagnostic performance of rolling bearings in high-noise environments. Li et al. [21] proposed a one-dimensional residual convolution neural network (1D-RCNN) that directly uses time-domain waveforms as input. Zhao et al. [22] used a residual network to fuse multiple sets of wavelet packet coefficients for fault diagnosis. Sun et al. [23] proposed a multi-scale cluster-graph convolution neural network with a multi-channel residual network (MR-MCGCN), which can eliminate the effect of noise efficiently.

Many researchers have found that the vibration signals of gearboxes have inherent multiscale features. Jiang et al. [24] proposed a new multiscale convolutional neural network (MSCNN) with coarse granularity strategy to effectively learn fault characteristics of different time scales. Chen et al. [25] proposed a multiscale convolutional neural network with feature alignment (MSCNN-FA). Liu et al. [26] proposed a multiscale kernel-based residual convolutional neural (MK-ResCNN) network. Chao et al. [27] proposed a multiscale cascaded midpoint residual convolutional neural network (MSC-MpResCNN).

The above methods focus on a single fault or a compound fault of a single part. In the actual industrial scenario, faults often occur in multiple parts of the entire mechanical system at the same time, which brings a great challenge to fault diagnosis. A few researchers have been conducted on this problem. Li et al. [28] proposed the multivariate variational mode decomposition (MVMD) method, Yuan et al. [29] proposed an Adaptive-Projection with intrinsically transformed multivariate Empirical Mode decomposition (APIT-MEMD), and Lonare et al. [30] proposed a new morphological joint time-frequency adaptive kernel-based semi-smart framework. However, multi-fault diagnosis based on deep learning is rarely studied at present.

In this paper, we propose an intelligent fault diagnosis method based on AMC-RSN to solve the above problems. Firstly, a channel attention mechanism module is constructed

in the residual block and a soft thresholding function is introduced for noise reduction in the original signal. Then, a multi-channel network is constructed to fuse the feature information of each channel to extract as many features as possible. Hereinto, adaptive weights are set in each channel to obtain the most important information, and these weights are updated adaptively as the model is trained to obtain the best values. Finally, the Meta-ACON activation function is used before the fully connected layer to decide whether to activate the neurons by the model outputs, which can improve the classification accuracy of the model. In summary, this paper has three main contributions as follows.

(1) A new multi-channel residual network is proposed for extracting richer features in the signal, which solves the problem of insufficient effective feature extraction;
(2) An adaptive learning method based on activation function is proposed to activate neurons adaptively, which can effectively avoid the interference of redundant features;
(3) In the case of multiple faults in gearboxes, the experimental results show that the proposed method can effectively extract the features of the target faults and classify them accurately.

The remainder of this paper is as follows. Section 2 presents the theory underlying the AMC-RSN of this paper. Section 3 presents the principles and architecture of the proposed AMC-RSN. In Section 4, the validation of the method and the comparison of related models are carried out and the results are discussed. Section 5 provides a full-text summary.

2. Basic Theory

This section introduces convolutional neural networks, then explains the residual blocks clearly. The ACON family of adaptive activation functions is described at the end.

2.1. Convolutional Neural Networks (CNN)

CNN has two prominent features, i.e., sparse connectivity and weight sharing. The two most important elements of CNN are the convolutional layer and pooling layer, while the fully connected layer will be classified.

(1) Convolutional Layer: There are several convolutional kernels in the convolutional layer, and the raw data are convolved to obtain data features by convolutional operations. The formula is described as follows:

$$S_j^l = f(\sum_{i \in N} S_i^{l-1} * \omega_{ij}^l + b_j^l) \tag{1}$$

where S_j^l is the output feature of the jth convolutional kernel in layer l, $f(\cdot)$ is the activation function, S_i^{l-1} is the output feature of the ith convolutional kernel in layer $l-1$, ω_{ij}^l is the weights, b_j^l is the bias, and N is the target convolution range.

(2) Pooling Layer: In pooling operations, the space size of the data body is reduced by using maximum pooling. Thus, the overfitting phenomenon of the network can be effectively controlled. The formula is described as follows:

$$x_{il} = Max(x_{i'(l-1)}), i \le i' \le i + m \tag{2}$$

where are all eigenvalues in the sampling range of layer $l-1$, m is the width of a pooling window, x_{il} is the corresponding maximum pooled sampling value, and Max $(\cdot)$ is the maximum value function.

2.2. Residual Block

The residual network is a variant of CNN. Due to its special structure, the residual network has applications in various fields. In the field of intelligent fault diagnosis, many researchers have verified the residual network. The residual network consists of several residual blocks, each of which contains two convolution pooling layers. Classical residuals change the output of the residuals block, namely the residuals function, through identify

shortcuts learning. Therefore, when the optimal residual network is obtained by training, the training error is close to zero and the identity mapping between input and output is formed. This procedure can solve the problem of gradient disappearance in back-propagation to update the whole network parameters. The residual block is described by the following formula:

$$H(x) = F(x) + x \tag{3}$$

where $H(x)$ is the output, $F(x)$ is the output mapping of residual blocks, and x is the feature vector.

2.3. Adaptive Activation Function

Ma et al. [31] proposed a practical activation function, i.e., ACON. The ACON function is a smoothed form of the ReLU function, which decides whether to activate the neurons by the model outputs. And the differential variant of the ACON function can be expressed as follows:

$$S_\alpha(x_1, \cdots, x_n) = \frac{\sum_{i=1}^{n} x_i e^{\alpha x_i}}{\sum_{i=1}^{n} e^{\alpha x_i}} \tag{4}$$

where α is the smoothing coefficient. When $\alpha \to \infty$, then $S_\alpha \to$ max. If $\alpha \to 0$, $S_\alpha \to$ mean value. The Formula (4) can be generalized with the model $[f(x), g(x)]$, and the model can be further refined to obtain as follows:

$$\begin{aligned} S_\alpha(f(x), g(x)) &= f(x) \cdot \frac{e^{\alpha f(x)}}{e^{\alpha f(x)} + e^{\alpha g(x)}} + g(x) \cdot \frac{e^{\alpha g(x)}}{e^{\alpha f(x)} + e^{\alpha g(x)}} \\ &= f(x) \cdot \frac{1}{1 + e^{-\alpha(f(x) - g(x))}} + g(x) \cdot \frac{1}{1 + e^{-\alpha(g(x) - f(x))}} \\ &= f(x) \cdot \sigma[\alpha(f(x) - g(x))] + g(x) \cdot \sigma[\alpha(g(x) - f(x))] \\ &= (f(x) - g(x)) \cdot \sigma[\alpha(f(x) - g(x))] + g(x) \end{aligned} \tag{5}$$

when $f(x) = x$, $g(x) = 0$, and it is equivalent to ReLU activation function.

In addition, Ma et al. [31] further proposed Meta-ACON, which can adaptively adjust activation or not according to the input. Meta-ACON is an application of ACON-C in neural networks. ACON-C can be described as follows:

$$f(x) = \lambda x, g(x) = \mu x (\lambda \neq \mu) \tag{6}$$

$$f_{\text{ACON-C}}(x) = S_\alpha(\lambda x, \mu x) = (\lambda - \mu) x \cdot \sigma[\alpha(\lambda - \mu) x] + \mu x. \tag{7}$$

λ and μ are channel constraints. The first-order derivative of the ACON-C is as follows:

$$\frac{d}{dx}[f_{\text{ACON-C}}(x)] = \frac{(\lambda - \mu)(1 + e^{-\alpha(\lambda x - \mu x)}) + \alpha(\lambda - \mu)^2 (1 + e^{-\alpha(\lambda x - \mu x)}) x}{(1 + e^{-\alpha(\lambda x - \mu x)})^2} + \mu \tag{8}$$

$$\lim_{x \to \infty} \frac{d}{dx}[f_{\text{ACON-C}}(x)] = \lambda, \lim_{x \to -\infty} \frac{d}{dx}[f_{\text{ACON-C}}(x)] = \mu (\alpha > 0) \tag{9}$$

The second-order derivatives are as follows:

$$\begin{aligned} \frac{d^2}{dx^2}[f_{\text{ACON-C}}(x)] &= \frac{\alpha(\mu - \lambda)^2 e^{\alpha(\lambda - \mu)x}}{(1 + e^{\alpha(p_1 - p_2)x})^3} \\ &\cdot \frac{(\alpha(\mu - \lambda)x + 2) e^{\alpha(\lambda - \mu)x} + \alpha(\lambda - \mu)x + 2}{(1 + e^{\alpha(\lambda - \mu)x})^3} \end{aligned} \tag{10}$$

Making its second-order derivative equal to 0, and the first-order derivative upper and lower bounds are obtained as follows:

$$\text{maxima}\left(\frac{df_{\text{ACON-C}}(x)}{dx}\right) \approx 1.0998\lambda - 0.0998\mu \tag{11}$$

$$\text{minima}\left(\frac{df_{\text{ACON-C}}(x)}{dx}\right) \approx 1.0998\mu - 0.0998\lambda \tag{12}$$

ACON-C has a threshold range of $(-0.0998, 1.0998)$ in the first-order derivative. These bounds are learnable and are determined by λ and μ. Based on ACON-C, Meta-ACON averages the channels and then passes two convolution layers so that all pixels in each channel share a weight.

3. Methodology

3.1. Proposed Adaptive Multi-Channel Residual Shrinkage Block

(1) Channel attention mechanism: In the general convolution process, all feature information is uniformly extracted, so the representation information in the original signal is easily ignored in the noisy environment. This section describes the channel attention mechanism to improve the performance of the fault diagnosis model. The attention mechanism is added to the channel (convolutional kernel) dimension to obtain the significance of each channel through a fully connected layer (FC). Then weights are assigned to each channel according to the significance level so that the neural network pay attention to the information of a certain channel. This mechanism can promote the information channels that are useful for the current task and suppress the channels that are not very useful for the current task.

This paragraph describes some specific operations of the attention mechanism. First, the size of the output data obtained after inputting data of size C × W × 1 to the two convolutional layers is unchanged, and the feature S of size C × 1 × 1 is obtained by global average pooling (GAP) compression. Then, feature S is divided into two branches, and in branch 1, the feature S does not undergo any transformation. In branch 2, feature S acquires channel-level weights through two FCs. The channel attention coefficients w_c are obtained after the sigmoid activation function as follows:

$$w_c = \sigma(\text{FC}_2 \cdot \text{ReLU}(\text{FC}_1 \cdot \eta)) \tag{13}$$

where η is the feature vector. Finally, as shown in Figure 1, the features from the two branches are multiplied to generate the channel attention feature S_c.

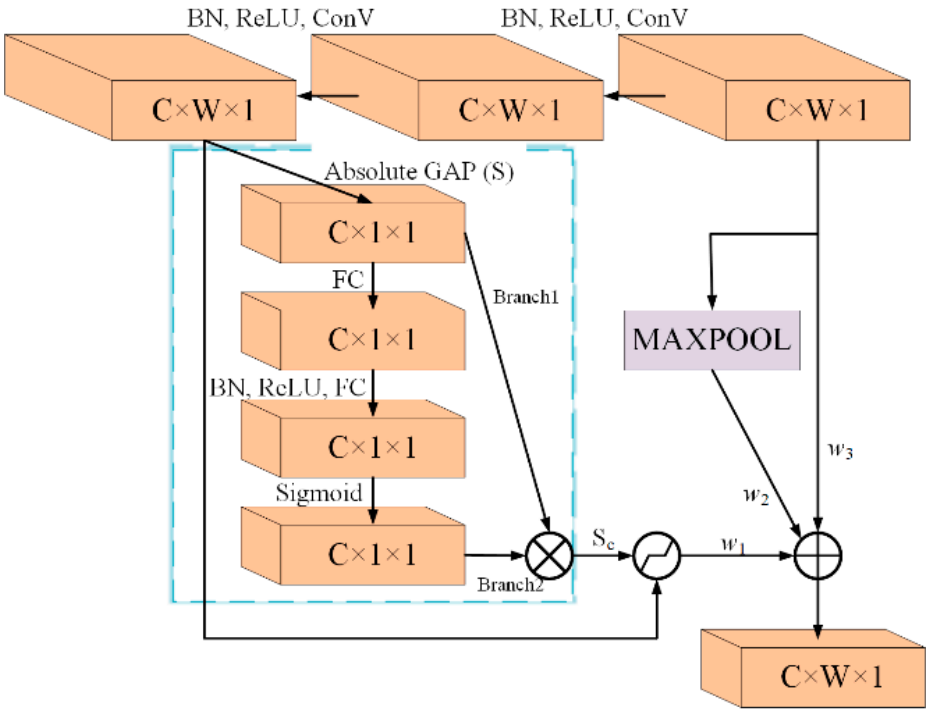


Figure 1. AMC-RSB, where the blue dotted line frame is the channel attention mechanism module.

(2) Soft threshold: To reduce the noise and improve the diagnostic performance of the network, the basic block inserts a soft threshold function [32] to eliminate irrelevant features of the output through the attention mechanism module. The formula for the soft threshold function is expressed as follows:

$$y = \begin{cases} \omega + T & \omega \leq -T, \\ 0 & -T < \omega \leq T, \\ \omega - T & \omega \geq T. \end{cases} \tag{14}$$

where ω is the input value, and T is the threshold. The idea of a deep residual shrinkage network [7] is used to construct the threshold, which is automatically determined according to the inputs. The threshold value is calculated by a series of matrices from the outputs of two convolution layers and the channel attention mechanism module.

(3) Adaptive multi-channel residual shrinkage block: The diagnostic model can resist interference, but it is not yet able to extract obvious features in fault signals. A maximum pooling branch is set at the input as a way to extract more obvious features of the input signal. Additionally, since each channel output feature has some ability to characterize the faulty features, weights are set in each branch. These weights are back-propagated to update the values adaptively based on the training of the model. The designed adaptive multi-channel residual shrinkage block (AMC-RSB) is shown in Figure 1.

3.2. Structure of Adaptive Multi-Channel Residual Shrinkage Network

The proposed diagnostic model was designed with four AMC-RSB based on the resnet18 [21]. To obtain more classification accuracy, the Meta-ACON activation function is used before the fully connected layer. Meta-ACON activation function can calculate whether to activate the channel by the output of each channel, can effectively filter the interference channel, thus improving the diagnostic accuracy. The overall architecture and specific parameters of the proposed network are shown in Figure 2.

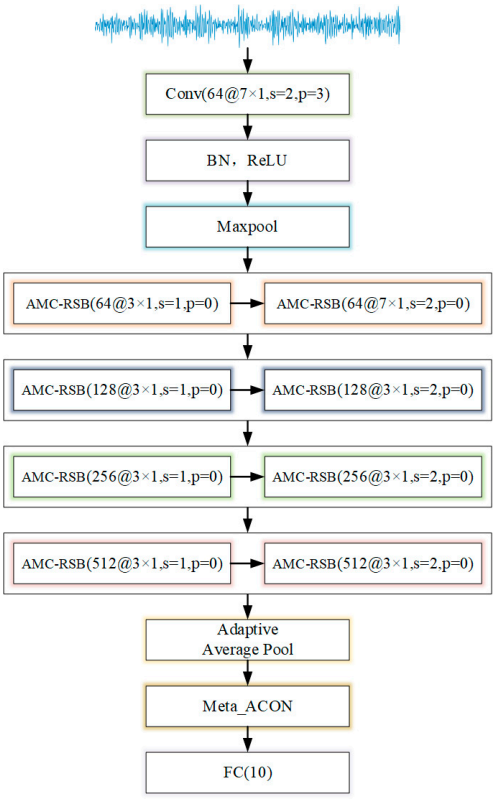


Figure 2. Overall architecture and specific parameters.

In this paper, an end-to-end fault diagnosis process is proposed. First, the time series signal is sampled overlappingly to obtain the data-enhanced samples, and the samples are divided into training set and test set. Then, the training set is fed into the model to train the diagnostic model. Finally, the trained model is tested with the test set to obtain the classification results and analyzed. The overall process of the proposed AMC-RSN is shown in Figure 3.

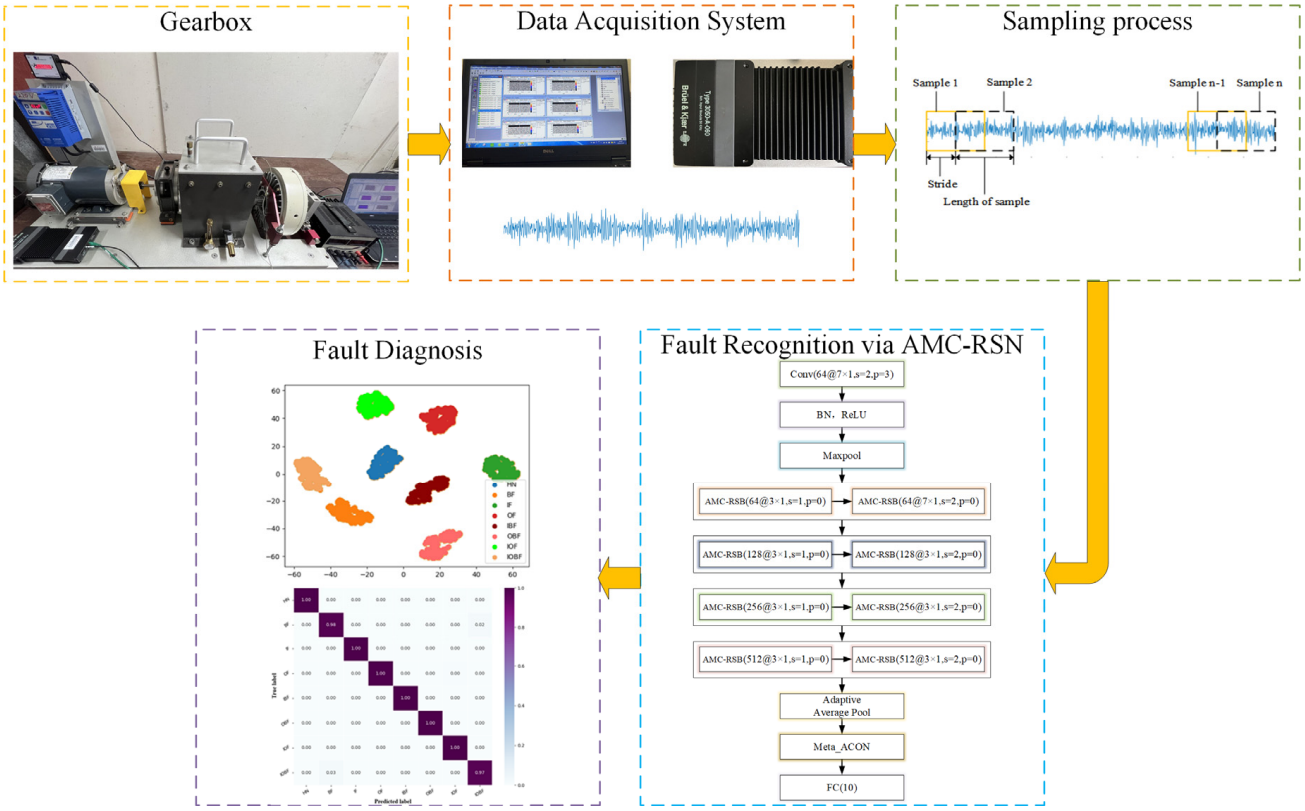


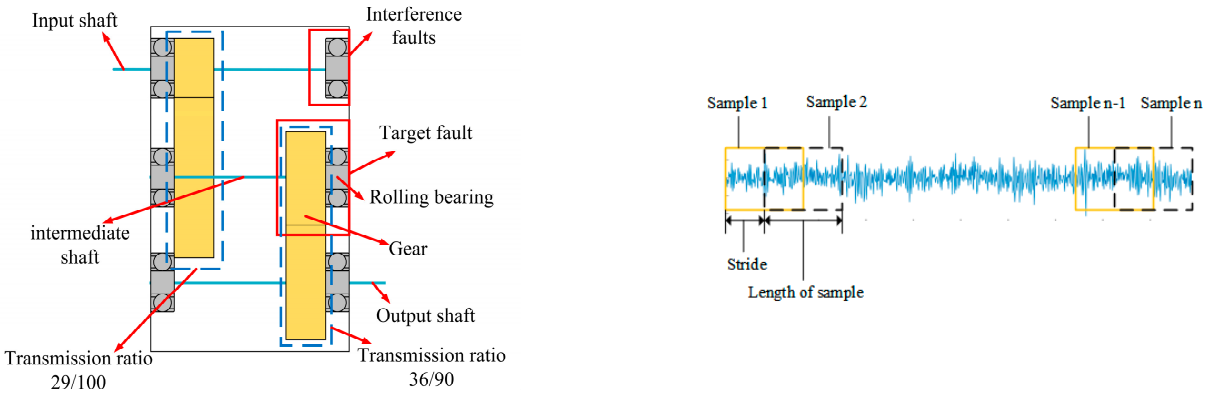
Figure 3. Diagnostic process of AMC-RSN.

4. Experimental Validation and Analysis

To test the validity of the proposed AMC-RSN, two case studies are conducted in this section. The two different cases used for testing include the rolling bearing and gear datasets. To verify the anti-interference of the proposed AMC-RSN, it is worth noting that the collected vibration signals contain two faulted parts, one of which act as interference item. The experiment setup is shown as Figure 4a. The parallel gearbox contains two stages of gear reduction, with a primary ratio of 0.29 and a secondary ratio of 0.4. Two cases are conducted to study the effects of the proposed method. In both cases, the right-side rolling bearing of the input shaft is pre-set with ball fault, treated as an interference faulted part. In Case I, different fault types of rolling bearings (gears are healthy) in the intermediate shaft are combined with the above interference faulted part to form the multi-part multi-fault experiment setup. In Case II, different fault types of gears (rolling bearings are healthy) in the intermediate shaft are combined with the above interference faulted part to form the multi-part multi-fault experiment setup. In Case III, different mixed fault types of bearings and gears are set up in the intermediate shaft, and no interference fault is set in the input axis. Due to this setup, along with the noise generated by the machine operation, it will inevitably pose a challenge to the detection and classification of target faults.

To enhance the dataset, all samples are overlapping samples with a sample length of 1024 and a stride of 512, the sampling form is shown in Figure 4b. Vibration data from the different datasets were collected using the SpectraQuest Drivetrain Dynamics Simulator (DDS). The data are collected by the accelerometer, and the arrangement of the sensor

is shown in Figure 5, No. 6. The sampling time is 43.5 s, and the sampling frequency is 12,800 Hz. The length of each fault signal is 556,800, so the number of samples for each fault type after overlapping sampling is 1086. In all experiments, 80% of the training samples and 20% of the test samples are used. The details of the DDS are shown in Figure 5.



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

Figure 4. Experiment preparation. (**a**) Details of the fault arrangement. (**b**) Overlap sampling.

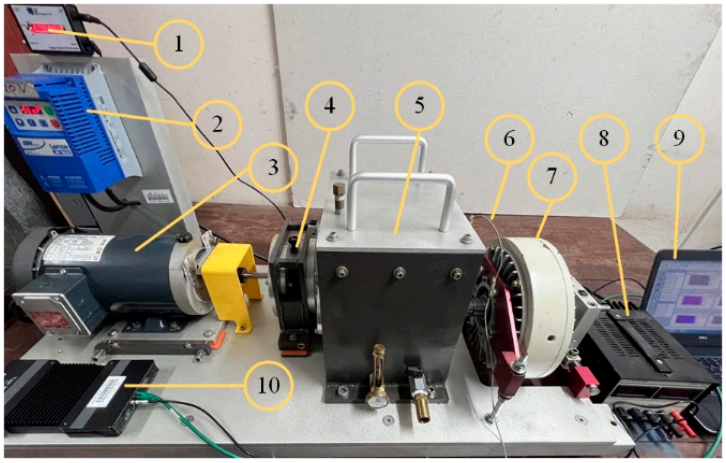


Figure 5. Specific arrangement of DDS. 1. Speedometer. 2. Converter. 3. Drive motor. 4. Planetary gearbox. 5. Parallel gearbox. 6. Accelerometer. 7. Load. 8. Load controller. 9. Laptop. 10. Acquisition system.

The program uses PyTorch DL software with the Python 3.7 language. The CPU of the computer is configured as Intel Core i7-9750H and the GPU is configured as NVIDIA RTX2060. In the validation session, each test model is the optimal model obtained by 100 epochs of training. To avoid randomness, each method was trained ten times and ten test models were obtained. The Adam optimizer is used to optimize the learning rate during the training process. The loss function is a cross-entropy function.

4.1. Case I

(1) Rolling Bearing Datasets: In the datasets, there are mainly eight fault states, namely health (HN), ball fault (BF), inner ring fault (IF), outer ring fault (OF), inner ring, and ball compound fault (IBF), outer ring and ball compound fault (OBF), inner and outer ring compound fault (IOF), and inner and outer ring and ball compound fault (IOBF). The datasets were collected at a motor speed of 3120 rpm with 2 V (Torque: 4.06 N·m) load. The datasets of rolling bearing failures for eight different failure types are shown in Table 1.

Table 1. Details of the rolling bearing datasets.

Condition	Fault Type	Training Set	Test Set
3120 rpm-2V	HN	868	218
	BF	868	218
	IF	868	218
	OF	868	218
	IBF	868	218
	OBF	868	218
	IOF	868	218
	IOBF	868	218

(2) Analysis of Fault Diagnosis Results in Rolling Bearing Datasets: To show the superiority of the proposed method, a comparison with other DL methods is conducted in this section. The methods used for comparison contain CNN, deep residual shrinkage network with channel-wise thresholds (DRSN-CW) [6], multiscale cascade midpoint residual convolutional neural network (MSC-MpResCNN) [27], and multiscale kernel-based ResCNN (MK-ResCNN) [26]. To verify the effectiveness of the multi-channel strategy and Meta-ACON, adaptive residual shrinkage network with Meta-ACON activation function (ARSN-M) and adaptive multi-channel residual shrinkage network (AMRS) without Meta-ACON are used for comparison.

In Figure 6, we observe that AMC-RSN has the highest classification accuracy in all methods. We can also find that CNN has the worst diagnosis accuracy. Among these different residual networks, AMC-RSN can obtain the best diagnostic results and the highest time consumption. Compared to AMRS and ARSN-M, AMC-RSN can learn the classifier weights adaptively to improve classification accuracy and extract richer features. The tested bearing is shown in Figure 7. From Table 2, it is obvious that the average classification diagnosis accuracy of AMC-RSN is the highest, i.e., 99.24%, which is better than the remaining methods. In addition, the proposed method has the smallest standard deviation of 0.11%, which reflect AMC-RSN has robustness.

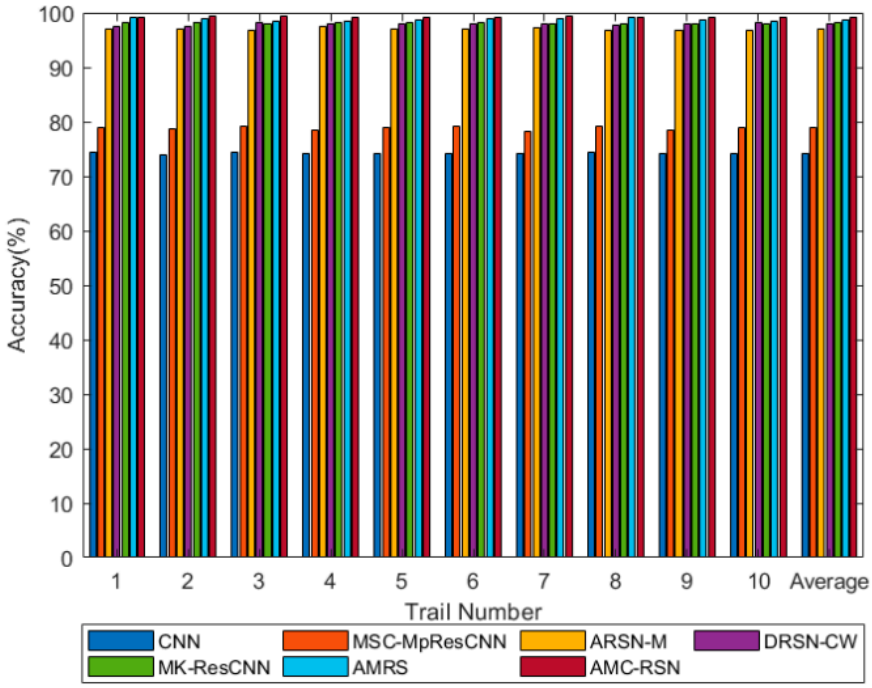


Figure 6. Diagnostic results of different models for the bearing datasets.



Figure 7. These three pictures show the front, back, and side of the test rolling bearing respectively.

Table 2. Fault diagnosis results of rolling bearing.

Datasets	Model	Min − acc (%)	Max − acc (%)	Avg − acc (%)	Avg-Time (s)
Rolling bearing datasets	CNN	73.95	74.34	74.23 ± 0.1	274
	MSC-MpResCNN	78.17	79.26	78.81 ± 0.32	562
	ARSN-M	96.73	97.42	97.00 ± 0.23	1729
	DRSN-CW	97.53	98.28	97.91 ± 0.22	1426
	MK-ResCNN	97.87	98.29	98.10 ± 0.13	1823
	AMRS	98.34	99.2	98.78 ± 0.30	1802
	AMC-RSN	99.08	99.43	99.24 ± 0.11	1922

4.2. Case II

(1) Gear datasets: To further verify the effectiveness of the proposed method, different fault types in the gear datasets are collected and used for validation. As shown in Figure 8, there are six different damaged types of gears.

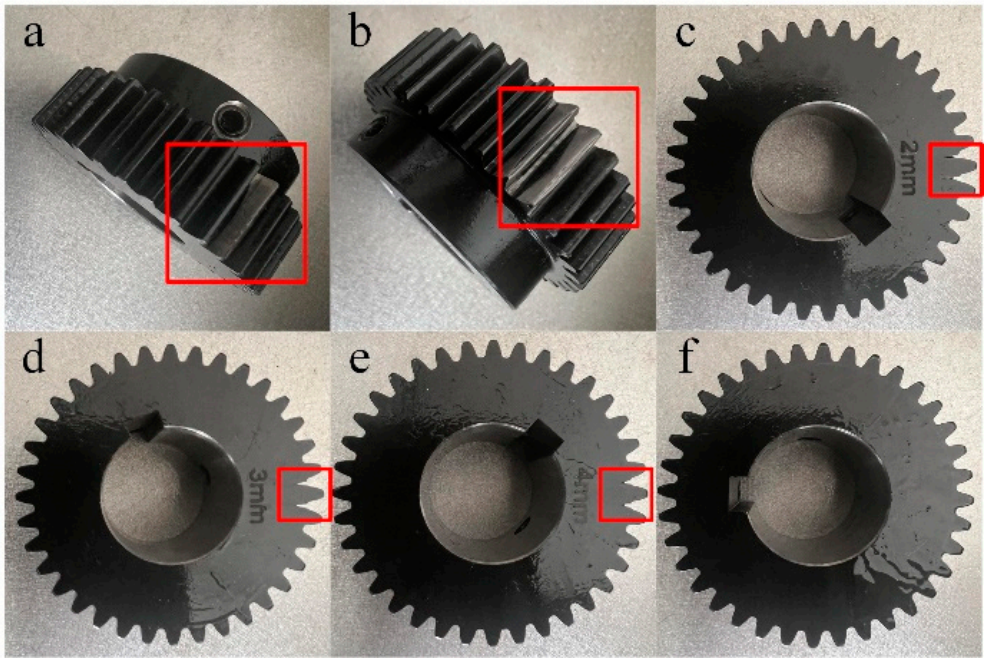


Figure 8. Gears with different failure modes. (**a**) 1-tooth worn. (**b**) 2-tooth worn. (**c**) 2 mm crack. (**d**) 3 mm crack. (**e**) 4 mm crack. (**f**) Healthy.

As is shown in Table 3, the gear datasets contain six different fault types, i.e., healthy gear, 1-tooth worn gear, 2-tooth worn gear, 2 mm crack gear, 3 mm crack gear, and 4 mm crack gear. The latter three damages are different crack lengths in the tooth root radial direction. The datasets were collected at a motor speed of 3120 rpm with 2 V (Torque: 4.06 N·m) load.

Table 3. Details of the gear datasets.

Condition	Fault Type	Training Set	Test Set
3120 rpm-2V	Healthy	868	218
	1-tooth worn	868	218
	2-tooth worn	868	218
	2 mm crack	868	218
	3 mm crack	868	218
	4 mm crack	868	218

(2) Analysis of Fault Diagnosis Results in Gear Datasets: As is shown in Figure 9, the proposed AMC-RSN has the best diagnostic effect and the lowest diagnostic accuracy of CNN compared with other methods. From Table 4, it is noteworthy that the proposed AMC-RSN can obtain the highest accuracy of 98.24% with a standard deviation of 0.16%, which reflects that AMC-RSN has good robustness. Compared with DRSN-CW, MK-ResCNN, and MSC-MpResCNN, the proposed AMC-RSN has significantly improved its ability in extracting the features of the original signal with the highest diagnosis accuracy. Besides, the AMC-RSN has superior performance when compared with the AMRS and ARSN-M.

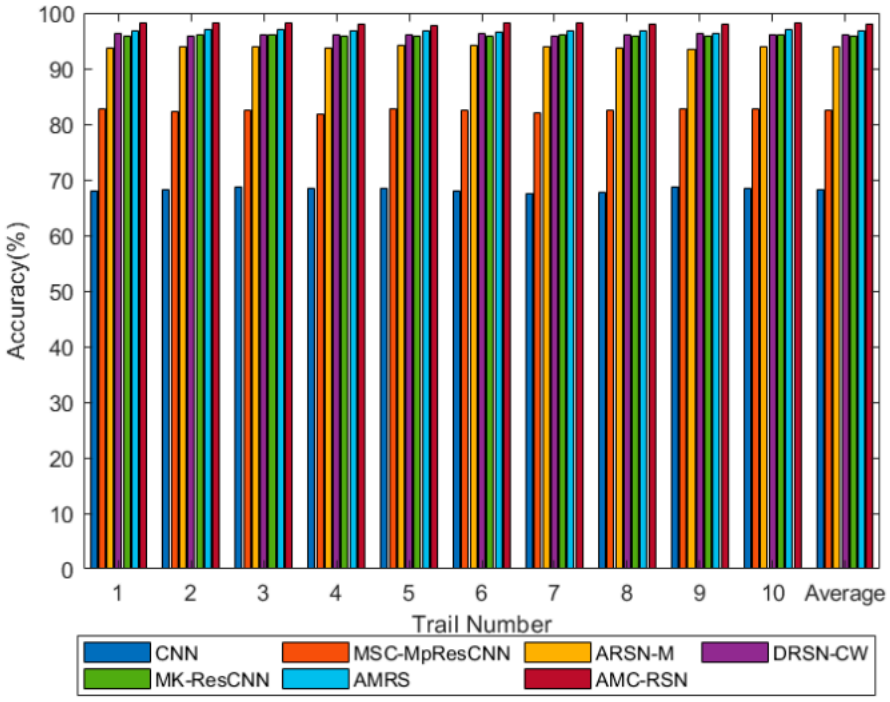


Figure 9. Diagnostic results of different models for the gear datasets.

Table 4. Fault diagnosis results in gear.

Datasets	Model	Min − acc (%)	Max − acc (%)	Avg − acc (%)	Avg-Time (s)
Gear datasets	CNN	67.59	68.78	68.24 ± 0.39	186
	MSC-MpResCNN	81.88	82.8	82.46 ± 0.31	434
	ARSN-M	93.5	94.19	93.85 ± 0.21	1377
	DRSN-CW	95.87	96.41	96.15 ± 0.15	1074
	MK-ResCNN	95.79	96.06	95.92 ± 0.10	1412
	AMRS	96.41	97.02	96.81 ± 0.18	1422
	AMC-RSN	97.78	98.24	98.06 ± 0.16	1521

4.3. Case III

(1) Mixed datasets: In order to further investigate the effectiveness of the proposed method in the diagnosis of mixed gear faults and bearing faults. As shown in Table 5, the mixed datasets contain six different mixed types of bearing faults and gear faults, including healthy bearing and healthy gear (HH), bearing with inner ring fault and 1-tooth worn gear (IWF), bearing with inner ring fault and 2 mm crack gear (ICF), bearing with outer ring fault and 1-tooth worn gear (OWF), bearing with outer ring fault and 2 mm crack gear (OCF), and bearing with ball fault and 3 mm crack gear (BCF). For consistency with the previous two cases, the datasets were collected at a motor speed of 3120 rpm with 2 V (Torque: 4.06 N·m) load.

Table 5. Details of the mixed datasets.

Condition	Fault Type	Training Set	Test Set
3120 rpm-2V	HH	868	218
	IWF	868	218
	ICF	868	218
	OWF	868	218
	OCF	868	218
	BCF	868	218

(2) Analysis of Fault Diagnosis Results in Mixed Datasets: As shown in Figure 10, our proposed AMC-RSN still has better results compared to other methods. From Table 6, the proposed AMC-RSN can obtain the highest accuracy of 99.49% with the lowest standard deviation of 0.11%, which has better robustness compared to other networks. Compared with AMRS, MK-ResCNN, DRSN-CW, and ARSN-M, the proposed AMC-RSN achieves a slight advantage, while DRSN-CW takes the shortest time. Meanwhile, the CNN and MSC-MpResCNN also achieve higher accuracy rates. Therefore, the fault features of the mixed dataset are more prominent, while the proposed AMC-RSN can learn deeper features and classify them accurately more, which takes relatively more time.

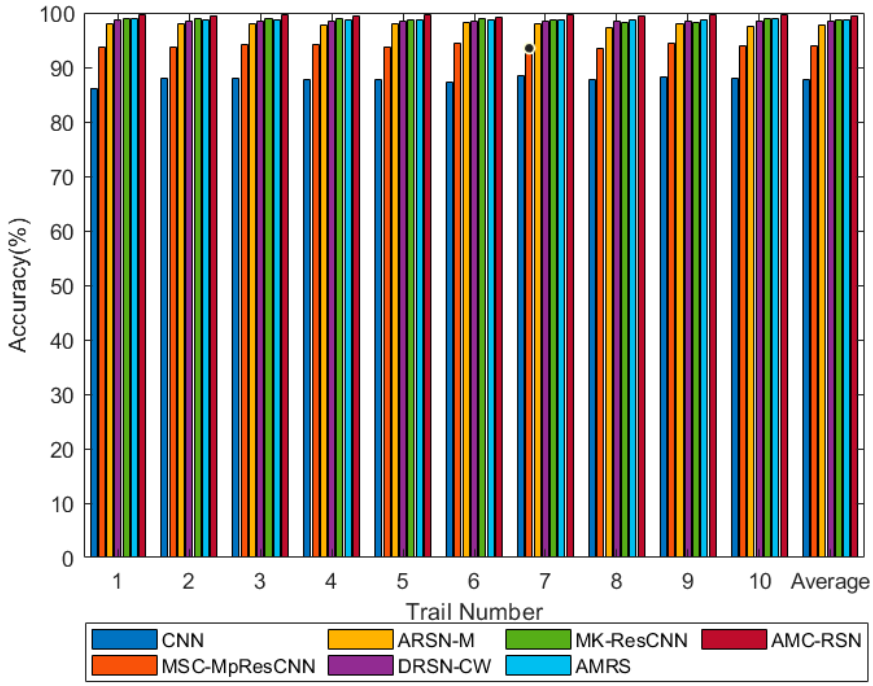


Figure 10. Diagnostic results of different models for the Mixed datasets.

Table 6. Fault diagnosis results in gear.

Datasets	Model	Min − acc (%)	Max − acc (%)	Avg − acc (%)	Avg-Time (s)
Mixed datasets	CNN	86.16	88.38	87.71 ± 0.59	184
	MSC-MpResCNN	93.38	94.42	93.90 ± 0.35	448
	ARSN-M	97.2	98.24	97.84 ± 0.28	1290
	DRSN-CW	97.98	98.62	98.44 ± 0.17	1021
	MK-ResCNN	98.13	99	98.72 ± 0.29	1357
	AMRS	98.07	98.85	98.67 ± 0.21	1421
	AMC-RSN	99.24	99.62	99.49 ± 0.11	1531

4.4. Visualization Results

In the test phase, t-SNE and confusion matrix visualization are used to illustrate the best model test results of AMC-RSN in the two cases. In Case I, the feature distribution of the classification results for rolling bearings visualized with t-SNE is shown in Figure 11a. It can be seen that the inter-class distance is small and the intra-class distance is large for each different type of data sample. Besides, the visualization result of the confusion matrix in Figure 12a shows that the classification of AMC-RSN is excellent. In Case II, the feature distribution of the classification results for gears visualized with t-SNE is shown in Figure 11b. The inter-class distances between 4 mm crack and 2 mm crack are relatively small, but other types are relatively large. According to the confusion matrix of gears in Figure 12b, the overall results are good. So, the different fault types of gear datasets can also be classified by AMC-RSN. In Case III, the feature distribution of the classification results for mixed datasets visualized with t-SNE is shown in Figure 11c. It can also be seen that the inter-class distance is small and the intra-class distance is large for each different type of data sample. From Figure 12c, the predicted and true values of the proposed method AMC-RSN are basically consistent.

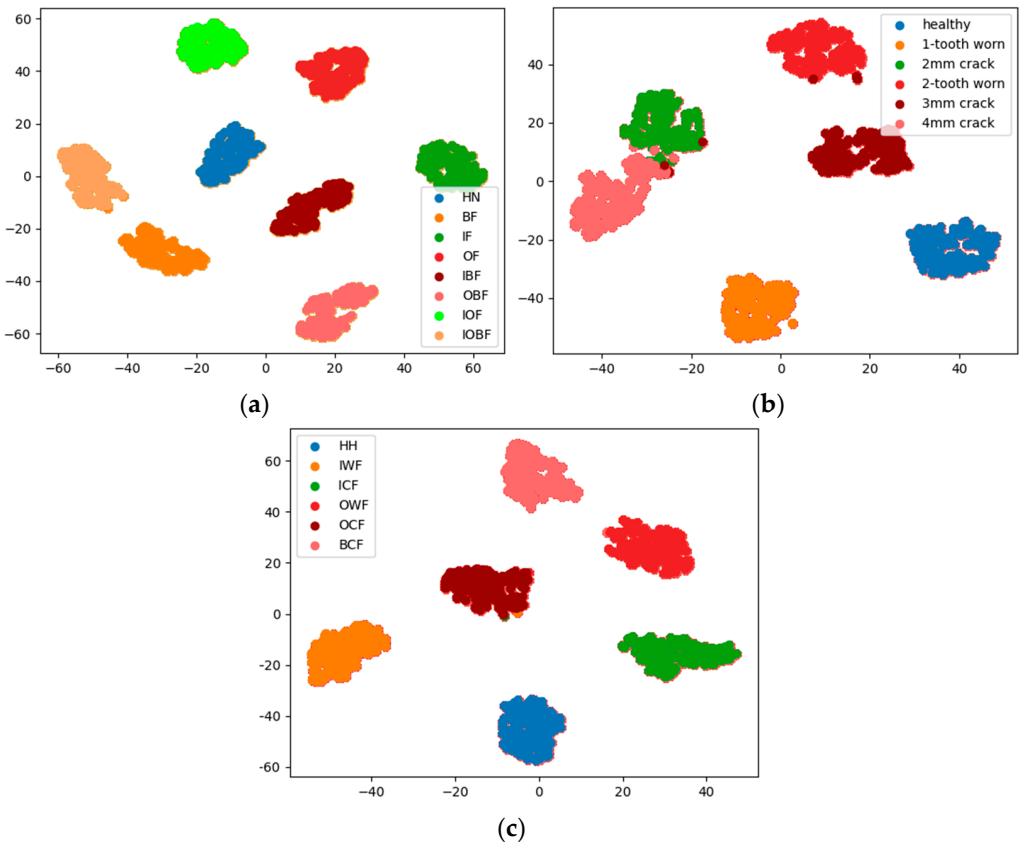


Figure 11. AMC-RSN with visualization of t-SNE. (**a**) Case I. (**b**) Case II. (**c**) Case III.

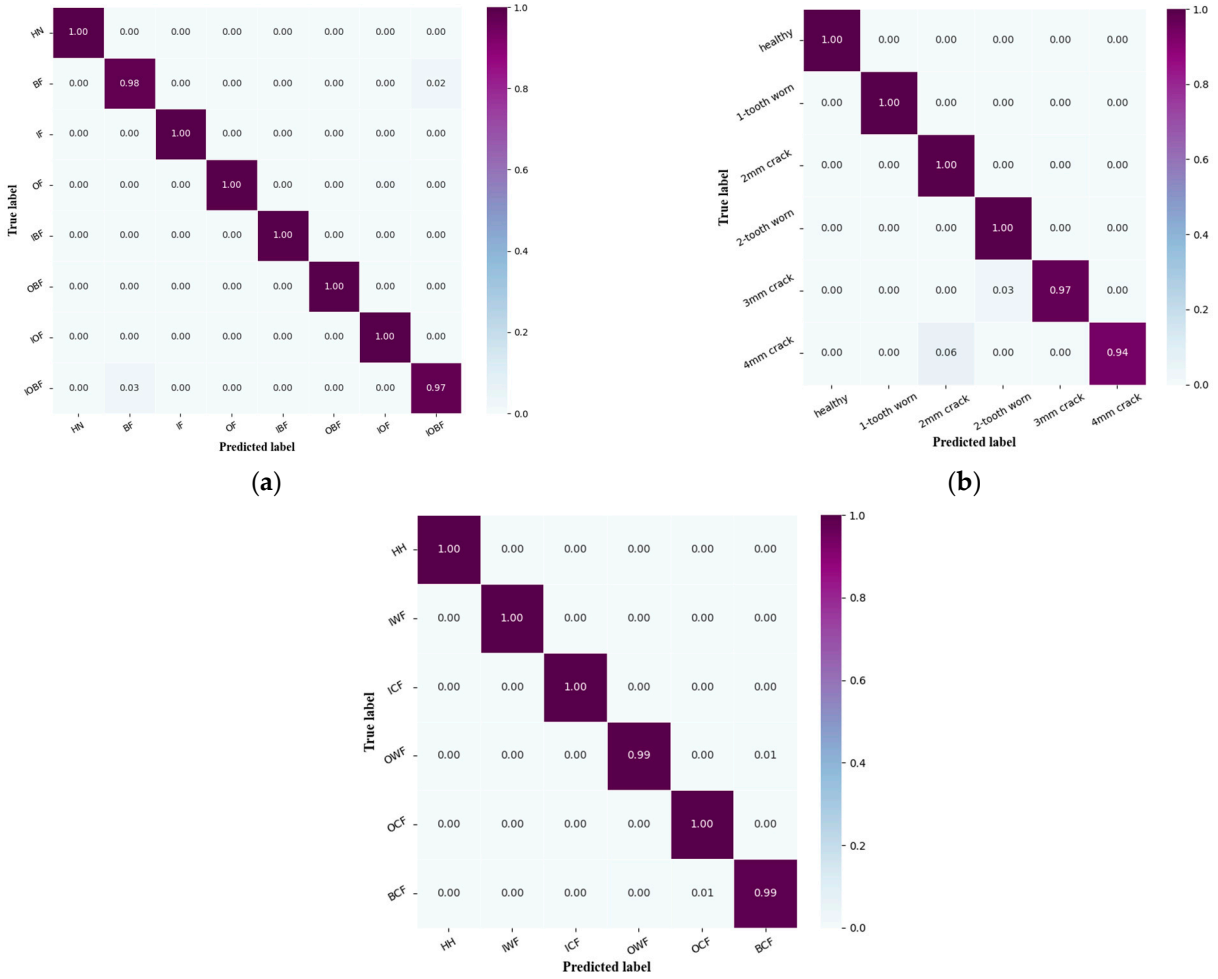


Figure 12. AMC-RSN with visualization of confusion matrix. (**a**) Case I. (**b**) Case. II. (**c**) Case III.

5. Conclusions

In this paper, fault diagnosis under multi-component faults in gearboxes is investigated and a new intelligent fault diagnosis method, namely AMC-RSN, is proposed. The effectiveness of AMC-RSN is verified by three different types of datasets. Compared with other residual networks, the experimental results show that AMC-RSN achieves the best diagnosis under different types of multi-fault classification and noise interference. The summary is as follows: (1) The maximum pooling channel can effectively extract the fault target features; (2) the introduction of the Meta-ACON activation function can effectively improve the classification accuracy; (3) through cross-comparison, it can be seen that the proposed method can extract richer features of the gearbox with multiple faults. The study is for the diagnosis of target faults under multiple faults, which can be interpreted as the diagnosis of important part faults under multiple fault conditions. It was also tested that AMC-RSN can have good diagnosis under mixed faults, but the limitation is that it has not yet distinguished the specific faults of mixed faults. We will continue to follow up with research on how to identify the specific faults of each component under multiple faults.

Author Contributions: Conceptualization, X.L.; methodology, W.C. and K.S.; software, W.C.; validation, W.C. and K.S.; formal analysis, X.L.; investigation, Y.X. and J.X.; data curation, H.M.; writing—original draft preparation, W.C.; writing—review and editing, W.C.; visualization, W.C.; supervision, X.L.; funding acquisition, H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the National Natural Science Foundation of China (51365006), National Natural Science Foundation of China (62073090), Natural Science Foundation of Guangxi Province (2018GXNSFAA180206) and Natural Science Foundation of Guangxi Province (2018GXNSFAA 281312).

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. Jin, X.; Cheng, F.; Peng, Y.; Qiao, W.; Qu, L. Drive-train gearbox fault diagnosis: Vibration-and current-based approaches. *IEEE Ind. Appl. Mag.* **2018**, *24*, 56–66. [CrossRef]
2. Li, X.; Zhang, W. Deep Learning-Based Partial Domain Adaptation Method on Intelligent Machinery Fault Diagnostics. *IEEE Trans. Ind. Electron.* **2020**, *68*, 4351–4361. [CrossRef]
3. Wang, X.; Shen, C.; Xia, M.; Wang, D.; Zhu, J.; Zhu, Z. Multiscale deep intra-class transfer learning for bearing fault diagnosis. *Reliab. Eng. Syst. Saf.* **2020**, *202*, 107050. [CrossRef]
4. Liu, Z.; Tang, X.; Wang, X.; Mugica, J.E.; Zhang, L. Wind Turbine Blade Bearing Fault Diagnosis Under Fluctuating Speed Operations via Bayesian Augmented Lagrangian Analysis. *IEEE Trans. Ind. Inform.* **2020**, *17*, 4613–4623. [CrossRef]
5. Sharma, V.; Parey, A. Frequency domain averaging based experimental evaluation of gear fault without tachometer for fluctuating speed conditions. *Mech. Syst. Signal Process.* **2017**, *85*, 278–295. [CrossRef]
6. Zhao, M.; Zhong, S.; Fu, X.; Tang, B.; Pecht, M. Deep Residual Shrinkage Networks for Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2019**, *16*, 4681–4690. [CrossRef]
7. Tran, V.T.; Yang, B.-S.; Gu, F.; Ball, A. Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis. *Mech. Syst. Signal Process.* **2013**, *38*, 601–614. [CrossRef]
8. Zhang, X.; Chen, W.; Wang, B.; Chen, X. Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization. *Neurocomputing* **2015**, *167*, 260–279. [CrossRef]
9. Lu, J.; Qian, W.; Li, S.; Cui, R. Enhanced K-Nearest Neighbor for Intelligent Fault Diagnosis of Rotating Machinery. *Appl. Sci.* **2021**, *11*, 919. [CrossRef]
10. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [CrossRef]
11. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
12. Su, K.; Liu, J.; Xiong, H. Hierarchical diagnosis of bearing faults using branch convolutional neural network con-sidering noise interference and variable working conditions. *Knowl. Based Syst.* **2021**, *230*, 107386. [CrossRef]
13. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [CrossRef]
14. Mo, Z.; Zhang, Z.; Tsui, K.-L. The Variational Kernel-Based 1-D Convolutional Neural Network for Machinery Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3105252. [CrossRef]
15. Yu, W.; Zhao, C. Broad Convolutional Neural Network Based Industrial Process Fault Diagnosis With Incremental Learning Capability. *IEEE Trans. Ind. Electron.* **2019**, *67*, 5081–5091. [CrossRef]
16. Li, G.; Wu, J.; Deng, C.; Chen, Z.; Shao, X. Convolutional Neural Network-Based Bayesian Gaussian Mixture for Intelligent Fault Diagnosis of Rotating Machinery. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3080402. [CrossRef]
17. Jiao, J.; Zhao, M.; Lin, J.; Ding, C. Deep Coupled Dense Convolutional Network With Complementary Data for Intelligent Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9858–9867. [CrossRef]
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the Computer Vision—ECCV, Amsterdam, The Netherlands, 11–14 October 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science. Springer: Cham, Switzerland, 2016; 9908, pp. 630–645.
20. Chen, Y.; Zhang, D.; Zhang, H.; Wang, Q.G. Dual-Path Mixed-Domain Residual Threshold Networks for Bearing Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2022**, *69*, 13462–13472. [CrossRef]
21. Li, C.; Yu, L.; Zhang, A.; He, Q.; Yang, W.; Duan, Z. A Novel Bearing Fault Diagnosis of Raw Signals Based on 1D Residual Convolution Neural Network. In Proceedings of the 2020 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), Shenzhen, China, 23 May 2020; pp. 1–6. [CrossRef]
22. Zhao, M.; Kang, M.; Tang, B.; Pecht, M. Multiple wavelet coefficients fusion in deep residual networks for fault diagnosis. *IEEE Trans. Ind. Electron.* **2019**, *66*, 4696–4706. [CrossRef]

23. Sun, K.; Huang, Z.; Mao, H.; Qin, A.; Li, X.; Tang, W.; Xiong, J. Multi-Scale Cluster-Graph Convolution Network With Multi-Channel Residual Network for Intelligent Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 3136264. [CrossRef]
24. 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]
25. 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. Ind. Electron.* **2021**, *70*, 3517010.
26. Liu, R.; Wang, F.; Yang, B.; Qin, S.J. Multiscale Kernel Based Residual Convolutional Neural Network for Motor Fault Diagnosis Under Nonstationary Conditions. *IEEE Trans. Ind. Inform.* **2019**, *16*, 3797–3806. [CrossRef]
27. Chao, Z.; Han, T. A novel convolutional neural network with multiscale cascade midpoint residual for fault diag-nosis of rolling bearings. *Neurocomputing* **2022**, *506*, 213–227. [CrossRef]
28. Li, Z.; Lv, Y.; Yuan, R.; Zhang, Q. Multi-fault diagnosis of rotating machinery via iterative multivariate variational mode decomposition. *Meas. Sci. Technol.* **2022**, *33*, 125104. [CrossRef]
29. Yuan, R.; Lv, Y.; Song, G. Multi-Fault Diagnosis of Rolling Bearings via Adaptive Projection Intrinsically Transformed Multivariate Empirical Mode Decomposition and High Order Singular Value Decomposition. *Sensors* **2018**, *18*, 1210. [CrossRef] [PubMed]
30. Lonare, S.; Fernandes, N.; Abhyankar, A. Rolling element bearing multi-fault diagnosis using morphological joint time–frequency adaptive kernel–based semi-smart framework. *J. Vib. Control* **2021**, *28*, 2940–2949. [CrossRef]
31. Ma, N.; Zhang, X.; Liu, M.; Sun, J. Activate or Not: Learning Customized Activation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 8028–8038. [CrossRef]
32. Fornasier, M.; Rauhut, H. Iterative thresholding algorithms. *Appl. Comput. Harmon. Anal.* **2008**, *25*, 187–208. [CrossRef]

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.