

Article

Fractional Diversity Entropy: A Vibration Signal Measure to Assist a Diffusion Model in the Fault Diagnosis of Automotive Machines

Baohua Wang ^{1,2,3,*} , Jiacheng Zhang ^{1,2,*} , Weilong Wang ^{1,2,4}  and Tingting Cheng ^{1,2}

¹ College of Automotive Engineering, Hubei University of Automotive Technology, Shiyan 442002, China; 202111136@huat.edu.cn (W.W.); chengtt_qc@huat.edu.cn (T.C.)

² Hubei Key Laboratory of Automotive Power Train and Electric Control, Hubei University of Automotive Technology, Shiyan 442002, China

³ Hubei Longzhong Laboratory, Xiangyang 441106, China

⁴ Department of Automotive, Hubei Hanjiang Technician College, Shiyan 442013, China

* Correspondence: 19950009@huat.edu.cn (B.W.); 202211134@huat.edu.cn (J.Z.); Tel.: +86-13997810948 (B.W.); +86-15771099091 (J.Z.)

Abstract: Real-world vibration signal acquisition of automotive machines often results in imbalanced sample sets due to restricted test conditions, adversely impacting fault diagnostic accuracy. To address this problem, we propose fractional diversity entropy (FrDivEn) and incorporate it into the classifier-guided diffusion model (CGDM) to synthesize high-quality samples. Additionally, we present a corresponding imbalanced fault diagnostic method. This method first converts vibration data to Gramian angular field (GAF) image samples through GAF transformation. Then, FrDivEn is mapped to the gradient scale of CGDM to trade off the diversity and fidelity of synthetic samples. These synthetic samples are mixed with real samples to obtain a balanced sample set, which is fed to the fine-tuned pretrained ConvNeXt for fault diagnosis. Various sample synthesizers and fault classifiers were combined to conduct imbalanced fault diagnosis experiments across bearing, gearbox, and rotor datasets. The results indicate that for the three datasets, the diagnostic accuracies of the proposed CGDM using FrDivEn at an imbalance ratio of 40:1 are 91.22%, 87.90%, and 98.89%, respectively, which are 7.32%, 11.59%, and 3.48% higher than that of the Wasserstein generative adversarial network (WGAN), respectively. The experimental results across the three datasets validated the validity and generalizability of the proposed diagnostic method.

Keywords: fractional order; diversity entropy; fault diagnosis; diffusion model; ConvNeXt model



Citation: Wang, B.; Zhang, J.; Wang, W.; Cheng, T. Fractional Diversity Entropy: A Vibration Signal Measure to Assist a Diffusion Model in the Fault Diagnosis of Automotive Machines. *Electronics* **2024**, *13*, 3155. <https://doi.org/10.3390/electronics13163155>

Academic Editor: Felipe Jiménez

Received: 14 July 2024

Revised: 6 August 2024

Accepted: 8 August 2024

Published: 9 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Automotive machines, such as bearings, gearboxes, and rotors, play an indispensable role in transportation vehicles. Motor bearing, gearbox, and rotor failures significantly impact vehicle driving safety. When a vehicle experiences a malfunction, pinpointing the exact failing component is often challenging, necessitating the disassembly of the assembly to identify the issue [1]. This process hinders the efficiency and convenience of fault determination. In addition, as technology advances, the structure of these automotive machines is becoming increasingly sophisticated, and their application scenarios are becoming more complex. This complexity increases the susceptibility to various failures or damages. Such failures can reduce the operational efficiency of mechanical equipment and cause vehicle shutdowns. Hence, studying fault diagnostic methods for automotive machines holds significant theoretical and engineering value.

Traditional methods using pure signal processing techniques are still theoretically valid, but they have significant room for improvement in intelligence [2,3]. Meanwhile, deep learning (DL) can extract features from complex vibration, sound, or other sensor data and learn patterns between different states of an automotive machine, leading to

more accurate and early diagnosis of automotive machine faults [4]. Deep neural networks for fault diagnosis mainly consist of the recurrent neural network (RNN), auto-encoder (AE), and convolutional neural network (CNN). These networks do not require empirical knowledge and can extract feature information from data or samples adaptively. Among these models, CNNs utilize multiple convolutional kernels to extract features from input samples or upper-layer features [5,6]. CNNs perform matrix element multiplication summation and accumulate the deviations of the input features. Compared to RNNs and AEs, CNNs have notable advantages due to their increased network depth and weight sharing. The former, with deep residual learning [7], allows CNNs to increase their depth without encountering the vanishing gradient problem common in RNNs [8], thus enabling effective feature extraction from sensor time series data over longer periods. The latter reduces the number of trainable parameters, lowering model complexity and mitigating the overfitting problem prevalent in AEs [9], which augments the generalization of models [10,11].

With outstanding advantages, CNNs are widely used in fault diagnosis. For instance, an improved CNN model incorporating empirical mode decomposition (EMD) was proposed to enable end-to-end diagnosis, enhancing accuracy and anti-interference capabilities [12]. Additionally, a diagnostic method based on persistence spectrum imaging and the residual network (ResNet) structure was proposed. The improved ResNet structure allows for direct connections between different feature maps, facilitating the extraction of discriminant features [13]. Moreover, the CBAM-ResNet, which comprised the convolutional block attention module (CBAM) and a modified ResNet, was created to improve network feature extraction efficiency while maintaining high accuracy [14].

These DL-based fault diagnostic methods perform well and can accurately recognize different types of faults. However, these methods depend on large and balanced samples. In real-world conditions, data acquisition for automotive machines is often limited due to practical constraints. Moreover, since automotive machines typically operate in a normal state most of the time, it is easier to collect a substantial amount of normal-state data, leading to an imbalance in fault-type data. This imbalance restricts the diagnostic accuracy [15].

To fill the imbalanced sample sets, DL-based generative models are employed to synthesize the missing samples for each faulty class, thereby achieving a balanced dataset. The generative models used for this purpose mainly include the variational auto-encoder (VAE), generative adversarial network (GAN), and diffusion model. Among these generative models, the diffusion model's noising process is a multi-step procedure that gradually applies noise to samples. Conversely, the reverse denoising process is multi-step, gradually removing noise from the sample. This dual process allows the diffusion model to fulfill two key objectives: (1) starting from random noise samples ensures the diversity of the synthesized samples; (2) the gradual denoising process allows for meticulous control, enhancing the fidelity of the synthetic samples and avoiding the low fidelity issues of VAE-synthesized samples and the GAN's collapse [16,17].

Currently, diffusion models are developing rapidly in sample synthesis tasks. Dhariwal et al. [18] introduced the classifier-guided mechanism to diffusion models, proposing the classifier-guided diffusion model (CGDM) in 2021. This model can outperform the GAN by using a gradient scale that weighs the focus on the diversity and fidelity. We aimed to augment the imbalanced fault sample set by synthesizing specific ones using the CGDM. The CGDM allows for a trade-off between the diversity and fidelity of synthetic samples by adjusting the gradient scale s , offering the potential to improve the quality of synthetic fault samples under varying imbalance ratios. Nevertheless, determining the proper value of the gradient scale as a hyperparameter remains an important question. Dhariwal et al. [18] suggested that the gradient scale should be set at an intermediate point where the overall sample quality, considering diversity and fidelity, is highest. We sought a clearer indicator for setting the gradient scale than the vague "intermediate point".

Since DivEn can reflect the diversity of time series [19], it and its variants have been applied in machine fault diagnosis by extracting time series' features [20–22]. However, they have not yet been reported in imbalanced fault diagnosis. We aimed to map the

DivEn of machine signals at different imbalance ratios to an appropriate gradient scale for high-quality sample syntheses. Nevertheless, DivEn is not sensitive to time series at different imbalance ratios, as illustrated in Section 2.2. To address this problem, we propose fractional diversity entropy (FrDivEn) by incorporating fractional order calculus into DivEn. Fractional calculus generalizes the classical theory of integer-order differentiation and integration. Both the theory and its applications demonstrate that the fractional calculus operator effectively describes many complex systems. Due to its numerous unique characteristics, fractional calculus is extensively studied in fields like signal processing [23,24], image processing [25,26], and machine learning [27,28]. We propose FrDivEn to sensitively reflect the vibration signals' diversity and to better balance the diversity and fidelity of the synthetic samples.

Furthermore, we present a novel imbalanced diagnostic method for automotive machines by integrating the CGDM with a gradient scale corresponding to FrDivEn, the Gramian angular field (GAF) transformation, and the fine-tuned pretrained ConvNeXt. Specifically, the method first transforms time series signals into GAF image samples using GAF transformation. Then, high-quality samples are synthesized using the CGDM with a gradient scale corresponding to FrDivEn. These synthetic samples are combined with imbalanced real samples to obtain a mixed sample set, which is subsequently fed to a fine-tuned pretrained ConvNeXt for fault diagnosis. The contributions of this paper are summarized as follows:

- (1) For fault diagnosis with imbalanced sample sets in automotive machines, it is necessary to balance the diversity and fidelity of synthetic samples. Here, we innovatively propose a novel vibration signal measure, fractional diversity entropy (FrDivEn), to reflect signal diversity and adjust the generative model's emphases on the diversity and fidelity of sample synthesis. The proposed FrDivEn differs from the traditional DivEn, which is insensitive to signal diversity at different imbalance ratios. FrDivEn can sensitively vary with the imbalance ratio of the signal, reflecting signal diversity more efficiently than the traditional DivEn.
- (2) To select the appropriate gradient scale of CGDM accordingly and achieve high-quality sample synthesis, we innovatively propose using FrDivEn to determine the ideal gradient scale. This approach results in better sample synthesis compared to other generative models.
- (3) To boost diagnostic accuracy in automotive machines, we present a fault diagnostic method. This method primarily uses CGDM with a gradient scale corresponding to FrDivEn as the sample synthesizer and a fine-tuned pretrained ConvNeXt as the fault classifier. Experiments show that this method can be extended to various automotive machines and achieves higher diagnostic accuracy compared to other sample synthesizer and fault classifier methods.

The remainder of this paper is arranged as follows: The algorithms of DivEn and FrDivEn are provided in Section 2. The proposed imbalanced diagnostic method is described in Section 3. The experiments designed to verify the validity and generalizability of the method are presented in Section 4. Finally, the conclusions of this study are drawn in Section 5.

2. Algorithms

In this section, the algorithms for the diversity entropy (DivEn) and the proposed fractional diversity entropy (FrDivEn) are presented.

2.1. Diversity Entropy

For a given time series $X = \{x_1, \dots, x_i, \dots, x_N\}$, the diversity entropy (DivEn) can be derived according to the following steps.

Step 1: Phase space reconstruction. The time series can be reconstructed into orbits using an embedding dimension m [29]. This reconstruction involves creating subsequences. It allows for analysis of the system's dynamics by examining the geometric properties of the reconstructed phase space. X is divided into $(N - m + 1)$ subsequences. Each subsequence

$y_i(m)$ is formed as $\{x_i, x_{i+1}, \dots, x_{i+m-1}\}$. The reconstructed matrix Y consists of rows that are segments of X . The matrix Y is structured with each row representing a segment of length m from the time series. This matrix can reveal patterns and structures that are not apparent in the original time series, facilitating further analysis.

The phase space reconstruction matrix Y is given by

$$Y(m) = \{y_1(m), y_2(m), \dots, y_{N-m+1}(m)\} \quad (1)$$

In matrix form,

$$Y(m) = \begin{bmatrix} x_1 & x_2 & \dots & x_m \\ x_2 & x_3 & \dots & x_{1+m} \\ \vdots & \vdots & \dots & \vdots \\ x_{N-m+1} & x_{N-m+2} & \dots & x_N \end{bmatrix} \quad (2)$$

The rows of this matrix correspond to the subsequences.

Step 2: Cosine similarity calculation. The similarity between each row and the next row in the phase space matrix is calculated to yield a set of similarities $\{d_1, d_2, \dots, d_{N-m}\}$. This series of similarities helps in understanding the relations between successive states in the reconstructed phase space. The cosine similarity d between adjacent rows is defined mathematically. It calculates the cosine of the angle between two non-zero vectors, reflecting their directional alignment.

The series of cosine similarities is given by

$$D(m) = \{d_1, d_2, \dots, d_{N-m}\} \quad (3)$$

where

$$D(m) = \{d(y_1(m), y_2(m)), d(y_2(m), y_3(m)), \dots, d(y_{N-m}(m), y_{N-m+1}(m))\} \quad (4)$$

The similarity between two rows $y_i(m)$ and $y_j(m)$ is defined as

$$d(y_i(m), y_j(m)) = \frac{\sum_{k=1}^m y_i(k) \times y_j(k)}{\sqrt{\sum_{k=1}^m y_i(k)^2} \times \sqrt{\sum_{k=1}^m y_j(k)^2}} \quad (5)$$

The cosine similarity d ranges from -1 to 1 . A value of 1 indicates that the two rows are identical, 0 indicates that they are orthogonal (no similarity), and -1 indicates that they are completely opposite. High cosine similarity values indicate similar dynamic changes between two rows, while low values indicate diverse dynamic behavior.

Step 3: State probability calculation. The range $[-1, 1]$ is partitioned into ε intervals denoted as $(I_1, I_2, \dots, I_\varepsilon)$. This partitioning allows for the categorization of cosine similarity values into discrete intervals, facilitating the calculation of state probabilities. The state probabilities $(p_1, p_2, \dots, p_\varepsilon)$ are calculated by determining the frequency of cosine similarity values and normalizing by the values. The sum of state probabilities is equal to 1 , i.e., $\sum_{k=1}^{\varepsilon} p_k = 1$. This ensures that the probabilities are properly normalized, making the distribution valid and interpretable.

Step 4: Diversity entropy calculation. DivEn is calculated based on the state probabilities obtained from the partitioned cosine similarities using the following formula:

$$DivEn = -\frac{1}{\ln \varepsilon} \sum_{k=1}^{\varepsilon} p_k \ln p_k \quad (6)$$

where ε means the number of intervals, and p_k are the elements of the state probability.

DivEn is the expectation of the diversity between the rows of the phase space matrix. It quantifies how evenly the cosine similarities are distributed. The range of DivEn is $[0, 1]$,

according to the original entropy theory [30]. When DivEn tends to 0, this indicates low complexity in the time series, suggesting a dynamic system with similar phenomena or repetitive patterns. When DivEn tends to 1, this indicates high complexity in the time series, suggesting a dynamic system with diverse phenomena or more varied behavior.

2.2. Fractional Diversity Entropy

DivEn can characterize the diversity of time series. However, we find limited differentiation in DivEn calculation results for vibrational signals with different imbalance ratios, which reduces DivEn’s effectiveness in measuring the diversity of time series.

To illustrate this problem, the CWRU bearing dataset with a motor load of 0 and a speed of 1797 rpm is used as an example. The vibration signals are cut and spliced at different ratios for the nine fault states to create five time series. These time series simulate scenarios where the occurrence of faults is not uniform, providing a basis for analyzing how imbalance affects system dynamics. The specific allocation of bearing faults in the imbalanced time series is detailed in Table A1, Appendix A. Following the methodology in the original research on DivEn [19], the embedding dimension m is set to 4. DivEn is calculated for each of the five time series, allowing for a comparative analysis of their complexity and diversity. Similar to the original research on DivEn [19], we set the embedding dimension m to 4 and calculated DivEn for each of the above five time series. The results of the DivEn calculations are shown in Table 1. It can be found that the DivEn results of two adjacent imbalance ratio vibration signals have a limited difference of no more than 0.01. The calculation results reflect the limitations of DivEn in characterizing imbalance ratio fault vibration data, and DivEn is ill-equipped to reflect the diversity of time series vibration signals clearly.

Table 1. The results of the DivEn calculations.

Imbalance Ratio	2:1	5:1	10:1	20:1	40:1
Measure	2:1	5:1	10:1	20:1	40:1
DivEn	0.9496	0.9413	0.9357	0.9287	0.9195
Difference	0.0083	0.0056	0.0070	0.0092	

To address this problem and make DivEn sensitive to different imbalance ratio vibration signals, we combine DivEn with fractional order calculus to propose fractional diversity entropy (FrDivEn), which measures the diversity of time series. The improved algorithm for DivEn, called FrDivEn, is derived from DivEn and Shannon entropy at a fractional order α . FrDivEn extends the concept of DivEn to incorporate fractional calculus, enhancing the measurement of system complexity.

Shannon entropy is extended to consider fractional calculus [31], referred to as ShannonEn $_{\alpha}$. This extension allows for a more flexible and detailed analysis of the underlying dynamics of time series data.

The generalized expression for ShannonEn $_{\alpha}$ is given by

$$ShannonEn_{\alpha} = - \sum_s \left\{ \frac{p_s^{-\alpha}}{\Gamma(\alpha + 1)} [\ln p_s + \psi(1) - \psi(1 - \alpha)] \right\} p_s \tag{7}$$

where α denotes the fractional order, $\Gamma(\cdot)$ represents the gamma function, $\psi(\cdot)$ represents the digamma function, and p_s are the elements of the state probability in Shannon entropy calculation. This formula introduces fractional exponents and special functions to adjust the traditional entropy calculation.

FrDivEn is generalized based on the generalized expression of Shannon entropy and fluctuation-based calculus [32]. This extension enhances the traditional entropy measures by incorporating the concept of fractional calculus, providing a more detailed analysis of time series data.

FrDivEn at fractional order α is written as FrDivEn_α , defined as

$$\text{FrDivEn}_\alpha = -D^\alpha \text{FrDivEn} \quad (8)$$

where $D^\alpha(\cdot)$ denotes the derivative of fractional order α , introducing the concept of fractional differentiation into the entropy calculation.

Combining with Equation (6), the FrDivEn_α for the original time series $\mathbf{X} = \{x_1, \dots, x_i, \dots, x_N\}$ is given by

$$\text{FrDivEn}_\alpha = -\sum_{k=1}^{\varepsilon} \left\{ \frac{p_k^{-\alpha}}{\Gamma(\alpha + 1)} [\ln p_k + \psi(1) - \psi(1 - \alpha)] \right\} p_k \quad (9)$$

where α denotes the fractional order ($-1 < \alpha < 1$), $\Gamma(\cdot)$ represents the gamma function, $\psi(\cdot)$ represents the digamma function, ε denotes the number of intervals, and p_k are the elements of the state probability. This formula allows for a more adaptable and comprehensive calculation of entropy, reflecting the patterns of diversity within the time series data.

3. Proposed Imbalanced Fault Diagnostic Method

To enhance automotive machine diagnostic accuracy on limited and imbalanced fault data, we propose an innovative fault diagnostic method that introduces FrDivEn to trade off the classifier-guided diffusion model's (CGDM) sample synthesis.

First, to fully utilize the advantages of convolutional neural networks (CNNs) in image classification [33] for fault diagnosis, Gramian angular field (GAF) transformation is employed to convert the raw vibration signals of automotive machines into GAF images. Then, to balance the number of GAF images for each fault state, the CGDM with the FrDivEn trade-off is applied to synthesize high-quality GAF images. Next, the real samples and synthetic samples are combined into a single balanced sample set. Finally, to achieve highly accurate fault diagnosis, a fine-tuned ConvNeXt model based on transfer learning is implemented. For a given automotive machine vibration signal collection platform, four processes need to be performed.

3.1. Preprocess

In automotive machine fault diagnosis, the raw vibration signals collected by the sensor are converted into GAF images through GAF transformation. The GAF transformation is a time series data analysis coding method that enhances tasks such as classification and imputation [34]. The basic idea of GAF involves combining the coordinate transformation and the Gramian matrix. The detailed derivation of GAF transformation is provided in the Appendix B section. In this fault diagnostic method, GAF transformation is used to extract the temporal and numerical relationships of the vibration signals, representing these relationship features as GAF images.

3.2. Sample Synthesis

After obtaining real image samples, to fill the imbalanced sample set, we use a classifier-guided diffusion model (CGDM) with the assistance of FrDivEn to synthesize samples. The CGDM comprises noising and denoising processes: (1) Noising process (q): The random noise β is added to the original image sample x_0 gradually, resulting in a purely noise image x_T after T steps. (2) Denoising process (p): The noise is progressively removed from the noise image according to the conditional distribution p_θ , yielding the synthetic image sample after T steps. As implied by its name, the CGDM incorporates a classifier ϕ to guide sample synthesis. By guiding the diffusion model, overall sample quality is enhanced by balancing diversity and fidelity. Dhariwal et al. [18] found that increasing the classifier's gradient scale s boosts fidelity at the cost of diversity in synthetic samples, introducing a trade-off between sample fidelity and diversity. For instance, high fidelity for bearings means that the CGDM can accurately synthesize specific fault samples, such as ball or race faults, but this reduces the overall diversity of synthetic samples. Therefore, adjusting the

gradient scale offers a trade-off between the diversity and fidelity of synthetic samples. The schematic diagram of CGDM is shown in Figure 1.

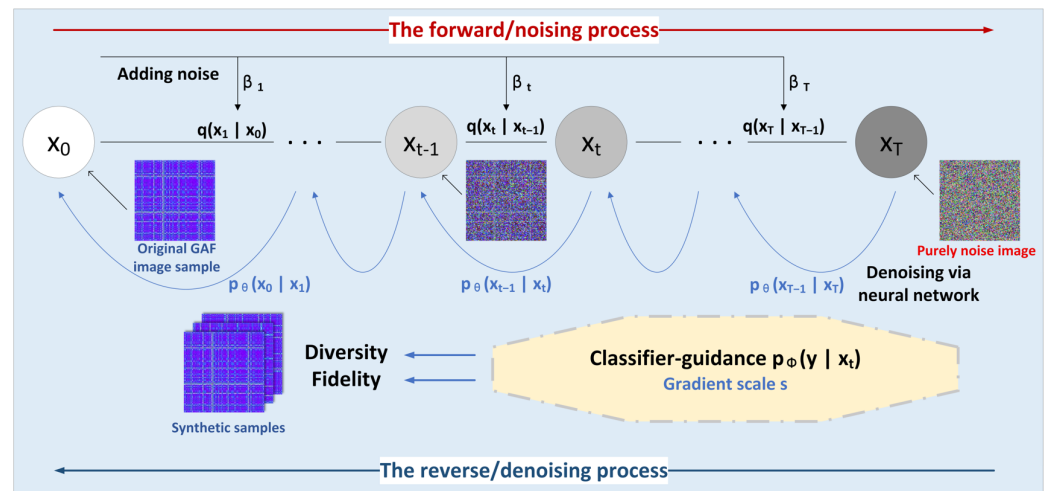


Figure 1. Schematic diagram of a classifier-guided diffusion model (CGDM).

However, there are no ideal measures to pick the appropriate gradient scale s . Given that the proposed FrDivEn can sensitively represent the diversities of vibration signals, we propose to tune the CGDM's gradient scale for high-quality sample synthesis using FrDivEn. The combination of FrDivEn and the CGDM aims to achieve a better trade-off in fault sample diversity and fidelity. The FrDivEn trade-off in diversity and fidelity can be described as the following three steps.

Step 1: Fractional order analysis. Machado [35] highlights that the fractional order enhances the description of system dynamics. This sensitivity adjustment can capture more subtle variations and patterns within the vibration signals. To choose suitable α values, we calculate FrDivEn results of the automotive machine vibration signal at different imbalance ratios and various α values, analyze the results, and summarize the laws associated with FrDivEn.

Step 2: Gradient scale analysis. To establish the correspondence between the FrDivEn of vibration signals and the appropriate gradient scale to trade off sample synthesis, we must find the gradient scale that yields the highest fault diagnostic accuracy at different imbalance ratios. We use the CGDM with varying gradient scales to synthesize fault samples, filling the imbalanced sample set at different imbalance ratios. Subsequently, we analyze the effect of gradient scales on the fault diagnostic accuracy. The gradient scale with the highest fault diagnostic accuracy is then chosen to correspond to the FrDivEn of the imbalance ratio.

Step 3: FrDivEn–gradient scale curve fitting. We map the FrDivEn and the appropriate gradient scale at different imbalance ratios onto Cartesian coordinates, obtain several corresponding points, and fit a FrDivEn–gradient scale curve through these points. With this fitted curve, we can find correspondence from any FrDivEn to the appropriate gradient scale. In this study, FrDivEn–gradient scale fitting curves are obtained from the CWRU bearing dataset. Theoretically, we can apply this fitting curve, i.e., the correspondence between FrDivEn and the gradient scale, to other automotive machines similar to the rolling bearing.

The comparison of the proposed FrDivEn with the existing DivEn, the selection process of the gradient scale in the CGDM, and the fitting process are elucidated in detail in Section 4.2. Additionally, we incorporate DivEn and FrDivEn into the CGDM, respectively, and demonstrate their effectiveness by validating them against other generative models through fault diagnostic experiments. The related experiments and discussions are included in Section 4.3.

3.3. Sample Mix

Synthetic fault samples are produced through the CGDM. This method synthesizes fault samples that can be used to supplement real-world samples, enhancing the sample set for CNN training. The mixer function \mathcal{M} is used to combine these samples into a cohesive set.

The mixed sample set, X_{mix} , is determined using the mixer function \mathcal{M} , integrating the real samples X_0 and synthetic samples \tilde{X}_0 :

$$X_{mix} = \mathcal{M}(X_0, \tilde{X}_0) \quad (10)$$

where \mathcal{M} denotes the sample mixing process; X_0 is the imbalanced real sample set; and \tilde{X}_0 is the supplementary synthetic sample set, which is produced by CGDM's synthesis technique. This integration aims to create a balanced sample set for fault diagnosis.

3.4. Fault Diagnosis

After obtaining the mixed sample set, features from the image samples are extracted to achieve accurate fault diagnosis. The ConvNeXt structure is combined with transfer learning to enhance fault diagnostic accuracy. ConvNeXt's excellent performance in image classification tasks was validated using the ImageNet sample set [36]. As a deep CNN model, ConvNeXt requires extensive parameter tuning after initialization. This tuning can lead to unsatisfactory fault diagnostic accuracy when trained over limited epochs, as it may not fully converge or learn the necessary features. Transfer learning significantly reduces the need for extensive parameter tuning [37]. The pretrained ConvNeXt model can leverage hierarchical representations learned from the ImageNet sample set, enhancing its performance on the target task.

The ConvNeXt model was pretrained on the ImageNet sample set, while the target dataset comprises GAF images. When there is dissimilarity between ImageNet images and GAF images, more layers should be fine-tuned for effective fault diagnosis. Fine-tuning adjusts the model to better recognize and classify the specific features of GAF images, which are different from natural images of ImageNet. The fine-tuning process involves adjusting ConvNeXt blocks and other layers at the lower level. The original output classes of the last linear layer, corresponding to ImageNet, are replaced with classes representing possible automotive machine working states. This ensures that the model's predictions are relevant to the fault diagnostic task. The fine-tuned ConvNeXt is illustrated in Figure 2, which shows how the model's architecture is adapted.

To sum up, the proposed method, shown in Figure 3, comprises a preprocessing module, a sample synthesis module based on CGDM with FrDivEn trade-off, a sample mix module, and a fault diagnostic module based on fine-tuning pretrained ConvNeXt.

Pretrained ConvNeXt model

Fault diagnosis model

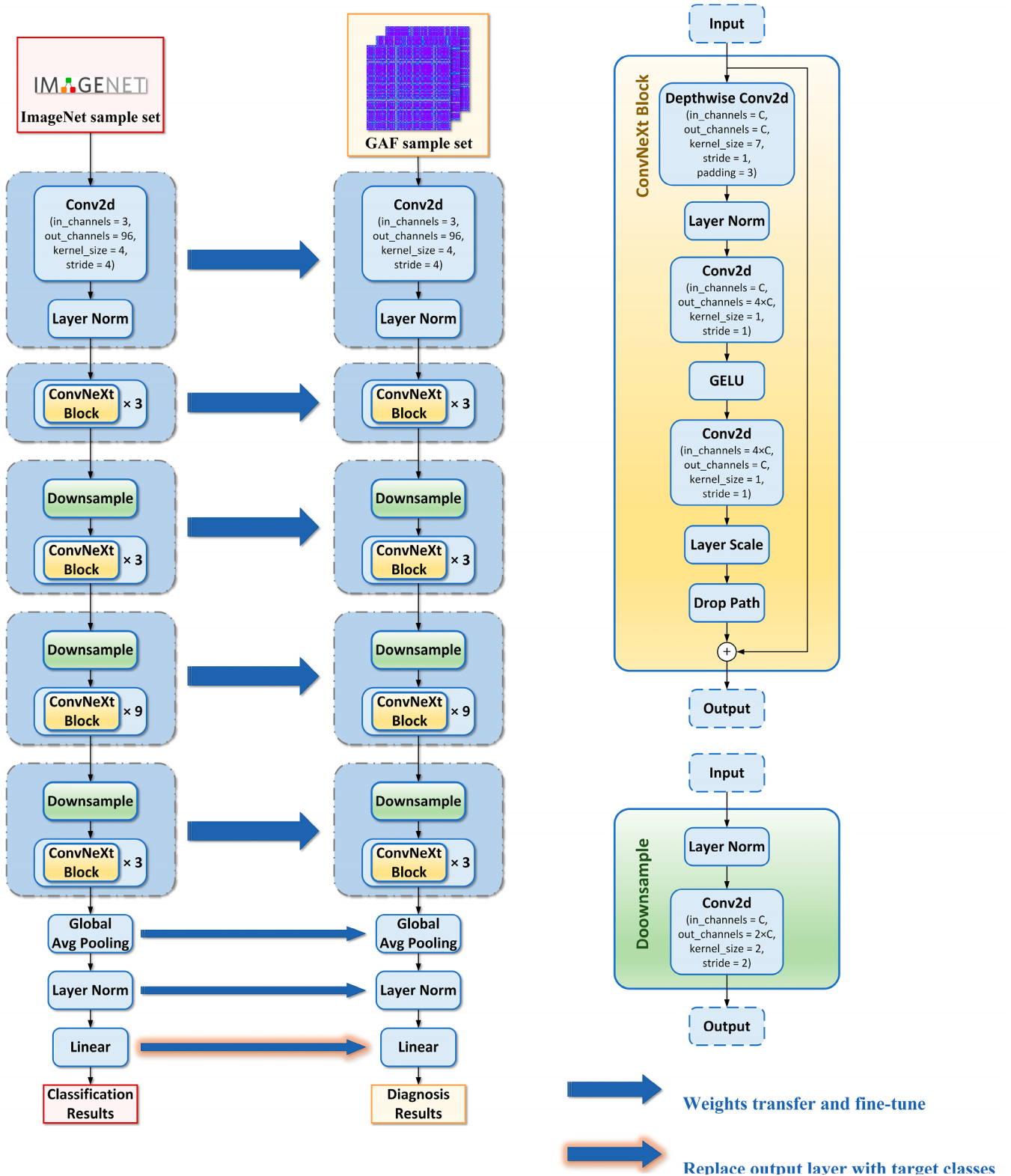


Figure 2. The fine-tuned ConvNeXt structure.

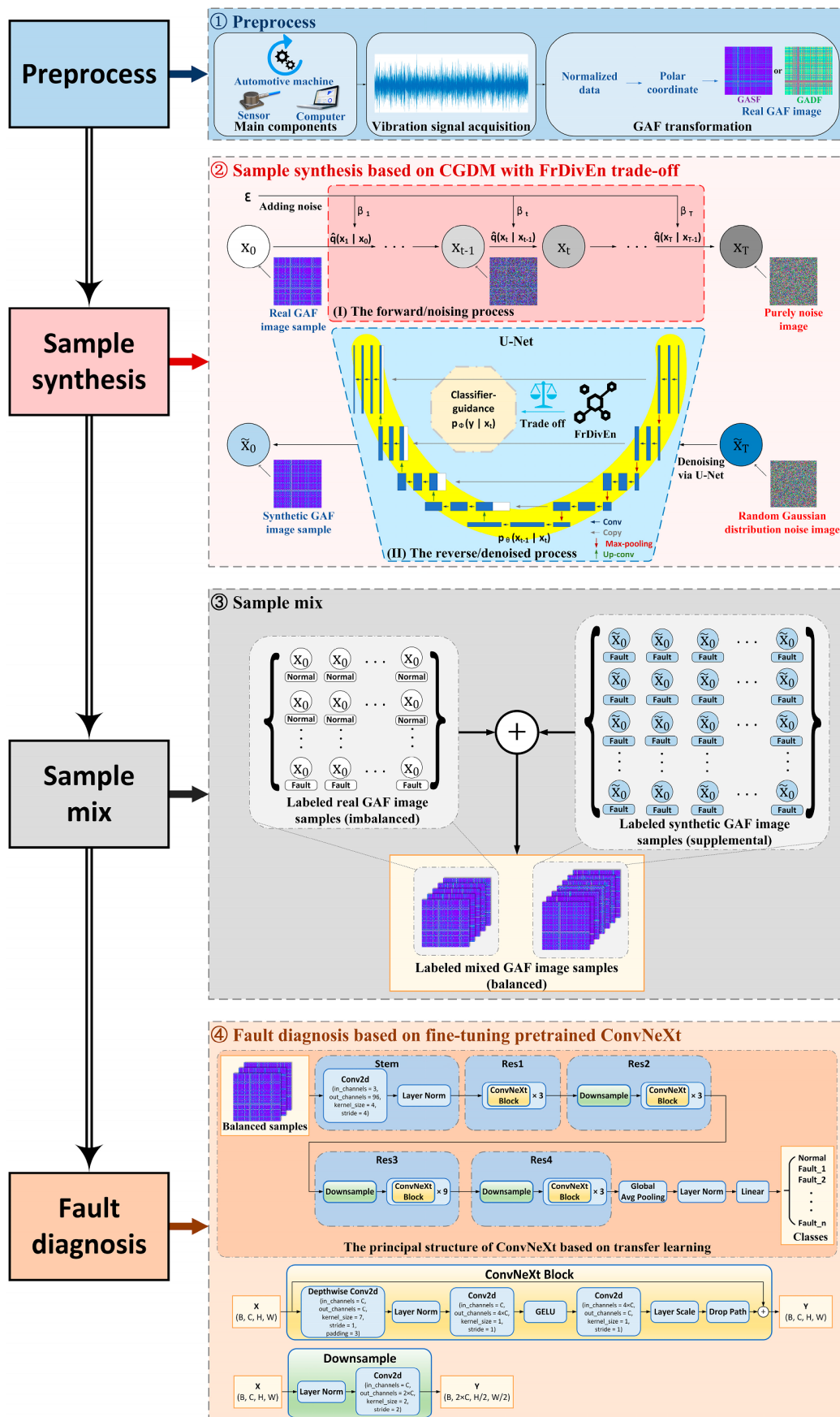


Figure 3. The proposed imbalanced fault diagnostic method for automotive machines.

4. Experiments and Discussion

To explore the correspondence between the proposed FrDivEn and the gradient scale, and to verify the effectiveness and generalization of the imbalanced fault diagnostic method, we conducted automotive machine fault diagnosis experiments.

4.1. Experimental Setup

In our experiments, the algorithms were implemented using PyTorch 2.2.1 and run on a platform equipped with an i9 12900 K CPU, 16 G × 2 of DDR5 RAM, and an NVIDIA GeForce RTX3090 GPU for training the proposed method.

Referring to the original paper [36,37] and considering the fault diagnosis effect as well as hardware constraints, the training settings were as follows. The AdamW optimizer was the network optimizer, and the cross-entropy loss function was selected as the loss function. The batch size was set to 16. The training process of the entire ConvNeXt model comprised 120 epochs, sequentially divided into two stages: 20 epochs for the training output layer and 100 epochs for training other layers: (1) Output layer training: The learning rate started at 0.01 and decayed to 0.001 after 10 epochs, continuing for another 10 epochs; (2) Other layer training: The learning rate was fixed to 0.0001 for fine-tuning, running for 100 epochs. To minimize the impact of stochasticity on the experiments, we carried out five identical fault diagnostic experiments for each method with varying settings or modules, such as imbalance ratios, sample synthesizers, and fault classifiers. The median of these five fault diagnostic accuracies was selected as the experimental result. The detailed training process setup is presented in Table 2.

Table 2. Training process setup.

Epoch	Training Layer	Learning Rate	Optimizer	Loss Function	Batch Size
1–10	Output layer	0.01	AdamW	Cross-Entropy	16
11–20		0.001			
21–120	Other layers	0.0001			

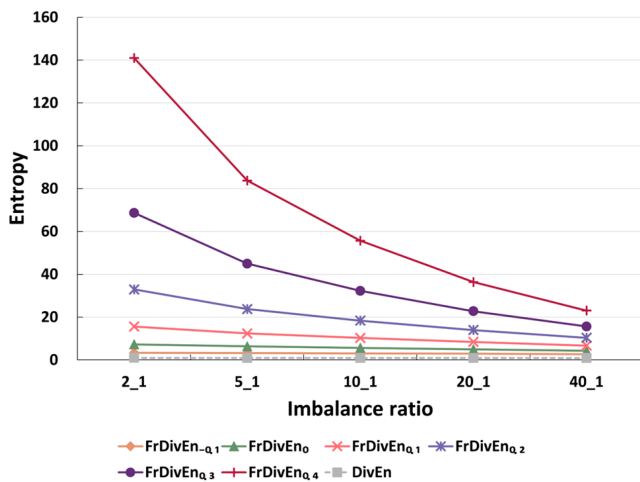
4.2. FrDivEn Trade-Off

To find a trade-off between synthetic samples' diversity and fidelity and achieve high-quality sample synthesis, we introduced FrDivEn, a sensitive measure of time series diversity. In exploring the correspondence between the proposed FrDivEn and the CGDM's appropriate gradient scale, we chose the CWRU bearing vibration data as the dataset. This was consistent with the calculation of DivEn in Section 2.2., ensuring a fair comparison. We proceeded sequentially with the three steps described in Section 3.2.

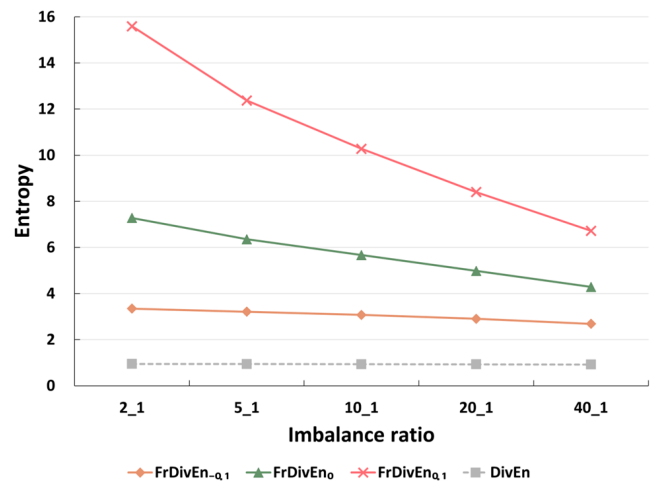
Step 1: Fractional order analysis. We calculated the FrDivEn results of the automotive machine vibration time series at different imbalance ratios and α values. The previously calculated DivEn and FrDivEn results at different fractional orders α are illustrated in Table 3 and Figure 4. From the calculation results, the following can be surmised: (1) When other conditions are constant, with the gradual increase in the imbalance ratio, FrDivEn gradually decreases. Taking FrDivEn_{0.1} as an instance, this decreases from 15.5899 at an imbalance ratio of 2:1 to 6.7208 at an imbalance ratio of 40:1, indicating that the diversity of the time series diminishes as the imbalance problem worsens. (2) When other conditions are constant, with the gradual increase in the fractional order α , FrDivEn drastically increases, with FrDivEn_{0.4} at an imbalance ratio of 2:1 even exceeding 140. (3) The difference between the FrDivEn results computed from two adjacent imbalance ratios is larger than that of DivEn. For example, the FrDivEn₀ difference between vibration signals at imbalance ratios of 2:1 and 5:1 is 0.9158, which is substantially larger than the DivEn difference of 0.0061. This indicates that the sensitivity of the proposed FrDivEn to the vibration signal is enhanced compared to DivEn.

Table 3. DivEn and FrDivEn at different fractional orders α .

		Imbalance Ratio				
		2:1	5:1	10:1	20:1	40:1
Entropy	FrDivEn _{-0.1}	3.3467	3.2108	3.0741	2.9016	2.6875
	FrDivEn ₀	7.2782	6.3517	5.6654	4.9789	4.2881
	FrDivEn _{0.1}	15.5899	12.3727	10.2736	8.4003	6.7208
	FrDivEn _{0.2}	32.9366	23.7600	18.3490	13.9453	10.3497
	FrDivEn _{0.3}	68.6517	44.9805	32.2624	22.7574	15.6337
	FrDivEn _{0.4}	140.9853	83.7834	55.6922	36.3769	23.0469
	DivEn	0.9496	0.9413	0.9357	0.9287	0.9195



(a)



(b)

Figure 4. DivEn and FrDivEn at different fractional orders α . (a) DivEn and FrDivEn when $\alpha = -0.1, 0, 0.1, 0.2, 0.3,$ and 0.4 ; (b) DivEn and FrDivEn when $\alpha = -0.1, 0,$ and 0.1 .

Step 2: Gradient scale analysis. After completing the analysis of FrDivEn, we employed the CGDM with differing gradient scales to select appropriate values. Specifically, for each imbalance ratio, finding the appropriate gradient scale was divided into coarse and fine sampling: (1) Coarse sampling of the gradient scale: we swept over the gradient scale values $[0.5, 1, 1.5, 2, 2.5, 3, 3.5]$, consistent with Dhariwal et al.’s [18] method when performing sample synthesis via the CGDM on ImageNet 256×256 (which is the same size as our selected sample size). (2) Fine sampling of the gradient scale: we denoted the gradient scale value that achieved the highest diagnostic accuracy in coarse sampling as s_c , and we swept over the interval $(s_c - 0.5, s_c + 0.5)$ at intervals of 0.1, taking the gradient scale that achieved the highest diagnostic accuracy here as the appropriate gradient scale. The imbalanced sample set was allocated one normal state and n fault states to classify, as shown in Table 4. The appropriate gradient scales selected at different imbalance ratios are shown in Table 5.

Step 3: FrDivEn–gradient scale curve fitting. To find the appropriate gradient scale value from any FrDivEn, we mapped the FrDivEn results calculated at different imbalance ratios in Step 1 to the appropriate gradient scale found in Step 2, plotting them as FrDivEn–gradient scale points in the coordinate plot. These points were then fitted to obtain the FrDivEn–gradient scale curve. Considering that Dhariwal et al. [18] did not set the gradient scale s of the CGDM smaller than 0, we used an exponential fit for the FrDivEn $_{\alpha}$ –gradient scale at different fractional orders α . For comparison, we applied the same method to fit the DivEn–gradient scale curve. The fitted entropy–gradient scale curves are illustrated in Figure 5.

Table 4. Imbalanced sample set allocation of automotive machine.

Class	Normal	Fault_1	Fault_2	...	Fault_n	Imbalance Ratio
No.	0	1	2	...	n	
Number of training samples	360	180	180	...	180	2:1
	360	72	72	...	72	5:1
	360	36	36	...	36	10:1
	360	18	18	...	18	20:1
	360	9	9	...	9	40:1
Number of test samples	90	90	90	...	90	1:1

Table 5. The appropriate gradient scales.

Imbalance Ratio	2:1	5:1	10:1	20:1	40:1
Gradient scale <i>s</i>	2.9	2.1	1.7	1.4	1.2

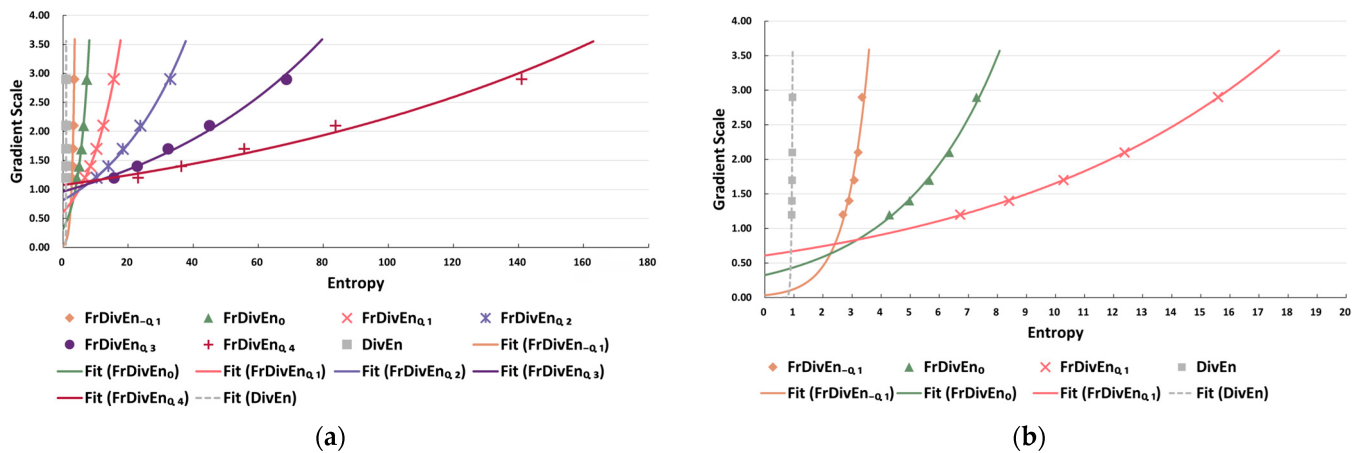


Figure 5. The fitted entropy–gradient scale curves. (a) DivEn– and FrDivEn–gradient scale when $\alpha = -0.1, 0, 0.1, 0.2, 0.3,$ and 0.4 ; (b) DivEn– and FrDivEn–gradient scale when $\alpha = -0.1, 0,$ and 0.1 .

From the entropy–gradient scale curves, the following can be surmised: (1) As the fractional order α gradually increases, the gradient scale at the initial point of the FrDivEn $_{\alpha}$ –gradient scale curves also increases. For instance, the gradient scale for FrDivEn $_{0.1}$ is consistently larger than the initial value of 0.61, while for FrDivEn $_{0.4}$, it is always larger than the initial value of 1.08. (2) As the fractional order α continues to increase, the FrDivEn $_{\alpha}$ –gradient scale curves tend to flatten. For FrDivEn $_{0.1}$, when it rises from 0 to 10, the corresponding gradient scale rises from 0.61 to 1.66, a change of 1.05, whereas for FrDivEn $_{0.4}$, the gradient scale increases by merely 0.08 for the same range. (3) When the entropy shifts from 0.7 to 0.9, the DivEn–gradient scale curve transitions rapidly from near-horizontal to near-vertical. In contrast, the FrDivEn $_0$ –gradient scale curve, for example, has a more stable slope of about 0.12. This indicates the instability in the gradient scale values derived from DivEn compared to FrDivEn.

Regarding the sample synthesizer, Dhariwal et al. [18] set the initial gradient scale value to 0.5 and incrementally increased it for sample synthesis. Combined with Figure 5, this led us to conclude that some of the minimum gradient scale values were too large to be desirable. For instance, the gradient scale corresponding to FrDivEn $_{0.4}$ could not be taken to a value below 1. Considering the gradient scale range and the smoothness of the FrDivEn–gradient scale curves, we selected FrDivEn results corresponding to the two curves with initial gradient scale values around 0.5, i.e., FrDivEn $_0$ and FrDivEn $_{0.1}$, as the basis for the gradient scale value in sample synthesis.

4.3. Applications of the Proposed Method

To test the validity and generalizability of the proposed imbalanced diagnostic method, we applied this method to (1) the CWRU bearing dataset with a motor load of 3 HP and a speed of 1730 rpm; (2) the University of Connecticut (UConn) gearbox dataset [38]; and (3) the Wuhan University (WHU) rotor dataset [39]. This allowed us to explore the diagnostic effect under a different load and speed of the same machine, as well as across different machines. After the GADF transformation, the imbalanced sample set of each automotive machine was allocated consistently with the process in Section 4.2, as shown in Table 4.

In addition to the CGDMs using FrDivEn₀ and FrDivEn_{0,1}, we included the Wasserstein generative adversarial network (WGAN) [40], the CGDM with a gradient scale consistent with the default value of 1 in the source code, and the CGDM using DivEn as the basis for the gradient scale value for comparison. Regarding the fault classifier, in addition to utilizing the pretrained ConvNeXt model based on transfer learning, we also incorporated the pretrained VGG model [41], GoogLeNet model [42], ResNet model [7], and DenseNet model [43] for comparison. The application of the proposed diagnostic method to the three different automotive machine datasets is demonstrated below.

4.3.1. Bearing (Motor Load: 3 HP; Speed: 1730 rpm)

We calculated the DivEn, FrDivEn₀, and FrDivEn_{0,1} of the time series at each imbalance ratio. The gradient scale s corresponding to each entropy was obtained by referencing the entropy–gradient scale curves. The entropy–gradient scale s and the diagnostic accuracy with ConvNeXt as the fault classifier are presented in Table 6. The imbalanced diagnostic accuracy with various sample synthesizers and fault classifiers of the bearing dataset is shown in Table 7.

The fault diagnostic results obtained using different sample synthesizers and fault classifiers provided a few key insights: (1) When other conditions are constant, the CGDM with a gradient scale corresponding to FrDivEn₀ achieves the highest fault diagnostic accuracy across all five imbalance ratios. For instance, with an imbalance ratio of 40:1 and the fine-tuned pretrained ConvNeXt as the fault classifier, the CGDM using FrDivEn₀ synthesizes samples achieving a fault diagnostic accuracy of 91.22%, which is 7.32% higher than the accuracy of samples synthesized using the WGAN. The samples synthesized from the CGDM with a gradient scale corresponding to FrDivEn₀ for each fault state are provided in Figure 6. (2) The diagnostic accuracy of samples synthesized using the CGDM with a gradient scale corresponding to DivEn is lower than those of FrDivEn₀ and FrDivEn_{0,1}. With an imbalance ratio of 40:1 and ConvNeXt as the fault classifier, the fault diagnostic accuracy of samples synthesized from the CGDM using DivEn is 87.67%, which is 3.89% lower than the accuracy of samples synthesized using FrDivEn₀, and even lower than the 90.22% accuracy of the default CGDM where the gradient scale is fixed to 1. (3) ConvNeXt consistently maintains a high accuracy advantage over other fine-tuned pretrained models under the same conditions. With an imbalance ratio of 40:1 and the sample synthesizer as the CGDM using FrDivEn₀, the fault diagnostic accuracy using ConvNeXt is 91.22%, which is 10.20% higher than the accuracy when using VGG and 3.40% higher than that with ResNet.

Table 6. The entropy–gradient scale s and the diagnostic accuracy of bearing data.

Entropy– s	Imbalance Ratio					
	2:1	5:1	10:1	20:1	40:1	
FrDivEn ₀ , s	7.3113, 2.84	6.3955, 2.17	5.7001, 1.76	5.0058, 1.44	4.3094, 1.17	
Accuracy	99.44%	98.56%	97.67%	95.56%	91.22%	
FrDivEn _{0,1} , s	15.6910, 2.93	12.4886, 2.12	10.3552, 1.72	8.4554, 1.42	6.7594, 1.20	
Accuracy	99.22%	98.33%	97.00%	95.11%	90.89%	
DivEn, s	0.9539, 3.39	0.9478, 2.83	0.9415, 2.35	0.9337, 1.87	0.9241, 1.41	
Accuracy	99.11%	98.11%	96.11%	94.44%	87.67%	

Table 7. Diagnostic accuracy with imbalanced bearing sample set (motor load: 3 HP; speed: 1730 rpm).

Imbalance Ratio	Synthesizer Classifier	N/A ¹	WGAN	CGDM			
				Default ²	DivEn ³	FrDivEn ₀ ⁴	FrDivEn _{0,1} ⁵
2:1	VGG	93.44%	96.44%	96.44%	96.56%	97.00%	96.67%
	GoogLeNet	95.44%	96.67%	97.11%	97.44%	97.67%	97.56%
	ResNet	96.00%	97.00%	97.44%	97.89%	98.00%	98.00%
	DenseNet	96.56%	97.44%	97.67%	98.11%	98.22%	98.11%
	ConvNeXt	97.89%	98.56%	98.67%	99.11%	99.44%	99.22%
5:1	VGG	90.56%	92.00%	93.56%	94.78%	96.78%	95.22%
	GoogLeNet	91.67%	94.33%	95.33%	96.33%	97.22%	97.11%
	ResNet	92.22%	94.78%	95.56%	96.67%	97.78%	97.22%
	DenseNet	92.89%	95.00%	96.44%	97.67%	97.78%	97.78%
	ConvNeXt	94.67%	97.33%	97.56%	98.11%	98.56%	98.33%
10:1	VGG	86.67%	89.89%	91.44%	91.89%	93.67%	93.00%
	GoogLeNet	88.89%	91.67%	92.78%	93.89%	94.78%	94.22%
	ResNet	89.22%	92.44%	93.78%	94.56%	95.00%	94.67%
	DenseNet	89.56%	92.56%	94.44%	94.56%	95.33%	95.11%
	ConvNeXt	92.11%	94.00%	95.78%	96.11%	97.67%	97.00%
20:1	VGG	81.33%	85.44%	88.33%	90.11%	91.11%	90.78%
	GoogLeNet	82.78%	88.56%	91.33%	92.00%	92.44%	92.33%
	ResNet	86.00%	89.00%	91.44%	92.44%	93.33%	92.89%
	DenseNet	86.56%	89.78%	91.67%	92.44%	93.56%	93.56%
	ConvNeXt	87.56%	92.11%	93.33%	94.44%	95.56%	95.11%
40:1	VGG	77.33%	79.33%	81.44%	80.67%	82.78%	82.44%
	GoogLeNet	77.67%	80.33%	84.00%	83.67%	86.78%	85.11%
	ResNet	81.11%	83.67%	87.00%	86.22%	88.22%	87.78%
	DenseNet	82.78%	84.56%	88.00%	86.56%	89.89%	88.44%
	ConvNeXt	83.44%	85.00%	90.22%	87.67%	91.22%	90.89%

¹ Without synthesizer; ² default CGDM with gradient scale fixed to 1; ³ CGDM with gradient scale corresponding to DivEn; ⁴ CGDM with gradient scale corresponding to FrDivEn₀; ⁵ CGDM with gradient scale corresponding to FrDivEn_{0,1}.

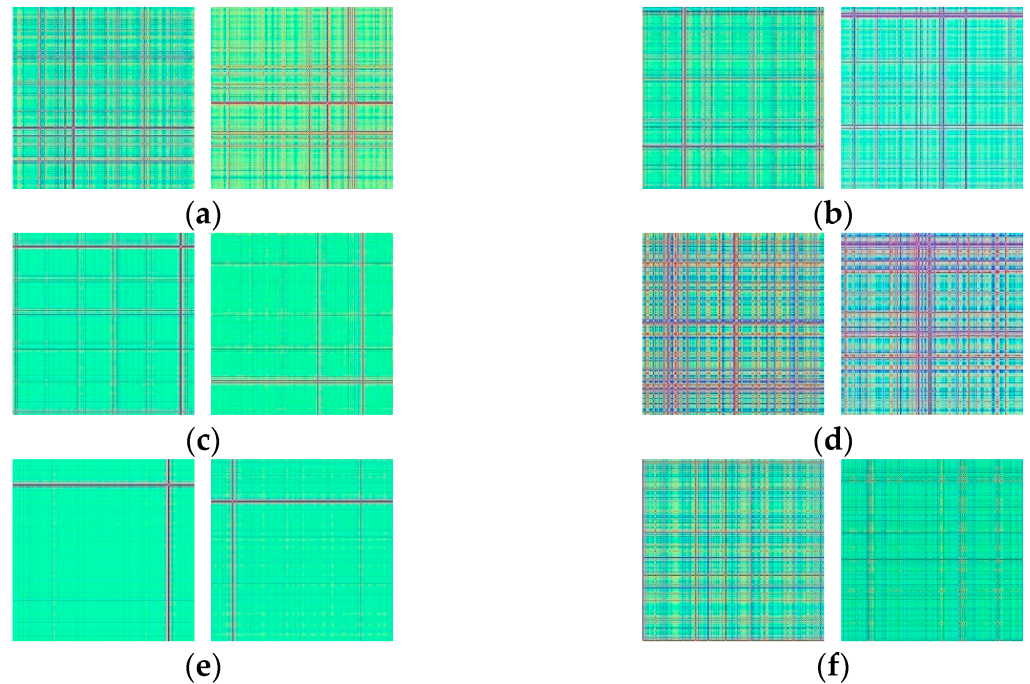


Figure 6. Cont.

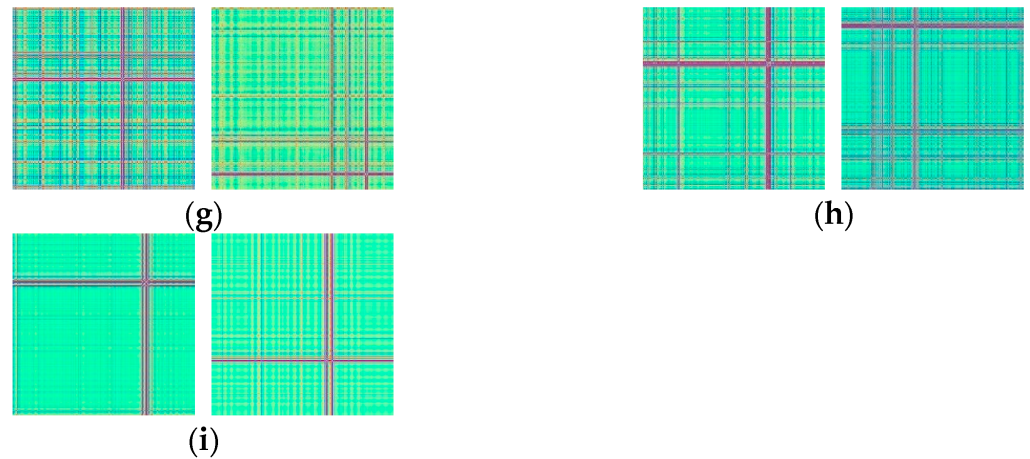


Figure 6. Real (left) and synthetic (right) samples synthesized from CGDM with a gradient scale corresponding to $FrDivEn_0$. (a) 7_BA; (b) 7_IR; (c) 7_OR; (d) 14_BA; (e) 14_IR; (f) 14_OR; (g) 21_BA; (h) 21_IR; (i) 21_OR.

T-distributed stochastic neighbor embedding (t-SNE) [44] was applied for the visualization and interpretation of latent features captured by the ConvNeXt models. T-SNE helped in visualizing high-dimensional data. The analysis compared the fine-tuned ConvNeXt model’s performance without a sample synthesizer and with different sample synthesizers at a moderate imbalance ratio of 10:1. Figure 7 shows the visualization of features. Compared to other sample synthesizers, the CGDM with a gradient scale corresponding to $FrDivEn_0$ results in better clustering of scatter points for each state. Taking the “7_OR” and “21_OR” states as examples, the CGDM using $FrDivEn_0$ enables ConvNeXt to cluster them separately without overlapping, unlike other sample synthesizers.

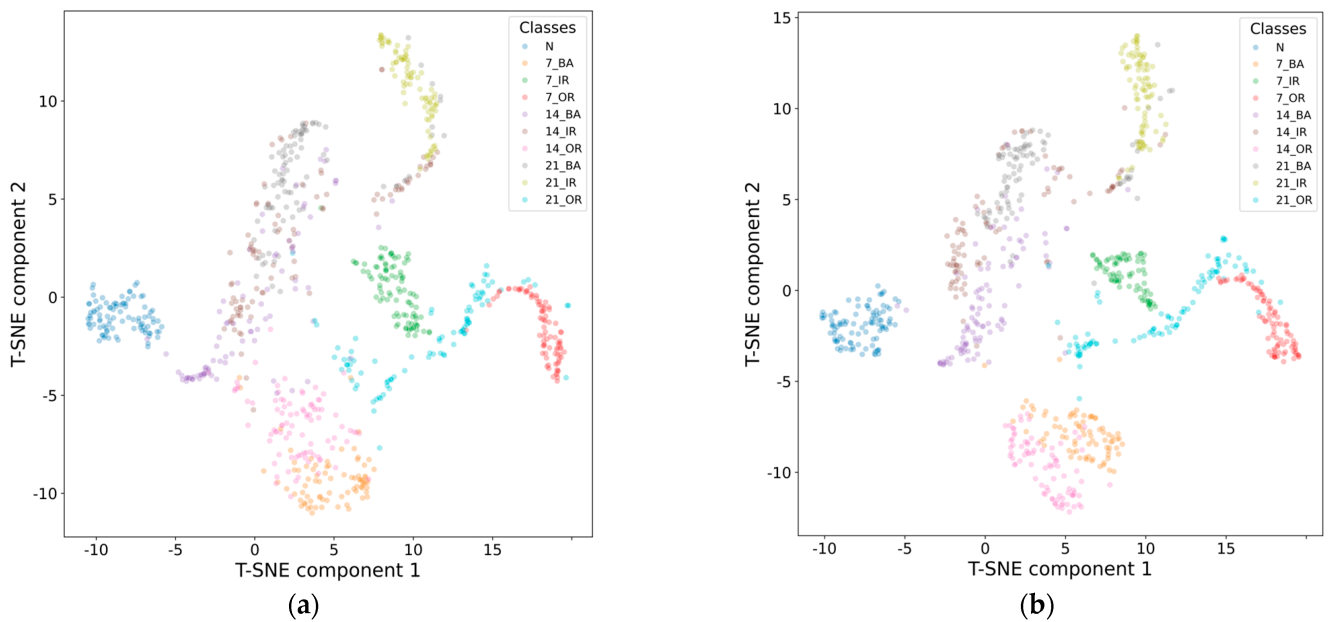


Figure 7. Cont.

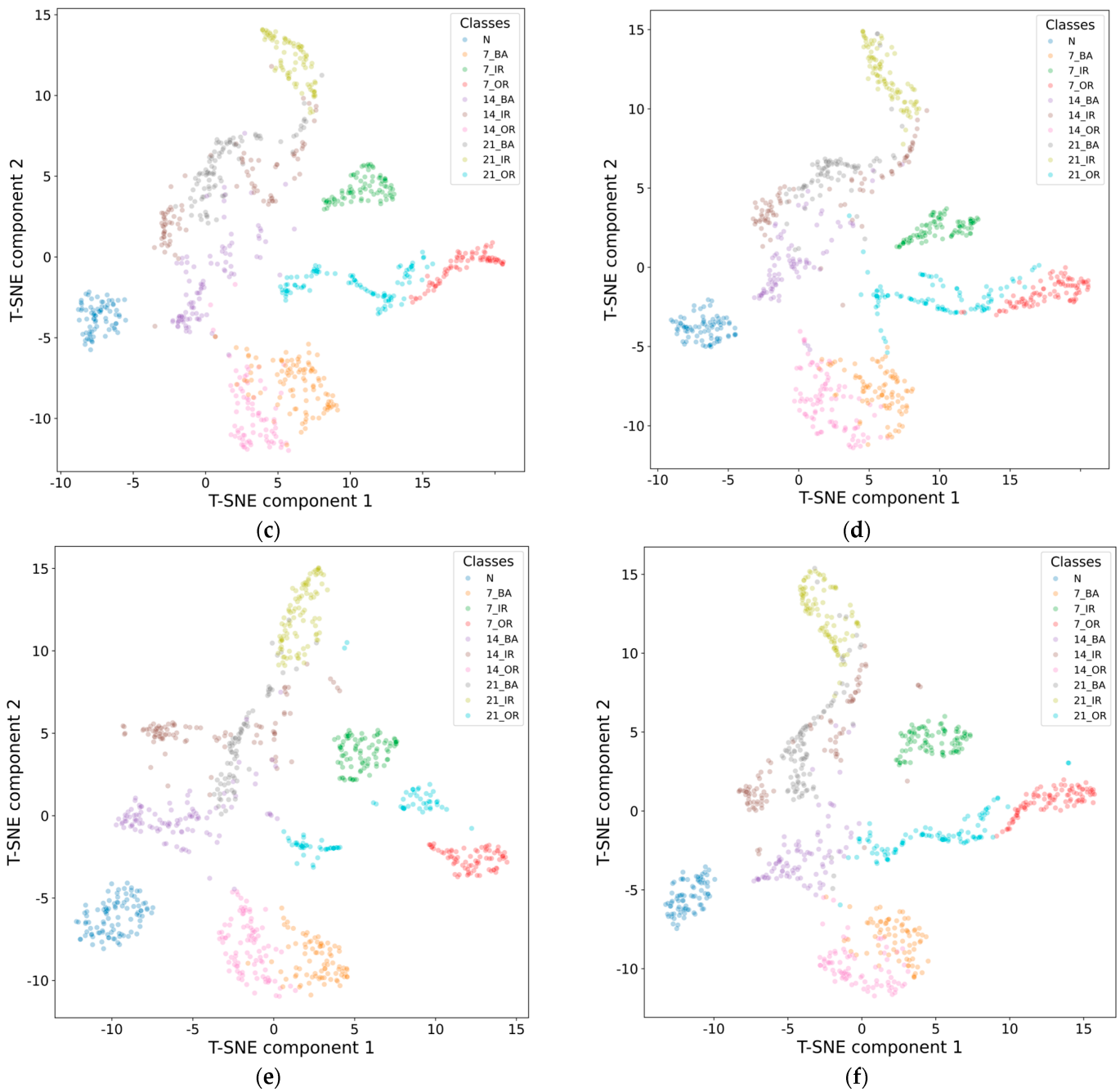


Figure 7. Feature visualization through t-SNE at bearing sample imbalance ratio of 10:1. (a) Without synthesizer; (b) WGAN; (c) default CGDM with gradient scale fixed to 1; (d) CGDM with gradient scale corresponding to DivEn; (e) CGDM with gradient scale corresponding to FrDivEn₀; (f) CGDM with gradient scale corresponding to FrDivEn_{0,1}.

A confusion matrix was employed to categorize the captured features into various labels, as shown in Figure 8. The diagonal elements from top left to bottom right indicate the number of correctly categorized samples in each class. The off-diagonal elements indicate the number of incorrectly categorized samples as other classes. In the test set, each bearing state contains 90 samples. The closer the value on the diagonal of a state class is to 90, the more accurately that state class is diagnosed. The CGDM with a gradient scale corresponding to FrDivEn₀ yields the greatest predictive accuracy.

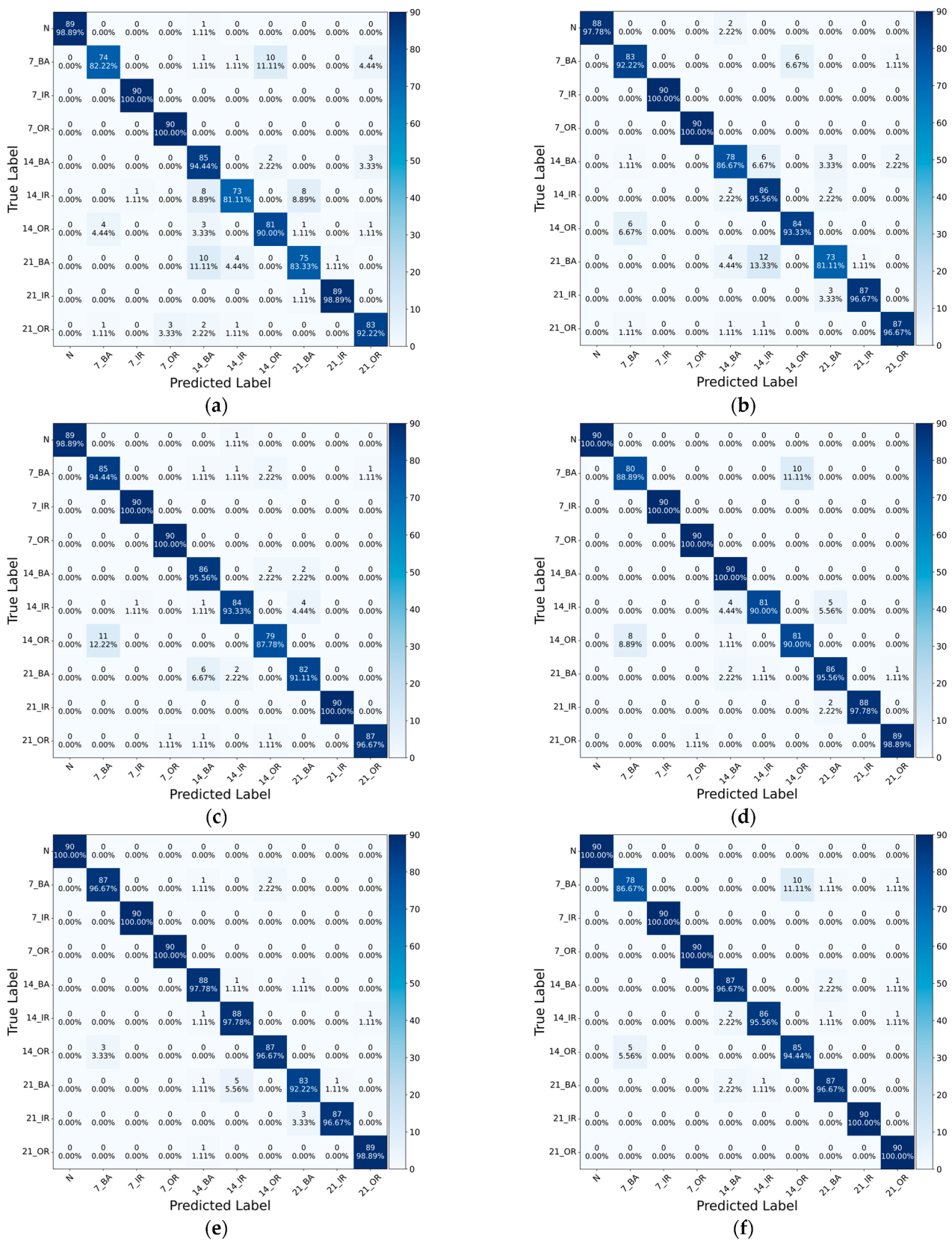


Figure 8. Confusion matrices at bearing sample imbalance ratio of 10:1. (a) Without synthesizer; (b) WGAN; (c) default CGDM with gradient scale fixed to 1; (d) CGDM with gradient scale corresponding to DivEn; (e) CGDM with gradient scale corresponding to FrDivEn₀; (f) CGDM with gradient scale corresponding to FrDivEn_{0.1}.

The receiver operating characteristic (ROC) curves are shown in Figure 9. In the ROC curves, “CGDM-Default” means “default CGDM with gradient scale fixed to 1”, “CGDM-DivEn” means “CGDM with gradient scale corresponding to DivEn”, “CGDM-FrDivEn₀” means “CGDM with gradient scale corresponding to FrDivEn₀”, “CGDM-FrDivEn_{0,1}” means “CGDM with gradient scale corresponding to FrDivEn_{0,1}”. Compared to other models, the ROC curve of the CGDM with a gradient scale corresponding to FrDivEn₀ is closest to the upper left corner, with an area under the curve (AUC) of 0.9962, indicating its excellent performance. The evaluation metrics are shown in Table 8. The CGDM with a gradient scale corresponding to FrDivEn₀ achieves the highest precision of 0.9768, recall of 0.9767, and F1-score of 0.9767, demonstrating its excellent performance. The errors made by the proposed method and other methods are shown in Table 9. The CGDM with a gradient scale corresponding to FrDivEn₀ achieves the lowest mean absolute error (MAE) of 0.0656 and root mean squared error (RMSE) of 0.4933. After combining the confusion matrix and comparing other methods, we believe that the misdiagnosis of some “21_BA” samples as “14_IR” is the main reason for the limitations of the proposed method. In addition, other failures such as pitting and multi-failure fusion, which are not considered in this bearing dataset, may also contribute to the limitations of the proposed method. The validity of the diagnostic method was tested by utilizing a dataset different from the previous bearing working conditions.

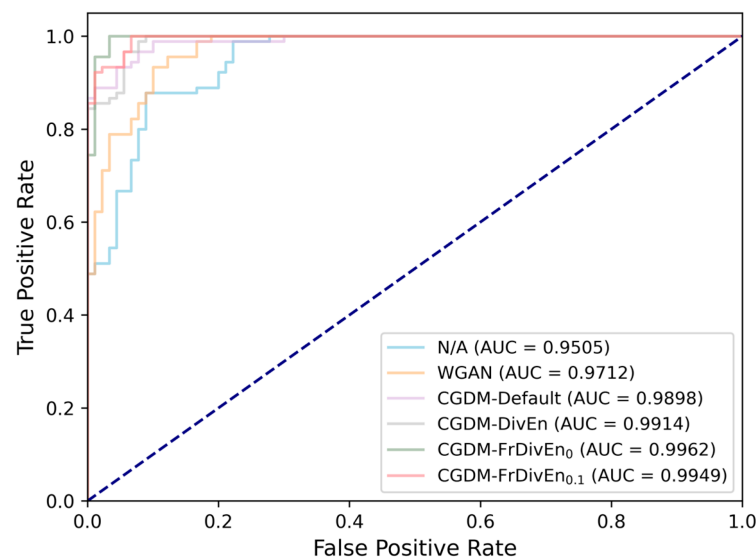


Figure 9. ROC curves at bearing sample imbalance ratio of 10:1.

Table 8. Evaluation metrics at bearing sample imbalance ratio of 10:1.

Evaluation Metric	N/A ¹	WGAN	CGDM			
			Default ²	DivEn ³	FrDivEn ₀ ⁴	FrDivEn _{0,1} ⁵
Macro-Precision	0.9245	0.9418	0.9584	0.9617	0.9768	0.9703
Macro-Recall	0.9211	0.9400	0.9578	0.9611	0.9767	0.9700
Macro-F1	0.9211	0.9401	0.9578	0.9610	0.9767	0.9699

¹ Without synthesizer; ² default CGDM with gradient scale fixed to 1; ³ CGDM with gradient scale corresponding to DivEn; ⁴ CGDM with gradient scale corresponding to FrDivEn₀; ⁵ CGDM with gradient scale corresponding to FrDivEn_{0,1}.

Table 9. Errors at bearing sample imbalance ratio of 10:1.

Evaluation Metric	N/A	WGAN	CGDM			
			Default	DivEn	FrDivEn ₀	FrDivEn _{0.1}
MAE	0.2767	0.1856	0.1589	0.1378	0.0656	0.1289
RMSE	1.1289	0.8762	0.8320	0.7760	0.4933	0.7902

4.3.2. Gearbox

The University of Connecticut (UConn) gearbox dataset includes nine different gear states: normal state (N), missing tooth (Miss), root crack (Crack), spalling (Spall), and five levels of chipping tip severity (Chip1, Chip2, Chip3, Chip4, and Chip5) [38]. The accelerometers, positioned on the input end of the gearbox housing, capture the gear vibration signals.

The entropy–gradient scale s and the diagnostic accuracy with ConvNeXt as the fault classifier are presented in Table 10. At an imbalance ratio of 40:1, given that the gradient scales corresponding to FrDivEn₀ and FrDivEn_{0.1} are both 1.39 to two decimal places, we take the CGDM with a gradient scale of 1.39 as the sample synthesizer’s fault diagnostic result for both FrDivEn₀ and FrDivEn_{0.1}. The diagnostic accuracy with different sample synthesizers and fault classifiers of the gearbox dataset is shown in Table 11.

The diagnostic results obtained using different sample synthesizers and fault classifiers provide a few key insights: (1) The similar fault diagnostic accuracies achieved by the CGDMs using FrDivEn₀ and FrDivEn_{0.1} are due to the minimal differences in the gradient scales. When the imbalance ratio is 40:1, the gradient scales for FrDivEn₀ and FrDivEn_{0.1} are identical, even to two decimal places, at 1.39. At this point, when using the fine-tuned pretrained ConvNeXt as the fault classifier and the CGDM with a gradient scale corresponding to FrDivEn as the sample synthesizer, the fault diagnostic accuracy reaches 87.90%, which is 11.59% higher than that of WGAN and 9.53% higher than the default CGDM. (2) At several imbalance ratios, the CGDM with gradient scales corresponding to FrDivEn₀ and FrDivEn_{0.1} achieves high fault diagnostic accuracy. For instance, with an imbalance ratio of 20:1 and ConvNeXt as the fault classifier, the CGDM using FrDivEn₀ achieves a fault diagnostic accuracy of 98.89%. This is 7.37% higher than the WGAN, 4.71% higher than the default CGDM, and 1.66% higher than the CGDM using DivEn. When comparing FrDivEn₀ and FrDivEn_{0.1}, the CGDM with a gradient scale corresponding to FrDivEn₀ achieves a slightly higher diagnostic accuracy. (3) As the imbalance problem worsens, the advantages of ConvNeXt over other CNN models become evident. With an imbalance ratio of 40:1 and the CGDM with a gradient scale corresponding to FrDivEn as the sample synthesizer, ConvNeXt achieves a diagnostic accuracy of 87.90%, which is 19.06% higher than VGG and 7.72% higher than ResNet.

Figure 10 illustrates a t-SNE visualization of the features. Compared to other sample synthesizers, the scatter points for each state are better clustered when the CGDM with a gradient scale corresponding to FrDivEn₀ is used as a synthesizer. For example, the CGDM using FrDivEn₀ enables ConvNeXt to cluster the scatter points for the “Chip1” state without overlap, unlike the other sample synthesizers.

Table 10. The entropy–gradient scale s and the diagnostic accuracy of gearbox data.

Entropy– s	Imbalance Ratio	2:1	5:1	10:1	20:1	40:1
	FrDivEn ₀ , s		7.9003, 3.39	6.9777, 2.58	6.2801, 2.09	5.5864, 1.71
Accuracy		100%	99.88%	99.63%	98.89%	87.90%
FrDivEn _{0.1} , s		18.1116, 3.73	14.5613, 2.61	12.2008, 2.06	10.1035, 1.67	8.2299, 1.39
Accuracy		100%	99.88%	99.51%	98.52%	87.90%
DivEn, s		0.9758, 6.47	0.9644, 4.63	0.9539, 3.40	0.9418, 2.37	0.9267, 1.52
Accuracy		100%	99.75%	99.26%	97.28%	86.05%

Table 11. Diagnostic accuracy with imbalanced gearbox sample set.

Imbalance Ratio	Synthesizer Classifier	N/A ¹	WGAN	CGDM			
				Default ²	DivEn ³	FrDivEn ₀ ⁴	FrDivEn _{0,1} ⁵
2:1	VGG	98.02%	98.15%	99.01%	99.26%	99.38%	99.38%
	GoogLeNet	99.63%	99.88%	99.88%	99.88%	100%	99.88%
	ResNet	99.75%	99.88%	99.88%	99.88%	100%	100%
	DenseNet	99.88%	99.88%	99.88%	100%	100%	100%
	ConvNeXt	99.88%	99.88%	100%	100%	100%	100%
5:1	VGG	92.96%	95.93%	98.52%	98.89%	99.51%	99.01%
	GoogLeNet	96.67%	98.27%	99.01%	99.51%	99.63%	99.51%
	ResNet	95.56%	98.77%	99.26%	99.75%	99.88%	99.88%
	DenseNet	97.65%	99.14%	99.51%	99.75%	99.88%	99.88%
	ConvNeXt	98.27%	99.14%	99.63%	99.75%	99.88%	99.88%
10:1	VGG	82.35%	88.15%	92.72%	94.94%	96.30%	96.05%
	GoogLeNet	88.27%	94.69%	96.54%	97.28%	98.52%	98.27%
	ResNet	89.26%	95.06%	97.78%	98.15%	99.14%	98.89%
	DenseNet	91.98%	96.79%	98.15%	98.89%	99.14%	99.14%
	ConvNeXt	94.32%	97.78%	99.01%	99.26%	99.63%	99.51%
20:1	VGG	68.40%	78.77%	85.68%	88.89%	92.22%	91.98%
	GoogLeNet	77.16%	85.68%	89.26%	92.22%	95.31%	95.06%
	ResNet	79.01%	87.53%	89.75%	93.33%	95.68%	95.56%
	DenseNet	82.59%	91.36%	93.58%	95.19%	98.02%	97.78%
	ConvNeXt	84.44%	92.10%	94.44%	97.28%	98.89%	98.52%
40:1	VGG	55.19%	60.62%	65.56%	68.89%	73.83%	
	GoogLeNet	59.38%	73.09%	75.06%	78.52%	81.23%	
	ResNet	63.21%	73.95%	77.65%	78.64%	81.60%	
	DenseNet	64.20%	78.27%	79.38%	84.81%	86.42%	
	ConvNeXt	68.15%	78.77%	80.25%	86.05%	87.90%	

¹ Without synthesizer; ² default CGDM with gradient scale fixed to 1; ³ CGDM with gradient scale corresponding to DivEn; ⁴ CGDM with gradient scale corresponding to FrDivEn₀; ⁵ CGDM with gradient scale corresponding to FrDivEn_{0,1}.

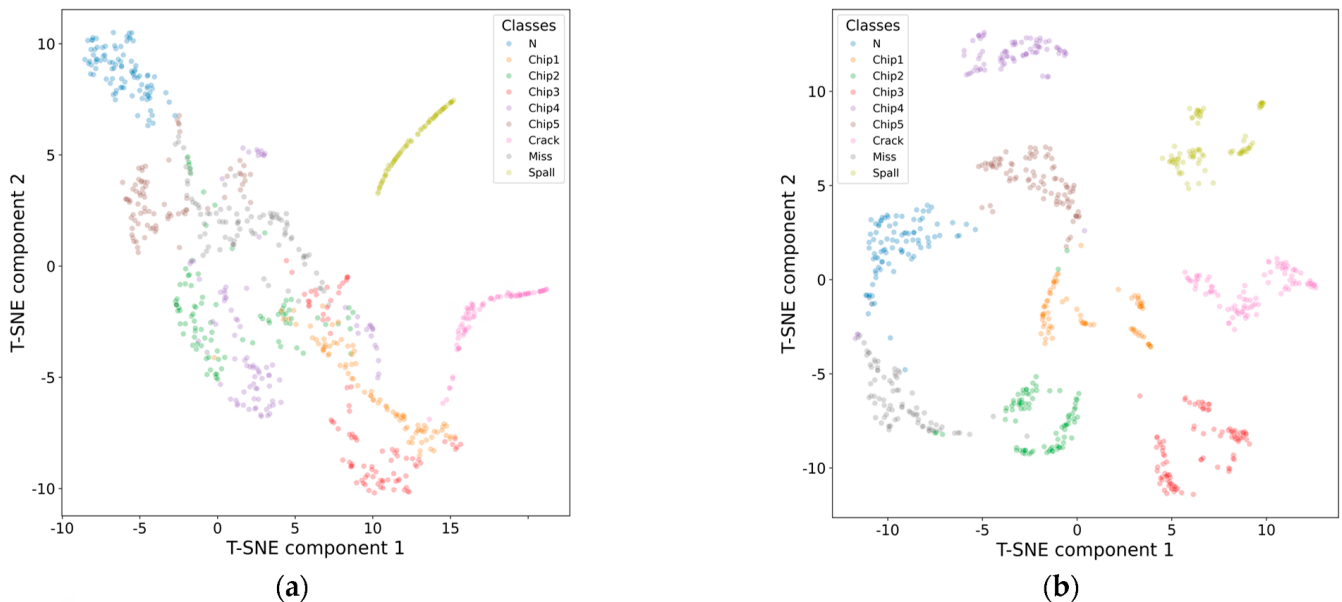


Figure 10. Cont.

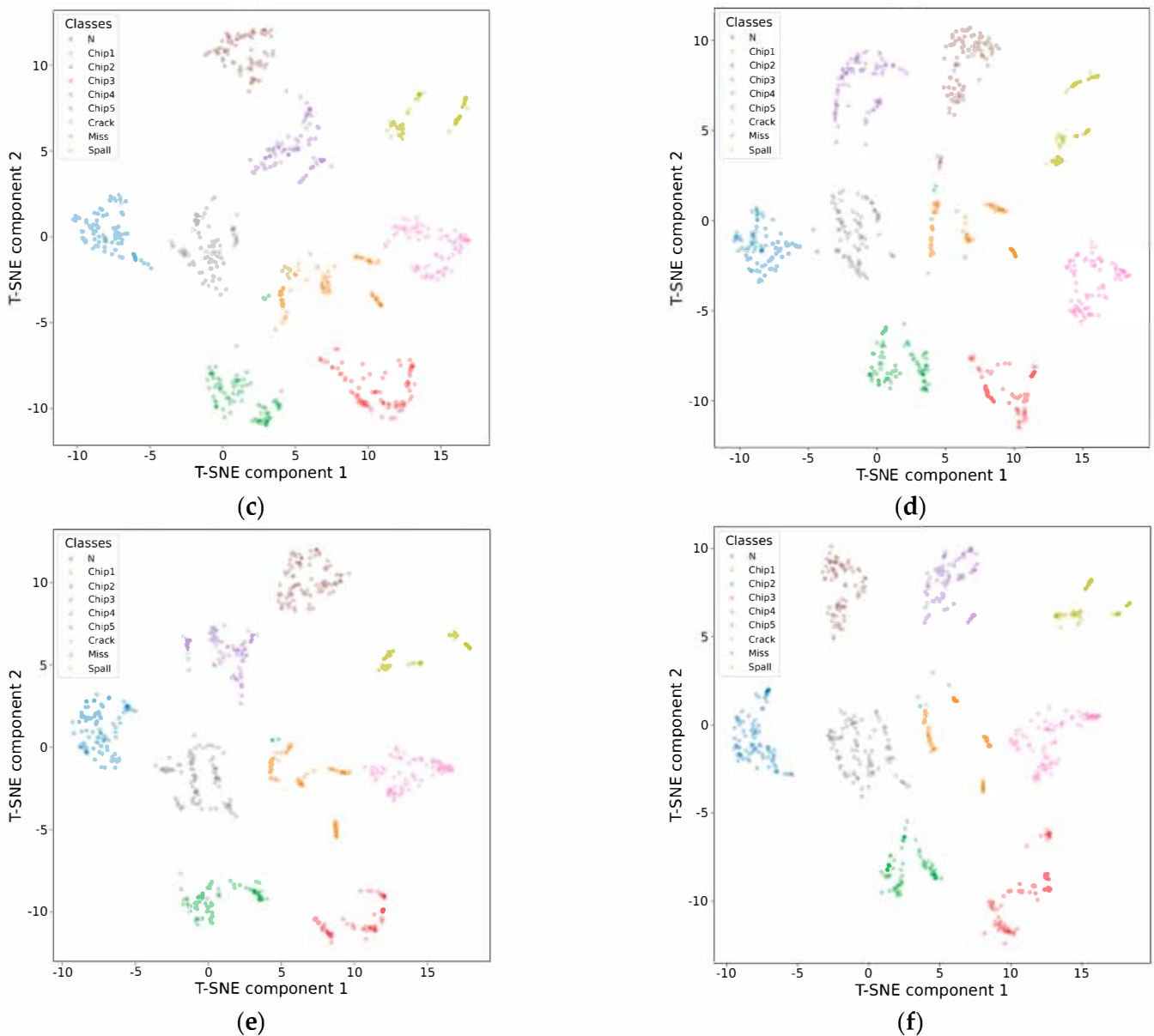


Figure 10. Feature visualization through t-SNE at gearbox sample imbalance ratio of 10:1. (a) Without synthesizer; (b) WGAN; (c) default CGDM with gradient scale fixed to 1; (d) CGDM with gradient scale corresponding to DivEn; (e) CGDM with gradient scale corresponding to FrDivEn₀; (f) CGDM with gradient scale corresponding to FrDivEn_{0.1}.

The confusion matrix was employed to further evaluate the sample synthesizers, as shown in Figure 11. In the test set, each gearbox state contains 90 samples. The closer the value on the diagonal of a state class is to 90, the more accurately that state class is diagnosed. The CGDM with a gradient scale corresponding to FrDivEn₀ achieves the highest predictive accuracy, while the CGDM with a gradient scale corresponding to FrDivEn_{0.1} is slightly less accurate.

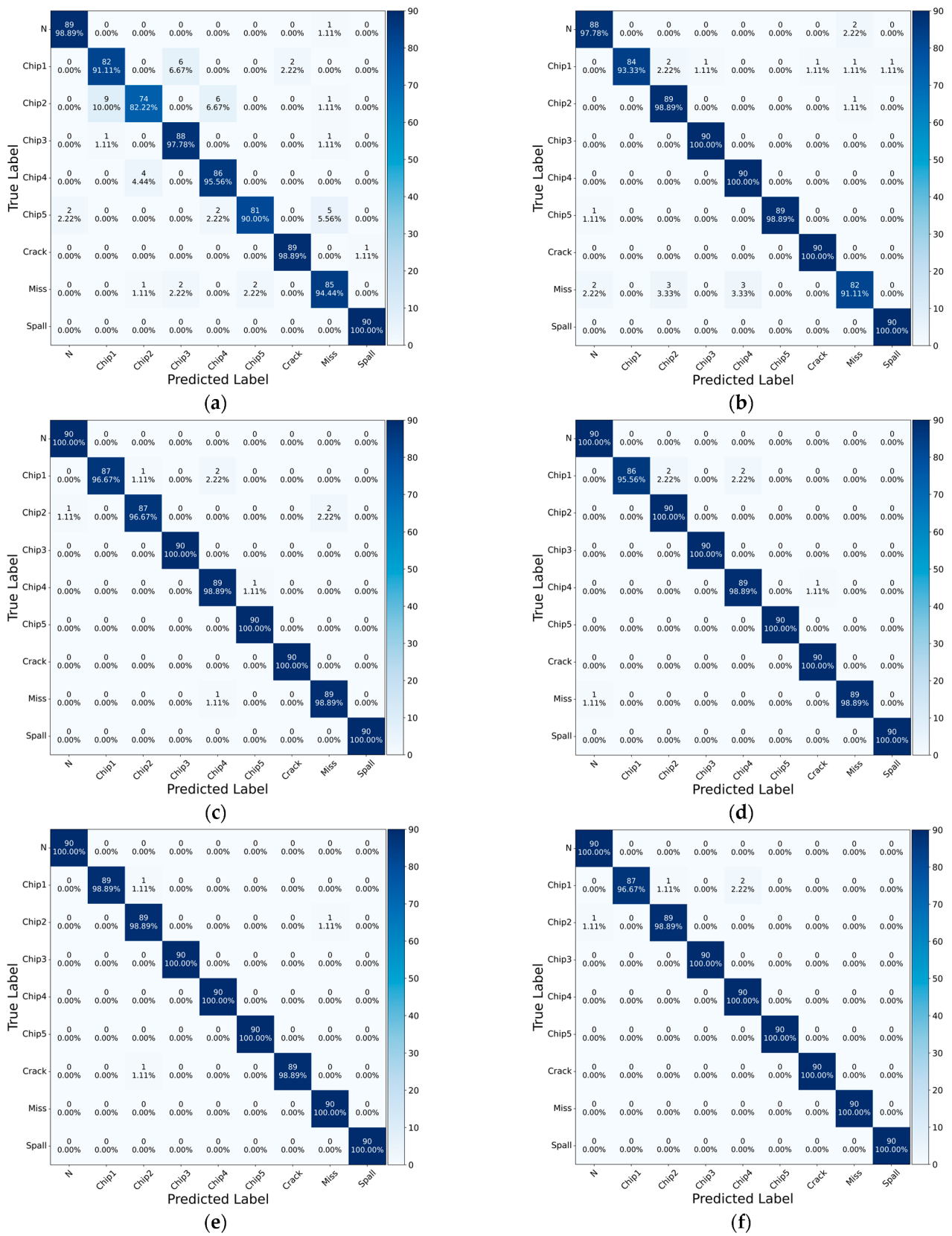


Figure 11. Confusion matrices at gearbox sample imbalance ratio of 10:1. (a) Without synthesizer; (b) WGAN; (c) default CGDM with gradient scale fixed to 1; (d) CGDM with gradient scale corresponding to DivEn; (e) CGDM with gradient scale corresponding to FrDivEn₀; (f) CGDM with gradient scale corresponding to FrDivEn_{0.1}.

The receiver operating characteristic (ROC) curves are shown in Figure 12. In different models, the CGDMs using gradient scales corresponding to FrDivEn₀ and FrDivEn_{0.1} achieve excellent performance levels, with their AUCs reaching 0.9998 and 1.0000, respectively. The evaluation metrics are shown in Table 12. The F1-score exceeds 0.99 when the CGDMs are used as the sample synthesizers, with the CGDMs using gradient scales corresponding to FrDivEn₀ and FrDivEn_{0.1} achieving F1-scores of 0.9963 and 0.9951, respectively. The confusion matrices and evaluation metrics indicate that both CGDMs using FrDivEn₀ and FrDivEn_{0.1} provide high-quality samples for the ConvNeXt model. The errors made by the proposed method and other methods are shown in Table 13. The CGDM with a gradient scale corresponding to FrDivEn_{0.1} achieves the lowest MAE of 0.0111 and RMSE of 0.1685. After combining the confusion matrix and comparing other methods, we believe that the limitation is mainly due to some of the samples being misdiagnosed as the “Chip 2” state. In addition, other faults that may occur in gearboxes but are not considered in this dataset, such as shaft bending and multi-fault fusion, may also contribute to the limitations of the proposed method. By using a different automotive machine dataset from the bearing in the previous analysis, the validity and generalizability of the imbalanced fault diagnostic method in automotive machines were tested.

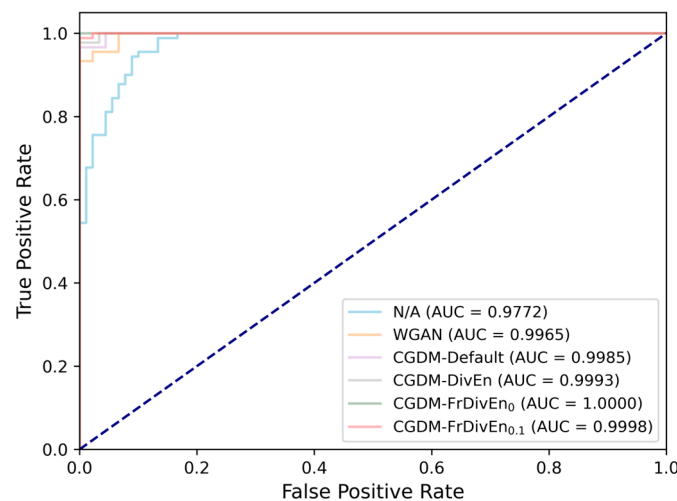


Figure 12. ROC curves at gearbox sample imbalance ratio of 10:1.

Table 12. Evaluation metrics at gearbox sample imbalance ratio of 10:1.

Evaluation Metric	N/A ¹	WGAN	CGDM			
			Default ²	DivEn ³	FrDivEn ₀ ⁴	FrDivEn _{0.1} ⁵
Macro-Precision	0.9438	0.9780	0.9902	0.9927	0.9963	0.9951
Macro-Recall	0.9432	0.9778	0.9901	0.9926	0.9963	0.9951
Macro-F1	0.9427	0.9776	0.9901	0.9926	0.9963	0.9950

¹ Without synthesizer; ² default CGDM with gradient scale fixed to 1; ³ CGDM with gradient scale corresponding to DivEn; ⁴ CGDM with gradient scale corresponding to FrDivEn₀; ⁵ CGDM with gradient scale corresponding to FrDivEn_{0.1}.

Table 13. Errors at gearbox sample imbalance ratio of 10:1.

Evaluation Metric	N/A	WGAN	CGDM			
			Default	DivEn	FrDivEn ₀	FrDivEn _{0.1}
MAE	0.1358	0.1037	0.0284	0.0210	0.0123	0.0111
RMSE	0.6648	0.7569	0.3201	0.3002	0.2277	0.1685

4.3.3. Rotor

The Wuhan University (WHU) rotor dataset was collected from an experimental automotive machinery system. Vibration signals were collected for four rotor states: normal (N), unbalanced (Unbal), misalignment (Misalign), and contact rubbing (Rub). The eddy current sensors, mounted on the sensor bracket, collected the vibration signals. The rotor vibration signals were denoised based on wavelet thresholding [39], resulting in samples with distinct characteristics and significant differences between states, thus reducing the difficulty of fault diagnosis. We used an image size of 32×32 , the same as the CIFAR-10 small-sized image sample set, for the input sample size of both the sample synthesizer and fault classifier. This approach was aimed at increasing the variability by using different models, and testing the proposed method’s effectiveness on small-sized samples.

The entropy–gradient scale s and the diagnostic accuracy with ConvNeXt as the fault classifier are presented in Table 14. The imbalanced diagnostic accuracy with various sample synthesizers and fault classifiers is shown in Table 15.

The entropy–gradient scale calculations and the fault diagnostic results provide a couple of key insights: (1) The DivEn, FrDivEn₀, and FrDivEn_{0.1} of the denoised rotor vibration signals are reduced compared to the previous bearing and gearbox datasets. For example, the FrDivEn₀ results of bearing and gearbox data at an imbalance ratio of 2:1 are 7.3113 and 7.9003, respectively, while the FrDivEn₀ of denoised rotor data is 3.6170. A decrease in the gradient scale of the CGDM accompanies this decrease in DivEn and FrDivEn. It should be noted that the gradient scale loses its ability to tune according to DivEn at five different imbalance ratios. Although DivEn still varies with the imbalance ratios, the gradient scale remains at 0. We believe this is due to the DivEn–gradient scale curve approximating the straight-line $y = 0$ when DivEn is less than 0.7 (as shown in Figure 5). The gradient scale being constantly 0 results in the CGDM using DivEn having the least diagnostic accuracy in the CGDMs. At an imbalance ratio of 40:1 and with the fine-tuned pretrained ConvNeXt as the fault classifier, the CGDM using DivEn as the sample synthesizer achieves a diagnostic accuracy of 97.22%, which is 1.69% lower than the CGDM using FrDivEn₀ and slightly lower than both the default CGDM and the CGDM using FrDivEn_{0.1}. This phenomenon reflects the limitations of setting the gradient scale according to DivEn, which may not be applicable to unfamiliar machines. (2) With all other conditions constant, ConvNeXt achieves the highest diagnostic accuracy compared to other CNN models on a sample set of size 32×32 . This is consistent with its excellent performance in the bearing and gearbox datasets. With an imbalance ratio of 40:1 and taking the CGDM using FrDivEn₀ as the sample synthesizer, ConvNeXt achieves a diagnostic accuracy of 97.50%, which is 12.50% higher than ResNet and 8.86% higher than DenseNet. Notably, the fine-tuned pretrained VGG is second only to ConvNeXt in diagnostic accuracy. We believe this is due to the stacked small-sized 3×3 convolutional kernels used in VGG, which make it suitable for small-sized sample classification tasks like those with the size of 32×32 .

Table 14. The entropy–gradient scale s and the diagnostic accuracy of rotor data.

Imbalance Ratio	2:1	5:1	10:1	20:1	40:1
Entropy–s					
FrDivEn ₀ , s	3.6170, 0.95	3.0164, 0.80	2.6692, 0.72	2.0982, 0.61	1.5783, 0.52
Accuracy	100%	100%	100%	99.72%	98.89%
FrDivEn _{0.1} , s	6.7020, 1.19	5.2597, 1.03	4.4257, 0.95	3.3008, 0.85	2.3345, 0.77
Accuracy	100%	100%	99.72%	99.44%	97.50%
DivEn, s	0.4321, 0.00	0.4046, 0.00	0.3948, 0.00	0.3458, 0.00	0.2936, 0.00
Accuracy	100%	100%	99.44%	98.61%	97.22%

Table 15. Diagnostic accuracy with imbalanced rotor sample set.

Imbalance Ratio	Synthesizer Classifier	N/A ¹	WGAN	CGDM			
				Default ²	DivEn ³	FrDivEn ₀ ⁴	FrDivEn _{0,1} ⁵
2:1	VGG	98.61%	98.89%	99.44%	99.17%	99.44%	99.44%
	GoogLeNet	95.28%	96.67%	98.06%	96.67%	98.06%	98.06%
	ResNet	96.67%	97.50%	98.33%	97.78	98.33%	98.06%
	DenseNet	97.78%	98.33%	98.61%	98.33%	98.89%	98.61%
	ConvNeXt	100%	100%	100%	100%	100%	100%
5:1	VGG	97.78%	98.61%	99.17%	98.61%	99.17%	98.89%
	GoogLeNet	91.94%	95.00%	96.67%	95.83%	96.67%	96.39%
	ResNet	93.06%	95.83%	96.67%	96.39%	97.50%	96.67%
	DenseNet	95.00%	96.11%	98.06%	96.94%	99.17%	97.78%
	ConvNeXt	99.72%	99.72%	100%	100%	100%	100%
10:1	VGG	97.22%	97.50%	98.06%	98.06%	98.61%	98.33%
	GoogLeNet	86.67%	91.94%	94.17%	92.78%	95.28%	94.44%
	ResNet	90.56%	93.89%	95.28%	94.17%	95.83%	95.56%
	DenseNet	91.94%	94.17%	96.11%	95.00%	96.67%	96.39%
	ConvNeXt	99.44%	99.44%	99.72%	99.44%	100%	99.72%
20:1	VGG	95.83%	97.22%	97.50%	97.22%	98.06%	97.78%
	GoogLeNet	81.11%	87.50%	90.28%	88.61%	93.33%	90.56%
	ResNet	84.17%	88.89%	91.67%	90.28%	93.61%	91.67%
	DenseNet	86.67%	91.94%	93.33%	92.78%	94.44%	93.61%
	ConvNeXt	97.78%	98.33%	99.17%	98.61%	99.72%	99.44%
40:1	VGG	86.94%	93.06%	93.61%	93.33%	96.11%	95.56%
	GoogLeNet	75.83%	76.39%	81.11%	80.56%	86.11%	84.17%
	ResNet	77.78%	78.06%	84.44%	83.06%	87.50%	86.67%
	DenseNet	81.67%	79.17%	84.72%	83.61%	88.89%	87.78%
	ConvNeXt	89.72%	95.56%	97.50%	97.22%	98.89%	97.50%

¹ Without synthesizer; ² default CGDM with gradient scale fixed to 1; ³ CGDM with gradient scale corresponding to DivEn; ⁴ CGDM with gradient scale corresponding to FrDivEn₀; ⁵ CGDM with gradient scale corresponding to FrDivEn_{0,1}.

The visualization through t-SNE is shown in Figure 13. Compared to other sample synthesizers, the scatter points for each state are better clustered when the CGDM using FrDivEn₀ is used as a synthesizer. Taking the “contact rubbing” and “unbalanced” states as examples, the CGDM using FrDivEn₀ enables ConvNeXt to cluster them separately without overlapping, unlike other sample synthesizers.

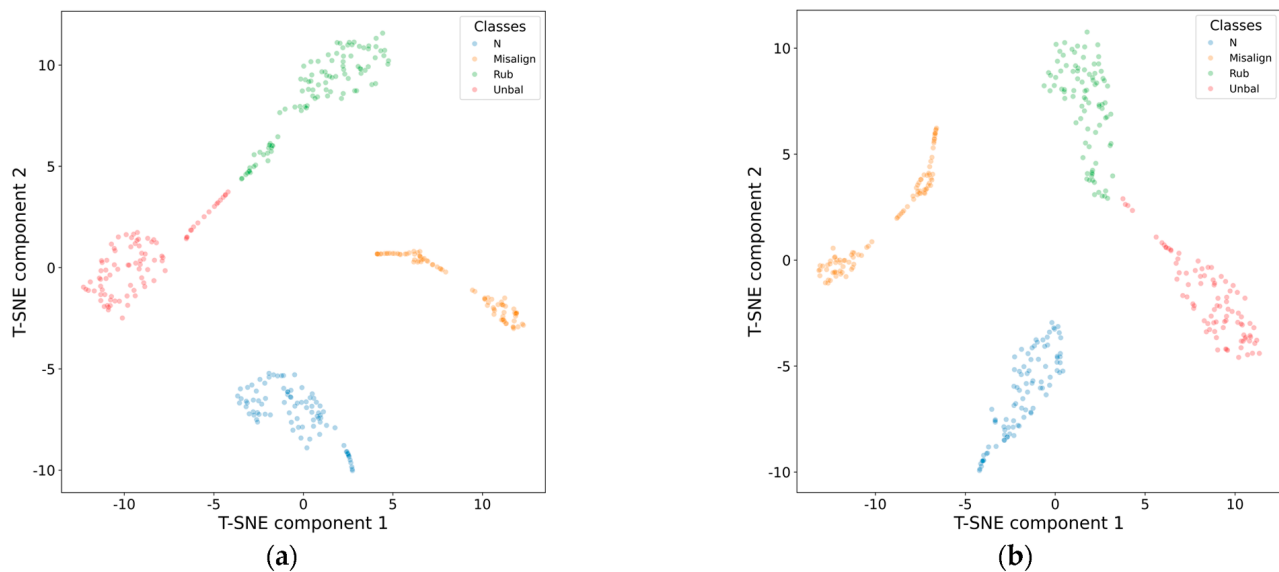


Figure 13. Cont.

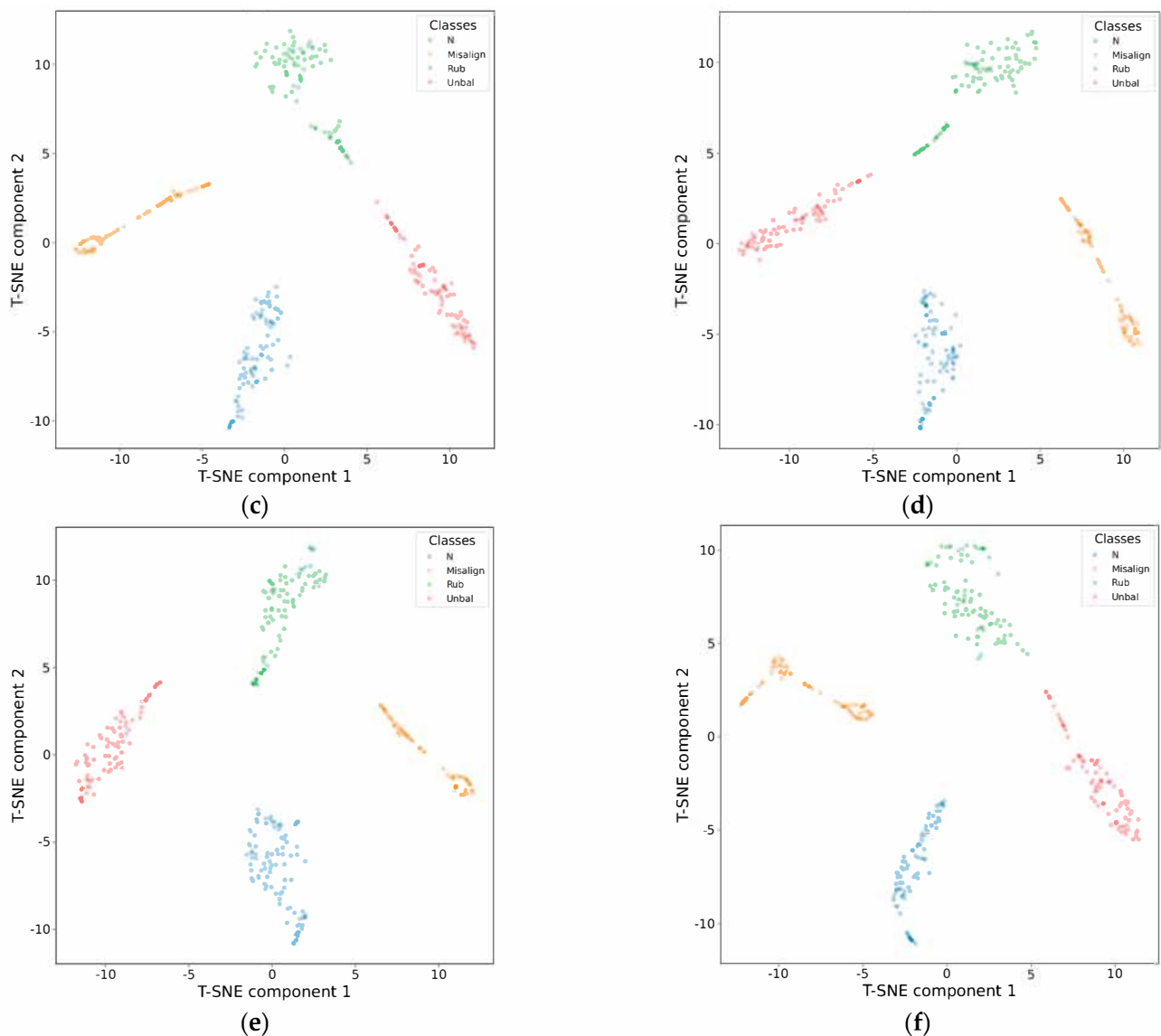


Figure 13. Feature visualization through t-SNE at rotor sample imbalance ratio of 10:1. (a) Without synthesizer; (b) WGAN; (c) default CGDM with gradient scale fixed to 1; (d) CGDM with gradient scale corresponding to DivEn; (e) CGDM with gradient scale corresponding to FrDivEn₀; (f) CGDM with gradient scale corresponding to FrDivEn_{0.1}.

The confusion matrix was employed to further evaluate the sample synthesizers, as shown in Figure 14. In the test set, each rotor state contains 90 samples. The closer the value on the diagonal of a state class is to 90, the more accurately that state class is diagnosed.

The evaluation metrics are presented in Table 16. The CGDM with a gradient scale corresponding to FrDivEn₀ achieves a predictive accuracy of 100% and an F1-score of 1, demonstrating its excellent performance in sample synthesis tasks. The errors made by the proposed method and other methods are shown in Table 17. The CGDM with a gradient scale corresponding to FrDivEn₀ achieves the lowest MAE of 0 and RMSE of 0. After combining the confusion matrix and comparing other methods, we believe that diagnosing the “rub” state is the most challenging, which may limit the effectiveness of diagnostic methods. In addition, other faults that may occur in the rotor, such as bar breaking and multi-fault fusion, which are not considered in this dataset, may also affect the limitations of the proposed method.

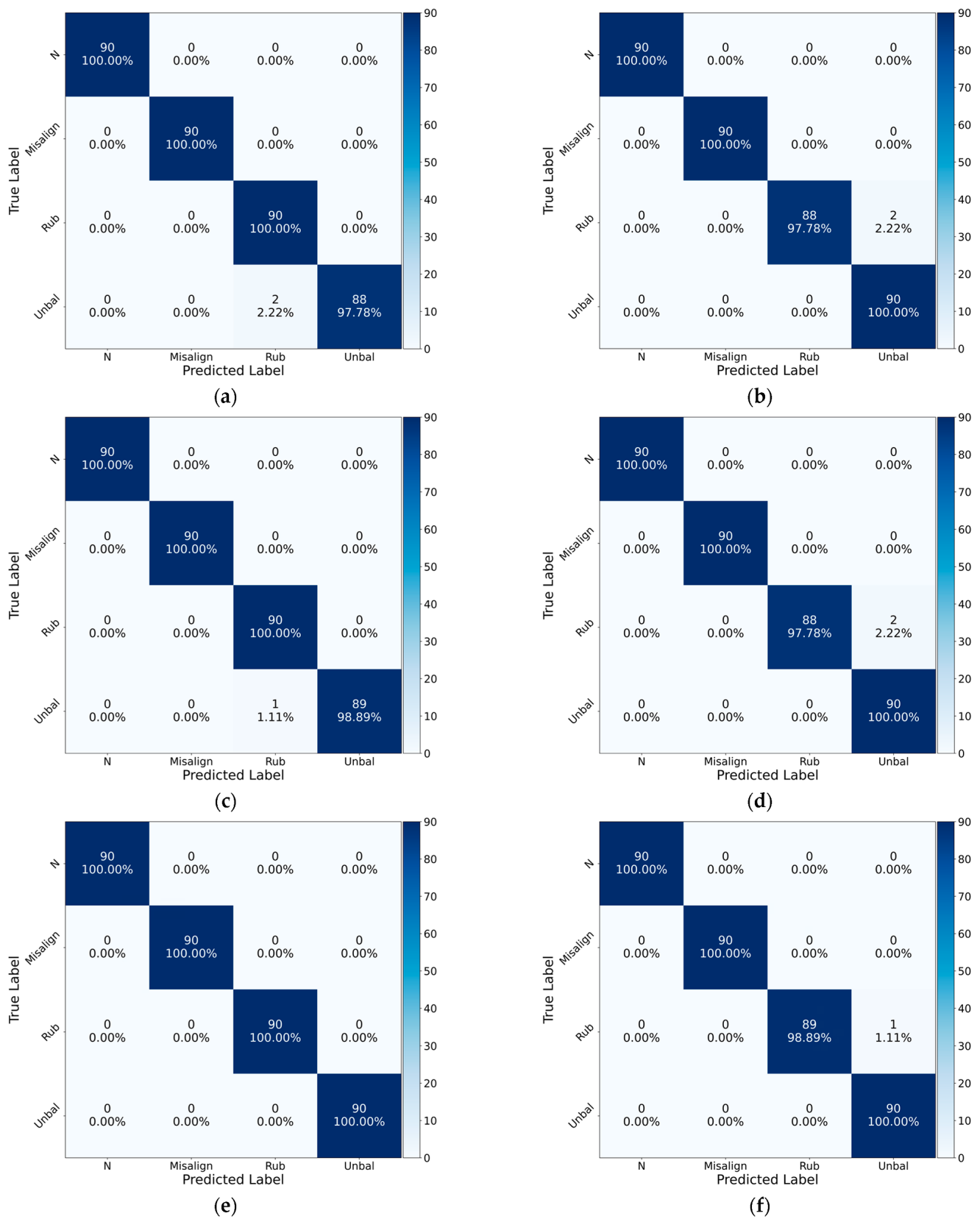


Figure 14. Confusion matrices at rotor sample imbalance ratio of 10:1. (a) Without synthesizer; (b) WGAN; (c) default CGDM with gradient scale fixed to 1; (d) CGDM with gradient scale corresponding to DivEn; (e) CGDM with gradient scale corresponding to FrDivEn₀; (f) CGDM with gradient scale corresponding to FrDivEn_{0,1}.

Table 16. Evaluation metrics at rotor sample imbalance ratio of 10:1.

Evaluation Metric	N/A ¹	WGAN	CGDM			
			Default ²	DivEn ³	FrDivEn ₀ ⁴	FrDivEn _{0.1} ⁵
Macro-Precision	0.9946	0.9946	0.9973	0.9946	1.0000	0.9973
Macro-Recall	0.9944	0.9944	0.9972	0.9944	1.0000	0.9972
Macro-F1	0.9944	0.9944	0.9972	0.9944	1.0000	0.9972

¹ Without synthesizer; ² default CGDM with gradient scale fixed to 1; ³ CGDM with gradient scale corresponding to DivEn; ⁴ CGDM with gradient scale corresponding to FrDivEn₀; ⁵ CGDM with gradient scale corresponding to FrDivEn_{0.1}.

Table 17. Errors at rotor sample imbalance ratio of 10:1.

Evaluation Metric	N/A	WGAN	CGDM			
			Default	DivEn	FrDivEn ₀	FrDivEn _{0.1}
MAE	0.0056	0.0056	0.0028	0.0056	0.0000	0.0028
RMSE	0.0745	0.0745	0.0527	0.0745	0.0000	0.0527

At the moderate imbalance ratio of 10:1 and with ConvNeXt as the fault classifier, the accuracy of the rotor fault diagnosis without the help of the sample synthesizer reaches 99.44%, which leads to a subtle difference in the diagnostic accuracy results after using different sample synthesizers to assist in the fault diagnosis. To further investigate the impacts of various synthesizers on fault diagnosis, we charted a line graph of the diagnostic accuracies over training epochs for the different synthesizers with a moderate imbalance ratio of 10:1, as shown in Figure 15.

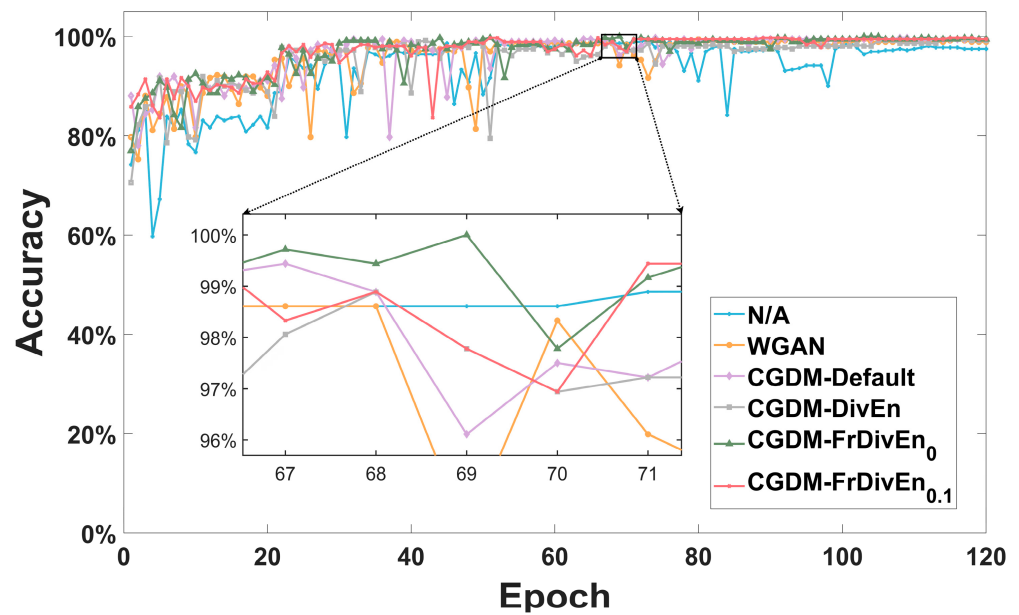


Figure 15. Diagnostic accuracy over training epochs for different synthesizers.

With a moderate imbalance ratio of 10:1 and ConvNeXt as the fault classifier, the rotor fault diagnostic accuracy without a sample synthesizer’s aid reaches 99.44%. This leaves little room for the different sample synthesizers to add boosts. The pretrained ConvNeXt’s output layer is trained during epochs 1–20, with the learning rate decaying to 0.001 between the 11th and 20th epochs, resulting in decreased oscillations in the accuracy lines compared to the first 10 epochs. The 21st epoch marks the beginning of training for ConvNeXt’s other layers. After 55 epochs, the accuracy lines for each CGDM stabilize, hovering around 98% or 99%. Throughout the entire training process, the accuracy lines of CGDMs are smoother

compared to the synthesizer-less models and the WGAN due to the higher-quality samples provided by the CGDMs. Notably, using the CGDM with FrDivEn₀ achieved the maximum diagnostic accuracy of 100% for the first time at the 69th epoch. These results reflect the advancement provided by the proposed FrDivEn trade-off for imbalanced fault diagnosis.

5. Conclusions

In the paper, to solve the problem of diversity entropy (DivEn) being insensitive to the diversity of time series, we combined DivEn with fractional order calculus to propose fractional diversity entropy (FrDivEn). Furthermore, we introduced FrDivEn to trade off the classifier-guided diffusion model's (CGDM) sample synthesis and presented an imbalanced diagnostic method for automotive machines. Specifically, this method first transforms the time series vibration signal into Gramian angular field (GAF) image samples using GAF transformation. Next, it synthesizes high-quality samples using the CGDM with a gradient scale corresponding to FrDivEn. These synthetic samples are then combined with imbalanced real samples to create a mixed sample set, which is finally input into the fine-tuned pretrained ConvNeXt for fault diagnosis.

The FrDivEn trade-off analysis, including the fractional order of FrDivEn and the gradient scale of the CGDM, was performed using the CWRU bearing dataset with a motor load of 0 and a speed of 1797 rpm. It should be noted that reliable diagnostic signals are a prerequisite for high-precision fault diagnosis. Suitable sensor selection and correct installation methods should not be overlooked, as they create the environment for obtaining reliable and high-quality diagnostic signals. In our study, we used three datasets to validate the effectiveness and generalizability of the proposed method: the CWRU bearing dataset and UConn gearbox dataset, which were not denoised, and the WHU rotor dataset, which was denoised. The results demonstrated the effectiveness of our method when using diagnostic signals with different processing methods. The main innovations and results were as follows:

- (1) For fault diagnosis in automotive machines with imbalanced sample sets, it is crucial to balance the diversity and fidelity of synthetic samples. We propose a novel signal measure called fractional diversity entropy (FrDivEn) to address this need. FrDivEn reflects vibration signal diversity at varying imbalance ratios and adjusts the generative model's emphases on diversity and fidelity during sample synthesis. Unlike the traditional DivEn, which is insensitive to signal diversity at different imbalance ratios, FrDivEn sensitively adapts to these ratios, providing a more effective reflection of signal diversity. In the CWRU bearing dataset, the differences in FrDivEn between vibration signals are significantly greater than those in DivEn. For example, the FrDivEn₀ difference between vibration data at imbalance ratios of 2:1 and 5:1 is 0.9158, which is substantially larger than the DivEn difference of 0.0061.
- (2) To select the appropriate gradient scale of the CGDM and achieve high-quality sample synthesis, we propose using FrDivEn to determine the ideal gradient scale. Utilizing the CWRU bearing dataset, we fit DivEn- and FrDivEn-gradient scale curves with various fractional orders. According to the fitting results, the FrDivEn₀- and FrDivEn_{0.1}-gradient scale curves exhibited a more suitable range of gradient scales and smoothness compared to the DivEn-gradient scale curve.
- (3) To enhance diagnostic accuracy in automotive machines, we propose a fault diagnostic method utilizing the CGDM with a gradient scale determined by FrDivEn as the sample synthesizer, and a fine-tuned pretrained ConvNeXt as the fault classifier. In an experiment using the CWRU bearing dataset with a motor load of 3 HP, a speed of 1730 rpm, and an imbalance ratio of 40:1, the diagnostic accuracies achieved using the CGDM with gradient scales corresponding to FrDivEn₀ and FrDivEn_{0.1} were 91.22% and 90.89%, respectively. These results represent improvements of 7.32% and 6.93% over the WGAN, and 4.05% and 3.67% over the CGDM with a gradient scale corresponding to DivEn. For the gearbox and rotor datasets, the diagnostic accuracies using the CGDM with FrDivEn at an imbalance ratio of 40:1 were 87.90% and 98.89%,

respectively, marking increases of 11.59% and 3.48% over the WGAN. Across these three imbalanced fault diagnosis experiments for various automotive machines, the CGDM with a gradient scale determined by FrDivEn consistently achieved a superior diagnostic accuracy compared to other sample synthesizers, with the CGDM using FrDivEn₀ performing slightly better than FrDivEn_{0.1}.

- (4) Across the three imbalanced fault diagnosis experiments for various automotive machines, the fine-tuned pretrained ConvNeXt consistently achieved the highest diagnostic accuracy compared to other fine-tuned pretrained CNN models. This was evident in both bearing and gearbox fault diagnoses with a sample size of 256×256 , as well as rotor fault diagnosis with a sample size of 32×32 . For instance, in the experiment using the CWRU bearing dataset with an imbalance ratio of 40:1 and the CGDM using FrDivEn₀ as the sample synthesizer, ConvNeXt achieved a fault diagnostic accuracy of 91.22%. This was 10.20% higher than the accuracy achieved using the pretrained VGG and 3.40% higher than that of the pretrained ResNet.

In summary, this study was focused on synthesizing high-quality samples by using the FrDivEn trade-off to achieve excellent imbalanced fault diagnostic accuracy. In the future, reducing the computational complexity and considering multi-fault fusion in the imbalanced fault diagnostic method could be the subjects of further research.

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

Funding: This research was funded by the Natural Science Foundation of China (Grant No. 52072116), the Key Research and Development Project of Hubei Province (Grant No. 2020BAB141), and the Special Fund of the Hubei Longzhong Laboratory of the Xiangyang Science and Technology Plan.

Data Availability Statement: The data supporting the conclusions of this article will be made available by the corresponding author upon request.

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

Abbreviations

AE	Auto-encoder
AUC	Area under curve
CBAM	Convolutional block attention module
CGDM	Classifier-guided diffusion model
CNN	Convolutional neural network
DDPM	Denosing diffusion probabilistic model
DivEn	Diversity entropy
DL	Deep learning
EMD	Empirical mode decomposition
FrDivEn	Fractional diversity entropy
GADF	Gramian angular difference field
GAF	Gramian angular field
GAN	Generative adversarial network
GASF	Gramian angular summation field
MAE	Mean absolute error
ResNet	Residual network
RMSE	Root mean squared error
RNN	Recurrent neural network
ROC	Receiver operating characteristic
VAE	Variational auto-encoder
WGAN	Wasserstein generative adversarial network

Appendix A

Case Western Reserve University (CWRU) bearing dataset: the bearing type chosen by CWRU Bearing Data Center is a deep groove ball bearing with the following dimensional specifications: an inside diameter of 0.9843 inches (25.0012 mm), an outside diameter of 2.0472 inches (51.9989 mm), and a ball diameter of 0.3126 inches (7.9400 mm). The accelerometers, situated on the drive end of the motor housing, were used to collect the vibration signals. To verify the applicability of the proposed methodology to different fault locations and degrees, data under the normal and fault states with varying locations and diameters are selected in this study. These include normal baseline (N), 7 mils ball fault (7_BA), 7 mils inner race fault (7_IR), 7 mils outer race fault (7_OR), 14 mils ball fault (14_BA), 14 mils inner race fault (14_IR), 14 mils outer race fault (14_OR), 21 mils ball fault (21_BA), 21 mils inner race fault (21_IR), and 21 mils outer race fault (21_OR). The CWRU bearing vibration signal acquisition platform is illustrated in Figure A1. The imbalanced time series allocation of bearing faults used to calculate the DivEn and the FrDivEn is shown in Table A1.

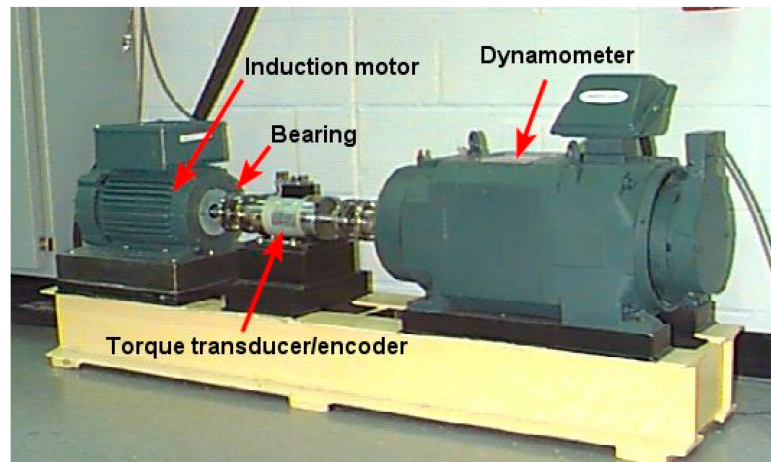


Figure A1. The CWRU bearing vibration data acquisition platform.

Table A1. Imbalanced time series allocation of bearing faults.

Fault Diameter	7 mils			14 mils			21 mils			Total	Imbalance Ratio	
	Label	7_BA ¹	7_IR ²	7_OR ³	14_BA	14_IR	14_OR	21_BA	21_IR			21_OR
Time series length		60,633	60,633	60,633	60,633	60,633	60,633	60,633	60,633	60,633	545,697	2:1
		24,253	24,253	24,253	24,253	24,253	24,253	24,253	24,253	24,253	218,277	5:1
		12,127	12,127	12,127	12,127	12,127	12,127	12,127	12,127	12,127	109,143	10:1
		6064	6064	6064	6064	6064	6064	6064	6064	6064	54,576	20:1
	3032	3032	3032	3032	3032	3032	3032	3032	3032	27,288	40:1	

¹ Ball; ² Inner race; ³ Outer race.

Appendix B

Here, we provide a detailed derivation of the Gramian angular field (GAF) transformation from Wang et al. [34]. For a given raw one-dimensional time series data $D = \{d_1, d_2, \dots, d_n\}$, which consists of n real-valued observations in chronological order, we normalize and rescale D to the interval $[-1, 1]$ by:

$$\tilde{d}_i = \frac{(d_i - \max(D)) + (d_i - \min(D))}{\max(D) - \min(D)} \tag{A1}$$

where d_i is the i -th real-valued observation of the original one-dimensional data; \tilde{d}_i is the value corresponding to the i -th element of the normalized time series data.

The Gramian matrix G is invoked to quantify the feature correlation between the encoded data [45]. From the normalized one-dimensional time series data \tilde{D} , the Gramian matrix G is defined as:

$$G = \tilde{D}^T \tilde{D} = \begin{bmatrix} \langle \tilde{d}_1, \tilde{d}_1 \rangle & \langle \tilde{d}_1, \tilde{d}_2 \rangle & \cdots & \langle \tilde{d}_1, \tilde{d}_n \rangle \\ \langle \tilde{d}_2, \tilde{d}_1 \rangle & \langle \tilde{d}_2, \tilde{d}_2 \rangle & \cdots & \langle \tilde{d}_2, \tilde{d}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \tilde{d}_n, \tilde{d}_1 \rangle & \langle \tilde{d}_n, \tilde{d}_2 \rangle & \cdots & \langle \tilde{d}_n, \tilde{d}_n \rangle \end{bmatrix} \quad (A2)$$

where \tilde{D} is the normalized one-dimensional time series data, $\tilde{D} = \{\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_n\}$.

The Gramian matrix G illustrates the degree of correlation between vectors by representing the inner product between them, which is determined by the angle between the vectors. However, the value \tilde{d}_i is not a vector. To address this, \tilde{d}_i is transformed into an angle representation ϕ_i through polar coordinate transformation. The angle ϕ_i is mapped from the value \tilde{d}_i , and the radius r_i is mapped from the timestamp tm_i corresponding to the i -th element of data. Thus, the normalized data \tilde{d}_i is represented in polar coordinates instead of the typical Cartesian coordinates by:

$$\begin{cases} \phi_i = \arccos(\tilde{d}_i), & -1 \leq \tilde{d}_i \leq 1, \tilde{d}_i \in \tilde{D} \\ r_i = \frac{tm_i}{N}, & tm_i \in \mathbb{N} \end{cases} \quad (A3)$$

where tm_i is the timestamp corresponding to the i -th element of the time series data, ensuring that the polar coordinate data retains temporal relationship; N is the normalization factor.

With the above polar transformation code, the normalized data element \tilde{d}_i and the timestamp tm_i are introduced into polar coordinates. The expression of polar coordinates is a bijective function, meaning there is a one-to-one correspondence between the independent and dependent variables. Over time, the data points unfold in polar coordinates, resembling the pattern of water ripples.

After transforming from the typical Cartesian coordinates to polar coordinates, the temporal correlation between the observed data at different time points is obtained according to the Gramian matrix G . The Gramian angular summation field (GASF) is given by:

$$GASF = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cos(\phi_1 + \phi_2) & \cdots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cos(\phi_2 + \phi_2) & \cdots & \cos(\phi_2 + \phi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cos(\phi_n + \phi_2) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix} \quad (A4)$$

$$= \tilde{D}^T \cdot \tilde{D} - \sqrt{I - \tilde{D}^T} \cdot \sqrt{I - \tilde{D}}$$

where I is a unit row vector of length n .

Similarly, we can calculate the sine of the angular difference between all polar data. The Gramian angular difference field (GADF) is given by

$$GADF = \begin{bmatrix} \sin(\phi_1 - \phi_1) & \sin(\phi_1 - \phi_2) & \cdots & \sin(\phi_1 - \phi_n) \\ \sin(\phi_2 - \phi_1) & \sin(\phi_2 - \phi_2) & \cdots & \sin(\phi_2 - \phi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \sin(\phi_n - \phi_1) & \sin(\phi_n - \phi_2) & \cdots & \sin(\phi_n - \phi_n) \end{bmatrix} \quad (A5)$$

$$= \sqrt{I - \tilde{D}^T} \cdot \tilde{D} - \tilde{D}^T \cdot \sqrt{I - \tilde{D}}$$

Each element of GASF or GADF is mapped to a pixel of an image to obtain a GASF image or GADF image, respectively.

References

1. Yan, S.; Sun, W.; Xia, Y. A joint fault-tolerant and fault diagnosis strategy for multiple actuator faults off full-vehicle active suspension systems. *IEEE Trans. Autom. Sci. Eng.* **2024**, *early access*. [[CrossRef](#)]
2. Singh, S.; Howard, C.Q.; Hansen, C.H.; Köpke, U.G. Analytical validation of an explicit finite element model of a rolling element bearing with a localised line spall. *J. Sound Vib.* **2018**, *416*, 94–110. [[CrossRef](#)]
3. Moshrefzadeh, A.; Fasana, A.; Antoni, J. The spectral amplitude modulation: A nonlinear filtering process for diagnosis of rolling element bearings. *Mech. Syst. Signal Process.* **2019**, *132*, 253–276. [[CrossRef](#)]
4. Li, X.; Zhang, W.; Ding, Q.; Sun, J.-Q. Intelligent automotive machinery fault diagnosis based on deep learning using data augmentation. *J. Intell. Manuf.* **2020**, *31*, 433–452. [[CrossRef](#)]
5. Yang, E.; Yi, O. Enhancing road safety: Deep learning-based intelligent driver drowsiness detection for advanced driver-assistance systems. *Electronics* **2024**, *13*, 708. [[CrossRef](#)]
6. Sree, S.R.; Lydia, E.L.; Anupama, C.S.S.; Nemani, R.; Lee, S.; Joshi, G.P.; Cho, W. A battle royale optimization with feature fusion-based automated fruit disease grading and classification. *AIMS Math.* **2024**, *9*, 11432–11451. [[CrossRef](#)]
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
8. Zhang, H.; Wang, Z.; Liu, D. A comprehensive review of stability analysis of continuous-time recurrent neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 1229–1262. [[CrossRef](#)]
9. Dong, S.; Wang, P.; Abbas, K. A survey on deep learning and its applications. *Comput. Sci. Rev.* **2021**, *40*, 100379. [[CrossRef](#)]
10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
11. 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 automotive machinery. *Measurement* **2023**, *206*, 112346. [[CrossRef](#)]
12. Zhou, H.; Chen, W.; Shen, C.; Cheng, L.; Xia, M. Intelligent Machine Fault Diagnosis with Effective Denoising Using EEMD-ICA-FuzzyEn and CNN. *Int. J. Prod. Res.* **2022**, *61*, 8252–8264. [[CrossRef](#)]
13. Lee, C.-Y.; Le, T.-A. Identifying Faults of Rolling Element Based on Persistence Spectrum and Convolutional Neural Network with ResNet Structure. *IEEE Access* **2021**, *9*, 78241–78252. [[CrossRef](#)]
14. Chang, M.; Yao, D.; Yang, J. Intelligent Fault Diagnosis of Rolling Bearings Using Efficient and Lightweight ResNet Networks Based on an Attention Mechanism. *IEEE Sens. J.* **2023**, *23*, 9136–9145. [[CrossRef](#)]
15. Zhao, Z.; Li, T.; Wu, J.; Sun, C.; Wang, S.; Yan, R.; Chen, X. Deep Learning Algorithms for Automotive machinery Intelligent Diagnosis: An Open Source Benchmark Study. *ISA Trans.* **2020**, *107*, 224–255. [[CrossRef](#)] [[PubMed](#)]
16. Yang, X.; Ye, T.; Yuan, X.; Zhu, W.; Mei, X.; Zhou, F. A novel data augmentation method based on denoising diffusion probabilistic model for fault diagnosis under imbalanced data. *IEEE Trans. Ind. Inform.* **2024**, *20*, 7820–7831. [[CrossRef](#)]
17. AlHalawani, S.; Benjdira, B.; Ammar, A.; Koubaa, A.; Ali, A.M. DiffPlate: A Diffusion Model for Super-Resolution of License Plate Images. *Electronics* **2024**, *13*, 2670. [[CrossRef](#)]
18. Dhariwal, P.; Nichol, A. Diffusion models beat GANs on image synthesis. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 8780–8794.
19. Wang, X.; Si, S.; Li, Y. Multiscale diversity entropy: A novel dynamical measure for fault diagnosis of automotive machinery. *IEEE Trans. Ind. Inform.* **2020**, *17*, 5419–5429. [[CrossRef](#)]
20. Li, Y.; Jiao, Z.; Wang, S.; Feng, K.; Liu, Z. Cross diversity entropy-based feature extraction for fault diagnosis of rotor system. *IEEE-ASME Trans. Mech.* **2023**, *early access*. [[CrossRef](#)]
21. Zhu, Z.; Cheng, J.; Wang, P.; Wang, J.; Kang, X.; Yang, Y. A novel fault diagnosis framework for automotive machinery with hierarchical multiscale symbolic diversity entropy and robust twin hyperdisk-based tensor machine. *Reliab. Eng. Syst. Safe.* **2023**, *231*, 109037. [[CrossRef](#)]
22. Xiao, Z.; Ma, H.; Lu, Y.; Zhang, G.; Liu, Z.; Song, Q. Real-Time milling tool breakage monitoring based on multiscale standard deviation diversity entropy. *Int. J. Mech. Sci.* **2023**, *240*, 107929. [[CrossRef](#)]
23. Fogret, É.; Pellat-Finet, P. A Light-Ray Approach to Fractional Fourier Optics. *Fractal Fract.* **2023**, *7*, 505. [[CrossRef](#)]
24. Sun, Y.; Cao, Y.; Li, P. Fault diagnosis for train plug door using weighted fractional wavelet packet decomposition energy entropy. *Accid. Anal. Prev.* **2022**, *166*, 106549. [[CrossRef](#)] [[PubMed](#)]
25. Huang, G.; Qin, H.-y.; Chen, Q.; Shi, Z.; Jiang, S.; Huang, C. Research on Application of Fractional Calculus Operator in Image Underlying Processing. *Fractal Fract.* **2024**, *8*, 37. [[CrossRef](#)]
26. Chen, L.; Gao, J.; Lopes, A.M.; Zhang, Z.; Chu, Z.; Wu, R. Adaptive fractional-order genetic-particle swarm optimization Otsu algorithm for image segmentation. *Appl. Intell.* **2023**, *53*, 26949–26966. [[CrossRef](#)]
27. Zhang, N.; Zhu, W.-Y.; Jin, P.; Huang, G.; Pu, Y.-F. Fractional Fuzzy Neural System: Fractional Differential-Based Compensation Prediction for Reputation Infringement Cases. *Fractal Fract.* **2024**, *8*, 172. [[CrossRef](#)]
28. Joshi, M.; Bhosale, S.; Vyawahare, V.A. A survey of fractional calculus applications in artificial neural networks. *Artif. Intell. Rev.* **2023**, *56*, 13897–13950. [[CrossRef](#)]

29. Takens, F. Detecting Strange Attractors in Turbulence. In *Dynamical Systems and Turbulence*; Rand, D., Young, L.S., Eds.; Springer: Warwick, UK, 1980; pp. 366–381.
30. Shannon, C.E. A mathematical theory of communication. *ACM SIGMOB. Mob. Comput. Commun. Rev.* **2001**, *5*, 3–55. [[CrossRef](#)]
31. Li, Y.; Mu, L. Particle Swarm Optimization Fractional Slope Entropy: A New Time Series Complexity Indicator for Bearing Fault Diagnosis. *Fractal Fract.* **2022**, *6*, 345. [[CrossRef](#)]
32. Zheng, J.; Pan, H. Use of generalized refined composite multiscale fractional dispersion entropy to diagnose the faults of rolling bearing. *Nonlinear Dyn.* **2021**, *101*, 1417–1440. [[CrossRef](#)]
33. Dong, Y.; Liu, Q.; Du, B.; Zhang, L. Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2022**, *31*, 1559–1572. [[CrossRef](#)] [[PubMed](#)]
34. Wang, Z.; Oates, T. Imaging time-series to improve classification and imputation. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.
35. Machado, J.T. Fractional Order Generalized Information. *Entropy* **2014**, *16*, 2350–2361. [[CrossRef](#)]
36. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. *arXiv* **2022**, arXiv:2201.03545.
37. Shao, S.; McAleer, S.; Yan, R.; Baldi, P. Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Trans. Ind. Inform.* **2018**, *15*, 2446–2455. [[CrossRef](#)]
38. Cao, P.; Zhang, S.; Tang, J. Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning. *IEEE Access* **2018**, *6*, 26241–26253. [[CrossRef](#)]
39. Liu, D.; Xiao, Z.H.; Hu, X.; Zhang, C.X.; Malik, O.P. Feature extraction of rotor fault based on EEMD and curve code. *Measurement* **2019**, *135*, 712–724. [[CrossRef](#)]
40. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875.
41. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
42. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 15 October 2015; pp. 1–9.
43. Huang, G.; Liu, Z.; Laurens, V.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
44. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
45. Almeida, M.C.; Asada, E.N.; Garcia, A.V. On the use of Gram matrix in observability analysis. *IEEE Trans. Power Syst.* **2008**, *23*, 249–251. [[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.