Wave-ConvNeXt: An Efficient and Precise Fault Diagnosis Method for IIoT Leveraging Tailored ConvNeXt and Wavelet Transform

Liangwei Zhang, *Member, IEEE*, Jing Lin, *Senior Member, IEEE*, Zhe Yang, Haidong Shao, *Senior Member, IEEE*, Biyu Liu, and Chuan Li, *Senior Member, IEEE*

***Abstract*—The burgeoning field of the Industrial Internet of Things (IIoT) necessitates advanced fault diagnosis methods capable of navigating the dual challenges of high predictive accuracy and the constraints of edge computing environments. Our study introduces Wave-ConvNeXt, a novel fault diagnosis model that seamlessly integrates the state-of-the-art ConvNeXt architecture with Wavelet Transform. This innovative model stands out for its lightweight design yet delivers exceptional accuracy in fault diagnosis. In Wave-ConvNeXt, we re-engineer the ConvNeXt model for IIoT applications by adopting one-dimensional convolution, tailored for processing high-frequency, non-periodic inputs. This adaptation is complemented by replacing the traditional "patchify" layer with a Wavelet transform layer, which simplifies input signals into sub-signals, thereby easing learning complexities and diminishing the dependence on elaborate deep architectures. Further enhancing this model, we incorporate a squeeze-and-excitation module, enriching its ability to prioritize channel-wise feature relevance, akin to self-attention mechanisms. This integration is rigorously validated through an ablation study. Wave-ConvNeXt epitomizes a holistic approach, enabling an end-to-end optimization of feature learning and fault classification. Our empirical analysis on two real-world IIoT datasets demonstrates Wave-ConvNeXt's superiority over existing models. It not only elevates prediction accuracy but also significantly curtails computational complexity. Additionally, our exploration into the impact of various mother wavelets reveals the effectiveness of using wavelet basis functions with smaller support, bolstering diagnostic precision. The source code of Wave-ConvNeXt is available at https://github.com/leviszhang/waveConvNeXt.**

The research was supported in part by the Guangdong Basic and Applied Basic Research Foundation (2022A1515140035, 2022A1515140093), the Research start-up funds of DGUT (GC300502-46), the National Natural Science Foundation of China (52275104, 71971064), and the Dongguan Scitech Commissioner Program (20231800500112). *(Corresponding author: Jing Lin)*.

Liangwei Zhang, Zhe Yang and Chuan Li are with the Department of Industrial Engineering, Dongguan University of Technology, Dongguan, 523808, China (email: liangwei.zhang@dgut.edu.cn, yangz@dgut.edu.cn, chuanli@dgut.edu.cn)

Jing Lin is with the Division of Operation and Maintenance, Luleå University of Technology, 97187, Luleå, Sweden, and Division of Product Realization, Mälardalen University, 63220, Eskilstuna, Sweden (email: janet.lin@ltu.se)

Haidong Shao is with the State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China (email: hdshao@hnu.edu.cn)

Biyu Liu is with the School of Economics and Management, Fuzhou University, Fuzhou, 350116, China (email: jasperseu@fzu.edu.cn)

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org

***Index Terms*—Industrial Internet of Things (IIoT); Fault Diagnosis; Wavelet Transform; ConvNeXt Architecture; Computational Efficiency**

I. Introduction

As the landscape of the fourth Industrial Revolution evolves, the Industrial Internet of Things (IIoT) emerges as a pivotal player, capturing the attention of researchers and industry practitioners alike [1]. IIoT transcends mere device connectivity, ushering in a new era of industrial applications characterized by sophisticated sensing, analysis, reasoning, and control mechanisms [2]. Within this spectrum, fault diagnosis emerges as a critical component, pivotal in identifying and characterizing potential faults in industrial apparatus [3]. Its significance extends beyond technical aspects, playing a crucial role in ensuring smooth operations and efficient maintenance, thereby upholding the productivity and reliability of industrial ecosystems [4].

In the realm of IIoT, the rapid identification and accurate diagnosis of faults are paramount [5]. Traditional methods have leaned on cloud-based computing for this task, but this approach often suffers from high network bandwidth demands and latency issues [6]. An alternative, more decentralized approach is edge computing, which, despite its benefits in reducing latency and enhancing data security, grapples with challenges such as limited resources and scalability [7]. These challenges necessitate a reimagining of fault diagnosis methods, particularly to suit the nuances of edge computing environments.

The advent of deep learning models has marked a significant milestone in the evolution of fault diagnosis methodologies, thanks to their robust representational capabilities [8]. However, the computational intensity and memory demands of these models pose a substantial challenge, especially in resource-constrained edge computing scenarios [9]. Addressing this challenge has led researchers to explore avenues like model compression, architectural optimization, and data reduction [10]. Each of these strategies aims to balance the trade-off between model complexity and performance, with a keen focus on adapting to the unique demands of IIoT environments.

The model compression process begins with the initial training of a typically complex model, purposefully designed

to capture intricate patterns and information. This sophisticated model serves as the foundation for subsequent stages, where the compressed model is meticulously crafted using an array of techniques. These techniques, including knowledge distillation [11], network pruning [12], quantization [13], and low-rank factorization [14] are strategically employed to systematically reduce the size and computational demands of the model. The artful application of these methods ensures that the compressed model maintains a delicate balance, preserving its overall performance even as it undergoes a process of size and computational optimization.

Architectural optimization involves two main methods: lightweight model design through heuristics [15] and Neural Architecture Search (NAS) for automated discovery of optimal architectures [16]. Lightweight models aim to reduce computational complexity and memory requirements while preserving performance, utilizing techniques like depthwise separable convolution, channel shuffling, squeeze-and-excitation modules, and compound scaling [17]. Corresponding to these techniques, architectures like MobileNet, ShuffleNet, SqueezeNet, and EfficientNet, initially introduced in computer vision, have found applications in fault diagnosis in resource-constrained environments [18]–[21]. NAS, on the other hand, automates the exploration of a vast architecture search space to identify the most suitable design for a given task [22]. Numerous studies in the literature apply NAS to fault diagnosis tasks [23]–[25].

Unlike model compression and architectural optimization, data reduction simplifies raw data for downstream fault diagnosis models [26]. Traditional methods rely on feature engineering to extract distinctive features, compressing them for resource-limited environments [27]. However, this can be laborious and suboptimal [28], prompting the use of deep neural networks for automated feature learning. Despite their effectiveness, deep learning involves substantial computations [29]. To address this, researchers use preprocessing techniques like Fourier Transform [30], Wavelet Transform [31], and Empirical Mode Decomposition (EMD) [32] to alleviate the learning difficulty. It's crucial to note that preprocessing doesn't always reduce input size; sometimes, transforming complex data into a larger, easier-to-learn domain serves edge-computing fault diagnosis.

In fault diagnosis, accuracy often overshadows computational and storage efficiency in edge computing. The evolution of data-driven approaches, notably deep learning, e.g., Deep Belief Networks (DBN), Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNN), prioritizes accuracy [33]. Studies commonly build complex neural networks that overfit, using regularization techniques like dropout, weight decay, and early stopping for generalization [34], [35]. However, this leads to redundancy, increasing unsuitable computational and storage demands for edge deployment [36]. Meticulous design limits scalability to varied complexities [37]. The analysis above uncovers the primary motivation of this study: crafting a lightweight, precise, and scalable fault diagnosis approach for IIoT.

In this context, we introduce Wave-ConvNeXt, a novel approach that synergizes the advanced capabilities of Wavelet Transform with the architectural innovations of the ConvNeXt model [38], [39]. This model is designed to be inherently lightweight, catering to the resource limitations of edge computing, while not compromising on diagnostic accuracy. Wave-ConvNeXt redefines the ConvNeXt architecture by incorporating one-dimensional convolutions, making it apt for handling high-frequency, non-periodic inputs typical in IIoT scenarios. The integration of a squeeze-and-excitation module further refines the model's focus on relevant features, enhancing its diagnostic precision. Additionally, the use of Wavelet Transform as a preprocessing step simplifies input signal complexities, paving the way for streamlined and efficient learning processes.

Our extensive empirical analysis on two real-world IIoT datasets positions Wave-ConvNeXt as a superior alternative to existing models, delivering heightened prediction accuracy with markedly reduced computational demands. We also delve into the effects of different mother wavelets, uncovering the advantages of using wavelet basis functions with smaller support for enhancing diagnostic accuracy. This exploration not only validates the efficacy of Wave-ConvNeXt but also enriches our understanding of the impact of wavelet selection on fault diagnosis performance.

Numerous studies have already explored the application of ConvNeXt for fault diagnosis. However, many involve transforming raw data into images to align with the expected input format of the ConvNeXt model. For instance, vibration signals were converted into images using techniques such as Symmetrized Dot Pattern [40], Gramian Angular Difference Field [41], Synchro-squeezed Wavelet Transform [42], and Continuous Wavelet Transform [43]. Although effective, these techniques may lead to information loss and introduce human subjectivity during the transformation process [44].

To the best of our knowledge, this is the first endeavor in the literature to build an end-to-end fault diagnosis model by utilizing the ConvNeXt model with vibration data. Envisioned as an end-to-end solution, Wave-ConvNeXt represents a significant stride forward in fault diagnosis for IIoT. The primary contributions of this research are threefold: (1) we propose an end-to-end, accurate, lightweight fault diagnosis approach using the ConvNeXt model; (2) the effectiveness and efficiency of the proposed approach are validated on two real-world IIoT datasets; (3) the effectiveness of the Squeeze-and-Excitation module and the preference for selecting mother wavelets with smaller support are confirmed through rigorous ablation studies.

From a pragmatic perspective, the proposed model's contribution extends to various critical aspects: (1) operational efficiency: it optimizes maintenance operations, resulting in more effective resource utilization and cost reduction; (2) adaptability and scalability: lightweight models facilitate seamless deployment across diverse IIoT devices and

networks, enhancing adaptability and scalability; (3) enhanced safety: accurate fault diagnosis anticipates hazardous conditions, preventing accidents and fostering a safer working environment; (4) data management: efficient processing and analysis of vast IIoT data yield valuable insights without overwhelming network or storage systems; (5) future readiness: with industries progressing towards automation, precise fault diagnosis becomes pivotal for predictive maintenance and automated decision-making systems. In essence, the proposed model serves as a foundational element of the Industrial Large Knowledge Model (ILKM) framework, empowering superior decision-making in smart manufacturing contexts [45].

The remainder of this paper is structured as follows: Section II lays the theoretical groundwork; Section III details the Wave-ConvNeXt approach and its architectural design; Section IV introduces the IIoT datasets used for validation; Section V discusses the results and implications of our model validation; and Section VI offers concluding remarks.

II. Theoretical Preliminaries

In this section, we delve into the foundational theories underpinning our research, laying out the essential concepts and models that form the basis of our approach to fault diagnosis in the Industrial Internet of Things (IIoT). Section II-A elaborates on how the Wavelet Transform facilitates detailed analysis of signals at various scales, emphasizing its capacity for time-frequency localization and its critical role in the effective identification of signal irregularities. Section II-B delineates how ConvNeXt, a modernized convolutional neural network architecture, integrates elements from the Transformer architecture and refines conventional CNN designs.

We commence by presenting the problem statement and notations. For a given set of real-valued training samples, denoted as $\boldsymbol{X}_{\text{train}} \in \mathbb{R}^{m\times c\times l}$, where m, c, and l represent the set's cardinality, the number of channels, and the length of each sample, respectively. Correspondingly, let $\boldsymbol{y}_{\text{train}} \in \mathbb{C}^m$ be a m- dimensional vector with categorical elements, representing the labels of the training set. Our objective is to construct a model that learns a mapping function capable of accurately associating training data with their respective labels, denoted as $f: \boldsymbol{X}_{\text{train}} \rightarrow \boldsymbol{y}_{\text{train}}$. While a moderately complex model might overfit the training data, achieving a perfect projection from $\boldsymbol{X}_{\text{train}}$ to $\boldsymbol{y}_{\text{train}}$, it may suffer from poor generalization to unseen samples. To address this, we introduce a validation set $\langle \boldsymbol{X}_{\text{val}}, \boldsymbol{y}_{\text{val}} \rangle$ for model selection. The ultimate objective is to enhance the model's generalization capability for testing samples $\boldsymbol{X}_{\text{test}}$, typically assessed through prediction accuracy derived from predicted labels $\hat{\boldsymbol{y}}_{\text{test}}$ and the true labels $\boldsymbol{y}_{\text{test}}$.

A. Wavelet Transform and Multi-resolution Analysis

The Wavelet Transform is a pivotal mathematical technique in signal processing, particularly beneficial in the domain of the Industrial Internet of Things (IIoT) [46], [47]. Its ability to dissect signals into various scales using wavelet functions grants invaluable insights into the signal's temporal characteristics and frequency content. Unlike traditional methods like the Fourier Transform, which offers only frequency domain representation, the Wavelet Transform excels in time-frequency localization. This unique characteristic allows for the simultaneous identification of both time and frequency information within a signal, making it an ideal tool for pinpointing specific time and frequency regions where faults or abnormalities manifest.

At the heart of the Wavelet Transform lies the decomposition of a signal into two components: the approximation coefficients representing low-frequency parts, and the detail coefficients indicative of high-frequency components. This decomposition is achieved through the convolution of the signal with a mother wavelet function, $\psi_{a,\tau}(t)$, defined as:

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right), \quad a,\tau \in \mathbb{R}; a > 0 \qquad (1)$$

where a is the scale parameter linked to frequency, and τ is the translation parameter associated with time. These wavelet functions are compactly supported, possessing specific properties that facilitate effective time-frequency localization.

In continuous Wavelet Transform (CWT), the wavelet coefficients are obtained by convolving the mother wavelet function with the signal $x(t)$:

$$W(a,\tau) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\psi^*(\frac{t-\tau}{a})dt, \quad a,\tau \in \mathbb{R}; a > 0 \qquad (2)$$

where $\psi^*(\cdot)$ denotes the complex conjugate of $\psi(\cdot)$. While CWT offers high resolution, it also comes with high computational complexity. To mitigate this, the Discrete Wavelet Transform (DWT) discretizes the scale and translation parameters, typically using dyadic discretization ($a = 2^j$ and $\tau = k2^j$, $j,k \in \mathbb{Z}$), thus reducing redundancy. The DWT is represented as:

$$W(j,k) = \frac{1}{\sqrt{2^j}}\int_{-\infty}^{+\infty} x(t)\psi^*(\frac{t-k2^j}{2^j})dt, \quad j,k \in \mathbb{Z} \qquad (3)$$

Through DWT, the energy distribution of the signal across different frequency bands is captured, allowing for signal reconstruction using the derived coefficients.

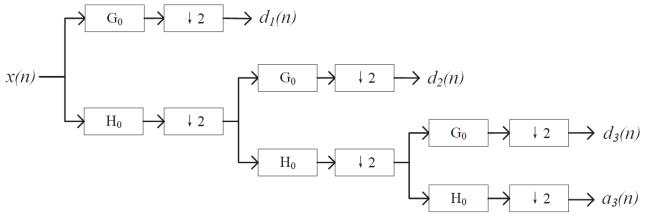


Fig. 1. Illustration of three-level multi-resolution analysis

Multi-resolution analysis (MRA) provides a structured framework for effectively implementing wavelet analysis algorithms. MRA systematically decomposes signals into various resolution levels or frequency bands through consecutive low-pass (H_0) and high-pass (G_0) filtering operations on the discrete time-domain signal (represented as $x(n)$, where n is an integer), as shown in **Fig. 1**. Each level of transformation yields detailed information $d(n)$ via the high

pass filter and coarse approximations $a(n)$ via the low pass filter. For instance, the first level of decomposition can be expressed as:

$$\begin{cases} a_1(k) = \sum_n x[n] \cdot H_0[2k-n] \\ d_1(k) = \sum_n x[n] \cdot G_0[2k-n] \end{cases} \quad (4)$$

It is important to highlight that MRA stands as a versatile framework designed for the representation and analysis of signals across multiple resolutions. This method intricately breaks down a signal into components of diverse scales or resolutions. One notable implementation of MRA is DWT, which employs wavelets to perform the decomposition and reconstruction of signals at various resolutions. Consequently, DWT emerges as a practical application situated within the expansive realm of multi-resolution analysis.

Each MRA level corresponds to a distinct resolution or frequency band, progressively capturing more refined frequency information. This capacity to detect faults from subtle changes to larger disturbances enhances fault isolation and analysis in signals. Notably, when conditions for aliasing-free and amplitude distortion elimination are met, MRA guarantees perfect signal reconstruction. From a machine learning perspective, the wavelet coefficients extracted through MRA encapsulate critical signal characteristics like transient behavior, frequency content, and energy distribution. By selectively analyzing these coefficients, we can extract meaningful, fault-related features for subsequent analysis and classification in IIoT fault diagnosis.

B. A Transformer-Inspired CNN Model: ConvNeXt

The ConvNeXt model, introduced in the paper "A ConvNet for the 2020s" by Liu et al. [38], represents a groundbreaking evolution in CNN architectures, primarily designed for computer vision tasks. Distinguished by its state-of-the-art performance across various benchmark datasets, ConvNeXt is a testament to the progressive strides in CNN development. Drawing inspiration from the acclaimed Transformer architecture renowned in Natural Language Processing (NLP), ConvNeXt aims to rejuvenate the conventional CNN structures, particularly the Residual Network, by integrating design strategies characteristic of vision transformers. The model's evolution from a standard ResNet architecture is marked by five significant innovations:

1). Macro Design Adjustments: Stemming from the VGG architecture's concept of dividing the network into blocks, ConvNeXt further refines this idea. By modifying the block ratio in ResNet-50 to 1:1:3:1, akin to the stage compute ratio in Swin Transformer, it enhances performance. For larger models, the ratio becomes 1:1:9:1. The model adopts a "patchify" strategy at its stem layer, using a 4×4, stride of 4 convolutional layer, which is non-overlapping to minimize redundancy and computational load.

2). ResNeXt-ify: In pursuit of efficiency, ConvNeXt incorporates depthwise separable convolution, a technique embraced by lightweight models like MobileNet and EfficientNet. This approach, which divides standard convolution into depthwise and point-wise convolutions, is balanced with an expanded network width, akin to the group convolution strategy in ResNeXt.

3). Inverted Bottleneck Integration: Borrowing from MobileNet-V2 and advanced ConvNet architectures, ConvNeXt adopts the inverted bottleneck design — a layout with a larger middle dimension flanked by smaller dimensions, believed to minimize information loss. The model also repositions the depthwise convolution layer to precede its layer, mirroring the Transformer architecture's multi-head self-attention block preceding multi-layer perceptron layers.

4). Adoption of Large Kernel Sizes: Challenging the norm of stacking layers with small (typically 3×3) kernel sizes, ConvNeXt opts for larger kernel sizes in its depthwise convolution operations, finding that a 7×7 kernel size optimizes accuracy without significantly impacting the network's computational load.

5). Micro Design Refinements: Inspired by design choices in Transformers, ConvNeXt replaces the traditional Rectified Linear Unit (ReLU) activation function with the smoother Gaussian Error Linear Unit (GELU). It also reduces the use of activation and normalization layers, replacing Batch Normalization (BN) with Layer Normalization (LN), and incorporates a spatial downsampling layer between stages.

By amalgamating these design elements, ConvNeXt surpasses its ResNet predecessor, achieving superior performance. Notably, while it doesn't explicitly incorporate attention-based modules, its use of depthwise convolution parallels the weight sum operation found in self-attention mechanisms. The model's simplicity and incorporation of lightweight design techniques make it a suitable candidate for developing accurate and computationally efficient fault diagnosis methods, particularly in the context of IIoT.

III. The Proposed Wave-ConvNeXt Fault Diagnosis Approach

In this section, we unfold the intricate details of our Wave-ConvNeXt model, a novel approach tailored for fault diagnosis within the Industrial Internet of Things (IIoT). Section III-A presents the modifications we implemented to adapt the ConvNeXt model for effective processing of sequential data. Section III-B delves into our novel approach of replacing the traditional "patchify" layer with a wavelet stem layer. Section III-C explores the integration of a channel-wise attention mechanism, i.e., the Squeeze-and-Excitation module. Section III-D provides the structure of the Wave-ConvNeXt model, highlighting the synergy between its various components.

A. Adaptation of the ConvNeXt Model for Sequential Input Processing

In its original form, the ConvNeXt model is tailored for image classification tasks, handling three-dimensional data inputs (channel, height, and width), as depicted in the left part of **Fig. 2**. This configuration, while effective for visual data, poses challenges when repurposed for fault diagnosis in industrial settings, particularly when dealing with sequential

data such as vibration signals. Previous applications of ConvNeXt to fault diagnosis have necessitated the transformation of raw inputs to fit the model's format. However, this additional data conversion step can lead to information loss and introduce human subjectivity, potentially impairing the model's performance.

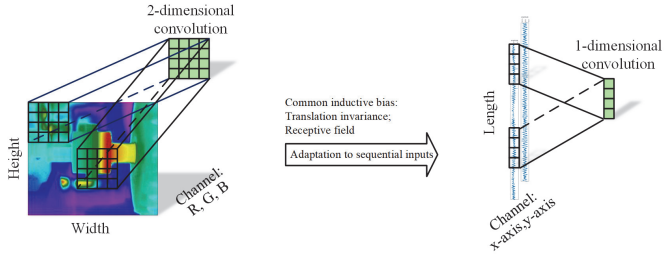


Fig. 2. Adaptation of the ConvNeXt model for sequential input processing

To overcome these limitations and harness the full potential of ConvNeXt in an industrial context, we have re-engineered the model to directly accommodate the sequential nature of common industrial data, particularly vibration signals. This adaptation primarily involves substituting the two-dimensional convolutions of the original network with one-dimensional counterparts. For a one-dimensional (single channel) real-valued vibration signal $\boldsymbol{x}$ of length l, i.e., $\{x(1), x(2), \dots, x(l)\}$, and a kernel $\boldsymbol{k} \in \mathbb{R}^{p}$, the i-th element of the output generated by a 1D convolution operation between $\boldsymbol{x}$ and $\boldsymbol{k}$ can be loosely defined as:

$$(\boldsymbol{x} \otimes \boldsymbol{k})(i) = \sum_{j=1}^{p} \boldsymbol{x}(i+j) \cdot \boldsymbol{k}(j) \qquad (5)$$

where j is a dummy variable, and $\otimes$ denotes the convolution operation. Intuitively, it represents a sliding dot product between the input signal and the kernel, commonly known as cross-correlation in Digital Signal Processing (DSP). It should be noted that this process differs from typical DSP convolution, which requires the kernel to be indexed in reverse order. In addition, the elements in the kernel are trainable parameters.

The rationale behind this adaptation is grounded in the proven efficacy of one-dimensional convolutions in extracting salient features from vibrational data, a success that can be attributed to similar underlying inductive biases shared with vision-based applications. For instance, vibrational signals, like images, exhibit translation-invariant properties. Furthermore, these signals display a receptive field pattern akin to that of images, where local receptive fields are adept at discerning high-frequency features and global receptive fields at capturing low-frequency characteristics.

This tailored approach, converting ConvNeXt into what we refer to as 1D-ConvNeXt, not only obviates the need for data conversion but also paves the way for an end-to-end, diagnostic-focused model. Such a model can optimize the feature learning process more effectively, aligning closely with the unique demands of fault diagnosis in industrial IoT settings.

B. Incorporating a Wavelet Stem Layer

The innovative concept of a "wavelet stem layer" stems from adapting and evolving the "patchify" stem layer used in vision transformers. In these transformers, the "patchify" layer segments the input image into distinct patches or local regions, which are then flattened into vector representations. These vectors serve as input tokens for the Transformer's subsequent layers, allowing the self-attention mechanism to discern relationships between different image patches. However, this approach, initially designed for image data, is not inherently aligned with the inductive bias of fault diagnosis in vibrational signals.

In the original ConvNeXt model, a "patchify" stem layer, comprising a 4×4 size with a stride of 4 convolutional layer, effectively downsamples the input by 16 times. This downsampling serves dual purposes: computational efficiency and addressing the redundancy typically found in natural images. For one-dimensional vibrational data, a similar approach would result in a fourfold reduction in input size. However, this could potentially lead to the omission of crucial low-frequency fault characteristics due to the principles of the Nyquist Sampling Theorem, potentially causing Type II errors, as these features often span longer in the temporal dimension. Type II error in statistics, also known as a false negative, occurs when a hypothesis test fails to reject a null hypothesis that is actually false [48]. In other words, it happens when the test incorrectly indicates that there is no fault when there actually is one. This error can lead to a failure to take necessary action or make a correct decision based on the IIoT measurements.

To address this challenge, we propose substituting the "patchify" stem layer with a wavelet stem layer. This layer employs one-dimensional discrete wavelet transform to iteratively decompose raw vibrational inputs into multiple sub-signals, see **Fig. 1** for a visual illustration. Unlike traditional wavelet transforms that decompose only the detail coefficients, our approach, akin to wavelet packet decomposition, also decomposes the approximation coefficients. Consequently, for an n-level decomposition, we obtain 2^n sets of coefficients. These coefficients are concatenated along the channel dimension for subsequent feature extraction and fault classification tasks.

Notably, the wavelet transform is reversible, ensuring that our wavelet stem layer preserves all information from the input. In other words, this decomposition process enables perfect reconstruction of the input signal without any information loss. The rate of input length reduction depends on the chosen level of decomposition, a parameter influenced by both signal length and hardware throughput. A higher level of decomposition offers greater computational savings, while a lower level increases computational demand. Yet, an excessively high level of decomposition can impede the learning of temporal dependencies within the signal using the ConvNeXt block, thus affecting the effectiveness of our Wave-ConvNeXt model. Adhering to the principle of quadruple reduction in the original "patchify" layer, we apply

two levels of decomposition, resulting in an output comprising four channels per input channel, with its length quartered.

The wavelet stem layer functions as a preprocessing step, devoid of learnable parameters. However, it necessitates careful design choices, such as the selection of decomposition levels and mother wavelet types. As discussed earlier, this layer simplifies raw data, easing the learning process for downstream fault diagnosis tasks. Moreover, it contributes to reducing the model's complexity and dependence on computational resources, aligning with our goal of creating an efficient and effective diagnostic tool for IIoT applications.

C. Channel-wise Attention in the Wave-ConvNeXt model

The ConvNeXt model, as discussed in Section II-B, compensates for the potential capacity loss inherent in depthwise convolution by broadening its network width. However, this increase in width can make the coordination and information exchange between channels more complex. In any feature map, not every channel contributes equally to the final representation — some channels carry more discriminative information, while others might be less informative or even introduce noise. To address this, we propose the integration of a channel-wise attention module following each depthwise convolution layer within our model.

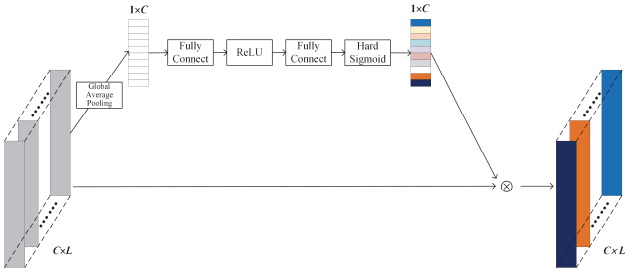


Fig. 3. Channel-wise Attention: the one-dimensional Squeeze-and-Excitation module

Channel-wise attention mechanisms typically employ a gating mechanism that dynamically assigns attention weights to each channel [49], [50]. This approach allows the model to prioritize the most informative channels and diminish the influence of less relevant ones, thereby enhancing its discriminative capability and robustness. The attention weights are computed by analysing channel-wise statistics, such as the mean and standard deviation, or through learned transformations of the feature maps. These weights are then applied to modulate the feature maps, either amplifying or attenuating the channel activations before they proceed to subsequent layers.

We have chosen to implement the Squeeze-and-Excitation module for channel-wise attention, as illustrated in **Fig. 3**. Here, a feature map $\boldsymbol{X}$ of size $C \times L$ (number of channels by length) undergoes a squeezing process through a 1D Global Average Pooling (GAP) operation, reducing it to a vector $\boldsymbol{z}$ of length C, i.e.,

$$\boldsymbol{z}_i = \frac{1}{L}\sum_{j=1}^{L} \boldsymbol{X}_{ij}\,, \quad i \in \{1,2,\dots,C\} \tag{6}$$

where $\boldsymbol{z}_i$ is the value in the i-th squeezed channel. This squeezed vector $\boldsymbol{z}$ is then passed through two fully connected layers: the first employs a ReLU activation function, while the second uses a hard sigmoid activation to compute attention weights vector $\boldsymbol{s}$:

$$\boldsymbol{s} = \sigma_{\text{hard}}\big(\boldsymbol{W_2} \cdot \text{ReLU}(\boldsymbol{W_1} \cdot \boldsymbol{z})\big) \tag{7}$$

where $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are weight matrices of the two fully connected layers, and $\sigma_{\text{hard}}(\cdot)$ is the hard sigmoid function, as defined below:

$$\sigma_{\text{hard}}(x) = \begin{cases} 0 & \text{if } x \le -3 \\ 1 & \text{if } x \ge 3 \\ x/6 + 1/2 & \text{otherwise} \end{cases} \tag{8}$$

The choice of hard sigmoid activation is made to streamline computations. The resulting C-dimensional excitation weights $\boldsymbol{s}$ are subsequently multiplied with the original inputs to yield a reweighted feature map.

$$\text{SE}(\boldsymbol{X}_i) = s_i \cdot \boldsymbol{X}_i, \qquad i \in \{1,2,\dots,C\} \tag{9}$$

GAP proves advantageous in computing channel-wise weights as it conducts temporal information reduction by averaging each feature map across all temporal locations. This yields a singular value per channel, effectively summarizing temporal information. Additionally, it introduces a form of translation invariance, rendering the model less sensitive to the precise feature location. This characteristic is pivotal in scenarios where the position of a point in a time series is less crucial than its mere presence. The averaged value for each channel essentially signifies the importance or activation strength of that channel across the entire temporal domain. This consolidated information enables the emphasis or suppression of specific channels, enabling the network to concentrate on the most informative channels during the learning process. In our Wave-ConvNeXt model, this channel-wise attention mechanism is a critical component, incorporated after every depthwise convolution layer to enhance the model's focus and efficacy.

D. Architectural Design of the Wave-ConvNeXt Model

The Wave-ConvNeXt model is the culmination of integrating various innovative components, as discussed in previous sections, into a cohesive architectural design. **Fig. 4** presents a detailed view of this architecture, revealing a multi-stage design where each stage comprises multiple Wave-ConvNeXt blocks. Following the design tradition outlined in Section II-B, the ratio of these blocks is set to 1:1:9:1. This ratio can be adjusted to suit the complexity of the specific diagnostic task at hand. For illustrative purposes, we demonstrate the dimensional changes of a sample input of size 1×5000, progressing through the model to an output of size 1×6.

At its core, each Wave-ConvNeXt block consists of several key elements: a depthwise convolution, a squeeze-and-excitation layer, two pointwise convolutions, and a skip connection. An additional layer scale operation, which is optional, is used to reweight its input. Additionally, the model incorporates a drop path layer — a regularization technique that randomly zeroes out some inputs. The rate of drop path is

adjustable and inversely proportional to the complexity of the model — a higher rate leads to greater regularization. Another critical component is the downsampling layer, which not only reduces the size of its input but also doubles the number of feature maps. This layer is strategically placed at the beginning of the last three stages of the model.

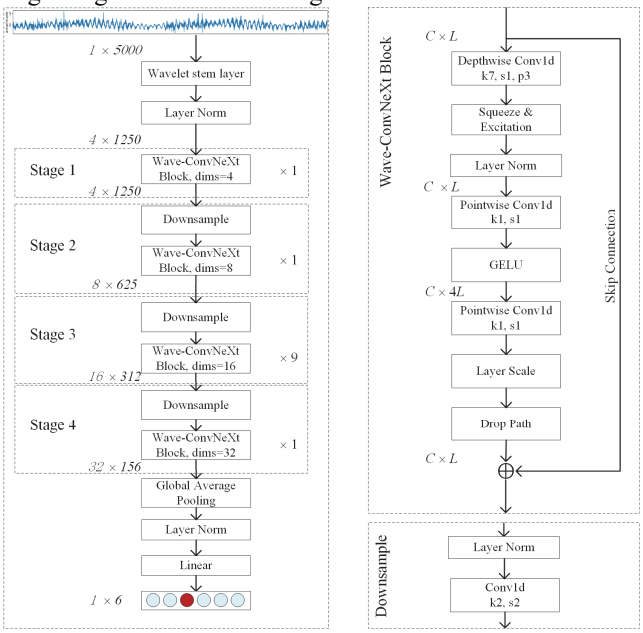


Fig. 4. Architectural design of the Wave-ConvNeXt model, where $k, s, and\ p$ denote the parameters for kernel size, stride, and padding, respectively.

Each of these components plays a vital role in enhancing the model's performance. The depthwise convolution and squeeze-and-excitation layers focus on extracting and emphasizing informative features from the input. The pointwise convolutions and skip connections aid in preserving the integrity of the signal throughout the network, while the layer scale and drop path layers contribute to the model's adaptability and robustness. The downsampling layer ensures that the model remains computationally efficient, even as it processes complex data. Our Wave-ConvNeXt model represents a significant advancement in fault diagnosis for IIoT, combining efficiency with accuracy. The model's code is openly accessible for the community, available on GitHub at: https://github.com/leviszhang/waveConvNeXt.

IV. Experimental Setup and Data Description

This section presents a comprehensive overview of the experimental framework and data characteristics central to validating the efficacy of the Wave-ConvNeXt model in fault diagnosis within the Industrial Internet of Things (IIoT). Section IV-A delves into the first case study involving a wind turbine gearbox. This section describes the experimental setup, the simulated fault conditions, and the methodology for data collection and processing, providing a nuanced understanding of the complexities involved in diagnosing gearbox faults. Section IV-B demonstrates the second case study, focusing on a dual-bearing dataset from an Automatic Washing Equipment. This section outlines the experimental procedures used to simulate various fault conditions, the data acquisition process, and the challenges faced in signal analysis for accurate fault diagnosis.

A. Case I: Wind Turbine Gearbox Dataset Analysis

In this section, we present our first case study, a detailed analysis of a gearbox dataset from a wind turbine test bed. This dataset simulates various fault conditions within a wind turbine gearbox, offering a rich source of data for evaluating our Wave-ConvNeXt model. As illustrated in **Fig. 5**, the test bed comprises a RiChuan Vertical-Axis wind power generator (RCVA-3000), with a 7.5kW axial-flow fan and a frequency converter to simulate different wind speeds. This setup allows for the replication of diverse operational conditions by varying wind speeds (20 Hz to 50 Hz) and applying three distinct levels of working load (low, medium, and high) on the wind turbine's rotor. Any stochastic vibration or speed perturbation on the input shaft can be transmitted through the gearbox to the output shaft, resulting in different fault characteristics of the gearbox.

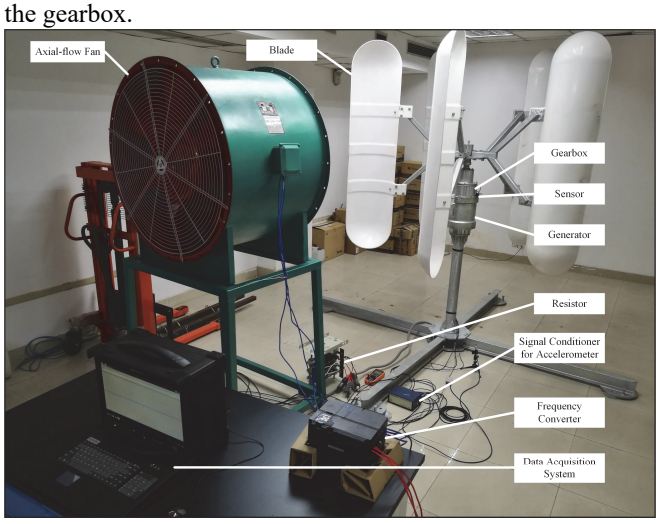


Fig. 5. Experimental setup of the wind turbine test bed

Our experimental setup utilizes a one-stage planetary gearbox with a specific configuration of gears (one carrier, one sun gear, one ring gear, and three planet gears). We induce two types of faults – broken tooth and cracked tooth root – into these gears, generating six unique health status classes for the gearbox, including a normal state (C0-C5). **Fig. 6** illustrates the five faulty states, corresponding to the first five classes in TABLE I.

We conduct independent trials for each of these six classes at three different working load levels, i.e., 18 trials in total. In each trial, we collect acceleration signals over 20 seconds. During this timeframe, the axial-flow fan undergoes the following speed regulation processes: 1) it operates at 20 Hz for 2 seconds; 2) it gradually ramps up to 50 Hz over 6 seconds; 3) it maintains a steady 50 Hz for 4 seconds; 4) it gradually ramps down to 20 Hz over 6 seconds; 5) it operates at 20 Hz for 2 seconds. Consequently, each trial generates an acceleration signal of length 2,000,000 (20 seconds at a

sampling frequency of 100 kHz). A concise summary of these working conditions is presented in TABLE I.

The acceleration signals from the 18 trials are sliced using the moving window method for data augmentation. The window length and step size are set to 5000 and 4000, respectively. This method not only generates an adequate number of training samples but also makes sure that each sample is associated with a certain speed mode. With these settings, the data augmentation technique produces 499 signal segmentations of length 5000 for each signal. The experiments are repeated four times: twice for training, once for validation, and once for testing. This results in 17,964 (2×18×499) training samples, 8,982 (18×499) validation samples, and 8,982 (18×499) testing samples.

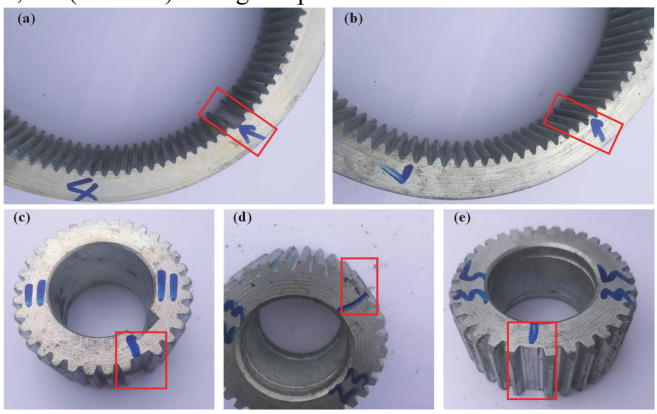


Fig. 6. The planetary gearbox is injected with five faulty states, which are: (a) ring gear with a broken tooth; (b) ring gear with a cracked tooth root; (c) sun gear with a broken tooth; (d) planet gear with a cracked tooth root; and (e) planet gear with a broken tooth. These states correspond to the first five classes of health status in TABLE I, i.e., C0-C4, respectively

TABLE I
PLANETARY GEARBOX HEALTH STATUS CLASSES AND WORKING CONDITIONS

Class	Fault location	Fault type	Working Condition
C0	Ring gear	Broken tooth	Load: High/ Medium/ Low. Speed regulation (axial-flow fan): 1) low speed for 2s; 2) acceleration for 6s; 3) high speed for 4s; 4) deceleration for 4s; 5) low speed for 2s.
C1	Ring gear	Cracked tooth root (width: 0.5 mm, depth: 0.3mm)	
C2	Sun gear	Broken tooth	
C3	Planet gear	Cracked tooth root (width: 0.5 mm, depth: 0.3mm)	
C4	Planet gear	Broken tooth	
C5	N/A (normal)	N/A (normal)	

To ensure uniformity across the dataset, we apply z-score normalization to the acceleration signals. A selection of these normalized signals, depicted in **Fig. 7**, shows no obvious visual distinction between the different fault states, underscoring the complexity and subtlety of the fault characteristics in the dataset. Further investigation into their frequency spectrums does not provide any additional insights either. This complexity is further compounded by the non-stationary conditions of the test bed, posing additional challenges in accurately annotating the testing samples. This case study provides a comprehensive and challenging dataset for validating the effectiveness of the Wave-ConvNeXt model in fault diagnosis, particularly in the complex and dynamic environment of a wind turbine gearbox.

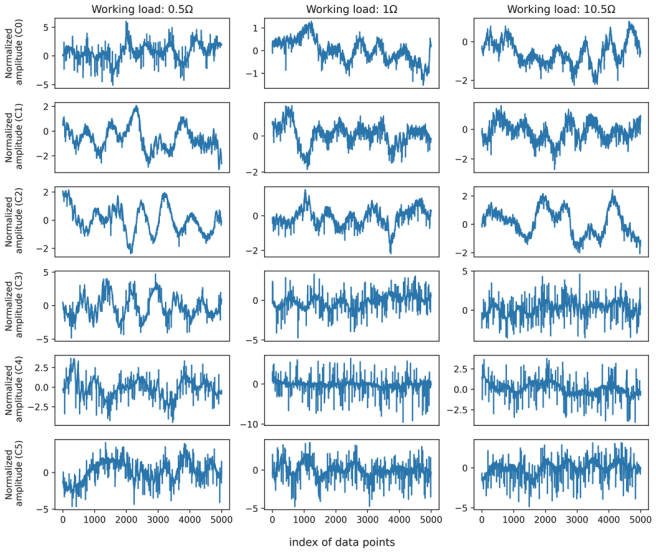


Fig. 7. Randomly selected samples from the testing set of Case I, showcasing the six health status classes under different working loads

B. Case II: Automatic Washing Equipment's Dual-bearing Dataset Analysis

Our second case study delves into a real-world dual-bearing dataset, open-sourced by the NGIT Laboratory of Beijing Jiaotong University, China [51], [52]. The dataset focuses on a dual-bearing component from an Automatic Washing Equipment (AWE), as depicted in **Fig. 8**, used for cleaning high-speed trains. The dual-bearing, connecting a vertical rotating shaft supporting the brush set, is subject to horizontal load and rotational force, making it susceptible to eccentricity issues. The purpose of this experiment is to diagnose to what extent the eccentricity level is in the dual-bearing component.

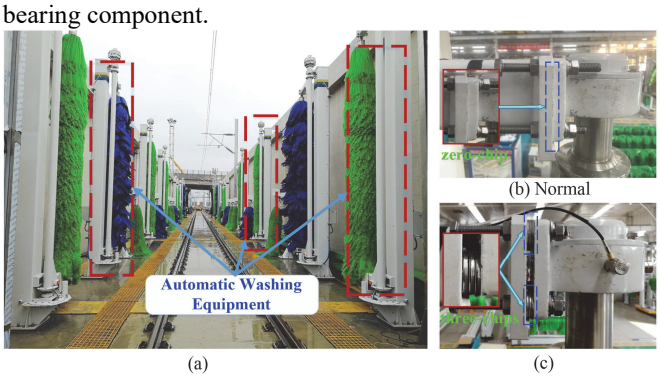


Fig. 8. (a) Overview of the Automatic Washing Equipment (AWE). (b) Normal condition setup. (c) Setup illustrating the C3 class with added chips [51], [52]

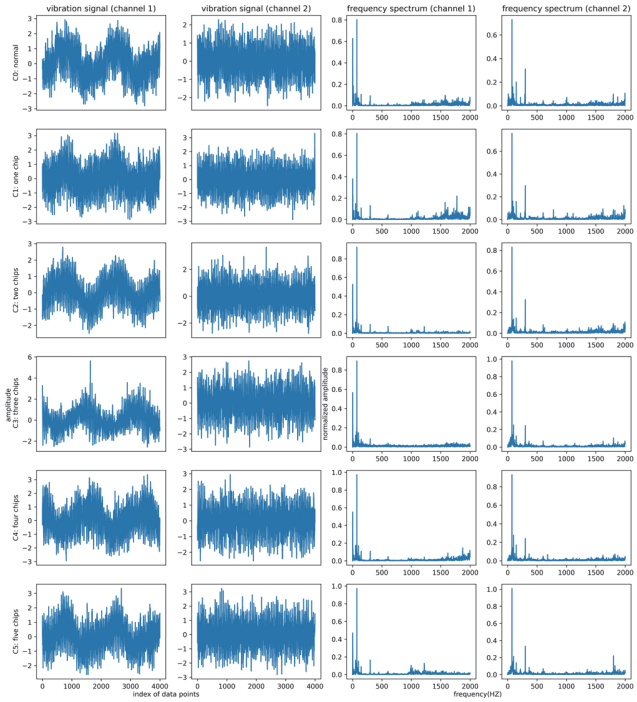


Fig. 9. Sample signals from the testing set, showing the dual-bearing component's different eccentricity levels

In this study, the eccentricity level of the dual-bearing component is experimentally varied to simulate different fault conditions. These variations were discretized into six distinct states, including the normal condition, by adding varying numbers of gaskets to the fastening screws of the upper bearing, as shown in **Fig. 8** (b) and (c). The maximum number and size of the gaskets were carefully chosen to prevent significant equipment damage. These six states of health conditions for the dual bearings are summarized in TABLE II. Two single-axis accelerometers were placed on the upper and lower bearings to record vibration signals at a sampling frequency of 4 kHz.

TABLE II
HEALTH STATUS CLASSES OF THE DUAL-BEARINGS IN THE AWE EXPERIMENT

Class	Fault location	Fault type
C0	N/A (normal)	N/A (normal)
C1	Fastening screws that attach the upper bearing to its supporting arm	1 chip added
C2		2 chips added
C3		3 chips added
C4		4 chips added
C5		5 chips added

For each of the six classes, seven independent experimental trials were conducted, each lasting 10 minutes. This generated a collection of $6\times7\times600\times4000\times2$ (classes, trials, sampling time in seconds, sampling frequency, and number of channels, respectively) digital numbers, which were then divided into training, validation, and testing sets. **Fig. 9** showcases time-domain vibration signals from these trials, illustrating the complexity of signal segmentation and the challenge in signal differentiation based on the Nyquist Sampling Theorem. Hence, the length of each signal segment was maintained at 4000. With these settings, the training set, validation set, and testing set are of size $14400\times2\times4000$, $7200\times2\times4000$, and $3600\times2\times4000$, respectively. Consistent with the first case study, we standardized all samples using the z-score normalization method.

While previous studies often utilized various sensor data, this study focuses solely on two-channel vibration signals to maintain consistency for comparison. The experimentally curated dataset provides a nuanced spectrum of concentricity deviations, offering a challenging yet informative platform for evaluating our model's performance. From **Fig. 9**, certain class characteristics are visible in their frequency spectrums, yet discerning each class remains challenging. This study also considers the fusion of two channels of vibration signals, which could provide additional insights for the diagnostic task.

V. RESULTS AND DISCUSSION

This section offers a thorough analysis and discussion of the results obtained from the deployment of the Wave-ConvNeXt model across our two case studies. Section V-A delves into the strategies and technical choices made during the training process of the Wave-ConvNeXt model. Section V-B presents a comparative analysis of the Wave-ConvNeXt model against other state-of-the-art deep learning models, highlighting its superior accuracy and computational efficiency. In Section V-C, we explore the features learned by the model and investigate specific instances of misclassification to understand the model's capabilities and limitations better. Section V-D provides an analysis of the individual components of the Wave-ConvNeXt model, particularly focusing on the squeeze-and-excitation module and the choice of mother wavelets, and their respective contributions to the model's performance.

A. Model Training and Validation for Wave-ConvNeXt

Training a deep neural network, particularly for complex high-frequency non-periodic signal data like ours, presents distinct challenges. To navigate these, we adopted several key strategies and techniques from the deep learning community. For efficient and effective model training, we employed Automatic Mixed Precision (AMP) and learning rate annealing. AMP combines single and half-precision representations to expedite training while minimizing memory requirements, all without sacrificing the accuracy typically achieved with single precision. Learning rate annealing, facilitated by the "ReduceLROnPlateau" scheduler, dynamically adjusts the learning rate to mitigate issues such as instability and overshooting, ensuring optimal learning rate throughout the training process.

Further, we made strategic engineering choices regarding the training parameters, including the number of epochs, batch size, initial learning rate, loss function, optimizer type, and drop path rate (detailed in TABLE III). These choices significantly impact the training process, influencing convergence speed, stability, and constringency. A trial-and-error approach helped fine-tune these parameters, particularly

focusing on minimizing average loss on the validation set. Regularization techniques such as label smoothing, weight decay, and drop path were implemented to prevent overfitting, with their coefficients tuned using a grid search method. All model development and training were conducted using the Pytorch framework.

TABLE III
ENGINEERING CHOICES IN MODEL TRAINING

Name	Case I	Case II
Number of epochs	200	30
Batch size	1200	500
Initial learning rate	0.01	0.001
Loss function	Cross entropy loss with a label smoothing of 0.2	Cross entropy loss with a label smoothing of 0.2
Optimizer	"AdamW" with a weight decay of 0.001	"AdamW" with a weight decay of 0.001
Drop path rate	0.2	0.2

The learning curves of our Wave-ConvNeXt model and the vanilla 1D-ConvNeXt model on the wind turbine gearbox dataset are depicted in **Fig. 10**. Both models exhibit a plateau in later training stages, yet our Wave-ConvNeXt model converges faster, around the 140th epoch. This accelerated convergence is likely due to the wavelet stem layer, which simplifies raw signals into sub-signals, aiding in quicker and more effective learning. Notably, while both models achieve training accuracies higher than 99%, indicating ample representation capabilities, our model demonstrates a smaller generalization gap. This indicates superior accuracy in predicting unseen data. This pattern of learning curve and performance is similarly observed in the automatic washing equipment dataset.

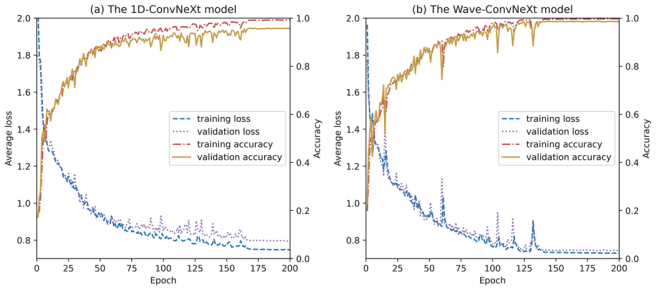


Fig. 10. Learning curve of (a) the vanilla 1D-ConvNeXt model, and (b) the Wave-ConvNeXt model on the wind turbine gearbox dataset

B. Model Comparison and Evaluation on Testing Sets

In our effort to validate the effectiveness and efficiency of the proposed Wave-ConvNeXt model, we compared it against several state-of-the-art deep learning models, including Bi-LSTM, 1D-CNN, and variants incorporating Empirical Mode Decomposition (EMD) as a preprocessing step [32], [53]. TABLE IV presents a comparative analysis of these models in terms of prediction accuracy on the wind turbine dataset's testing set, along with the number of trainable parameters and total multiply-add operations for a standard input sample of size $1\times1\times5000$.

Our Wave-ConvNeXt model achieved the highest prediction accuracy at 98.64%, outperforming the other models, including the vanilla 1D-ConvNeXt. An interesting observation from the results is the beneficial impact of preprocessing layers, like EMD and wavelet transform, in boosting diagnostic accuracy. Moreover, the ConvNeXt-based models, particularly our Wave-ConvNeXt, demonstrated significantly fewer trainable parameters and multiply-add operations than their counterparts. This reduction in computational complexity makes them more suitable for edge-computing solutions and reflects the efficiency gains from depthwise separable convolutions and group convolutions used in these models. The addition of Squeeze-and-Excitation blocks in the Wave-ConvNeXt model slightly increases the number of trainable parameters compared to the 1D-ConvNeXt model. However, this increase is negligible when considering the significant improvement in testing accuracy.

We also conducted a thorough comparison between our model and a cutting-edge model designed for long-term time series forecasting tasks known as PatchTST [54]. This chosen model, PatchTST, exhibits similarities with our approach through its utilization of channel-independent patching operations and explicit incorporation of transformer backbones. To tailor the model to our specific case study, we adjusted both the input (set to 5000) and prediction sequence lengths (set to 6). The original paper introducing PatchTST proposes two model variants, PatchTST-64 and PatchTST-42, differing in the number of patches (64 and 42, respectively). The authors achieve this variation by selecting appropriate patch lengths and stride values. In our comparative study, we similarly redesigned these parameters to achieve the desired number of patches. Specifically, PatchTST-64 was configured with a patch length of 156 and a stride of 78, while PatchTST-42 adopted a patch length of 236 and a stride of 118. Notably, the dimension of the fully connected layer was reduced to 64 to accommodate the decreased output dimension.

TABLE IV
COMPARISON OF MODELS ON THE TESTING SET OF THE WIND TURBINE DATASET

Model	Testing set accuracy	Number of trainable parameters	Total Mult-adds (M)
Bi-LSTM	94.01%	970182	8.75
EMD-Bi-LSTM	94.6%		
1D-CNN	94.2%	145742	2.94
EMD-1DCNN	96.93%		
PatchTST-64	74.63%	2547334	2.51
PatchTST-42	75.28%	2509446	2.49
1D-ConvNeXt	95.2%	32018	0.83
Wave-ConvNeXt	98.64%	35382	0.83

During the training of the PatchTST model, notable success was achieved in terms of high training accuracy, indicative of the model's substantial representational capacity. However, despite employing diverse regularization techniques, a significant generalization gap persisted, resulting in notably low accuracy on the testing set. TABLE IV presents the best testing accuracies for both variants, showcasing a substantial

shortfall compared to alternative models, despite their low complexity in terms of the total Mult-adds. This observation underscores the challenge of handling high-frequency non-periodic signals, emphasizing a preference for signal processing methods like wavelet transforms over direct segmentation for preprocessing purposes.

Fig. 11 showcases the confusion matrices for the six models when tested on the wind turbine gearbox dataset. The testing set has an equal number of samples for each target class, which are 1497 (three levels of working load times 499 signal segmentation). The true positives, represented by the darker diagonal elements, indicate the model's accuracy in correctly classifying samples into their respective classes. The off-diagonal elements reveal the classification errors, with higher values indicating more misclassifications.

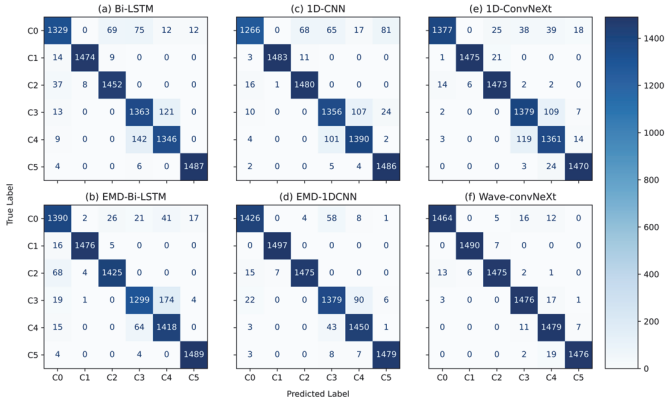


Fig. 11. Confusion matrices for different models on the wind turbine gearbox dataset's testing set

To extend our evaluation, we compared the Wave-ConvNeXt model with several traditional lightweight models in the second case study involving the automatic washing equipment dataset. These models, adapted for vibrational inputs through one-dimensional convolutions, include SqueezeNet, ShuffleNet, MobileNet V3, EfficientNet V2, and 1D-ConvNeXt. As seen in TABLE V, while all models exhibit high testing accuracy, they differ significantly in efficiency metrics such as the number of trainable parameters, total multiply-add operations, and inference time. Note the inference time was measured using a single instance of size 1×2×4000. These results, derived from an average of 300 independent runs on a specific hardware setup (Intel Core i7-9750H CPU and an Nvidia RTX-3000 GPU), highlight the superior balance of accuracy and efficiency offered by our model.

We compared our model with the 1D-ConvNeXt model, which, despite having fewer trainable parameters, surpasses our model in total Mult-adds operations. The efficiency gain is attributed to replacing the patchify stem layer with the wavelet stem layer, leading to significant computational savings that offset the additional cost of the Squeeze-and-Excitation layer. Unlike the prior single-channel input case study, the 1D-ConvNeXt model's patchify stem layer necessitates increased computation for the two-channel input in the subsequent case study. Notably, the 1D-ConvNeXt model exhibits a shorter inference time on CPU compared to GPU, suggesting that the CPU's efficiency outperforms the GPU in handling data I/O.

TABLE V
COMPARISON OF LIGHTWEIGHT MODELS ON THE AUTOMATIC WASHING EQUIPMENT DATASET

Model	Testing set accuracy	Number of trainable parameters	Total Mult-adds (M)	Inference time on GPU/CPU (ms)
SqueezeNet	100%	362278	340.2	7.21/36.75
ShuffleNet	99.97%	341246	50.33	13.47/26.99
MobileNet V3	99.97%	1477041	67.16	11.3/36.36
EfficientNet V2	100%	19445470	2600	40.47/264.03
1D-ConvNeXt	100%	32598	0.79	11.94/11.3
Wave-ConvNeXt	100%	36062	0.75	10.86/11.35

C. Analysing the Learned Features and Misclassifications

To gain insight into what our deep learning models have learned, we analyzed the activations from the second-to-last layer of the trained models. We used t-distributed Stochastic Neighbor Embedding (t-SNE) to map these activations into a two-dimensional space. This technique is effective in preserving the intricate structures of high-dimensional data. **Fig. 12** illustrates the two-dimensional embeddings of the testing set as evaluated by the six models, with different colors and markers representing the health status classes of the wind turbine gearbox. A clear separation of the classes in this space indicates a model's ability to distinguish between them. The subplots in **Fig. 12** reveal a progressive separation from (a) to (f), aligning with the testing set evaluation results and suggesting that our proposed model learns more discriminative features for fault diagnosis.

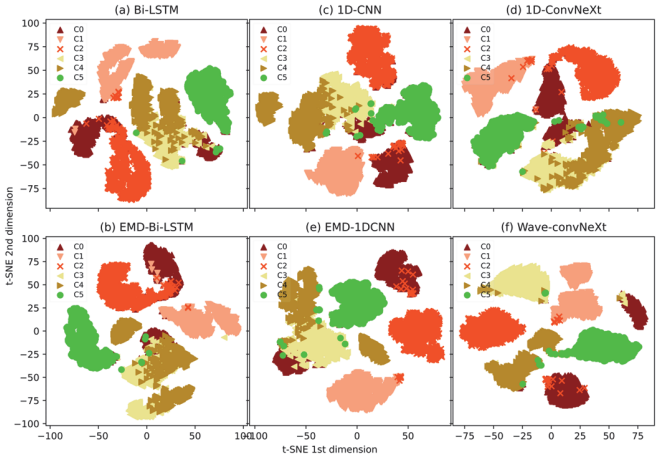


Fig. 12. Two-dimensional embeddings of the penultimate layer activations for different models on the wind turbine gearbox dataset

Despite the Wave-ConvNeXt model's superior accuracy, as shown in TABLE IV and **Fig. 11**, it still encountered 122 misclassifications. To understand these errors, we conducted an in-depth analysis. We randomly selected a misclassified sample (index 3763) from class "C2" that was incorrectly labeled as "C0" under a high-speed and medium-load condition. For comparison, we also chose two other samples under similar conditions: one correctly classified as "C2"

(index 3704) and another correctly labeled as "C0" (index 706).

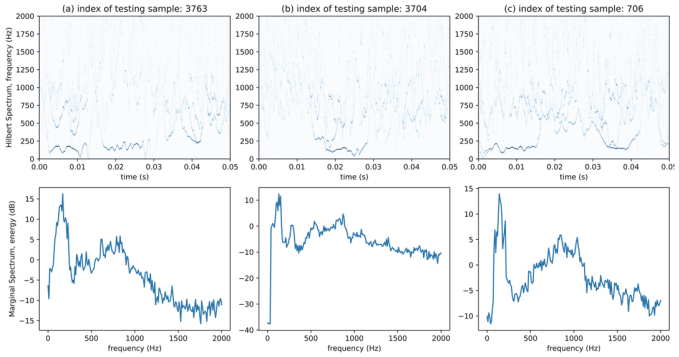


Fig. 13. Hilbert spectrum and marginal spectrum analysis of three testing samples (indices 3763, 3704, and 706)

Considering the nonstationary nature of the wind turbine gearbox dataset, we applied Hilbert spectral analysis to examine the time-frequency domain of these signals. **Fig. 13** displays the Hilbert and marginal spectrums of the selected samples. A Hilbert spectrum represents the amplitude and frequency components of a signal over time, while a marginal spectrum, derived by integrating the Hilbert spectrum over time, shows the total energy distribution across frequency components. Intriguingly, the marginal spectrums of samples 3763 and 706 share similarities in energy scale and patterns, differing from sample 3704. Notably, sample 3763 exhibits characteristics of both 3704 and 706 in the low-frequency band as shown in their Hilbert spectrums. This observation leads us to speculate that the misclassification could be attributed to the Heisenberg uncertainty principle, which limits the simultaneous compactness of continuous-time signals in both time and frequency domains. This principle presents a fundamental challenge in fault diagnosis using time-varying signals and warrants further investigation.

D. Ablation Study of the Wave-ConvNeXt Model

We now delve into an ablation study to discern the impact of various components on the performance of our Wave-ConvNeXt model. First, we examine the role of the squeeze-and-excitation module. We randomly selected six testing samples from the AWE dataset, each representing a distinct health status, and analyzed the activations from the squeeze-and-excitation module of a trained Wave-ConvNeXt model. Specifically, we focused on the re-weighting activations of 32 channels in stage 4 of the model, as illustrated in **Fig. 14**. The varying shades in the figure indicate the weightage of each channel, with darker colors representing higher weights. A significant observation from **Fig. 14** is the prevalence of zero weights for certain classes, particularly "C0" and "C2", indicating the effective suppression of less critical channels by the channel-wise self-attention mechanism. When the squeeze-and-excitation module was removed from the model, the testing accuracy slightly decreased from 100% to 99.94%, underscoring its critical role in the model's performance.

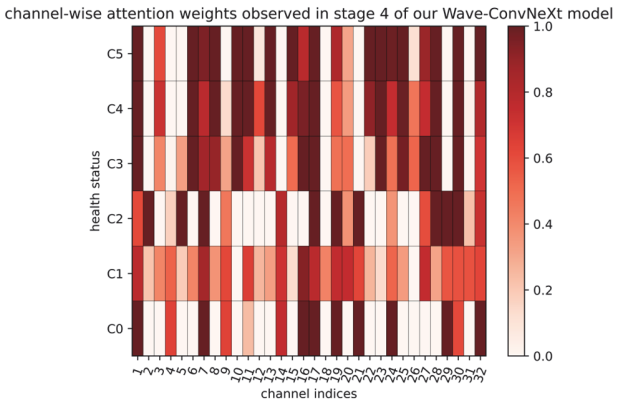


Fig. 14. Channel-wise attention weights in stage 4 for six distinct health statuses from the AWE dataset

Next, we explored the influence of different types of mother wavelets on our model's performance, using the wind turbine gearbox dataset as a case study. We compared the model's effectiveness with various mother wavelets from seven distinct families: Haar, Daubechies, Symlets, Coiflets, Biorthogonal, Reverse Biorthogonal, and Discrete Meyer, particularly focusing on wavelets with smaller support to capture high-frequency features effectively. The selected wavelets were "haar", "db2", "sym2", "bior1.3", "rbio1.3", "coif1", and "dmey". The testing accuracies for these wavelets were 98.64%, 97.09%, 97.26%, 93.32%, 96.35%, 92.71%, and 90.98%, respectively, as shown in **Fig. 15**.

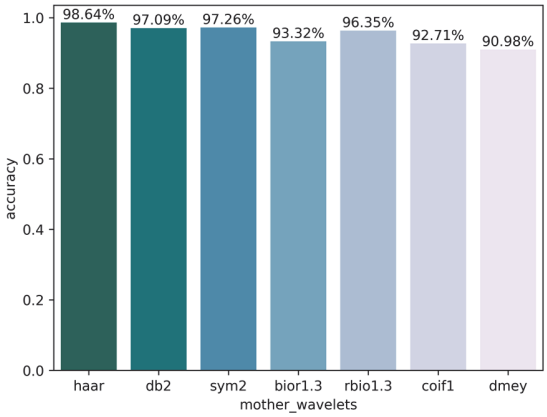


Fig. 15. Testing set accuracy comparison for various mother wavelets in the wind turbine gearbox's dataset

The choice of mother wavelet significantly influenced the model's accuracy, with higher accuracy observed for wavelets with smaller support. This finding corroborates our initial hypothesis favoring such wavelets, suggesting that wavelets with larger support might overly preprocess low-frequency signals, hindering the learning of these features via layer-by-layer abstraction in the Wave-ConvNeXt backbone. Interestingly, despite variations in testing accuracy, all mother wavelets achieved high training set accuracies, surpassing 98%, indicating the robust representational capacity of our model. However, the generalization gap observed in some wavelets with lower testing accuracies suggests a potential need for further hyperparameter tuning.

VI. Conclusion

This research presents Wave-ConvNeXt, an efficient and precise approach to fault diagnosis in the rapidly evolving domain of the Industrial Internet of Things (IIoT). Addressing two primary challenges in IIoT fault diagnosis — the need for timely and accurate detection and the constraints of edge computing environments — we successfully harness and adapt deep learning models to meet these demands. Wave-ConvNeXt, a synergy of Wavelet Transform and ConvNeXt, exemplifies a lightweight yet scalable solution, effectively balancing the computational and memory limitations of edge-computing devices.

The innovative architecture of Wave-ConvNeXt integrates the ConvNeXt model, a contemporary convolutional neural network (CNN) that merges convolutional operations with attention mechanisms, and tailors it to process high-frequency, non-periodic inputs. This is achieved by replacing traditional two-dimensional convolutions with one-dimensional convolutions. The inclusion of the squeeze-and-excitation module further refines the model, sharpening its focus on the most informative features. Additionally, the utilization of Wavelet Transform as a preprocessing step simplifies input signals into more manageable sub-signals, thereby reducing the dependence on complex deep learning architectures.

Our extensive experiments with two real-world IIoT datasets have demonstrated the effectiveness and efficiency of the Wave-ConvNeXt approach. By employing an end-to-end architecture, our approach minimizes information loss and curtails human subjectivity in the feature learning process. The results from these experiments, including model training and validation, comparative analyses with state-of-the-art models, and an in-depth ablation study, affirm the superior accuracy and computational efficiency of Wave-ConvNeXt.

The two case studies primarily focus on categorizing established faults acquired through fault injection experiments. Conducting such destructive experiments is typically expensive and may not be viable in certain contexts. Besides, simulating all potential compound faults is impractical due to the issue of combinatorial explosion. In real-world scenarios, detecting the onset of unknown faults, referred to as fault detection, can be even more crucial than fault diagnosis. When faulty data are available, fault detection transforms into a binary classification task. Our Wave-ConvNeXt model excels in addressing such challenges, but adapting it to situations lacking faulty data requires further investigation. We posit that integrating the concept of One-class classification with the architecture of the Wave-ConvNeXt model is a promising avenue for tackling such issues.

In conclusion, Wave-ConvNeXt stands as a promising solution for fault diagnosis in IIoT, adeptly navigating the challenges of resource-constrained environments. It epitomizes the potential of combining advanced deep learning techniques with Wavelet Transform and architectural optimizations. This research not only offers a robust and scalable approach for fault diagnosis but also paves the way for future advancements in IIoT fault diagnosis methodologies.

References

[1] Y. Liao, E. D. F. R. Loures, and F. Deschamps, "Industrial Internet of Things: A Systematic Literature Review and Insights," IEEE Internet Things J., vol. 5, no. 6, pp. 4515–4525, 2018.

[2] S. K. Jagatheesaperumal, M. Rahouti, K. Ahmad, A. Al-Fuqaha, and M. Guizani, "The Duo of Artificial Intelligence and Big Data for Industry 4.0: Applications, Techniques, Challenges, and Future Research Directions," IEEE Internet Things J., vol. 9, no. 15, pp. 12861–12885, 2022.

[3] Y. Chi, Y. Dong, Z. J. Wang, F. R. Yu, and V. C. M. Leung, "Knowledge-Based Fault Diagnosis in Industrial Internet of Things: A Survey," IEEE Internet Things J., vol. 9, no. 15, pp. 12886–12900, 2022.

[4] Q. Cao et al., "KSPMI: A Knowledge-based System for Predictive Maintenance in Industry 4.0," Robot. Comput. Integr. Manuf., vol. 74, p. 102281, 2022.

[5] J. Yu and Y. Zhang, Challenges and opportunities of deep learning-based process fault detection and diagnosis: a review, vol. 35, no. 1. Springer London, 2023.

[6] S. Lu, J. Lu, K. An, X. Wang, and Q. He, "Edge Computing on IoT for Machine Signal Processing and Fault Diagnosis: A Review," IEEE Internet of Things Journal, vol. 10, no. 13. IEEE, pp. 11093–11116, 2023.

[7] T. Qiu, J. Chi, X. Zhou, Z. Ning, M. Atiquzzaman, and D. O. Wu, "Edge Computing in Industrial Internet of Things: Architecture, Advances and Challenges," IEEE Commun. Surv. Tutorials, vol. 22, no. 4, pp. 2462–2488, 2020.

[8] L. Zhang, J. Lin, B. Liu, Z. Zhang, X. Yan, and M. Wei, "A Review on Deep Learning Applications in Prognostics and Health Management," IEEE Access, vol. 7, pp. 162415–162438, 2019.

[9] H. Hussain, P. S. Tamizharasan, and C. S. Rahul, "Design possibilities and challenges of DNN models: a review on the perspective of end devices," Artif. Intell. Rev., vol. 55, pp. 5109–5167, 2022.

[10] N. Li, L. Ma, G. Yu, B. Xue, M. Zhang, and Y. Jin, "Survey on Evolutionary Deep Learning: Principles, Algorithms, Applications, and Open Issues," ACM Comput. Surv., vol. 56, no. 2, pp. 1–34, 2024.

[11] M. Ji et al., "A neural network compression method based on knowledge-distillation and parameter quantization for the bearing fault diagnosis," Appl. Soft Comput., vol. 127, p. 109331, 2022.

[12] A. Ding, Y. Qin, B. Wang, L. Jia, and X. Cheng, "Lightweight Multiscale Convolutional Networks With Adaptive Pruning for Intelligent Fault Diagnosis of Train Bogie Bearings in Edge Computing Scenarios," IEEE Trans. Instrum. Meas., vol. 72, p. 3502813, 2023.

[13] N. D. Thuan, T. P. Dong, H. Thi Nguyen, and H. S. Hoang, "Efficient bearing fault diagnosis with neural network search and parameter quantization based on vibration and temperature," Eng. Res. Express, vol. 5, no. 2, p. 025044, 2023.

[14] Z. Xing, Y. He, and W. Zhang, "An Online Multiple Open-Switch Fault Diagnosis Method for T-Type Three-Level Inverters Based on Multimodal Deep Residual Filter Network," IEEE Trans. Ind. Electron., vol. 70, no. 10, pp. 10669–10679, 2023.

[15] S. Duan et al., "Distributed Artificial Intelligence Empowered by End-Edge-Cloud Computing: A Survey," IEEE Commun. Surv. Tutorials, vol. 25, no. 1, pp. 591–624, 2023.

[16] Y. Liu, Y. Sun, B. Xue, M. Zhang, G. G. Yen, and K. C. Tan, "A Survey on Evolutionary Neural Architecture Search," IEEE Trans. Neural Networks Learn. Syst., vol. 34, no. 2, pp. 550–570, 2023.

[17] K. T. Chitty-Venkata and A. K. Somani, "Neural Architecture Search Survey: A Hardware Perspective," ACM Comput. Surv., vol. 55, no. 4, pp. 1–36, 2022.

[18] D. Yao, G. Li, H. Liu, and J. Yang, "An intelligent method of roller bearing fault diagnosis and fault characteristic frequency visualization based on improved MobileNet V3," Meas. Sci. Technol., vol. 32, no. 12, p. 124009, 2021.

[19] W. Wang et al., "Intelligent Fault Diagnosis Method Based on VMD-Hilbert Spectrum and ShuffleNet-V2: Application to the Gears in a Mine Scraper Conveyor Gearbox," Sensors, vol. 23, no. 10, p. 4951, 2023.

[20] J. Zhang, L. Duan, S. Luo, and K. Li, "Fault diagnosis of reciprocating machinery based on improved MEEMD-SqueezeNet," Measurement, vol. 217, p. 113026, 2023.

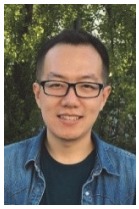
[21] Z. Liu, W. Sun, S. Chang, K. Zhang, Y. Ba, and R. Jiang, "Corn Harvester Bearing Fault Diagnosis Based on ABC-VMD and Optimized EfficientNet," Entropy, vol. 25, no. 9, p. 1273, 2023.

[22] Y. Liu, X. Li, and Y. Hu, "Differentiable neural architecture search for domain adaptation in fault diagnosis," Mech. Syst. Signal Process., vol. 202, p. 110639, 2023.
[23] X. Li, J. Zheng, M. Li, W. Ma, and Y. Hu, "One-shot neural architecture search for fault diagnosis using vibration signals," Expert Syst. Appl., vol. 190, p. 116027, 2022.
[24] J. Zhou, L. Zheng, Y. Wang, C. Wang, and R. X. Gao, "Automated Model Generation for Machinery Fault Diagnosis Based on Reinforcement Learning and Neural Architecture Search," IEEE Trans. Instrum. Meas., vol. 71, pp. 1–12, 2022.
[25] Y. Wu, B. Tang, L. Deng, and Q. Li, "Distillation-enhanced fast neural architecture search method for edge-side fault diagnosis of wind turbine gearboxes," Expert Syst. Appl., vol. 208, p. 118049, 2022.
[26] X. Wang, S. Lu, W. Huang, Q. Wang, S. Zhang, and M. Xia, "Efficient Data Reduction at the Edge of Industrial Internet of Things for PMSM Bearing Fault Diagnosis," IEEE Trans. Instrum. Meas., vol. 70, pp. 1–12, 2021.
[27] C. Chen, H. Fu, Y. Zheng, F. Tao, and Y. Liu, "The advance of digital twin for predictive maintenance: The role and function of machine learning," J. Manuf. Syst., vol. 71, pp. 581–594, 2023.
[28] L. Zhang, J. Lin, H. Shao, Z. Zhang, X. Yan, and J. Long, "End-to-end unsupervised fault detection using a flow-based model," Reliab. Eng. Syst. Saf., vol. 215, p. 107805, 2021.
[29] G. Xu, C. Huang, D. S. Da Silva, and V. H. C. De Albuquerque, "A Compressed Unsupervised Deep Domain Adaptation Model for Efficient Cross-Domain Fault Diagnosis," IEEE Trans. Ind. Informatics, vol. 19, no. 5, pp. 6741–6749, 2023.
[30] J.-G. Jang, C.-M. Noh, S.-S. Kim, S.-C. Shin, S.-S. Lee, and J.-C. Lee, "Vibration data feature extraction and deep learning-based preprocessing method for highly accurate motor fault diagnosis," J. Comput. Des. Eng., vol. 10, no. 1, pp. 204–220, 2023.
[31] N. Jia, Y. Cheng, Y. Liu, and Y. Tian, "Intelligent Fault Diagnosis of Rotating Machines Based on Wavelet Time-Frequency Diagram and Optimized Stacked Denoising Auto-Encoder," IEEE Sens. J., vol. 22, no. 17, pp. 17139–17150, 2022.
[32] L. Zhang, Q. Fan, J. Lin, Z. Zhang, X. Yan, and C. Li, "A Nearly End-to-End Deep Learning Approach to Fault Diagnosis of Wind Turbine Gearboxes Under Nonstationary Conditions," Eng. Appl. Artif. Intell., vol. 119, p. 105735, 2023.
[33] B. A. Tama, M. Vania, S. Lee, and S. Lim, Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals, vol. 56, no. 5. Springer Netherlands, 2023.
[34] T. Zhang, C. Li, J. Chen, S. He, and Z. Zhou, "Feature-level consistency regularized Semi-supervised scheme with data augmentation for intelligent fault diagnosis under small samples," Mech. Syst. Signal Process., vol. 203, p. 110747, 2023.
[35] T. Han, T. Zhou, Y. Xiang, and D. Jiang, "Cross-machine intelligent fault diagnosis of gearbox based on deep learning and parameter transfer," Struct. Control Heal. Monit., vol. 29, no. 3, pp. 1–21, 2022.
[36] Q. Huang, Y. Han, X. Zhang, J. Sheng, Y. Zhang, and H. Xie, "FFKD-CGhostNet: A Novel Lightweight Network for Fault Diagnosis in Edge Computing Scenarios," IEEE Trans. Instrum. Meas., vol. 72, pp. 1–10, 2023.
[37] H. Xu, J. Wu, Q. Pan, X. Guan, and M. Guizani, "A Survey on Digital Twin for Industrial Internet of Things: Applications, Technologies and Tools," IEEE Commun. Surv. Tutorials, vol. 25, no. 4, pp. 1–30, 2023.
[38] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2022, vol. 2022-June, pp. 11966–11976.
[39] S. Woo et al., "ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 16133–16142.
[40] S. Yang, Y. Xiang, Z. Long, X. Ma, Q. Ding, and J. Jia, "Fault Diagnosis of Harmonic Drives Based on an SDP-ConvNeXt Joint Methodology," IEEE Trans. Instrum. Meas., vol. 72, pp. 1–8, 2023.
[41] P. Guo, B. Cui, and W. Zhang, "Cross-device target migration intelligent diagnosis method of wind power," Meas. Sci. Technol., vol. 35, no. 2, p. 025025, 2024.
[42] D. Yu, H. Fu, Y. Song, W. Xie, and Z. Xie, "Deep transfer learning rolling bearing fault diagnosis method based on convolutional neural network feature fusion," Meas. Sci. Technol., vol. 35, no. 1, p. 015013, 2024.
[43] Z. Chao, Q. Feifan, Z. Wentao, L. Jianjun, and L. Tongtong, "Research on Rolling Bearing Fault Diagnosis Based on Digital Twin Data and Improved ConvNext," Sensors, vol. 23, no. 11, p. 5334, 2023.
[44] X. Xu, Z. Tao, W. Ming, Q. An, and M. Chen, "Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion," Measurement, vol. 165, p. 108086, 2020.
[45] J. Lee and H. Su, "A Unified Industrial Large Knowledge Model Framework in Smart Manufacturing," arXiv:2312.14428, pp. 1–6, 2023.
[46] 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, 2019.
[47] 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, 2021.
[48] M. Gupta, R. Wadhvani, and A. Rasool, "A real-time adaptive model for bearing fault classification and remaining useful life estimation using deep neural network," Knowledge-Based Syst., vol. 259, p. 110070, 2023.
[49] K. Zhao, F. Jia, and H. Shao, "Unbalanced fault diagnosis of rolling bearings using transfer adaptive boosting with squeeze-and-excitation attention convolutional neural network," Meas. Sci. Technol., vol. 34, no. 4, p. 044006, 2023.
[50] Y. Wang, M. Yang, Y. Zhang, Z. Xu, J. Huang, and X. Fang, "A Bearing Fault Diagnosis Model Based on Deformable Atrous Convolution and Squeeze-and-Excitation Aggregation," IEEE Trans. Instrum. Meas., vol. 70, pp. 1–10, 2021.
[51] H. Wang, W. Zhang, D. Yang, and Y. Xiang, "Deep-Learning-Enabled Predictive Maintenance in Industrial Internet of Things: Methods, Applications, and Challenges," IEEE Syst. J., vol. 17, no. 2, pp. 2602–2615, 2023.
[52] W. Zhang, D. Yang, Y. Xu, X. Huang, J. Zhang, and M. Gidlund, "DeepHealth: A Self-Attention Based Method for Instant Intelligent Predictive Maintenance in Industrial Internet of Things," IEEE Trans. Ind. Informatics, vol. 17, no. 8, pp. 5461–5473, 2021.
[53] Z. Zhao et al., "Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study," ISA Trans., vol. 107, pp. 224–255, 2020.
[54] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A Time Series is Worth 64 Words: Long-term Forecasting with Transformers," *arXiv Prepr. arXiv2211.14730*, pp. 1–24, 2022.

Liangwei Zhang (Member, IEEE) received the Ph.D. degree in operation and maintenance engineering from Luleå University of Technology, Luleå, Sweden, in 2017. He is currently an Associate Professor with the Department of Industrial Engineering, Dongguan University of Technology, Dongguan, China. His research interests include machine learning, fault detection, and prognostics and health management.

Jing (Janet) Lin (Senior Member, IEEE) is a guest professor in Division of Product Realization, Mälardalen University, Sweden, and also an associate professor in Division of Operation and Maintenance, Luleå University of Technology, Sweden. Her research interests are mainly on PHM, Asset Management, RAM4S (Reliability, Availability, Maintainability, Safety, Sustainability, Security, Supportability), e-Maintenance.

Currently, Professor Lin is the Vice President of IEEE Reliability Society and successfully initiate the IEEE RS Sweden and Norway joint section chapter since May 2021 and become as its Chair since then.



Zhe Yang received the B.E. degree in Measurement & Control & Instrument and M.Sc. degree in Mechanical Engineering from Xi'an Jiaotong University, Xi'an, China, in 2012 and 2015, respectively. He received the Ph.D. degree in Energy and Nuclear Science and Technology from Politecnico di Milano, Milan, Italy, in 2020. He is currently a Teacher of Industrial Engineering with the School of Mechanical Engineering, Dongguan University of Technology, Dongguan, China. His research interests include the development of methods and techniques for prognostics and health management of industrial components.

Haidong Shao (Senior Member, IEEE) received the B.S. degree in Electrical Engineering and Automation and the Ph.D. degree in Vehicle Operation Engineering from Northwestern Polytechnical University, Xi'an, China, in 2013 and 2018, respectively. He is currently an Associate Professor in the College of Mechanical and Vehicle Engineering at Hunan University, Changsha, China. From 2019 to 2021, he was a Postdoctoral Fellow with the Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden. His current research interests include operation and maintenance, data mining, information fusion, and industrial internet.

Biyu Liu received the Ph.D. degree in management science and engineering from Southeast University, Nanjing, China, in 2014. She is currently a professor with the Department of Logistics, School of Economics and Management, Fuzhou University, Fuzhou, China. Her research interests include maintenance management, closed-loop supply chain, lease service supply chain, and sustainable operations management.

Chuan Li (Senior Member, IEEE) received his Ph.D. degree from Chongqing University, China, in 2007. He was a Postdoctoral Fellow with the University of Ottawa, a Research Professor with Korea University, a Senior Research Associate with the City University of Hong Kong, and a Prometeo with the Universidad Politecnica Salesiana. He is a Distinguished Professor with Dongguan University of Technology, China. His research interests include fault diagnostics and intelligent systems.