machines

MDPI

Article

A Data Augmentation Method for Shearer Rocker Arm Bearing Fault Diagnosis Based on GA-WT-SDP and WCGAN

Zhaohong Wu [1], Shuo Wang [1,*], Chang Liu [2], Haiyang Wu [1], Jiang Yi [1], Yusong Pang [3] and Gang Cheng [1]

[1] School of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China; wzhaohong@163.com (Z.W.); ts22050183p31@cumt.edu.cn (H.W.); ts22050086a31@cumt.edu.cn (J.Y.); chg@cumt.edu.cn (G.C.)
[2] School of Mechanical and Electrical Engineering, Xuzhou University of Technology, Xuzhou 221018, China; liuchang@xzit.edu.cn
[3] Faculty of Mechanical Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands; y.pang@tudelft.nl
* Correspondence: ws_8989@163.com

Abstract

This work addresses the challenges of inadequate data acquisition and the limited availability of labeled samples for shearer rocker arm bearing faults by developing a data augmentation methodology that synergistically incorporates the Genetic Algorithm-optimized Wavelet Transform Symmetrical Dot Pattern (GA-WT-SDP) with a Wasserstein Conditional Generative Adversarial Network (WCGAN). In the initial step, the Genetic Algorithm (GA) is employed to refine the mapping parameters of the Wavelet Transform Symmetrical Dot Pattern (WT-SDP), facilitating the transformation of raw vibration signals into advanced and discriminative graphical representations. Thereafter, the Wasserstein distance in conjunction with a gradient penalty mechanism is introduced through the WCGAN, thereby ensuring higher-quality generated samples and improved stability during model training. Experimental results validate that the proposed approach yields accelerated convergence and superior performance in sample generation. The augmented data significantly bolsters the generalization ability and predictive accuracy of fault diagnosis models trained on small datasets, with notable gains achieved in deep architectures (CNNs, LSTMs). The research substantiates that this technique helps overcome overfitting, enhances feature representation capacity, and ensures consistently high identification accuracy even in complex working environments.

Keywords: rearing; data augmentation; fault diagnosis; genetic algorithms; wasserstein conditional gAN

Check for updates

Academic Editor: Ahmed Abu-Siada

Received: 1 December 2025
Revised: 14 January 2026
Accepted: 21 January 2026
Published: 26 January 2026

Copyright: © 2026 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.

1. Introduction

Within fully mechanized coal mining operations, the shearer plays an indispensable role, wherein the rocker arm bearing represents one of its fundamental and indispensable components. Bearings subjected to sustained heavy loading and high-speed rotation exhibit a heightened susceptibility to faults, which in turn results in rocker arm damage of the shearer and significantly compromises the operational reliability of the equipment [1–3]. Accordingly, the accurate identification of bearing faults serves as a key prerequisite for guaranteeing the reliability and operational stability of equipment [4–7]. Despite this, the intricate and severe operational conditions in coal mining render the acquisition of high-quality labeled samples challenging, whereas unlabeled data can be collected with

comparatively greater ease. Consequently, the availability of labeled fault data for bearings is rather limited [8,9]. In practical engineering contexts, obtaining labeled fault data is both time-consuming and costly, with the majority of samples remaining unlabeled. Under such conditions, the direct application of traditional data-driven fault diagnosis methods may lead to insufficient training, thereby undermining the generalization capability and accuracy of the diagnostic models [10]. Consequently, a growing body of research has turned to machine learning–based data-driven approaches, in which conditional generative adversarial networks (CGANs) are widely recognized as an effective data augmentation strategy for small-sample scenarios, facilitating the expansion of available datasets [11–13].

In order to enhance fault classification of rolling bearings, Wang et al. [14] combined a Conditional Generative Adversarial Network (CGAN) with a Stacked Autoencoder (SAE), both constructed using stacked one-dimensional fully connected layers. First, the CGAN was employed to generate artificial samples, with class labels used as conditional information to produce signals corresponding to different categories. Meanwhile, spectral normalization was introduced into the discriminator of the CGAN to enhance training stability. Subsequently, the augmented training samples were fed into the SAE for feature extraction and fault classification. Finally, experiments conducted on a bearing dataset verified the effectiveness of the proposed CGAN–SAE-based method. Zhan et al. [15] proposed a data augmentation method based on an improved Generative Adversarial Network (GAN). First, real fault data were used to train the model and obtain a mature and stable GAN. Then, the trained model was employed to test the remaining data and generate pseudo-samples. Finally, the validity of the generated samples was evaluated using two-dimensional data analysis and time-domain analysis. Experimental results demonstrated that the improved Wasserstein Generative Adversarial Network (WCGAN) exhibited superior capability in generating high-quality synthetic data and effectively addressed the poor convergence issue commonly observed in traditional GANs. Yan et al. [16] presented an AFDD approach combining a Transformer-integrated Conditional WCGAN with Deep Reinforcement Learning (TCWCGAN-DRL) for fault data synthesis and selection of high-quality samples. Despite achieving improvements in fault recognition accuracy of 4.9%, 3.66%, and 4.02% with only 15, 20, and 30 real samples, the method remains practically limited under extremely small-sample conditions. Furthermore, the training procedure is considerably time-intensive due to the inherent complexity of integrating GAN and DRL frameworks. Sun et al. [17] devised an innovative methodology to tackle the scarcity of labeled data in slewing bearing fault diagnosis, which synergizes a Time Series Generative Adversarial Network (Time GAN) and a Tabular Prior-Data Fitted Network (TabPFN). The data scarcity issue is mitigated via Time GAN, which creates synthetic samples with temporal dependencies from easily acquired audible sound signals. The TabPFN classifier then executes the fault classification without the need for hyperparameter adjustment. The method attained a 96.5% accuracy in small-sample scenarios, demonstrating notable superiority over established benchmarks and robust generalizability. Notwithstanding its promising outcomes, the approach has room for enhancement in the temporal alignment of generated data, its applicability to compound faults, and the interpretability of the model. Rathore and Harsha [18] introduced WCGAN for imbalanced data augmentation. By utilizing a Stacked Autoencoder (SAE) to derive latent features from raw vibration signals and feeding them as noise vectors into WCGAN, the method achieved higher-quality data generation than traditional approaches such as VAE and GAN. Liao et al. [19] developed a Conditional Auxiliary Class Cycle-consistent GAN with a Gradient-Penalized Wasserstein constraint (CAC-CycleGAN-WGP), which successfully generates fault signals and markedly improves the accuracy of fault diagnosis.

Huo et al. [20] introduced a fault diagnosis framework based on a Residual Mixed Self-Attention Wasserstein Conditional GAN coupled with a 1D Convolutional Neural Network (RMA-WCGAN-1DCNN). The approach demonstrates significant improvements in classification performance under imbalanced dataset conditions and achieves high-precision diagnostic outcomes. Nevertheless, owing to the experimental nature of the dataset, the technique may encounter practical challenges in industrial environments, where data acquisition is often hindered by noise and incomplete sensor measurements.

Based on the analysis of existing research, although Conditional Generative Adversarial Networks (CGANs) can generate virtual samples for dataset augmentation, most existing work validates these methods exclusively on widely adopted benchmark datasets (CWRU). The absence of comparative analyses across diverse datasets results in insufficiently demonstrated generalization capability. In addition, under conditions of limited real samples, the network is susceptible to overfitting, making it challenging for the model to meet practical performance requirements. Furthermore, the training process is often time-consuming and computationally inefficient, restricting its effectiveness in producing high-quality samples.

Therefore, this paper proposes an integrated and innovative framework to address the problem of intelligent fault diagnosis of rolling bearings under small-sample conditions. The core of the framework lies in the combination of genetic algorithm–optimized feature mapping with Wasserstein conditional generative adversarial network (WCGAN) data augmentation (GA-WT-SDP-WCGAN), enabling an end-to-end solution from feature optimization and data generation to diagnostic enhancement. Specifically, at the methodological level, the genetic algorithm is applied for the first time to optimize the parameters of wavelet transform–symmetric dot pattern (WT-SDP) feature maps, extracting highly discriminative high-order vibration features in a data-driven manner and overcoming the limitations of manual parameter tuning in traditional approaches. At the procedural level, the optimized high-quality feature maps serve as anchor samples and are deeply integrated with the improved WCGAN, enhancing the quality of generated samples and accelerating model convergence through refined loss functions and training objectives. At the validation level, cross-validation on the CWRU dataset and a self-built experimental platform demonstrates the stable generalization capability of the proposed method across different data distributions, effectively mitigating common issues in small-sample training such as overfitting, performance degradation, and domain shift. Overall, this study not only introduces an innovative and practical paradigm for rolling bearing fault diagnosis but also provides a reliable technical approach to improve the performance of intelligent diagnostic models under small-sample conditions.

2. Methodology

The proposed methodology integrates four core components into a logically coherent pipeline. First, a Genetic Algorithm (GA) optimizes the parameters of the subsequent Wavelet Transform-based Symmetrical Dot Pattern (WT-SDP) transformation to maximize feature discriminability. The optimized WT-SDP then converts 1D vibration signals into 2D feature images, which serve as the visual input for a Wasserstein Conditional Generative Adversarial Network (WCGAN). Leveraging these images, the WCGAN generates augmented, category-specific samples to mitigate data scarcity. Finally, the enriched dataset supports the robust training of downstream fault diagnosis models. This sequence ensures that each stage directly prepares the optimal input for the next, forming an end-to-end solution from feature optimization and data generation to enhanced diagnostic performance. Based on the above analysis, this study is conducted under the following working hypotheses:

(1) The GA-optimized WT-SDP feature mapping is capable of extracting highly discriminative feature representations that are insensitive to variations in operating conditions;
(2) The implicit domain-adversarial mechanism based on Wasserstein conditional adversarial learning can effectively alleviate distribution discrepancies across different operating conditions and datasets;
(3) Under small-sample conditions, data augmented through domain-adversarial generation can significantly improve the generalization performance and robustness of downstream fault diagnosis models.

2.1. Genetic Algorithm

To establish a robust foundation for feature representation, a Genetic Algorithm (GA) is first employed to optimize the key parameters of the subsequent feature mapping stage.

Genetic Algorithm (GA) [21,22], a heuristic optimization technique inspired by natural selection and genetic inheritance, operates on the principles of Darwinian evolution. By mimicking the evolutionary dynamics of species, populations iteratively improve their adaptability via selection, crossover, and mutation, thereby attaining globally optimal solutions.

The procedure is outlined as follows:

Initialization of the Population: An initial population is established through random sampling of the solution space, wherein each individual encodes a candidate solution within the search domain.

Fitness Assessment: Individuals are evaluated based on a fitness function, which quantifies the quality and effectiveness of each candidate solution.

The fitness function for the proposed algorithm is formulated as follows in Equation (1).

$$F(x) = \frac{1}{1 + f(x)}, \tag{1}$$

where $f(x)$ is the objective function value of solution x, equivalent to the logarithm of the inter-class Euclidean distance. The fitness value $F(x)$ is positively correlated with solution quality, meaning larger values indicate better solutions. The objective function is defined as follows in Equation (2).

$$f(x) = lnd(x, y) = ln\sqrt{\sum_{i=1}^{n}\left(\frac{x_i - y_i}{s_i}\right)^2}, \tag{2}$$

where x_i and y_i denote the values of the i-th feature dimension of individuals x and y, respectively. s_i represents the standard deviation of the component.

Selection Procedure: Parent individuals are selected in accordance with their fitness values, where individuals exhibiting higher fitness possess an increased likelihood of selection.

The probability formula is expressed as follows in Equation (3).

$$P(x_i) = \frac{F(x_i)}{\sum_{i=1}^{N} F(x_i)}, \tag{3}$$

In this equation, $P(x_i)$ indicates the selection probability of an individual x_i, $F(x_i)$ stands for the fitness value of the individual x_i, and N is the size of the population.

Crossover Procedure: Two parent individuals are selected, and genetic crossover is performed to produce offspring individuals, emulating the gene recombination process characteristic of biological reproduction.

Mutation Procedure: Individual genes undergo random modifications with a small probability to maintain diversity within the population and to mitigate the risk of the algorithm becoming trapped in local optimal solutions.

Replacement Procedure: The fitness of parents and offspring is evaluated, and the most fit individuals are selected to propagate into the next generation, ensuring the continued improvement of the population.

The parameter combinations optimized by the GA (the time interval l_i and the angular amplification factor ξ) are subsequently employed for WT-SDP feature mapping, ensuring that the generated two-dimensional images exhibit maximal inter-class discriminability and provide high-quality structured inputs for the WCGAN.

2.2. Wavelet Transform (WT)

Building upon the optimized parameters from the GA, the Wavelet Transform-based Symmetrical Dot Pattern (WT-SDP) method converts the raw vibration signals into two-dimensional feature images with enhanced discriminability.

The Wavelet Transform (WT) [23,24] is a time-frequency analytical method designed for non-stationary signal processing. It constructs wavelet basis functions by applying scaling and translation to a "mother wavelet," facilitating both signal decomposition and reconstruction. After wavelet transform decomposition, the original signal can be decomposed into m layers, with each layer containing A_i data points. The decomposed signal points at each layer are expressed by Equation (4).

$$X(t) = A_i + \sum_{i=1}^{m} D_i, \tag{4}$$

where $X(t)$ is the reconstructed signal, A_i is the wavelet approximation component of the i layer, representing the low-frequency part and the main signal trend retained in the current layer; constant frequency resolution. D_i is the wavelet detail component of the i layer, representing the high-frequency part, containing detailed information such as local mutations and noise.

The decomposed layers from the wavelet transform are fed into the SDP feature map, producing the WT-SDP, as illustrated in Equations (5)–(7).

$$R_i[k] = \frac{D_{in} - D_{imin}}{D_{imax} - D_{imin}}, \tag{5}$$

$$\theta_{1i}[k] = \theta + \frac{D_{i,n+1} - D_{imin}}{D_{imax} - D_{imin}}\xi, \tag{6}$$

$$\theta_{2i}[k] = \theta - \frac{D_{i,n+1} - D_{imin}}{D_{imax} - D_{imin}}\xi, \tag{7}$$

In the equation, $R_i[k]$ is the mapped radius of the wavelet detail component in the polar coordinate space. D_{in} represents the value of the n-th data point in the i-th level wavelet detail component. D_{imin}, D_{imax} represents the minimum value and maximum value of the i-th level wavelet detail component. $\theta_{1i}[k]$ denotes the counterclockwise mirrored angle associated with the i-th frequency of the k-th set of wavelet detail components. $\theta_{2i}[k]$ denotes the clockwise mirrored angle associated with the i-th frequency of the k-th set of wavelet detail components. ξ is the angular amplification factor, which controls the scaling range of the angle mapping and is optimized via the genetic algorithm.

According to the preceding formulas, the time-domain signal is decomposed using wavelet transform into approximation and detail coefficients corresponding to different frequency bands. Detailed information from each level is subsequently extracted, catego-

rized, and mapped into feature maps via the SDP method. These feature maps enhance the visualization of the signals and contribute to increased classification accuracy. The overall procedure is depicted in Figure 1.

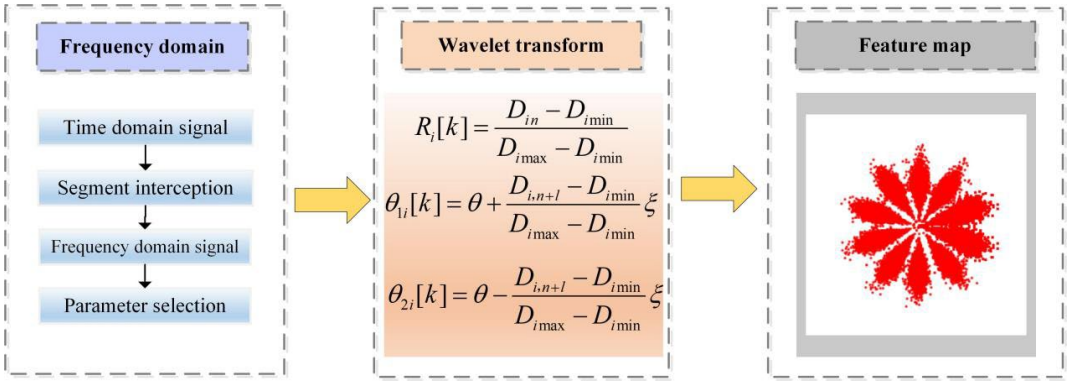

Figure 1. Wavelet transform SDP image plotting flowchart.

The feature maps obtained through WT-SDP transformation not only preserve the time–frequency characteristics of the vibration signals but also enhance visual discriminability via the symmetric dot pattern. These images are used as "real samples" and fed into the WCGAN to guide the generation of synthetic images with similar class-wise feature distributions.

2.3. Wasserstein Conditional GAN (WCGAN)

Leveraging the high-quality feature images generated by WT-SDP, a Wasserstein Conditional Generative Adversarial Network (WCGAN) is then designed to synthesize realistic and diverse samples, thereby addressing the data scarcity challenge.

The Wasserstein Conditional GAN (WCGAN) [25,26] adopts the Wasserstein distance as its loss function to facilitate improved training performance. This metric reliably assesses the discrepancy between the generated and real data distributions, even when significant differences exist. The computation is expressed in Equation (8).

$$W(P_H, P_G) = \inf_{\gamma \sim \Pi(P_H, P_G)} E_{(x,y)\sim\gamma}[\|x-y\|], \tag{8}$$

In this equation, $W(P_H, P_G)$ represents the Wasserstein distance between the real distribution P_H and the generated distribution P_G. $\Pi(P_H, P_G)$ represents the collection of all possible joint distributions formed by combining the historical distribution P_H and the generated distribution P_G. For each possible joint distribution γ, a pair of samples x and y can be obtained from the sampling of $(x, y) \sim \gamma$, and their distance $\|x-y\|$ can be calculated. The lower bound of the expected value $E_{(x,y)\sim\gamma}[\|x-y\|]$ of the sample pair distance under the joint distribution γ corresponds to the Wasserstein distance.

The WCGAN is trained using feature maps generated by the GA-WT-SDP framework. By incorporating the Wasserstein distance and a gradient penalty mechanism, the training process is stabilized, enabling the generation of synthetic samples that closely approximate the real data distribution. These samples are directly used to augment the training set, thereby enhancing the generalization capability of subsequent diagnostic models.

2.4. Model Development and Data Augmentation Pipeline

Finally, by integrating the GA-optimized feature mapping and WCGAN-based augmentation, a complete data augmentation pipeline is constructed to generate sufficient training samples for downstream fault diagnosis models.

The proposed GA-WT-SDP-WCGAN framework is implemented using TensorFlow 2.0. The framework mainly deals with the experimental measured vibration signals, as shown in Figure 2. The model consists of two main parts: a generator and a discriminator classifier.

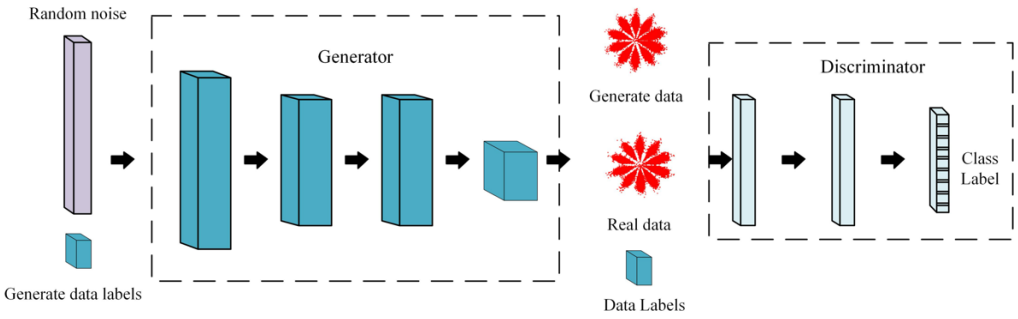

Figure 2. Structure of the WCGAN Based on GA-WT-SDP.

The process begins by feeding random noise data combined with conditional labels into the Generator, which then produces synthetic samples through multiple layers of nonlinear transformations. The generated and real samples are subsequently input into the Discriminator, which is trained to classify them accurately. Through iterative adversarial training between the Generator and Discriminator, the framework refines the generation process, progressively aligning the generated data distribution with that of the real data.

The proposed data augmentation algorithm, based on the GA-WT-SDP-enhanced Wasserstein Conditional Generative Adversarial Network, is structured around three key components, data preprocessing, model training, and generation of virtual samples, as depicted in Figure 3.

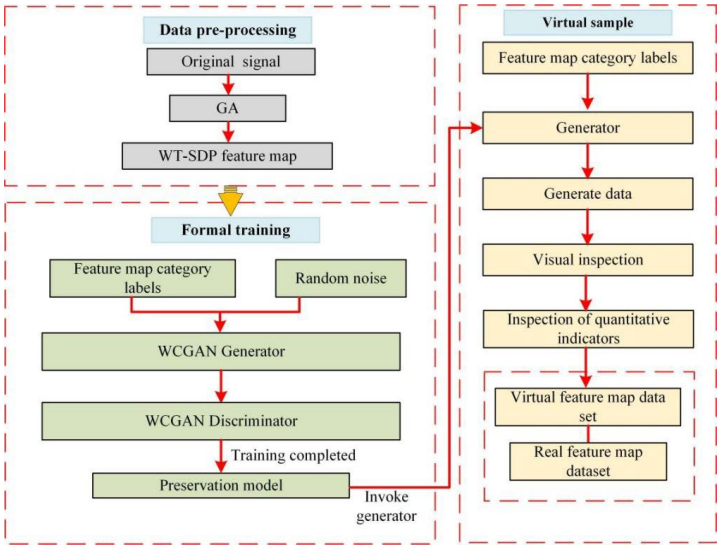

Figure 3. Data augmentation process based on GA-WT-SDP in WCGAN.

(1) Data Pre-processing: The raw dataset is segmented into samples of 2048 points each, and the SDP parameters are optimized through a Genetic Algorithm to produce the optimal WT-SDP feature maps. For each class, 10 representative feature maps are selected as training samples to accurately represent essential features, thereby enhancing the model's generalization capability and robustness under complex operational scenarios.

(2) Model Training: During the training process, the Generator and Discriminator are adversarially trained using real samples, random noise, and corresponding labels. Over repeated iterations, the Generator acquires latent representations of different

data categories, thereby enhancing the authenticity and variability of the generated samples, while the Discriminator strengthens its ability to differentiate real from synthetic data, ensuring high-quality data generation.

(3) Data Augmentation: Following training, the GA-WT-SDP-WCGAN model is preserved to facilitate efficient deployment and repeated use in downstream tasks such as sample generation, feature extraction, and fault diagnosis. The model is able to generate high-quality feature map samples under varying operational conditions, reducing the necessity for repeated training, enhancing diagnostic efficiency and robustness, and laying the groundwork for future model transfer and optimization.

3. Experiments and Analysis

3.1. Dataset Selection

Given the practical challenges in obtaining vibration data from real shearer rocker arm bearings, the present study reproduces their vibrational characteristics under different operational conditions via experimental simulation. The experimental dataset consists of radial vibration signals acquired from the Dynamic Diagnosis System (DDS) testbed at China University of Mining and Technology. Controlled noise at designated signal-to-noise ratios is introduced to increase the realism and complexity of the dataset. For further validation and comparison, the widely recognized Case Western Reserve University (CWRU) bearing fault dataset is also incorporated into the analysis. As shown in Figures 4 and 5 and Table 1.

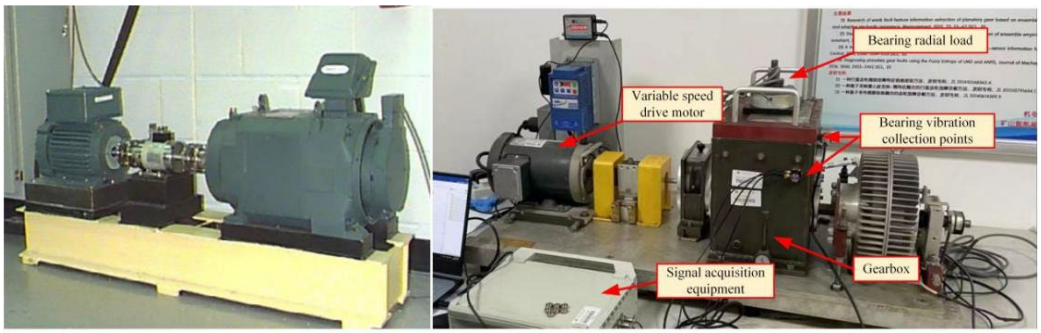

Figure 4. DDS and CWRU Testbed.

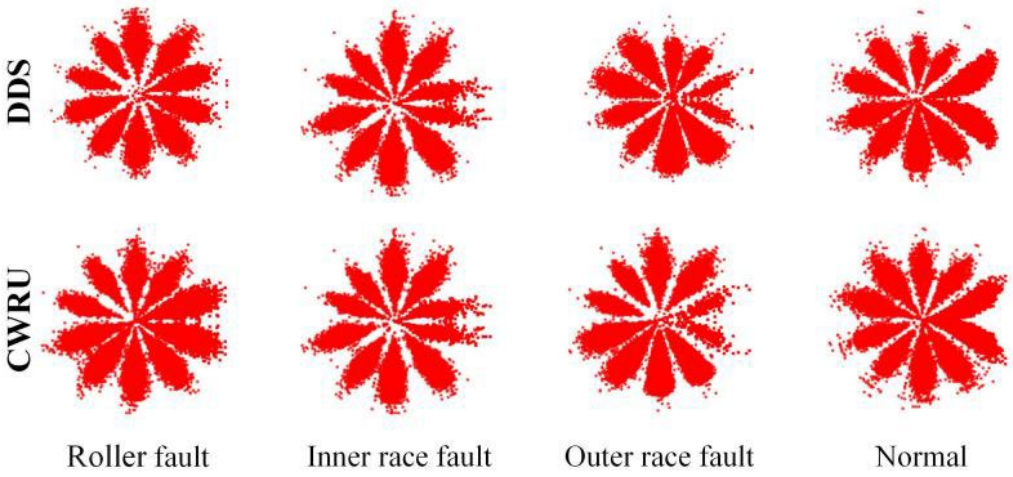

Figure 5. Example of WT-SDP dataset.

Table 1. Experimental dataset.

Dataset Codename	Load	Acquired Rotational Speed (rpm)	Normal	Outer Race	Inner Race	Roller Fault
A	3HP	1730	100	100	100	100
B	Onload	1650	100	100	100	100

3.2. Experimental Comparison and Analysis

Three sets of comparative experiments, accompanied by a detailed examination of the results, are performed to comprehensively validate the effectiveness and superiority of the proposed model for bearing fault diagnosis.

(1) Effectiveness experiments of the GA-WT-SDP feature map fusion method. The performance benefits of GA-WT-SDP in feature representation are evaluated through a comparison of multiple feature extraction and fusion strategies.

(2) Data augmentation experiments based on WCGAN. The robustness and diversity of the proposed approach for small-sample data augmentation are evaluated by analyzing the distribution properties and quality of generated samples across different augmentation methods.

(3) Fault recognition performance experiments based on generated images. Both generated and real samples are employed to train the classification model, and classification accuracy under different generation strategies is compared to validate the effectiveness of the proposed model in enhancing fault recognition performance.

3.2.1. Effectiveness Experiment of the GA-WT-SDP Feature Map Fusion Method

This section presents comparative experiments between GA-WT-SDP and ACO-WT-SDP. Firstly, GA is employed to optimize the parameters of feature maps generated by SDP. The critical parameters (*l_value* denoting the time interval l and *xi_value* representing the angular amplification factor ξ) are dynamically adjusted to enhance the quality of the feature maps. The logarithmic values of the Euclidean distances between generated images of different categories are employed to assess their quality, enabling a progressive approximation to the global optimal solution. The corresponding parameters are listed in Table 2.

Table 2. Genetic algorithm parameter settings.

Parameter	Value
Time interval	0–15
Angular amplification factor	0.1–0.628
Number of iterations	50
Optimization objective	Logarithmic value of the Euclidean distance between different categories
Crossover rate	0.7
Mutation rate	0.1
Population size	20

As illustrated in Figure 6, in the vicinity of the parameter combination ($l_{value} = 1$, $xi_{value} = 0.1$), the image quality score attains its lowest value of 9.28. The optimization process progressively converges through continuous iterations, improving the objective function value. When the parameter configuration approaches ($l_{value} = 14$, $xi_{value} = 0.5$), the image quality score achieves its highest value of 9.98. This demonstrates that the method exhibits strong global search ability and favorable convergence characteristics.

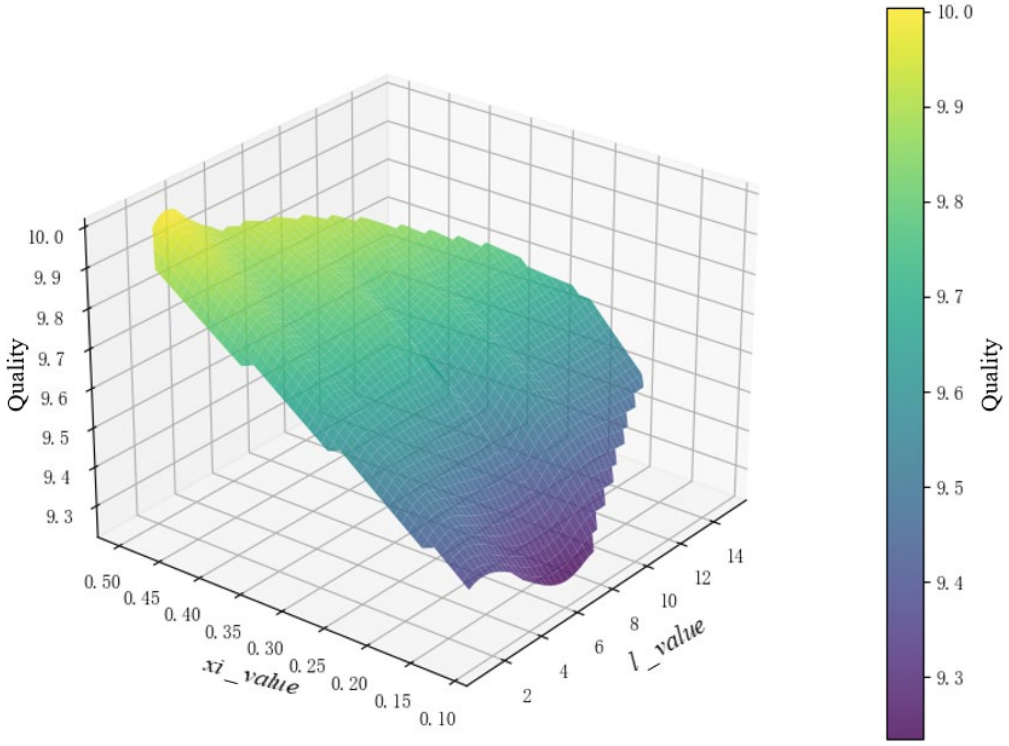

Figure 6. Iterative experimental results of genetic algorithm optimization for SDP.

Subsequently, Ant Colony Optimization (ACO) was evaluated against Genetic Algorithm (GA). By emulating ant foraging behavior, ACO efficiently explores the solution space, showing superior local search ability. The associated parameters are summarized in Table 3.

Table 3. Parameter Settings for Ant Optimization Algorithm.

Parameter	Value
Number of ants	20
Pheromone influence factor	1
Heuristic factor	2
Pheromone evaporation factor	0.1
Pheromone constant	100
Number of iterations	50

As shown in Figure 7, in the vicinity of the parameter set ($l_{value} = 3$, $xi_{value} = 0.15$), the image quality score attains the lowest value of 9.28. The optimization process progressively converges through continuous iterations, improving the objective function value. When the parameter configuration approaches ($l_{value} = 4$, $xi_{value} = 0.487$), the image quality score achieves its highest value of 9.87. This demonstrates that the method exhibits strong global search ability and favorable convergence characteristics.

In conclusion, while Ant Colony Optimization (ACO) is capable of iteratively approaching the global optimum, its convergence rate is comparatively slow, and it is susceptible to entrapment in local optima for complex, high-dimensional SDP problems, thereby reducing optimization efficiency. Conversely, Genetic Algorithm (GA) possesses remarkable global search capabilities, enabling it to overcome local optima and ACO's limitations. By integrating superior solutions over multiple generations, GA can achieve a global or near-global optimum within fewer iterations.

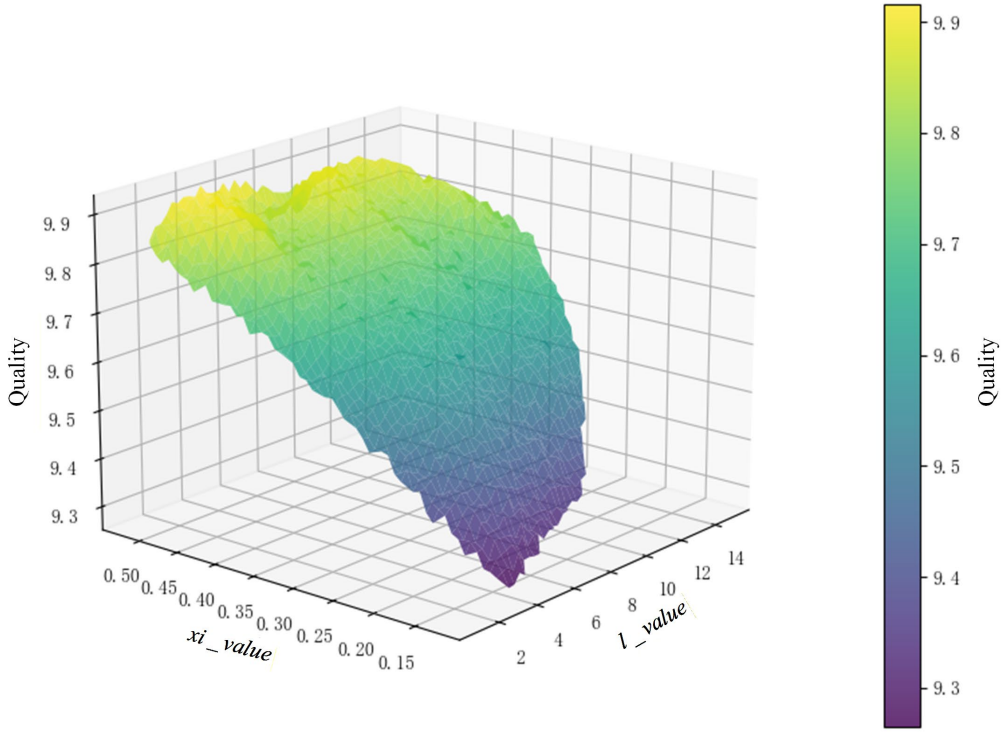

Figure 7. Iterative Optimization Results of Ant Colony Algorithm for SDP.

3.2.2. Data Augmentation Experiments Based on WCGAN

In order to evaluate the efficacy of the Wasserstein Conditional Generative Adversarial Network (WCGAN) within a transfer learning framework, a comparative analysis is carried out between the traditional conditional GAN and the enhanced WCGAN model. The evaluation highlights the quality of generated images, training stability, and convergence speed. The corresponding parameter configurations are summarized in Tables 4 and 5.

Table 4. Model parameters of GA-WT-SDP-CGAN.

Model		
CGAN	Generator learning rate	0.0005
	Discriminator learning rate	0.00006
	Loss function	Least Squares Cross-Entropy
	Number of training epochs	1000
	Number of generator network layers	4
	Number of discriminator network layers	6
	Gradient penalty coefficient ratio	/

Table 5. Model parameters of GA-WT-SDP-WCGAN.

Model		
WCGAN	Generator learning rate	0.0005
	Discriminator learning rate	0.00006
	Loss function	Wasserstein distance
	Number of training epochs	1000
	Number of generator network layers	4
	Number of discriminator network layers	6
	Gradient penalty coefficient ratio	10

The input to the generator consists of a latent vector and label information. First, the label information is embedded through a fully connected (FC) layer and reshaped into a $16 \times 16 \times 1$ tensor. Then, it is combined with the latent vector via another FC layer, mapping the noise data into a $16 \times 16 \times 128$ tensor. The generator input is formed by concatenating the mapped noise data with the embedded label tensor. During the decoding phase, the generator sequentially passes the input through multiple transposed convolution layers, each followed by a Batch Normalization (BN) layer and a ReLU activation to restore the spatial dimensions of the feature maps progressively. Finally, a 3-channel RGB image is generated. A Sigmoid activation function is applied to ensure that the pixel values are normalized within the range [0, 1], as shown in Table 6.

Table 6. Generator architecture and parameters.

Model Components	Layer Type	Output Shape	Parameters	Activation Function
Generator	Input Vector	(latent_dim,)	-	-
	Input Label	(num_labels,)	-	-
	FC Layer	(16, 16, 1)	16 ×16 × 1	-
	Transposed Convolution	(16, 16, 128)	kernel_size = (5, 5), strides = (2, 2), padding = 'same'	ReLU
	BN	(16, 16, 128)	-	-
	Transposed Convolution	(32, 32, 64)	kernel_size = (5, 5), strides = (2, 2), padding = 'same'	ReLU
	BN	(32, 32, 64)	-	-
	Transposed Convolution	(64, 64, 32)	kernel_size = (5, 5), strides = (2, 2), padding = 'same'	ReLU
	BN	(64, 64, 32)	-	-
	Transposed Convolution	(128, 128, 3)	kernel_size = (5, 5), strides = (2, 2), padding = 'same'	Sigmoid

The discriminator receives image data and corresponding label information as inputs. Similarly, labels are embedded via a fully connected (FC) layer and reshaped into a $256 \times 256 \times 1$ tensor, then merged with the input image data. The merged tensor is fed into multiple convolutional layers, where each layer employs a LeakyReLU activation to extract hierarchical image features while downsampling the spatial dimensions progressively. The network also includes a Dropout layer to prevent overfitting. The extracted features are compressed in the final fully connected layer, and a single value is output to indicate the image's authenticity, as shown in Table 7.

The CWRU dataset was adopted for data augmentation in the experimental setup, and the iterative generation results of the two models are illustrated in Figure 8.

As illustrated in Figure 8, the enhanced WCGAN, through the integration of the Wasserstein distance and a gradient penalty strategy, substantially improves training stability and alleviates prevalent issues in conventional GANs, such as gradient vanishing and gradient explosion. The generator attains stable convergence in fewer iterations, yielding higher-quality images characterized by reduced noise, smoother color transitions, and enhanced detail representation. By optimizing the Wasserstein loss, the improved WCGAN effectively broadens the distribution coverage of generated samples, preserves data diversity, and alleviates mode collapse. Consequently, it exhibits substantial advantages over conventional GANs in terms of image quality, training stability, and convergence efficiency.

Table 7. Discriminator architecture and parameters.

Model Components	Layer Type	Output Shape	Parameters	Activation Function
Discriminator	Input Image	(256, 256, 3)	-	-
	Input Label	(num_labels,)	-	-
	FC Layer	(256, 256, 1)	256 × 256 × 1	-
	Conv Layer	(128, 128, 128)	kernel_size = (3, 3), strides = (2, 2), padding = 'same'	LeakyReLU
	Conv Layer	(64, 64, 64)	kernel_size = (3, 3), strides = (2, 2), padding = 'same'	LeakyReLU
	Conv Layer	(32, 32, 32)	kernel_size = (3, 3), strides = (2, 2), padding = 'same'	LeakyReLU
	Conv Layer	(16, 16, 16)	kernel_size = (3, 3), strides = (2, 2), padding = 'same'	LeakyReLU
	Dropout Layer	(16, 16, 16)	-	-

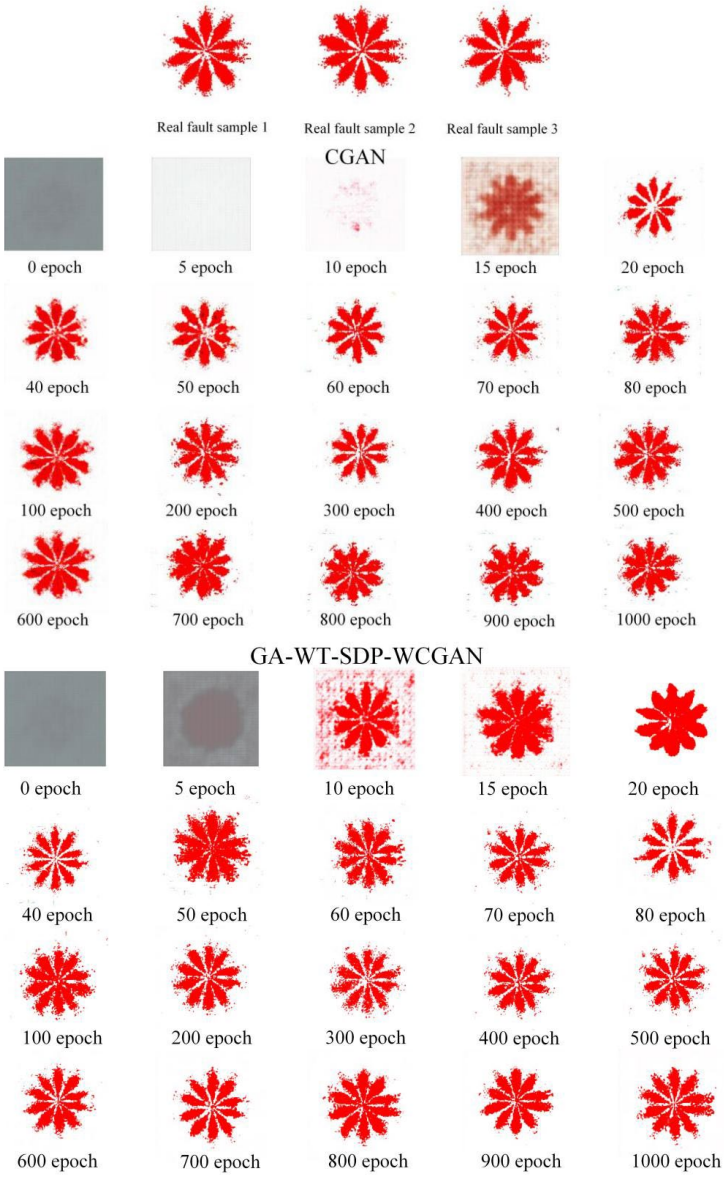

Figure 8. Generator generation results.

To further evaluate the similarity between synthesized and real images, this section introduces the Structural Similarity (SSIM) index as a quantitative metric for assessing the structural consistency of generated images relative to authentic images. This metric effectively quantifies the similarity between images in terms of luminance, contrast, and structural features, providing an intuitive and reliable basis for evaluating the performance of the generative model. The corresponding experimental results are presented in Figures 9 and 10.

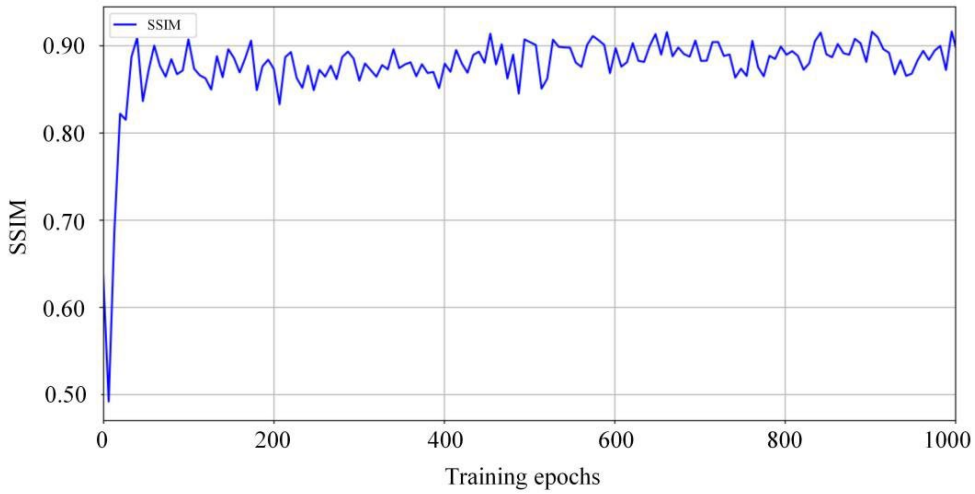

Figure 9. Improved WCGAN bearing fault dataset generated sample SSIM curve.

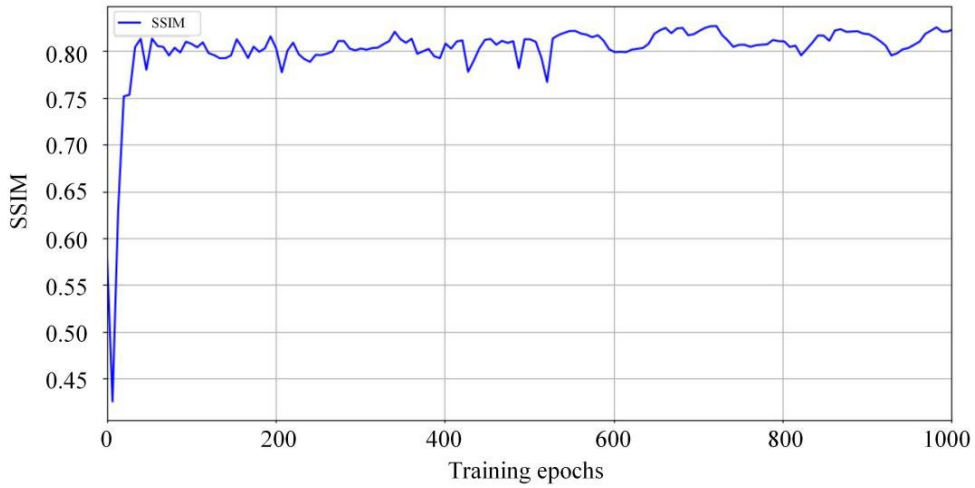

Figure 10. CGAN bearing fault dataset generated sample SSIM curve.

As illustrated in Figures 9 and 10, the SSIM value of the CGAN model steadily decreases from 0.80 to 0.45 as training epochs progress from 200 to 800, indicating a continuous loss of structural fidelity in the generated images and highlighting clear signs of overfitting and performance deterioration. A particularly notable drop to 0.60 is observed at 600 epochs, followed by a further decrease to 0.45 at 800 epochs, highlighting the difficulty that conventional CGAN models face in maintaining consistent output quality during extended training. In contrast, the Wasserstein Conditional Generative Adversarial Neural Network (WCGAN) model based on GA–WT–SDP demonstrates superior structural preservation capability during training. Its SSIM value consistently fluctuates within the range of 0.9 to 0.5, which is significantly higher than that of the CGAN model. The model retains an SSIM above 0.5 throughout 1000 training epochs, demonstrating high robustness under challenging training conditions. Furthermore, the model rapidly achieves a high

similarity of SSIM > 0.8 even in the early training stage (200–400 epochs), demonstrating its strong capability in capturing structural features of images.

3.2.3. Fault Recognition Performance Experiments Based on Generated Images

This section presents the relevant experiments to further validate the practical effectiveness of the generated samples in enhancing fault diagnosis performance.

First, the t-SNE algorithm is employed in this section to perform nonlinear dimensionality reduction and visualization of the augmented image dataset in order to examine the distribution of generated samples and real samples in the feature space. By applying the t-SNE dimensionality reduction method, high-dimensional features are mapped into a two-dimensional space, allowing intuitive visualization of the clustering patterns and class separability among different fault categories, thereby enabling an evaluation of the effectiveness of generated samples in enhancing the original data distribution and improving class boundaries. In the original dataset, each class contains 20 baseline samples. Data augmentation is performed using the proposed GA-WT-SDP-based Wasserstein conditional generative adversarial network, increasing the effective number of samples per class to 40. In the t-SNE visualization, the embedding dimension is set to 3, the perplexity is fixed at 30, and the learning rate is set to 200. The algorithm converges after 100 iterations, and the corresponding analysis results are shown in Figure 11.

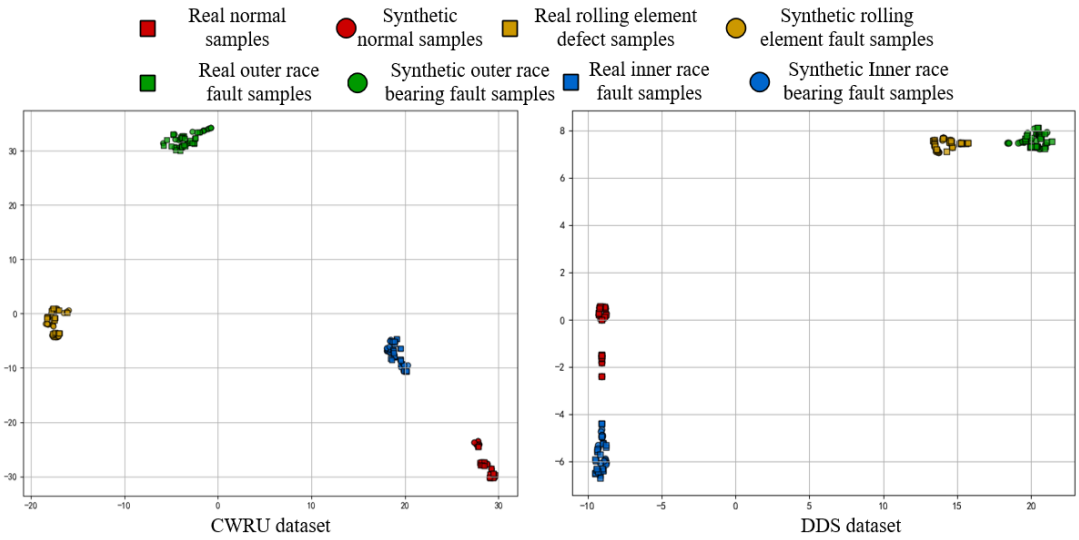

Figure 11. Analysis of t-SNE for GA-TW-SDP feature maps.

The visualization results indicate that the average intra-class silhouette coefficient reaches 0.68 ± 0.05, which is significantly higher than that of the dataset before augmentation (0.52 ± 0.07). Further analysis shows that the intra-class Mahalanobis distance between generated–real sample pairs (0.41 ± 0.09) accounts for only 29.9% of the inter-class distance (1.37 ± 0.12), confirming that the augmented data strictly preserve the class-discriminative characteristics of the original dataset. These results validate the effectiveness of the proposed data augmentation strategy in feature-space alignment and provide reliable augmented data for small-sample learning tasks.

Moreover, by augmenting the original dataset with varying numbers of generated images, the recognition performance of four representative classification models—Support Vector Machine (SVM), Least Squares SVM (LS-SVM), Convolutional Neural Network (CNN), and Long Short-Term Memory network (LSTM)—was comparatively analyzed. Independent validation sets constructed from the DDS and CWRU datasets were used for evaluation, and the results are presented in Figures 12 and 13.

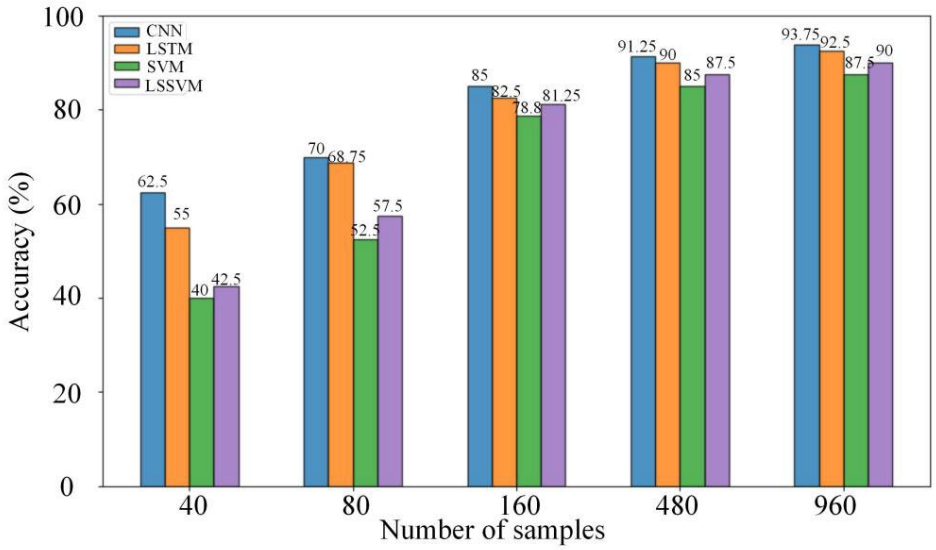

Figure 12. DDS: Classification accuracy under different numbers of generated samples.

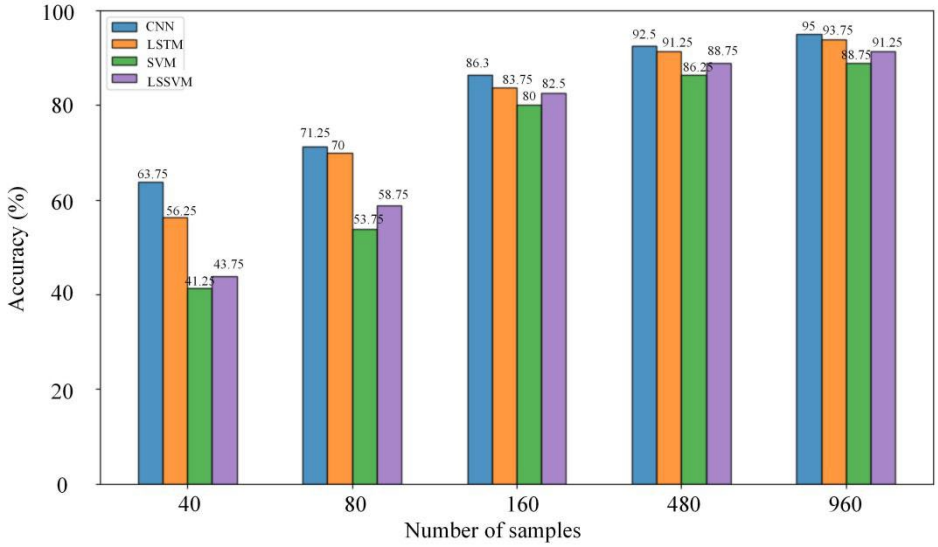

Figure 13. CWRU: Classification accuracy under different numbers of generated samples.

The results indicate that increasing the sample size leads to a substantial improvement in accuracy across all models. For the DDS dataset, accuracy rose from 62.5% to 93.75% for CNN, 55.0% to 92.5% for LSTM, 40.0% to 87.5% for SVM, and 42.5% to 90.0% for LS-SVM. On the CWRU dataset, CNN increased from 63.75% to 95.0%, LSTM from 56.3% to 93.75%, SVM from 41.25% to 88.75%, and LS-SVM from 43.75% to 91.25%. CNN and LSTM exhibited the highest performance, indicating that deep models possess superior feature learning capabilities with larger datasets. Across all sample sizes, the CWRU dataset consistently achieved higher classification accuracy than DDS. The difference is particularly pronounced for small sample sizes but persists as the dataset grows, suggesting that CWRU features are clearer and less noisy, facilitating model training and convergence.

To further investigate model performance under limited sample conditions, this experiment visualized the recognition results for 40 and 960 samples using confusion matrices, as illustrated in Figures 14–17.

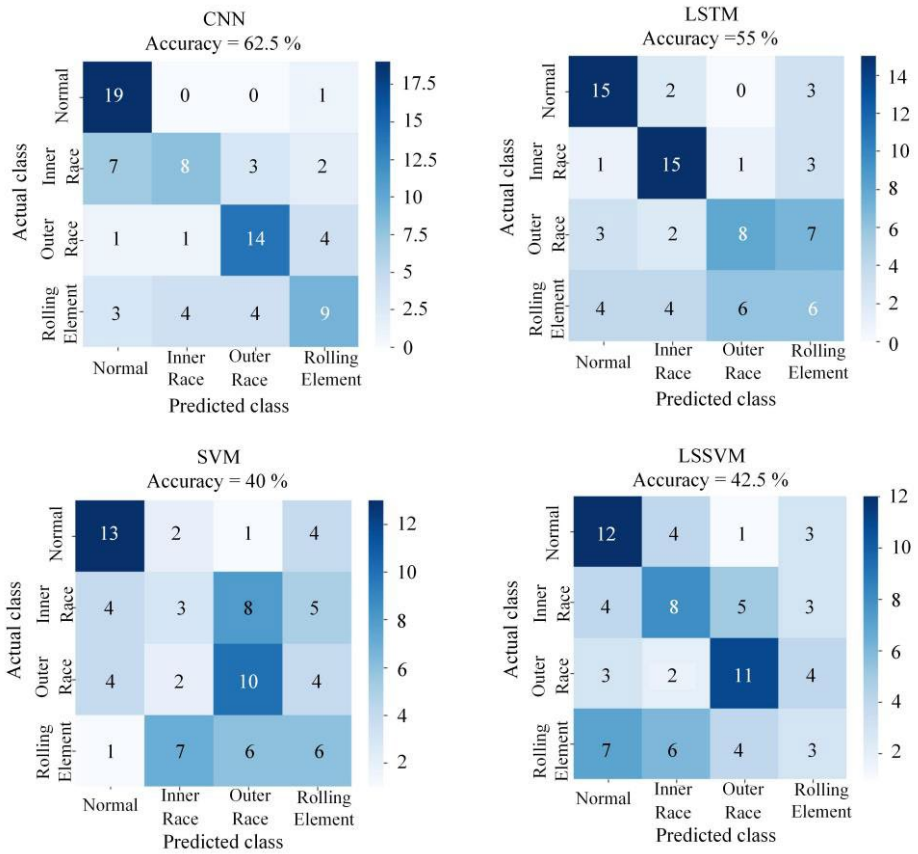

Figure 14. DDS: Confusion matrix of validation set with 40 training samples.

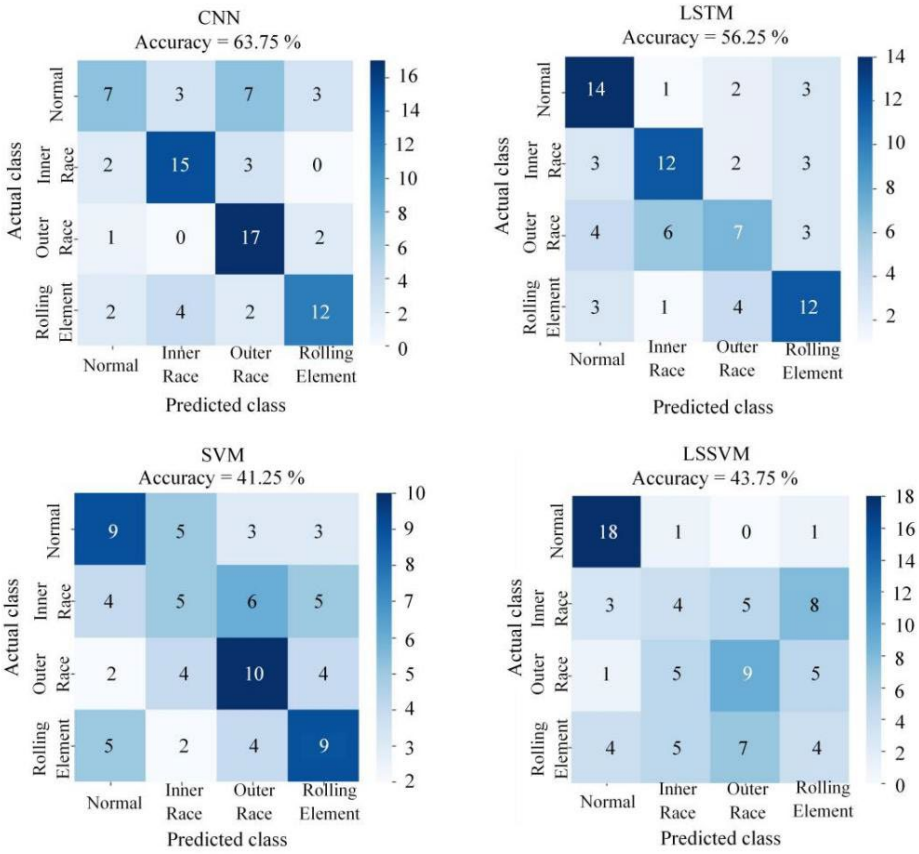

Figure 15. CWRU: Confusion matrix of validation set with 40 training samples.

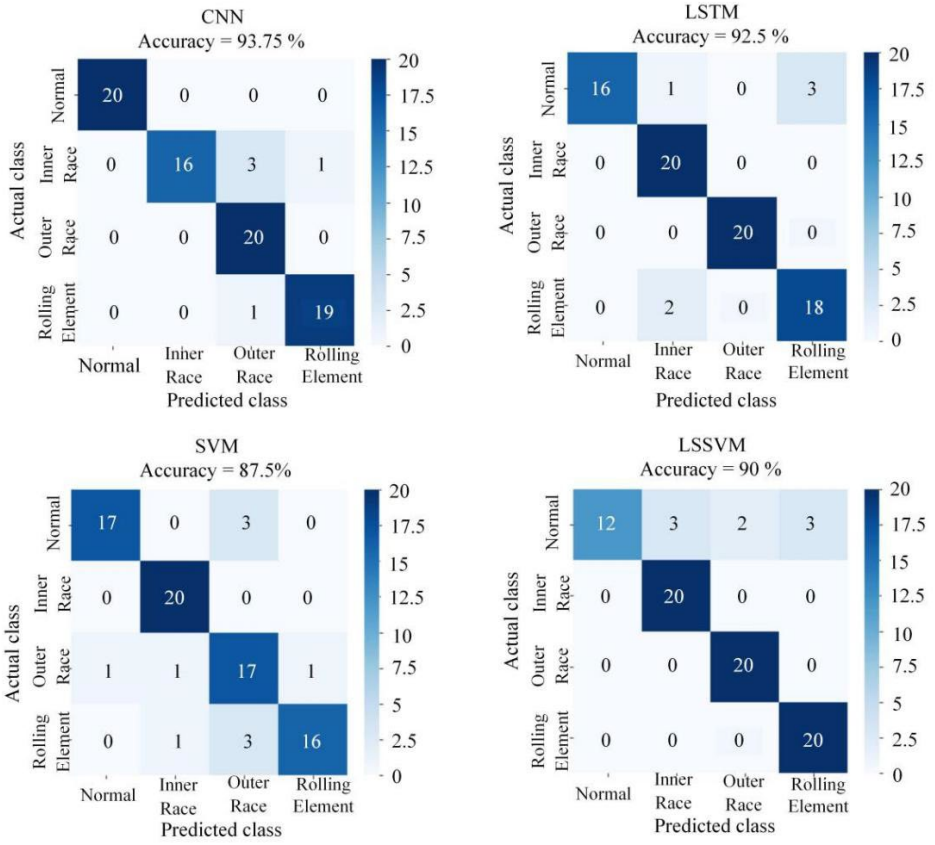

Figure 16. DDS: Confusion matrix of validation set with 960 training samples.

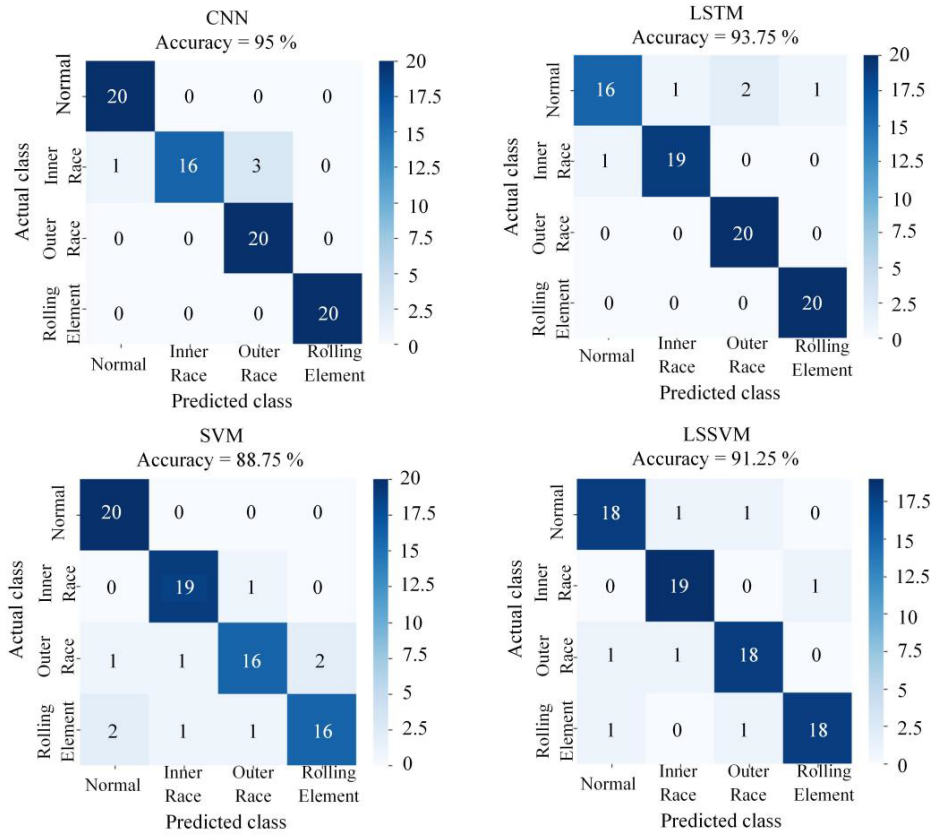

Figure 17. CWRU: Confusion matrix of validation set with 960 training samples.

When the number of training samples is 40, all models exhibit relatively low recognition accuracy, particularly for classes with uneven distribution or high complexity, making complex categories difficult to identify. This is primarily because the limited sample size restricts feature learning, hindering the extraction of effective discriminative information and resulting in blurred classification boundaries. Moreover, small sample conditions easily lead to overfitting or underfitting, especially when classes are imbalanced or noise levels are high, making the models highly sensitive to anomalous samples and reducing stability and robustness. This demonstrates the poor adaptability of traditional supervised learning models in data-scarce scenarios and highlights the importance of few-shot learning and transfer learning approaches.

When the sample size increased to 960, the recognition accuracy of all models improved significantly to over 85%. This indicates that sufficient data enables the models to better learn discriminative features between classes, suppress overfitting, enhance feature representation, and improve generalization performance.

In conclusion, the performance of traditional fault diagnosis models in complex, multi-class tasks is highly data-dependent, with larger datasets improving the recognition of boundary and minority samples, as well as enhancing classification balance and robustness.

4. Conclusions

To address the challenges of acquiring shearer rocker arm bearing fault data and the scarcity of labeled samples, this paper proposes a data augmentation approach based on the GA-WT-SDP with Wasserstein Conditional Generative Adversarial Neural Network (WC-GAN). The WT-SDP feature map parameters were optimized using a Genetic Algorithm, and the incorporation of the Wasserstein distance with a gradient penalty mechanism substantially improved the quality of generated samples and training stability. The method was validated on the CWRU and DDS datasets, leading to the following conclusions:

(1) The GA-WT-SDP approach demonstrates superior adaptability and discriminative power in feature extraction. Compared with the Ant Colony Optimization (ACO) algorithm, the Genetic Algorithm (GA) offers enhanced global search capability and faster convergence.

(2) WCGAN incorporates the Wasserstein distance and a gradient penalty mechanism to enhance training stability, alleviating the gradient vanishing and exploding issues commonly observed in conventional GANs. The generator achieves stable convergence with fewer iterations, producing higher-quality images with reduced noise, more natural color transitions, and finer details.

(3) The experimental results indicate that even using a limited number of synthetic samples generated solely by this framework for data augmentation, the recognition performance of various diagnostic models can be significantly and consistently improved. This confirms that the data generated by GA-WT-SDP-WCGAN possesses a high degree of authenticity and information fidelity. Such augmented data not only effectively expands the sample space and alleviates overfitting under small-sample conditions, but, more importantly, it substantially enhances the models' ability to represent and discriminate complex fault patterns and class boundaries.

Author Contributions: Conceptualization, Z.W.; writing—original draft preparation, S.W.; methodology, C.L.; software, H.W.; validation, J.Y.; resources, Y.P.; supervision, G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (No. 52304179).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: All authors have no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACO	Ant Colony Optimization
BN	Batch Normalization layer
CWRU	Case Western Reserve University
CGAN	Conditional Generative Adversarial Network
CNN	Convolutional Neural Network
DDS	Dynamic Diagnosis System
FC	Fully connected layer
GA	Genetic Algorithm
LSTM	Long Short-Term Memory Network
LS-SVM	Least Squares Support Vector Machine
SDP	Symmetrical Dot Pattern
SSIM	Structural Similarity
SVM	Support Vector Machine
WT	Wavelet Transform
WCGAN	Wasserstein Conditional Generative Adversarial Network

References

1. Yan, R.; Shang, Z.; Xu, H.; Wen, J.; Zhao, Z.; Chen, X.; Gao, R. Wavelet transform for rotary machine fault diagnosis:10 years revisited. *Mech. Syst. Signal Process.* **2023**, *200*, 110545. [CrossRef]
2. Liu, Y.; Jiang, H.; Yao, R. Counterfactual-augmented few-shot contrastive learning for machinery intelligent fault diagnosis with limited samples. *Mech. Syst. Signal Process.* **2024**, *16*, 11507. [CrossRef]
3. Mali, A.R.; Shinde, P.V.; Patil, A.P.; Salunkhe, V.G.; Desavale, R.G.; Jadhav, P.S. A Novel Method for Bearing Fault Diagnosis Based on Novel Feature Sets with Machine Learning Technique. *ASME. J. Tribol.* **2025**, *147*, 24301. [CrossRef]
4. Liu, Y.; Jiang, H.; Liu, C. Data-augmented wavelet capsule generative adversarial network for rolling bearing fault diagnosis. *Knowl.-Based Syst.* **2022**, *252*, 1094. [CrossRef]
5. Ruan, D.; Zhang, F.; Zhang, L. Optimal modifications in CNN for bearing fault classification and adaptation across different working conditions. *J. Vib. Eng. Technol.* **2024**, *12*, 4075–4095. [CrossRef]
6. Liu, G.; Wu, L. Incremental bearing fault diagnosis method under imbalanced sample conditions. *Comput. Ind. Eng.* **2024**, *192*, 110203. [CrossRef]
7. Salunkhe, V.G.; Khot, S.M.; Yelve, N.P.; Jagadeesha, T.; Desavale, R.G. Rolling Element Bearing Fault Diagnosis by the Implementation of Elman Neural Networks with Long Short-Term Memory Strategy. *ASME J. Tribol.* **2025**, *147*, 84301. [CrossRef]
8. Qin, Y.; Fu, X.; Li, X. ADAMS simulation and HHT feature extraction method for bearing faults of coal shearer. *Processes* **2024**, *12*, 164. [CrossRef]
9. Lu, J.; Wu, W.; Huang, X. A modified active learning intelligent fault diagnosis method for rolling bearings with unbalanced samples. *Adv. Eng. Inform.* **2024**, *60*, 102397. [CrossRef]
10. Fu, W.; Jiang, X.; Tan, C. Rolling bearing fault diagnosis in limited data scenarios using feature enhanced generative adversarial networks. *IEEE Sens. J.* **2022**, *22*, 8749–8759. [CrossRef]
11. Dai, J.; Wang, J.; Yao, L. Categorical feature GAN for imbalanced intelligent fault diagnosis of rotating machinery. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3525212. [CrossRef]
12. Fu, W.; Chen, Y.; Li, H. Imbalanced fault diagnosis using conditional Wasserstein generative adversarial networks with switchable normalization. *IEEE Sens. J.* **2023**, *23*, 29119–29130. [CrossRef]
13. Liu, J.; Xu, K.; Cai, B.; Guo, Z. Fault Prediction of On-Board Train Control Equipment Using a CGAN-Enhanced XGBoost Method with Unbalanced Samples. *Machines* **2023**, *11*, 114. [CrossRef]
14. Wang, J.; Han, B.; Bao, H. Data augment method for machine fault diagnosis using conditional generative adversarial networks. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2020**, *234*, 2719–2727. [CrossRef]
15. Zhan, L.; Xu, X.; Qiao, X. Fault data expansion method of permanent magnet synchronous motor based on Wasserstein-generative adversarial network. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2022**, *238*, 6242–6255. [CrossRef]

16. Yan, L.; Lu, C.; Ma, X. Intelligent fault diagnosis for air handing units based on improved generative adversarial network and deep reinforcement learning. *Expert Syst. Appl.* **2024**, *240*, 122545. [CrossRef]
17. Sun, L.; Wu, J.; Wang, J. Fault diagnosis of slewing bearing using audible sound signal based on time generative adversarial network–TabPFN method. *J. Vib. Acoust.-Trans. ASME.* **2025**, *147*, 41002. [CrossRef]
18. Rathore, M.; Harsha, S. Intelligent fault detection scheme for rolling bearing based on generative adversarial network and autoencoders using convolutional neural network. *J. Vib. Eng. Technol.* **2024**, *12*, 8979–8991. [CrossRef]
19. Liao, W.; Wu, L.; Xu, S. A novel approach for intelligent fault diagnosis in bearing with imbalanced data based on cycle-consistent GAN. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 3525416. [CrossRef]
20. Huo, J.; Qi, C.; Li, C. Data augmentation fault diagnosis method based on residual mixed self-attention for rolling bearings under imbalanced samples. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3528914. [CrossRef]
21. Can, H.; Xing, F.; Pan, S. Fault diagnosis of rolling bearings based on variational mode decomposition and genetic algorithm-optimized wavelet threshold denoising. *Machines* **2022**, *10*, 649. [CrossRef]
22. Fan, Q.; Yu, F.; Xuan, M. Transformer fault diagnosis method based on improved whale optimization algorithm to optimize support vector machine. *Energy Rep.* **2021**, *7*, 856–866. [CrossRef]
23. Zhang, C.; He, Y.; Yang, T. An analog circuit fault diagnosis approach based on improved wavelet transform and MKELM. *Circuits Syst. Signal Process.* **2022**, *41*, 1255–1286. [CrossRef]
24. Zeng, Q.; Liu, X.; Zhu, X. Data-driven structural topology optimization method using conditional Wasserstein generative adversarial networks with gradient penalty. *Comput. Model. Eng. Sci.* **2024**, *141*, 2065–2085. [CrossRef]
25. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. *J. Mach. Learn. Res.* **2017**, *70*, 214–223.
26. Ferracuti, F.; Freddi, A.; Monteriù, A. Fault diagnosis of rotating machinery based on Wasserstein distance and feature selection. *IEEE Trans. Autom. Sci. Eng.* **2022**, *19*, 1997–2007. [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.