

Article

Generative Dual-Modal Data Augmentation for Motor Fault Diagnosis Under Sample Imbalance

Ganxin Jie ^{1,4}, Cailiang Zhang ^{2,3,*}, Junqing Ma ⁵, Yang Yang ^{1,4} and Chuan Chen ^{1,4}¹ State Key Laboratory of Environmental Adaptability for Industrial Products, China National Electric Apparatus Research Institute Co., Ltd., Guangzhou 510663, China² Yangjiang Offshore Wind Power Laboratory, Yangjiang 529500, China³ School of Software & Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang 330013, China⁴ General Machinery Key Core Components Innovation Center (Anhui) Co., Ltd., Hefei 230094, China⁵ Ocean College, Zhejiang University, Zhoushan 316021, China

* Correspondence: zhangcl1027@163.com

Abstract

This study investigates class imbalance in motor fault diagnosis. Fault samples, especially those at different severity levels, are often much fewer than healthy samples. To address this issue, a self-attention guided Wasserstein conditional GAN with gradient normalization (SWGAN) is proposed. The method is based on synchronized three-phase current and vibration measurements. It separately generates label-conditioned current spectra and vibration spectra to supplement minority fault classes. Self-attention is used to capture long-range spectral dependencies. Gradient normalization is introduced to improve adversarial training stability. The generated current and vibration spectra are then fused at the feature level and fed into a stacked autoencoder (SAE)-based multi-modal classifier. Experiments were conducted on a PMSM stator fault dataset and a variable-speed three-phase asynchronous motor dataset. On the PMSM dataset, SWGAN achieved highest accuracies of 98.90% and 97.81% under two fault-category imbalance settings. On the variable-speed motor dataset, the proposed method achieved accuracies of 98.10% and 97.65%, respectively. These results show that SWGAN can provide effective supplementary samples for minority fault classes. They also indicate that the proposed method improves diagnostic performance under both fixed-speed and variable-speed conditions.

Keywords: fault diagnosis; electric motor; conditional generative adversarial networks; gradient normalization; self-attention mechanism; multi-source fusion



Academic Editors: Athanasios Karlis and Tarek Berghout

Received: 12 April 2026

Revised: 8 May 2026

Accepted: 11 May 2026

Published: 1 June 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Fault diagnosis and maintenance operations of electric motors are crucial for the safe and stable operation of industrial production machinery [1]. Therefore, many scholars have increasingly paid attention to the field of motor fault diagnosis, such as the motor failure of unmanned aerial vehicles [2,3], wind turbines [4], etc. Among them, rapid condition monitoring and accurate fault identification for motors are crucial [5]. As a result of advancements in computer technology, data-driven algorithms such as neural networks [6], support vector machines [7], clustering analysis [8] and autoencoders [9] have significantly expanded the range of fault diagnosis techniques available for analyzing electric motors. Deep learning-based methodologies exhibit robust feature-learning capabilities, through

the construction of deep networks that capture intricate internal data representations, consequently leading to precise fault diagnosis outcomes [10].

Recent research has extensively focused on data-driven motor fault diagnosis methods, with various proposed approaches such as time–frequency dual-domain contrast and fusion (TFDDCF) [11], a motor fault diagnosis method based on a spiking convolutional neural network with multi-scale decomposition local features [12]; a novel modified conditional generative adversarial network (MCGAN) based on memristive hyperchaotic sequences and a mixed-dimensional convolutional neural network (MCNN) [13]; notch filter augmentation-based multi-channel self-supervised learning (NFA-MSSL) for motor fault diagnosis [14]; and self-supervised learning model that leverages time–frequency dual-domain prediction (TFDDP) for bearing fault diagnosis within motor systems [15]. Many deep learning algorithms are designed to work with balanced datasets and uniform class distributions. They typically require a substantial amount of fault data for training and adhere strictly to the distribution of training and testing datasets [16]. However, in practical engineering settings, motors rarely operate in a failure state, making the acquisition of motor fault data both costly and risky. As a result, data collected during normal motor operation usually greatly outnumber fault data, and motor faults tend to be diverse and random, with some rare fault types represented by only a few samples in the dataset. Over time, variations in motor operating conditions and environments can cause shifts in data distribution, leading to inconsistencies that adversely affect the performance of deep learning models. In the field of motor fault diagnosis, sample imbalance is therefore a prevalent and challenging issue. Moreover, modern motor monitoring systems typically integrate multiple sensors to capture complementary equipment states. While fusing such multi-source data can provide richer fault signatures and partially alleviate the scarcity of samples for certain fault classes, it also introduces significant challenges, including heterogeneity across data modalities, misalignment of different sampling rates, and distribution gaps among sensor sources, which further complicate the development of reliable diagnostic models.

To exploit the useful complementary information provided by multiple sensors, a range of deep fusion architectures have been investigated for motor fault diagnosis. Yi et al. [17] proposed a novel cross-modal vector fusion induction motor fault diagnosis and classification model (CNN-ELMNet), which not only preserves the model's FE capability but also enhances the convergence speed and classification recognition rate of the model. Dehnavi et al. [18] presented a comprehensive approach for data measurement and fault diagnosis of motor systems using novel model-based and data-driven strategies and data fusion techniques. Li et al. [19] proposed a novel hydraulic motor fault diagnosis method based on weighted fusion of multi-channel data and migration learning to solve low data utilization, poor reliability, and insufficient generalization ability under variable working conditions. Huang [20] proposed an interpretable convolutional neural network with multiple wavelet kernel fusion, which includes a novel superlet convolutional layer and an effective channel attention module. Nevertheless, the vast majority of existing fusion-based diagnosis methods are evaluated under the assumption of balanced class distributions [21]. The interaction between severe sample imbalance and multi-modal fusion remains largely unexamined; minority fault features risk being suppressed in the fused representation, offsetting the benefits of sensor diversity. Consequently, developing a robust motor fault diagnosis method that can handle imbalanced samples and effectively fuse heterogeneous multi-sensor data is essential.

Numerous scholars have employed oversampling methods to address the challenge of data imbalance. These techniques initiate from minority class samples, identify adjacent samples, and synthesize new minority class samples to achieve a balance with the

majority class samples. Wang et al. [22] enhanced the original SMOTE method by introducing boundary enhancement and a Gaussian mixture model. Wei et al. [23] proposed a novel multi-class imbalanced fault diagnosis method utilizing the Sample-Characteristic Oversampling Technique (SCOTE) and multi-class LS-SVM. Han et al. [24] put forward a feature-weighted oversampling method named AFS-O (Attention Features Selection Oversampling Technique), applied to the synthesis of accelerated life data for rolling bearings. Tian et al. [25] proposed a weighted modified conditional variational auto-encoder (WM-CVAE) as a novel data augmentation technique to improve the data quality and learning capability in class-imbalance fault diagnosis.

In 2014, Goodfellow et al. [26] introduced generative adversarial networks (GANs) as a way to generate new samples in unsupervised learning while capturing data distributional properties. Since then, various GAN variants have emerged [27,28] and found applications in image recognition [29,30]. Recently, GAN models have been extended to mechanical equipment fault diagnosis. For example, Dixit et al. [31] proposed a GAN framework combined with model-agnostic meta-learning for conditionally assisted classifiers. Chen et al. [32] proposed a cross-domain fault diagnosis method based on an enhanced unsupervised generative adversarial network (GAN) and fine-tuning techniques by employing an improved adversarial loss aligned with the statistical characteristics of real data. Liu et al. [33] proposed an improved multiscale residual-based GAN and a feature enhancement-driven capsule network for imbalanced fault diagnosis. Wang et al. [34] introduced a Fourier transform-like transform based on a self-encoder and proposed an FTGAN method for aligning time-domain and frequency-domain distributions of real and generated data. Liu et al. [35] integrated the Haar wavelet into GAN, forming a Wavelet Transform GAN. They replaced some CNNs in WTGAN with capsule networks, creating a Wavelet Capsule Generative Adversarial Network.

GAN-based methods have been widely used to generate minority fault samples and reduce class imbalance in fault diagnosis. However, their performance remains limited in practical motor applications. Most existing studies are conducted under steady operating conditions, whereas real motor systems often operate with speed changes, load fluctuations, and environmental disturbances. These factors make current and vibration signals highly non-stationary. As a result, conventional GANs may generate samples with similar statistical characteristics, but they may fail to preserve fault-related temporal patterns and global dependencies. In addition, many GAN-based methods focus on a single signal source. Motor faults are reflected in both electrical and mechanical responses. Current signals are more sensitive to electromagnetic abnormalities, while vibration signals describe mechanical degradation and structural response. Generating only one modality may lead to incomplete fault information. Generating current and vibration samples separately may also weaken their cross-modal consistency. Training stability is another concern. Traditional GANs are prone to mode collapse and poor convergence. Although Wasserstein distance has been introduced to improve adversarial training [36], the Lipschitz constraint still needs to be enforced. Spectral normalization (SN) [37] constrains each network layer and may reduce model flexibility. Gradient penalty (GP) [38] is more flexible, but it introduces additional regularization and may affect training efficiency. These limitations become more evident when minority fault samples are scarce and operating conditions vary.

This study proposes a GAN-based current–vibration fusion framework for imbalanced motor system fault diagnosis. The method is developed to address two issues in practical diagnosis: the limited number of fault samples and the insufficient use of complementary electrical and mechanical information. In the proposed framework, stator current spectra and vibration envelope spectra are first extracted from synchronized measurements. A self-attention guided Wasserstein conditional GAN with gradient normalization (SWGAN) is

then used to generate label-conditioned samples for minority fault classes. The augmented current–vibration features are finally fed into an SAE-based classifier for fault identification. Compared with existing GAN-based fault diagnosis methods that mainly focus on single-modal data generation or steady operating conditions, the proposed framework considers class imbalance, current–vibration feature fusion, and training stability under both fixed-speed and variable-speed motor conditions. The main contributions are summarized as follows:

- (1) An SWGAN-based data augmentation method is proposed for imbalanced motor stator fault diagnosis. Unlike conventional oversampling methods and single-modal generation schemes, the proposed model generates both stator current spectra and vibration envelope spectra. This design supplements minority fault classes while retaining electrical and mechanical fault information.
- (2) Self-attention and gradient normalization are incorporated into the Wasserstein conditional GAN. The self-attention module improves the modeling of global dependencies in fault-related spectral features. Gradient normalization improves adversarial training stability without imposing strict layer-wise constraints. These designs help improve the quality and consistency of the generated minority class samples.
- (3) A current–vibration fusion diagnosis scheme is constructed on the augmented feature space and validated on both the fixed-speed PMSM stator fault dataset and the variable-speed asynchronous motor dataset. Results under different imbalance ratios, fault types, fault severities, sensor combinations, and operating conditions show that the proposed method outperforms conventional data augmentation methods and single-modal diagnosis schemes.

Section 2 introduces the theoretical background of the proposed method. Section 3 presents the SWGAN-based current–vibration fusion framework and its training procedure. Section 4 describes the experimental datasets and validation results, including the PMSM stator fault dataset and the variable-speed asynchronous motor dataset. Section 5 provides comparative analysis, ablation studies, and discussions on the diagnostic performance of the proposed method. For Section 6, the conclusions and findings are succinctly summarized.

2. Theoretical Background

2.1. Conditional Generative Adversarial Networks

GANs are implemented by making the discriminator(D) and generator(G) influence and game each other. Specifically, the goal of G is to generate the input random noise vectors as much as possible to have a similar distribution to the target data, while D is responsible for distinguishing the similar data generated by G from the target data. In a continuous iterative journey, both sides will keep optimizing until the Nash equilibrium is reached. Its overall training objective function is shown below:

$$\min_G \max_D \mathbb{E}_{x \sim p_r(x)} [\log(D(x))] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

where $p_r(x)$ is the pattern of original data points and p_z is the pattern of the stochastic noise vector. $D(*)$ and $G(*)$ represent the output of the discriminator and generator respectively.

The unconditionally constrained traditional GAN can only perform unsupervised learning and cannot control the pattern of data it generates; however, it is possible to guide the process of data generation by constraining the model with additional information. Based on this we introduce conditional information on top of the GAN to assist in controlling the effect of data generation, i.e., conditional GAN (CGAN). Figure 1 demonstrates the specific difference from the traditional GAN. Specifically, we add a condition, i.e., additional

information y , on top of the G-generated data and D-discriminated data, where y adopts the category label c . Additional information y is thrown into G and D as an additional input layer, so that the original unsupervised learning network becomes a supervised learning network, which is theoretically capable of generating or discriminating different categories of sample data in a better way.

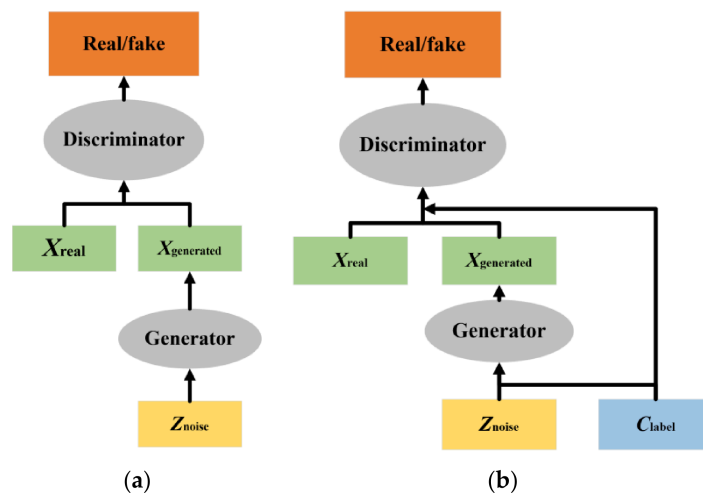


Figure 1. Basic constructs of network: (a) GAN, (b) CGAN.

The generator's objective is to make the generated data closer to real data given specific conditions. The objective functions of the G and D are, respectively,

$$L_G = \mathbb{E}_{z \sim p_z(z), c} [\log(1 - D(G(z, c)))] \quad (2)$$

$$L_D = -\mathbb{E}_{x \sim p_r(x), c} [\log(D(x, c))] - \mathbb{E}_{z \sim p_z(z), c} [\log(1 - D(G(z, c)))] \quad (3)$$

where c is the data category label.

The overall loss function becomes

$$\min_G \max_D \mathbb{E}_{x \sim p_r(x), c} [\log(D(x, c))] + \mathbb{E}_{z \sim p_z(z), c} [\log(1 - D(G(z, c)))] \quad (4)$$

2.2. Self-Attention Mechanism

The self-attention mechanism was proposed as an attempt to solve the problem of neural networks' inability to establish correlations for multiple inputs. By incorporating the self-attention mechanism, our objective is to address the challenge posed by conventional GANs in accurately capturing inter-location feature relationships within extensive time series data. The integration of the self-attention mechanism empowers the network to flexibly distribute attention weights, facilitating enhanced concentration on pivotal areas present in the extensive time series data, like Figure 2 shows. In simple terms, the principle of the self-attention mechanism is to obtain the relationship between different vectors.

First, we input the vibration signal into the network to obtain a feature vector $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{C \times N}$, which contains N elements, each with information of C channels. This feature vector X is processed through several hidden layers by three different linear transformation functions $f(x)$, $g(x)$ and $h(x)$, respectively.

$$f(x) = W_f(x) \quad (5)$$

$$g(x) = W_g(x) \quad (6)$$

$$h(x) = W_h(x) \quad (7)$$

where $W_f \in \mathbb{R}^{\hat{C} \times C}$, $W_g \in \mathbb{R}^{\hat{C} \times C}$, $W_h \in \mathbb{R}^{\hat{C} \times C}$ are the weight matrices obtained by the convolutional neural network, $\hat{C} = C/8$, which acts on the vectors X , generating the feature spaces with different dimensions f , g and h . The feature spaces f and g both have vector dimensions of size $\hat{C} \times N$, and h has dimension size $C \times N$.

Next, computing the corresponding attention scores between these vectors, the f -space vector (set as x_i) is transposed, and its dimensions are changed to $N \times \hat{C}$, and it is dot-multiplied with the g -space vector (set as x_j) to get the matrix $a_{i,j}$ with the dimension of $N \times N$; $a_{i,j}$ represents the degree of distribution of the un-normalized attention distribution, i.e.,

$$a_{i,j} = f(x_i)^T g(x_j) \quad (8)$$

Then $a_{i,j}$ is normalized using the softmax function to get the attention weight $b_{i,j}$, i.e.,

$$b_{i,j} = \frac{\exp(a_{i,j})}{\sum_{i=1}^N (a_{i,j})} \quad (9)$$

This final attention distribution graph is used to show the level of attention paid to x_i when x_j is generated, i.e., the distribution of input weights for x_i .

Applying the attention weights to the element's feature representation $h(x_i)$ and performing a weighted summation generates the final output o_j , this output o_j is obtained by combining the weight matrix W_v with the element's transformation $v(x_i)$.

$$v(x_i) = W_v x_i \quad (10)$$

$$o_j = v \left(\sum_{i=1}^N b_{i,j} h(x_i) \right) \quad (11)$$

where $W_v \in \mathbb{R}^{C \times \hat{C}}$ is also the weight matrix as above. The network is allowed to learn local information first, and then to slowly learn information at a distance.

Finally, we can set a scale parameter γ that can be learnt to automatically measure the output weights and add the output of the attention mechanism to the original input vector, with a final output of

$$y_i = \gamma o_i + x_i \quad (12)$$

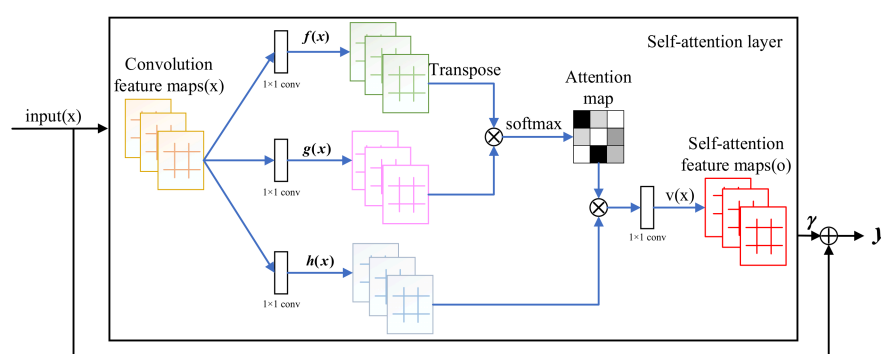


Figure 2. Self-attention layer.

2.3. Gradient Normalization Module

Our goal is to use neural networks to generate data samples with the same distribution but with diversity to aid in pattern recognition. Although GANs have a great advantage in feature generation and reinforcement of data, they still face two problems: one is pattern collapse during generator training due to training instability; the other is that the data generated by unsupervised networks do not have the same distribution as the original target data. Recent research has focused on these two aspects.

Wasserstein was initially introduced into GANs [26] to achieve the goal of making the distribution of the generated data $p_g(x)$ as biased as possible towards the real data $p_r(x)$, i.e.,

$$\min_G \max_{D, L_D \leq 1} \mathbb{E}_{x \sim p_r(x)} [D(x)] - \mathbb{E}_{\tilde{x} \sim p_g(x)} [D(\tilde{x})] \quad (13)$$

where $p_g(x)$ is the distribution of generated data and L_D represents the Lipschitz constant of D . L_D can be defined as the following:

$$L_D := \inf \{L \in \mathbb{R} : |D(x) - D(y)| \leq L \|x - y\|, \forall x, y \in \mathbb{R}^n\} \quad (14)$$

where $\|\cdot\|$ can be the norm of the vector.

Ensuring a proper equilibrium between the Lipschitz constraint and network capacity poses a significant challenge when dealing with neural networks. Although methods like spectral normalization (SN) [37] and gradient penalty (GP) [38] offer partial solutions, they still exhibit certain limitations. SN, for instance, can lead to considerable parameter reductions in deep multilayer networks. Meanwhile, GP might struggle to provide precise values at specific locations. An alternative approach, gradient normalization (GN), shows promise in addressing these issues. Typically, activation functions like ReLU or LeakyReLU are favored. Under these activations, the functions behave as “piecewise linear functions,” demonstrating linearity within local continuous regions except at the boundaries. Consequently, they exhibit characteristics of constant vectors. Bhaskara et al. [39] proposed a gradient normalization method to improve the reliability of the training process and the usability of the generated data by introducing a gradient normalization operation between different layers of the network with the following objective function:

$$\hat{D}(x) = \frac{D(x) \cdot \|\nabla D(x)\|}{\|\nabla D(x)\|^2 + \epsilon} \quad (15)$$

The gradient normalization method proposed by Wu et al. [40] aims to solve the gradient explosion and gradient disappearance problems in GAN training by normalizing the gradients in the generator and discriminator networks with the following functions:

Both methods are designed to enhance the stability of training by limiting the gradient to a suitable range. To be specific, for each network parameter, its gradient is divided by the gradient’s parametric number to normalize the gradient to unit length. The loss function is as follows:

$$\hat{D} = D(x) / (\|\nabla D(x)\| + |D(x)|) \quad (16)$$

$$L(D) = \mathbb{E}_{x \sim p_r(x)} [\hat{D}(D(x))] - \mathbb{E}_{\tilde{x} \sim p_g(x)} [\hat{D}(D(\tilde{x}))] \quad (17)$$

$$L(G) = -\mathbb{E}_{x \sim p_z(x)} [\hat{D}(G(x))] \quad (18)$$

where $\hat{D}$ represents functions that can be redefined and set, $D(x)$ serves as a variable depending on the situation (like $G(x)$ or $D(x)$), and $\|\nabla D(x)\|$ denotes the norm of the vector.

Compared with SN and GP, GN is used here as a training-level stabilization strategy rather than a layer-wise structural constraint. It normalizes the gradient scale during discriminator optimization and avoids directly restricting the spectral norm of each layer. In addition, it does not require interpolation sampling as in GP. This reduces the dependence on penalty-coefficient tuning and helps preserve the flexibility of the discriminator when learning multi-modal spectral distributions.

2.4. Spectral Features of Vibration and Current

To provide compact and fault-sensitive inputs for the proposed SWGAN–SAE framework, the raw time-domain vibration and stator current signals are transformed into one-dimensional spectral features with clear physical meaning. Specifically, the vibration channel adopts an envelope spectrum representation to enhance modulation components excited by mechanical impacts, whereas the electrical channel uses a space-vector current spectrum derived from Clarke-transformed three-phase currents to capture distortions and unbalance caused by stator faults.

For the vibration signal $x(t)$, a band-pass filter $\mathcal{B}\{\cdot\}$ is first applied around a selected resonance band to extract the high-frequency response excited by periodic impacts:

$$x_f(t) = \mathcal{B}\{x(t)\} \quad (19)$$

The analytic signal $z(t)$ associated with $x_f(t)$ is then constructed via the Hilbert transform $\mathcal{H}[\cdot]$:

$$z(t) = x_f(t) + j\mathcal{H}[x_f(t)] = a(t)e^{j\varphi(t)} \quad (20)$$

where $a(t)$ and $\varphi(t)$ denote the instantaneous amplitude (envelope) and instantaneous phase, respectively. The envelope is given by

$$a(t) = |z(t)| = \sqrt{x_f^2(t) + \left(\mathcal{H}[x_f(t)]\right)^2} \quad (21)$$

The Fourier transform of the envelope yields the envelope spectrum:

$$E(f) = |\mathcal{F}\{a(t)\}| \quad (22)$$

where $\mathcal{F}\{\cdot\}$ denotes the Fourier transform. The spectrum $E(f)$ suppresses irrelevant carrier frequencies while highlighting fault-related modulation structures, and is therefore well suited as a vibration-domain input feature.

For the stator current, the three-phase currents $i_a(t)$, $i_b(t)$ and $i_c(t)$ are first mapped into the stationary $\alpha\beta$ reference frame using the Clarke transformation:

$$\begin{bmatrix} i_\alpha(t) \\ i_\beta(t) \end{bmatrix} = \frac{2}{3} \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} i_a(t) \\ i_b(t) \\ i_c(t) \end{bmatrix} \quad (23)$$

The corresponding space-vector magnitude is defined as

$$i_s(t) = \sqrt{i_\alpha^2(t) + i_\beta^2(t)} \quad (24)$$

which compresses the three-phase information into a single scalar signal while preserving waveform distortions and asymmetries induced by stator faults. Its Fourier transform defines the current spectrum

$$S(f) = |\mathcal{F}\{i_s(t)\}| \quad (25)$$

In summary, the envelope spectrum $E(f)$ and current spectrum $S(f)$ convert high-dimensional raw time-domain data into low-dimensional spectral features that emphasize fault-related harmonic and modulation patterns while reducing redundant information and noise.

3. Proposed Method and System Framework

To address the fault sample imbalance problem in motor system fault diagnosis, we propose a data augmentation framework termed SWGAN. It consists of two main components: (1) an improved self-attention guided Wasserstein conditional GAN with gradient normalization, which generates realistic fault samples of current and vibration signals with distributions closely matching those of real data; and (2) a stacked autoencoder (SAE)-based fault classification network, which is used to validate the usefulness of the generated samples and to evaluate diagnostic performance on the augmented dataset.

3.1. Feature Generator Combining Self-Attention

CNN networks are able to extract features from certain segments, but when faced with long-sequence data samples, they are often unable to accurately identify the dependencies between different segments of the same sequence of data. In this paper, we introduce the self-attention mechanism. Algorithm 1 shows the pseudo-code of the redesigned generator, which shows the operation of the feature generator. At each iteration, noise samples z and fault labels c are first drawn and fed into attention-augmented layers. Within these layers, the input undergoes convolutional projections to obtain query f , key g , and value h representations, followed by scaled dot-product attention to capture long-range dependencies across spectral positions. The attention output is then merged with the convolutional feature map and progressively refined through subsequent layers to produce synthetic multi-modal spectral samples.

Algorithm 1. Feature generator combining self-attention

Input: Noise samples z from distribution $p_g(z)$, conditional labels c .
Output: Updated generator parameters θ_g .

- 1: **for** each training iteration **do**:
- 2: Sample minibatch of m noise samples from noise prior $p_g(z)$,
- 3: denote as $\{z(1), \dots, z(m)\}$;
- 4: Sample corresponding minibatch of labels $\{c(1), \dots, c(m)\}$;
- 5: **for** sample $z(i), c(i)$ in each attention layer **do**:
- 6: Apply convolutional transformations to the input;
- 7: Derive f, g, h from the transformed output;
- 8: Compute attention:

$$\text{Attention_Output} = \text{softmax}\left(\frac{g(x) * f(x)^T}{\sqrt{d_k}}\right) * h(x).$$
- 9: Combine Attention_Output with the layer's output;
- 10: Apply subsequent layers to enhance feature representation;
- 11: Generate samples $G(z^{(i)}, c^{(i)})$ using the enhanced features;
- 11: **end**
- 12: Compute the generator loss using Wasserstein distance:

$$L_G = -\frac{1}{m} \sum_{i=1}^m D(G(z^i, c^i)).$$
- 13: Compute gradient of L_G with respect to θ_G :

$$g = \nabla_{\theta_G} L_G.$$
- 14: Update θ_G using Adaptive Moment Estimation;
- 14: **end**

Figure 3 shows its schematic diagram in the network. The path below is the base CNN path in its generator. The top is the introduced self-attention layer. The two paths simultaneously accept data input from the previous layer, the self-attention layer calculates

the weight distribution of different features based on the input data, and the convolutional layer obtains features from them for local aggregation. γ is a trainable parameter.

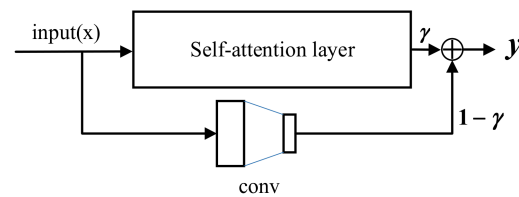


Figure 3. Feature generator containing self-attention.

3.2. System Framework Design

The proposed fault diagnosis framework comprises three tightly integrated components. For data augmentation, a self-attention enhanced generator synthesizes multi-modal spectral samples conditioned on fault labels. As detailed in Table 1, it is built with two convolutional, two deconvolutional, and one flattening layer; a 1×2400 random noise vector is concatenated with an 8-dimensional label to form a 2408-dimensional input, and self-attention modules are inserted into the first three layers to model long-range dependencies across spectral positions. The resulting 2400-dimensional output is then passed as a generated sample to the discriminator.

Table 1. Specific parameters of the proposed method.

Layers	SWGAN
Generator	
Input-G	G-in = input (noise-dim = 2400)
Label	in-dim = 2400, out-dim = 2408, label-dim = 8
Conv 1	in = 1, out = 64, k = 3, s = 1, LeakReLU
Conv 2	in = 64, out = 128, k = 3, s = 1, LeakReLU
Deconv 1	in = 128, out = 64, k = 3, s = 2, LeakReLU
Deconv 2	in = 64, out = 1, k = 3, s = 2, LeakReLU
Fc 1	Out-dim = 2400, LeakReLU
Flatten	Out-dim = 1×2400
Discriminator	
Input-D	D-in = input (dim = 2400)
Label	in-dim = 2400, out-dim = 2408, label-dim = 8
Fc 1	Out-dim = 512, LeakReLU
Fc 2	Out-dim = 128, LeakReLU
Fc 3	Out-dim = 32, LeakReLU
Fc 4	Out-dim = 1
Classifier	
Input-V	V-in = input (dim = 2400)
Input-I	I-in = input (dim = 2400)
Fc 1-V	Out-dim = 512, ReLu
Fc 1-C	Out-dim = 512, ReLu
Fusion-Fc	In-dim = 512 + 512, Out-dim = 512, ReLu
Fc 2	Out-dim = 256, ReLu
Fc 3	Out-dim = 128, ReLu
Fc 4	Out-dim = num-class, SoftMax

The discriminator is a fully connected network with layer sizes of 2408, 512, 128, 32, and 1, receiving both real samples and the synthetic ones in the same format. To stabilize adversarial training without introducing extra hyperparameters, a GN module is embedded into the discriminator. The training procedure is illustrated in Algorithm 2. After the Wasserstein loss is computed over a mini-batch, the gradients are normalized,

which keeps the gradient magnitude within a stable range without requiring per-layer weight constraints or a hand-tuned penalty coefficient. The normalized gradients are then used to update the parameters via Adam.

Algorithm 2. Discriminator with gradient normalization

Input: Real samples from distribution $p_r(x)$, conditional labels c , noise samples z from distribution $p_g(z)$.

Output: Updated discriminator parameters θ_d .

```

1:  for each training iteration do:
2:      Sample minibatch of  $m$  real samples from real data  $p_r(x)$ ,
        denote as  $\{x(1), \dots, x(m)\}$ ;
3:      Sample minibatch of  $m$  noise samples from noise prior  $p_g(z)$ ,
        denote as  $\{z(1), \dots, z(m)\}$ ;
4:      Using a minibatch of fake samples from the feature generator
         $\{G(z(1), c(1)), \dots, G(z(m), c(m))\}$ ;
5:      Sample corresponding minibatch of labels  $\{c(1), \dots, c(m)\}$ ;
6:      Compute the discriminator loss using Wasserstein distance:

$$L_D = \frac{1}{m} \sum_{i=1}^m D(x^i, c^{(i)}) - \frac{1}{m} \sum_{i=1}^m D(G(x^i, c^{(i)}))$$

7:      Compute the gradients of  $L_D$  with respect to  $\theta_d$ :

$$g = \nabla_{\theta_d} L_D$$

8:      Normalize the gradients:

$$\text{norm} = \sqrt{\sum_j g_j^2 + \epsilon}$$

        Calculate normalized gradients:

$$\dot{g} = \frac{g}{\text{norm} + |D(X)|}$$

9:      Update  $\theta_d$  using Adaptive Moment Estimation.
10: end

```

For fault classification, a two-branch stacked autoencoder is employed to process the augmented data. The network includes an input layer, multiple feature-extraction layers, a feature-fusion module, and a final classification layer. The envelope spectra of vibration signals and the space-vector spectra of stator currents are first fed into two modality-specific encoding branches to extract latent features. Vibration envelope spectra and current space-vector spectra are encoded separately into 512-dimensional features. These features are then fused through a nonlinear fully connected module with learnable fusion weights. Followed by two hidden layers of 256 and 128 dimensions, the fused representation is further transformed into higher-level features and finally fed into a SoftMax classifier to identify fault categories. Network parameters are updated by backpropagation using the Adam optimizer.

3.3. System Training Procedures

The complete fault diagnosis procedure is illustrated in Figure 4, and the main steps are as follows:

- (1) Multi-channel vibration acceleration signals and three-phase stator currents are collected from the experimental platform under different motor stator fault conditions and health states.

- (2) The acquired signals are segmented by a sliding window. For each segment, the envelope spectrum of the vibration signal and the space-vector spectrum of the stator current are computed and then normalized to the range $[-1, 1]$.
- (3) An SWGAN model is constructed and trained on the imbalanced spectral datasets to generate additional fault samples of both vibration and current, conditioned on the corresponding fault labels.
- (4) The generated samples are combined with the original spectral data to form an enhanced, approximately balanced multi-modal dataset.
- (5) The enhanced datasets are fed into the SAE-based classification network, where fused current–vibration features are learned and the final fault categories are obtained.

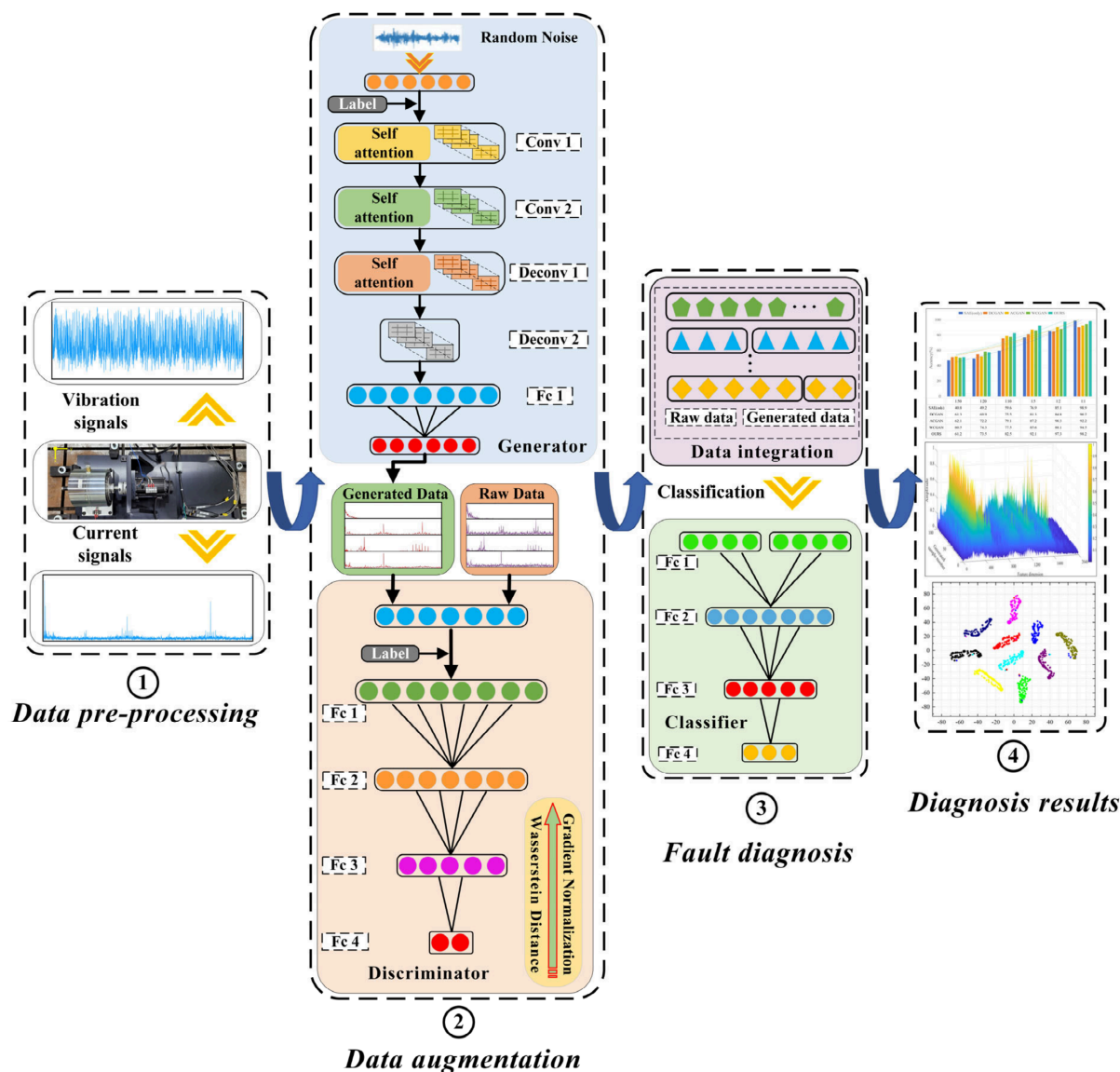


Figure 4. SWGAN fault diagnosis framework.

All models were implemented in Python 3.10 using PyTorch 2.11. The Adam optimizer was used for the generator, discriminator, and SAE classifier. The learning rate was set to 1×10^{-4} , the batch size was 100, and the number of training epochs was 200. Each experiment was repeated 20 times. The experiments were conducted on a workstation equipped with an i7 14700 kf CPU, 64 GB RAM, and an NVIDIA 5060Ti GPU with 16 GB memory. The proposed framework is mainly designed for offline augmentation and

classifier training. In online diagnosis, the trained SAE classifier can be used directly after spectral feature extraction, while SWGAN does not need to run during inference.

4. Experimental Verification

4.1. Dataset A

4.1.1. Description of PMSM Dataset

In this section, the proposed method is evaluated using a multi-channel vibration and current Dataset A of three-phase motors with stator faults [41], and the testbed is shown in Figure 5. Three motors with rated powers of 1.0 kW, 1.5 kW, and 3.0 kW are tested under a constant operating condition of 3000 rpm and a rated load corresponding to 15% of the allowable torque limit (1.5 N·m), and the details are illustrated in Table 2. Vibration signals are measured in the axial direction using an accelerometer mounted on the motor bearing, and are acquired at a sampling frequency of 25.6 kHz. Three-phase stator currents are measured using current probes clamped on the U-, V-, and W-phases of the motor and are collected at a sampling frequency of 100 kHz.

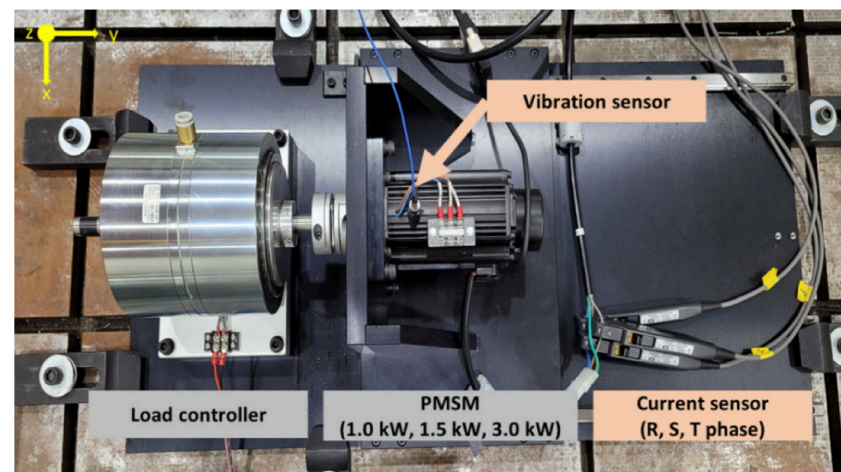


Figure 5. Description of PMSM testbed of Casa A.

Table 2. Parameters of PMSM.

Parameters	A Motor	B Motor	C Motor
Manufacturing company	Higen Motors	Higen Motors	Higen Motors
Rated power (W)	1000	1500	3000
Input voltage (V)	380	380	380
Frequency (Hz)	60	60	60
Number of phases	3	3	3
Number of poles	4	4	4
Rated torque (Nm)	3.18	4.77	9.55
Rated speed (RPM)	3000	3000	3000
Magnetic flux (mT)	400	350	300
Rotor inertia ($\text{kg}\cdot\text{m}^2$)	2.07	7.48	14.34
Inter-turn resistance value R_{it} (Ω)	0.1385	0.0958	0.1087
Inter-coil resistance value R_{cc} (Ω)	0.0409	0.3021	0.1534

Stator short-circuit faults are artificially introduced on the U-phase windings. The physical locations of the faults for the three motors are shown in Figure 6, while Figure 7 depicts the equivalent short-circuit scheme for the inter-turn and inter-coil branches. Both fault types are realized by inserting a bypass resistor R_{bypass} between different turns of the same coil (inter-turn) or between adjacent coils (inter-coil), which forms a parallel

path with respect to the original stator resistance R . The fault severity is quantified by the fault ratio:

$$\text{FaultRatio}(\text{FR}) = \left(\frac{R}{R + R_{\text{bypass}}} \right) \times 100 \quad (26)$$

where a smaller R_{bypass} corresponds to a larger fault ratio and hence a more severe short circuit. As summarized in Table 3, the bypass resistance takes different values for both inter-turn and inter-coil faults, and the corresponding fault ratios for the three motors are reported.

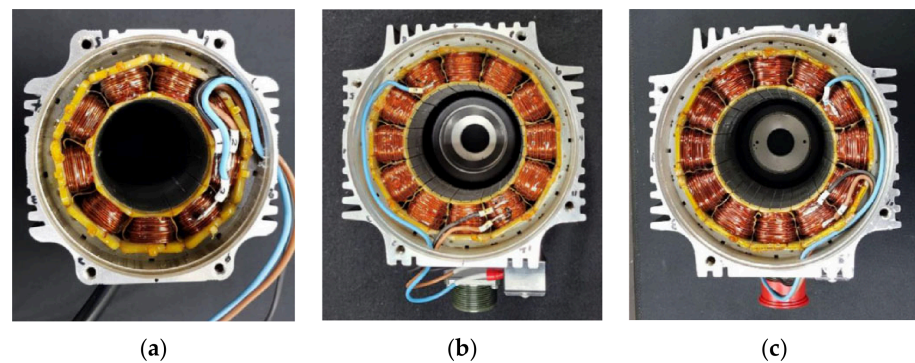


Figure 6. Fault on U-phase of PMSMs with different powers: (a) 1.0 kW, (b) 1.5 kW, (c) 3.0 kW.

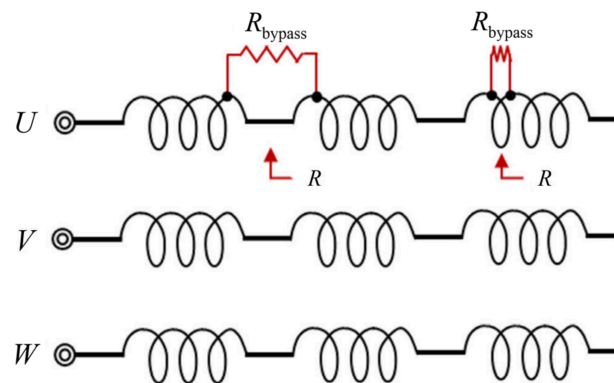


Figure 7. Stator winding of short-circuit fault scheme.

Table 3. Description of dataset fault types.

Fault Type	Condition	Bypass Resistance R_{bypass} (Ω)	Motor A	Motor B	Motor C
Normal	NC	–(no bypass)	0 (normal)	0 (normal)	0 (normal)
Inter-turn short circuit	IT-1	6	2.26	1.57	1.78
	IT-2	5	2.70	1.88	2.13
	IT-3	4	3.35	2.34	2.65
	IT-4	3	4.41	3.10	3.50
	IT-5	2	6.48	4.57	5.16
	IT-6	1	12.17	8.74	9.81
	IT-7	0.5	21.69	16.08	17.86
Inter-coil short circuit	IC-1	6	0.68	4.79	2.49
	IC-2	5	0.81	5.70	2.98
	IC-3	4	1.01	7.02	3.69
	IC-4	3	1.34	9.15	4.86
	IC-5	2	2.00	13.12	7.12
	IC-6	1	3.93	23.20	13.30
	IC-7	0.5	7.56	37.66	23.48

For each motor, we consider 10 operating conditions derived from Table 3. The healthy state corresponds to the case where no bypass resistor is connected in either the inter-turn or the inter-coil branch, whereas the other fault states are obtained by inserting bypass resistors within the inter-turn or inter-coil paths, covering a wide range from mild to severe damage. Vibration acceleration and three-phase current data under healthy, inter-coil short-circuit, and inter-turn short-circuit conditions are segmented by a sliding window to construct sample-level datasets.

4.1.2. Evaluation of Sample Similarity Based on Visualization of Dataset A

The proposed SWGAN model was employed to generate additional spectral samples for ten operating conditions of the PMSM stator. The synthetic samples were then fused with the original vibration–current spectra to balance the training set, ensuring that the proportion of each fault class reached the same level as the test set.

Figure 8 presents the two-dimensional projections of both real and generated samples for the vibration and current spectra under inter-turn and inter-coil short-circuit faults. This statement suggests that SWGAN has effectively captured the statistical characteristics of real data, allowing it to generate samples that closely resemble real data samples. This improvement enhances the method's resilience and its ability to produce results that are representative of the broader data distribution. To put it differently, SWGAN preserves the essential attributes of the original signal while eliminating some extraneous or noisy elements that could impede the process of identifying patterns within the data.

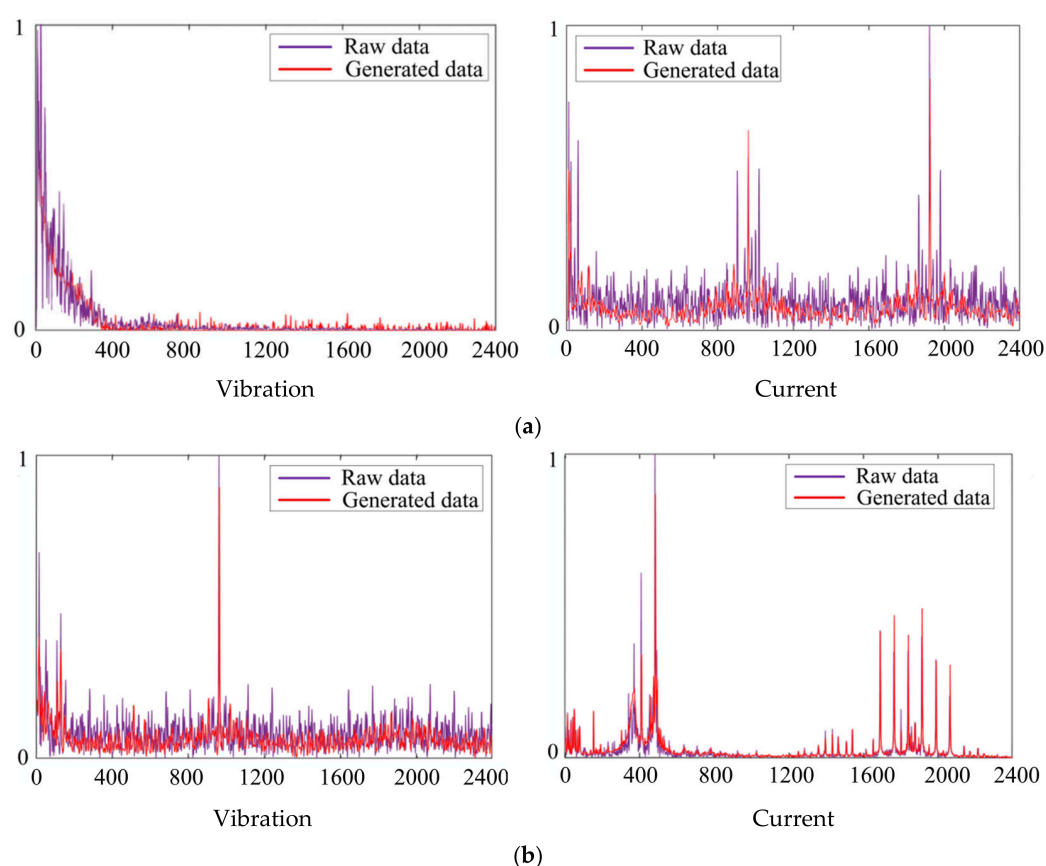


Figure 8. Spectral comparison under 2-D of Dataset A: (a) IT1, (b) IC3.

4.1.3. Evaluation of Sample Similarity Based on Statistical Indicators of Dataset A

Jensen–Shannon divergence and Cosine Distance are employed to assess the similarity between real and generated data from a statistical point of view, which can also demonstrate

the average performance of the generative model. Jensen–Shannon (JS) divergence is a symmetric measure of statistical similarity between two probability distributions. It quantifies the average Kullback–Leibler (KL) divergence between each distribution and their average, providing a comprehensive assessment of overall dissimilarity. With a range between 0 and 1. The smaller its value, the higher the similarity. And it is calculated as follows:

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log \left(\frac{P(i)}{Q(i)} \right) \quad (27)$$

$$D_{JS}(P \parallel Q) = \frac{1}{2} D_{KL} \left(P \parallel \frac{P+Q}{2} \right) + \frac{1}{2} D_{KL} \left(Q \parallel \frac{P+Q}{2} \right) \quad (28)$$

Cosine distance measures the dissimilarity between two vectors in vector space by computing the cosine of the angle between them. The distance is defined as one minus the cosine similarity. Its value ranges from 0 to 2, where a smaller value (closer to 0) indicates a higher degree of similarity, as the vectors are oriented more closely in the same direction. The cosine distance is calculated by the following formula:

$$d_{CS}(x, y) = 1 - \left[\frac{\sum_{i=1}^m x_i y_i}{\sqrt{\sum_{i=1}^m x_i^2} \sqrt{\sum_{i=1}^m y_i^2}} \right] \quad (29)$$

A violin plot is a graph used to show the distribution of data, which can help to understand the central tendency of the data, the degree of dispersion, and possible outliers. Figure 9 shows violin plots giving similarity metrics for 20 experiments for faults IT1, IT3 and IC3 under different methods respectively. It can be seen in the figure that the distance between the data generated by DCGAN, ACGAN and WCGAN and the original data is significantly larger than the data generated by SWGAN. And the similarity between the data generated by SWGAN and the original data stays in a stable interval, indicating that the method proposed in this paper can generate high-quality usable data more stably. In summary, the SWGAN model exhibits robustness and superior multi-category data generation capabilities, outperforming other models in performance.

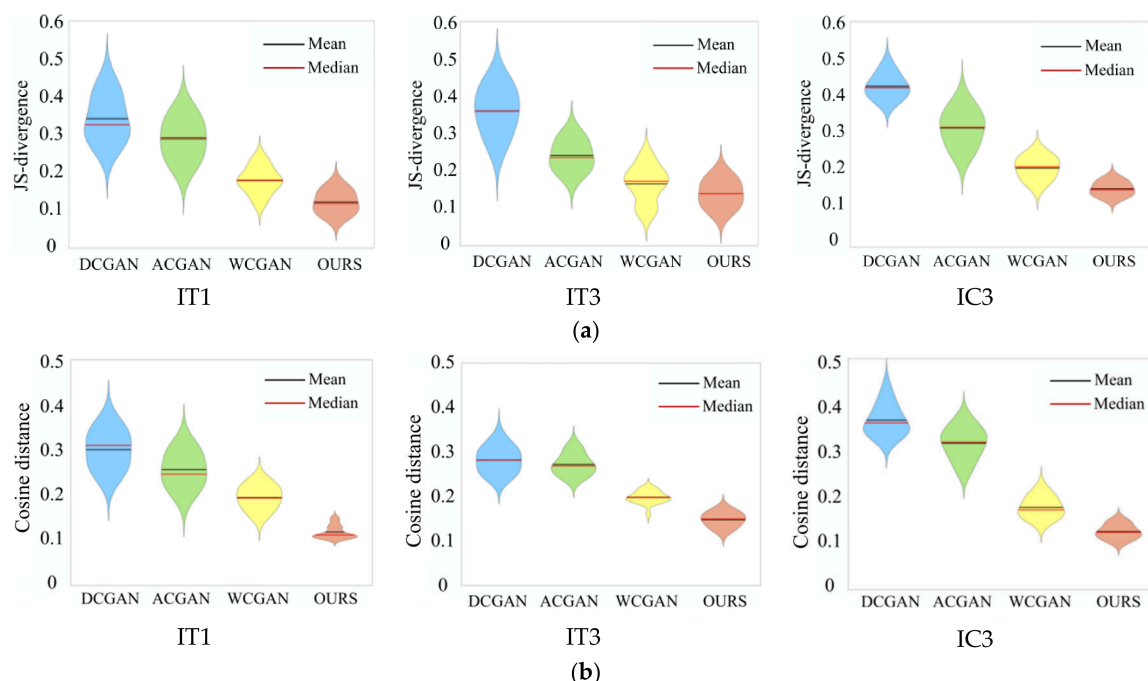


Figure 9. Violin plot of statistical indicators of Dataset A: (a) JS divergence, (b) cosine distance.

4.1.4. Diagnosis Results of Dataset A

(a) Fault diagnosis under limited training set

The fault types considered in this study include N, IT1, IT3, IT4, IT5, IT6, IC1, IC3, IC4, and IC5. For each health condition, the continuous raw signals are first divided into non-overlapping segments. Each sample is then constructed from a 4800-point time-domain segment. For current signals, the Clarke transformation is applied to obtain the α - β stationary components and their space-vector magnitude, followed by FFT to generate a 2400-point current spectrum. For vibration signals, envelope extraction and spectral analysis are performed, yielding a 2400-point vibration envelope spectrum. Training and testing samples are selected from different non-overlapping segments to avoid sample leakage caused by highly overlapped windows. A total of 400 samples is collected for each health state, with 200 used for training and 200 for testing. The dataset composition is summarized in Table 4.

Table 4. Fault-type situations in the datasets.

Label	Health Condition	Rated Power (W)	Fault Severity (%)	Data Length (Points)	Training Samples	Testing Samples
0	N	1000	–	2400	200	200
1	IT1	1000	2.26	2400	200	200
2	IT3	1000	3.35	2400	200	200
3	IT4	1000	4.41	2400	200	200
4	IT5	1000	6.48	2400	200	200
5	IT6	1000	12.17	2400	200	200
6	IC1	1000	0.68	2400	200	200
7	IC3	1000	1.34	2400	200	200
8	IC4	1000	2.00	2400	200	200
9	IC5	1000	3.93	2400	200	200

In this experiment, we simulate a scenario of motor fault diagnosis under a severely limited overall training set, rather than class imbalance. Specifically, the training set is constructed by severely restricting the total number of available samples across all classes. All classes without the healthy state are set to the same small number of raw training samples, including 4, 10, 20, 40, 100, or 200 per class. These correspond to increasingly scarce training regimes. The test set remains balanced with 200 samples per class for fair evaluation. The specific information is shown in Table 5. The aim is to assess whether the proposed method can effectively learn from only a handful of total training examples and still achieve reliable diagnosis, thus serving as a preliminary investigation into data-scarce fault diagnosis before addressing the more complex class-imbalance problem.

Table 5. Status of datasets under limited training set.

Data Type	Dominant Class			Minority Class			
	0	1~9	1~9	1~9	1~9	1~9	1~9
Imbalance ratio	-	1:50	1:20	1:10	1:5	1:2	1:1
Training samples	200	4	10	20	40	100	200
Testing samples	200	200	200	200	200	200	200

To evaluate the effectiveness of the proposed SWGAN, three representative GAN models, including a WCGAN (i.e., SWGAN without GN and self-attention), a DCGAN [42], and an ACGAN [43] are selected for comparison. Each method independently augments datasets under different imbalance ratios, after which the augmented data are classified

using SAE. In addition, a baseline experiment without data augmentation (SAE-only) is conducted. All results are averaged over 20 runs.

As shown in Figure 10, the classification accuracy of all GAN-based methods consistently increases as the available training samples become more abundant. Under severe data scarcity ratios of 1:50 and 1:20, the differences among the various GAN variants remain small, whereas the SAE-only classifier performs markedly poorly, because the extremely limited number of fault samples hinders effective feature extraction. When the training set ratio reaches 1:10, the proposed SWGAN begins to show a clear advantage over the compared methods. At the 1:1 ratio, where the training set is fully balanced in size, SWGAN achieves over 98% accuracy; notably, in this case, generated samples are used to replace real ones during training, and the resulting accuracy approaches the upper-bound performance of an SAE-only model trained with sufficient real data.



Figure 10. Diagnostic accuracy with different imbalance ratios.

To further examine the classification behavior, Figure 11 presents the confusion matrices obtained under two representative training set ratios, 1:5 and 1:1. The diagonal entries are substantially higher at the 1:1 ratio, confirming that when an adequate number of training samples is available, the network can effectively learn discriminative features and produce high-quality synthetic samples, leading to more reliable fault identification.

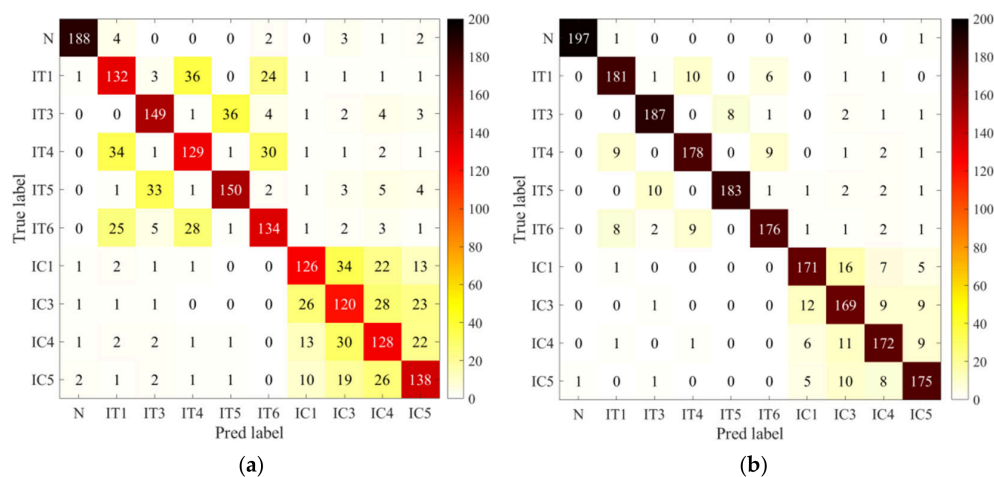


Figure 11. Categorical confusion matrix of Dataset A: (a) Imbalance ratio of 1:5, (b) imbalance ratio of 1:1.

(b) Fault diagnosis under training set class imbalance

Data imbalance among different PMSM stator fault types is another key issue in intelligent fault diagnosis. The imbalance across fault categories is more complex and more common in real industrial scenarios. As shown in Table 6, two imbalance settings—Case A and Case B—are designed in this study. Case A simulates moderate imbalance: the sample proportions of the ten PMSM conditions (N, IT1, IT3, IT4, IT5, IC1, IC3, IC4, IC5, IC6) are configured so that the normal state accounts for the largest portion, while different inter-turn and inter-coil fault levels contain progressively fewer samples. Case B further exaggerates this imbalance by significantly reducing the number of samples of minority fault classes.

Table 6. Situation of the imbalanced Cases A and B.

Label	Health Condition	Rated Power (W)	Training Samples		Testing Samples Case A/B
			Case A	Case B	
0	N	2000	200	200	200
1	IT1	2000	60	30	200
2	IT3	2000	100	50	200
3	IT4	2000	120	60	200
4	IT5	2000	160	80	200
5	IC1	2000	60	30	200
6	IC3	2000	100	50	200
7	IC4	2000	120	60	200
8	IC5	2000	160	80	200
9	IC6	2000	180	90	200

All imbalanced datasets are augmented using the SWGAN and comparison GAN methods to achieve a balanced ratio of 1:1. The classifier is then trained using the mixed dataset. The test set still contains 10 categories with 200 samples per category, ensuring a consistent evaluation framework across all methods.

Figure 12 reports the classification accuracy on Cases A and B over twenty independent runs. On Case A, the SWGAN achieves a mean accuracy of 97.81% with a standard deviation of 0.53%. The WCGAN attains $94.05\% \pm 1.24\%$. The ACGAN yields $87.39\% \pm 1.69\%$. The DCGAN obtains $82.32\% \pm 2.13\%$. The SAE-only baseline reaches only $65.50\% \pm 2.20\%$. On Case B, where class imbalance is more severe, all methods suffer performance drops. The DCGAN decreases to $78.67\% \pm 3.02\%$, the ACGAN to $81.17\% \pm 2.05\%$, the WCGAN to $88.76\% \pm 1.78\%$, and the SAE-only to $58.78\% \pm 6.47\%$. In contrast, the SWGAN maintains $93.77\% \pm 1.12\%$. As the imbalance intensifies, the effective training data becomes scarcer, limiting the feature information available to the generative models and degrading the quality of the generated samples.

To visualize the learned representations, Dimension reduction through t-SNE (t-distributed stochastic neighbor embedding) [44] and UMAP (uniform manifold approximation and projection) [45] are applied to the samples from Case A, with the results shown in Figures 13 and 14. The DCGAN-generated samples display severe inter-class overlap and numerous outliers, indicating poor cluster separation. The ACGAN provides marginal improvement, yet substantial overlap between categories remains. The WCGAN succeeds in separating most classes and forms clearer clusters, but scattered outliers are still observable, suggesting that its fidelity is not fully satisfactory. In sharp contrast, the SWGAN produces compact, well-separated clusters for each fault type, with only sporadic misclassifications. These visualization results strongly confirm the effectiveness of the proposed method.

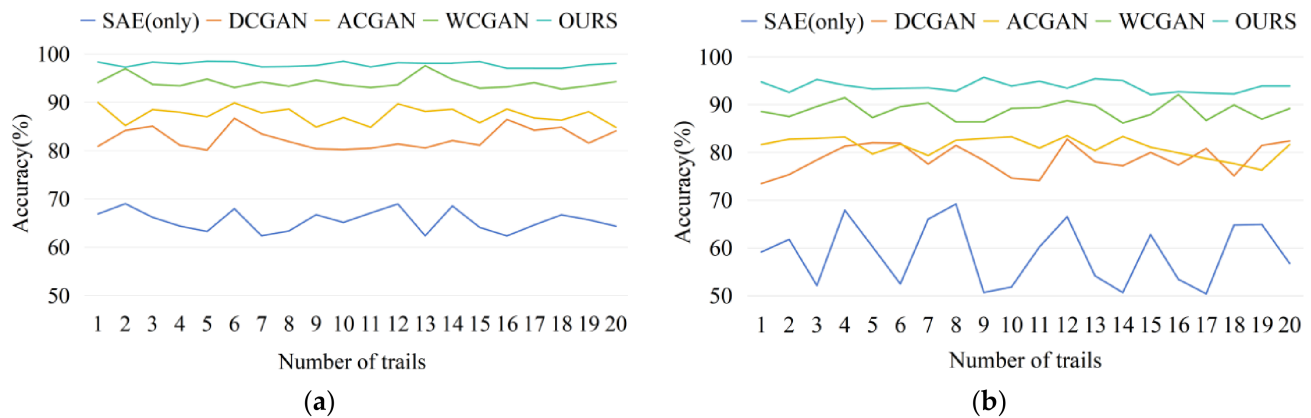


Figure 12. Performance comparison of five methods on two datasets: (a) Case A, (b) Case B.

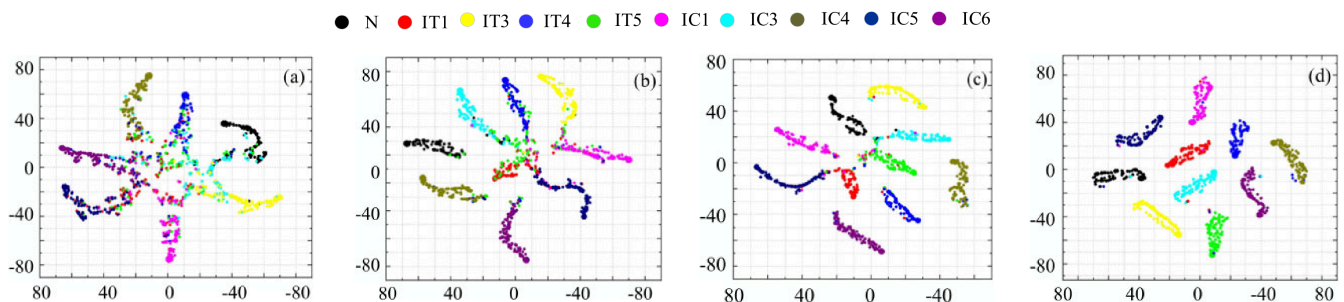


Figure 13. Visualization of T-sne on Case A: (a) DCGAN, (b) ACGAN, (c) WCGAN, (d) Ours.

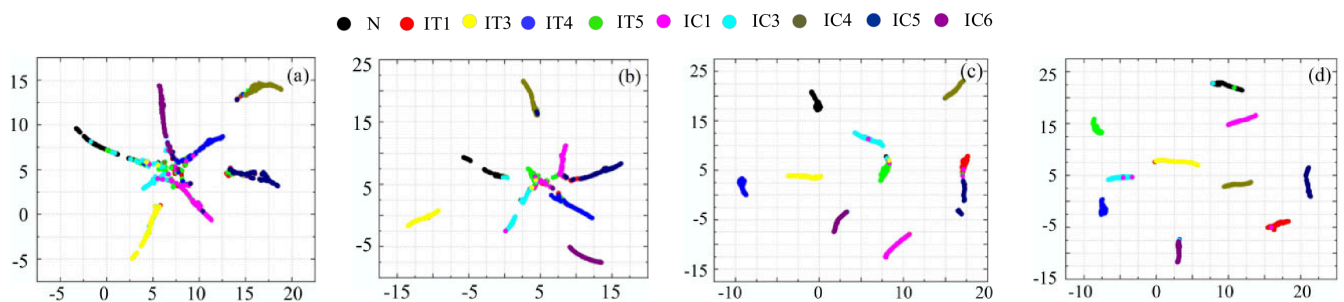


Figure 14. Visualization of UMAP on Case A: (a) DCGAN, (b) ACGAN, (c) WCGAN, (d) Ours.

4.2. Dataset B

4.2.1. Description of Induction Motor Dataset

To verify the generalization capability and robustness of the proposed method under variable-speed conditions, a public multi-channel dataset of a three-phase asynchronous motor, denoted as Dataset B [46], was used for further evaluation. The test rig is shown in Figure 15. It consists of a 2.2 kW three-phase asynchronous motor, a torque sensor, a two-stage parallel gearbox, and a magnetic powder brake. A triaxial acceleration sensor was mounted on the drive end of the motor. During the test, triaxial vibration signals, three-phase current signals, torque signals, and key-phase signals were synchronously collected using a multi-channel data acquisition system. The sampling frequency was 12.8 kHz. The main motor parameters are listed in Table 7. This dataset contains variable-speed and variable-load operating conditions. In this study, the variable-speed condition was selected for evaluation. The corresponding speed profile is shown in Figure 16. The constant-load variable-speed condition with a load torque of 20 Nm and a speed profile of 0–1000 rpm was selected for evaluation.

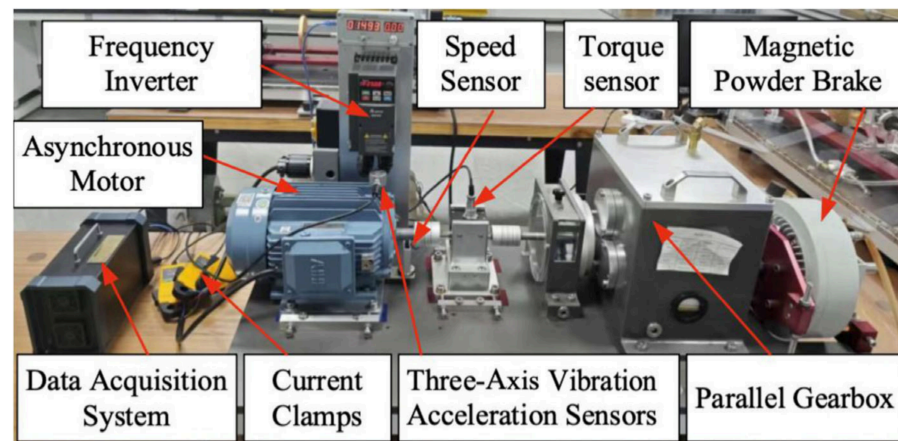


Figure 15. Description of induction motor testbed.

Table 7. Parameters of main motor.

Parameter	Value
Motor pattern	QABP-90L2
Poles of pair	2
Single-side air gap length	0.5 mm
Bearings	6205.2Z-C3/6205.2Z-C3
Stator inner diameter	72 mm
Stator outer diameter	130 mm
Rotor inner diameter	25 mm
Rotor outer diameter	71 mm

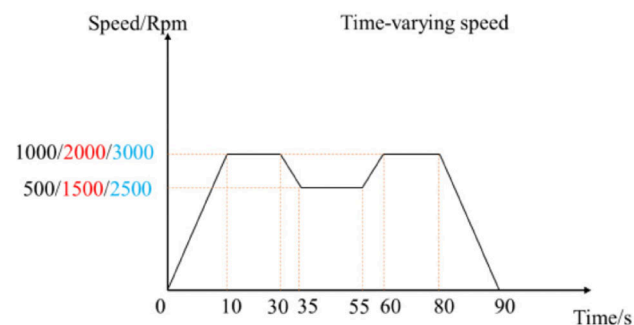


Figure 16. Time-varying speed curve.

To simulate representative electrical and mechanical faults, several controlled fault settings were introduced. Inter-turn faults were produced by connecting a variable shunt resistor R_f between two taps of the target phase winding, as shown in Figure 17. Two severity levels were obtained by generating circulating currents equal to 5% and 10% of the rated phase current. Shaft bending was produced by a four-point bending test, resulting in a permanent bend of 0.2 mm. Voltage unbalance was introduced by connecting an external resistor in series with one phase of the power supply, as shown in Figure 18. Two calibrated settings were used, corresponding to voltage unbalance levels of 4% and 8%. Static eccentricity was set with radial offsets of 0.125 mm and 0.250 mm, as shown in Figure 19. Rotor unbalance was introduced by mounting an unbalanced mass on the rotor shaft, as shown in Figure 20. Broken rotor bars were simulated by removing three consecutive rotor bars, as shown in Figure 21. Counterweights were added to the end ring to avoid mechanical imbalance. All fault types are listed in Table 8.

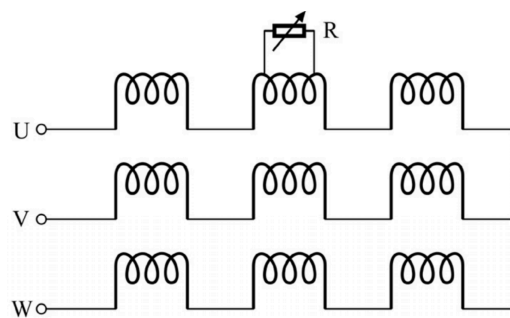


Figure 17. Inter-turn fault simulation scheme.

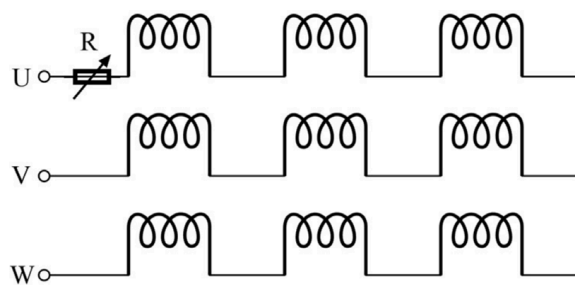


Figure 18. Voltage unbalance fault simulation scheme.

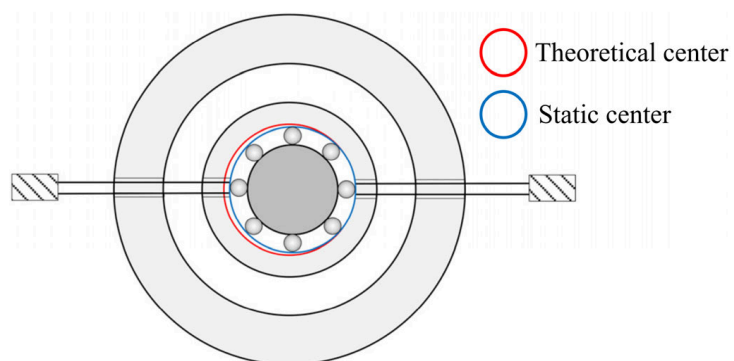


Figure 19. Static eccentricity setting.

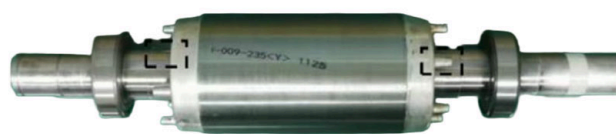


Figure 20. Rotor unbalance setting.



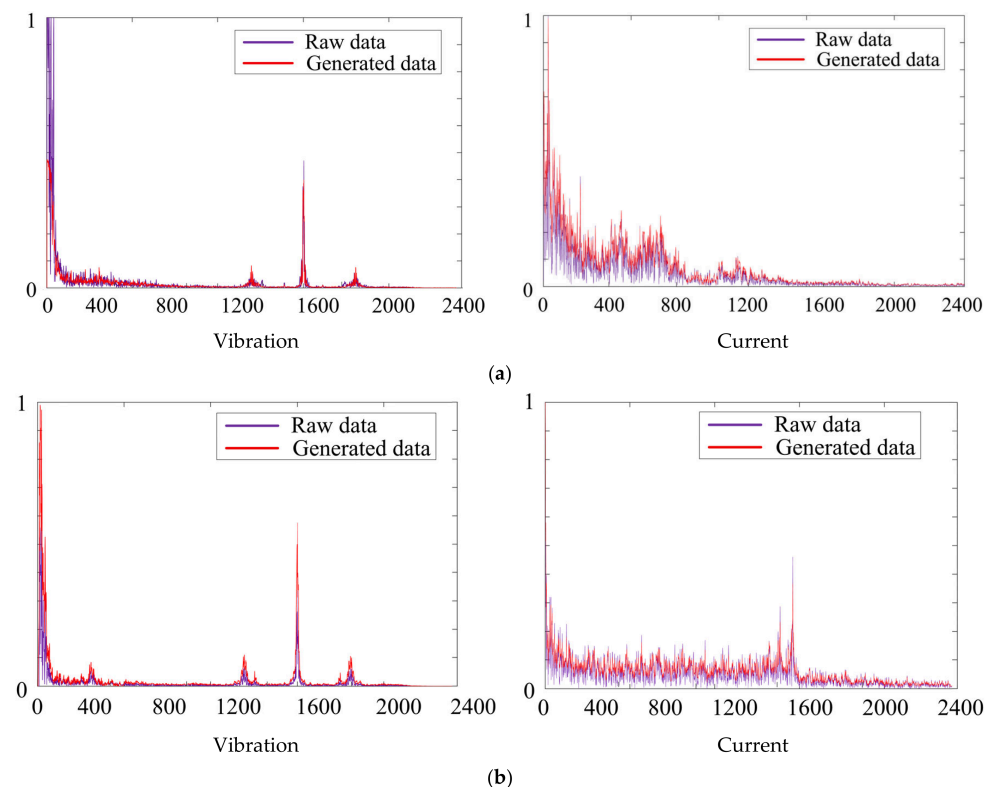
Figure 21. Broken rotor bar fault.

Table 8. Fault types of Dataset B.

Fault Type	Condition	Severity	Value
Healthy	NC	—	—
Inter-turn short circuit	IC1	Light	5% rated current shunt
	IC2	High	10% rated current shunt
Rotor bending	RB	—	0.2 mm permanent bend
Voltage unbalance	VU1	Light	4% voltage unbalance
	VU2	High	8% voltage unbalance
Static eccentricity	SE1	Light	0.125 mm radial offset
	SE2	High	0.250 mm radial offset
Rotor unbalance	RU	—	Additional imbalance mass
Broken rotor bars	RR	—	3 bars removed, counterweights added

4.2.2. Evaluation of Sample Similarity Based on Visualization of Dataset B

Figure 22 presents the two-dimensional projections of real and generated samples in the vibration and current spectra under inter-turn short-circuit and voltage-unbalance fault conditions. The generated samples show a generally consistent distribution trend with the real samples, indicating that the SWGAN can learn the main distribution characteristics of multi-modal spectral features under different fault states. Meanwhile, a small number of scattered points and local noisy samples can still be observed in the generated data. These points are partly consistent with the random fluctuations in real measurements. They also indicate that the generative model does not simply copy the original samples during data augmentation. These results suggest that the generated samples can expand the minority fault sample space and provide richer training information for the subsequent classification model.

**Figure 22.** Spectral comparison under 2-D of Dataset B: (a) IC2, (b) VU2.

4.2.3. Evaluation of Sample Similarity Based on Statistical Indicators of Dataset B

Figure 23 shows the JS divergence and cosine distance results for IC2, VU2, and SE2 under different generative methods. Each method was repeated 20 times. The SWGAN shows the lowest JS divergence and a narrower distribution range. This indicates a smaller distribution difference between the generated and real samples. In contrast, the DCGAN, ACGAN, and WCGAN show higher JS divergence, with larger fluctuations in some cases. For cosine distance, the SWGAN also remains at a lower level. This indicates that the generated samples are closer to the real samples in terms of spectral feature direction. For IC2, VU2, and SE2, the SWGAN shows stable generation performance. These results suggest that the proposed method can generate relatively stable supplementary samples under variable-speed conditions.

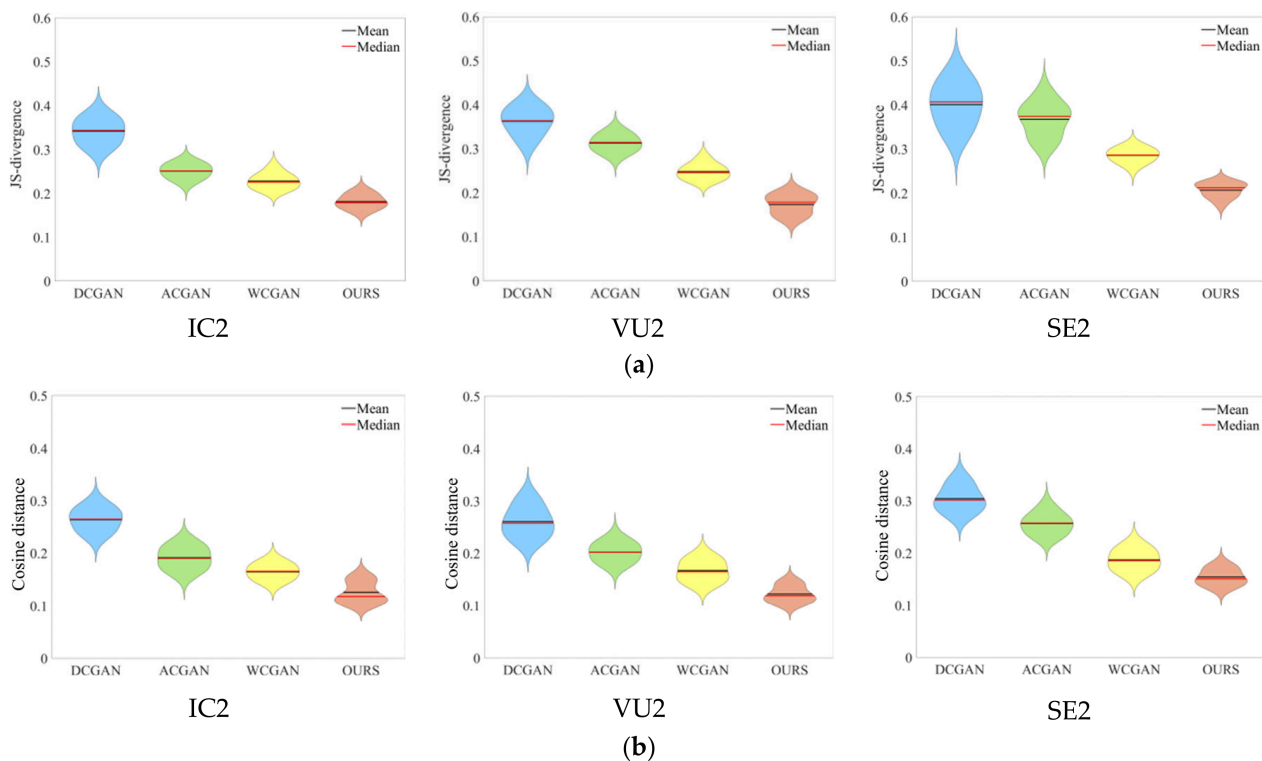


Figure 23. Violin plot of statistical indicators of Dataset B: (a) JS divergence, (b) cosine distance.

4.2.4. Diagnosis Results

(a) Fault diagnosis under limited training set

The proposed method was further evaluated on Dataset B under variable-speed conditions. In this study, the selected operating condition was a constant load of 20 Nm with a speed profile of 0–1000 rpm. The fault categories include NC, IC1, IC2, RB, VU1, VU2, SE1, SE2, RU, and RR. For each fault state, 400 samples were selected. Half of them were used for training and the other half were used for testing. Each sample was constructed from a 4800-point time-domain segment. The final spectral length of each modality was 2400. The dataset composition is summarized in Table 9.

To simulate class imbalance in the training set, the healthy class was kept with 200 training samples, while the number of training samples in each fault class was gradually reduced. As shown in Table 10, six imbalance ratios were considered, ranging from 1:50 to 1:1. Specifically, the minority fault classes contained 4, 10, 20, 40, 100, and 200 training samples, respectively. The test set remained balanced, with 200 samples for each class.

Table 9. Fault-type situations of Dataset B.

Label	Health Condition	Operating Condition	Fault Severity	Spectral Length (Points)	Training Samples	Testing Samples
0	NC	20 Nm, 0–1000 rpm	—	2400	200	200
1	IC1	20 Nm, 0–1000 rpm	Light	2400	200	200
2	IC2	20 Nm, 0–1000 rpm	High	2400	200	200
3	RB	20 Nm, 0–1000 rpm	—	2400	200	200
4	VU1	20 Nm, 0–1000 rpm	Light	2400	200	200
5	VU2	20 Nm, 0–1000 rpm	High	2400	200	200
6	SE1	20 Nm, 0–1000 rpm	Light	2400	200	200
7	SE2	20 Nm, 0–1000 rpm	High	2400	200	200
8	RU	20 Nm, 0–1000 rpm	—	2400	200	200
9	RR	20 Nm, 0–1000 rpm	—	2400	200	200

Table 10. Status of datasets under training set class imbalance.

Data Type	Dominant Class			Minority Class			
	0	1~9	1~9	1~9	1~9	1~9	1~9
Imbalance ratio	-	1:50	1:20	1:10	1:5	1:2	1:1
Training samples	200	4	10	20	40	100	200
Testing samples	200	200	200	200	200	200	200

To verify the effectiveness of the SWGAN, three representative GAN models, namely the DCGAN, ACGAN, and WCGAN, were selected for comparison. SAE without data augmentation was also used as the baseline. Each method first augmented the minority classes and then the augmented data were classified by SAE. All results were averaged over 20 runs.

As shown in Figure 24, the accuracy of all GAN-based methods increases as the imbalance ratio decreases. When the imbalance is severe, the performance gap among different GAN methods is relatively small, and SAE-only performs poorly because the number of minority fault samples is too limited. As the number of training samples increases, the advantage of SWGAN becomes gradually clearer. In particular, SWGAN achieves the best performance under the ratios of 1:10, 1:5, and 1:2. Under the balanced condition of 1:1, SWGAN still maintains the highest accuracy among all compared methods. These results indicate that the proposed method can provide more effective supplementary samples under variable-speed conditions.

To further analyze the classification performance, Figure 25 shows the confusion matrices under the imbalance ratios of 1:5 and 1:1. At the ratio of 1:5, most fault categories can already be identified correctly, but some confusion still exists between faults with similar characteristics, especially between IC1 and IC2, VU1 and VU2, SE1 and SE2, and RU and RR. When the ratio reaches 1:1, the diagonal values increase further and the off-diagonal errors are reduced. This indicates that the balanced training data allow the classifier to learn more discriminative features, thereby improving the fault diagnosis performance.

(b) Fault diagnosis under training set class imbalance

Class imbalance across fault categories was further evaluated on Dataset B. Two imbalance settings were constructed, denoted as Case C and Case D. Case C represents a moderate imbalance condition. Case D represents a more severe imbalance condition. The training samples of different fault categories were set to different numbers, while the test set was kept balanced with 200 samples for each class. The detailed settings are listed in Table 11.

Figure 26 shows the classification accuracy of five methods over 20 repeated trials. The SWGAN achieves the highest accuracy in both Case C and Case D. It also shows smaller fluctuations across repeated runs. When the imbalance becomes more severe in Case D, the performance of the comparison methods decreases more clearly. In contrast, the SWGAN maintains relatively stable accuracy. This indicates that the proposed method can provide more effective supplementary samples for minority fault categories under variable-speed conditions.



Figure 24. Diagnostic accuracy of Dataset B under different imbalance ratios.

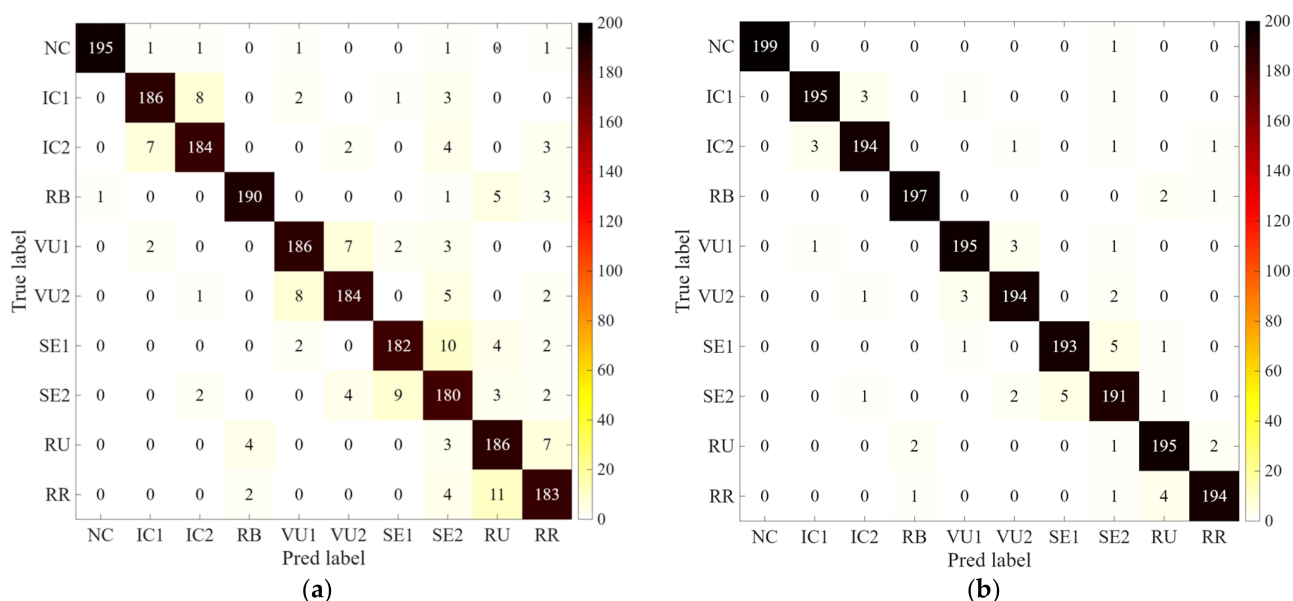


Figure 25. Categorical confusion matrix: (a) Imbalance ratio of 1:5, (b) imbalance ratio of 1:1.

Table 11. Situation of imbalanced Cases C and D.

Label	Health Condition	Operating Condition	Training Samples		Testing Samples
			Case C	Case D	Case C/D
0	NC	20 Nm, 0–1000 rpm	200	200	200
1	IC1	20 Nm, 0–1000 rpm	60	30	200
2	IC2	20 Nm, 0–1000 rpm	120	60	200
3	RB	20 Nm, 0–1000 rpm	100	50	200
4	VU1	20 Nm, 0–1000 rpm	60	30	200
5	VU2	20 Nm, 0–1000 rpm	120	60	200
6	SE1	20 Nm, 0–1000 rpm	60	30	200
7	SE2	20 Nm, 0–1000 rpm	120	60	200
8	RU	20 Nm, 0–1000 rpm	100	50	200
9	RR	20 Nm, 0–1000 rpm	100	50	200

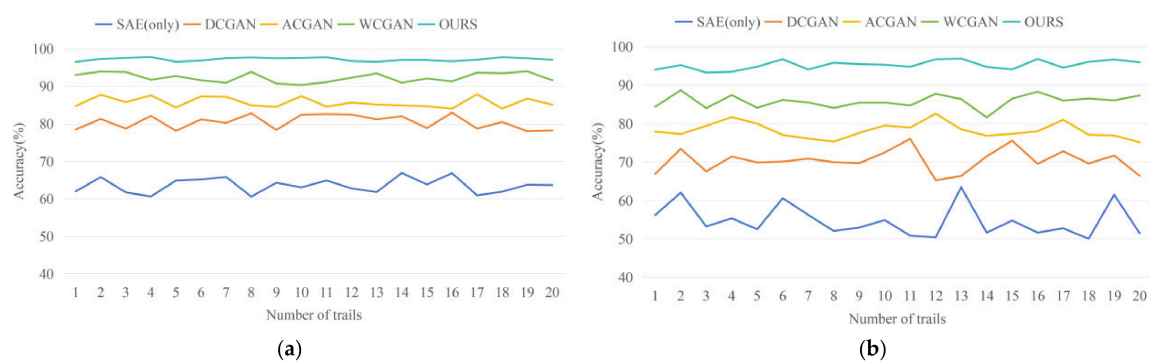
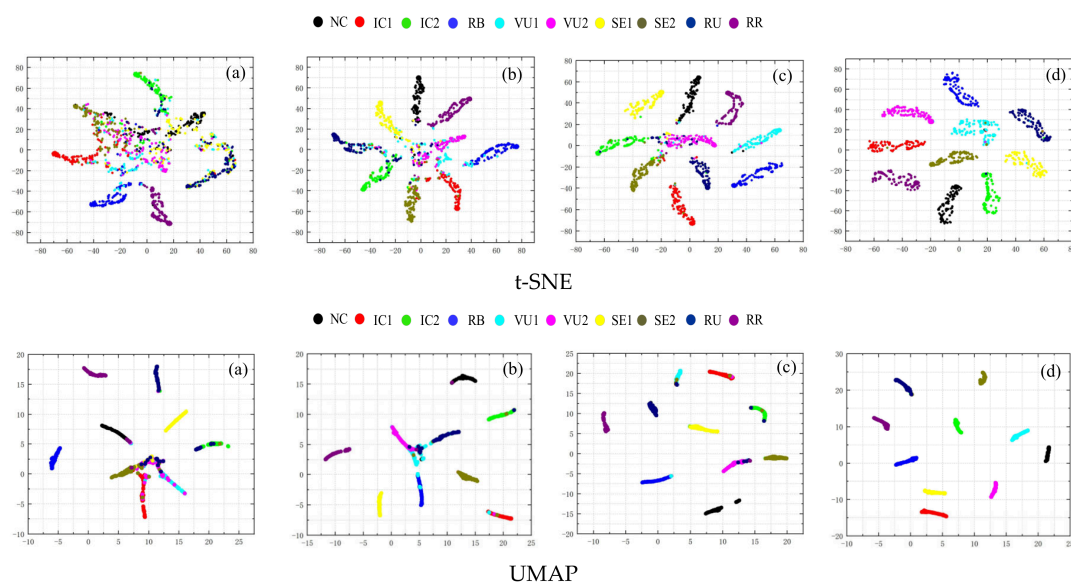
**Figure 26.** Performance comparison of five methods on two cases: (a) Case C, (b) Case D.

Figure 27 presents the t-SNE and UMAP visualization results under Case C. The DCGAN shows obvious overlap among several fault categories. The ACGAN improves the distribution to some extent, but the class boundaries are still unclear. The WCGAN shows better clustering results, although some scattered samples remain. In contrast, the SWGAN forms more compact clusters and clearer separation among different fault categories. This result is consistent with the classification results in Figure 25.

**Figure 27.** Visualization outcomes of dimensionality reduction on Case C: (a) DCGAN, (b) ACGAN, (c) WCGAN, (d) Ours.

5. Discussion

To further assess the robustness of the proposed method, an additional experiment was conducted on Dataset A using five representative conditions (NC, IT1, IT3, IC2, IC5). Six imbalance levels ranging from 1:50 to 1:1 were designed to reflect realistic variations in fault occurrence. Five comparison methods were used: (1) the proposed SWGAN with gradient normalization and self-attention; (2) a variant replacing GN with the conventional gradient penalty; (3) a version without self-attention; (4) a baseline using vibration-only data without augmentation; and (5) a baseline using current-only data without augmentation. For each imbalance setting, ten independent experiments were performed, and SAE was used to classify the augmented datasets. The average accuracy across trials was reported as the final evaluation metric. The specific experimental setups are shown in Tables 12–14.

Table 12. Situation of the imbalanced bearing datasets.

Label	Health Condition	Speed Condition (rpm)	Fault Severity (%)	Data Length (Points)	Training Samples	Testing Samples
0	NC	3000	-	2400	200	200
1	IT1	3000	2.36	2400	200	200
2	IT3	3000	3.35	2400	200	200
3	IC2	3000	0.81	2400	200	200
4	IC5	3000	2.00	2400	200	200

Table 13. Imbalanced status of bearing datasets.

Data Type	Dominant Class		Minority Class				
	0	1~4	1~4	1~4	1~4	1~4	1~4
Imbalance ratio	-	1:50	1:20	1:10	1:5	1:2	1:1
Training samples	200	4	10	20	40	100	200
Testing samples	200	200	200	200	200	200	200

Table 14. Experimental comparison methods.

NO.	Method
Model 1	SA + GN + Base
Model 2	SA + GP + Base
Model 3	GN + Base
Model 4	Vibration
Model 5	Current

Figure 28 shows the diagnostic accuracy of these five methods at different imbalance ratios. When the imbalance ratio is large, the diagnostic accuracies of all the methods are low because there are few valid available samples, but when the imbalance becomes gradually lower, the accuracy of the proposed method becomes gradually higher and outperforms all the other methods.

The loss curves of the discriminator and generator during the training process are shown in Figure 29; Figure 29a represents the variation in the loss value of Model 2 during the training process and Figure 29b represents the variation in the loss value of the proposed method during the training process. The loss curves give a rough indication of the performance of the network. In the first 120 steps, the discriminator and generator in the two models are all in an intense confrontation of continuous optimization, and the loss values constantly oscillate in a large range. However, after 120 steps, the loss values of the method proposed in this paper quickly reach dynamic equilibrium and converge—i.e., the

discriminator and generator reach “Nash equilibrium”—but Model 2 does not gradually equilibrate until after 160 steps. It can be seen that both models reach equilibrium within 200 steps of training, but the method proposed in this paper is significantly faster and more stable.

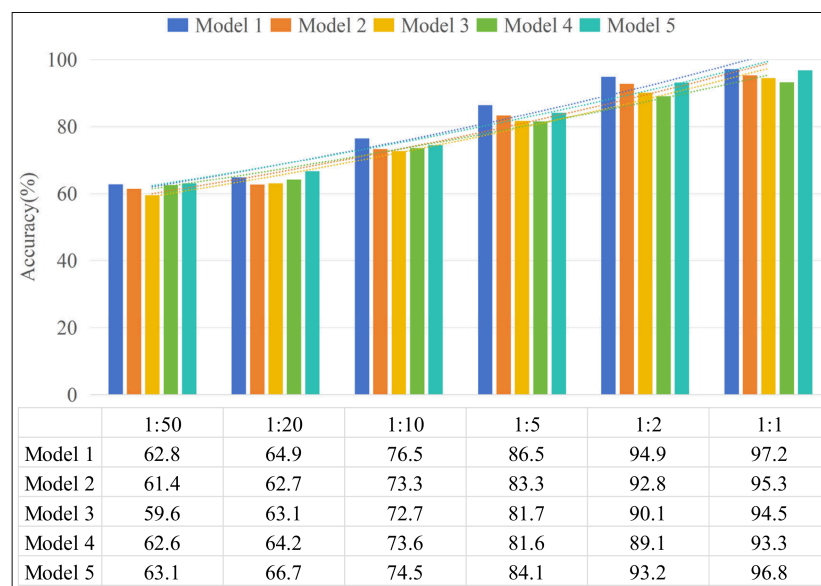


Figure 28. Diagnostic accuracy of five methods under different imbalance ratios.

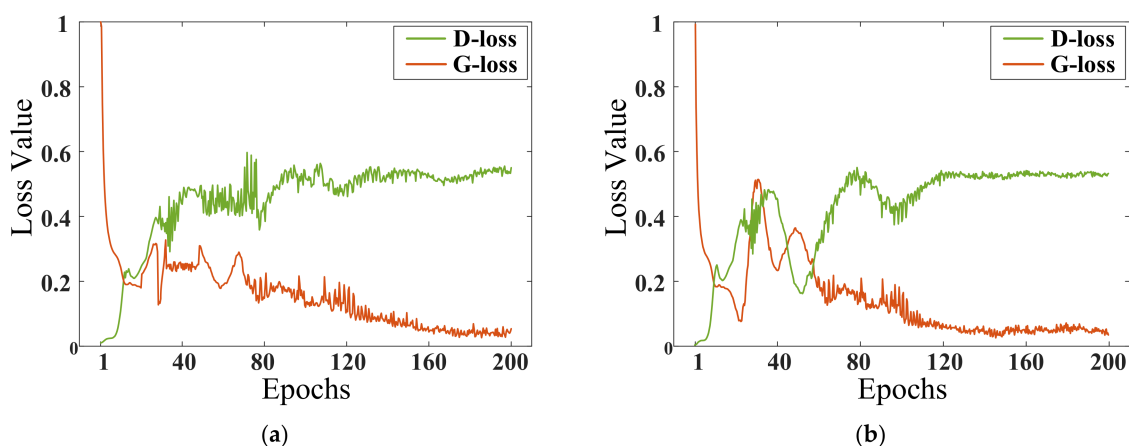


Figure 29. Changes in loss values under different methods: (a) Model 2, (b) Model 1.

6. Conclusions

This study investigates the problem of imbalanced data in fault diagnosis for electric motors and develops a multi-modal diagnostic framework based on an SWGAN. By introducing self-attention into a Wasserstein conditional generative adversarial network and incorporating gradient normalization, the model generates label-conditioned current and vibration spectral samples with distributions close to those of real fault data. The augmented dataset is then used to train an SAE-based classifier that exploits complementary information from electrical and mechanical measurements.

Experimental results on two distinct motor datasets, a fixed-speed PMSM dataset and a variable-speed induction motor dataset, demonstrate that the generated samples preserve the key statistical and structural characteristics of the original data. This is confirmed by both feature visualization and quantitative similarity analysis. The augmented data are evaluated under various data-scarce settings, including limited overall training samples

and class-level imbalance, and consistently improve diagnostic accuracy across different operating conditions while mitigating the performance degradation observed in conventional methods. Comparative and ablation experiments further reveal that self-attention enhances the capture of global spectral dependencies, while gradient normalization improves training stability and accelerates convergence. These findings confirm that the proposed framework effectively enhances fault diagnosis performance under severe sample imbalance and enables more reliable identification of minority fault classes.

Future work will focus on real-time online deployment and embedded hardware implementation of the proposed framework, so as to extend its applicability to practical industrial scenarios.

Author Contributions: Conceptualization, J.M. and C.Z.; methodology, J.M.; software, J.M.; validation, C.Z.; resources, C.C.; writing—original draft preparation, J.M.; writing—review and editing, C.Z.; supervision, G.J.; project administration, Y.Y.; funding acquisition, G.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the State Key Laboratory of Environmental Adaptability for Industrial Products, China National Electric Apparatus Research Institute Co., Ltd, under Grants 2025EASKJ-010.

Data Availability Statement: The dataset A used in this study was obtained from the publicly available dataset, released by a research team from the Korea Advanced Institute of Science and Technology (KAIST), the dataset is available at: <https://data.mendeley.com/datasets/rgn5brgrn/5>, accessed on 10 May 2026. The dataset B used in this study was obtained from the publicly available dataset, released by a research team from Tsinghua University, the dataset is available at: <https://data.mendeley.com/datasets/6s3dggj9mw/1>, accessed on 10 May 2026.

Conflicts of Interest: Authors Ganxin Jie, Yang Yang and Chuan Chen were employed by the company China National Electric Apparatus Research Institute Co., Ltd. and General Machinery Key Core Components Innovation Center (Anhui) Co., Ltd. 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. The authors declare that this study received funding from China National Electric Apparatus Research Institute Co., Ltd. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

1. Jieyang, P.; Kimmig, A.; Dongkun, W.; Niu, Z.; Zhi, F.; Jiahai, W.; Liu, X.; Ovtcharova, J. A systematic review of data-driven approaches to fault diagnosis and early warning. *J. Intell. Manuf.* **2023**, *34*, 3277–3304. [\[CrossRef\]](#)
2. Ahmadi, K.; Asadi, D.; Nabavi-Chashmi, S.-Y.; Tutsoy, O. Modified adaptive discrete-time incremental nonlinear dynamic inversion control for quad-rotors in the presence of motor faults. *Mech. Syst. Signal Process.* **2023**, *188*, 109989. [\[CrossRef\]](#)
3. Tutsoy, O.; Asadi, D.; Ahmadi, K.; Nabavi-Chasmi, S.-Y. Robust reduced order thau observer with the adaptive fault estimator for the unmanned air vehicles. *IEEE Trans. Veh. Technol.* **2023**, *72*, 1601–1610. [\[CrossRef\]](#)
4. Badihi, H.; Zhang, Y.; Jiang, B.; Pillay, P.; Rakheja, S. A comprehensive review on signal-based and model-based condition monitoring of wind turbines: Fault diagnosis and lifetime prognosis. *Proc. IEEE* **2022**, *110*, 754–806. [\[CrossRef\]](#)
5. Huang, R.; Xia, J.; Zhang, B.; Chen, Z.; Li, W. Compound fault diagnosis for rotating machinery: State-of-the-art, challenges, and opportunities. *J. Dyn. Monit. Diagn.* **2023**, *2*, 13–29. [\[CrossRef\]](#)
6. Han, B.; Zhang, X.; Wang, J.; An, Z.; Jia, S.; Zhang, G. Hybrid distance-guided adversarial network for intelligent fault diagnosis under different working conditions. *Measurement* **2021**, *176*, 109197. [\[CrossRef\]](#)
7. Tan, C.; Yang, L.; Chen, H.; Xin, L. Fault diagnosis method for rolling bearing based on VMD and improved SVM optimized by METLBO. *J. Mech. Sci. Technol.* **2022**, *36*, 4979–4991. [\[CrossRef\]](#)
8. Fang, H.; Liu, H.; Wang, X.; Deng, J.; An, J. The method based on clustering for unknown failure diagnosis of rolling bearings. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3509508. [\[CrossRef\]](#)
9. Zhao, X.; Yao, J.; Deng, W.; Jia, M.; Liu, Z. Normalized conditional variational auto-encoder with adaptive focal loss for im-balanced fault diagnosis of bearing-rotor system. *Mech. Syst. Signal Process.* **2022**, *170*, 108826. [\[CrossRef\]](#)

10. Zhu, Z.; Lei, Y.; Qi, G.; Chai, Y.; Mazur, N.; An, Y.; Huang, X. A review of the application of deep learning in intelligent fault diagnosis of rotating machinery. *Measurement* **2023**, *206*, 112346. [\[CrossRef\]](#)
11. He, D.; Xu, Y.; Sun, H.; Jin, Z.; Zhao, M. Self-supervised learning for vehicle bearing fault diagnosis based on time–frequency dual-domain contrast and fusion. *Nonlinear Dyn.* **2025**, *113*, 17385–17412. [\[CrossRef\]](#)
12. Wu, G.; Huang, Z.; Long, Z.; Huang, F.; Wang, M.-H.; Zhang, X. Motor fault diagnosis method based on spiking convolutional neural network with multi-scale decomposition local features. *ISA Trans.* **2025**, *164*, 271–283. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Li, H.; Yao, L. Motor fault diagnosis based on generative adversarial network using hyperchaotic sequences and mixed-dimensional network. *IEEE Trans. Ind. Inform.* **2025**, *21*, 3216–3225. [\[CrossRef\]](#)
14. Lee, S.K.; Kim, H.; Chae, M.; Oh, H.J.; Yoon, H.; Youn, B.D. Self-supervised feature learning for motor fault diagnosis under various torque conditions. *Knowl.-Based Syst.* **2024**, *288*, 111465. [\[CrossRef\]](#)
15. Xu, Y.; He, D.; Sun, H.; Jin, Z.; Zhao, M. Self-supervised learning for train bearing fault diagnosis based on time–frequency dual domain prediction. *Struct. Health Monit.* **2026**, 14759217251405584. [\[CrossRef\]](#)
16. Hamani, K.; Kuchar, M.; Kubatko, M.; Kirschner, S. Advancements in Induction Motor Fault Diagnosis and Condition Monitoring: A Comprehensive Review. *Sensors* **2025**, *25*, 5942. [\[CrossRef\]](#)
17. Yi, L.; Huang, Y.; Zhan, J.; Wang, Y.; Sun, T.; Long, J.; Liu, J.; Chen, B. CNN-ELMNet: Fault diagnosis of induction motor bearing based on cross-modal vector fusion. *Meas. Sci. Technol.* **2024**, *35*, 115114. [\[CrossRef\]](#)
18. Dehnavi, V.S.; Shafiee, M. Fault diagnosis of induction motors using novel measurement techniques and data fusion. *Measurement* **2025**, *256*, 118135. [\[CrossRef\]](#)
19. Li, T.; Qiao, Z.; Kumar, A.; Xie, C.; Zhang, C.; Lai, Z. A hydraulic motor fault diagnosis method based on weighted mul-ti-channel information fusion. *Meas. Sci. Technol.* **2025**, *36*, 015120. [\[CrossRef\]](#)
20. Huang, X.; Jin, T.; Liu, Z.; Chen, C.; Hua, C.; Zhang, L. An interpretable multi-superlet kernel fusion convolutional neural network for rotating machinery fault diagnosis. *Expert Syst. Appl.* **2025**, *290*, 128484. [\[CrossRef\]](#)
21. Gawde, S.; Patil, S.; Kumar, S.; Kotecha, K. A scoping review on multi-fault diagnosis of industrial rotating machines using multi-sensor data fusion. *Artif. Intell. Rev.* **2023**, *56*, 4711–4764. [\[CrossRef\]](#)
22. Wang, Z.; Liu, T.; Wu, X.; Liu, C. A diagnosis method for imbalanced bearing data based on improved SMOTE model combined with CNN-AM. *J. Comput. Des. Eng.* **2023**, *10*, 1930–1940. [\[CrossRef\]](#)
23. Wei, J.; Huang, H.; Yao, L.; Hu, Y.; Fan, Q.; Huang, D. New imbalanced bearing fault diagnosis method based on Sam-ple-characteristic Oversampling TechniquE (SCOTE) and multi-class LS-SVM. *Appl. Soft Comput.* **2021**, *101*, 107043. [\[CrossRef\]](#)
24. Han, Z.; Wang, H.; Shen, C.; Song, X.; Cao, L.; Yu, L. Attention features selection oversampling technique (AFS-O) for rolling bearing fault diagnosis with class imbalance. *Meas. Sci. Technol.* **2023**, *35*, 035002. [\[CrossRef\]](#)
25. Tian, J.; Jiang, Y.; Zhang, J.; Luo, H.; Yin, S. A novel data augmentation approach to fault diagnosis with class-imbalance problem. *Reliab. Eng. Syst. Saf.* **2024**, *243*, 109832. [\[CrossRef\]](#)
26. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*; NeurIPS: San Diego, CA, USA, 2014.
27. Chen, X.; Duan, Y.; Houthoofd, R.; Schulman, J.; Sutskever, I.; Abbeel, P. Infogan: Interpretable representation learning by in-formation maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems 29*; NeurIPS: San Diego, CA, USA, 2016.
28. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784. [\[CrossRef\]](#)
29. Feng, J.; Yu, H.; Wang, L.; Cao, X.; Zhang, X.; Jiao, L. Classification of hyperspectral images based on multiclass spatial–spectral generative adversarial networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5329–5343. [\[CrossRef\]](#)
30. Jiang, K.; Xie, W.; Li, Y.; Lei, J.; He, G.; Du, Q. Semisupervised spectral learning with generative adversarial network for hy-perspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 5224–5236. [\[CrossRef\]](#)
31. Dixit, S.; Verma, N.K.; Ghosh, A.K. Intelligent fault diagnosis of rotary machines: Conditional auxiliary classifier GAN coupled with meta learning using limited data. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3517811. [\[CrossRef\]](#)
32. Chen, X.; Chen, Z. Cross-domain fault diagnosis of induction motor based on an improved unsupervised generative adversarial network and fine-tuning under limited labeled data. *Measurement* **2025**, *249*, 116988. [\[CrossRef\]](#)
33. Liu, J.; Zhang, C.; Jiang, X. Imbalanced fault diagnosis of rolling bearing using improved MsR-GAN and feature enhancement-driven CapsNet. *Mech. Syst. Signal Process.* **2022**, *168*, 108664. [\[CrossRef\]](#)
34. Wang, H.; Li, P.; Lang, X.; Tao, D.; Ma, J.; Li, X. FTGAN: A novel GAN-based data augmentation method coupled time–frequency domain for imbalanced bearing fault diagnosis. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3502614. [\[CrossRef\]](#)
35. Liu, Y.; Jiang, H.; Liu, C.; Yang, W.; Sun, W. Data-augmented wavelet capsule generative adversarial network for rolling bearing fault diagnosis. *Knowl.-Based Syst.* **2022**, *252*, 109439. [\[CrossRef\]](#)
36. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning*; PMLR: New York, NY, USA, 2017.

37. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. *arXiv* **2018**, arXiv:1802.05957. [[CrossRef](#)]
38. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems 30*; NeurIPS: San Diego, CA, USA, 2017.
39. Bhaskara, V.S.; Aumentado-Armstrong, T.; Jepson, A.D.; Levinshtein, A. GraN-GAN: Piecewise gradient normalization for generative adversarial networks. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; IEEE: New York, NY, USA, 2022.
40. Wu, Y.-L.; Shuai, H.-H.; Tam, Z.-R.; Chiu, H.-Y. Gradient normalization for generative adversarial networks. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2021.
41. Jung, W.; Yun, S.-H.; Lim, Y.-S.; Cheong, S.; Park, Y.-H. Vibration and current dataset of three-phase permanent magnet syn-chronous motors with stator faults. *Data Brief* **2023**, *47*, 108952. [[CrossRef](#)] [[PubMed](#)]
42. Zhong, H.; Yu, S.; Trinh, H.; Lv, Y.; Yuan, R.; Wang, Y. Fine-tuning transfer learning based on DCGAN integrated with self-attention and spectral normalization for bearing fault diagnosis. *Measurement* **2023**, *210*, 112421. [[CrossRef](#)]
43. Li, W.; Zhong, X.; Shao, H.; Cai, B.; Yang, X. Multi-mode data augmentation and fault diagnosis of rotating machinery using modified ACGAN designed with new framework. *Adv. Eng. Inform.* **2022**, *52*, 101552. [[CrossRef](#)]
44. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
45. Healy, J.; McInnes, L. Uniform manifold approximation and projection. *Nat. Rev. Methods Primers* **2024**, *4*, 82. [[CrossRef](#)]
46. Chen, S.; Liu, Z.; Li, C.; Zou, D.; He, X.; Zhou, D. Multi-mode fault diagnosis datasets of three-phase asynchronous motor under variable working conditions. *Data Brief* **2026**, *65*, 112583. [[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.