ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems Part B: Mechanical Engineering
Online journal at:
asmedigitalcollection.asme.org/risk

Chaojun Xu
School of Qilu Transportation,
Shandong University,
17923 Jingshi Road, Lixia District,
Jinan, Shandong 250061, China

Yaqiang Jin[1]
School of Qilu Transportation,
Shandong University,
17923 Jingshi Road, Lixia District,
Jinan, Shandong 250061, China
e-mails: yaqiang.jin@sdu.edu.cn; yaqiang.jin@outlook.com

Junxiao Ma
College of Engineering,
Shantou University,
243 Daxue Road, Jinping District,
Shantou, Guangdong 515063, China

Yuhao Wu
College of Engineering,
Shantou University,
243 Daxue Road, Jinping District,
Shantou, Guangdong 515063, China

Peng Chen[1]
College of Engineering,
Shantou University,
243 Daxue Road, Jinping District,
Shantou, Guangdong 515063, China
e-mails: pengchen@alu.uestc.edu.cn; dr.pengchen@foxmail.com

Ge Xin
School of Traffic and Transportation,
Beijing Jiaotong University,
No. 3 Shangyuan Village, Haidian District,
Beijing 100091, China

Changbo He
College of Electrical Engineering and Automation,
Anhui University,
No. 3 Feixi Road, Shushan District,
Hefei, Anhui 230039, China

CyclicalNet: A Dense-Connected Architecture for Multiscale Cyclic Feature Learning in Planetary Gearbox Fault Diagnosis

Planetary gearboxes, crucial for industrial applications like wind turbines, automobiles, and robotics, are susceptible to failures due to complex vibrations and harsh conditions, making the development of effective fault diagnosis methods essential. While general-purpose deep learning models—such as Transformers, ResNets, DenseNets, and generative adversarial networks (GANs)—have shown promise in identifying intricate patterns within time series data for fault diagnosis, conventional one-dimensional convolutional kernels have inherent limitations, as they can only capture features from adjacent time series points. Additionally, methods that utilize attention mechanisms often struggle to uncover consistent dependencies among sparse data points that may be concealed within complex patterns. This paper introduces CyclicalNet, an innovative deep learning architecture designed to address these fundamental limitations in planetary gearbox fault diagnosis. The architecture uniquely integrates cyclical blocks, encoder blocks, and dense connections to process high amplitude–frequency components, traditionally overlooked in conventional models. A strategic reshape operation transforms one-dimensional sequential data into two-dimensional representations, enabling more structured and efficient feature extraction while facilitating simultaneous modeling of temporal and periodic features. The incorporation of dense connections promotes multiscale feature learning and enhances information flow across the network, resulting in richer and more discriminative feature representations. Through comprehensive experimental evaluation on planetary gearbox datasets, CyclicalNet demonstrates superior performance in fault diagnosis accuracy compared to existing methods, establishing a new benchmark in predictive maintenance technology. [DOI: 10.1115/1.4069653]

Keywords: planetary gearbox, fault diagnosis, vibration analysis, feature extraction, deep learning

1 Introduction

Planetary gearboxes are essential components in wind turbines, cars, and robotics, but when they fail, it can lead to costly repairs and extended downtime [1–3]. Their failure can come from several factors, such as the complex vibration patterns created by varying loads, the interactions among multiple planet gears that can lead to interference and modulated vibrations, adverse operating conditions due to high loads, elevated temperatures, and contamination, as well as the effects of faults interacting with other components in the drivetrain.

Traditional condition monitoring of gearboxes has heavily relied on various vibration analysis techniques for fault diagnosis. These techniques include frequency-domain methods, such as fast Fourier transform [4], time–frequency approaches like wavelet transforms [5], and cepstrum analysis [6]. However, conventional signal processing methods exhibit limitations in managing the complex

[1]Corresponding authors.

Manuscript received July 2, 2025; final manuscript received August 19, 2025; published online March 5, 2026. Assoc. Editor: Ke Feng.

nonlinear and nonstationary vibration signals generated in the presence of faults [7]. In particular, the hand-crafted feature extraction associated with these methods often lacks robustness under varying loading conditions. Furthermore, simplified physics-based models [8] struggle to accurately represent the intricate fault-induced vibrations observed in gearboxes. In conclusion, while existing vibration signal processing methods play a crucial role in condition monitoring, they are constrained by their reliance on hand-crafted feature engineering and overly simplistic physics-based models. These limitations could potentially be addressed through the application of advanced techniques capable of identifying reliable characteristics while more accurately representing the underlying physics of gearbox operation.

Moreover, traditional machine learning approaches, such as artificial neural networks, have frequently been employed for gearbox fault diagnosis due to their pattern recognition capabilities. Techniques such as convolutional neural networks and deep belief networks have been utilized to extract features and diagnose faults specifically in planetary gearboxes [9]. Nonetheless, these methods significantly depend on hand-crafted feature engineering, which often lacks resilience to changes in operational scenarios. In addition, oversimplified approximations of gearbox dynamics are commonly employed alongside physics-based models to aid in the interpretation of vibration measurements. This practice further limits the model's ability to accurately correlate vibration signals with component degradation [10,11]. Overall, despite the widespread application of traditional machine learning approaches, these methods continue to be constrained by the necessity of manual feature extraction and simplified representations of gearbox physics. To mitigate these shortcomings and provide a reliable diagnosis of planetary gearbox defects, there is a pressing need for more sophisticated techniques.

Deep learning has the potential to mitigate existing obstacles in gearbox fault diagnosis by automatically learning robust characteristics associated with gearbox defects. By directly utilizing raw vibration data for feature extraction, deep learning networks are capable of modeling the intricate fault signatures that occur under varying load conditions. This capacity for automated feature learning enables deep learning algorithms to detect subtle changes in vibration patterns that may signal the early stages of failure. Therefore, deep learning demonstrates significant promise for achieving accurate and early fault diagnosis in gearboxes, thereby addressing the challenges that traditional methods encounter when handling complex fault-induced vibration signals. The data-driven feature extraction approach from raw data provides deep learning networks with a distinct advantage in managing nonlinear and nonstationary vibrations. In their study, Li et al. [12] proposed a joint deep learning model that integrates multidomain features derived from vibration, acoustic emission, and other relevant sources to learn reliable gear defect characteristics. The fusion of data from multiple sources enhances diagnostic accuracy when compared to single-domain models. Similarly, Koutsoupakis et al. [13] utilized raw supervisory control and data acquisition (SCADA) and vibration data to train deep neural networks, thus achieving robust feature extraction that is capable of handling complex fault signatures under fluctuating loads. Furthermore, Perez-Sanjines et al. [14] developed a convolutional neural network-long short-term memory (CNN–LSTM) model that leverages SCADA data to accurately predict impending failures in wind turbines, such as cracked gears and bearing faults. Various additional deep network architectures, including autoencoders [15–17] and generative adversarial networks (GANs) [18,19], have also been investigated for gearbox fault diagnosis. Specifically, Guo et al. employed a stacked autoencoder model combined with a softmax classifier to perform unsupervised feature learning from vibration data, which led to improved defect detection for sun, planet, and ring gears. GANs have also shown promise in addressing issues related to imbalanced fault data, as demonstrated by Wang et al. [19]. In addition to these neural network approaches, deep transfer learning has garnered significant interest for fault diagnosis purposes. For instance, Chen et al. [20] developed a transfer learning-based methodology utilizing CNN and deep belief network models that were pretrained on image data.

Artificial general network architectures, such as Transformers [21–24], ResNets [25], DenseNets [26], and GANs [18], have exhibited promising capabilities for fault diagnosis applications. Jiang et al. [27] developed a transformer-based model known as SAnD, which applies self-attention mechanisms to extract spatio-temporal features from multisensor data for bearing fault classification. In a related effort, Wu et al. [28] introduced a visual–audio transformer that integrates both image data and acoustic signals to facilitate robust gear fault diagnosis. Moreover, Han et al. [29] introduced a convolutional transformer with novel squeeze-and-excitation network (Convformer-NSE) architecture that enhances end-to-end fault diagnosis of gearboxes under heavy noise by integrating both global and local feature extraction through a combination of Convformer and a newly designed Senet for improved channel adaptivity. Additionally, Beitao et al. [30] proposed a novel architecture called TimesNet, which is a deep learning model designed for multivariate time series forecasting; TimesNet effectively captures both local and global temporal dependencies in time series data through the implementation of dilated convolutions and self-attention mechanisms. Deep convolutional networks, such as ResNets and DenseNets, have also been applied to fault diagnosis tasks. ResNets have demonstrated notable performance in fault classification from vibration data, as evidenced by the findings of Wen et al. [31]. The implementation of residual connections within ResNets addresses issues like vanishing gradients, facilitating improved learning outcomes. Similarly, DenseNets have proven effective in extracting multiscale feature maps from raw vibration signals, which leads to accurate fault identification, outperforming traditional CNNs and LSTMs [32]. The dense connections within DenseNets enhance information flow and feature reuse, thereby contributing to improved diagnostic performance.

General-purpose deep learning models [21,33–36] have demonstrated significant potential in recognizing complex patterns within time series data, particularly in the domain of fault diagnosis tasks. Nevertheless, substantial efforts are necessary to enhance the capabilities and performance of these deep learning models in order to facilitate more robust diagnoses of intricate faults present in complex, real-world industrial gearboxes. Currently, one-dimensional convolutional kernels exhibit a limited scope because they are designed to capture patterns exclusively from adjacent time series data points, which restricts their effectiveness. Additionally, attention mechanisms can struggle to discern consistent dependencies among sparse data points that are concealed within complex patterns. This issue is particularly pertinent when considering the standard convolutional layers utilized in deep neural networks, which possess restricted receptive fields that limit their ability to model long-range dependencies in time series data effectively. Moreover, attention-based methods, such as transformers [21], may encounter challenges in accurately correlating distant vibration periods, especially when these signals are obscured by complex noise or intricate patterns. Sequential vibration signals obtained from rotating machinery are inherently rich in temporal information, as they contain clues such as continuity, frequency, cyclicity, deviations, fluctuations, and trends. These temporal characteristics can provide valuable insights into the machinery's inherent condition and operational status. Furthermore, the presence of complicated overlapping temporal fluctuations in real-world vibration time series—such as those generated by planetary gearboxes—poses significant modeling challenges. Therefore, there is a pressing need for advanced techniques capable of effectively capturing global context and latent patterns within the data. Such innovations hold promise for enhancing deep learning-based diagnosis and prognosis processes that leverage sequential vibration signals, ultimately leading to improved predictive maintenance and operational reliability in industrial applications. In summary, the limitations of the methods discussed above can be outlined as follows:

(1) Limited scope of one-dimensional convolutional kernels: One-dimensional convolutional kernels primarily focus on

extracting features from adjacent time series data points. This design restricts their ability to capture broader patterns across longer sequences, which is crucial for effective fault diagnosis in complex industrial applications.

(2) Challenges with attention mechanisms: While attention mechanisms offer potential for modeling dependencies in the data, they often struggle with sparse data points and complex patterns. This limitation is particularly evident in standard convolutional layers, which have restricted receptive fields and are ineffective at modeling long-range dependencies in time series data, leading to difficulties in accurately correlating distant signals.

(3) Restricted receptive fields of convolutional layers: Standard convolutional layers in deep neural networks have limited receptive fields, which hinder their capacity to model long-range dependencies in time series data. This is particularly problematic for analyzing the temporal dynamics of vibration signals from machinery.

(4) Complexity of temporal fluctuations: Real-world vibration signals contain intricate temporal fluctuations and overlapping patterns, which pose significant challenges for modeling. Capturing the rich temporal information present in these signals—such as continuity, frequency, and trends—requires advanced techniques that can effectively account for global context and latent patterns, which are currently lacking in existing deep learning models.

To address the limitations of existing approaches, a novel architecture named CyclicalNet is proposed to enhance the extraction of periodic and temporal features from vibration signals. CyclicalNet integrates cyclical blocks, encoder blocks, and dense connections, enabling the effective capture of inherent cyclicity and complex temporal dynamics within time series data. The architecture begins by utilizing high-amplitude components in the frequency domain—often neglected in traditional models—as input to the cyclical blocks, thereby focusing on segments with high diagnostic relevance. A reshape operation is then employed to transform the one-dimensional sequential data into two-dimensional representations, allowing for more structured and efficient feature extraction through a two-dimensional network architecture. This design facilitates the simultaneous modeling of temporal and periodic features, as well as long-range dependencies across distant time points, which are essential for uncovering complex, latent patterns in the data. Finally, dense connections are incorporated to promote multiscale feature learning, enhance information flow, and support feature reuse across the network. These components collectively improve the model's capacity for accurate and robust fault diagnosis by enabling the extraction of richer and more discriminative representations from the input signals.

The primary contributions of this paper can be summarized as follows:

(1) Enhanced periodicity feature extraction with the CyclicalNet architecture: The CyclicalNet architecture addresses the inherent limitations of one-dimensional convolutional kernels—particularly their limited receptive fields—by incorporating cyclical blocks, encoder blocks, and dense connections. This design facilitates the extraction of long-range dependencies and complex periodic patterns in time series data, which are critical for accurate fault diagnosis in industrial systems. Consequently, CyclicalNet significantly enhances the model's ability to comprehensively represent features, effectively mitigating the shortcomings associated with traditional one-dimensional convolutional approaches.

(2) Robust temporal pattern recognition: CyclicalNet leverages the inherent cyclical characteristics of vibration data and employs an advanced dependency modeling mechanism, effectively addressing the limitations of conventional attention mechanisms in processing sparse data points and intricate temporal patterns. This architectural design enhances the model's capacity to address such complexities, thereby improving its diagnostic performance.

(3) Adaptive receptive fields and multiscale feature extraction: CyclicalNet integrates dense connections to adaptively expand the network's receptive fields, thereby overcoming the limitations of the restricted receptive scope in conventional convolutional layers. This enhancement enables the model to capture long-range dependencies more accurately, thereby facilitating comprehensive analysis of temporal dynamics in vibration signals and enhancing the extraction of critical diagnostic features.

(4) Advancements in predictive maintenance techniques: CyclicalNet employs a novel data transformation approach that reshapes one-dimensional signals into two-dimensional representations, thereby enhancing its ability to capture intricate multiscale temporal characteristics. Through the integration of time–frequency analysis with deep learning, the model significantly improves the accuracy and robustness of fault diagnosis in predictive maintenance. This architecture enables the extraction of global contextual information and the identification of latent patterns, effectively overcoming the limitations of conventional models in interpreting overlapping and nonstationary temporal structures commonly present in real-world vibration signals.

The remainder of this paper is structured as follows: Section 2 reviews the related theory behind TimeNet. Section 3 then provides a detailed illustration of the proposed CyclicalNet architecture, highlighting key innovations and capabilities. Following this, Sec. 4 presents the experimental results from two fault diagnosis case studies, providing a detailed analysis and comparison of the outcomes. Finally, Sec. 5 summarizes the conclusions of this work and discusses potential avenues for future research to further advance vibration analysis using deep learning.

2 Related Theory

In this section, the primary theories underlying TimesNet [30] are presented. TimesNet's architecture comprises multiple interconnected components, referred to as Times-blocks, each of which performs an essential function in the overall operation of the system. Within each Times-block, the Fourier transform is applied, thereby extracting the amplitude–frequency properties of the input sequence. Subsequently, the reshape operation transforms the sequential one-dimensional representation into a two-dimensional format, which facilitates further manipulation and analysis of the data. After this transformation, the two-dimensional data is converted back into a one-dimensional output. Through this comprehensive process, TimesNet effectively decomposes the time series signal into its constituent frequencies, thereby providing insights that enable more sophisticated analysis and modeling.

The construction of TimesNet is illustrated in Fig. 1, which involves transforming the sequence into the deep feature after passing it through several TimesBlocks. The time series data $\boldsymbol{Y}_{1D}^{0}$ has a length of L and C channels, which can be represented as $\boldsymbol{Y}_{1D}^{0} \in \mathbb{R}^{L\times C}$. The Fourier transform, as described in Eq. (1), is utilized in $\boldsymbol{Y}_{1D}^{0}$ to obtain the top k frequencies with the highest amplitudes. These frequencies are then employed to compute k periods

$$\begin{aligned} k_i, f_i|_{i=1}^{k} &= \text{TopkAmp}\left(\text{FFT}\left(\boldsymbol{Y}_{1D}^{0}\right)\right) \\ p_i|_{i=1}^{k} &= \frac{L}{f_i} \end{aligned} \tag{1}$$

where f_i and k_i represent the fast Fourier transform result and the amplitude value calculation result, respectively. The highest k amplitudes of $\boldsymbol{Y}_{1D}^{0}$ are denoted as k_i, with their corresponding frequencies f_i and periods p_i. By reshaping the sequence using these periods, k feature maps are generated, as demonstrated in the following equation:

$$\boldsymbol{Y}^{i}_{2D} = \text{Reshape}_{p_i \times n_i}(\text{Padding}\,(\boldsymbol{Y}^{0}_{1D})), i \in \{1, \ldots, k\} \tag{2}$$

where n_i is defined as the number of sequences after cutting. The time series is given zeros along the temporal dimension using Padding$(\cdot)$ to correctly format it for use with Reshape$_{p_i \times n_i}(\cdot)$.

Once the 1D time series is converted into 2D tensors, each row represents the temporal connections of the signal, while each column captures its periodic connections. The convolutional kernel can simultaneously extract both temporal and periodic features of the signal, enabling the identification of underlying feature relationships. The above process can be formally expressed as follows:

$$\begin{aligned} \hat{\boldsymbol{Y}}^{i}_{2D} &= \text{Inception}(\boldsymbol{Y}^{i}_{2D}), i \in \{1, \ldots, k\} \\ \hat{\boldsymbol{Y}}^{i}_{1D} &= \text{Trunc}(\text{Reshape}_{1 \times L}(\hat{\boldsymbol{Y}}^{i}_{2D})), i \in \{1, \ldots, k\} \end{aligned} \tag{3}$$

where the inception block denoted as, which is parameter-efficient, is known for providing informative representations, as stated in the research by Szegedy et al. [37]. To revert the padded series with length $(p_i \times n_i)$ to its original length L, the Trunc function is utilized. This allows the learned 2D representation, denoted as $\hat{\boldsymbol{Y}}^{i}_{2D}$, to be mapped back to $\hat{\boldsymbol{Y}}^{i}_{1D}$ for aggregation.

Subsequently, the 1D-representations, which consist of $\{\hat{\boldsymbol{Y}}^{1}_{1D}, \ldots, \hat{\boldsymbol{Y}}^{k}_{1D}\}$ and are derived from the amplitudes, can be calculated in the following manner:

$$\begin{aligned} k_i^{'} &= \text{Softmax}\ (k_i) \\ \boldsymbol{Y}^{'}_{1D} &= \sum_{i=1}^{k} k_i^{'} \times \hat{\boldsymbol{Y}}^{i}_{1D} \end{aligned} \tag{4}$$

Finally, output $\boldsymbol{Y}^{1}_{1D}$ is obtained using the residual method

$$\boldsymbol{Y}^{1}_{1D} = \boldsymbol{Y}^{0}_{1D} + \boldsymbol{Y}^{'}_{1D} \tag{5}$$

3 CyclicalNet

This section presents a novel architecture known as CyclicalNet, which is specifically designed to explore potential patterns within temporal vibration signals. In addition to its primary goal of enhancing the analysis of vibration signals, CyclicalNet introduces a new architectural framework that diverges from traditional methods. It accomplishes this through the extraction of periodic features, as well as the integration of cyclical blocks, encoder blocks, and dense connections. By leveraging the cyclic nature inherent in vibration signals, CyclicalNet is intentionally engineered to capture and extract intricate temporal patterns that might otherwise remain undetected. The implementation of this innovative approach has the potential to revolutionize several fields, particularly in industrial machinery monitoring, and fault detection systems. By unveiling hidden patterns within temporal vibrations, CyclicalNet empowers stakeholders to refine their predictive maintenance strategies, thereby enhancing overall system reliability—especially within wind turbine systems. Furthermore, by elucidating critical insights into vibration behavior, CyclicalNet facilitates proactive measures aimed at preventing potential failures and optimizing operational efficiency.

The collaborative interaction among the various components of the CyclicalNet architecture fosters a robust framework for the efficient extraction of both temporal and periodic features from vibration signals. This enhanced capability not only enriches the analysis but also has the potential to significantly improve the performance of various machine learning models. Figure 2 provides a visual representation of the overall framework, illustrating the interconnections and interactions among the primary components. This diagram serves to offer a comprehensive overview of the proposed approach and its salient elements, thereby facilitating a deeper understanding of the architecture's operational mechanisms and the synergy among its different components.

In this context, the input vibration signal, denoted as $\boldsymbol{Y}^{0}_{1D}$, is recorded by an accelerometer. Specifically, $\boldsymbol{Y}^{0}_{1D}$ represents an individual sample within the dataset, characterized by a length L and a number of channels C. The value of each data point is represented by R, such that $\boldsymbol{Y}^{0}_{1D}|R^{L \times C}$.

The data processing procedure commences with the Fourier transformation of $\boldsymbol{Y}^{0}_{1D}$. This mathematical operation decomposes the signal into its individual frequency components. By analyzing the spectrum derived from the Fourier transform, significant insights into the frequency content and characteristics of the vibration signal are obtained. This analysis offers valuable insights into the underlying distribution of frequencies within the signal, including the identification of principal frequency components. The amplitude, frequency content, and associated characteristics are determined from Eq. (6). In order to capture the dynamic information present in vibration signal, the top k amplitudes are selected and denoted as $k_i|_{i=1}^{k}$. The top k amplitude components are extracted because they represent the most energetic frequency components in the spectrum, which are typically associated with the dominant fault-induced vibration mechanisms. Corresponding cycles are marked by are marked by $p_i|_{i=1}^{k} = L/f_i$, while the associated frequencies are designated by $f_i|_{i=1}^{k}$. The parameter k is a hyperparameter that must be appropriately chosen based on the characteristics of the data. These values can be systematically calculated and are utilized to estimate the size of the sequence reshaping, thereby optimizing subsequent analysis processes. The cyclical block leverages the frequency-domain transformation $F(\omega)$ (Eq. (1)) to capture periodic features, which are critical for distinguishing different fault types

$$\begin{aligned} k_i, f_i|_{i=1}^{k} &= \text{Topk Amp}\big(\text{FFT}\big(\boldsymbol{Y}^{0}_{1D}\big)\big) \\ p_i|_{i=1}^{k} &= \frac{L}{f_i} \end{aligned} \tag{6}$$

where TopkAMP denoted as the process of selecting the first k largest magnitude components from the input signal's spectrum obtained via Fourier transform

$$\boldsymbol{Y}^{i}_{2D}|_{i=1}^{k} = \text{Reshape}_{p_i \times n_i}(\boldsymbol{Y}^{0}_{1D}) \tag{7}$$

The input components $\boldsymbol{Y}^{0}_{1D}$, k_i, and p_i are processed within the cyclical block, where a transformation is applied to convert the one-dimensional data $\boldsymbol{Y}^{0}_{1D}$ into two-dimensional format, specifically $\boldsymbol{Y}^{i}_{2D}|_{i=1}^{k}$. This transformation process involves segmenting $\boldsymbol{Y}^{0}_{1D}$ into sequences of length p_i, resulting in a total of $n_i|_{i=1}^{k}$ sequences that

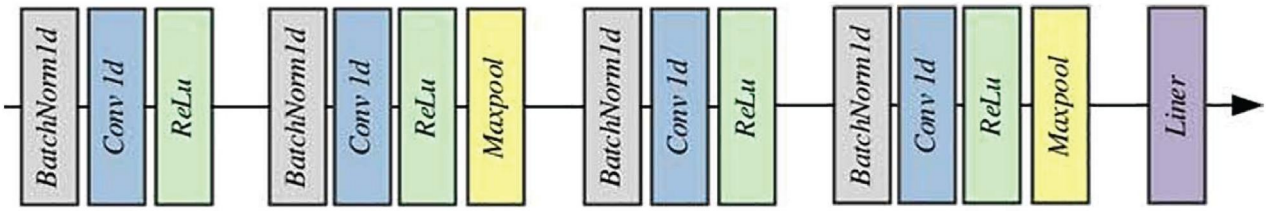

Fig. 1 Architecture of TimesNet

maintain uniform length. It is important to note that if the final sequence is shorter than p_i, it is padded with zeros to ensure consistent sequence length across all segments. Zero-padding is employed to prevent the introduction of extraneous artificial features that could distort the representation of the vibration signal's intrinsic periodic patterns, thereby ensuring that the learning process focuses on the signal's inherent characteristics. Following this segmentation, the sequences are concatenated vertically to form the two-dimensional data matrix $\boldsymbol{Y}_{2D}^{i}$, where $\boldsymbol{Y}_{2D}^{i} \in p_i \times n_i$. In this matrix, each row corresponds directly to the original time series data while each column effectively captures periodic relationships. This transformation not only allows for the representation of the time series data in a two-dimensional framework but also ensures that key frequency information is preserved during the preprocessing stage. As a result, the inherent cyclical features of the data become more pronounced, thereby facilitating the more efficient extraction of these periodic characteristics in subsequent analytical phases.

Moreover, this design enables the kernel to capture features from neighboring time points as well as corresponding points in adjacent time intervals, thereby encompassing a broader temporal context. Simultaneously, $\boldsymbol{Y}_{2D}^{i}$ retains the ability to restore the original one-dimensional vibration signal, leveraging the parameters p_i and n_i. The specific details are defined as follows:

$$\begin{aligned} \hat{\boldsymbol{Y}}_{2D}^{i}|_{i=1}^{k} &= \text{Inception}(\boldsymbol{Y}_{2D}^{i}) \\ \hat{\boldsymbol{Y}}_{1D}^{i}|_{i=1}^{k} &= \text{Reshape}_{1\times L}(\hat{\boldsymbol{Y}}_{2D}^{i}) \end{aligned} \tag{8}$$

To extract relevant time and cycle information at the same scale, the function of Inception$(\cdot)$ is employed, which operates without altering the dimensions of the feature maps. The function of Inception$(\cdot)$ executes a two-dimensional convolution operation on $\boldsymbol{Y}_{2D}^{i}$ with a stride of 1. This means that the sizes of $\hat{\boldsymbol{Y}}_{2D}^{i}$ and $\boldsymbol{Y}_{2D}^{i}$ remain unchanged. This design allows $\hat{\boldsymbol{Y}}_{2D}^{i}$ to utilize p_i and n_i for the purpose of restoring one-dimensional data.

In evaluating the significance of each frequency, the corresponding amplitude function is employed to determine their respective weights. This approach effectively addresses potential overflow issues that may arise from high amplitudes. Subsequently, the k feature sequences are fused using these calculated weights, resulting in the formation of $\boldsymbol{Y}_{1D}^{1}$, as defined by the following equation:

$$\begin{aligned} w_i|_{i=1}^{k} &= \frac{k_i}{\sum_{i=1}^{k} k_i} \\ \boldsymbol{Y}_{1D}^{1} &= \frac{\sum_{i=1}^{k} w_i \times \boldsymbol{Y}_{1D}^{i}}{k} \end{aligned} \tag{9}$$

The introduction of $\boldsymbol{Y}_{1D}^{0}$, k_i, and p_i as dense connections prior to each block plays a pivotal role in mitigating the issue of vanishing gradients. This strategic design facilitates effective feature transfer and promotes feature reuse within the model. Following this, both $\boldsymbol{Y}_{1D}^{0}$ and $\boldsymbol{Y}_{1D}^{1}$ are then passed into the second cyclical block. Given that all cyclical blocks share an identical structure, $\boldsymbol{Y}_{1D}^{0}$ is transformed into $\boldsymbol{Y}_{1D}^{1}$, which is subsequently processed into $\boldsymbol{Y}_{1D}^{2}$. Here, $\boldsymbol{Y}_{1D}^{2}$ represents $\boldsymbol{Y}_{1D}^{0}$ after passing through two layers of cyclical blocks. These cyclical blocks are specifically designed to capture both temporal and cyclical dependencies; thus, $\boldsymbol{Y}_{1D}^{1}$ reflects one layer of extracted temporal and cyclical features, while $\boldsymbol{Y}_{1D}^{2}$ represents two layers of such features. Subsequently, $\boldsymbol{Y}_{1D}^{0}$, $\boldsymbol{Y}_{1D}^{1}$, and $\boldsymbol{Y}_{1D}^{2}$ are fed into the third cyclical block, resulting in outputs $\boldsymbol{Y}_{1D}^{1}$, $\boldsymbol{Y}_{1D}^{2}$, and $\boldsymbol{Y}_{1D}^{3}$, which represent the temporal and periodic features extracted across multiple layers from $\boldsymbol{Y}_{1D}^{0}$.

Following this, as depicted, the input vectors $(\boldsymbol{Y}_{1D}^{0}, \boldsymbol{Y}_{1D}^{1}, \boldsymbol{Y}_{1D}^{2}, \boldsymbol{Y}_{1D}^{3})$ are processed within the encoder block. The encoder block comprises several key modules as illustrated in Fig. 3, including 1D batch normalization layers, 1D convolution layers, rectified linear unit layers, max pooling layers, and a linear layer. These modules function collaboratively to first normalize the distribution of the input data via the batch normalization layer, thereby enhancing the stability of training in subsequent layers. The 1D convolution layer then captures local temporal or spatial features, while the rectified linear unit layer introduces nonlinearity, enabling the model to represent complex relationships. The max pooling layer further down-samples the data, reducing dimensionality and computational complexity while preserving critical feature information. Finally, the linear layer projects the extracted high-dimensional features onto a low-dimensional feature space, forming the feature manifold. Through this systematic process of feature extraction and dimensionality reduction, the model effectively identifies the most discriminative and expressive features from the original signal

$$\begin{aligned} d_i &= \text{Softmax}\left(\text{Encoder}\left(\boldsymbol{Y}_{1D}^{0}, \boldsymbol{Y}_{1D}^{1}, \boldsymbol{Y}_{1D}^{2}, \boldsymbol{Y}_{1D}^{3}\right)\right) \\ L_s &= -\frac{1}{N}\sum_{i=1}^{N} l_i^{\text{label}} \cdot \log(d_i) \end{aligned} \tag{10}$$

The encoder block generates a probability distribution vector, representing the predicted health state. These probabilities are normalized using the SoftMax function, ensuring that the output probabilities are comparable across different classes. Finally, the model's performance is evaluated using cross-entropy loss L_s, which is calculated based on the predicted d_i and the true label l_i^{label}, where N denotes the number of dataset samples.

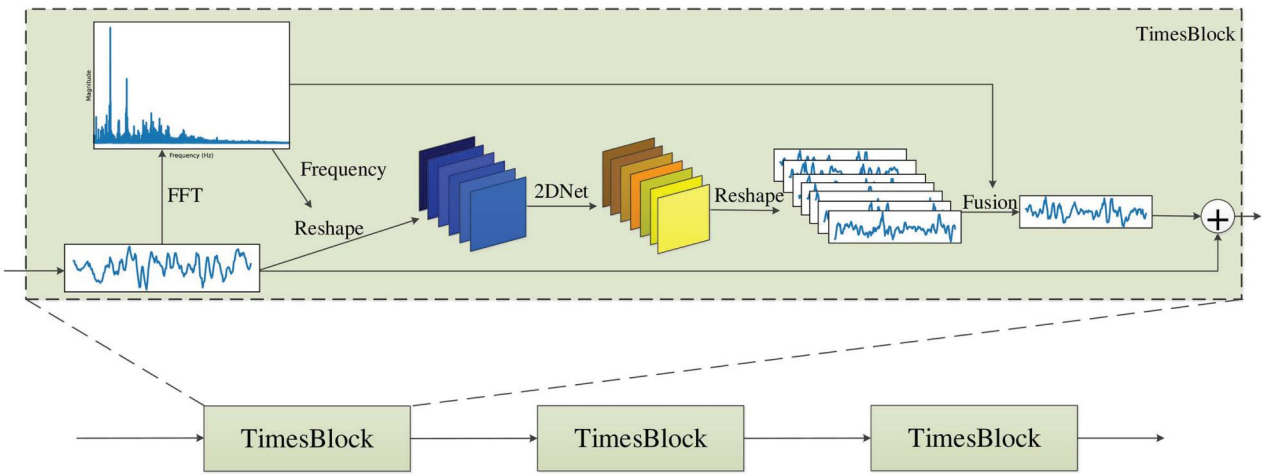

Fig. 2 Architecture of proposed CyclicalNet

4 Experimental Validation and Comparative Analysis

4.1 Case Study I

4.1.1 Specifications for Data Description and Test-Rig. The gearbox dataset originates from a gear drive system that comprehensively encompasses multiple operating conditions and fault types. As illustrated in Fig. 4, the system comprises several key components, including a tachometer, a drive motor, a torque sensor, a two-stage parallel gearbox, a load gearbox, and a load motor. The placement of the accelerometer is particularly significant, as it is affixed to a separate disk, a configuration that is shown in greater detail in the zoomed-in view of Fig. 4. To achieve high-precision sampling of the system's dynamic behavior, data are recorded at a frequency of 12.8 kHz. Moreover, the dataset spans a broad range of operating conditions, with rotational speeds systematically varying between 1600 and 2400 rpm. In addition to normal operation, the dataset encompasses five common gear failure types, which are depicted in Fig. 5: miss (missing tooth), chipped (cracked teeth), surface (wear on gear surface), and root (crack at tooth root). Furthermore, the meshing structure of the gears is illustrated in Fig. 6(*a*), while detailed internal structural information about the parallel gearbox system is provided in Fig. 6(*b*), where the faulty gear is highlighted with dotted lines for ease of identification.

To facilitate a comprehensive analysis of gear fault diagnosis, vibration data are collected along the *X*-axis of the accelerometer while the gear operates at a constant speed of 1600 rpm. Each state, including the healthy condition, comprises 768,000 data points collected over a duration of 60 s. Consequently, this large-scale dataset serves as a robust foundation for developing and validating fault diagnosis algorithms, thus enabling researchers to investigate various operating states and fault types under controlled experimental conditions.

4.1.2 Comparative Networks and Results Analysis. Case study I aims to analyze the performance variations of a gearbox under typical operating conditions, as well as under different fault scenarios. To ensure the reliability of the experimental results and maintain the integrity of the testing process, the dataset is systematically divided into two independent subsets: 80% of the data is allocated to the training set for model learning and parameter optimization, while the remaining 20% is reserved as a test set to assess the model's generalization performance. This data partitioning strategy effectively minimizes data bias and ensures that the model demonstrates strong predictive capability on previously unseen data, thereby enhancing the reliability of fault diagnosis outcomes. Furthermore, an experimental comparison is conducted between the proposed architecture and current state-of-the-art models for fault diagnosis in gearboxes. The benchmark models selected for comparison include TimesNet [30], Vanilla Transformer [21], Autoformer [23], hierarchical multiscale dense (HMSD) [26], FEDformer [24], and AD-ResNet [25]. By evaluating performance using metrics such as accuracy, precision, and the confusion matrix, the objective is to quantitatively assess the capabilities of the proposed approach in relation to existing leading methods for this task. This metrics-driven comparison will facilitate an objective evaluation of the improvements that the proposed architecture offers over previous state-of-the-art models. Rigorously comparing quantitative metrics is believed to provide valuable insights into the advancements enabled by the proposed approach compared to current methods.

Overall, this experimental benchmarking process allows for an impartial, metrics-based assessment of the proposed architecture against established models for gearbox fault diagnosis. The successful training of the proposed architecture heavily relies on the precise determination of appropriate hyperparameters. In the case of CyclicalNet, this architecture encompasses five topmost amplitudes and three cyclical blocks, along with an encoding depth of 128 in the encoding segment. To optimize the training process, several critical decisions have been made concerning the hyperparameters. Firstly, the initial learning rate is set to 0.0001, which is common practice to initiate training with a small learning rate and subsequently adjust it during the process. The learning rate is decreased with each epoch, with the learning rate for each epoch being half of that of the previous epoch. This gradual decrease enables the model to converge effectively, while also potentially avoiding overshooting the optimal values. Additionally, the

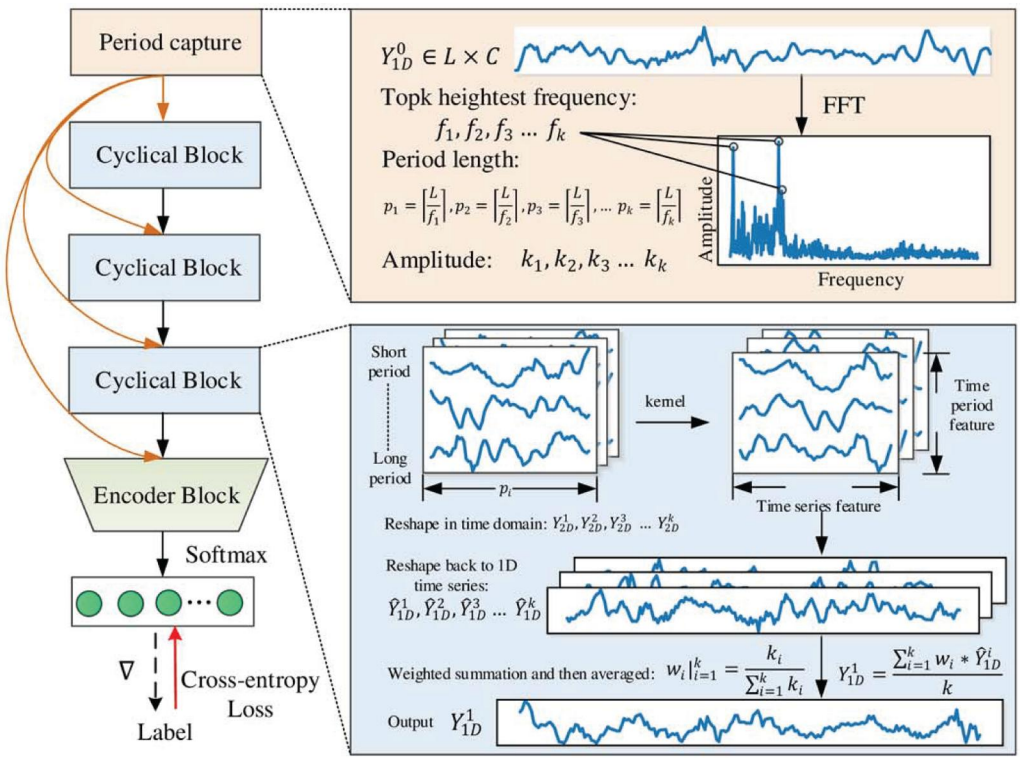

Fig. 3 Architecture of encoder block

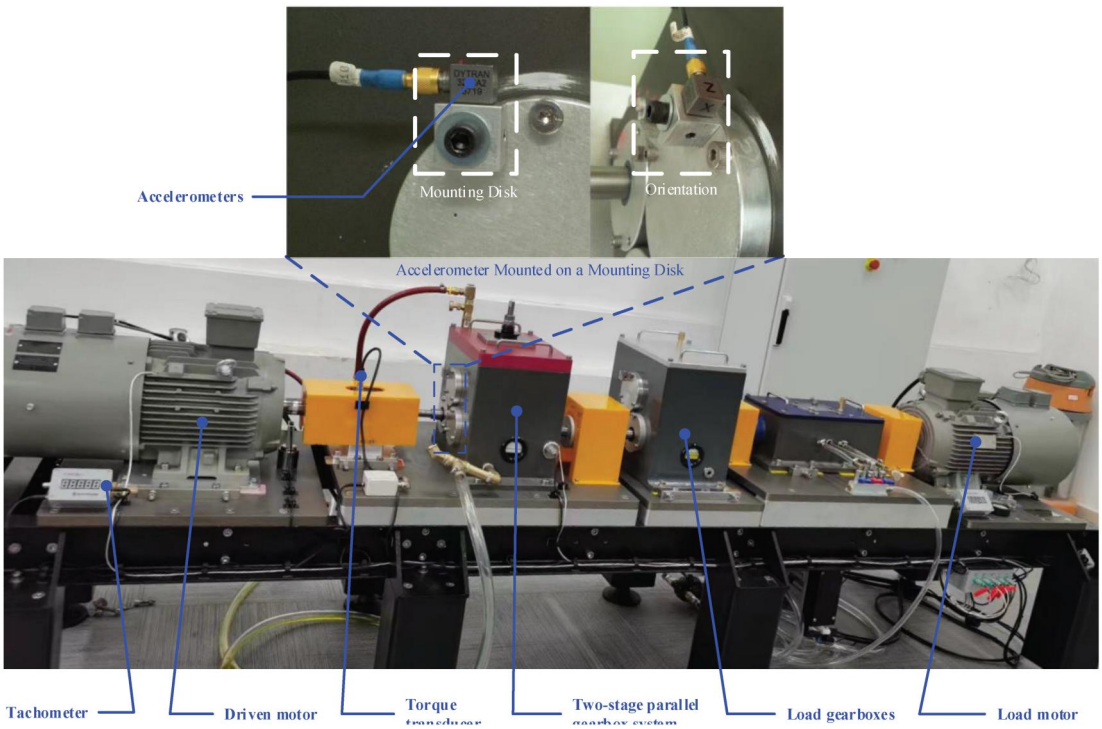

Fig. 4 Experimental test-rig of gear transmission system

optimizer employed in this architecture is Adam, a widely used optimization algorithm noted for its ability to adaptively adjust learning rates for different parameters. Adam incorporates principles from both RMSProp and momentum-based optimization, thereby making it suitable for a variety of deep learning tasks. The chosen loss function for the training process is CrossEntropyLoss, which is commonly employed in classification problems, where the goal is to minimize the discrepancy between predicted and actual class labels. This function measures the dissimilarity between probability distributions, rendering it an appropriate choice for training classification models. Moreover, the activation function utilized in CyclicalNet is Gaussian error linear unit, which is a nonlinear activation function that has gained prominence in recent years due to its ability to enhance the performance of deep neural networks. To mitigate overfitting and improve generalization, a dropout rate of 0.1 is applied. Dropout is a regularization technique that randomly sets a portion of input units to zero during training, thereby preventing the model from over-relying on specific features and improving its capacity to generalize to unseen data. Finally, the training process is conducted over 50 epochs; an epoch denotes a complete pass through the entire training dataset. Training for multiple epochs allows the model to iteratively update its parameters and learn from the data more comprehensively.

The experimental verification results, as illustrated in Fig. 7(*a*), demonstrate that the evaluated transformer-based time series models exhibit only moderate performance in the fault diagnosis task for the gearbox dataset. Among these models, the Vanilla Transformer emerges as the best performer, achieving a maximum test accuracy of 57%. However, this accuracy, which hovers around 60%, highlights a significant need for improvement in the context of sequential fault diagnosis applications. In contrast, other models, including Autoformer, TimesNet, and FEDformer, perform even less effectively, with test accuracies ranging from 40% to 45%. Notably, Autoformer achieves the lowest accuracy, approximately 40%, while TimesNet exhibits slightly better performance at 45%. These relatively low results underscore the limitations of the current Autoformer, TimesNet, and FEDformer architectures in effectively learning failure patterns and making accurate predictions within transmission datasets. The overall mediocre test accuracies, which fall between 40% and 60%, indicate that the transformer models evaluated are not well-suited for this specific fault diagnosis task, thereby indicating a clear potential for improvement through the development of more advanced transformer architectures for such applications.

In comparison to the aforementioned models, AD-ResNet demonstrates a significant improvement, maintaining an accuracy consistently around 82% and thereby surpassing the performance of Autoformer, TimesNet, FEDformer, and the Vanilla Transformer. The architecture of HMSD, which employs a feedforward mechanism to directly connect each layer, facilitates effective feature flow and gradient propagation, thus rendering it well-suited for diagnostic tasks. Consequently, hierarchical multiscale dense network (HMSDN) achieved a maximum test accuracy of approximately 90%, a performance that can be primarily attributed

Fig. 5 (*a*) Miss (missing tooth), (*b*) chipped (cracked teeth), (*c*) surface (wear on gear surface), and (*d*) root (crack at tooth root)

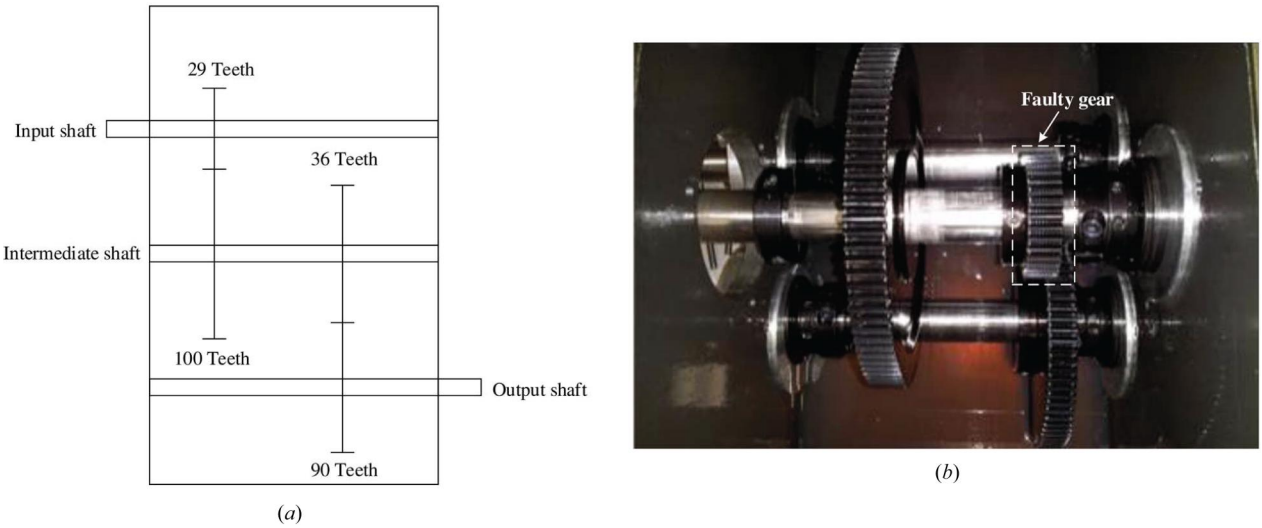

Fig. 6 (*a*) Gear meshing and (*b*) internal configuration of parallel gearbox system

to its hierarchical multiscale dense connection architecture, which effectively facilitates comprehensive feature fusion across multiple scales. Nevertheless, it is important to note that HMSDN does not explicitly emphasize the extraction of periodic features—such as gear meshing harmonics and fault-induced periodic impulses—which are essential for accurate fault diagnosis in planetary gearboxes. The absence of a targeted mechanism for capturing these periodic characteristics may limit its capability to fully exploit the underlying cyclic patterns inherent in gearbox vibration signals. However, as depicted in Fig. 7(*a*), CyclicalNet outperforms the typical performance of HMSD by attaining a test accuracy of around 98%, reflecting its strong capability on the test dataset. The black line in the figure indicates the model's standard deviation, revealing that while HMSD's accuracy fluctuates significantly during experiments, CyclicalNet maintains a relatively stable accuracy. Even when accounting for fluctuations, HMSD's accuracy rarely exceeds CyclicalNet's minimum accuracy, which underscores CyclicalNet's superior and stable fault identification capability compared to the other evaluated models.

The receiver operating characteristic (ROC) curves [38], presented in Fig. 7(*b*), provide a visual representation of the tradeoff between true positive rates and false positive rates across different diagnostic thresholds for each model. CyclicalNet achieves a higher area under the curve in its ROC curve compared to the other models, indicating superior sensitivity across thresholds while maintaining a lower false alarm rate. This finding demonstrates CyclicalNet's capability to balance detection accuracy and misdiagnosis across a range of operating points. The improved performance of CyclicalNet can be attributed to its multidimensional feature learning architecture, which links pertinent features across layers and extracts both temporal and cyclical fault characteristics. By combining thorough diagnostic patterns throughout the temporal, cycle, and depth dimensions, CyclicalNet enhances its capacity for discrimination. Overall, the ROC analysis reveals that CyclicalNet is able to acquire more comprehensive gearbox fault characteristics, which enables it to surpass other models in terms of fault identification performance across a variety of decision thresholds.

4.1.3 Comparative Analysis of Visualizations. A more rigorous quantitative analysis of the diagnosis is conducted using the confusion matrix. The results, depicted in Fig. 8, provide a visual representation of the model's performance across various fault types. Upon careful examination of the data, it becomes evident that the proposed CyclicalNet successfully identifies all fault types with very high accuracy, as shown in Fig. 8(*g*). Specifically, the model achieves over 99% accuracy for chipped, miss, and root failures, and over 95% accuracy for health and surface failures, thereby demonstrating its superior diagnostic capabilities.

In contrast, TimesNet (Fig. 8(*a*)) and transformer-based models—Vanilla Transformer (Fig. 8(*b*)), Autoformer (Fig. 8(*c*)), and FEDformer (Fig. 8(*e*))—exhibit significant limitations in accurately identifying gear faults. Although AD-ResNet (Fig. 8(*f*)) shows a basic ability to distinguish gear failures, its diagnostic accuracy is limited to chipped failures and health states, highlighting its shortcomings in a comprehensive fault diagnosis scenario.

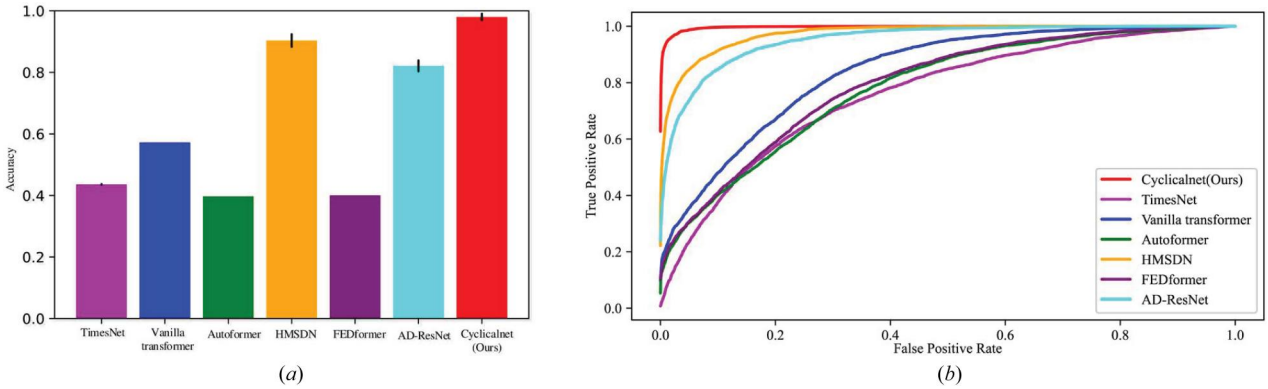

Fig. 7 (*a*) Evaluating accuracy performance across various models and (*b*) analyzing ROC curve across diverse models

HMSDN (Fig. 8(*d*)) presents better fault diagnosis capabilities; however, its overall accuracy remains below 91%, with only chipped faults exceeding 95% diagnostic accuracy, indicating room for improvement.

To conduct a comprehensive comparative analysis of the features captured by different models, the input feature data of all models are uniformly scaled to the surface of a unit sphere, as illustrated in Fig. 9. This advanced visualization technique is employed because it provides a more intuitive and clear understanding of high-dimensional feature spaces, allowing for a better assessment of how various models interpret and process input data. In this context, Figs. 9(*a*)–9(*c*), 9(*e*), and 9(*f*) present the results for five advanced methods. These models exhibit low feature clustering on the spheres, which indicates inherent limitations in their capacity for effective feature extraction and domain adaptation. Such low clustering indicates that these models may struggle to differentiate or generalize between fault categories in complex diagnostic scenarios.

In contrast, Fig. 9(*d*), which corresponds to the HMSDN method, reveals somewhat improved feature clustering. This improvement suggests that HMSDN is more effective at capturing essential diagnostic features, thereby potentially enhancing model accuracy. Nevertheless, the boundaries of HMSDN's feature clusters remain indistinct, leading to some samples being misclassified, which indicates room for further optimization. However, as illustrated in Fig. 9(*g*), CyclicalNet demonstrates a markedly superior clustering effect compared to the aforementioned models. Specifically, the feature distributions exhibit high intraclass compactness and well-defined interclass boundaries, indicating a clear separation between different fault categories. This distinct clustering pattern underscores the enhanced capability of CyclicalNet in capturing discriminative cyclic features, thereby leading to significant performance improvements in transmission fault diagnosis. The model's ability to effectively extract and integrate both periodic and temporal features simultaneously enables it to perform complex diagnostic tasks with greater precision and reliability, highlighting its potential as a leading solution in the field. Overall, this analysis illustrates that while existing models exhibit various levels of clustering capabilities, CyclicalNet distinctly excels, thereby offering significant improvements in the accuracy and reliability of fault diagnosis.

4.2 Case Study II

4.2.1 Details for Data Description and Test Setup. In order to further validate and rigorously test the proposed method, an experimental apparatus known as the Drivetrain Prognostics Simulator, as illustrated in Fig. 10, is employed for case study II. This experimental setup includes two 10-hp, three-phase induction motors, each with two pairs of poles; one motor functions as the driving motor, while the other serves as the load motor. The Drivetrain Prognostics Simulator drivetrain is composed of several gearboxes: a two-stage planetary test gearbox, a two-stage parallel shaft test gearbox, and two load gearboxes, all equipped with either rolling or sleeve bearings. The two-stage planetary gearbox features a gear ratio of 27:1 and consists of the first stage with four planetary gears and the second stage with three planetary gears. Under investigation is the planetary gearbox, which incorporates four distinct types of damaged sun gears, specifically sun gears with surface wear, tooth crack, chipped tooth, and missing tooth, as detailed in Fig. 11. Vibration measurements are systematically conducted in both the horizontal and vertical directions using two piezoelectric ceramic-based accelerometers mounted on the gearbox. This comprehensive setup ensures a robust analysis of the method's performance under various conditions.

The purpose of this research is to investigate how planetary gearbox performance varies under typical operating circumstances and various malfunctioning scenarios. For analysis, time-domain data from the gearbox is collected across five distinct damage stages. This dataset includes five types of vibration signals associated with sun gear malfunctions. For each damage condition, a two-channel dataset is recorded, encompassing both the tacho signal and vibrations detected from the planetary gearbox shell. Data collection is executed via a computer with a National Instruments acquisition card, operated at a sampling frequency of 30,720 Hz. Postdata acquisition, a total of 761 samples for each fault condition is compiled, leading to a cumulative total of 3805 samples. These samples are then randomly divided into training samples (80%) and validation samples (20%).

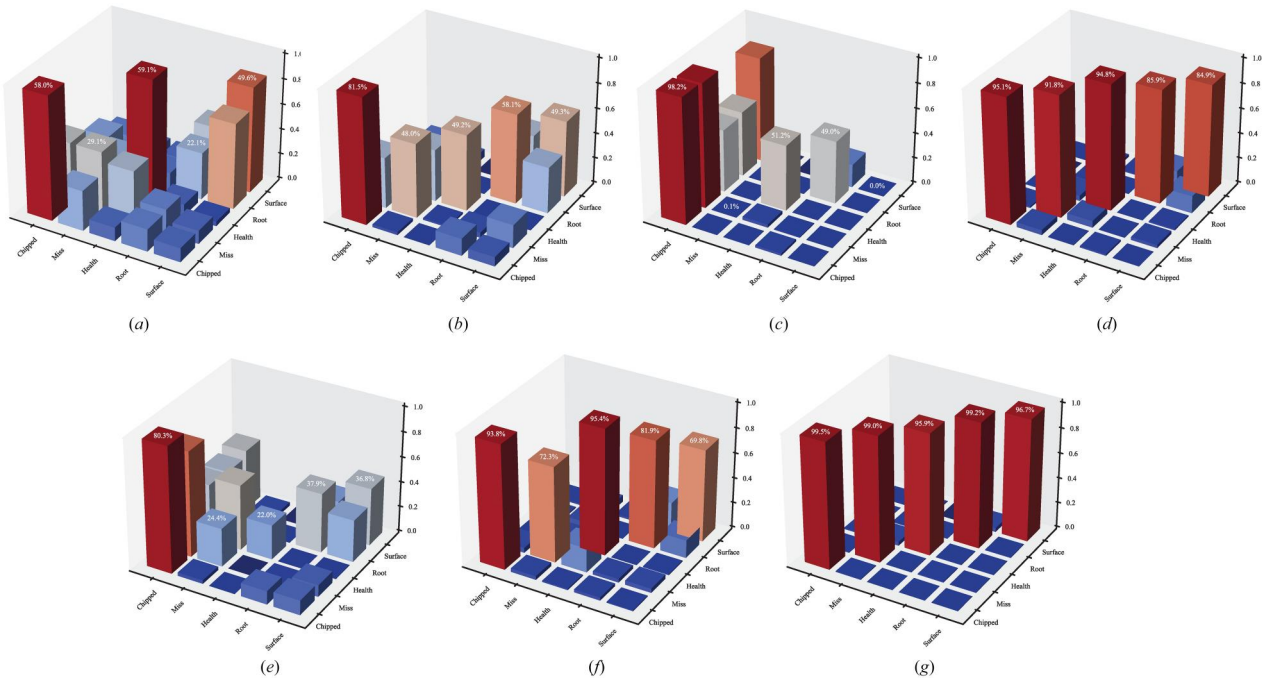

Fig. 8 Classification performance via confusion matrix: (*a*) TimesNet, (*b*) Vanilla Transformer, (*c*) Autoformer, (*d*) HMSDN, (*e*) FEDformer, (*f*) AD-ResNet, and (*g*) CyclicalNet (Ours)

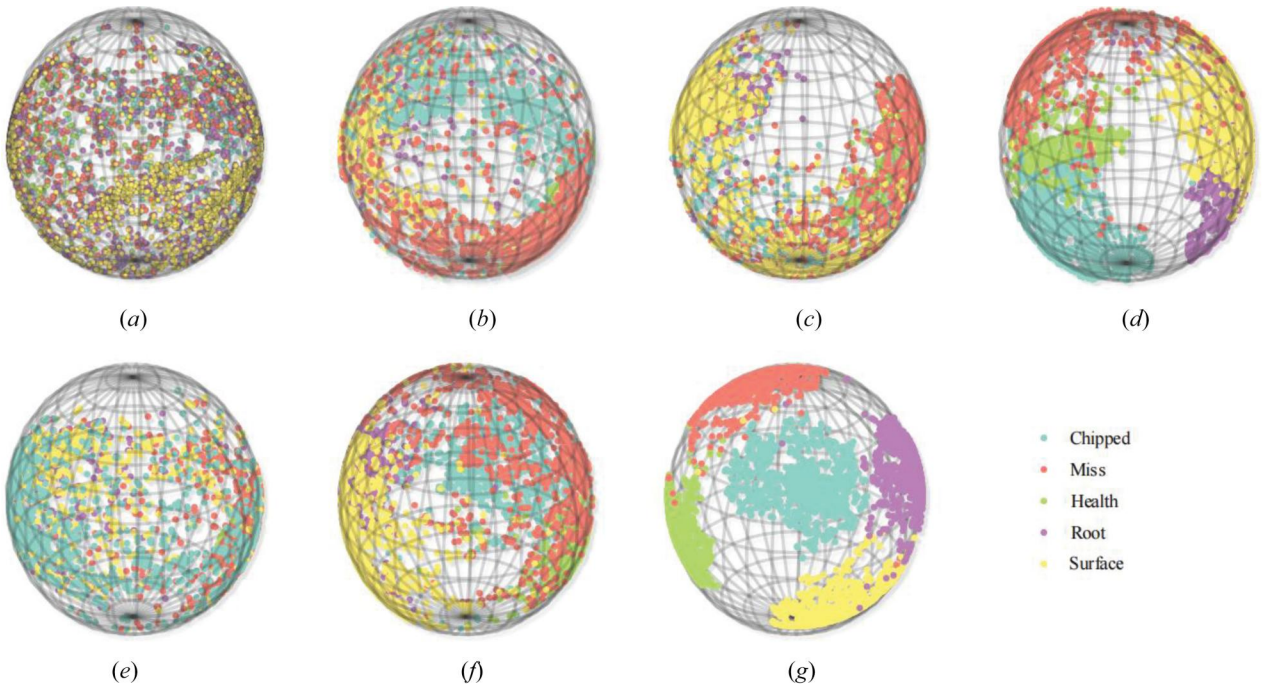

Fig. 9 Visualization results of unit sphere surface features: (*a*) TimesNet, (*b*) Vanilla Transformer, (*c*) Autoformer, (*d*) HMSDN, (*e*) FEDformer, (*f*) AD-ResNet, and (*g*) CyclicalNet (Ours)

4.2.2 Analysis of Comparative Networks and Result Evaluations. In this analysis, the proposed architecture is experimentally compared to current leading models for planetary gearbox fault diagnosis. The benchmark models, aligning with those documented in case study I, comprise TimesNet, Vanilla Transformer, Autoformer, HMSD, FEDformer, and AD-ResNet. As depicted in Fig. 12(*a*), the evaluation indicates that transformer-based time series models exhibit only moderate performance on the fault diagnosis task, with test accuracies for Vanilla Transformer, Autoformer, TimesNet, and FEDformer ranging between 35% and 70%. These comparatively low results imply that these transformer-based architectures and TimesNet may not be optimally designed for this specific diagnostic application.

Conversely, AD-ResNet achieves a notable enhancement, maintaining a consistent accuracy around 89%, while HMSD reaches a peak test accuracy of approximately 95%. However, as illustrated in Fig. 12(*a*), CyclicalNet surpasses HMSD with a test accuracy of roughly 97%, demonstrating its effective performance on the test dataset. The standard deviation of each model's test accuracy, indicated by a black line, reveals significant fluctuations for AD-ResNet and HMSD during the experiment. In stark contrast, CyclicalNet demonstrates relatively stable accuracy, emphasizing its superior and reliable fault diagnosis capabilities compared to other assessed models.

The ROC curves in Fig. 12(*b*) illustrate the balance between true positive rate and false positive rate across various diagnostic thresholds for each model. CyclicalNet's ROC curve displays a

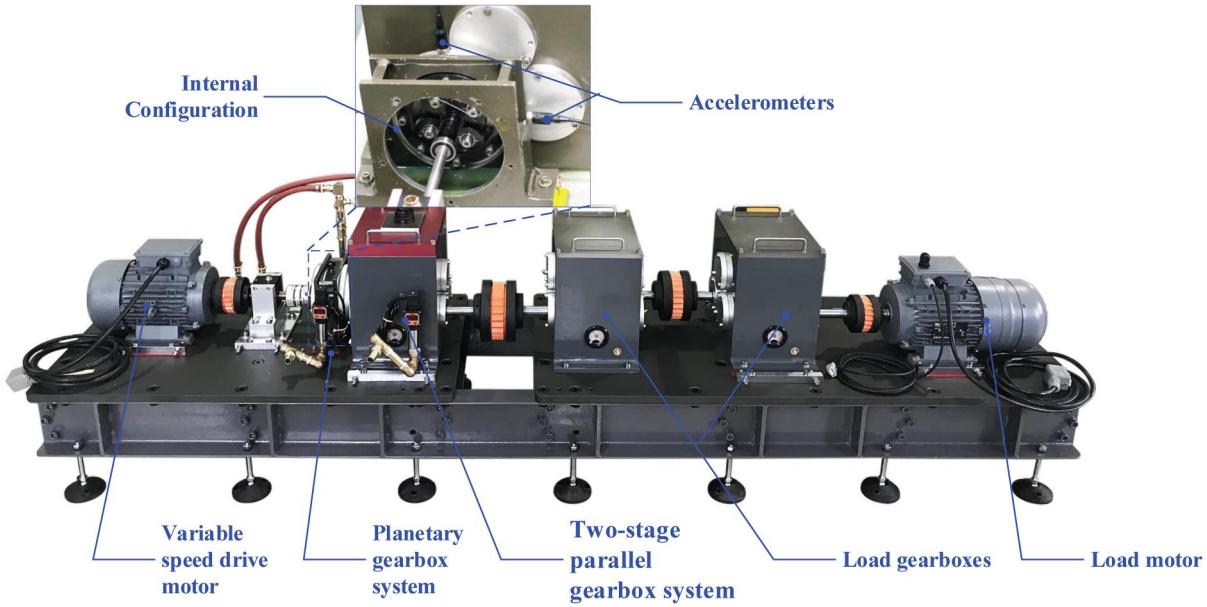

Fig. 10 Industrial drivetrains involve the configuration of gear transmission systems

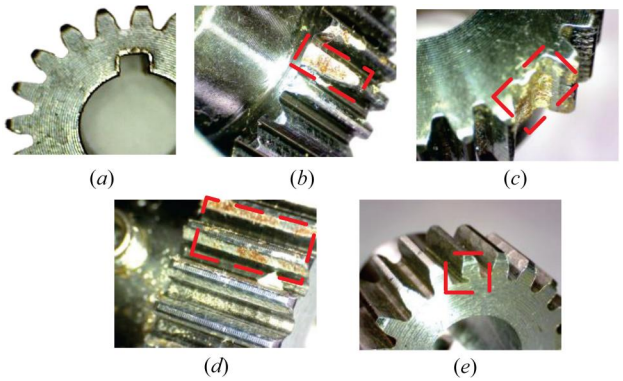

Fig. 11 The planetary gearbox exhibits various defect modes, including (*a*) sun gear with an intact tooth, (*b*) sun gear with a surface wear tooth, (*c*) sun gear with a crack tooth, (*d*) sun gear with a chipped tooth, and (*e*) sun gear with a missing tooth

higher area under the curve than other models, indicating its enhanced capacity to balance detection accuracy while minimizing misdiagnoses across diverse operating conditions. This capability enables CyclicalNet to achieve a more thorough gearbox fault profile, thereby surpassing other models in fault identification performance across different decision thresholds.

4.2.3 Comparative Analysis of Visual Representations. Confusion matrices serve as a quantitative tool for evaluating the accuracy of different models in classifying captured fault characteristics. When examining Figs. 13(*a*)–13(*c*) and 13(*e*), it becomes evident that transformer-based models like Autoformer, TimesNet, FEDformer, and the Vanilla Transformer struggle to correctly classify the majority of faults. Conversely, as depicted in Fig. 13(*f*), AD-ResNet distinguishes sun gear failures more effectively. Moreover, HMSDN, shown in Fig. 13(*d*), performs adeptly by accurately identifying most failures. Notably, the proposed CyclicalNet architecture, illustrated in Fig. 13(*a*), achieves approximately 97% classification accuracy across nearly all failure types. This represents a significant improvement over the other models, with only a few misidentified cracks and chips, underscoring

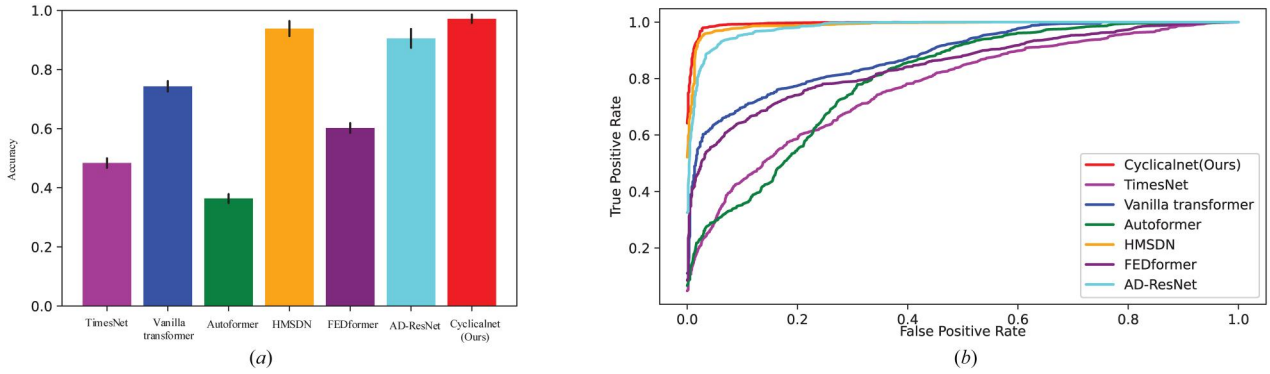

Fig. 12 (*a*) Evaluating accuracy performance across various models and (*b*) analyzing ROC curve across diverse models

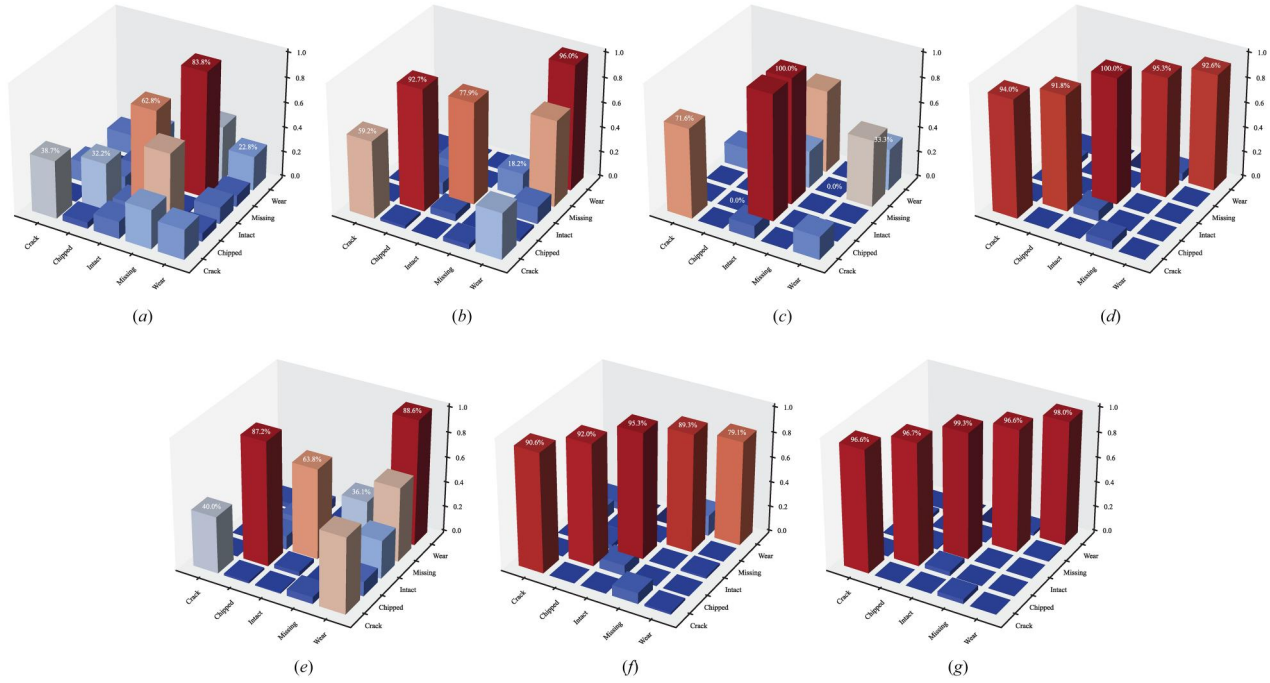

Fig. 13 Classification performance via confusion matrix: (*a*) TimesNet, (*b*) Vanilla Transformer, (*c*) Autoformer, (*d*) HMSDN, (*e*) FEDformer, (*f*) AD-ResNet, and (*g*) CyclicalNet (Ours)

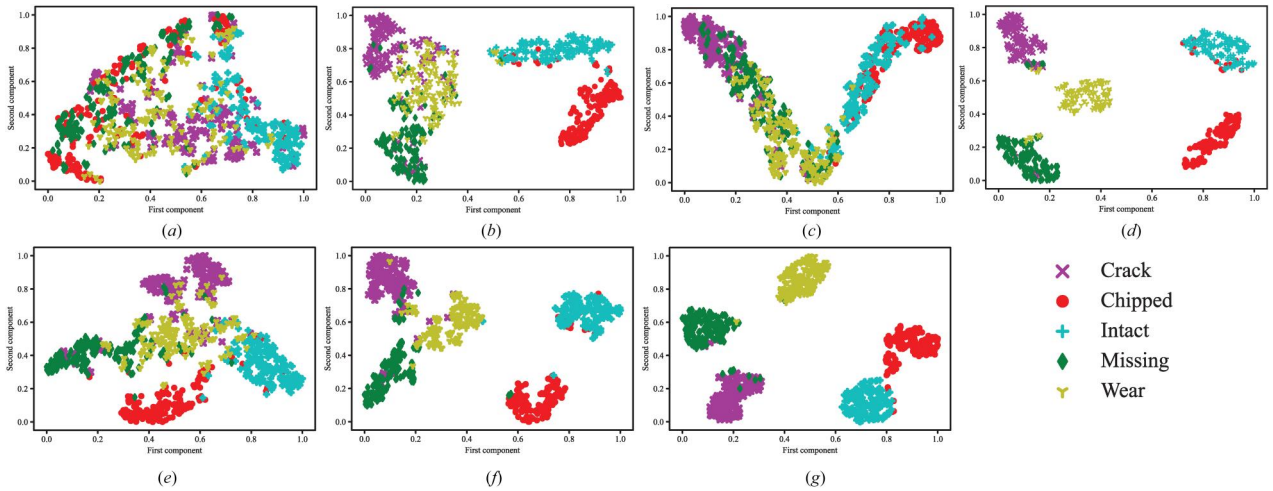

Fig. 14 The feature visualization through t-SNE: (*a*) TimesNet, (*b*) Vanilla Transformer, (*c*) Autoformer, (*d*) HMSDN, (*e*) FEDformer, (*f*) AD-ResNet, and (*g*) CyclicalNet (Ours)

CyclicalNet's superior fault diagnosis capabilities. By harnessing the intrinsic cyclic nature of vibration signals, CyclicalNet captures intricate temporal patterns that other approaches may overlook. Therefore, the results underscore CyclicalNet's effectiveness in exploiting cyclic redundancies to learn nuanced diagnostic features, leading to state-of-the-art performance in gearbox fault classification.

The latent diagnostic features captured by these models are visualized and interpreted using t-distributed stochastic neighbor embedding (t-SNE) for dimensionality reduction [39,40]. t-SNE, an unsupervised deep learning technique, reduces high-dimensional feature representations into fewer dimensions while preserving local proximity relationships, thus enabling visualization of inherent data grouping and separation. The t-SNE projections for each model are shown in Fig. 14, offering insights into how effectively each model's representations categorize and segregate various fault types when examining these reduced feature spaces in detail.

For the transformer-based models, such as Autoformer, TimesNet, FEDformer, and Vanilla Transformer, the feature manifolds exhibit significant overlap and clustering, hindering correct fault separation, as evidenced in Figs. 14(*a*)–14(*c*) and 14(*e*). In contrast, the feature manifold captured by AD-ResNet, shown in Fig. 14(*f*), displays improved separation, with most fault features independently clustered. The performance of HMSDN surpasses that of AD-ResNet, as its feature manifolds distinctly separate all fault types. This ability stems from HMSDN's layered feedforward connections, which facilitate the learning of hidden characteristics that are advantageous for diagnosis.

However, CyclicalNet leverages the naturally cyclic nature of vibration signals to discern subtle temporal patterns that elude other models. As demonstrated in Fig. 14(*g*), the feature representations extracted by CyclicalNet exhibit a high degree of intraclass compactness, forming a concentrated, approximately circular cluster in the feature space. This clustering effect is notably more cohesive compared to the relatively dispersed distributions observed in HMSDN and other benchmark models, reflecting the superior feature discrimination capability of CyclicalNet. The optimized feature clustering signifies CyclicalNet's superiority in extracting the most discriminative fault characteristics among all evaluated models.

5 Conclusion

This paper presented the innovative CyclicalNet architecture, a novel approach to diagnosing faults in the gearboxes. Current deep learning models have shown potential but are inherently constrained by their inability to identify consistent dependencies within complex patterns and limited scope of one-dimensional convolution kernels. The proposed CyclicalNet architecture effectively transcends these limitations by the extraction of periodic features and the integration of cyclical blocks, encoder blocks, and dense connections. It uniquely leverages the top amplitude–frequency components as input to cyclical blocks, surpassing conventional strategies. The introduction of a reshape operation to transform the 1D sequential representation into 2D maps broadens the scope of feature extraction, enabling improved information flow and feature re-usability. The use of dense connections between blocks further contributes to the identification of underlying patterns within the vibration sequences, thereby enhancing multiscale feature extraction. In conclusion, the CyclicalNet architecture introduces a robust, data-driven solution to model complex vibration signals, paving the way for a more reliable and comprehensive fault diagnosis for planetary gearboxes.

Future research will prioritize broadening the applicability of the innovative CyclicalNet architecture, extending it beyond planetary gearboxes to encompass a diverse range of complex industrial systems. This endeavor will involve fine-tuning the architecture and optimizing its components to cater to different types of machinery, aiming to uncover its universal potential in fault diagnosis. Thorough testing and evaluation under varied operational conditions and disturbances will be essential to assess the architecture's ability to generalize across different scenarios. It will be important to incorporate noisy and nonstationary vibration signals during testing to thoroughly examine the architecture's performance in challenging real-world conditions. In summary, future work will primarily focus on generalizing the CyclicalNet architecture, validating its performance across diverse machinery types, and elucidating its decision-making process. This research trajectory aims to unlock the architecture's full potential as a reliable and effective fault diagnosis tool for complex machinery.

Funding Data

- National Natural Science Foundation of China (Grant Nos. 52105111, 52375078, U23B20104, and 52305085; Funder ID: 10.13039/501100001809).
- Guangdong Basic and Applied Basic Research Foundation (Grant No. 2025A1515012256).
- Shantou University (STU) Scientific Research Initiation Grant (NTF21029; Funder ID: 10.13039/100009047).
- China Postdoctoral Science Foundation (Grant No. 2023M740021; Funder ID: 10.13039/501100002858).
- Natural Science Foundation of Anhui Province (Grant No. 2108085QE229; Funder ID: 10.13039/501100003995).

Data Availability Statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

References

[1] Zhang, Y., Zhao, X., Liang, H., and Chen, P., 2024, "Multiscale Dilated Convolution and Swin-Transformer for Small Sample Gearbox Fault Diagnosis," Appl. Intell., **54**(17–18), pp. 7716–7732.

[2] Maestro-Prieto, J. A., Ramírez-Sanz, J. M., Bustillo, A., and Rodriguez-Díez, J. J., 2024, "Semi-Supervised Diagnosis of Wind-Turbine Gearbox Misalignment and Imbalance Faults," Appl. Intell., **54**(6), pp. 4525–4544.

[3] Chen, P., Zhang, R., Fan, S., Guo, J., and Yang, X., 2025, "Step-Wise Contrastive Representation Learning for Diagnosing Unknown Defective Categories in Planetary Gearboxes," Knowl.-Based Syst., **309**, p. 112863.

[4] Randall, R., and Antoni, J., 2011, "Rolling Element Bearing Diagnostics—A Tutorial," Mech. Syst. Signal Process., **25**(2), pp. 485–520.

[5] Chen, P., Wang, K., Zuo, M. J., and Wei, D., 2019, "An Ameliorated Synchroextracting Transform Based on Upgraded Local Instantaneous Frequency Approximation," Measurement, **148**, p. 106953.

[6] Dalpiaz, G., Rivola, A., and Rubini, R., 2000, "Effectiveness and Sensitivity of Vibration Processing Techniques for Local Fault Detection in Gears," Mech. Syst. Signal Process., **14**(3), pp. 387–412.

[7] Jia, F., Lei, Y., Lin, J., Zhou, X., and Lu, N., 2016, "Deep Neural Networks: A Promising Tool for Fault Characteristic Mining and Intelligent Diagnosis of Rotating Machinery with Massive Data," Mech. Syst. Signal Process., **72**, pp. 303–315.

[8] Rafiee, J., Arvani, F., Harifi, A., and Sadeghi, M., 2007, "Intelligent Condition Monitoring of a Gearbox Using Artificial Neural Network," Mech. Syst. Signal Process., **21**(4), pp. 1746–1754.

[9] Chen, P., Gao, J., Zhang, R., Jin, Y., Yu, R., He, C., and Qi, J., 2025, "Metric-Guided Graph Contrastive Learning: An Unsupervised Approach for Few-Shot Gearbox Fault Diagnosis," Meas. Sci. Technol., **36**(7), p. 076110.

[10] Lei, Y., Li, N., Guo, L., Li, N., Yan, T., and Lin, J., 2018, "Machinery Health Prognostics: A Systematic Review From Data Acquisition to RUL Prediction," Mech. Syst. Signal Process., **104**, pp. 799–834.

[11] Chen, P., Wu, Y., Xu, C., Jin, Y., and Zhou, C., 2024, "Markov Modeling of Signal Condition Transitions for Bearing Diagnostics Under External Interference Conditions," IEEE Trans. Instrum. Meas., **73**(3518308), pp. 1–8.

[12] Li, B., Tang, B., Deng, L., and Wei, J., 2022, "Joint Attention Feature Transfer Network for Gearbox Fault Diagnosis With Imbalanced Data," Mech. Syst. Signal Process., **176**, p. 109146.

[13] Koutsoupakis, J., Seventekidis, P., and Giagopoulos, D., 2023, "Machine Learning Based Condition Monitoring for Gear Transmission Systems Using Data Generated by Optimal Multibody Dynamics Models," Mech. Syst. Signal Process., **190**, p. 110130.

[14] Perez-Sanjines, F., Peeters, C., Verstraeten, T., Antoni, J., Nowé, A., and Helsen, J., 2023, "Fleet-Based Early Fault Detection of Wind Turbine Gearboxes Using Physics-Informed Deep Learning Based on Cyclic Spectral Coherence," Mech. Syst. Signal Process., **185**, p. 109760.

[15] Chen, P., Ma, J., He, C., Jin, Y., and Fan, S., 2025, "Semi-Supervised Consistency Models for Automated Defect Detection in Carbon Fiber Composite Structures With Limited Data," Meas. Sci. Technol., **36**(4), p. 046109.

[16] Chen, P., Li, Y., Wang, K., and Zuo, M. J., 2021, "An Automatic Speed Adaption Neural Network Model for Planetary Gearbox Fault Diagnosis," Measurement, **171**, p. 108784.

[17] Chen, P., Li, Y., Wang, K., and Zuo, M. J., 2020, "A Novel Knowledge Transfer Network With Fluctuating Operational Condition Adaptation for Bearing Fault Pattern Recognition," Measurement, **158**, p. 107739.

[18] Chen, P., Li, Y., Wang, K., Zuo, M. J., Heyns, P. S., and Baggeröhr, S., 2021, "A Threshold Self-Setting Condition Monitoring Scheme for Wind Turbine Generator Bearings Based on Deep Convolutional Generative Adversarial Networks," Measurement, **167**, p. 108234.

[19] Wang, Z., Wang, J., and Wang, Y., 2018, "An Intelligent Diagnosis Scheme Based on Generative Adversarial Learning Deep Neural Networks and its Application to Planetary Gearbox Fault Pattern Recognition," Neurocomputing, **310**, pp. 213–222.

[20] Chen, H., Shi, L., Zhou, S., Yue, Y., and An, N., 2022, "A Multi-Source Consistency Domain Adaptation Neural Network MCDANN for Fault Diagnosis," Appl. Sci., **12**(19), p. 10113.

[21] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I., 2017, "Attention Is All You Need," Adv. Neural Inf. Process. Syst., **30**, pp. 5998–6008.

[22] Chen, P., Zhang, R., He, C., Jin, Y., Fan, S., Qi, J., Zhou, C., and Zhang, C., 2025, "Progressive Contrastive Representation Learning for Defect Diagnosis in Aluminum Disk Substrates With a Bio-Inspired Vision Sensor," Expert Syst. Appl., **289**, p. 128305.

[23] Wu, H., Xu, J., Wang, J., and Long, M., 2021, "Autoformer: Decomposition Transformers With Auto-Correlation for Long-Term Series Forecasting," Adv. Neural Inf. Process. Syst., **34**, pp. 22419–22430.

[24] Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., and Jin, R., 2022, "Fedformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting," International Conference on Machine Learning, Baltimore, MA, July 17–23, pp. 27268–27286.

[25] Xiong, H., Wang, Z., Wu, G., Pan, Y., Yang, Z., and Long, Z., 2022, "Steering Actuator Fault Diagnosis for Autonomous Vehicle With an Adaptive Denoising Residual Network," IEEE Trans. Instrum. Meas., **71**, pp. 1–13.

[26] Xu, Y., Yan, X., Sun, B., and Liu, Z., 2022, "Hierarchical Multiscale Dense Networks for Intelligent Fault Diagnosis of Electromechanical Systems," IEEE Trans. Instrum. Meas., **71**, pp. 1–12.

[27] Jiang, K., Liu, H., Ruan, H., Zhao, J. and Lin, Y., 2023, "ALAE: Self-Attention Reconstruction Network for Multivariate Time Series Anomaly Identification," Soft Comput, **27**(15), pp.10509–10519.

[28] Wu, R., Liu, C., Han, T., Yao, J., and Jiang, D., 2022, "A Planetary Gearbox Fault Diagnosis Method Based on Time-Series Imaging Feature Fusion and a Transformer Model," Meas. Sci. Technol., **34**(2), p. 024006.

[29] Han, S., Shao, H., Cheng, J., Yang, X., and Cai, B., 2023, "Convformer-NSE: A Novel End-to-End Gearbox Fault Diagnosis Framework Under Heavy Noise Using Joint Global and Local Information," IEEE/ASME Trans. Mechatron., **28**(1), pp. 340–349.

[30] Beitao, L., Guangquan, Z., Nana, Y., Chenglin, W., Shuai, F., Lu, P., and Xiaojun, C., 2022, "TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, June 18–24, pp. 12033–12042.

[31] Wen, L., Li, X., Gao, L., and Zhang, Y., 2020, "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method," IEEE Trans. Ind. Electron., **68**(7), pp. 6552–6562.

[32] Yu, Y., Si, X., Hu, C. and Zhang, J., 2019, "A review of recurrent neural networks: LSTM cells and network architectures," Neural computation, **31**(7), pp.1235-1270.

[33] He, K., Zhang, X., Ren, S., and Sun, J., 2016, "Deep Residual Learning for Image Recognition," Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, June 27–30, pp. 770–778.

[34] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q., 2017, "Densely Connected Convolutional Networks," Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, July 21–26, pp. 4700–4708.

[35] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., 2014, "Generative Adversarial Nets," Adv. Neural Inf. Process. Syst., **27**, pp. 2672–2680.

[36] Chen, P., Ma, Z., Xu, C., Jin, Y., and Zhou, C., 2024, "Self-Supervised Transfer Learning for Remote Wear Evaluation in Machine Tool Elements With Imaging Transmission Attenuation," IEEE Internet Things J., **11**(13), pp. 23045–23054.

[37] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., 2015, "Going Deeper With Convolutions," Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, June 7–12, pp. 1–9.

[38] Fawcett, T., 2006, "An Introduction to ROC Analysis," Pattern Recognit. Lett., **27**(8), pp. 861–874.

[39] Chen, P., Xu, C., Ma, Z., and Jin, Y., 2023, "A Mixed Samples-Driven Methodology Based on Denoising Diffusion Probabilistic Model for Identifying Damage in Carbon Fiber Composite Structures," IEEE Trans. Instrum. Meas., **72**, pp. 1–11.

[40] Van der Maaten, L., and Hinton, G., 2008, "Visualizing Data Using t-SNE," J. Mach. Learn. Res., **9**(11), pp. 2579–2605.