Digital twin-assisted imbalanced fault diagnosis framework using subdomain adaptive mechanism and margin-aware regularization

Shen Yan[1], Xiang Zhong[1], Haidong Shao[1*], Yuhang Ming[1], Chao Liu[2], Bin Liu[3]

1 College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China

2 College of Engineering and Physical Sciences, Aston University, Birmingham B47ET, UK

3 Department of Management Science, University of Strathclyde, Glasgow G1 1XQ, UK

Corresponding author: Haidong Shao (hdshao@hnu.edu.cn)

Abstract: The current data-level and algorithm-level based imbalanced fault diagnosis methods have respective limitations such as uneven data generation quality and excessive reliance on minority class information. In response to these limitations, this study proposes a novel digital twin-assisted framework for imbalanced fault diagnosis. The framework begins by analyzing the nonlinear kinetic characteristics of the gearbox and establishing a dynamic simulation model assisted by digital twin technology to generate high-fidelity simulated fault data. Subsequently, a subdomain adaptive mechanism is employed to align the conditional distribution of the subdomains by minimizing the dissimilarity of fine-grained features between the simulated and real-world fault data. To improve the fault tolerance of the model's diagnosis, margin-aware regularization is designed by applying significant regularization penalties to the fault data margins. Experimental results from two gearboxes demonstrate that, compared to the recent data-level and algorithm-level based imbalanced fault diagnosis methods, the proposed framework holds distinct advantages under the influence of highly imbalanced data, offering a fresh perspective for addressing this challenging scenario. In addition, the effectiveness of subdomain adaptive mechanism and margin-aware regularization is verified through the ablation experiment.

Keywords: Digital twin; Imbalanced fault diagnosis; Subdomain adaptive mechanism; Margin-aware regularization; Gearbox.

1. Introduction

Research on the health monitoring of mechanical equipment has effectively enhanced production reliability and reduced maintenance costs [1-4]. The gearbox, an essential mechanical transmission system widely used in various mechanical devices, inevitably generates various fault modes due to its long-term operation under complex conditions [5, 6]. Therefore, studying advanced fault diagnosis techniques for gearboxes is crucial for efficient and safe operation of equipment.

Data-driven methods based on deep learning have received substantial attention in the field of fault diagnosis in recent years due to their powerful feature extraction capabilities [7-10]. However, the general requirements of deep learning methods for balanced data distribution limit their deployment and application in industrial scenarios [11]. Gearboxes usually operate under the normal condition, with the low frequency and short duration of faults. Therefore, the volume of fault data collected by sensors is far less than that of normal data, leading to the data imbalance. When diagnostic models are trained by imbalanced data, the overall loss minimization forces the network to recognize the majority class's normal state while ignoring the minority class's fault state, as shown in **Fig. 1**. In recent years, a

series of imbalanced fault diagnosis methods have emerged to improve the recognition accuracy of minority class fault states [12], which can be broadly divided into two categories: data-level based methods and algorithm-level based methods.

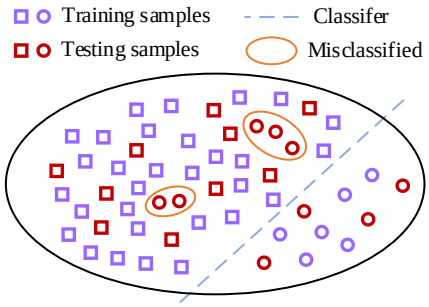


Fig.1. The challenge in imbalance fault diagnosis.

Data-level methods primarily encompass resampling strategies and data generating strategies, aimed at reducing the data imbalance rate to achieve a balanced distribution across different categories. Resampling strategies usually oversample minority and undersample majority. For instance, classic resampling strategies such as synthetic minority oversampling technique (SMOTE) [13] and SMOTE-Tomek (SMOTE-T) [14] have been researched in fault diagnosis. However, oversampling is prone to overfitting due to the recursive effect of samples, while undersampling may result in underfitting due to the loss of valid information from samples [15]. Data generating strategies based on generative adversarial network (GAN) inherit the powerful feature learning capabilities of deep learning. They adaptively generate new data through data generation and adversarial learning mechanisms. Liu *et al.* [16] introduced the variational autoencoder GAN (VAE-GAN), incorporating the encoder into the GAN to enhance the quality of generated rolling bearing data. Li *et al.* [17] designed a revised auxiliary classifier GAN (ACGAN) framework that can utilize random noise and label information to generate multi-modal fault samples. Yu *et al.* [18] presented a modified gradient penalty Wasserstein GAN (WGAN-GP) model, leveraging healthy samples to generate fault samples, thereby reducing the data imbalance rate. However, the majority of existing data generation strategies based on GANs still suffer the following limitations: 1) The unsmooth training process makes the model easy to pattern collapse, and the prolonged iterative process results in high training difficulty [19]. 2) The limited fault information in minority classes would cause GANs to generate interfering features, leading to a certain discrepancy between the generated data and experimental data [20].

Algorithm-level methods are mainly dominated by cost-sensitive strategies, which try to balance the feature learning effect among classes by enhancing the penalty cost for minority classes in the loss function. For example, focal loss (FL) [21] reduces the weight of easy-to-classify samples, enabling the model to focus on hard-to-classify minority class samples during training. Class-balanced loss (CBL) [22] reweights the loss for each class according to the number of valid samples. Dynamically weighted balanced loss (DWBL) [23] adapts class weights based on prediction scores. In the field of fault diagnosis, Jia *et al.* [24] incorporated weighted loss (WL) into convolutional neural networks to address the challenges of imbalanced data. Duan *et al.* [25] introduced FL for imbalanced fault diagnosis tasks. He *et al.* [26] employed dynamic penalty factors to tackle class imbalance issues. Yu *et al.* [27] designed an adaptive loss function that assigns different costs based on class quantity

differences, thereby highlighting the minority class. Hou *et al.* [28] implemented an imbalance strategy by weighting the similar loss on the number of valid samples to better learn the fault feature. However, algorithm-level imbalance fault diagnosis methods generally suffer the following limitations: 1) When data is severely imbalanced, models tend to overfit minority class samples, rendering the classification boundary unreliable [29]. 2) The model performance heavily relies on the feature information embedded in the minority class samples, limiting the generalizability of the algorithm [30].

To deal with the limitations of the aforementioned data-level and algorithm-level methods, this paper proposes a new digital twin-assisted framework for imbalanced fault diagnosis. Specifically, we first analyze the nonlinear kinetic characteristics of the gearbox and establish its dynamic simulation model with the aid of digital twin technology, to obtain high-fidelity simulated fault data. Subsequently, a subdomain adaptive mechanism (SAM) is employed, which aligns the conditional distribution of the subdomains by minimizing the dissimilarity of fine-grained features between the simulated and real-world fault data. Finally, a margin-aware regularization (MAR) is designed, which imposes significant regularization penalties on the fault data margin to enhance the fault tolerance of the model's diagnosis. Experimental results from two gearboxes show that compared to the recent data-level and algorithm-level imbalanced fault diagnosis methods, the proposed framework demonstrates clear advantages under the setting of highly imbalanced data, offering a new perspective addressing this challenging scenario. Additionally, the effectiveness of SAM and MAR is tested by the ablation experiment.

The structure of this paper is organized as follows: Section 2 introduces the related works of digital twin and imbalanced transfer learning in fault diagnosis. Section 3 explains the details of the proposed method. Section 4 describes the digital twin-assisted imbalanced fault diagnosis framework. Section 5 provides the experimental validation. Section 6 concludes the paper.

2. Related works

2.1. Digital twin-aided fault diagnosis

Digital twin technology, which integrates real-world systems with their digital representations to create dynamic simulation models, has gradually become a crucial technology in industry [31]. It provides new avenues for equipment monitoring processes from the perspective of physical information coupling. Specifically, a kinetic model that maps the mechanical system parametric characteristics is first constructed. Then, the operating parameters of the corresponding components are updated dynamically based on the real system data, obtaining simulation data under different system operating conditions. Finally, data-driven methods are combined to analyze the real system's health status, thereby implementing the monitoring and control process for the equipment.

In 2023, Wang *et al.* [32] proposed a digital twin-aided adversarial transfer learning method for fault diagnosis in three-cylinder pumps. In the same year, Zhang *et al.* [33] established a virtual representation model for bearings to generate simulation data and employed Transformer to learn feature information. Wang *et al.* [34] also leveraged digital twin technology to create a bearing simulation model and machine learning methods to predict the likelihood of bearing failure. In 2023, Xiang *et al.* [35] constructed a rotor dynamic model with crack faults and used domain-adaptive networks to learn transferable features of rotor cracks. These recent fault diagnosis studies have

successfully integrated digital twin technology with data-driven methods, effectively utilizing the abundant fault information in simulation data. However, they all assume that the distribution of real-world training data categories is balanced and do not consider the impact of imbalanced data on the fault diagnosis performance of the models.

2.2. Imbalanced transfer learning-based fault diagnosis

Transfer learning is generally applicable to fault diagnosis under multiple operating conditions or multiple devices. It narrows the feature distribution differences between source domain data and target domain data, thereby utilizing the empirical knowledge of the source domain to improve feature learning in the target domain. Among these, the mapping-based metric method, a classic transfer learning strategy, has received extensive research attention [36]. It computes the inter-domain statistical moment as regularization for the loss function to learn transferable features, with the core idea expressed as follows:

$$\mathcal{L}_{\mathcal{H}}(P,Q) = \left\| E_P[\psi(x^s)] - E_Q[\psi(x^t)] \right\|_{\mathcal{H}}^2 \tag{1}$$

where x^s and x^t are the samples from the source and target domains, respectively, which follow the probability distributions P and Q. E [·] represents the mathematical expectation operation; $\mathcal{H}$ is the feature mapping space, often utilizing the reproducing kernel Hilbert space (RKHS); $\psi(\cdot)$ denotes the data feature mapping process. However, the majority of transfer learning-based fault diagnosis methods are rooted in the ideal assumption of implicitly class-balanced data.

In the recent two years, imbalanced transfer learning has gained some attention in fault diagnosis, primarily for addressing scenarios with imbalanced label distribution in the target domain, as shown in **Fig. 2**. In 2022, Kuang *et al.* [37] constructed a class-imbalanced adversarial transfer learning network to learn domain-invariant features. In the same year, Liu *et al.* [38] designed a meta-data-based transfer residual network and a weight allocation strategy for imbalanced cross-domain fault diagnosis. In 2022, Wu *et al.* [39] embedded a cost-sensitive strategy into the adversarial transfer learning model to achieve imbalanced fault diagnosis in the cross-domain bearing scenario. In 2023, Ding *et al.* [40] proposed an imbalanced domain adaptation fault diagnosis framework to deal with the scenario where both feature distribution and label distribution differ under different operating conditions. These recent studies have reduced the model's requirements for label distribution, expanding the use scenarios of transfer learning. However, their source domain all come from real-world testbed data, leading to high data acquisition costs.

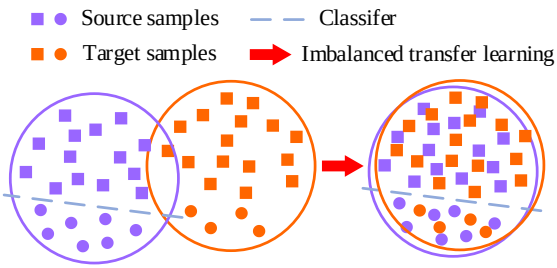


Fig.2. Imbalance transfer learning-based fault diagnosis.

3. The details of the proposed method

3.1. Digital twin-assisted gearbox simulation

To address challenges such as incomplete minority class fault information and high costs of obtaining real-world fault data, we analyze the nonlinear kinetic characteristics of the gearbox and leverage digital twin technology to establish its dynamic simulation model, obtaining simulation data containing rich fault feature information.

First, the lumped mass method is used to analyze the nonlinear kinetic characteristics of the planetary gearbox, and its kinetic model can be simplified as shown in **Fig. 3**. In this model, two rigid circles represent meshing gear pair, with the meshing point replaced by a spring damper set along the common tangent. The shaft is supported by a pair of flexible ball bearings that can be modeled as linear springs and viscous dampers [41]. When neglecting the gear mass eccentricity, the kinetic equations for this gear pair can be expressed as:

$$m_1\ddot{x}_1 + c_{x1}\dot{x}_1 + k_{x1}x_1 + F_m \sin\alpha = 0 \tag{2}$$

$$m_1\ddot{y}_1 + c_{y1}\dot{y}_1 + k_{y1}y_1 + F_m \cos\alpha = 0 \tag{3}$$

$$J_1\ddot{\theta}_1 + F_m R_1 = T_1 \tag{4}$$

$$m_2\ddot{x}_2 + c_{x2}\dot{x}_2 + k_{x2}x_2 - F_m \sin\alpha = 0 \tag{5}$$

$$m_2\ddot{y}_2 + c_{y2}\dot{y}_2 + k_{y2}y_2 - F_m \cos\alpha = 0 \tag{6}$$

$$J_2\ddot{\theta}_2 - F_m R_2 = -T_2 \tag{7}$$

where m_i represents the mass of gear i (i=1, 2); R_i is the base circle radius of gear i; $\ddot{\theta}_i$ is the angular acceleration of gear i; x_i and y_i are the transverse and longitudinal displacements of gear i, respectively; $\dot{x}_i$ and $\dot{y}_i$ are the corresponding velocities; $\ddot{x}_i$ and $\ddot{y}_i$ are the corresponding accelerations; c_{xi} and c_{yi} are the transverse and longitudinal bearing damping of gear i, respectively; k_{xi} and k_{yi} are the transverse and longitudinal bearing stiffness of gear i, respectively; T_i is the torque applied to gear i; J_i is the moment of inertia of gear i relative to the centroid; α is the pressure angle; F_m is the meshing force, which is related to the static transmission error $e(t)$, the initial backlash b_0, the stiffness k_m, and the damping c_m. The constraint conditions for the gearbox are the same as those in Ref. [42].

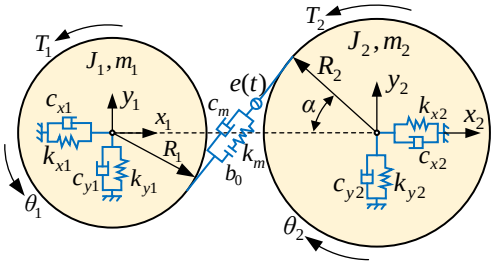


Fig.3. The kinetic model of the gear pair.

The kinetic equations are solved using the implicit integration method, and the relevant

parameters of the gear components are iteratively updated based on the gearbox's operational data to obtain a dynamic simulation model of the planetary gearbox, which can be simplified as **Fig.4 (a)**. **Table 1** lists the basic parameters of each gear. By adjusting the stiffness and damping of the sun gear shaft in the simulation model, the vibration response of the gearbox under different fault modes can be simulated, thus obtaining high-fidelity simulation fault data. **Fig.4 (b)** shows root crack and missing tooth gear faults, and in addition, gear broken tooth is also simulated. The centroid angular acceleration signals in the transverse and longitudinal directions of the planetary gearbox are collected as simulation fault data; the rotation frequency of the input shaft is set to 30 Hz, and the sampling frequency is 12.8 kHz.

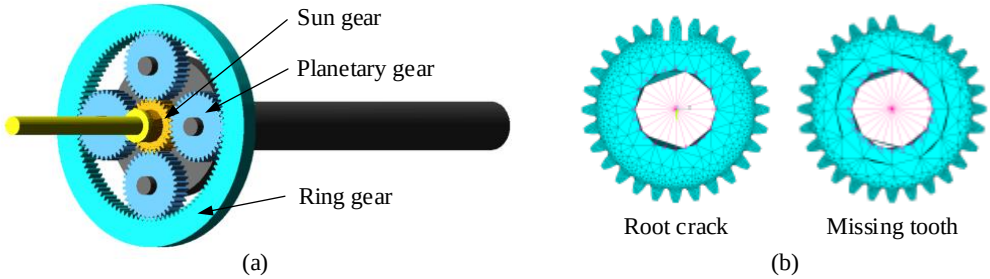


Fig.4. The dynamic simulation model: (a) Planetary gearbox; (b) Two types of faulty gears.

Table 1
Parameters of the gear part.

Part	Number of teeth	Modulus	Pressure angle	Tooth width (mm)
Sun gear	28	1	20°	10
Planetary gear	36	1	20°	10
Ring gear	100	1	20°	10

3.2. Subdomain adaptive mechanism

To reduce the distribution space differences between fault modes in the simulation domain and the real-world domain, we employ the SAM based on the local maximum mean discrepancy metric [43]. By minimizing the dissimilarity of fine-grained features between simulation and real-world fault data, we align the conditional distribution of fault data in the two domains, as shown in **Fig.5**.

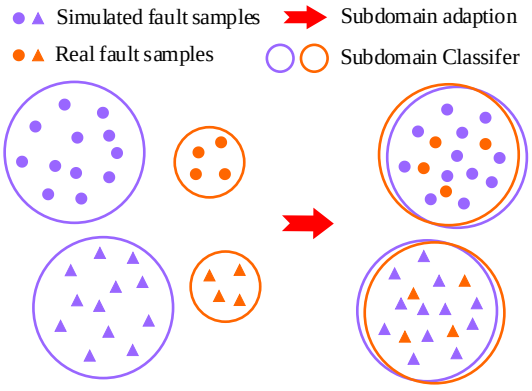


Fig.5. Subdomain adaptive mechanism.

Specifically, the local maximum mean discrepancy metric can be expressed as follows:

$$\mathcal{L}_{\mathcal{H}}(P_c, Q_c) = E_c \left\| E_{P_c}[\psi(x^s)] - E_{Q_c}[\psi(x^t)] \right\|^2_{\mathcal{H}} \tag{8}$$

where $\mathcal{H}$ is the RKHS; P_c and Q_c represent the probability distributions of fault category c in the simulated and real domains, respectively. Unlike Eq. (1), which measures global distribution differences, Eq. (8) is concerned with the expectation of the local distance for each fault mode. If each fault sample belongs to fault category c with weight w^c, the empirical estimate of $\mathcal{L}_{\mathcal{H}}(P^c, Q^c)$ can be expressed as:

$$\hat{\mathcal{L}}_{\mathcal{H}}(P_c, Q_c) = \frac{1}{C}\sum_{c=1}^{C}\left\| \sum_{i=1}^{n_s} w_i^{sc}\psi(x_i^s) - \sum_{j=1}^{n_t} w_j^{tc}\psi(x_j^t) \right\|_{\mathcal{H}}^2 \tag{9}$$

where n_s and n_t are the number of fault samples in the simulated and real domains, respectively; w_i^{sc} and w_j^{tc} represent the likelihood of samples x_i^s and x_j^t belonging to fault category c, satisfying $\sum_{i=1}^{n_s} w_i^{sc} = 1$ and $\sum_{j=1}^{n_t} w_j^{tc} = 1$. For each sample x_i in both the simulated and real domains, we have its one-hot vector label $y_i = \left[y_i^1, \cdots, y_i^C \right]$, and the weight w_i^c for each sample can be computed as follows:

$$w_i^c = \frac{y_i^c}{\sum_{i=1}^{n} y_i^c} \tag{10}$$

If the feature extractor is represented as $G_f(\cdot)$, the SAM employed can be expanded as follows:

$$\begin{aligned} \hat{\mathcal{L}}_{\mathcal{H}f}(P_c, Q_c) = \frac{1}{C}\sum_{c=1}^{C}\Bigg[& \sum_{i=1}^{n_s}\sum_{j=1}^{n_s} w_i^{sc} w_j^{sc} k\left(G_f(x_i^s), G_f(x_j^s)\right) \\ & + \sum_{i=1}^{n_t}\sum_{j=1}^{n_t} w_i^{tc} w_j^{tc} k\left(G_f(x_i^t), G_f(x_j^t)\right) \\ & - 2\sum_{i=1}^{n_s}\sum_{j=1}^{n_t} w_i^{sc} w_j^{tc} k\left(G_f(x_i^s), G_f(x_j^t)\right) \Bigg] \end{aligned} \tag{11}$$

where $k(\cdot)$ represents the kernel function used for data feature mapping. By minimizing Eq. (11), the conditional distribution alignment of each subdomain is achieved.

3.3. Margin-aware regularization

Although the simulation data enriches the fault feature information, the employed SAM projects it as much as possible into the limited distribution space of real-world fault data. It results in the decision boundary based on these feature distributions remaining relatively fragile. Therefore, we design the MAR based on the real-world label distribution, which applies significant regularization penalties to the fault data margins, in order to enhance the model's fault diagnosis robustness.

First, the margin of a sample x_i can be expressed as follows:

$$\gamma_{y_i} = G_c\left(G_f\left(x_i\right)\right)_{y_i} - \max_{c \neq y_i} G_c\left(G_f\left(x_i\right)\right)_c \tag{12}$$

where y_i is the label value of the sample x_i, and $G_c(\cdot)_{y_i}$ is the output value of the classifier for class y_i. Then, the margin of a class c can be expressed as follows:

$$\gamma_c = \min_{i \in n_c} \gamma_{y_i} \tag{13}$$

where n_c is the samples' number of the class c in real world. Relevant research shows that the generalization error is minimized when $\gamma_c \propto n_c^{-1/4}$ is satisfied [44]. In other words, the more samples a class has, the smaller distance between the class and the decision boundary should be, which can be simplified as shown in **Fig.6**. Therefore, the MAR based on the real-world label distribution can be expressed as follows:

$$\mathcal{L}(x, y) = -\sum_{i=1}^{n} \ln \frac{e^{z_{y_i} - \Delta_{y_i}}}{e^{z_{y_i} - \Delta_{y_i}} + \sum_{c \neq y_i} e^{z_c}} \tag{14}$$

$$\Delta_c = \frac{\max(n_c^{1/4})}{2n_c^{1/4}} \quad c \in \{1, \cdots, C\} \tag{15}$$

where n represents all samples in the real world; $z_{y_i} = G_c\left(G_f\left(x_i\right)\right)_{y_i}$ represents the output value of class y_i; $z_c == G_c\left(G_f\left(x_i\right)\right)_c$ represents the output value of class c; Δ_c is the regularization penalty coefficient for class c, which is proportional to $n_c^{-1/4}$. The MAR is essentially a cross-entropy loss function with an optimized decision boundary.

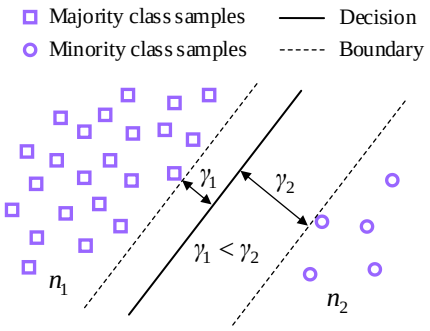


Fig.6. Decision boundary for imbalanced samples

4. Digital twin-assisted imbalanced fault diagnosis framework

By integrating the details of the proposed method, this paper proposes a novel digital twin-assisted framework, providing a fresh perspective for dealing with the imbalanced fault diagnosis scenario, as shown in **Fig.7**.

Data balancing: For real-world gearbox, sufficient normal data and little fault data are collected through sensors to simulate the imbalanced scenario. By utilizing digital twin technology, a dynamic simulation model of the gearbox is established to obtain sufficient simulated fault data to balance the training data.

Model training: First, the simulated fault data and real imbalanced data are input into a shared feature extractor $G_f(\cdot)$ in parallel, with the 1D-ResNet18 network [36] selected as $G_f(\cdot)$. Next, the data features and corresponding labels from the simulated and real domains are input into Eq. (11) in the SAM to obtain $\hat{\mathcal{L}}_{\mathcal{H}}(P_c, Q_c)$. Subsequently, the features of all data are input into the classifier $G_c\left(\cdot\right)$ to obtain predicted labels; we choose two linear layers as $G_c\left(\cdot\right)$, which are the same as those in Ref. [36]. Then $\mathcal{L}(x, y)$ can be obtained by Eq. (14) and Eq. (15) in the MAR. Therefore, the loss function

of the model can be expressed as:

$$\mathcal{L}oss(\theta_f,\theta_c)=\mathcal{L}(x,y)+\lambda_t\hat{\mathcal{L}}_{\mathcal{H}}(P_c,Q_c) \tag{16}$$

where $\lambda_t=2\big/(1+e^{-10t/t_m})-1$ is a time-varying weight coefficient; t and t_m are the current and maximum number of iterations, respectively. θ_f and θ_c are the learning parameters of $G_f(\cdot)$ and $G_c(\cdot)$. During the training iteration process, the loss $\mathcal{L}oss(\theta_f,\theta_c)$ is minimized using the Adam optimizer with weight decay regularization [45] to optimize the network learning parameters θ_f and θ_c. The model with the lowest training set loss value during the iteration process is chosen as the well-trained model.

Test analysis: The test set from the real-world gearbox is input into the trained model to obtain the diagnosis results, upon which multi-dimensional visualization analysis is conducted.

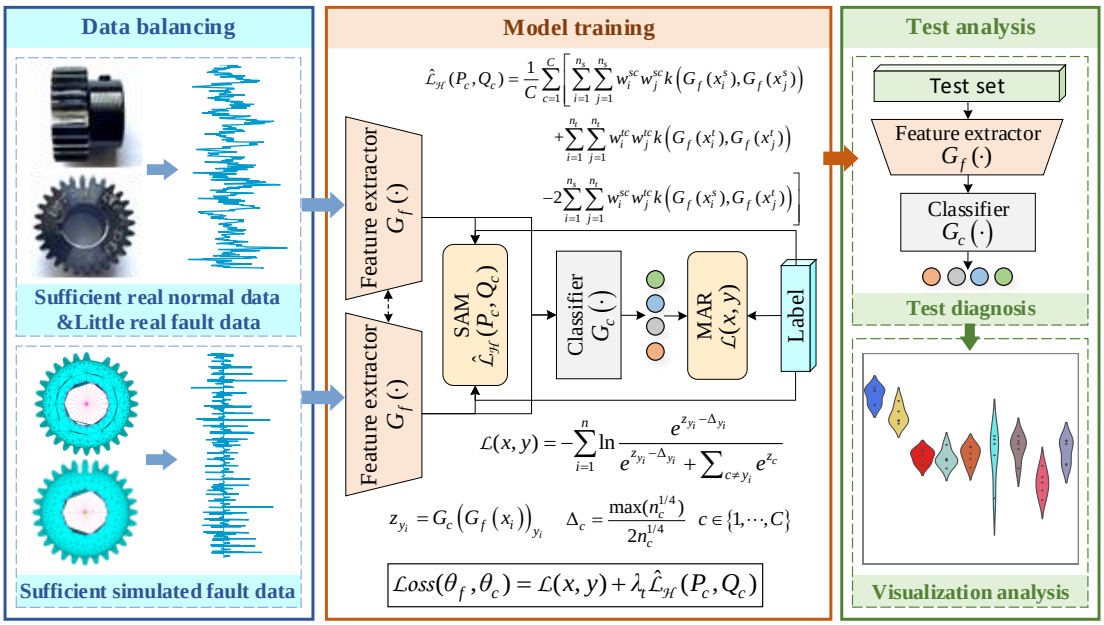


Fig.7. Digital twin-assisted imbalanced fault diagnosis framework.

5. Experiment verification

Section 5.1 and 5.2 demonstrate the advantages of the proposed framework in imbalanced datasets of two gearboxes, while Section 5.3 examines the effectiveness of SAM and MAR through the ablation experiment.

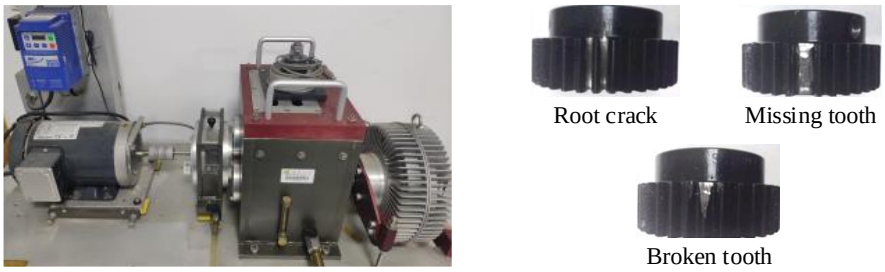
5.1. Case 1: Planetary gearbox from Southeast University

5.1.1. Data preparation of Case 1

The datasets of Case 1 are derived from the drivetrain diagnostic simulator (DDS) of Southeast University [46], which includes a motor, braking system, parallel gearbox, planetary gearbox, and other components, as shown in **Fig.8 (a)**. Four gearbox health states are selected for experimental verification, including normal state and three sun gear fault states shown in **Fig.8 (b)**. Torque and vibration acceleration signals of the motor are selected as the raw data, with the motor speed of 1800 rpm, workload of 2 V, and sampling frequency of 5 kHz.

Three datasets with different imbalanced ratios, namely S1, S2, and S3, are constructed, as listed

in **Table 2**. Taking dataset S1 as an example, the training set contains 488 normal state samples and 8 samples for each of the three fault states. For dataset S3, the imbalanced ratio of the dataset is as high as 253 (506/2) times. In the test process of the three datasets, there are 200 samples for each health state. Each sample consists of 1024 sampling points, and to avoid test leakage, there is no overlap of sampling points between samples. It should be noted that the training set of the proposed method also includes simulated fault samples, with 512 simulation samples for each of the three fault states in all three datasets.



(a) (b)

Fig.8. The experimental setup of Case 1: (a) The test bench; (b) Three types of faulty gears.

Table 2
Imbalanced datasets of Case 1.

Health states	Labels	Datasets (No. of training samples)			No. of testing samples
		S1	S2	S3	
Normal	0	488	500	506	200
Missing tooth	1	8	4	2	200
Root crack	2	8	4	2	200
Broken tooth	3	8	4	2	200

5.1.2. Result analysis of Case 1

A total of 8 methods are selected for comparison with the proposed method, including 4 data-level methods: ACGAN [17], VAE-GAN [16], WGAN-GP [18], and SMOTE-T [14], and 4 algorithm-level methods: CBL [22], DWBL [23], FL [21], and WL [24]. To reduce experimental errors, each group of experiments is repeated 5 times, and the feature extractor $G_f(\cdot)$ and classifier $G_c(\cdot)$ are the same for all methods. The main hyperparameters are set as follows: total 100 iterations, batch size of 128, the initial learning rate of 0.001 with adaptive decay mode. The selection rules of the well-trained models are also consistent with the proposed method.

To record the convergence of each method during the iteration process, the training loss and test accuracy variation of all methods are averaged over 5 repeated experiments in the S1 dataset, as shown in **Fig.9**. The legend "Proposed" represents the proposed method. From **Fig.9 (a)**, after training for 60 epochs, the training losses of all methods show convergence to low values. Similarly, for **Fig.9 (b)**, the test accuracies of all methods show stable high values after approximately 60 epochs of training. This indicates that the setting of the iteration count is reasonable. Upon further observation, the proposed method has the highest test accuracy and the fastest convergence. This preliminarily demonstrates the advantage of the proposed digital twin-assisted fault diagnosis framework.

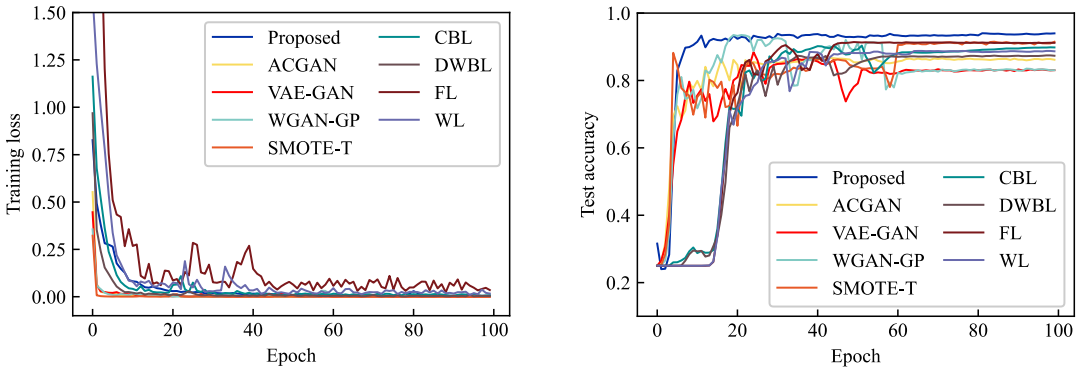


Fig.9. The average training loss and test accuracy of each method in the S1 dataset: (a) Training loss; (b) Test accuracy.

Table 3 lists the diagnostic accuracy of each method in Case 1. In the three datasets with different imbalanced ratios, the mean accuracies of the proposed method are the highest among all methods. Moreover, as the dataset imbalanced ratio increases, the diagnostic advantage of the proposed method over the comparison methods becomes more significant. In the S1 dataset with 61 times the imbalance ratio, although the minimum accuracy of the proposed method is 0.50% lower than that of the second-best WGAN-GP, the mean and the maximum accuracies are 2.18% and 2.62% higher, respectively. In the S2 dataset with 125 times the imbalance ratio, although the maximum accuracy of the proposed method is 0.24% lower than that of ACGAN, the mean and the minimum accuracies are 16.20% and 26.75% higher, respectively, and the stability is clearly better than ACGAN; compared with the second-best FL, the mean, maximum and minimum accuracies of the proposed method are 12.90%, 16.26%, and 7.88% higher, respectively. In the S3 dataset with 253 times the imbalance ratio, the proposed method has the mean, maximum and minimum accuracies of 17.30%, 7.12%, and 33.00% higher than the second-best ACGAN, respectively. As the imbalanced ratio increases, the diagnostic results of the 4 data-level methods show a distinct downward trend. This is because the extremely limited fault information of the minority class can easily lead to a large number of interfering features in the generated data.

Table 3
Diagnostic accuracy of each method in Case 1.

Method	S1-Accuracy (%)			S2-Accuracy (%)			S3-Accuracy (%)		
	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min
Proposed	**93.53**	**98.00**	86.25	**85.50**	94.88	**68.88**	**71.33**	**76.50**	**66.25**
ACGAN	86.97	91.38	80.00	69.30	**95.12**	42.13	54.03	69.38	33.25
VAE-GAN	81.85	97.88	68.12	58.02	74.12	38.00	37.98	47.12	26.00
WGAN-GP	91.35	95.38	**86.75**	58.65	79.50	48.50	35.30	45.25	25.87
SMOTE-T	83.67	97.00	56.12	46.67	78.25	28.25	27.07	32.75	25.00
CBL	89.67	94.38	83.12	66.67	87.38	53.25	49.98	64.50	42.00
DWBL	87.17	92.50	76.75	67.80	81.00	48.38	51.80	59.13	44.88
FL	91.15	94.25	**86.75**	72.60	78.62	61.00	46.75	60.25	38.88
WL	88.55	92.00	84.38	67.35	79.88	55.12	52.40	61.38	47.75

Remarks: S1-Accuracy (%) represents the accuracy in the S1 dataset. Mean is the mean accuracy of the 5 repeated experiments; Max is the maximum accuracy and Min is the minimum accuracy.

Fig.10 summarizes the diagnostic accuracies of all methods for 15 experiments in the three

datasets. The solid triangle represents the average accuracy, and the black horizontal line in the box represents the median accuracy. The average accuracy and median accuracy of the proposed method are significantly better than those of the comparison methods. Specifically, the average and median accuracies of the proposed method are more than 10% higher than those of the second-best ACGAN. Furthermore, the proposed method can maintain excellent stability during repeated experiments. This further validates the diagnostic advantage of the proposed digital twin-assisted framework under the influence of highly imbalanced data.

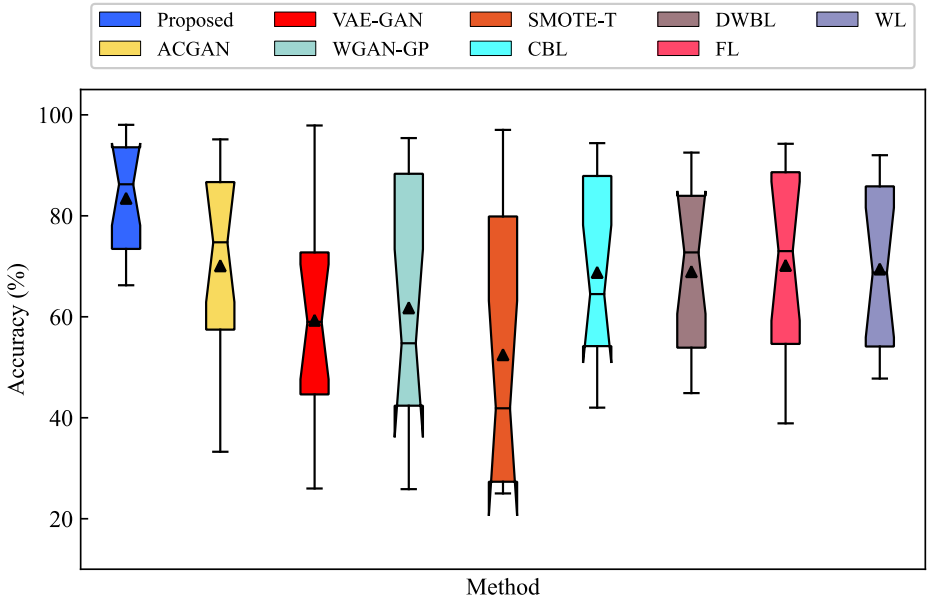


Fig.10. The all accuracies in S1, S2 and S3 datasets of each method.

In order to quantitatively compare the output feature quality of different methods, a feature quantitative evaluation index J is introduced, expressed as follows:

$$J = \frac{\|S_b\|}{\|S_w\|} \tag{17}$$

$$S_w = \sum_{c=1}^{C}\sum_{i=1}^{N_c}(f_i^c - \hat{f}^c)(f_i^c - \hat{f}^c)^T \tag{18}$$

$$S_b = \sum_{c=1}^{C} N_c(\hat{f}^c - \hat{f})(\hat{f}^c - \hat{f})^T \tag{19}$$

$$\hat{f}^c = \frac{1}{N_c}\sum_{i=1}^{N_c} f_i^c,\ \ \hat{f} = \frac{1}{N}\sum_{i=1}^{N} f_i \tag{20}$$

where S_w is the within-class covariance; S_b is the between-class covariance; $\|\cdot\|$ is the second-order norm; C is the number of classes; N is the total number of samples in the test set; f_i is the output feature of the ith sample; and $\hat{f}$ is the average output feature of all samples. **Table 4** lists the feature quantitative evaluation of each method in Case 1. In the three datasets, the mean J values of the proposed method are the highest among all methods, which is consistent with the results in **Table 3**. In the S1 dataset, the mean, maximum and minimum J values of the proposed method are approximately 78%, 100%, and 78% higher than those of the second-best WGAN-GP, respectively. In the S2 dataset,

although the maximum value of the proposed method is 5% lower than that of the second-best ACGAN, the mean and the minimum values are 29% and 60% higher, respectively, and the stability is superior to ACGAN. In the S3 dataset, although the minimum value of the proposed method is 19% lower than that of the second-best DWBL, the mean and the maximum values are 15% and 21% higher, respectively. Finally, the average *J* value of the proposed method on the three datasets is also superior to that of the comparison methods. Compared with the second-best DWBL, the average value of the proposed method is 64% higher. This indirectly indicates that the proposed framework can effectively distinguish the differences between class features in the highly imbalanced scenario.

Table 4
Feature quantitative evaluation of each method in Case 1.

Method	S1-*J*			S2-*J*			S3-*J*			Avg
	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	
Proposed	**6.32**	**7.91**	**4.99**	**4.39**	5.72	**2.84**	**3.22**	**4.43**	2.04	**4.63**
ACGAN	3.01	3.44	2.76	3.40	**6.04**	1.78	2.03	2.28	1.73	2.81
VAE-GAN	3.19	5.68	1.99	2.07	2.46	1.75	1.79	1.92	1.64	2.35
WGAN-GP	3.55	3.96	2.81	1.97	2.58	1.53	2.00	2.43	1.79	2.51
SMOTE-T	3.04	4.07	2.10	2.01	2.17	1.91	1.67	1.84	1.36	2.24
CBL	2.92	3.50	2.39	2.17	3.11	1.72	1.61	2.10	1.34	2.23
DWBL	2.98	3.34	2.35	2.69	3.13	1.83	2.79	3.67	**2.51**	2.82
FL	2.82	3.31	2.31	1.83	1.94	1.70	1.24	1.61	0.85	1.96
WL	2.69	3.41	2.14	1.80	2.28	1.40	1.44	2.06	1.12	1.98

Remarks: S1-J represents the feature quantitative evaluation index J on the S1 dataset, where a higher value indicates better output feature quality. Mean is the mean J value of the 5 repeated experiments; Max is the maximum J value and Min is the minimum J value. Avg denotes the average J value of the S1, S2 and S3 datasets.

To further compare the output feature quality of different methods for samples from different health states, we use t-distributed stochastic neighbor embedding (t-SNE) to reduce the output features of the test set to two dimensions for visualization, as shown in **Fig.11**. It should be noted that we select 6 methods with better mean *J* values in the S2 dataset for comparison and average the output features of 5 repeated experiments to reduce the error caused by a single experiment. Compared with other methods, the proposed method can accurately identify the feature information of samples with four different health states, effectively clustering samples with the same health state, and creating clear boundaries between different health states. The missing tooth and root crack classes in VAE-GAN show the overlapping boundary, while DWBL has difficulty defining the boundary between root crack and broken tooth. Although the comparison methods ACGAN, SMOTE-T, and CBL have relatively clear boundaries between different health states, ACGAN clusters some missing tooth and broken tooth samples into the root crack feature space; SMOTE-T clusters part of root crack samples into the missing tooth and broken tooth feature spaces; CBL clusters some missing tooth samples into the root crack and broken tooth feature spaces, all of which pose a risk of misdiagnosis. In addition, all methods have good clustering effects in the feature space of sufficient normal samples.

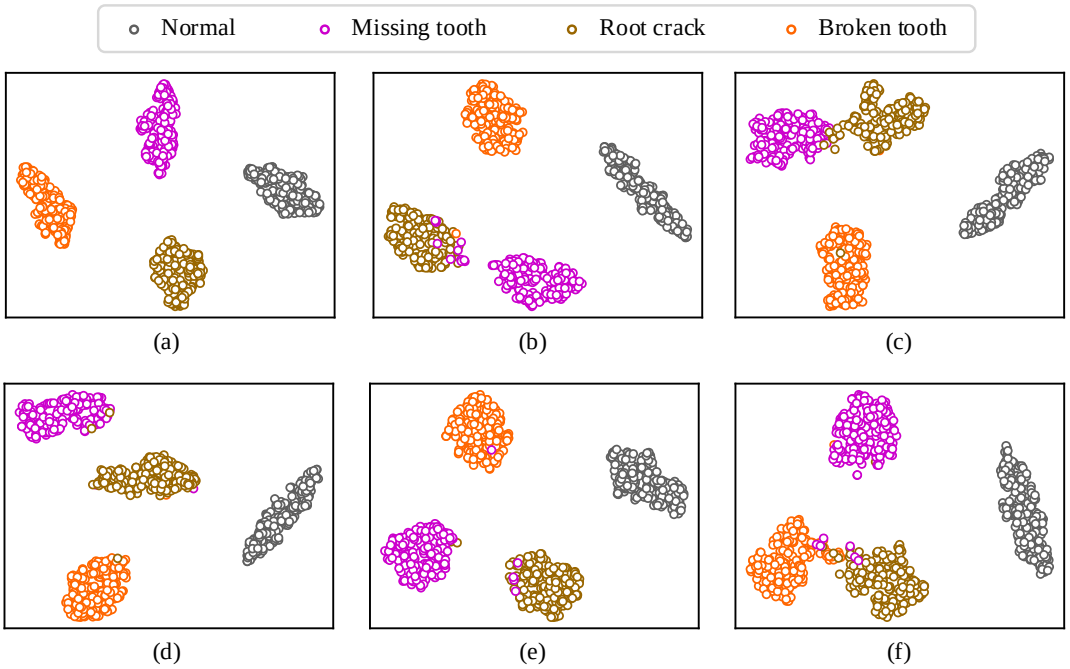


Fig.11. Feature visualization via t-SNE of the S2 dataset: (a) Proposed; (b) ACGAN; (c) VAE-GAN; (d) SMOTE-T; (e) CBL; (f) DWBL.

5.2. Case 2: Planetary gearbox from Xi'an Jiaotong University

5.2.1. Data preparation of Case 2

The datasets of Case 2 are derived from the DDS [47] of Xi'an Jiaotong University, with the test bench shown in **Fig.12 (a)**. Similarly, four gearbox health states are selected for experimental validation, including the normal state and three planetary gear fault modes depicted in **Fig.12 (b)**. During experiments, two accelerometers are installed in the X and Y directions of the planetary gearbox to collect vibration signals, with the motor speed of 1800 r/min and sampling frequency of 20,480 Hz. Three datasets with different imbalanced ratios, namely X1, X2, and X3, are constructed as listed in **Table 5**. The dataset configuration rules are consistent with **Table 2**, but the number of training samples and imbalanced ratios vary slightly. In the three datasets, the number of simulated training samples for various fault modes used in the proposed method is set to 384.

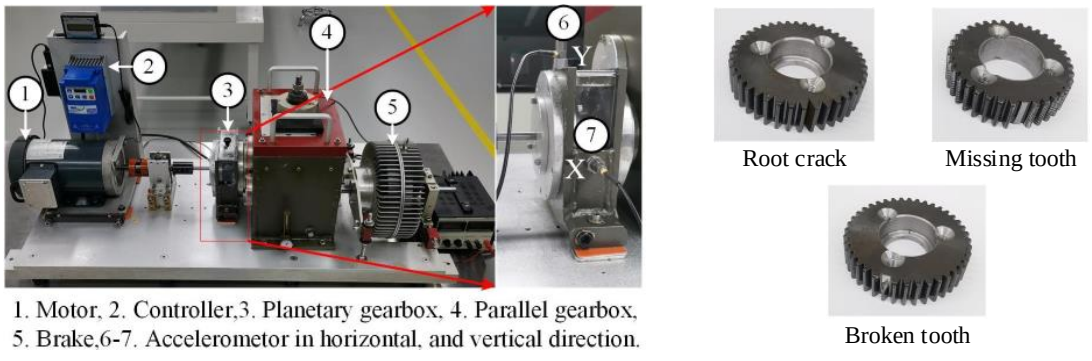


Fig.12. The experimental setup Case 2: (a) The test bench; (b) Three types of faulty gears.

Table 5
Imbalanced datasets of Case 2.

Health states	Labels	Datasets (No. of training samples)			No. of testing samples
		X1	X2	X3	
Normal	0	336	348	360	200
Missing tooth	1	16	12	8	200
Root crack	2	16	12	8	200
Broken tooth	3	16	12	8	200

5.2.2. Result analysis of Case 2

Table 6 lists the diagnostic accuracy of each method in Case 2. In the X1 dataset with 21 times the imbalanced ratio, ACGAN achieves the best diagnostic accuracy. This indicates that when a certain number of fault samples are available, ACGAN can effectively utilize fault information to generate multi-state fault samples. The cost-sensitive methods CBL and WL, by balancing the feature learning effect between classes, also exhibit satisfactory diagnostic results in the X1 dataset. However, as the imbalanced ratio increases, these two-categories imbalanced methods are constrained by extremely limited fault information, and the diagnostic results show a clear declining trend. In the X2 dataset with 29 times the imbalanced ratio, ACGAN experiences mode collapse during the training process, leading to significant fluctuations in diagnostic results. The diagnostic results of cost-sensitive methods also show noticeable fluctuations and declines. In contrast, the proposed method's mean accuracy only decreases by 2.14% compared to the X1 dataset, effectively mitigating the downward trend and demonstrating the best mean and minimum accuracy. In the X3 dataset with 45 times the imbalanced ratio, although the maximum accuracy of proposed method is 3.50% lower than ACGAN, its mean and minimum accuracies are 2.65% and 13.12% higher, respectively, outperforming ACGAN in stability. Compared to other benchmark methods, the proposed method also exhibits obvious diagnostic advantages.

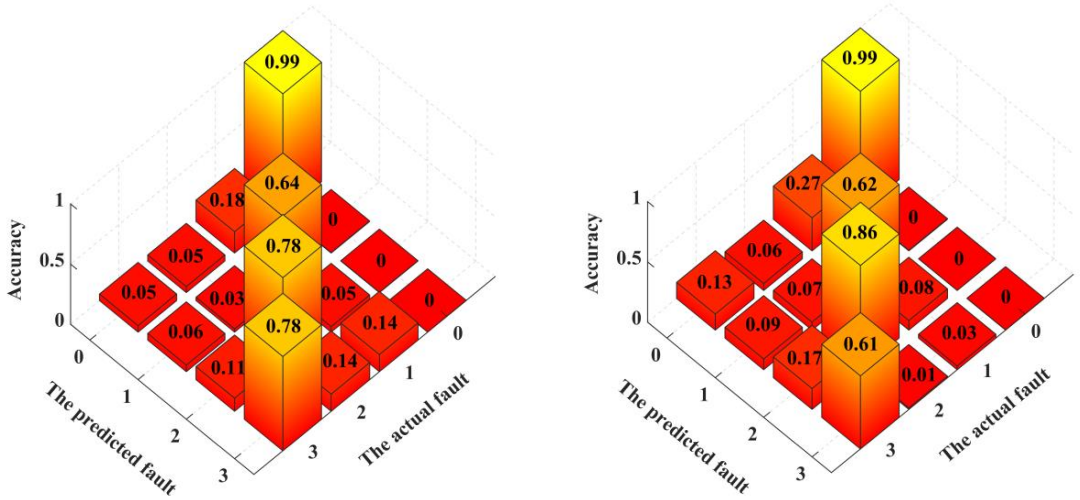
Table 6
Diagnostic accuracy of each method in Case 2.

Method	X1-Accuracy (%)			X2-Accuracy (%)			X3-Accuracy (%)			Avg (%)
	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	
Proposed	81.97	83.75	78.62	**79.83**	83.50	**75.25**	**73.50**	76.38	**70.50**	**78.43**
ACGAN	**85.47**	**87.00**	**84.12**	77.03	**87.00**	48.62	70.85	**79.88**	57.38	77.78
VAE-GAN	71.72	75.00	64.88	65.53	70.00	60.38	55.80	57.00	54.00	64.35
WGAN-GP	71.90	76.25	68.50	66.00	73.25	58.13	58.67	63.00	52.62	65.52
SMOTE-T	76.03	86.00	70.00	66.20	75.50	54.00	60.33	75.12	40.62	67.52
CBL	80.65	84.38	76.62	73.67	81.50	66.00	65.35	73.12	56.88	73.22
DWBL	75.35	79.75	70.62	70.83	75.25	66.38	62.95	66.75	60.50	69.71
FL	77.15	82.50	75.62	68.88	74.25	63.38	63.45	65.38	62.25	69.83
WL	81.85	86.62	77.25	72.75	77.38	62.62	61.92	70.25	54.62	72.17

Remarks: X1-Accuracy (%) represents the accuracy in the X1 dataset. Mean is the mean accuracy of the 5 repeated experiments; Max is the maximum accuracy and Min is the minimum accuracy. Avg denotes the average accuracy of the X1, X2 and X3 datasets.

To further compare the diagnostic results of the methods in each health state, we visualize the three-dimensional confusion matrix in the X3 dataset, as shown in **Fig.13**. It should be noted that the two methods with the best diagnostic results in **Table 6** are selected for comparison, and the average of

5 repeated experiment results is taken. The horizontal coordinates 0-3 correspond to the four health conditions in **Table 5**. Although the proposed method has an 8% lower diagnostic result for root crack than ACGAN, its discrimination effects for broken tooth and missing tooth are superior, with a 17% and 2% higher rate, respectively. Both methods exhibit high recognition accuracy for the normal state, as the feature information of normal samples is abundant. Based on the height of bars on the diagonal, it can be intuitively observed that the classification stability of the proposed method for the three fault modes is better than that of ACGAN.



(a) (b)

Fig.13. Three-dimensional confusion matrix in the X3 dataset: (a) Proposed; (b) ACGAN.

Fig.14 visualizes the *J* values of each method in the three datasets, with black dots representing the *J* values of experiments. As the imbalanced ratio increases, the *J* values of all methods show varying degrees of fluctuation and decline, consistent with the pattern presented in **Table 6**. In the S1 dataset, the mean value of the proposed method is slightly better than that of ACGAN, achieving the best result. In the S2 dataset, the mean value of the proposed method only declines by about 4% compared to S1 dataset, while the mean value of ACGAN decreases by about 12% with severe fluctuations; the best cost-sensitive method DWBL also declines by about 10%. In the S3 dataset, the mean value of the proposed method still maintains above 2 with a smaller volatility range, while the mean values of other methods are all noticeably below 2. This further verifies the clustering and discrimination advantages of the proposed method for sample features in the imbalanced scenario.

To further compare the feature extraction performance of the methods for different health states, the receiver operating characteristic (ROC) curves for all health states in the X2 dataset are plotted, comparing the cost-loss method DWBL with the proposed method as shown in **Fig.15**. Similar to **Fig.13**, the mean of 5 repeated experiment results is taken to reduce error interference. The macro-average is the average false positive rate and true positive rate of the four health states; the micro-average calculation involves adding the true positive, false positive, true negative, and false negative counts for each health condition, and then calculating the false positive rate and true positive rate. The area under the curve (AUC) is used to measure the feature quality of each health condition, listed in the parentheses in the bottom right corner. It can be intuitively seen that the ROC curves of the proposed method are superior to the benchmark method DWBL. For the four health states, the AUC values of the

proposed method are 0.80%, 5.96%, 2.55%, and 2.40% higher than DWBL, respectively; the macro-average and micro-average AUC values are also 2.92% and 6.42% higher, respectively. This indicates that the proposed method has good feature extraction capabilities for samples from different health states and better stability, further validating the results of **Fig.13**.

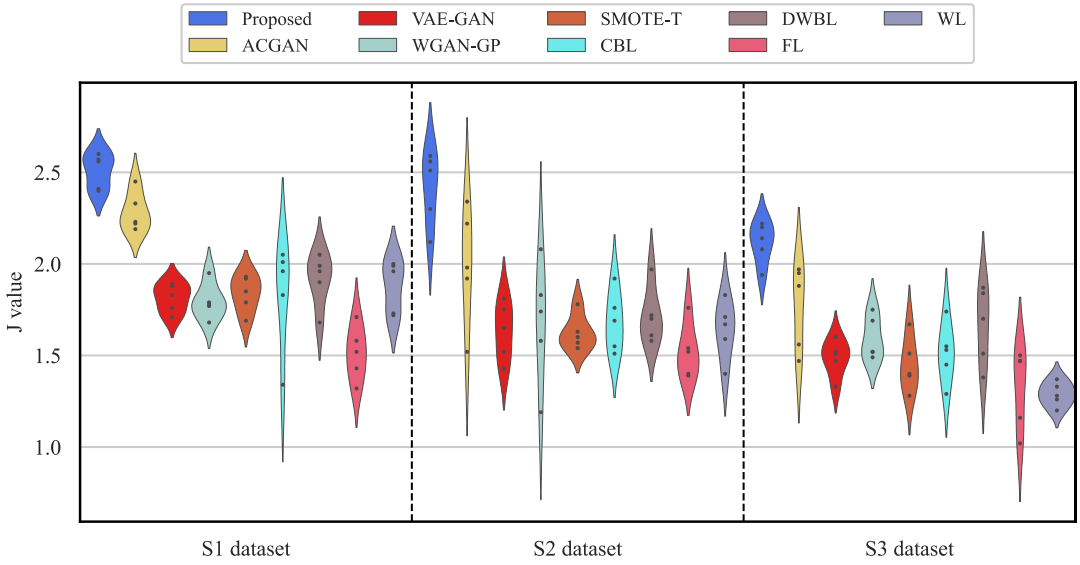


Fig.14. Feature quantitative evaluation of each method in Case 2.

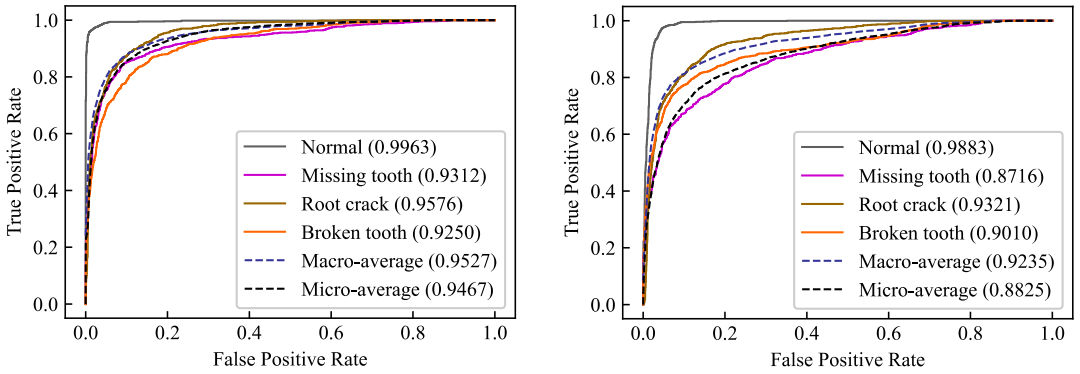


Fig.15. ROC curves in the X2 dataset: (a) Proposed; (b) DWBL.

5.3. The ablation experiment

To further test the effectiveness of SAM and MAR, the ablation experiment is conducted in the six datasets of the two cases, with mean diagnostic accuracy shown in **Fig.16**. - MAR represents the proposed method minus MAR; - SAM represents the proposed method minus SAM; and - (MAR+SAM) represents both being subtracted, with only simulated fault data added for model training. In the three datasets of Case 1, as the imbalanced ratio increases, the effectiveness of SAM and MAR is demonstrated. In the S2 dataset with an imbalanced ratio of 125 times, the mean accuracy increases by approximately 6% after only adding MAR; then adding SAM, the mean accuracy increases by another 3%. In the S3 dataset with an imbalanced ratio of 253 times, the mean accuracy decreases by approximately 3% after subtracting MAR from the proposed method; then subtracting SAM, the accuracy decreased by another 7%. In the three datasets of Case 2, SAM plays a more significant role.

In the X1 dataset, the mean accuracy increases by approximately 8% when only adding SAM; in the X2 dataset, the mean accuracy quickly drops by about 10% after subtracting SAM from the proposed method; in the X3 dataset, the accuracy even drops by about 14% after subtracting SAM from the proposed method. This is because the distribution spaces of fault data in different categories in the simulated and the real domain have certain discrepancies. SAM can effectively reduce the dissimilarity of fine-grained features between simulated and real fault data, aligning the conditional distribution of fault data in both domains. Overall, MAR can improve the stability of model's diagnosis to some extent by imposing significant regularization penalties on the margins of minority class samples.

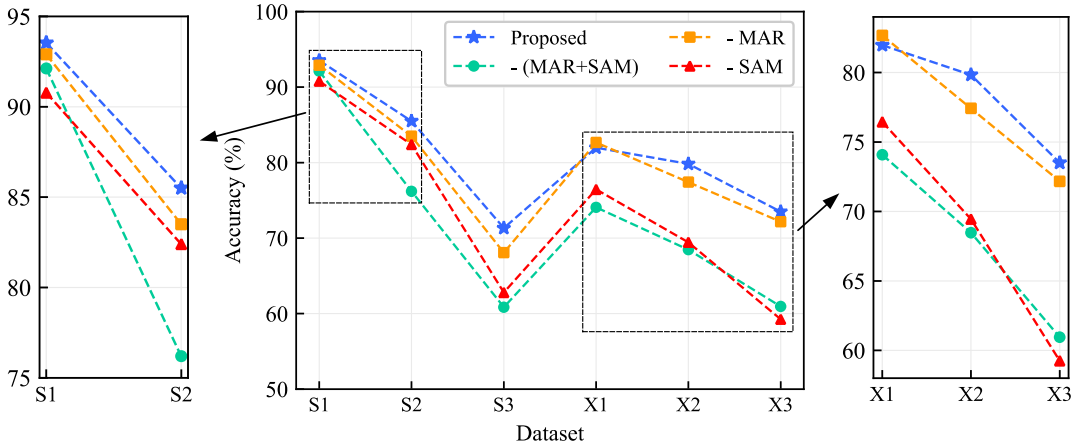


Fig.16. Mean diagnostic accuracy of the ablation experiment.

6. Conclusions

To address the limitations of data-level and algorithm-level based imbalanced fault diagnosis methods, this study proposes a novel digital twin-assisted imbalanced fault diagnosis framework. Firstly, the nonlinear kinetic characteristics of gearbox are analyzed and its dynamic simulation model is established using digital twin technology to obtain high-fidelity simulated fault data. Secondly, the SAM is employed to align the conditional distribution of subdomains by minimizing the dissimilarity of fine-grained features between simulated and real fault data. Finally, the MAR is designed to improve the fault tolerance of model's diagnosis by imposing significant regularization penalties on the margins of fault data.

The imbalanced fault diagnosis results of two gearboxes show that: (1) Compared to the popular data-level and algorithm-level imbalanced fault diagnosis methods, the proposed method is superior under the influence of highly imbalanced data. (2) By effectively integrating the data-level digital twin technology and the algorithm-level SAM and MAR, the framework provides a new perspective for dealing with the imbalanced fault diagnosis scenario. In line with the characteristics of fault diagnosis tasks, we will continue to conduct in-depth research on the lightweight implementation and interpretability of the method, facilitating its deployment and improving its reliability.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (No. 52275104) and the Natural Science Fund for Excellent Young Scholars of Hunan Province (No. 2021JJ20017).

References

[1] Kumar A, Kumar R. Least square fitting for adaptive wavelet generation and automatic prediction of defect size in the bearing using levenberg–marquardt backpropagation. Journal of Nondestructive Evaluation, 2017, 36(1): 7.

[2] Kumar A, Kumar R. Enhancing weak defect features using undecimated and adaptive wavelet transform for estimation of roller defect size in a bearing. Tribology Transactions, 2017, 60(5): 794-806.

[3] Kumar A, Kumar R. Oscillatory behavior-based wavelet decomposition for the monitoring of bearing condition in centrifugal pumps. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2018, 232(6): 757-772.

[4] Yan S, Shao H, Min Z, et al. FGDAE: A new machinery anomaly detection method towards complex operating conditions. Reliability Engineering & System Safety, 2023, 236: 109319.

[5] Gu J X, Albarbar A, Sun X, et al. Monitoring and diagnosing the natural deterioration of multi-stage helical gearboxes based on modulation signal bispectrum analysis of vibrations. International Journal of Hydromechatronics, 2021, 4(4): 309-330.

[6] Hilbert M, Smith W A, Randall R B. The effect of signal propagation delay on the measured vibration in planetary gearboxes. Journal of Dynamics, Monitoring and Diagnostics, 2022, 1(1): 9-18.

[7] Hahn T V, Mechefske C K. Self-supervised learning for tool wear monitoring with a disentangled-variational-autoencoder. International Journal of Hydromechatronics, 2021, 4(1): 69-98.

[8] Wang W, Lei Y, Yan T, et al. Residual convolution long short-term memory network for machines remaining useful life prediction and uncertainty quantification. Journal of Dynamics, Monitoring and Diagnostics, 2022, 1(1): 2-8.

[9] Lin J, Shao H, Zhou X, et al. Generalized MAML for few-shot cross-domain fault diagnosis of bearing driven by heterogeneous signals. Expert Systems with Applications, 2023: 120696.

[10] Kumar A, Berrouche Y, Zimroz R, et al. Non-parametric Ensemble Empirical Mode Decomposition for extracting weak features to identify bearing defects. Measurement, 2023, 211: 112615.

[11] Chen X, Shao H, Xiao Y, et al. Collaborative fault diagnosis of rotating machinery via dual adversarial guided unsupervised multi-domain adaptation network. Mechanical Systems and Signal Processing, 2023, 198: 110427.

[12] Ren Z, Lin T, Feng K, et al. A systematic review on imbalanced learning methods in intelligent fault diagnosis. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 3508535.

[13] Sun M, Qian H, Zhu K, et al. Ensemble learning and SMOTE based fault diagnosis system in self-organizing cellular networks. GLOBECOM 2017 - 2017 IEEE Global Communications Conference, 2017: 1-6.

[14] Swana E F, Doorsamy W, Bokoro P. Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset. Sensors, 2022, 22(9): 3246.

[15] Chen M, Shao H, Dou H, et al. Data augmentation and intelligent fault diagnosis of planetary gearbox using ILoFGAN under extremely limited samples. IEEE Transactions on Reliability, DOI 10.1109/TR.2022.3215243.

[16] Liu S, Jiang H, Wu Z, et al. Rolling bearing fault diagnosis using variational autoencoding generative adversarial networks with deep regret analysis. Measurement, 2021, 168: 108371.

[17] Li W, Zhong X, Shao H, et al. Multi-mode data augmentation and fault diagnosis of rotating machinery using modified ACGAN designed with new framework. Advanced Engineering Informatics, 2022, 52: 101552.

[18] Yu Y, Guo L, Gao H, et al. PCWGAN-GP: A new method for imbalanced fault diagnosis of machines. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 3515711.

[19] Zhao X, Yao J, Deng W, et al. Normalized Conditional Variational Auto-Encoder with adaptive Focal loss for imbalanced fault diagnosis of Bearing-Rotor system. Mechanical Systems and Signal Processing, 2022, 170: 108826.

[20] Peng Y, Wang Y, Shao Y. A novel bearing imbalance Fault-diagnosis method based on a Wasserstein conditional generative adversarial network. Measurement, 2022, 192: 110924.

[21] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[22] Cui Y, Jia M, Lin T Y, et al. Class-balanced loss based on effective number of samples. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 9268-9277.

[23] Fernando K R M, Tsokos C P. Dynamically weighted balanced loss: class imbalanced learning and confidence calibration of deep neural networks. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(7): 2940-2951.

[24] Jia F, Lei Y, Lu N, et al. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. Mechanical Systems and Signal Processing, 2018, 110: 349-367.

[25] Duan A, Guo L, Gao H, et al. Deep focus parallel convolutional neural network for imbalanced classification of machinery fault diagnostics. IEEE Transactions on Instrumentation and Measurement, 2020, 69(11): 8680-8689.

[26] He Z, Shao H, Cheng J, et al. Support tensor machine with dynamic penalty factors and its application to the fault diagnosis of rotating machinery with unbalanced data. Mechanical systems and signal processing, 2020, 141: 106441.

[27] Yu Y, Guo L, Gao H, et al. Pareto-optimal adaptive loss residual shrinkage network for imbalanced fault diagnostics of machines. IEEE Transactions on Industrial Informatics, 2021, 18(4): 2233-2243.

[28] Hou R, Chen J, Feng Y, et al. Contrastive-weighted self-supervised model for long-tailed data classification with vision transformer augmented. Mechanical Systems and Signal Processing, 2022, 177: 109174.

[29] Ren Z, Zhu Y, Kang W, et al. Adaptive cost-sensitive learning: Improving the convergence of intelligent diagnosis models under imbalanced data. Knowledge-Based Systems, 2022, 241: 108296.

[30] Liu H, Liu Z, Jia W, et al. A novel imbalanced data classification method based on weakly supervised learning for fault diagnosis. IEEE Transactions on Industrial Informatics, 2021, 18(3): 1583-1593.

[31] Feng K, Ji J C, Zhang Y, et al. Digital twin-driven intelligent assessment of gear surface degradation. Mechanical Systems and Signal Processing, 2023, 186: 109896.

[32] Wang J, Zhang Z, Liu Z, et al. Digital twin aided adversarial transfer learning method for domain adaptation fault diagnosis. Reliability Engineering & System Safety, 2023, 234: 109152.

[33] Zhang Y, Ji J C, Ren Z, et al. Digital twin-driven partial domain adaptation network for intelligent fault diagnosis of rolling bearing. Reliability Engineering & System Safety, 2023, 234: 109186.

[34] Wang H, Zheng J, Xiang J. Online bearing fault diagnosis using numerical simulation models and machine learning classifications. Reliability Engineering & System Safety, 2023, 234: 109142.

[35] Xiang L, Zhang X, Zhang Y, et al. A novel method for rotor fault diagnosis based on deep transfer learning with simulated samples. Measurement, 2023, 207: 112350.

[36] Zhao Z, Zhang Q, Yu X, et al. Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 3525828.

[37] Kuang J, Xu G, Tao T, et al. Class-imbalance adversarial transfer learning network for cross-domain fault diagnosis with imbalanced data. IEEE Transactions on Instrumentation and Measurement, 2021, 71: 3501111.

[38] Liu X, Chen J, Zhang K, et al. Cross-domain intelligent bearing fault diagnosis under class imbalanced samples via transfer residual network augmented with explicit weight self-assignment strategy based on meta data. Knowledge-Based Systems, 2022, 251: 109272.

[39] Wu Z, Zhang H, Guo J, et al. Imbalanced bearing fault diagnosis under variant working conditions using cost-sensitive deep domain adaptation network. Expert Systems with Applications, 2022, 193: 116459.

[40] Ding Y, Jia M, Zhuang J, et al. Deep imbalanced domain adaptation for transfer learning fault diagnosis of bearings under multiple working conditions. Reliability Engineering & System Safety, 2023, 230: 108890.

[41] Yi Y, Huang K, Xiong Y, et al. Nonlinear dynamic modelling and analysis for a spur gear system with time-varying pressure angle and gear backlash. Mechanical Systems and Signal Processing, 2019, 132: 18-34.

[42] Song M M, Xiong Z C, Zhong J H, et al. Research on fault diagnosis method of planetary gearbox based on dynamic simulation and deep transfer learning. Scientific Reports, 2022, 12(1): 17023.

[43] Zhu Y, Zhuang F, Wang J, et al. Deep subdomain adaptation network for image classification.

IEEE transactions on neural networks and learning systems, 2021, 32(4): 1713-1722.

[44] Cao K, Wei C, Gaidon A, et al. Learning imbalanced datasets with label-distribution-aware margin loss. Advances in neural information processing systems, 2019, 1567-1578.

[45] Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101, 2017.

[46] Shao S, McAleer S, Yan R, et al. Highly accurate machine fault diagnosis using deep transfer learning. IEEE Transactions on Industrial Informatics, 2018, 15(4): 2446-2455.

[47] Li T, Zhou Z, Li S, et al. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. Mechanical Systems and Signal Processing, 2022, 168: 108653.