IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 January 2024, accepted 13 January 2024, date of publication 17 January 2024, date of current version 26 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3355268

RESEARCH ARTICLE

Physics-Informed Time-Frequency Fusion Network With Attention for Noise-Robust Bearing Fault Diagnosis

YEJIN KIM[1] AND YOUNG-KEUN KIM[2], (Member, IEEE)
[1]Department of Mechanical and Control Engineering, Handong Global University, Pohang 37554, South Korea
[2]School of Mechanical and Control Engineering, Handong Global University, Pohang 37554, South Korea

Corresponding author: Young-Keun Kim (ykkim@handong.edu)

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) funded by the Ministry of Trade, Industry & Energy (MOTIE), Republic of Korea, under Grant 20214000000010.

ABSTRACT We propose an accurate and noise-robust deep learning model to diagnose bearing faults for practical implementation in industry. To achieve high classification accuracy in a noisy environment, we designed a time-frequency multi-domain fusion block, incorporated bearing-fault physics into the model parameters, and employed attention modules. The proposed model individually extracts essential features from the time-domain vibration signal and the corresponding spectrum in a parallel pipeline. Subsequently, multi-domain feature maps are fused to capture a wider representation of bearing fault signals. The performance was enhanced by incorporating physical knowledge of fault frequencies in the design of the frequency-domain feature extraction network. The employment of an attention mechanism to selectively focus on high-importance fault characteristics on the multi-domain feature maps further improved the accuracy under high noise levels. Experiments on bearing datasets with artificially added noise demonstrated the effectiveness of the proposed model compared to other benchmark models.

INDEX TERMS Additive white Gaussian noise, ball bearing, condition monitoring, convolutional neural networks, fault diagnosis, feature extraction, machine learning algorithms, prognostic and health management, reliability, robustness.

I. INTRODUCTION

Bearings are core components of rotational machinery. Under harsh industrial environments, bearing failure is one of the main reasons for machinery breakdown. To ensure effective maintenance and reduce the cost of unexpected breakdowns of machinery, the monitoring of machinery health is a vital factor in industry, and research on fault monitoring and diagnosis of bearings has attracted specific attention.

Bearings are encased inside the structures of rotational machinery, making direct inspections of bearing conditions highly complex. Various methods of indirect sensing of bearing conditions have been developed, including thermal imaging [1], acoustic emission [2], and vibration [3]. Vibration signals are commonly used in monitoring because of their rapid and high sensitivity to fault detection.

Traditional machine learning methods for diagnosing bearing failure have consisted of the standard sequence of raw signal preprocessing, feature extraction, and classification. The raw vibration signals are contaminated by resonance and noise from other machinery; therefore, the signals are generally prepared using several signal processing techniques. The predominant techniques for preprocessing vibration signals are envelope analysis [4], the wavelet packet transform [5], and the Hilbert-Huang transform [6]. The fault characteristic features are extracted by using statistical parameters such as the root mean square, kurtosis, multi-scale root mean square, and principal component analysis [7], [8], [9]. The features are subsequently incorporated in a machine-learning model for classifying bearing faults. Traditionally, machine

The associate editor coordinating the review of this manuscript and approving it for publication was Gerard-Andre Capolino.

learning models such as support vector machines [4], [5], [9], the K-nearest neighbors algorithm [10], and linear discriminant analysis [11] have been utilized in studies of bearing diagnostics.

Recently, the rapid growth of computing power and the exceptional performance levels facilitated by advanced model architectures have brought deep learning models to greater prominence in studies of bearing diagnostics [12]. In contrast to traditional machine learning approaches that require the processes of manual feature extraction and classification to be conducted separately, deep learning models can perform an end-to-end diagnosis directly from the input signal. Furthermore, it has been demonstrated that deep learning models can outperform traditional machine learning models because of their high level of abstraction in capturing the complex and non-linear features of fault characteristics [12], [13], [14].

Among several deep learning methods, convolutional neural networks (CNNs) have been extensively applied in studies of bearing fault diagnostics. In these networks, the input time-series vibration signal is processed directly by 1D CNNs for a cost-effective classification of bearing faults [15], [16], [17], [18], [19]. Zhang et al. designed a 1D CNN with a wide first layer kernel size to diagnose bearing faults [15]. Song et al. applied a wide-kernel 1D CNN for a wide receptive field [16], whereas Huang et al. used the varied scale of the filters of a 1D CNN model to obtain more valuable information for diagnosing the bearing states [17], and He et al. proposed multi-scale kernels with varied dilation rates for the multi-scale features of a 1D CNN [18]. In addition, Chen et al. proposed a 1D bearing fault diagnosis model that integrated CNN with a long short-term memory (LSTM) network [19]. The model extracts features covering the low- and high-frequency regions of the input signal with a multi-scale kernel. However, the aforementioned 1D CNN models with time-domain vibration input may become compromised when the input signals are corrupted by resonances, noise, and other disturbance signals associated with other machinery.

In contrast with analysis in the time domain, the representation of vibration signals in the frequency domain has the advantage of easily separating signals associated with fault characteristics from background and noise signals [20]. However, the fault frequency characteristic may not be readily visible in the early stages of bearing decay [21]. Therefore, approaches (using 2D CNN) that capture feature representations from both the time and frequency domains have been proposed. The input 1D time-domain vibration signal is processed to transform it into a 2D time-frequency signal that is then applied in a powerful 2D CNN model [22]. Several methods have been used to transform bearing signals, including the wavelet packet energy [23], short-time Fourier transform (STFT) [24], [25], and continuous wavelet transform (CWT) [26], [27]. Huang et al. proposed a 2D multi-scale CNN that extracted local and global information from the feature map using a multi-scale kernel, and focused on important information using channel and spatial attention [26]. However, 2D CNNs are usually heavier and slower than 1D CNN models because of the 2D models' larger number of trainable parameters.

Recently, a light-weighted and faster 1D CNN model using multi-transformation domain signals has been proposed for monitoring bearing health. Liu et al. proposed a parallel CNN model combining an STFT-based 2D CNN and a time domain 1D CNN [28]. The model used multi-domain data but was constrained by the large number of trainable parameters resulting from the inclusion of a 2D CNN architecture. Dong et al. designed a 1D dilated CNN with multi-domain inputs, consisting of the manually prepared statistical features of the vibration signals, to diagnose bearing conditions [29]. Sun et al. proposed a 1D CNN of residual dense network time that used multi-domain data [30]. The time and the transformed frequency signals were first concatenated and then applied as input to the CNN model.

Similarly to the study by Sun et al., the present study is focused on the design of a light-weighted 1D CNN model based on multi-domain signal information. However, in contrast with previous studies, the proposed model contains independent time-domain and frequency-domain feature extraction layers, to capture a wider representation of fault features in a parallel pipeline.

None of the aforementioned published studies considered the noisy and harsh environments in which rotating machines operate in real-world industrial sites. For example, the vibration signals emanating from a particular set of machinery in a heavy-industry manufacturing environment are susceptible to various shocks and disturbances emanating from other machinery. As a further consideration, many previous studies have been developed using open datasets applied to bearings, such as the Case Western Reserve University (CWRU) datasets [31], which are prepared in a laboratory environment.

Because the noise contained in a vibration signal can diminish the accuracy of bearing fault diagnosis [32], various noise-robust bearing diagnosis models that account for high noise levels have recently been proposed. Li et al. [33] proposed a deep distance metric learning method based on a CNN to address the domain shift problem caused by environmental noise. The method used the frequency features as input and was tested under noise levels ranging from an SNR of -8 dB to 8 dB. In [34], a deep transfer learning method was proposed that applied the bearing diagnosis model to the target domain after being trained on multiple source domains. The method was tested under various noise levels ranging from an SNR of -4 dB to 8 dB. Zhang et al. applied adaptive batch normalization to a 1D CNN with a wide first-layer kernel (WDCNN) to obtain a robust diagnosis of bearing faults in a noisy environment with SNR values ranging from -4 dB to 10 dB [15]. A transformer-based convolutional model with linear self-attention (CLFormer) was designed to boost the accuracy of a lightweight model [35]. The model

achieved a high accuracy with SNR values ranging from −8 dB to 8 dB with the self-made dataset. A composite of 1D CNN and LSTM model is proposed in MCNN-LSTM [19] that extracted high-frequency and low-frequency components of the vibration signal with multi-scale kernels. Another 1D CNN used both the frequency spectrum and the phase information of bearing vibration signals for noise robustness [36]. However, the aforementioned studies were designed for either time or frequency domain input only, not for multi-domain signals; such signals can effectively enhance the robustness against noise, as we also demonstrate in the present study.

Furthermore, our proposed model implements an attention mechanism to boost the accuracy of bearing diagnosis in a strong-noise environment. A number of studies have also applied attention mechanisms in deep learning models to achieve robustness against noise. Wang et al. [37] applied multi-head attention to a 2D CNN, and Peng et al. [38] proposed a 1D CNN with multiple residual mixed domain attention modules to extract semantic information in the time and channel domain. Yin et al. employed a Squeeze-and-Excitation network [39] on a 1D CNN with a time domain signal [40]. Zhong et al. proposed an ensemble model consisting of SqueezeNet based on the 2D image of a continuous wavelet transformation with self-attention [41]. An adaptive model for anti-noise bearing diagnosis (CLFormer) [35] applied linear attention on a GRU block and exhibited a slight enhancement in accuracy. However, all the aforementioned attention-based models were based on time-domain input signals. In contrast to these studies, we propose an attention module for rescaling the essential weights of features in time-frequency multi-domains.

A number of research groups have attempted to employ prior physical knowledge of fault bearings to the deep learning model as a particular approach for improving bearing diagnosis [42], [43], [44]. A neural network whose learning is guided by physical information is known as a physics-informed neural network (PINN) [45]. Shen et al. designed a bearing diagnosis model that applied a physics-informed threshold algorithm to the input signal to increase the reliability of the diagnosis [42]. Sadoughi et al. proposed a physics-based CNN that extracts periodic fault signals using a kernel design based on bearing fault physics [43]. Lu et al. improved the performance and reliability of a data-driven bearing diagnosis model by applying high weights to the bearing fault key frequencies in the proposed CNN model [44]. Although the employment of bearing fault physics has shown promising results, only a few such studies have been conducted. Consequently, our proposed model employs bearing-fault physics using a different approach for designing kernels in the frequency-domain CNN to enhance the model's diagnostic performance further.

In summary, we propose a highly accurate bearing fault diagnosis model that accommodates a strong noise environment and is based on a physics-informed 1D CNN of time-frequency multi-domain fusion combined with a Squeeze-and-Excitation attention block (MDFN+SE). The main contributions of this study are:

1) We propose an accurate and light-weighted 1D CNN that is composed of a parallel network for extracting each time- and frequency-domain feature, and a multi-domain fusion block for capturing a wider representation of bearing fault characteristics across multi-domains.
2) Physical knowledge of bearing fault frequencies was incorporated in the design of the model parameters of the frequency-domain network to enhance the robustness against noise effectively.
3) To facilitate a further increase in model accuracy under strong noise, a channel-wise attention mechanism was employed to assist the model in selectively focusing on the essential fault-related features in multi-domain representations.
4) Ablation studies and experiments on multiple bearing datasets with added noise at SNR levels ranging from −6 dB to 6 dB were conducted and the results were compared with those of related models to confirm the effectiveness of the proposed model in diagnosing bearing faults.

The remainder of this paper is organized as follows. The bearing physics-informed preprocessing methods are explained in Section II. In Section III, the architecture of the proposed model is described in detail. Sections IV and V include the experimental results for commonly-used bearing datasets, in order to verify the proposed model in comparison with similar models. Finally, the conclusions of this study are presented in Section VI.

II. DOMAIN-INFORMED PREPROCESSING

The raw bearing vibration signal exists in the time domain; however, we also used the transformed frequency domain signal in this study, to capture a wider representation of fault characteristics.

The transformed frequency-domain spectrum is a mixture of fault-specific frequencies, inter-machine resonance, and other sources of noise. Consequently, we applied the bearing physics-informed preprocessing to the model to prepare for an enhanced form of frequency-domain signals.

A. BEARING FAILURE PHYSICS

An understanding of the physics of bearing faults facilitates the insight into signal processing methods to prepare datasets of higher quality. A bearing is composed of several main parts: the inner race, outer race, rolling elements, and cage elements. If a defect occurs in the bearing, a periodic impulse signal is generated when the fault area collides with the rolling elements. Depending on the types of bearing faults, the impulse signals have specific fault frequency characteristics that can be estimated from the bearing model specifications.

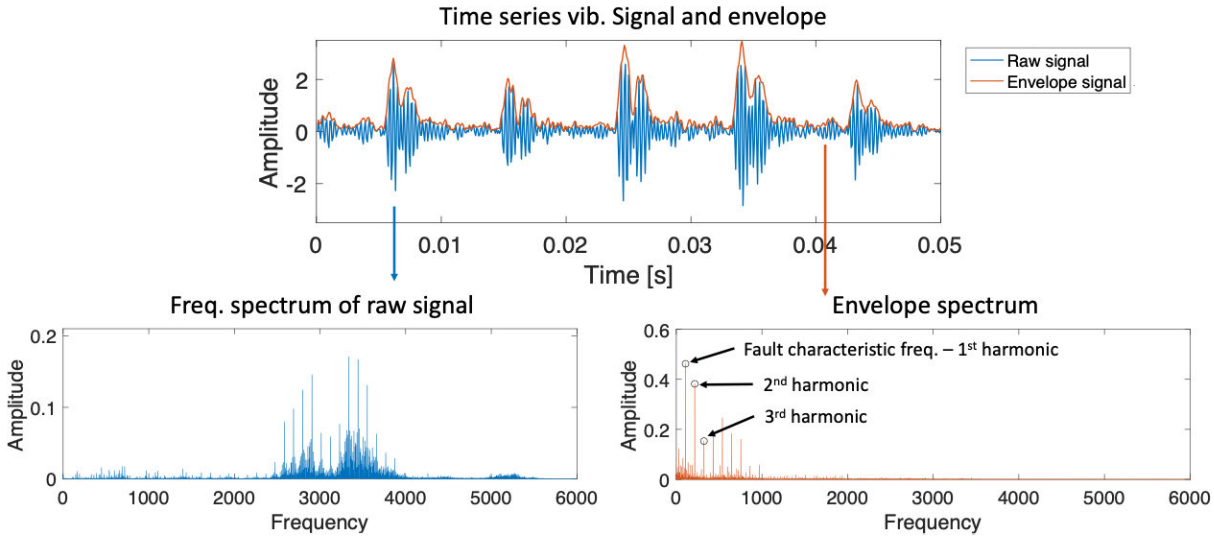

FIGURE 1. Comparison of frequency spectrum from raw signal and envelope extracted signal. Fault characteristic frequencies are more clearly visible by envelope extraction.

This study concentrated on three types of bearing faults: inner race faults, outer race faults, and ball faults. The fault characteristic frequencies of each fault type can be estimated from the bearing physics [46]:

Ball pass frequency for inner race (BPFI):

$$f_{IR} = \frac{Z\left(1 + \cos\alpha \times \frac{d}{D}\right)}{2} \times f_r \tag{1}$$

Ball pass frequency for outer race (BPFO):

$$f_{IR} = \frac{Z\left(1 - \cos\alpha \times \frac{d}{D}\right)}{2} \times f_r \tag{2}$$

Ball spin frequency (BSF):

$$f_{BS} = \frac{1}{2}\frac{D}{d}\left(1 - \left(\frac{d}{D} \times \cos\alpha\right)^2\right) \times f_r \tag{3}$$

where

f_r: shaft rotational speed
Z: number of rolling elements
D: pitch circle diameter of the bearing
d: ball diameter
α: contact angle.

B. ENVELOPE ANALYSIS

The purpose of signal preprocessing is to highlight the main fault characteristic frequencies in order to enhance the feature extractions. The bearing fault impulse signal is amplitude-modulated by the carrier signals of noise and machinery resonance. When the modulated fault signal is transformed to a frequency spectrum, as shown in the left side of Fig. 1, the fault characteristic frequencies are masked. Therefore, the raw signal needs to be demodulated to extract the defect fault signals appropriately with the aid of the bearing domain information.

The amplitude demodulation is processed by the envelope analysis, which constructs the vibration signal as a complex analytic signal with the imaginary component consisting of the Hilbert transform of the real part. The mathematical expression of the analytic signal $x_h(t)$ is

$$x_h(t) = x(t) + j\hat{x}(t) \tag{4}$$

where $\hat{x}(t)$ is Hilbert transform of raw signal, $x(t)$.

The absolute value of the analytic signal $|x_h(t)|$ is the envelope of the raw signal, the demodulated fault characteristic signal harmonics, which are clearly visible in the frequency domain, as shown in Fig. 1.

C. DATASET AUGMENTATION BASED ON FAULT FREQUENCIES

The raw vibration signal input data are the time-series vibration signals obtained through monitoring the bearings. The dataset applied in this study is the Case Western Reserve University (CWRU) benchmark [31], which contains bearing vibration signals sampled at a rate of 12 kHz. The CWRU dataset contains only one set of vibration data of approximately 10 s duration for each fault. To acquire a sufficient number of datasets, a data augmentation method was applied by using the overlapping window as illustrated in Fig. 2.

The length of the overlapping window was designed by considering the fault frequency analysis of the vibration signals. The primary BPFI, BPFO, and BSF were all located in a relatively low frequency range below 600 Hz. Therefore, the frequency resolution of the vibration spectrum was selected to be approximately 1 Hz, so that fault types could be distinguished. Using the relationship expressed in Eq. 5, the

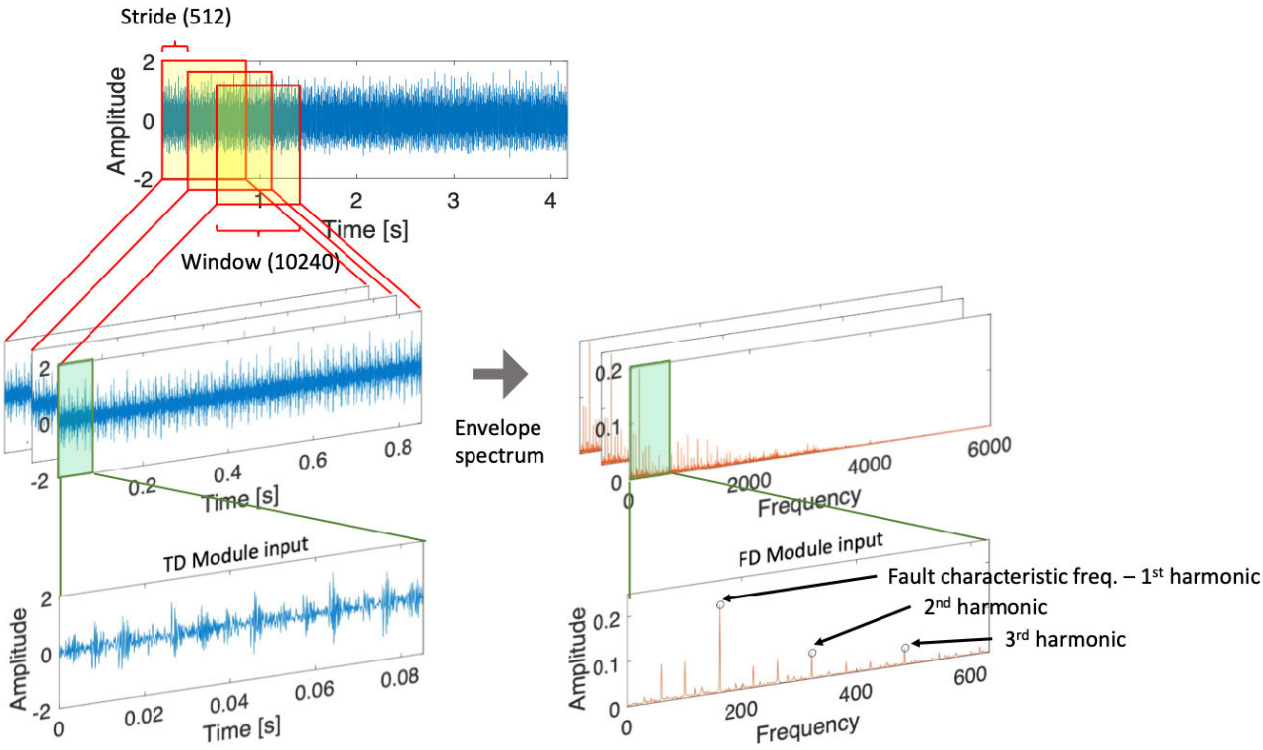

FIGURE 2. Data augmentation method for time-domain and frequency-domain signals.

window length L_{window} was set to 10240 – as shown in Fig. 2.

$$f_{res} = f_{sampling}/L_{window} \tag{5}$$

where f_{res} is the frequency resolution and $f_{sampling}$ is the sampling frequency of the accelerometer.

The time-domain input signal is generated from the augmented dataset by reducing the data length tenfold to 1024 points. Considering the primary fault frequencies, the reduced time-domain signal includes approximately nine fault periodic signals, which is sufficient for feature extractions.

D. PREPARATION FOR FREQUENCY-DOMAIN

Using the augmented dataset, the frequency spectrum is prepared with the physics-informed preprocessing described in II-A and II-B.

The demodulated vibration signal obtained through the envelope analysis is transformed to the frequency representation using the fast Fourier transform (FFT). The sampling frequency of the CWRU dataset is 12 kHz; accordingly, the spectrum signal has a range from 0 to 6 kHz. Since the bearing fault frequencies are located at a low bandwidth, we extracted the spectrum signal from 0 to 600 Hz. The spectra included fault signals up to the third harmonics. Accordingly, the frequency-domain dataset contained 540 points with a frequency resolution of approximately 1 Hz. The frequency-domain signal is normalized by subtracting its average power and re-scaling.

III. PROPOSED MODEL: MDFN+SE

A. NETWORK ARCHITECTURE

As shown in Fig. 3, the proposed multi-domain fusion network with a Squeeze-and-Excitation block (MDFN+SE) is composed of three main blocks: the time-domain feature extraction module (TD module), the frequency-domain feature extraction module (FD module), and the multi-domain fusion module (MDF module).

The TD module is a component of the feature extraction process that is not well understood as yet. The FD module employs more explicit information on the periodic fault components using the frequency representation of the raw signal. In addition, the design of the kernel size and stride of the FD module was based on a physics-informed method so that features of each fault type could be extracted. At the conclusion of the feature extraction process in each module, the extracted features are concatenated into the multi-domain fusion layer in the MDF module. Since the important feature domain differs for each fault type, the SE block (a widely used attention module) is used to direct the focus of the network toward the appropriate feature channels in a multi-domain fusion layer in the MDF module. The SE block enables the model to attend to significant channels by strengthening the essential features and suppressing irrelevant information.

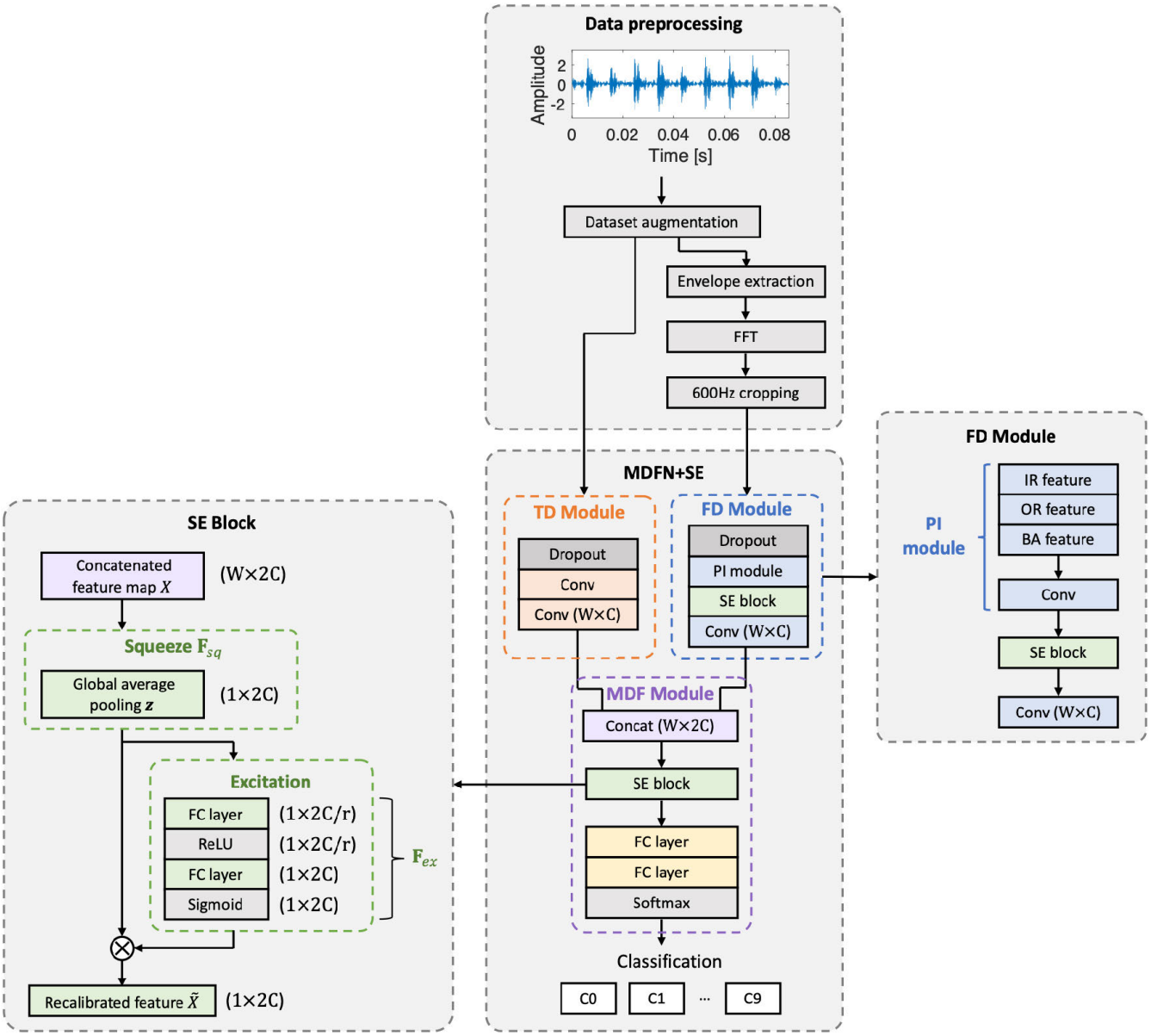

FIGURE 3. Architecture of the proposed MDFN+SE model with preprocessing.

Finally, the feature layer is forwarded to the MLP layer to predict the health state of the bearing. Each of these process steps is explained in detail in the next section.

1) TIME-DOMAIN FEATURE EXTRACTION: TD MODULE

To improve the generalization performance of the model, a 0.5 percentage dropout layer is inserted before the first convolutional layer. In addition, a 0.2 percentage dropout layer is inserted after each convolutional layer. The TD module is based on a 1D CNN architecture as specified in Table 5. The input is the time-domain vibration signal of 2048 data points. The TD module consists of two convolutional layers with a rectified linear unit (ReLU) activation function. The batch normalization layers are applied at each convolutional layer to ensure the stability of the training process. The feature map from the TD module is passed to the global averaging pooling layer to generate a 1D vector input to the fully connected layer. It does not use the classical feature map flattening, which can create a computational bottleneck and overfitting. In contrast, the global average pooling (GAP) of the channel-wise direction can significantly reduce the number of trainable parameters to avoid overfitting and exhibit improved performance [47].

2) FREQUENCY-DOMAIN FEATURE EXTRACTION: FD MODULE

As with the TD module, a 0.5 percentage dropout layer is added before the first layer, and a 0.2 percentage dropout

TABLE 1. Detailed description of the CWRU dataset for case study 1.

Fault location		Normal	Inner race			Outer race			Ball			Motor load [HP]
Fault size [inch]		-	0.007	0.017	0.021	0.007	0.017	0.021	0.007	0.017	0.021	
Fault label		0	1	2	3	4	5	6	7	8	9	
Dataset A	Train	306	306	306	306	306	306	306	306	306	306	0
	Test	102	102	102	102	102	102	102	102	102	102	
	valid	102	102	102	102	102	102	102	102	102	102	
Dataset B	Train	306	306	306	306	306	306	306	306	306	306	1
	Test	102	102	102	102	102	102	102	102	102	102	
	valid	102	102	102	102	102	102	102	102	102	102	
Dataset C	Train	306	306	306	306	306	306	306	306	306	306	2
	Test	102	102	102	102	102	102	102	102	102	102	
	valid	102	102	102	102	102	102	102	102	102	102	
Dataset D	Train	306	306	306	306	306	306	306	306	306	306	3
	Test	102	102	102	102	102	102	102	102	102	102	
	valid	102	102	102	102	102	102	102	102	102	102	

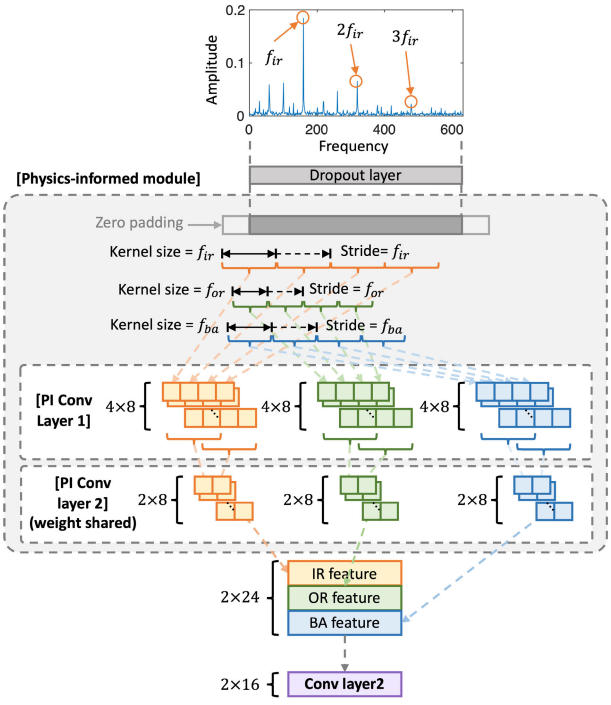

FIGURE 4. Architecture of the proposed physics-informed feature extracting layer.

layer is added after all convolutional layers, to generalize the model. The FD module extracts features from the spectrum input of 540 points. The FD module has one physics-informed layer and two convolutional layers with the ReLU activation function, batch normalization, and a global average pooling layer. The vibration signal acquired from a defective bearing has a periodic fault characteristic frequency, and its harmonics appear at regular intervals in the frequency domain. Because the fault characteristic frequency differs for each fault location, the kernel and stride design was based on physical information to capture features for each fault type. Features of the inner race fault (IR), outer race fault (OR), and ball fault (BA) are respectively extracted through three parallel convolutional layers as shown in Fig. 4. Each layer containing IR, OR, and BA fault-related features with size 4×8 goes through a convolutional layer again (PI conv layer 2). In this layer, three feature blocks share the same kernel weights to extract features up to the third harmonic of each fault characteristic frequency. Subsequently, the features associated with each fault type are passed through one more convolutional layer and are concatenated channel-wise. The entire feature combining all fault types is finally passed through a convolutional layer.

3) MULTI-DOMAIN FUSION NETWORK

The 1D feature vectors from the time- and frequency-domain feature extraction modules are passed to the multi-domain fusion multi-layer perception (MLP). The multi-domain feature vectors are concatenated and passed to the Squeeze-and-Excitation attention block. This channel-wise attention block assists in assigning a higher weight to essential features in the different domains, as explained in detail in Section III-B. Subsequent to a series of fully connected layers, the softmax output nodes are generated in the number of fault classes for the classification.

B. ATTENTION MODULE FOR NOISE ROBUSTNESS

Industrial data are often corrupted with various unpredictable types of noise, depending on the working environment. Therefore, we applied an attention module to the multi-domain fusion module to improve the overall accuracy of the proposed network in a noisy environment.

The attention mechanism has been widely used in natural language processing to enhance the performance of seq-to-seq models by focusing on the essential parts in the input sequence [48]. As the attention mechanism has been shown to enhance model performance, a number of studies have applied it in bearing diagnosis [37], [38], [40], [41]. However,

TABLE 2. Parameters of the model variants of time-domain module network with different layer depth.

TD-L				TD-M				TD-S			
Layer	Kernel /stride	Kernel no.	Output shape	Layer	Kernel /stride	Kernel no.	Output shape	Layer	Kernel /stride	Kernel no.	Output shape
Dropout	p=0.5		2048×1	Dropout	p=0.5		2048×1	Dropout	p=0.5		2048×1
Conv1	256/3	4	683×4	Conv1	256/8	8	256×8	Conv1	256/16	16	128×16
Maxpool	2/2		341×4	Maxpool	2/2		128×8	Maxpool	2/2		64×16
Conv2	19/3	8	114×8	Conv2	19/4	16	32×16	Conv2	19/5	16	13×16
Conv3	7/3	16	38×16	Conv3	7/2	32	16×32	GAP			16×1
Conv4	5/2	32	19×32	Conv4	5/1	64	16×64	fc1			1×16
Conv5	3/1	64	19×64	GAP			1×64	fc2			1×10
Conv6	3/1	128	19×128	fc1			1×16				
GAP			1×128	fc2			1×10				
fc1			1×16								
fc2			1×10								

TABLE 3. Parameters of the model variants of freuency-domain module network with different layer depth.

FD-L				FD-M				FD-S			
Layer	Kernel /stride	Kernel no.	Output shape	Layer	Kernel /stride	Kernel no.	Output shape	Layer	Kernel /stride	Kernel no.	Output shape
Dropout	p=0.5		540×1	Dropout	p=0.5		540×1	Dropout	p=0.5		540×1
Conv1	256/3	4	180×4	Conv1	256/8	8	68×8	conv1	256/16	16	34×16
Conv2	19/3	8	60×8	Conv2	19/4	16	17×16	conv2	19/5	16	5×16
Conv3	7/3	16	20×16	Conv3	7/2	32	9×32	GAP			1×16
Conv4	5/2	32	10×32	Conv4	5/1	64	9×64	fc1			1×16
Conv5	3/1	64	10×64	GAP			1×64	fc2			1×10
Conv6	3/1	128	10×128	fc1			1×16				
GAP			1×128	fc2			1×10				
fc1			1×16								
fc2			1×10								

the attention modules used in those studies were not designed for multi-domain fused features.

The input data for the proposed model are the vibration time signal and its frequency spectrum, which are related to the bearing condition. Consequently, the overall shape of the spectrum and features of the local parts of each domain signal are both important information. Among various types of attention modules, we selected channel-wise attention to capture the essential global features of fault characteristics in both the time and frequency domains. Also, channel-wise attention can analyze the interdependencies between the channels to highlight the essential fault features across the different domains.

This study applied the Squeeze-and-Excitation block (SE block) in Squeeze-and-Excitation networks (SENet) [39] that performs dynamic modeling of channel-wise feature dependencies. The architecture of the SE block, as illustrated in Fig. 3, directs the model to focus on important features in the channel direction by squeezing and exciting the feature map of the model.

The convolution of a filter operates in conjunction with a local receptive field that is unable to exploit features outside the region. This limitation can be eased by squeezing global spatial information using GAP to generate channel-wise statistics. The input feature map $\mathbf{u_c}$ of size $W \times 2C$ is squeezed to $1 \times 2C$ by the squeeze function $\mathbf{F}_{sq}$. The c-th element of the squeezed vector from the c-th element of input $\mathbf{u_c}$ is calculated by:

$$z_c = \mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{W}\sum_{i=1}^{W} u_c(i) \tag{6}$$

The excitation operation captures the channel-wise dependencies of the compressed information generated by the squeeze operation. This constitutes a simple bottleneck with two fully connected layers with sigmoid activation. The dimension of the input channel, $2C$, is reduced by the ratio of r with ReLU activation. Subsequently, it is returned to the input channel dimension by the second fully-connected layer (FC layer) with the sigmoid activation. The output of the excitation process is the vector $\mathbf{s_c}$ of $1 \times 2C$:

$$\mathbf{s_c} = \mathbf{F}_{ex}(\mathbf{z_c}, \mathbf{W}) = \sigma(\mathbf{W}_2\delta(\mathbf{W}_1\mathbf{z_c})) \tag{7}$$

where σ is ReLU and δ sigmoid activation function, and weight $W_1 \in \mathbb{R}^{\frac{2C}{r}\times 2C}$, and $W_2 \in \mathbb{R}^{2C\times\frac{2C}{r}}$.

The vector $\mathbf{s_c}$ is the attention weight that emphasizes significant features of the input feature map. Therefore, the input feature map $\mathbf{u}_c$ is rescaled by the element-wise multiplication with the vector $\mathbf{s_c}$, producing the attention weight applied feature map as follows:

$$\tilde{\mathbf{x}} = \mathbf{s_c} \odot \mathbf{u}_c \tag{8}$$

where $\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_c]$

TABLE 4. Classification accuracy of ablation study with different SNR values on CWRU dataset B.

Model name	TD module	FD module	MDF module	SE block in MDF	SNR[dB]							
					None	6	4	2	0	-2	-4	-6
TDN	✓				99.8	99.8	99.8	99.8	99.8	99.7	96.63	80.69
FDN		✓			**100**	**100**	99.8	98.32	95.94	89.31	77.43	57.82
MDFN	✓	✓	✓		**100**	**100**	**100**	**100**	**100**	99.9	97.82	86.34
MDFN+SE	✓	✓	✓	✓	**100**	**100**	**100**	**100**	**100**	**100**	**99.21**	**93.47**

IV. CASE STUDY 1: CWRU DATASET

A. DATASET DESCRIPTION

The proposed model was initially designed for the bearing dataset provided by Case Western Reserve University (CWRU) [31]. The dataset consists of vibration signals acquired at a sampling rate of 12 kHz on rotating machinery that uses a deep groove ball bearing (SKF 6205-2RS). It is driven by a 2 HP motor and operated under various motor load conditions: 0, 1, 2, and 3 HP. The dataset was divided by each motor load as shown in Table 1. Bearing faults are artificially processed by electro-discharge machining (EDM) and are classified as normal, inner race fault, outer race fault, or ball fault. Each fault has three fault depth sizes of 0.007, 0.014, and 0.021 inches, respectively.

This study classified the state of bearing health according to the types and severity of the failures as summarized in Table 1. There were a total of ten classes, and each class dataset included all motor load (Dataset A+B+C+D) conditions in the same ratio. After the aforementioned data augmentation process, a total of 8600 datasets were prepared, which were divided into training, validation, and testing data in a ratio of 6:2:2.

B. TRAINING CONFIGURATION

The model loss was calculated using the cross-entropy, and the adaptive moment estimation (Adam) optimizer had an initial learning rate of 0.001. Model training was stopped early if the epoch number reached 60 or the validation accuracy decreased continuously, to avoid overfitting.

C. ABLATION STUDY

A series of ablation studies of the network design was conducted to select an optimal parameter configuration.

1) NOISE ENVIRONMENT

The model variants were trained under no-noise conditions and tested under various noise conditions to evaluate robustness and accuracy in noisy environments. The dataset used for the ablation study is CWRU dataset B. For all experiments, the performance of the model was evaluated using the median result from five repetitive tests.

Although it is challenging to mimic the real noise environment, additive white Gaussian noise is widely used in noise robustness studies [15], [19], because it can closely resemble many random noise in nature. The test dataset was mixed with Gaussian white noise, which has zero mean, and the standard normal distribution is based on a random generator. The power of noise is calculated from the power of the input signal to satisfy the desired signal-to-noise ratio (SNR ratio). The SNR is defined as the ratio of the signal power (P_s) to the noise power (P_n), i.e., $SNR = 10log_{10}\frac{P_s}{P_n}$. The model was trained without noise and evaluated with the noise-added test data with different SNR levels from -6 dB to 6 dB.

2) THE NUMBER OF CONVOLUTIONAL LAYERS

The feature extraction blocks of the time and frequency domains consist of 1D convolutional layers and ReLU activation functions. Padding is applied to all layers to prevent the loss of features at the ends. The effect of the number of convolutional layers in the CNN model was analyzed to design an optimal number of layers that would enhance the representation of the characteristics of the input signals.

In the analysis, the time-domain and frequency-domain feature extraction modules were individually tested by eliminating the other domain modules to minimize their mutual effect. Furthermore, the SE block was removed in this analysis. Each model for setting the number of layers is summarized in Table 2. The time domain models with 6, 4, and 2 convolutional layers are named TD-A, TD-B, and TD-C, respectively. Similarly, the frequency domain models are named FD-A, FD-B, and FD-C, respectively, as listed in Table 3. By setting the stride for each convolutional layer, the feature-length was reduced as the convolution progressed. For models with shallower, the stride was set wider so that the final feature lengths of the models were similar.

The test results are shown in Fig. 5. All the model variants achieved performance levels close to 100% under noise-free conditions. As the noise level increased, the performance of the time-domain feature extraction module started to decrease with varying severity, according to the number of convolutional layers. For both time- and frequency-domain feature extraction, the model with the two convolutional layers displayed superior performance across all noise conditions. Therefore, the number of convolutional layers for both the time and frequency-domain feature extractions in the final model was selected to be two.

3) EFFECTIVENESS OF PHYSICS-INFORMED NETWORK

An experiment was conducted to evaluate the performance of the purely data-driven layer and the physics-informed layer in the FD module. As shown in Section IV-C2, the optimal

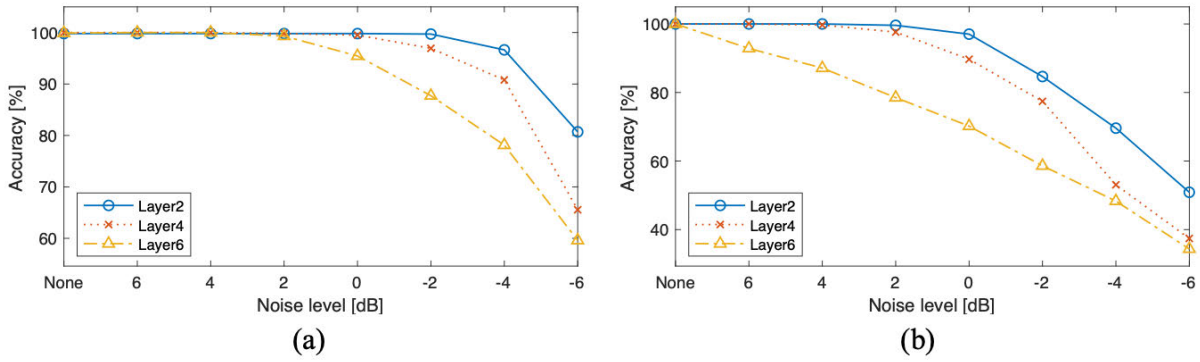

FIGURE 5. Effect of convolutional layer numbers for (a) TD module and (b) FD module with different SNR values.

TABLE 5. Parameter description of the proposed network MDFN+SE.

TD module					FD module				
Layer	kernel /stride	kernel no.	output shape	padding	Layer	kernel /stride	kernel no.	output shape	padding
Dropout		p=0.5			Dropout		p=0.5		
Conv1	256/16	16	128×16	127	PI conv1-ir1	138/138	8	4×8	69
Maxpool	2/2		64×16		PI conv1-or1	90/90	8	4×8	45
Conv2	19/5	16	13×16	9	PI conv1-ba1	120/120	8	4×8	60
GAP			16×1		Conv2	3/2	8	2×8	7
					Concat			2×24	
					SE Block	1/4		2×24	
					Conv2	3/1	16	2×16	1
					GAP			1×16	

MDF module				
Layer	Reduction ratio	Input shape	Output shape	Activation
SE Block	1/4	1×32	1×32	ReLU
FC1		1×32	1×16	ReLU
FC2		1×16	1×10	Softmax

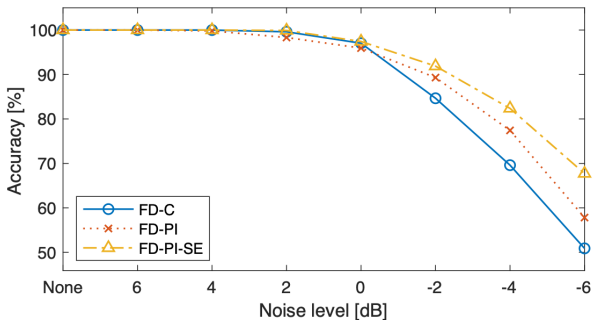

FIGURE 6. Effect of physics-informed parameter design on the accuracy of FD module.

performance exhibited by the purely data-driven FD modules was that of FD-C. It was compared with the FD-PI model, in which the first convolutional layer of the FD module was replaced with a physics-informed layer. The performance of the FD-PI-SE module, which added SE blocks to the FD-PI module to improve prediction performance, was also evaluated. All models displayed prediction performance close to 100% in environments with no or weak noise; the performance degradation as noise levels increased are shown in Fig. 6. The reduction ratio of the SE block was set to 4. In Fig. 6, when the first layer is replaced with a physics-informed layer, the noise robustness of the module is improved. When the SE block is added to the FD-PI module, the prediction performance in a noisy environment increases slightly. We ascribe the improvement in prediction performance to the SE module's emphasizing of appropriate channels that include the features of each fault type.

4) REDUCTION RATIO OF SE BLOCK

The SE block was included to emphasize the essential part of the multi-domain fused features in the MDF module. In order to highlight the channel-wise features effectively, the SE block needs to be configured with the appropriate reduction ratio (r). It has been previously demonstrated that a reduction ratio of 16 achieves an optimal balance between accuracy and complexity for 2D image processing [39]. However, the optimal ratio value should be re-analyzed for the 1D vibration signals of fault bearings.

Several reduction ratios, ranging from 2 to 16, were evaluated to determine the most appropriate reduction ratio under noisy conditions. Fig. 7 shows that a reduction ratio of

TABLE 6. Classification accuracy[%] of the proposed model compared to different methods under various SNR values on CWRU dataset.

Model	Train/test dataset	SNR [dB]								No. of params
		none	6	4	2	0	-2	-4	-6	
WDCNN [15]	B+C+D	-	99.77	99.23	97.52	90.15	80.81	66.95	-	54,510
MCNN-LSTM [19]	D	98.46	90.11	88.19	85.90	81.41	77.27	-	-	73,480
FD 1D-CNN [36]	A+B+C+D	-	98.12	98.37	97.87	97.12	95.37	92.64	84.87	**9,220**
TICNN [49]	A+B+C+D	-	99.59	99.61	99.27	98.22	96.47	82.05	-	66,182
MDFN (ours)	A+B+C+D	99.94	99.97	**100**	**100**	**99.97**	99.79	98.91	90.18	15,770
MDFN+SE (ours)	A+B+C+D	**100**	**100**	**100**	**100**	99.94	**99.83**	**99.07**	**93.26**	

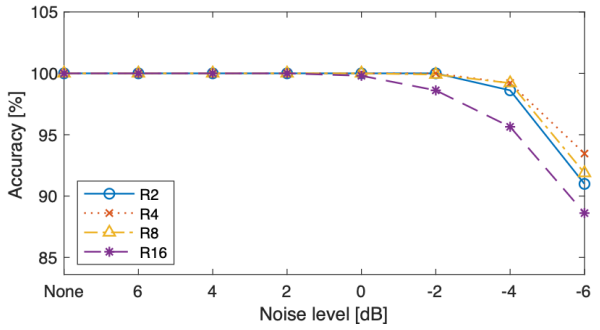

FIGURE 7. Effect of reduction ratio of SE block on the model accuracy.

4 exhibited the highest performance at all noise levels. Under the highest noise condition of $SNR = -6$ dB, this setting demonstrated the strongest robustness, achieving an accuracy of approximately 93%.

5) EFFECT OF MULTI-DOMAIN AND ATTENTION MODULE

The overall effect of applying a multi-domain model and attention module was analyzed. The compared model variants are: (1) time-domain module only network (TDN), (2) frequency-domain module only network (FDN), (3) multi-domain fusion network without attention (MDFN), and (4) multi-domain fusion network with attention (MDFN+SE).

The effectiveness of multi-domain fusion under a strong noise is clearly visible in Table 4. Under noise level of less than -2 dB, the time-domain network (TDN) had a similar performance as the multi-domain fusion network (MDFN). As the noise level was increased, MDFN outperformed either TDN or FDN networks. When the noise SNR is -6 dB, the accuracy of MDFN dropped only to 86.34% while the accuracy of TDN dropped at a higher rate to 80.69%. From the results, it can be observed that extracting and fusing features across different domains helps in capturing comprehensive bearing fault information, which could not be obtained by using only one domain.

When including the attention block in the MDFN model, the overall performance was enhanced significantly. For all noise conditions, the multi-domain fusion network with SE attention (MDFN+SE) exhibited the highest accuracy compared to other model variants, as shown in Table 4. From this analysis, it can be concluded that an attention mechanism enabled the model to focus on appropriate components in the multi-domain features, with the aim of improving the diagnostic performance in a noisy environment.

From this analysis, it can be concluded that an attention mechanism enabled the model to focus on appropriate parts in the multi-domain features to improve the diagnosing performance in a noisy environment.

D. COMPARATIVE STUDY

The proposed model is also compared with the following similar models that used the same dataset of CWRU.

- WDCNN [15] is a 1D CNN model whose input is a time-domain vibration signal and which is often used as a comparison model in studies of bearing diagnostics.
- MCNN-LSTM [19] is a composite of 1D CNN and LSTM model, in which the CNN extracts high-frequency and low-frequency components of the vibration signal with multi-scale kernels.
- FD 1D-CNN [36] is a 1D CNN using both the frequency spectrum and the phase information of bearing vibration signals. The model that was trained without noise was used in the analysis.
- TICNN [49] is a 1D CNN similar to WDCNN, but it employed several additional algorithms to diagnose the bearing states under noisy conditions.

Although all the compared models used the same CWRU dataset, each model selected different data subsets of motor loads. The trained and tested dataset of motor loads were indicated in Table 6. FD 1D-CNN [36] showed comparatively high performance for all noise levels. Under the noise SNR of -6 dB, the accuracy decreased to 84.87%. Under the strong noise level, the proposed model could maintain high accuracy that outperform other methods. In addition, the proposed model is moderately light-weighed in terms of parameter numbers but could achieved the highest accuracy to all noise levels.

E. t-SNE ANALYSIS: FEATURE VISUALIZATION

The effectiveness of feature extractions for diagnostic performance can be visualized and analyzed using t-distributed stochastic neighbor embedding (t-SNE). t-SNE is a method that projects a high-dimensional feature map to a two-dimensional map while preserving the significant relationships among the features. The distance between the

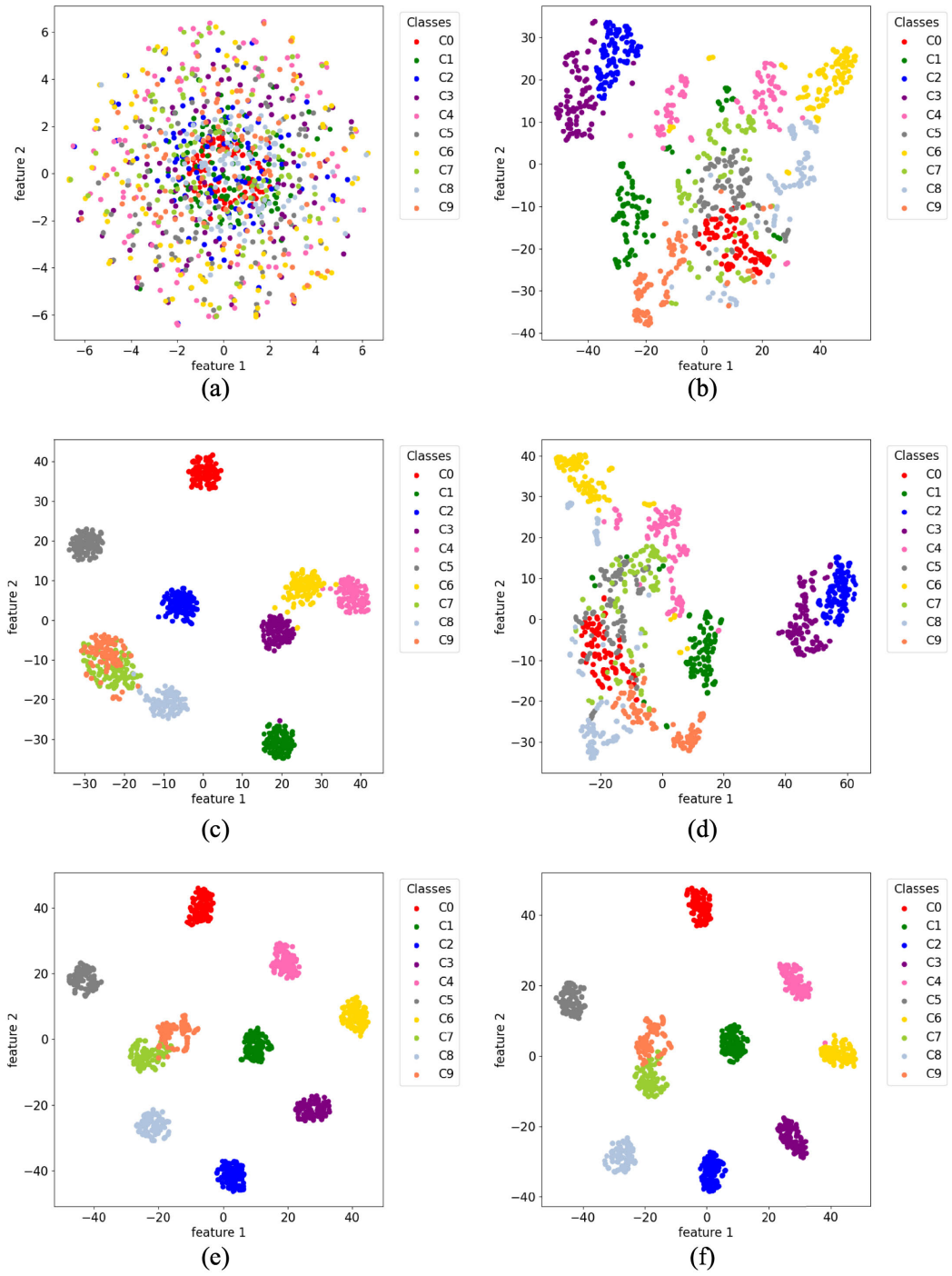

FIGURE 8. Feature visualization of the proposed model (MDFN+SE) using t-SNE with noise SNR of −6: (a) first convolutional layer of TD module; (b) first convolutional layer of FD module; (c) second convolutional layer of TD module; (d) second convolutional layer of FD module; (e) multi-domain fusion layer, (f) SE layer.

features in the plot indicates the similarity and separability among the features of the fault classes.

The proposed model was designed to extract significant and relevant fault features along the network layers, to ensure high performance in fault-type classifications. Thus, at the last layer of the network, the features of the same class on the t-SNE plot should be clustered tightly while remaining distinguishable from other inter-class features.

The t-SNE analysis was applied to the MDFN+SE model against the test dataset with an added noise level (SNR = −6 dB). Fig. 8 shows how the distribution of the extracted features is changed from the first to the last layer

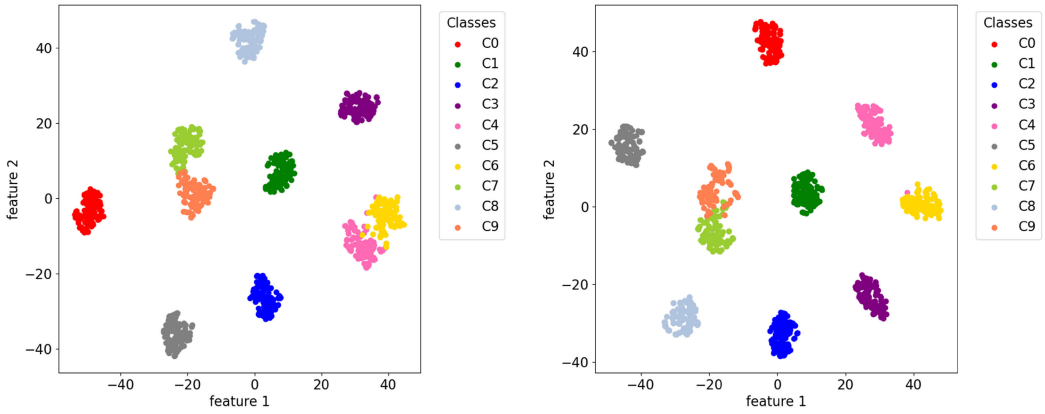

FIGURE 9. t-SNE: dimension-reduced features extracted from the multi-domain fusion layer of the proposed model (a) without attention (b) and with attention.

of the MDFN+SE model. The initial feature distribution at the input stage in the time-domain module (Fig. 8 (a)) and in the frequency-domain module (Fig. 8 (b)) are all interspersed, with low separability. As the convolution operation progresses, the same-class features in each time and frequency domain are clustered closer compared to the first convolutional layers, as shown in Fig. 8 (c) and (d), respectively. Nevertheless, these features are insufficient to make an accurate diagnosis.

After the attention block and multidomain fusion module in Fig. 8 (e)∼(f), the boundary between classes of features is clearly visible for all classes that explain the high accuracy performance even with strong noised signals.

After introducing the attention block and the multi-domain fusion module, as shown in Fig. 8 (e)∼(f), the inter-class feature boundaries are more clearly visible with a much improved intra-class feature clustering. The distinct boundaries for all classes in the feature extraction performed by the proposed model explain the high accuracy performance, even with strong-noise signals.

The effectiveness of the attention for classification enhancement is visualized by comparing the t-SNE plot for the proposed model with and without the attention block, as shown in Fig. 9. The t-SNE plot at the multi-domain fusion layer indicates that applying the attention module yielded a slightly higher degree of separation and clear boundaries among different classes. In particular, the distance between classes C0 and C5 increased and the boundary between classes C4 and C6 became more distinctive. Therefore, it can be concluded that the attention module is effective in improving the robustness and accuracy of the model in noisy environment applications.

V. CASE STUDY 2: PADERBORN DATASET

A. DATASET PREPARATION

For the second case study, the bearing vibration dataset provided by Paderborn University (PU) [50] was analyzed.

TABLE 7. Description of the Paderborn(PU) dataset for case study 2.

Class no.	Bearing code	Fault location	Fault level	No. of samples
C0	K001	Normal	-	1260
C1	KI01	Inner race fault	1	1260
C2	KI18	Inner race fault	2	1260
C3	KA01	outer race fault	1	1260
C4	KA15	outer race fault	1	1260

This dataset consists of acceleration data acquired at 64 kHz from a bearing mounted on a rotating machine driven by a 425 W motor rotating at 1500 rpm; its specification is as follows:

- Nominal torque: 1.35 [Nm]
- Nominal speed: 3, 000 [rpm]
- Nominal current: 2.3 [A]
- Pole pair: 4

The bearing used in this machinery is the 6203 ball-bearing type. The PU dataset has the following varied operating conditions: (1) load torque: 0.1 or 0.7 [Nm], (2) Radial force: 400 or 1000 [N]. Bearing fault types used in this study are summarized in Table 7. The PU dataset was downsampled by 1/5 since its sampling frequency is five times larger than the CWRU dataset. The total size of the dataset is 10260. It was split in a ratio of 6:2:2 and used for training, validation, and testing, respectively. In addition, similar signal preprocessing to that discussed in Section II was applied, based on the physics of the bearing fault frequencies using Eq. 1 and 2. The kernel size of the PI module in the FD module was modified according to the specifications of the bearing used in the experiment.

B. RESULT ANALYSIS

The model was trained in a noise-free environment and tested in environments under white Gaussian noise by using the same procedure as explained above. Similarly, the model

TABLE 8. Classification accuracy under various noise conditions of different model variants with PU dataset [%].

Model	SNR[dB]							
	None	6	4	2	0	-2	-4	-6
TDN	98.89	99.76	99.92	99.92	**100**	99.92	89.84	54.37
FDN	**100**	**100**	**100**	99.92	98.97	92.94	75.0	55.0
MDFN	99.44	**100**	**100**	**100**	**100**	**100**	98.1	73.1
MDFN-SE	**100**	**100**	**100**	**100**	**100**	**100**	**98.65**	**85.16**

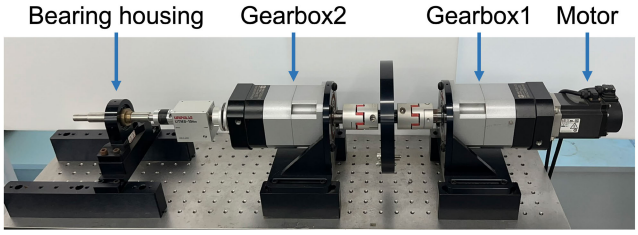

FIGURE 10. Testbench of bearing fault diagnosis used in case study 3.

TABLE 9. Description of our testbench dataset for case study 3.

Class no.	Bearing spec.	Fault location	Fault size	No. of samples
C0	7004	Normal	1 [*mm*]	2100
C1	7004	Inner race fault	1 [*mm*]	2100
C2	7004	Outer race fault	1 [*mm*]	2100

was evaluated with the noise-added test data with different SNR levels from −6 dB to 6 dB. Since the other models did not apply the PU dataset in their studies, we evaluated the effectiveness of MDFN+SE with the backbone network variants of TDN, FDN, and MDFN.

The test results are summarized in Table 8. Without noise addition, MDFN+SE and other variant models showed similar performance. As the noise level increased, the superiority of the proposed model was clearly demonstrated, achieving the highest accuracy for all noise level ranges. Under the noise level of −6dB, the accuracy of the proposed model dropped only to 85.16%, while MDFN dropped the accuracy to 73.1 %.

VI. CASE STUDY 3: LAB-MADE BEARING DIAGNOSIS TESTBENCH DATASET

A. DATASET PREPARATION

To evaluate the generality of the proposed method (MDFN+SE), another experiment is conducted using the dataset from our bearing fault diagnosis testbed, as shown in Fig. 10. This testbed is designed to emulate a small-scaled wind turbine and it is composed of a motor, two gearboxes, and a bearing box. The motor power is 1.4 [*kW*] with a rotating speed of 1200 [*rpm*]. The first gearbox reduces the speed by 1/100 ratio, and then the second gearbox increases the speed by a factor of 50. The shaft connected to the bearing is rotated at a rate of 600 rpm, and the vibration on the bearing housing is collected at a sampling rate of 10 [*kHz*]. The bearing specifications and the bearing fault classes are explained in in Table 9. The artificial faults on the bearing inner race and outer race are generated by electric discharge machining with a fault size of 1 mm.

B. RESULT AND ANALYSIS

All the compared models were trained in a noise-free environment and were evaluated with the noise-added test data with different SNR levels from −6 dB to 6 dB. We evaluated the effectiveness of multi-domain fusion with attention by comparing the performance with the backbone network variants of TDN, FDN, and MDFN. All models diagnosed the bearing states with near 100% accuracy for low noise levels less than 0dB, as shown in Table 10. As the noise SNR increased, using only a single domain(TDN, FDN) started to show much higher degradation compared to the multi-domain network (MDFN). Under the noise level of −6dB, the accuracy of MDFN is 81.88 %, while FDN degraded to 77.29%. The results showed that multi-domain based network outperforms a single-domain network for bearing conditions. Employing attention mechanism as in MDFN+SE, further improved the accuracy from 81.88 % to 94.38 % under the noise SNR of −6dB. In accordance with previous case studies, the proposed model of MDFN+SE showed the highest robustness performance under all noise conditions by having the least accuracy degradation rate.

VII. CONCLUSION

In this study, we developed a noise-robust bearing diagnosis model for practical industrial implementation in a heavy noise environment. The proposed model is based on a time-frequency multi-domain fusion 1D CNN model with the Squeeze-and-Excitation (SE) attention block. A high accuracy in classifying the bearing fault class and the severity (fault sizes) under a strong noise level could be achieved by designing a time-frequency multi-domain fusion block, incorporating bearing-fault physics into the model parameters, and employing attention modules. By extracting features in the time domain, in the frequency domain, and in the fused time-frequency domain, a comprehensive set of essential information relating to bearing faults and sizes could be obtained. Furthermore, applying the bearing fault physics in designing the appropriate kernel size in the CNN model and adding the channel-wise attention mechanism further boosted the performance of the proposed model. The effectiveness

TABLE 10. Classification accuracy under various noise conditions of different model variants with our testbench dataset [%].

Model	SNR[dB]							
	None	6	4	2	0	-2	-4	-6
TDN	99.79	**100**	**100**	**100**	**100**	99.79	97.5	76.46
FDN	**100**	**100**	**100**	**100**	**100**	99.58	95.0	77.29
MDFN	**100**	**100**	**100**	**100**	**100**	**100**	96.25	81.88
MDFN-SE	**100**	**100**	**100**	**100**	**100**	**100**	**99.58**	**94.38**

of the proposed model was validated on two datasets with artificial noise added. Under the most substantial noise level of SNR −6 dB on the CWRU dataset, the propoed model maintained a high accuracy of 93.27%. Compared to similar bearing diagnosis models, the proposed model exhibited the highest noise robustness for all noise levels.

This study applied lab-environment datasets with added Gaussian white noise, which may not represent the actual industrial environment. Thus, for future work, fault-bearing vibration signals from industrial sites should be collected and analyzed to verify the practical implementations.

REFERENCES

[1] A. Choudhary, T. Mian, and S. Fatima, "Convolutional neural network based bearing fault diagnosis of rotating machine using thermal images," *Measurement*, vol. 176, May 2021, Art. no. 109196.

[2] B. Van Hecke, J. Yoon, and D. He, "Low speed bearing fault diagnosis using acoustic emission sensors," *Appl. Acoust.*, vol. 105, pp. 35–44, Apr. 2016.

[3] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.

[4] L. Guo, J. Chen, and X. Li, "Rolling bearing fault classification based on envelope spectrum and support vector machine," *J. Vibrat. Control*, vol. 15, no. 9, pp. 1349–1363, Sep. 2009.

[5] M. Kedadouche and Z. Liu, "Fault feature extraction and classification based on WPT and SVD: Application to element bearings with artificially created faults under variable conditions," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 231, no. 22, pp. 4186–4196, Nov. 2017.

[6] V. K. Rai and A. R. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2607–2615, Aug. 2007.

[7] S. Sassi, B. Badri, and M. Thomas, "A numerical model to predict damaged bearing vibrations," *J. Vibrat. Control*, vol. 13, no. 11, pp. 1603–1628, Nov. 2007.

[8] S.-D. Wu, C.-W. Wu, T.-Y. Wu, and C.-C. Wang, "Multi-scale analysis based ball bearing defect diagnostics (don't short) using Mahalanobis distance and support vector machine," *Entropy*, vol. 15, no. 2, pp. 416–433, Jan. 2013.

[9] L. Shuang and L. Meng, "Bearing fault diagnosis based on PCA and SVM," in *Proc. Int. Conf. Mechatronics Autom.*, Aug. 2007, pp. 3503–3507.

[10] S. Dong, T. Luo, L. Zhong, L. Chen, and X. Xu, "Fault diagnosis of bearing based on the kernel principal component analysis and optimized K-nearest neighbour model," *J. Low Freq. Noise, Vibrat. Act. Control*, vol. 36, no. 4, pp. 354–365, Dec. 2017.

[11] C. P. Mbo'o and K. Hameyer, "Fault diagnosis of bearing damage by means of the linear discriminant analysis of stator current features from the frequency selection," *IEEE Trans. Ind. Appl.*, vol. 52, no. 5, pp. 3861–3868, Sep. 2016.

[12] M. Hamadache, J. H. Jung, J. Park, and B. D. Youn, "A comprehensive review of artificial intelligence-based approaches for rolling element bearing PHM: Shallow and deep learning," *JMST Adv.*, vol. 1, nos. 1–2, pp. 125–151, Jun. 2019.

[13] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.

[14] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier," *J. Signal Process. Syst.*, vol. 91, no. 2, pp. 179–189, Feb. 2019.

[15] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017.

[16] X. Song, Y. Cong, Y. Song, Y. Chen, and P. Liang, "A bearing fault diagnosis model based on CNN with wide convolution kernels," *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 8, pp. 4041–4056, Aug. 2022.

[17] W. Huang, J. Cheng, Y. Yang, and G. Guo, "An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis," *Neurocomputing*, vol. 359, pp. 77–92, Sep. 2019.

[18] J. He, P. Wu, Y. Tong, X. Zhang, M. Lei, and J. Gao, "Bearing fault diagnosis via improved one-dimensional multi-scale dilated CNN," *Sensors*, vol. 21, no. 21, p. 7319, Nov. 2021.

[19] X. Chen, B. Zhang, and D. Gao, "Bearing fault diagnosis base on multi-scale CNN and LSTM model," *J. Intell. Manuf.*, vol. 32, no. 4, pp. 971–987, Apr. 2021.

[20] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," *Mech. Syst. & Signal Process.*, vol. 25, no. 2, pp. 485–520, 2011.

[21] S. F. A. Note, "Rolling element bearings," REB, Sales Technol., League City, TX, USA, Tech. Rep., 2012.

[22] D. Wang, Q. Guo, Y. Song, S. Gao, and Y. Li, "Application of multiscale learning neural network based on CNN in bearing fault diagnosis," *J. Signal Process. Syst.*, vol. 91, no. 10, pp. 1205–1217, Oct. 2019.

[23] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 8, pp. 1926–1935, Aug. 2017.

[24] H.-Y. Chen and C.-H. Lee, "Vibration signals analysis by explainable artificial intelligence (XAI) approach: Application on bearing faults diagnosis," *IEEE Access*, vol. 8, pp. 134246–134256, 2020.

[25] D. Zhong, W. Guo, and D. He, "An intelligent fault diagnosis method based on STFT and convolutional neural network for bearings under variable working conditions," in *Proc. Prognostics Syst. Health Manag. Conf. (PHM-Qingdao)*, Oct. 2019, pp. 1–6.

[26] T. Huang, S. Fu, H. Feng, and J. Kuang, "Bearing fault diagnosis based on shallow multi-scale convolutional neural network with attention," *Energies*, vol. 12, no. 20, p. 3937, Oct. 2019.

[27] D. Neupane, Y. Kim, and J. Seok, "Bearing fault detection using scalogram and switchable normalization-based CNN (SN-CNN)," *IEEE Access*, vol. 9, pp. 88151–88166, 2021.

[28] X. Liu, W. Sun, H. Li, Z. Hussain, and A. Liu, "The method of rolling bearing fault diagnosis based on multi-domain supervised learning of convolution neural network," *Energies*, vol. 15, no. 13, p. 4614, Jun. 2022.

[29] K. Dong and A. Lotfipoor, "Intelligent bearing fault diagnosis based on feature fusion of one-dimensional dilated CNN and multi-domain signal processing," *Sensors*, vol. 23, no. 12, p. 5607, Jun. 2023.

[30] J. Sun, J. Wen, C. Yuan, Z. Liu, and Q. Xiao, "Bearing fault diagnosis based on multiple transformation domain fusion and improved residual dense networks," *IEEE Sensors J.*, vol. 22, no. 2, pp. 1541–1551, Jan. 2022.

[31] *Case Western Reserve University Bearing Data Center*. Accessed: Jan. 22, 2024. [Online]. Available: https://engineering.case.edu/bearingdatacenter

[32] C. Yang, Z. Qiao, R. Zhu, X. Xu, Z. Lai, and S. Zhou, "An intelligent fault diagnosis method enhanced by noise injection for machinery," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[33] X. Li, W. Zhang, and Q. Ding, "A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning," *Neurocomputing*, vol. 310, pp. 77–95, Oct. 2018.

[34] X. Li, W. Zhang, Q. Ding, and X. Li, "Diagnosing rotating machines with weakly supervised data using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1688–1697, Mar. 2020.
[35] H. Fang, J. Deng, Y. Bai, B. Feng, S. Li, S. Shao, and D. Chen, "CLFormer: A lightweight transformer based on convolutional embedding and linear self-attention with strong robustness for bearing fault diagnosis under limited sample conditions," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–8, 2022.
[36] M. Hakim, A. A. B. Omran, J. I. Inayat-Hussain, A. N. Ahmed, H. Abdellatef, A. Abdellatif, and H. M. Gheni, "Bearing fault diagnosis using lightweight and robust one-dimensional convolution neural network in the frequency domain," *Sensors*, vol. 22, no. 15, p. 5793, Aug. 2022.
[37] H. Wang, J. Xu, R. Yan, C. Sun, and X. Chen, "Intelligent bearing fault diagnosis using multi-head attention-based CNN," *Proc. Manuf.*, vol. 49, pp. 112–118, Jan. 2020.
[38] D. Peng, H. Wang, W. Desmet, and K. Gryllias, "RMA-CNN: A residual mixed-domain attention CNN for bearings fault diagnosis and its time-frequency domain interpretability," *J. Dyn., Monitor. Diag.*, vol. 2, no. 2, pp. 115–132, Apr. 2023.
[39] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[40] J. Yin and G. Cen, "Intelligent motor bearing fault diagnosis using channel attention-based CNN," *World Electric Vehicle J.*, vol. 13, no. 11, p. 208, Nov. 2022.
[41] H. Zhong, Y. Lv, R. Yuan, and D. Yang, "Bearing fault diagnosis using transfer learning and self-attention ensemble lightweight convolutional neural network," *Neurocomputing*, vol. 501, pp. 765–777, 2022.
[42] S. Shen, H. Lu, M. Sadoughi, C. Hu, V. Nemani, A. Thelen, K. Webster, M. Darr, J. Sidon, and S. Kenny, "A physics-informed deep learning approach for bearing fault detection," *Eng. Appl. Artif. Intell.*, vol. 103, Aug. 2021, Art. no. 104295.
[43] M. Sadoughi and C. Hu, "Physics-based convolutional neural network for fault diagnosis of rolling element bearings," *IEEE Sensors J.*, vol. 19, no. 11, pp. 4181–4192, Jun. 2019.
[44] H. Lu, V. P. Nemani, V. Barzegar, C. Allen, C. Hu, S. Laflamme, S. Sarkar, and A. T. Zimmerman, "A physics-informed feature weighting method for bearing fault diagnostics," *Mech. Syst. Signal Process.*, vol. 191, May 2023, Art. no. 110171.
[45] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019.
[46] S. Poddar and M. L. Chandravanshi, "Ball bearing fault detection using vibration parameters," *Int. J. Eng. Res. Technol.*, vol. 2, no. 12, pp. 1239–1244, 2013.
[47] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.
[48] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[49] 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.
[50] *Paderborn University Bearing Data Center*. Accessed: Jan. 22, 2024. [Online]. Available: https://mb.uni-paderborn.de/kat/forschung/kat-datacenter/bearing-datacenter/data-sets-and-download

YEJIN KIM received the B.E. degree in mechanical engineering from Handong Global University, Pohang, South Korea, in 2021. Her research interests include prognostics and health management for rotationary machines and deep learning for industrial AI.

YOUNG-KEUN KIM (Member, IEEE) received the B.E. degree from Handong Global University, Pohang, South Korea, in 2008, and the M.Sc. (Eng.) and Ph.D. degrees in mechanical engineering from the Korea Advanced Institute of Science & Technology (KAIST), Daejeon, South Korea, in 2010 and 2014, respectively. Since 2014, he has been with Handong Global University. Currently, he is an Associate Professor and the Head of the Industrial Intelligence Laboratory, Department of Mechanical and Control Engineering. His research interests include industrial AI, 2D image machine vision, and 3D LiDAR signal processing.