

# Deep Squeeze-enhanced Axial Transformer SMOTE: A Novel Approach For Imbalance-bearing Fault Diagnosis

**Abstract**—In industrial production, the imbalance between sparse failure signals and abundant normal signals biases fault diagnosis models towards the normal class, reducing accuracy and reliability. Existing synthetic data methods help mitigate this but often neglect spatiotemporal long-distance dependencies and local feature similarities, limiting their effectiveness. To address these limitations, we propose Deep Squeeze-enhanced Axial Transformer Synthetic Minority Oversampling Technique (DSEA-SMOTE). This method integrates a specially designed continuous wavelet transform data filtering preprocessing technology module that converts one-dimensional time-domain data into a two-dimensional feature map, enhancing model performance while simplifying feature learning. It then captures spatiotemporal long-distance dependencies and feature similarities in space-time slices through the novel Squeeze-enhanced Axial Attention mechanism and Auxiliary Feature Classifier. A Multi-category Sample Feature Filtering Technology module is also introduced to further improve synthesis quality. Additionally, we refine the loss function based on the Auxiliary Feature Classifier to enhance generation quality. Experimental evaluations on two real-world datasets show that DSEA-SMOTE outperforms recent techniques. Ablation experiments further verify the effectiveness of each component in our design.

**Index Terms**—Data imbalance, Fault diagnosis, Synthetic Minority Oversampling Technique(SMOTE), Deep Learning, Oversampling.

## I. INTRODUCTION

IN industrial production, class imbalance between normal and sparse abnormal signals biases fault diagnosis models toward the normal class, reducing accuracy and stability. Additionally, obtaining enough abnormal signals for effective diagnosis requires costly, long-term experiments with specialized equipment, which is often impractical [1]. To address these challenges, researchers have used synthetic data, with methods like SMOTE, GANs, and VAEs showing success [2] [3] [4].

Traditional methods like SMOTE balance datasets by interpolating the minority class, with recent studies introducing various improvements, mainly combining interpolation with new technologies [5] [6] [7] [8]. For example, Shen et al. [9] proposed SSMOTE-STCN, which combines SMOTE with a self-attention-based Temporal Convolutional Network (TCN) to generate synthetic fault data for chiller diagnosis.

While easy to use, these methods often suffer from inaccurate synthesis, high costs, and poor generalization [10], and none address spatiotemporal long-distance dependencies or the similarity of overall and local category features.

Generative models like VAEs and GANs are used to address class imbalance by generating realistic synthetic samples for fault diagnosis. However, recent studies mainly focus on improving loss functions and adding new components [11] [12] [13] [14] [15]. For example, Huo et al. [16] proposed RMA-WGAN-1DCNN for industrial bearing fault diagnosis, which incorporates multiple components for improvement.

These models are successful in generating samples but rely on a generator-discriminator framework, making training unstable. They focus on overall sample similarity, which can blur classification boundaries and ignore long-distance dependencies and local category features in spatiotemporal slices, reducing diagnostic performance [17] [18]. GAN-based models also face challenges like mode collapse, catastrophic forgetting, and convergence issues, requiring extensive adjustments to model parameters [19] [20], with inconsistent sample quality despite efforts.

Overall, most studies based on these methods and variants suffer from poor generalization, high data requirements, and training challenges due to their inherent limitations. They also fail to effectively capture spatiotemporal long-range dependencies in signals, such as periodicity and causal correlations, as well as the relationships between space-time slices.

To address the challenges above, we propose a novel framework: Deep Squeeze-enhanced Axial Transformer Synthetic Minority Oversampling Technique (DSEA-SMOTE). This method effectively addresses fault class imbalance by synthesizing high-quality data. DSEA-SMOTE overcomes problems of poor generalization, high data volume requirements, and training difficulties found in SMOTE, VAE, GANs and their variants. Its unique architecture enhances spatiotemporal long-distance dependencies and the similarity of overall and local features in spatiotemporal slices, filling gaps in previous research. The main contributions of this paper are as follows:

- 1) This paper presents a novel minority class oversampling framework, DSEA-SMOTE, which integrates several innovative components: a continuous wavelet transform-based data filtering module, a squeeze-enhanced Axial attention mechanism module, an auxiliary feature classifier with an improved loss function, and a multi-category sample feature filtering module. These innovations effectively address class imbalance in fault diagnosis. The method outperforms previous approaches on two real-world datasets, with ablation studies confirming the effectiveness of the proposed components.

- 2) Based on the framework, the squeeze-enhanced Axial attention mechanism module (SEA-module) and the auxil-

iary feature classifier module (AFC-module) are designed, along with an improved loss function and the new auxiliary classification feature loss function. This design enables our model framework to effectively capture spatiotemporal long-distance dependencies and the similarities between overall and local features of spatiotemporal slices, resulting in more comprehensive and balanced high-quality samples.

3) By incorporating continuous wavelet transform data filtering preprocessing technology (CWT-module) into the framework, the difficulty of learning features from the original samples is reduced, while enhancing feature extraction.

4) Previous studies lacked clear criteria for retaining or discarding synthetic samples, leading to inconsistent quality. To address this, we introduce a multi-category sample feature filtering method (Filtering-module) within DSEA-SMOTE. This method uses quantifiable quality indicators as thresholds, incorporating a time-frequency domain-based filtering approach to improve synthetic sample quality and reduce feature degradation.

The paper is organized as follows: Section II and Section III cover the theory and model framework. Section IV presents the experimental setup, while Sections V and VI provide experimental results on two real-world datasets. Section VII discusses the ablation study, and Section VIII concludes the paper.

## II. THEORETICAL KNOWLEDGE

### A. Continuous Wavelet Transform data filtering preprocessing technology(CWT-module)

Continuous Wavelet Transform (CWT) is a time-frequency domain analysis tool used to process non-stationary signals, such as seismic, medical, and audio signals [21]. Unlike traditional time or frequency domain methods (e.g., FFT, DFT), CWT combines both domains, providing richer spatiotemporal features that help models extract more relevant information for learning. For a signal  $x(t)$ , the CWT at scale  $a \in R^{+*}$  and position  $b \in R$  is defined as:

$$X_w(a, b) = \frac{1}{|a|^{\frac{1}{2}}} \int_{-\infty}^{\infty} x(t) \overline{\Psi}(\frac{t-b}{a}) dt \quad (1)$$

Here,  $X_w(a, b)$  is the CWT result,  $\Psi(t)$  is the mother wavelet,  $\overline{\Psi}(t)$  is its conjugate. In this paper, CWT is used for data filtering preprocessing, enhancing feature extraction and simplifying the learning process. The original signal,  $X = \{x_1, x_2, \dots, x_N\}$ , is a 1D time-domain signal. After CWT transformation, a 1D time-frequency signal  $X' = \{x'_1, x'_2, \dots, x'_T\}$ . This is processed by an energy contour heat map module  $C(x)$ , which filters out irrelevant information, producing a 2D feature map  $X''$ :

$$X'' = C(X') \quad (2)$$

### B. SEA Attention Mechanism Theory of SEA-module

Squeeze-enhanced Axial attention mechanism(SEA), proposed by Wan et al. [22], improves upon window and axial attention mechanisms in the context of global attention. It is a novel lightweight attention mechanism primarily used in

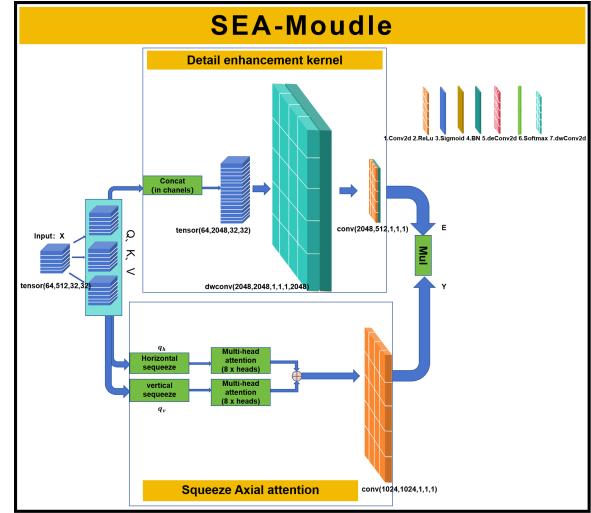


Fig. 1: Squeeze axial attention Mechanism module(SEA-module).

computer vision. SEA consists of two key components: the Squeeze Axial attention mechanism and the detail enhancement kernel. The squeeze axial process is expressed as follows:

$$q_{(h)} = \frac{1}{W}(q A_W^T) \quad (3)$$

$$q_{(v)} = \frac{1}{H}(q A_H^T) \quad (4)$$

Here,  $A_W$  and  $A_H$  are weight matrices.  $q_{(v)}$  and  $q_{(h)}$  represent the squeezed input vectors  $X = q = (Q, K, V)$  along the vertical and horizontal directions, respectively. After the squeeze axial processing, the resulting outputs are passed through an 8-head attention mechanism (Transformer) and summed to obtain the squeeze axial attention representation  $Y$ . To mitigate the information loss from the squeeze axial process, SEA introduces a set of detail enhancement kernels. These kernels provide a detailed supplement  $E$  to the input vector via a series of concatenation and convolution operations. Finally, the output  $O$  is obtained by applying a sigmoid weight mapping to both  $Y$  and  $E$ , followed by element-wise multiplication:

$$O = \text{sigmoid}(Y) \odot E \quad (5)$$

This formula demonstrates how SEA enhances local details while capturing both long-distance and local dependencies via squeeze axial attention. The entire process of SEA module can be summarized as shown in the Fig.1.

### C. Auxiliary Feature Classifier Module(AFC-module) and Improved Loss Function

We design an Auxiliary Feature Classifier (AFC) module in the decoder, which classifies synthesized samples based on their features. During training, the AFC loss is calculated for both synthetic and real samples, strengthening class boundaries and improving feature resolution, thereby enhancing spatiotemporal slice quality and local category feature similarity. The AFC loss is expressed as follows:

$$L_{AFC} = E_{g \sim P(G)}[\log(P(Y = y|S_g))] + E_{o \sim P(O)}[\log(P(Y = y|S_o))] \quad (6)$$

Here,  $P(G)$  and  $P(O)$  represent the distributions of the generated and real sample spaces, respectively, and  $Y = y$  denotes the classification label.  $g$  represents a synthetic sample, while  $o$  represents an original real sample. This loss helps align the generated space with the real space in terms of both overall and local feature similarities. Additionally, a distance loss is introduced to approximate the real space based on spatiotemporal long-distance dependencies. The distance loss is defined as:

$$L_D = E_{g \sim P(G)} [\log(1 - D(E(g)))] + E_{o \sim P(O)} [\log D(O)] \quad (7)$$

Here,  $D(g)$  and  $E(g)$  represent the decoder and encoder, respectively. The total loss function, combining AFC loss and distance loss, is:

$$L_{total} = L_{AFC} + L_D \quad (8)$$

#### D. Multi-category sample feature filtering technology (Filtering-module)

To improve synthesis quality and prevent underfitting and overfitting, we introduce a multi-category sample feature filtering technique. This method collects high-quality synthetic samples while filtering out low-quality ones, based on a quality threshold for different class imbalance ratios. First, we define the category feature recognition accuracy  $G_{acc}$ , which measures the recognition accuracy of the generated samples:

$$G_{acc} = \frac{G_{pred\_true}}{G_{total}} \quad (9)$$

where  $G_{pred\_true}$  is the number of generated samples whose labels match the real sample labels, and  $G_{total}$  is the total number of generated samples. The calculation of  $G_{acc}$  depends on the output of the AFC module. This metric evaluates the similarity between generated and real samples in terms of category features. Next, we define the generation quality metric  $G_{quality}$ , calculated by:

$$G_{quality} = Q(G) \quad (10)$$

Here,  $G_{quality}$  is measured using similarity metrics such as Pearson Correlation Coefficient (PCC).  $G$  represents a synthetic sample. We also define the class Balance Ratio (BR) to measure class imbalance:

$$BR = \frac{\text{Minority}_{num}}{\text{Majority}_{num}} \quad (11)$$

where  $\text{Minority}_{num}$  and  $\text{Majority}_{num}$  are the number of samples in the minority and majority classes, respectively. Finally, the multi-category sample feature filtering function is defined as:

$$F(G) = \begin{cases} 1, & \text{if } G_{acc}, G_{quality} > \beta(BR) \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where  $\beta(BR)$  increases with BR and can be taken as a constant for convenience. We use  $\beta(BR) = 0.6$  as a threshold. If  $F(G) = 1$ , the generated sample  $G$  is accepted.

### III. PROPOSED APPROACH

#### A. Overall method flow

The proposed method can be summarized in three parts:

Part 1: Data preprocessing includes normalization, slicing, dataset division, class imbalance construction, and undersampling alignment. The signal data is normalized in space-time to reduce complexity, sliced with a sliding window, and split into training and test sets. Training datasets with varying BRs are then created and aligned before being input to DSEA-SMOTE.

Part 2: The DSEA-SMOTE framework processes the pre-processed data for training, sample synthesis, and collection under specific class imbalance conditions.

Part 3: Downstream diagnosis classification involves adding synthetic samples to the unbalanced training set to balance the classes, followed by training a diagnosis classifier and testing it on the test set.

#### B. DSEA-SMOTE

The DSEA-SMOTE framework, based on an encoder-decoder architecture, is designed to address class imbalance in fault diagnosis. It consists of the following 5 steps:

Step 1: The 1D input  $X$  is processed by CWT, transforming it into the time-frequency domain, then converted into a three-channel 2D color map via  $C(X)$ , and finally into a single-channel 2D grayscale map  $X''$ .

Step 2: The encoder upsamples the result from Step 1 to obtain more details and features. At the same time, a synthetic tensor  $Y$  is generated based on the extracted feature information.

Step 3: The SEA module further processes the tensor from the encoder to obtain a synthetic tensor  $O$  with long-distance dependencies in space-time and overall and local feature details.

Step 4: The decoder downsamples  $O$  to extract features for synthesis, using the embedded AFC module to enhance feature similarity in the spatiotemporal slices. The process is guided by an improved loss function, and the synthesized samples are output.

Step 5: The Filtering module selects qualified synthetic samples to improve the final synthesis quality.

A schematic diagram of the framework is provided in Fig. 2 for better clarity.

### IV. EXPERIMENTAL SETUP

#### A. Evaluation Metrics Settings

We use various metrics for objective evaluation at each stage. For dataset construction, class imbalance is measured by Balance Ratio (BR). During sample generation, we assess performance using PCC ( $[-1, 1]$ , higher is better), generated category accuracy ( $G_{acc}$ ) ( $[0, 1]$ , higher is better), and Jensen-Shannon(JS) and Kullback-Leibler(KL) divergences. JS ranges from 0 (identical) to 1 (completely different), while KL ranges from 0 to infinity (higher indicates greater distribution differences). As these metrics may overlook nonlinearities and long-distance spatiotemporal dependencies, final quality is evaluated based on downstream classifier performance.

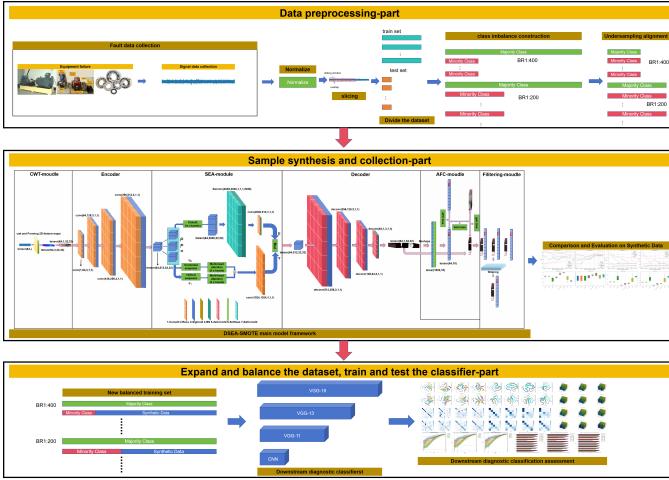


Fig. 2: Overall Framework diagram of this article.

For the diagnostic classifier, we use accuracy, F1-score, and G-mean. All metrics range from 0 to 1, with higher values indicating better performance. For a comprehensive evaluation, results are obtained on downstream classifiers of varying types and depths: VGG-16, VGG-13, VGG-11, and a simple 3-layer CNN.

This paper compares the proposed method with LSGAN [23], ACGAN, SAGAN [24], DCGAN, ACWGAN-GP [25], CVAE, and Deep-SMOTE [8].

### B. Device and hyperparameter settings

The GPU used in the experimental equipment is RTX-3080ti (12GB) and the memory is 90GB. The optimizer and batch size of DSEA-SMOTE and downstream diagnostic classifier are Adam and 64. The learning rate and epoch of DSEA-SMOTE are set to 0.0005 and 1500, and those of the downstream classifier are set to 0.0001 and 200.

## V. EXPERIMENTAL VERIFICATION AND ANALYSIS ON THE CWRU BEARING DATASET

### A. Dataset description and processing

The CWRU bearing dataset in this section, from the Power Engineering Laboratory of CWRU University, is widely used in bearing fault and class imbalance research [26]. It is available for download on the CWRU website. For this study, only relevant data files are used, focusing on multi-class imbalanced conditions. Fault data from the DE at 12 kHz with 0.007", 0.014", and 0.021" diameters and 2hp load (1750rpm) are selected, along with corresponding normal data. The data is divided into 10 categories (normal and faults at various positions). Then we perform a normalization on the data. We then apply a sliding window slicing method, creating 256-point time domain slices. The datasets are split into 80% training and 20% testing sets, with 2000 training slices and 500 test slices per class. To simulate multi-class imbalance, the normal class remains unchanged, and fault class samples are randomly reduced to achieve different BRs. Undersampling alignment is applied to balance the training set before feeding

it into DSEA-SMOTE. After balancing, majority class samples are no longer used for filling. The preprocessed data status is shown in Table I. Finally, the preprocessed data is passed through the CWT-module of DSEA-SMOTE to generate a two-dimensional feature map.

### B. Data generation and evaluation

The processed data, as a two-dimensional grayscale feature map, is sent to the DSEA-SMOTE encoder for training. Both model training and generation results are output as grayscale feature maps. The quality is evaluated using metrics like PCC,  $G_{acc}$ , KL, and JS to assess similarity to real samples. Comparison results for different models and BRs are shown in Fig. 3a (line graphs), with variations across models and BRs illustrated in Fig. 3b (box plots).

Fig. 3a shows that DSEA-SMOTE excels in PCC and  $G_{acc}$ , achieving the best generation quality across most BRs. For example, under the PCC index of BR 1:5, DSEA-SMOTE (0.9533) outperforms LSGAN (0.8895), ACGAN (0.9031), SAGAN (0.8248), DCGAN (0.8957), ACWGAN-GP (0.8723), CVAE (0.8801), and Deep-SMOTE (0.9198) by 0.0638, 0.0502, 0.1285, 0.0576, 0.081, 0.0732, and 0.0335. While DSEA-SMOTE doesn't lead in KL and JS metrics, it consistently ranks second. CVAE performs better in these metrics, but this may indicate overfitting, where the model captures background noise that could hurt downstream classification, as confirmed by classifier results. Fig. 3b further demonstrates DSEA-SMOTE's robustness across different BRs, highlighting its strong generalization ability, while other models show greater variability, indicating sensitivity to class imbalance changes.

### C. Performance of downstream fault diagnosis classifiers

DSEA-SMOTE generates samples to balance multi-class imbalanced training sets, which are used to train downstream fault diagnosis classifiers. Multiple classifiers with varying depths are trained to avoid bias. Results, averaged over three runs, are evaluated using accuracy, F1-score, and G-mean for objective assessment. Fig. 4a shows the performance of different generation models on classifiers (VGG-16, VGG-13, VGG-11, CNN) in terms of accuracy, F1-score, and G-mean on the test set.

Our method outperforms others in most conditions. For example, under BR 1:400, DSEA-SMOTE achieves 74.64% accuracy on VGG-16, surpassing LSGAN (73.34%) by 1.3%, ACGAN (46.60%) by 28.04%, and others by up to 34.66%. DSEA-SMOTE shows better stability and is less sensitive to classifier depth and type. Fig. 5 illustrates model stability with shaded curves, where larger areas indicate higher sensitivity and poorer generalization.

To gain deeper insights, t-SNE and confusion matrices are used to visualize classifier performance. Fig. 6a shows t-SNE plots and confusion matrices for VGG-16 at BR 1:400. These results demonstrate clear classification boundaries with DSEA-SMOTE, capturing long-distance spatiotemporal dependencies and feature similarities. In contrast, other models show confusion and misclassifications, as seen in the confusion matrix.

TABLE I: CWRU dataset:Training and Testing Sample Size Settings under Different Balance Ratios

Data Type	Class Type	Construction of training and test sets for different categories under different BRs						
		1:400 (train/test)	1:200 (train/test)	1:100 (train/test)	1:50 (train/test)	1:25 (train/test)	1:10 (train/test)	1:5 (train/test)
Majority Class	0	2000/500	2000/50	2000/500	2000/500	2000/500	2000/500	2000/500
Minority Class	1-9(per class)	5/500	10/500	20/500	40/500	80/500	200/500	400/500
Minority class Balance Difference		1995	1990	1980	1960	1920	1800	1600

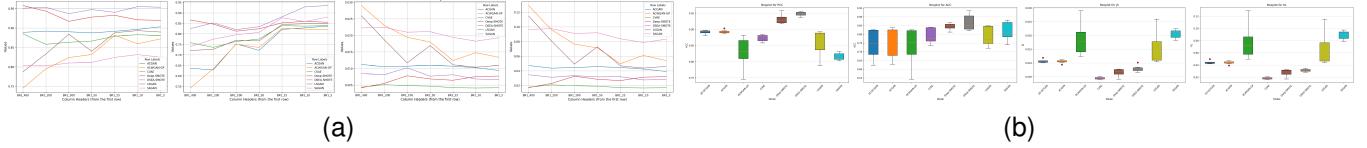
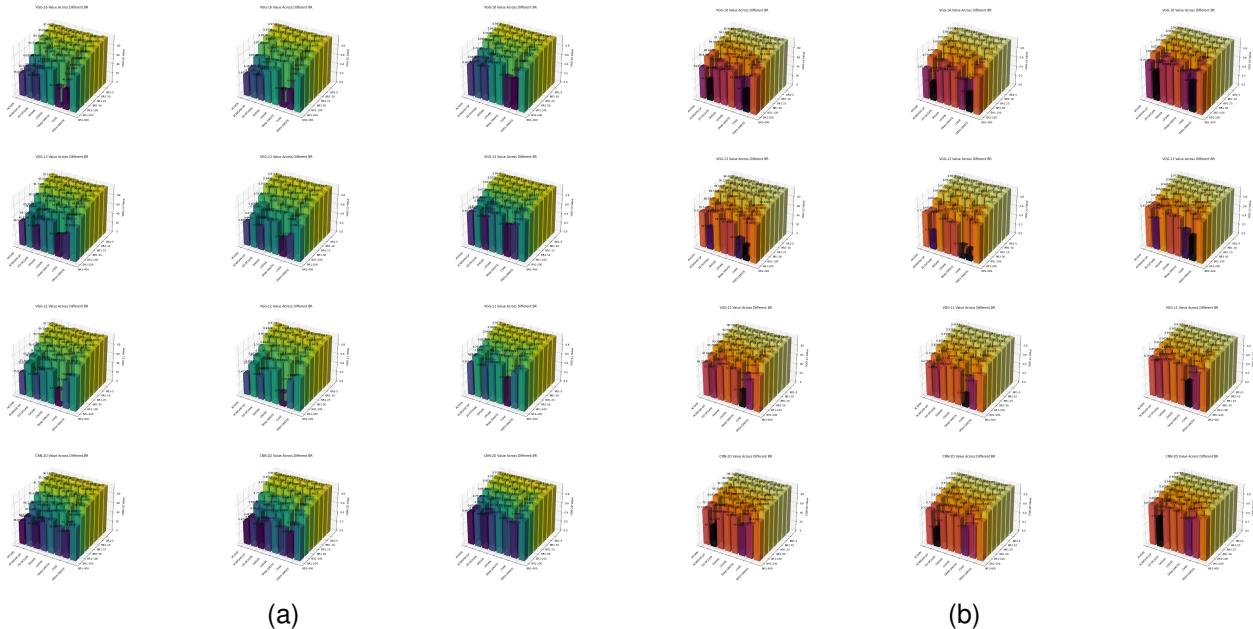
Fig. 3: Generation Performance and Discreteness of Different Models on CWRU Dataset under Various BRs(from left to right: PCC,  $G_{acc}$ , KL, JS). (a)line graphs. (b)box plots.

Fig. 4: Performance evaluation with accuracy, F1-score, and G-mean (columns 1-3), and classifiers VGG-16, VGG-13, VGG-11, CNN (rows 1-4). (a) CWRU dataset. (b) SEU dataset.

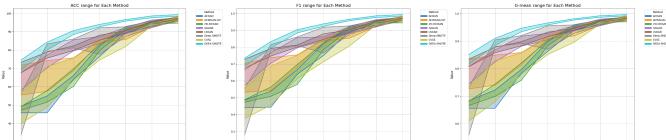


Fig. 5: CWRU dataset:The range of indicator changes under different BRs.

#### D. Compared with some recent novel methods

We compare our method with recent models (BCTGAN (2024) [18], RMA-WCGAN (2023) [16], REF-DDPM (2024) [27], DDPM (2020)) using the CWRU dataset and a CNN classifier. Fig. 7 shows that our model outperforms some of the latest new models.

## VI. EXPERIMENTAL VERIFICATION AND ANALYSIS ON THE SEU GEARBOX DATASET

### A. Dataset description and processing

To further evaluate the generalization of our model on other datasets, we tested it on the SEU gearbox dataset, provided by Southeast University. The dataset, available on their website and GitHub project [28], consists of two sub-datasets: gearbox and bearing faults. We used the bearing fault data with a 2 horsepower load at 30 kHz, which includes normal and four types of faults (labeled 0-4, the label 1 is compound fault type). Preprocessing follows a similar approach to the CWRU dataset, with 5 categories and the normal class as the majority. The training set contains 4000 slices per class, and the test set has 1000 slices per class. Testing on the SEU dataset allows us to evaluate our method's generalization performance and its ability to detect compound faults. The dataset construction

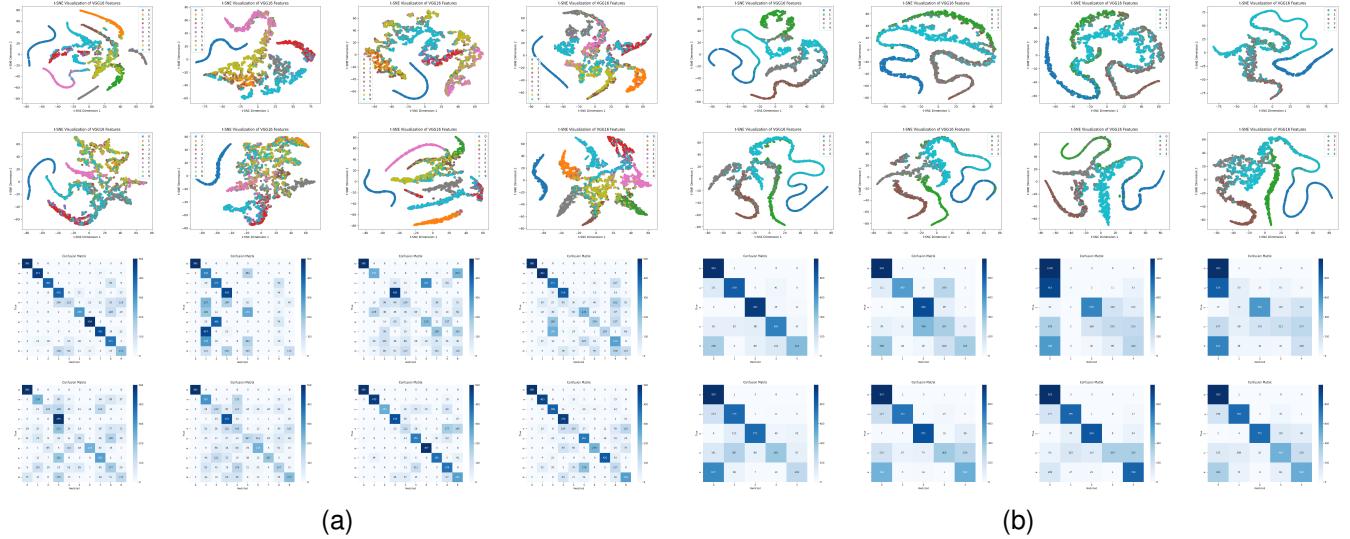


Fig. 6: t-SNE plots and Confusion matrix diagram of VGG-16 classifiers at BR 1:400 . Every two rows from top left to bottom right are DSEA-SMOTE, Deep-SMOTE, CVAE, ACWGAN-GP, ACGAN, 2D-DCGAN, LSGAN and SAGAN. (a) CWRU dataset. (b) SEU dataset.

TABLE II: SEU dataset: Training and Testing Sample Size Settings under Different BRs

Data Type	Class Type	Construction of training and test sets for different categories under different Balance Ratios						
		1:400 (train/test)	1:200 (train/test)	1:100 (train/test)	1:50 (train/test)	1:25 (train/test)	1:10 (train/test)	1:5 (train/test)
Majority Class	0	4000/1000	4000/1000	4000/1000	4000/1000	4000/1000	4000/1000	4000/1000
Minority Class	1-4(per class)	10/1000	20/1000	40/1000	80/1000	160/1000	400/1000	800/1000
Minority Class Balance Difference		3990	3980	3960	3920	3840	3600	3200

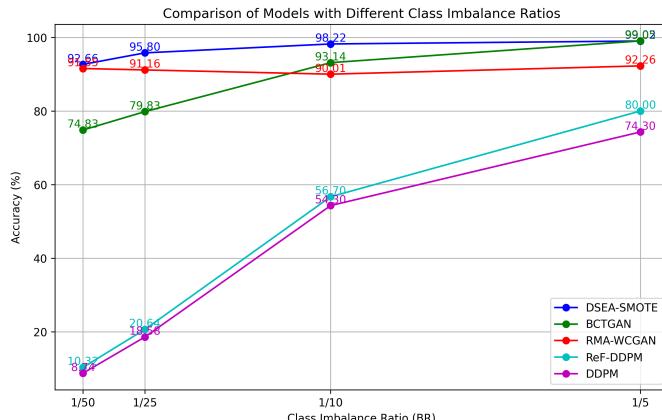


Fig. 7: CWRU dataset:Compared with some recent novel methods.

details are shown in Table II.

### B. Data generation and evaluation

After processing, the two-dimensional feature map samples are sent to the model's encoder to generate synthetic samples. Unlike the CWRU dataset, the SEU dataset has 5 categories, so the generated samples also reflect this. The results are evaluated and presented in Fig. 8a and Fig. 8b.

As shown in the figures, our method performs well and stably on the SEU dataset, achieving the best PCC and  $G_{acc}$ .

in most cases. For instance, under BR 1:400, DSEA-SMOTE (0.9604) outperforms LSGAN (0.8519), ACGAN (0.8662), SAGAN (0.7949), DCGAN (0.8813), ACWGAN-GP (0.7660), CVAE (0.8980), and Deep-SMOTE (0.9477) by significant margins. While CVAE excels in JS and KL, it tends to overfit, as observed in later experiments. Our model exhibits good stability and generalization across indicators, as shown by concentrated results in the box plot.

### C. Performance of downstream fault diagnosis classifiers

We further evaluate the downstream classifier. The evaluation results are shown in Fig. 4b.

As shown in Fig. 4b, our model outperforms others on the SEU dataset across most conditions. For instance, under BR 1:400, DSEA-SMOTE achieves 78.14% accuracy on the VGG-16 classifier, surpassing LSGAN (75.52%) by 2.62%, ACGAN (64.00%) by 14.14%, and others. Our model excels in performance stability, remaining robust across different classifiers. In contrast, CVAE and Deep-SMOTE show significant overfitting and instability. The performance range curve in Fig. 9 further illustrates our model’s stability. Fig. 9 shows that our method consistently outperforms most models, with a narrow performance range that often exceeds the upper limits of other models. As scale effects take place, all models converge to a similar range with minimal differences.

Next, we present the confusion matrix and t-SNE plots for the VGG-16 classifiers at BR 1:400 in Fig. 6b to further

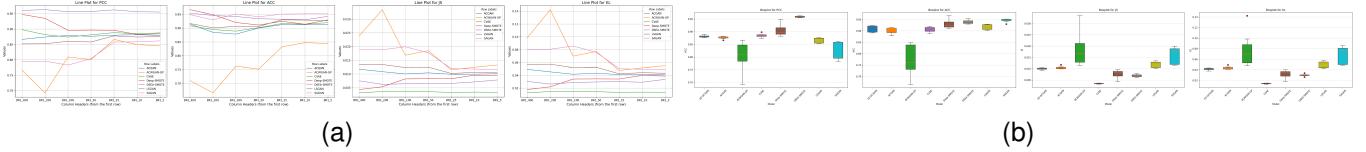


Fig. 8: Generation Performance and Discreteness of Different Models on SEU Dataset under Various BRs(from left to right: PCC,  $G_{acc}$ , KL, JS). (a)line graphs. (b)box plots.

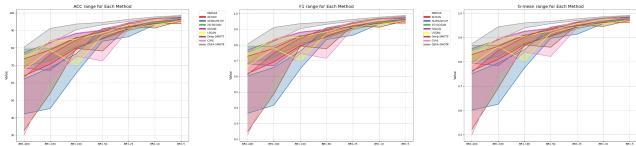


Fig. 9: SEU dataset: The range of indicator changes under different BRs.

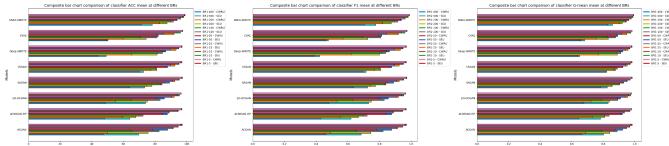


Fig. 10: Comparison of composite bar charts of the mean values of classifiers of various models at different BRs on the CWRU and SEU datasets.

analyze the specific performance. The t-SNE and confusion matrix results highlight our model's strong performance on the SEU dataset, maintaining clear classification boundaries even under extreme class imbalance. It excels in classifying compound faults (category 1) with minimal misclassifications, while other models show significant confusion. Our model also performs stably across different classifiers, unlike others that are highly sensitive.

#### D. A brief comparison between CWRU and SEU

After completing experiments on CWRU and SEU, we compare the performance on the CWRU and SEU datasets by averaging the indicators of four classifiers for each model across the same dataset and BR. The results are presented in a composite bar chart in Fig. 10. Fig. 10 shows that our model maintains excellent stability across different datasets, with minimal performance variation. In contrast, some models, like ACGAN, ACWGAN-GP, and 2D-DCGAN, show poor stability, with performance differences of up to 50% in extreme imbalance (e.g., BR 1:400). Our model consistently performs well, achieving results at BR1:400 comparable to those of unstable models at BR 1:200 or BR 1:100 (e.g., CVAE).

## VII. ABLATION EXPERIMENT

### A. Purpose of ablation experiment

Experiments on CWRU and SEU show that our model outperforms conventional and recent methods in most fault diagnosis imbalance tasks, demonstrating exceptional performance and generalization. This section further validates the effectiveness of the key components in our model.

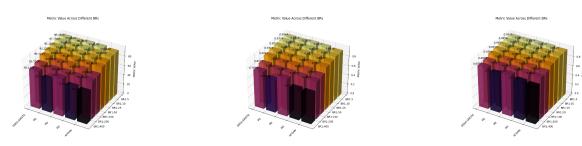


Fig. 11: Ablation experiment:The performance of each model on the VGG-16 classifier. Each row from left to right is measured by accuracy, F1-score, and G-mean indicators.

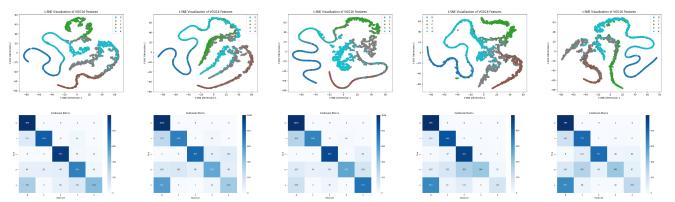


Fig. 12: Ablation experiment:t-SNE and confusion matrix plots of each model on the VGG-16 classifier at BR 1:400. From left to right, each row is DSEA-SMOTE, M1, M2, M3, ACGAN.

### B. Experimental design and results analysis

This section ablates the SEA and AFC modules integral to the design. The experiment setup includes DSEA-SMOTE, M1 (SEA ablation), M2 (AFC ablation), M3 (both SEA and AFC ablated), and ACGAN as the baseline. All experiments will use the SEU dataset with VGG-16 as the downstream classifier to evaluate diagnostic performance. This ablation verifies the effectiveness of the DSEA-SMOTE design.

As shown in Fig. 11, Fig. 12, and Table III, DSEA-SMOTE outperforms all models in the ablation experiment. For example, under BR 1:100, DSEA-SMOTE achieves 90.42% accuracy on the VGG-16 classifier, surpassing M1 (85.96%), M2 (87.54%), M3 (84.26%), and ACGAN (80.00%) by significant margins. DSEA-SMOTE performs best, followed by M2 (AFC removed) and M1 (SEA removed). M3, with both modules removed, performs worst but still outperforms ACGAN. These results emphasize the SEA module's role in capturing long-distance dependencies and space-time features, while the AFC module follows closely. Even without both modules, DSEA-SMOTE remains superior, showcasing the robustness of our model framework.

## VIII. CONCLUSION

In this paper, we propose DSEA-SMOTE, a novel framework for fault data synthesis to address class imbalance in fault diagnosis. The framework enhances synthetic data

TABLE III: Ablation experiment:Comprehensive performance evaluation table of each ablation model under different BR on VGG-16

Model	BR1:400 (ACC/F1/G-mean/Rank)	BR1:200 (ACC/F1/G-mean/Rank)	BR1:100 (ACC/F1/G-mean/Rank)	BR1:50 (ACC/F1/G-mean/Rank)	BR1:25 (ACC/F1/G-mean/Rank)	BR1:10 (ACC/F1/G-mean/Rank)	BR1:5 (ACC/F1/G-mean/Rank)
DSEA-SMOTE	78.14%/0.7699/0.8595/1	82.70%/0.8229/0.8895/1	90.42%/0.9041/0.9394/1	93.22%/0.9322/0.9573/1	95.18%/0.9517/0.9697/1	97.74%/0.9774/0.9858/1	98.48%/0.9848/0.9905/1
M1	67.44%/0.6699/0.7885/3	79.94%/0.7962/0.8714/3	85.96%/0.8586/0.9107/3	90.26%/0.9027/0.9384/3	93.58%/0.9355/0.9596/3	95.88%/0.9589/0.9741/3	97.50%/0.9749/0.9843/3
M2	77.83%/0.7704/0.8544/2	80.78%/0.7993/0.8769/2	87.54%/0.8714/0.9209/2	91.28%/0.9130/0.9449/2	94.44%/0.9443/0.9650/2	96.64%/0.9663/0.9789/2	98.38%/0.9838/0.9899/2
M3	67.06%/0.6251/0.7845/4	79.42%/0.7918/0.8680/4	84.26%/0.8414/0.8997/4	88.44%/0.8846/0.9267/5	93.36%/0.9322/0.9582/4	94.42%/0.9438/0.9649/4	97.42%/0.9742/0.9838/4
ACGAN	64.00%/0.6175/0.7632/5	78.82%/0.7784/0.8640/5	80.00%/0.7955/0.8718/5	89.40%/0.8927/0.9329/4	91.12%/0.9109/0.9439/5	94.14%/0.9413/0.9631/5	93.96%/0.9398/0.9620/5

quality by incorporating the SEA attention mechanism and Auxiliary Feature Classifier modules, which improve long-distance dependencies and feature similarity in space-time slices. Additionally, CWT data filtering preprocessing module, multi-category sample feature filtering module, and an improved loss function further boost data quality and simplify model learning.

We evaluate DSEA-SMOTE through experiments on two real-world datasets and ablation studies, drawing the following conclusions: (1) Our model shows excellent and stable performance across various metrics. (2) Even under extreme imbalance ratios (e.g., BR 1:400), synthetic data quality remains high, with accuracy exceeding 68% across all classifiers, and over 98% for BR 1:5, outperforming similar models. (3) Visualization results (t-SNE and confusion matrix) show clear classification boundaries, accurately identifying both single-class and compound faults, even under extreme imbalance. Future work will extend DSEA-SMOTE to medical image synthesis for rare diseases, remote sensing image synthesis for extreme natural disasters, radar signal synthesis for aviation objects, and explore its use with other modalities.

## REFERENCES

- [1] Z. Ren, Y. Zhu, Z. Liu, and K. Feng, "Few-shot gan: Improving the performance of intelligent fault diagnosis in severe data imbalance," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023.
- [2] X. Chen, R. Yang, Y. Xue, M. Huang, R. Ferrero, and Z. Wang, "Deep transfer learning for bearing fault diagnosis: A systematic review since 2016," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–21, 2023.
- [3] S. Gawde, S. Patil, S. Kumar, P. Kamat, K. Koteka, and A. Abraham, "Multi-fault diagnosis of industrial rotating machines using data-driven approach : A review of two decades of research," *Eng. Appl. Artif. Intell.*, vol. 123, p. 106139, 2023.
- [4] Z. Ren, T. Lin, K. Feng, Y. Zhu, Z. Liu, and K. Yan, "A systematic review on imbalanced learning methods in intelligent fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–35, 2023.
- [5] J. Wang, J. Wei, H. Huang, L. Wen, Y. Yuan, H. Chen, R. Wu, and J. Wu, "Imwmote: A novel oversampling technique for fault diagnosis in heterogeneous imbalanced data," *Expert Syst. Appl.*, vol. 251, p. 123987, 2024.
- [6] X. Gu, Y. Zhao, G. Yang, and L. Li, "An imbalance modified convolutional neural network with incremental learning for chemical fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 3630–3639, 2022.
- [7] D. Liu, S. Zhong, L. Lin, M. Zhao, X. Fu, and X. Liu, "Deep attention smote: Data augmentation with a learnable interpolation factor for imbalanced anomaly detection of gas turbines," *Comput. Ind.*, vol. 151, p. 103972, 2023.
- [8] D. Dablain, B. Krawczyk, and N. V. Chawla, "Deepsmote: Fusing deep learning and smote for imbalanced data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 6390–6404, 2023.
- [9] C. Shen, H. Zhang, S. Meng, and C. Li, "Augmented data driven self-attention deep learning method for imbalanced fault diagnosis of the hvac chiller," *Eng. Appl. Artif. Intell.*, vol. 117, p. 105540, 2023.
- [10] D. Liu, S. Zhong, L. Lin, M. Zhao, X. Fu, and X. Liu, "Feature-level smote: Augmenting fault samples in learnable feature space for imbalanced fault diagnosis of gas turbines," *Expert Syst. Appl.*, vol. 238, p. 122023, 2024.
- [11] X. Wang, H. Jiang, Y. Liu, S. Liu, and Q. Yang, "A dynamic spectrum loss generative adversarial network for intelligent fault diagnosis with imbalanced data," *Eng. Appl. Artif. Intell.*, vol. 126, p. 106872, 2023.
- [12] M. Zareapoor, P. Shamsolmoali, and J. Yang, "Oversampling adversarial network for class-imbalanced fault diagnosis," *Mech. Syst. Signal Process.*, vol. 149, p. 107175, 2021.
- [13] L. Tan, T. Huang, J. Liu, Q. Li, and X. Wu, "Deep adversarial learning system for fault diagnosis in fused deposition modeling with imbalanced data," *Comput. Ind. Eng.*, vol. 176, p. 108887, 2023.
- [14] P. Lyu, Y. Cheng, M. Zhang, W. Yu, L. Xia, and C. Liu, "Gpsc-gan: A data enhanced model for intelligent fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–16, 2024.
- [15] X. Huang, X. Zhang, Y. Xiong, B. Fan, and F. Dai, "Highly imbalanced fault diagnosis of turbine blade cracks via deep focal dynamically weighted conditional variational auto-encoder network," *Adv. Eng. Inform.*, vol. 62, p. 102612, 2024.
- [16] J. Huo, C. Qi, C. Li, and N. Wang, "Data augmentation fault diagnosis method based on residual mixed self-attention for rolling bearings under imbalanced samples," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023.
- [17] Y. Liu, H. Jiang, R. Yao, and H. Zhu, "Interpretable data-augmented adversarial variational autoencoder with sequential attention for imbalanced fault diagnosis," *J. Manuf. Syst.*, vol. 71, pp. 342–359, 2023.
- [18] H. Chen, J. Wei, H. Huang, L. Wen, Y. Yuan, and J. Wu, "Novel imbalanced fault diagnosis method based on generative adversarial networks with balancing serial cnn and transformer (bctgan)," *Expert Syst. Appl.*, vol. 258, p. 125171, 2024.
- [19] R. LI, S. LI, K. XU, M. ZENG, X. LI, J. GU, and Y. CHEN, "Auxiliary generative mutual adversarial networks for class-imbalanced fault diagnosis under small samples," *Chin. J. Aeronaut.*, vol. 36, no. 9, pp. 464–478, 2023.
- [20] S. Liu, H. Jiang, Z. Wu, and X. Li, "Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis," *Mech. Syst. Signal Process.*, vol. 163, p. 108139, 2022.
- [21] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zi, J. Yuan, B. Chen, and Z. He, "Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 70–71, pp. 1–35, 2016.
- [22] Q. Wan, Z. Huang, J. Lu, G. Yu, and L. Zhang, "Seaformer++: Squeeze-enhanced axial transformer for mobile visual recognition," *arXiv preprint arXiv:2301.13156*, 2023.
- [23] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2794–2802.
- [24] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Int. Conf. Mach. Learn.* PMLR, 2019, pp. 7354–7363.
- [25] Z. Li, T. Zheng, Y. Wang, Z. Cao, Z. Guo, and H. Fu, "A novel method for imbalanced fault diagnosis of rotating machinery based on generative adversarial networks," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–17, 2021.
- [26] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64–65, pp. 100–131, 2015.
- [27] T. Yu, C. Li, J. Huang, X. Xiao, X. Zhang, Y. Li, and B. Fu, "Ref-ddpm: A novel ddpm-based data augmentation method for imbalanced rolling bearing fault diagnosis," *Reliab. Eng. Syst. Saf.*, vol. 251, p. 110343, 2024.
- [28] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, 2019.