International Journal of Mechanical and Electrical Engineering
ISSN: 3005-9615 (Print), ISSN: 3005-7132 (Online) | Volume 3, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijmee.v3n2.04
Journal homepage: https://wepub.org/index.php/IJMEE/index

CLS GAN: Integrating Autoencoders and Transformers for Enhanced Bearing Fault Diagnosis

Mengmeng Ping *

College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, China

* Corresponding author: Mengmeng Ping

ABSTRACT

Bearing fault diagnosis with limited samples is a key challenge in the field of intelligent manufacturing, necessitating the development of models capable of accurate learning from constrained data with strong generalization capabilities. This study proposes a novel framework combining autoencoders and generative adversarial networks, termed the Conditional Latent Space Generative Adversarial Network (CLS GAN), which utilizes autoencoders to learn the latent data distribution of signals, effectively capturing and reproducing the complexity of fault signals. Enhanced with an improved Transformer structure, this model is able to process and recognize long temporal features between signal segments, thereby boosting the accuracy and efficiency of fault diagnosis. Through the architecture of a Conditional GAN, a multi-class task discriminator is implemented, enabling effective fault type discrimination under conditions of limited samples. In situations where samples are restricted, the proposed CLS GAN model achieved an accuracy of 75% on the CWRU dataset, demonstrating the efficacy and practicality of an integrated framework that combines advanced generative adversarial networks and Transformer technology in mechanical fault diagnosis.

KEYWORDS

Few-shot Learning; Generative Adversarial Network; Autoencoder; Transformer.

1. INTRODUCTION

As the proportion of modern large-scale equipment in the manufacturing industry continues to rise, the increasing scale of this equipment also enhances the complexity involved in analyzing and diagnosing faults when they occur. Intelligent fault diagnosis is a technology that determines whether the whole or a part of a machine is functioning normally or abnormally based on its operational state. Bearings represent a crucial component within the study of rolling element fault diagnosis[1]. Although structurally simple, the causes of bearing failures are multifaceted, including uncertain position fractures of the inner and outer rings, wear between rolling elements and raceways, and variations in bearing materials. Early researchers primarily relied on empirical methods, diagnosing faults through monitoring temperature, noise, deformation, and other characteristics. However, these methods often struggle to detect minor internal faults in bearings and frequently lead to false alarm issues.

In recent years, with the rapid development of deep learning, breakthroughs have been made in the research on fault diagnosis of rolling element equipment. Zhang et al. [2] proposed a method of converting original signals into two-dimensional images and then introduced an intelligent diagnostic algorithm based on convolutional neural networks (CNNs). Sun et al. [3] transformed vibration

signals into polar coordinate symmetric images using the Symmetric Dot Pattern (SDP) principle, and then input these SDP images into the input layer of a convolutional neural network to determine the CNN model based on new metrics involving accuracy and time ratio. Song et al. [4] generated more input data by expanding and used wide kernels in the first two convolutional layers to rapidly extract features to enhance efficiency. Han et al. [5] combined the excellent feature processing capabilities of convolutional neural networks with the superior generalization ability of support vector machines (SVMs). Sun et al. [6] proposed a novel bearing fault diagnosis method based on empirical mode decomposition (EMD) and an improved Chebyshev distance. However, in practical environments, the labeled bearing fault data is very limited, and the aforementioned diagnostic methods are all based on models trained with a sufficient amount of data.

The task of bearing fault diagnosis with few samples is currently an important research objective. The core challenge of this task is to require the model to not only learn the fault pattern features from limited data but also to possess strong generalization capabilities to adapt to varied operating conditions and environments, which is the main challenge faced by the task of bearing fault diagnosis with few samples. Li et al. [7] proposed a novel model-agnostic meta-learning fault diagnosis method (MLFD) to address this problem. Wang et al. [8] used the similarity of sample pairs for classification rather than end-to-end classification. Wang et al. [9] introduced an Automatic Embedding Transformer (AET) method to achieve interpretable multiple fault diagnoses for rolling bearings. Zhang et al. [10] proposed a small sample learning framework for bearing fault diagnosis based on model-agnostic meta-learning, aimed at using limited data to train effective fault classifiers. Ma et al. [11] proposed an Unsupervised Domain Adaptation (UDA) method with enhanced transferability and discriminability (ETDS-UDA) for few-sample diagnosis of high-speed train bearing faults. Che et al. [12] presented an Integrated Meta-Learning (EML) model for few-sample fault diagnosis of rolling bearings. Fu et al. [13] introduced a semi-supervised prototype network with a dual-stream wavelet scattering convolutional encoder based on roller state signals (TWSCE-SSPN). These methods can achieve a certain degree of fault diagnosis accuracy with few samples, yet we aim to find a method that learns the latent distribution of fault sample feature space to achieve effective fault diagnosis with fewer samples.

Based on the aforementioned research and analysis, we propose a Conditional Latent Space Generative Adversarial Network (CLS GAN) method for small-sample fault diagnosis. In the generator, an autoencoder is used to learn the latent data distribution of signals, and in the discriminator, an improved Transformer structure captures long temporal feature information between signal segments. Further, a multi-class discriminator structure is implemented through the architecture of Conditional GAN. Overall, our work is mainly reflected in the following points:

(1) For bearing signal characteristics, a U-net-based autoencoder structure is proposed, which can effectively learn the latent distribution of the signal feature space.

(2) For bearing signal data, a multi-classifier discriminator is established using an improved Transformer structure, providing reliable fault diagnosis during adversarial training.

(3) This method has been validated on the CWRU dataset.

2. RELATED WORK

2.1. Fault Diagnosis

Currently, researchers have applied various deep learning methods to the problem of fault diagnosis, achieving excellent results. Among them, data-driven fault diagnosis involves inputting data collected by sensors into a diagnostic model to determine the labels (types of faults) of the signals. Definition: A sequence of sampled points of length n obtained by sensors is used as time-series data

$X:\{x_1,x_2,x_3,\cdots,x_n\}$; the mapping $f:y'=f(X)$, where x_i represents the data collected by the sensor at moment i, and the number of channels of x_i depends on the number of sensors, with vibration sensors in this article generally having one channel (unless otherwise specified); y' is the predicted label output by the model.

Common deep learning units used to construct fault diagnosis models include linear (Linear) fully connected layers of multilayer perceptron (MLP), layers of convolutional neural network (CNN) such as convolution layers and pooling layers, and units of recurrent neural network (RNN) such as basic RNN cells, long short-term memory (LSTM), and gated recurrent units (GRU).

The states of bearings include: Normal (N) state and three fault states as shown in Figure 1. In the implementation of deep learning methods for bearing fault diagnosis, it is usually necessary to preprocess the collected vibration signal data to enhance the performance and accuracy of the diagnostic model. Preprocessing steps include denoising, standardization, or normalization, which help eliminate noise interference and standardize the range of input data to make it suitable for the input requirements of deep learning models.

(a) outer rate(OR) (b) inner rate(IR) (c) ball(B)

Figure 1. Diagram of three types of bearing faults.

2.2. Generative Adversarial Network

A Generative Adversarial Network (GAN) consists of a generator and a discriminator. The generator creates realistic fault samples from noise input, while the discriminator is used to distinguish between the generated samples and real samples. The two components engage in a game-theoretic contest where the generator continuously improves the realism of the samples, ultimately producing high-quality fault data.

Based on this, GANs can enhance the diversity of fault data and reduce the impact of data imbalance on diagnostic models. Moreover, they do not rely on real fault data and can simulate rare or unprecedented types of faults. This generative technique can address the scarcity of fault data by creating fault samples, thereby improving the training effectiveness of machine learning models.

3. CONDITIONAL LATENT SPACE GENERATIVE ADVERSARIAL NETWORK

In scenarios with limited fault data, models tend to memorize specific features of these samples rather than learning more general, generalizable patterns. Small datasets may not cover all possible manifestations of faults, making it difficult for the model to respond to new or slightly different fault conditions. Generative Adversarial Networks (GANs) can generate additional, realistic training samples, helping to expand the fault dataset. In this way, GANs can also learn and extract complex distributions within the data, deriving useful feature representations from limited samples.

To address this, the study proposes a novel framework called Conditional Latent Space Generative Adversarial Network (CLS GAN) that combines autoencoders with generative adversarial networks to learn the latent data distribution of signals. This model integrates the traditional GAN architecture with the recently popular Transformer model to process and generate signal data. By iteratively training the generator and discriminator, this architecture effectively enhances model performance.

3.1. Generator Structure

As shown in the Figure 2, the structure of the generator is described. The generator takes noise $\mathbf{z}$ as input, which is fed into a linear layer that maps the noise vector to a higher-dimensional space. The output of the linear layer is reshaped by a fully connected layer to adapt to subsequent convolutional layers. The convolution module, consisting of multiple layers of convolution and upsampling layers, gradually constructs the output of the target dimension. Here, the upsampling layers increase the dimensions to the target size, using Batch Normalization (BatchNorm1d) and LeakyReLU activation functions to enhance the stability and efficiency of model training. The output is finally processed through a Tanh activation function to ensure that the output values are between -1 and 1, matching the preprocessing range of the signal data.

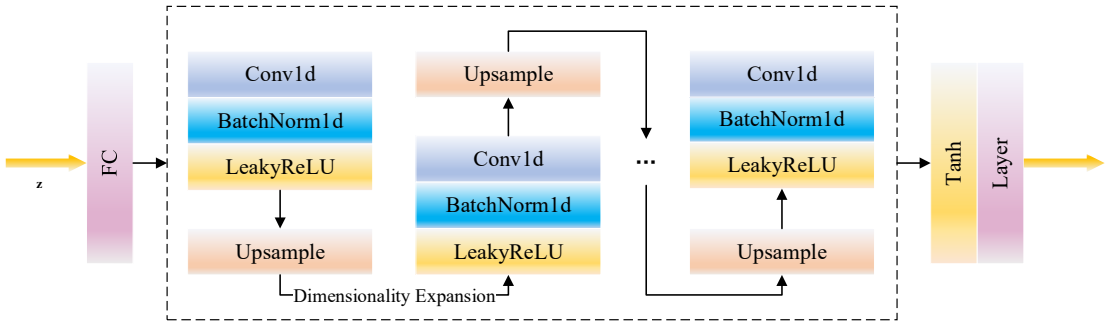

Figure 2. Generator structure diagram

The objective of the generator is to minimize the loss between the generated data and the real data, as shown in equation (1).

$$L_G = -E_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[\log D(G(\mathbf{z}))] \tag{1}$$

The data generated by the generator from the noise $\mathbf{z}$ is denoted $G(\mathbf{z})$, and $p(\mathbf{z})$ represents the distribution of the noise. As the generator improves over time, it becomes better at understanding and utilizing the latent space to produce increasingly realistic data. This improvement is reflected in the fluctuations of the generator's loss, which may indicate that the generator is exploring new effective areas in the latent space to deceive the discriminator.As an emerging form of digital currency, the independent operational architecture of digital renminbi is still evolving. There are many legal gaps in regulatory measures at the legal level in its issuance and circulation. Currently, the legal tender in China is mainly traditional paper currency, and digital renminbi has not yet been fully incorporated into the legal currency regulatory framework. For the counterfeiting of paper currency and coins, banks have established a comprehensive identification and tracking mechanism.

3.2. Discriminator Structure

As depicted in the Figure 3, the structure of the discriminator is comprised of convolutional layers and a Transformer. The convolutional layers are made up of multiple stacked layers of one-dimensional convolutions and max pooling. After inputting generated and real samples, a series of convolutional layers are used to extract features from the input signals. The max pooling layers reduce the dimensionality of the features, which facilitates the Transformer's understanding and representation of key segments.

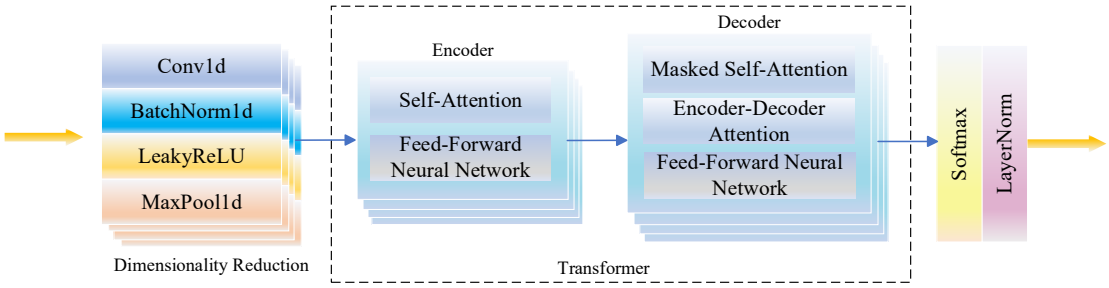

Figure 3. Discriminator structure diagram

The Transformer employs a self-attention mechanism to process the features, enhancing the understanding and expression of critical parts in the signal. The true or false nature of the signal is output through a linear layer followed by a sigmoid activation function, which also predicts the category of the output signal.

The discriminator's objective is to minimize the error rate, as shown in equation (2).

$$L_D = -E_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log D(\mathbf{x})] - E_{\mathbf{z}\sim p_z(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \tag{2}$$

The output $D(\mathbf{x})$ represents the discriminator's judgment result for real data, where p_{data} is the distribution of the real data. The discriminator, by identifying the differences between real and generated data, forces the generator to continuously improve its ability to represent data in the latent space. As the discriminator's ability enhances, the generator must seek more complex or indistinguishable ways of data representation in the latent space.

3.3. Training

Generative Adversarial Networks (GANs) consist of two parts: the Generator and the Discriminator. The objective of the Generator is to produce fake data that is as close as possible to real data, while the objective of the Discriminator is to distinguish between real data and the fake data generated by the Generator. These two networks enhance each other's performance through an adversarial process, with the Generator attempting to "deceive" the Discriminator, and the Discriminator striving not to be deceived.

The training of a GAN can be viewed as a two-player zero-sum game, where the Discriminator aims to maximize its ability to differentiate between real and generated data, and the Generator aims to minimize the Discriminator's ability to distinguish. This involves both adversarial and cooperative interactions between the Discriminator and the Generator. The training process of a GAN is described through a minimax game.

The objective of the Discriminator is to maximize the accuracy of discriminating real data while minimizing the misjudgment rate of generated data. The loss function of the Discriminator is composed of two parts: one is the loss on real data, and the other is the loss on generated data.

The objective of the Generator is to minimize the probability that its generated data is mistakenly identified as fake by the Discriminator. During the training process of a GAN, the optimization goals of the Generator and the Discriminator are opposed to each other. The success of the Generator (producing realistic data) implies the failure of the Discriminator (unable to differentiate between real and generated data), and vice versa. The training of a GAN can be seen as a minimax problem, where the Generator tries to minimize the probability of being identified as fake by the Discriminator, while the Discriminator tries to maximize its ability to distinguish between real and generated data.

This framework ensures that the Generator and Discriminator drive each other during the training process, eventually reaching a dynamic equilibrium where the Generator can produce high-quality data, and the Discriminator can effectively differentiate between real and generated data.

4. EXPERIMENT

4.1. Data

We utilized the publicly available dataset from Case Western Reserve University (CWRU), which provides vibration signal data of bearings operating under various working conditions, for the development and testing of bearing fault detection and diagnostic algorithms. The data acquisition platform and process are illustrated in Figure 4.

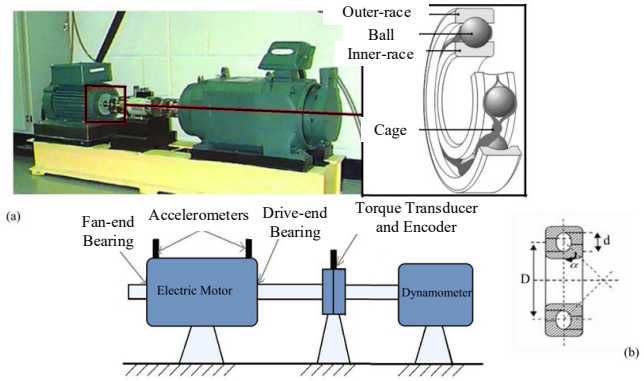

Figure 4. CWRU Data Collection Platform

The experimental platform is equipped with a 2-horsepower motor, precisely controlled by an electronic controller. Key components in the experiment include a torque sensor/decoder and a power meter, placed on either side of the motor, used to measure torque and motor efficiency, respectively. Additionally, the test bench also examined two types of bearings: SKF6205 at the drive end and SKF6203 at the fan end, with vibration data sampled at frequencies of 12 kHz and 48 kHz, respectively.

The parameter settings of the platform are shown in Table 1.

Table 1. Dataset parameters

Fault type	Fault diameter (10^{-3} inch)	Number of samples
Normal samples	0	4
Outer race fault	7	77
	14	
	21	
Inner race fault	7	40
	14	
	21	
Rolling element fault	7	40
	14	
	21	

The dataset encompasses samples of the three main types of bearing faults: outer race, inner race, and rolling element faults. These fault samples are characterized by fault diameters of 0.007 inches, 0.014 inches, and 0.021 inches, providing 77 outer race fault samples, 40 inner race fault samples, and 40 rolling element fault samples. In addition, there are 4 normal operation samples included for comparative analysis.

4.2. Training

During the diagnostic process, to better facilitate the adversarial interplay between the generator and discriminator for improved diagnostic results, the following parameter settings were implemented.

Table 2. Training parameters

Parameters	value
input_dim	401
num_channels	1
num_classes	4
latent_dim	100
output_dim	401
seq_len	401
d_model	128
nhead	8
num_epochs	200
batch_size	32

As shown in Table 2, the latent dimension is set to 100, allowing the generator to learn complex data distributions from high-dimensional space and enhance the diversity of generated data. The output dimension matches the signal length of 401, ensuring consistency in the structure of generated data. A feature count of 128 allows the Transformer network to capture complex features of the data, while 200 training epochs ensure the model has sufficient time to adapt to and learn the adversarial structure, balancing learning depth with computational efficiency.

In this section of the experiment, we calculated the losses for both the generator and discriminator and visualized these in a loss graph. The training loss graph is shown in Figure 5, where the red dashed line represents the Generator Loss and the blue line represents the Discriminator Loss.

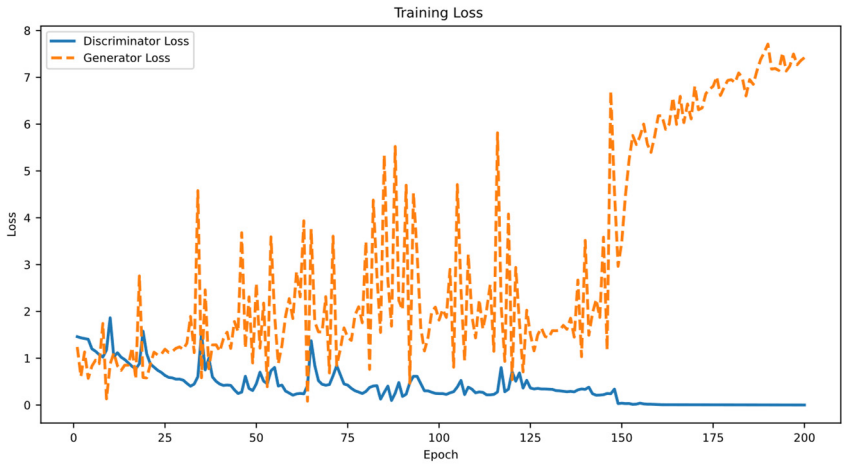

Figure 5. CWRU data collection platform

In the displayed Figure 5, the discriminator's loss (blue line) generally shows a downward trend, despite some fluctuations, indicating a gradual improvement in the discriminator's performance. Meanwhile, the generator's loss (orange dashed line) displays significant variability, particularly with several spikes around the 75th, 100th, and 150th cycles. These spikes may indicate that the generator has discovered new strategies to deceive the discriminator at these points, necessitating an adjustment in the discriminator's strategy to accommodate the generator's new tactics, temporarily increasing the loss.

These fluctuations and spikes exemplify the typical adversarial training process in GANs, with both models gradually improving through constant competition. The discriminator shows a more stable progression relative to the generator, which exhibits greater variability. This dynamic reflects the expected characteristics of GAN training, where each model continuously adapts to the other's strategies.

The continuous decline in discriminator loss depicted in the graph demonstrates the discriminator's gradual enhancement in distinguishing between real and generated data. Conversely, the variability in generator loss reflects how the generator explores the latent space to improve the quality of its generated data. This interactive learning and adaptation process highlights the powerful capabilities of GANs in complex data generation tasks, especially in applications requiring the simulation of highly nonlinear and variable data distributions.

To optimize payment networks, collaboration with other central banks and payment clearing institutions is essential to jointly promote the construction of the digital renminbi cross-border payment network. Establishing bilateral or multilateral central bank digital currency payment systems to provide legal protection for digital renminbi cross-border payments is crucial. Finally, to mitigate the impact on traditional financial systems, promoting the integration of traditional banks with digital renminbi and establishing mechanisms for payment system reform and integration to ensure the smooth transition of payment systems and stable operation of financial markets are necessary.

4.3. Experimental Results

To enhance computational efficiency and prevent any modification to the weights during the evaluation process, a gradient computation disabling context manager is used to halt gradient calculations for any computation involved in the model's forward propagation. For each batch's output, the classification loss is computed using Cross-entropy loss. The loss measures the distance between the actual class labels and the predicted probability distributions of class labels, as shown in the formula (3). The accumulated loss values are then used to assess the average performance of the entire test set.

$$L = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c}) \tag{3}$$

In the formula, N represents the number of samples, and C denotes the total number of categories. $y_{i,c}$ is a one-hot encoded vector where if sample i belongs to category c, then $y_{i,c}=1$; otherwise, $y_{i,c}=0$. Accuracy is a commonly used metric to measure model performance, which represents the ratio of the number of samples correctly predicted by the model to the total number of samples.

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} 1(y_i = \hat{y}_i) \tag{4}$$

The curves depicting changes in loss and accuracy are shown in Figure 6.

This graph displays the changes in test loss and accuracy of the model over 200 training epochs. Initially, the test loss exhibits significant fluctuations but generally shows a downward trend as training progresses, especially stabilizing after 100 epochs. Between 100 and 125 epochs, there was a notable spike in loss, possibly reflecting temporary instability in the model due to certain parameter adjustments or specific test data. Simultaneously, the accuracy gradually improved from the start of training and stabilized at a higher level after 75 epochs, demonstrating the model's gradual maturation and adaptation process, consistently remaining around 75%. This performance reflects the effectiveness of model optimization and parameter adjustments, and despite some fluctuations, the overall trend indicates continuous improvement in performance.

5. CONCLUSION

This study discusses the transition from traditional monitoring methods to fault diagnosis techniques based on deep learning models, with a particular emphasis on how to effectively enhance diagnostic accuracy and the generalizability of models under limited sample conditions. By integrating GANs

and Transformer architectures, this research demonstrates the practical application and potential of this approach in improving the efficiency of fault diagnosis with few samples.

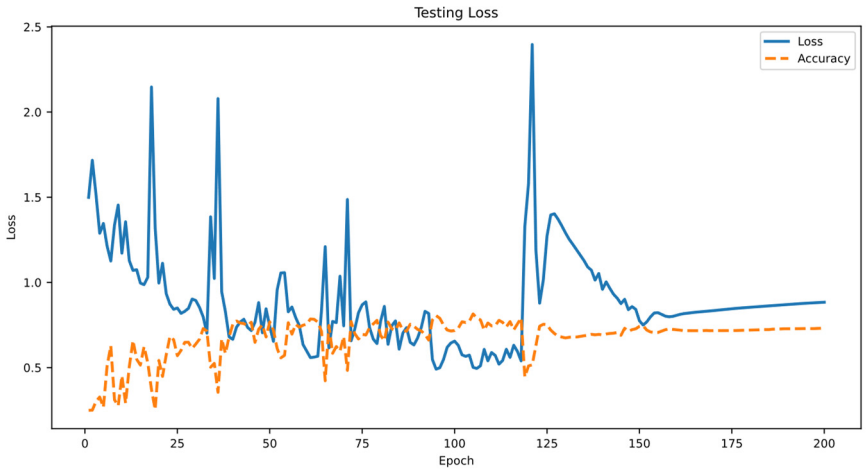

Figure 6. Cross-entropy loss and accuracy curves over training

Experimental results using the CWRU dataset have validated the proposed model and indicated that it can significantly optimize the performance of fault diagnosis. These findings not only confirm the effectiveness of deep learning technologies in handling complex and dynamic fault pattern recognition but also showcase their practical value in industrial applications.

The deep learning framework, especially when integrated with cutting-edge algorithms and model structures, provides an efficient and reliable solution for mechanical system fault diagnosis. Looking forward, the paper suggests continuing to optimize the algorithmic architecture and enhance model training efficiency, while also exploring intelligent fault diagnosis technologies applicable to other types of mechanical systems.

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] ZHANG X, ZHAO B, LIN Y. Machine learning based bearing fault diagnosis using the case western reserve university data: a review [J]. IEEE Access, 2021, 9: 155598-608.

[2] ZHANG J, YI S, LIANG G, et al. A new bearing fault diagnosis method based on modified convolutional neural networks [J]. Chinese Journal of Aeronautics, 2020, 33(2): 439-47.

[3] SUN Y, LI S. Bearing fault diagnosis based on optimal convolution neural network [J]. Measurement, 2022, 190: 110702.

[4] SONG X, CONG Y, SONG Y, et al. A bearing fault diagnosis model based on CNN with wide convolution kernels [J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(8): 4041-56.

[5] HAN T, ZHANG L, YIN Z, et al. Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine [J]. Measurement, 2021, 177: 109022.

[6] SUN Y, LI S, WANG X. Bearing fault diagnosis based on EMD and improved Chebyshev distance in SDP image [J]. Measurement, 2021, 176: 109100.

[7] LI C, LI S, ZHANG A, et al. Meta-learning for few-shot bearing fault diagnosis under complex working conditions [J]. Neurocomputing, 2021, 439: 197-211.

[8] WANG C, XU Z. An intelligent fault diagnosis model based on deep neural network for few-shot fault diagnosis [J]. Neurocomputing, 2021, 456: 550-62.

[9] WANG G, LIU D, CUI L. Auto-embedding transformer for interpretable few-shot fault diagnosis of rolling bearings [J]. IEEE Transactions on Reliability, 2023.

[10] ZHANG S, YE F, WANG B, et al. Few-shot bearing fault diagnosis based on model-agnostic meta-learning [J]. IEEE Transactions on Industry Applications, 2021, 57(5): 4754-64.

[11] MA W, ZHANG Y, MA L, et al. An unsupervised domain adaptation approach with enhanced transferability and discriminability for bearing fault diagnosis under few-shot samples [J]. Expert Systems with Applications, 2023, 225: 120084.

[12] CHE C, WANG H, XIONG M, et al. Few-shot fault diagnosis of rolling bearing under variable working conditions based on ensemble meta-learning [J]. Digital Signal Processing, 2022, 131: 103777.

[13] FU X, TAO J, JIAO K, et al. A novel semi-supervised prototype network with two-stream wavelet scattering convolutional encoder for TBM main bearing few-shot fault diagnosis [J]. Knowledge-Based Systems, 2024, 286: 111408.