Physically Interpretable Wavelet-Guided Networks With Dynamic Frequency Decomposition for Machine Intelligence Fault Prediction

Wang, Huan; Li, Yan Fu; Men, Tianli; Li, Lishuai

DOI
10.1109/TSMC.2024.3389068

Publication date
2024

Document Version
Final published version

Published in
IEEE Transactions on Systems, Man, and Cybernetics: Systems

Citation (APA)
Wang, H., Li, Y. F., Men, T., & Li, L. (2024). Physically Interpretable Wavelet-Guided Networks With Dynamic Frequency Decomposition for Machine Intelligence Fault Prediction. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *54*(8), 4863 - 4875. https://doi.org/10.1109/TSMC.2024.3389068

Physically Interpretable Wavelet-Guided Networks With Dynamic Frequency Decomposition for Machine Intelligence Fault Prediction

Huan Wang, Yan-Fu Li, *Senior Member, IEEE*, Tianli Men, and Lishuai Li, *Senior Member, IEEE*

***Abstract*—Machine intelligence fault prediction (MIFP) is crucial for ensuring complex systems' safe and reliable operation. While deep learning has become the mainstream tool for MIFP due to its excellent learning abilities, its interpretability is limited, and it struggles to learn frequencies, making it challenging to understand the physical knowledge of signals at the frequency level. Therefore, this article proposes a physically interpretable wavelet-guided network (WaveGNet) with deep frequency separation for MIFP, inspired by the sound theoretical basis and physical meaning of discrete wavelet transform (DWT). WaveGNet expands the feature learning space of CNN into the frequency domain, allowing for a better understanding of the physical insights behind the frequency level. Specifically, WaveGNet involves a derivable and learnable frequency learning layer (FL-Layer) consisting of a wavelet-driven frequency decomposition module and a convolution-driven feature learning module. Multiple DWT-driven FL-Layers are used in WaveGNet to achieve deep frequency decomposition and multiresolution frequency feature learning in a coarse-to-fine manner. The effectiveness of WaveGNet was evaluated in real high-speed train wheel wear monitoring and high-speed aviation bearing fault diagnosis cases. Experimental results showed that WaveGNet outperforms cutting-edge deep learning algorithms and has excellent fault diagnosis and prediction abilities. Furthermore, an in-depth analysis of the learning mechanism of wavelet-driven CNN from the frequency domain perspective was conducted.**

***Index Terms*—Bearings, convolutional neural network, fault prediction, high-speed train (HST).**

I. Introduction

MODERN mechanical systems, such as aircraft engines and high-speed trains (HSTs), operate in harsh environments and require high levels of reliability and safety. However, these systems can experience functional failures due to the fatigue degradation of their core components. These failures can have severe consequences, such as train derailment accidents and sudden engine shutdowns [1]. Machine intelligence fault prediction (MIFP) can predict and diagnose these faults by monitoring the system's health status [2], [3]. Therefore, monitoring the health status of key components is crucial for ensuring the safe and reliable operation of mechanical systems.

Deep learning [4] can extract valuable information from large datasets and learn to accurately predict outcomes in high-dimensional spaces. It has revolutionized many fields by automating feature engineering and decision-making processes [5], [6], [7]. In the context of ever-increasing data, deep learning has made it possible to process vast amounts of information without manual intervention [8]. In MIFP, deep learning has shown great promise and achieved significant breakthroughs, particularly in the field of CNN-based fault prediction [9], [10], [11]. However, most existing research has been conducted in controlled laboratory environments, without considering the challenges posed by the complex and harsh environments in which mechanical systems operate. Therefore, to promote deep-learning-based MIFP technology from laboratory to practical application, the following serious challenges need to be solved.

1) Monitoring signals in mechanical systems contain complex frequency components due to their dynamic environment and multiple information sources, which can cause valuable frequency components to mix with irrelevant information.
2) Mechanical systems are often exposed to harsh working environments that can cause severe noise pollution in monitoring signals. Different types of noise, such as background noise, collision noise, resonance, and impact noise, can significantly contaminate monitoring signals.
3) Health-related features of mechanical systems are usually weak, and faults often start as mild and progress to severe. Early fault identification and prediction are crucial, but in the early stages of faults, fault-related features are weak and difficult to detect.

To meet the above challenges, CNN-based MIFP algorithms also have the following shortcomings that need to be addressed.

1) *Limited Frequency Learning Ability:* The CNN only consists of limited filters that are optimized with random initialization, leading to poor frequency learning

Manuscript received 26 October 2023; revised 16 January 2024; accepted 8 April 2024. Date of publication 2 May 2024; date of current version 19 July 2024. This work was supported in part by the Beijing Municipal Natural Science Foundation-Rail Transit Joint Research Program under Grant L231020, and in part by the National Natural Science Foundation of China through a Key Project under Grant 71731008. This article was recommended by Associate Editor Z. Gao. *(Corresponding author: Yan-Fu Li.)*

Huan Wang, Yan-Fu Li, and Tianli Men are with the Department of Industrial Engineering, Tsinghua University, Beijing 100084, China (e-mail: liyanfu@tsinghua.edu.cn).

Lishuai Li is with the Faculty of Aerospace Engineering, Delft University of Technology, 2600 AA Delft, The Netherlands, and also with the School of Data Science, City University of Hong Kong, Hong Kong, SAR.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSMC.2024.3389068.

Digital Object Identifier 10.1109/TSMC.2024.3389068

ability. Understanding signal physical knowledge at the frequency level becomes difficult.

2) *Limited Noise Robustness:* CNN is highly sensitive to noise, and its performance drops significantly in a noisy environment, as reported in [12] and [13].
3) *Limited Interpretability:* In the industrial domain, understanding what features the algorithm has learned and its learning mechanisms are crucial when applying deep models. However, CNNs still remain as black-box models, making it exceptionally challenging to conduct interpretable analysis on them.

The above limitations challenge the practical application prospects of CNN. Traditionally, researchers have employed various signal analysis techniques, such as fast Fourier transform (FFT) [14], discrete wavelet transform (DWT) [15], etc., to remove irrelevant components from the frequency domain and extract valuable fault features from noisy signals. Therefore, scholars have combined time–frequency analysis methods with deep learning to enhance diagnostic performance [16], [17]. For instance, using wavelet transform, Guo et al. [18] captured multiscale information from sensor signals and then utilized CNN to extract fault-related features and identify sensor faults in aircraft engines. These methods typically involve time–frequency analysis of signals, followed by inputting the obtained features into deep models to enhance learning capabilities. However, in these approaches, the integration of time–frequency methods and deep models is relatively simple, and their advantages have not been fully exploited. Additionally, the preprocessing in the time–frequency domain might discard critical features, thereby limiting the information acquired by deep models.

Based on the above discussion and the recognized benefits of DWT, this article explores a DWT-guided CNN architecture design scheme for MIFP. Driven by DWT, the CNN model can effectively capture the physical knowledge of monitoring signals from the frequency domain. Moreover, DWT is known for its signal analysis and denoising capabilities, which can enhance the model's noise robustness. Finally, it also enhances the interpretability of the diagnostic model, allowing for analysis of the feature learning of CNN at the frequency level.

Specifically, this article introduces a physically interpretable wavelet-guided network (WaveGNet) with deep frequency separation for MIFP. The study explores a paradigm for designing the deep model driven by DWT for frequency feature learning. It proposes the wavelet-driven CNN, integrating a multiresolution decomposition mechanism for learning frequency information from coarse to fine levels. However, DWT and CNN represent significantly different technical systems. The key challenge in their integration lies in developing a differentiable DWT layer that efficiently transfers gradient information. To this end, this article introduces a derivable and learnable DWT-driven frequency learning layer (FL-Layer). This FL-Layer comprises a DWT-driven frequency decomposition module and a convolution-driven feature learning module, working efficiently to conduct frequency feature learning. It decomposes frequency and learns features layer by layer in a coarse-to-fine manner, capturing valuable frequency features that are often challenging for CNNs to learn in the time domain.

The contributions of this article are summarized as follows.

1) This study explores and develops deep learning frameworks dominated by frequency domain analysis, breaking away from simple combinations of DWT and CNN. This framework surpasses the limitations of traditional deep learning models dominated by time-domain analysis. It enables the exploration of fault-related features in the frequency space and allows for physical property analysis.
2) Extending the deep model's feature space from the time to the frequency domain is a significant breakthrough. This provides a novel design perspective for deep model algorithms, enabling deep diagnostic models to integrate effectively with real-world fault physics and mechanisms.
3) This article proposes a differentiable and trainable DWT-driven FL-Layer that suppresses frequency aliasing problems through multilevel frequency decomposition and thus builds a physically interpretable WaveGNet for MIFP.
4) The proposed approach was tested and studied in two real-world cases. Particularly, it used on-track signal data collected over half a year for train wheel wear prediction, demonstrating accurate prediction of wheel wear conditions. This provides valuable information for the operational maintenance of HSTs, ensuring their safety.

This article is organized as follows. Section II introduces the related work. Section III introduces the proposed method in detail. Section IV verifies the effectiveness and superiority of the proposed method in two cases. Section V provides an in-depth analysis of the interpretability of the proposed method. Section VI summarizes this article.

II. Related Works

A. Deep-Learning-Based MIFP Methods

Deep learning has made significant strides in MIFP thanks to its outstanding ability to learn features and make automated decisions [19], [20], [21]. For example, Peng et al. [22] employed Deep SVDD in a semi-supervised manner to predict faults in wind turbines with high accuracy. Wang et al. [23] proposed a multitask attention CNN, which realizes the HST-bearing health status monitoring by mining the complementary information between different tasks. Wei and Wang [24] improved vanilla convolution to propose a defect recognition framework based on dynamic convolution, which achieved accurate recognition of composite wafer defects. However, existing algorithms lack consideration for frequency domain feature learning, causing deep models to have performance bottlenecks when processing data containing complex frequencies. Therefore, this study aims to rethink feature learning within the CNN framework by using the wavelet domain. It explores a DWT-guided CNN architecture design paradigm to extract valuable frequency features that are challenging to learn in the temporal space.

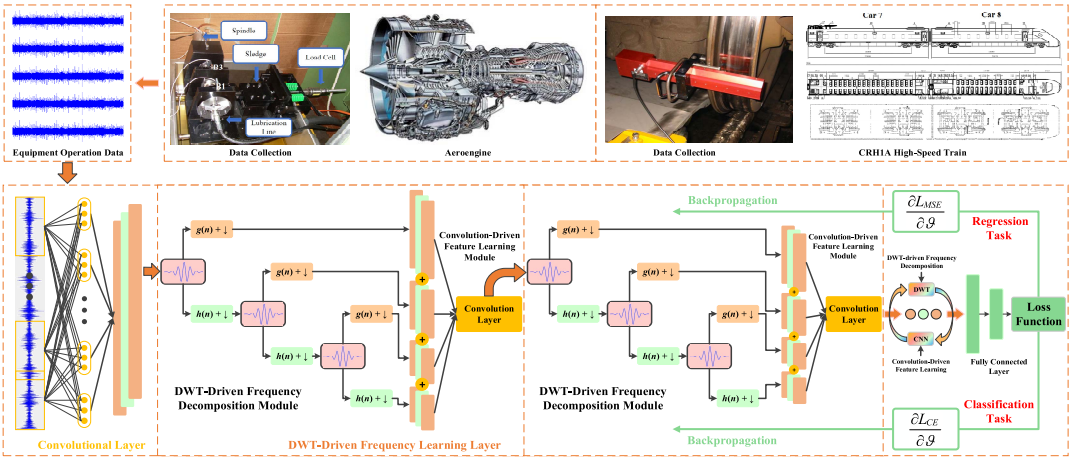

Fig. 1. Proposed wavelet-driven convolutional neural network with deep frequency separation for MIFP.

B. MIFP Methods Using Wavelet Analysis

Wavelet analysis is an excellent time–frequency analysis method that can extract valuable frequency information. It has been frequently used as a feature extraction technique in previous works to assist deep learning models. For example, Liang et al. [16] transformed vibration signals into two-dimensional (2-D) time–frequency features and used a multilabel CNN model for composite fault diagnosis of gearboxes. Tran et al. [25] utilized continuous wavelet transform (CWT) for signal transformation into time–frequency features, followed by a convolutional attention network model for motor fault identification. Recently, researchers have attempted to integrate wavelet transforms into CNN architectures to improve interpretability and harness their benefits. Li et al. [26] proposed a wavelet-kernel-net for MIFP, where they replaced the first convolution layer of the network with a continuous wavelet convolution layer. This method combines wavelet and convolution modules to enable effective frequency-domain feature extraction for MIFP. However, these methods do not deeply and properly integrate frequency learning capabilities throughout the deep model, so the frequency analysis capabilities of the model are limited.

C. Interpretability Studies of MIFP Methods

Interpretability analysis of deep learning for MIFP has been a major research focus. Wang et al. [27] utilized temporal and channel attention to investigate the feature learning mechanism of fault diagnosis models. Zhou et al. [28] used physical knowledge of mechanical systems to guide the design of CNN models for interpretability analysis of diagnostic models. Grezmak et al. [29] proposed an interpretable deep CNN model for gearbox fault diagnosis by applying layer-wise relevance propagation. However, elaborating and analyzing the interpretability of deep models from a frequency perspective is still lacking. Moreover, frequency analysis is very important for equipment fault diagnosis, because fault characteristics will be reflected in specific frequency bands. Xie et al. [30] proposed a deep learning framework that integrates fault frequency priors, including fault frequency priors learning branch and self-learning branch and achieved excellent offshore wind turbine fault diagnosis performance. Zhang et al. [31] proposed a hierarchical cognitive architecture based on domain knowledge and data fusion to achieve accurate fault diagnosis of interconnected systems.

To this end, this study proposes a novel solution for deep learning interpretability analysis. The approach involves utilizing a signal analysis method that has physical meaning and theoretical basis to improve the CNN model's interpretability. By doing so, it helps to explain the preference and learning mechanism of its frequency feature learning.

III. Methodology

A. Wavelet-Driven Convolutional Neural Network

This article proposes a wavelet-guided CNN (WaveGNet) with deep frequency separation for MIFP. The network architecture of WaveGNet is illustrated in Fig. 1. DWT is innovatively incorporated into the CNN framework, thus expanding CNN's feature learning capacity from the temporal domain to the frequency domain. This enables CNNs to acquire valuable physical knowledge from the frequency level. WaveGNet consists of multiple derivative and learnable DWT-driven FL-Layers. FL-Layer performs multiresolution frequency decomposition on the input signal and then uses the feature learning module to perform feature learning on the decomposed results. In WaveGNet, the input signal is subjected to frequency decomposition and feature learning layer by layer, which constructs a deep frequency decomposition mechanism to realize multiresolution frequency feature learning in a coarse-to-fine manner. In the following, the proposed DWT-driven FL-Layer and deep frequency separation are introduced, and the physical meaning and interpretability of the wavelet-driven CNN are analyzed.

B. DWT-Driven Frequency Learning

1) Frequency Learning Layer: The monitoring data of mechanical systems are usually signal data, so 1-D CNN is generally used for feature extraction. In 1DCNN, the information flow is represented by a 2-D feature map, which

consists of multiple feature signals that carry different feature information. Given a feature map $\mathcal{Z} \in \mathbb{R}^C \times L$ of a 1DCNN, it consists of C feature signals, each of length L. The mathematical description of the convolution operation is as follows:

$$Y[i] = (X * \mathcal{K})[i] = \sum_{j=0}^{\zeta-1} X[i+j] \times \mathcal{K}[j] \tag{1}$$

where X represents the input sequence, Y represents the output sequence, $\mathcal{K}$ represents the convolution kernel, and ζ represents the convolution kernel size. For the convolutional layer, it generally also includes the bias term ς and the activation function σ which is expressed as

$$Y[i] = \sigma((X * \mathcal{K} + \varsigma)[i]). \tag{2}$$

Wavelet analysis is an excellent time–frequency analysis method, which can analyze the valuable information of the signal from the frequency level. Let $\int(x)$ be a square-integrable function, that is, $\int(t) \in L^2(R)$; $\psi(t)$ is the mother wavelet, if $\psi(t)$ satisfies the following properties:

$$C_\psi = \int_0^\infty \frac{|\widehat{\psi}(\omega)|^2}{\omega} d\omega < \infty. \tag{3}$$

Then, the wavelet transform can be expressed as

$$Wf(a,b) = \langle f, \psi_a, b\rangle = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{a}} \psi * \left(\frac{t-b}{a}\right) dt \tag{4}$$

where a is the scale factor and b is the displacement factor. When the two parameters a and b of the basis function ψ_a, b are continuous variables, the above formula is called CWT.

In order to process discrete signals, the wavelet transform needs to be discretized. A common method is to discretize the scale factor a according to the power series, and the displacement factor b is uniformly discretized within the scale, and has a power relationship between the scales, that is, $a = a_0^j, b = nb_0a_0^j; a_0 > 1, b_0 > 0, j, n \in \mathrm{Z}$. The discrete wavelet basis function is expressed as follows:

$$\psi_{j,n}(t) = \frac{1}{\sqrt{a_0^j}} \psi\left(\frac{t - nb_0a_0^j}{a_0^j}\right) = a_0^{-j/2} \psi\left(a_0^{-j}t - nb_0\right). \tag{5}$$

At this time, the DWT of any function $\int(x)$ is

$$\begin{aligned} Wf(j,n) &= \left\langle f(t), \psi_{j,n}(t)\right\rangle = \int_{-\infty}^{\infty} f(t)\psi_{j,n}(t)dt \\ &= a_0^{-j/2} \int_{-\infty}^{\infty} f(t)\psi\left(a_0^{-j}t - nb_0\right)dt. \end{aligned} \tag{6}$$

This study uses the Mallat algorithm [32] to realize the multilevel wavelet transform in the CNN model. Given the scale function $\phi(t)$ and wavelet function $\psi(t)$ of wavelet transform, the intrinsic relationship between them can be expressed as a dual-scale equation

$$\phi(t) = \sum_n h(n)\phi_{-1,n}(t) = \sqrt{2}\sum_n h(n)\phi(2t-n) \tag{7}$$

$$\psi(t) = \sum_n h(n)\phi_{-1,n}(t) = \sqrt{2}\sum_n g(n)\phi(2t-n). \tag{8}$$

It can be obtained

$$h(n) = \left\langle \phi, \phi_{-1,n}\right\rangle; g(n) = \left\langle \psi, \phi_{-1,n}\right\rangle. \tag{9}$$

The filter banks $h(n)$ and $g(n)$ are determined by the scale function and the wavelet function of the wavelet transform, which have low-pass and high-pass properties, respectively. <·> means inner product operation. Based on the Mallat algorithm, given any function $x(t)$ the following formula can be derived:

$$\begin{aligned} a_{j,k} &= \left\langle x(t), \phi_{j,k}(t)\right\rangle = \int_{-\infty}^{\infty} x(t)2^{-j/2}\phi\left(2^{-j}t - k\right)dt \\ &= \sum_m h(m-2k) \int_{-\infty}^{\infty} x(t)2^{(-j+1)/2}\phi\left(2^{-j+1}t - m\right)dt. \\ &= \sum_m h(m-2k)\ a_{j-1,m} \end{aligned} \tag{10}$$

The same can be obtained

$$\begin{aligned} d_{j,k} &= \left\langle x(t), \psi_{j,k}(t)\right\rangle = \int_{-\infty}^{\infty} x(t)2^{-j/2}\psi\left(2^{-j}t - k\right)dt \\ &= \sum_m g(m-2k) \int_{-\infty}^{\infty} x(t)2^{(-j+1)/2}\psi\left(2^{-j+1}t - m\right)dt. \\ &= \sum_m g(m-2k)\ a_{j-1,m}. \end{aligned} \tag{11}$$

a_j, k is the scale coefficient, and d_jk is the wavelet coefficient. Suppose $\mathcal{Z}^i(n)$ is the ith feature signal of feature map $\mathcal{Z}$ which is a discrete signal. Set the start of the algorithm as$A_0^i(n) = \mathcal{Z}^i(n)$ the recursive formula of the above process is expressed as

$$\begin{cases} A_\tau^i(n) = \sum_m h(m-2n)\ A_{\tau-1}^i(m) \\ D_\tau^i(n) = \sum_m g(m-2n)\ A_{\tau-1}^i(m) \end{cases}, i = 1, 2, \ldots, C \tag{12}$$

where $A_\tau^i(n)$ is the approximation coefficient (low-frequency component) of the signal $\mathcal{Z}^i(n)$ and $D_T^i(n)$ is the detail coefficient (high-frequency component) of the signal $\mathcal{Z}^i(n)$. To integrate this process into the CNN model, a DWT-driven FL-Layer is proposed whose detailed structure is shown in Fig. 2. As Fig. 2, each feature signal of the feature map $\mathcal{Z}$ is decomposed by DWT in parallel and independently, and this process can be quickly implemented on the GPU. To achieve multiresolution analysis, the three-level DWT in FL-Layer is implemented, i.e., (12) is iterated three times. Each level of DWT decomposition will decompose the low-frequency components of the previous level to obtain more detailed low-frequency components and high-frequency components. Finally, feature maps (D_1, D_2, D_3, and A_3) with different frequency components can be obtained. Then, these four feature maps are spliced into a new mixed frequency feature map $\mathcal{L}$ according to the channel, and $\mathcal{L}$ can be expressed as

$$\begin{aligned} \mathcal{L} &= (D_1, D_2, D_3, A_3) \in \mathbb{R}^{C\times L} \\ &\text{where } D_1 \in \mathbb{R}^{C\times L/2}, D_2 \in \mathbb{R}^{C\times L/4}, D_3, A_3 \in \mathbb{R}^{C\times L/8}. \end{aligned} \tag{13}$$

The feature learning module of the FL-Layer is a convolution module with a group normalization (GN) layer [33] and a ReLU function, and valuable frequency information is obtained after the feature learning of the convolution layer.

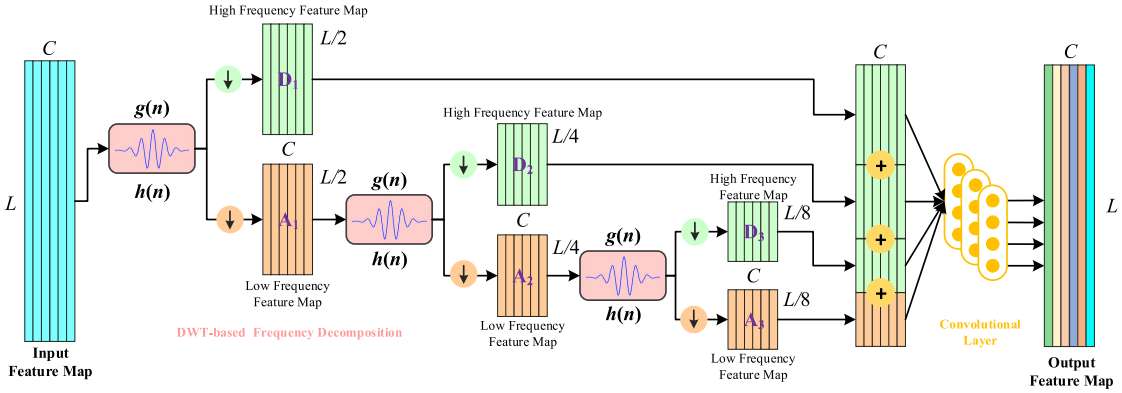

Fig. 2. Proposed DWT-driven FL-Layer, which consists of the wavelet-driven frequency decomposition module and the convolution-driven feature learning module.

2) DWT-Driven Backpropagation: FL-Layer can be integrated into deep learning models for backpropagation and parameter optimization. Next, the wavelet-driven backpropagation is analyzed. Assume the error that passes to $\mathscr{L}$ is $\delta_{\mathscr{L}} \varepsilon \mathbb{R}^{C \times L}$ Since $\mathscr{L}$ is produced by concatenating and D_1, D_2, D_3, and A_3 according to their channels, these four feature maps' corresponding errors are cut from $\delta_{\mathscr{L}}$ the same sequence, which results in: $\delta_D 1 \varepsilon \mathbb{R}^{C \times L/2}$, $\delta_D 2 \varepsilon \mathbb{R}^{C \times L/4}$, $\delta_{D3} \varepsilon \mathbb{R}^{C \times L/8}$, and $\delta_{A3} \varepsilon \mathbb{R}^{C \times L/8}$ First, for the third DWT that produces $_{D3}$ and $_{A3}$ since they are produced by (12), which is a downsample operation through convolution with a factor of 2, the backpropagation first requires to upsample δ_{A3} and δ_{D3} to recover the original length

$$\tilde{\delta}_{D_3,A_3}(n) = \begin{cases} \delta_{D_3,A_3}(n/2), & n/2 \text{ are integers} \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Note that $\tilde{\delta}_{D_3,A_3} \varepsilon \mathbb{R}^C \times L/4$. Then, the error is passed to the input feature map of the third DWT by

$$\delta^i_{A_2} = \tilde{\delta}^i_{D_3} * \text{ reverse } (g_3) + \tilde{\delta}^i_{A_3} * \text{ reverse } (h_3). \tag{15}$$

δ_{A_2} is the error that passed to the input of third DWT, which is also the low-frequency part from the second DWT. h_3 and g_3 are the low-pass and high-pass filters for the third DWT. reverse means the vector is reversed, i denotes the ith channel, $*$ convolution. Note that for backpropagation, we take every channel of the upsampled error and do convolution with the low-pass or high-pass filters individually. The results are concatenated to form a new error map.

For the second DWT, the propagation process is related to δ_{A_2} and δ_{D_2} Similarly, they are first upsampled by a factor of 2

$$\tilde{\delta}_{D_2,A_2}(n) = \begin{cases} \delta_{D_2,A_2}(n/2), & n/2 \text{ are integers} \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

Then by the low-pass filter h_3 and high-pass filter g_3 for the second DWT, the error is passed to the low-frequency output of the first DWT by

$$\delta^i_{A_1} = \tilde{\delta}^i_{D_2} * \text{reverse}(g_2) + \tilde{\delta}^i_{A_2} * \text{reverse}(h_2). \tag{17}$$

δ_{A_1} is the error of the low-frequency output of the first DWT.

For the first DWT, δ_{A_1} and δ_{D_1} are first upsampled with the same procedure by

$$\tilde{\delta}_{D_1,A_1}(n) = \begin{cases} \delta_{D_1,A_1}(n/2), & n/2 \text{ are integers} \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

Then, the error is passed to the input of the DWT-driven FL-Layer by

$$\delta^i_Z = \tilde{\delta}^i_{D_1} * \text{reverse}(g_1) + \tilde{\delta}^i_{A_1} * \text{reverse}(h_1). \tag{19}$$

h_1 and g_1 the low-pass and high-pass filters for the first DWT, δ_Z the error of the input of the DWT-driven FL-Layer.

C. Deep Frequency Separation

The advantage of DWT is its ability to perform multiresolution analysis of the signal, separating frequency components in a complex signal layer by layer. This facilitates the feature learning module in capturing information hidden in the frequency domain. In order to take advantage of DWT as much as possible, the deep frequency separation mechanism is proposed, which is denoted in Fig. 3. DWT-based multiresolution frequency decomposition and convolution-based feature learning are performed repeatedly. This hierarchical frequency separation and feature learning integrates the advantages of DWT and CNN as much as possible. DWT decomposes the different frequency components of signals layer by layer and sends them to the convolution layer. The convolution layer learns valuable information layer by layer and inputs the learned information into DWT for more detailed decomposition. By decomposing and discarding various useless frequency components and noises of complex signals layer by layer, WaveGNet retains weak feature information related to the mechanical system's state. The training algorithm of WaveGNet is shown in Algorithm 1. $D = \{\Gamma^i, \Im^i\}_i^N = 1$ represents the dataset, Γ^i represents the data sample, and its corresponding label is Γ^i FC indicates the fully connected layer. In this study, E was set to 200 and H was set to 5.

D. Implementation Details

The proposed model was implemented using PyTorch and Python and trained on a server with an NVIDIA GeForce RTX 3090 GPU and Intel 10900K CPU. The Adam optimization

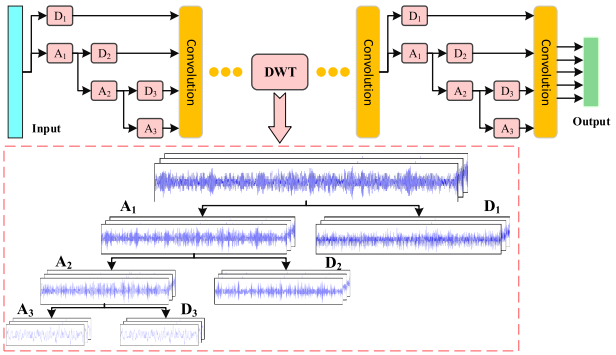

Fig. 3. Deep frequency decomposition of signals.

Algorithm 1 Wavelet-Driven Convolutional Neural Network

Input: DatasetD = $D = \{\Gamma^i, \Im^i\}_i^N = 1$ trained by mini-batch
Output: Optimized Predictor $\mathcal{F}(v)$
1: Set $e = 0$ and epoch E ; set H.
2: Initialize the ϑ; Normalize the dataset D
3: **while** $e \leq E$ **do**
4: Sample a batch of data B from D; set $\hbar = 0$.
5: Perform convolution to extract features: $A_0^0 = ReLU(\mathbf{W}^T B + b)$
6: **while** $\hbar < H$ **do**
7: Calculate DWT according to Eq. (12), obtained $A_1^\hbar$ and $D_1^\hbar$
8: Calculate DWT according to Eq. (12), obtained $A_2^\hbar$ and $D_2^\hbar$
9: Calculate DWT according to Eq. (12), obtained $A_3^\hbar$ and $D_3^\hbar$
10: Get mixed frequency features $\mathscr{L} = (D_1^\hbar, D_2^\hbar, D_3^\hbar, A_3^\hbar)$
11: Perform convolution to extract features: $A_0^\hbar$ ReLU $(\mathbf{W}^\hbar * \mathcal{L}^\hbar + b^\hbar)$
12: Set $\hbar =^\hbar +1$
13: **end while**
14: Calculate the output of $\mathcal{F}(\vartheta)$: $\Im_p = \mathcal{F}(B|\vartheta) = FC(A_0^H)$
15: Calculate $\partial L(\Im, \Im_p)/\partial\vartheta$, update ϑ by back propagation
16: If an epoch is completed, set $e = e + 1$
17: **end while**

algorithm was employed with a learning rate of 0.0001 and a batch size of 64. Z-score normalization was applied to the data samples to ensure stable training. These parameter values were chosen based on empirical knowledge and grid search, and experimental results showed that it can effectively utilize the model's performance. The db12 wavelet, commonly used in vibration-based condition monitoring applications, was chosen for the experimentation. In Section IV-B4, experiments were conducted to analyze the performance of ten different wavelet functions to select the optimal one. To minimize accidental results, each experiment was repeated four times.

During the model design, a grid search algorithm revealed that employing five DWT-driven FL-Layers could achieve optimal performance. Hence, the proposed WaveGNet in this experiment comprises five DWT-driven FL-Layers, a global average pooling layer, and a fully connected layer.

The design concept of the DWT-driven FL-Layer was inspired by the multiresolution wavelet decomposition algorithm in the signal analysis field. When performing frequency analysis on vibration signals, a commonly used practice is a three-level multiresolution decomposition, which is an empirical choice. A three-level decomposition can offer sufficient frequency resolution while avoiding excessive decomposition.

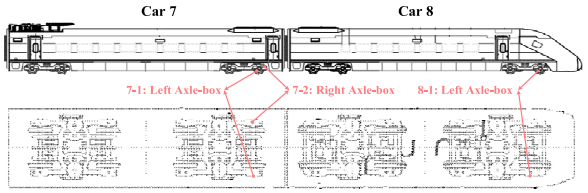

Fig. 4. Installation location of the acceleration sensors. Car 7 is a trail car, and Car 8 is a motor car.

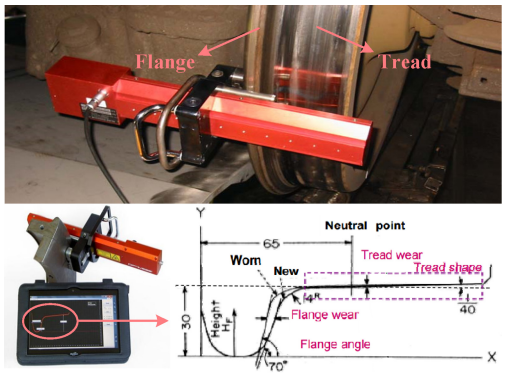

Fig. 5. Measured wear profile via measurement instrument.

While additional decomposition levels might provide finer frequency information, they also increase computational complexity. In most practical scenarios, three-level decomposition balances preserving information and computational efficiency. Therefore, each FL-Layer includes a 3-level DWT, a convolution module, and a dropout layer (dropout rate: 0.1) to prevent overfitting. The convolution modules consist of a convolutional layer, a GN layer, and a ReLU activation function, utilizing a 5×1 convolution kernel to enhance the model's feature perception ability without significantly increasing the number of parameters. The model is a versatile framework that delivers excellent results for both classification and regression tasks. It employs cross-entropy (CE) loss for classification and mean-squared error (MSE) loss for regression, which are two classic and practical loss functions.

IV. Experimental Validation

A. Case 1: High-Speed Train Wheel Wear Prediction

1) Data Description: The study conducted signal acquisition over six months on a CRH1A HST, which typically operates at 200 km/h and reaches a maximum speed of 250 km/h. The train traveled on three different lines: 1) Changsha Huaihua Line; 2) Guangzhou Zhuhai Line; and 3) Sanya Haikou Line. Vibration signals were collected using on-board acceleration sensors installed on the axle boxes of the 7th and 8th cars (refer to Fig. 4), and three wheel vibration signals were collected (7–1, 7–2, and 8–1). Since the data was collected from actual train operations, the signals were complex and contained noise components. Wheel wear was measured approximately every week during signal acquisition. Wheel tread and flange wear were calculated separately using a mobile laser-based wheel profile measurement system, as shown in Fig. 5. The degree of wheel tread and flange wear

TABLE I
TREAD AND FLANGE WEAR PREDICTION RESULTS OF WK-1DCNN, OW-1DCNN, MW-1DCNN, AND WAVEGNET

Wheel	Method	Tread Wear			Flange Wear		
-	-	MSE ($\times 10^{-3}$)	MAE ($\times 10^{-2}$)	R^2-Score	MSE ($\times 10^{-3}$)	MAE ($\times 10^{-2}$)	R^2-Score
7-1	WK-1DCNN	3.578 ± 0.038	4.447 ± 0.066	0.888 ± 0.001	7.292 ± 0.420	5.416 ± 0.173	0.915 ± 0.004
	OW-1DCNN	1.466 ± 0.058	2.780 ± 0.076	0.955 ± 0.002	3.361 ± 0.287	3.455 ± 0.099	0.961 ± 0.004
	MW-1DCNN	1.606 ± 0.128	2.646 ± 0.181	0.950 ± 0.004	4.344 ± 0.367	3.123 ± 0.170	0.952 ± 0.006
	WaveGNet	**0.927 ± 0.057**	**1.988 ± 0.080**	**0.971 ± 0.002**	**2.270 ± 0.138**	**2.394 ± 0.065**	**0.975 ± 0.001**
7-2	WK-1DCNN	6.914 ± 0.057	6.064 ± 0.055	0.743 ± 0.002	7.452 ± 0.395	5.133 ± 0.076	0.879 ± 0.008
	OW-1DCNN	2.544 ± 0.171	3.793 ± 0.133	0.906 ± 0.006	6.616 ± 0.813	4.967 ± 0.387	0.896 ± 0.013
	MW-1DCNN	2.128 ± 0.297	3.183 ± 0.235	0.921 ± 0.011	5.228 ± 0.676	3.859 ± 0.234	0.918 ± 0.008
	WaveGNet	**1.560 ± 0.033**	**2.775 ± 0.074**	**0.941 ± 0.001**	**5.041 ± 0.653**	**3.510 ± 0.104**	**0.919 ± 0.012**
8-1	WK-1DCNN	10.086 ± 0.302	7.561 ± 0.212	0.679 ± 0.011	6.619 ± 0.333	5.823 ± 0.105	0.899 ± 0.004
	OW-1DCNN	2.593 ± 0.178	3.670 ± 0.113	0.917 ± 0.006	2.184 ± 0.074	3.468 ± 0.036	0.967 ± 0.001
	MW-1DCNN	1.786 ± 0.114	2.659 ± 0.083	0.943 ± 0.003	1.810 ± 0.346	2.489 ± 0.145	0.973 ± 0.005
	WaveGNet	**1.518 ± 0.105**	**2.513 ± 0.109**	**0.951 ± 0.003**	**1.101 ± 0.184**	**2.270 ± 0.109**	**0.983 ± 0.002**

are primary indicators of wheel health. The study also found that wear-related features are mostly contained in the low-frequency components of the signal, based on the wheel polygon wear mechanism [34]. By mining this information, wheel wear predictions can be made.

The dataset covers the period after all train wheels were lathed and ends after traveling 185 000 km. For the purpose of analysis, it is assumed that the wheel was in its initial state after being lathed and had zero wear at that time. To avoid the influence of track irregularity, acceleration, and deceleration on the prediction results, the smooth running data between the two stations are selected to train the proposed model. Sliding segmentation is used to obtain the required training samples. The length of each sample is set to 3072 $\times 1$, and finally 39 532 training samples and 8764 testing samples are obtained. This article uses three regression prediction performance metrics. They are: MSE, mean absolute error (MAE), and coefficient of determination (R^2-Score).

2) Effectiveness of the Proposed Method: This section conducts ablation experiments and compares the results of four network models. The first model, WaveGNet, is the proposed wavelet-driven CNN model. The second model, WK-1DCNN, is WaveGNet with all DWT-driven FL-Layers removed. The third model, OW-1DCNN, is a model where a 3-level wavelet transform is added to WK-1DCNN. Finally, MW-1DCNN is a model where the 3-level wavelet transform of FL-Layer in WaveGNet is replaced with a 1-level wavelet transform. All four models have the same convolution and classification layer parameters except for the wavelet transform. The performance of these networks is shown in Table I.

According to Table I, integrating DWT into the CNN model significantly improves the model's prediction performance. Specifically, WK-1DCNN performs the worst on the three wheels, and when a 3-level DWT is added, the performance is significantly improved. For instance, OW-1DCNN outperforms WK-1DCNN significantly, with MSE and R^2-Scores of 2.593 $\times 10^{-3}$ and 0.917, respectively, for tread wear prediction on wheel 8-1. This shows that DWT can greatly improve the feature learning ability of CNN. This study proposes a creative integration of DWT into the CNN model through the DWT-driven FL-Layer, which effectively combines the strengths of wavelet transform and CNN, resulting in further improvements in the model's predictive performance. WaveGNet outperforms WK-1DCNN and OW-1DCNN, demonstrating the effectiveness of wavelet-driven CNNs. The proposed algorithm employs frequency decomposition and feature learning to obtain wear-related features from complex signals, resulting in accurate prediction of wheel wear. Furthermore, the comparison of WaveGNet with MW-1DCNN also validates the effectiveness of deep frequency separation. WaveGNet decomposes the signal layer by layer, expanding the feature learning space of CNN into the frequency domain, thereby enhancing the model's ability to extract valuable frequency information. This innovative learning mechanism significantly improves the performance of conventional CNNs.

3) Compared With Existing Prediction Methods: This study compares the proposed WaveGNet method with several existing deep-learning-based prediction methods, including 1D-VGG16 [35], 1D-ResNet18 [7], Bi-LSTM [36], and Transformer [37]. In addition, this study compared three novel comparison methods: 1) the improved multi-LSTM model (IM-LSTM); 2) the LGF-Trans model based on Transformer [11]; and 3) the Wavelet LSTM model. The LGF-Trans model is a multiscale local-global time fusion framework based on Transformer. Wavelet LSTM combines wavelet transform and LSTM. Due to the challenge of integrating wavelet transform into the internal structure of LSTM, the signal is first decomposed by DWT and then input into LSTM in this architecture. The performance of these methods on three wheels is shown in Table II.

The results show that WaveGNet outperforms other deep-learning-based prediction methods on the three wheels. While Transformer and Bi-LSTM, which have long-term feature learning ability, generally perform better than 1D-VGG16 and 1D-ResNet18, this suggests that the CNN-based models have limited time-series signal processing capability. However, WaveGNet overcomes this limitation by significantly enhancing the model's time-series prediction capability. For instance, on wheel 7-1, the MSE and R^2-Score of WaveGNet for tread wear prediction are 0.927×10^{-3} and 0.971, respectively. In comparison, the MSE and R^2-Score of 1D-VGG16 are only 4.120×10^{-3} and 0.875, and those of Bi-LSTM are only 1.811×10^{-3} and 0.945, respectively. Similarly, for flange wear prediction, the MSE and R^2-Score of WaveGNet on wheel 7−1 are 4.383 $\times 10^{-3}$ and 0.975, while those of 1D-VGG16 are only 9.586×10^{-3} and 0.898, and those of Bi-LSTM are only 4.383×10^{-3} and 0.953. Moreover, Wavelet LSTM shows performance improvement compared to LSTM-based model, indicating the benefits of wavelet transform. However, since wavelet and LSTM are not deeply and closely integrated

TABLE II
TREAD AND FLANGE WEAR PREDICTION RESULTS OF 1D-VGG16, 1D-RESNET18, TRANSFORMER, BI-LSTM, AND WAVEGNET

Wheel	Method	Tread Wear			Flange Wear		
-	-	MSE ($\times 10^{-3}$)	MAE ($\times 10^{-2}$)	R^2-Score	MSE ($\times 10^{-3}$)	MAE ($\times 10^{-2}$)	R^2-Score
7-1	1D-VGG16	4.120 ± 0.650	5.035 ± 0.383	0.875 ± 0.019	9.586 ± 1.222	7.031 ± 0.942	0.898 ± 0.013
	1D-ResNet18	3.220 ± 0.159	4.360 ± 0.137	0.903 ± 0.004	10.047 ± 0.344	5.690 ± 0.125	0.894 ± 0.004
	Transformer	2.588 ± 0.639	3.444 ± 0.642	0.922 ± 0.019	8.913 ± 1.090	4.288 ± 0.915	0.905 ± 0.012
	Bi-LSTM	1.811 ± 0.077	2.309 ± 0.021	0.945 ± 0.002	4.383 ± 0.601	2.398 ± 0.046	0.953 ± 0.006
	IM-LSTM	1.906 ± 0.053	2.380 ± 0.084	0.941 ± 0.002	4.490 ± 0.233	2.423 ± 0.068	0.951 ± 0.003
	Wavelet LSTM	1.537 ± 0.064	2.018 ± 0.087	0.953 ± 0.004	3.802 ± 0.193	**2.311 ± 0.073**	0.954 ± 0.005
	LGF-Trans	2.346 ± 0.147	3.075 ± 0.106	0.929 ± 0.005	5.932 ± 0.732	3.642 ± 0.311	0.936 ± 0.008
	WaveGNet	**0.927 ± 0.057**	**1.988 ± 0.080**	**0.971 ± 0.002**	**2.270 ± 0.138**	2.394 ± 0.065	**0.975 ± 0.001**
7-2	1D-VGG16	7.262 ± 1.664	6.831 ± 0.922	0.736 ± 0.061	10.211 ± 1.830	6.718 ± 0.697	0.846 ± 0.026
	1D-ResNet18	7.061 ± 0.098	6.296 ± 0.093	0.744 ± 0.001	8.507 ± 0.704	5.851 ± 0.072	0.868 ± 0.010
	Transformer	5.804 ± 0.290	5.461 ± 0.248	0.789 ± 0.011	6.591 ± 0.725	4.776 ± 0.295	0.896 ± 0.016
	Bi-LSTM	5.030 ± 0.111	4.563 ± 0.036	0.818 ± 0.003	5.605 ± 0.126	3.754 ± 0.066	0.912 ± 0.002
	IM-LSTM	5.179 ± 0.093	4.657 ± 0.032	0.809 ± 0.002	6.616 ± 0.155	4.014 ± 0.069	0.893 ± 0.002
	Wavelet LSTM	3.792 ± 0.048	4.403 ± 0.091	0.886 ± 0.003	5.261 ± 0.219	3.528 ± 0.173	0.918 ± 0.003
	LGF-Trans	4.290 ± 0.376	4.732 ± 0.285	0.844 ± 0.013	7.570 ± 0.323	5.048 ± 0.205	0.883 ± 0.007
	WaveGNet	**1.560 ± 0.033**	**2.775 ± 0.074**	**0.941 ± 0.001**	**5.041 ± 0.653**	**3.510 ± 0.104**	**0.919 ± 0.012**
8-1	1D-VGG16	6.310 ± 0.679	6.021 ± 0.527	0.804 ± 0.022	6.344 ± 1.115	5.930 ± 0.490	0.910 ± 0.016
	1D-ResNet18	10.535 ± 0.568	7.755 ± 0.201	0.673 ± 0.017	6.558 ± 0.430	6.143 ± 0.226	0.905 ± 0.006
	Transformer	6.697 ± 0.093	5.526 ± 0.137	0.790 ± 0.004	5.095 ± 0.059	4.988 ± 0.071	0.925 ± 0.001
	Bi-LSTM	6.080 ± 0.121	4.909 ± 0.026	0.811 ± 0.003	5.085 ± 0.185	4.577 ± 0.020	0.926 ± 0.002
	IM-LSTM	6.407 ± 0.202	5.227 ± 0.119	0.799 ± 0.006	5.711 ± 0.167	4.776 ± 0.088	0.917 ± 0.001
	Wavelet LSTM	5.281 ± 0.237	4.592 ± 0.178	0.816 ± 0.007	4.771 ± 0.172	4.216 ± 0.083	0.937 ± 0.003
	LGF-Trans	4.043 ± 0.392	4.132 ± 0196	0.874 ± 0.012	3.346 ± 0.091	3.954 ± 0.121	0.952 ± 0.002
	WaveGNet	**1.518 ± 0.105**	**2.513 ± 0.109**	**0.951 ± 0.003**	**1.101 ± 0.184**	**2.270 ± 0.109**	**0.983 ± 0.002**

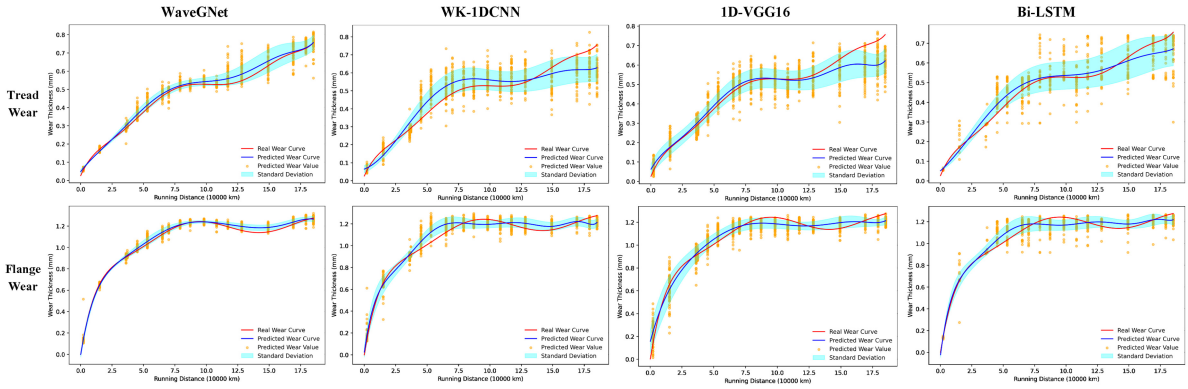

Fig. 6. Tread wear curves and flange wear curves were predicted by WaveGNet, WK-1DCNN, 1D-VGG16, and BI-LSTM. The actual wear curve is represented by the red line, and the blue line represents the predicted wear curve obtained by averaging the predicted wear values. The orange dots are the predicted wheel wear values. The cyan bar represents the standard deviation of the predicted wear values.

in Wavelet LSTM, its performance is much lower than the proposed WaveGNet. The results confirm the effectiveness of integrating wavelet transform into CNN, which allows learning signal features from the frequency domain and enhances the model's ability to extract valuable information from complex signals with significant noise. The use of deep frequency decomposition enables accurate modeling of the nonlinear mapping relationship between vibration signals and wheel wear, as wear-related features are extracted layer by layer.

4) Visualization of Predicted Wear Curves: The wear curves predicted by different models for wheel 8–1 are presented and analyzed in this section, as shown in Fig. 6. The wear curves predicted by WaveGNet, WK-1DCNN, 1D-VGG16, and Bi-LSTM are compared. The predicted wear curve of WaveGNet closely matches the actual wear curve, indicating that WaveGNet outperforms the other models in accurately modeling the relationship between vibration signals and wheel wear. In addition, the standard deviation of WaveGNet is smaller than that of the other models, indicating that the predicted wear values are more tightly clustered around the actual values, and the predictions are more stable.

Comparing the wear curves predicted by WaveGNet and WK-1DCNN, it is intuitive to see the improvement brought by the wavelet transform to the CNN model. After integrating the wavelet transform, the predicted wear curves of the CNN model are more accurate, and the scatter of predicted values is significantly reduced. This suggests that the wavelet transform can enhance the prediction accuracy and stability of the model by expanding the feature learning space of the CNN to the frequency level.

5) Computational Burden Analysis: This section analyzes the computational burden of our proposed method. Pytorch's built-in functions and ptflops library were used to obtain the parameters and calculation amount of the model, respectively. Table III displays the parameters and multiply–accumulate operations (MACs) of WaveGNet and four comparison methods. WaveGNet requires only 0.496M parameters, which is significantly less than the other models. This highlights that deep learning models need a large number of parameters to achieve feature learning through random initialization. In contrast, the wavelet transform has strong theoretical foundations and signal analysis abilities. WaveGNet leverages the wavelet transform to achieve excellent diagnostic performance with very few parameters. Furthermore, the MACs of WaveGNet are only 0.08G, demonstrating that WaveGNet achieves high performance with low computational requirements.

TABLE III
PARAMETERS AND COMPUTATIONS FOR THE FIVE MODELS

-	WaveGNet	Bi-LSTM	Transformer	1D-ResNet18	1D-VGG16
Parameters	0.496M	5.26M	19.01M	3.88M	109.68M
MACs	0.08G	0.33G	0.14G	0.08G	1.0G

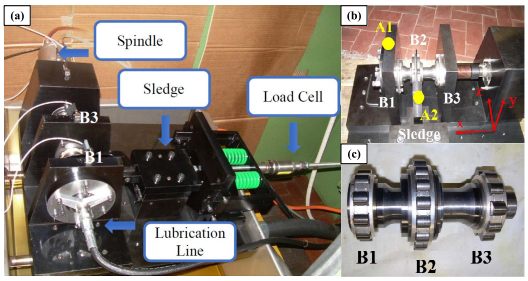

Fig. 7. High-speed aeronautical bearings test rig. (a) General view of the test rig. (b) Positions of the accelerometers and the reference system. (c) Shaft with its three roller bearings.

TABLE IV
DESCRIPTION OF THE HSA BEARING DATASET INFORMATION

Defect	Dimension	Load	Speed	Label
No defect	–	0N-1800N	100Hz-500Hz	F1
Diameter of an indentation on the inner ring	450μm	0N-1800N	100Hz-500Hz	F2
Diameter of an indentation on the inner ring	250μm	0N-1800N	100Hz-500Hz	F3
Diameter of an indentation on the inner ring	150μm	0N-1800N	100Hz-500Hz	F4
Diameter of an indentation on a roller	450μm	0N-1800N	100Hz-500Hz	F5
Diameter of an indentation on a roller	250μm	0N-1800N	100Hz-500Hz	F6
Diameter of an indentation on a roller	150μm	0N-1800N	100Hz-500Hz	F7

B. Case 2: High-Speed Aviation Bearings Fault Diagnosis

1) Data Description: In this case, the health status of bearings were monitored using a high-speed aviation (HSA) bearing test rig [38], as shown in Fig. 7. Two acceleration sensors were installed at A1 and A2, respectively, and the inner rings of bearings B1, B2, and B3 were connected to a specially designed short and thick hollow shaft capable of handling speeds up to 35 000 rpm. The B1 bearing, which includes one normal state and six fault states listed in Table IV, was chosen as the monitoring object. Data samples were collected under different speed and load conditions, with 100 Hz corresponding to 6000 rpm. The sliding segmentation method was used for data augmentation, with a sample length of 4096. A total of 22 134 training samples and 7259 testing samples were obtained, and the performance of the network model was evaluated using the accuracy (ACC) metric and confusion matrix. To verify the model's performance in a noisy environment, Gaussian white noise with SNR levels of 4, 0, −2, and −4 dB was added to the vibration signal.

2) Effectiveness of the Proposed Method: This section compares and analyzes four network models: 1) WK-1DCNN; 2) OW-1DCNN; 3) MW-1DCNN; and 4) WaveGNet. The parameter configuration of each model is the same as in Section IV-B2, with the CE Loss used to adapt to the fault diagnosis application. The results under SNR levels of 4, 0, −2, and −4 dB noise are presented in Table V.

Table V demonstrates that WK-1DCNN achieves 92.67% accuracy at SNR = 4 dB but only 52.72% at SNR = −4 dB, indicating poor noise robustness of the common CNN model. This observation has also been reported in previous studies [12], [13]. This experiment shows that the wavelet transform can significantly improve the noise robustness of the CNN model. For instance, OW-1DCNN achieves nearly 20% higher accuracy than WK-1DCNN when SNR = −4 dB after adding a 3-level DWT. This finding strongly supports the notion that feature learning from the frequency level can considerably enhance CNN performance. Moreover, WaveGNet fully integrates the advantages of DWT and CNN, achieving the best performance across all four noise situations. Specifically, WaveGNet achieves 99.13% accuracy at SNR = 4 dB and still obtains 88.79% accuracy at SNR = −4 dB. These results suggest that wavelet-driven CNN can filter out irrelevant noise information and capture fault-related information from noisy signals, significantly enhancing the diagnostic model's performance and noise robustness.

TABLE V
EXPERIMENTAL RESULTS OF WK-1DCNN, OW-1DCNN, MW-CNN, AND WAVEGNET UNDER FOUR KINDS OF NOISE

Noise	WK-1DCNN	OW-1DCNN	MW-1DCNN	WaveGNet
4 dB	92.67 ± 0.64	97.24 ± 0.36	95.93 ± 0.31	**99.13 ± 0.14**
0 dB	81.95 ± 0.77	90.84 ± 0.47	92.59 ± 0.32	**96.81 ± 0.07**
−2 dB	67.87 ± 1.29	82.73 ± 0.57	87.74 ± 0.38	**93.76 ± 0.48**
−4 dB	52.72 ± 1.37	73.48 ± 0.54	78.88 ± 0.55	**88.79 ± 0.50**

3) Compared With Existing Diagnostic Methods: In this experiment, WaveGNet is compared with various deep-learning-based fault prediction methods, including MA1DCNN [27], Bi-LSTM [36], Wen-CNN [39], 1D-VGG16 [35], and 1D-ResNet18 [7]. MA1DCNN is a multiattention CNN-based model, while Wen-CNN is a 2DCNN-based model that first converts the signal into a 2-D representation. In addition, this study compared three novel comparison methods: 1) the hybrid model based on CNN and transformer (CNN-Trans); 2) Mexhat-Net; and 3) Morlet-Ne. CNN-Trans fully integrates the advantages of CNN in local feature learning and Transformer in global feature learning. Both Mexhat-Net and Morlet-Net were proposed by Li et al. [26], integrating different wavelet basis functions into the first convolutional layer of the deep model to achieve better signal decomposition and feature learning. The fault diagnosis results are summarized in Table VI.

Table VI shows that WaveGNet outperforms this fault prediction models across all noise levels (−4, −2, 0, and 4 dB). WaveGNet achieves 99.13% accuracy when SNR is 4 dB and 96.81% accuracy when SNR is 0 dB, which indicates its ability to skillfully deal with noise interference in real industrial environments. In comparison, when SNR is 0 dB, MA1DCNN achieves only 84.11% accuracy, Bi-LSTM achieves only 76.64% accuracy, and 1D-VGG16 achieves only 79.18% accuracy. Moreover, the eight comparison methods are highly sensitive to strong noise, leading to a significant decline in their performance. For instance, at SNR = −4 dB, the MA1DCNN's accuracy reduces to 64.81%, Bi-LSTM's accuracy drops to 58.26%, and 1D-VGG16's accuracy decreases to 48.61%. This indicates that these methods lack the ability to counter noise, making it challenging to extract valuable features from noisy signals. WaveGNet, on the

TABLE VI
EXPERIMENTAL RESULTS OF WAVEGNET, MA1DCNN, BI-LSTM, WEN-CNN,1D-VGG16, AND 1D-RESNET18 UNDER FOUR KINDS OF NOISE

Noise	WaveGNet	Mexhat-Net	Morlet-Net	CNN-Trans	MA1DCNN	Bi-LSTM	Wen-CNN	1D-VGG16	1D-ResNet18
4 dB	**99.13±0.14**	94.13±0.52	95.76±0.37	92.47±0.61	94.06±0.77	89.90±0.39	90.44±0.96	91.55±0.48	85.31±0.67
0 dB	**96.81±0.07**	86.82±0.81	87.14±0.45	85.05±0.93	84.11±0.90	76.64±0.38	78.91±0.76	79.18±0.42	63.58±0.56
−2 dB	**93.76±0.48**	80.05±0.73	80.71±0.47	74.69±2.17	76.72±0.89	67.81±0.86	72.10±0.54	65.50±0.82	48.06±1.03
−4 dB	**88.79±0.50**	71.44±0.76	71.49±1.06	62.08±2.01	64.81±1.13	58.26±0.60	61.18±2.06	48.61±1.45	38.50±0.87

TABLE VII
EXPERIMENTAL RESULTS OF THE SELECTION OF WAVELET BASIS FUNCTIONS (SNR = 4 DB)

Method	WaveGNet (db4)	WaveGNet (db8)	WaveGNet (db12)	WaveGNet (db16)	WaveGNet (db20)
ACC	98.42±0.08	98.99±0.15	99.13±0.14	99.03±0.12	99.07±0.10
Method	WaveGNet (db24)	WaveGNet (haar)	WaveGNet (dmey)	WaveGNet (coif8)	WaveGNet (sym8)
ACC	99.08±0.09	97.87±0.35	99.14±0.10	99.03±0.11	98.73±0.26

		Predicted Label								
		F1	F2	F3	F4	F5	F6	F7	Recall	Test Number
True Label	F1	1036	0	0	0	0	1	0	**99.90%**	1037
	F2	0	1034	0	3	0	0	0	**99.71%**	1037
	F3	0	0	1030	0	0	7	0	**99.32%**	1037
	F4	3	1	0	1015	0	18	0	**97.88%**	1037
	F5	0	0	0	5	1032	0	0	**99.52%**	1037
	F6	1	0	25	2	0	1009	0	**97.30%**	1037
	F7	1	0	2	1	0	0	1033	**99.61%**	1037
	Precision	**99.52%**	**99.90%**	**97.45%**	**98.93%**	**100%**	**97.49%**	**100%**	—	7259

Fig. 8. Confusion matrix of WaveGNet on HSA bearing dataset (SNR = 4 dB).

		Predicted Label								
		F1	F2	F3	F4	F5	F6	F7	Recall	Test Number
True Label	F1	951	15	14	32	2	15	8	**91.71%**	1037
	F2	16	925	3	39	13	20	21	**89.20%**	1037
	F3	18	5	976	0	0	23	15	**94.12%**	1037
	F4	34	34	5	844	0	102	18	**81.39%**	1037
	F5	0	34	1	30	958	3	11	**92.38%**	1037
	F6	21	37	61	50	2	851	15	**82.06%**	1037
	F7	14	18	21	26	0	9	949	**91.51%**	1037
	Precision	**90.23%**	**86.61%**	**90.29%**	**82.66%**	**98.26%**	**83.19%**	**91.51%**	—	7259

Fig. 9. Confusion matrix of WaveGNet on HSA bearing dataset (SNR = −4 dB).

other hand, leverages the DWT and deep frequency decomposition mechanism to decompose the signal layer by layer, learning useful frequency features while discarding irrelevant noise information. This drastically enhances the model's anti-noise ability, leading to exceptional performance in noisy environments.

Figs. 8 an 9 illustrate the confusion matrices of WaveGNet's prediction results at SNR = 4 dB and SNR = −4 dB, respectively. The diagonal lines represent the correct predictions for each category, and the last column represents the total number of test samples for each category. At low-noise intensity (SNR = 4 dB), WaveGNet achieved high prediction accuracy for each category. Categories F5 and F7 had 100% precision, F2 had 99.90% precision, and F1 had a recall rate of 99.90%. However, when the SNR changed from 4 to −4 dB, the prediction performance for each category degraded. The recall rate of F1, F2, F3, F5, and F7 remained high, approaching or exceeding 90%. However, the recall rate of F4 dropped significantly, possibly because its fault diameter is only 150 um, making the fault degree light and the fault features weak.

4) Discuss the Selection of Wavelet Basis Functions: The proposed method has the flexibility to use various wavelet basis functions to cater to different application scenarios. This section analyzes the impact of the wavelet basis function and its parameters on the proposed method's performance on the HSA dataset with SNR = 4 dB. First, this section takes dbN wavelet as an example to discuss the effect of the value of vanishing moments on the model. This section conducts experiments on six models using db4, db8, db12, db16, db20, and db24 wavelets, respectively, and the results are shown in Table VII. The findings reveal that as the vanishing moments increase from 4 to 12, the diagnostic performance of WaveGNet also increases. Furthermore, as the vanishing moments increase from 12 to 24, the performance of WaveGNet remains mostly the same. Therefore, in this study, the db12 wavelet basis function is adopted. Moreover, different wavelet basis functions is analyzed, and the results in Table VII show that the db12 wavelet and dmey wavelet achieve the best diagnostic performance, while the haar wavelet performs poorly due to its low frequency resolution caused by its rectangular shape.

V. FREQUENCY INTERPRETABILITY ANALYSIS

A. Feature Visualization Details

To understand the learning process of the convolutional layers in WaveGNet, the output features of the convolution module of the DWT-driven FL-Layer on the CWRU dataset is visualized. Fig. 10 displays the features output by the first (F), second (S), and third (T) FL-layers, with the first layer outputting 24 channel feature signals denoted as F_1-1, $F_1-2,\ldots, F_1-24$. Each feature signal is represented as a color band consisting of small grids, with the color of each grid indicating the magnitude of the value. Brighter colors indicate larger values. The second and third FL-layers have 48 and 96 channels, respectively, and we randomly selected 24 feature signals from each for visualization, named as F_2-1, $F_2-2,\ldots$, F_2-24 and F_3-1, $F_3-2,\ldots$, F_3-24. The input features to the convolution module are A_3, D_3, D_2, and D_1 frequency components, with their frequencies increasing from low to high. In Fig. 10, the A_3, D_3, D_2, and D_1 frequency components are marked by red straight lines on the feature maps.

To further analyze the low-frequency components (A_3) of the 24 features output by the first FL-layer, their squared

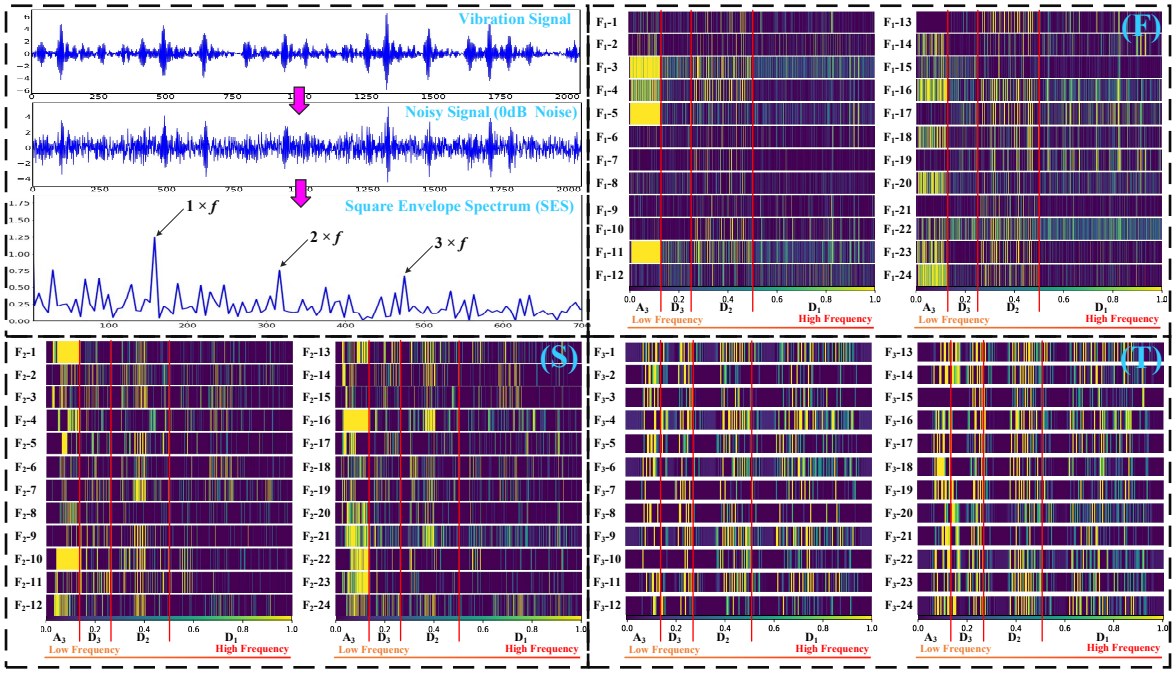

Fig. 10. Features output by the first (F), second (S), and third (T) FL-layers. Each feature signal is represented as a color band consisting of small grids, with the color of each grid indicating the magnitude of the value. It also displays the input signal and noisy signal along with its SES.

envelope spectrum (SES) are obtained and shown in Fig. 11, where $\boldsymbol{f}$ represents the fault characteristic frequency. Fig. 10 also displays the input signal and noisy signal along with its SES.

B. Interpretability Analysis of Feature Learning

Assumption: In the CNN model, features with higher response values output by the convolutional layer are considered more important. This is intuitive as a response close to 0 adds no valuable information, and responses below 0 are discarded by the ReLU function. The interpretability analysis presented below relies on this assumption.

From Fig. 10(F), it is evident that the convolutional layer follows a clear pattern in learning features from the mixed feature map input. Specifically, the layer mainly focuses on learning and retaining features related to the A3 frequency component among the 24 feature signals. Additionally, the convolutional layer exhibits distinct boundaries in learning different frequency components, indicating its ability to effectively separate valuable frequency information from irrelevant ones with the aid of DWT's multiresolution decomposition. Notably, the convolutional layer has a preference for low-frequency features, suggesting that it responds differently to varying frequency components.

Fig. 10(S) displays the output features from the second FL-Layer. Upon passing through the first FL-Layer, the original signal retains mid- and low-frequency features. These features are then fed into the second FL-Layer for frequency decomposition, followed by convolutional layer feature learning. As shown in Fig. 10(S), the convolutional module of the second FL-Layer effectively learns and retains more refined valuable frequency features. In particular, the layer focuses on A_3 frequency components, while partially retaining D_2 frequency components. The second FL-Layer further discards irrelevant information from the original signal, while aggregating and condensing valuable frequency features.

In Fig. 10(T), the features output by the third FL-Layer are presented. Following two rounds of frequency decomposition and feature learning, the valuable frequency features become highly condensed, with a relatively narrow frequency range. The distinctions among the four frequency components are no longer apparent. Thus, the convolution module expands its scope to capture more sophisticated fault-related features from all four frequency components (A_3, D_3, D_2, and D_1).

In summary, the three FL-Layers learn fault-related features from mid-low frequency layers in a hierarchical manner for bearing fault diagnosis. According to the failure mechanism regular low-frequency fault information will be generated when the bearing fails. The fault category can be identified by determining the fault characteristic frequency. This shows that FL-Layer's feature learning adheres to the physical mechanism and proves the effectiveness of wavelet-driven CNN approach.

C. Interpretability Analysis by Squared Envelope Spectrum

In the previous section, it noted that the proposed FL-Layers tend to learn low-frequency features, but do not directly prove their relationship to faults. This section utilizes SES to analyze the low frequency component (A_3) and provide evidence for this claim. Fig. 11 shows that the A_3 component features high response value, and its fault characteristic frequency $\boldsymbol{f}$ in SES is typically clearly visible. However, the A_3 frequency component with low response value is often challenging to find

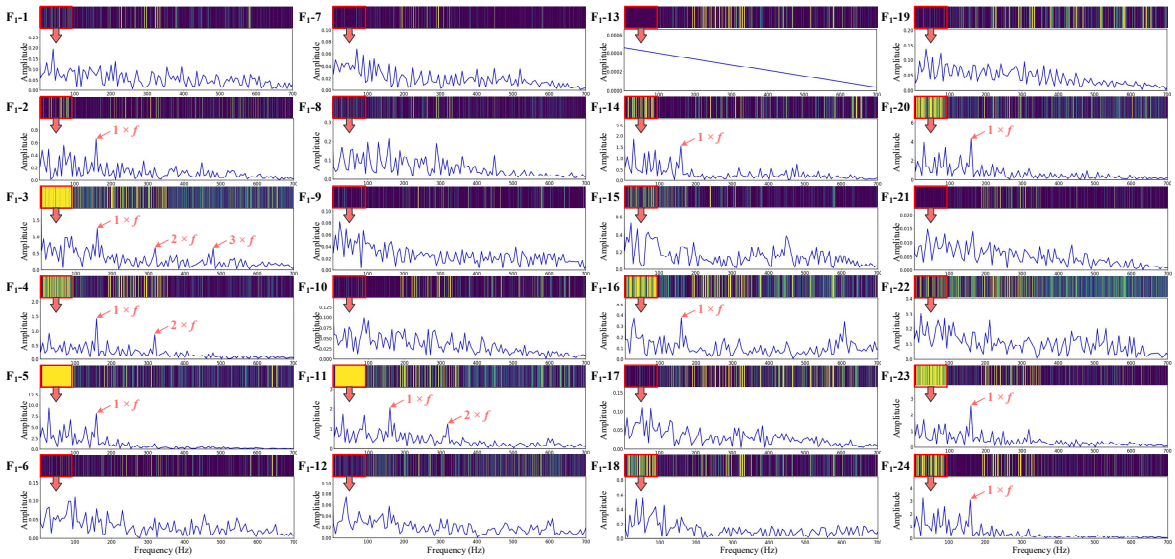

Fig. 11. Low-frequency components (A_3) of the 24 features output by the first FL-layer, their SES.

the $\boldsymbol{f}$. For instance, the $\boldsymbol{f}$ in the SES of A_3 of Feature F_1-3 is apparent, whereas the $\boldsymbol{f}$ in the SES of A_3 of Feature F_1-9 is indistinguishable. These findings support that FL-Layer learns high-value fault-related features from the frequency level, and its feature learning mechanism is well suited to the bearing fault mechanism.

D. Interpretability Analysis of WaveGNet

In summary, WaveGNet achieves accurate learning of valuable key frequency features without prior knowledge, mainly due to the collaborative fusion and optimization mechanism of CNN and DWT. WaveGNet integrates CNN and wavelet transform perfectly, and its feature learning process involves decomposing and learning frequency features from coarse to fine. The wavelet transform performs multiresolution frequency decomposition on the signal with excellent theoretical basis, and the convolution layer automatically learns and discards frequency features with excellent adaptive feature learning ability. Additionally, WaveGNet establishes a network design paradigm based on frequency feature learning, following a relatively clear rule for feature learning. From the perspective of frequency learning, the convolutional modules of WaveGNet exhibit interpretable feature learning preferences and learned features, which follow the physical mechanism of bearing faults as demonstrated by experiments.

E. Discuss Interpretability

For deep learning, the interpretability of models is a highly subjective characteristic, and defining it rigorously through precise mathematical expressions poses a challenge. Generally, the interpretability of deep learning can be viewed as the "degree to which humans understand the decisions or predictions made by a model." This means users can more easily comprehend the decisions and predictions made by models with higher interpretability. "Interpretability" aims to describe the internal structure of a system in a way understandable to humans; it is closely related to human cognition, knowledge, and biases. The main approaches to interpretability include intrinsic interpretability and post hoc interpretability. Intrinsic interpretability involves constraining the model's architecture to make its operations and intermediate results more understandable to people. On the other hand, post hoc interpretability refers to interpreting a trained model using various statistical measures, visualization techniques, and causal reasoning methods.

This study contributes to both intrinsic and post hoc interpretability. First, the study integrates DWT deeply into CNN, involving it in the training and optimization processes of CNN, achieving mutual learning and enhancement. This imposes constraints on the feature learning and structure of CNNs, enabling an understanding of their properties in the frequency domain feature learning. Specifically, based on convolution theory, the proposed model employs learnable convolutional kernels and physically meaningful wavelet filters for extracting features from vibration signals. Furthermore, wavelet transform has a rich theoretical foundation and physical significance. Incorporating it into the convolutional neural network model allows humans to analyze and understand some feature processing and learning.

Additionally, this study conducts post hoc interpretability analysis through visualization methods. The preceding sections provide clear results and compelling evidence demonstrating the mechanism of the proposed model in frequency learning. It elucidates the correlation between the feature learning of the proposed method and the physical mechanisms of actual bearing faults. This renders it trustworthy and interpretable for practical engineering applications.

VI. Conclusion

This article proposes a new design paradigm for CNN architectures, using DWT for deep frequency separation. By

integrating DWT and CNN, a WaveGNet, called WaveGNet, is proposed for MIFP, which improves feature learning from a frequency perspective. The proposed hierarchical DWT-driven FL-Layer enables deep frequency decomposition and discriminative feature learning in a coarse-to-fine mode. The theoretical foundation of DWT provides excellent interpretability analysis of WaveGNet, and its feature learning mechanism in terms of frequency. In real-world applications, WaveGNet shows excellent fault diagnosis and prediction capabilities, outperforming state-of-the-art deep learning methods. The interpretability analysis of WaveGNet and the frequency feature learning mechanism of convolutional layers demonstrate the contribution of this proposed method.

In our future work, we aim to explore the fusion design paradigm of signal analysis methods and CNN models to achieve superior performance. We will also further investigate the model's interpretability research and analyze its feature learning mechanism with the help of signal analysis methods.

REFERENCES

[1] M. Kordestani, M. Saif, M. E. Orchard, R. Razavi-Far, and K. Khorasani, "Failure prognosis and applications—A survey of recent literature," *IEEE Trans. Reliab.*, vol. 70, no. 2, pp. 728–748, Jun. 2021.

[2] P. Henriquez, J. B. Alonso, M. A. Ferrer, and C. M. Travieso, "Review of automatic fault diagnosis systems using audio and vibration signals," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 5, pp. 642–652, May 2014.

[3] D. Codetta-Raiteri and L. Portinale, "Dynamic Bayesian networks for fault detection, identification, and recovery in autonomous spacecraft," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 1, pp. 13–24, Jan. 2015.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[5] Z.-H. Liu, B.-L. Lu, H.-L. Wei, L. Chen, X.-H. Li, and M. Rätsch, "Deep adversarial domain adaptation model for bearing fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 7, pp. 4217–4226, Jul. 2021.

[6] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 1, pp. 136–144, Jan. 2019.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, 2016, pp. 770–778.

[8] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.

[9] J. Deutsch and D. He, "Using deep learning-based approach to predict remaining useful life of rotating components," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 1, pp. 11–20, Jan. 2018.

[10] T. Han, Y.-F. Li, and M. Qian, "A hybrid generalization network for intelligent fault diagnosis of rotating machinery under unseen working conditions," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, Jun. 2021.

[11] H. Wang, T. Men, and Y.-F. Li, "Transformer for high-speed train wheel wear prediction with multiplex local–global temporal fusion," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, Feb. 2022.

[12] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.

[13] D. Peng, H. Wang, Z. Liu, W. Zhang, M. J. Zuo, and J. Chen, "Multibranch and multiscale CNN for fault diagnosis of wheelset bearings under strong noise and variable load condition," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4949–4960, Jul. 2020.

[14] Z. Zhang, Y. Wang, and K. Wang, "Fault diagnosis and prognosis using wavelet packet decomposition, fourier transform and artificial neural network," *J. Intell. Manuf.*, vol. 24, pp. 1213–1227, May 2013.

[15] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Process.*, vol. 96, pp. 1–15, Mar. 2014.

[16] P. Liang, C. Deng, J. Wu, Z. Yang, J. Zhu, and Z. Zhang, "Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform," *Comput. Ind.*, vol. 113, Dec. 2019, Art. no. 103132.

[17] R. Chen, X. Huang, L. Yang, X. Xu, X. Zhang, and Y. Zhang, "Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform," *Comput. Ind.*, vol. 106, pp. 48–59, Apr. 2019.

[18] L. Gou, H. Li, H. Zheng, H. Li, and X. Pei, "Aeroengine control system sensor fault diagnosis based on CWT and CNN," *Math. Probl. Eng.*, vol. 2020, Jan. 2020, Art. no. 5357146.

[19] H. Wang, Z. Liu, Y. Ge, and D. Peng, "Self-supervised signal representation learning for machinery fault diagnosis under limited annotation data," *Knowl. Based Syst.*, vol. 239, Mar. 2022, Art. no. 107978.

[20] H. Wang, Z. Liu, D. Peng, and Z. Cheng, "Attention-guided joint learning CNN with noise robustness for bearing fault diagnosis and vibration signal denoising," *ISA Trans.*, vol. 128, pp. 470–484, Sep. 2022.

[21] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, Apr. 2021, Art. no. 107398.

[22] D. Peng, C. Liu, W. Desmet, and K. Gryllias, "Semi-supervised CNN-based SVDD anomaly detection for condition monitoring of wind turbines," in *Proc. Int. Conf. Offshore Mech. Arctic Eng.*, 2022, Art. no. 98704.

[23] H. Wang, Z. Liu, D. Peng, M. Yang, and Y. Qin, "Feature-level attention-guided multitask CNN for fault diagnosis and working conditions identification of rolling bearing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4757–4769, Sep. 2022.

[24] Y. Wei and H. Wang, "Mixed-type wafer defect pattern recognition framework based on multifaceted dynamic convolution," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, May 2022.

[25] M.-Q. Tran, M.-K. Liu, Q.-V. Tran, and T.-K. Nguyen, "Effective fault diagnosis based on wavelet and convolutional attention neural network for induction motors," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.

[26] T. Li et al., "WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 4, pp. 2302–2312, Apr. 2022.

[27] H. Wang, Z. Liu, D. Peng, and Y. Qin, "Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5735–5745, Sep. 2020.

[28] D. Zhou, Q. Yao, H. Wu, S. Ma, and H. Zhang, "Fault diagnosis of gas turbine based on partly interpretable convolutional neural networks," *Energy*, vol. 200, Jun. 2020, Art. no. 117467.

[29] J. Grezmak, P. Wang, C. Sun, and R. X. Gao, "Explainable convolutional neural network for gearbox fault diagnosis," *Procedia CIRP*, vol. 80, pp. 476–481, Jan. 2019.

[30] T. Xie, Q. Xu, C. Jiang, S. Lu, and X. Wang, "The fault frequency priors fusion deep learning framework with application to fault diagnosis of offshore wind turbines," *Renew. Energy*, vol. 202, pp. 143–153, Jan. 2023.

[31] T. Zhang et al., "Hierarchical cognize framework for the multi-fault diagnosis of the interconnected system based on domain knowledge and data fusion," *Expert Syst. Appl.*, vol. 193, May 2022, Art. no. 116503.

[32] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[33] Y. Wu and K. He, "Group normalization," *Int. J. Comput. Vis.*, vol. 128, pp. 742–755, 2019.

[34] Y. Wu et al., "Polygonal wear mechanism of high-speed wheels based on full-size wheel-rail roller test rig," *Wear*, vol. 494, Apr. 2022, Art. no. 204234.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015, pp. 1–14.

[36] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[37] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. ICLR*, 2021, pp. 1–22.

[38] A. P. Daga, A. Fasana, S. Marchesiello, and L. Garibaldi, "The Politecnico di Torino rolling bearing test rig: Description and analysis of open access data," *Mech. Syst. Signal Process.*, vol. 120, pp. 252–273, Apr. 2019.

[39] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.