lubricants

MDPI

Article

Few-Shot Fault Diagnosis of Rolling Bearings Using Generative Adversarial Networks and Convolutional Block Attention Mechanisms

Yong Chen [1,2,*], Xiangrun Pu [1,2], Guangxin Li [3], Yunhui Bai [4] and Lijie Hao [5]

[1] Guangxi Key Laboratory of New Energy Vehicle Power Battery and Green Powertrain Domain, Nanning 530004, China
[2] Engineering Research Center of New Energy Vehicle Advanced Powertrains, University of Guangxi, Nanning 530004, China
[3] Tianjin Key Laboratory of Power Transmission and Safety Technology for New Energy Vehicles, School of Mechanical Engineering, Hebei University of Technology, Tianjin 300130, China
[4] Aurobay (Ningbo) Intelligent Technology Co., Ltd., Ningbo 315336, China
[5] AECC Beijing Hangke Engine Control System Science & Technology Co., Ltd., Beijing 102200, China
* Correspondence: chenyong1585811@163.com

Abstract

In modern industrial systems, diagnosing faults in the rolling bearings of high-speed rotating machinery remains a considerable challenge due to the scarcity of reliable fault samples and the inherent complexity of the diagnostic task. To address these limitations, this study proposes an intelligent fault diagnosis method that integrates a generative adversarial network (GAN) with a convolutional block attention mechanism (CBAM). First, after systematically evaluating several loss functions, a GAN based on the Wasserstein distance loss function was adopted to generate high-quality synthetic vibration samples, effectively augmenting the training dataset. Subsequently, a convolutional block attention mechanism-based convolutional neural network (CBAM-CNN) was developed. By adaptively emphasizing salient features through channel and spatial attention modules, the CBAM-CNN improves feature extraction and recognition performance under limited-sample conditions. To validate the proposed method, an experimental platform for a two-speed automatic mechanical transmission (2AMT) of an electric vehicle was developed, and diagnostic experiments were conducted on high-speed rolling bearings. The results indicate that, under extremely severe conditions, CBAM-CNN achieves a diagnostic accuracy of 96.64% for rolling element pitting defects using only 10% of authentic samples. For composite faults, the model maintains an average accuracy above 97%, demonstrating strong generalization capability. These findings provide solid theoretical support and practical engineering guidance for rolling bearing fault diagnosis under few-shot conditions.

Keywords: few-shot learning; generative adversarial networks (GAN); convolutional block attention module (CBAM); fault diagnosis; rolling bearing

check for updates

Received: 15 October 2025
Revised: 21 November 2025
Accepted: 24 November 2025
Published: 25 November 2025

Citation: Chen, Y.; Pu, X.; Li, G.; Bai, Y.; Hao, L. Few-Shot Fault Diagnosis of Rolling Bearings Using Generative Adversarial Networks and Convolutional Block Attention Mechanisms. *Lubricants* **2025**, *13*, 515. https://doi.org/10.3390/lubricants13120515

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the advancement of high-speed rotating machinery toward the era of intelligent manufacturing, the condition monitoring and intelligent fault diagnosis of key transmission components, such as rolling bearings, have become essential techniques for ensuring the reliable operation of transmission systems [1]. In modern industrial applications such as aeroengines, new energy vehicles, and wind turbines, the reliability of rolling bearings

critically influences the overall performance and operational safety of mechanical systems. Within these critical transmission systems, severe bearing failures may lead to unplanned downtime, substantial maintenance costs, and even serious safety incidents [2].

In high-speed transmission equipment featuring adequate lubrication and high machining precision, rolling bearings gradually develop surface contact fatigue damage with prolonged operation. This damage accumulation propagates as abnormal vibration and noise, leading to progressive performance degradation and eventual failure [3]. Therefore, accurate monitoring of the real-time status of rolling bearings is essential for timely maintenance and risk prevention [4]. By analyzing real-time sensor data, intelligent fault diagnosis can assess the operational status of bearings and issue early warnings for necessary maintenance or component replacement. This capability significantly enhances industrial equipment utilization efficiency and reduces maintenance costs, leading to its widespread application in industrial health monitoring systems [5].

Fault data in rotating machinery are often characterized by high uncertainty, data imbalance, difficulty in data acquisition, and complex multi-source signals, which represent a typical few-shot problem. Consequently, there is a pressing need for intelligent, efficient, and reliable fault diagnosis models capable of operating with few-shot training. It is within this engineering context that deep learning-based intelligent diagnostics have garnered attention, due to their capabilities in addressing such challenges [6]. Currently, fault diagnosis models for rotating mechanical components can be generalized into model-based and data-driven methods [7]. The model-based fault diagnosis approaches typically require substantial prior knowledge to develop accurate fault mechanistic models. For instance, Bera et al. [8] proposed an adaptive model-based method to evaluate imbalance faults in rotor-bearing systems. Zhao et al. [9] addressed the limited-data problem in bearing diagnosis by augmenting and reconstructing the dataset, employing a Support Vector Machine (SVM) for fault classification. Although their model achieved high accuracy, its diagnostic capability under non-stationary conditions remains limited. Due to the structural complexity, highly variable operating conditions, and strong environmental disturbances in rotating machinery, traditional model-based methods often fail to meet the practical diagnostic requirements [10].

In contrast, data-driven fault diagnosis techniques, which require minimal reliance on precise analytical models or expert knowledge, have emerged as the predominant trend in the intelligent development of modern rotating machinery diagnosis [11].

The data-driven approaches, particularly deep learning and reinforcement learning, have been extensively applied to bearing fault diagnosis. Prominent techniques include convolutional neural networks (CNNs) [12], generative adversarial networks (GANs) [13], transfer learning (TL) [14], and manifold learning [15]. These methods excel at capturing critical information from datasets and establishing mappings between raw sensory data and the discriminative features.

CNN models have been extensively utilized in the domain of fault diagnostics of high-speed rotating machinery, owing to their local connectivity and weight-sharing properties [16]. Ongoing research has focused on enhancing CNN architectures to enable direct input of raw vibration signals for end-to-end intelligent diagnosis. Zhang et al. [17] developed a deep CNN with wide convolutional kernels based on domain adaptation principles, achieving robust fault diagnosis under varying operation conditions using raw bearing vibration signals. Similarly, Jiang et al. [18] proposed an intelligent diagnosis model that systematically extracts discriminative fault characteristics from raw vibration of wind turbine gearboxes. Meanwhile, the attention mechanisms have been incorporated into the latest CNNs, allowing models to emphasize the most informative features. Wang et al. [19] designed an attention-based multi-layer fusion CNN (AMFCNN) to construct a lightweight

multi-sensor gear fault diagnosis model. Zhou et al. [20] proposed a frequency-domain attention CNN to mitigate the environmental noise interference. The data-driven fault diagnosis models inherently depend on a substantial quantity of labelled samples [21]. However, the operational conditions of high-speed rotating machinery are highly variable and complex, making it difficult to acquire sufficient bearing fault data [22]. Consequently, relying solely on CNN models for fault diagnosis carries a high risk of overfitting. This data scarcity hinders the model's ability to generalize, compromising diagnostic accuracy and failing to meet the monitoring requirements.

To address the issue of insufficient sample size, the latest data-driven fault diagnosis models have been optimized at the algorithm and data levels, respectively [23]. A representative algorithmic method is the TL [24]. To address the challenge of data scarcity, Ai et al. [25] pioneered a TL framework based entirely on simulated data. Their approach employs a domain-invariant transformation technique to enable effective feature sharing between simulation and real-world domains. Wang et al. [26] proposed the rolling bearing fault diagnosis method based on a dynamic simulated model from the source domain to the target domain with improved alternating TL. However, the performance of TL strongly depends on the availability of sufficient source-domain samples. When the source data are limited, the model's generalization ability in the target domain is reduced [27].

The data-level approaches enhance training data diversity by transforming existing samples, but they must be tailored to the specific signal characteristics of the application scenario [28]. Among these, the GAN models are particularly effective, as they can generate new samples while learning the intrinsic data distribution. The GAN comprises a generator, which produces synthetic data, and a discriminator, which distinguishes between real and generated samples. These two components are trained adversarially until they reach a Nash equilibrium [9]. Huang et al. [29] employed an enhanced GAN to improve diagnostic accuracy for wind turbine gearboxes under actual in-service conditions. Miao et al. [30] proposed a data augmentation strategy based on an improved variational autoencoder GAN. Zhuang et al. [31] developed a two-stage feature extractor with residual attention mechanisms and corresponding GAN-based generative modules. Owing to the nonlinear and non-stationary nature of vibration signals, the generator often struggles to accurately capture the underlying distribution of raw data. Furthermore, the generator itself exhibits deficiencies in its capacity for feature extraction and instability in its model framework [32].

The few-shot diagnosis of rolling bearings in high-speed rotating machinery faces dual challenges: the degraded accuracy of a standalone CNN due to insufficient training samples and the limited feature representation capability of GAN. To address these challenges, this study introduces a novel few-shot intelligent diagnosis framework that synergistically integrates GAN with the convolutional block attention module (CBAM). The specifics of this framework are detailed as follows:

(1) The limited and imbalanced fault data were augmented using a GAN with the Wasserstein distance loss function, thereby generating a diverse training dataset with increased sample sizes for each bearing fault category to resolve the data imbalance problem.
(2) To address the challenge of few-shot fault diagnosis, a novel CNN architecture incorporating a convolutional block attention module (CBAM-CNN) was proposed. This model was designed to fuse convolutional operations with an attention mechanism, facilitating a focus on salient feature regions and the capture of both local and global dependencies.
(3) This study mainly focused on the pitting failure of rolling bearings in a typical high-speed rotating machine, specifically the two-speed automatic mechanical transmission (2AMT). Experimental results confirm that CBAM-CNN delivers high-precision fault identification and diagnosis even under few-shot conditions.

The proposed method synergistically combines the strengths of the CBAM and GAN models, thereby effectively compensating for their respective limitations. The novel method has the potential to reduce the reliance of fault diagnosis models on large-scale datasets, thereby enabling effective fault diagnosis of rolling bearings even with limited sample sizes.

2. Methodology

2.1. Fundamental Principles of GAN

The GAN is an unsupervised learning framework specifically designed to model complex data distributions. Its core architecture consists of a generator and a discriminator. This process allows the model to estimate the underlying distribution of input data and produce novel samples that are highly consistent with real observations.

A key advantage of GANs over traditional generative models is their freedom from restrictive prior assumptions regarding the data distribution, which facilitates efficient training and optimization. In practice, the generator creates synthetic samples through forward propagation, enabling the realistic simulation of experimental data [33]. Essentially, the training process of GANs is formulated as a minimax (zero-sum) game between the generator and the discriminator, as depicted in Figure 1.

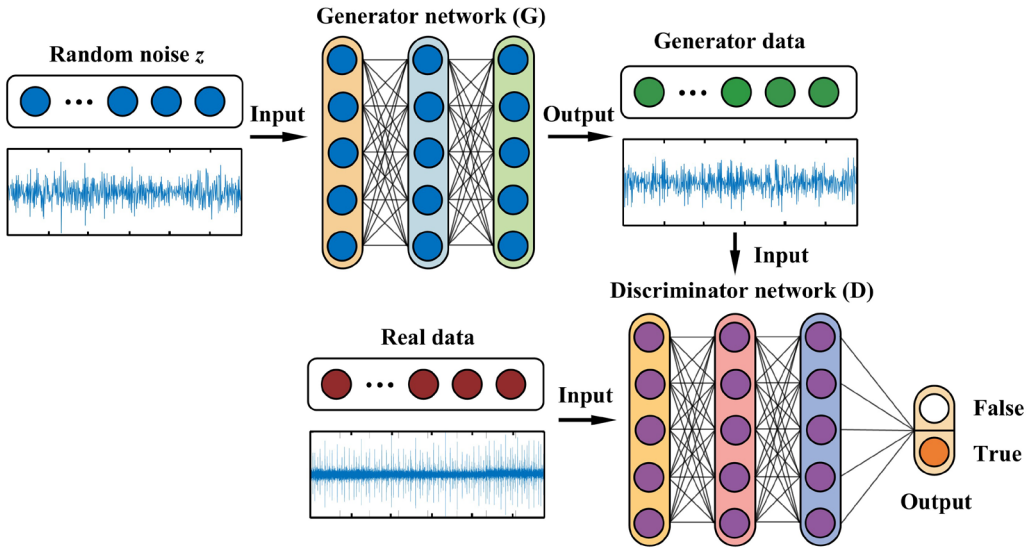

Figure 1. The principle of GAN.

The generator accepts random noise z sampled from a Gaussian distribution as input and transforms it into synthetic data. The discriminator, structured as a binary classification network, evaluates whether an input originates from real data or has been produced by the generator. When presented with authentic data, the discriminator outputs a probability close to 1; conversely, it approaches 0 for generated samples.

Throughout the training process, the generator refines its parameters based on feedback from the discriminator, progressively improving the realism of its outputs. Through this adversarial interaction and iterative parameter optimization, the generator gradually develops its capacity to produce high-quality data, while the discriminator simultaneously enhances its capacity to distinguish authenticity. When the discriminator and the generator achieve a Nash equilibrium, it indicates that the GAN has effectively captured the inherent distribution of the underlying data.

Therefore, GAN optimization may be defined as an extremum problem, with the objective function defined as follows:

$$\min_G \min_D V(D,G) = E_{x\sim P_{data(x)}}[log((D(x)))] + E_{Z\sim P_z(z)}[log1 - D(G(z))] \quad (1)$$

where $V(D,G)$ represents the divergence, which can be formulated as a cross-entropy loss function. $\log(D(x))$ represents the discriminator's assessment of authentic data. $\log(1-D(G(z)))$ reflects its evaluation of the generated data.

The training of GAN is typically performed through an alternating optimization strategy. Specifically, the generator is first fixed while the discriminator is optimized to maximize its objective function, with parameters updated using mini-batch gradient descent. The following is a concise summary of the process:

$$\min_D V(D,G) = -\nabla_{\theta_G} \frac{1}{m}\sum_{i=1}^{m} logD\left(x^{(i)}\right) - log\left[1 - D\left(G\left(z^{(i)}\right)\right)\right] \quad (2)$$

The optimization of the discriminator aims to increase the probability of correctly classifying real samples as "True", while simultaneously maximizing the probability of classifying generated samples as "False". Its loss function is typically computed based on the outputs of the discriminator for both real and generated data.

Subsequently, the discriminator is set and the generator is optimized to minimize the objective function of the discriminator, as expressed below:

$$\min_D V(D,G) = -\nabla_{\theta_G} \frac{1}{m}\sum_{i=1}^{m}\left[D\left(G\left(z^{(i)}\right)\right)\right] \quad (3)$$

The objective of the generator is to maximize the possibility that the discriminator will classify its produced samples as "True". Accordingly, the loss function of the generator is usually formulated in relation to the discriminator output.

In practical bearing monitoring applications, vibration signal data can be incomplete or contain gaps. The GAN can mitigate these issues by leveraging generative capabilities to reconstruct missing samples, thereby enabling the model to utilize a more complete dataset and preventing the performance degradation associated with incomplete data.

2.2. Principles of Attention Mechanisms

The attention mechanism is derived by simulating the human optical system. Its core idea is to enable the model to autonomously determine which regions of the input should receive more focus, thereby prioritizing the allocation of limited computational resources to the most informative parts. The attention mechanism has been shown to concentrate the model on critical information more precisely by adaptively weighting input features, thus leading to improved overall performance [34].

2.2.1. Channel Attention Mechanism

The channel attention mechanism (CAM) is designed to perform feature selection and weighting across channels. By applying global pooling operations to compress spatial information, the model can capture the global characteristics of each channel. Subsequently, deep learning algorithms are employed to explore inter-channel dependencies and adaptively adjust the weights of individual channels.

Based on the learned channel weights, the original input features are reweighted: channels that contribute significantly to the diagnostic task are enhanced, while those of lesser relevance are suppressed [35].

The configuration of CAM is presented in Figure 2. First, spatial dimensionality reduction is performed through global max pooling and average pooling, producing two feature maps that focus on channel-specific information. Next, channel feature learning is conducted by feeding the two feature maps into shared convolutional layers to perform nonlinear transformations, thereby generating importance weights for each channel. These weights are then normalized using a Sigmoid function to produce attention weights that reflect the importance of each channel. Finally, the original features are recalibrated according to attention weights, highlighting task-relevant features.

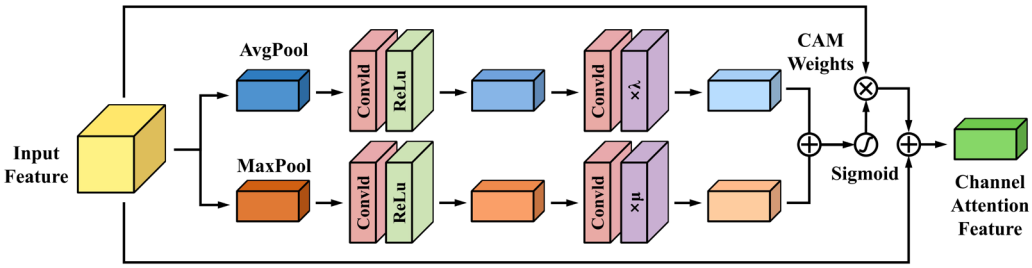

Figure 2. CAM schematic diagram.

The CAM can be expressed as follows:

$$\omega_1 = \delta_2(\lambda \times F_2(\delta_1(F_1(Avepool(x))))) + \mu \times F_2(\delta_1(F_1(Maxpool(x)))) \quad (4)$$

where δ_1 denotes the ReLU function and δ_2 denotes the Sigmoid function, which generates a vector in the range [0, 1]. λ and μ denote the average and max pooling weights, respectively. These are learnable parameters that can be adapted through training. F_1 and F_2 denote 1×1 convolutions with one channel.

2.2.2. Spatial Attention Mechanism

The spatial attention mechanism (SAM) is inspired by the human visual system's ability to dynamically focus on salient regions within a scene. By adaptively generating spatial attention weights based on input features, SAM enables the model to emphasize informative spatial areas while suppressing less relevant ones [36]. Figure 3 illustrates the architecture of SAM. Specifically, max and average pooling are first deployed along the channel dimension, and the resulting feature maps are concatenated to form a two-channel representation. A subsequent 1×1 convolution is then employed to further extract spatial information. Finally, the Sigmoid function is applied to produce the spatial attention map ω_2, which highlights the most informative regions. Compared with CAM, SAM places emphasis on the identification of spatial regions that exert a substantial influence on the input features. This process enables the capture of temporal variations in fault signals across disparate time periods.

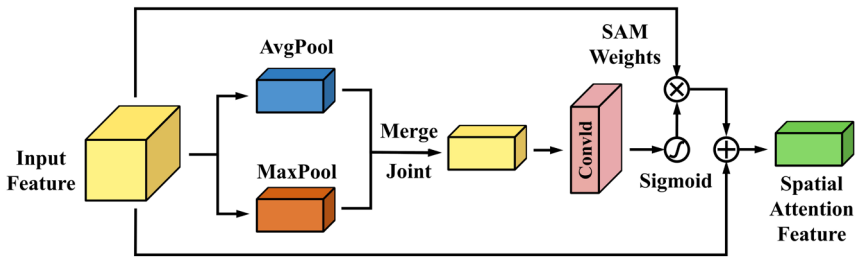

Figure 3. SAM Schematic Diagram.

SAM can be expressed as follows:

$$\omega_2 = \delta_2(F_3(Avepool(x); Max(x))) \tag{5}$$

where F_3 represents a 1×1 convolution with a single channel. As with CAM, residual connections are introduced to prevent performance degradation.

2.2.3. Convolutional Block Attention Module

Convolution operations primarily focus on capturing local features of the input data, whereas attention mechanisms are capable of modeling global contextual information. The combination of these two techniques enables the model to both extract fault-related local features and understand the overall data distribution. The CBAM integrates both the CAM and the SAM. As illustrated in Figure 4, the output is weighted through the CAM and spatial SAM to achieve the result, thereby augmenting the model's perception of and utilizing pivotal information [37].

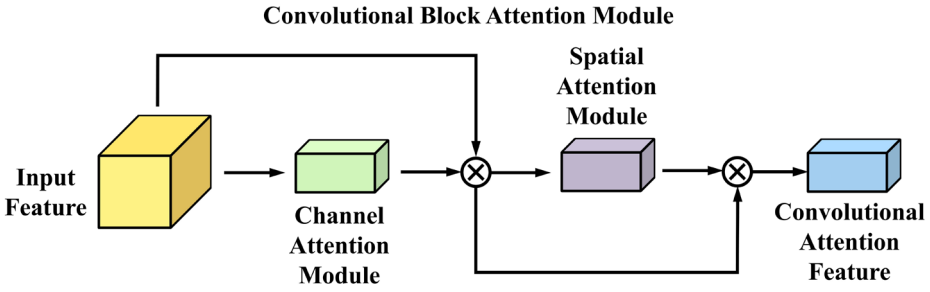

Figure 4. CBAM schematic diagram.

Within CBAM, the convolutional layers are responsible for extracting local features, while the attention modules assign different weights to features according to their relative importance. This integration helps the model more accurately identify and interpret fault-related features in few-shot bearing datasets, thereby improving overall diagnostic performance. Moreover, by enabling the model to concentrate on informative regions and suppress irrelevant information, CBAM enhances computational efficiency. The CBAM provides an effective and reliable solution for identifying faults in complex conditions with scarce data.

3. GAN Data Augmentation

3.1. GAN Architecture

The configuration of the constructed GAN is shown in Figure 5, where both the generator and discriminator adopt CNN structures. The performance of the GAN relies on the dynamic adversarial interaction and alternating optimization between the generator and discriminator. Consequently, a degree of architectural symmetry is maintained between the two networks to facilitate this process.

Specifically, the discriminator comprises 3 convolutional layers, 3 pooling layers, and 2 fully connected layers. In contrast, the generator takes a random noise vector as input and progressively synthesizes data through three upsampling layers followed by three transposed convolution layers.

The overall GAN structure is shown in Figure 6, and the detailed network parameters are listed in Table 1. The overall GAN structure is shown in Figure 6. The input sample size for the discriminator is 24×24, while the generator takes a 256-dimensional vector of random noise, which is reshaped into a $2 \times 2 \times 64$ matrix prior to being input into the

CNN for alternating training with the discriminator. The ReLU function is used, and the average pooling is applied for sampling.

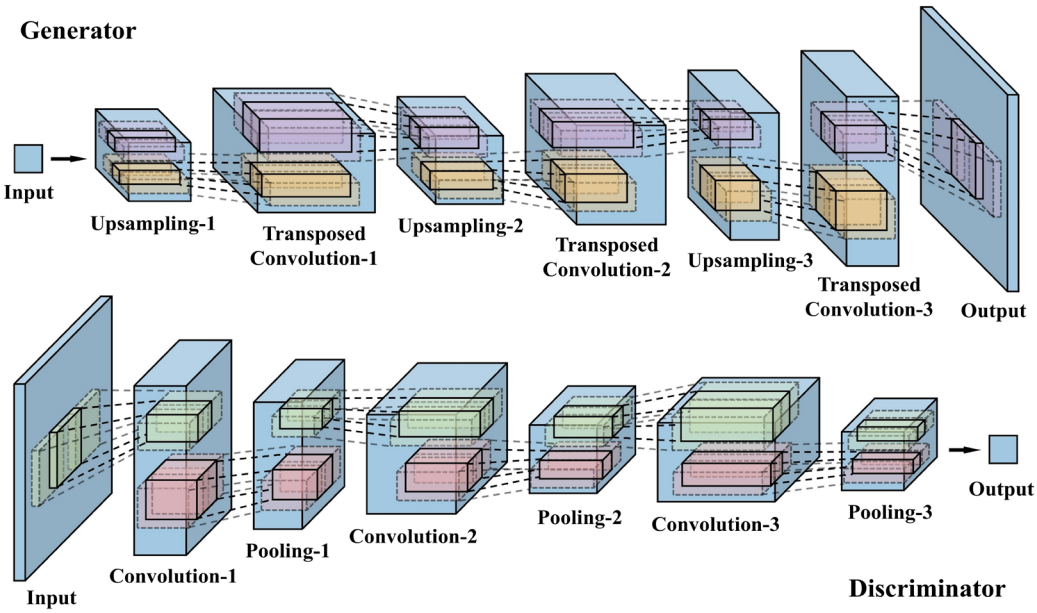

Figure 5. The architecture of the constructed GAN.

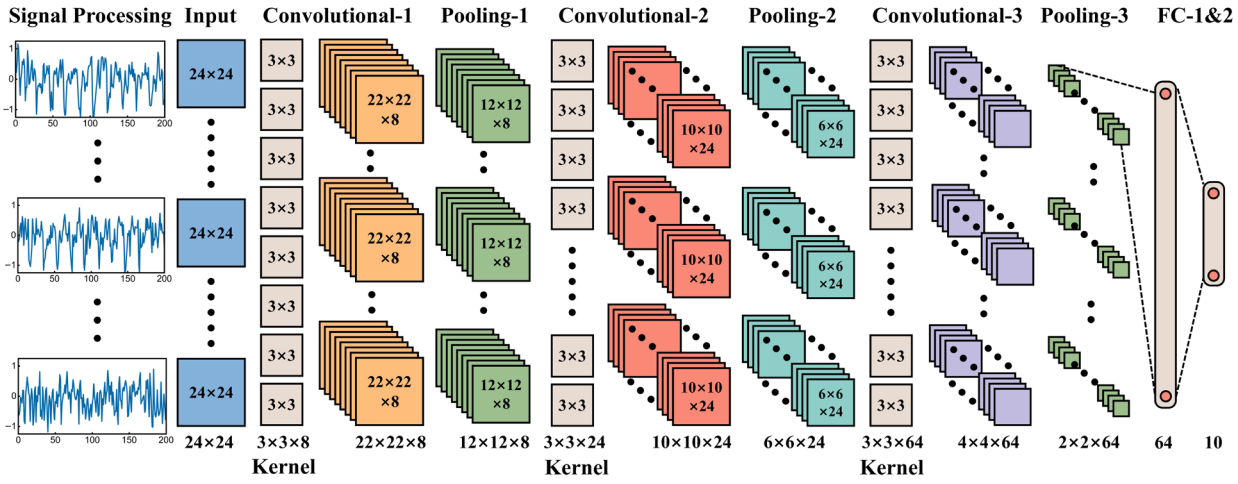

Figure 6. The overall GAN structure.

Table 1. Comparison of sample hardness.

Network Layer	Parameters
Input layer	Input matrix 24 × 24
Convolutional layer_1	Convolution kernel 3 × 3, quantity 8, stride 1
Activation function	ReLU
Pooling layer_1	Average pooling 2 × 2, padding 1
Convolutional layer_2	Convolution kernel 3 × 3, quantity 24, stride 1, padding 1
Activation function	ReLU
Pooling layer_2	Average pooling 2 × 2, padding 1
Convolutional layer_3	Convolution kernel 3 × 3, quantity 64, stride 1
Activation Function	ReLU
Pooling layer_3	Pooled layer 2 × 2
Fully connected layer	64, 10

3.2. Loss Function Selection

The loss function plays a critical role in the generation and optimization of GANs, as it directly affects both the quantity and the stability of the training process. The following loss functions are commonly employed: the least squares loss, the cross-entropy loss, and the Wasserstein distance loss [38].

Different loss functions are suitable for different types of datasets. To identify the most appropriate loss function for fault data from high-speed rotating machinery bearings, this study uses pitting fault samples from the rolling elements of bearings in a 2AMT as an example to systematically compare the effects of these three loss functions on the characteristics of generated data.

Each pitting fault sample consists of 576 data points, forming an input matrix of size 24 × 24, as illustrated in Figure 7. In light of the limited availability of rolling element pitting fault samples, 20 real samples were used as the input data for GAN training. Under different loss function settings, new synthetic samples were generated with a learning rate of 0.001, a batch size of 2, and 50 training iterations. The generated samples were then combined with experimental data from three other bearing health states to form the diagnostic dataset, as shown in Table 2. Among them, the number of real samples for the normal, outer ring pitting, and inner ring pitting conditions was 200 each, while for the rolling element pitting condition, 20 samples were real, and 180 were generated.

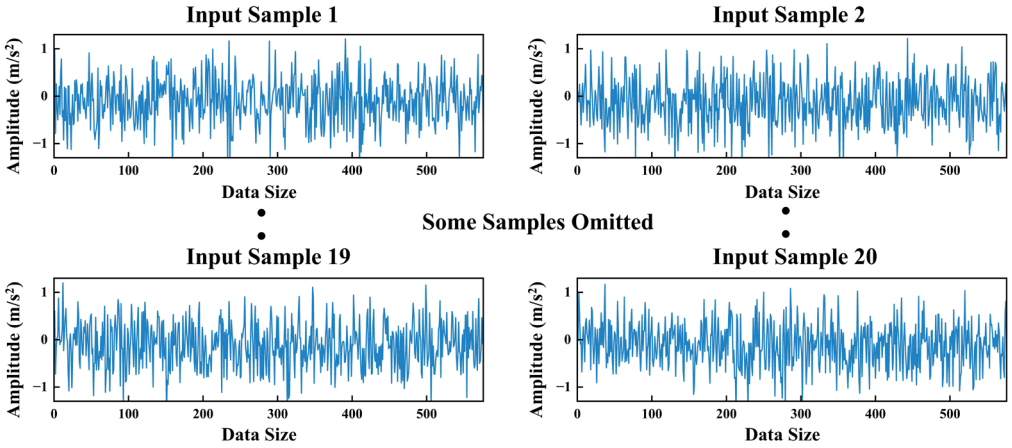

Figure 7. Actual data samples of pitting failures in bearing rolling elements.

Table 2. Rolling bearing few-shot dataset for loss function selection.

Label	Type	Experimental Data Sample Size	Generate Data Sample Size
N	Normal	200	0
OP	Outer ring pitting	200	0
IP	Inner ring pitting	200	0
BP	Rolling element pitting	20	180

The intelligent model based on raw vibration data (RVDCNN) [23] was adopted as the diagnostic framework for testing. Additionally, the dataset in Table 2 was partitioned into training (80%) and testing (20%) subsets, and fault diagnosis experiments were repeated 10 times. The accuracies pertaining to disparate loss functions are depicted in Figure 8, while the mean accuracy and the standard deviation are summarized in Table 3.

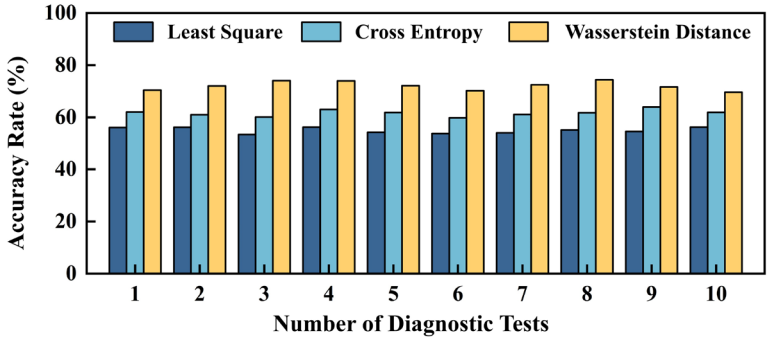

Figure 8. Bearing pitting formation data: accuracy results of 10 failure diagnoses.

Table 3. Comparison of diagnostic accuracy under different loss functions.

Loss Function	Average Recognition Accuracy (%)	Standard Deviation
Least squares loss function	54.90	1.03
Cross-entropy loss function	61.70	1.19
Wasserstein distance loss function	72.05	1.63

The experimental results indicate that among all the loss functions, the Wasserstein distance loss achieves the highest diagnostic performance, with an average recognition accuracy of 72.05%. The confusion matrix of the fifth experiment, which corresponds to the median accuracy, is presented in Figure 9.

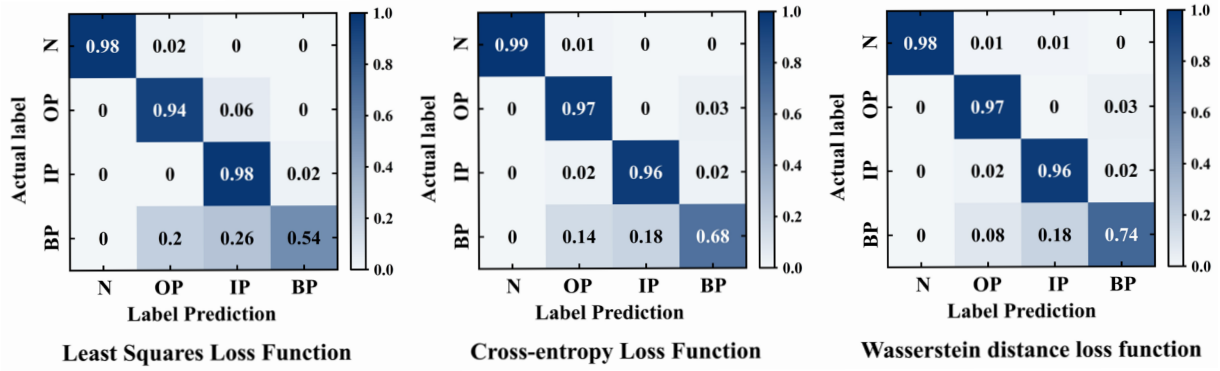

Figure 9. Diagnosis confusion matrix plots of generated data using different loss functions.

For newly generated samples, the diagnostic accuracies for the normal, outer ring, and inner ring pitting conditions range between 0.94 and 0.99, markedly exceeding those for the rolling-element pitting condition. Most of the generated rolling element pitting samples were misclassified as inner or outer ring pitting, leading to a substantial decrease in diagnostic accuracy.

In summary, the samples generated using the Wasserstein distance loss are closer to the true data distribution. Nevertheless, when both generated and real samples are evaluated jointly, the single CNN-based diagnostic model does not achieve satisfactory recognition accuracy. This indicates a need to further enhance the adaptive feature extraction capability of the CNN to improve diagnostic performance under mixed-data conditions.

4. Few-Shot Intelligent Diagnosis Model Based on CBAM

The data generated by the GAN alleviates the issue of insufficient effective data acquisition caused by the short fault-evolution duration in rotating machinery bearings. To

further strengthen the model's capability in identifying and extracting salient features, the CBAM-CNN architecture is proposed.

4.1. Model Structural Parameters

The architecture of the proposed CBAM-CNN model is presented in Figure 10. For few-shot real data, large convolution kernels are initially employed to extract the global characteristics of the raw vibration data, followed by the global average pooling layer for feature compression. Subsequently, the CBAM is incorporated into the feature extraction stage to enhance feature depiction. Thereafter, a 1 × 1 convolution is applied to fuse and calibrate the features derived from both real and generated samples.

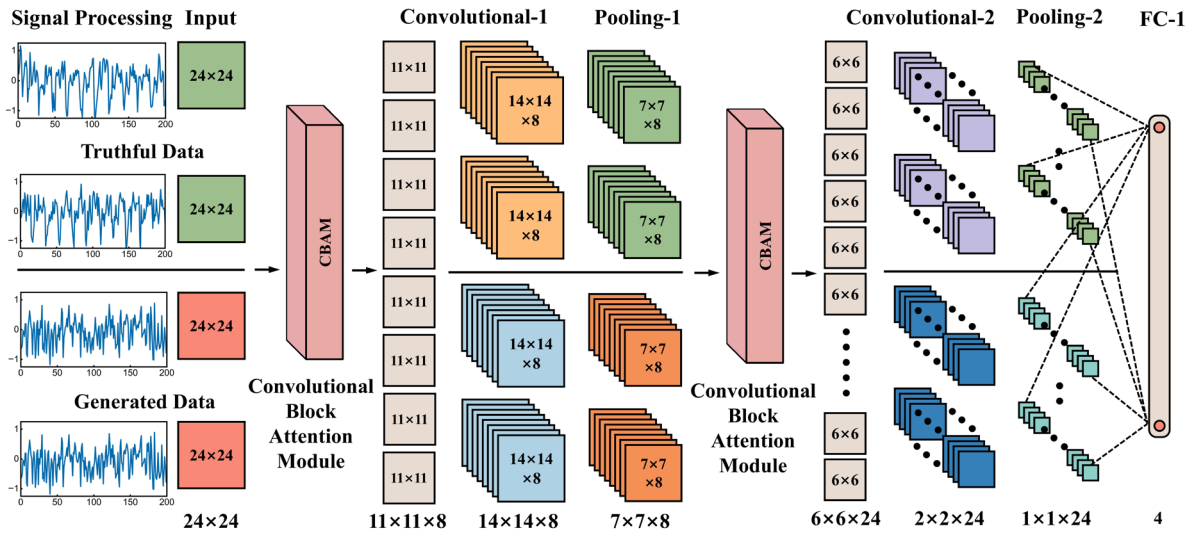

Figure 10. The overall structure of CBAM-CNN.

The detailed configuration of the model's structure and parameters is enumerated in Table 4. The input sample size is 24 × 24, and the network includes two convolutional layers. The first layer adopts a large convolution kernel of 11 × 11 with eight channels, while the second layer employs a 6 × 6 kernel with the channel number expanded to 24. The use of large convolution kernels provides a wide receptive field, covering larger regions of the input data and permitting the model to effectively apprehend large-scale features while reducing the need for deeper convolutional layers. The pooling layers utilize 2 × 2 average pooling to accomplish dimensionality reduction and feature aggregation, thereby enhancing the model's ability to emphasize dominant trends in fault-related features.

Table 4. Pre-trained model structural parameters.

Number of Layers	Network Layer	Parameters
1	Input layer	Input Matrix 24 × 24
2	Convolutional layer_1	Convolution kernel 11 × 11, quantity 8, stride 1
3	Batch normalization	Channel number 8
4	Activation function	ReLU
5	Pooling layer_1	Average pooling 2 × 2
6	Convolutional layer_2	Convolution kernel 6 × 6, quantity 24, stride 1
7	Batch normalization	Channel number 6
8	Activation function	ReLU
9	Pooling layer_2	Average pooling 2 × 2
10	Fully connected layer	4

4.2. 2AMT Bearing Failure Diagnosis Experiment

The target bearing application is associated with the 2AMT used in pure electric vehicles, as shown in Figure 11a. The 2AMT system consists of three shafts, front and rear housings, gears, and rolling bearings.

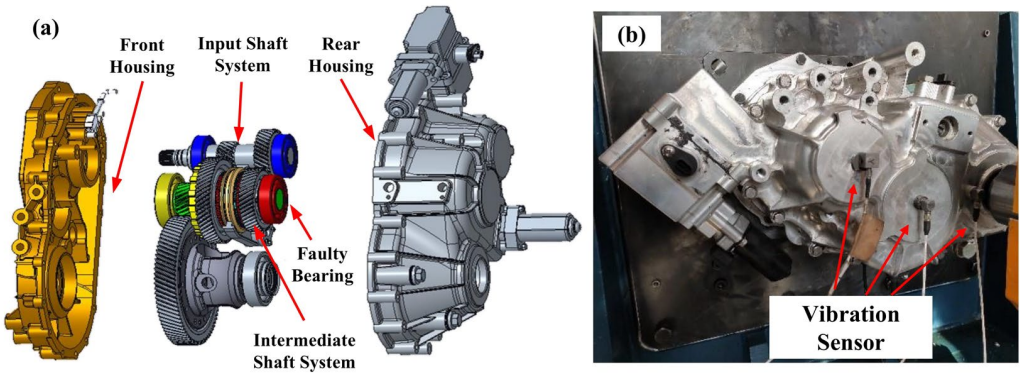

Figure 11. Schematic diagram of the 2AMT: (**a**) 3D structure diagram; (**b**) sensor mounting positions.

The bearing highlighted in Figure 11a serves as the target bearing in the experimental setup, and the corresponding sensor arrangement is shown in Figure 11b. The experimental bearings are deep groove ball bearings. To establish a comprehensive fault dataset, six types of defective bearings were fabricated using laser machining technology. The surface pitting faults are indicated by the red circles in Figure 12. These include single faults occurring on the inner ring, outer ring, and rolling element, as well as compound faults combining inner & outer rings, inner & rolling elements, and outer & rolling elements. The pitting diameter is 0.53 mm for single faults and 0.18 mm for compound faults.



Figure 12. Single and combined failures in experimental faulty bearings.

Vibration data were collected using the LMS Test Lab system. Vibration signal acquisition was performed for 7 bearing types: one healthy bearing, three single pitting fault types, and three compound pitting fault types. Each bearing type was tested for 30 s at a sampling frequency of 16,384 Hz. For vibration-signal analysis, the axial vibration signals were selected under an operating condition of 5000 r/min and 32 Nm. This experimental condition was established based on the China Light-Duty Vehicle Test Cycle (CLTC-P) and formulated according to the comprehensive shift logic of 2AMT together with the operating

speed-torque range of the second gear. Partial vibration acceleration signal segments for the seven bearing types are presented in Figure 13.

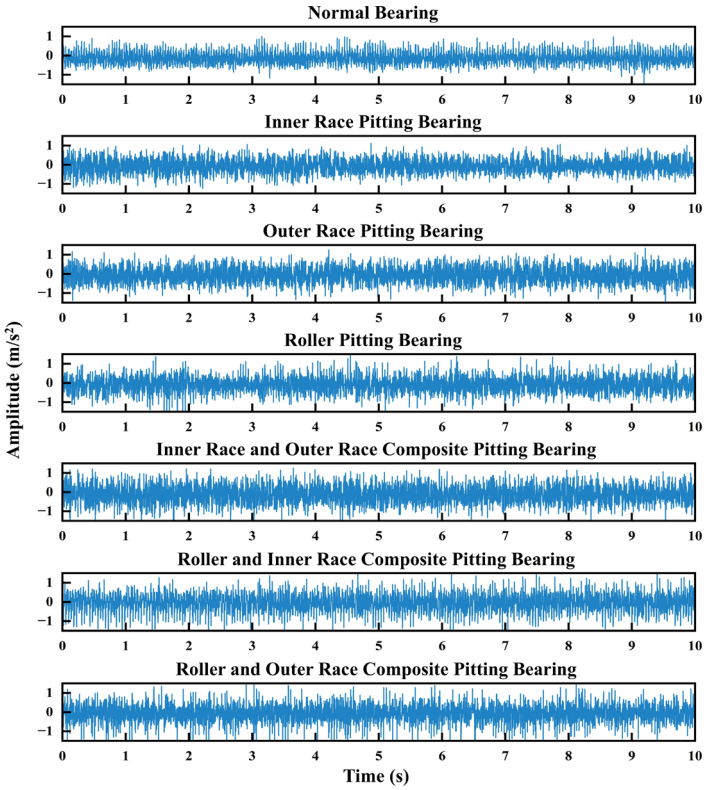

Figure 13. Time-domain diagrams of vibration acceleration for 6 defect states of bearings.

4.3. Diagnostic Results for Rolling Bearing Few-Shot Dataset

To validate the CBAM-based few-shot intelligent diagnostic model, the 2AMT single-point pitting bearing dataset was first analyzed. Each fault condition contained 200 samples, including three pitting fault types: inner ring, rolling element, and outer ring. For each fault type, they randomly selected 10~50% of the samples as real data, while they generated the remaining samples using the GAN to form a hybrid dataset, as Table 5 summarizes. The four bearing conditions—normal, inner ring pitting, ball pitting, and outer ring pitting—were labeled as N, IP, BP, and OP, respectively.

Table 5. 2AMT single-point pitting bearing dataset.

Label	Type	Sample Size (Units)	Actual Data Ratio (%)
N	Normal	200	100
OP	Outer ring pitting	200	10, 20, 30, 40, 50
IP	Inner ring pitting	200	10, 20, 30, 40, 50
BP	Rolling element pitting	200	10, 20, 30, 40, 50

Fifteen datasets were constructed based on the three pitting fault types (outer ring, inner ring, and rolling element) under different real-data proportions, and multiple diagnostic experiments were performed for each case. The overall distribution of recognition accuracies is shown in Figure 14. When the rolling element pitting data were used as few-shot data, the recognition accuracy for all data proportions exceeded 96.32%, representing a significant improvement compared with the RVDCNN baseline. Furthermore, within

each fault category, the recognition accuracy consistently increased as the proportion of real samples increased.

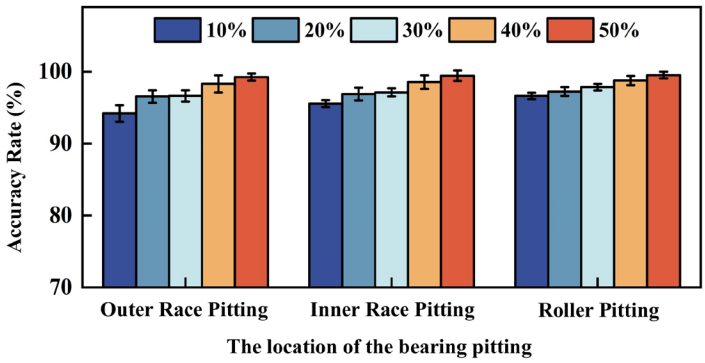

Figure 14. Diagnostic accuracy rates under different proportions of real fault data.

The mean and standard deviation of the ten repeated diagnostic results are summarized in Table 6. Under the same few-shot ratio, increasing the severity of pitting faults enhanced the fault-related features in vibration signals, thereby improving local feature learning through the CBAM. When the proportion of real data reached 50%, the average diagnostic accuracies for the outer ring, inner ring, and rolling element pitting faults were 99.26%, 99.46%, and 99.54%, respectively, indicating that the model could accurately identify fault types under this condition.

Table 6. Specific diagnostic accuracy rates under different proportions of real fault data.

Real Data Proportion (%)	Inner Ring Pitting Accuracy (%)	Outer Ring Pitting Accuracy (%)	Rolling Element Pitting Accuracy (%)
10	94.21 ± 0.81	95.57 ± 0.34	96.64 ± 0.32
20	96.56 ± 0.62	96.89 ± 0.63	97.25 ± 0.44
30	96.64 ± 0.56	97.14 ± 0.41	97.86 ± 0.31
40	98.32 ± 0.84	98.57 ± 0.66	98.78 ± 0.46
50	99.26 ± 0.36	99.46 ± 0.52	99.54 ± 0.33

When the proportion of real data was limited to 10%, confusion matrices were employed to characterize the diagnostic performance across the three pitting fault types, as shown in Figure 15. In these matrices, the horizontal and vertical axes correspond to the predicted and actual labels, and the matrix values represent the normalized proportion of predicted samples.

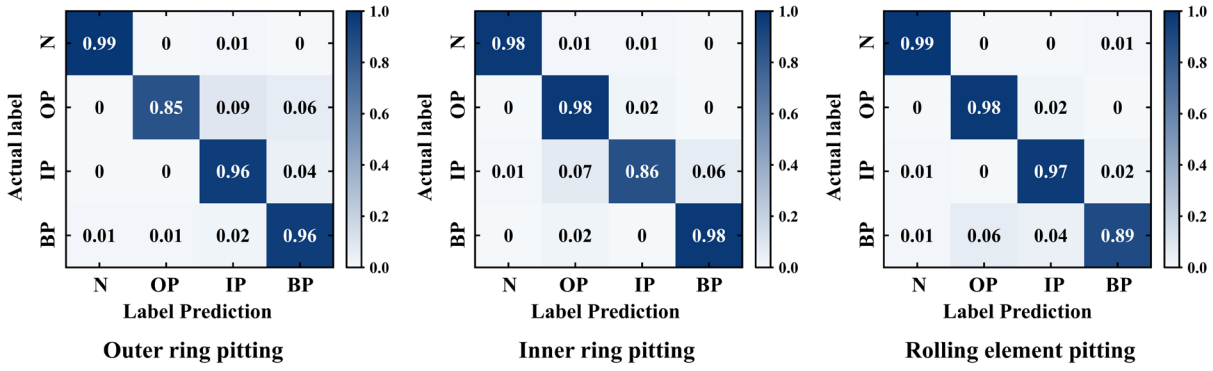

Figure 15. Confusion matrix diagram of the single-point pitting bearing dataset.

When the outer ring pitting data was used as the few-shot samples, the overall diagnostic accuracy was 94.00%, with an 85.00% accuracy for outer ring pitting samples; 9% and 6% of these were misclassified as "IP" and "BP", respectively.

When the inner ring pitting data served as few-shot samples, the overall accuracy was 95.00%, and the diagnostic accuracy for inner ring pitting was 86.00%, with 8% and 6% of samples misclassified as "BP" and "BP", respectively.

When the rolling element pitting data served as few-shot samples, the overall diagnostic accuracy reached 96.25%, and the accuracy for rolling element pitting was 90.00%, with 6% and 4% misclassified as "BP" and "IP", respectively.

The confusion matrices indicate that, when the real data of outer ring and inner ring pitting faults were scarce, the augmented samples generated by GAN were more susceptible to misclassification into other pitting categories. In contrast, for few-shot rolling element pitting faults, the CBAM effectively enhanced the extraction and recognition of pitting features through its adaptive weighting, thereby yielding superior diagnostic performance.

4.4. Diagnostic Results for Rolling Bearing Mixed Sample Dataset

As the number of transmission components increases, the rolling bearings inside the gearbox are exposed to various excitation sources and ambient noise during operation, which substantially degrades the signal-to-noise ratio. To further evaluate the generalization capability of the proposed CBAM-CNN model, a hybrid bearing dataset was constructed, as summarized in Table 7.

Table 7. Rolling bearing mixed pitting dataset.

Label	Type	Fault Dimensions (mm)	Dataset Name								
			A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**J**
N	Normal	--									
IP	Inner ring pitting	0.53	√ [1]						√ [2]	√	
BP	Rolling element pitting	0.53		√					√ [2]		√
OP	Outer ring pitting	0.53			√					√	√
IOP	Inner & outer rings pitting	0.18				√					
IBP	Inner ring & rolling element pitting	0.18					√				
OBP	Outer ring & rolling element pitting	0.18						√			

[1] The dataset A contains samples of inner ring pitting faults. [2] The dataset G contains samples of inner ring pitting and rolling element pitting faults.

The dataset consisted of 7 bearing types: the bearing in normal condition; single-point pitting faults located on the inner ring, rolling element, and outer ring (fault size: 0.53 mm); and compound pitting faults occurring on inner & outer rings (IOP), inner ring & rolling element (IBP), and outer ring & rolling element (OBP) (fault size: 0.18 mm). Based on these seven fault types, nine datasets (labeled A–J) were generated.

In dataset A, the inner ring pitting fault (IP) was defined as the few-shot category, with a real-to-generated sample ratio of 2:8 and a total of 200 samples per fault type. In dataset G, both IP and BP were considered few-shot categories, while the remaining five fault types each contained 200 samples. Following this criterion, six datasets were designed with single few-shot labels, and three datasets contained dual few-shot labels.

Each of the nine few-shot bearing datasets was subjected to 10 repeated diagnostic experiments using the CBAM-CNN model. The mean accuracy and standard deviation were adopted as evaluation metrics, and the results are presented in Figure 16. As illustrated by the error bars, diagnostic accuracy exhibited slight fluctuations across datasets but remained generally stable. Datasets D~J showed slightly lower diagnostic accuracy

compared to A~C, reflecting the increased difficulty introduced by compound faults and limited quantities of real samples.

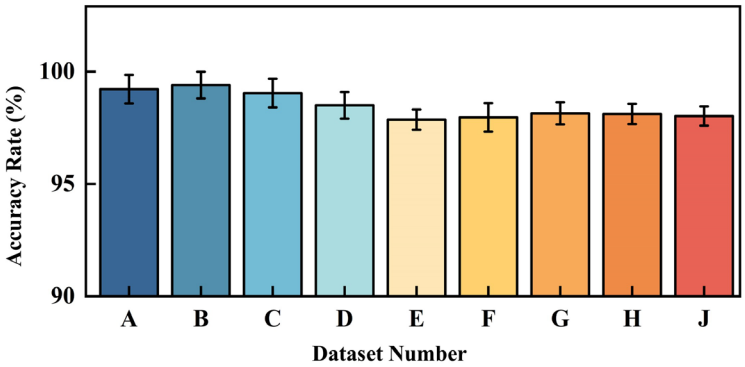

Figure 16. Diagnostic results for the few-shot sample mixed dataset of rolling bearings.

The detailed results are summarized in Table 8. The CBAM-CNN model demonstrated consistently high and stable diagnostic performance across all fault types. The accuracies for datasets A~C were 98.36%, 98.27%, and 98.31%, indicating stable performance for single fault types. Dataset E, which contained few-shot compound faults involving the inner ring and rolling element, achieved the lowest accuracy at 96.90%. Datasets G~J, each containing dual few-shot labels, achieved accuracies of 97.65%, 97.55%, and 97.59%, approximately 0.72% lower than datasets A~C, demonstrating the model's adaptability to complex labeling scenarios. As datasets D~F involved compound few-shot categories and datasets G~J contained multiple few-shot labels, the increased complexity in data distribution and temporal sequence structure posed greater challenges to both the generator and discriminator, thereby slightly reducing diagnostic accuracy.

Table 8. Average accuracy and standard deviation of different few-shot sample datasets.

Dataset	A	B	C	D	E	F	G	H	J
Average precision (%)	98.36	98.27	98.31	97.14	96.90	97.07	97.65	97.55	97.59
Standard deviation	0.45	0.44	0.43	0.42	0.39	0.40	0.37	0.34	0.53

To further investigate the classification performance, confusion matrices for datasets D~J were constructed, as shown in Figure 17, where the horizontal and vertical axes represent the predicted and actual fault labels, and each matrix entry number denotes the normalized proportion of predicted samples.

For dataset D, the overall diagnostic accuracy was 97.14%, with 12 misclassified compound fault samples for inner outer ring pitting. For dataset E, the overall accuracy was 96.57%, with 22 misclassified samples for inner-rolling element pitting. For dataset F, the accuracy was 96.43%, with 18 misclassified samples for outer-rolling element pitting. For dataset G, the accuracy was 97.57%, with 10 misclassified inner ring and 10 misclassified rolling element pitting samples. For dataset H, the accuracy was 97.21%, with 11 misclassified inner ring and 8 misclassified outer ring samples. For dataset J, the accuracy was 97.14%, with 9 misclassified outer ring and 8 misclassified rolling element samples.

Overall, the misclassified samples were predominantly concentrated in the few-shot fault categories, indicating that data scarcity remained the main challenge. Nevertheless, the CBAM-CNN model maintained high robustness and diagnostic reliability under these complex few-shot and compound fault conditions, confirming its potential applicability in real-world transmission systems.

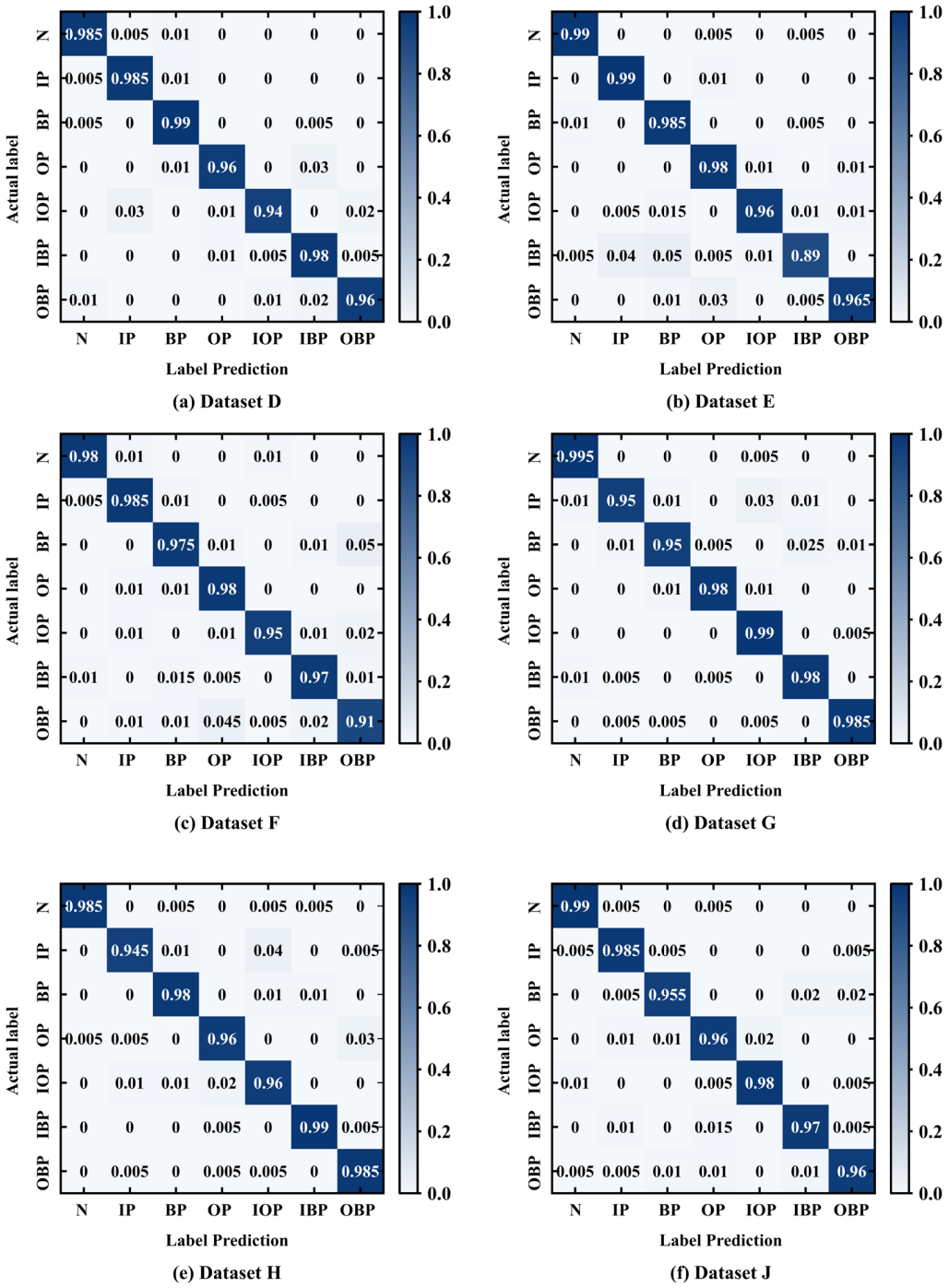

Figure 17. Confusion matrix results for the few-shot sample dataset.

4.5. Model Core Ablation and Comparison of Different Diagnostic Models

The objective of the ablation study is to quantify the functional contribution of each core module to the overall diagnostic performance. To systematically evaluate the independent effects of the CBAM-CNN and the GAN, the ablation experiments were conducted using dataset C. Similar to the hyperparameter tuning process, all ablation configurations were trained under identical optimization settings to ensure fair comparability. The diagnostic accuracies of ablation are summarized in Table 9, and the corresponding confusion matrices are illustrated in Figure 18.

Table 9. Average accuracy and standard deviation for different ablation models.

Model	CNN	GAN + CNN	CBAM-CNN	Proposed Model
Average precision (%)	84.57	91.39	92.71	98.31
Standard deviation	1.15	0.82	0.79	0.43

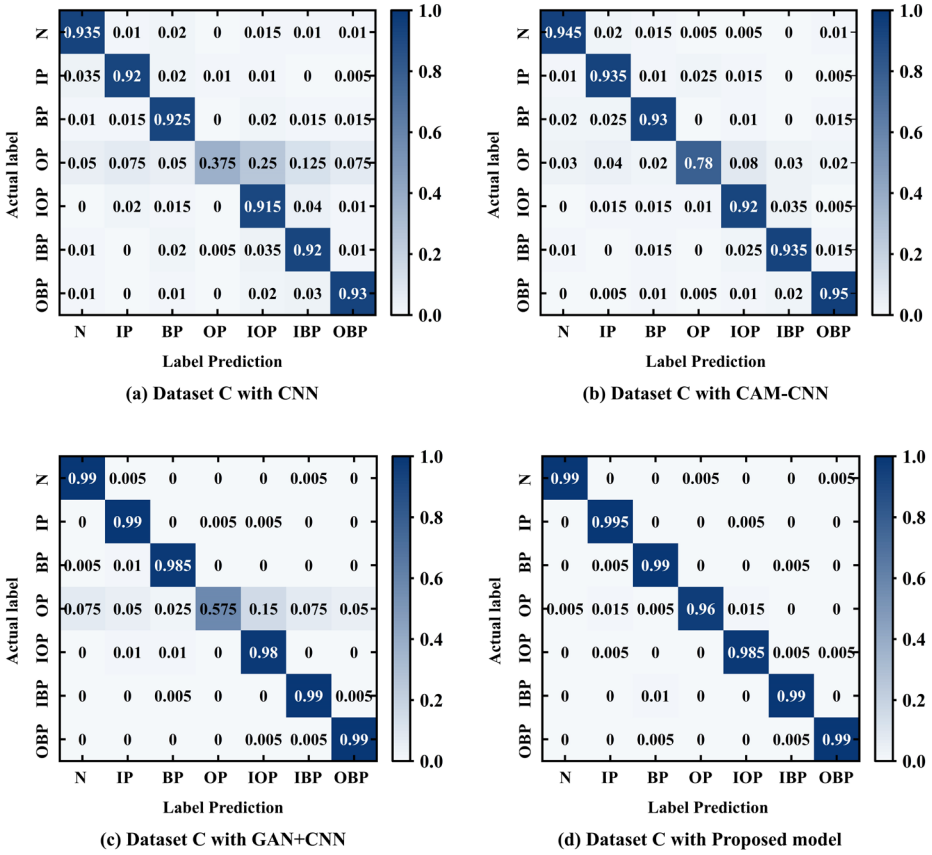

Figure 18. Confusion matrix results for different ablation models with dataset C.

As shown in Table 9, the baseline CNN model that excludes both GAN and CBAM modules achieves the lowest performance across all evaluation metrics, with the average accuracy rate of 85.47%. This performance degradation is attributable to limited original training, data imbalance, and the restricted capability of traditional convolutional layers to capture complex fault patterns.

When the GAN module is introduced, the model performance improves significantly. This improvement is primarily attributed to the GAN's ability to generate class-consistent synthetic samples, thereby enriching the training distribution, mitigating data imbalance, and reducing overfitting. When the CBAM module was further incorporated, the diagnostic accuracy improved relative to the GAN + CNN model. The CBAM adaptively refines both channel and spatial attention distributions, enabling the model to emphasize salient fault features and suppress irrelevant information. This enhances the model's ability to focus on and represent key features. Furthermore, by combining the GAN with the CBAM, the diagnostic performance of the model has been further enhanced.

The evaluation of each module's contribution to the enhancement of diagnostic accuracy was conducted through quantification. The contribution degree of the GAN was 8.06%; the contribution degree of the CBAM was 9.63%; and the contribution degree of the GAN and CBAM-CNN was 16.25%. The contribution of the CBAM module was greater, and there was a certain degree of functional overlap between the two modules.

To further demonstrate the proposed model's reliability, a recent related model was included for comparison. The baseline model integrates an asymmetric convolutional network (AC-Net) with a multi-head attention mechanism (MHA) and utilizes TL and GAN to generate fault data [39]. The confusion matrices are illustrated in Figure 19. The diagnostic accuracies are summarized in Table 10.

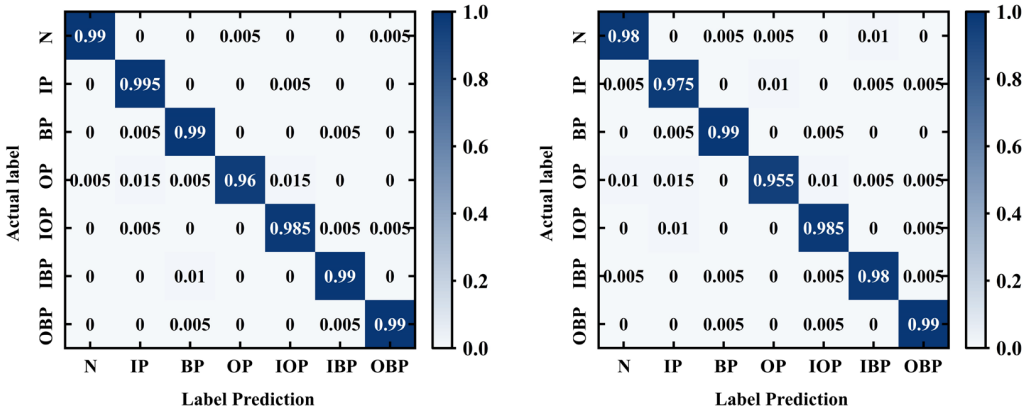

Figure 19. Confusion matrix results for baseline models with dataset C.

Table 10. Average accuracy and standard deviation for different baseline models.

Model	Proposed Model	Baseline Model
Average precision (%)	98.31	97.94
Standard deviation	0.43	0.49

The comparative results reveal that the proposed model delivers superior diagnostic accuracy in few-shot fault diagnosis tasks. The CBAM enhances the model's capacity to focus on discriminative and fine-grained fault features, thereby improving classification precision. Owing to the scarcity of fault samples in the source domain, the baseline model's capability for fault detection was suboptimally developed. Consequently, after fine-tuning on the target domain data, its diagnostic accuracy was lower than that of the CBAM-CNN model.

Overall, the results demonstrate that GAN-based data augmentation broadens the training data distribution, while CBAM preserves essential diagnostic semantics. These mechanisms play a pivotal role in improving the generalization capability and diagnostic reliability of the proposed model.

5. Discussion

The GAN module, founded upon the Wasserstein distance loss function, is demonstrated to effectively address issues such as data imbalance and the paucity of training samples, thereby synthesizing a high-precision fault dataset. Concurrently, the convolutional attention mechanism module is introduced, thereby enabling the model to focus on both local features and global information of the input data. The combination of convolutional operations and the attention mechanism enables the model to extract fault feature information from the data and comprehend the data distribution structure, thereby enhancing the model's capacity to identify bearing faults at various scales. Furthermore, the findings of the recognition results of bearing faults on small sample datasets demonstrate the efficacy of the proposed fault diagnosis model. The model receives the vibration acceleration signals from the sensors of high-speed equipment. It detects the operating status of

key components in real time, judges their operating conditions, and sends out information, such as fault alerts.

However, in practical engineering applications, the model parameters require fine-tuning based on the operating conditions and noise levels of the target equipment to achieve optimal diagnostic accuracy. Additionally, it should be noted that there is still room for improvement in the final diagnosis results due to certain deviations between the generated data by GAN and the real data, and the limited training dataset of CBAM-CNN, leading to overfitting.

6. Conclusions

This study addresses the challenges of limited fault samples, insufficient feature extraction, and poor diagnostic performance in high-speed rotating machinery under real-world conditions. A data-driven CBAM-CNN model for few-shot intelligent fault diagnosis is proposed, enabling high-precision identification and diagnosis of rolling bearing faults. The proposed model integrates the high-quality sample generation capability of GAN with the feature extraction and recognition capabilities of CBAM. The key conclusions are as follows:

1. To determine the most suitable loss function for high-speed rotating machinery bearing fault data, few-shot pitting fault data from the 2AMT were analyzed. A systematic comparison of different loss functions for GANs revealed that employing the Wasserstein distance loss significantly improved the quality and diversity of the generated data, yielding the highest diagnostic accuracy.
2. To address the insufficient feature extraction capability of few-shot and generated data, the CBAM was incorporated into the CNN framework. By combining CAM and SAM, the model adaptively weighted key features across different bearing fault categories, thereby enabling the model to focus on both local features and global information of the input data and enhancing its diagnostic accuracy and efficiency.
3. The CBAM-CNN model was validated using a hybrid pitting fault dataset of intermediate shaft bearings from a 2AMT. When the proportion of real samples was only 10%, the diagnostic accuracy for rolling element pitting faults was 96.64%. The diagnostic accuracy increased steadily with a higher proportion of real data. When the real sample ratio reached 20%, the average diagnostic accuracy across nine few-shot hybrid pitting datasets ranged from 96.90% to 98.36%. The ablation experiments and comparison with advanced models also demonstrated the adaptability of the proposed model.

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

Funding: This research was funded by the Guangxi Major Science and Technology Special Project through grants No. GuiKeAA24263068 and the Guangxi Major Science and Technology Special Project through grants No. GuiKeAA24206038.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: Guangxi University and Hebei University of Technology are gratefully acknowledged for their support in producing samples and technical services, respectively.

Conflicts of Interest: Author Yunhui Bai was employed by the company Aurobay (Ningbo) Intelligent Technology Co., Ltd. Author Lijie Hao was employed by the company AECC BeijingHangke Engine Control System Science & Technology Co., Ltd. The authors declare no conflicts of interest. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

CNN	Convolutional neural network
GAN	Generative adversarial network
SVM	Support vector machine
CAM	Channel attention mechanism
SAM	Spatial attention mechanism
CBAM	Convolutional block attention module
CBAM-CNN	CNN with a convolutional block attention module
AMFCNN	Attention-based multi-layer fusion CNN
TL	Transfer learning
RVDCNN	CNN based on raw vibration data
2AMT	Two-speed automatic mechanical transmission
AC-Net	Asymmetric convolutional network
MHA	Multi-head attention mechanism

References

1. Chen, Y.; Pu, X.R.; Hao, L.J.; Li, G.X.; Luo, L. Study on the Friction Characteristics and Fatigue Life of Carbonitriding-Treated Needle Bearings. *Lubricants* **2024**, *12*, 234. [CrossRef]
2. Qin, Z.H.; Huang, F.G.; Pan, J.F.; Niu, J.L.; Qin, H.H. Improved Generative Adversarial Network for Bearing Fault Diagnosis with a Small Number of Data and Unbalanced Data. *Symmetry* **2024**, *16*, 358. [CrossRef]
3. Hao, L.J.; Chen, Y.; Li, G.X.; Zhang, M.; Wu, Y.M.; Liu, R.; Chen, G. Study on the friction characteristics and fatigue life of manganese phosphate coating bearings. *Lubricants* **2023**, *11*, 99. [CrossRef]
4. An, Y.Y.; Li, Z.F.; Li, Y.Y.; Zhang, K.; Zhu, Z.Q.; Chai, Y. Few-shot learning-based fault diagnosis using prototypical contrastive-based domain adaptation under variable working conditions. *IEEE Sens. J.* **2024**, *24*, 25019–25029. [CrossRef]
5. Dou, S.S.; Li, F.D.; Chang, Y.H.; Chen, J.L.; Zheng, W.G.; Li, A.M. Few-shot fault identification of complex equipment via metric-based features capture GAN combining prior knowledge-augmented strategy. *J. Manuf. Syst.* **2023**, *71*, 238–256. [CrossRef]
6. Chen, Z.R.; Wang, R.C.; Ding, R.K.; Liu, B.; Liu, W.; Sun, D.; Guo, Z.Y. Research Progress and Future Prospects of Brake-by-Wire Technology for New Energy Vehicles. *Energies* **2025**, *18*, 2702. [CrossRef]
7. Jardine, A.K.S.; Lin, D.M.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [CrossRef]
8. Lin, C.L.; Liang, J.W.; Huang, Y.M.; Huang, S.C. A novel model-based unbalance monitoring and prognostics for rotor-bearing systems. *Adv. Mech. Eng.* **2023**, *15*, 1–16. [CrossRef]
9. Coronel, E.; Barán, B.; Gardel, P. A Survey on Data Mining for Data-Driven Industrial Assets Maintenance. *Technologies* **2025**, *13*, 2227–7080. [CrossRef]
10. Tamilselvan, P.; Wang, P.F. Failure diagnosis using deep belief learning based health state classification. *Reliab. Eng. Syst. Safe* **2013**, *115*, 124–135. [CrossRef]
11. Neupane, D.; Bouadjenek, M.R.; Dazeley, R.; Aryal, S. Data-driven machinery fault diagnosis: A comprehensive review. *Neurocomputing* **2025**, *627*, 129588. [CrossRef]
12. Yu, G.Y.; Wu, P.; Lv, Z.; Hou, J.J.; Ma, B.; Han, Y.M. Few-shot fault diagnosis method of rotating machinery using novel MCGM based CNN. *IEEE Trans. Ind. Inform.* **2023**, *19*, 10944–10955. [CrossRef]
13. Shang, Z.W.; Zhang, J.; Li, W.X.; Qian, S.Q.; Liu, J.Y.; Gao, M.S. A novel small samples fault diagnosis method based on the self-attention Wasserstein generative adversarial network. *Neural Process. Lett.* **2023**, *55*, 6377–6407. [CrossRef]
14. Zhao, Y.P.; Cai, W. The perceptron algorithm with uneven margins based transfer learning for turbofan engine fault detection. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107249. [CrossRef]
15. Yao, L.H.; Wang, H.W.; Tao, L.; Fang, Z.W.; Wang, H.R.; Liu, Y.; Wang, H.L. Bearing fault diagnosis based on transfer learning with dual-flow manifold ResNet and improved CapsNet. *Meas. Sci. Technol.* **2024**, *35*, 076123. [CrossRef]

16. Soomro, A.A.; Muhammad, M.B.; Mokhtar, A.A.; Saad, M.H.M.; Lashari, N.; Hussain, M.; Sarwar, U.; Palli, A.S. Insights into modern machine learning approaches for bearing fault classification: A systematic literature review. *Results Eng.* **2024**, *23*, 102700. [CrossRef]
17. Zhang, W.; Peng, G.L.; Li, C.H.; Chen, Y.H.; Zhang, Z.J. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* **2017**, *17*, 425. [CrossRef]
18. Jiang, G.Q.; He, H.B.; Yan, J.; Xie, P. Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. *IEEE Trans. Ind. Electron.* **2018**, *66*, 3196–3207. [CrossRef]
19. Wang, M.Y.; Yang, Y.X.; Wei, L.X.; Li, Y.S. A lightweight gear fault diagnosis method based on attention mechanism and multilayer fusion network. *IEEE Trans. Instrum. Meas.* **2023**, *73*, 3503011. [CrossRef]
20. Zhou, H.; Liu, R.D.; Li, Y.X.; Wang, J.C.; Xie, S.C. A rolling bearing fault diagnosis method based on a convolutional neural network with frequency attention mechanism. *Struct. Health Monit.* **2024**, *23*, 2475–2495. [CrossRef]
21. Zhang, G.Y.; Li, Y.T.; Jiang, W.B.; Shu, L.J. A fault diagnosis method for wind turbines with limited labeled data based on balanced joint adaptive network. *Neurocomputing* **2022**, *481*, 133–153. [CrossRef]
22. Chen, Y.; Li, G.X.; Li, A.H.; He, B.L. End-to-End Intelligent Fault Diagnosis of Transmission Bearings in Electric Vehicles Based on CNN. *Lubricants* **2024**, *12*, 364. [CrossRef]
23. Zhan, F.; Hu, L.K.; Huang, W.K.; Dong, Y.K.; He, H.; Wu, G.J. Category knowledge-guided few-shot bearing fault diagnosis. *Eng. Appl. Artif. Intell.* **2025**, *139*, 109489. [CrossRef]
24. Chen, X.H.; Yang, R.; Xue, Y.H.; Huang, M.J.; Ferrero, R.; Wang, Z.D. Deep transfer learning for bearing fault diagnosis: A systematic review since 2016. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3508221. [CrossRef]
25. Ai, T.; Liu, Z.L.; Zhang, J.Y.; Liu, H.H.; Jin, Y.Q.; Zuo, M.J. Fully simulated-data-driven transfer-learning method for rolling-bearing-fault diagnosis. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3526111. [CrossRef]
26. Wang, H.; Wang, P.; Wang, S.Y.; Li, D.Q. Rolling bearing fault diagnosis method based on dynamic simulated model from source domain to target domain with improved alternating transfer learning. *Nonlinear Dyn.* **2025**, *113*, 4485–4510. [CrossRef]
27. Fang, H.R.; Liu, H.; Wang, X.; Deng, J.; An, J.L. The method based on clustering for unknown failure diagnosis of rolling bearings. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3509508. [CrossRef]
28. Yu, T.Z.; Jiang, Z.Y.; Ren, Z.H.; Zhang, Y.C.; Zhou, S.H.; Zhou, X. A Cooperative Convolutional Neural Network Framework for Multisensor Fault Diagnosis of Rotating Machinery. *IEEE Sens. J.* **2024**, *24*, 38309–38317. [CrossRef]
29. Huang, N.T.; Chen, Q.Z.; Cai, G.W.; Xu, D.G.; Zhang, L.; Zhao, W.G. Fault diagnosis of bearing in wind turbine gearbox under actual operating conditions driven by limited data with noise labels. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 3502510. [CrossRef]
30. Miao, J.G.; Wang, J.Y.; Zhang, D.C.; Miao, Q. Improved generative adversarial network for rotating component fault diagnosis in scenarios with extremely limited data. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 3500213. [CrossRef]
31. Zhuang, J.C.; Yan, J.H.; Huang, C.G.; Jia, M.P. Residual attention temporal recurrent network for fault diagnosis of gearboxes under limited labeled data. *Eng. Appl. Artif. Intell.* **2024**, *129*, 107539. [CrossRef]
32. 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]
33. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434. [CrossRef]
34. Ghaffarian, S.; Valente, J.; Van Der Voort, M.; Tekinerdogan, B. Effect of attention mechanism in deep learning-based remote sensing image processing: A systematic literature review. *Remote Sens.* **2021**, *13*, 2965. [CrossRef]
35. Hu, J.Q.; Liu, Y.; Wu, K.S. Neural network pruning based on channel attention mechanism. *Connect. Sci.* **2022**, *34*, 2201–2218. [CrossRef]
36. Liu, T.L.; Luo, R.H.; Xu, L.Q.; Feng, D.C.; Cao, L.; Liu, S.Y.; Guo, J.J. Spatial channel attention for deep convolutional neural networks. *Mathematics* **2022**, *10*, 1750. [CrossRef]
37. Shi, H.T.; Sun, H.Y.; Bai, X.T.; Song, Z.L.; Gao, T.H. Research on rolling bearing fault diagnosis method based on improved multi-source fusion convolutional neural network. *Meas. Sci. Technol.* **2024**, *36*, 015142. [CrossRef]
38. Shao, H.D.; Li, W.; Cai, B.P.; Wan, J.F.; Xiao, Y.M.; Yan, S. Dual-threshold attention-guided GAN and limited infrared thermal images for rotating machinery fault diagnosis under speed fluctuation. *IEEE Trans. Ind. Inform.* **2023**, *19*, 9933–9942. [CrossRef]
39. Zhang, B.J.; Wang, W.; He, Y. A hybrid approach combining deep learning and signal processing for bearing fault diagnosis under imbalanced samples and multiple operating conditions. *Sci. Rep.* **2025**, *15*, 13606. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.