Article

MCBA-MVACGAN: A Novel Fault Diagnosis Method for Rotating Machinery Under Small Sample Conditions

Wenhan Huang [ID], Xiangfeng Zhang *, Hong Jiang, Zhenfa Shao and Yu Bai [ID]

School of Mechanical Engineering, Xinjiang University, Urumqi 830047, China; 107552204190@stu.xju.edu.cn (W.H.)
* Correspondence: xjuzxf@xju.edu.cn

Abstract: In complex industrial scenarios, high-quality fault data of rotating machinery are scarce and costly to collect. Therefore, small sample fault diagnosis needs further research. To solve this problem, in this work is proposed a minimum variance auxiliary classifier generation adversarial network based on a multi-scale convolutional block attention mechanism. Firstly, the multi-scale convolutional block attention mechanism is designed to extract multi-scale information and perform weighted fusion to enhance the ability of the model to capture effective features. Secondly, the minimum variance term is designed to minimize the variance of sample distribution, so that the generated samples are distributed more evenly in the feature space, avoiding the problem of pattern collapse. Finally, the objective function is reconstructed by independent classification loss to improve the ability of model data generation. Experimental results on CWRU and gearbox datasets validate the effectiveness and reliability of the proposed method.

Keywords: fault diagnosis; generative adversarial network; small sample; attention mechanism; minimum variance

1. Introduction

Rotary machinery is extremely important in industrial applications, but due to its harsh working environment, it is prone to unexpected failures that can lead to serious production losses and human accidents. Therefore, the troubleshooting of rotating machinery has been widely considered by various researchers [1,2].

Traditional signal processing methods include wavelet transform [3], the time–frequency analysis method [4], and empirical mode decomposition [5]. These techniques extract periodic impulse features from vibration signals to determine the fault location and type. However, their effectiveness is limited when dealing with unclear fault mechanisms and invalid fault-related features. Therefore, machine learning such as support vector machines [6], naive Bayes [7], and neural networks [8,9] has attracted a lot of attention. By analyzing patterns in historical failure data, these methods identify early signs of equipment failure, predict potential failures, and provide strong support for maintenance decisions. However, their diagnostic performance may decline when faced with low-quality or insufficient data. Traditional machine learning methods need to choose the right features and rely on expert experience.

Deep learning is widely used for its unique advantages, such as automatic feature learning, handling large-scale data, modeling complex nonlinear relationships, and not relying on prior knowledge [10–13]. Xu et al. [14] improve the fault diagnosis accuracy of rotating machinery by improving the convolutional neural network (CNN). Zhou et al. [15] designed a multi-channel wide-kernel convolution and integrated discrete wavelet transform blocks into the CNN for gearbox fault diagnosis. Ye et al. [16] introduced a capsule

layer into the convolutional neural network to vectorize features and embedded an attention mechanism to address the issue of small sample fault diagnosis with noise. Wang et al. [17] developed a neural transformer with strong robustness, and the proposed multi-head spatio-temporal peak self-attention mechanism abandoned the tedious and costly multiplication operation, reduced the calculation amount, and realized high-precision fault diagnosis. Zhou et al. [18] proposed a kind of unbalanced depth subdomain adaptive network to achieve high-precision fault diagnosis in complex cross-domain scenarios. Wang et al. [19] proposed a lightweight model based on a progressive joint transfer integrated network to realize high-precision unbalanced fault diagnosis. However, the above methods still have some limitations. The effective training of these supervision methods involving a large number of parameters requires sufficient labeled fault samples, and the acquisition and labeling of large amounts of data is time-consuming and costly [20,21]. However, in actual industrial production, it is very difficult to obtain enough effective labeling data for rotating machinery [22]. Therefore, the use of limited labeled fault samples to achieve high-precision fault diagnosis remains to be further studied.

Due to the limitations of obtaining labeled fault data in real-world scenarios, research on unsupervised learning models has gained extensive attention. Unsupervised learning is a method that does not require labeled fault data. Instead, it trains by analyzing the intrinsic structure and patterns within the data, effectively reducing the cost and time required for data preparation [23,24]. Generative Adversarial Network (GAN) is an unsupervised generation model that can generate new data similar to existing data without requiring any labeled data [25]. Therefore, GAN-based methods are widely used in various fields [26–28]. In recent years, GAN-based methods have been widely used in rotating machinery fault diagnosis. Gu et al. [29] used Wasserstein to introduce the attention mechanism in gradient penalty GAN and designed the cosine similarity loss function for high-precision fault diagnosis under unbalanced conditions. Zhang et al. [30] developed a GAN combined with a gradient penalty to study the performance of a multi-structure GAN. Chen et al. [31] integrated a Pre-Trained Network (PT) into WGAN-GP to achieve efficient feature extraction, thus improving fault diagnosis accuracy. Fu et al. [32] combined ACGAN with transformer networks, avoiding traditional iterative and convolution structures and improving motor bearing fault diagnosis accuracy. Zhang et al. [33] embedded the Convolution Block Attention Mechanism (CBAM) module into the least square GAN and introduced conditional regularization loss into the least square loss function to effectively extract data feature extraction. In summary, the application results of the GAN-based method in rotating machinery fault diagnosis are encouraging. However, there is still significant potential for further exploration and application of GAN's data generation capabilities, especially in small sample scenarios.

Although many of the aforementioned studies have utilized Auxiliary Classifier GAN (ACGAN) in certain fields, ACGAN, despite its ability to improve the ability of the model to capture effective features of real samples by incorporating label information and auxiliary classifiers, still has several limitations. (1) Most ACGAN models use Jensen–Shannon (JS) divergence and ignore its limitations. Because of its discretization, the gradient disappears when the real and generated distributions hardly overlap [34]. (2) Faced with complex data with high dimensions, or when the discriminator's performance is much stronger than that of the generator, ACGAN is prone to pattern and generate only a subset of some training samples while ignoring others. As a result, the diversity of the generated data is low. (3) In ACGAN, the discriminator plays the role of both classification and discrimination. When the discriminator produces an incorrect output, the classification task takes up too much performance, or the resulting sample quality may decrease [35]. ACWGAN-GP uses Wasserstein distance instead of JS divergence to directly represent the minimum cost of

converting the generated sample distribution to the real sample distribution. It can provide effective gradient information when the distribution is very different, and the gradient penalty can keep the gradient norm of the discriminator in the range of close to one, so as to avoid the gradient disappearing or exploding due to being too large or too small [36]. However, ACWGAN-GP still has the limitation of mode collapse in the face of the scarcity of effective training data.

In order to solve these problems, a fault diagnosis method of rotating machinery based on CBAM-MVACGAN is proposed in this paper. The performance of the proposed method is verified on CWRU and SQI datasets. The results show that this method can generate high-quality fault samples under the condition of a small number of training samples, and can realize high precision small sample fault diagnosis.

The main contributions of this paper are as follows:

(1) A minimum variance term is designed to make the generated samples evenly distributed and make up for the mode collapse problem of ACWGAN-GP.

(2) A multi-scale convolution block attention mechanism (MCBA) is designed to reconstruct channel attention blocks through multi-scale information, and combine the calculated weights with multi-scale information through sigmoid function to retain CBAM spatial attention blocks and obtain new effective features.

(3) Improve the model architecture through an independent classifier, reconstruct the classification loss generator loss, and improve the quality of generated samples.

(4) A minimum variance auxiliary classifier generation adversarial network (MCBA-MVACGAN) based on a multi-scale convolution block attention mechanism is designed for small sample rotating machinery fault diagnosis. This method can effectively extract data features and improve the quality of generated samples.

Section 2 introduces important relevant theories. Section 3 discusses the MCBA-MVACGAN. In Section 4, the performance of MCBA-MVACGAN is verified by a large number of experiments. Finally, Section 5 concludes the paper.

2. Relevant Theory

2.1. GAN-Based Method

Generators and discriminators are trained alternately to competitively optimize their parameters and optimize the model performance [25], whose objective function is:

$$\min_G \max_D L_{GAN}(D,G) = E_{x \sim P_r(x)}[\log D(x)] + E_{x \sim P_r(z)}[(\log(1 - D(G(z))))] \tag{1}$$

where z is the random noise vector, and E represents the mathematical expectation.

To solve the problems of training instability and mode collapse in GAN, WAGN introduces Wasserstein distance as a measure of distributed distance, as follows:

$$W(P_r, P_g) = \inf_{\gamma \in \prod(P_r, P_g)} E_{(x, \tilde{x}) \sim y}[\|x - \tilde{x}\|] \tag{2}$$

where inf is the infimum, and $\tilde{x}$ denotes the generated samples. To optimize the objective $W(P_r, P_g)$, where a smaller value indicates that P_g is closer to P_r, and to ensure that the function's gradient does not change too drastically, thereby allowing the model to update iteratively with normal gradients and addressing the performance limitations and instability issues caused by weight clipping in WGAN, a K-Lipschitz constraint and gradient penalty term are introduced [31]. The expression is as follows:

$$W(P_r, P_g) = \frac{1}{K} \sup_{\|f\| L \leq K} E_{x \sim P_r}[f(x)] - E_{x \sim P_g}[f(x)] + \lambda E_{\hat{x} \sim P_{\hat{x}}}[\|\nabla \hat{x} D(\hat{x})\|_2 - K]^2 \tag{3}$$

where is the supremum, λ is the regularization coefficient, and K is a constant representing the range within which the gradient is constrained. $P_{\hat{x}}$ is the uniform sampling along the straight line between the point pairs of the actual data distribution P_r and the generated distribution. The purpose of this is to apply the constraint evenly along the straight line because the optimal discriminator D consists of a straight line connecting two distributions.

ACGAN is an improved GAN that can be conditionally generated using class label information to generate samples with specified class characteristics. The discriminator improves the learning effect of the generator through confrontation and classification tasks [37]. Figure 1 shows the model structure of ACGAN, and the formula is as follows:

$$L_{Source} = E_{x \sim P_x(x)}[\log D(x)] + E_{z \sim P_z(z)}[\log(1 - D(G(z, c)))] \tag{4}$$

$$L_{Class} = E_{x \sim P_r(x)}[-\log P(class = c \mid x)] + E_{z \sim P_z(z)}[-\log P(class = c \mid G(z))] \tag{5}$$

where L_{Source} is the discriminator G loss function, L_{Class} is the classifier loss function, G generates sample $G(z, c)$ using noise z and labels c. Through continuous adversarial training, the model capability is optimized to eventually generate fault samples with specific labels.
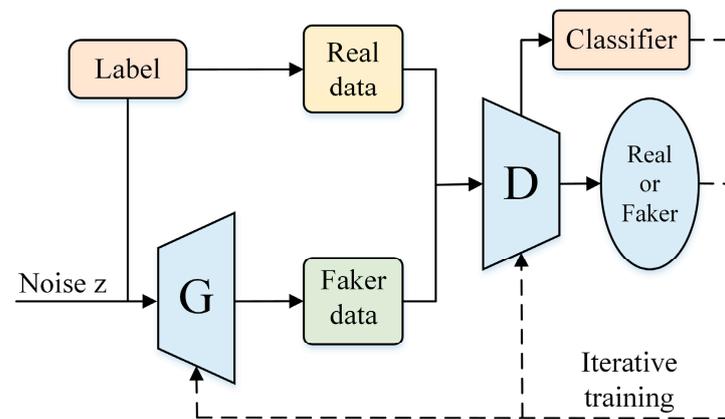

Figure 1. The Structure of ACGAN.

2.2. Convolution Block Attention Mechanism

Convolutional block attention module (CBAM) improves the feature extraction process of convolutional neural networks by integrating channel attention mechanisms and spatial attention mechanisms [38]. Its structure is shown in Figure 2. Unlike traditional attention mechanisms that typically focus on a single dimension, CBAM constructs a sequential structure from the channel dimension to the spatial dimension, providing a comprehensive multi-dimensional enhancement of feature maps within convolutional neural networks. In the channel attention stage, the importance weight of each channel is obtained. By multiplying the original feature graph by channel, the dependence on the important channel is enhanced, and the influence of the weaker channel is reduced. In the spatial attention stage, the attention map is obtained, weighted, and fused with the original feature map, and the position with the higher response is highlighted to generate a new feature map. CBAM is as follows:

$$U = \varphi_{sa}(Y) \otimes Y = \varphi_{sa}[\varphi_{ca}(X) \otimes X] \otimes [\varphi_{ca}(X) \otimes X] \tag{6}$$

where $\otimes$ is the element multiplication symbol, φ_{sa} is the spatial attention mechanism, and φ_{ca} is the channel attention mechanism.
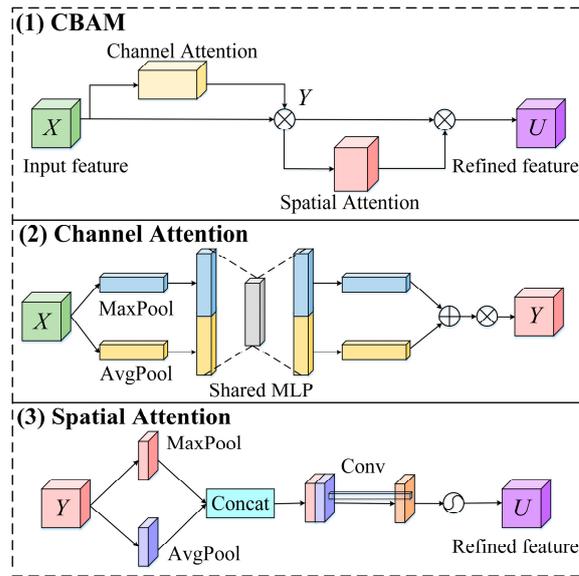
Figure 2. The structure of convolutional block attention mechanism.

3. Proposed Method

3.1. MCBA-MVACGAN

3.1.1. Multi-Scale Convolution Block Attention Mechanism

The traditional CBAM's channel attention block overlooks features at different scales, and single-scale convolutional kernels are insufficient to capture diverse features, thereby limiting the model's expressive capacity. At the same time, a select kernel network (SKN) can capture image information more comprehensively through a multi-scale convolution kernel for feature extraction [39]. However, the softmax function of SKN has the limitation of ignoring non-optimal features, and overreliance on a certain feature may occur in the case of data scarcity [40]. Therefore, this paper uses the sigmoid function instead of the softmax function to improve the feature extraction capability of SKN, and uses it to replace the channel attention block in CBAM. A multi-scale convolution block attention mechanism (MCBA) is proposed. The structure is shown in Figure 3. First, given any input feature map, multi-scale information is extracted through convolution operations with kernel sizes of 3 and 5, respectively: $f_{\text{Conv}}^{3\times3} : X \to X_{\text{Conv}}^{3\times3}$ and $f_{\text{Conv}}^{5\times5} : X \to X_{\text{Conv}}^{5\times5}$. Secondly, in order to obtain the attention weight of multi-scale information, the features of different branches are combined:

$$X^{\text{Mult}} = X_{\text{Conv}}^{3\times3} \oplus X_{\text{Conv}}^{5\times5} \tag{7}$$

where $\oplus$ represents the addition of elements.

Then, the combined feature X^{Mult} is globally average pooled to compress the information of each channel: $f_{\text{Avg}}^{\text{Ch}} : X^{\text{Mult}} \to X_{\text{Avg}}^{\text{Mult}}$. For the nth element in $X_{\text{Avg}}^{\text{Mult}}$, the expression is as follows:

$$X_{\text{Avg}}^{\text{Mult}}n = f_{\text{Avg}}^{\text{Ch}}(X_n^{\text{Mult}}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_n^{\text{Mult}}(i,j) \tag{8}$$

where i and j denote the position index in the height and width dimensions of the feature map, respectively.

Then, the attention weight of multi-scale information is learned through two fully connected layers, W_1 and W_2:

$$Z_{\text{Ch}}^{\text{Mult}} = W_2(\delta(B(W_1 X_{\text{Avg}}^{\text{Mult}}))) \tag{9}$$

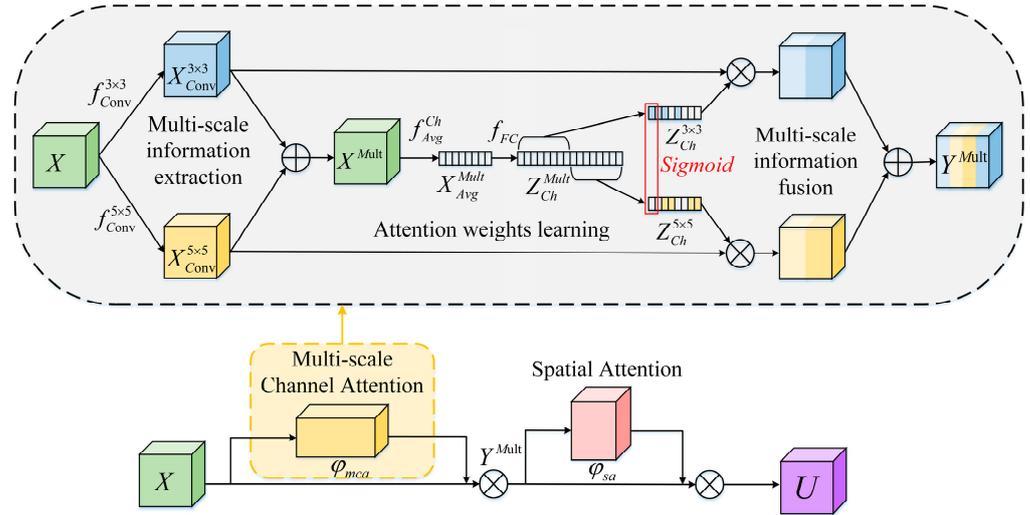where B is the batch normalization, and δ is the ReLU function.

Figure 3. The structure of Multi-scale Convolution Block Attention mechanism.

Then, the channel attention weight $Z_{\text{Ch}}^{\text{Mult}}$ is divided into two equal components: $Z_{\text{Ch}}^{3\times3}$ and $Z_{\text{Ch}}^{5\times5}$, and the sigmoid function is applied to the corresponding positions of $Z_{\text{Ch}}^{3\times3}$ and $Z_{\text{Ch}}^{5\times5}$ to achieve further integration of multi-scale information. The weighted fusion of the attention weight information of each branch of the final channel attention feature graph U^{Mult} on the channel is calculated, which can be expressed as:

$$Y^{\text{Mult}} = \varphi_{\text{mca}}(X) \otimes X = \left(\sum_{n=1}^{N} (X_{\text{Conv}}^{3\times3}n \otimes Z_{\text{Ch}}^{3\times3}n \oplus X_{\text{Conv}}^{5\times5}n \otimes Z_{\text{Ch}}^{5\times5}n) \right) \otimes X \tag{10}$$

According to Section 2.2, the MCBA is as follows:

$$U = \varphi_{\text{sa}}\left(Y^{\text{Mult}} \right) \otimes Y^{\text{Mult}} \tag{11}$$

3.1.2. Minimum Variance-Assisted Classification Generates Adversarial Networks

In order to alleviate the mode collapse problem existing in ACWGAN-GP, the proposed method adds a minimum variance term to the discriminator loss, which can be expressed as:

$$L_{\text{var}} = \text{Var}_{x\sim P_{r(x)}}[D(x)] - \text{Var}_{z\sim P_{z(z)}}[D(G(z))]/2 \tag{12}$$

where Var is the variance, $D(x)$ is the output of the discriminator against the real sample x, and $D(G(z))$ is the output of the discriminator D on the sample $G(z)$ generated by the generator G from random noise z. Sample variance $\text{Var}(X)$ and sample mean μX are introduced to calculate the formula:

$$\text{Var}(X) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu X) \tag{13}$$

$$\mu X = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{14}$$

where x_i is the i sample value. Combined with Formula (12), which is expressed as:

$$L_{\text{var}} = E_{x\sim P_{r(x)}}\left[D(x)^2 \right]/2 - (E_{x\sim P_{r(x)}}[D(x)])^2/2 - E_{x\sim P_{r(x)}}\left[D(G(z))^2 \right]/2 + E_{x\sim P_{r(x)}}[D(G(z))]^2/2 \tag{15}$$

In this case, it can be known that when L_{var} is positive, the generator reduces the variance of the fake sample. Instead, it increases the output variance of the generated sample, bringing it closer to the true distribution. Through back propagation and gradient descent, parameters are constantly adjusted parameters to align distribution distances. Therefore, minimum variance loss can alleviate overfitting and gradient instability to a certain extent.

Further, classification losses are considered to improve model performance [35], expressed as:

$$L_C = \left(L_C^G + L_C^R\right)/2 = E_{x \sim Pr(x)}[-\log P(c = c_r|x)]/2 + E_{z \sim P_z}\left[-\log P(c = c_g|G(z, c_g))\right]/2 \tag{16}$$

where L_C^R and L_C^G are the classification loss functions of real samples and generated samples:
Finally, the MCBA-MVACGAN objective function is as follows:

$$\min_D V_{\text{MVACGAN}}(D) = L_{\text{Wasserstein}} - \lambda L_{GP} + \lambda_1 L_{\text{Var}} = E_{x \sim P_{r(x)}}[D(x)] - E_{z \sim P_{z(z)}}\left[D(G(z, c_g))\right] - \lambda E_{\hat{x} \sim P_{\hat{x}}}\left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2\right] + \lambda_1 L_{\text{var}} \tag{17}$$

$$\min_G V_{\text{MVACGAN}}(G) = E_{z \sim P_{z(z)}}\left[D(G(z, c_g))\right]/2 + L_C/2 \tag{18}$$

where L_{GP} is the gradient penalty term, λ is the gradient penalty factor, and λ_1 is the minimum variance factor.

Figure 4 shows the diagnostic flow of the proposed method. Algorithm 1 shows the training algorithm of the proposed method.

Algorithm 1. MCBA-MVACGAN Training Process

Input: Number of iterations N, batch size m, learning rate s α_G, α_D, α_C. Adam hyperparameters β_1, β_2, ratio factors λ, λ_1
Parameter initialization: Initial discriminator parameters θ_D, generator parameters θ_G, and classifier parameters θ_C
1: **For** k = 1 to N **do**
2:     **For** i = 1 to n **do**
3:         Sample real data $\{x_i\}_{i=1}^m$ and noise $\{z_i\}_{i=1}^m$ with labels $\{c_i\}_{i=1}^m$
4:         Generate fake data x_i', $c_i' \leftarrow G_{\theta_G}(z_i, c_i)$
5:         Calculate D's output on real and fake data: $D(x_i)$, $D(x_i')$
6:         $L_{GP} \leftarrow E_{\hat{x} \sim P_{\hat{x}}}\left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2\right]$
7:         $L_{\text{var}} \leftarrow \left(\text{Var}_{x \sim P_{r(x)}}[D(x)] - \text{Var}_{z \sim P_{z(c)}}\left[D(G(z, c_i'))\right]\right)/2$
8:         $L_D \leftarrow D(x_i) - D(x_i') - \lambda L_{GP} - \lambda_1 L_{\text{var}}$
9:         Update discriminator: $\theta_D \leftarrow \text{Adam}(\nabla_{\theta_D} L_D, \theta_D, \alpha_D, \beta_1, \beta_2)$
10:     **End for**
11:     Sample real data $\{x_i\}_{i=1}^m$ and noise $\{z_i\}_{i=1}^m$ with labels $\{c_i\}_{i=1}^m$
12:     Generate fake data x_i', $c_i' \leftarrow G_{\theta_G}(z_i, c_i)$
13:     Calculate C's output on real and fake data: $P_{\theta_C}(c = c_i^r|x_i)$, $P_{\theta_C}(c = c_i'|x_i')$
14:     $L_C \leftarrow \frac{1}{2}\left(-\log P_{\theta_C}(c = c_i^r|x_i) - \log P_{\theta_C}(c = c_i'|x_i')\right)$
15:     Update classifier: $\theta_C \leftarrow \text{Adam}(\nabla_{\theta_C} L_C, \theta_C, \alpha_C, \beta_1, \beta_2)$
16:     $L_G \leftarrow D(x_i') + L_C$
17:     Update generator: $\theta_G \leftarrow \text{Adam}(\nabla_{\theta_G} L_G, \theta_G, \alpha_G, \beta_1, \beta_2)$
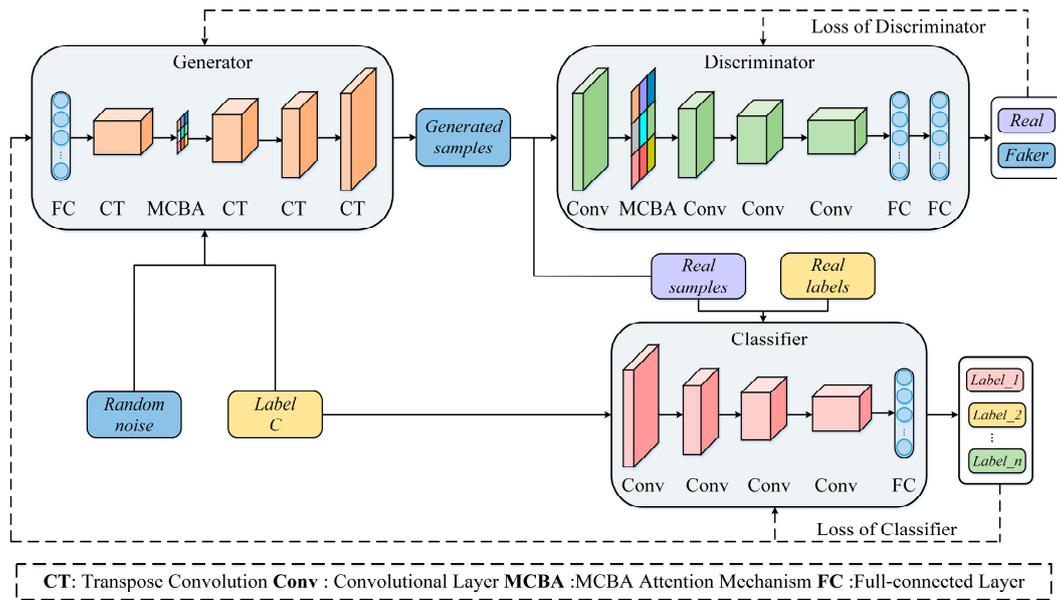18: **End for**

Figure 4. The structure of MCBA-MVACGAN.

3.2. Intelligent Fault Diagnosis Based on Deep Convolutional Neural Network

Deep Convolutional neural network (DCNN) has been widely used in fault diagnosis because of its excellent feature extraction capability and end-to-end learning framework. Therefore, this paper uses DCNN as the fault diagnosis classification model. The MCBA-MVACGAN generated fake samples and actual samples as a training set for training, and other real samples are taken as the test set and input into the classifier DCNN. First, the input training sample is convolved to reduce the size to achieve dimensionality reduction. Batch standardization is then carried out to accelerate training and reduce the risk of overfitting. Use the LeakyReLU activation function to add nonlinearity. After such an operation is carried out four times, the feature map is flattened into a one-dimensional feature vector through the fully connected layer, and fault identification is finally realized. The structure of DCNN is shown in Figure 5.
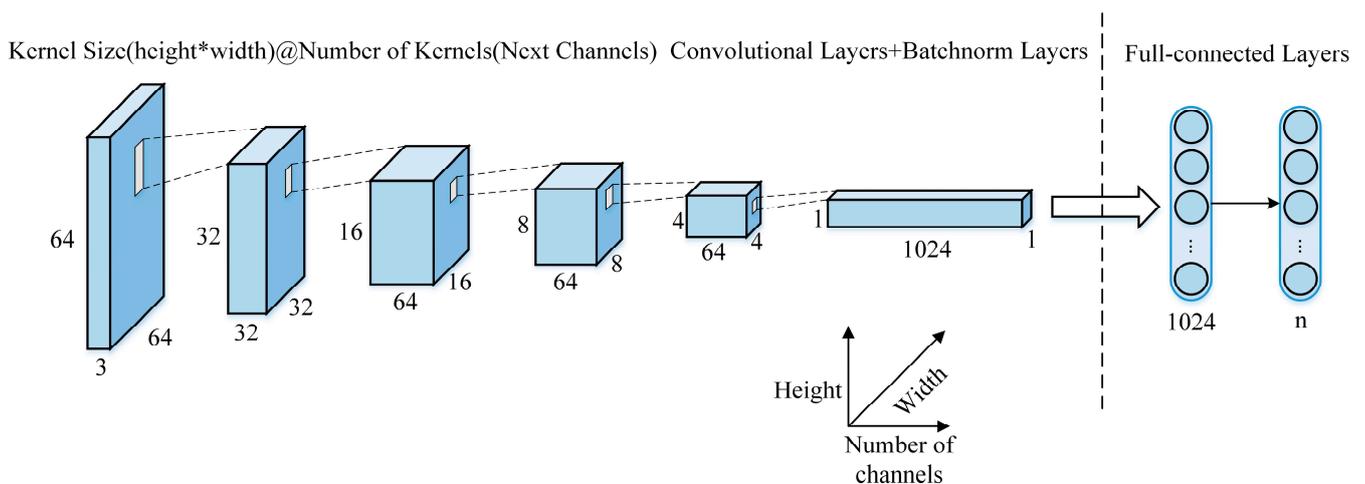

Figure 5. The Structure of DCNN.

3.3. MCBA-MVACGAN Fault Diagnosis Procedure

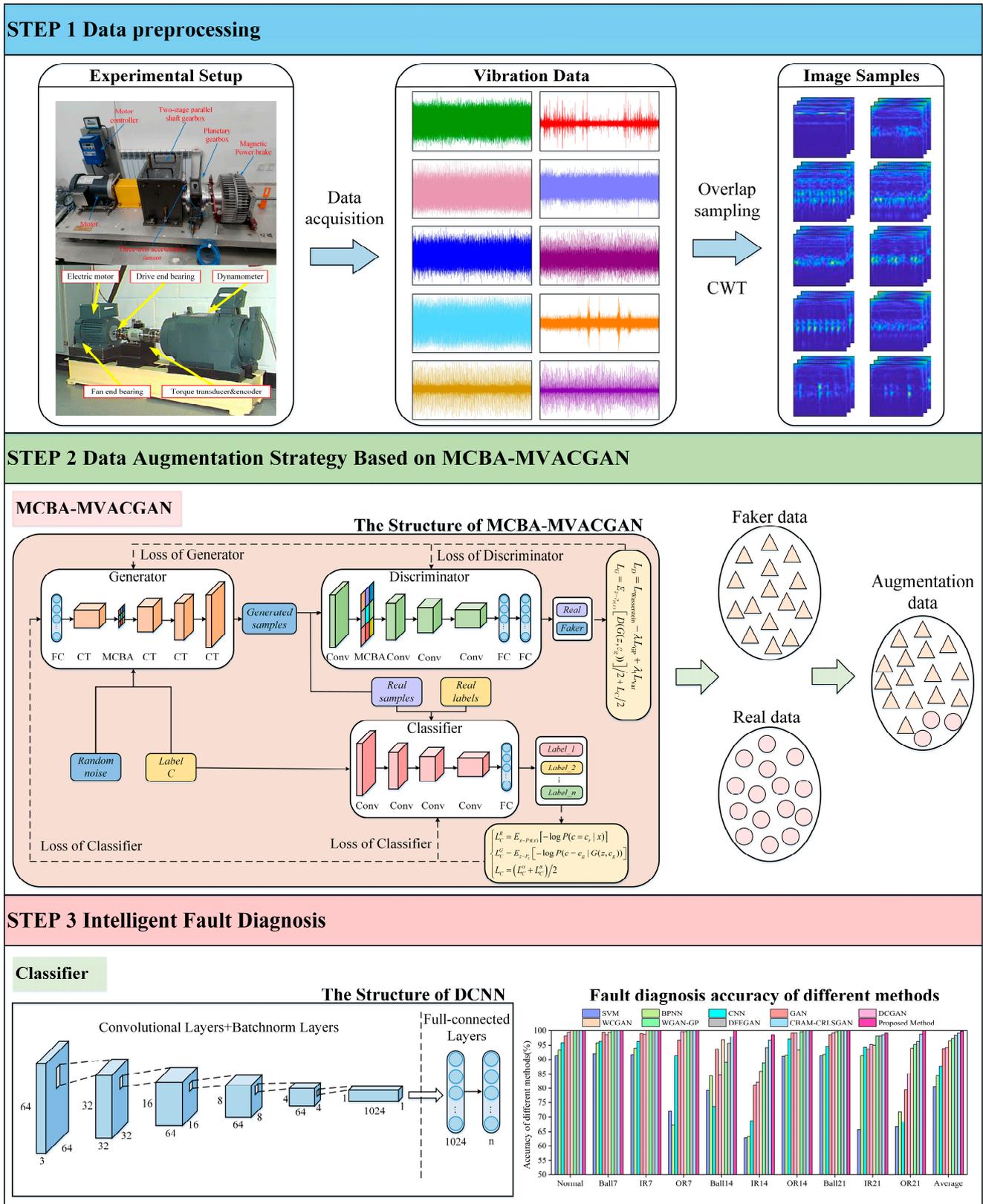Figure 6 shows the overall diagnosis process of MCBA-MVACGAN, as follows:

Figure 6. Overall flow chart of rotating machinery fault diagnosis method based on MCBA-MVACGAN.

Step 1 Time–frequency sample generation method based on CWT: collect bearing vibration signals of various fault types, convert them into time–frequency images, and then divide them into a small number of training samples and enough test samples.

Step 2 Data enhancement strategy based on MCBA-MVACGAN: a small number of samples are input into the proposed MCBA-MVACGAN model, and a large number of fake samples similar to real samples are generated to form a new training set for intelligent fault diagnosis in the next step.

Step 3 Intelligent fault diagnosis: the new training set is input into the DCNN model for training. The test set evaluates the diagnostic performance of MCBA-MVACGAN.

4. Experimental Verifications

4.1. Dataset Description

Dataset A: The Case Western Reserve University (CWRU) [41] dataset collects vibration signals from rolling bearings under multiple working conditions and failure modes. The test stand is composed of a motor, a torque transducer/encoder, and a dynamometer. The testing bearings of the drive end and the fan end are SKF deep groove ball bearings: 6205-2RS JEM and 6203-2RS JEM. The experimental platform is shown in Figure 7. In order to simulate different faults, a single-point defect is created on the outer ring, inner ring, and roller of the bearing by means of EDM. The fault levels are 0.007, 0.014, 0.021, and 0.04. Each bearing was tested with four different loads (0, 1, 2, 3 hp) at a constant speed (approximately 1720–1797 r/min). In the process of data acquisition, the accelerometer is respectively installed at the 12 o'clock position of the fan end and the drive end of the motor housing to collect vibration signals. A 16-channel data logger was used to collect signals with sampling frequencies of 12 kHz and 48 kHz, respectively. Torque sensors/encoders are used to collect horsepower and speed data [41].
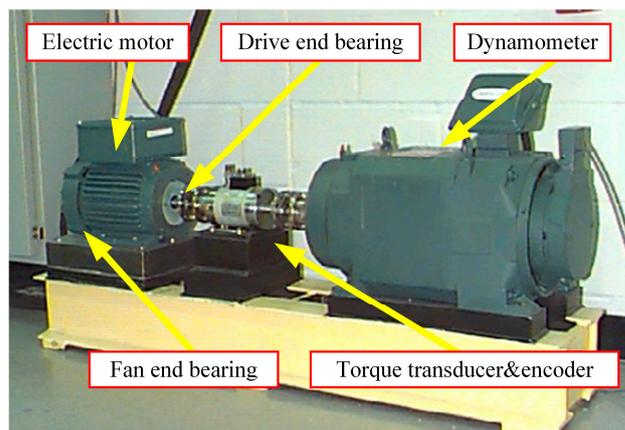

Figure 7. Experimental setup of bearing dataset.

Dataset B: The structure of the SQI wind turbine drive system fault diagnosis test bench is shown in Figure 8, which is mainly composed of a motor, motor controller, two-stage parallel shaft gearbox, planetary gearbox, magnetic powder brake, etc. The number of gear teeth on the input shaft and output shaft is 36 and 100, respectively, and the number of gear teeth on the idler shaft is 90 and 28, respectively. The gear module is 1.5 and the transmission ratio is 8.92. The experimental data were collected under 298.28 W, 1500 r/min constant speed transmission. The sensor used for data acquisition is a piezoelectric three-axis acceleration sensor. The sensor has a sensitivity of 95.8 mV/g and an output bias of 10.9 VDC. The bearing used for testing is a deep groove ball bearing with the model number ER16KCL. The failure components of the gearbox bearing used in the experiment are shown in Figure 8, including the failure of the bearing roller, the failure of the bearing outer ring, and the failure of the bearing inner ring. After data collection, the faulty bearing is installed in a parallel shaft gearbox. The data contain four transmission states, including normal state and three single fault states. The vibration signal of the gearbox is collected

by the acceleration sensor placed on the auxiliary parallel shaft gearbox, the sampling frequency is 20,480 Hz, and the sampling time is 50 s.
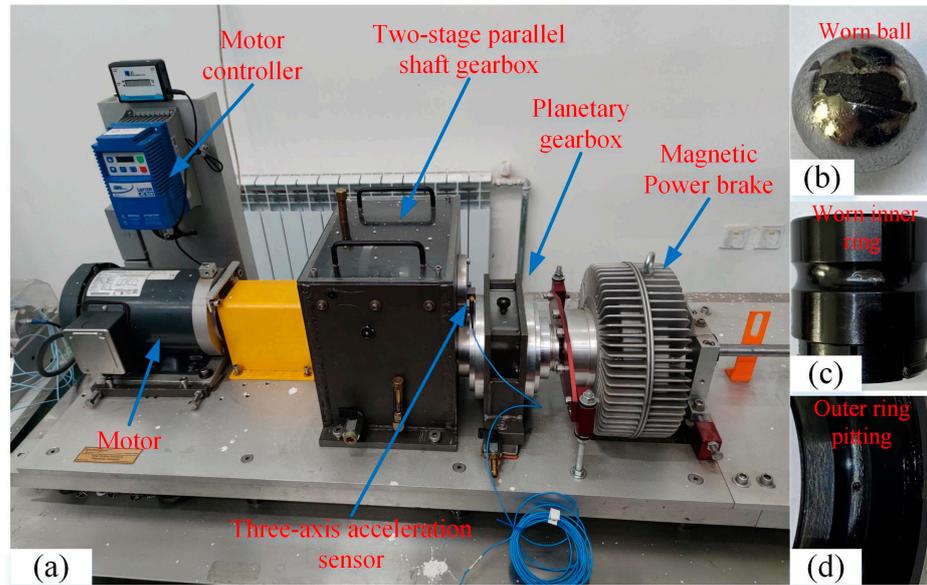

Figure 8. Parallel shaft gearbox experimental platform and faulty parts: (**a**) Experimental platform, (**b**) BF, (**c**) IF, (**d**) OF.

For every state of health, we adopted a sliding window to intercept non-overlapping vibration signals, selected 2048 points as sampling Windows, and randomly selected 120 samples. Finally, the time–frequency graph samples of each health state are obtained by continuous wavelet transform. The specific dataset partition is shown in Table 1. Figure 9 shows the time-domain waveform of dataset A and dataset B.
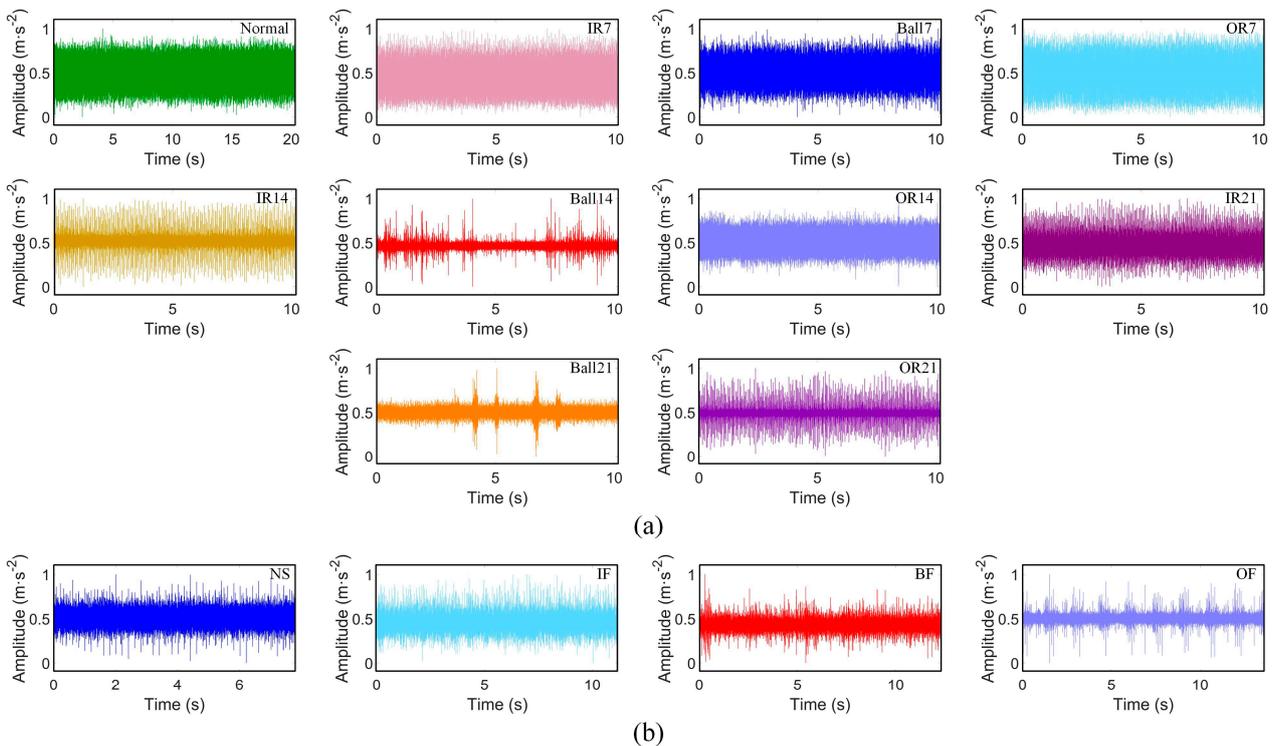

Figure 9. Time-domain waveform of dataset A and dataset B: (**a**) CWRU; (**b**) SQI.

Dataset	Health State	Training Sample	Testing Sample	Class Labels
Dataset A	Normal	20	100	Normal
	Inner-ring fault 0.007	20	100	IR7
	Ball fault 0.007	20	100	Ball7
	Outer-ring fault	20	100	OR7
	Inner-ring fault 0.014	20	100	IR14
	Ball fault 0.014	20	100	Ball14
	Outer-ring fault 0.014	20	100	OR14
	Inner-ring fault 0.021	20	100	IR21
	Ball fault 0.021	20	100	Ball21
	Outer-ring fault 0.021	20	100	OR21
Dataset B	Normal	20	100	NS
	Outer-ring fault	20	100	OF
	Inner-ring fault	20	100	IF
	Ball fault	20	100	BF

4.2. Data Preprocessing

In order to increase the number of samples, this paper conducted overlapping sampling on the original data, and the overlap rate was 0.5. Considering the high efficiency and strong adaptability of Complex Morlet wavelets in the time–frequency field, it is selected as the wavelet basis to generate a time–frequency image through continuous wavelet transform [42]. Its function expression is:

$$W_f(s, \tau) = \int_{-\infty}^{\infty} f(t) \cdot \pi^{-\frac{1}{4}} e^{-j\omega_0(t-\tau)} e^{-\frac{(t-\tau)^2}{2s^2}} dt \tag{19}$$

where $f(t)$ is the signal, ω_0 is the angular frequency parameter of the wavelet, and τ is the translation parameter. $W_f(s, \tau)$ is the time–frequency representation of the Morlet wavelet function convolution with the signal at different scales and shifts. The data preprocessing process is shown in Figure 10.
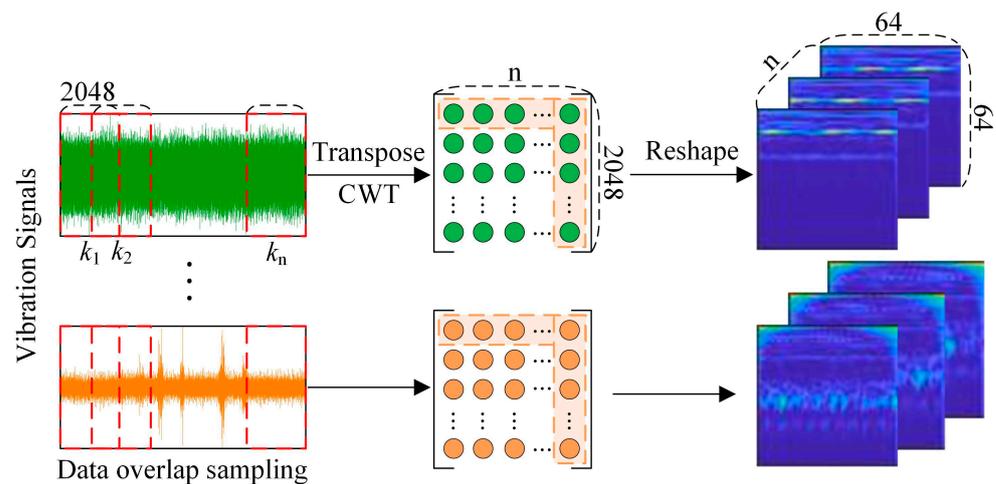

Figure 10. Data preprocessing.

4.3. Experimental Parameter

Table 2 shows the MCBA-MVACGAN network structure. The input shape is (20, 3, 64, 64), that is, 20 samples, 3 channels, image length, and width are 64. A Gaussian noise of input shape (1, 100) is fed into a generator, which synthesizes a false sample of shape (3, 64, 64) and feeds the false sample into a discriminator.

Table 2. Structure of the designed MCBA-MACGAN.

Network Structure	Layers	Convolution Kerne Size/Step	Output Size
Generator	Linear	/	512@4 × 4
	Transposed convolution	4 × 4/2	256@8 × 8
	CBAM	/	256@8 × 8
	Transposed convolution	4 × 4/2	128@16 × 16
	Transposed convolution	4 × 4/2	64@32 × 32
	Transposed convolution	4 × 4/2	3@64 × 64
Discriminator	Convolution	4 × 4/2	64@32 × 32
	CBAM	/	64@32 × 32
	Convolution	4 × 4/2	128@16 × 16
	Convolution	4 × 4/2	256@8 × 8
	Convolution	4 × 4/2	512@4 × 4
	Flatten	/	8192
	Linear	/	1024
	Linear	/	1
Classifier	Convolution	4 × 4/2	32@32 × 32
	Convolution	4 × 4/2	64@16 × 16
	Convolution	4 × 4/2	128@8 × 8
	Convolution	4 × 4/2	256@4 × 4
	Flatten	/	4096
	Linear	/	10

In order to correctly determine the training epoch of MCBA-MVACGAN, Figure 11a plots the loss of generator G, discriminator D, and classifier C with the epoch. In the initial training phase, D and C compete with each other, and the loss of G and D fluctuates wildly. After 1200 rounds of training, the loss of G, D, and C converges. Therefore, to ensure better fake samples are generated, the training epoch is set to 2000. For the learning rate of the training, the generator is set to 0.0001, the discriminator is set to 0.0001, and the classifier is set to 0.0001. Update network parameters using the Adam optimizer.

(a)

(b)
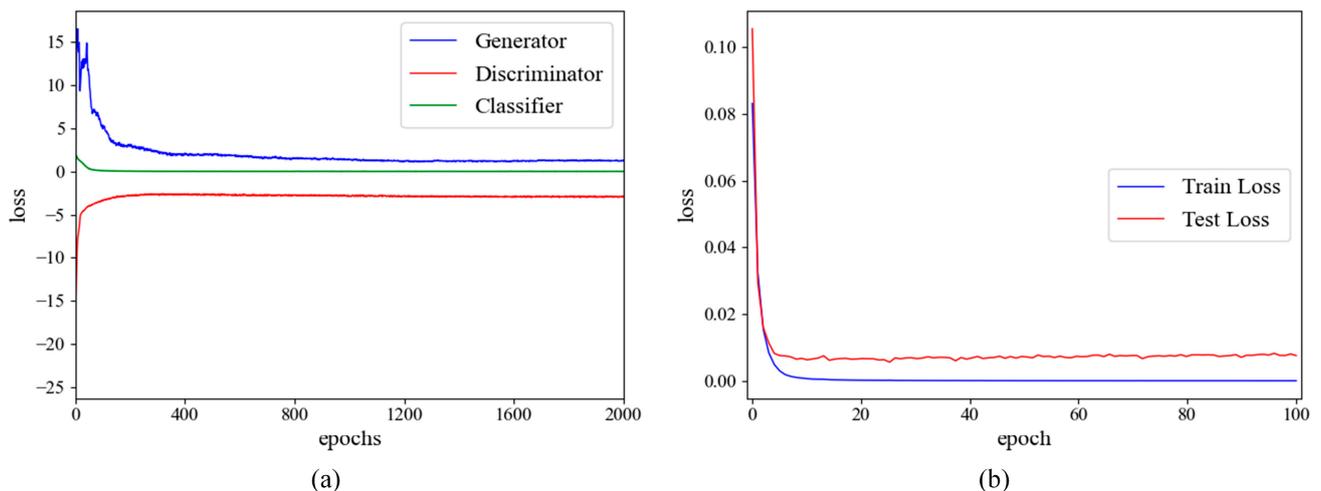
Figure 11. Loss curve of the model: (**a**) G, D, and C losses for MCBA-MVACGAN, (**b**) Training and Testing Losses of Classifier DCNN.

The DCNN network structure is shown in Table 3. To determine the training epochs of DCNN, the training and test losses with epochs are plotted in Figure 11b. To ensure that

the DCNN classifier is trained sufficiently, the epoch training and learning rates were 100 and 0.0001, respectively.

Table 3. Structure of DCNN.

Layers	Convolution Kernel Size/Step	Output Size
Convolution	$4 \times 4/2$	$32@32 \times 32$
Convolution	$4 \times 4/2$	$64@16 \times 16$
Convolution	$4 \times 4/2$	$64@8 \times 8$
Convolution	$4 \times 4/2$	$64@4 \times 4$
Flatten	/	1024
Linear	/	10

Under the guidance of reference [31], the gradient penalty factor λ is set to 10. In order to correctly determine the optimal hyperparameters of the minimum variance factor λ_1, the hyperparameter experiment results of the SQI dataset were conducted in this paper under the condition that 20 real samples and 200 fake samples were used as training sets, as shown in Figure 12. The change in hyperparameters has a great influence on the accuracy of fault diagnosis of the model, but when λ_1 is too large, the accuracy of fault diagnosis of the model will be reduced. As can be seen from Figure 12, when $\lambda_1 = 0.3$, the fault diagnosis accuracy of the proposed method is the highest, reaching 98.25%.
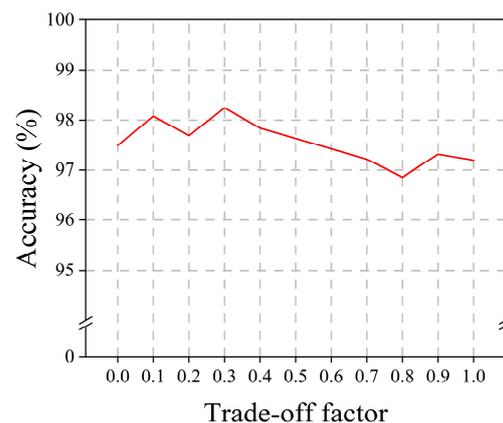

Figure 12. MCBA-MVACGAN performance with trade-off factors λ_1.

4.4. Effect of Real Training Sample Number on Diagnostic Accuracy

In this Section, four small sample tests with 5, 10, 15, and 20 real training samples were used to study the influence of different numbers of real training samples on fault diagnosis accuracy. Among them, the training set consisted of real training samples and 200 generated fake samples, and the test set consisted of 100 independent real samples. To account for randomness, each experiment was conducted 10 times, and the average calculated. The experimental results are shown in Figure 13.

For dataset A, the experimental results were 97.15%, 97.92%, 98.91%, and 99.78%, respectively. When the number of real training samples was reduced from 20 to 5, the fault diagnosis accuracy rate was reduced from 99.78% to 97.15%, a decrease of 2.63%. For dataset B, the experimental results are 95.47%, 96.18%, 96.84%, and 97.82%, respectively. When the number of real training samples is reduced from 20 to 5, the fault diagnosis accuracy rate is reduced from 97.82% to 95.47%, a decrease of 2.35%. The experimental analysis shows that the effective fault information learned by the proposed MCBA-MVACGAN can achieve the effect of data enhancement to a certain extent and realize the small sample fault diagnosis even when the data are seriously insufficient.
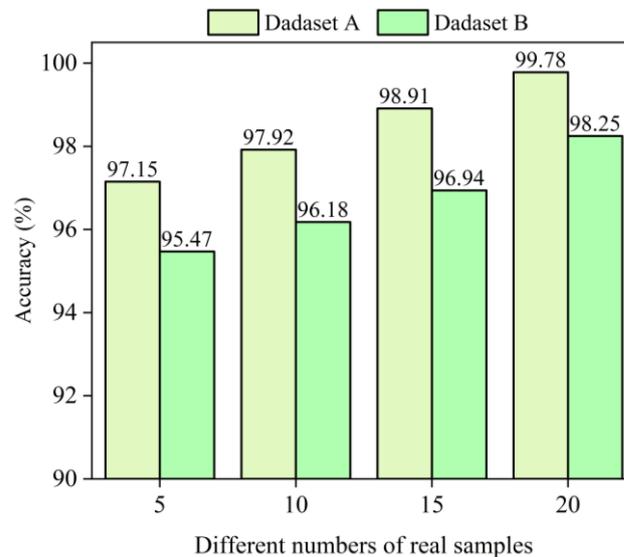
Figure 13. Influence of different numbers of real training samples on dataset A and dataset B.

4.5. Ablation Experiments

In this Section, the validity of each module is verified. ACWGAN-GP, independent classifier (IC), minimum variance (MV), and MCBA were used to conduct ablation experiments on dataset A and dataset B, respectively, and the comparison results are shown in Table 4.

Table 4. Results of ablation experiments on dataset A and dataset B.

Method	Name	Accuracy (%)	
		Dataset A	Dataset B
Method 1	ACWGAN-GP	97.62	95.69
Method 2	IC + ACWGAN-GP	98.57	96.38
Method 3	MV + ACWGAN-GP	98.43	96.51
Method 4	MV + IC + ACWGAN-GP	99.12	97.23
Method 5	MCBA + ACWGAN-GP	98.57	96.21
Method 6	MCBA + IC + ACWGAN-GP	99.31	97.49
Proposed Method	**MCBA + MV + IC + ACWGAN-GP**	**99.78**	**98.25**

Experimental analysis shows that the introduction of an independent classifier, minimum variance loss, and MCBA block can effectively improve the performance of the model in the case of small samples. Compared to method 4 without MCBA, the diagnostic accuracy is increased by 0.66% and 1.02%, respectively. Therefore, the introduction of MCBA can extract fault information more effectively. Compared to method 6, the accuracy is improved by 0.47% and 0.76%, respectively. Therefore, the minimum variance term can effectively alleviate the problem of model collapse. Methods 1 and 2 prove that the introduction of independent classifiers can improve the quality of generated samples and improve the accuracy of model diagnosis.

4.6. Comparative Experiments

4.6.1. Comparison of Model Diagnostic Performance

Three classification methods, SVM [43], BPNN [44], and CNN [11], were used in this experiment, and six enhancement methods, GAN [25], DCGAN [45], WCGAN [46], WGAN-GP [47], DFEGAN [48], and CBAM-CRLSGAN [33] were used as comparison methods. The experimental data of the proposed method are compared. All the data enhancement methods use DCNN as the classifier and add 200 fake training samples.

Considering randomness, all methods were tested 10 times, respectively, and the average value was calculated. Experimental analysis results are shown in Tables 5 and 6.

Table 5. Comparison results of fault diagnosis accuracy on dataset A.

Method	Health State										
	Normal	Ball7	IR7	OR7	Ball14	IR14	OR14	Ball21	IR21	OR21	Average
SVM	91.33	92.00	91.67	72.00	79.25	62.84	91.15	91.33	65.74	66.69	80.50
BPNN	93.25	95.75	94.00	67.17	84.39	63.25	91.50	91.75	91.33	71.78	84.42
CNN	95.80	96.25	96.20	91.33	73.75	68.67	97.00	94.60	94.37	68.25	87.62
GAN	98.20	99.40	99.00	96.60	93.75	81.17	99.20	98.62	93.75	79.43	93.91
DCGAN	99.40	98.6	98.75	99.60	84.73	82.24	99.20	99.20	95.25	85.23	94.22
WCGAN	100.00	99.40	100.00	99.60	96.75	86.00	93.25	99.60	95.00	94.00	96.36
WGAN-GP	100.00	100.00	100.00	100.00	88.85	88.75	99.60	100.00	98.20	95.20	97.06
DFEGAN	100.00	100.00	100.00	100.00	95.60	94.20	100.00	100.00	98.20	96.20	98.42
CBAM-CRLSGAN	100.00	100.00	100.00	100.00	97.80	96.60	100.00	100.00	98.60	98.80	99.18
Proposed Method	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**98.60**	**100.00**	**100.00**	**99.20**	**100.00**	**99.78**

Table 6. Comparison results of fault diagnosis accuracy on dataset B.

Method	Health State				
	NS	BF	IF	OF	Average
SVM	71.68	69.75	54.84	92.37	72.16
BPNN	75.37	78.74	57.67	91.46	75.81
CNN	82.37	81.83	58.57	93.75	79.13
GAN	93.25	75.16	80.52	95.75	86.17
DCGAN	89.77	86.82	88.21	98.20	90.75
WCGAN	95.40	88.75	87.13	97.25	92.13
WGAN-GP	96.95	90.25	88.60	98.20	93.50
DFEGAN	96.25	95.15	93.25	98.40	95.75
CBAM-CRLSGAN	98.00	95.75	96.25	99.00	97.25
Proposed Method	**98.25**	**98.75**	**97.00**	**99.00**	**98.25**

To better show the effect of the performance comparison of various methods, Figures 14 and 15 visualize the fault diagnosis accuracy of the nine comparison methods and the proposed methods on dataset A and dataset B. Obviously, due to insufficient model training, the problem of small samples greatly affects the high diagnostic accuracy obtained by conventional methods, and the diagnostic effect of SVM, DPNN, and CNN is not ideal. The average accuracy of this method is 99.80% in dataset A and 98.25% in dataset B. The experimental results show that the performance of the proposed method is significantly better than that of the other nine methods, and the fault diagnosis accuracy can be effectively improved even in the case of less data.

To visualize the ability of different methods to extract effective features, t-SNE [49] was used to simplify the feature vector output of the last layer network work into two dimensions. Figure 16 is a visualization of the results on dataset A. In the figure, IR14 overlaps OR21 in all but the proposed method, DFEGAN, and CBAM-CRLSGAN. Compared with DFEGAN and CBAM-CRLSGAN, the proposed method has a better clustering effect on IR14 and OR21, and the overlap between Ball14 and Ball21 is minimal.

In order to more accurately compare the diagnostic capabilities of all methods, accuracy rate, recall rate, and F1-score [50] were introduced for quantitative analysis. The expression is as follows:

$$\begin{cases} Precion = TP/(TP + FP) \\ Recall = TP/(TP + FN) \\ F1 - score = \frac{(\beta^2 + 1) * Precision * Recall}{\beta^2 * Precision + Recall} \end{cases} \tag{20}$$

where *TP*, *TN*, *FP*, and *FN* are true positive, true negative, false positive, and false negative, respectively. To ensure an even F-score, *β* is set to one.

In order to visually reflect the quantitative analysis results of different methods in dataset B, Figure 17 visualizes using the confusion matrix. The diagonal section shows the correct classification of the sample size, and the non-diagonal section shows the sample size of misclassification. For SVM, BPNN, and CNN, it was clear that many samples were misclassified, with the IF being the most obvious, and the highest having only 57 correct predictions. Therefore, these three classification methods have great limitations when only a few real training samples are available. For all data enhancement methods, the correct predictions of BF and IF for GAN, DCGAN, and WCGAN were only 75 and 80 times, 89 and 86 times, and 88 and 87 times, which is unacceptable. Although WGAN-GP has 100 correct predictions, OF, BF, and IF have only 90 and 88 correct predictions. The correct prediction times of DFEGAN were 96, 95, 95, and 98 times, and the correct prediction times of CBAM-CRLSGAN were 99, 94, 96, and 100 times, respectively, and the correct prediction times of the method in this paper were 98, 99, 97, and 99 times, respectively. Obviously, compared with the advanced method DFEGAN, the proposed method can also be comprehensively superior to DFEGAN. Compared with the advanced method CBAM-CRLSGAN, the correct prediction times OF NS and OF are less than one time, but the correct prediction times of BF and IF are more than five times and one time respectively. Therefore, the comprehensive performance of the proposed method is stronger than that of CBAM-CRLSGAN, and it can achieve more accurate predictions. Table 7 shows the accuracy, recall, and F1-score of all comparison methods on dataset A and dataset B. The scores of the proposed method are 99.78%, 99.80%, 99.79%, 98.25%, 98.23%, and 98.24%, respectively, higher than those of the other nine methods, which verifies the superiority of the method in diagnostic performance.
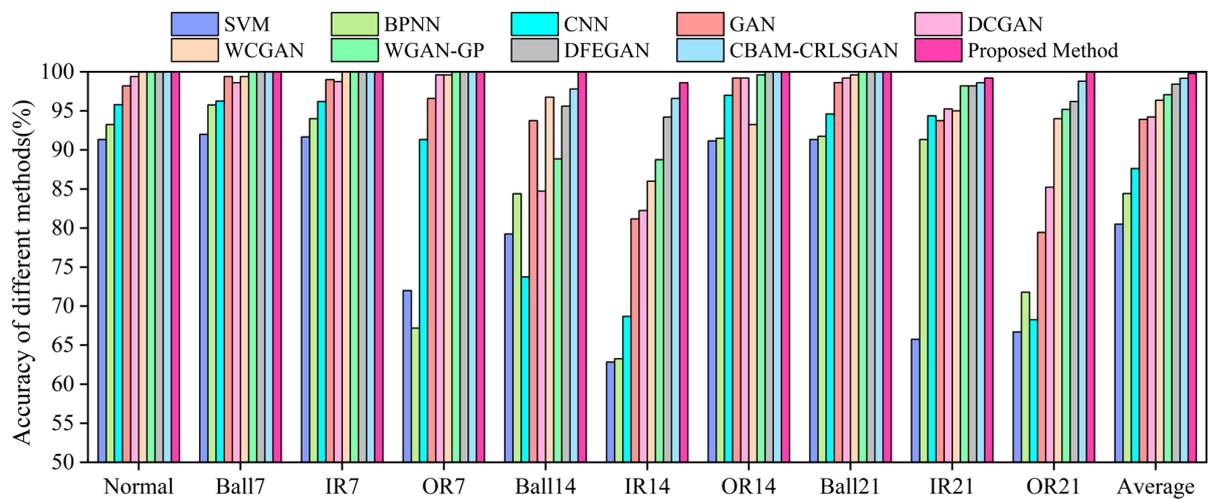

Figure 14. Comparative experimental results on dataset A.

Table 7. Average precision, recall rate, and F1-Score comparison of different methods.

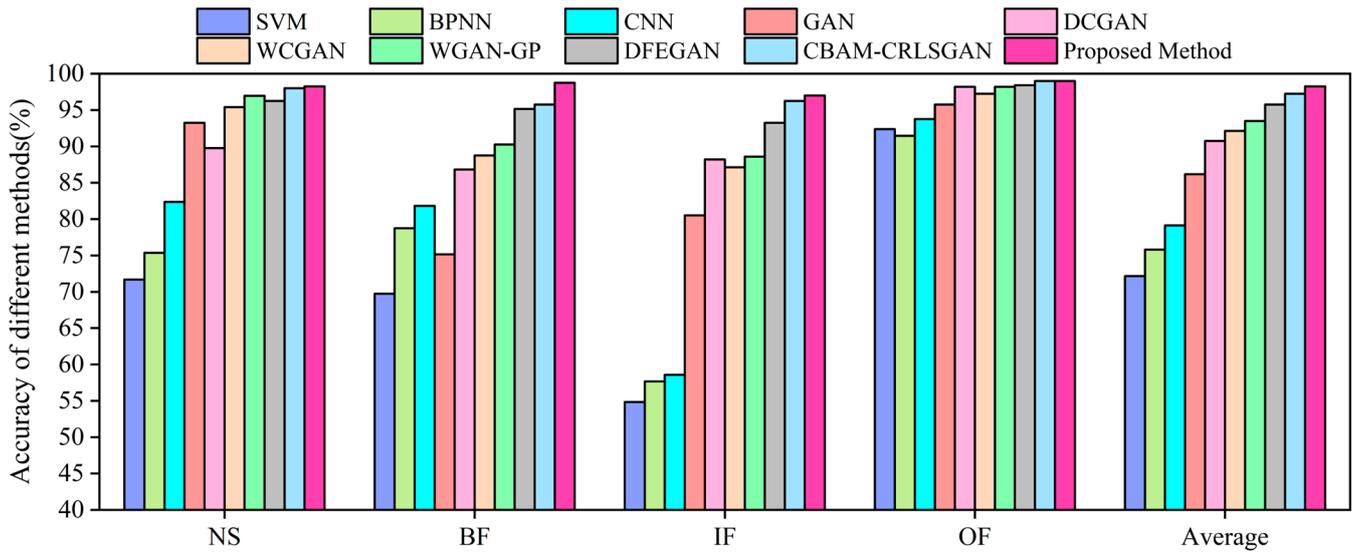| Method	Dataset A			Dataset B		
	Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
SVM	82.32	80.50	82.41	72.14	72.16	72.15
BPNN	87.00	84.30	85.63	75.41	75.81	75.61
CNN	87.70	87.60	87.65	79.12	79.05	79.08
GAN	95.07	93.90	94.48	85.87	86.02	85.94
DCGAN	94.72	94.22	94.47	90.71	90.69	90.70
WCGAN	96.57	96.30	96.43	92.18	92.00	92.09
WGAN-GP	97.36	97.12	97.26	93.45	93.50	93.47
DFEGAN	98.52	98.50	98.51	95.75	95.75	95.75
CBAM-CRLSGAN	99.18	99.20	99.19	97.25	97.31	97.28
Proposed Method	**99.78**	**99.80**	**99.79**	**98.25**	**98.23**	**98.24**

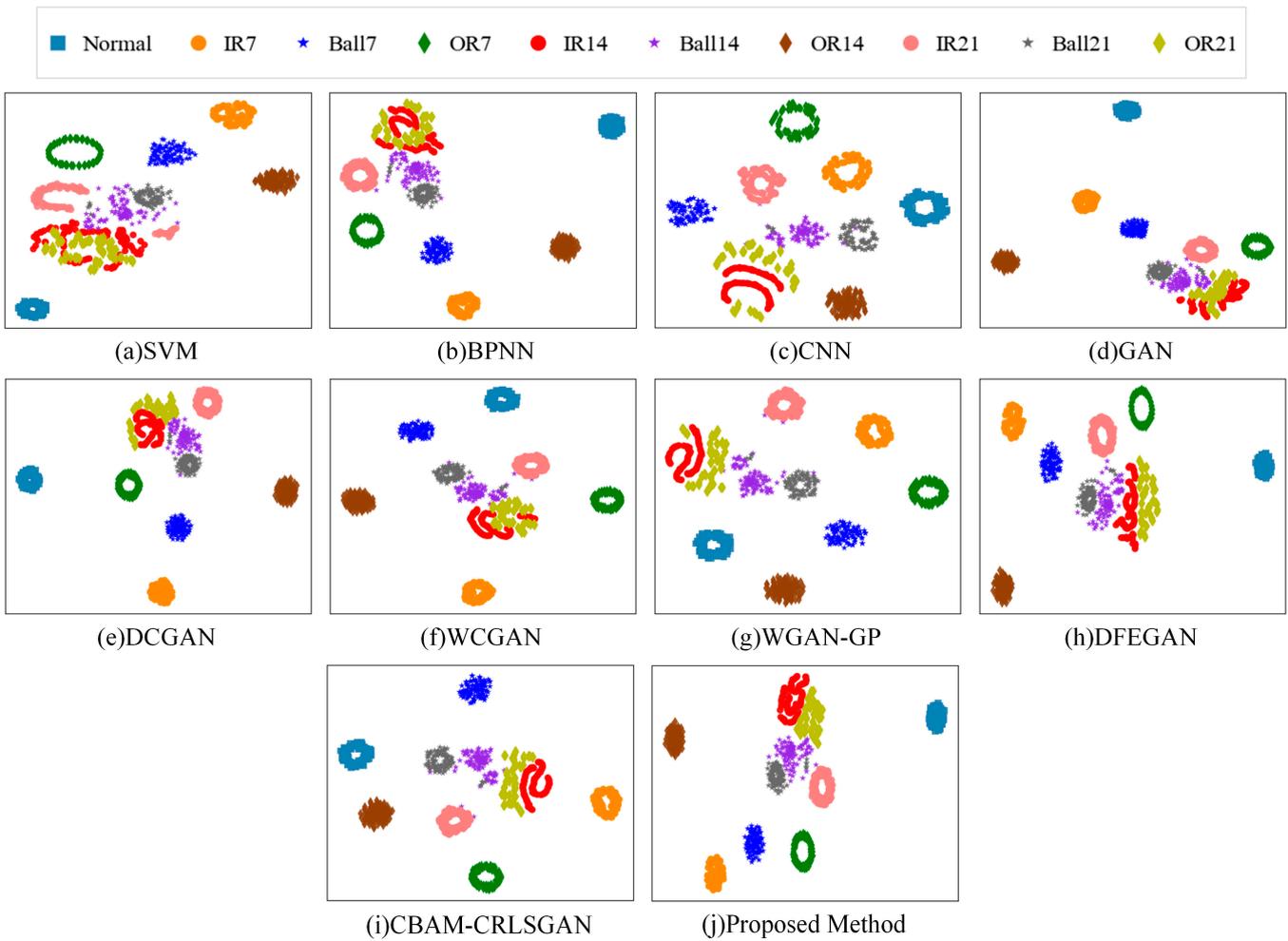**Figure 15.** Comparative experimental results on dataset B.

Figure 16. Visualization of fault diagnosis t_SNE by different methods on dataset A.
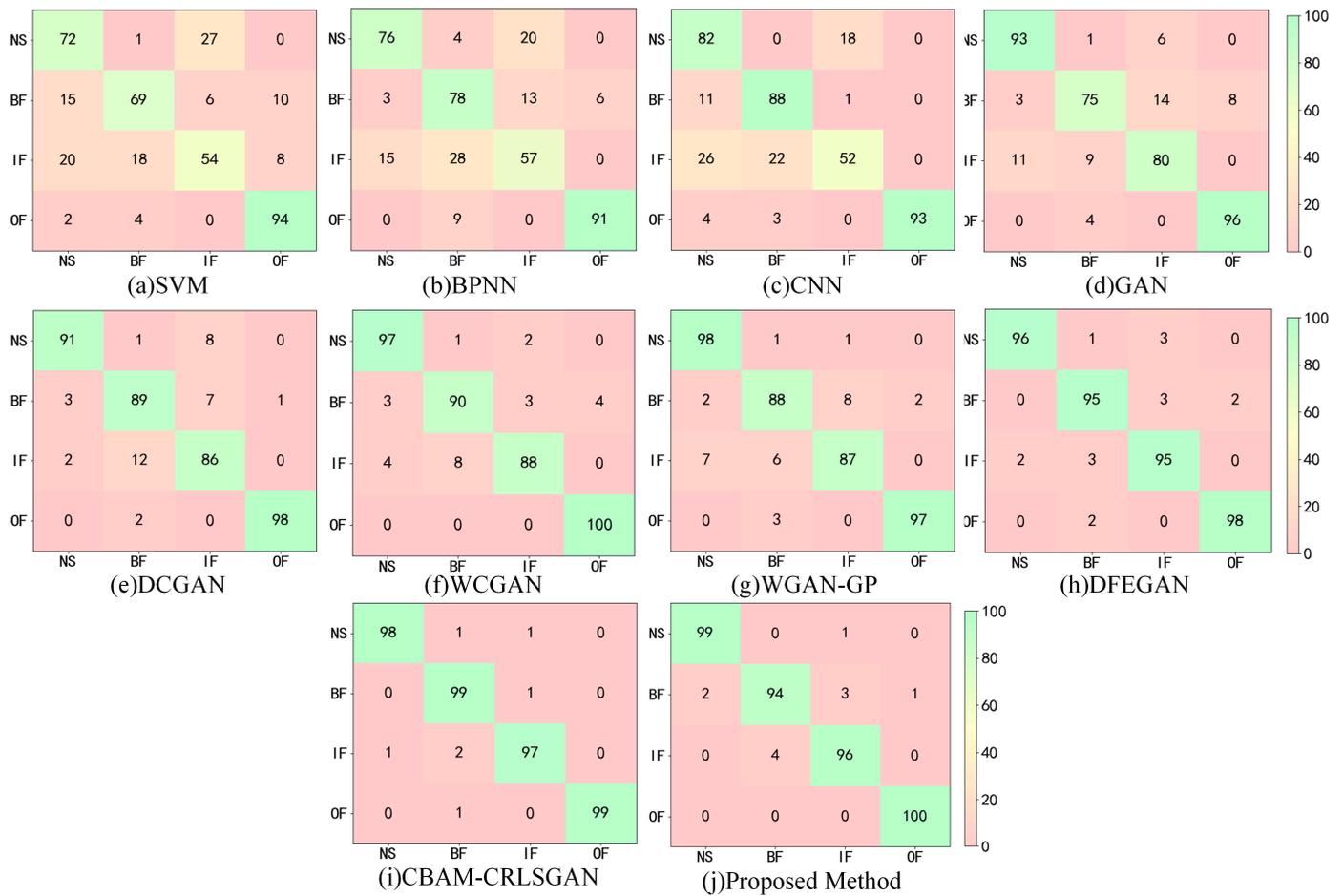
Figure 17. Fault diagnosis confusion matrix of different methods on dataset B.

4.6.2. Image Quality Evaluation

In order to visually show the similarity between true and false samples, this paper compares true and false samples of 10 health states in dataset A, as shown in Figure 18. A comparison of the true and false samples of four health states in dataset B is shown in Figure 19. The fake sample generated by MCBA-MVACGAN is highly similar to the real sample and can effectively generate important features of the real sample. Therefore, this method can realize effective fault diagnosis data enhancement.
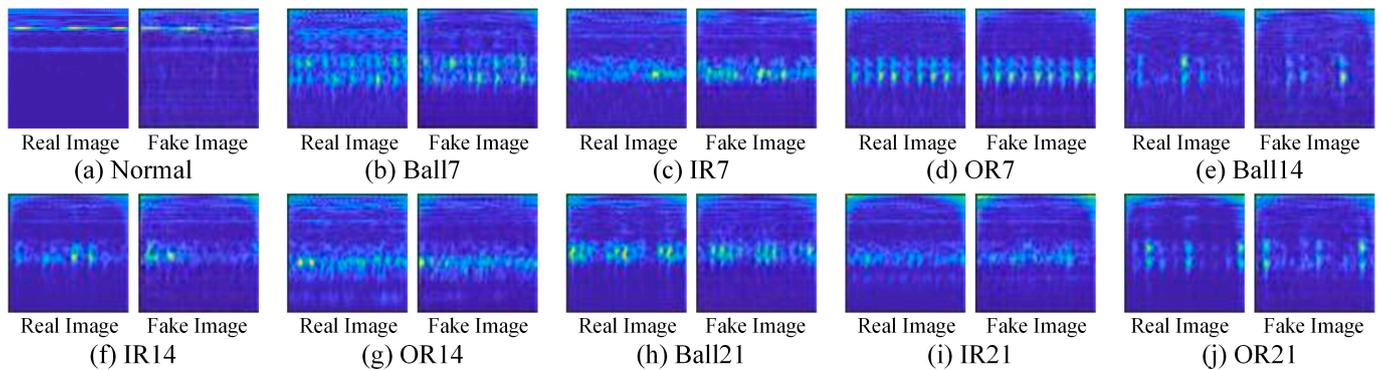

Figure 18. Comparison of the real sample with the generated fake sample for dataset A.

Real Image Fake Image   Real Image Fake Image   Real Image Fake Image   Real Image Fake Image

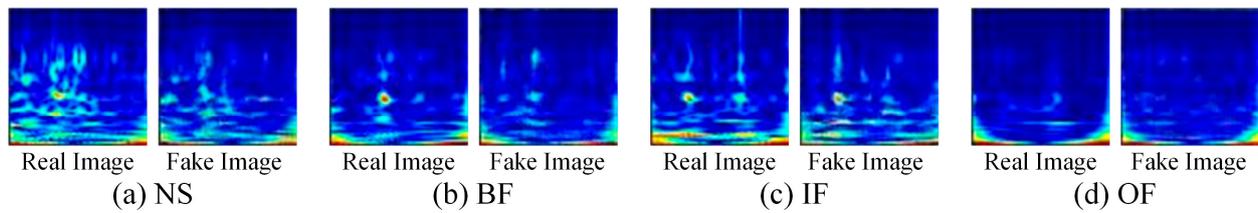(a) NS   (b) BF   (c) IF   (d) OF

Figure 19. Comparison of the real sample with the generated fake sample for dataset B.

To effectively evaluate the fake sample performance of different methods, the following methods are considered from a quantitative point of view: structural similarity (SSIM) [51], maximum mean difference (MMD) [35], and FID-Score [52]. SSIM ranges from 0 to 1. The higher the SSIM score, the better the fake sample performance. The lower the MMD value and FID-Score, the higher the quality of the fake image. The expressions of SSIM, MMD, and FID-Score are as follows:

$$SSIM(x,y) = \frac{\left((2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2) \right)}{\left(\left(\mu_x^2 + \mu_y^2 + c_1 \right)\left(\sigma_x^2 + \sigma_y^2 + c_2 \right) \right)} \tag{21}$$

$$MMD^2(x,y) = \frac{1}{m^2}\sum_{i,j} k(x_i, x_j) + \frac{1}{n^2}\sum_{i,j} k(y_i, y_j) - \frac{2}{mn}\sum_{i,j} k(x_i, y_j) \tag{22}$$

$$FID - Score = \left\| \mu_x - \mu_y \right\| + Tr\left(\sigma_x + \sigma_y - 2\sqrt{\sigma_x\sigma_y} \right) \tag{23}$$

where μ_x and μ_y, and σ_x^2 and σ_y^2 represent the mean and variance of real and fake samples, respectively, σ_{xy} are the covariances, and c_1 and c_2 are the constants. In Formula (22), x_i and x_j are samples taken from real images, y_i and y_j are samples taken from generated images, m and n are the sizes of the real and fake sample sets, and k is the kernel function. In Equation (23), σ_x and σ_y represent the variance of the real samples and the generated samples, respectively.

Table 8 shows the scores of FID-Score, SSIM, and FID for datasets A and B using different methods. In the table, the FID-Score of GAN is 98.28 and 141.15, respectively, indicating that the distribution of fake samples is far from the real samples. The FID-Score of the method in this paper is 28.97 and 48.14, respectively, indicating that the distribution of fake samples differs very little from that of real samples. In the SSIM index, the maximum values of 0.9834 and 0.9776 can be achieved by this method, indicating that the structural information of the fake samples is very similar to the real samples. In the MMD index, the proposed method can achieve the lowest 0.1122 and 0.1347, indicating that the distribution of fake samples differs very little from that of real samples. Experimental analysis shows that the proposed method can extract important features of real samples, retain the basic structure of real samples, generate fake samples with strong performance, and realize effective data enhancement.

Table 8. Generate fake sample quality comparisons.

Method	Dataset A			Dataset B		
	SSIM	MMD	FID-Score	SSIM	MMD	FID-Score
GAN	0.3374	0.7215	98.28	0.2714	0.7731	141.15
DCGAN	0.6583	0.6427	82.37	0.5126	0.6814	88.33
WCGAN	0.8237	0.4295	68.52	0.7833	0.4075	70.52
WGAN-GP	0.8868	0.3344	60.19	0.8641	0.3742	65.11
DFEGAN	0.9693	0.1868	46.31	0.9513	0.1949	53.03
CBAM-CRLSGAN	0.9532	0.1523	34.40	0.9609	0.1638	49.82
Proposed Method	**0.9834**	**0.1122**	**28.97**	**0.9776**	**0.1347**	**48.14**

4.6.3. Noise Immunity Experiment

In actual industrial scenarios, data collection is easily interference by noise. Therefore, it is extremely significant to improve the anti-noise performance of the model. In this experiment, the signal-to-noise ratio (SNR) is introduced to verify the noise immunity performance of different methods, as defined as follows:

$$SNR = 10\log\left(\sum_{i=1}^{N} s_i^2 / \sum_{i=1}^{N} n_i^2\right) \tag{24}$$

where s_i and n_i respectively represent the i decimal point of the noiseless signal s, and random noise n.

On the basis of the original data, random noise with SNR of 0 dB, 2 dB, 4 dB, 6 dB, 8 dB, and 10 dB was added, and the experiment was repeated 10 times under different SNR scenarios. The diagnostic accuracy rates of various comparison methods under different SNRs are shown in Figure 20 and Table 9. The experimental results show that with the increase in the signal-to-noise ratio, the diagnostic accuracy of various methods is also improved. For datasets A and B, when the SNR is 0 dB, 2 dB, 4 dB, 6 dB, 8 dB, and 10 dB, the accuracy of the proposed method is 94.31% and 92.50%, 95.83% and 94.16%, 97.31% and 95.82%, 98.64% and 97.33%, 99.54% and 98.14%, and 100% and 98.56%, respectively. Even in the case of a strong noise level (signal-to-noise ratio = 0 dB), the method can still obtain 94.31% and 92.50% satisfactory diagnostic accuracy. Under the condition of low noise level (signal-to-noise ratio of 10 dB), the diagnostic accuracy of this method reaches 100% and 98.56%. In addition, regardless of the noise level, the method achieves the highest diagnostic accuracy compared to other methods. The experimental results show that the method has good anti-noise performance.
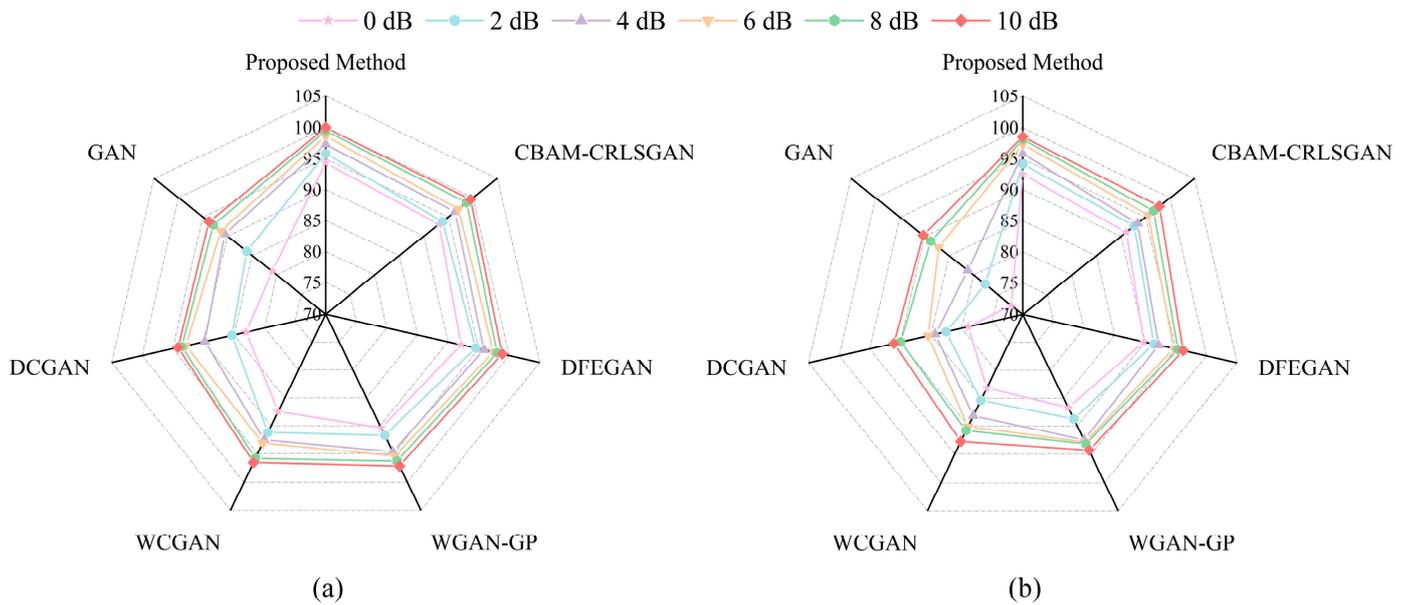

Figure 20. Diagnostic accuracies of various comparison methods under different SNRs: (**a**) CWRU, (**b**) SQI.

Table 9. Diagnostic accuracies of various comparison methods under different SNRs.

Dataset	Method	SNE (dB)					
		0 dB	2 dB	4 dB	6 dB	8 dB	10 dB
Dataset A	GAN	80.89	86.09	90.56	91.37	92.90	93.77
	DCGAN	83.01	85.32	89.75	92.84	93.56	94.13
	WCGAN	87.45	91.20	92.53	93.12	95.80	96.52
	WGAN-GP	90.39	91.73	94.66	95.41	96.26	97.12
	DFEGAN	92.15	94.52	95.88	97.41	98.04	98.86
	CBAM-CRLSGAN	93.07	93.81	96.35	97.04	98.70	99.52
	Proposed Method	**94.31**	**95.83**	**97.31**	**98.64**	**99.54**	**100**
Dataset B	GAN	72.35	77.66	81.20	87.14	88.75	90.28
	DCGAN	78.92	82.53	84.31	85.46	89.93	91.04
	WCGAN	83.18	85.37	88.15	90.12	90.81	92.76
	WGAN-GP	86.69	88.74	92.49	92.86	93.18	94.33
	DFEGAN	89.79	91.41	92.18	94.65	95.34	96.19
	CBAM-CRLSGAN	91.06	92.69	93.37	95.53	96.63	97.81
	Proposed Method	**92.50**	**94.16**	**95.82**	**97.33**	**98.14**	**98.56**

5. Conclusions

A new MCBA-MVACGAN algorithm is proposed to solve the problem of the lack of high-quality fault data of rotating machinery. The method is divided into three stages: data processing, data enhancement, and fault intelligent diagnosis. The proposed MCBA-MVACGAN can overcome the mode collapse problem of ACWGAN-GP, improve the quality and feature differentiation of generated fake samples, achieve effective data enhancement, and greatly improve the accuracy of small sample fault diagnosis. Compared with other GAN-based methods, this method can achieve higher fault diagnosis accuracy. The main summary is as follows:

1. A minimum variance term is designed, and an independent classifier is added to establish the MVACGAN model. The objective function of the model is reconstructed, which alleviates the problem of mode collapse in the classical ACWGAN-GP model.

2. A multi-scale convolutional block attention mechanism is designed to achieve multi-scale information extraction, attention weight learning, and multi-scale information weighting fusion, so as to capture multi-scale feature information and improve the diversity of generated samples. By integrating these advantages into MVACGAN, MCBA can effectively improve the model's ability to generate high-quality fake samples.

3. In the small sample scenario, MCBA-MVACGAN is significantly superior to other GAN-based data enhancement methods in considering the diagnostic performance of quantitative analysis and the quality of fault samples generated.

Although the MCBA-MVACGAN method proposed in this paper achieves satisfactory results in the fault diagnosis of bearing and gearbox datasets in small sample scenarios, it still has the problem of long training time. However, the lightweight of models is very important for industrial applications, and for the field of small sample fault diagnosis, it is a challenge. Therefore, we will consider the lightweight and interpretability of the model in the future. In addition, we plan to apply MCBA-MVACGAN to industrial equipment to address the scarcity of valid fault data in complex industrial environments.

Author Contributions: W.H. is responsible for methodology, validation, writing—original draft, writing—review and editing, visualization. X.Z. is responsible for supervision and project management. H.J. is responsible for project management and fund acquisition. Z.S. is responsible for writing—reviewing and editing. Y.B. is responsible for writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All the data are shown in the tables and figures of this paper.

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

References

1. Zhu, Z.; Lei, Y.; Qi, G.; Chai, Y.; Mazur, N.; An, Y.; Huang, X. A review of the application of deep learning in intelligent fault diagnosis of rotating machinery. *Measurement* **2023**, *206*, 112346. [CrossRef]
2. Zhang, K.; Tang, B.; Deng, L.; Liu, X. A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox. *Measurement* **2021**, *179*, 109491. [CrossRef]
3. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15. [CrossRef]
4. Feng, Z.; Liang, M.; Chu, F. Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mech. Syst. Signal Process.* **2013**, *38*, 165–205. [CrossRef]
5. Flandrin, P.; Rilling, G.; Goncalves, P. Empirical Mode Decomposition as a Filter Bank. *IEEE Signal Process. Lett.* **2004**, *11*, 112–114. [CrossRef]
6. Zhang, J.; Zhang, Q.; Qin, X.; Sun, Y. A two-stage fault diagnosis methodology for rotating machinery combining optimized support vector data description and optimized support vector machine. *Measurement* **2022**, *200*, 111651. [CrossRef]
7. 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]
8. Unal, M.; Onat, M.; Demetgul, M.; Kucuk, H. Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network. *Measurement* **2014**, *58*, 187–196. [CrossRef]
9. Liu, H.; Zhang, J.; Cheng, Y.; Lu, C. Fault diagnosis of gearbox using empirical mode decomposition and multi-fractal detrended cross-correlation analysis. *J. Sound Vib.* **2016**, *385*, 350–371. [CrossRef]
10. Zhu, H.; Cheng, J.; Zhang, C.; Wu, J.; Shao, X. Stacked pruning sparse denoising autoencoder based intelligent fault diagnosis of rolling bearings. *Appl. Soft Comput.* **2020**, *88*, 106060. [CrossRef]
11. Wang, J.; Mo, Z.; Zhang, H.; Miao, Q. A Deep Learning Method for Bearing Fault Diagnosis Based on Time-Frequency Image. *IEEE Access* **2019**, *7*, 42373–42383. [CrossRef]
12. Li, Y.; Zou, L.; Jiang, L.; Zhou, X. Fault Diagnosis of Rotating Machinery Based on Combination of Deep Belief Network and One-dimensional Convolutional Neural Network. *IEEE Access* **2019**, *7*, 165710–165723. [CrossRef]
13. Chen, L.; Xu, G.; Zhang, S.; Yan, W.; Wu, Q. Health indicator construction of machinery based on end-to-end trainable convolution recurrent neural networks. *J. Manuf. Syst.* **2020**, *54*, 1–11. [CrossRef]
14. Xu, Y.; Yan, X.; Feng, K.; Sheng, X.; Sun, B.; Liu, Z. Attention-based multiscale denoising residual convolutional neural networks for fault diagnosis of rotating machinery. *Reliab. Eng. Syst. Saf.* **2022**, *226*, 108716. [CrossRef]
15. Zhou, J.; Zhang, X.; Jiang, H.; Shao, Z.; Ma, B.; Zhou, R. MC-WDWCNN: An interpretable multi-channel wide-kernel wavelet convolutional neural network for strong noise-robust fault diagnosis. *Meas. Sci. Technol.* **2024**, *35*, 096125. [CrossRef]
16. Ye, X.; Yan, J.; Wang, Y.; Lu, L.; He, R. A novel capsule convolutional neural network with attention mechanism for high-voltage circuit breaker fault diagnosis. *Electr. Power Syst. Res.* **2022**, *209*, 108003. [CrossRef]
17. Wang, C.; Tian, B.; Yang, J.; Jie, H.; Chang, Y.; Zhao, Z. Neural-transformer: A brain-inspired lightweight mechanical fault diagnosis method under noise. *Reliab. Eng. Syst. Saf.* **2024**, *251*, 110409. [CrossRef]
18. Zhou, J.; Zhang, X.; Jiang, H.; Li, J.; Shao, Z. Cross-domain transfer fault diagnosis by class-imbalanced deep subdomain adaptive network. *Measurement* **2025**, *242*, 115901. [CrossRef]
19. Wang, C.; Yang, J.; Jie, H.; Tao, Z.; Zhao, Z. A lightweight progressive joint transfer ensemble network inspired by the markov process for imbalanced mechanical fault diagnosis. *Mech. Syst. Signal Process.* **2025**, *224*, 111994. [CrossRef]
20. Mao, W.; Feng, W.; Liu, Y.; Zhang, D.; Liang, X. A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. *Mech. Syst. Signal Process.* **2021**, *150*, 107233. [CrossRef]
21. Li, Y.; Wang, Y.; Zhang, Y.; Zhang, J. Diagnosis of Inter-turn Short Circuit of Permanent Magnet Synchronous Motor Based on Deep learning and Small Fault Samples. *Neurocomputing* **2021**, *442*, 348–358. [CrossRef]
22. Wang, Z.; Xuan, J. Intelligent fault recognition framework by using deep reinforcement learning with one dimension convolution and improved actor-critic algorithm. *Adv. Eng. Inform.* **2021**, *49*, 101315. [CrossRef]
23. Li, J.; Li, X.; He, D.; Qu, Y. Unsupervised rotating machinery fault diagnosis method based on integrated SAE–DBN and a binary processor. *J. Intell. Manuf.* **2020**, *31*, 1899–1916. [CrossRef]

24. Zhang, L.; Lin, J.; Shao, H.; Zhang, Z.; Yan, X.; Long, J. End-to-end unsupervised fault detection using a flow-based model. *Reliab. Eng. Syst. Saf.* **2021**, *215*, 107805. [CrossRef]

25. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]

26. Zhang, L.; Duan, L.; Hong, X.; Liu, X.; Zhang, X. Imbalanced data enhancement method based on improved DCGAN and its application. *J. Intell. Fuzzy Syst.* **2021**, *41*, 3485–3498. [CrossRef]

27. Liu, S.; Chen, J.; He, S.; Xu, E.; Lv, H.; Zhou, Z. Intelligent fault diagnosis under small sample size conditions via Bidirectional InfoMax GAN with unsupervised representation learning. *Knowl. Based Syst.* **2021**, *232*, 107488. [CrossRef]

28. Zhu, Q.-X.; Hou, K.-R.; Chen, Z.-S.; Gao, Z.-S.; Xu, Y.; He, Y.-L. Novel virtual sample generation using conditional GAN for developing soft sensor with small data. *Eng. Appl. Artif. Intell.* **2021**, *106*, 104497. [CrossRef]

29. Gu, X.; Yu, Y.; Guo, L.; Gao, H.; Luo, M. CSWGAN-GP: A new method for bearing fault diagnosis under imbalanced condition. *Measurement* **2023**, *217*, 113014. [CrossRef]

30. Zhang, T.; Chen, J.; Li, F.; Pan, T.; He, S. A Small Sample Focused Intelligent Fault Diagnosis Scheme of Machines via Multimodules Learning with Gradient Penalized Generative Adversarial Networks. *IEEE Trans. Ind. Electron.* **2021**, *68*, 10130–10141. [CrossRef]

31. Chen, J.; Yan, Z.; Lin, C.; Yao, B.; Ge, H. Aero-engine high speed bearing fault diagnosis for data imbalance: A sample enhanced diagnostic method based on pre-training WGAN-GP. *Measurement* **2023**, *213*, 112709. [CrossRef]

32. Fu, Z.; Liu, Z.; Ping, S.; Li, W.; Liu, J. TRA-ACGAN: A motor bearing fault diagnosis model based on an auxiliary classifier generative adversarial network and transformer network. *ISA Trans.* **2024**, *149*, 381–393. [CrossRef] [PubMed]

33. Zhang, J.; Kong, Y.; Chen, Z.; Han, T.; Han, Q.; Dong, M.; Chu, F. CBAM-CRLSGAN: A novel fault diagnosis method for planetary transmission systems under small samples scenarios. *Measurement* **2024**, *234*, 114795. [CrossRef]

34. Lee, M.; Seok, J. Controllable Generative Adversarial Network. *IEEE Access* **2019**, *7*, 28158–28169. [CrossRef]

35. Li, W.; Zhong, X.; Shao, H.; Cai, B.; Yang, X. Multi-mode data augmentation and fault diagnosis of rotating machinery using modified ACGAN designed with new framework. *Adv. Eng. Inform.* **2022**, *52*, 101552. [CrossRef]

36. Li, X.; Yue, C.; Liu, X.; Zhou, J.; Wang, L. ACWGAN-GP for milling tool breakage monitoring with imbalanced data. *Robot. Comput. Integr. Manuf.* **2024**, *85*, 102624. [CrossRef]

37. Wang, C.; Ma, J.; Jin, H.; Wang, G.; Chen, C.; Xia, Y.; Gou, J. ACGAN and BN based method for downhole incident diagnosis during the drilling process with small sample data size. *Ocean. Eng.* **2022**, *256*, 111516. [CrossRef]

38. Zhu, Y.; Pei, Y.; Wang, A.; Xie, B.; Qian, Z. A partial domain adaptation scheme based on weighted adversarial nets with improved CBAM for fault diagnosis of wind turbine gearbox. *Eng. Appl. Artif. Intell.* **2023**, *125*, 106674. [CrossRef]

39. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective kernel networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.

40. Azizzadenesheli, K.; Anandkumar, A. Efficient exploration through bayesian deep q-networks. In Proceedings of the 2018 Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 11–16 February 2018.

41. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64–65*, 100–131. [CrossRef]

42. Wei, H.; Zhang, Q.; Shang, M.; Gu, Y. Extreme learning Machine-based classifier for fault diagnosis of rotating Machinery using a residual network and continuous wavelet transform. *Measurement* **2021**, *183*, 109864. [CrossRef]

43. Wenyi, L.; Zhenfeng, W.; Jiguang, H.; Guangfeng, W. Wind turbine fault diagnosis method based on diagonal spectrum and clustering binary tree SVM. *Renew. Energy* **2013**, *50*, 1–6. [CrossRef]

44. Li, J.; Yao, X.; Wang, X.; Yu, Q.; Zhang, Y. Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis. *Measurement* **2020**, *153*, 107419. [CrossRef]

45. Viola, J.; Chen, Y.; Wang, J. FaultFace: Deep Convolutional Generative Adversarial Network (DCGAN) based Ball-Bearing failure detection method. *Inf. Sci.* **2021**, *542*, 195–211. [CrossRef]

46. Peng, Y.; Wang, Y.; Shao, Y. A novel bearing imbalance Fault-diagnosis method based on a Wasserstein conditional generative adversarial network. *Measurement* **2022**, *192*, 110924. [CrossRef]

47. Gao, X.; Deng, F.; Yue, X. Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty. *Neurocomputing* **2020**, *396*, 487–494. [CrossRef]

48. Liu, S.; Jiang, H.; Wu, Z.; Li, X. Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis. *Mech. Syst. Signal Process.* **2022**, *163*, 108139. [CrossRef]

49. Vander Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

50. Sokolova, M.; Japkowicz, N.; Szpakowicz, S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In *AI 2006: Advances in Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1015–1021.

51. Zhong, G.; Gao, W.; Liu, Y.; Yang, Y.; Wang, D.H.; Huang, K. Generative adversarial networks with decoder-encoder output noises. *Neural Netw.* **2020**, *127*, 19–28. [CrossRef]

52. Zhang, K.; Xu, G.; Han, Z.; Ma, K.; Zheng, X.; Chen, L.; Duan, N.; Zhang, S. Data Augmentation for Motor Imagery Signal Classification Based on a Hybrid Neural Network. *Sensors* **2020**, *20*, 4485. [CrossRef] [PubMed]