IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 July 2024, accepted 2 August 2024, date of publication 5 August 2024, date of current version 13 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3439017

RESEARCH ARTICLE

Fault Diagnosis of Tractor Transmission System Based on Time GAN and Transformer

LIYOU XU, (Member, IEEE), GUODONG ZHANG, (Member, IEEE), SIXIA ZHAO, (Member, IEEE), YIWEI WU, (Member, IEEE), AND ZHIQIANG XI, (Member, IEEE)

Vehicle and Transportation Engineering Institute, Henan University of Science and Technology, Luoyang 471003, China
State Key Laboratory of Intelligent Agricultural Power Equipment, Luoyang 471003, China

Corresponding author: Sixia Zhao (zhao_sixia@haust.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFD2001203 and Grant 2022YFD2001201B, in part by the Central Plains Technology Leading Talent Support Program Project under Grant 244200510043, in part by Henan Provincial University Science and Technology Innovation Team Support Program under Grant 24IRTSTHN029, and in part by Henan Province Science and Technology Research Projects under Grant 222102110233.

ABSTRACT The transmission system of a tractor is a crucial component, so it is crucial to promptly identify and correctly diagnose faults in it. However, due to the limited samples of faults occurring during its operational processes, employing existing fault diagnosis methods directly yields unsatisfactory results. This paper proposes a fault diagnosis model combining Time Generative Adversarial Networks (Time GAN) and Transformer. To enhance diagnostic accuracy, we first employ Time GAN for data augmentation, addressing the issue of imbalanced fault samples in practical scenarios. Then, we integrate a Transformer network with improved multi-head self-attention mechanisms, leveraging the advantages of the Transformer's encoder-decoder architecture and attention mechanism to enhance diagnostic performance. Bearing data from Case Western Reserve University (CWRU) was used to validate the diagnostic performance of the proposed model, while gear data from an experimental rig built by the author was used to validate the model's generalization capability. Experimental results indicate that the accuracy reached 98.96% and 95.36% in CWRU Dataset and Self-made Dataset respectively. In strong noise environments, the accuracy remains above 93%. In conclusion, the diagnostic model presented in this paper can reliably diagnose tractor transmission system problems in few-sample conditions and noise environments compared to traditional machine learning models.

INDEX TERMS Transformer model, Time GAN, data augmentation, fault diagnosis.

I. INTRODUCTION

The tractor gearbox, as a critical component of the mechanical transmission system, plays a vital role in power transmission [1]. During operation, it constantly rotates and is subjected to variable speed loads and severe impacts, leading to damage and issues such as inner race fault in bearing, outer race fault, miss tooth, chipped tooth in the gearbox, and other faults. Statistics show that approximately 18% of tractor failures in agricultural operations are attributed to transmission system faults. These faults result in prolonged downtime, significantly affecting the harvesting and transportation of crops, thereby causing substantial losses in agricultural production. To mitigate the severe impact of these failures, timely and reliable condition monitoring and fault diagnosis of the tractor transmission system are essential.

Existing diagnostic methods can generally be classified into two categories. One category is non-intelligent fault diagnosis, which primarily relies on time-frequency domain processing of vibration signals [2], [3], and typically does not involve deep learning. These methods study fault signals using signal processing techniques such as wavelet analysis [4], empirical mode decomposition [5], et al, to extract fault characteristic frequencies. For instance, Zhang et al. [6] suggested a technique for feature extraction using a combination of time-domain and frequency-domain data

The associate editor coordinating the review of this manuscript and approving it for publication was Guillermo Valencia-Palomo.

with variational mode decomposition. Initially, the original vibration signals undergo variational mode decomposition to obtain modal functions. Subsequently, singular value decomposition is employed to further extract the modal features of the modal functions. These extracted modal features are combined with the time-domain and frequency-domain features of the original signals to generate hybrid features, and finally, SVDD kernel functions are used for parameter optimization. However, these methods have the following shortcomings, (1) Non-intelligent fault diagnosis methods rely on physical information and domain expertise for fault diagnosis, resulting in a relatively low degree of automation. (2) Massive fault data will not be effectively learned due to manual feature extraction. (3) Rich prior knowledge is necessary for diagnostic professionals to perform the feature extraction method [7].

The second category of fault diagnosis techniques, referred to as ''intelligent fault diagnosis,'' has also evolved and is maturing with the emergence of artificial intelligence. The majority of these techniques rely on deep learning, which can eliminate the need for specialist knowledge and manual feature extraction procedures. In intelligent fault diagnosis methods, Convolutional Neural Network (CNN)–based diagnostic approaches are commonly employed. For instance, a fault detection model based on Google Net and Gramian Angular Field (GAF) was presented by Huang et al. [8]. You et al. [9] introduced a fault detection model based on PCA-CNN. Chen et al. [10] combined Smoothed Pseudo Wigner-Ville Distribution (SPWVD) with CNN for fault diagnosis. Although the aforementioned studies have achieved good results, they are mostly constructed and validated on balanced datasets with rich typical fault information and sufficient healthy labeled data, overlooking the problem of sample class imbalance. Moreover, CNN has limitations in modeling dependencies over a wide range of low-level features [11]. The aforementioned intelligent fault diagnosis methods require a sufficient variety of fault types to achieve high accuracy. When the labeled sample size is adequate and balanced, the diagnostic accuracy of the network model is high. However, when labels are missing, the sample size is small, or fault samples are imbalanced, the network training becomes inadequate, leading to suboptimal accuracy and significant degradation in algorithm performance. In practical diagnostic processes, although a large amount of monitoring data for tractor transmission systems can be obtained through signal acquisition, most of this data represents normal conditions. The scarcity of fault data results in an imbalanced sample set, preventing the neural network from being fully trained and consequently affecting the subsequent diagnostic accuracy. Therefore, effectively diagnosing faults in tractor transmission systems using traditional diagnostic methods is extremely challenging.

The multidimensionality, class imbalance, and concurrency of fault data present three major challenges for fault diagnosis [12]. In fault diagnosis, the lack of sufficient sample data to train intelligent diagnostic models often results in sacrificing the accuracy of minority classes to achieve higher accuracy across the entire dataset, greatly reducing diagnostic accuracy. Therefore, many scholars have begun researching data augmentation methods to enhance data. Yu et al. presented a technique for sophisticated few-fault diagnosis in rotating machinery based on Convolutional Neural Networks (CNN), which utilizes Mechanism Characteristic Generation Models (MCGM) integrated with Generative Adversarial Networks (GAN) to generate virtual samples [13]. A gearbox defect diagnosis technique based on Wasserstein distance enhanced Auxiliary Classifier Generative Adversarial Network (AC-GAN) models and Bayesian optimization was presented by Li et al., which simultaneously optimizes the generator and discriminator through adversarial gaming mechanisms, significantly enhancing the model's generalization and fault feature extraction capabilities [14]. Qin et al. proposed a model named FBC-GAN, which stands for Frequency-Domain Bidirectional Long Short-Term Memory (Bi-LSTM) Cycle Generative Adversarial Network (CycleGAN). In this model, Bi-LSTM is used to enhance feature extraction capabilities. By using FBC-GAN and Fourier Transform, simulated fault location signals that closely resemble actual signals are generated, thereby achieving data augmentation [15]. Ruan et al. suggested a novel deep learning-based FDD method under the condition of imbalanced samples that converts the time-series signals into an image signal to extract the timing and coupling features, and then applies the improved conditional variational autoencoder-generative adversarial network (CVAE-GAN) to generate fault samples [16]. Liao et al. introduced a conditional auxiliary classier cycle-consistent generative adversarial network restrained by Wasserstein distance with gradient penalty (CAC-CycleGAN-WGP). This model can generate superior-quality signals of the minority classes with stability from the majority class [17]. Qin et al. proposed a digital twin–based fault data–generation method that inverses physics–informed neural network (PINN) built to recognize dynamic model parameters by embedding a bearing dynamic model into a neural network to produce fault samples under multiple working conditions [18]. Pei et al. utilized an improved few-shot Wasserstein auto-encoder (fs-WAE), which can strengthen the representational power of the encoder and provide a flexible optimization function for fault data–generation [19]. Wang et al. presented an effective deep learning method, namely, domain adaptive efficient sub-pixel network (DAESPN) that can enhance the resolution of the original sample for data augmentation [20]. Data augmentation models can produce a large amount of new data that resembles the distribution of the original data, effectively addressing the issue of the scarcity of fault samples. Additionally, these models can specifically resolve the problem of sample imbalance. However, the generator and discriminator in the network still face mode collapse issues during training, necessitating optimization and improvement.

Based on the analysis above, to address the imbalance issue of fault samples in practical scenarios and the limitations of existing neural networks in handling data, we propose a fault diagnosis technique tailored for few-fault data. The aim is to enhance diagnostic accuracy from multiple perspectives, leveraging data augmentation with Time GAN and an improved self-attention mechanism within a Transformer classifier. The following is a summary of this paper's primary contributions.

(1) Utilizing the Time GAN equipped with an auto-encoding module, consisting of embedded and recovery functions, this module, compared to traditional GAN networks, generates higher-quality samples to tackle the issue of scarce fault samples in practical scenarios.

(2) An improved Transformer fault diagnosis method is introduced. This method transforms time-step attention into sensor attention, addressing the issue of weak global information modeling capacity in CNN. Enhanced diagnostic accuracy and convergence speed.

(3) The Time GAN data augmentation model generates fault samples through its inherent adversarial mechanism and unique auto-encoding module, thereby balancing the dataset. The augmented and balanced dataset is then sent into the Transformer model, which utilizes the advantages of the Transformer's encoder-decoder architecture and improved attention structure for feature extraction and fault classification. Experimental validation demonstrates that the model exhibits high accuracy and feasibility.

The paper's remaining portions are arranged as follows. In Section II, the basic principle is presented. Section III delineates the proposed fault diagnosis process and the framework of the diagnostic model. The dataset and network settings utilized in the fault diagnosis experiments are shown in Section IV. Section V comprehensively analyzes the experimental results from various aspects and compares them with other models, verifying the suggested method's performance. Finally, a summary of the entire content is provided.

II. PRELIMINARIES

The fundamental structure and essential elements of the Time GAN and Typical Transformer will be covered in this part.

A. DATA AUGMENTATION

Deep learning network models, which can automatically extract multiple complex features from input samples without human intervention, have been extensively applied in the fault diagnosis field. However, from a mathematical perspective, their use of gradient descent optimization can easily lead to the problem of local optima, and their learning style and generalization ability are closely related to the selection of samples. Due to the inherent characteristics of imbalanced sample classes and complex distributions in tractor gearbox data, directly inputting the data into deep learning network models may not effectively diagnose and classify them. In this article, we employ Time Series Generative Adversarial Networks (Time GAN) to alter the data distribution, balancing the various data samples to prevent skewness and thereby enhancing the diagnostic capability and generalization of the model.

The steps for data augmentation using Generative Adversarial Networks (GANs) are as follows:

To address the issue of model overfitting or underfitting due to the small training sample size, this paper employs the Time GAN to augment the fault samples. Time GAN is a branch of Generative Adversarial Networks (GAN) [21]. Typically, the main component of a GAN is an adversarial module consisting of two neural networks: the generator and the discriminator. These two neural networks continuously compare generated data with real data to improve the quality of the generated data. Equation (1) represents the value function for evaluating the generator and discriminator.

$$\mathcal{L}_U = {}_{s,x_{1:t}\ p}\left[\|s-\tilde{s}\|_2 + \sum_t x_t - \tilde{x}\|_2\right] \tag{1}$$

where: $s, x_{(1:t)}$ represent the original dataset, $\tilde{s}, \tilde{x}$ denote the reconstructed form of the original dataset, and the conditional distribution that corresponds to the original data is represented by p.

The Time GAN network features not only the typical adversarial module found in conventional GANs but also an auto-encoding module [22]. The primary role of the autoencoding module is data dimensionality reduction. Two neural networks make up this module: the embedding function and the recovery function, connected via a latent function. The embedding function transforms the data into latent codes h(h∈ H), which are then fed into the discriminator for data selection. Subsequently, the recovery function performs the inverse transformation, ultimately outputting the enhanced dataset. An illustrative diagram of data generation in Time GAN is depicted in Figure 1. In Time GAN, the neural network formed by the embedding function and the recovery function is evaluated using equation (2).

$$\mathcal{L}_R = {}_{x_{1:t\sim p}}\left[\sum_i \| x_t - r(e)(x_t)\Big) \|_2\right] \tag{2}$$

where: r represents the reconstruction function, and e represents the embedding function.

Generating time-correlated data is inherently challenging for GANs, especially when the input data consists of long sequences with multiple features. To incorporate the temporal relationships between data into the learning architecture, Time GAN employs a supervised loss function based on autoregressive learning algorithms. This allows the GAN network to introduce time-conditioned probability.

The real datasets collected in this study consist of time-series data, specifically vibration signals from the tractor gearbox collected every minute during operation. Therefore, the data in the dataset exhibit characteristics of long sequences with multiple features. Consequently, collecting

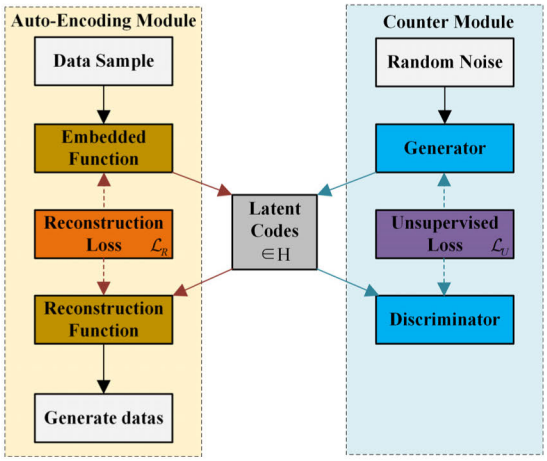


FIGURE 1. The process of data generation in Time GAN.

real datasets aligns well with the applicability of Time GAN for data augmentation purposes.

B. TRANSFORMER MODEL

The Transformer model is based on the attention mechanism, featuring an encoder-decoder architecture [23], [24], [25]. The main component of the Transformer network is stacked encoders and decoders. A typical Transformer network is illustrated in Figure 2.

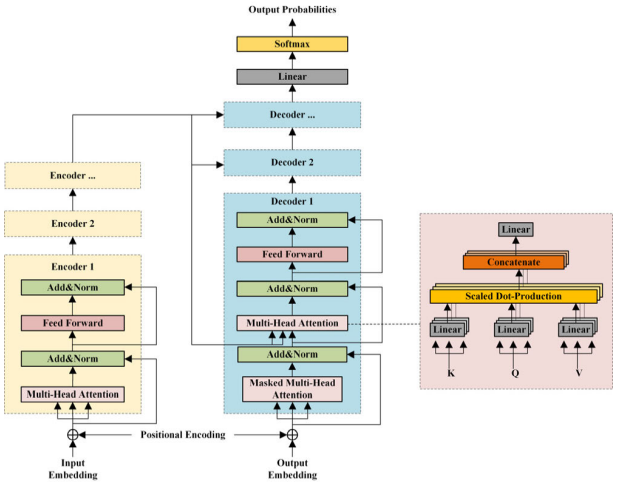


FIGURE 2. Typical transformer network.

1) ENCODER

The encoder consists of a positional encoding layer, feedforward neural network layer, residual, layer normalization layer, and multi-head self-attention layer. The core of the encoder is the multi-head self-attention mechanism, primarily used to allocate attention (weights) to the input feature sequence, enabling the model to focus more on important information within the input feature sequence. The feedforward neural network layer is mainly responsible for transforming the output of the multi-attention layer. Additionally, residual and layer normalization layers are added after the multi-head self-attention layer and feedforward neural network layer to address issues such as gradient vanishing caused by excessive network depth, thereby accelerating network convergence speed and enhancing network generalization ability. The main components of the encoder are described as follows.

a: POSITION ENCODING

The attention mechanism is the primary component of the Transformer network [26]. However, the attention layer alone cannot learn the positional information of sequences. Therefore, positional encoding is introduced to incorporate positional information into the sequences. Positional encoding applies functions to alternate even and odd dimensions of input feature samples, as shown in equations (3) and (4).

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{3}$$

$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{4}$$

where: pos represents the position of input feature samples, i denotes the dimension of input features,$2i$ and $2i+1$ correspond to even and odd dimensions, $d_{\mathrm{mod}\,el}$ represents the dimensionality of positional encoding, which is equivalent to the dimensionality of input features.

b: MULTI-HEAD SELF-ATTENTION

This layer calculates the similarity between input feature sequences to allocate weights, representing the importance of information [27], [28], [29]. Firstly, the input feature sequence $\boldsymbol{X}$ is multiplied by three distinct weight matrices to get the query matrix $\boldsymbol{Q}$, key matrix $\boldsymbol{K}$, and value matrix $\boldsymbol{V}$. Then, the query matrix is used to compute the dot product with the key matrix $\boldsymbol{K}$ and *SoftMax* normalization to calculate attention weights. Finally, the output feature sequence is obtained by weighting the value matrix $\boldsymbol{V}$ based on the attention weights, as shown in equations (5) to (8).

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}_q \tag{5}$$

$$\boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}_k \tag{6}$$

$$\boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}_v \tag{7}$$

$$Self - Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = SoftMax(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_{model}}})\boldsymbol{V} \tag{8}$$

where: $\boldsymbol{W}_q, \boldsymbol{W}_k, \boldsymbol{W}_v$ are the weight matrices corresponding to $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}$ respectively.

Equations (9) and (10), which show the concatenation and linear transformation of several self-attention processes, represent the multi-head self-attention mechanism.

$$Multi - Head(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = Concat(H_1, H_2, \cdots, H_n)\boldsymbol{W} \tag{9}$$

$$H_i = Self - Attention(\boldsymbol{X}\boldsymbol{W}_{\boldsymbol{Q}_i}, \boldsymbol{X}\boldsymbol{W}_{\boldsymbol{K}_i}, \boldsymbol{X}\boldsymbol{W}_{\boldsymbol{V}_i}) \tag{10}$$

where: $H_i (i \in \{1, 2, \cdots, n\})$ represents the *ith* self-attention head, n is the number of self-attention heads, the *Concat*

function concatenates the outputs of multiple self-attention heads. $\boldsymbol{W}$ is the weight matrix for multi-head self-attention and $\boldsymbol{W}_{\boldsymbol{Q}_i}, \boldsymbol{W}_{\boldsymbol{K}_i}, \boldsymbol{W}_{\boldsymbol{V}_i}$ represent the $\boldsymbol{Q}, \boldsymbol{K}$ and $\boldsymbol{V}$ weight matrices, respectively, for the *ith* self-attention head.

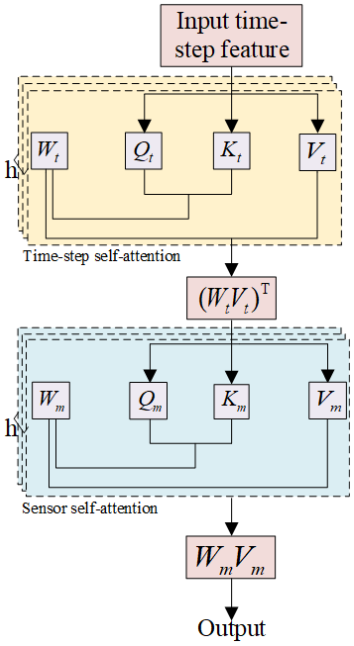


FIGURE 3. Improved multi-head self-attention.

2) DECODER

The decoder consists of a masked multi-head self-attention layer, an encoder-decoder multi-head self-attention layer, and a residual and layer normalization layer [30]. The masked multi-head self-attention is utilized to allocate attention (weights) to the input label sequence [31]. Subsequently, the encoder-decoder multi-head self-attention mechanism is employed to learn the dependency between the intermediate vectors from the encoder's output and the input label sequence [32]. The main components of the decoder are described as follows.

(1)Masking Multi-Head Self-Attention

The masked multi-head self-attention mechanism can prevent the model from learning future label sequence information at each time step [33], [34], [35]. The masking operation involves introducing a lower triangular unit matrix $\boldsymbol{M}$ (with elements above the main diagonal being 0, elements on the main diagonal, and below being 1) and performing element-wise multiplication with $\boldsymbol{QK}$ during the computation of scaled dot-product self-attention. This operation zeros out future label sequence information, as shown in equation (11).

$$Masked(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = SoftMax(\frac{\boldsymbol{QK}^T\boldsymbol{M}}{\sqrt{d_{\bmod el}}}) \quad (11)$$

(2) The encoder-decoder self-attention represents the label sequence information and is denoted as $\boldsymbol{Q}_D$ [36], [37]. The key matrix $\boldsymbol{K}$ and value matrix $\boldsymbol{V}$ are derived from the intermediate vectors of the encoder's output, representing the feature sequence information, and are respectively denoted as $\boldsymbol{K}_E$ and $\boldsymbol{V}_E$, as shown in equation (12).

$$E - Dncoder(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = SoftMax(\frac{\boldsymbol{Q}_D\boldsymbol{K}_E^T}{\sqrt{d_{\bmod el}}})\boldsymbol{V}_E \quad (12)$$

III. FAULT DIAGNOSIS MODEL

We will present our fault diagnostic model in this part. The model is trained, validated, and tested on the bearing and gear dataset. The fault diagnosis process is shown in Figure 4, while the architecture of the fault diagnosis model is depicted in Figure 5.

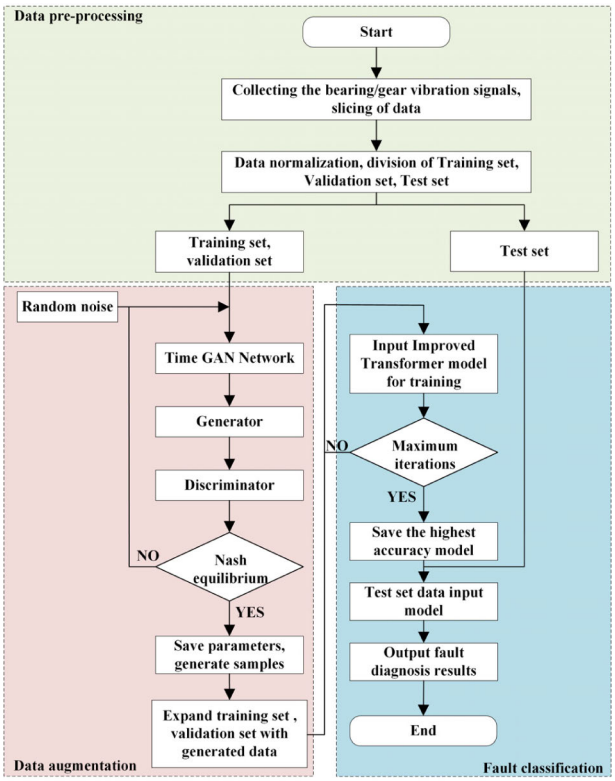


FIGURE 4. Fault diagnosis process.

The paper is based on one-dimensional raw vibration signals and completes the entire process of the fault diagnosis model from pre-training to practical application. The steps involved are as follows:

(1) Collect vibration signal characteristics of the tractor transmission system (bearing/gear) and divide them proportionally into training, validation, and testing sets. Normalize the data and complete the pre-training process of the vibration data.

(2) Input the training and testing set samples along with random noise into the Time GAN. Alternately train the generator and discriminator until reaching Nash equilibrium. Retain the parameters and generate synthetic samples similar to the original samples, enhancing the dataset through data augmentation.

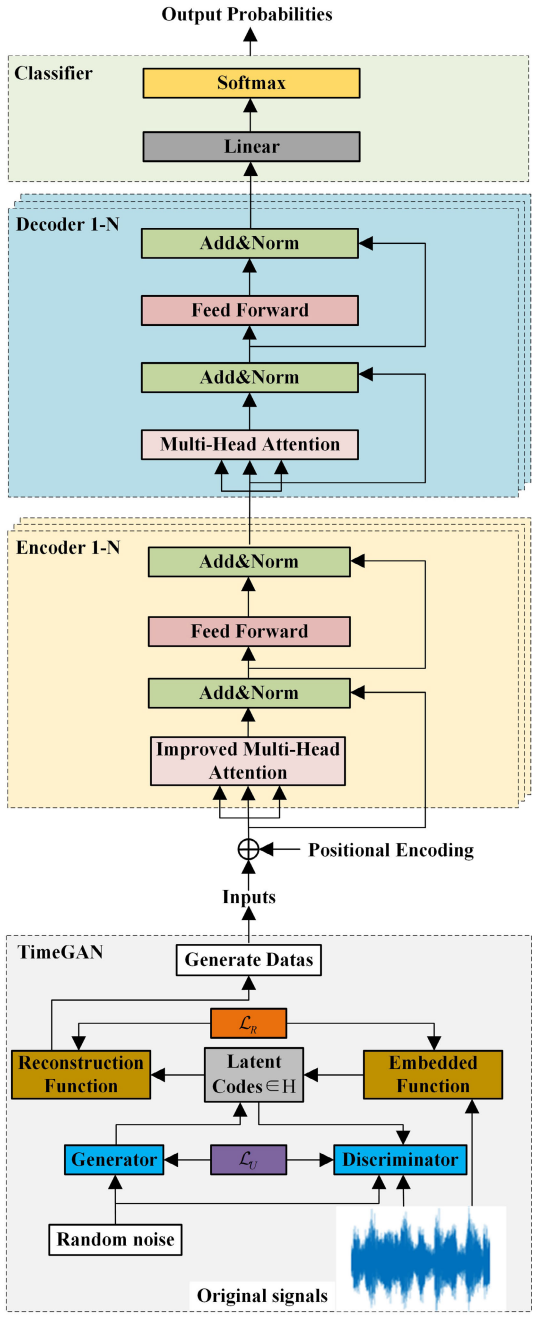


FIGURE 5. Proposed fault diagnosis structure.

(3) Input the augmented samples into the Transformer model improved with a multi-head self-attention mechanism for training. Extract crucial features from the input feature sequences for multi-label fault diagnosis, considering different time steps and sensors. Save the diagnostic model with the highest accuracy after reaching the maximum iteration count.

(4) Use the trained Transformer model with the highest accuracy to diagnose and classify various fault types by inputting the testing set. Obtain the final fault diagnosis results.

(5) Utilizing four assessment metrics, assess the improved model's diagnostic performance.

A. IMPROVED MULTI-HEAD SELF-ATTENTION FOR TRANSFORMER

In domains like natural language processing, the Transformer has had great success. However, its unique structure poses limitations in the fault diagnosis field. Therefore, addressing the multi-label fault diagnosis problem, this paper proposes an improved Transformer model.

The traditional Transformer model focuses solely on attention (weights) across different time steps in the input sequence, disregarding the importance of different sensors within the input sequence. Therefore, this paper introduces an improved Transformer model. This enhancement involves integrating the improved attention mechanism into the encoder of the Transformer model to extract crucial features from different time steps and different sensors within the input feature sequence.

The improved self-attention of the Transformer model is illustrated in Figure 3. First, the time-step feature sequence $\boldsymbol{Q}_t, \boldsymbol{K}_t, \boldsymbol{V}_t$is input and processed through multi-head time-step self-attention to acquire the weighted time-step feature sequence $\boldsymbol{W}_t\boldsymbol{V}_t$, as shown in Equation 13. By applying multi-headed attention to the feature sequences at different time steps, a weighted feature sequence of time steps is obtained. Subsequently, this weighted feature sequence of time steps is transposed to yield a feature sequence of sensors $\boldsymbol{Q}_m, \boldsymbol{K}_m, \boldsymbol{V}_m$, as shown in Equation 14. Finally, the sensor feature sequence is processed through multi-head sensor self-attention to acquire the weighted sensor feature sequence $\boldsymbol{W}_m\boldsymbol{V}_m$, as shown in Equation 15.

$$Time(\boldsymbol{Q}_t, \boldsymbol{K}_t, \boldsymbol{V}_t) = SoftMax(\frac{\boldsymbol{Q}_t\boldsymbol{K}_t^T}{\sqrt{d_{model}}})\boldsymbol{V}_t = \boldsymbol{W}_t\boldsymbol{V}_t \tag{13}$$

$$(\boldsymbol{W}_t\boldsymbol{V}_t)^T = \boldsymbol{Q}_m = \boldsymbol{K}_m = \boldsymbol{V}_m \tag{14}$$

$$Sensor(\boldsymbol{Q}_m, \boldsymbol{K}_m, \boldsymbol{V}_m) = SoftMax(\frac{\boldsymbol{Q}_m\boldsymbol{K}_m^T}{\sqrt{d_{model}}})\boldsymbol{V}_m = \boldsymbol{W}_m\boldsymbol{V}_m \tag{15}$$

where: $\boldsymbol{Q}_t, \boldsymbol{K}_t, \boldsymbol{V}_t, \boldsymbol{W}_t$ show the weight matrices for the time step features,$\boldsymbol{Q}_m, \boldsymbol{K}_m, \boldsymbol{V}_m, \boldsymbol{W}_m$represent the weight matrices for the sensor features.

Therefore, the improved self-attention not only extracts important features from different time steps but also different sensors. Enhanced attention mechanisms in Transformer models enable better extraction of global information, thereby improving the model's capability to learn various fault categories and enhancing classification accuracy.

B. TIME GAN DATA AUGMENTATION

Before inputting the original dataset into Time GAN for training, it needs to be sliced into three-dimensional data with certain step sizes. This process can be considered as grouping

the data to different degrees, and the generation process involves training and generating based on each group of data. This preprocessing step will determine the multiplication factor for augmenting the data. Due to the varying sizes of the original dataset, different numbers of slices will to some extent determine the quality of the augmented dataset.

After slicing the data, normalization is uniformly applied to the data [38]. This is beneficial for reducing model training time and improving training accuracy. The normalization method chosen is min-max normalization, as shown in equation (16), which scales the data range to (0, 1). Compared to mean normalization, min-max normalization is more suitable for time series data and data with fewer extreme values.

$$x' = \frac{\bar{x} - min(x)}{max(x) - min(x)} \tag{16}$$

where: $\bar{x}$ represents the sample mean,$min(x)$ means the minimum value of the sample set x and $max(x)$ represents the maximum value of the sample set x.

Finally, to ensure the relative independence of the sliced data before generation, the segmented slices are shuffled before inputting them into Time GAN to generate data.

IV. EXPERIMENTAL DATASET AND NETWORK PARAMETERS

This section introduces the sources of experimental data and the configuration of parameters for the fault diagnosis models utilized.

A. CASE WESTERN RESERVE UNIVERSITY(CWRU) DATASET

To assess the proposed model, experiments were conducted using a bearing dataset provided by the laboratory at CWRU, USA. The bearing test rig is depicted in Figure 6. The drive-end bearing is of the SKF6205 model, while the fan-end bearing is of the SKF6203 model. Various conditions of the bearings were simulated using electrical discharge machining. Accelerometers were positioned to capture vibration acceleration signals of faulty bearings. A 16-channel data logger was used to record vibration signals at sampling frequencies of 12 kHz and 48 kHz [39].

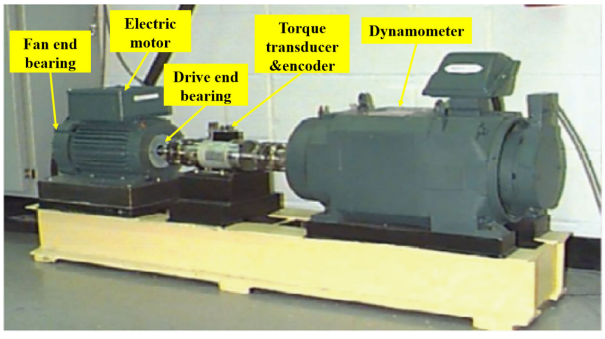


FIGURE 6. CWRU rolling bearing experimental rig.

The dataset comprises four different bearing health conditions: normal, inner race fault, outer race fault, and ball fault. Each fault condition has three fault diameters: 7 mils, 14 mils, and 21 mils. Vibration signals were collected from the motor under four different loads (0HP, 1HP, 2HP, and 3HP) and at four speeds(1730r/min, 1750r/min, 1772r/min, and 1797r/min).

This paper focuses on the 6205-2RS JEM SKF deep groove ball bearing as the research subject. The damage to the bearing is created using electrical discharge machining to induce single-point damage with the motor load at 0 horsepower and the bearing speed at 1797 rpm, with a sampling frequency of 12 kHz. Vibration signals from 10 different categories of data collected at the drive end are selected for analysis. The classification of labels is presented in Table 1.

TABLE 1. Classification of bearing vibration signals.

Fault diameter	Speed(r/min)	Fault severity	Location	Label
/	1797	/	Normal	0
7mils	1797	Mild	Ball	1
14mils	1797	Moderate	Ball	2
21mils	1797	Severe	Ball	3
7mils	1797	Mild	Inner race	4
14mils	1797	Moderate	Inner race	5
21mils	1797	Severe	Inner race	6
7mils	1797	Mild	Outer race	7
14mils	1797	Moderate	Outer race	8
21mils	1797	Severe	Outer race	9

B. SELF-MADE DATASET

The dataset collected in the laboratory consists of gear vibration signals acquired from a tractor transmission system loaded on an experimental test rig. The physical representation of the laboratory tractor transmission system test rig is depicted in Figure 7.

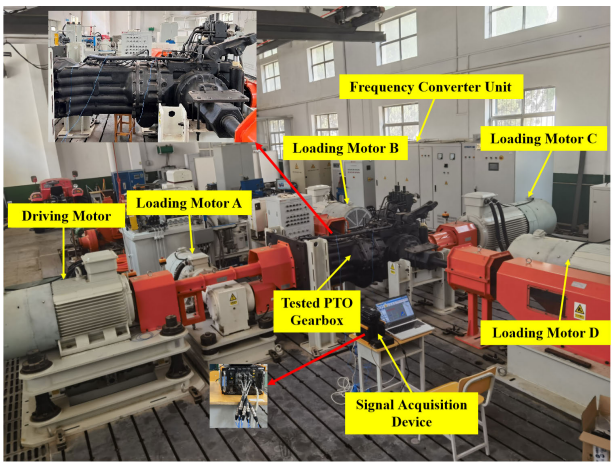


FIGURE 7. Tractor transmission system loading experimental rig.

The tractor transmission system test rig mainly consists of the driving unit, front axle loading unit (Motor A), right rear

axle loading unit (Motor B), PTO (Power Take-Off) loading unit (Motor C), left rear axle loading unit (Motor D), PTO gearbox, and frequency converter unit. This experimental rig simulates the operating conditions of the tractor transmission gearbox and primarily collects vibration signals from various fault types of gear.

The vibration data from the gearbox is collected using three channels, with acceleration sensors mounted on the gearbox to capture vibration signals in the X, Y, and Z directions of the planetary gears, with a frequency of 5 kHz. The experimental study focuses on a condition with a speed-load combination of 30 Hz and 2V. The fault states of the gears are categorized into five classes: normal, chipped tooth, miss tooth, root fault, and surface fault. Therefore, the fault diagnosis task based on the dataset collected in the tractor transmission system experimental rig is a 5-class classification problem, the labels as shown in Table 2.

TABLE 2. Classification of gearbox vibration signals.

Location	Type	Label
Gearbox	Normal	0
	Chipped tooth	1
	Miss tooth	2
	Root fault	3
	Surface fault	4

C. EXPERIMENTAL DATASET SETTING

The fault diagnosis dataset mentioned in the subsequent sections of this paper simulates the imbalance of fault classes in real-world conditions before augmentation. For each type of fault, 20 real samples were extracted, and each sample had 1024 sampling points for diagnostic testing. The training, validation, and testing sets of these samples are split up in a 3:1:1 ratio. After augmentation using the Time GAN network, 100 mixed samples are extracted from both real and synthetic samples for each fault type, with 1024 sampling points per sample. In a similar vein, these enhanced samples undergo preprocessing and are split into 3:1:1 training, validation, and testing sets. Table 3 shows the quantity of samples both before and after augmentation.

D. NETWORK PARAMETERS

The experimental environment utilizes Python 3.10 and the PyTorch 1.13.1 deep learning framework. The experiment is divided into two parts: Time GAN data augmentation and Transformer fault classification. In the experiment, the Time GAN network is first used to generate the training, testing, and validation sample sizes, as shown in Table 3. Then, the improved Transformer model and other comparative experimental models are trained and evaluated. During the training of the machine learning models, the grid search method is employed to determine the optimal hyperparameters for each machine learning model, ensuring their best performance.

TABLE 3. Experimental fault samples.

Dataset/Samples	Before data augmentation	After data augmentation (Mixed samples)	Label
CWRU Dataset	20	100	0
	20	100	1
	20	100	2
	20	100	3
	20	100	4
	20	100	5
	20	100	6
	20	100	7
	20	100	8
	20	100	9
Samples	200(120/40/40)	1000(600/200/200)	
Self-Made Dataset	20	100	0
	20	100	1
	20	100	2
	20	100	3
	20	100	4
Samples	100(60/20/20)	500(300/100/100)	

The model parameters selected through grid search are as follows: The parameters for the Time GAN model are listed in Table 4, while the parameters for the improved Transformer model are shown in Table 5, and other comparative model parameters are presented in Table 6 of Section V.

V. EXPERIMENTAL RESULTS AND ANALYSIS

This section validates the performance of the proposed diagnostic model through experimentation and compares it with other diagnostic algorithms.

A. FAULT DATA AUGMENTATION

The batch size and iteration count in the Time GAN network are set to 128 and 5000, respectively, during the data production phase. The learning rate is 0.001. As shown in Figure 8, the discriminator's loss plot from the training phase is used to more clearly visualize the data that the network generates.

From the above figure, it can be observed that the Time GAN network used in this paper achieves convergence within a short training time in both the CWRU dataset and the self-made dataset, with minimal network fluctuations. This suggests that Time GAN pays closer attention to the temporal

TABLE 4. Time GAN parameters.

Model	Parameters	
Time GAN	Batch size	128
	Iterations	5000
	Learning rate	0.001
	Hidden size	24
	Optimizer	Adam
	Activation function	Sigmoid

TABLE 5. Improved Transformer model parameters.

Model	Parameters	
Improved Transformer	Batch size	64
	Model dimension	64
	Encoders	4
	Decoders	4
	Improved Multi-Head Self-Attention	4
	Attention dropout	0.1
	Hidden layers	128
	Dropout	0.1
	Learning rate	0.0005

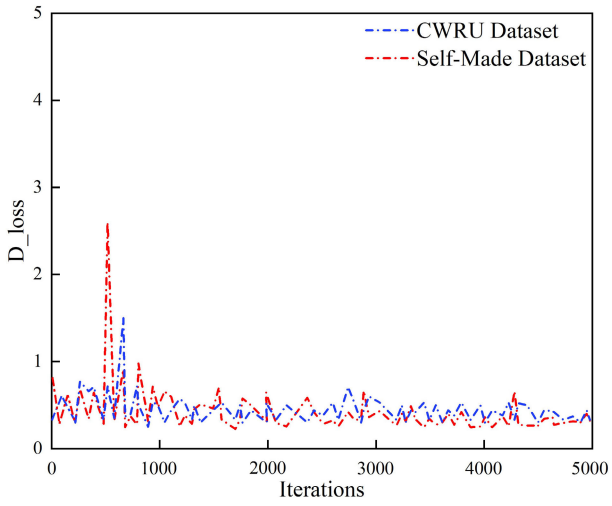


FIGURE 8. Time GAN discriminator loss curves.

characteristics of the data and benefits from the incorporation of the auto-encoder module, which enables the generation of higher-quality samples. Consequently, Time GAN exhibits a significant advantage in data generation, producing more effective results.

To assess the quality of the data generated by Time GAN in two datasets, this paper introduces t-distribution Stochastic Neighbor Embedding (t-SNE) [40] to visualize and compare the distributions of the original data and the data that was generated. The visualizations of the distributions of the two experimental datasets before and after data augmentation are illustrated in Figure 9. In Figures 9 (a) and (b), the points corresponding to colors 0-9 on the right side represent the ten labels for bearing classification. In Figures 9 (c) and (d), the points corresponding to colors 0-4 on the right side represent the five labels for gear classification. The labels are defined in Table 1 and Table 2 in Section IV.

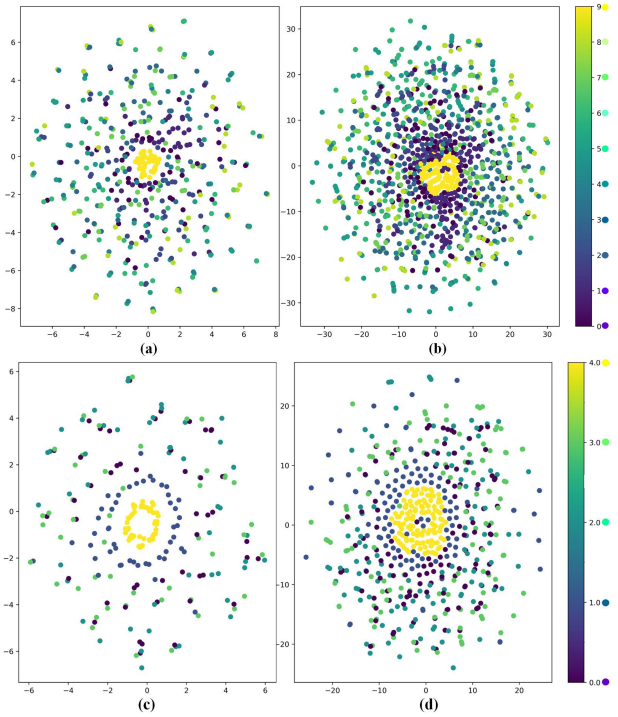


FIGURE 9. Visualization of synthetic data via t-SNE. (a) before data augmentation in CWRU dataset, (b) after data augmentation in CWRU dataset, (c) before data augmentation in self-made dataset, (d) after data augmentation in self-made dataset.

Observing the visualizations, it is apparent that the data generated by Time GAN closely resembles the distribution of the original data in both datasets. Furthermore, the feature points of the fault samples appear to be more abundant and concentrated, enhancing the distinguishability between categories and facilitating the network's learning of data features. Therefore, it can be concluded that the Time GAN data augmentation model effectively addresses the issue of insufficient fault samples.

B. FAULT DIAGNOSIS EVALUATION METRICS

We utilize four metrics to assess the fault diagnosis model performance.

Accuracy is commonly the most prevalent evaluation metric, defined as the percentage of properly classified

samples out of all samples, as shown in equation (17).

$$Accuracy = \frac{\sum_{n=1}^{q} tp_n + \sum_{n=1}^{q} tn_n}{\sum_{n=1}^{q} tp_n + \sum_{n=1}^{q} tn_n + \sum_{n=1}^{q} fp_n + \sum_{n=1}^{q} fn_n} \quad (17)$$

Precision refers to the number of real positive samples contained within the predicted positive results. Precision is defined as shown in equation (18).

$$Precision = \frac{\sum_{n=1}^{q} tp_n}{\sum_{n=1}^{q} tp_n + \sum_{n=1}^{q} fp_n} \quad (18)$$

Recall, also known as sensitivity, refers to the number of accurately predicted positive samples out of all samples in the dataset. Recall is defined as shown in equation (19).

$$Recall = \frac{\sum_{n=1}^{q} tp_n}{\sum_{n=1}^{q} tp_n + \sum_{n=1}^{q} fn_n} \quad (19)$$

where: $n \in \{1, 2, \cdots, 10\}$ represents fault types; $q = 10$ is the total number of fault types, tp_n indicates how many real positive samples were accurately predicted, fp_n represents the number of false positive samples where positive samples are incorrectly predicted, tn_n shows how many true negative samples accurately predicted, fn_n signifies the number of false negative samples where negative samples are incorrectly predicted.

Due to the occasional negative correlation between precision and recall, it is necessary to introduce the F1-score to further assess the model performance. The F1-score, which represents their weighted harmonic mean, is derived from recall and precision. The F1-score is defined as shown in equation (20).

$$F1 = 2 \times \frac{\mathrm{Pr}\,ecision \times Recall}{\mathrm{Pr}\,ecision + Recall} \quad (20)$$

C. RELATED ALGORITHMS AND PARAMETERS

Three machine-learning fault diagnostic models and the typical Transformer model are compared to the Improved Transformer in order to assess the effectiveness of our method. Table 6 provides an outline of its parameters.

TABLE 6. Model parameters.

Model	Kernel Function	Parameters
CNN-BiLSTM	-	Batch size=64 CNN kernels =64 CNN kernel size=3 CNN pool size=2 BiLSTM units=64
WDCNN	-	Batch size=64 First CNN kernels =16 First CNN kernel size=64 Other CNN kernels =64 Other CNN kernels size =3 CNN pool size=2
SVM	Sigmoid	C=100，gamma=1
	Linear	C=100，gamma=1
	Rbf	C=1000，gamma=0.01
	Poly	C=10，gamma=3
Typical Transformer	-	Batch size =64 Model dimension =64 Encoders =4 Decoders =4 Multi-Head Self-Attention =4 Attention dropout =0.1 Hidden layers =128 Dropout =0.1 Learning rate =0.0003

1) CNN-BiLSTM [41]

Compared to traditional LSTM, the BiLSTM model enhances the ability to capture bidirectional information within time series. When combined with CNN, the network can effectively extract key features from the input data, improving the model's understanding of the data.

2) WDCNN [42]

The Wide Deep Convolutional Neural Network (WDCNN) effectively addresses multi-condition problems. Its main feature is a large convolution kernel in the first layer, which is trained using optimization algorithms. This allows it to automatically learn diagnostic-relevant features while removing those that do not aid diagnosis.

3) SVM [43]

Support Vector Machine (SVM) is a classifier that categorizes predictions into different classes by learning the distinctive features of the training dataset. It is a classic machine-learning model. As SVM has evolved, four commonly used kernel functions have emerged: Sigmoid, Linear, RBF (Radial Basis Function), and Poly. This study conducts experiments sequentially for each of these four kernel functions to validate their performance.

4) THE TYPICAL TRANSFORMER

Following feature input, an attention mechanism is incorporated to provide multiple representation subspaces, allowing the Transformer model to concentrate on feature information from distinct representation subspaces at various positions.

D. CASE 1: CWRU DATASET

In this section, various evaluation metrics are utilized as criteria, and the proposed diagnostic model's performance in bearing fault diagnosis is validated through comparisons with other algorithms.

1) RESULTS

First, imbalance samples without data augmentation are inputted into the Transformer model for training and classification. The obtained accuracy curve and loss curve are depicted in Figure 10 (a). From the graph, it can be observed that due to the imbalance in the dataset, the network fails to adequately learn the corresponding features of each fault type. The network rapidly converges, reaching stability after 50 iterations. However, the final training set accuracy only reaches 94.5%, and the test set accuracy only reaches 93.7%. Additionally, the training loss remains relatively high.

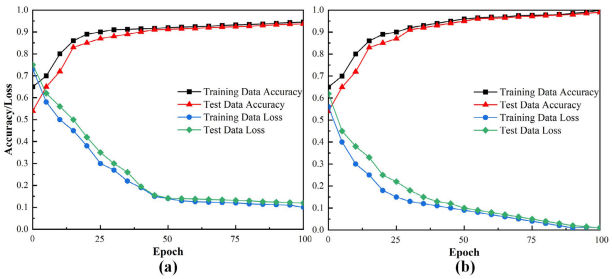


FIGURE 10. Training accuracy and loss curves (a) Before data augmentation, (b)After data augmentation.

Next, Time GAN is utilized to generate data, and the generated samples for each fault type are added to the original imbalanced sample set, thus forming a balanced sample set. This balanced sample set is then inputted into the Transformer model for training and fault classification. The resulting accuracy curve and loss curve are depicted in Figure 10 (b).

From Figure 10 (b), it is evident that after data augmentation, the network can adequately learn the feature distributions of various fault types. There is a significant improvement in training performance. After iterating 50 times, the network stabilizes, achieving a final training set accuracy of 99.3%, while the test set accuracy stabilizes at 98.9%.

In Figures 11 (a) and 11 (b), the points corresponding to colors 0-9 on the right side represent the ten labels for bearing classification.

From Figure 11 (a), it can be observed that although there are discernible boundaries for fault type recognition before sample augmentation, the accuracy of fault recognition is compromised due to the scarce samples, resulting in overlaps among some points. In Figure 11 (b), after augmentation, clear boundaries between ten classes are established, facilitating efficient fault classification.

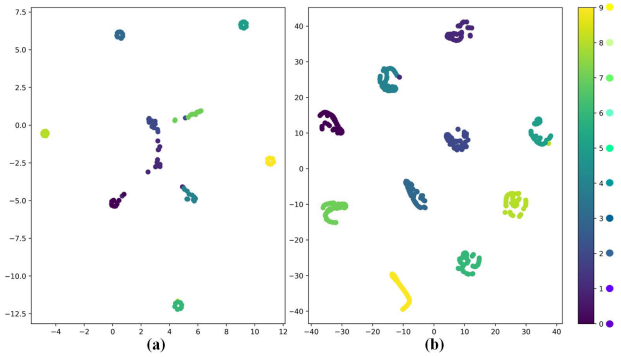


FIGURE 11. Visualization of CWRU dataset the classification results via t-SNE (a) Before data augmentation, (b)After data augmentation.

Confusion matrices are introduced to assess the experimental findings and provide a more lucid demonstration of the recognition results of the proposed model for each category in the test set. Confusion matrices are plotted for the classification results before and after data augmentation, presenting the classification results for each fault category before and after data augmentation. The confusion matrices are depicted in Figure 12.

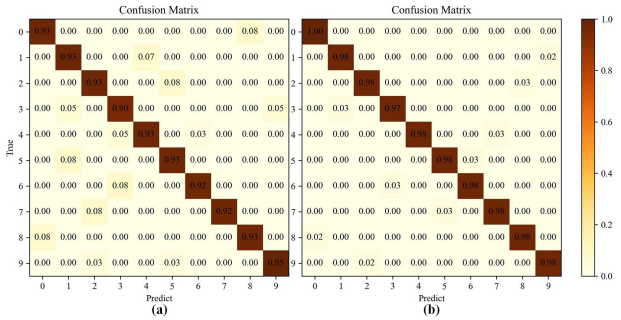


FIGURE 12. Classification confusion matrix of CWRU dataset:(a) Before data augmentation, (b)After data augmentation.

From Figure 12, it is evident that after data augmentation, the classification accuracy of the network for normal bearings increased from 93% to 100%. Moreover, the classification accuracy for the three fault types (namely inner race faults, outer race faults, and ball faults) with three different fault depths each (labels 1-9), all improved from an average of 93% to over 97%, achieving excellent classification performance. Comparing the results before and after data augmentation, there is a significant improvement in both overall classification accuracy and individual fault classification accuracy.

2) METHOD COMPARISON

The experiment employed the same training set as the improved Transformer model and tested four types of machine-learning models: CNN-BiLSTM, WDCNN,

TABLE 7. Diagnostic results of different fault diagnosis models after Time GAN data augmentation.

Model	Data augmentation	Kernel Function	Accuracy/%	Precision/%	Recall/%	F1-score/%
CNN-BiLSTM	NO	-	88.36	88.01	87.33	89.25
	YES		90.62	90.33	89.25	90.18
WDCNN	NO	-	90.42	89.52	88.68	84.20
	YES		91.58	90.46	90.35	89.68
SVM	NO	Sigmoid	91.32	89.24	90.74	89.98
	YES		93.53	92.36	92.52	92.44
	NO	Linear	88.57	90.36	85.48	87.85
	YES		93.22	91.24	87.54	89.35
	NO	Rbf	90.36	90.33	87.49	88.89
	YES		93.41	92.76	89.53	91.12
	NO	Poly	90.24	93.28	89.34	91.27
	YES		92.32	94.36	91.34	92.83
Typical Transformer	NO	-	92.66	93.83	91.46	92.95
	YES		93.68	94.72	92.23	93.24
Improved Transformer	NO	-	93.76	94.77	92.16	93.45
	YES		98.96	95.00	94.23	94.61

SVM, and the Typical Transformer. Each sample contains 1024 sampling points, each model was run ten times and averaged. Evaluation metrics for fault diagnosis results before and after dataset augmentation are presented in Table 7.

(1) Compared to the composite neural network CNN-BiLSTM, the Transformer network model with improved attention mechanism and encoder-decoder architecture outperforms the above neural networks in all aspects of fault diagnosis tasks. After dataset augmentation, the accuracy of the Transformer model is 11.34% higher than that of the CNN-BiLSTM model; the precision is increased by 9.67%; the recall is improved by nearly 7.98%; and the F1-score is enhanced by 8.43%.

(2) Compared to the neural network WDCNN, the improved Transformer model outperforms the WDCNN model in all evaluation metrics, indicating that the multi-head attention enables the Transformer to concentrate on more critical fault information. In multi-label fault diagnosis tasks, single-head self-attention cannot improve the recurrent neural network as effectively as multi-head attention does.

(3) Compared to the classical machine learning method SVM with four different kernel functions, the improved Transformer network demonstrates superior performance, achieving the best results in all four evaluation metrics. This showcases the model's strong capability in handling multi-class classification problems.

(4) The improved Transformer model incorporates better encoder attention compared to the Typical Transformer model. This enhancement aims to extract critical features from input feature sequences with varying time steps and from different sensors, thereby enhancing the model's fault diagnosis performance. All evaluation metrics indicated that the improved diagnostic model outperforms the typical Transformer model.

From Table 7, it is evident that there are differences in classification accuracy among various networks before and after data augmentation. This discrepancy stems from two main reasons: (1) when the input training samples are limited, the networks fail to sufficiently learn the sample features, leading to overfitting and ultimately reducing the classification accuracy of the networks; (2) the preprocessing of vibration data is relatively simple and direct, overlooking the correlations among data points, thereby missing some features. Therefore, conducting data augmentation before diagnosing faults with limited samples is crucial. It effectively enhances diagnostic performance.

From Figures 13 and 14, it is evident that the overall fault diagnosis performance of the WDCNN model is the worst. However, after data augmentation, all evaluation metrics improved. The Transformer fault diagnosis model, after improvements, demonstrated the best performance, with all four evaluation metrics surpassing those of the other four comparative models both before and after data augmentation. Experimental results indicate that when an adequate number of training samples are available, the proposed diagnostic models in this study achieved fault diagnosis accuracy reaching as high as 98.9%.

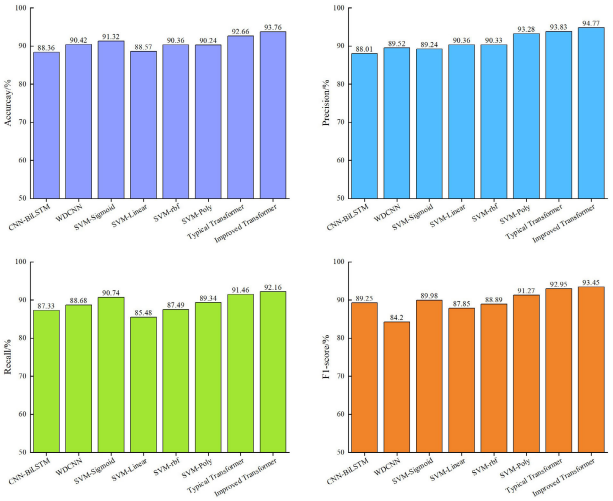


FIGURE 13. Diagnostic performance of different models before data augmentation in CWRU dataset.

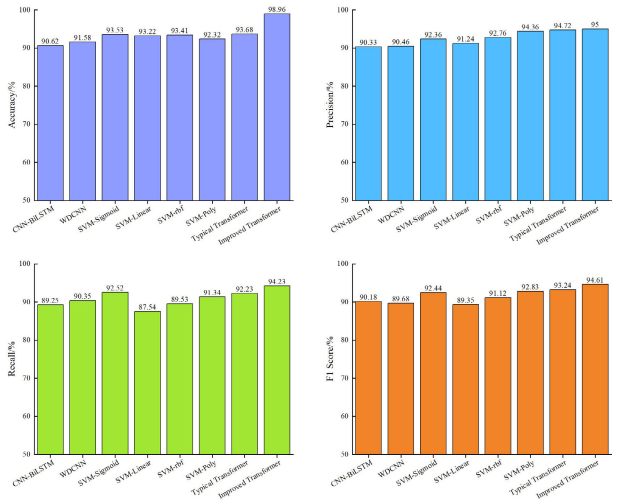
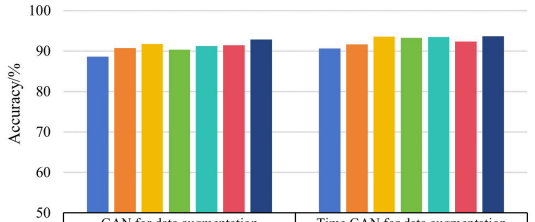


FIGURE 14. Diagnostic performance of different models after Time GAN data augmentation in CWRU dataset.

To further demonstrate the superiority of the proposed model in diagnosing faults in tractor transmission systems, we conducted an ablation experiment. We employed a traditional GAN network for data augmentation, followed by fault diagnosis experiments using the aforementioned comparative models. The results are as follows.

As shown in Figure 15, the diagnostic accuracy of each model after data augmentation with the traditional GAN network is lower than that achieved with the Time GAN network (results from Table 7). This is because the proposed Time GAN network can generate higher-quality samples during data augmentation, enabling the diagnostic models to learn more useful fault features and thus improve the models' accuracy.



	GAN for data augmentation	Time GAN for data augmentation
CNN-BiLSTM	88.65	90.62
WDCNN	90.67	91.58
SVM-Sigmoid	91.78	93.53
SVM-Linear	90.32	93.22
SVM-Rbf	91.25	93.41
SVM-Poly	91.38	92.32
Typical Transformer	92.87	93.68

FIGURE 15. Diagnostic accuracy of different models after GAN and Time GAN data augmentation.

Figures 14 and 15 illustrate that the fault diagnosis model proposed in this paper, which combines Time GAN and Improved Transformer, can generate high-quality samples while also achieving excellent classification accuracy in fault diagnosis. Therefore, even in cases with a limited number of fault samples, the model proposed in this paper can accurately diagnose faults in the tractor transmission system.

To quantitatively illustrate the diagnostic efficiency of the proposed model, the training and testing times for each model after data augmentation were recorded. The results are shown in Table 8.

TABLE 8. Different model training and testing time and accuracy.

Model	Training time/s	Testing time/s	Accuracy/%
CNN-BiLSTM	365.3	22.7	90.62
WDCNN	387.7	23.3	91.58
SVM-Sigmoid	325.5	10.7	93.53
SVM-Linear	316.9	9.3	93.22
SVM-Rbf	318.7	10.7	93.41
SVM-Poly	332.6	10.5	92.32
Typical Transformer	423.6	26.8	93.68
Improved Transformer	413.8	26.3	98.96

After Time GAN data augmentation, the training time for each model increases due to the larger volume of data input into each network. As shown in Table 8, the training time of the proposed model is slightly slower than that of other models. However, due to its excellent processing and classification capabilities of data features, it achieves the highest diagnostic accuracy, offering a new solution for fault diagnosis in few-sample conditions. Therefore, the computational efficiency of the proposed method is acceptable.

To verify the noise robustness of our model, this study added Gaussian white noise with five different signal-to-noise ratios (SNR $= -10$dB, -5dB, 0 dB, 5 dB, 10dB) to the original bearing signals. This simulates the real-world scenario where the vibration signals of the tractor are easily affected by strong environmental noise interference. The SNR is shown in equation (21).

$$SNR = 10 \times \lg \frac{P_{signal}}{P_{noise}} \tag{21}$$

where: P_{signal} represents the power of signal and P_{noise} means the power of noise.

As seen from equation (21), the smaller the SNR, the greater the noise power.

After data augmentation under the four different signal-to-noise ratios, the training methods and parameters for each model remained the same as previously described, each model was run ten times and averaged. The diagnostic accuracy of each model is shown in Table 9. The experiments demonstrate that as the SNR decreases, the accuracy of all methods declines. This is due to the greater variance in data distribution and positional information caused by the lower SNR. However, even when the SNR

is −10dB, the accuracy of the proposed method remains above 93%.

TABLE 9. Accuracy of each model under different signal-to-noise ratios.

Model	SNR=-10	SNR=-5	SNR=0	SNR=5	SNR=10
CNN-BiLSTM	83.62	85.26	88.57	89.32	90.32
WDCNN	81.58	85.78	87.65	90.23	90.88
SVM-Sigmoid	62.65	65.35	88.73	89.26	91.25
SVM-Linear	63.11	66.22	89.67	90.64	92.36
SVM-Rbf	62.32	68.78	87.65	91.34	92.65
SVM-Poly	61.15	65.32	86.46	91.63	92.15
Typical Transformer	90.25	91.03	92.06	92.13	93.26
Improved Transformer	93.32	94.23	96.78	97.25	98.58

To visually present the test diagnosis results of different methods, a corresponding bar chart was created based on Table 9, as shown in Figure 16. This figure reflects the diagnostic capability of the proposed method under the five different SNR.

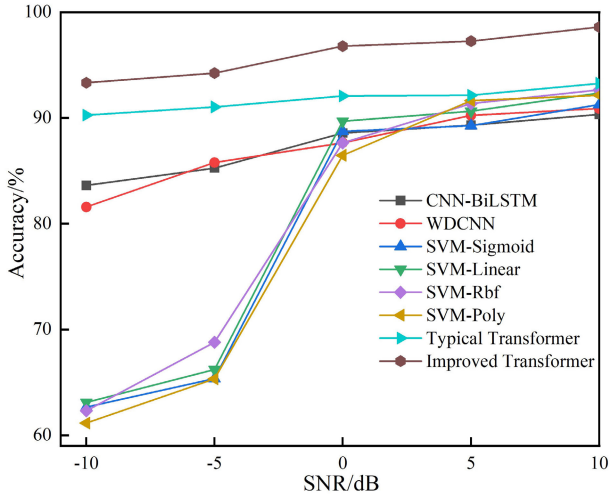


FIGURE 16. Classification accuracy of each model under different signal-to-noise ratios.

It can be seen that the SVM model performs poorly in classification under strong noise environments, whereas the model proposed in this paper exhibits a higher noise robustness. Compared to other models, the method presented in this paper improves the attention mechanism, facilitating the acquisition of more data features and demonstrating better performance. Therefore, the proposed method can extract more effective features from noisy samples and is more robust than other networks in different noisy environments.

E. CASE 2: SELF-MADE DATASET

In this section, gear data collected from a self-made dataset was used for experiments, and the proposed diagnostic model's generalization ability is validated through comparisons with other algorithms.

1) RESULTS

The generalization ability of the combined approach of the Time GAN and Transformer model proposed in this study was validated through testing against several classical machine learning fault diagnosis methods on the self-made dataset. To lessen the impact of unpredictability in the data and outcomes, each model was run five times and averaged.

We first utilized t-SNE visualization of extracted features and confusion matrices to illustrate the performance of the proposed diagnostic model, as shown in Figures 17 and 18. In Figures 17 (a) and 17 (b), the points corresponding to colors 0-4 on the right side represent the five labels for gear classification.

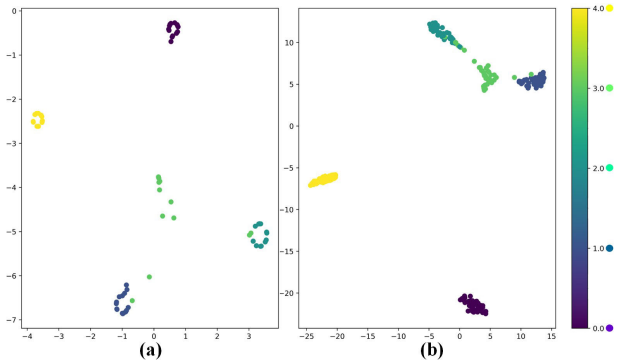


FIGURE 17. Visualization of self-made dataset classification results via t-SNE (a) Before data augmentation, (b)After data augmentation.

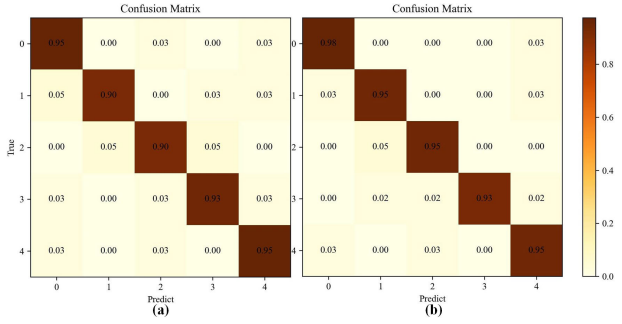
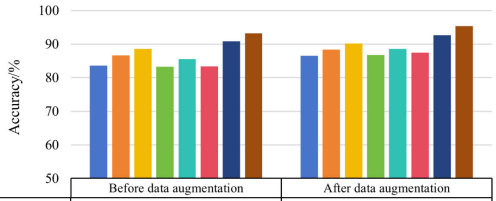


FIGURE 18. Classification confusion matrix of self-made dataset:(a) Before data augmentation, (b)After data augmentation.

As shown in Figure 17(a), before sample augmentation, the accuracy of fault identification is affected due to the scarcity of samples. In Figure 17(b), after augmentation, clear boundaries are established between the five categories, enabling effective fault classification. The results indicate



	Before data augmentation	After data augmentation
CNN-BiLSTM	83.63	86.52
WDCNN	86.65	88.36
SVM-Sigmoid	88.53	90.15
SVM-Linear	83.22	86.78
SVM-Rbf	85.47	88.63
SVM-Poly	83.37	87.42
Typical Transformer	90.87	92.63
Improved Transformer	93.25	95.36

FIGURE 19. Diagnostic accuracy of different models before and after data augmentation in self-made dataset.

that after data augmentation, the proposed model demonstrates robust learning capabilities and effectively learns the hidden features that distinguish between the five different types of faults, significantly improving fault classification performance.

From Figure 18, it is evident that after data augmentation, the classification accuracy of the network for normal bearings increased from 95% to 98%. Moreover, the classification accuracy for the four fault types, namely chipped tooth, missing tooth, root faults, and surface faults, all improved from an average of 92% to over 94.5%, achieving excellent classification performance. Comparing the results before and after data augmentation, there is a significant improvement in both overall classification accuracy and individual fault classification accuracy.

2) METHOD COMPARISON

To further elucidate the diagnostic performance of our method in detecting faults in tractor gearbox gears, comparative experiments were conducted with other diagnostic models. The average accuracies of different diagnostic models on the test set before and after data augmentation are depicted in Figure 19.

It is evident that the diagnostic model applied in this study performs similarly on the CWRU dataset as it does on the self-made dataset. After data augmentation, the diagnostic accuracy significantly improves compared to before augmentation. Following data augmentation, the Transformer model with improved multi-head self-attention can emerge as a significant solution for fault diagnosis of gear. This highlights the combined approach of the Time GAN and Transformer model proposed in this study has strong generalization ability in accurately diagnosing faults in tractor transmission systems even with limited fault samples.

VI. CONCLUSION

To address the problem of the tractor transmission system's low fault diagnosis accuracy resulting from a lack of fault samples during actual operating conditions, this paper proposes a fault diagnosis model based on the fusion of Time GAN and Transformer. The effectiveness of this method was validated using the CWRU bearing dataset and the self-made gear dataset. Comparative analysis was conducted in the experiments, contrasting the diagnostic results of existing neural networks such as CNN-BiLSTM, WDCNN, and SVM, serving as control groups for the proposed fault diagnosis method. The following are the primary results:

(1) Under the constraint of limited real fault sample data, the proposed method in this paper demonstrates excellent performance. Experimental results indicate that the Time GAN network can generate higher-quality data without substantially increasing the amount of data, thereby enhancing the fault diagnosis performance of various models under small-sample conditions.

(2) Through comparative experiments with models such as CNN-BiLSTM, WDCNN, and SVM, the improved Transformer diagnostic model proposed in this paper exhibits stronger capabilities in acquiring global information. Although the training time of our model is slightly longer than that of other models, it can accurately identify fault types under varying levels of noise and minimizes the misclassification of normal samples as certain types of fault samples, thereby enabling accurate and rapid diagnosis of faults occurring in tractor transmission systems during operation.

(3) The model for fault diagnostics put forth in this article combines Time GAN with Transformer, The Time GAN sample augmentation model generates high-quality fault samples through its excellent adversarial mechanism, resulting in balanced data. The augmented and balanced dataset is then fed into the Transformer classification model, which leverages the advantages of the Transformer model's encoder-decoder architecture and improved attention mechanism for feature extraction and pattern recognition. This process yields the final fault diagnosis classification results and addresses the issue of limited fault samples in actual tractor operating conditions while also exploring the potential of the Transformer model in the fault diagnosis field. In future research, we will continue to delve into the network optimization of sample generation models and diagnostic models, aiming to enhance the diagnostic accuracy and the model generalization ability under the condition of scarce samples in tractor transmission systems.

In conclusion, combining Time GAN with Transformer provides an accurate and efficient approach for fault diagnosis in tractor transmission systems. After data augmentation, the diagnostic accuracy for all fault types of bearings and gears listed in this paper has improved. This indicates that the method proposed in this paper has strong generalization ability and can more accurately identify fault types and locate fault positions in the fault diagnosis of tractor transmission systems, thereby improving diagnostic efficiency. Its excellent performance offers a new solution for fault diagnosis when fault data are scarce.

REFERENCES

[1] A. A. Dubaish and A. A. Jaber, "Comparative analysis of SVM and ANN for machine condition monitoring and fault diagnosis in gearboxes," *Math. Model. Eng. Problems*, vol. 11, no. 4, pp. 976–986, Apr. 2024.

[2] Y. Lei, B. Yang, and Z. Du, "Depth transfer diagnosis method of mechanical equipment failure under big data," *J. Mech. Eng.*, vol. 55, no. 7, pp. 1–8, 2019.

[3] A. A. Jaber, "Diagnosis of bearing faults using temporal vibration signals: A comparative study of machine learning models with feature selection techniques," *J. Failure Anal. Prevention*, vol. 24, no. 2, pp. 752–768, Apr. 2024.

[4] B. Huang, Y. Zhang, and L. Zhao, "Rolling bearing fault diagnosis method based on cuckoo search algorithm and maximum second-order cyclisation blind deconvolution," *J. Mech. Eng.*, vol. 57, no. 9, pp. 99–107, 2021.

[5] S. Dong, X. Pei, and W. Wu, "Fault diagnosis method of rolling bearings based on multi-layer denoising technology and improved convolutional neural network," *Mech. Eng.*, vol. 57, no. 1, pp. 148–156, 2021.

[6] X. Zhang, Y. Yuan, and X. Li, "Fault diagnosis of rolling bearings based on CF feature extraction and MBA-SVDD," *Mach. Tool Hydraulic Transmiss.*, vol. 50, no. 1, pp. 182–188, 2022.

[7] T. Zhou, T. Han, and E. L. Droguett, "Towards trustworthy machine fault diagnosis: A probabilistic Bayesian deep learning framework," *Rel. Eng. Syst. Saf.*, vol. 224, Aug. 2022, Art. no. 108525.

[8] L. Huang, S. Ma, and Y. H. Cao, "Research on bearing fault diagnosis based on GAF and Google Net," *Mach. Tool Hydraulic Transmiss.*, vol. 50, no. 1, pp. 193–198, 2022.

[9] D. Z. You, L. Chen, and Y. Zhang, "Research on fault diagnosis method of rolling bearings based on PCA-CNN," *Mach. Tool Hydraulic Transmiss.*, vol. 49, no. 19, pp. 172–177, 2021.

[10] W. X. Chen, X. X. Sun, and T. Wang, "Fault diagnosis for rolling bearings based on SPWVD-CNN," *Mach. Tool Hydraulics*, vol. 48, no. 12, pp. 147–154, 2020.

[11] Y. Tian and X. Bu, "Lumbar spine image segmentation method based on attention mechanism and Swin Transformer model," *Metrol. Meas. Technol.*, vol. 48, no. 12, pp. 57–61, 2021.

[12] Y. Wang, S. Ji, and B. Du, "Target detection algorithm for water column at sea impact point based on improved faster R-CNN," *Ordnance Equip. Eng.*, vol. 43, no. 6, pp. 182–189, 2022.

[13] G. Yu, P. Wu, Z. Lv, J. Hou, B. Ma, and Y. Han, "Few-shot fault diagnosis method of rotating machinery using novel MCGM based CNN," *IEEE Trans. Ind. Informat.*, vol. 19, no. 11, pp. 10944–10955, Nov. 2023.

[14] D. Li, Y. Liu, Y. Zhao, and Y. Zhao, "Fault diagnosis method for wind turbine planetary gearbox based on improved generative adversarial network," *Proc. Chin. Soc. Electr. Eng.*, vol. 41, no. 21, pp. 7496–7507, 2021.

[15] Y. Qin, H. Liu, and Y. Mao, "Faulty rolling bearing digital twin model and its application in fault diagnosis with imbalanced samples," *Adv. Eng. Informat.*, vol. 61, Aug. 2024, Art. no. 102513.

[16] Y. Ruan, M. Zheng, F. Qian, H. Meng, J. Yao, T. Xu, and D. Pei, "Fault detection and diagnosis of energy system based on deep learning image recognition model under the condition of imbalanced samples," *Appl. Thermal Eng.*, vol. 238, Feb. 2024, Art. no. 122051.

[17] W. Liao, L. Wu, S. Xu, and S. Fujimura, "A novel approach for intelligent fault diagnosis in bearing with imbalanced data based on cycle-consistent GAN," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–16, 2024.

[18] Y. Qin, H. Liu, Y. Wang, and Y. Mao, "Inverse physics–informed neural networks for digital twin–based bearing fault diagnosis under imbalanced samples," *Knowl.-Based Syst.*, vol. 292, May 2024, Art. no. 111641.

[19] Z. Pei, H. Jiang, X. Li, J. Zhang, and S. Liu, "Data augmentation for rolling bearing fault diagnosis using an enhanced few-shot wasserstein auto-encoder with meta-learning," *Meas. Sci. Technol.*, vol. 32, no. 8, Aug. 2021, Art. no. 084007.

[20] X. Wang, Z. Chu, B. Han, J. Wang, G. Zhang, and X. Jiang, "A novel data augmentation method for intelligent fault diagnosis under speed fluctuation condition," *IEEE Access*, vol. 8, pp. 143383–143396, 2020.

[21] G. Baasch, G. Rousseau, and R. Evins, "A conditional generative adversarial network for energy use in multiple buildings using scarce data," *Energy AI*, vol. 5, Sep. 2021, Art. no. 100087.

[22] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 11567–11573.

[23] D. Yang, X. Pan, and L. Mao, "Improved dense video description transformer decoding algorithm," *Comput. Eng. Appl.*, vol. 22, pp. 1–12, Nov. 2023.

[24] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need," in *Proc. 31st Conf. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, Dec. 2017.

[25] A. Dosovitskiy, L. Beyer, and A. Kolesnikov, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, New York, NY, USA, 2021.

[26] Z. Jiao, L. Pan, W. Fan, Z. Xu, and C. Chen, "Partly interpretable transformer through binary arborescent filter for intelligent bearing fault diagnosis," *Measurement*, vol. 203, Nov. 2022, Art. no. 111950.

[27] C. R. Chen, Q. Fan, and R. Panda, "CrossViT: Cross-attention multi-scale vision transformer for image classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, New York, NY, USA, Oct. 2021, pp. 347–356.

[28] Y. Ding, M. Jia, Q. Miao, and Y. Cao, "A novel time–frequency transformer based on self–attention mechanism and its application in fault diagnosis of rolling bearings," *Mech. Syst. Signal Process.*, vol. 168, Apr. 2022, Art. no. 108616.

[29] Y. Jin, L. Hou, and Y. Chen, "A time series transformer based method for the rotating machinery fault diagnosis," *Neurocomputing*, vol. 494, pp. 379–395, Jul. 2022.

[30] S. Zhu, B. Liao, Y. Hua, C. Zhang, F. Wan, and X. Qing, "A transformer model with enhanced feature learning and its application in rotating machinery diagnosis," *ISA Trans.*, vol. 133, pp. 1–12, Feb. 2023.

[31] Z. Fu, Z. Fu, Q. Liu, W. Cai, and Y. Wang, "SparseTT: Visual tracking with sparse transformers," in *Proc. 31st Int. Joint Conf. Artif. Intell.*, New York, NY, USA, Jul. 2022.

[32] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, U.K., Aug. 2020, pp. 213–229.

[33] P. Niki, A. Vaswani, J. Uszkoreit, L. Kaiser, N. Shazeer, A. Kolesnikov, and D. Tran, "Image transformer," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Stockholm, Sweden, Jul. 2018, pp. 4055–4064.

[34] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 9992–10002.

[35] H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, and H. Jégou, "Going deeper with image transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 32–42.

[36] X. Du, L. Jia, and I. Ul Haq, "Fault diagnosis based on SPBO-SDAE and transformer neural network for rotating machinery," *Measurement*, vol. 188, Jan. 2022, Art. no. 110545.

[37] A. Zhao, K. Subramani, and P. Smaragdis, "Optimizing short-time Fourier transform parameters via gradient descent," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Toronto, ON, Canada, Jun. 2021, pp. 736–740.

[38] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Lille, France, 2015, pp. 448–456.

[39] Y. Li, X. Wang, S. Si, and S. Huang, "Entropy based fault classification using the Case Western Reserve University data: A benchmark study," *IEEE Trans. Rel.*, vol. 69, no. 2, pp. 754–767, Jun. 2020.

[40] K. X. Cao and C. Li, "Diagnosis and analysis of spindle faults of power plant units based on WPD-T-SNE-SVM method," *Machinery Manuf. Autom.*, vol. 52, no. 6, pp. 226–228, 2023.

[41] B. Song, Y. Liu, J. Fang, W. Liu, M. Zhong, and X. Liu, "An optimized CNN-BiLSTM network for bearing fault diagnosis under multiple working conditions with limited training samples," *Neurocomputing*, vol. 574, Mar. 2024, Art. no. 127284.

[42] J. Zhou, X. Zhang, H. Jiang, Z. Shao, B. Ma, and R. Zhou, "MC-WDWCNN: An interpretable multi-channel wide-kernel wavelet convolutional neural network for strong noise-robust fault diagnosis," *Meas. Sci. Technol.*, vol. 35, no. 9, Sep. 2024, Art. no. 096125.

[43] X. Yang, T. Zhang, and Y. Li, "Fault diagnosis of motor rolling bearings based on ISSA optimized SVM," *Electron. Meas. Technol.*, vol. 46, no. 15, pp. 186–192, 2023.

LIYOU XU (Member, IEEE) received the B.S. degree in mechanical manufacturing technology and equipment from Jiaozuo Institute of Technology, Jiaozuo, China, in 1998, the M.S. degree in vehicle engineering from Luoyang Institute of Technology, Luoyang, China, in 2001, and the Ph.D. degree in vehicle engineering from Xi'an University of Technology, Xi'an, in 2007. Since 2013, he has been a Professor with the College of Vehicle and Traffic Engineering, Henan University of Science and Technology, Luoyang. His research interests include new transmission theory and control technology, vehicle performance analysis method and simulation technology, and low-speed electric vehicle transmission technology.

GUODONG ZHANG (Member, IEEE) received the B.S. degree in vehicle engineering from Henan University of Engineering, Zhengzhou, China, in 2022. He is currently pursuing the master's degree in vehicle engineering with Henan University of Science and Technology, Luoyang, China. His research interest includes fault diagnosis of agricultural machinery.

SIXIA ZHAO (Member, IEEE) received the B.S. degree in mechanical engineering and automation from Henan Polytechnic University, Jiaozuo, China, in 2013, and the M.S. degree in mechanical engineering and the Ph.D. degree in vehicle engineering from Henan University of Science and Technology, Luoyang, China, in 2018 and 2022, respectively. Since 2022, he has been a Lecturer with the College of Vehicle and Traffic Engineering, Henan University of Science and Technology. His research interests include the fault diagnosis of agricultural machinery and the new transmission system of agricultural vehicles.

YIWEI WU (Member, IEEE) received the Ph.D. degree in vehicle engineering from Henan University of Science and Technology, Luoyang, China, in 2022, where he has been a Lecturer with the College of Vehicle and Traffic Engineering, since 2023. His research interests include digital design and virtual test of agricultural vehicles, man-machine systems designs, collaborative modeling, and simulation of multidisciplinary virtual prototyping. He is a member of the CSAE and CSAM.

ZHIQIANG XI (Member, IEEE) received the Ph.D. degree in mechanical engineering from Xi'an University of Technology, Xi'an, China. Since 2016, he has been a Lecturer with the College of Vehicle and Traffic Engineering, Henan University of Science and Technology. His research interests include vehicle transmission systems and their intelligent control technology. He is a member of the CSAE and CSAM.