processes

Article

Fault Diagnosis of Wind Turbine Main Bearing in the Condition of Noise Based on Generative Adversarial Network

Zhixin Fu, Zihao Zhou * and Yue Yuan

College of Energy and Electrical Engineering, Hohai University, Nanjing 211100, China
* Correspondence: 201306080047@hhu.edu.cn

Abstract: In order to solve the problem that the fault classification accuracy of the main bearing of the wind turbine is not high due to the unbalanced vibration signal data of the main bearing of the wind turbine under the background of noise, this article proposes a double-layer fault diagnosis model for the main bearing of the wind turbine that combines the auxiliary classifier generation adversarial network (ACGAN) and the deep residual shrinkage network (DRSN). First, the wind turbine main bearing data is sent into the ACGAN to learn the distribution features of fault data, and a particular type of fault data is generated to expand the original dataset to achieve balance conditions, and then the expanded dataset is sent to the DRSN to reduce noise to improve the fault classification accuracy. The simulation results show that, compared with the traditional deep learning model, the model proposed in this article can significantly improve the classification accuracy of the main bearing fault of wind turbines under noise conditions, and also has a strong diagnosis ability in a state of datasets with different loads.

Keywords: wind turbine; main bearing; fault diagnosis; noise; deep residual shrinkage network; auxiliary classifier generative adversarial network

check for updates

Citation: Fu, Z.; Zhou, Z.; Yuan, Y. Fault Diagnosis of Wind Turbine Main Bearing in the Condition of Noise Based on Generative Adversarial Network. *Processes* **2022**, *10*, 2006. https://doi.org/10.3390/pr10102006

Academic Editor: Marco S. Reis

Received: 29 August 2022
Accepted: 1 October 2022
Published: 4 October 2022

1. Introduction

As an important green renewable energy, wind energy has become a research hotspot in recent years, and the wind power industry has also developed rapidly. By the end of 2021, the cumulative installed capacity of wind turbines in the world has reached 328 million kilowatts, of which the installed capacity of offshore wind power has reached 26.39 million kilowatts [1]. If calculated according to the 5-year warranty period of wind turbines, about 3 GW capacity of offshore wind turbines are about to or have already gone out of the warranty period [2]. Therefore, the market potential for wind turbine operation and maintenance is huge. The operating environment of wind turbines is harsh, and the wind turbines are exposed to sand and snow for a long time [3]. In addition, there is also severe weather such as thunderstorms and fog at sea. According to statistics, the operation and maintenance cost of onshore wind farms is as high as 15% to 20% of the total wind farm revenue, while the operation and maintenance cost of offshore wind farms is much higher than that of onshore wind farms, accounting for about 20% to 25% of the total wind farm revenue [4]. The high fail rate of wind turbines brings great difficulties to the operation and maintenance of wind farms, and the failure to discover potential faults in time and repeated maintenance of components with a high fail rate will increase the operation and maintenance costs of wind farms.

During the operation stage of the wind farm, the faults of the generator, the gearbox, the transmission system, and the blades are the most common [5], of which the main bearing of the wind turbine plays a role in transmitting energy to the wind turbine [6]. As a rotating component, the main bearing is more prone to failure, and the entire unit will stop running after the failure, causing huge economic losses. The vibration signal contains all the useful information about the components and it is also one of the important indicators for analyzing the operating state. The fault diagnosis technology for analyzing the vibration signal of the main bearing is currently the most effective and widely used [7]. In recent years, some domestic and foreign scholars have introduced artificial intelligence methods into the field of fault diagnosis of rotating components of wind turbines, such as deep learning and so on. Compared with traditional fault diagnosis methods based on statistical analysis methods, the fault diagnosis method based on the neural network does not rely on a large number of signal processing related knowledge and rich expert experience, but the essential characteristics of faults are extracted from massive historical data, avoiding the randomness of manual selection of parameters, and the diagnosis process is more intelligent [8]. Cao et al. [9] use Long-Short Term Memory (LSTM) neural network to extract the fault characteristics of vibration signal of wind turbines and perform fault classification and compare this method with the support vector machine method to verify the superiority of the algorithm, the method proposed in this article achieves 97.2% of the classification accuracy of the gearbox. Wu et al. [10] adopt a convolutional neural network to study the one-dimensional vibration signal of the planetary gearbox. The conclusion shows that the accuracy of the one-dimensional convolutional neural network model for fault diagnosis of planetary gearbox is higher than that of traditional diagnosis methods. Yao et al. [11] propose a fault diagnosis method for rolling bearings based on a convolutional neural network and recurrent neural network.

The one-dimensional vibration signal is converted into a two-dimensional image signal by the Gram angle field method, and the image signal is input into the model for training, which has a higher fault classification accuracy, experiments show that the method proposed has an accuracy of more than 98.15% for the classification of rolling bearing faults. However, these references ignore that the fault data of offshore wind turbines is often difficult to be obtained, and there is a general problem of insufficient fault samples. In particular, the main bearing fault data of wind turbines accounts for a relatively low proportion of all fault data, and there is a serious unbalanced dataset problem. Therefore, it is difficult for deep learning methods to achieve high fault classification accuracy in this case. Zhou et al. [12] believe that when most classification algorithms classify unbalanced data, the obtained classification hyperplane will be biased toward a few types of data, which leads to the algorithm misjudging the minority type of data as the majority type of data. In order to obtain sufficient and balanced vibration signal samples, some scholars refer to generative adversarial networks in the field of rolling bearing fault diagnosis. Lu et al. [13] propose a data enhancement method for the vibration signal of the main bearing of wind turbines based on an auxiliary classification generation adversarial network, which can effectively extract the original data distribution characteristics and generate high-quality vibration signal samples, after using ACGAN to expand the original dataset, the fault classification accuracy of various models is improved by about 2%. Li et al. [14] improve the auxiliary classification generative adversarial network based on Bayesian optimization and Wasserstein distance, realize data enhancement, and obtained a higher fault classification accuracy of wind turbine planetary gearboxes. The classification accuracy of WAC-GAN could remain above 94% for various types of failures. In addition, due to the harsh operating environment of offshore wind turbines, the signal samples collected by sensors often contain noises. These noises will affect the feature extraction performance of neural networks during training. Traditional signal denoising methods often require a lot of statistical knowledge. Different noise interference is targeted for different noise reduction processing. Zhao et al. [15] propose a deep learning-based feature learning algorithm for noisy data, which integrates the attention mechanism and the idea of a soft threshold, effectively reducing the impact of noise interference on the model, experiments show that in the case of inserting various types of noise, the accuracy of DRSN with channel-wise thresholds (DRSN-CW) is about 3.32% higher than that of ResNet. In [16], a deep residual shrinkage network is added to the convolutional neural network to achieve signal noise reduction and solve the degradation problem of the multi-layer model. This method has a higher fault classification accuracy reaching 99.5% than the traditional neural network method. However, the above studies do not take into account the unbalanced data of the main bearing of the actual wind turbine, which limited the fault diagnosis capability of the model.

In order to improve the performance of the fault diagnosis model in practical applications, this article focuses on the research on the fault diagnosis method of wind turbine main bearing under noise conditions and proposes a fault diagnosis method of the main bearing of the wind turbine based on the auxiliary classification generate adversarial network and the deep residual shrinkage network. First, the auxiliary classification generative adversarial network is adopted to learn the data distribution of vibration signal samples with different signal-noise ratios, and the datasets of each fault are expanded. Then, the expanded dataset is sent to the deep residual shrinkage network for training. Finally, the test set is fed into the trained deep residual shrinkage network to test the fault classification accuracy of the model. Experimental results show that the proposed method has good fault diagnosis performance in the face of the vibration signal sample of the main bearing of the wind turbine when the actual operation contains noise interference, and the data is unbalanced.

The first chapter of this article is an introduction, the second chapter gives the structure of the model proposed in this article, the third chapter describes the basic principle of the auxiliary classification generative adversarial network, and the fourth chapter describes the basic principle of the deep residual shrinkage network, the fifth chapter uses two experiments to verify the effectiveness of the method proposed in this article, and the sixth chapter gives some conclusions and suggestions.

2. Fault Diagnosis Model of Wind Turbine Main Bearing

In this article, a fault diagnosis model of the main bearing of a wind turbine with a double-layer network structure is used, and the model is shown in Figure 1. The upper layer is a generative network based on the auxiliary classification generative adversarial network. The generator learns the data distribution characteristics of the original vibration signal dataset during training. Then, particular types of fault data are generated to expand the dataset to a balanced state. That is, the ratio of the sample of each type of fault data to health status data is 1:1; then the expanded dataset is fed into the classification network based on the deep residual shrinkage network, and the attention mechanism and soft threshold are used in the classification network to reduce the redundant noise in the signal adaptively, and the classifier can accurately identify the fault samples of the main bearing of the wind turbine through the training of the expanded dataset.

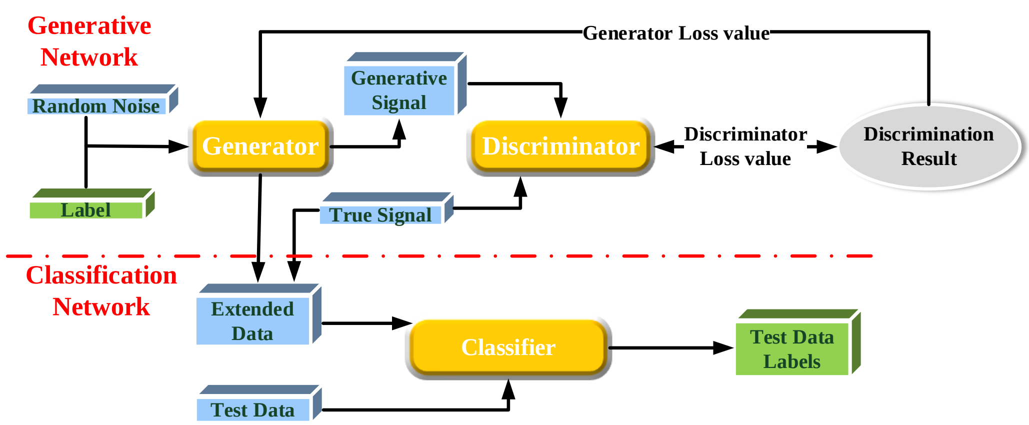


Figure 1. Fault Diagnosis Model of the Main Bearing of Wind Turbine.

3. Generative Adversarial Network

In 2014, Lan Goodfellow et al. proposed Generative Adversarial Networks (GAN) [17]. Since GAN can generate data with a specific distribution, it is an unsupervised deep learning model, so it is widely used in image inpainting, text generation, audio generation, and other fields, and is the research hotspot in the image field in recent years. Because the fault classification algorithm based on deep learning needs a large amount of fault data, and the actual wind turbine main bearing fault data samples are scarce, there is a serious sample imbalance problem between fault data and normal data. In order to improve the accuracy of fault classification, some scholars have introduced the generative adversarial network into the field of fault diagnosis. The generative adversarial network can learn the data distribution of real fault samples to generate new fault samples to supplement the original dataset. Studies have shown that [18], this method has great potential in the application of time series data generation.

3.1. Generative Adversarial Network Principle

GAN contains two networks, Generator and Discriminator. The application of the generator is to convert the random noise into data that is close to the one-dimensional vibration signal data distribution of the main bearing of the wind turbine as much as possible. The function of the discriminator is to judge the authenticity of the input samples. During the training, the generator and the discriminator game alternately. The game mechanism continuously improves the generation ability of the generator, so that the data generated by the generator is as real as possible and deceives the discriminator to achieve the purpose of being a genuine one.

The training of GAN mainly includes two stages: discriminator training and generator training. In each round of iteration, the generator and the discriminator compete with each other and finally reach the Nash equilibrium, that is, the discriminator classification accuracy rate reaches 50%. The real fault data of the main bearing of a wind turbine or the fake data generated by the generator has a 50% chance of being misjudged, and the generator completes the training. The structure of GAN is shown in Figure 2.

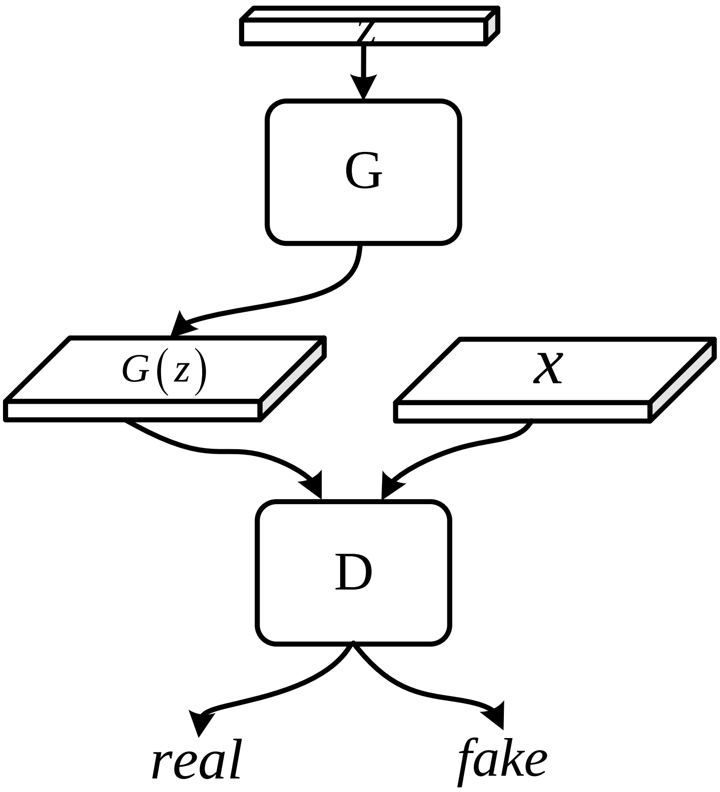


Figure 2. Generative Adversarial Network Structure Diagram.

In the initial stage of training, the capabilities of the generator and the discriminator are very weak. First, the random noise vector is sent to the generator to generate fake samples, fake samples, and the real signal samples are passed through the discriminator to generate scores. The loss function consists of samples and the labels corresponding to these samples, and the gradient is calculated to update the discriminator; then, the parameters of the discriminator are fixed, and fake samples will be sent to the discriminator to get the score since it is hoped that the samples generated by the generator are as real as possible and deceive the discriminator, The optimization goal at this time is to make the score reach 1 to update the generator by calculating the gradient. In the iteration, the generator and the discriminator have trained alternately, and finally, the two networks reach a Nash equilibrium state. At this time, the discriminator cannot distinguish between real samples and fake samples, and the accuracy of the discriminator is 50%. The network objective function is (1):

$$\min_{G}\max_{D} V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))] \tag{1}$$

$$D(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \tag{2}$$

where E is the mathematical expectation, $p_{data}(x)$ and $p_g(x)$ is the probability that the sample is true or false, respectively, z is the random noise vector, and $G(z)$ is the fake sample generated by the generator. Our mission is to train the discriminator to maximize $\log D(x)$ and $\log(1 - D(G(z)))$, and train the generator to minimize $\log(1 - D(G(z)))$.

The generative model is essentially a maximum likelihood estimation. It is assumed that the initial distribution of the generator is $P_g(x|\theta)$, where θ is the parameter of the distribution. In order to make the generated data distribution close to the real data distribution, it is necessary to calculate the value $\hat{\theta}$ to maximize (3). Therefore, the calculation formula is (4):

$$L_g = \prod_{i}^{n} p_g\left(x^i \middle| \theta\right) \tag{3}$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \prod\nolimits_{i=1}^{n} p_g\left(x^i \middle| \theta\right) = \underset{\theta}{\operatorname{argmin}} KL\left(P_{data}(x) \parallel P_g(x|\theta)\right) \tag{4}$$

where $p_g\left(x^i \middle| \theta\right)$ is the likelihood function of the real data, $P_{data}(x)$ is the real data distribution, and $P_g(x|\theta)$ is the generated data distribution. GAN adopts KL divergence to measure the distance between two distributions. If the KL divergence reaches the minimum value 0, then the distribution $P_{data}(x)$ and $P_g(x|\theta)$ are equal everywhere.

3.2. Auxiliary Classifier Generative Adversarial Networks

Unlike traditional GAN, Auxiliary Classifier Generative Adversarial Networks (ACGAN) add labels to the random noises which are input to the generator and generate fake fault samples with a specific type of label. Then the true and false fault data samples are input into the discriminator to get the output results, and the output results include both true or false labels and classification labels. The network can be used to generate different types of wind turbine main-bearing fault data in a targeted manner, and the original fault dataset can be expanded into a balanced dataset.

It can be seen from Figure 3 that ACGAN not only outputs the probability that the fault sample is real data or not but also outputs the fault class probability of the sample. Since ACGAN has category labels when generating and judging samples, it makes the generated fault samples more controllable. The true or false judgment and classification loss functions are (5) and (6), respectively:

$$L_s = E_{x\sim P_{data}}[\log_2 D(x)] + E_{z\sim P_z}[\log_2(1 - D(G(z)))] \tag{5}$$

$$L_c = E_{c\sim P_{data}}[\log_2 D(c)] + E_{c\sim P_z}[\log_2(1 - D(G(c)))] \tag{6}$$

where, L_s is the probability that the sample is real, and L_c is the probability that the sample is correctly classified. Therefore, in training, the discriminator is trained to maximize $L_s + L_c$, and the generator is trained to maximize $L_c - L_s$.

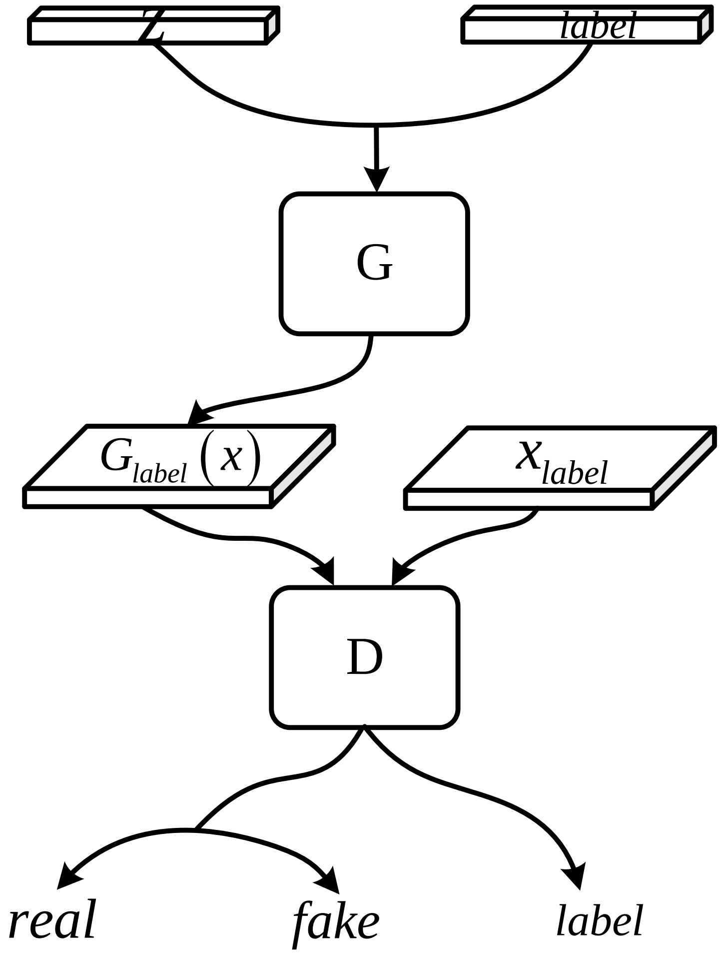


Figure 3. ACGAN Structure Diagram.

4. Deep Residual Shrinking Networks

The collected vibration signals of the main bearing of the wind turbine are often accompanied by noise actually, and the fault signal features are difficult to be extracted. The traditional signal noise reduction method is to transform the noisy signal (wavelet transform, empirical mode decomposition, etc.), and then use a soft threshold to reduce the noise, and finally, the signal is inverse transformed to obtain the signal after noise reduction. However, the noise signal of wind turbines may be different under different working conditions, and the selection of threshold is more complicated and requires a lot of relevant knowledge of signal processing. Therefore, this article selects the Deep Residual Shrinkage Network [15] as the classifier to diagnose the fault of the main bearing of the wind turbine. With the deepening of network layers, the ordinary convolutional neural network model training is difficult, the accuracy may be reduced. The residual network introduces the idea of an identity shortcut, the output of the previous layer of the network is directly transmitted to the next layer to achieve a smaller training error in the case of a larger number of network layers [19]. The DRSN is improved on the basis of the residual network to realize the function of noise reduction.

4.1. Attention

Attention Mechanism is widely used in the fields of natural language processing and pattern recognition. Its essence is similar to the human visual attention mechanism, that is, it selects the key information of the target task from many targets and suppresses useless information. Since the one-dimensional vibration data of the main bearing of the wind turbine is highly time-varying and the composition is complex, the introduction of an attention mechanism can adaptively weight different feature channels to highlight useful information. In this article, the channel threshold attention mechanism is used to apply weights to the signals on each channel to improve the fault classification ability of the network under various working conditions. For example, Squeeze-and-Excitation Networks (SENet) is a network that sets an attention mechanism for channels.

Figure 4 is a schematic diagram of the Squeeze-and-Excitation Networks module. The number of input channels is c_1, the number of output channels of the second layer is c_2, and then be compressed into a feature map of size $c_2 \times 1 \times 1$ by global average pooling. Finally, the $Softmax$ activation function is used to obtain the weight of each channel and the second layer is weighted to obtain the output result. (7) is the $Softmax$ function expression.

$$f(x_k) = \frac{\exp(x_k)}{\sum_{k=1}^{c} \exp(x_k)} \tag{7}$$

where $f(x_k)$ represents the weight prediction value of the kth channel by the activation function $Softmax$, and c is the number of channels.

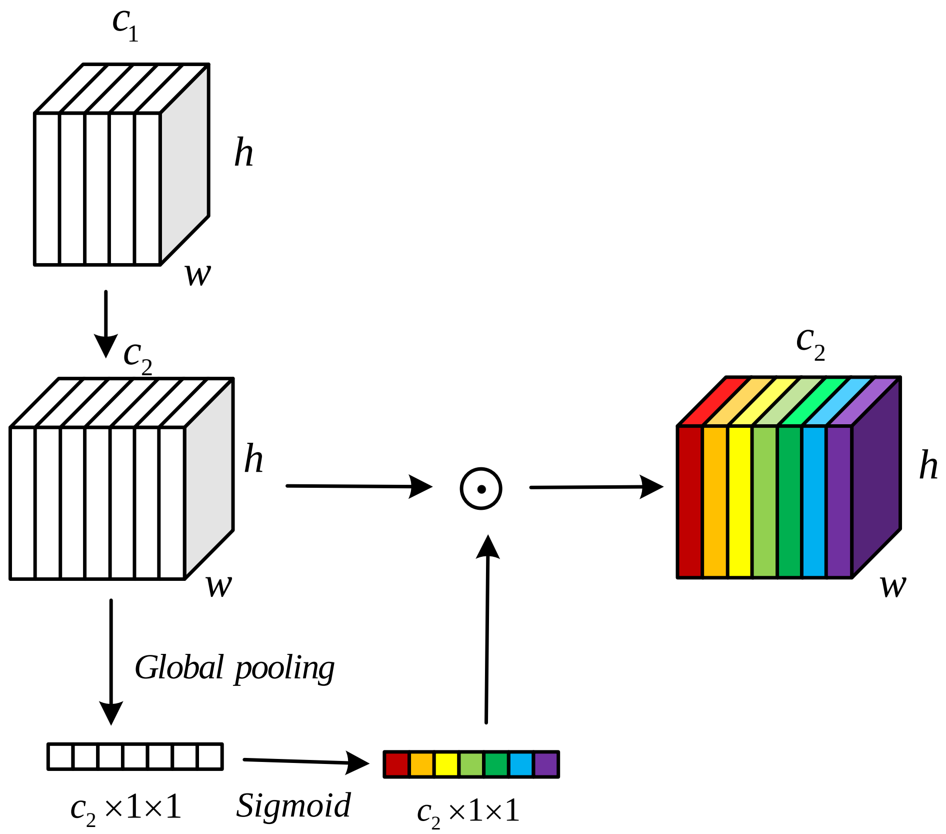


Figure 4. SENet Structure Diagram.

4.2. Soft Threshold Noise Reduction

The noise in the actual vibration signal greatly reduces the ability of model feature extraction. On the basis of the residual network, the DRSN adopts the method of soft threshold to denoise the signal. The soft threshold is to set a threshold, set the signal below the threshold to 0, and adjust the signal above the threshold to 0, that is, "shrink". The core of the DRSN is to notice the vibration of the main bearing of the wind turbine through the attention mechanism. The unimportant features in the signal are set to zero by the soft threshold, which enhances the neural network's ability to extract fault features from noisy signals. The selection of the threshold size has a direct impact on the noise reduction effect. (8) is the expression of the soft threshold.

$$S(x,\tau) = \begin{cases} x-\tau & x>\tau \\ 0 & -\tau<x<\tau \\ x+\tau & x<-\tau \end{cases} \tag{8}$$

where τ represents the size of the threshold, it can be seen from (8) that when the signal is within the threshold, the derivative is 0, otherwise, the derivative is 1.

4.3. Residual Shrinkage Module

In this article, a DRSN with a channel-wise thresholds module is used to build a classifier network. Different from ordinary DRSN, this module has independent thresholds in each channel [15]. The overall structure of the residual shrinkage module is shown in Figure 5.

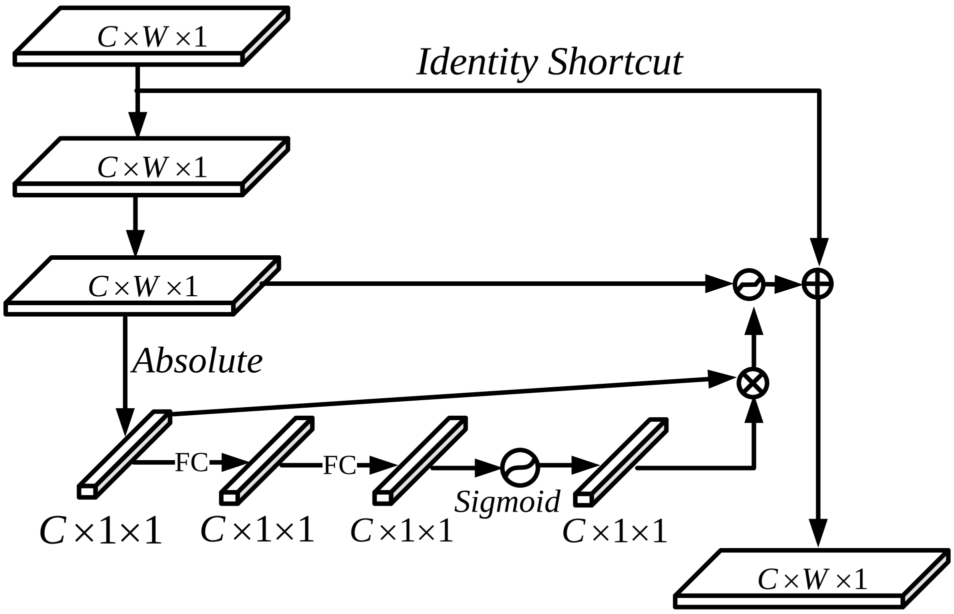


Figure 5. DRSN Module.

This module adds a denoising function to the traditional deep residual module. After the input data is passed through two layers of convolution, a one-dimensional vector is obtained after the global mean pooling layer. The attention weight of each channel is obtained by this vector through the two fully connected layers and the *Sigmoid* activation function, and the threshold is obtained by multiplying it with the corresponding average value of each channel. (9) is the expression of the threshold of the deep residual shrinkage module.

$$\tau_c = \omega_c \cdot average|x_c| \tag{9}$$

where τ_c is the threshold of channel c, ω_c is the weight of channel c, and $average|x_c|$ is the average value of the absolute value of each element of channel c.

5. Example Analysis

5.1. Model Framework

In order to test the effectiveness of the method mentioned above, the rolling bearing dataset [20] of Case Western Reserve University (CWRU) was selected as the simulation analysis object. The dataset comes from the sampled vibration data of the driving end of the wind turbine. The sampling frequency is 12 kHz, and each sample in the dataset contains 2048 sampling points. The final data is shown in Table 1. The simulation includes one normal state and nine different types of fault states, each state sample contains four different operating motor speeds. The dataset is randomly shuffled and divided into training and test sets.

Table 1. CWRU experimental data classification.

Fault Label	Bearing Status (Wear)	Fault Location	Fault Diameter/mm	Motor Load	Number of Training Samples	Number of Test Samples
0	Normal	/	0	0~3	700	120
1	Slight	Inner Race	0.18	0~3	160	70
2	Moderate		0.36	0~3	160	70
3	Heavy		0.54	0~3	160	70
4	Slight	Ball	0.18	0~3	160	70
5	Moderate		0.36	0~3	160	70
6	Heavy		0.54	0~3	160	70
7	Slight	Outer Race	0.18	0~3	160	70
8	Moderate		0.36	0~3	160	70
9	Heavy		0.54	0~3	160	70

Four different fault diagnosis models were implemented using TensorFlow 2.8.0, which is a machine learning toolkit released by Google. Experiments are conducted on the computer in which the CPU is AMD Ryzen 6 4800H, the GPU is Nvidia GeForce RTX 2060, and the memory is 32 GB.

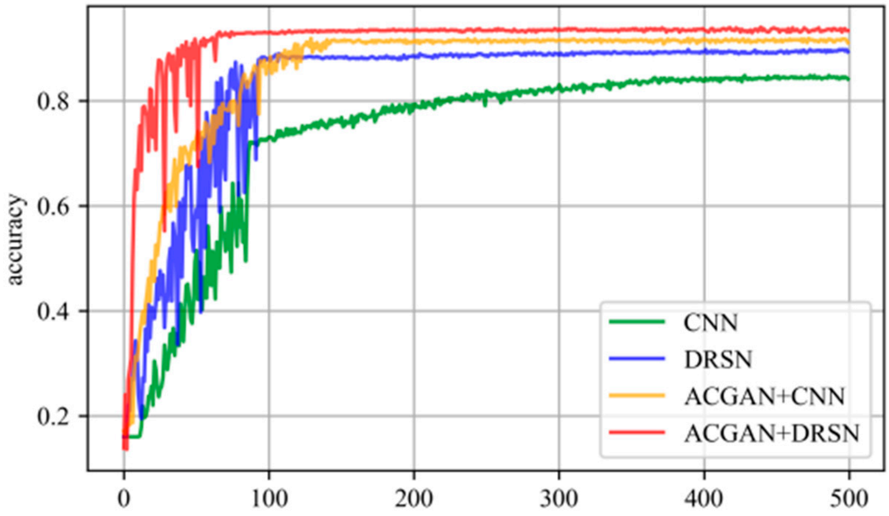
In Table 1, load 0 represents the motor speed of 1797 rpm, load 1 represents the motor speed of 1772 rpm, load 2 represents the motor speed of 1750 rpm, and load 3 represents the motor speed of 1730 rpm.

The input of the generator in the generation network is a 100-dimensional normally distributed random vector and a label value. The label and the random vector are sent to the generator at the same time and become a fake sample of the same size as the input signal through a series of one-dimensional convolution operations. Each one-dimensional convolutional layer in the generator uses $LeakyReLU$ as the activation function, adding a Dropout layer and a batch normalization layer to prevent overfitting and make the network easier to be trained. The structure of the discriminator is basically identical to the generator. The input data is a vibration signal sample, and the data after multi-layer one-dimensional convolution is passed through the fully connected layer and the $Softmax$ activation function to obtain the true or false probability and the sample category probability respectively. The classification network structure is obtained by improvement on the basis of literature [15]. The improvement ideas are: (1) Appropriately reduce the depth of the network, which can significantly improve the training speed of the model and prevent the model from overfitting to a certain extent; (2) Increase the network width. It is found that increasing the number of channels of one-dimensional convolution can effectively improve the accuracy of the model. The RMSprop optimization algorithm is used in the model training, the hyperparameter is set to 0.9, and the adjustable learning rate with a lower limit of 0.00001 is used to speed up the convergence of the model.

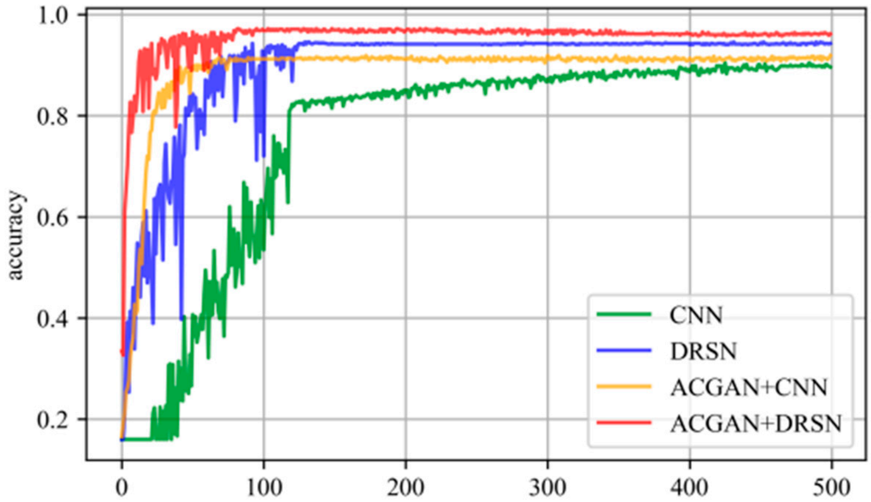
5.2. Fault Diagnosis Ability under Noise Conditions

In order to test the denoising ability of the model in the condition of the unbalanced dataset, this experiment adds Gaussian white noise with different signal-noise ratios of −5 db, −2 db, 0 db, 2 db, and 5 db on the basis of the original dataset. CNN trained on the original dataset, DRSN trained on the original dataset, CNN trained on the dataset expanded by ACGAN, and DRSN trained on the dataset expanded by ACGAN are used to compare the classification accuracy for the test set. The CNN, DRSN, and ACGAN structures of different experimental groups in the simulation are the same, respectively.

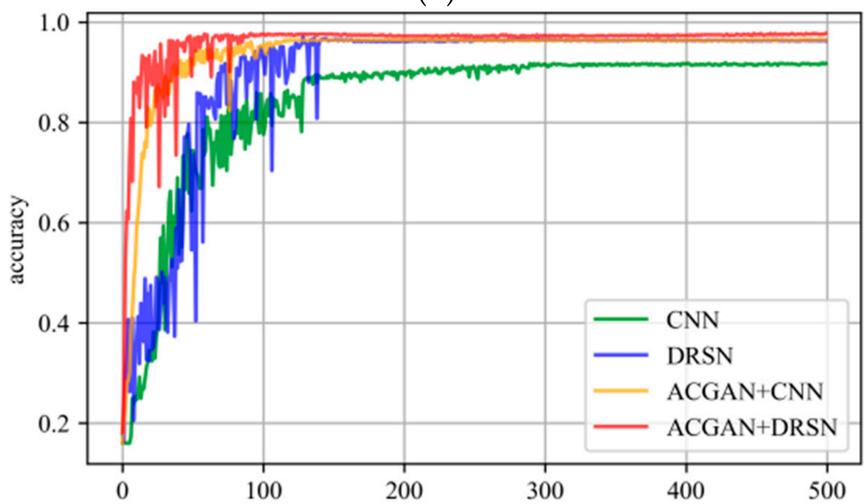
The simulation results are shown in Figure 6a–e. The ordinate in the figure represents the model accuracy, and the abscissa represents the number of iterations.



(**a**)

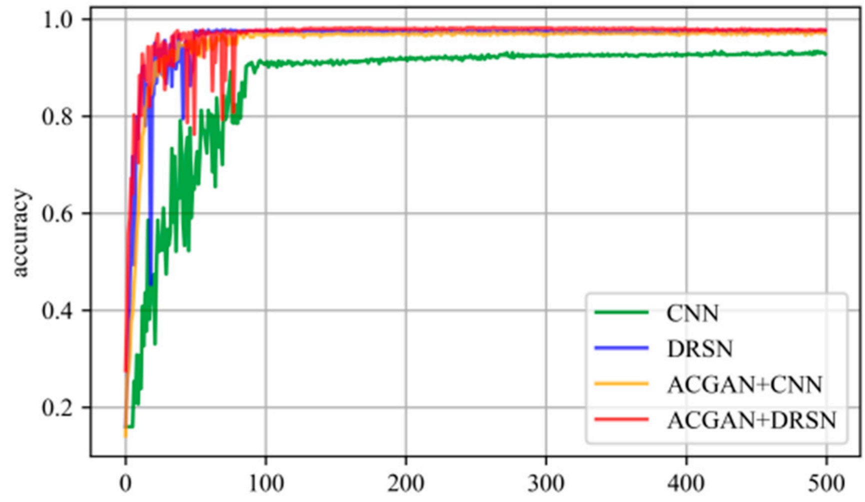


(**b**)

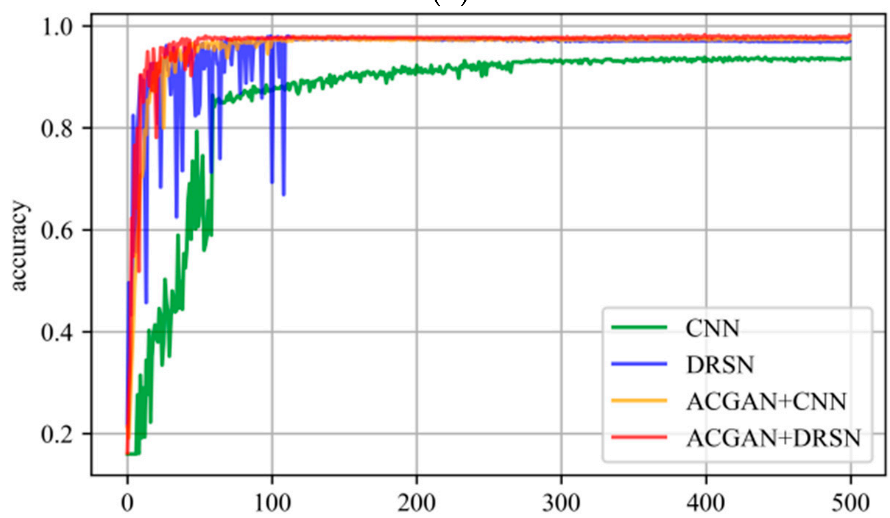


(**c**)

Figure 6. *Cont*.



(**d**)



(**e**)

Figure 6. (**a**) Model fault classification accuracy in the condition of −5 db noise; (**b**) Model fault classification accuracy in the condition of −2 db noise; (**c**) Model fault classification accuracy in the condition of 0 db noise; (**d**) Model fault classification accuracy in the condition of 2 db noise; (**e**) Model fault classification accuracy in the condition of 5 db noise.

As can be seen from Figure 6a–e, in the initial stage of training, the training speed of the dataset improved by ACGAN is significantly faster than that of the ordinary dataset. When the noise in the original signal reaches −5 db, the fault classification accuracy of the model proposed in this article is improved by about 9.20% compared with the accuracy of the method only using CNN, and when the noise in the original signal is 5 db, the fault classification accuracy of the model proposed in this article is improved by about 4.53% compared with the accuracy of the method only use CNN. Therefore, the higher SNR of the noise contained in the original signal, the more obvious the improvement of the fault classification accuracy of the model proposed in this article compared with the model based on ordinary CNN only.

It can be seen from Figure 7 that when the vibration signal contains more noise, the classification accuracy of the three networks all has different degrees of degeneration in general. First, using DRSN to train a noisy fault dataset has a significant improvement in classification accuracy compared to the traditional CNN-based method. The classification accuracy of using traditional CNN to train the dataset expanded by ACGAN is significantly better than that of directly training the original dataset by CNN, and it implies the effectiveness of training the dataset expanded by ACGAN; In addition, using DRSN and CNN to train the dataset expanded by ACGAN are comparable in fault classification accuracy with different signal-noise ratios. Reference [21] added noise with a signal-to-noise ratio of −5~5 db to the vibration signal to simulate the complex working environment of rolling bearings in industrial production. It had concluded that when the signal-noise ratio is higher than 0 db, the model with DRSN and the ordinary model had a good performance of noise reduction, and when the signal-to-noise ratio reaches −5 db, the classification accuracy of CNN is only 79%, which is much lower than 86% of DRSN. Reference [22] proved by experiments that the expansion of the original unbalanced dataset by ACGAN can reduce the influence of unbalanced data on the classification accuracy and the misjudgment rate of fault diagnosis. Therefore, previous studies are consistent with the experimental results in this article. In the case of training the dataset expanded by AC-GAN, the classification accuracy is obviously better compared with using CNN when selecting the DRSN as a classifier. For vibration signals with different signal-noise ratio noises, the classification accuracy of the model proposed in this article changes relatively gently. The accuracy can be maintained above 90%. Generally, the classification accuracies of the other three methods are inferior to the method proposed in this article. Table 2 shows the fault classification accuracy of each model in conditions of different noises. The fault classification accuracy of the model proposed in this article is the highest under different noises, and the fault classification accuracy changes smoothly in the condition of −5~5 db noises. In conclusion, the ACGAN + DRSN model has good classification accuracy and stability.

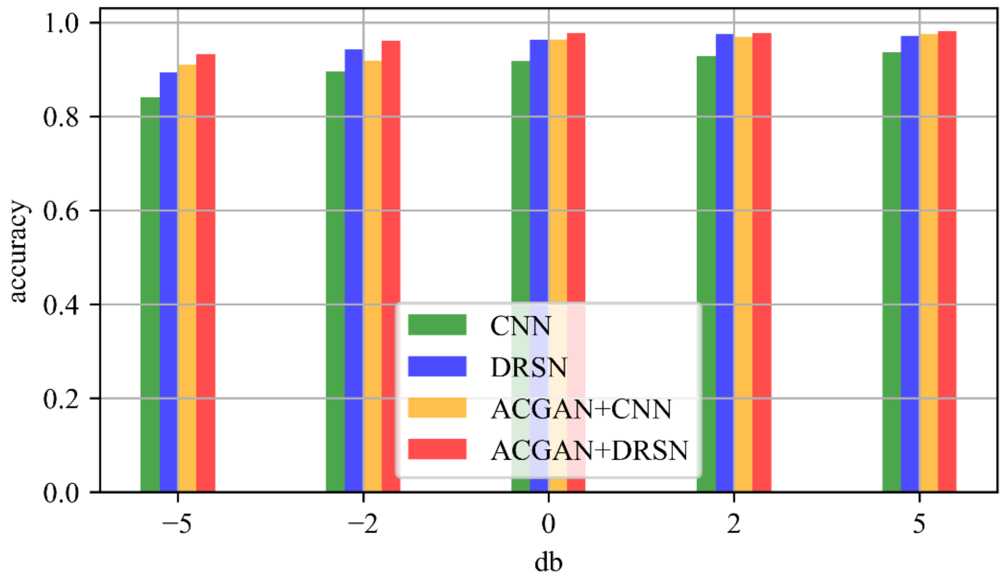


Figure 7. Fault classification ability under the circumstance of different noise.

Table 2. Fault classification accuracy in the condition of different noises.

SNR	Mean Failure Recognition Rate		
	CNN	**ACGAN + CNN**	**ACGAN + DRSN**
−5 db	84.133%	91.067%	93.333%
−2 db	89.600%	91.867%	96.133%
0 db	91.733%	96.400%	97.733%
2 db	92.800%	96.933%	97.733%
5 db	93.600%	97.467%	98.133%

5.3. Fault Diagnosis Capability under Variable Working Conditions

In this example, a high amount of noise by −10∼−5 db is randomly added to the original vibration signal to simulate the actual working conditions of the main bearing of the offshore wind turbine [6]. The vibration data with the motor speed of 1797 rpm, 1772 rpm, and 1750 rpm are used as the training set, and the vibration data with the motor speed of 1730 rpm are used as the test set to compare the generalization ability of the three fault classification models in the noise background. Figure 8a–d is the accuracy confusion matrices of the four models which use traditional CNN to train the original dataset, using DRSN to train the original dataset, using traditional CNN to train ACGAN-expanded data, and using DRSN to train ACGAN-expanded data.

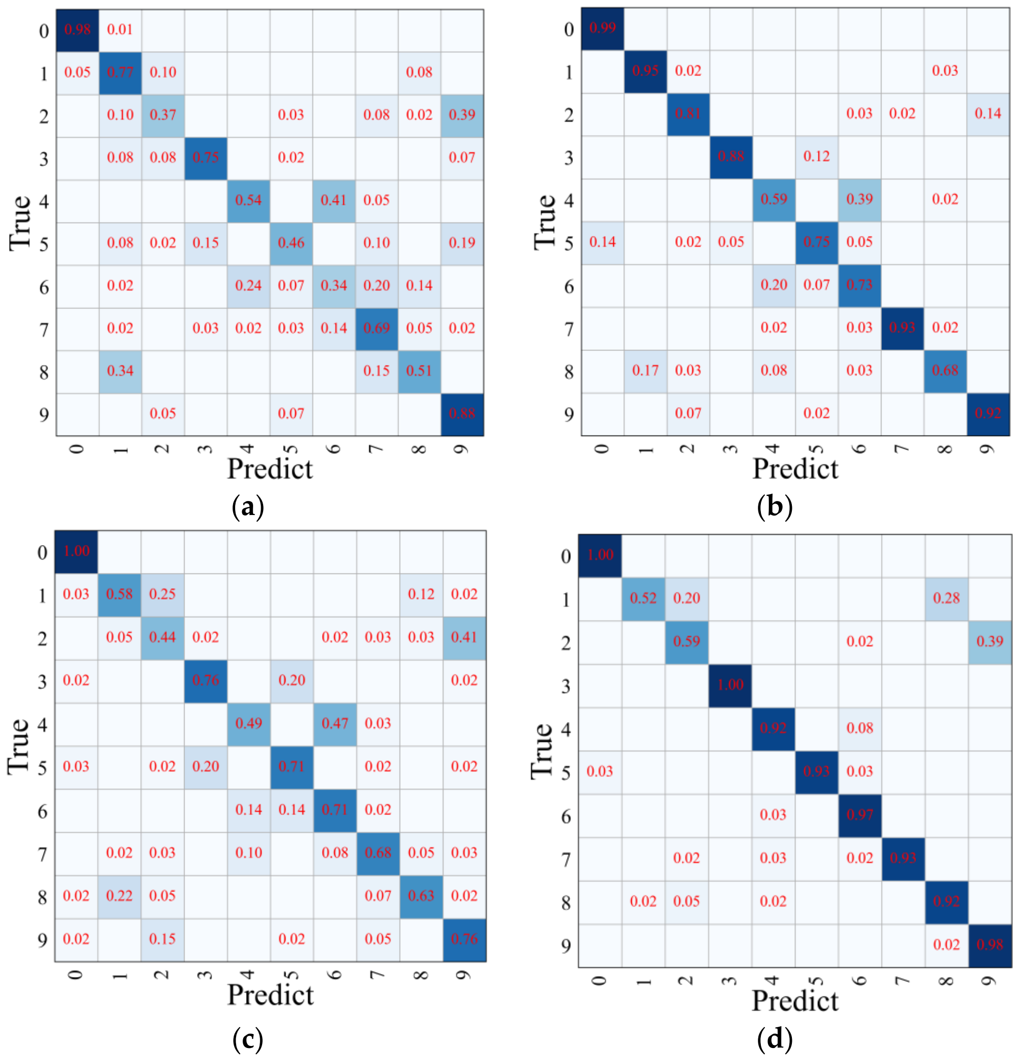


Figure 8. (**a**) CNN model fault classification ability; (**b**) DRSN model fault classification ability; (**c**) ACGAN + CNN model fault classification ability; (**d**) ACGAN + DRSN model fault classification ability. The darker the blue squares in these figures, the higher the probability that the fault will be correctly classified.

It can be seen that when training the unbalanced dataset of the vibration signal of the main bearing of the wind turbine under the noise conditions, the classification accuracy of the health state samples of the main bearing of the wind turbine is the highest, and the classification accuracy of the remaining samples are low. When using CNN alone to diagnose a test set, the accuracy of each fault sample is difficult to maintain above 50%, and the accuracy of the test set is about 71%; Compared with the method only using traditional CNN, the classification accuracies of the other two methods have been significantly improved; While the accuracy of each fault classification of the model proposed in this article is higher than 52%, and the accuracy of the test set is higher than 89%. Therefore, the ACGAN + DRSN model still has stronger fault classification ability when tested under different working conditions, and the classification accuracy is significantly higher than that of the ordinary CNN model. In order to visually demonstrate the fault classification ability of ACGAN+DRSN, the dimensionality reduction visualization of the model feature extraction effect is carried out.

Figure 9a–d is the effect diagrams of dimensionality reduction visualization using t-SNE for the original dataset, the output dataset of the last layer of the traditional CNN model, the traditional DRSN model, the ACGAN + CNN model, and the ACGAN + DRSN model, respectively. In Figure 9a, all the fault states of the main bearings of wind turbines are crossed and difficult to be classified; Figure 9b shows that after using the traditional CNN model, the health data of the main bearing of the wind turbine has been effectively classified, and the fault data initially shows the boundary, but it is still difficult to be classified; Figure 9c,d show that various types of fault samples have obvious boundaries, but each cluster obviously contains more than two different fault states, and it's hard to see which approach is better; In Figure 9e, with the model proposed in this article, various states clustering is enhanced. The red area is the healthy state of the main bearing of the wind turbine, this area has the best clustering effect. The slight wear and heavy wear of the rolling elements are not very separable in this example, but the rest of the fault states can be well classified effectively.

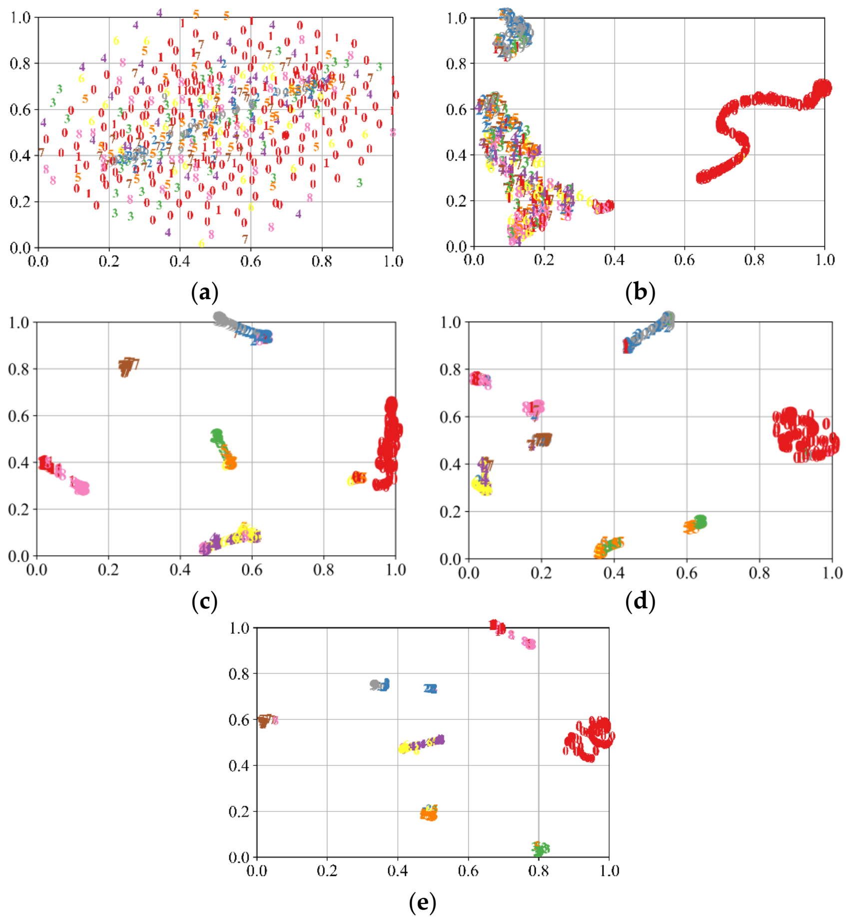


Figure 9. (**a**) Visualization of Raw dataset; (**b**) Visualization of CNN model dataset; (**c**) Visualization of DRSN model dataset; (**d**) Visualization of the dataset of ACGAN + CNN; (**e**) Visualization of the dataset of ACGAN + DRSN.

Table 3 shows the fault classification accuracy under different motor speeds. The motor speed of the test set of Experiment 1, Experiment 2, Experiment 3, and Experiment 4 are 1730 rpm, 1750 rpm, 1772 rpm, and 1798 rpm, respectively. The other three motor speeds are used as the training set. The accuracy of the model proposed in this article is higher than the other models obviously. The accuracy is always maintained above 85%.

Table 3. Fault classification accuracy under variable working conditions.

Experiment	CNN	DRSN	ACGAN + CNN	ACGAN + DRSN
1	71.13%	86.22%	75.16%	89.34%
2	74.71%	87.35%	75.23%	91.13%
3	72.49%	86.57%	79.53%	88.66%
4	78.62%	83.85%	81.85%	87.38%

6. Discussion

In order to solve the problem that the vibration fault data of wind turbine main bearing is difficult to be obtained in the condition of noise, resulting in the low fault diagnosis accuracy, this article proposes a fault classification method for wind turbine main bearing based on ACGAN and DRSN. This method has the following advantages:

(1) In view of the problem that the model training accuracy is not high due to insufficient fault data of the main bearing of the wind turbine, this article uses ACGAN to learn the distribution characteristics of fault data from the limited vibration signal samples, and generate high-quality fault samples to achieve data enhancement and improve the classification accuracy of the model;

(2) The use of attention mechanism and soft threshold of deep residual shrinkage network as a classification network can effectively reduce the different degrees of noise interference contained in the dataset, and fully explore the data fault characteristics of the main bearing of the wind turbines;

(3) Compared with the traditional CNN model, the ACGAN + DRSN model still has a stronger fault classification ability in the face of an unbalanced dataset containing noise under the variable working motor speed of the main bearing of the wind turbine.

The limitations of the methodological approach adopted are as follows:

(1) The research in this article is based on ACGAN. Many scholars have found that the original GAN has the problem of unstable training and poor ability to generate data.

(2) This model needs to train two models successively, which takes a longer time, so the training efficiency needs to be improved.

The simulation data of this experiment is still based on laboratory data and artificially added random noise. Considering the slight error of the wind turbine itself and changes in the operating environment, it is recommended that the wind turbine operation and maintenance manufacturers fully collect the different types of actual vibration data of the main bearing of the wind turbine when using the model proposed by this article. In addition, since the fault classification accuracy of the model decreases when the wind turbine runs at a new motor speed, the different working conditions of the wind turbine should be fully considered based on the model to improve its robustness of the model.

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

Funding: This research was funded by [Jiangsu Coastal Development and Collaborative Innovation Center Project] grant number [B2106118].

Institutional Review Board Statement: The study did not require ethical approval.

Informed Consent Statement: Not applicable.

Data Availability Statement: The study did not report any data.

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

References

1. Polaris Wind Power Network. In-Depth Report on China's Wind Power Industry in 2021. Available online: https://news.bjx.com.cn/html/20220512/1224507.shtml (accessed on 31 May 2022).
2. Polaris Wind Power Network. A Weekly Review of Wind Power. Available online: https://news.bjx.com.cn/html/20220114/1199608.shtml (accessed on 25 January 2022).
3. Long, X.; Yang, P.; Guo, H.; Wu, X. Review of Fault Diagnosis Methods for Large Wind Turbines. *Power Syst. Technol.* **2017**, *41*, 3480–3491.
4. Chen, X.F.; Guo, Y.J.; Xu, C.; Shang, H. Review of Fault Diagnosis and Health Monitoring for Wind Power Equipment. *China Mech. Eng.* **2020**, *31*, 175–188.
5. Zheng, X.X.; Ye, C.J.; Yang, F. Research and Development of Operation and Maintenance for Offshore Wind Farms. *Power Syst. Energy* **2012**, *28*, 90–94.
6. Zeng, Z. Research on Fault Diagnosis Method of Wind Turbine Transmission System Based on Deep Learning. Master's Thesis, Northeast Electric Power University, Jilin, China, 2021.
7. Zhang, L.; Lang, Z.Q. Wavelet Energy Transmissibility Function and its Application to Wind Turbine bearing Condition Monitoring. *IEEE Trans. Sustain. Energy* **2018**, *9*, 1833–1843. [CrossRef]
8. Wang, J. Research on Fault Diagnosis Technology of Planetary Gearboxes Based on Deep Learning. Master's Thesis, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 2018.
9. Cao, L.; Zhang, J.; Wang, J.; Qian, Z. Intelligent fault diagnosis of wind turbine gearbox based on Long short-term memory networks. In Proceedings of the IEEE International Symposium on Industrial Electronics, Vancouver, BC, Canada, 12–14 June 2019; pp. 890–895.
10. Wu, C.Z.; Jiang, P.C.; Ding, C.; Feng, F.; Chen, T. Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network. *Comput. Ind.* **2019**, *108*, 53–61. [CrossRef]
11. Yao, L.; Sun, J.J. Fault Diagnosis Method for Rolling Bearing Based on Gramian Angular Fields and CNN-RNN. *Bearing* **2022**, *2*, 61–67.
12. Zhou, Y.; Sun, H.Y.; Fang, Q.; Xia, H. Review of imbalanced data classification methods. *Appl. Res. Comput.* **2022**, *39*, 1615–1621.
13. Lu, J.L.; Zhang, X.G. Fault Diagnosis of Main Bearing based of Wind Turbine Based on Improved Auxiliary Classifier Generative Adversarial Network. *Autom. Electr. Power Syst.* **2021**, *45*, 148–154.
14. Li, D.D.; Liu, Y.H. Fault Diagnosis Method of Wind Turbine Planetary Gearbox Based on Improved Generative Adversarial Network. *Proc. CSEE* **2021**, *41*, 7496–7506.
15. Zhao, M.; Zhong, S. Deep Residual Shrinkage Networks for Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 4681–4689. [CrossRef]
16. Che, C.C.; Wang, H.W. Fault diagnosis of rolling bearing based on deep residual shrinkage network. *J. Beijing Univ. Aeronaut. Astronaut.* **2021**, *47*, 1399–2021.
17. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *3*, 2672–2680.
18. Wang, P.; Hou, B.R. ECG arrhythmias detection using auxiliary classifier generative adversarial network and residual network. *IEEE Access* **2019**, *7*, 100910–100922. [CrossRef]
19. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June 2016.
20. Xu, M.; Wang, P. Research on Rolling Bearing Fault Diagnosis Based on FB-LSTM Resnet. Available online: http://kns.cnki.net/kcms/detail/41.1148.th.20220507.1708.006.html (accessed on 22 July 2022).
21. Yang, S.L. Bearing Fault Diagnosis Based on Hybrid Domain Attention Mechanism Convolutional Network and Residual Contraction Network. Master's Thesis, School of Information and Communication Engineering, Beijing, China, 2020.
22. Lu, J.L.; Zhu, C.F. Power Transformer Fault Diagnosis Technology Based on Improved ACGAN Data Augmentation. *Electr. Power Sci. Eng.* **2021**, *37*, 42–51.