Adaptive Wavelet-Informed Physics-Based CNN for Bearing Fault Diagnosis

Reza Hassannejad, Mir Mohammad Ettefagh, and Yousef Bahrami Mossayebi

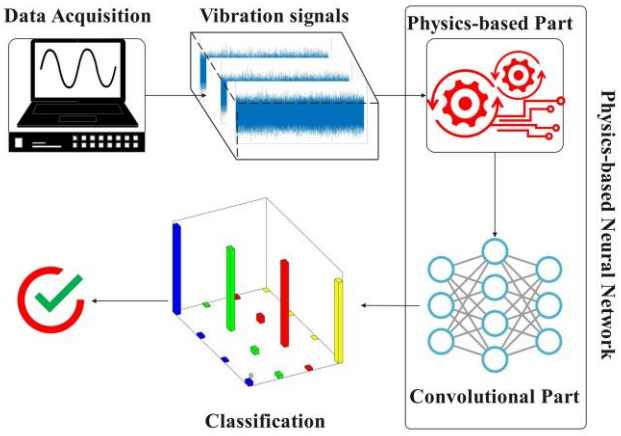
Faculty of Mechanical Engineering, University of Tabriz, Tabriz, 51666-14766, Iran

hassannejhad@tabrizu.ac.ir
ettefagh@tabrizu.ac.ir
yousefbahrami98@ms.tabrizu.ac.ir

ABSTRACT

With the increasing expansion of data science into various fields, the application of deep neural networks in the fault diagnosis of rotating machines has attracted significant attention from researchers. However, in the methods available in the literature, the physical characteristics of the problem are not incorporated into the structure of deep networks. In most existing methods, fault diagnosis is performed solely based on features extracted by convolutional layers, with no additional layers utilized to enhance or refine these features. This work introduces a novel physics-based neural network for bearing fault diagnosis, in which specific layers are designed based on signal processing methods to extract the physical features of faults. These layers, referred to as physics-based layers, are constructed using adaptive analytical wavelet filterbanks. The features extracted by these layers are then classified using convolutional layers, enabling the diagnosis of bearing faults. A key advantage of this physics-based network is that it does not rely on a fixed architecture for feature extraction and classification. Instead, the characteristics of the network layers adapt to the fault characteristics present in the bearing vibration signals. The classification accuracy of the proposed method has been evaluated using experimental data from two studied cases. The results demonstrate that the newly introduced network achieves higher accuracy in classifying bearing signals with different faults compared to similar methods.

Keywords: Filterbank, CNN, Physics-based Neural Networks, Ball Bearing, Wavelet Transform, Fault Diagnosis.



1. INTRODUCTION

Bearings are among the most widely used mechanical components in rotating machines. Fault diagnosis of bearings is crucial to prevent irreparable damage (Mohanty, 2014). Vibration signal processing has emerged as a prominent method in the field of condition monitoring for rotating machinery, garnering significant attention from researchers (Ahmed & Nandi, 2020; Randall, 2021). Existing methods for bearing fault detection often rely on signal processing techniques such as wavelet transform, empirical and variational mode decomposition, and time-frequency analysis(Dibaj et al., 2020; Feldman, 2009; Maruthi & Hegde, 2015; Rai & Upadhyay, 2016). Among these, wavelet transform stands out as one of the most widely used methods for time-frequency analysis of vibration signals, including bearing signals, due to its high frequency-time resolution(Chen et al., 2016; Sharma et al., 2021).

In recent years, with the growing application of artificial intelligence across various fields, intelligent fault diagnosis of rotating machines using machine learning methods and neural networks has garnered significant attention from researchers (Asr et al., 2017; Lei et al., 2020; Yang & Delpha,

https://doi.org/10.36001/IJPHM.2025.v16i1.4234

2022a). One of the main challenges in fault diagnosis for rotating machines is the design of artificial neural networks capable of extracting the physical features of vibration signals and classifying faults accurately (Qian et al., 2022; Zhang et al., 2019). In most existing methods, neural networks are fed with physical features extracted using signal processing techniques (Cheng et al., 2021; Shen et al., 2015). Meanwhile, other researchers have attempted to integrate the physical characteristics of the problem directly into the structure of neural networks (Sun et al., 2019; Yang et al., 2019). Given that each problem has its unique characteristics, incorporating physical properties into the design of artificial neural networks can significantly enhance their efficiency (Pestana-Viana et al., 2016). For instance, a key feature of vibration signals from faulty bearings is the presence of harmonic impulsive peaks in the frequency domain (Randall, 2021). Consequently, a neural network that is more sensitive to this feature will be more effective for bearing fault diagnosis. Below, some of the studies that have explored the integration of physical knowledge with neural networks for bearing fault diagnosis are briefly summarized.

Tian et al. (Tian et al., 2015) utilized the K-Nearest Neighbor (KNN) algorithm to classify healthy and faulty bearing vibration signals. They employed Spectral Kurtosis (SK) to extract features from the analyzed signals of a filterbank. These extracted features were then fed into the KNN classifier to classify the input signals. In another study, Sadoughi and Hu (Sadoughi & Hu, 2019) introduced a physics-based convolutional neural network for bearing fault diagnosis. The network proposed in their research incorporates several layers dedicated to physical feature extraction prior to the convolutional layers. These feature extraction layers include SK analysis, envelope spectrum analysis, a convolutional kernel, and fast Fourier transform.

In another study, Guo et al. (Guo et al., 2018) employed transfer learning to classify unlabeled bearing vibration data. Unlike other methods, they did not use any signal processing techniques for feature extraction and instead utilized raw bearing vibration signals as the input for two neural networks. In their approach, they first trained a deep neural network using bearing data from one rotating machine and then applied transfer learning to detect bearing faults in another rotating machine. Similarly, Dibaj et al. (Dibaj, Ettefagh, et al., 2021) combined Variable Mode Decomposition (VMD) with Convolutional Neural Networks (CNNs) for bearing fault diagnosis. In their research, they initially analyzed bearing vibration signals using a fine-tuned VMD method. The time-frequency images of the decomposed modes were subsequently fed into a CNN for classification. Additionally, Li et al. (Li et al., 2020) proposed a method based on back-propagation neural networks to classify bearing vibration signals. Their approach involved applying Discrete Wavelet Transform (DWT) to the vibration signals and extracting local features from the resulting sub-signals. These features were then input into a back-propagation neural network, and the outputs were classified using a Support Vector Machine (SVM) classifier.

In another study conducted by Mao et al. (Mao et al., 2022), a deep denoising autoencoder and joint adversarial learning were used for bearing fault diagnosis. In their work, fault features were first extracted using a deep autoencoder, and then convolutional networks with adversarial learning were employed to classify the extracted features. Similarly, Yang and Delpha (Yang & Delpha, 2022b) utilized the local Mahalanobis distance for bearing fault diagnosis. They combined the local Mahalanobis index with its anchor generation algorithm to isolate faulty variables and classify bearing signals.

Ma et al. (Ma et al., 2022) also employed autoencoders and deep networks for bearing fault diagnosis. In their method, bearing vibration signals were fed into a deep autoencoder, after which the encoder's output was classified using a deep neural network. Additionally, a fully connected neural network was trained through transfer learning. In another study, Shen et al. (Shen et al., 2021) proposed a physics-informed deep learning method for bearing fault diagnosis. Their approach combined a simple threshold model with a deep CNN to classify sub-bands obtained from envelope spectrum analysis. They also trained the network using a novel loss function that incorporated the physical characteristics of the bearings.

In another study, Kim and Kim (Kim & Kim, 2024) introduced a noise-robust fault diagnosis method specifically designed for bearings. Their approach utilizes an adversarial physics-based neural network that integrates the time-frequency characteristics of vibration signals. The method involves extracting multi-domain features along a complex parallel pipeline, which are then processed by a CNN incorporating a Squeeze-and-Excitation block. These features are ultimately classified to identify the type of the bearing fault. Liao et al. (Liao et al., 2025) proposed a physics-informed fault diagnosis method for bearings that leverages blind deconvolution. Their approach integrates deep classifier neural networks with blind deconvolution to develop a supervised deep network tailored for fault diagnosis. In their method, time-frequency features of vibration signals are first extracted using blind deconvolution. These features are then fed into a deep neural network, which employs a physics-informed loss function specifically designed for the bearing fault diagnosis problem. Ma et al. (Ma et al., 2025) proposed a method for bearing fault diagnosis in few-shot conditions by combining physics-based sample generation with CNNs. Their method uses dynamic modeling to create additional synthetic samples of vibration signals from faulty bearings. These generated samples are merged with experimental data and input into a CNN, which classifies the type of fault in the bearing.

To advance and expand upon prior research, this study introduces a novel adaptive filterbank based on wavelet

transforms. The parameters of this filterbank are optimized for each input signal through two separate optimization processes. This adaptive filterbank is then integrated with the Hilbert envelope spectrum and a kernel designed based on the physical characteristics of bearing faults to form the physics-based component of the proposed neural network. Finally, by combining this component with a CNN-based module, an adaptive physics-based neural network is developed for bearing fault diagnosis. The proposed network is designed such that the primary features of bearing faults —specifically, the presence of impulsive peaks in the frequency spectrum of vibration signals—are directly embedded in its structure. The initial layers of the network are tailored to the physical characteristics of bearing faults and adapt to the input signal. These are followed by convolutional layers, a fully connected layer, and a classifier. The proposed method can be summarized as follows:

1) Feeding the bearing signals into an adaptive layer (a new analytical filterbank).
2) Entering the outputs of the filterbank into a mode selection layer.
3) Feeding the selected mode into a convolution layer, whose kernel is designed based on the ballpass frequency of the bearing faults.
4) Calculating the Hilbert envelope spectrum of the input.
5) Feeding the calculated spectra into several convolution layers.
6) Entering the output of the convolutional layers into the fully connected layer and then the classification layer.

Figure (1) provides an illustration of the proposed method's flowchart. The rest of the paper is organized as follows: Section 2 provides an explanation of the adaptive analytical filterbank. Section 3 offers a basic description of designing the physics-based layers, and Section 4 discusses the theory of convolutional neural networks. The experimental evaluation of the proposed bearing fault diagnosis method is presented in Section 5. Finally, the conclusion is drawn in Section 6. Additionally, Section 7 discusses the limitations of the proposed method and similar methods, along with suggestions for improving these methods in future work.

2. ADAPTIVE ANALYTICAL FILTERBANK

In this chapter, the principles of the proposed filterbank theory are presented, followed by an explanation of the optimization process for its parameters.

2.1. Proposed filterbank

Flexible Analytical Wavelet Transform (FAWT) is a flexible filterbank, introduced by Byram (Bayram, 2012). Some of its key characteristics are as follows:

1- The coefficients of the up/down samplers are chosen as arbitrary rational values.
2- For the definition of filters, Hilbert bases, and Daubechies wavelets are employed.
3- Filters are defined in the frequency domain (Bayram, 2012; Zhang et al., 2015).

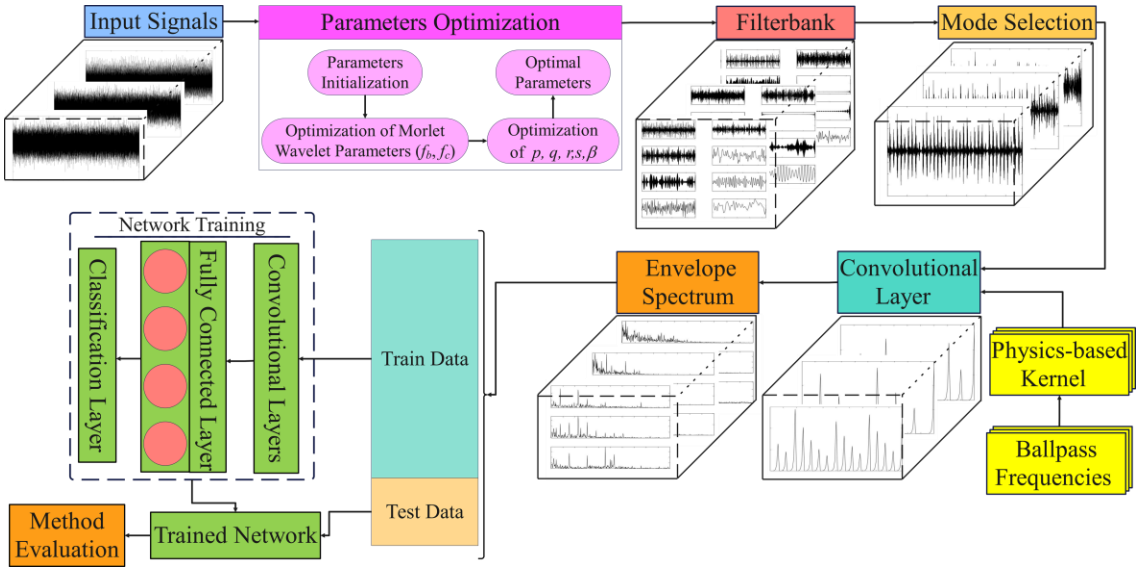


Figure 1. Flowchart of the proposed method.

Due to the aforementioned characteristics, FAWT has flexible time-frequency coverage. The first application of the mentioned wavelet transform was to analyze oscillatory signals, particularly audio signals (Bayram, 2012). The aim of this study is to develop a filterbank capable of effectively extracting bearing fault features from vibration signals. To achieve this, an optimized filterbank for bearing signal analysis has been designed by implementing the following modifications:

1- implementing the Morlet wavelet to define filters of the filterbank.

2- removing filters that extract the negative frequency of the input (in the primary filter bank, two filters are employed to analyze the high-frequency channel. One for positive frequencies and the other for negative frequencies).

These modifications were made due to the absence of negative frequencies in mechanical signals and the effectiveness of the Morlet wavelet in diagnosing faults in rotating machines, such as bearings (Albezzawy et al., 2019; Qin et al., 2016).

Figure (2) illustrates the ideal frequency response of the low-pass (*H(ω)*) and high-pass (*G(ω)*) filters within the described filterbank (Zhang et al., 2015).

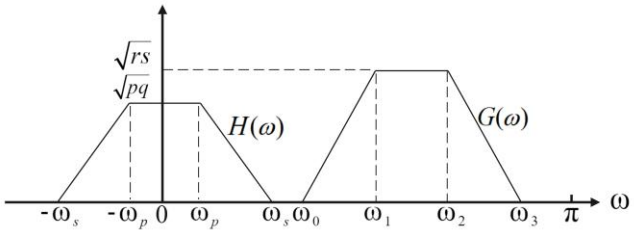


Figure 2. frequency response of FAWT filters.

The high-pass (*G(ω)*) and low-pass (*H(ω)*) filters of the proposed filterbank are mathematically defined in the frequency domain as follows (Bayram, 2012):

$$H(\omega)=\begin{cases}\sqrt{pq} & |\omega|<\omega_p \\ \sqrt{pq}\left[\left(\omega_s-\omega_p\right)^{-1}\left(\omega-\omega_p\right)\right] & \omega_p\le\omega\le\omega_s \\ \sqrt{pq}\,\theta\left[\left(\omega_s-\omega_p\right)^{-1}\left(\pi-\omega+\omega_p\right)\right] & -\omega_s\le\omega\le-\omega_p \\ 0 & |\omega|>\omega_s\end{cases} \quad (1)$$

$$G(\omega)=\begin{cases}\sqrt{rs}\,\theta\left[\left(\omega_1-\omega_0\right)^{-1}\left(\pi-\omega-\omega_0\right)\right] & \omega_0\le\omega<\omega_1 \\ \sqrt{rs} & \omega_1\le\omega<\omega_2 \\ \sqrt{rs}\,\theta\left[\left(\omega_3-\omega_2\right)^{-1}\left(\omega-\omega_2\right)\right] & \omega_2\le\omega\le\omega_3 \\ 0 & \omega\in[0,\omega_0)\cup(\omega_2,2\pi)\end{cases} \quad (2)$$

, where *p, q, r,* and *s* are related to sampling factors of filterbank and:

$$\omega_p=\frac{(1-\beta)\pi}{q},\ \omega_s=\frac{\pi}{q},\ \omega_0=\frac{(1-\beta)(\pi+\varepsilon)}{r},$$

$$\omega_1=\frac{p\pi}{qr},\ \omega_2=\frac{(p-\varepsilon)}{r},\ \omega_3=\frac{(\pi+\varepsilon)}{r}$$

$$\theta(\omega)=\exp(-\pi f_b(\frac{\omega}{2\pi}-f_c)^2)\qquad for\quad \omega\in[0,\pi] \quad (3)$$

$$\varepsilon=\frac{1}{32}\left(\frac{p-q+\beta q}{p+q}\right)\pi \quad (4)$$

, where $\beta<1$ is a nonnegative constant and θ, f_b, and f_c are frequency response, bandwidth parameter, and center frequency of Morlet wavelet, respectively. In Figure (3), two levels of filterbank decomposition are shown.

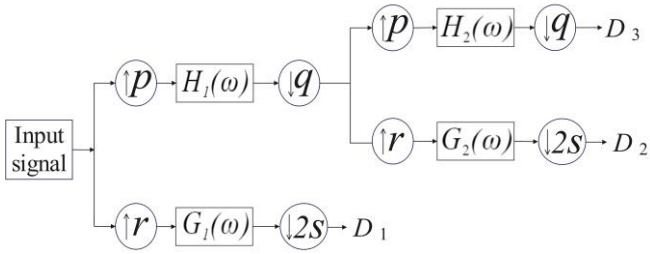


Figure 3. Two levels of the introduced filterbank.

2.2. Parameter optimization

To make the filterbank adaptable, the optimal parameters of the filterbank should be determined based on any arbitrary input signal *x(t)*. There are seven adjustable parameters in the proposed filterbank: the Morlet wavelet transform parameters (f_b, f_c), which form the core of the filters, and the filterbank sampling parameters (*p, q, r, s, β*). These parameters must be optimized to ensure the filterbank has a high capability for extracting bearing fault features. Given that bearing faults are characterized by the presence of harmonic impulsive peaks in the frequency spectrum of vibration signals, a two-stage optimization process is proposed:

The first stage involves determining the optimal parameters of the Morlet wavelet (f_b, f_c). The Gini index (GI) is selected as the objective function for this purpose due to its

high sensitivity to repetitive impulses in signals (Wang, 2018). The Gini index ranges between zero and one, with higher values indicating a greater number of repetitive impulse peaks in a signal. The effectiveness of the Gini index in optimizing the coefficients of the complex Morlet wavelet has been demonstrated in prior research, such as in (Albezzawy et al., 2020). To solve the optimization problem for the Morlet wavelet parameters, the particle swarm optimization (PSO) algorithm is employed. PSO is a metaheuristic algorithm designed to find the global minimum efficiently.

The second stage of the filterbank optimization process involves determining the optimal sampling coefficients (*p, q, r, s, β*). To optimize these coefficients, the weighted Kurtosis index (WKI) is chosen as the objective function. The WKI, an enhanced version of the Kurtosis index, exhibits high sensitivity to impact peaks in the frequency spectrum of signals (Dibaj, Hassannejad, et al., 2021). Its effectiveness in extracting bearing fault features and diagnosing bearing faults has been demonstrated in prior research, such as in (Zhang et al., 2021). Since the filterbank output is non-differentiable with respect to its sampling coefficients, metaheuristic algorithms are employed to optimize these coefficients. In this work, the Genetic Algorithm (GA) is utilized. Additionally, the parameters of the optimization algorithms (PSO and GA) are determined using a grid search approach. The formulation of each stage in the optimization process is outlined below.

The first stage of the parameter optimization process is Morlet wavelet parameters $\left(f_b, f_c\right)$ optimization. The GI is employed as the objective function to optimize these parameters. For a vector like $y[n]=\left[y(1), y(2), ..., y(N)\right]$ with its elements re-ordered and represented by: $y_{(K)}$ for $K=1,2,...,N$, where $\left|y_{(1)}\right| \le \left|y_{(2)}\right| \le ... \le \left|y_{(N)}\right|$, the GI is defined as follows (Albezzawy et al., 2020; Hurley & Rickard, 2009):

$$GI(y)=1-2\sum_{K=1}^{N}\frac{\left|y_{(K)}\right|}{y_{L1}}\times\left(\frac{N-K+\frac{1}{2}}{N}\right) \tag{5}$$

, where y_{L1} is the l_1 norm of $y[n]$.

To optimize the parameters of the Morlet wavelet, first, the wavelet coefficients are calculated from the equation of wavelet transform which is as follows:

$$w_x(t,a)=\left|\frac{1}{\sqrt{a}}\right| x(t) * \bar{\psi}\left(\frac{-t}{a}\right) \tag{6}$$

, where a and * are the scale factor and convolution operator, respectively, and $\bar{\psi}$ is complex conjugate of the mother wavelet ψ. The mother wavelet function for the Morlet wavelet is as follows (Jafarizadeh et al., 2008):

$$\psi=\frac{1}{\sqrt{\pi f_b}} e^{j2\pi f_c t} e^{\left(-t^2/f_b\right)} \tag{7}$$

, where f_b, and f_c are bandwidth parameter, and center frequency of Morlet wavelet, respectively. Then, the GI value is calculated for the squared envelope of the complex Morlet Wavelet coefficients.

The constraints used in (Albezzawy et al., 2019) are the constraints of the problem. Finally, the optimization problem of Morlet wavelet parameters is as follows:

$$\begin{aligned} &Maximize: \quad GI(x) \\ &s.t. \begin{cases} f_c/f_b > 3.5 \\ f_c + f_b/2 < f_s/2 \\ f_c - f_b/2 \le 30 f_r \end{cases} \end{aligned} \tag{8}$$

, where f_s and f_r are sampling frequency and rotational frequency, respectively.

The second stage of the optimization process involves finding the parameters *p, q, r, s,* and *β*. To achieve this, a fault mode selection index is defined. Specifically, the index value is calculated for each decomposed mode, and the mode yielding the highest index value is selected as the fault mode. Given that the primary feature of vibration signals from a faulty bearing is associated with impulsive peaks in the frequency domain, the WKI is employed for fault mode selection. The WKI is defined as follows: (Dibaj, Hassannejad, et al., 2021):

$$WKI=|C|.Ku\left[e_s\right] \tag{9}$$

, where *C*, e_s, *Ku*[.] are correlation between the input signal and decomposed mode, Hilbert envelope spectrum, and kurtosis index, respectively. The Kurtosis index of a signal $x(t)$ is as follows:

$$Ku[x(t)]=\frac{E\left[(x(t)-u)^4\right]}{\sigma^4} \tag{10}$$

, where u and σ are the mean and standard deviation of the input signal, respectively. $E[.]$ also represents the expected value.

Now, to optimize the parameters p, q, r, s, and β the objective function is the maximum WKI value among the calculated WKIs for decomposed modes and the constraints extracted in (Zhang et al., 2015) are used. So, the optimization problem after M decomposition level of the filterbank will be as follows:

$$\begin{aligned} &Maximize: \max\{WKI(D_1), WKI(D_2), ...WKI(D_{M+1})\} \\ &s.t. \begin{cases} 1-\frac{p}{q} \le \beta \le \frac{r}{s}, \\ 1 \le s/r \le 3, \\ \frac{1}{2} \le p/q \le 1 \\ 1 \le (p,q,r,s) \le 20, \ \beta = k\, r/s, \ 0 < k \le 1 \end{cases} \end{aligned} \tag{11}$$

, Where D_i 's are decomposed modes.

The number of decomposition levels, M is also determined for each input signal as follows (Zhang et al., 2020):

$$\begin{cases} M_{\max} = \dfrac{\log\left\{\frac{s}{r}.\frac{l_*}{N}\right\}}{\log(p/q)} + 1 \\ M = \min(M_{\max}, 20) \end{cases} \tag{12}$$

, where N is the length of the input signal and l_* is the length of M^{th} decomposed mode, which in the present study, is a number greater than 4.

3. PHYSICS-BASED NEURAL NETWORK

The network, introduced in the present work consists of two main parts: a physics-based part and a convolutional part. The physics-based part relies on physical equations and the extraction of the input signal features. In this part, firstly, vibration signals of the bearing are decomposed to sub-signals (modes) with the adaptive analytical filterbank, introduced in section 2. Secondly, the WKI is calculated for all decomposed modes, and the mode with the greatest index value is selected as the fault-sensitive mode. In the third step, the preprocessing kernel described in (Sadoughi & Hu, 2019) is convolved with the selected mode. This kernel is defined as follows:

$$\phi = A_0 \sum_{v=0}^{V} \left[U(\tau - \frac{v}{f_d}).\exp(-\xi(\tau - \frac{v}{f_d})) \right] \tag{13}$$

, where A_0 is a constant amplitude related to radial load and fault severity. U, ξ and f_d are unit step function, damping coefficient, and fault ballpass frequency, respectively. The ballpass frequencies of the bearing faults can be calculated from the relations, mentioned in (Randall, 2021). Also τ is the time index τ = 1,2, 3, …, τ_W, which τ_W is the kernel width and is much less than the length of the input signal. The upper limit of sigma, denoted as V, is assumed to be 20 in this study. The kernel diagram corresponding to the values $A_0 = 1$, $f_d = 105$, $\tau_W = 200$, and $\xi = 1500$ is illustrated in Figure (4)

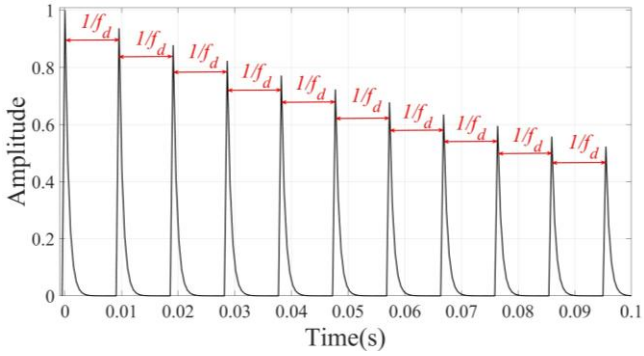


Figure 4. A sample of the physics-based kernel.

As shown in Figure (4), Eq. (13) generates harmonic impulsive peaks corresponding to the ball-pass frequencies of the bearing faults. Consequently, the convolution of the kernel with the fault-sensitive mode effectively extracts the physical features associated with the bearing faults. Thus, this kernel is employed to further expand and amplify the frequency characteristics of faulty bearings, which have been initially extracted by the optimal filterbank.

In the fourth step of the physics-based part of the proposed method, Hilbert envelope spectrum analysis is employed. The Hilbert transform has effective capabilities for frequency demodulation, which makes the fault's ballpass frequencies in the spectrum domain more prominent (Dibaj, Hassannejad, et al., 2021). As the final step, the output of the physics-based part is fed into the convolutional part, where the signals are classified according to the type of fault occurring in the bearing.

It should be noted that there are no learnable parameters in the physics-based component. Therefore, this part does not participate in network training and does not add any new

weights to the neural network. As a result, the physics-based component enhances the efficiency of the CNNs without increasing the learning time.

4. CONVOLUTIONAL NEURAL NETWORK

The convolutional part of the network is used to extract additional features and classify the input signals. In each convolutional layer (CL), the weights of the layer, known as the "kernel," are convolved with the input to calculate the output of the layer. Compared to fully connected layers, convolutional layers (CL) have fewer learnable weights, yet they are more effective at feature extraction (Sadoughi & Hu, 2019). Additionally, each CL has an activation function, which is applied to the layer's output. The output of a CL for the input P_L is as follows:

$$out = F(P_L * W + B) \tag{14}$$

, where W, B, and F are layer weights(kernel), layer bias, and activation function, respectively. The Rectified Linear Unit (ReLU) function is typically employed as a CL's activation function in classification tasks. The ReLU function is defined as follows:

$$\mathrm{Re}\,LU(x) = \max(0, x) \tag{15}$$

The structure of the CNN also includes additional layers. Input features are extracted through several convolutional layers (CLs) and pooling layers. Pooling layers are used to reduce the number of learnable weights between CLs. The output from these layers is then fed into the fully connected layers and classifier layer(s), producing the final output of the network. The classifier layer is responsible for classifying the network inputs into the desired categories. Layer Normalization Layers (LNLs) are also employed when the neural network inputs are vectors, helping to decrease training time and the sensitivity of the CNN to the initialization of the weights. LNLs are placed after the learnable layers, and by applying learnable scale and shift factors to the input, these layers normalize the input to the subsequent layer. Additionally, Residual Networks, a widely used deep neural network architecture, incorporate shortcut connections between CLs (He et al., 2016).

In this study, the bearing fault diagnosis performance of CNN and Residual architectures will be examined. Briefly, the output of the physics-based component is first normalized and then fed into the convolutional part for classification. The normalization of the physics-based output is performed using the z-score method as preprocessing for the convolutional part's inputs. The z-score method measures the distance of the data from the mean in terms of standard deviation ((Fei et al., 2021). Therefore, the input of the convolutional part will have a mean of zero and a standard deviation of 1. Since the inputs are vibration signals, LNLs are employed in place of middle pooling layers in this study, and a global average pooling layer will be applied after the last LNL. Also, the SoftMax classifier is used for the classification layer. Additionally, the He initialization method (He et al., 2015) was used for the weight's initialization, and the activation function of CLs will be the ReLU function. The two architectures used in the present research can be seen in Figure (5).

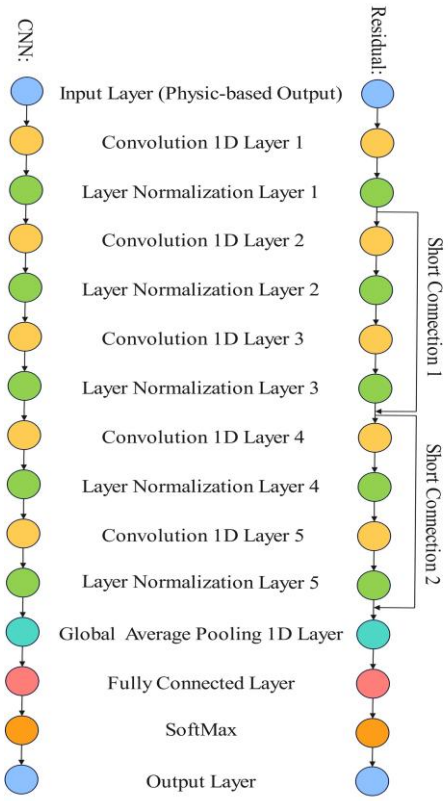


Figure 5. Architectures used for the convolution part.

The backpropagation algorithm is used to train the convolutional neural network. In brief, the algorithm works as follows: the input data is fed into the neural network, whose weights are initialized beforehand, and the network's final output is computed. Next, the loss value is calculated by comparing the predicted output with the expected output. This loss is then backpropagated through the network, and the gradient of each layer is computed with respect to its weights. Finally, gradient-based optimization algorithms are employed to update the weights of each layer to their optimal values.

In this work, the cross-entropy loss function and the Adam optimization algorithm were used for the backpropagation process. Training deep neural networks involves tuning several hyperparameters, such as the learning rate, regularization rate, batch size, and others, which must be carefully determined. In this study, a grid search was utilized to identify the optimal values for these hyperparameters.

5. Experimental RESULTS

In this section, the proposed fault diagnosis method is evaluated by two experimental bearing datasets. Case Western Reserve University (CWRU) bearing data (Lou & Loparo, 2004) and Paderborn University bearing dataset (Lessmeier et al., 2016).

5.1 Case Study #1: CWRU bearing dataset

The CWRU dataset comprises data from bearings under various conditions. For this study, healthy bearing data (H) and a subset of Drive End Bearing Fault Data, sampled at 12 kHz, were utilized. The dataset includes three fault states for faulty bearings: inner ring fault (IR), ball fault (B), and outer ring fault (OR). Additionally, three modes of signal collection are defined for outer ring faults based on the fault position relative to the load zone: 'centered' (6.00 o'clock position), 'orthogonal' (3.00 o'clock), and 'opposite' (12.00 o'clock). This work focuses exclusively on data from the 'centered' mode. Since the CWRU dataset uses numerical file names to denote different tests, the specifications of the data employed in this study are detailed in Table (1). For each test, acceleration was recorded in the vertical direction on the housings of the drive-end bearing (DE) and fan-end bearing (FE), as well as on the motor supporting base plate (BA). The bearing model used is also SKF 6205-2RS JEM. Figure (6) shows the experimental test equipment.

It should be noted that Smith and Randall (Smith & Randall, 2015) conducted a benchmark study on the CWRU dataset, identifying datasets that could not be detected using conventional methods. Their work serves as a key reference for utilizing the CWRU dataset in subsequent research. Table (2) lists the data files where Smith and Randall's methods failed to detect bearing faults but which were included in the evaluation of the proposed method in this study.

In this case study, the length of the selected signals is 4096 samples. Additionally, the ballpass frequencies of the faults, as reported by CWRU, are provided in Table 3. Since the rotational speeds range from 1730 to 1797 RPM (28.83 to 29.95 Hz), the average ballpass frequencies for all three fault types were used to calculate the kernel ϕ (Eq. (13)). Subsequently, vibration signals from bearings with varying rotational speeds and fault diameters were utilized simultaneously to evaluate the performance of the proposed neural network. This way of selecting input signals, reflects the real-world operating conditions of industrial rotating machinery, where rotational speed can vary within small tolerances. faults in such machinery can lead to faults of varying severity.

It should be noted that, based on the number of available samples, there are 825 signals from healthy bearings. To ensure homogeneity across different classes, the same number of signals was selected for bearings with different fault conditions. As a result, the total dataset consists of 3,300 signals. These signals were divided such that 70% were used for training, 15% for validation, and 15% for testing. Table 4 summarizes the number of signals corresponding to each class, while Table 5 provides details on the network architecture and the specifications of each layer.

Fault width (mm)	load (hp)	Shaft speed (rpm)	Dataset file name for each fault type			
			H	IR	B	OR
-	0	1797	97	-	-	-
	1	1772	98	-	-	-
	2	1750	99	-	-	-
	3	1730	100	-	-	-
0.18	0	1797	-	105	118	130
	1	1772	-	106	119	131
	2	1750	-	107	120	132
	3	1730	-	108	121	133
0.36	0	1797	-	169	185	197
	1	1772	-	170	186	198
	2	1750	-	171	187	199
	3	1730	-	172	188	200
0.53	0	1797	-	209	222	234
	1	1772	-	210	223	235
	2	1750	-	211	224	236
	3	1730	-	212	225	237

Table 1. CWRU dataset specifications used in this study.

Partially diagnosable Dataset file name	**Non-diagnosable Dataset file name**
169BA, 170FE, 170BA, 171FE, 171BA, 172FE, 172BA, 120FE, 121FE, 185FE, 185BA, 186DE, 186BA, 187BA, 188, 222FE, 223DE, 224FE, 225BA, 199BA	118, 119, 120DE, 120BA, 121BA, 187FE, 224DE, 224BA, 225DE, 225FE 197FE, 197BA, 198FE, 198BA, 199FE, 200

Table 2. Non-diagnosable and partially diagnosable data from the CWRU dataset using the benchmark method (Smith & Randall, 2015).

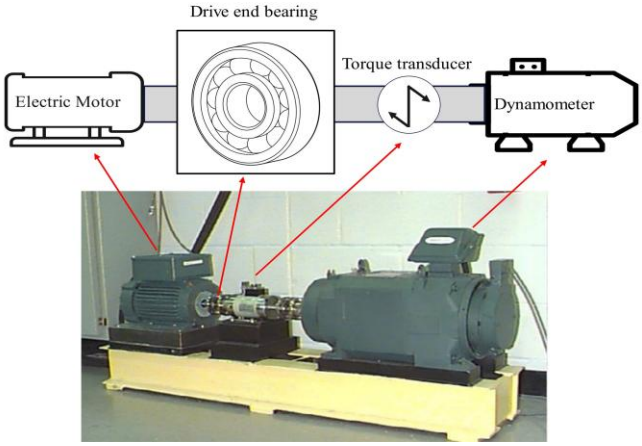


Figure 6. Experimental test equipment of case study #1 (Lou & Loparo, 2004).

Inner Ring	Outer Ring	Rolling Element
$5.4152\times f_r$	$3.5848\times f_r$	$4.7135\times f_r$

Table 3. The ballpass frequency of bearing model 6205.

	Number of signals		
	Train	**Val**	**Test**
H	577	124	124
IR	577	124	124
B	577	124	124
OR	577	124	124
Sum	2308	496	496

Table 4. Number of signals per class in the case study #1.

The first five layers in Table (5) correspond to the physics-based part of the network. The number of kernels in the filterbank represents the number of filters through which the input signal passes. Additionally, the length of the output signals from the filterbank depends on the parameters of the up-sampler and down-sampler, which are adaptively determined for each input signal. To standardize the output, all signals from the filterbank are zero-padded on the right side to achieve a uniform length of 4096 samples.

To calculate ϕ (layer No. 4), the ballpass frequencies of bearing faults, as provided in Table (3), were utilized. This layer generates three output vectors of identical length for each input vector of a given length. Following the Hilbert envelope spectrum layer (layer No. 5), the primary features are concentrated at the beginning of the output signals, while the remaining values are close to zero. As a result, only the first 500 samples of the output vectors are retained, and the subsequent samples are disregarded.

In the convolutional part of the network, the size of all kernels is set to one, except for the fully connected layer. The kernel size in the fully connected layer (layer No. 17) is designed to produce an output vector with a length equal to the number of classes. In this case study, there are four classes: healthy, outer race fault, inner race fault, and ball fault. For training the network, mini-batches of size 105 were used, along with the Adam optimization algorithm, a learning rate of 0.001, and a regularization rate of 0.01.

The proposed physics-based network was evaluated for fault diagnosis in Case Study #1, and the classification results are presented in Table (6). For comparison, the classification results using CNN and Residual networks without the physics-based part were also evaluated and are included in Table (6). Additionally, several fault diagnosis methods proposed by researchers in recent years were evaluated using the CWRU dataset under the same conditions outlined in this paper, and their results are also summarized in Table (6). The compared methods include:

1) A CNN structure introduced by (Ji et al., 2022), (we call it JI CNN)
2) A Multiscale Local Feature Learning method based on Back-Propagation Neural Network (MLFL + BPNN) proposed by (Li et al., 2020)
3) A Physics-based Convolutional Neural Network (PCNN) presented by (Sadoughi & Hu, 2019).

As evident from the results in Table (6), the inclusion of physics-based layers has significantly improved the accuracy of deep neural networks for bearing fault classification. Furthermore, the physics-based CNN network demonstrates slightly higher accuracy for both training and test data compared to the physics-based Residual network. In addition, the accuracy of other similar methods for bearing fault diagnosis is considerably lower than that of the proposed method. These comparable methods only address fault diagnosis for bearings with a specific rotational speed. As a result, they perform poorly in classifying bearing signals with multiple rotational speeds. The confusion matrix for the best-performing proposed neural network is illustrated in Figure (7).

Upon careful examination of Figure (7), it is evident that the data related to the bearing's outer race fault are accurately classified during both the training and testing stages. This outcome suggests that the outer race fault data contain less noise, which aligns with practical observations, as the proximity of the outer race to the data acquisition sensor typically results in clearer signals. Additionally, the healthy bearing data are well-classified during the training stage, though some misclassification occurs during the testing stage. This discrepancy may arise because the network's physics-based foundation is primarily designed using features from unhealthy bearings rather than healthy ones.

No.	Layer type	Number of kernels		stride		Output shape (width × depth)		padding	Learnable parameters	
1	Input	-		-		4096×1		False	-	
Physics-based Part										
2	Adaptive Filterbank	2*M*		1		4096×(*M*+*1*)		True	-	
3	Mode selection	-		-		4096×1		False	-	
4	ϕ	3		1		4096×3		False	-	
5	Envelope spectrum	-		-		500×3		False	-	
Convolutional Part										
-	-	CNN	Residual	CNN	Residual	CNN	Residual	-	CNN	Residual
6	CL1	300	300	3	3	167×300	167×300	True	3×300 +300	3×300 +300
7	LNL1	300	300	1	1	167×300	167×300	False	300+300	300+300
8	CL2	250	250	3	1	56×250	167×250	True	300×250 +250	300×250 +250
9	LNL2	250	250	1	1	56×250	167×250	False	250+250	250+250
10	CL3	125	300	1	1	56×125	167×300	True	250×125 +125	250×300 +300
11	LNL3	125	300	1	1	56×125	167×300	False	125+125	300+300
12	CL4	100	100	1	1	56×100	167×100	True	125×100 +100	300×100 +100
13	LNL4	100	100	1	1	56×100	167×100	False	100+100	100+100
14	CL5	50	300	1	1	56×50	167×300	True	100×50 +50	100×300 +300
15	LNL5	50	300	1	1	56×50	167×300	False	50+50	300+300
16	GAP*	50	300	1	1	56×1	167×1	False	-	-
17	FC*	1	1	1	1	1×4	1×4	False	4×50+4	4×300+4
18	SoftMax	1	1	-	-	1×4	1×4	-	-	-
									$\sum = 127329$	$\sum = 215854$

* GAP: global average poling, FC: Fully connected.

Table 5. Architecture and details of presented physics-based neural networks.

	Training	validation	test
Proposed Physics-based CNN	**99.57**	91.31	**92.53**
Proposed Physics-based Residual Neural Network	98.09	**92.32**	92.12
CNN	87.92	87.07	85.45
Residual Neural Network	91.52	88.48	87.71
JI CNN(Ji et al., 2022)	54.85	52.93	53.54
MLFL + BPNN (Li et al., 2020)	52.42	49.90	51.52
PCNN (Sadoughi & Hu, 2019)	63.81	59.39	57.17

Table 6. Classification results of different methods for case study#1.

For further investigation, Figure (8) illustrates the initial raw data, the outputs of the physics-based part, and the final output of the neural network for both the training and test data of Case Study #1, presented in three dimensions. To achieve this visualization, the t-distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction method was applied to reduce the data dimensions to three dimensions. It is important to note that t-SNE is not a part of the proposed method; rather, it is solely employed for dimensionality reduction to enable the visualization of outputs from different layers.

A careful examination of Figure (8) reveals that the initial input data is completely mixed and inseparable (left column). However, after applying the physics-based layers, the input data appears to be grouped into smaller categories with similar characteristics (middle column). Finally, by feeding the extracted features from the physics-based layers into the convolutional layers, deeper features are learned, resulting in a well-separated distribution of data for each class (right column).
Similarly, the final output of other methods for the test data is plotted in Figure (9). From Figure (9), it is evident that none of the comparable methods achieve proper separation of the input data. Additionally, convolutional neural networks without the physics-based component fail to produce acceptable results and do not achieve good separation. These findings highlight the critical role of physics-based layers in creating an initial separation of the data. Deep neural networks perform significantly better in classifying inputs that have undergone multiple stages of pre-classification using physics-based layers.

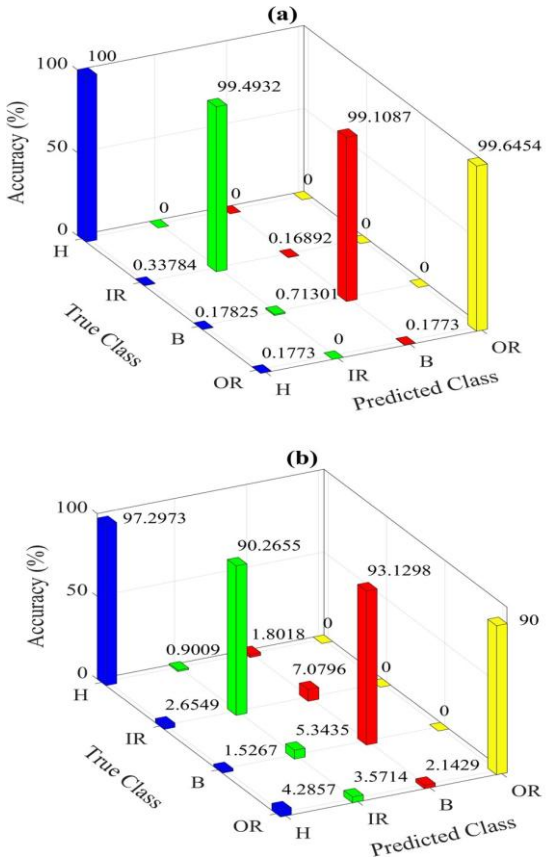


Figure 7. Confusion matrix of proposed physics-based CNN for case study#1 a) training data b) test data. (H: Healthy, IR: Inner Race Fault, B: Ball Fault, OR: Outer Race Fault)

5.2 Case Study #2: Paderborn University bearing dataset

The Paderborn University bearing dataset includes vibration signals obtained from a bearing model 6203. When recording vibration signals from the bearing, three states—healthy, outer ring fault, and inner ring fault—are considered. In this dataset, the signals of faulty bearings are measured under two different scenarios. In one scenario, the bearing damage is artificially induced, while in the other, damage occurs due to an accelerated lifetime test. The faulty bearing rotates at 1500 rpm. Additionally, radial forces of 400 N and 1000 N are applied with load torques of 0.1 N·m and 0.7 N·m. A sampling frequency of 64 kHz was used to record vibration data from bearings with different fault severities.

Figure 10 displays the experimental equipment diagram for Case Study #2. The geometric specifications of the bearing are listed in Table 7.

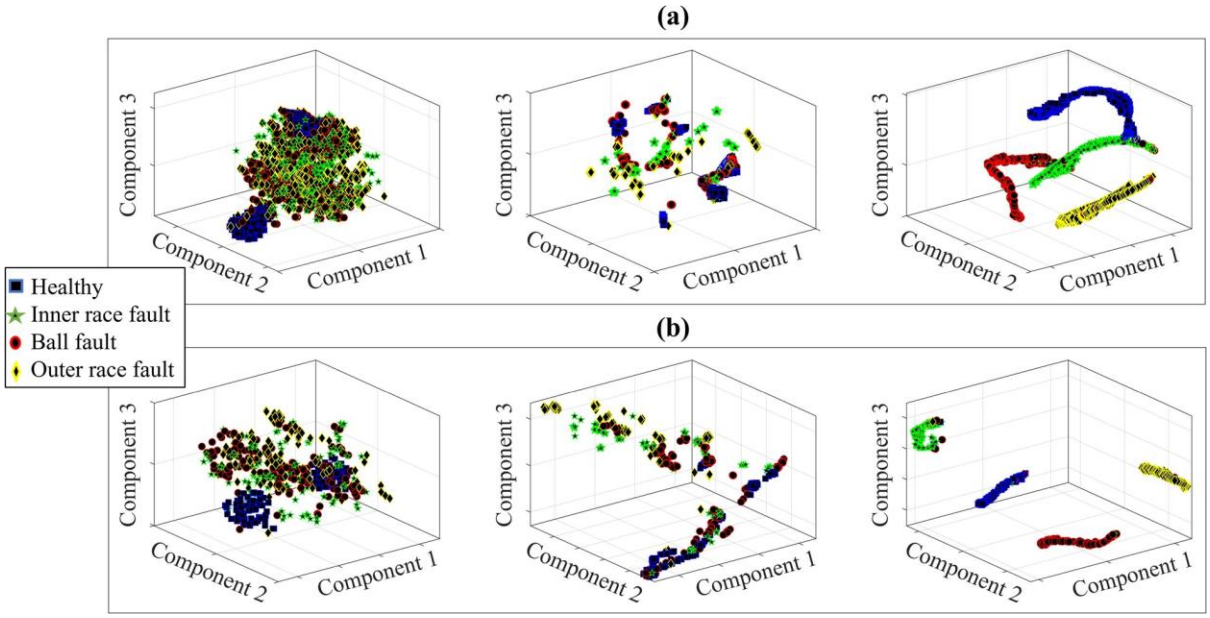


Figure 8. 3D representation of the three main layers of the proposed physics-based CNN for a) training data and b) test data of case study#1.
(**Left:** raw input data. **Middle:** output of the physics-based layers. **Right:** final output of the neural network).

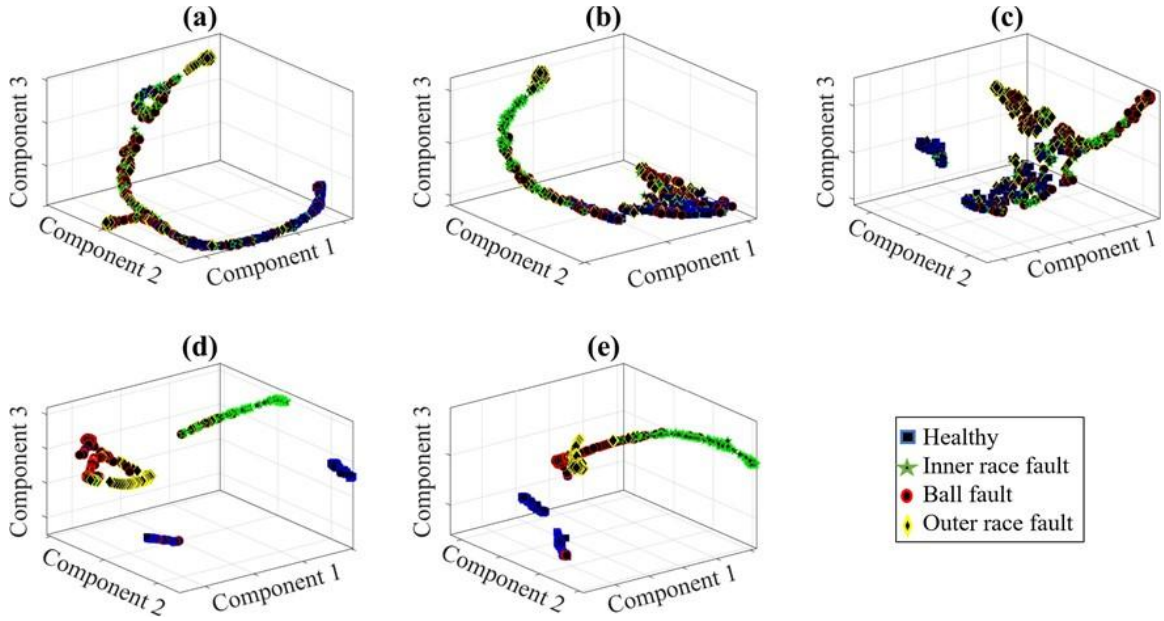


Figure 9. 3D representation of the outputs of the compared methods (a) JI CNN(Ji et al., 2022), (b) Multiscale local feature learning + BPNN (Li et al., 2020), (c) PCNN (Sadoughi & Hu, 2019), (d) CNN and (e) Residual neural network, after applying test data of the case study#1.

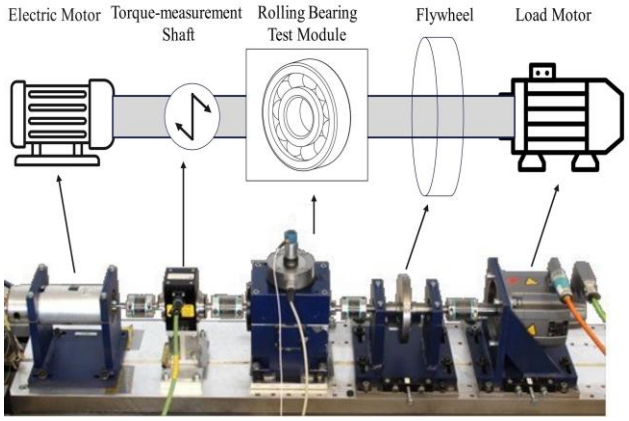


Figure 10. Experimental equipment of Paderborn University bearing dataset.

Diameter of inner raceway	24.0 mm
Diameter of outer raceway	33.1 mm
Pitch diameter	28.55 mm
Number of rolling elements	8
Rolling element diameter	6.75
Contact angle	0º

Table 7. Geometric specifications of bearing model 6203.

Therefore, according to the geometric characteristics of the bearing in Table (7) and the ball pass frequency formulas provided in (Randall, 2021), the ballpass frequency of the outer race fault (f_{OR}) and inner race fault (f_{IR})of the bearing used in the case study#2 is:

$$f_{OR} = 76.3573Hz \quad , \quad f_{IR} = 123.6427Hz$$

Due to the absence of a ball fault in Case Study #2, only the ball-pass frequencies of the outer race fault and the inner race fault will be present in the kernel ϕ calculation for this case. Additionally, the length of the input signals has been set to 64,000 samples. Therefore, the only difference between the neural network used for Case Study #2 and Case Study #1 is the size of the first four layers in the physics-based part. Specifically, in layers 1 to 4 of Table (5), the value of 4096 (input signal length) is replaced with 64,000, and the number of kernels in the fourth layer is set to 2. According to the number of available signals and the selected length for the input signals, there will be 1200 signals for each class (healthy, outer race fault, and inner race fault) in each scenario related to artificial damage and real damage. So, the total number of data is 3600 signals for each scenario. Among these, 75% of the data are considered for training, 12.5% for validation, and 12.5% for the test process. Table (8) summarizes the number of signals corresponding to each class.

	Number of signals		
	Train	**Val**	**Test**
H	900	150	150
IR	900	150	150
OR	900	150	150
Sum	2700	450	450

Table 8. Number of signals per class in the case study #2.

A mini-batch size of 100, a learning rate of 0.001, and a regularization rate of 0.01 were used to train the physics-based neural networks in Case Study #2. After applying the proposed method, the classification accuracy of the introduced approach, along with the compared methods, is listed in Table (9).

Table (9) demonstrates that the accuracy of the proposed physics-based method in diagnosing and classifying faulty bearings is significantly higher than that of other methods. Notably, the accuracy of the introduced method in diagnosing bearings with real damage is higher than its accuracy in diagnosing bearings with artificial damage. This outcome can be attributed to the fact that faults in real applications tend to be more severe than those artificially induced under controlled conditions. Regardless, the high accuracy of the proposed method in diagnosing and classifying bearing faults in real-world scenarios highlights its effectiveness for industrial and practical applications.

The confusion matrix of the introduced physics-based neural network for classifying vibration signals of bearings with artificial and real damage is shown in Figures (11) and (12), respectively. A careful examination of Figures (11) and (12) reveals that the proposed method successfully distinguishes healthy bearings from faulty ones with an accuracy of approximately 100%. The minor classification errors in the introduced neural network stem from occasional misclassification of signals related to inner race faults and outer race faults. In other words, the proposed method achieves a fault detection accuracy of 100% and fault diagnosis accuracy of 90% and 95% for artificially damaged and real damaged bearings, respectively.

In the following, the 3D representations of the input data, the output of the physics-based part, and the final output of the introduced neural network for both scenarios in Case Study #2 (bearings with artificial damage and bearings with real damage) are shown in Figures (13) and (14). Additionally, the 3D representations of the final outputs of the compared methods on the test data are presented in Figures (15) and (16).

Bearing fault condition	Test bearings with artificial damage			Test bearings with real damages		
	Training	Validation	Test	Training	Validation	Test
Proposed Physics-based CNN	96.41	85.56	83.11	**100**	**100**	**95.56**
Proposed Physics-based Residual Neural Network	**92.89**	**92.71**	**90.44**	**100**	**100**	**95.56**
CNN	65.63	60.22	61.78	81.63	80.00	80.00
Residual Neural Network	84.69	82.67	79.78	84.26	83.78	83.11
JI CNN(Ji et al., 2022)	69.33	69.11	61.15	48.41	45.78	44.67
MLFL + BPNN (Li et al., 2020)	49.41	49.33	46.44	55.07	52.22	51.11
PCNN (Sadoughi & Hu, 2019)	77.93	76.00	74.22	73.56	68.67	64.40

Table 9. Classification results of different methods for case study#2.

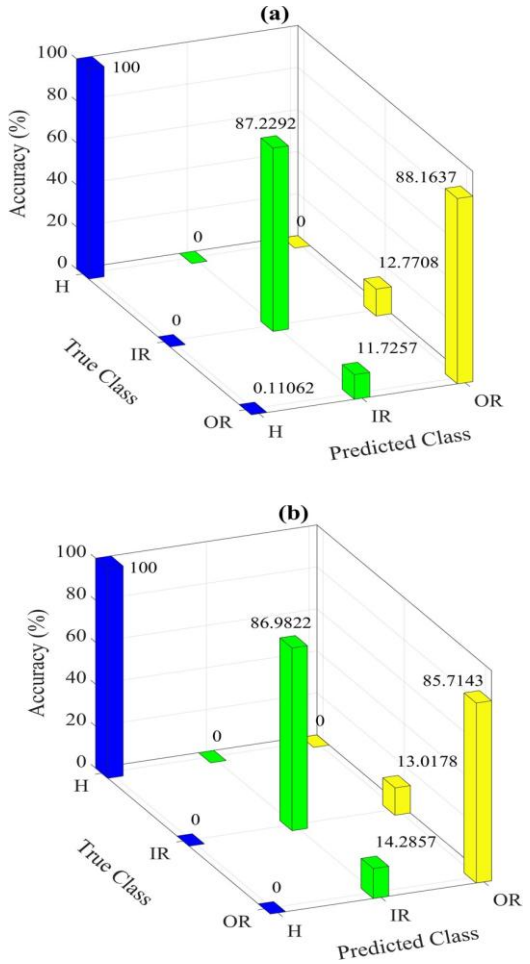


Figure 11. Confusion matrix of the proposed physics-based Residual neural network for the **artificially damaged** bearing of case study#2, a) training data b) test data
(H: Healthy, IR: Inner Race Fault, OR: Outer Race Fault).

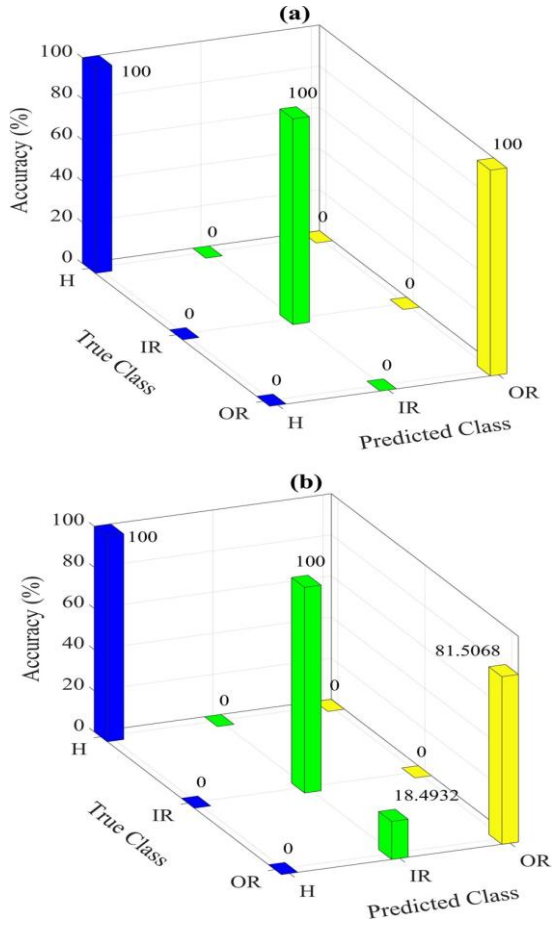


Figure 12. Confusion matrix of the proposed physics-based CNN for the **really damaged** bearing of case study#2
a) training data, b) test data
(H: Healthy, IR: Inner Race Fault, OR: Outer Race Fault).

By closely examining the middle columns of Figures (13) and (14), it is evident that the introduced physics-based part has effectively separated the initially mixed-up input data (left columns of Figures (13) and (14)) through a pre-classification process. Subsequently, the CNN further classifies this pre-classified data, with the final output shown in the right columns of Figures (13) and (14). This pre-classification process is the key advantage of the introduced physics-based part, significantly enhancing the proposed method compared to other approaches. Comparing Figure (13.b, right) with Figure (15) and Figure (14.b, right) with Figure (16) clearly demonstrates that the proposed method achieves superior classification of the desired fault classes compared to other similar methods.

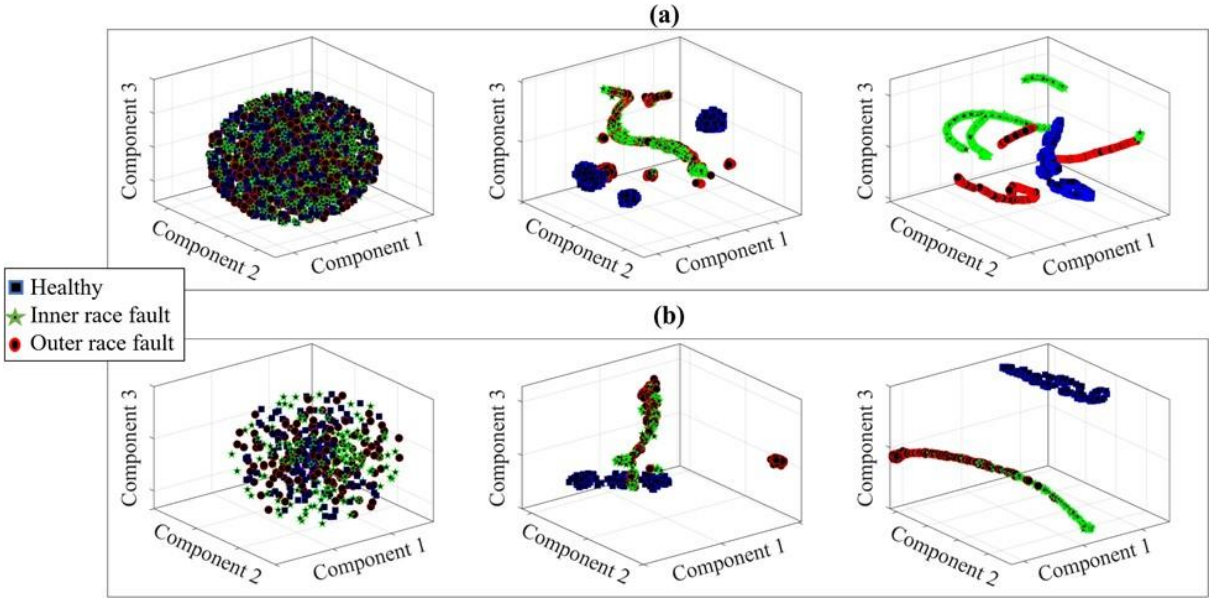


Figure 13. 3D representation of the three main layers of the proposed physics-based CNN for a) training data and b) test data of the **artificially damaged** bearing of case study#2.
(**Left:** raw input data. **Middle:** output of the physics-based layers. **Right:** final output of the neural network)

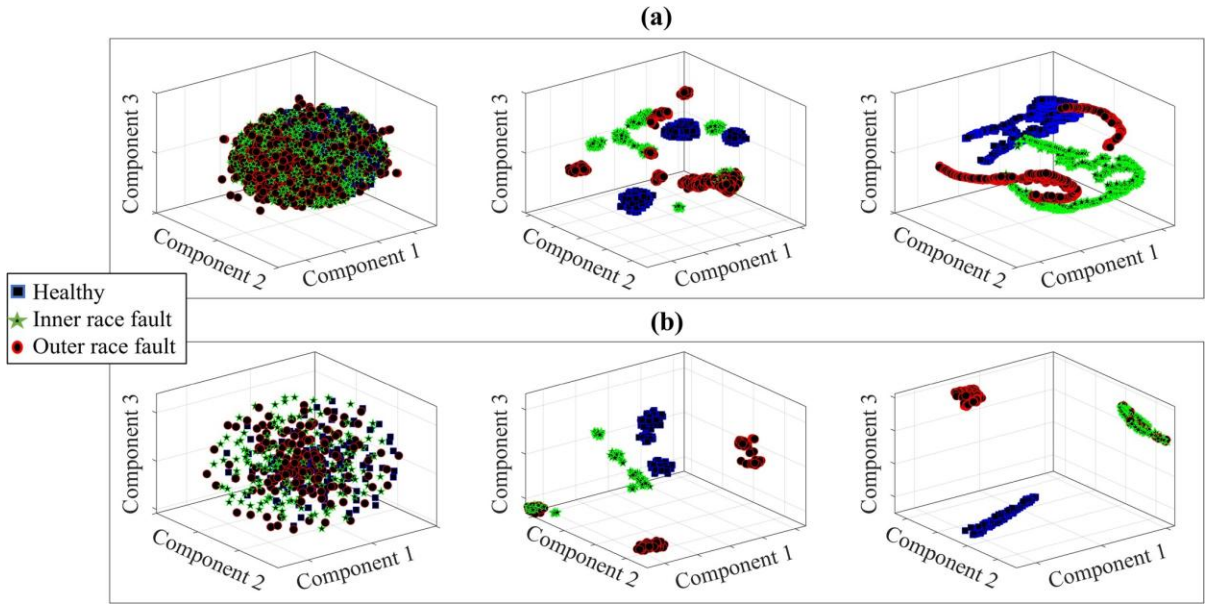


Figure 14. 3D representation of the three main layers of the proposed physics-based CNN for a) training data and b) test data of the **really damaged** bearing of the case study#2.
(**Left:** raw input data. **Middle:** output of the physics-based layers. **Right:** final output of the neural network)

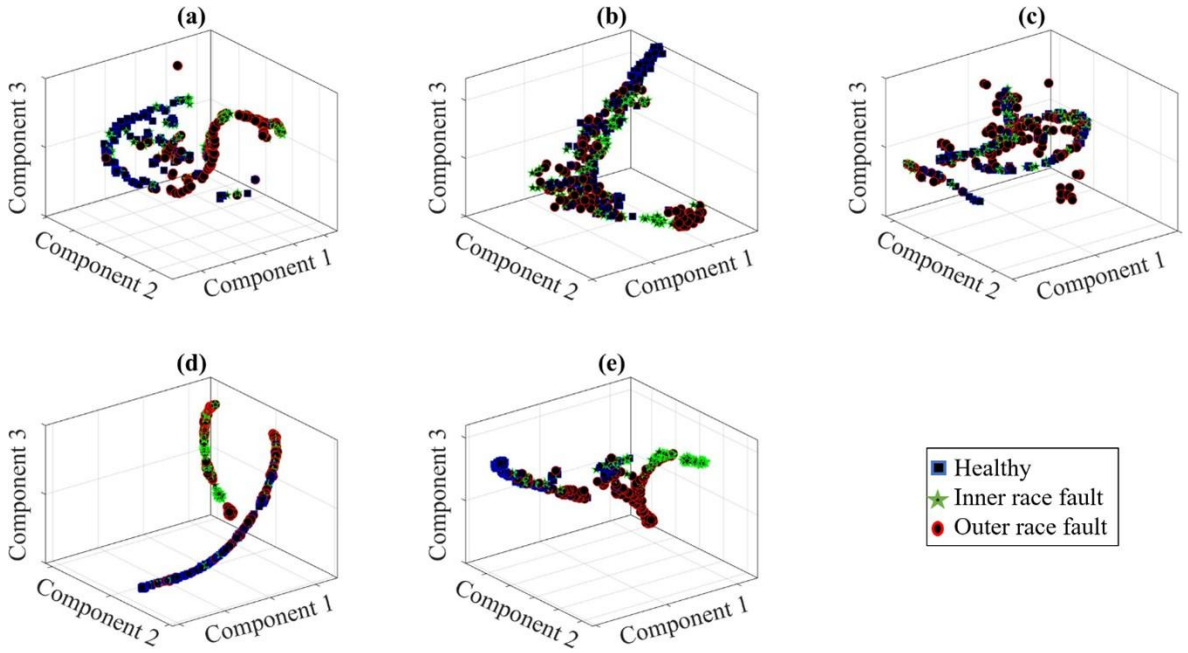


Figure 15. 3D representation of the outputs of the compared methods (a) JI CNN (Ji et al., 2022), (b) Multiscale local feature learning + BPNN (Li et al., 2020), (c) PCNN (Sadoughi & Hu, 2019), (d) CNN and (f) Residual neural network, after applying test data of the **artificially damaged** bearing of case study#2.

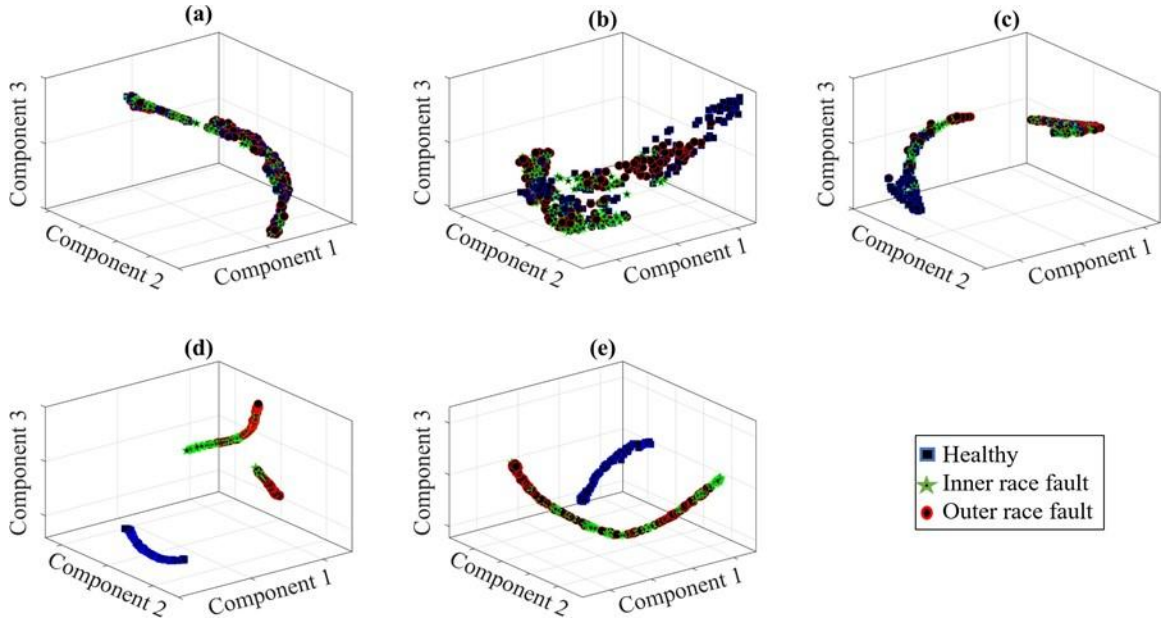


Figure 16. 3D representation of the outputs of the compared methods (a) JI CNN (Ji et al., 2022), (b) Multiscale local feature learning + BPNN (Li et al., 2020), (c) PCNN (Sadoughi & Hu, 2019), (d) CNN and (f) Residual neural network, after applying test data of the **really damaged** bearing of case study#2.

6. Conclusion

This research proposes a physics-based neural network architecture for bearing fault diagnosis. In the proposed method, vibration signals are first fed into the physics-based layers, which include an adaptive analytical filterbank, a proper mode selector, a convolution layer with a physics-based kernel, and Hilbert envelope spectrum analysis. In this stage, the main physical features of bearing faults are extracted from the vibration signals. Next, the outputs of the physics-based part are passed to the convolutional part, which consists of convolutional layers, layer normalization layers, average pooling, a fully connected layer, and a SoftMax classifier. This part further extracts features and classifies the signals into four categories: healthy bearings, bearings with inner race faults, bearings with outer race faults, and bearings with ball faults. The proposed physics-based neural network is evaluated using two experimental datasets.

The results indicate that the proposed network achieves high accuracy in diagnosing and classifying bearing faults. Additionally, the physics-based part performs a pre-classification process on the raw input data, leading to higher classification accuracy compared to similar methods that rely solely on deep neural networks. The results of Case Study #1 (bearing data with different fault severities and rotational speeds) demonstrate the effectiveness of the proposed method in industrial and experimental fault diagnosis. Moreover, the network's sensitivity to key bearing fault features, such as impulsive peaks in the frequency domain, is evident. The results from Case Study #2 confirm that the proposed method achieves higher classification accuracy than other similar approaches, both for artificially induced faults and faults caused by lifetime testing. Furthermore, the method exhibits near-perfect fault detection performance, with minor errors arising only in distinguishing between different faulty bearing conditions. Additionally, the use of one-dimensional convolutional layers instead of two-dimensional convolutions significantly reduces computational costs compared to networks that process images. Overall, the findings confirm that the introduced method is highly accurate and efficient for the experimental diagnosis of bearing faults.

7. Limitations and Future Work

Although the proposed method achieves acceptable results in bearing fault diagnosis, it is not without limitations. The first limitation, common to all physics-based fault diagnosis methods, is their generalizability. Since physics-based fault diagnosis methods are designed based on the physics of a specific component, they cannot be directly applied to diagnose faults in other components. The second limitation is that implementing physics-based methods for complex mechanical systems can be challenging. In complex systems, the physical properties of mechanical faults may depend on many factors, and fault characteristics may not always be well-defined. As a result, incorporating the physical features of the problem into the fault diagnosis method introduces significant complexity. The third limitation is that the proposed fault diagnosis method in this work is designed based on the ball-pass frequency of bearing faults, which is directly related to the rotational speed of the bearing. This poses a challenge for the effectiveness of the proposed method under variable speed conditions.

In this regard, potential areas for future work include generalizing physics-based models to different rotating machinery and components, extending these models for diagnosing combined faults in rotating machinery, and developing the proposed model for variable-speed and low-speed conditions.

References

Ahmed, H., & Nandi, A. K. (2020). *Condition monitoring with vibration signals: Compressive sampling and learning algorithms for rotating machines*. John Wiley & Sons.

Albezzawy, M., Nassef, M., Elsayed, E., & Elkhatib, A. (2019). Early rolling bearing fault detection using a Gini index guided adaptive Morlet wavelet filter. 2019 IEEE 10th International Conference on Mechanical and Aerospace Engineering (ICMAE),

Albezzawy, M. N., Nassef, M. G., & Sawalhi, N. (2020). Rolling element bearing fault identification using a novel three-step adaptive and automated filtration scheme based on Gini index. *ISA transactions*, *101*, 453-460. https://doi.org/https://doi.org/10.1016/j.isatra.2020.01.019

Asr, M. Y., Ettefagh, M. M., Hassannejad, R., & Razavi, S. N. (2017). Diagnosis of combined faults in rotary machinery by non-naive Bayesian approach. *Mechanical Systems and Signal Processing*, *85*, 56-70. https://doi.org/https://doi.org/10.1016/j.ymssp.2016.08.005

Bayram, I. (2012). An analytic wavelet transform with a flexible time-frequency covering. *IEEE Transactions on Signal Processing*, *61*(5), 1131-1142. https://doi.org/https://doi.org/10.1109/TSP.2012.2232655

Chen, J., Li, Z., Pan, J., Chen, G., Zi, Y., Yuan, J., Chen, B., & He, Z. (2016). Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, *70*, 1-35. https://doi.org/https://doi.org/10.1016/j.ymssp.2015.08.023

Cheng, Y., Lin, M., Wu, J., Zhu, H., & Shao, X. (2021). Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network. *Knowledge-Based Systems*, *216*, 106796. https://doi.org/https://doi.org/10.1016/j.knosys.2021.106796

Dibaj, A., Ettefagh, M. M., Hassannejad, R., & Ehghaghi, M. B. (2020). Fine-tuned variational mode decomposition for fault diagnosis of rotary machinery. *Structural Health Monitoring*, *19*(5), 1453-1470. https://doi.org/https://doi.org/10.1177/1475921719887496

Dibaj, A., Ettefagh, M. M., Hassannejad, R., & Ehghaghi, M. B. (2021). A hybrid fine-tuned VMD and CNN scheme for untrained compound fault diagnosis of rotating machinery with unequal-severity faults. *Expert Systems with Applications*, *167*, 114094. https://doi.org/https://doi.org/10.1016/j.eswa.2020.114094

Dibaj, A., Hassannejad, R., Ettefagh, M. M., & Ehghaghi, M. B. (2021). Incipient fault diagnosis of bearings based on parameter-optimized VMD and envelope spectrum weighted kurtosis index with a new sensitivity assessment threshold. *ISA transactions*, *114*, 413-433. https://doi.org/https://doi.org/10.1016/j.isatra.2020.12.041

Fei, N., Gao, Y., Lu, Z., & Xiang, T. (2021). Z-score normalization, hubness, and few-shot learning. Proceedings of the IEEE/CVF International Conference on Computer Vision,

Feldman, M. (2009). Analytical basics of the EMD: Two harmonics decomposition. *Mechanical Systems and Signal Processing*, *23*(7), 2059-2071. https://doi.org/https://doi.org/10.1016/j.ymssp.2009.04.002

Guo, L., Lei, Y., Xing, S., Yan, T., & Li, N. (2018). Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics*, *66*(9), 7316-7325. https://doi.org/10.1109/TIE.2018.2877090

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Proceedings of the IEEE international conference on computer vision,

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition,

Hurley, N., & Rickard, S. (2009). Comparing measures of sparsity. *IEEE Transactions on Information Theory*, *55*(10), 4723-4741.

Jafarizadeh, M., Hassannejad, R., Ettefagh, M., & Chitsaz, S. (2008). Asynchronous input gear damage diagnosis using time averaging and wavelet filtering. *Mechanical Systems and Signal Processing*, *22*(1), 172-201. https://doi.org/https://doi.org/10.1016/j.ymssp.2007.06.006

Ji, M., Peng, G., Li, S., Cheng, F., Chen, Z., Li, Z., & Du, H. (2022). A neural network compression method based on knowledge-distillation and parameter quantization for the bearing fault diagnosis. *Applied Soft Computing*, *127*, 109331. https://doi.org/https://doi.org/10.1016/j.asoc.2022.109331

Kim, Y., & Kim, Y.-K. (2024). Physics-informed time-frequency fusion network with attention for noise-robust bearing fault diagnosis. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3355268

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, *138*, 106587. https://doi.org/https://doi.org/10.1016/j.ymssp.2019.106587

Lessmeier, C., Kimotho, J. K., Zimmer, D., & Sextro, W. (2016). Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. PHM Society European Conference,

Li, J., Yao, X., Wang, X., Yu, Q., & Zhang, Y. (2020). Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis. *Measurement*, *153*, 107419. https://doi.org/https://doi.org/10.1016/j.measurement.2019.107419

Liao, J.-X., He, C., Li, J., Sun, J., Zhang, S., & Zhang, X. (2025). Classifier-guided neural blind deconvolution: A physics-informed denoising module for bearing fault diagnosis under noisy conditions. *Mechanical Systems and Signal Processing*, *222*, 111750. https://doi.org/https://doi.org/10.1016/j.ymssp.2024.111750

Lou, X., & Loparo, K. A. (2004). Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical Systems and Signal Processing*, *18*(5), 1077-1095. https://doi.org/https://doi.org/10.1016/S0888-3270(03)00077-3

Ma, Y., Yang, J., & Li, L. (2022). Collaborative and adversarial deep transfer auto-encoder for intelligent fault diagnosis. *Neurocomputing*, *486*, 1-15. https://doi.org/https://doi.org/10.1016/j.neucom.2022.02.050

Ma, Z., Fu, L., Xu, F., & Zhang, L. (2025). A physics-based sample generation method for few-shot bearing

condition monitoring. *Knowledge-Based Systems*, *310*, 112952. https://doi.org/https://doi.org/10.1016/j.knosys.2024.112952

Mao, W., Ding, L., Liu, Y., Afshari, S. S., & Liang, X. (2022). A new deep domain adaptation method with joint adversarial training for online detection of bearing early fault. *ISA transactions*, *122*, 444-458. https://doi.org/https://doi.org/10.1016/j.isatra.2021.04.026

Maruthi, G., & Hegde, V. (2015). Application of MEMS accelerometer for detection and diagnosis of multiple faults in the roller element bearings of three phase induction motor. *IEEE Sensors Journal*, *16*(1), 145-152. https://doi.org/10.1109/JSEN.2015.2476561

Mohanty, A. R. (2014). *Machinery condition monitoring: Principles and practices*. CRC Press.

Pestana-Viana, D., Zambrano-López, R., De Lima, A. A., Prego, T. d. M., Netto, S. L., & da Silva, E. A. (2016). The influence of feature vector on the classification of mechanical faults using neural networks. 2016 IEEE 7th Latin American Symposium on Circuits & Systems (LASCAS),

Qian, C., Zhu, J., Shen, Y., Jiang, Q., & Zhang, Q. (2022). Deep transfer learning in mechanical intelligent fault diagnosis: application and challenge. *Neural Processing Letters*, *54*(3), 2509-2531. https://doi.org/https://doi.org/10.1007/s11063-021-10719-z

Qin, Y., Xing, J., & Mao, Y. (2016). Weak transient fault feature extraction based on an optimized Morlet wavelet and kurtosis. *Measurement Science and Technology*, *27*(8), 085003. https://doi.org/https://doi.org/10.1088/0957-0233/27/8/085003

Rai, A., & Upadhyay, S. H. (2016). A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International*, *96*, 289-306. https://doi.org/https://doi.org/10.1016/j.triboint.2015.12.037

Randall, R. B. (2021). *Vibration-based Condition Monitoring: Industrial, Automotive and Aerospace Applications*. John Wiley & Sons.

Sadoughi, M., & Hu, C. (2019). Physics-based convolutional neural network for fault diagnosis of rolling element bearings. *IEEE Sensors Journal*, *19*(11), 4181-4192. https://doi.org/10.1109/JSEN.2019.2898634

Sharma, S., Tiwari, S., & Singh, S. (2021). Integrated approach based on flexible analytical wavelet transform and permutation entropy for fault detection in rotary machines. *Measurement*, *169*, 108389. https://doi.org/https://doi.org/10.1016/j.measurement.2020.108389

Shen, F., Chen, C., Yan, R., & Gao, R. X. (2015). Bearing fault diagnosis based on SVD feature extraction and transfer learning classification. 2015 prognostics and system health management conference (PHM),

Shen, S., Lu, H., Sadoughi, M., Hu, C., Nemani, V., Thelen, A., Webster, K., Darr, M., Sidon, J., & Kenny, S. (2021). A physics-informed deep learning approach for bearing fault detection. *Engineering Applications of Artificial Intelligence*, *103*, 104295. https://doi.org/https://doi.org/10.1016/j.engappai.2021.104295

Smith, W. A., & Randall, R. B. (2015). Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mechanical Systems and Signal Processing*, *64*, 100-131. https://doi.org/https://doi.org/10.1016/j.ymssp.2015.04.021

Sun, M., Wang, H., Liu, P., Huang, S., & Fan, P. (2019). A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings. *Measurement*, *146*, 305-314. https://doi.org/https://doi.org/10.1016/j.measurement.2019.06.029

Tian, J., Morillo, C., Azarian, M. H., & Pecht, M. (2015). Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis. *IEEE Transactions on Industrial Electronics*, *63*(3), 1793-1803. https://doi.org/10.1109/TIE.2015.2509913

Wang, D. (2018). Some further thoughts about spectral kurtosis, spectral L2/L1 norm, spectral smoothness index and spectral Gini index for characterizing repetitive transients. *Mechanical Systems and Signal Processing*, *108*, 360-368. https://doi.org/https://doi.org/10.1016/j.ymssp.2018.02.034

Yang, B., Lei, Y., Jia, F., & Xing, S. (2019). An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mechanical Systems and Signal Processing*, *122*, 692-706. https://doi.org/https://doi.org/10.1016/j.ymssp.2018.12.051

Yang, J., & Delpha, C. (2022a). An incipient fault diagnosis methodology using local Mahalanobis distance: Detection process based on empirical probability density estimation. *Signal Processing*, *190*, 108308. https://doi.org/https://doi.org/10.1016/j.sigpro.2021.108308

Yang, J., & Delpha, C. (2022b). An incipient fault diagnosis methodology using local Mahalanobis distance: Fault isolation and fault severity estimation. *Signal Processing*, *200*, 108657. https://doi.org/https://doi.org/10.1016/j.sigpro.2022.108657

Zhang, C., Li, B., Chen, B., Cao, H., Zi, Y., & He, Z. (2015). Weak fault signature extraction of rotating machinery using flexible analytic wavelet transform. *Mechanical Systems and Signal Processing*, *64*, 162-187. https://doi.org/https://doi.org/10.1016/j.ymssp.2015.03.030

Zhang, C., Yuling, L., Fangyi, W., Binqiang, C., Jie, L., & Bingbing, H. (2020). Multi-faults diagnosis of rolling bearings via adaptive customization of flexible analytical wavelet bases. *Chinese Journal of Aeronautics*, *33*(2), 407-417. https://doi.org/https://doi.org/10.1016/j.cja.2019.03.014

Zhang, X., Liu, Z., Wang, J., & Wang, J. (2019). Time–frequency analysis for bearing fault diagnosis using multiple Q-factor Gabor wavelets. *ISA transactions*, *87*, 225-234. https://doi.org/https://doi.org/10.1016/j.isatra.2018.11.033

Zhang, Y., Lv, Y., & Ge, M. (2021). Time–frequency analysis via complementary ensemble adaptive local iterative filtering and enhanced maximum correlation kurtosis deconvolution for wind turbine fault diagnosis. *Energy Reports*, *7*, 2418-2435. https://doi.org/https://doi.org/10.1016/j.egyr.2021.04.045