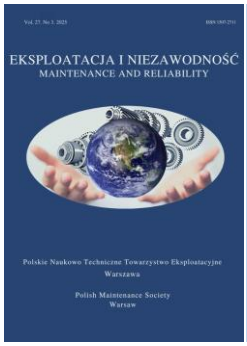


Eksploatacja i Niezawodnosc – Maintenance and Reliability
Volume 27 (2025), Issue 3
journal homepage: http://www.ein.org.pl

Article citation info:
Yang N, Liu J, Ma H, Zhao W, Gao Y, Fault diagnosis of gear based on URP-CVAE-MGAN under imbalanced and small sample conditions, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2025: 27(3) http://doi.org/10.17531/ein/199417

Fault diagnosis of gear based on URP-CVAE-MGAN under imbalanced and small sample conditions

Indexed by: Web of Science Group

Na Yang[a], Jie Liu[a,*], Hui Ma[b], Weiqiang Zhao[a], Yu Gao[a]

[a] Shenyang University of Technology, China
[b] Northeastern University, China

Highlights

- The Un-threshold Recurrence Plots (URP) and Vision Transformer (ViT) improved by Dropkey are used for diagnosing gear fault types and severities.
- A new generated model, the Variational Autoencoder added Conditional variable (CVAE) combined with Generative Adversarial Network improved Mean feature difference function (MGAN), is used for data augmentation of un-threshold recurrence plots from gear under imbalanced and small sample conditions.
- Dropkey-ViT has more advantages in comprehensively capturing fault information compared with the comparison method.

Abstract

To improve diagnosis accuracy for gear fault diagnosis under imbalanced and small sample conditions, a method combining the Un-threshold Recurrence Plots - Conditional Variational Autoencoder-Mean Generative Adversarial Network (URP-CVAE-MGAN) combined with Dropkey-Vision Transformer (DViT) is proposed. First, gear vibrational signals are transformed into Recurrence Plots (RP) images to extract more fault features without threshold effect. Then, a conditional variable and mean feature difference function are incorporated into VAE-GAN to improve the quality and diversity of generated samples, balancing the imbalanced and small sample sets. Dropkey is applied to the diagnosis model Vision Transformer to capture more fault information, improving diagnosis accuracy across various fault types and severities for gear. Finally, the proposed method is verified based on two datasets, demonstrating a significant accuracy improvement of up to 7.84% under the imbalanced and small samples, and confirming its feasibility and superiority.

Keywords
gear fault diagnosis, conditional variational autoencoder, dropkey, generative adversarial network, mean feature matching, vision transformer

1. Introduction

In recent years, with the rapid advancement of industrial intelligence, increasing attention has been placed on the operation and maintenance of rotating machinery. As critical components of rotating machinery, gears, and bearings operate under harsh conditions and within complex structures. Therefore, efficient and accurate fault diagnosis is essential for the health management of rotating machinery. Prolonged operation can lead to various faults in gears, including Worn, Cracked, Missing, Chipped, and so on. If these faults are not detected and addressed promptly, significant losses in personnel and property may occur [1,2]. The diagnosis of gear faults primarily relies on vibration signals. Initially, time-domain [3,4], frequency--domain [5], and time-frequency analysis [6,7] were employed, alongside feature extraction methods such as Singular Value Decomposition (SVD) [8,9], Wavelet Transform (WT) [10] and Empirical Mode Decomposition (EMD) [11], often integrated with machine learning methods. Zhao et al. analyzed the vibrational properties of pitting-faulty gears using

(*) Corresponding author.
E-mail addresses: N. Yang (ORCID: 0000-0001-7733-5384) skyyangna@126.com, J. Liu (ORCID: 0000-0002-9772-2039) liuj@sut.edu.cn, H. Ma (ORCID: 0000-0002-9812-0806) mahui_2007@163.com, W. Zhao (ORCID: 0000-0003-1881-5155) weiq.zhao@foxmail.com, Y. Gao (ORCID:0009-0005-9328-8289) gaoyu_2235@163.com

time-frequency, frequency-domain, and time-domain methods, examining the impact of the pitting area [12]. Wang et al. applied sparse filtering to extract fault features from the frequency domain and classified different fault types using SoftMax regression [13]. Gradually, deep learning methods have gained widespread attention for their end-to-end manner of achieving fault diagnosis and eliminating human factors [14,15]. A common practice involves transforming vibration signals into time-frequency images by WT, which are then used as inputs to deep learning models. This mainly utilizes the powerful capabilities of deep learning in the field of image recognition by extracting features from time-frequency images for fault diagnosis. WT enables multi-resolution analysis, capturing both the time-domain and frequency-domain features of signals, and is suitable for non-stationary gear vibration signals. However, deep learning models require sufficient fault samples, which are often limited and unevenly distributed, particularly in rotating equipment that primarily operates in healthy states [16]. Currently, under small and imbalanced conditions, sufficient characterization of fault characteristics is crucial, and the global feature structure becomes crucial beyond just time-frequency domain features. Another primary problem to be solved is for augmenter imbalanced and small samples. The current data augmentation methods include rotating, translation, scaling, flipping, cropping, and noise-adding [17]. Yu et al. proposed seven augmentation strategies for small sample one-dimensional monitoring data [18]. Easy to implement, but alters signal characteristics like temporality, periodicity, and amplitude. Deep generative models, such as Generative Adversarial Network (GAN), can enhance datasets while preserving key signal features. Through adversarial training, samples with distributions similar to real data are generated, addressing imbalanced and small sample conditions. However, adversarial training between two networks often results in instability and collapse, necessitating successive improvements. In response, a new Sparse Constraint Generative Adversarial Network model (SC-GAN) was proposed in [19], incorporating sparse constraints to make signals more interpretable, thus generating more stable vibration signals. Wang et al. employed a GAN variant optimized by Wasserstein to generate samples. In GAN and its variants, such as Stacked Generative Adversarial Networks (SGAN), Wasserstein Generative Adversarial Network (WGAN), Auxiliary Classifier Generative and Adversarial Network (ACGAN), data features are effectively learned, and samples are expanded. However, issues such as gradient vanishing or explosion persist [20]. Similarly, the Variational Auto-Encoder (VAE) introduces latent variables to capture the underlying data structure, generating new samples through a decoder [21]. By applying variational inference, VAEs learn complex features and structures, generating samples with distributions similar to the original data. These samples typically exhibit better interpretability and stability. Zhao et al used VAE to generate more vibrational signals for machines [22,23]. Alfredo et al established a new VAEs structure to process incomplete and heterogeneous datasets, suitable for both supervised and unsupervised scenarios [24]. However, when dealing with high-dimensional data and intricate distributions, VAEs may struggle to fully capture detailed information, resulting in lower quality and realism of the generated samples. Ensuring diversity and quality becomes challenging, especially when samples are scarce for certain categories, limiting the model's ability to generate samples with specific attributes, and lacking generalization. In addition to generating sufficient samples, establishing accurate models is equally critical. Convolutional Neural Networks (CNN) [25], AlexNet [26], GoogLeNet [27], Deep Belief Network [28] and their variants [29-31] are widely employed in fault diagnosis. These models, composed of fully connected layer, convolution, and pooling operations, have demonstrated significant diagnostic success due to the powerful feature extraction ability of convolutional layers. However, when processing complex fault features, long-distance dependencies, or diverse data forms (e.g., two-dimensional gray maps, time-frequency maps, etc.) from gear signals, CNNs often only capture partial features due to the limitations of convolutional kernels. Global features, particularly those from long-term series information, are often overlooked, especially under imbalanced and small sample conditions, leading to overfitting. To address these shortcomings, the Vision Transformer has been introduced. Tang et al. transformed one-dimensional bearing signals into time-frequency images using WT, constructed multiple parallel ViT models, and proposed a soft voting method to fuse diagnostic results [32]. Zhou et al. extracted time-frequency features of bearings using CNN and implemented the

final diagnosis with ViT [33]. With the attention mechanisms, ViT captures global dependencies in signals and dynamically adjusts focus on different regions, effectively handling complex, non-stationary, and nonlinear gear signals. While Dropout prevents overfitting by randomly discarding activation values, its randomness may result in the loss of important fault features, impacting feature extraction and diagnostic accuracy.

To overcome the limitations of gear fault diagnosis methods under imbalanced and small sample conditions, a method integrating UPR-CVAE-MGAN with an improved ViT using Dropkey is proposed. This method aims to enhance fault information representation, generate high-quality and diverse samples, and improve the diagnostic network's ability to capture gear fault-related details. Accurate diagnosis of various types and severities of gear faults is achieved in the final. The key contributions are shown:

(1) URP has been proposed to transform complex, non-stationary gear vibration signals into two-dimensional images, capturing richer nonlinear and complex dynamic features compared to traditional time-frequency maps. Without relying on specific thresholds, the converted images more comprehensively reflect potential fault features.

(2) By combining VAE and GAN, conditional variables are introduced to enable the generation of samples corresponding to specific fault modes. To ensure the correlation between generated samples and fault types, a mean feature difference loss function is applied, enhancing the quality of generated samples and their similarity to real samples. This reduces feature bias and improves the realism and diversity of the samples.

(3) To optimize the ViT's extraction capabilities and prevent the loss of important information, DropKey is introduced as a replacement for traditional Dropout. This technique selectively discards irrelevant or redundant features, enhancing the model's ability to capture fault-related information.

(4) The URP recursive graph enhances gear fault features, while the CVAE-MGAN generates high-quality extended samples, mitigating issues related to small sample sizes and data imbalance. The Dropkey- ViT further improves the model' s classification performance in complex fault diagnosis. The integration of these components enables the diagnostic model to effectively address various types and severities of gear faults at different stages, including data input, feature enhancement, sample expansion, and feature extraction thereby demonstrating excellent accuracy and robustness.

The paper is organized as follows: Section 2 presents the theoretical backdrop, Section 3 elaborates on the proposed method, Section 4 compares the method's performance to prevalent approaches, and Section 5 concludes with key findings.

2. Background

2.1. Un-threshold Recurrence Plots (URP)

The URP is primarily used to analyze the non-stationarity, chaos, and periodicity of time-series data [34]. A two-dimensional matrix (that is recursive plot) is constructed by comparing data from different time points, where each element represents the similarity or recurrence between corresponding points in the time series. Finally, the time-series data is encoded by PR to two-dimensional images to enhance feature information. The RP usually contains a traditional threshold recursive graph and a threshold-free recursive graph. Traditional RP often relies on threshold selection, which can lead to distortion, particularly for non-stationary signals. In contrast, threshold-free RP can automatically determine node connections based on the data characteristics, retaining more information [35,36]. In this paper, the URP is adopted to transform complex, non-stationary gear vibration signals into two-dimensional images, capturing richer nonlinear and dynamic features compared to traditional time-frequency mapping methods without relying on any specific threshold. The main process is recorded as follows.

The time series signal/data is provided as $u_k(k = 1,2,\dots,n)$ and the sampling time interval is determined as Δt. The embedding dimensions m and latency time τ are used to reconstruct the phase space x_i.

$$x_i = [u_i, u_{i+\tau}, \dots, u_{i+(m-1)\tau}] \tag{1}$$

where, $i = 1,2,\dots,n-(m-1)\tau$.

The distance between the i point x_i and the j point x_j from the phase space is calculated after reconstruction.

$$S_{ij} = \|x_i - x_j\| \tag{2}$$

where, $i,j = 1,2,\dots,n-(m-1)\tau$, $\|\ \ \|$ representatives norm.

Finally, the recursive value is calculated as:

$$R_{ij} = \theta(\varepsilon_i - S_{ij}) \tag{3}$$

where, R_{ij} is a $N \times N$ square matrix; N is the number of

condition vectors x_i; The threshold ε represents a preset critical distance; $\theta(\cdot)$representatives Heaviside function, its expression is :

$$\theta(r) = \begin{cases} 1 & r \geq 0 \\ 0 & r < 0 \end{cases} \quad (4)$$

2.2. Principle of the proposed CVAE-MGAN

VAE, introduced by Kingma and Welling in 2013, consists of an encoder and decoder [37]. VAE is applied to generate new data through potential representation. GAN, a generation model introduced by Ian Goodfellow in 2014, generates new data via adversarial training between a generator and a discriminator [38]. Both models have limitations, and their combination of VAE-GAN has been applied in sample augmentation, where VAE provides a robust initial latent space representation, and GAN refines the generated samples to enhance their realism [39]. However, the training process remains unstable, the generated samples lack clarity, and sample diversity is limited, despite the integration of GAN. Additionally, gradient vanishing and explosion frequently occur due to the Cross-Entropy loss in GAN. Although the VAE-GAN combination partially mitigates some of their individual shortcomings, it exhibits poor adaptability to different fault types and fails to generate targeted samples for specific fault conditions. This is primarily because VAE-GAN lacks sufficient conditional constraints for various fault types, leading to generated samples that may not accurately reflect actual fault characteristics. Furthermore, the cross-entropy loss function performs inadequately when there are significant feature differences. When the generated samples deviate significantly from the real fault signal in feature space, the model may fail to capture key features, thus impacting the accuracy of fault diagnosis. Thus, the conditional variable (e.g., label) and a mean feature matching function are introduced to improve VAE-GAN [40]. Additional conditional information is incorporated to generate samples that better align with the expectations of the encoder. The adversarial training process helps mitigate the issue of mode collapse, enhances training stability, and improves both the diversity and clarity of generated samples. The improved model, named CVAE-MGAN, comprises an encoder, generator, discriminator, and classifier. Input data with conditional encoding is transformed into distribution variables (mean and standard deviation) in the latent space by the encoder, replacing the random variable input in the original generator. The generator then reconstructs the original pixels, aligning the characteristics of the original image with a given latent vector. The discriminator is presented with both real samples and generated samples, while the classifier calculates the class probability of the input data. The improved CVAE-MGAN's specific structure is illustrated in Fig.1.

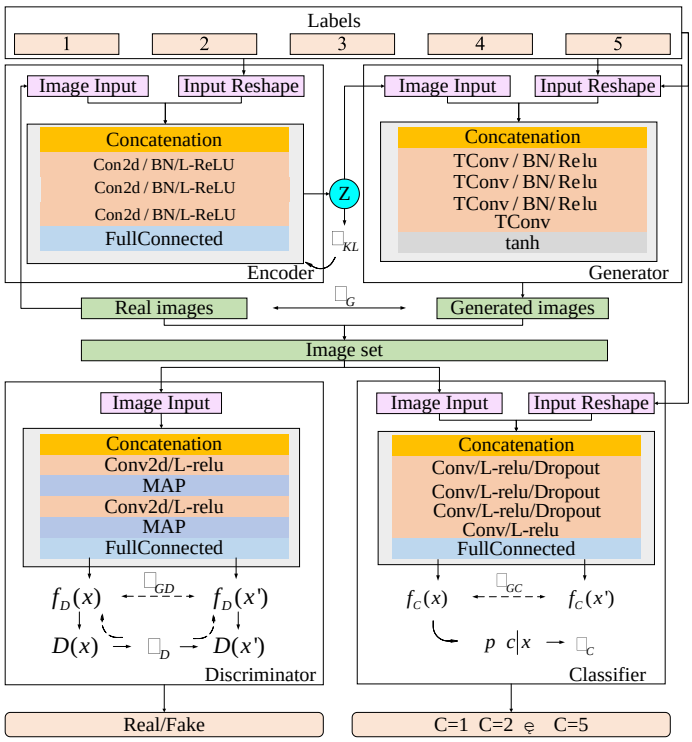


Fig. 1. The structure of the proposed CVAE-MGAN.

Especially, the goal of the improved method is to minimize the following loss function.

$$\mathcal{L} = \mathcal{L}_D + \mathcal{L}_C + \lambda_1 \mathcal{L}_{KL} + \lambda_2 \mathcal{L}_G + \lambda_3 \mathcal{L}_{GD} + \lambda_4 \mathcal{L}_{GC} \quad (5)$$

$$\mathcal{L}_D = -\mathbb{E}_{x\sim P_r}[log\ D\ (x)] - \mathbb{E}_{z\sim P_z}[log(\ 1 - D(G(z)))] \quad (6)$$

$$\mathcal{L}_{GD} = \frac{1}{2}\left\|\mathbb{E}_{x\sim P_r} f_D(x) - \mathbb{E}_{z\sim P_z} f_D\big(G(z)\big)\right\|_2^2 \quad (7)$$

$$\mathcal{L}_C = -\mathbb{E}_{x\sim P_r}[log\ P\ (c|x)] \quad (8)$$

$$\mathcal{L}_{GC} = \frac{1}{2}\sum_c \left\|\mathbb{E}_{x\sim P_r} f_C(x) - \mathbb{E}_{z\sim P_z} f_C\big(G(z,c)\big)\right\|_2^2 \quad (9)$$

$$\mathcal{L}_{KL} = \frac{1}{2}(\mu^T\mu + sum(exp(\ \varepsilon) - \varepsilon - 1)) \quad (10)$$

$$\mathcal{L}_G = \frac{1}{2}(\|x - x'\|_2^2 + \|f_D(x) - f_D(x')\|_2^2 + \|f_C(x) - f_C(x')\|_2^2) \quad (11)$$

where, $\mathcal{L}_D$ represents the loss function of the discriminator, G represents a generative network, D represents a discriminative network, x represents a real sample, p_r represents the real samples' probability distribution, z represents random noises being input to G network (Generally followed by Gaussian distribution). p_z represents the probability distribution of

random noises, $G(z)$ represents the samples generated by G, $D(x)$ represents the probability that D determines whether the real sample is true or not, $D(G(z))$ represents the probability that D determines whether the generative sample by G is true or not. $\mathcal{L}_{GD}$ represents the discriminator's loss function improved by the mean feature matching function. $f_D$represents a feature centre for discrimination. $\mathcal{L}_C$ represents the classifier's loss function, $P(c|x)$ represents the probability of some x as some class classed by the classifier. $\mathcal{L}_{GC}$ represents the generator's loss function improved by the mean feature matching function. f_C represents the feature centre similarly to the classifier. $\mathcal{L}_{KL}$represents the encoder's loss function. μ and ε represent the discriminator mean and covariance for the potential variables. $\mathcal{L}_G$ represents the loss function added an L2 reconstruction loss and feature matching loss based on the loss functions x and the generated x' for the generator.

2.3. Principle of the improved Vision Transformation

In 2021, Google introduced ViT at ICLR, marking a significant advancement by applying the transformer framework to the Computer Vision (CV) domain. [41]. Vit processes one-dimensional tokens (vectors) as input, requiring the input image to be transformed into a sequence of patches through an embedding layer, as shown in the following [42].

If the input image is $O \in \mathbb{R}^{m\times m\times c}$, it will be divided into $O_i \in \mathbb{R}^{k\times k\times c}$ $(i = 1,2,\dots,N)$ which are several equally sized $k\times k$ patches with the number of N by the Embedding layer, $Nk^2 = m^2$, m represents width, height, and c represents number of channels for the original image. Then, each patch O_i becomes a one-dimensional vector $I_i \in \mathbb{R}^{1,d}$ $(d = c \times k^2)$ after being exhibited and mapping a linear layer to a high-dimensional space form patching embedding (PE_i).

Subsequently, a class token $CLS \in \mathbb{R}^{1,de}$ is added into PE_i jointly form expansion embedding of the first dimension $EE \in \mathbb{R}^{N+1,de}$, and then positional information $PoE \in \mathbb{R}^{N+1,de}$ is added. The accurate input is informed as FE.

$$PE_i = I_i W_{PE} \tag{12}$$

$$FE = dropout(EE + PoE, 0.1) \tag{13}$$

where, W_{PE} is the parameter matrix, $dropout$is a regularization technique.

Through self-attention mechanism, the Transformer block converts the given hidden layer input into outputs of the same dimension, incorporating Multi-Headed Self Attention (MSA) to perform self-attention and a Feed Forward Network (FFN) to update weights. In MSA, the input tokens are transformed into three crucial vectors respectively D-dimensional Keys (K), Values (V), and Queries (Q). If there is only one head, the dimensional is $(N + 1) \times d$. Else the dimensional is $(N + 1) \times a$ and $n \times a = d$, the output dimensional is same as $(N + 1) \times d$. Finally, the result is obtained by multi-layer perceptron (MLP).

$$Attention(Q, K, V) = soft\,max(\frac{QK^T}{\sqrt{d_k}})V \tag{14}$$

$$FFN(x) = \sigma((xW_1 + b_1)W_2 + b_2) \tag{15}$$

where, W_1, W_2 are respectively weights of two linear transformations, $\sigma(\cdot)$is the nonlinear activation function.

However, regularization in ViT is performed through Dropout, where the same attention dropout is uniformly applied across different layers during the attention computation. Although simple, this method frequently leads to local features and instability during training. To address this, Li [43] proposed a novel regularization technique called DropKey. The key concept of DropKey is to treat the Key as the dropout target, allowing the model to adjust attention weights adaptively. By penalizing overly attended regions and redistributing attention towards other relevant areas, the model's capacity to capture global information is enhanced. As the layers deepen, the drop probability decreases, leading to more stable training. The details are detailed below:

The system calculates the dot products between each key and the query when a feature map consists of a query $Q \in \mathbb{R}^{n_h n_w\times n_c}$, keys $K \in \mathbb{R}^{n_h n_w\times n_c}$, and values $V \in \mathbb{R}^{n_h n_w\times n_c}$. Each result is scaled down by a normalizing factor $\sqrt{d_k}$. Subsequently, a random mask matrix $M \in \mathbb{R}^{1\times(n_h n_w\times n_h n_w)}$ is created, and its specifics are determined as outlined below:

$$m_j = \begin{cases} 0 & with probability \quad 1-m \\ -\infty & with probability \quad m \end{cases} \tag{16}$$

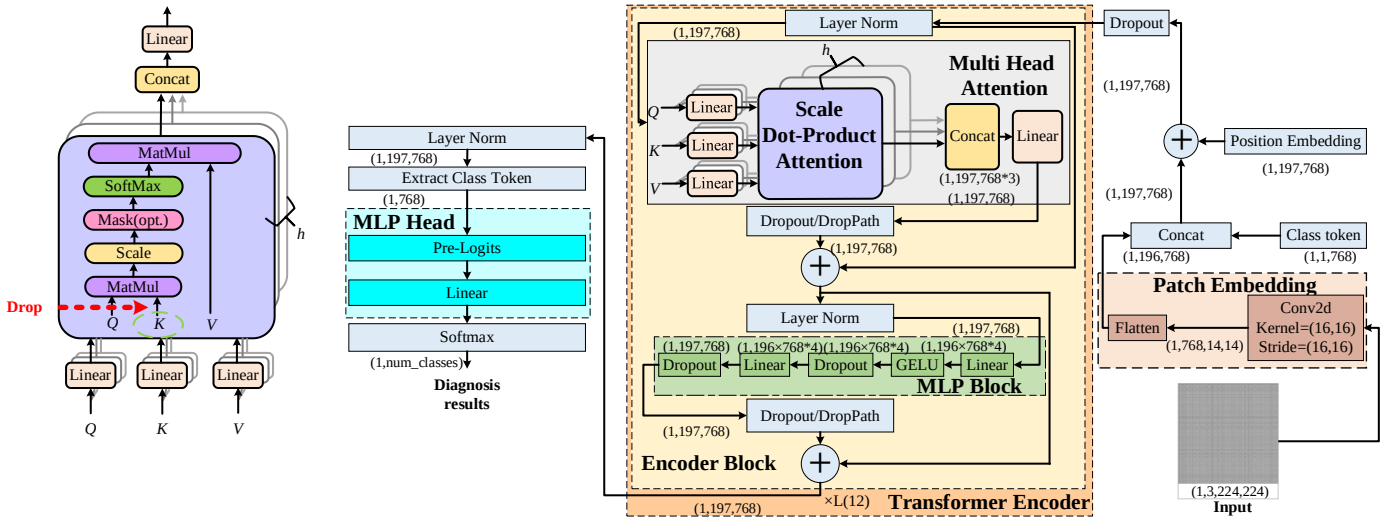
The attention weight matrix is derived in part from the mask matrix. As a result, the following is how each patch's output within the attention layer is calculated:

$$o = \sum_{j=1}^{n_h n_w}(\frac{m_j p_j}{\sum_{j=1}^{n_h n_w} m_j p_j})\, v_j = \sum_{j=1}^{n_h n_w} p_j\, v_j \tag{17}$$

$$p_j = \frac{exp(\frac{qk_j^T}{scale})}{\sum_{j=1}^{n_h n_w} exp(\frac{qk_j^T}{scale})} = \frac{exp(m_j+\frac{qk_j^T}{\sqrt{d_k}})}{\sum_{j=1}^{n_h n_w} exp(m_j+\frac{qk_j^T}{\sqrt{d_k}})} \tag{18}$$

where, q_i stands up query of i^{th} patch, k_i corresponds key of j^{th} patch, v_j represents value of j^{th} patch.

During training, the fixed probability is replaced with the number of dropped keys decreases as the layers deepen. The enhanced details and the structure of DViT are illustrated in Fig. 2.



(a) (b)

Fig.2. The details of the improved ViT.

3. The proposed method

The proposed URP-CVAE-MGAN method combined with DViT for diagnosing gear faults is illustrated in Fig.3.

Step 1 Gear signal collection and normalization: One-dimensional vibration signals for gear in health and multi-fault states are collected using sensors during the gear fault experiment. The signals, which generally vary in scale, are normalized to improve diagnostic efficiency and to handle various feature types.

Step 2 Signal transforming into the recursive image: The one-dimensional vibration signals for gear, both in health and multi-fault states, are transformed into two-dimensional recursive images using URP. This helps remain more fault-related information and enhances feature extraction.

Step 3 Image sets expansion: The URP images are divided into a training set, which is used to train the proposed CVAE-MGAN model, generating an adequate number of fault samples for model training.

Step 4 Fault diagnosing: The generated samples, along with the original training set, are split in a 7:2:1 ratio into the training set, validation set, and testing set. These sets are then input into the DViT model for training and validation. The optimal model is saved and employed in the testing set, ultimately outputting the diagnostic results.

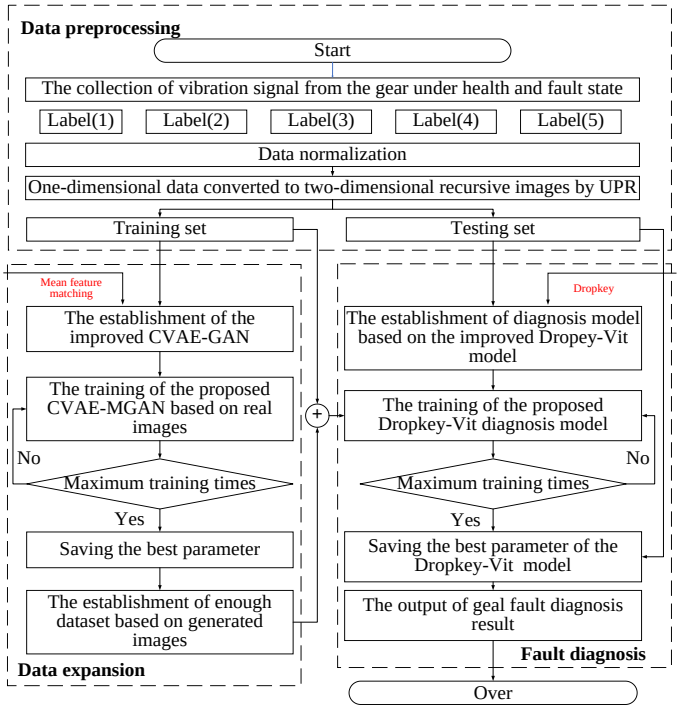


Fig.3. The implementation details of the URP-CVAE-MGAN DViT method.

4. Experiment

To assess the performance and superiority of the proposed method, two comprehensive case studies are conducted. The method is implemented on a computing platform equipped with an Intel Core i5-14400F 10-core CPU operating at 2.5GHz, an NVIDIA GeForce RTX 4060Ti GPU, and utilizing software such as Matlab R2022a, Python 3.8.

4.1. Case1: analysis of the SEU datasets

1) Experiment details and dataset preprocessing

The experiment data are obtained from a gear fault experiment conducted at Southeast University (SEU), and the related details of the experiment are demonstrated in Fig.4 [44].

The gear fault types are labeled as 1, 2, 3, 4 and 5 corresponding to Health, Chipped, Missing, Cracked, and Worn respectively. They primarily occur in parallel-axis gearboxes. The sampling frequency is 5120Hz, and the gear operates under a 20HZ-0V condition with an input shaft speed of 1200 rpm and a torque of 0 Nm. Three-channel signals are collected using 608A11 sensors and the second channel signal is used in this study. For each gear state, the vibration signals are normalized based on 784000 sampling points.

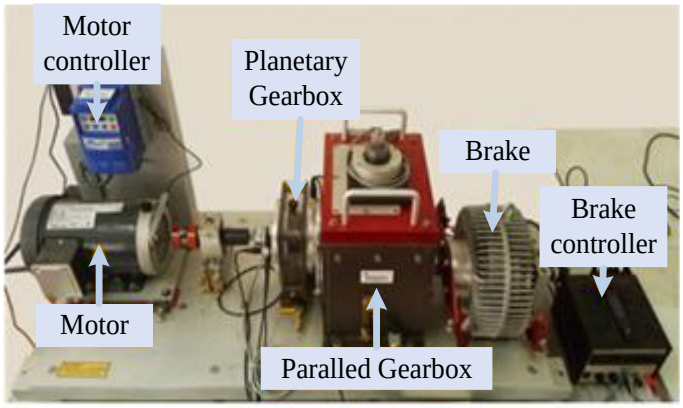


Fig.4. The specific setup for the experiment about the gearbox dataset.

2) Fault feature enhancement based on URP

The one-dimensional normalized vibration signals under different fault states are transformed into two-dimensional grey-scale images using URP. This conversion improves processing speed and highlights key features. Where, according to the gear input shaft speed and sampling frequency, every 784 points, containing at least two fault features, are converted into a URP image, resulting in a total of 1000 samples

3）Sample generation based on the URP-CVAE-MGAN

a) Creation of imbalanced and small sample sets

In practical applications, gear faults are rare, with varying types and degrees of severity. The number of samples in the healthy state often far exceeds the number of faulty samples, leading to an imbalanced dataset. The URP-CVAE-MGAN is used to generate additional samples to address the small and imbalanced sample problem. The accuracy of fault diagnosis is verified by Dropkey ViT. In this part, the creation of small and imbalanced sample sets is illustrated in the Table. 1.

Table 1. The creation of sample sets.

	Sample set	Training sample					P	Testing sample
		Labels						
		1	2	3	4	5		
Imbalanced sample	1	800	800	800	800	800	100%	200
	2	800	600	600	600	600	75%	200
	3	800	400	400	400	400	50%	200
	4	800	200	200	200	200	25%	200
Small sample	5	500	500	500	500	500	100%	200
	6	300	300	300	300	300	100%	200
	7	100	100	100	100	100	100%	200

Note: P represents the degree of imbalanced samples. Its definition is the proportion of small samples to normal samples.

b) Parameter setting of the proposed generated model

The CVAE-MGAN model parameters are set as follows, the batch size is 128, the noise vector is 100, the epoch is 500, the learning ratio is 0.0001, and Adam is selected as the optimized function. The iteration process is drawn in Fig.5.

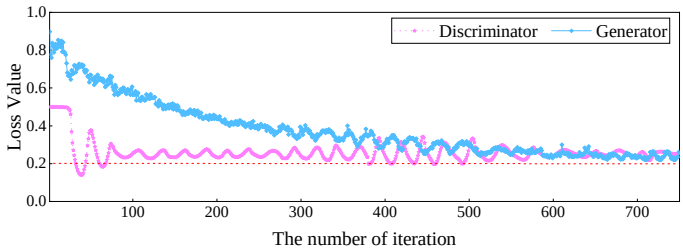
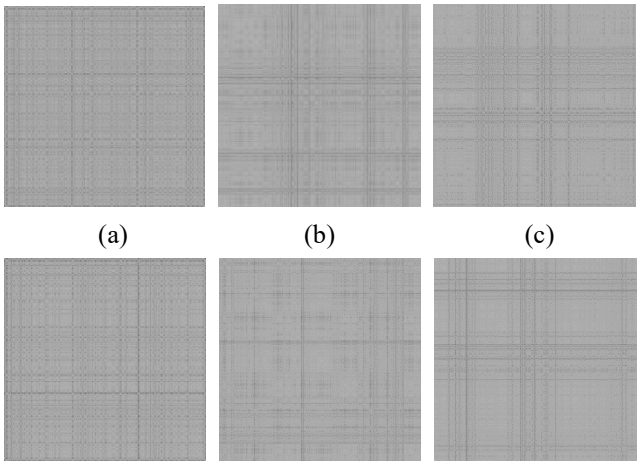


Fig.5. The training details of the CVAE-MGAN model.

From Fig.5, the losses of both the generator and the discriminator gradually decrease as iterations increase. By the 500th iteration, the generator's loss approaches 0.4. while the discriminator's loss stabilizes near 0.2, both remaining below 0.5. These results indicate that the generator can produce fault samples with probability distribution closely resembling real samples after multiple adversarial iterations. The generated samples using the best-generation model are shown in Fig.6.



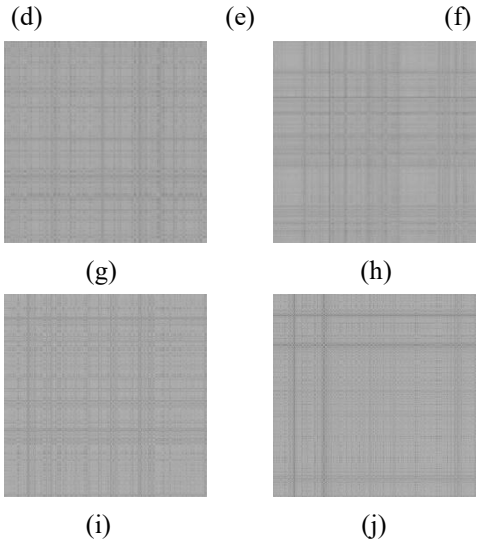


Fig.6. URP diagrams for five types of gear faults based on the CVAE-MGAN for generated and original samples. (a)-(c) Label 1, 2, 3 from the original samples. (d)-(f) Label 1, 2, 3 from the generated samples. (g)-(h) Label 4,5 from the original samples. (i)-(j) Label 4,5 from the generated samples.

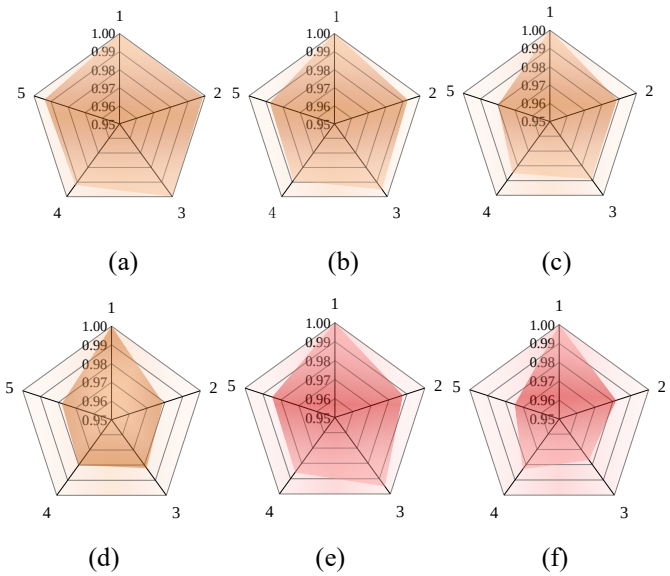
URP shows the periodicity, stability, and complexity of gear vibration signals, as well as fault state information. In healthy states, it exhibits a regular and uniform grid-like structure, reflecting strong signal periodicity and repeatability due to proper gear meshing and minimal dynamic disturbances. Under faulty states, the overall regional density becomes uneven as faults disrupt the signal's periodicity. For example, in the chipped state, the upper left corner of the diagonal retains a uniform grid structure, but the stripes in the diagonal area appear slightly blurred due to irregular high-frequency components introduced by gear damage, reducing periodicity and signal repeatability. In the missing state, symmetry is clearly broken, and the grid structure becomes highly irregular, as imbalanced gear meshing causes irregular impacts, almost eliminating periodicity. In the cracked state, a fuzzy gridline structure emerges, indicating gear operation instability. In the worn state, the overall structure becomes relatively chaotic, with the introduction of low-frequency vibration components disrupting signal stability and enhancing nonlinearity.

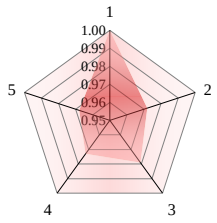
After expanding the original samples using CVAE-MGAN, the generated samples show a high degree of structural similarity to the original samples. The generated healthy state samples retain a regular and uniform grid-like structure. In the chipped state, while the distribution of points in the diagonal area of the generated samples is not as clear as that in the original, it still reserves a somewhat grid-like pattern resembling the original state. In the missing state, the generated samples exhibit earlier concentrated vertical and horizontal line densities, maintaining an irregular point distribution. The generated cracked state samples reflect the same fuzzy network-like structure as the original samples. Similarly, the generated worn state samples preserve the chaotic structure seen in the original, especially in the upper left region of the diagonal. Overall, the CVAE-MGAN model demonstrates strong performance in sample feature extraction. The generated samples maintain the essential structure and key features of the original samples with only minor differences in certain states. This fully confirms the effectiveness of the CVAE-MGAN model in expanding gear samples under small sample imbalance conditions, ensuring high similarity and accuracy.

c) Analysis under imbalanced and small sample conditions

The DViT parameters are configured as follows: batch size is sixteen, learning rate is 0.001, and weight decay is zero. The optimizer used is SGD, with 50 epochs. Patch Size is 16×16, while the Layers parameter is 12. The Hidden size is 768. MLP size is 3072, and the number of attention heads is 12. Additionally, the multi-head attention mechanism incorporates a novel Dropkey regularization technique for the first time.

Based on imbalanced sample sets detailed in Table 1, the CVAE-MGAN is trained on seven sample sets to address the imbalanced and small sample issues, splitting the dataset into a 7:2:1 ratio for training, validation, and testing. The accuracy of the DViT diagnosis model is evaluated using an adequate number of samples, as illustrated in Fig.7.





(g)

Fig.7. Accuracy in identifying faults under imbalanced sample sets and small sets based on case 1. (a) Set1. (b) Set2. (c) Set3. (d) Set4. (e) Set5. (f) Set6. (g) Set7.

d) Comparative analysis with other data generation methods

The performance of the proposed CVAE-MGAN is also evaluated using five indicators: Root Mean Squared Error (RMSE), Structural Similarity (SSIM) [45], Peak Signal-to-Noise Ratio (PSNR) [46], Fréchet Inception Distance (FID), Inception Score (IS) [47]. Several widely used, representative, and closely related generative models are used as comparative methods including CVAE-GAN (which adds conditional variables to the combination of VAE and GAN), VAE-GAN (which is the combination of VAE and GAN), VAE, GAN, and CGAN (which introduces conditional variables into GAN) [48-51]. These models have demonstrated strong performance in data generation and have been successfully applied to solve gear fault diagnosis problems under imbalanced and small sample conditions. The mean results for the five indicators between the original and the generated images are summarized in Table 2.

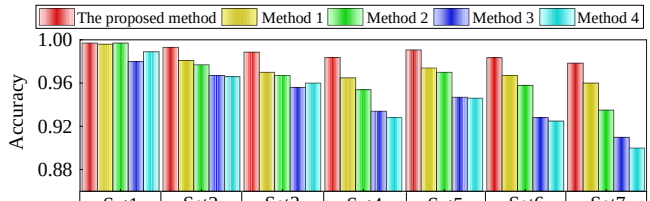
Table 2. The comparison results for different generation methods based on Case1.

Methods	RMSE	SSIM	PSNR	FID	IS
The proposed method	2.9344	0.9410	38.7803	0.0008	3.0140
CVAE-GAN	3.1467	0.9341	35.7896	0.0103	2.7895
VAE-GAN	3.2678	0.9067	30.8970	0.1090	2.2345
VAE	4.7896	0.9134	20.9809	0.1134	0.8890
GAN	3.5689	0.8990	22.3423	0.1506	0.9998
CGAN	3.4896	0.9209	33.4576	0.0388	2.5689

From Table 2, it is evident that the samples generated by the CVAE-MGAN model outperform those produced by the CVAE-GAN, VAE-GAN, VAE, GAN, and CGAN models across all five evaluation indicators. The generated samples from the CVAE-MGAN contain less noise, convey more useful information, and exhibit higher quality and diversity. They also show a closer distribution to the original images, with greater similarity. The effectiveness of introducing conditional variables and the mean feature difference function into VAE and GAN, as well as combining VAE and GAN, is fully verified. Additionally, the sample generation capability of the CVAE-MGAN model demonstrates a clear advantage over other Conditional Generative Adversarial Networks (CGAN) methods. The experiment confirms that the CVAE-MGAN model effectively addresses the challenges posed by imbalanced and small samples.

e) Comparison with other diagnosis models

The comparison includes the ViT before improvement, MLP-Mixer, AlexNet, and ResNet, which are referenced as Method 1, 2, 3, and 4, respectively [52-54]. The samples are divided into training, validation, and testing sets following a 7:2:1 ratio across all seven sets (Set1 to Set7). The average diagnostic accuracy of each model is shown in Fig.8.



	Set1	Set2	Set3	Set4	Set5	Set6	Set7
The proposed method	0.9970	0.9930	0.9886	0.9838	0.9908	0.9834	0.9784
Method 1	0.9960	0.9810	0.9700	0.9650	0.9770	0.9670	0.9600
Method 2	0.9970	0.9770	0.9670	0.9540	0.9700	0.9580	0.9350
Method 3	0.9800	0.9670	0.9560	0.9340	0.9470	0.9280	0.9100
Method 4	0.9890	0.9660	0.9600	0.9280	0.9460	0.9250	0.9000

Fig.8. The diagnosis accuracy from different models.

From Fig.8, the proposed method consistently maintains high accuracy, achieving at least 97.84%, even under conditions of extreme imbalance or scarcity of samples. Compared to the original ViT, DViT captures more fault information in limited sample conditions, resulting in a 1.784% improvement in diagnostic accuracy. The highest improvement, relative to other methods, reaches 7.84%. In contrast, the diagnostic performance of other methods is heavily influenced by sample conditions. Even when additional samples are generated, these models overly rely on the original data, limiting their fault-detection capabilities. As a result, their accuracy remains around 90% under extremely small and imbalanced conditions.

Five evaluation indicators—*Accuracy*, *Precision*, *Recal*l, and *F1*-score—are used to further compare the method's performance with those approaches [55]. Fig. 9 presents the final comparative results.

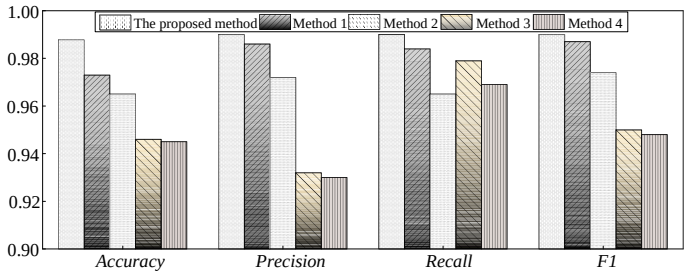


Fig.9. The performance comparison results for all methods.

From Fig.9, the DViT exhibits the highest accuracy, precision, recall, and F1 score among all methods, fully validating the model's diagnosis performance. For example, *Precision* improves by 0.4% compared to the original ViT model and by up to 6% compared to other methods. The effectiveness of the proposed diagnosis model has been thoroughly confirmed.

f) Ablation Studies

Table 3 The experiment results of ablation studies

Method	Accuracy/(%)
WT-CVAE-MGAN-DVit	97.23
URP-VAE-MGAN-DVit	97.86
URP-CVAE-GAN-DVit	98.02
URP-CVAE-MGAN-Vit	97.70
URP-CVAE-MGAN-Dvit	99.08

To verify the contribution of each component in the proposed method, ablation experiments are conducted, with results shown in Table 3:

(1) WT-CVAE-WGAN-DVit: Converts 1D gear signals into common time-frequency maps based on wavelet transform

(2) URP-VAE-WGAN-DVit: Omits conditional variables in the autoencoder

(3) URP-CVAE-GAN-Dvit: Retains the original GAN loss function without switching to the mean feature difference function.

(4) URP-CVAE-WGAN-Vit: Retains Dropout instead of the Dropkey regularization method in Vision Transformer.

(5) URP-CVAE-WGAN-Dvit: The method proposed in this article

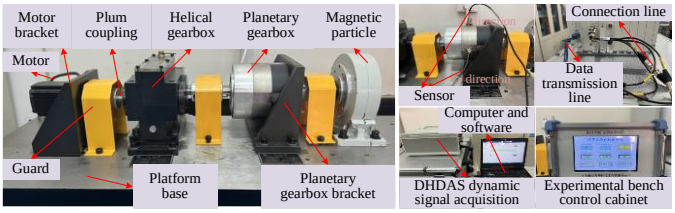
As shown in Table 3, traditional time-frequency maps derived from WT capture only local features in time and frequency, leading to a 1.85% reduction in classification accuracy. This demonstrates the importance of URP in extracting the nonlinear features from gear fault signals. When conditional variables are excluded from CVAE-MGAN, the generator fails to associate specific fault types with conditional information, reducing the diversity of generated samples and their correlation with specific fault types, ultimately resulting in a 1.22% decrease in diagnostic accuracy. This indicates the necessity of conditional variables in maintaining the correlation between generated samples and actual fault conditions. The removal of the mean average feature difference loss function further increases the discrepancy between some generated samples and actual fault samples, reducing sample quality and decreasing diagnostic accuracy by 1.06%. Retaining the original Dropout regularization method weakens the model's ability to capture comprehensive fault features, resulting in a 1.38% decrease in accuracy. These findings demonstrate that Dropkey regularization is more effective than traditional Dropout for representing global information.

The ablation experiments confirm that each component of the proposed method plays a critical role in enhancing performance. URP improves nonlinear feature extraction, conditional variables ensure sample specificity, the mean feature difference loss maintains sample quality, and Dropkey improves global feature representation.

4.2. Case2: analysis of the SUT datasets

1) The experiment's detail and the dataset preprocessing

The Shenyang University of Technology (SUT) datasets are obtained from a gear fault experiment conducted on the model LY-SCL-04 experiment bench, as illustrated in Fig.10.



(a) (b)

Fig.10. The specific setup for the experiment about planetary parallel axis gearbox faults. (a) Basic structure (b) Details.

The different fault types and severities, including cracked faults of 3mm and 6mm and light and severe corrosion, are labeled as 2, 3, 4, and 5, respectively. Health is labeled as 1. The faults occur in the sun gear of the ZLS120-5-S planetary gearbox. The motor speed is controlled by the HZXT-008 system, ranging in [2000, 4000]rpm with increments of 100rpm, and loads are set at 0%, 20%, and 50%. The faulty vibrational signals from x and y direction of faulty gear are collected using the Jiangsu DonghHua DHDAS dynamic system and two CA-

YD-182 piezoelectric acceleration sensors. Based on prior experience, a sampling frequency of 5120 Hz is used to capture fault characteristics effectively adhere to the sampling theorem, and prevent signal distortion. To avoid external influences, the experiment is conducted in a soundproof and shockproof environment. Similarly, the second channel signal is applied in this part. The vibrational signals are normalized based on 784000 sampling points for each gear state.

3）Sample generation based on the URP-CVAE-MGAN

a) The setting of the imbalanced and small sample and the parameter setting of the proposed generated model

In this part, small and imbalanced samples occur not only with different faulty types but also with different faulty severities. Given the same number of labels as in Case 1, the same sample sets are employed. To ensure accuracy, and repeatability and to more comprehensively evaluate the method's performance, the same parameters are used as in case 1. Finally, URP images are shown in Fig.11 for original and generated samples after obtaining the best-generation model under different faulty types and severities.

From Fig.11, distinct structural patterns are observed for different gear fault types, while similar structures appear across varying fault severities. In the healthy state, a very regular and uniform grid-like structure is displayed, which is replicated in the generated sample. In the cracked state, the 3mm crack shows an uneven distribution between the stripes, while the 6mm crack exhibits denser, more blurred horizontal and vertical lines. The generated samples accurately reflect the fuzzy point distribution and capture the periodic disturbances caused by varying degrees of crack faults. In the corrosion state, the structure remains mostly uniform, except for a blurred region above the center of the symmetrical point. As the corrosion deepens, the grid-like structure becomes more indistinct. The generated samples of the lightly corroded state maintain a relatively regular grid-like structure, with the concentration of blurring in the lower right corner of the diagonal. In the heavy corrosion state, the grid structure in the generated samples deteriorates significantly, indicating the increased complexity of the fault. Although subtle differences exist, overall, the generated samples still preserve the key structural features and patterns of the original samples, demonstrating the generative mode's high effectiveness in sample generation.

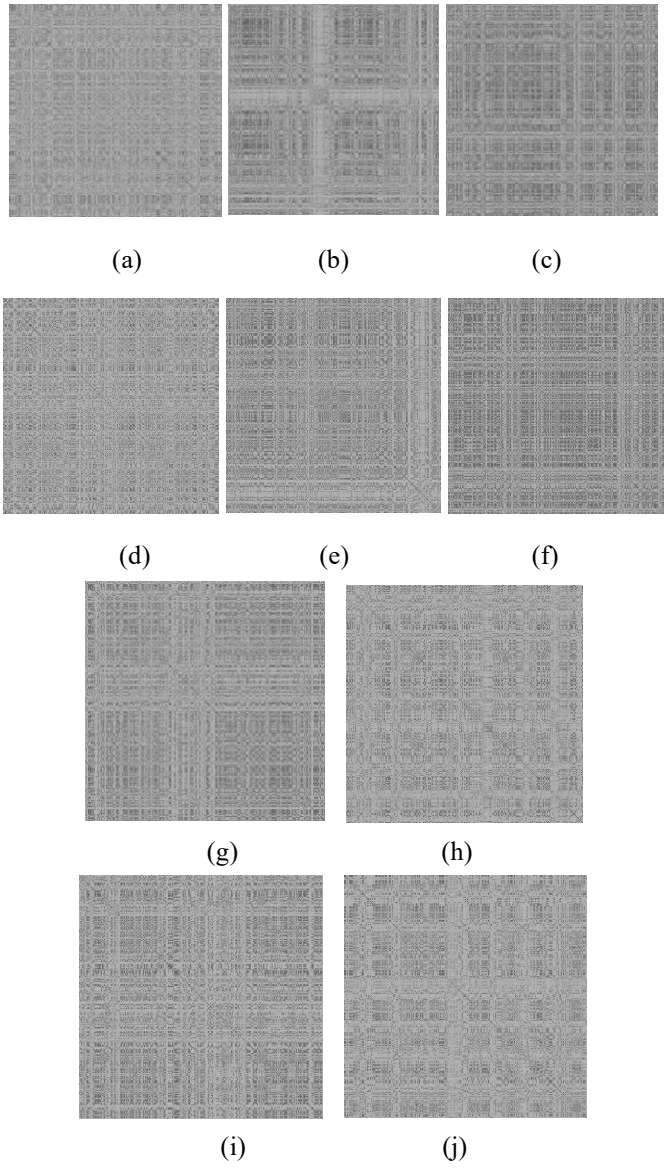


Fig.11. URP diagrams for five types and degrees of gear faults. (a)-(c) Label 1, 2, 3 from the original samples. (d)-(f) Label 1, 2, 3 from the generated samples. (g)-(h) Label 4,5 from the original samples. (i)-(j) Label 4,5 from the generated samples.

b) Analysis under imbalanced and small sample conditions

Based on the imbalanced sample sets details in Table I, the trained CVAE-MGAN model is applied to supplement imbalanced samples. Sufficient samples are input to the Dropkey-ViT diagnosis model to assess the accuracy of fault diagnosis, using the same parameters, training, validation, and testing sets as Case1. The final diagnosis accuracy across seven sample sets is shown in Fig.12.

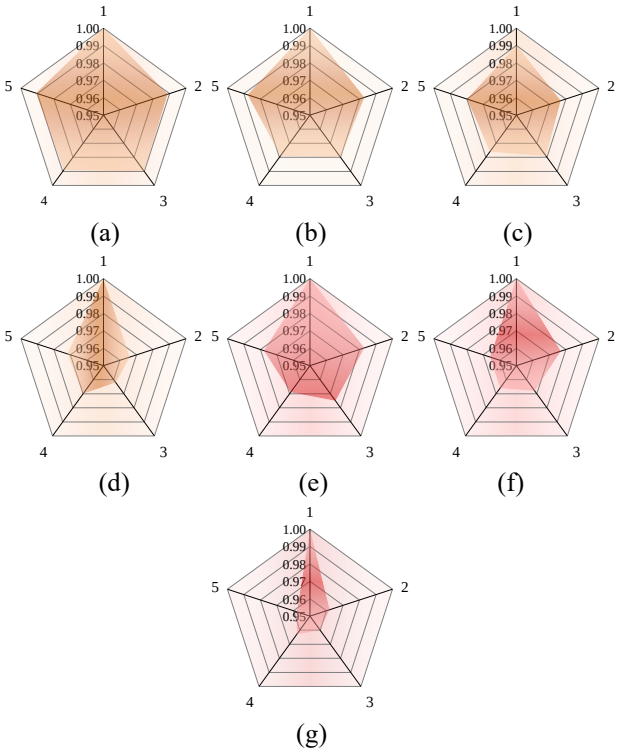


Fig.12. Fault identification accuracy of imbalanced sample sets and small sets based on case 2. (a) Set1. (b) Set2. (c) Set3. (d) Set4. (e) Set5. (f) Set6. (g) Set7.

From Fig.12, Label 1 consistently achieves 100% accuracy. While the diagnosis accuracy remains high across all data states, a slight decrease is observed as the sample set becomes increasingly imbalanced and reduced. Nevertheless, the diagnostic accuracy remains at or above 96%.

c) Comparison with other data generation methods

The average results for five indicators RMSE, SSIM, PSNR, FID, and IS are shown in Table 4 between the original images and the generated images. These results are based on the ablation experiments and comparison models.

Table 4. The comparison results for different generation methods based on Case2.

Method	RMSE	SSIM	PSNR	FID	IS
The proposed method	3.6997	0.8821	36.7676	0.0162	3.4043
CVAE-GAN	4.0090	0.8090	34.0980	0.0189	3.0908
VAE-GAN	4.2343	0.7589	33.4454	0.2367	2.7912
VAE	5.7789	0.7092	30.6733	0.3567	2.5547
GAN	5.0909	0.7123	31.9998	0.2498	2.0089
CGAN	3.8003	0.8222	32.0078	0.0879	2.9098

From Table 4, it is evident that the proposed method outperforms other models across all indicators except for FID, where the CVAE-MGAN model shows a slightly higher value than the CVAE-GAN. This suggests that the introduction of the mean feature difference function improves the sample distribution for the SUT dataset. Although the SSIM only reaches 0.8821, the highest among the compared methods. It indicates a discrepancy between the self-made testing system dataset and publicly available datasets, likely due to differences in experimental environments. Nonetheless, the results confirm the proposed strategy effectively generates high-quality samples. The CVAE-MGAN model demonstrates strong capabilities in addressing issues related to unbalanced samples and small samples.

d) Comparison with other diagnosis models

The same comparison methods are applied to verify the proposed method's diagnosis performance across various fault types, particularly at different fault severities. Using Set3 from imbalanced samples and Set6 from small samples as examples, the confusion matrices in Fig.13 and Fig.14 illustrate the results.

The diagnosis accuracy of the proposed method consistently surpasses the comparison methods, with only a few misclassifications, mostly among different fault severities (Fig.13). This may be attributed to limitations in experimental equipment and sensor performance in the self-designed experiments compared to publicly available datasets. The diagnosis results based on DViT show minimal misclassifications, highlighting its improved global fault information capture and more accurate diagnosis. By contrast, the last three methods exhibit more misclassifications. In Fig.14, the proposed method achieves the highest accuracy across all faulty types and severities, maintaining at least 97.0% accuracy even with small sample conditions. The strong diagnostic performance of the proposed method under unbalanced and small sample conditions has been once again verified.

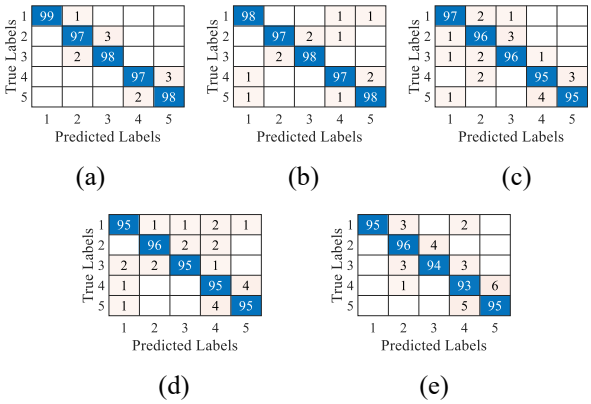


Fig.13. Test results from different methods for the imbalanced sample set. (a) The proposed method. (b) Method 1. (c) Method 2. (d) Method 3. (e) Method 4.

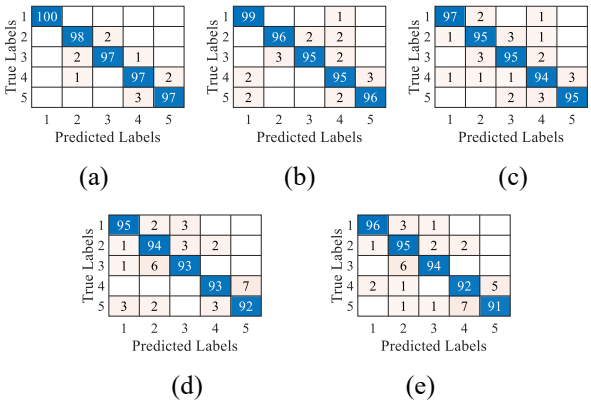


Fig.14. Test results from different methods for a small sample set. (a) The proposed method. (b) Method 1. (c) Method 2. (d) Method 3. (e) Method 4.

5. Conclusions

The URP-CVAE-MGAN-DViT method is introduced in this article, and the effectiveness of the proposed method is verified with different imbalances and small sample conditions across two datasets containing different gear faulty types or severities. The main conclusions drawn from the experimental results are as follows.

(1) The proposed method outperforms other generative models in terms of the quality and diversity of generated samples. The similarity between generated samples and original samples, as measured by RMSE, SSIM, PSNR, and FID, is the highest level among all methods. Even under extremely unbalanced conditions, the proposed method maintains a diagnostic accuracy of 97%, fully demonstrating its effectiveness in addressing sample imbalance problems.

(2) The proposed method outperforms the performance than other diagnosis models as evidenced by indicators $Accuracy$, $Pr\,e\,cision$, $Re\,c\,all$, and $F1$, all indicators have improved to varying degrees, fully confirming the effectiveness and superiority of this method in comprehensively capturing features compared to the previous model.

(3) The ablation studies verify the importance of each component of the proposed method. The threshold-free recursive graph analysis, conditional variables, and mean feature loss function all contribute significantly to the model's performance. These components work together to improve both the quality and diversity of sample generation, improving the ability to capture fault characteristic information, but also enhancing the diagnostic ability of the model when dealing with imbalanced samples, demonstrating its potential for practical applications.

The proposal is also applicable to other rotating machines. But there are also some limitations in this study, the imbalance between different types has not been fully considered yet. Due to constraints such as computational resources, existing research focus, and experimental conditions, we have only chosen a more classical method for comparison. Future work will consider the imbalance between different categories and explore in depth the applicability and potential of more advanced models in this field.

References

1. Liu R, Yang B, Zio E, Chen X. Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing 2018; 108: 33-47, https://doi.org/10.1016/j.ymssp.2018.02.016.
2. Ning J, Chen Z, Wang Y, Wang Y, Li F, Zhai W. Vibration feature of spur gear transmission with non-uniform depth distribution of tooth root crack along tooth width. Engineering Failure Analysis 2021; 129: 105713, https://doi.org/10.1016/j.engfailanal.2021.105713.
3. Parey A, Pachori R. Variable cosine windowing of intrinsic mode functions: Application to gear fault diagnosis. Measurement 2012; 45(3): 415-426, https://doi.org/10.1016/j.measurement.2011.11.001.
4. Praveenkumar T, Sabhrish B, Saimurugan M, Ramachandran K. Pattern recognition based on-line vibration monitoring system for fault diagnosis of automobile gearbox. Measurement 2018; 114: 233-242, https://doi.org/10.1016/j.measurement.2017.09.041.
5. Li D Z, Wang W, Ismail F. An enhanced bispectrum technique with auxiliary frequency injection for induction motor health condition monitoring. IEEE Transactions on Instrumentation Measurement 2015; 64(10): 2679-2687, https://doi.org/10.1109/TIM.2015.2419031.
6. Chaari F, Bartelmus W, Zimroz R, Fakhfakh T, Haddar M. Gearbox vibration signal amplitude and frequency modulation. Shock and Vibration 2012; 19(4): 635-652, https://doi.org/10.3233/SAV-2011-0656.
7. Yu K, Lin T, Ma H, Li H, Zeng J. A combined polynomial chirplet transform and synchro extracting technique for analyzing nonstationary signals of rotating machinery. IEEE Transactions on Instrumentation and Measurement 2019; 69(4): 1505-1518, https://doi.org/10.1109/TIM.2019.2913058.

8. Banerjee M, Pal N. Feature selection with SVD entropy: Some modification and extension. Information Sciences 2014; 264: 118-134, https://doi.org/10.1016/j.ins.2013.12.029.
9. Li H, Liu T, Wu X, Chen Q. A bearing fault diagnosis method based on enhanced singular value decomposition. IEEE Transactions on Industrial Informatics 2020; 17(5): 3220-3230, https://doi.org/ 10.1109/TII.2020.3001376.
10. Peng Z, Chu F. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. Mechanical Systems and Signal Processing 2004; 18(2): 199-221, https://doi.org/10.1016/S0888-3270(03)00075-X.
11. Lei Y, Lin J, He Z, Zuo M. A review on empirical mode decomposition in fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing 2013; 35(1-2): 108-126, https://doi.org/10.1016/j.ymssp.2012.09.015.
12. Zhao W, Liu J, Zhao W, Zheng Y. An investigation on vibration features of a gear-bearing system involved pitting faults considering effect of eccentricity and friction. Engineering Failure Analysis 2022; 131: 105837, https://doi.org/10.1016/j.engfailanal.2021.105837.
13. Wang J, Li S, Xin Y, An H. Gear fault intelligent diagnosis based on frequency-domain feature extraction. Journal of Vibration Engineering & Technologies 2019; 7(2): 159-166, https://doi.org/ 10.1007/s42417-019-00089-1.
14. Wang H, Zhou Z, Zhang L, Yan R. Multiscale deep attention Q network: A new deep reinforcement learning method for imbalanced fault diagnosis in gearboxes. IEEE Transactions on Instrumentation and Measurement 2024; 73: 3503512, https://doi.org/10.1109/TIM.2023.3338664.
15. Ye Z, Yue S, Yang P, Zhou R, Yu J. Deep morphological shrinkage convolutional auto-encoder-based feature learning of vibration signals for gearbox fault diagnosis. IEEE Transactions on Instrumentation and Measurement 2024; 73: 3510712, https://doi.org/10.1109/TIM.2024.3366570.
16. Zhang J, Xu B, Wang Z, Zhang J. An FSK-MBCNN based method for compound fault diagnosis in wind turbine gearboxes. Measurement 2021; 172: 108933, https://doi.org/10.1016/j.measurement.2020.108933.
17. Jia S, Wang P, Jia P, Hu S. Research on data augmentation for image classification based on convolution neural networks. In 2017 Chinese automation congress (CAC) IEEE 2017; 4165-4170, https://doi.org/ 10.1109/CAC.2017.8243510.
18. Yu K, Lin T, Ma H, Li X, Li X. A multi-stage semi-supervised learning approach for intelligent fault diagnosis of rolling bearing using data augmentation and metric learning. Mechanical Systems and Signal Processing 2021; 146: 107043, https://doi.org/10.1016/j.ymssp.2020.107043.
19. Ma L, Ding Y, Wang Z, Wang C, Ma J, Lu C. An interpretable data augmentation scheme for machine fault diagnosis based on a sparsity-constrained generative adversarial network. Expert Systems with Applications 2021; 182: 115234, https://doi.org/10.1016/j.eswa.2021.115234.
20. Wang J, Li S, Han B, An Z, Bao H, Ji S. Generalization of deep neural networks for imbalanced fault classification of machinery using generative adversarial networks. IEEE Access 2019; 7: 111168-111180, https://doi.org/10.1109/ACCESS.2019.2924003.
21. Moreno-Barea F J, Jerez J M, Franco L. Improving classification accuracy using data augmentation on small data sets. Expert Systems with Applications 2020; 161: 113696, https://doi.org/10.1016/j.eswa.2020.113696.
22. Zhao D, Liu S, Gu D, Sun X, Wang L, Wei Y, Zhang H. Enhanced data-driven fault diagnosis for machines with small and unbalanced data based on variational auto-encoder. Measurement Science and Technology 2019; 31(3): 035004, 2019 https://doi.org/ 10.1088/1361-6501/ab55f8.
23. Khan M A, Asad B, Vaimann T, Kallaste A, Pomarnacki R, Hyunh V K, Improved fault classification and localization in power transmission networks using VAE-generated synthetic data and machine learning algorithms. Machines 2023; 11(10): 963, https://doi.org/10.3390/machines11100963.
24. Nazabal A, Olmos P M, Ghahramani Z, Valera I. Handling incomplete heterogeneous data using vaes. Pattern Recogn 2020; 107: 107501, https://doi.org/ https://doi.org/10.1016/j.patcog.2020.107501.
25. Yao Y, Wang H, Li S, Liu Z, Gui G, Dan Y, Hu J. End-to-end convolutional neural network model for gear fault diagnosis based on sound signals. Applied Sciences 2018; 8(9): 1584, https://doi.org/ 10.3390/app8091584.
26. Ghulanavar R, Dama K, Jagadeesh A. Diagnosis of faulty gears by modified AlexNet and improved grasshopper optimization algorithm (IGOA). Journal of Mechanical Science and Technology 2020; 34: 4173-4182, https://doi.org/ 10.1007/s12206-020-0909-6.
27. Yang G, Wei Y, Li H. Acoustic diagnosis of rolling bearings fault of CR400 EMU traction motor based on XWT and GoogleNet. Shock

and Vibration 2022; 2022: 2360067, https://doi.org/10.1155/2022/2360067.

28. Hu H, Feng F, Jiang F, Zhou X, Zhu J, Xue J, Jiang P, Li Y, Qian Y, Sun G, Chen C. Gear fault detection in a planetary gearbox using deep belief network. Mathematical Problems in Engineering 2022; 2022: 9908074, https://doi.org/10.1155/2022/9908074.
29. Shi H, Huang C, Zhang X, Zhao J, Li S. Wasserstein distance based multi-scale adversarial domain adaptation method for remaining useful life prediction. Applied Intelligence 2023; 53(3): 3622-3637, https://doi.org/ 10.1007/s10489-022-03670-6.
30. Yan X, Liu Y, Jia M. Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions. Knowledge-Based Systems 2020; 193: 105484, https://doi.org/ https://doi.org/10.1016/j.knosys.2020.105484.
31. Hu A, Sun J, Xiang L, Xu Y. Rotating machinery fault diagnosis based on impact feature extraction deep neural network. Measurement Science and Technology 2022; 33(11): 114004, https://doi.org/10.1088/1361-6501/ac7eb1.
32. Tang X, Xu Z, Wang Z. A novel fault diagnosis method of rolling bearing based on integrated vision transformer model. Sensors 2022; 22(10): 3878, https://doi.org/10.3390/s22103878.
33. Zhou Z, Ai Q, Lou P, et al. A novel method for rolling bearing fault diagnosis based on gramian angular field and CNN-ViT. Sensors 2024; 24(12): 3967, https://doi.org/10.3390/s24123967
34. Bai R, Meng Z, Xu Q, Fan F. Fractional fourier and time domain recurrence plot fusion combining convolutional neural network for bearing fault diagnosis under variable working conditions. Reliability Engineering & System Safety 2023; 232: 109076, https://doi.org/ https://doi.org/10.1016/j.ress.2022.109076.
35. Liu X P, Xia L J, Shi J, Zhang L J, Bai L Y, Wang S P. A fault diagnosis method of rolling bearing based on improved recurrence plot and convolutional neural network. IEEE Sensors Journal 2023; 23(10): 10767-10775, https://doi.org/10.1109/JSEN.2023.3265409.
36. Yan J, Huang Y, Yuan S, Lu Y, Yu Z. Open‐circuit fault analysis and recognition in three‐level inverters based on recurrence plot and convolution neural network. International Transactions on Electrical Energy Systems 2023; 2023(1): 4755960, https://doi.org/10.1155/2023/4755960.
37. Kingma D P, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv 2013; 1312: 6114, https://doi.org/https://doi.org/10.48550/arXiv.1312.6114.
38. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In Advances in Neural Information Processing Systems 2012; 2672‐2680, https://doi.org/ 10.3156/JSOFT.29.5_177_2.
39. Larsen A B L, Sønderby S K, Winther O. Autoencoding beyond pixels using a learned similarity metric. arXiv preprint arXiv 2015; 1512: 09300, https://doi.org/ 10.48550/arXiv.1512.09300.
40. Bao J, Chen D, Wen F, Li H, Hua G. CVAE-GAN: Fine-grained image generation through asymmetric training. In Proceedings of the IEEE international conference on computer vision 2017; 2745-2754, https://doi.org/ 10.1109/ICCV.2017.299.
41. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J. An image is worth 16x16 words: Transformers for image recognition at scale. arxiv preprint arxiv: 2020, 2010: 11929, https://doi.org/ 10.48550/arXiv.2010.11929.
42. Yang N, Liu J, Zhao W Q, Tan Y T. Fault diagnosis of gear based on multi-channel feature fusion and Dropkey-Vision Transformer. IEEE Sensors Journal 2024; 24(4): 4758-4770, https://doi.org/10.1109/JSEN.2023.3344999.
43. Li B, Hu Y, Nie X, Han C, Jia X, Guo T, Liu L. DropKey. Presented at proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2023; 2023, https://doi.org/10.48550/arXiv.2208.02646.
44. Chen C, Shen F, Xu J, Yan R. Model parameter transfer for gear fault diagnosis under varying working conditions. Chinese Journal of Mechanical Engineering 2021; 34(1): 13, https://doi.org/10.1186/s10033-020-00520-9.
45. Meng Z, He H H, Cao W, Li J M, Cao L X, Fan J J, Zhu M, Fan F J. A novel generation network using feature fusion and guided adversarial learning for fault diagnosis of rotating machinery. Expert Systems with Applications 2023; 234: 121058, https://doi.org/10.1016/j.eswa.2023.121058.
46. Rathore M S, Harsha S P. Non-linear vibration response analysis of rolling bearing for data augmentation and characterization. Journal of Vibration Engineering & Technologies 2023; 11(5): 2109-2131, https://doi.org/10.1007/s42417-022-00691-w.
47. Zhang L, Duan L, Hong X, Liu X, Zhang X. Imbalanced data enhancement method based on improved DCGAN and its application. Journal of Intelligent & Fuzzy Systems 2021; 41(2): 3485-3498, https://doi.org/ 10.3233/JIFS-210843.

48. Zhang L, Zhang H, Cai G. The multiclass fault diagnosis of wind turbine bearing based on multisource signal fusion and deep learning generative model. IEEE Transactions on Instrumentation and Measurement 2022; 71: 1-12, 2022 https://doi.org/10.1109/TIM.2022.3178483.
49. Wang Y, Sun G, Jin Q. Imbalanced sample fault diagnosis of rotating machinery using conditional variational auto-encoder generative adversarial network. Applied Soft Computing 2020; 92: 106333, https://doi.org/10.1016/j.asoc.2020.106333.
50. Gao Y, Liu X, Xiang J. Fault detection in gears using fault samples enlarged by a combination of numerical simulation and a generative adversarial network. IEEE/ASME Transactions on Mechatronics 2021; 27(5): 3798-3805, https://doi.org/10.1109/TMECH.2021.3132459.
51. Zhou C, Wang H, Hou S, et al. A hybrid physics-based and data-driven method for gear contact fatigue life prediction. International Journal of Fatigue 2023; 175: 107763, https://doi.org/10.1016/j.ijfatigue.2023.107763.
52. Targ S, Almeida D, Lyman K. Resnet in resnet: Generalizing residual architectures. ArXiv Preprint ArXiv 2016; 1603: 08029, https://doi.org/10.48550/arXiv.1603.08029.
53. Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks. Advances in Nneural Information Processing Systems 2017; 6(60): 84-90, https://doi.org/10.1145/3065386.
54. Maaten L V D, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research 2008; 9(11): 2579-2605.
55. Chen X, Zhang B, Gao D. Bearing fault diagnosis base on multi-scale CNN and LSTM model, Journal of Intelligent Manufacturing 2021; 32: 971-987, https://doi.org/10.1007/s10845-020-01600-2.