

Physics-Informed Residual Network (PIResNet) for rolling element bearing fault diagnostics

Qing Ni^a, J.C. Ji^a, Benjamin Halkon^a, Ke Feng^{b,*} and Asoke K. Nandi^c

^a*School of Mechanical and Mechatronic Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia*

^b*School of Engineering, University of British Columbia, Kelowna, BC V1V 1V7, Canada*

^c*Department of Electronic and Electrical Engineering, Brunel University London, Uxbridge, UB8 3PH, United Kingdom*

ARTICLE INFO

Keywords:

deep learning
bearing fault diagnostics
speed and load variations
Physics-Informed Residual Network
modal-property-dominant-generated
layer

ABSTRACT


Various deep learning methodologies have recently been developed for machine condition monitoring recently, and they have achieved impressive success in bearing fault diagnostics. Despite the capability of effectively diagnosing bearing faults, most deep learning methods are tremendously data-dependent, which is not always available in industrial applications. In practical engineering, bearings are usually installed in rotating machinery where speed and load variations frequently occur, resulting in difficulty in collecting large training datasets under all operating conditions. Additionally, physical information is usually ignored in most deep learning algorithms, which sometimes leads to the generated results of low compliance with the physical law. To tackle these challenges, a novel Physics-Informed Residual Network (PIResNet) is proposed for learning the underlying physics that is embedded in both training and testing data, thus providing a physical consistent solution for imperfect data. In the proposed method, a physical modal-property-dominant-generated layer is adopted at first to generate the modal-property-dominant feature. Then, a domain-conversion layer is constructed to enable the feasibility of extracting the discriminative bearing fault features under varying operating speed conditions. Lastly, a parallel bi-channel residual learning architecture that can automatically extract the bearing fault signatures is meticulously established to incorporate the bearing fault characteristics. Experimental datasets under variable operating speeds and loads, and time-varying operating speeds are utilized to demonstrate the superiority of the PIResNet under non-stationary operating conditions.

1. Introduction

An ever-increasing demand for the reliability and safety of machinery has been driven in the modern industry due to the requirement for automation, integration, and precision for industrial innovation. Bearings are crucial components in rotating machinery, which are commonly used in various modern industrial applications (some examples of bearing's applications are shown in Figure 1) owing to their ability to support the shafts, constrain relative motion, and alleviate friction between rotating components. In practice, the degradation and failure of bearings are inevitable as a result of continuous long operating duration, improper maintenance, suffering dynamic stress induced by the dynamic loads of rotating parts, etc. Thus, bearings are one of the most frequent failure components in rotating machinery. For example, according to a motor reliability survey, approximately 40% failure of the induction motor originates from the bearings faults [1]. And in one statistical report conducted by the National Renewable Energy Laboratory of America, the bearing faults account for 76.2 % of wind turbine gearbox failures [2]. The deterioration or failure of bearings can result in unscheduled shutdowns of the whole mechanical system, leading to enormous costs due to productivity loss, maintenance, or more seriously, fatal accidents. Therefore, it is of great significance for modern industry health management to accurately diagnose bearing faults.

Generally, bearing fault diagnosis is realized through three steps: data acquisition, feature extraction, and fault detection and/or isolation, among which feature extraction has garnered the greatest attention. Classical feature extraction is usually achieved through signal processing approaches, primarily utilizing the time domain, frequency domain, and time-frequency domain techniques. Time domain feature extraction is intuitive and often realized through statistical algorithms. Typical time domain extracted indicators include mean value, mean square value, root mean square value, variance value, peak value, peak-to-peak value, impulse factor, crest factor, shape factor, skewness,

*Corresponding author

 ke.feng@outlook.com.au (K. Feng)

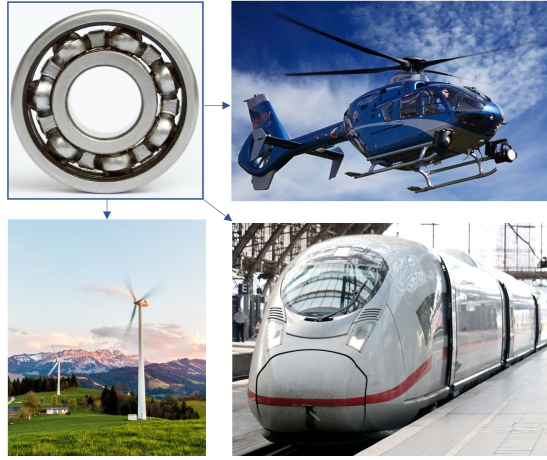


Figure 1: Examples of bearing's applications

kurtosis, higher-order cumulants, etc. [3]. Frequency domain feature extraction is essentially achieved through the Fourier Transform (FT), generating frequency domain-based indicators from the raw spectrum, power spectrum, holographic spectrum, cepstrum, (squared) envelope spectrum [4], and some improved spectra such as correntropy-derived energy spectrum [5]. Time-frequency feature extraction techniques have the capability of simultaneously expressing the joint distribution information of the time and frequency domains. Widely utilized time-frequency feature extraction techniques include Short-time Fourier Transform, Winger Ville distribution, Wavelet Transform, synchrosqueezing transform, etc. [6].

Based on the extracted feature, the bearing healthy status and fault locations can be diagnosed by professional knowledge derived from the bearing kinetics such as the amplitude of fault characteristic frequency (FCF). Besides incorporating fault-mechanism-based professional knowledge, machine learning-based methodologies, including k-nearest neighbours, random forest, decision tree, support vector machine, etc., have increasingly been focused for bearing fault detection and/or isolation [7]. These traditional machine learning methods typically require to perform feature engineering (feature extraction) at first. However, different from the traditional professional-knowledge-derived methods, they can establish beyond simple linear relationships to reflect the bearing status by several uncomplicated existing features through which the bearing status sometimes can not be straightforwardly determined. In addition to the aforementioned traditional machine learning methods, deep learning methods have evolved dramatically as a new branch of machine learning. Deep learning generically refers to algorithms that adopt hierarchical architectures, in which layers are usually higher than three, to learn discriminative features from intricate high-dimensional input data [8]. The low/no requirement for manual feature engineering makes deep learning algorithms distinguishable from traditional machine learning methods. Typical deep learning architectures include Deep Neural Network, Deep Belief Network, Convolutional Neural Network (CNN), Recurrent Neural Network, and their variants [9]. For example, in [10], a convolutional deep belief network-based method was proposed to implement fault diagnosis of machinery, in which the diagnostic performance improvement was achieved based on multimodal data isolation and automatic feature extraction.

Despite its impressive success in bearing fault diagnostics, the underlying assumption that there is no distribution discrepancy between the training and testing data can inhibit the further application of deep learning in complicated practical industries. The bearings are usually utilized in critical industrial components, such as gearbox, motor, and machine tool principal axis, which frequently operate under speed and load variations. Therefore, it is difficult to collect training data under all operating conditions, leading to the discrepancy between the training and testing data distributions in some situations. Transfer learning has been demonstrated to provide a feasible solution for this challenge by learning the common knowledge embedded in one or more scenarios that are relevant but not identical to the target data [9]. In [11], a supervised contrastive learning-based transfer learning method was proposed to obtain class-specific knowledge. With the proposed supervised contrastive loss, the classification accuracy for the target domain where distribution differences exist was vastly improved. Feng et al. [12] developed a transfer learning algorithm to leverage the information learned from the established digital twin model to the real industry structure, achieving the transmission

system degradation assessment with high accuracy and practicality. In [13], the marginal distribution adaptation was extended to the joint distribution adaptation to achieve domain adaptation. Experimental results demonstrated that the proposed method can better match the real distribution compared with conventional maximum mean discrepancy.

Although being able to deal with the distribution discrepancy challenge, transfer learning still does not incorporate physical knowledge, leading to difficulty in producing physically interpretable results and sometimes generating results against physical law. In contrast to pure data-driven methodologies, the Physics-Informed deep learning methods provide a promising alternative that embeds the prior understanding derived from observation, experience, physics, and mathematics to the deep learning architecture [14]. The integration of physical information contributes to the deep learning model learning not only rules from the existing observational (training) data but also fundamental physical or domain knowledge from all data (including the testing data), thus providing a physical consistent solution for imperfect data (such as training and testing data collected from discrepant operating conditions, noisy data, etc.). Physics-Informed deep learning has attracted exploded attention since it was proposed. One of the most pioneered applications might be solving nonlinear partial differential equations. In research [15], Raissi et al. integrated the underlying physical law into loss function, successfully constructing a computationally efficient neural network that can deal with partial differential equations. In addition, Physics-Informed deep learning concepts have been studied and embedded in various applications such as the augmenting turbulence models [16], power system [17], weather and climate modelling [18], radiative transfer [19], topology optimization [20], etc.

Notwithstanding the ever-changing applications of Physics-Informed deep learning, the realization of embedding physical knowledge into deep learning architecture relies on introducing three biases, namely observational biases, inductive biases, and learning bias [14]. Observational biases could be introduced by embedding underlying physics into the data. It might be the most intuitive approach to incorporate physics information into the network for bearing diagnosis. To improve the generalization ability, the authors in research [21] preprocessed the collected signal based on the rotational speed and geometry of fault bearing. Through resampling the signal from the time domain to the angle domain using rotational frequency, ball pass frequency of inner race (BPFI), ball pass frequency of outer race (BPFO), and ball spinning frequency (BSF), as references respectively, and then inputting the envelope of the resampled signal into a deep learning network, the proposed method was able to deal with cross-domain bearing diagnostic issues.

The introduction of inductive biases refers to embedding the physical information into deep learning architecture, usually in the form of customized layers. Sadoughi and Hu [22] developed a physics-based CNN by integrating several physics-based customized layers. Taking advantage of the impulsiveness, cyclostationarity, and FCF, a Spectral kurtosis-based layer was established for enhancing impulsiveness, an envelope-based layer was constructed for demodulating, a physics-based convolutional layer was established for extracting local similarities between FCF-based convolutional kernels and input, and a Fast Fourier Transformation (FFT)-based layer was constructed for frequency-domain transforming in this paper. In research [23], a continuous wavelet convolutional layer was embedded to replace a deep CNN's first layer. The proposed continuous wavelet convolution-based layer enabled the extraction of scale and translation parameters from raw data, thus generating a physically interpretable network for bearing diagnostics. Lu et al. [24] proposed a Physics-Informed bearing fault diagnostic network by introducing a weighting layer. In the proposed method, the envelope of the signal was calculated and then transformed to the order domain. This was followed by the development of a Physics-Informed feature weighting layer, which assigned higher weights to features that are closer to the FCFs. Utilizing the weighting layer, the proposed network was proved to be sensitive to the bearing faults while less affected by speed variations. Furthermore, algorithm unrolling, an emerging methodology that unfolds iterative algorithms of a pre-established model to network architecture and executes the algorithm's parameters determination process as the network training rather than analytical derivations or cross-validation, might provide a promising alternative for integrating the physical information into the network. For example, an adversarial algorithm unrolling network (AAU-Net) was proposed for interpretable mechanical anomaly detection in [25]. Based on the prior knowledge of the mechanical component's distribution characteristics, a sparse coding model was constructed in the AAU-Net for feature encoding and decoding, enabling the proposed network mechanism-driven and interpretable. Experimental studies demonstrated that AAU-Net can effectively learn signal features that match the dynamic mechanism of the mechanical system.

Learning bias could be embedded by introducing algorithms that can constrain the convergence of training progress towards the direction with specific hidden physical information. The customized loss function is usually employed for this purpose, through which the underlying physical constraints could be approached. In research [26], a customized loss function based on the cross-entropy between the detection result of one threshold model and the real label was proposed and integrated into a deep CNN. The proposed threshold model employed the order spectrum's amplitude at

FCF and its harmonics, aiming to distinguish the heavy, light, and healthy conditions by amplitude threshold. Utilizing the customized loss function, the proposed Physics-Informed deep learning methodology was able to alleviate the misclassification for different fault severity. To tackle the issue that Generative Adversarial Network (GAN) merely takes the overall similarity between the input data and generated data into consideration during the data generation process, a loss function based on the amplitudes and frequencies of real and generated data was established in [27]. Upon utilizing the frequencies at FCF and its first five harmonics, the GAN was able to generate data based on the similarity of fault characteristics. Adopting the theory that the remaining useful life of bearing follows Weibull distribution, a Weibull-based loss function was proposed in research [28]. By calculating the mean squared error, root mean squared error, and root mean squared log error between the real and estimated cumulative distribution value of a two-parameter Weibull distribution function at each time, the proposed network could generate outputs whose distribution aligned with domain knowledge. In addition, Physics-Informed concepts can be employed in other processes of deep learning for better diagnostic accuracy. For example, based on the fault mechanism, Chen et al. [29] proposed a Physics-Informed hyperparameters selection strategy for a long short-term memory network, in which the discrepancy between the faulty and healthy states was maximized, thus enabling the construction of a high detection accuracy network with physical consistency.

Although the Physics-Informed deep learning network is gradually becoming a promising option to address the issue that bearing training and testing data under different operating conditions. Challenges to establish such a network, such as extracting the physical information that is robust under different operating conditions (the operating speeds and loads in this research) and embedding this physical information into the network architecture, still exist. To this end, a Physics-Informed deep learning network incorporating the modal properties is proposed in this paper. This paper has the following contributions:

- A physical modal-property-dominant-generated layer is established based on cepstrum exponential filtering [30], which enables the modal properties to dominate in the processed signal. The modal properties are only related to the system properties, indicating that they are not sensitive to the changes in the operating conditions but sensitive to the changes of the system, possibly the variations from healthy states to faulty states or variations of fault locations, thus enabling the feasibility to establish network work robustly under varying operating conditions.
- A domain-conversion layer is established based on computed order tracking (COT) to transform the time domain signal into the angle domain signal, which is not tied to the rotational speed. The domain-conversion layer, therefore, is able to alleviate the influence brought by speed variations and thus extract the discriminative bearing fault features under variable or time-varying speed conditions.
- A parallel bi-channel residual learning architecture is established after the first two layers. The two channels can automatically extract the intricate high-dimensional bearing fault characteristics from the raw vibration and modal-property-dominant signal parallelly, which can improve the diagnostic accuracy when the distribution discrepancy exists between the training and testing data.

The subsequent sections of this paper are structured as follows. Section 2 presents the methodology of the proposed Physics-Informed parallel bi-channel Residual Network. Section 3 includes two experimental validations under variable and time-varying operating conditions to validate the effectiveness of the proposed method. Section 4 summarizes the work completed in this research.

2. Methodology

2.1. Convolutional Neural Network and Residual Network

Among various deep learning methodologies, CNN has been extensively studied and utilized in bearing condition monitoring, owing to its unique advantage of automatically extracting the discriminative bearing fault characteristics through convolution and pooling operations from raw data. This relieves people from manually feature extracting, rendering the possibility of decision-making for the on-site operator without professional bearing fault mechanism knowledge. Since the raw bearing vibration is usually one dimension, 1-dimensional (1D) CNN is adopted in this paper and will be briefly introduced below. The feature extraction of CNN is achieved mainly through the convolutional and pooling layers [31].

The convolutional layer utilizes multiple convolution kernels to extract the features, and the 1D convolutional layer employs 1D convolution kernels to convolute the 1D input. If there are N convolution kernels, the convolution will be

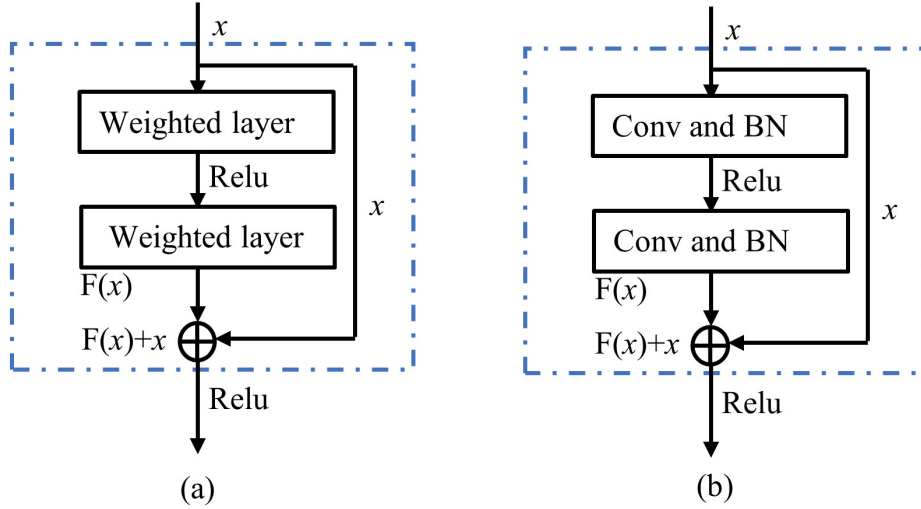


Figure 2: (a) A typical residual building block; (b) The residual building block used in this paper.

executed N times, and each convolution operation will use the same convolution kernel to convolute the input so as to generate one feature map. In the l^{th} layer, convolution operation for the i^{th} convolution kernel can be expressed as:

$$feature_{i,j}^l = \omega_i^l * feature_j^{l-1} + b_i^l \quad (1)$$

where ω_i^l and b_i^l represent the weight and bias matrix of the i^{th} convolution kernel in the l^{th} layer, respectively; $feature_j^{l-1}$ denotes the j^{th} output feature map in the $(l-1)^{th}$ layer; $feature_{i,j}^l$ denotes the j^{th} output feature map of the i^{th} convolution kernel in the l^{th} layer.

The pooling layer aims at condensing the size of the local feature and extracting the most important information. Commonly used pooling layers include the max pooling layer and average pooling layer. The max pooling layer is used in this paper. If the length of the pooling operator is s , the max pooling function can be expressed as:

$$\mathbf{p}_j^l = [p_j^{l,1}, p_j^{l,2}, \dots, p_j^{l,k}, \dots, p_j^{l,N}] \quad (2)$$

$$p_j^{l,k} = \max_{(k-1)s+1 \leq n \leq ks} \{feature_j^{l-1}(n)\} \quad (3)$$

where \mathbf{p}_j^l represents the output after the pooling operation of the j^{th} feature map in the l^{th} layer.

Residual Network, also known as ResNet, has gained significant interest in solving the gradient vanishing issue for deep learning [32]. The ResNet is a type of neural network architecture with residual building block (RBB) being the most essential element, which facilitates the direct transfer of features from the previous layer to the subsequent layers through shortcut connections. A typical RBB proposed in research [32] is shown in Figure 2(a), and the RBB adopted in this paper is depicted in Figure 2(b).

As shown in Figure 2(a), a RBB can be formulated as:

$$y = F(x) + x \quad (4)$$

where x and y symbolize the input vector and output vector of the RBB, respectively; $F(\cdot)$ stands for the residual function. The utilization of a shortcut connection x allows for a significant increase in the depth of the network. Additionally, as demonstrated in [32], the removal of a single layer has little impact on the overall testing performance of the residual network. Note that the $F(\cdot)$ and x are supposed to have the same dimension, if they are not in the same dimension, the shortcut x can be replaced by the alternative convolution layer of x to guarantee the same dimension.

2.2. Physical modal-property-dominant feature extraction

The bearing fault signals are typically a series of 'almost-periodic' impulses, which are essentially compound responses of the forcing function and transfer function. The forcing function repeats itself at a rate tied to the rotational speed but has a slight deviation from the deterministic FCF. The transfer function is related to the modal properties with fixed resonance frequencies which are independent of the magnitude and frequency of inputs but can result in the amplification of the inputs [33]. Therefore, the bearing fault information is carried and amplified by the resonance frequencies, and a modal-property-dominant-generated layer is designed to enable the modal properties that carry bearing fault information to be dominant. The modal-property-dominant-generated layer is derived from cepstrum exponential filtering[30], with specifics outlined below.

The cepstrum was initially developed for speech recognition. Owing to its capability to separate the forcing and transfer functions, it was subsequently used for operational modal analysis [34]. In general, the most commonly used cepstrum is the complex cepstrum and unless specifically mentioned, the cepstrum in the following paper refers to the complex cepstrum. One of the most widely utilized forms of cepstrum is given by:

$$\hat{y}_c(\tau) = \mathfrak{F}^{-1}[\log(\mathfrak{F}(y(t)))] = \mathfrak{F}^{-1}[\log(|Y(f)|) + j\angle Y(f)] \quad (5)$$

where $y(t)$ represents the time signal; $\hat{y}_c(\tau)$ represents the cepstrum of $y(t)$; $\mathfrak{F}^{-1}(\cdot)$ and $\mathfrak{F}(\cdot)$ denote the inverse FT (IFT) operator and FT operator, respectively; $|Y(f)|$ and $\angle Y(f)$ are the amplitude and phase of $Y(f)$, where $Y(f) = \mathfrak{F}[y(t)] = |Y(f)|e^{j\angle Y(f)}$. If the amplitude information is merely retained, the result $\mathfrak{F}^{-1}[\log(|Y(f)|)]$ will yield the 'real cepstrum', which is irreversible to the raw time signal.

For a single input and multiple outputs (SIMO) system, the response function is the convolution of the forcing function and transfer function in the time domain:

$$y(t) = x(t) * h(t) \quad (6)$$

where the $x(t)$ and $h(t)$ represent the forcing and transfer functions, respectively.

Performing FT for $y(t)$ yields:

$$Y(f) = \mathfrak{F}[x(t)] \times \mathfrak{F}[h(t)] \quad (7)$$

And if the logarithm operation is then executed for $Y(f)$, the result will be:

$$\log[Y(f)] = \log(\mathfrak{F}[x(t)]) + \log(\mathfrak{F}[h(t)]) \quad (8)$$

The cepstrum of $y(t)$ is the IFT of $\log[Y(f)]$ and can be formulated as:

$$\hat{y}_c(\tau) = \mathfrak{F}^{-1}[\log(Y(f))] = \mathfrak{F}^{-1}[\log(\mathfrak{F}(x(t)))] + \mathfrak{F}^{-1}[\log(\mathfrak{F}(h(t)))] = \hat{x}_c(\tau) + \hat{h}_c(\tau) \quad (9)$$

Hence, the cepstrum of the response signal $y(t)$ is the sum of the forcing function's cepstrum and the transfer function's cepstrum, thus allowing the separation of the two. Note that this is applicable for SIMO systems, in the case of multiple input and multiple output systems, the response function is a sum of convolutions, requiring signal processing techniques, such as blind source separation [35], to process the signal to a single excitation prior to any further processing [36]. Since this topic falls outside the scope of this paper, the relevant research will be conducted in further work.

In the case of sampled sequence, the general form of transfer function $h(n)$ can be represented in the z plane using the pole-zero model through [37]:

$$H(z) = \frac{C Z^r \prod_{k=1}^{N_{zi}} (1 - a_k z^{-1}) \prod_{k=1}^{N_{zo}} (1 - b_k z)}{\prod_{k=1}^{N_{pi}} (1 - c_k z^{-1}) \prod_{k=1}^{N_{po}} (1 - d_k z)} \quad (10)$$

where C is a constant scaling factor; N_{zi} , N_{zo} , N_{pi} , and N_{po} represent the number of zeros inside the unit circle, the number of zeros outside the unit circle, the number of poles inside the unit circle, and the number of poles outside the unit circle, respectively; a_k , $1/b_k$, c_k , and $1/d_k$ denote the zeros inside the unit circle, the zeros outside the unit circle, the poles inside the unit circle, and the poles outside the unit circle, respectively ($|a_k|, |b_k|, |c_k|, |d_k| < 1$). Replacing

the FT and IFT by z-transform and inverse z-transform, respectively, the analytical form of the cepstrum for a stable minimum phase system's transfer function can be formulated as [37]:

$$\hat{h}_c(n) = \begin{cases} 0 & n < 0 \\ \ln |C| & n = 0 \\ \sum_{k=1}^{N_{pi}} \frac{c_k^n}{n} - \sum_{k=1}^{N_{zi}} \frac{a_k^n}{n} & n > 0 \end{cases} \quad (11)$$

For $n > 0$, the $\hat{h}_c(n)$ can be expressed by the poles and zeros inside the unit circle. Replacing the poles and zeros with complex conjugate pairs in Equation (11) yields:

$$\hat{h}_c(n) = 2 \sum_{k=1}^{N_{pi}/2} \frac{A_{ck}^n}{n} \cos(\omega_{ck}n) - 2 \sum_{k=1}^{N_{zi}/2} \frac{A_{ak}^n}{n} \cos(\omega_{ak}n), n > 0 \quad (12)$$

where $A_{ck} = |c_k|$, $\omega_{ck} = \angle c_k$, $A_{ak} = |a_k|$, and $\omega_{ak} = \angle a_k$, respectively. In the case of a single-degree-of-freedom system with a pair of conjugate poles, based on Equation (12), the cepstrum of the transfer function can be expressed as [37]:

$$\hat{h}_c(n) = 2 \frac{A_{ck}^n}{n} \cos(\omega_{ck}n) = \frac{2}{n} e^{-\sigma_{ck}n\Delta t} \cos(\omega_{ck}n), n > 0 \quad (13)$$

where σ_{ck} (in rad/s) denotes half of the corresponding resonance peak's 3dB bandwidth, Δt represents the time resolution and $t = n \times \Delta t$. The cepstrum of the transfer function with respect to each pole thus has a similar form to the impulse response with the same pole but is scaled by an additional factor of $1/n$, indicating that the cepstrum of the transfer function decays at a rapid pace and thus tends to be located in the low quefrency area. It is worth noting that the zeros exhibit the similar form but with a negative sign, as they can be regarded as negative poles. As for the multi-degree-of-freedom system, the transfer function's cepstrum can be represented as the sum of the zeros and poles [30]. If the cepstrum of the response function is multiplied by an exponential window $e^{-\sigma_0 t}$, i.e.,

$$\hat{y}_c(n) \times e^{-\sigma_0 t} = [\hat{x}_c(n) + \hat{h}_c(n)] \times e^{-\sigma_0 t} = \hat{x}_c(n) \times e^{-\sigma_0 t} + \hat{h}_c(n) \times e^{-\sigma_0 t} \quad (14)$$

It is equivalent to low (quefrency)-pass filter the forcing function's cepstrum. While for modal properties. While for modal properties, it only adds a known damping σ_0 to each pole and zero for the cepstrum of the transfer function, which can be compensated for. Therefore, the exponential window is able to make the modal properties, usually located in the low quefrency range, more dominant, and eliminate much forcing function information which is often distributed in a relatively higher quefrency region. This enables the possibility of applying cepstrum filtering for operational mode analysis [38], and an exponential filter is introduced in the cepstrum to construct a modal-property-dominant-generated layer in this paper. Based on research [30], the modal-property-dominant-generated layer used in this paper is summarized in Figure 3.

For illustration purpose, the signal before and after the modal-property-dominant-generated layer in the time domain, together with their frequency spectra, are depicted in Figure 4. The forcing function will change according to the speed and load variations while the transfer function is only tied to the system properties. The extraction of modal-property-dominant information, therefore, generates features that are less sensitive to the speed and load variations but sensitive to the changes in the system, such as the changes from healthy states to faulty states or fault locations. This enables the feasibility of establishing the network less affected by the operating conditions. Note that, the extracted feature is not the pure transfer function but the feature dominated by modal properties that carry bearing fault information. The forcing function located at low quefrency is still retained, making the extracted feature not completely operating-condition-affected-free.

2.3. Alleviation of speed variations

The bearing fault is usually located on the inner race, outer race, and ball. These three types of faults are considered in this paper. As mentioned in the last section, the local fault can generate a series of 'almost-periodic' impulses, and the fault information is embedded in this periodic frequency (i.e. FCF). Through bearing kinematics, the FCF can be calculated through:

$$f_{bpf_i} = \frac{n}{2} \left(1 + \frac{D_r}{D_p} \cos(\theta) \right) f_r \quad (15)$$

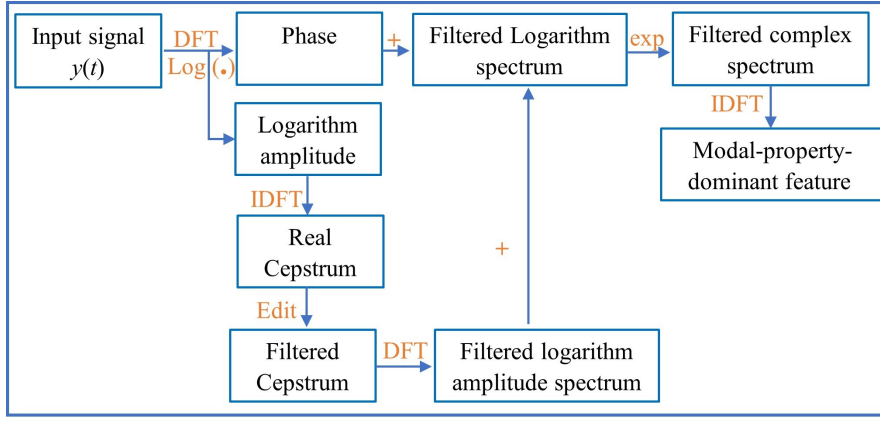


Figure 3: The modal-property-dominant-generated layer[36]

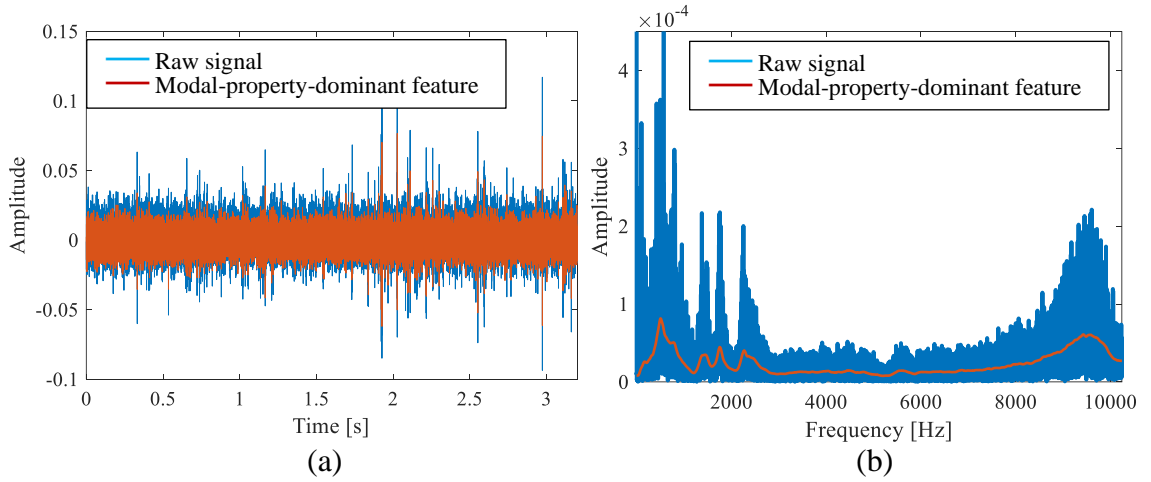


Figure 4: (a) Raw signal and feature after the modal-property-dominant-generated layer in time domain; (b) Raw signal and feature after the modal-property-dominant-generated layer in frequency domain.

$$f_{bpf o} = \frac{n}{2} \left(1 - \frac{D_r}{D_p} \cos(\theta) \right) f_r \quad (16)$$

$$f_{bsf} = \frac{D_p}{2D_r} \left(1 - \left(\frac{D_r}{D_p} \cos(\theta) \right)^2 \right) f_r \quad (17)$$

where n represents the rolling elements number, D_r denotes the rolling element diameter, D_p represents the pitch diameter, θ denotes the contact angle, and f_r represents the shaft rotating frequency. From Equation (15) - Equation (17), the bearing FCF can be regarded as the product of a constant which is determined by the bearing geometry and fault location, and the shaft rotating frequency. With the increasing complexity of the applications and operating conditions, there has been growing research on bearings operating under varying speeds. If the bearing operates under variable speeds, the FCF will change accordingly. One fundamental approach for dealing with variable speed scenarios is COT [39], which is performed by resampling the time signal with constant time increments to the angle signal with constant angle increments. If the rotating shaft is utilized as the reference, i.e. converting the sampling rate from the

number of samples per second to the number of samples per shaft revolution, the order in order spectrum (spectrum of the angle signal) can be defined as the multiple of the reference frequency:

$$o = f/f_r \quad (18)$$

From Equation (18), the conversion from the time domain to the angle domain can remove the frequency modulation caused by the rotational speed variations. Therefore, a domain-conversion layer is established based on COT to convert the raw signal to the angle domain, through which the influence of speed variations can be alleviated. Details about COT can refer to [39].

2.4. The Physics-Informed Residual Network architecture

Based on the physical modal-property-dominant-generated layer and domain-conversion layer, a Physics-Informed Residual Network is proposed in this paper. The architecture of the proposed PIResNet is depicted in Figure 5. As shown in Figure 5, the proposed PIResNet is composed of two parallel channels, in which the left channel is similar to a traditional Residual Network. The left channel begins with the domain-conversion layer, followed by a wide kernel CNN layer to suppress high-frequency noise [40], after which two RBBs and max pooling layers are sequentially added. The right channel starts with a modal-property-dominant-generated layer and is used to enhance the modal properties which are related to the system's intrinsic property and less affected by the operating condition. The rest of the right channel is the same as the left channel, targeting to automatically extract the intricate high-dimensional features from the modal-property-dominant signal. The left channel and right channel are then flattened and concatenated together, after which fully connected and softmax layers are utilized for classification. The detailed parameters used in the proposed PIResNet are listed in Table 1. Note that the Batch Normalization (BN) and Rectified Linear Unit (Relu) do not change the size of the feature and thus the parameters of BN and Relu are omitted in Table 1. Furthermore, the cross-entropy is used as the loss function to learn class discriminant features. Some important hyper-parameters utilized in the proposed PIResNet include: Adam Optimizer, mini-batch size of 128, training epochs of 100, dropout rate of 0.5, and initial learning rate of 0.001 (multiplied by 0.5 every twenty training epochs).

Table 1
Summary of the parameters used in the PIResNet.

Layer Type	Kernels Size/ Stride	Kernel Number	Padding	Output Size (width × depth)
Input	-	-	No	-
Model-property-dominant-generated layer	-	-	No	-
Domain-conversion layer	-	-	No	2048 × 1
Convolution 1	64 × 1/16 × 1	64	Yes	128 × 64
Pooling 1	2 × 1/2 × 1	64	No	64 × 64
Convolution 2 in RBB1	3 × 1/1 × 1	64	Yes	64 × 64
Convolution 3 in RBB1	3 × 1/1 × 1	64	Yes	64 × 64
Pooling 2	2 × 1/2 × 1	64	No	32 × 64
Convolution 4 in RBB2	3 × 1/1 × 1	64	Yes	32 × 64
Convolution 5 in RBB2	3 × 1/1 × 1	64	Yes	32 × 64
Pooling 3	2 × 1/2 × 1	64	No	16 × 64
Flatten and Concatenate	-	-	No	2048 × 1
Linear 1	-	-	No	256 × 1
Linear 2	-	-	No	32 × 1
Linear 3	-	-	No	4 × 1
Softmax	-	-	No	4

3. Experimental validation

To demonstrate the effectiveness of the proposed PIResNet, validation is conducted on two distinct experiments, including the bearings operating under variable speeds and loads, and bearings operating under time-varying speeds.

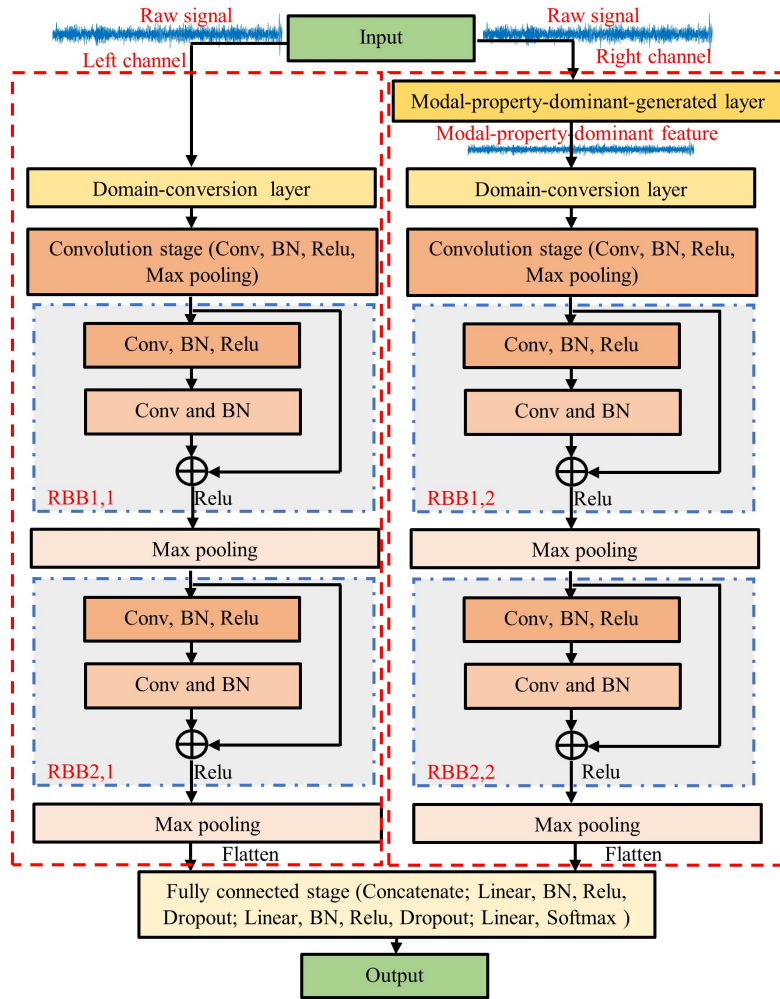


Figure 5: Architecture of the Physics-Informed Residual Network.

3.1. Experimental validation on bearings operating under variable speeds and loads

3.1.1. Description of data

The data collected from a drivetrain diagnostics simulator (DDS), as shown in Figure 6, is used in this section for validating the superiority of the PIResNet. The DDS is composed of four main components, including a drive motor, a two-stage planetary gearbox, a two-stage parallel gearbox and a magnetic power brake [41]. The DDS is driven by the drive motor, by adjusting which the speed variations can be simulated. The load is provided by the magnetic power brake, with different loads being generated by altering its current. In this paper, three distinct speeds and three different loads were set, as listed in Table 2, yielding nine different operating conditions in total.

Table 2

Summary of the DDS operating conditions.

Speed Load	20 Hz	35 Hz	40 Hz
0.0 A	Condition 1	Condition 2	Condition 3
0.3 A	Condition 4	Condition 5	Condition 6
0.6 A	Condition 7	Condition 8	Condition 9

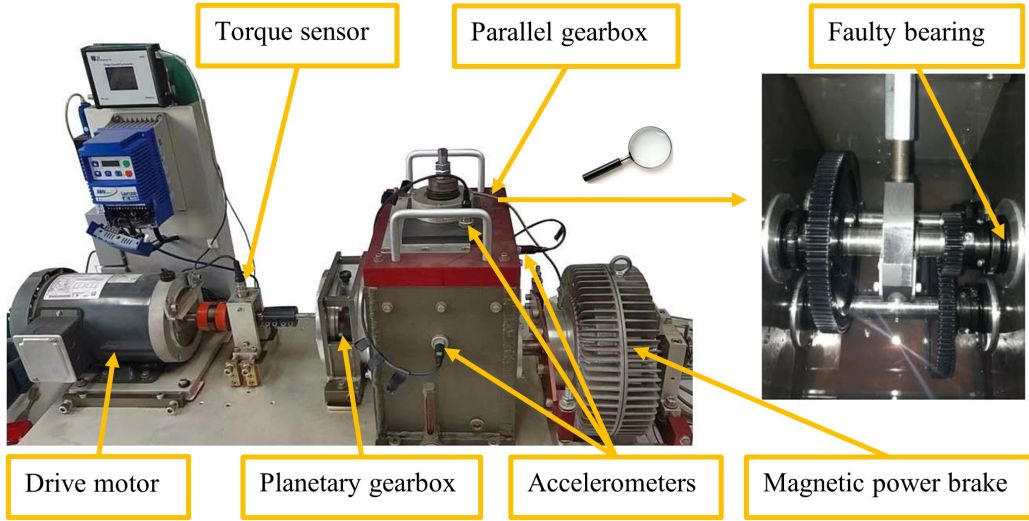


Figure 6: The layout of DDS test rig.

The bearings of healthy, outer race fault, inner race fault, and ball fault conditions were installed on the intermediate shaft's right side of the parallel gearbox, respectively. The vibration signals were collected by three accelerometers with a sampling frequency of 12800 Hz and a sampling time of 80 s. Since the bearing under inspection is located on the right side of the parallel gearbox, the vibration from the accelerometer placed on the corresponding side of the parallel gearbox is adopted for analysis. The vibration signal is normalized through Equation (19) at first. After the physical modal-property-dominant-generated layer and the domain-conversion layer (resampling frequency set to be 512), the signal is split into samples of 2048 length with an overlap of 0.5. Note that the construction of multiple samples is executed after the domain-conversion layer rather than the raw signal. This is because if the raw signal is split into samples with the same length, the sample length will be changed accordingly with speed after the angle resampling.

$$y_{normalized} = \frac{y - \text{mean}(y)}{\text{std}(y)} \quad (19)$$

3.1.2. Comparison settings

In this study, the PIResNet is implemented based on the PyTorch library. Four bearing conditions, including the healthy (Class 0), outer race fault (Class 1), inner race fault (Class 2), and ball fault (Class 3), are set. The performance of the proposed method is quantified by classification accuracy. A series of comprehensive comparisons are conducted to demonstrate the effectiveness of the proposed method. More specifically, the proposed PIResNet is compared with the following methods:

- 1) The methods with a similar network architecture to PIResNet but without physical modal properties embedding. Two networks are established for this purpose. The first network (denoted as Method 1 and will be used in the following paper) is established with the right channel of the PIResNet being deleted while keeping other parameters the same as the PIResNet. The second network (denoted as Method 2 and will be used in the following paper) is established by replacing the right channel with the same structure as the left channel and keeping other parameters the same as the PIResNet. A comparison of the PIResNet and Method 1 can demonstrate the superiority of the proposed method to the traditional pure data-driven method. However, there are two different factors between the PIResNet and Method 1, namely the network architecture (from one channel to two channels) and the embedding of physical modal properties. To guarantee the performance improvement is mainly caused by the Physics-Informed factor, Method 2 is designed with only one difference, i.e. physical modal properties, which is an ablation study to demonstrate the superiority of introducing the modal-property-dominant-generated layer.
- 2) Three state-of-the-art methods, including the DCNN [26], ResNet[42], and WDCNN [40]. The DCNN used in this paper is with the same architecture of research [26], which mainly adopts five convolution stages and one

fully-connected stage. The ResNet utilized in this paper is with the same architecture of research [42], which is mainly composed of one convolution stage, two RBBs, and one fully-connected stage. The WDCNN employed in this paper is with the same architecture of research [40], which mainly involves a wide kernel first convolution stage, four common convolution stages, and one fully-connected stage.

3.1.3. Demonstration and analysis of results

In this section, the performance of the PIResNet is evaluated from three different aspects, including 1) when all the operating conditions are included in the training data, 2) when the operating condition of the testing data is not included in the training data, and 3) when different levels of Gaussian white noise are added. The tasks in these scenarios are summarized in Table 3. For all these tasks, the computation is repeated five times. The mean accuracy and standard deviation of these five computations are recorded. To avoid misunderstanding, the accuracy discussed in the following of this paper refers to the mean accuracy of each task unless otherwise specified. Detailed results and discussion are shown below.

Table 3

Summary of the diagnostic tasks for experiments of DDS.

Task	Training data	Testing data
A	$0.7 \times$ Condition 1,2,3,4,5,6,7,8,9	$0.3 \times$ Condition 1,2,3,4,5,6,7,8,9
B1	Condition 2,3,4,5,6,7,8,9	Condition 1
B2	Condition 1,3,4,5,6,7,8,9	Condition 2
B3	Condition 1,2,4,5,6,7,8,9	Condition 3
B4	Condition 1,2,3,5,6,7,8,9	Condition 4
B5	Condition 1,2,3,4,6,7,8,9	Condition 5
B6	Condition 1,2,3,4,5,7,8,9	Condition 6
B7	Condition 1,2,3,4,5,6,8,9	Condition 7
B8	Condition 1,2,3,4,5,6,7,9	Condition 8
B9	Condition 1,2,3,4,5,6,7,8	Condition 9
C1	$0.7 \times$ Condition 1,2,3,4,5,6,7,8,9 and SNR=5dB	$0.3 \times$ Condition 1,2,3,4,5,6,7,8,9 and SNR=5dB
C2	$0.7 \times$ Condition 1,2,3,4,5,6,7,8,9 and SNR=0dB	$0.3 \times$ Condition 1,2,3,4,5,6,7,8,9 and SNR=0dB
C3	$0.7 \times$ Condition 1,2,3,4,5,6,7,8,9 and SNR=-5dB	$0.3 \times$ Condition 1,2,3,4,5,6,7,8,9 and SNR=-5dB

- 1) Diagnostic accuracy when all operating conditions are included. This case evaluates the performance when there is no distribution discrepancy between the training and testing data, which is the underlying assumption of success for most deep learning methods. In this case, data from all operating conditions are employed and shuffled, with a random 70 per cent being the training data and the rest 30 per cent as the testing data. The accuracy using the proposed method and other comparative methods is presented in Table 4.

Table 4

Diagnostic accuracy (%) for Task A (The highest accuracy in the task is marked in bold numerals).

	PIResNet	Method 1	Method 2	DCNN	ResNet	WDCNN
Task A	99.59 ± 0.16	99.54 ± 0.09	99.53 ± 0.10	97.75 ± 2.20	97.18 ± 2.58	99.40 ± 0.16

- 2) Diagnostic accuracy when the operating condition of testing data is missing. This case aims at exploring the performance in the presence of distribution discrepancy between the training data and testing data, which is a common phenomenon during bearing operation. In practice, massive data under various operating conditions may have been stored, but it seems impossible to collect the data under all operating conditions. Since bearings usually run under complex operating conditions and a slight variation in speed, load, or other factors can result in a new operating condition. In this case, data from one operating condition are utilized as the testing data and the data from the other operating conditions are employed for training, details can refer to Table 3. The testing accuracy of the proposed method and other methods are enumerated in Table 5, which can reflect the domain adaptability under variable loads and speeds.

Table 5

Diagnostic accuracy (%) for Task B (The highest accuracy in each task is marked in bold numerals).

	PIResNet	Method 1	Method 2	DCNN	ResNet	WDCNN
Task B1	80.19 ± 2.24	76.66 ± 1.76	74.50 ± 2.50	63.67 ± 7.72	60.23 ± 3.56	68.09 ± 4.71
Task B2	91.66 ± 1.35	88.92 ± 1.17	90.48 ± 3.02	86.78 ± 2.70	93.61 ± 2.01	81.35 ± 2.44
Task B3	98.49 ± 0.37	97.84 ± 0.33	97.68 ± 1.13	86.27 ± 1.91	96.52 ± 0.92	92.41 ± 1.96
Task B4	98.00 ± 0.68	96.35 ± 1.22	96.85 ± 1.01	74.20 ± 7.59	94.60 ± 0.94	90.70 ± 7.42
Task B5	96.17 ± 0.53	95.28 ± 1.15	95.54 ± 1.71	89.31 ± 3.37	93.34 ± 2.26	86.99 ± 3.17
Task B6	99.10 ± 0.29	98.39 ± 0.41	98.36 ± 0.90	91.22 ± 1.22	92.08 ± 2.62	94.36 ± 1.89
Task B7	99.63 ± 0.15	99.56 ± 0.17	99.03 ± 0.42	89.33 ± 1.75	89.59 ± 9.38	96.06 ± 3.21
Task B8	97.57 ± 1.09	96.23 ± 1.29	97.16 ± 1.04	92.49 ± 3.12	87.70 ± 3.03	92.73 ± 1.91
Task B9	99.76 ± 0.10	99.79 ± 0.07	99.66 ± 0.25	99.17 ± 0.88	99.98 ± 0.03	99.69 ± 0.22

- 3) Diagnostic accuracy under different levels of signal-to-noise ratios (SNR). This case targets to investigate the robustness against the noise of the PIResNet and other methods. Different levels of Gaussian white noise are added to the raw signal to generate different levels of SNR, which is defined as:

$$SNR_{dB} = 10 \times \log_{10} \frac{P_{raw}}{P_{noise}} \quad (20)$$

where P_{raw} represents the power of the raw signal and P_{noise} denotes the power of Gaussian white noise. The testing accuracy when SNR equals 5dB, 0dB, and -5dB is listed in Table 6.

Table 6

Diagnostic accuracy (%) for Task C (The highest accuracy in each task is marked in bold numerals).

	PIResNet	Method 1	Method 2	DCNN	ResNet	WDCNN
Task C1	98.85 ± 0.17	98.69 ± 0.06	98.54 ± 0.31	93.77 ± 2.25	94.38 ± 2.76	97.24 ± 2.22
Task C2	95.96 ± 0.31	95.62 ± 0.18	95.45 ± 0.41	83.00 ± 2.61	82.38 ± 2.76	92.98 ± 2.00
Task C3	82.89 ± 0.25	82.23 ± 0.60	82.40 ± 0.50	54.04 ± 1.66	57.80 ± 2.31	74.14 ± 1.74

In addition to the testing accuracy, the confusion matrices using the proposed PIResNet for Task A, Task B, and Task C are displayed in Figure 7. In the confusion matrix, each cell is normalized by the number of the same true class. Note that the tables present the mean value and standard deviation value of five trials while each confusion matrix displays the results with the highest accuracy among these five trials.

From Table 4, it can be concluded that if there is no distribution discrepancy, all methods can achieve high diagnostic accuracy, among which the lowest diagnostic accuracy is 97.18%. The proposed PIResNet has the highest accuracy, which is 99.59%. From Table 5, when the operating condition of the testing set is not included in the training set, the diagnostic accuracy of all methods deteriorates. In Task B1, the accuracy of all the methods is less than 80% except for the PIResNet, partly because the vibration signal is weak in essence at a zero-load and relatively low-speed operating condition. In Task B, the proposed PIResNet demonstrates superior performance, attaining the highest accuracy on 7 out of 9 tasks. The ResNet outperforms the PIResNet in Task B2 and Task B9. Note that in Task B2, although lower than the accuracy of the ResNet, the PIResNet can achieve the second-highest accuracy. And in Task B9, the proposed PIResNet can also attain a remarkable accuracy (99.76%), which is acceptable in the industry. In addition, the overall mean accuracy of the PIResNet in Task B is 95.62%, which is much higher than that of the ResNet. Consequently, the PIResNet exhibits excellent adaptability across variable loads and speeds. From Table 6, the proposed PIResNet achieves the highest diagnostic accuracy in various SNR scenarios, showing high robustness against the noise.

Furthermore, the proposed PIResNet offers higher accuracy than Method 1 and Method 2 in all tasks for the DDS dataset. Method 2 achieves higher accuracy than Method 1 in Task B2, B4, B5, B8, and C3, with an overall mean accuracy of 94.24% for all tasks of the DDS dataset. While Method 1 attains higher accuracy than Method 2 in the

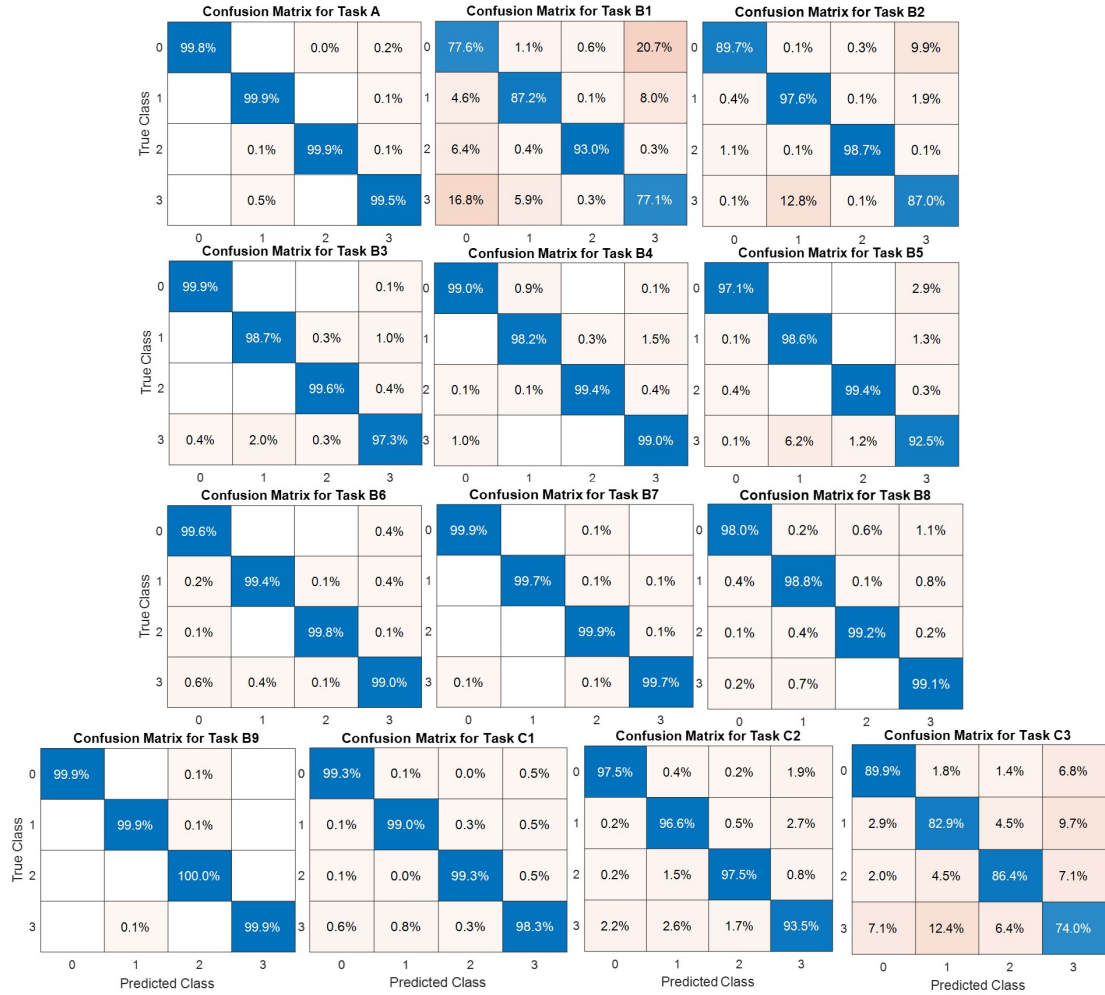


Figure 7: The confusion matrices using PIResNet for DDS dataset.

remaining tasks, with a similar overall mean accuracy of 94.24% for all tasks. As a result, the alteration of the network architecture from one channel to two parallel channels does not affect the diagnostic accuracy significantly. It is the embedding of a modal-property-dominant-generated layer that primarily improves diagnostic performance.

3.2. Experimental validation on bearings operating under time-varying speeds

3.2.1. Description of data

This section investigates the effectiveness of the proposed method under time-varying speed operating conditions. The data collected from a bearing test rig at Korea Advanced Institute of Science and Technology (KAIST) [43], as shown in Figure 8, is used in this section. The KAIST bearing test rig mainly consists of one brake, two bearings, one mass wheel, one gearbox, one torque meter, and a motor. This experiment simulated the time-varying speed conditions by adjusting the rotational speed of the motor, and the motor's output speed was randomly varied from 680 rev/min to 2460 rev/min. Figure 9 shows a random 300-second segment of the speed in this experiment. The bearings were seeded with outer race fault, inner race fault, and ball fault by electrical discharge machining, generating four different bearing conditions (including the healthy condition). The vibration under time-varying operating conditions was collected by the accelerometers installed horizontally and vertically on Bearing B's bearing house. The sampling frequency was set as 25600 Hz and a total of 2100 seconds of vibration was collected for each condition. More details about this experiment can refer to [43].

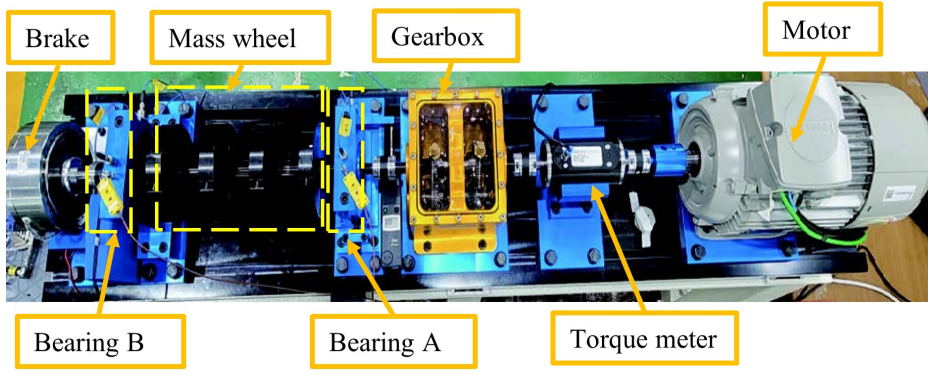


Figure 8: The layout of KAIST bearing test rig [43].

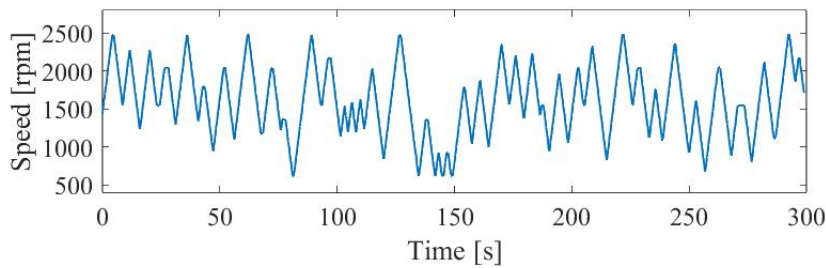


Figure 9: A random 300-second segment of the speed.

3.2.2. Demonstration and analysis of results

This section validates the superiority of the PIResNet from two tasks, including 1) Task D: when a random 70 per cent of the time-varying samples is utilized as the training data and the rest 30 per cent of samples is employed as the testing data. The data are processed using the same procedure as that of the first experiment, and the accuracy using the proposed method and other comparison methods is presented in Table 7. And 2) Task E: when different levels of Gaussian white noise are added. The results under different SNRs (5dB, 0dB, -5dB) are summarized in Table 8. Furthermore, the confusion matrices utilizing PIResNet for Task D and Task E are displayed in Figure 10.

Table 7

Diagnostic accuracy (%) for Task D (The highest accuracy in the task is marked in bold numerals).

	PIResNet	Method 1	Method 2	DCNN	ResNet	WDCNN
Task D	97.78 ± 0.18	97.11 ± 0.27	97.02 ± 0.35	95.50 ± 0.14	90.96 ± 0.89	96.91 ± 0.29

Table 8

Diagnostic accuracy (%) for Task E (The highest accuracy in each task is marked in bold numerals).

	PIResNet	Method 1	Method 2	DCNN	ResNet	WDCNN
Task E1	90.26 ± 0.18	88.08 ± 0.29	88.00 ± 0.19	82.29 ± 0.26	80.31 ± 0.52	82.50 ± 0.25
Task E2	81.28 ± 0.36	79.77 ± 0.19	79.78 ± 0.40	70.46 ± 0.60	71.28 ± 0.36	76.11 ± 0.50
Task E3	71.67 ± 0.20	71.36 ± 0.40	71.32 ± 0.29	59.00 ± 0.20	61.03 ± 0.33	68.78 ± 0.20

From Table 7, the PIResNet produces the highest accuracy of 97.78%, followed by Method 1 and Method 2, with accuracy at 97.11 % and 97.02 %, respectively. The WDCNN, DCNN, and ResNet achieve accuracy at 96.91 %, 95.50%,

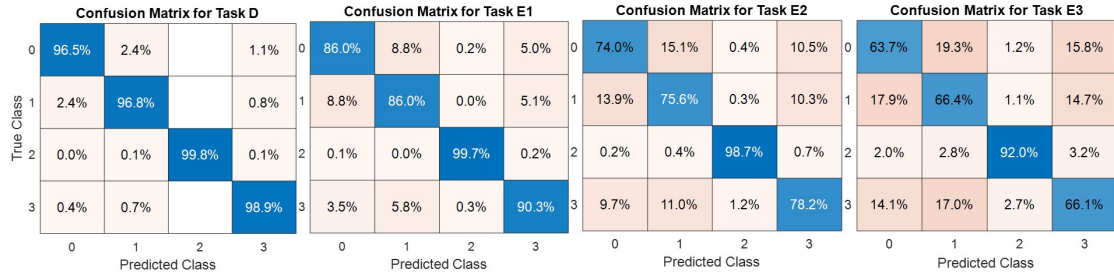


Figure 10: The confusion matrices using PIResNet for KAIST dataset.

and 90.96%, respectively. Therefore, the PIResNet outperforms other methods under time-varying speed conditions. From Table 8, the PIResNet has the highest diagnostic accuracy when SNR equals 5dB, 0dB, and -5dB, demonstrating it can perform robustly against noise in time-varying speed conditions. Furthermore, Method 2 has a higher accuracy than Method 1 in Task E2, with an overall mean accuracy of 84.03 % for Task D and Task E. While Method 1 outperforms Method 2 in Task D, Task E1, and Task E3, with an overall mean accuracy of 84.08 % for Task D and Task E. In this experiment, changing from one channel to two parallel channels in the network architecture seems not to increase diagnostic accuracy. This once again reveals that the incorporation of physics information contributes to improving diagnostic accuracy as opposed to the network architecture alteration.

3.3. Summary of results

In this section, two datasets from the DDS test rig (under variable speeds and loads) and the KAIST test rig (under time-varying speeds) are employed to demonstrate the superiority of the proposed PIResNet. The two experiments involve five types of tasks with 17 sub-tasks in total. More specifically, in Task A, a random 70 per cent of samples under all operating conditions is utilized as training data (DDS dataset); in Task B, the operating condition of testing data is not included in the training data (DDS dataset); in Task C, data are under different levels of SNR (DDS dataset); in Task D, a random 70 per cent of the time-varying samples is utilized as the training data (KAIST dataset); and in Task E, data are under different levels of SNR (KAIST dataset). From Table 4 to Table 8, it can be concluded that 1) the highest accuracy in each task is obtained from either the proposed PIResNet or ResNet; 2) out of the 17 sub-tasks, the proposed PIResNet offers the highest accuracy in 15 sub-tasks; 3) out of the 17 sub-tasks, in comparison of the proposed PIResNet with the ResNet, the proposed PIResNet offers the lower standard deviations in 15 sub-tasks and one same standard deviation in Task E2. Therefore, the proposed PIResNet achieves the highest accuracy in most sub-tasks and offers lower standard deviations in the majority of cases, demonstrating its ability to efficiently handle distribution discrepancy and noise scenarios under variable or time-varying operating conditions.

4. Conclusion

This paper has developed an accurate and robust diagnostic framework for bearings, in which a PIResNet with high physical compliance is proposed. In the proposed PIResNet framework, a novel physical modal-property-dominant-generated layer is constructed. The physical modal-property-dominant-generated layer is developed on the basis of the cepstrum exponential filtering, which generates the feature that is dominated by modal properties. By integrating the modal-property-dominant-generated layer, the PIResNet is able to learn features not only from the existing training data but also the modal properties embedded in both training and testing data, thus providing a physical consistent solution that works efficiently for imperfect data such as distribution discrepancy. In addition, a domain-conversion layer is introduced to transform the time signal into the angle signal to further alleviate the influence of any speed variations in bearing diagnostics. The intricate high-dimensional features of the raw signal and modal-property-dominant signal are automatically extracted and fused in the proposed PIResNet, generating results with high accuracy and in alignment with physics. The performance of the proposed method is validated by experimental results of bearings operating under variable speeds and loads, and time-varying speed conditions. The outstanding performance of the PIResNet is confirmed by comprehensive comparisons, demonstrating that the proposed PIResNet is an intelligent diagnostic method with high accuracy, domain adaptability of variable loads and speeds, and robustness to noise, offering great promise as a Physics-Informed solution for industrial bearing condition monitoring. In this paper, the proposed

PIResNet framework has been developed and validated for bearing diagnostics. However, it would be interesting to investigate the generalization ability to other types of mechanical systems, such as gears and shafts. The potential applications to other mechanical systems could facilitate the proposed method to be more broadly applied in industrial condition monitoring.

Acknowledgement

This research is partially supported by the Australian Research Council under Linkage Project [LP220100389]. The authors would also like to appreciate the KAIST for publishing the data [43] for analysis.

References

- [1] Pinjia Zhang, Yi Du, Thomas G Habetler, and Bin Lu. A survey of condition monitoring and protection methods for medium-voltage induction motors. *IEEE Transactions on Industry Applications*, 47(1):34–46, 2010.
- [2] The National Renewable Energy Laboratory. Wind turbine gearbox damage distribution statistics 2016. <https://grd.nrel.gov/#/stats>. Accessed 15/01/2023.
- [3] Hosameeldin Ahmed and Asoke K Nandi. *Condition monitoring with vibration signals: Compressive sampling and learning algorithms for rotating machines*. John Wiley & Sons, 2020.
- [4] Pankaj Gupta and MK Pradhan. Fault detection analysis in rolling element bearing: A review. *Materials Today: Proceedings*, 4(2):2085–2094, 2017.
- [5] Qing Ni, JC Ji, Ke Feng, and Benjamin Halkon. A novel correntropy-based band selection method for the fault diagnosis of bearings under fault-irrelevant impulsive and cyclostationary interferences. *Mechanical Systems and Signal Processing*, 153:107498, 2021.
- [6] Shubin Wang, Xuefeng Chen, Ivan W Selesnick, Yanjie Guo, Chaowei Tong, and Xingwu Zhang. Matching synchrosqueezing transform: A useful tool for characterizing signals with fast varying instantaneous frequency and application to machine fault diagnosis. *Mechanical Systems and Signal Processing*, 100:242–288, 2018.
- [7] Yaguo Lei, Bin Yang, Xinwei Jiang, Feng Jia, Naipeng Li, and Asoke K Nandi. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, 138:106587, 2020.
- [8] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [9] Weihua Li, Ruyi Huang, Jipu Li, Yixiao Liao, Zhuyun Chen, Guolin He, Ruqiang Yan, and Konstantinos Gryllias. A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. *Mechanical Systems and Signal Processing*, 167:108487, 2022.
- [10] Ying Zhang and Jinchun Ji. Intelligent fault diagnosis of a reciprocating compressor using mode isolation convolutional deep belief networks. *IEEE/ASME Transactions on Mechatronics*, 26(3):1668–1677, 2020.
- [11] Yongchao Zhang, Zhaohui Ren, Shihua Zhou, Ke Feng, Kun Yu, and Zheng Liu. Supervised contrastive learning-based domain adaptation network for intelligent unsupervised fault diagnosis of rolling bearing. *IEEE/ASME Transactions on Mechatronics*, 27(6):5371–5380, 2022.
- [12] Ke Feng, JC Ji, Yongchao Zhang, Qing Ni, Zheng Liu, and Michael Beer. Digital twin-driven intelligent assessment of gear surface degradation. *Mechanical Systems and Signal Processing*, 186:109896, 2023.
- [13] Te Han, Chao Liu, Wenguang Yang, and Dongxiang Jiang. Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application. *ISA transactions*, 97:269–281, 2020.
- [14] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.
- [15] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.
- [16] Jin-Long Wu, Heng Xiao, and Eric Paterson. Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework. *Physical Review Fluids*, 3(7):074602, 2018.
- [17] George S Misyris, Andreas Venzke, and Spyros Chatzivasileiadis. Physics-informed neural networks for power systems. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2020.
- [18] Karthik Kashinath, M Mustafa, Adrian Albert, JL Wu, C Jiang, Soheil Esmaeilzadeh, Kamyar Azizzadenesheli, R Wang, A Chattopadhyay, A Singh, et al. Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, 379(2194):20200093, 2021.
- [19] Siddhartha Mishra and Roberto Molinaro. Physics informed neural networks for simulating radiative transfer. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 270:107705, 2021.
- [20] Lu Lu, Raphael Pestourie, Wenjie Yao, Zhicheng Wang, Francesc Verdugo, and Steven G Johnson. Physics-informed neural networks with hard constraints for inverse design. *SIAM Journal on Scientific Computing*, 43(6):B1105–B1132, 2021.
- [21] Huailiang Zheng, Yuantao Yang, Jiancheng Yin, Yuqing Li, Rixin Wang, and Minqiang Xu. Deep domain generalization combining a priori diagnosis knowledge toward cross-domain fault diagnosis of rolling bearing. *IEEE Transactions on Instrumentation and Measurement*, 70:1–11, 2020.
- [22] Mohammadkazem Sadoughi and Chao Hu. Physics-based convolutional neural network for fault diagnosis of rolling element bearings. *IEEE Sensors Journal*, 19(11):4181–4192, 2019.
- [23] Tianfu Li, Zhibin Zhao, Chuang Sun, Li Cheng, Xuefeng Chen, Ruqiang Yan, and Robert X Gao. Waveletkernelnet: An interpretable deep neural network for industrial intelligent diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(4):2302–2312, 2021.

- [24] Hao Lu, Venkat Pavan Nemani, Vahid Barzegar, Cade Allen, Chao Hu, Simon Laflamme, Soumik Sarkar, and Andrew T Zimmerman. A physics-informed feature weighting method for bearing fault diagnostics. *Mechanical Systems and Signal Processing*, 191:110171, 2023.
- [25] Botao An, Shibin Wang, Fuhua Qin, Zhibin Zhao, Ruqiang Yan, and Xuefeng Chen. Adversarial algorithm unrolling network for interpretable mechanical anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–16, 2023.
- [26] Sheng Shen, Hao Lu, Mohammadkazem Sadoughi, Chao Hu, Venkat Nemani, Adam Thelen, Keith Webster, Matthew Darr, Jeff Sidon, and Shawn Kenny. A physics-informed deep learning approach for bearing fault detection. *Engineering Applications of Artificial Intelligence*, 103:104295, 2021.
- [27] Diwang Ruan, Xinzhou Song, Clemens Gühmann, and Jianping Yan. Collaborative optimization of cnn and gan for bearing fault diagnosis under unbalanced datasets. *Lubricants*, 9(10):105, 2021.
- [28] Tim von Hahn and Chris K Mechefske. Knowledge informed machine learning using a weibull-based loss function. *arXiv preprint arXiv:2201.01769*, 2022.
- [29] Yuejian Chen, Meng Rao, Ke Feng, and Ming J Zuo. Physics-informed lstm hyperparameters selection for gearbox fault detection. *Mechanical Systems and Signal Processing*, 171:108907, 2022.
- [30] RB Randall, Michael D Coats, and Wade A Smith. Repressing the effects of variable speed harmonic orders in operational modal analysis. *Mechanical Systems and Signal Processing*, 79:3–15, 2016.
- [31] Qinghua Wang, Chenguang Yang, Hongqiang Wan, Donghua Deng, and Asoke K Nandi. Bearing fault diagnosis based on optimized variational mode decomposition and 1d convolutional neural networks. *Measurement Science and Technology*, 32(10):104007, 2021.
- [32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [33] RB Randall and WA Smith. New cepstral methods for the diagnosis of gear and bearing faults under variable speed conditions. In *ICSV23 conference, Athens*, 2016.
- [34] Y Gao and RB Randall. Determination of frequency response functions from response measurements—i. extraction of poles and zeros from response cepstra. *Mechanical systems and signal processing*, 10(3):293–317, 1996.
- [35] Asoke Nandi. *Blind estimation using higher-order statistics*. Springer Science & Business Media, 1999.
- [36] RB Randall. A history of cepstrum analysis and its application to mechanical problems. *Mechanical Systems and Signal Processing*, 97:3–19, 2017.
- [37] Alan V Oppenheim. *Discrete-time signal processing*. Pearson Education India, 1999.
- [38] RB Randall, Jérôme Antoni, and WA Smith. A survey of the application of the cepstrum to structural modal analysis. *Mechanical Systems and Signal Processing*, 118:716–741, 2019.
- [39] KR Fyfe and EDS Munk. Analysis of computed order tracking. *Mechanical systems and signal processing*, 11(2):187–205, 1997.
- [40] Wei Zhang, Gaoliang Peng, Chuanhao Li, Yuanhang Chen, and Zhujun Zhang. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, 17(2):425, 2017.
- [41] Qing Ni, JC Ji, Ke Feng, and Benjamin Halkon. A fault information-guided variational mode decomposition (fivmd) method for rolling element bearings diagnosis. *Mechanical Systems and Signal Processing*, 164:108216, 2022.
- [42] Wei Zhang, Xiang Li, and Qian Ding. Deep residual learning-based fault diagnosis method for rotating machinery. *ISA transactions*, 95:295–305, 2019.
- [43] Wonho Jung, Seong-Hu Kim, Sung-Hyun Yun, Jaewoong Bae, and Yong-Hwa Park. Vibration, acoustic, temperature, and motor current dataset of rotating machine under varying operating conditions for fault diagnosis. *Data in Brief*, page 109049, 2023.