

## A wavelet neural network informed by time-domain signal preprocessing for bearing remaining useful life prediction



Kai Zhou <sup>a,b,\*</sup>, Jiong Tang <sup>c</sup>

<sup>a</sup> Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, China

<sup>b</sup> Department of Mechanical Engineering and Engineering Mechanics, Michigan Technological University, USA

<sup>c</sup> Department of Mechanical Engineering, University of Connecticut, USA

### ARTICLE INFO

#### Article history:

Received 18 December 2022

Revised 10 May 2023

Accepted 30 May 2023

Available online 3 June 2023

#### Keywords:

Bearing fault prognosis

Bearing remaining useful life (RUL)

prediction

Time-domain signal processing

Empirical mode decomposition (EMD)

Wavelet neural network (WNN)

### ABSTRACT

Rolling bearings are important components in rotating machinery in various industries. Conducting intelligent prognostics for bearing remaining useful life (RUL) prediction plays an important role in the health management and reliability assessment of these machinery systems. While the bearing fault prognosis using measured system response through machine learning techniques has attracted significant attention and demonstrated promising potential, completely data-driven approaches face some challenges in understanding complex domains with data efficient learning. In this research, we develop an integral bearing fault prognosis framework informed by the well-designed time-domain signal preprocessing to conduct the bearing RUL prediction. This framework is built upon physical feature-oriented signal preprocessing and an associated wavelet neural network (WNN). Sequential procedures of time-domain analyses are proposed to extract the physical features of bearing degradation. Empirical mode decomposition (EMD) is specifically chosen owing to its capability of handling bearing fault signals that are nonstationary with underlying nonlinearities. The WNN is a new class of neural networks that combines the classic neural network and wavelet analysis. Here in this research a WNN model built upon B-spline mother wavelet to suit the preceding EMD-based signal preprocessing is constructed to process the bearing degradation features extracted and then identify their correlation with bearing RUL. The combination of EMD-based signal preprocessing and B-spline mother wavelet in WNN enables the network to learn the input-output correlation in a physical sense. Case studies formulated upon multiple bearing datasets and multiple benchmark methods are carried out to systematically validate the proposed framework. The results consistently demonstrate prediction accuracy and performance robustness.

© 2023 Elsevier Inc. All rights reserved.

## 1. Introduction

Rotating machinery is commonly used in various industries involving manufacturing, aerospace, civil infrastructure, and transportation etc. The rolling bearings are one of the essential components in rotating machinery that supports the shaft and reduces friction. They are prone to various failures/faults as they operate under harsh conditions. Unexpected failure of in-service bearings increases the downtime and maintenance cost, resulting in significant economic and manpower losses

\* Corresponding author at: Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, China.  
E-mail address: [cee-kai.zhou@polyu.edu.hk](mailto:cee-kai.zhou@polyu.edu.hk) (K. Zhou).

[1]. The concept of prognostics and health management (PHM) has been introduced in order to reduce the probability of failure events. Predicting the remaining useful life (RUL) is a critical process in bearing PHM.

Existing prognostic approaches are often classified into two main categories, i.e., the model-based approaches and the data-driven ones. The model-based approaches characterize the inherent degradation behavior based on physics modeling and then estimate the RUL of the target system by utilizing the mathematical models that incorporate the damage and fatigue mechanisms [2,3]. Establishing a high-fidelity physical model that can account for the machinery degradation process, nevertheless, is difficult, especially when the operating conditions are complex. Even if a physical model may be theoretically reasonable, extensive effort is required to tune the model parameters using the experimental measurement [4]. On the other hand, the data-driven approaches directly predict the RUL of the target system by building the correlation of RUL with respect to the run-to-failure data collected [5]. Owing to the concurrent advancements of computing power and computational intelligence algorithms, a variety of machine learning approaches have recently been explored for bearing fault prognosis. It is worth noting that many earlier machine learning approaches possessed limited capacity in handling large number of input features in the data. Fundamentally, their relatively simple architectures cannot ensure the adequate learning of intrinsic correlation between the high-dimensional features and RUL. Hence, they were often performed upon the salient features with dimensional reduction. Benkedjouh et al. proposed to use the isometric feature mapping reduction technique and support vector regression to estimate the RUL of bearings [6]. Zhao et al. proposed a two-step supervised technique, i.e., a combination of principal component analysis (PCA) and linear discriminant analysis (LDA), to achieve the feature dimension reduction, which facilitates the RUL prediction using a subsequent linear regression model [7]. Gao et al. developed a data-driven Bayesian model built upon the Metropolis-Hastings algorithm to estimate the RUL of bearings from the dimensionally reduced time-domain features [8]. Lyn et al. developed a similarity-based method to estimate the RUL of the system that can utilize its multiple past events and corresponding impacts [9]. This method has been further improved by involving a self-adaptive weight allocation approach. Besides these conventional machine learning approaches, more recent efforts focus on deep learning that can fully exploit the neural network architecture to utilize a large amount of available data for machinery fault diagnosis and prognosis [10–14]. By nature, deep learning can minimize or even eliminate the data pre-processing effort while maintaining the desired predictive performance.

Despite their significant progresses, most of the data-drive approaches lack the physical interpretation of the entire prognostic processes as the neural networks inherently handle the features in an implicit manner. Aiming at overcoming this drawback and elucidating the failure occurrence via the temporal data in a physical sense, physics-based signal processing have been increasingly adopted to extract the fault-related system characteristics and improve the data-driven technique for characterizing bearing degradation behavior. It is noted that here the physics-based signal processing differs from the data-driven feature extraction and dimension reduction which is built upon the statistical data analysis to cluster/group the salient/pilot features for discriminating different bearing degradation conditions. A variety of signal processing methods have been developed for RUL estimation of bearings. One class of these methods is to extract the explicit and essential features in the time domain, frequency domain, or time-frequency domain that are sensitive to bearing degradation. Mosallam et al. developed a nonparametric trend modeling method for multidimensional feature extraction of sensory data followed by unsupervised feature selection. The empirical mode decomposition (EMD) was then employed on the projected features to reflect the health evolution of bearing over time [15]. Motahari-Nezhad and Jafari adopted sixty time-domain features extracted from acoustic emission data and then used the improved distance evaluation (IDE) algorithm to select the best features for bearing RUL prediction [16]. Liu et al. used the discrete wavelet transform (DWT) technique to conduct the multiresolution analysis of bearing vibration data [17]. Moreover, some other methods aim at constructing the new health indicator/index to assess the bearing degradation status. Qiu et al. constructed a new health indicator using the structural information of the spectrum (SIOS) algorithm [18]. Xu et al. proposed a new health index evaluated by the moving average cross-correlation of power spectral density of vibration signals [19]. Cheng et al. processed the raw vibration data using the Hilbert–Huang transform (HHT) to construct a novel nonlinear degradation energy indicator to evaluate the bearing health status [20]. Additionally, there is also effort to enhance the signal processing and data-driven approaches to fulfill the practical but challenging bearing fault prognosis mission, such as the prognosis under time-varying conditions and ambient uncertainties [21–23].

The overarching goal of this research is to apply carefully designed signal preprocessing techniques to machine learning as a performance multiplier to realize an integrated physics-informed data-driven framework for bearing fault prognosis. This unified framework with seamless integration of all tailored techniques provides another path for bearing RUL prediction, which is considered as the main contribution of this research. To thoroughly retrieve the key degradation features of a bearing and subsequently facilitate the establishment of an associated data-driven model, sequential signal preprocessing using statistical analysis of segmented time-domain responses followed by empirical mode decomposition (EMD) is synthesized. We first segment the original raw data and compute key statistical metrics of the segmented data, and form new times series of these statistical metrics. EMD is an adaptive multiresolution technique that decomposes a time series into physically meaningful components. It has demonstrated its effectiveness in analyzing the nonlinear and nonstationary bearing fault signals [24–26]. While the EMD oftentimes is applied in conjunction with the Hilbert spectral analysis (HSA) to achieve the so-called Hilbert–Huang transform (HHT) [27–30], in this research we only employ EMD to further process the segmented time-domain statistical metrics. Since the signal transformation only occurs in the time domain, EMD-based time-domain analysis possesses less computational complexity than the conventional Hilbert–Huang transform (HHT) for joint time-frequency analysis. The detailed investigation of the computational complexity of the EMD can be referred to

[31]. Treating the degradation features obtained from EMD and the associated bearing RUL as input and output, respectively, we then establish a wavelet neural network (WNN) model for conducting the bearing RUL estimation. WNN inherits the architecture of the classical neural network and meanwhile harnesses the wavelet analysis within the network, which generally can be achieved by utilizing the wavelet function as the activation function [32]. While various options of wavelets exist, in this research we specifically chose the B-spline mother wavelet to construct the WNN, since this mother wavelet fits very well with the nature of the preceding EMD based signal preprocessing that is built upon splines. The combination of EMD and B-spline mother wavelet in WNN allows the built-in wavelet transform that enables the physical interpretation of the network learning process, into RUL prediction. The pivot degrading features extracted from EMD generally are low-dimensional, and will be fed into WNN as input information. Hence, WNN possesses low computational complexity. In addition to WNN, it is interesting to note that the other variant termed as WaveletKernelNet recently has emerged for intelligent machinery PHM. Different from WNN that modifies the activation function, WaveletKernelNet adopts the wavelet function as the kernel/filter to convolve the input signal for accomplishing the wavelet transform during the learning process. With this, the essential semantic information with a clear physical context can be extracted from high-dimensional input [33].

The remainder of this paper is organized as follows. In Section 2, the proposed methodology for bearing fault prognosis is outlined, in which the EMD-based time-domain signal preprocessing, WNN, and fault prognosis procedures are consecutively introduced. In Section 3 we implement comprehensive case studies to validate the methodology using publicly accessible bearing datasets, including the NASA Ames prognostics dataset and the XJTU-SY bearing dataset, followed by the conclusions summarized in Section 4.

## 2. Formulation of fault prognosis framework

In this section, the integrated bearing fault prognosis framework that leverages upon the sequential time-domain signal preprocessing and feature extraction, and physics-based wavelet neural network is presented.

### 2.1. Sequential time-domain signal preprocessing

The bearing degradation is reflected essentially in the time domain responses acquired. To facilitate the degradation characterization, time-domain signal processing and feature extraction has been a common technique. However, it is well known that simple time-domain signal processing may overlook the rich and subtle underlying features in bearing degradation which is a highly nonlinear and relatively slow process. Although the joint time-frequency domain approaches have been attempted, they usually require extensive empirical knowledge for decision making which is subjective. Fundamentally, these signal processing techniques lack the learning capability thus rely heavily on empirical experience. In this research, we aim at establishing a new framework for bearing prognosis, in which we intend to leave the mapping of hidden features of bearing degradation and the RUL to a wavelet neural network (WNN) (to be explained) to take advantage of the concept of machine learning. Before applying WNN, a sequential signal pre-processing procedure is formulated to extract the time-domain essential features/characteristics which are then fed to the WNN. As the first step of signal preprocessing, we segment the available time domain responses and, for each segment, compute several prominent statistical metrics from the raw vibration data, including [34]:

a) *Mean* indicates the average of acceleration signals in the segment and can be calculated as

$$\mu = \frac{1}{N} \sum_{i=1}^N y_i^{(a)} \quad (1)$$

where  $N$  denotes the length of segment (i.e., number of data points) and  $y^{(a)}$  denotes the measured acceleration signals.

a) *Variance (VAR)* measures the data distribution around the segment mean that can be expressed as

$$\text{VAR} = \frac{1}{N} \sum_{i=1}^N |y_i^{(a)} - \mu|^2 \quad (2)$$

b) *Mean absolute value (MAV)* is the absolute average of acceleration signals in the segment that is defined as

$$\text{MAV} = \frac{1}{N} \sum_{i=1}^N |y_i^{(a)}| \quad (3)$$

c) *Simple sign integral (SSI)* determines the energy of signals in the segment and can be computed as

$$\text{SSI} = \sum_{i=1}^N |y_i^{(a)}|^2 \quad (4)$$

d) *Root mean square* (RMS) is the amplitude of modulated random process that can be described as

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N |y_i^{(a)}|} \quad (5)$$

e) *Peak to peak* (P2P) value represents the difference between the maximum and minimum in the form of

$$\text{P2P} = \max([y_1^{(a)}, y_2^{(a)}, \dots, y_N^{(a)}]) - \min([y_1^{(a)}, y_2^{(a)}, \dots, y_N^{(a)}]) \quad (6)$$

f) *Kurtosis* (KURT) is the scaled form of the fourth statistical moment to assess the degree of tailedness in the probability distribution curve, given as

$$\text{KURT} = \frac{\frac{1}{N} \sum_{i=1}^N |y_i^{(a)} - \mu|^4}{\left(\frac{1}{N} \sum_{i=1}^N |y_i^{(a)} - \mu|^2\right)^2} \quad (7)$$

g) *Skewness* (SKW) is the third statistical moment to evaluate the degree of asymmetry of the probability distribution curve, given as

$$\text{SKW} = \frac{\frac{1}{N} \sum_{i=1}^N |y_i^{(a)} - \mu|^3}{\left(\sqrt{\frac{1}{N} \sum_{i=1}^N |y_i^{(a)} - \mu|^2}\right)^3} \quad (8)$$

For each segment, we can compute the aforementioned statistical metrics. When we consider all segments, these segmented statistical metrics then form new time-series statistical metrics data. As the second step of signal preprocessing, we employ the empirical mode decomposition (EMD) technique to these time-series statistical metrics data to further extract the critical degradation-related features which are then fed to the subsequent WNN for RUL prediction. The underlying principle of EMD is to locally identify the most rapid oscillations in the time series, which is defined as a waveform interpolating local maxima and minima [35]. Specifically, the local maxima and minima are interpolated through a cubic spline, resulting in the upper and lower bounds, i.e., the envelope. The mean of the envelope can then be subtracted from the original time series. The same interpolation procedure is iterated on the remainder of the time series, which is called the sifting process. This process stops when the mean of the envelope approaches zero over the entire domain. The resulting time series is defined as the first intrinsic mode function (IMF). Removing the preceding IMFs from the original time series and applying the same sifting process iteratively can yield higher-order IMFs. Theoretically, a time series can be classified as IMF only when two criteria hold concurrently. The first is that the number of local maxima and the number of local minima must differ by at most one, and the second is that the mean of the envelope must be zero. When the EMD is employed, a 1-dimensional discrete time series can be represented as

$$S = \sum_{i=1}^K \text{imf}_k + \text{res} \quad (9)$$

where  $\text{imf}_k$  is the decomposed  $k$ -th order IMF, and  $\text{res}$  is the residual i.e., a low-order polynomial component that cannot be further decomposed. Fundamentally, a finite (and limited) number of IMFs are nearly orthogonal to each other. Inherently, this technique can decompose the time series into different fundamental components, each with a distinct time scale. As can be seen above, the time scale of the component increases when the decomposition proceeds. Correspondingly, the oscillation frequency of the component decreases. As pointed out in the literature [24–26], the main strength of EMD is to deal with nonstationary behaviors of systems with nonlinearities. This appears to be very promising to deal with bearing fault degradation, which indeed features nonstationary responses with system nonlinearities. For this reason, we will employ EMD to the time-series statistical metrics data and use the resulting fundamental components to conduct the bearing fault prognosis analysis.

## 2.2. Physics-based wavelet neural network

In this section, the wavelet neural network (WNN) employed in the proposed framework is outlined. WNN is a feedforward neural network with the theoretical formulation in terms of the wavelet decomposition. It is a generalization of the radial basis function network (RBFN) [32,36]. The underlying idea of a WNN is to adopt the wavelet basis to the training data, i.e., utilize the wavelet function as the activation function. This is the main difference between the WNN and the conventional neural networks that use classical activation functions such as sigmoid, and ReLU etc. [36]. A WNN usually

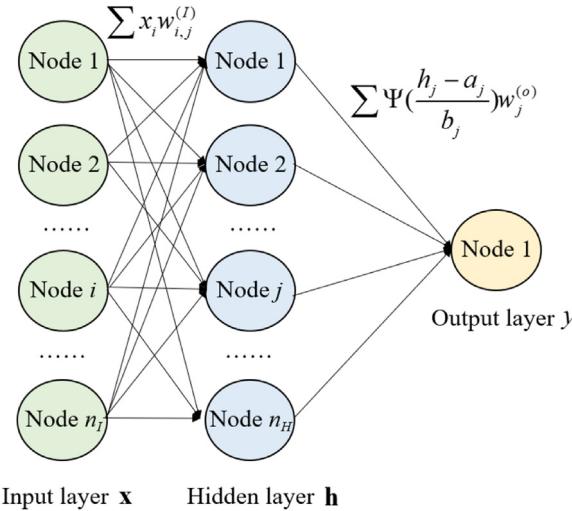


Fig. 1. WNN architecture.

consists of three layers, including an input layer, a hidden layer, and an output layer, as shown in Fig. 1 [37]. The input layer intends to introduce the explanatory variables into the WNN. The hidden layer consists of hidden neurons/nodes referred to as wavelons. The wavelons can transform the input variables into dilated and translated form of the mother wavelet. The output layer approximates the target value using the information from the hidden layer.

Since the wavelet activation function integrated into the hidden layer is incorporated with multi-scaled analysis and scale translation, WNN usually has high precision and resolution. The unique structure of WNN also leads to its excellent capability in analyzing the local information of signals and performing the function approximation, especially for the single variable function approximation. It is generally of small scale, enabling computationally efficient model training. More importantly, compared to the traditional activation function, the built-in wavelet activation function in WNN allows the implementation of wavelet transform, to establish an interpretable neural network, as will be shown subsequently. Owing to these features, the concept of WNN has been employed in diverse applications [32,38–41]. Nevertheless, in the past due to structural limitations, WNN has not been formulated to fulfill the complex learning tasks by handling a large number of features simultaneously. This shortcoming can be tackled by reconfiguring the network architecture, i.e., appending the convolutional layers at the beginning of the network to establish the so-called convolutional wavelet neural network (CWNN) [42], or using the wavelet function as the kernel to convolve the input signals in a WaveletKernelNet [33].

Training the WNN model resorts to backpropagation optimization [43]. The basic idea of backpropagation optimization is to find the percentage of contribution of each weight to the error. The error generally is formulated as a cost function to be minimized via model training. For a regression problem, the cost function is typically defined as the mean squared error (MSE) [44],

$$L = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (10)$$

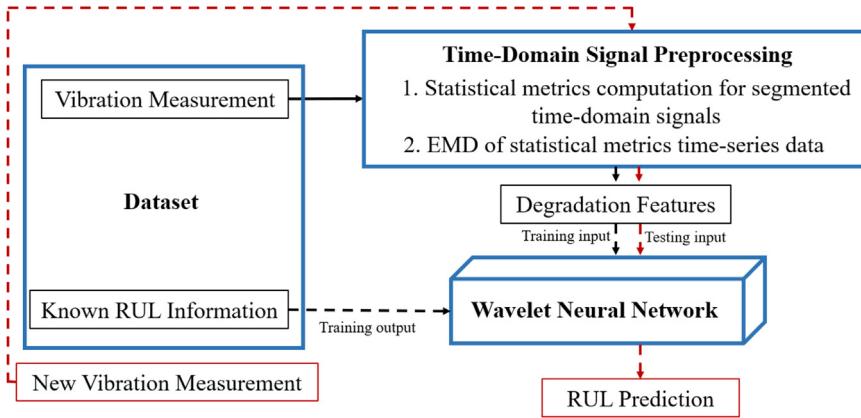
The contribution of unknown weight to the cost function can be evaluated through the partial derivative below,

$$\frac{\partial L}{\partial \theta} = -\frac{2}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \frac{\partial y_i}{\partial \theta} \quad (11)$$

where  $\theta$  represents any type of unknown weight. It is noted that the computation of the above partial derivative requires implementing the succeeding gradient descent algorithm for WNN weight optimization [45]. As shown in Fig. 1, different types of weights are involved in a WNN, including the traditional weights  $w_{i,j}^{(I)}$  and  $w_j^{(O)}$  in the input and output layers, and wavelet weights  $a_j$  and  $b_j$  (i.e., translation and dilation parameters) in the hidden layer. Let  $\theta$  be denoted as a possible set of  $[w_{i,j}^{(I)}, a_j, b_j, w_j^{(O)}]$ ,  $\partial y_i / \partial \theta$  can be derived in different forms,

$$\frac{\partial y_i}{\partial w_{i,j}^{(I)}} = \Psi\left(\frac{h_j - a_j}{b_j}\right) x_i \quad (12a)$$

$$\frac{\partial y_i}{\partial a_j} = -\Psi\left(\frac{h_j - a_j}{b_j}\right) \frac{1}{b_j} \quad (12b)$$



**Fig. 2.** Bearing fault prognosis implementation workflow.

$$\frac{\partial y_i}{\partial b_j} = -\Psi\left(\frac{h_j - a_j}{b_j}\right)\left(\frac{h_j - a_j}{b_j^2}\right) \quad (12c)$$

$$\frac{\partial y_i}{\partial w_j^{(O)}} = \Psi\left(\frac{h_j - a_j}{b_j}\right) \quad (12d)$$

where  $\Psi(\cdot)$  is the generic form of the mother wavelet. Based on the above partial derivatives, the gradient descent algorithm is adopted to adjust the weights of WNN in each training increment. The new weights that are updated by the gradient descent algorithm are described as [45],

$$w_{i,j}^{(l)}(t+1) = w_{i,j}^{(l)}(t) - \gamma^{(l)} \frac{\partial L}{\partial w_{i,j}^{(l)}} \quad (13a)$$

$$a_j(t+1) = a_j(t) - \gamma^{(H)} \frac{\partial L}{\partial a_j} \quad (13b)$$

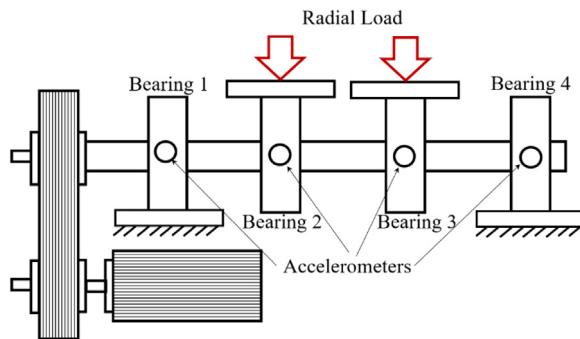
$$b_j(t+1) = b_j(t) - \gamma^{(H)} \frac{\partial L}{\partial b_j} \quad (13c)$$

$$w_j^{(O)}(t+1) = w_j^{(O)}(t) - \gamma^{(O)} \frac{\partial L}{\partial w_j^{(O)}} \quad (13d)$$

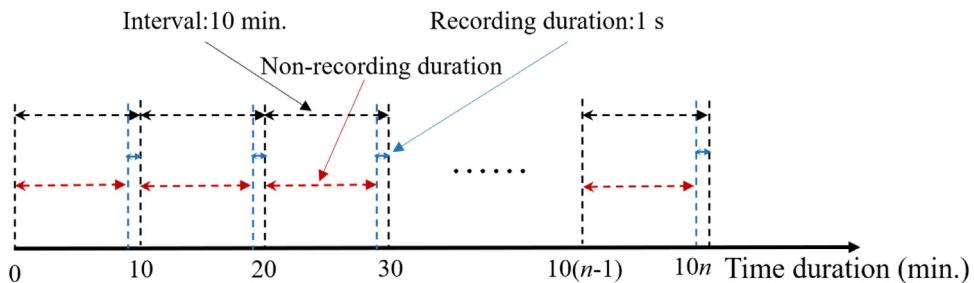
where  $t$  indicates the  $t$ -th training increment, and  $\gamma$  is the learning rate. Different learning rates can be assigned for the gradient descent algorithm to optimize the weights at different layers. The termination of model training can be determined by the specified threshold of variation of gradient or weights. However, for the implementation convenience, we use the number of epochs to manage the training duration in this research.

### 2.3. Integrated bearing fault prognosis framework

In the proposed new framework, we first segment the original time-domain data and compute respectively the statistical metrics of each segment when then form new time-series data. As can be seen, the underlying principle of empirical mode decomposition (EMD) is to perform the envelope interpolation using a cubic spline and then obtain the fundamental components. We plan to apply WNN to analyze these fundamental components. While many mother wavelets may be used, in this research we adopt the B-spline wavelet in the construction of WNN to specifically fit the nature of EMD-based signal preprocessing and unleash the potential of the network learning capability. The EMD-based signal preprocessing and the WNN can be seamlessly synthesized as an integrated framework for bearing fault prognosis. Specifically, the EMD-based time-series analysis is performed to extract the key bearing degradation features, i.e., intrinsic mode functions (IMFs) and residual of the statistical metrics data calculated from the raw vibration signals. The WNN model is subsequently trained by the relations between the degradation features (i.e., the input of WNN) and respective RUL information (i.e., the output of WNN). Once the WNN model is established, it can be used for bearing RUL prediction given any measured vibration. The implementation workflow of the proposed bearing fault prognosis is schematically illustrated in Fig. 2.



**Fig. 3.** Test rig for NASA Ames data [46].



**Fig. 4.** Illustration of data acquisition ( $n$  denotes the number of data recordings).

**Table 1**  
Overview of datasets.

	Time Duration	Failure Occurrence and Type	Instrumentation
Dataset 1	21,340 (min.)	Bearing 3 with inner race defect; Bearing 4 with roller element defect	Accelerometers placed at all bearings to measure both horizontal and vertical vibrations
Dataset 2	9840 (min.)	Bearing 1 with outer face defect	Accelerometers placed at all bearings to measure only horizontal vibration
Dataset 3	63,230 (min.)	Bearing 3 with outer face defect	Accelerometers placed at all bearings to measure only horizontal vibration

Note: datasets with gray highlights are used in this research.

### 3. Case illustration for bearing remaining useful life (RUL) prediction

In this section, we conduct implementation and demonstration using publicly accessible bearing fault datasets, illustrating the effectiveness of the methodology. Specifically, we use the run-to-failure experimental data from the NASA Ames Prognostics and the XJTU-SY data repositories for case illustrations [46,47].

#### 3.1. NASA Ames prognostic data

We first use the NASA Ames data to investigate the vibration characteristics along the bearing degradation history. The experimental test rig is shown in Fig. 3. In this test rig, four test bearings are installed on the shaft with a constant rotational speed of 2000 rpm. These four testing bearings all belong to the type of Rexnord ZA-2115 double row bearings. A radial load of 6000 lbf is directly applied to bearings 2 and 3 through a spring mechanism. Accelerometers are placed at all bearings to record the vibration data with 20 kHz sampling rate/frequency. Because of the slow varying nature of bearing vibration response under fault condition, in data acquisition, for every 10 min, signals were collected for 1 s duration until the bearing fault occurred. In other words, here 10 min are the interval between two adjacent data recordings as shown in Fig. 4. Three different run-to-failure experiments are performed, in which different types of failures on bearings are captured. The corresponding three datasets collected are shown in Table 1. In this research, datasets 2 and 3 are used to implement the subsequent bearing fault prognosis analysis.

### 3.2. Problem set-up, feature extraction, and model development

The objective of this research is to establish a methodology to perform the robust and reliable prediction of the remaining useful life (RUL) of in-service bearings given the measured vibration time-series data. As such, the vibration time-series data are the input, and the bearing RUL is the output. For the datasets shown in [Table 1](#), the RUL can be simply determined according to the running time, which is expressed as

$$\text{RUL} = T_{\text{total}} - t_{\text{current}} \quad (14)$$

where  $T_{\text{total}}$  is the total time duration of the experiment required to initiate the bearing failure, and  $t_{\text{current}}$  is the current time instant. In the case studies, we treat the vibration data measured in the recording duration (i.e., 1 s) of every 10 min. as one data subset (i.e., segment mentioned in [Section 2](#)) [Fig. 4](#)). Our hypothesis is that this can capture the periodic vibration characteristics of rotating machinery and thus can elucidate the underlying degradation behavior due to the fault. According to the data acquisition setup ([Fig. 4](#)), the total numbers of data subsets are identical to the number of sampling durations, which are 2134, 984, and 4448 for datasets 1, 2, and 3, respectively. The time-domain statistical metrics of all data subsets will be extracted accordingly. While there are many statistical metrics available to elucidate the temporal characteristics of raw vibration data, as shown in [Section 2](#), in this research, we select the metrics with trial-and-error nature, resulting in four representative metrics, including root mean square (RMS), peak-to-peak (P2P) value, kurtosis (KURT), and skewness (SKW) ([Eqs. \(5\)–\(8\)](#)). These metrics have been proven effective in machinery fault diagnosis and RUL prediction [48–52]. To perform the rigorous metric selection, some well-established methods in the feature engineering domain need to be adopted, in which the available metrics are treated as possible features [53].

Since only a small number of time-domain metrics are involved, it is unnecessary to carry out the metric selection procedure. In other words, the information of all metrics will contribute to the succeeding WNN model learning. Because the bearing fault signals inherently are nonstationary and nonlinear, the empirical mode decomposition (EMD) technique is then applied to those time-domain statistical metrics data to retrieve the more explicit features which can elucidate the bearing degradation tendency. These newly generated features generally include the intrinsic mode functions (IMFs) with different orders and residuals of the statistical metrics data that cannot be further decomposed. In this research, the residuals of statistical metrics data are particularly utilized to feed the proposed wavelet neural network (WNN) because they match the slowly progressive nature of bearing degradation.

With the input and output as residuals of the statistical metrics data and RUL of bearing, respectively, a WNN model can be established to fulfill the regression type of analysis. More specifically, a WNN model consists of a 4-node input layer carrying the residuals, and a 1-node output layer carrying the RUL information. The other hyperparameters subject to tuning include the number of nodes in the hidden layer, and the type of mother wavelet and associated wavelet parameters. The hyperparameter tuning generally is performed based upon empirical experience. In this research, we finalize the hidden layer with 30 neurons/nodes and select the B-spline wavelet as the mother wavelet [54]. It is worth noting that the B-spline mother wavelet embedded in WNN can well suit the EMD that inherently adopts the cubic spline for envelope interpolation. As will be shown later, the B-spline mother wavelet in WNN also can yield excellent prediction performance. The B-spline wavelet, i.e.,  $\Psi(\cdot)$  in [Fig. 1](#) can be represented as

$$\Psi(h, a_j, b_j) = \sqrt{f_b} \left( \sin c \left( f_b \frac{h - a_j}{mb_j} \right) \right)^m e^{(2\pi f_c \frac{h - a_j}{b_j})i} \quad (15)$$

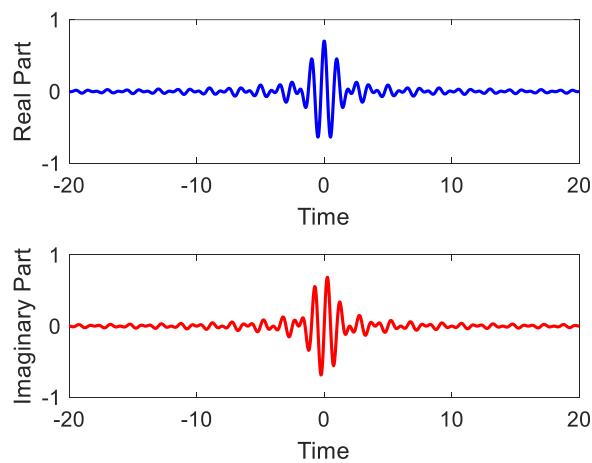
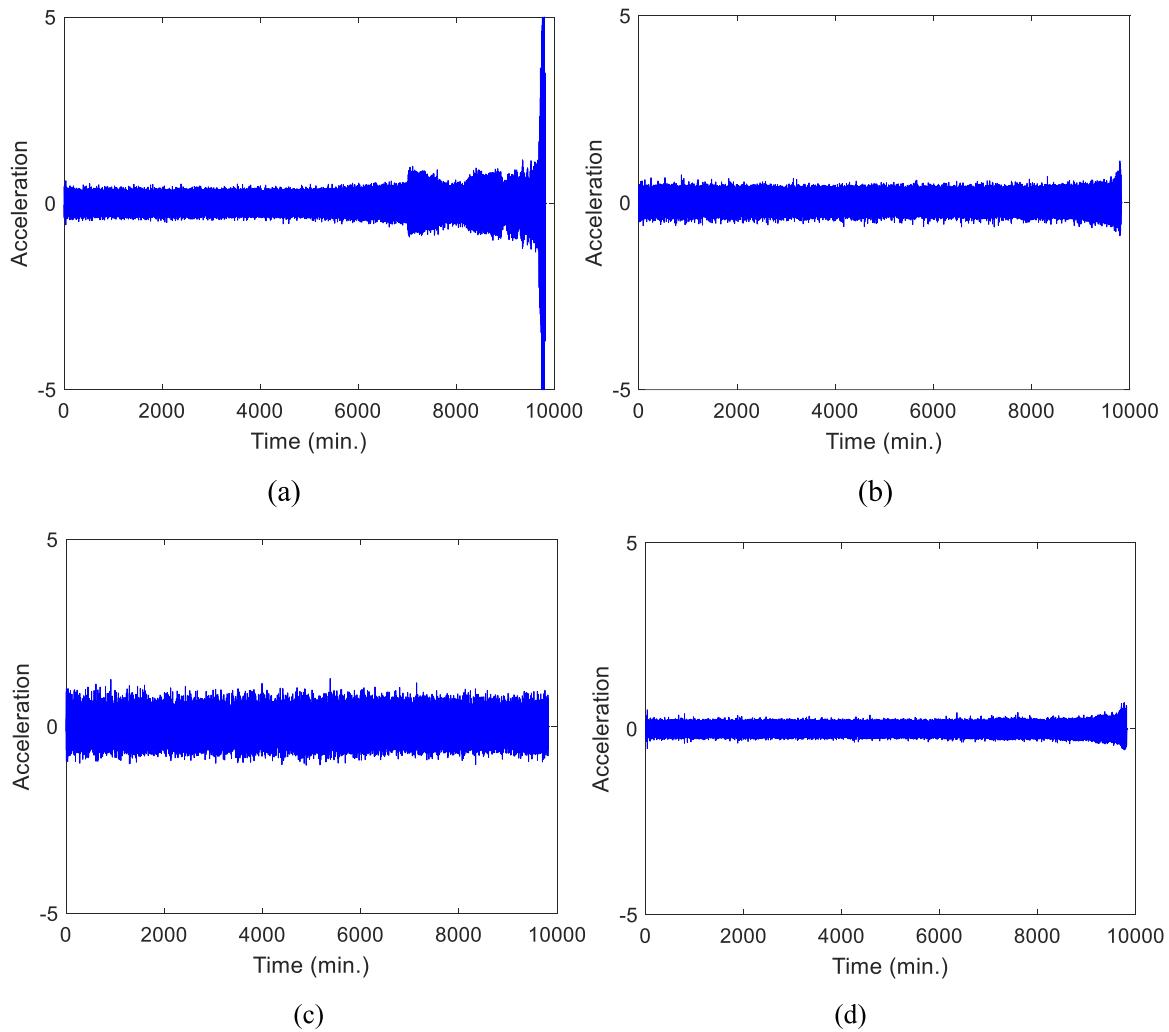
where  $m$ ,  $f_b$  and  $f_c$  denote the order parameter, bandwidth parameter, and wavelet center frequency parameter, respectively. They are the hyperparameters that are determined as  $m = 1$ ,  $f_b = 0.5$ ,  $f_c = 1$  in this research.  $h$  is the input of the hidden layer, and  $a_j$  and  $b_j$  represent the translation and dilation parameters, which can be treated as unknown weights in the neural network to be optimized via model training. A B-spline wavelet essentially is a continuous complex-valued wavelet, the profile of which is illustrated in [Fig. 5](#). Combining all the above information, the configuration of WNN architecture can be finalized. The total number of unknown weights in this WNN model is 210 (i.e.,  $4 \times 30$  ( $w_{i,j}^{(l)}$ ) + 30 ( $a_j$ ) + 30 ( $b_j$ ) + 30 ( $w_j^{(0)}$ )), indicating a small-scale WNN model with low training cost.

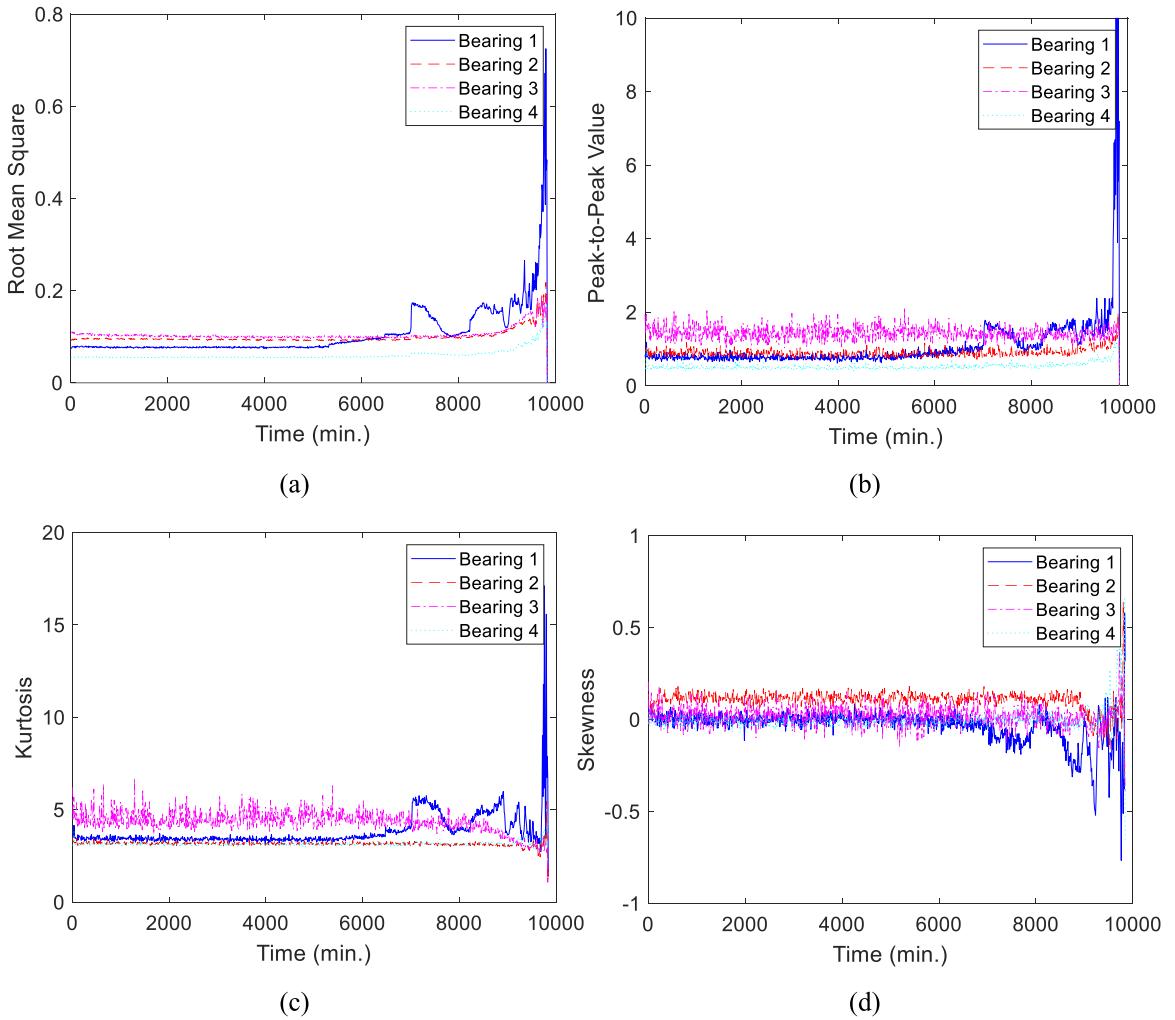
### 3.3. Predictive performance investigation and discussion

Following the set-up formulated in the previous section, the bearing fault prognosis is carried out using NASA Ames prognostics data. We first conduct the case study on dataset 2 shown in [Table 1](#).

#### 3.3.1. Case 1 – RUL prediction on dataset 2 ([Table 1](#))

The raw vibration time-series data acquired from different bearings are shown in [Fig. 6](#). Since bearing 1 eventually has outer race failure, its vibration amplitudes at the end-of-life cycle become noticeably large. On the other hand, a very insignificant increase in amplitudes over time can be observed for other bearings. To highlight the temporal characteristics of raw data that is expected to reflect the degradation nature of bearing, we first conduct the time-domain signal processing to compute the statistical metrics of all data subsets and yield the results shown in [Fig. 7](#). Compared to the raw vibration signals in [Fig. 6](#), those statistical metrics are more capable of accounting for the effect of bearing fault occurrence. Particularly,

**Fig. 5.** Illustration of B-spline wavelet.**Fig. 6.** Vibration data (a) measured on bearing 1; (b) measured on bearing 2; (c) measured on bearing 3; (d) measured on bearing 4.

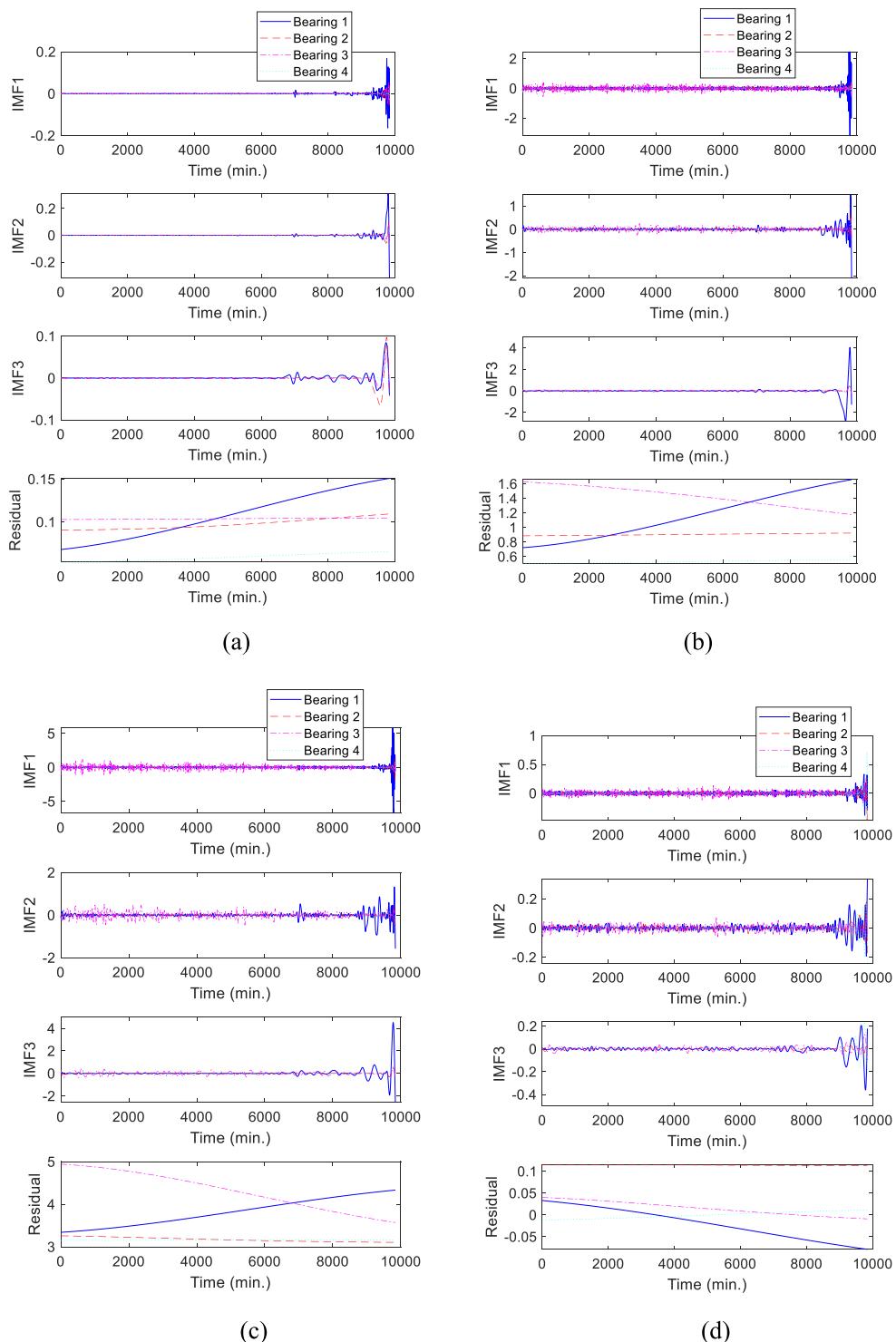


**Fig. 7.** Time-domain statistical metrics (a) root mean square; (b) peak-to-peak value; (c) kurtosis; (d) skewness.

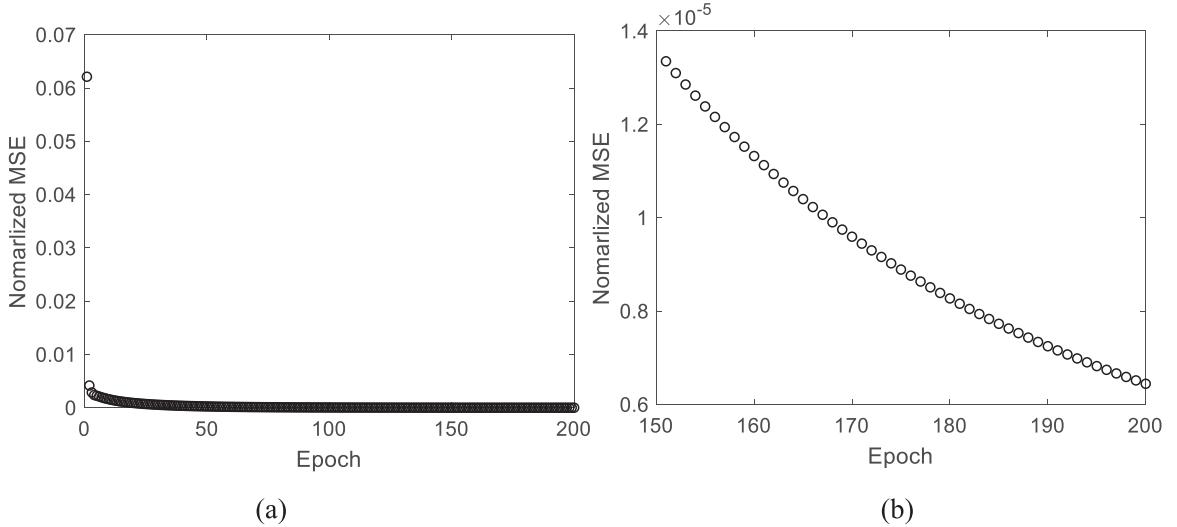
the statistical metrics on bearings 2, 3, and 4 at the end-of-life cycle have notable variations due to the fault occurred on bearing 1. The statistical metrics on bearing 1 are most informative to the bearing fault occurrence. Moreover, the skewness has the opposite trend with respect to other statistical metrics for all datasets.

While the statistical metrics data indicate the bearing degradation behavior to a certain extent, they appear to be indiscriminative at the early and middle stages of the life cycle. The fundamental reason is that the statistical metrics data essentially are superimposed with many high-frequency contents, i.e., fast oscillating components that exhibit strong non-linearity with respect to fault. To extract the key degradation features that can elucidate the slowly progressing nature of bearing degradation, empirical mode decomposition (EMD) is further performed to decompose the statistical metrics data into the components with different resolutions. The first three intrinsic mode functions (IMFs) and residuals extracted from the statistical metrics data are given in Fig. 8. It can be clearly seen that the higher-order the intrinsic mode function (IMF), the slower the dominant oscillation. The residuals are relatively proportional/linear with respect to the running time. Among them, the residuals of the bearing 1 dataset exhibit a more evident trend versus running time than that of other bearings because of the fault occurrence on bearing 1. Consistent with the observation above (Fig. 7), the residuals of the skewness show an adverse tendency. As reported in the study associated with the battery failure prediction, it was found that the residuals of EMD were strongly correlated with the ground truth battery RUL data [55]. Therefore, in this research we also use residuals of statistical metrics data as inputs to train the subsequent WNN model for bearing RUL prediction.

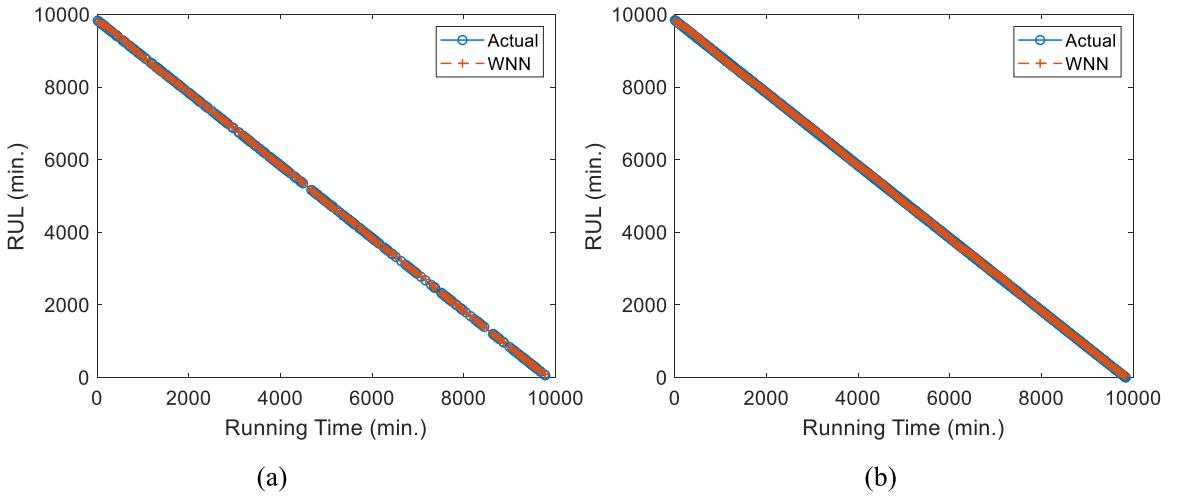
Recall that the signals within every recording duration, i.e., 1 s (Fig. 4) are treated as one data subset, upon which the respective statistical metrics and residual are retrieved. Let a sample be defined as the mapping/relation between the residuals of one data subset and associated RUL. We thus have a total of 984 samples for each dataset measured from each bearing. Here we use 984 samples of bearing 1 dataset for analysis. 30% and 70% of all samples are used as training and testing samples, respectively. A mean squared error (MSE) is employed as a metric to direct the model training and



**Fig. 8.** Intrinsic mode functions (IMFs) and residuals retrieved from different statistical metrics (a) root mean square; (b) peak-to-peak value; (c) kurtosis; (d) skewness.



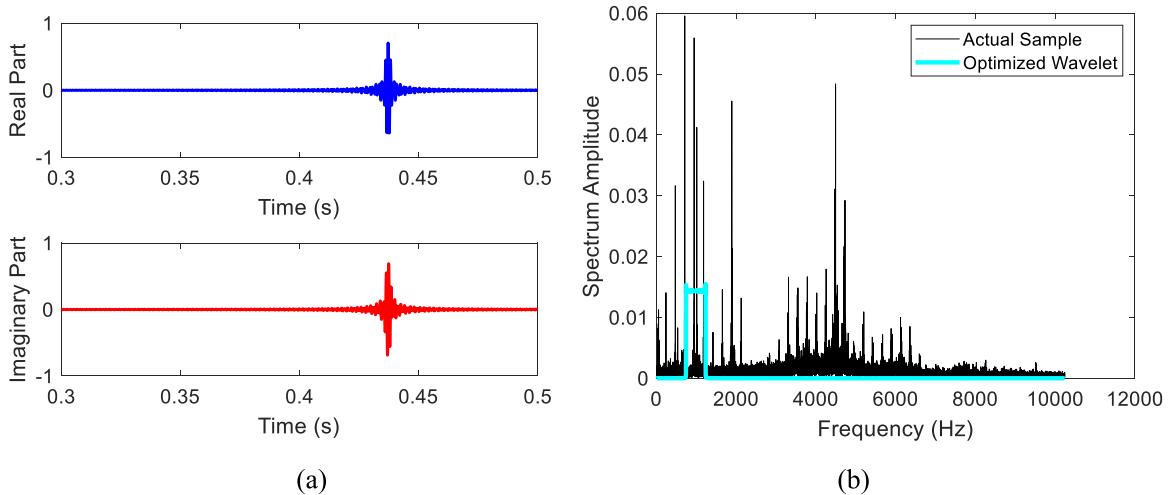
**Fig. 9.** Training history (a) full epochs; (b) last 50 epochs.



**Fig. 10.** RUL prediction accuracy (a) training samples; (b) testing samples.

evaluation. It is worth noting that the MSE essentially is normalized since the outputs, i.e., RUL values of samples, are mapped to unit space  $[0, 1]$  before we feed them into the model training. Regarding the emulation parameters, 200 epochs are adopted. The learning rate for input and output layer weights optimization is set as 0.01 ( $\gamma^{(I)}$  and  $\gamma^{(O)}$  in Eq. (13a, d)), and the learning rate for hidden layer weight optimization is set as 0.001 ( $\gamma^{(H)}$  in Eq. (13b, c)). The model training is very efficient, and only costs several seconds due to the small scale of the network model. The entire training history is shown in Fig. 9a. It is observed that the training performance improves drastically at the first several epochs. While the improvement of training performance becomes slower as the training proceeds, it does exist, as can be seen in the zoom-in view of training history at the last 50 epochs (Fig. 9b). Once the WNN model is well-trained, we can use it to directly predict the RUL of bearing for both the training and testing samples and obtain the training and testing accuracy shown in Fig. 10. The nominal relation between the running time and RUL is assumed to be linear with a monotonical decrease according to the definition of RUL given in Eq. (14). Clearly, both the training and testing accuracies are desired because the predicted RUL values match the actual RUL values extremely closely. This also indicates the adequate model training without overfitting and underfitting. The results validate the effectiveness of the proposed methodology which leverages upon the combination of signal preprocessing and WNN.

In contrast to the “black box” nature of traditional neural network, the physical knowledge incorporated makes the WNN model physically interpretable. After the WNN model is adequately trained, the optimized wavelet functions in the hidden layer of the WNN model can locate the salient features of vibration signals. To verify this, we select one of the optimized wavelet functions in the trained WNN model, the profiles of which are displayed in Fig. 11a. Additionally, we choose a data



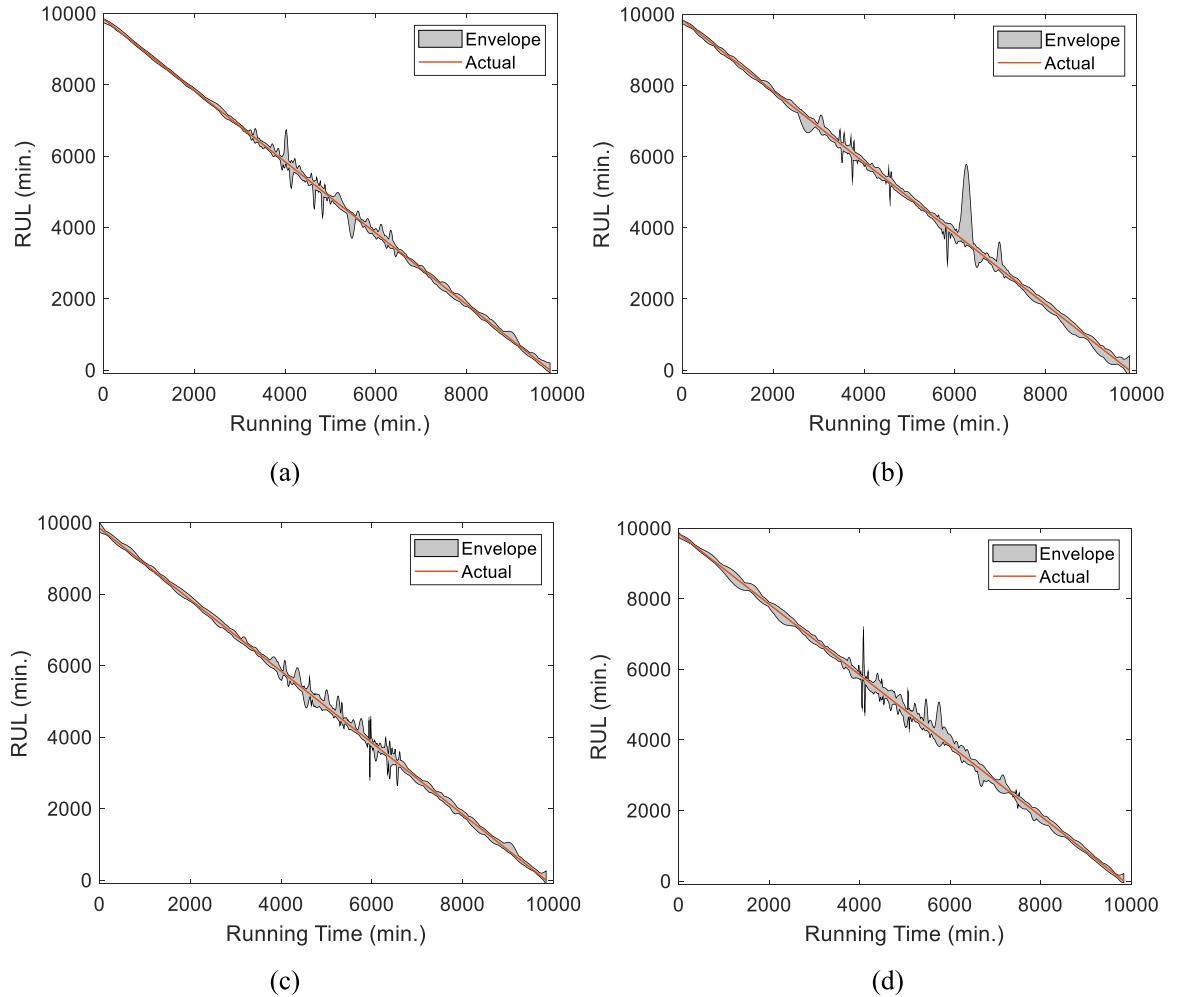
**Fig. 11.** Physical interpretation of the trained WNN model (a) optimized B-spline wavelet; (b) frequency spectrums comparison.

subset with signals collected at 9500 min. running time. We then obtain the frequency spectrums of the above optimized wavelet function and selected data subset by performing the fast Fourier transform (FFT), as shown in Fig. 11b. While the bearing characteristic frequencies are separated, and their bands cannot be all captured, the frequency spectrum (i.e., spectral envelope) of the wavelet function can locate one resonance frequency band of signals, which is centered around 1000 Hz with a larger amplitude and a narrower bandwidth. Such agreement indicates the WNN learning in a physical context. It is noted that the spectral envelope of the wavelet function has the “square” shape because the first-order B-spline wavelet is adopted (i.e.,  $m$  is set as 1 in Eq. (15)).

Considering the randomness in the data split and stochastic optimization for model training, 50 emulation runs based upon the random data splits are carried out to assess the robustness of the proposed methodology. The results are given in Fig. 12a. It is noted that the result comparison in Fig. 12a is conducted over the entire sample space, and thus the testing and training samples are not necessarily separated. The gray region represents the envelope of the prediction embraced by the upper and lower bounds. A very narrow envelope along the actual RUL line indicates the excellent robustness of the proposed methodology. The noticeable variation of predictions mainly occurs in the middle range of the bearing life cycle. We also examine the prediction envelops using other bearing datasets following the same analysis procedures and obtain the respective results shown in Fig. 12b–d, which also illustrate the excellent performance robustness. While the raw vibration data collected on other bearings cannot directly point to the fault occurrence (Fig. 6), good predictive capability also can be ensured by harnessing the proposed signal preprocessing and WNN collectively.

To further verify the efficacy of the proposed framework for RUL prediction, we conduct the performance investigation and comparison with respect to other benchmark methods. In this study, we first introduce two general machine learning models, i.e., support vector machine (SVM) and random forest (RF), which utilize the same input of the WNN. Comparing these two methods with the WNN aims to highlight the advantage of the WNN in predictive accuracy. We then develop a deep long short-term memory (DLSTM) network to directly handle the raw vibration signals for RUL prediction to validate the entire proposed framework. For rigorous comparison, hyperparameter tuning is carried out to optimize the performance of the benchmark methods. Specifically, the kernel functions, e.g., linear, polynomial, and RBF kernels of SVM, and the number of bagged decision trees in RF are subject to tuning. Moreover, the architecture of DLSTM network is carefully configured with the layer details shown in Table 2. Recall that a total of 984 samples are segmented in this dataset, each of which records 1 s duration of vibration. According to the specified sampling frequency, each sample contains 20,480 data points. The underlying strength of DLSTM is to characterize the data input-output correlation while accounting for the temporal dependence of data points within each single sample. To take such temporal nature into account during the implementation of DLSTM network, we transform each data sample with size  $1 \times 20,480$  to 20 time sequences, each of which is of size  $1 \times 1024$  ( $20,480 = 20 \times 1024$ ). Therefore, the input layer of DLSTM carries the sequence with the size, i.e.,  $1 \times 1024$ . The epoch size and batch size of DLSTM network are set as 100 and 10 respectively. “Adam” optimizer is used for model training. All the hyperparameters of DLSTM are determined by the combination of empirical experience, and training and validation history monitoring.

The same training and testing split ratio is adopted for all methods. As RUL prediction essentially belongs to the regression analysis, mean squared error (MSE) is originally used as a metric to assess the accuracy. It is noticed that the MSE magnitude order is significant due to the large RUL value, leading to the dramatic discrepancy in MSE among different methods. To facilitate the result comparison, we take the logarithm of MSE, i.e.,  $\log_{10}^{\text{MSE}}$  as a new accuracy metric. With the above data split ratio, 50 emulation runs are carried out to obtain the statistical distribution of this new accuracy metric, as shown in Fig. 13. As can be seen, the DLSTM network consistently shows more significant errors than the rest of the methods for



**Fig. 12.** RUL prediction envelopes (lower and upper bounds) (a) data collected from bearing 1; (b) data collected from bearing 2; (c) data collected from bearing 3; (d) data collected from bearing 4.

all the data collected from the different bearings. This clearly indicates the positive contribution of proposed signal preprocessing on the predictive accuracy improvement. With the same signal preprocessing employed, WNN outperforms RF and SVM in terms of accuracy, even though it has relatively larger variability of accuracy over different emulation runs. Because of the substantial reduction in data dimension, the WNN, RF, and SVM feature significantly less computational complexity than DLSM. Such observation illustrates that WNN appears to be the most tailored method. Furthermore, the data collected from bearing 1 especially favors the predictive performance of WNN and RF because of the respective reduced errors. This may be due to the fact that actual failure occurs at bearing 1. Overall, this set of results adequately verifies the effectiveness of the proposed framework for bearing RUL prediction.

### 3.3.2. Case 2 – RUL prediction on dataset 3 (Table 1)

In this section, we conduct another case study based on the NASA Ames prognostics data repository, specifically utilizing the dataset with a much longer duration with a fault occurrence on bearing 3, i.e., dataset 3 in Table 1. Upon the raw vibration signals, we conduct the time-domain statistical metric extraction followed by the empirical mode decomposition (EMD) to identify the fundamental components shown in Fig. 14.

Compared with Fig. 8, most of the intrinsic function modes (IMFs) extracted remain quite constant with a small oscillation amplitude before the fault occurrence. Additionally, it can be found that the residuals from the bearing 3 dataset exhibit a more prominent trend with respect to time than that from other bearing datasets. In this dataset, more samples are generated using the same data acquisition setup (Fig. 4) and signal preprocessing procedure because of the longer duration. The total number of samples is 6323. With the same training-testing data split ratio and other relevant operating parameters, we can identify the prediction envelopes from the collected results of 50 emulation runs, as shown in Fig. 15. Compared

**Table 2**  
Configuration of DLSTM network architecture.

Layer ID	Layer Type	Property
1	Sequence Input	
2	Convolutional	20 filters with size $1 \times 5$ ; stride $1 \times 1$ same padding
3	BatchNormalization	
4	ReLU	
5	MaxPooling	$1 \times 3$ Max pooling; stride $1 \times 3$ ; no padding
6	Convolutional	20 filters with size $1 \times 5$ ; stride $1 \times 1$ same padding
7	BatchNormalization	
8	ReLU	
9	MaxPooling	$1 \times 3$ Max pooling; stride $1 \times 3$ ; no padding
10	Convolutional	20 filters with size $1 \times 5$ ; stride $1 \times 1$ same padding
11	BatchNormalization	
12	ReLU	
13	MaxPooling	$1 \times 3$ Max pooling; stride $1 \times 3$ ; no padding
14	Flatten	
15	LSTM	200 hidden units
16	Fully connected	2000 nodes
17	Fully connected	2000 nodes
18	Output/regression	1 node

Note: the total number of learnable parameters is around 5.1 million.

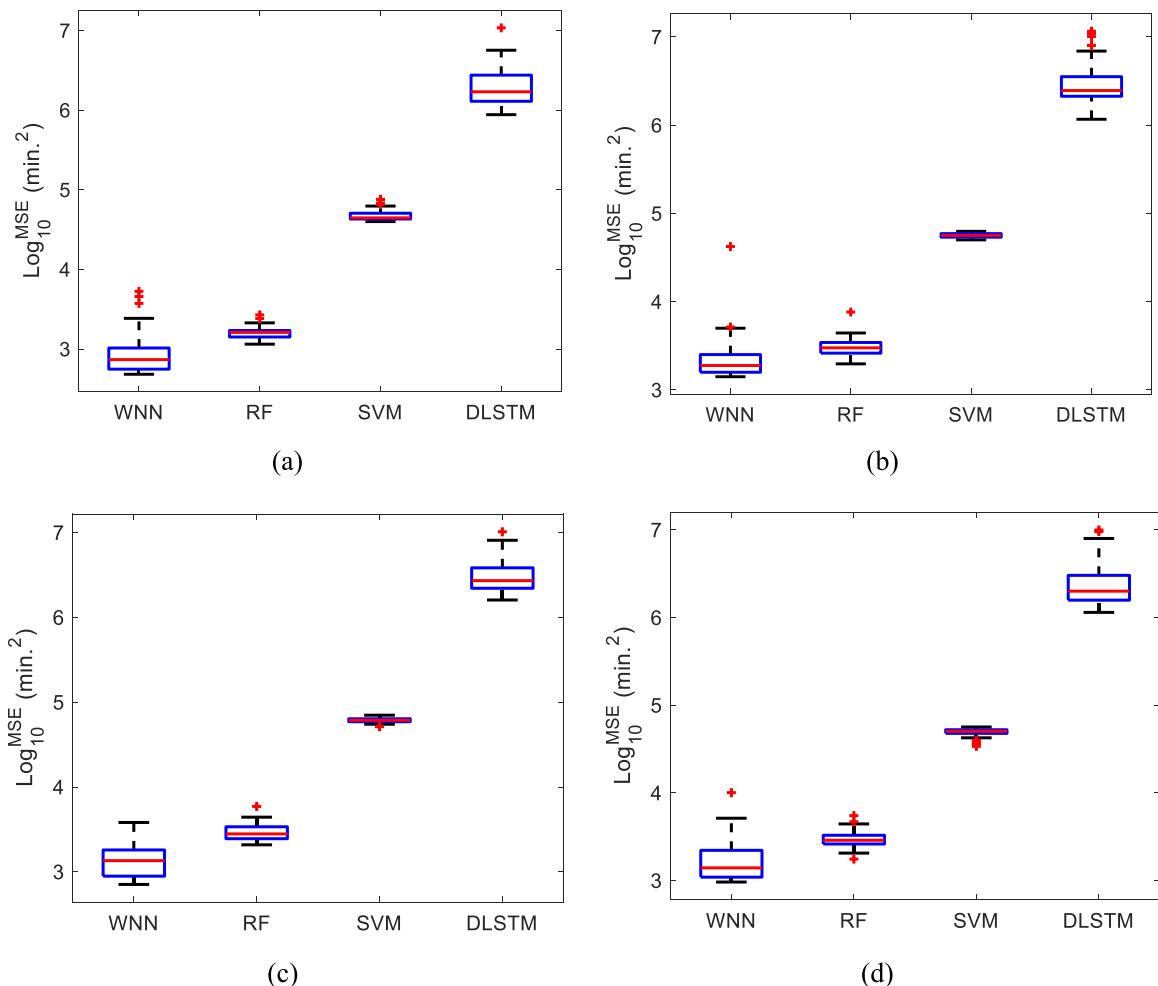
with the result in Case 1, the predictive accuracy and performance robustness in this case are considerably enhanced due to the reduction of envelope width. Regardless of the bearing dataset, excellent performance robustness can be ensured. Such enhancement may be attributed to the more samples involved in training, which can adequately learn the inherent correlation between the residuals of statistical metrics data and bearing RUL. The additional case investigation provided once again verifies the effectiveness of the proposed methodology. It is worth noting that in this case, we do not carry out the investigation of method performance because this dataset, i.e., Dataset 3, and the above dataset, i.e., Dataset 2 in Table 1, are acquired under the same experimental setup. A similar trend of method performance comparison shown in Fig. 13 is expected to be observed. Besides, this dataset contains an increasing number of data samples, which leads to the extremely computational cost in DLSTM network training with many emulation runs involved.

### 3.4. Further validation using the XJTU-SY bearing datasets

In practice, bearing data are acquired upon different experimental setups, bearing types, fault types etc. To systematically assess the proposed methodology, we also implement the bearing fault prognosis on the datasets from another public data repository, i.e., XJTU-SY bearing datasets. As shown in Fig. 16, the bearing testbed established to collect the XJTU-SY bearing datasets mainly consists of an alternating current (AC) induction motor with a motor speed controller, a support shaft, two support bearings, a testing bearing, and a hydraulic loading system. This testbed is designed to conduct accelerated degradation tests of rolling bearings under different operations [47]. Specifically, three operation conditions are classified according to the radial load applied and rotational shaft speed. Five run-to-failure experiments on different single bearings are conducted at each operation condition. The sampling frequency of data acquisition is 25.6 kHz, and the interval between two adjacent recordings is 1 min. The corresponding data are collected as given in Table 3. In the XJTU-SY bearing datasets, both the horizontal and vertical vibrations are measured by properly placing the dual accelerometers. Among the total fifteen datasets, we specifically select three datasets with different bearing fault types, which we perceive are sufficient to validate the methodology. Those datasets utilized are highlighted in the gray background (Table 3).

After implementing the procedures mentioned above, we obtain the results shown in Fig. 17. It is found that the direction of vibration measurements plays a negligible role in dictating prediction accuracy. Additionally, the bearing fault type appears to be uninfluential with respect to the prediction accuracy. On the other hand, the number of samples dependent on the time duration impacts the accuracy and performance robustness. When more samples are generated due to the longer duration, more training samples will contribute to the model training given the same training-testing data split, thereby improving the predictive robustness under the inevitable randomness. This statement is clearly illustrated in the results that the performance using dataset #11 tops that using other datasets. Overall, the prediction envelope is very narrow, consistently indicating the good performance robustness of the proposed methodology. As a new testing case built upon the new bearing testbed, it is essential to conduct the performance investigation of different methods following the similar procedures adopted in Case 1. The results are given in Fig. 18, which clearly demonstrate the underlying advantage of the proposed methodology once again.

Noteworthy, enhancing the generalization capability of the bearing RUL prediction method is equally important to pursue the success of the practical applications, which we deem a future endeavor. Moreover, in the proposed method, EMD IMFs

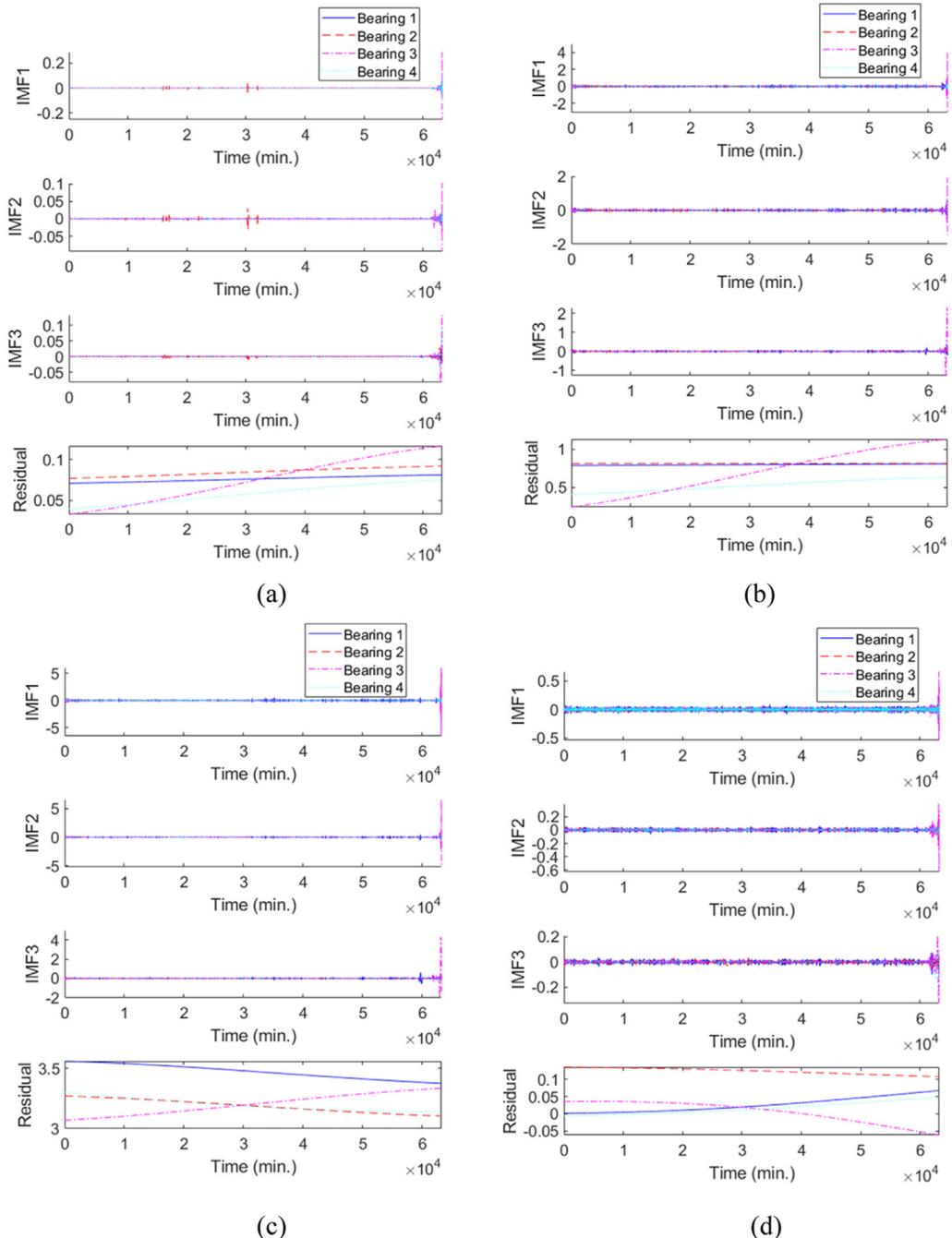


**Fig. 13.** Comparison of prediction accuracy over 50 emulation runs among different methods (a) data collected from bearing 1; (b) data collected from bearing 2; (c) data collected from bearing 3; (d) data collected from bearing 4.

**Table 3**  
Overview of datasets.

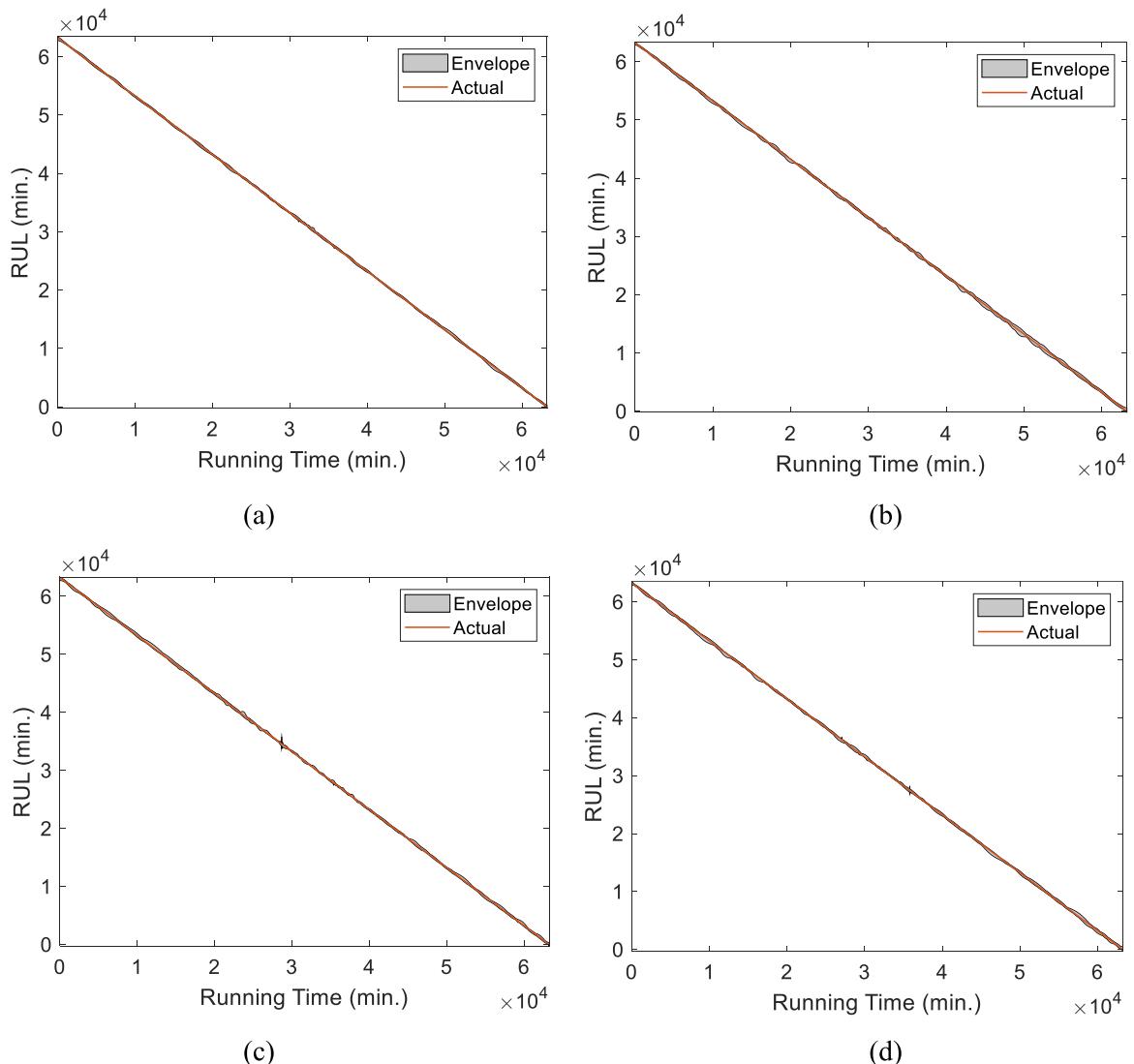
Operation Condition	Dataset Index	Fault type	Time Duration
Condition 1 (2100 rpm/12 kN)	#1	Outer race	123 min.
	#2	Outer race	161 min.
	#3	Outer race	158 min.
	#4	Cage	122 min.
	#5	Inner race and outer face	52 min.
Condition 2 (2250 rpm/11 kN)	#6	Inner race	491 min.
	#7	Outer race	161 min.
	#8	Cage	533 min.
	#9	Outer race	42 min.
	#10	Outer race	339 min.
Condition 3 (2400 rpm/10 kN)	#11	Outer race	2538 min.
	#12	Inner race, outer race, ball and cage	2496 min.
	#13	Inner race	371 min.
	#14	Inner race	1515 min.
	#15	Outer race	114 min.

Note: datasets with gray highlights are used in this research.

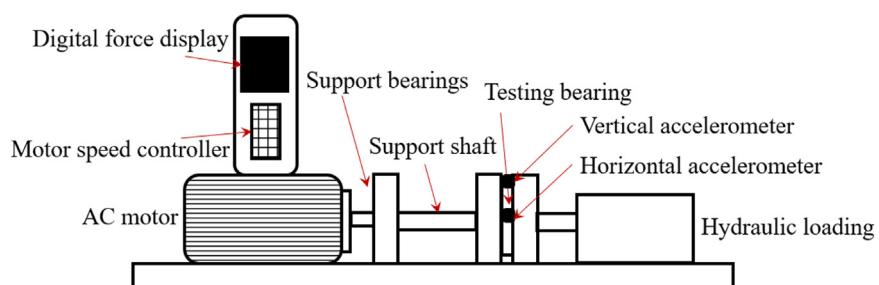


**Fig. 14.** Intrinsic mode functions (IMFs) and residuals retrieved from different statistical metrics (a) root mean square; (b) peak-to-peak value; (c) kurtosis; (d) skewness.

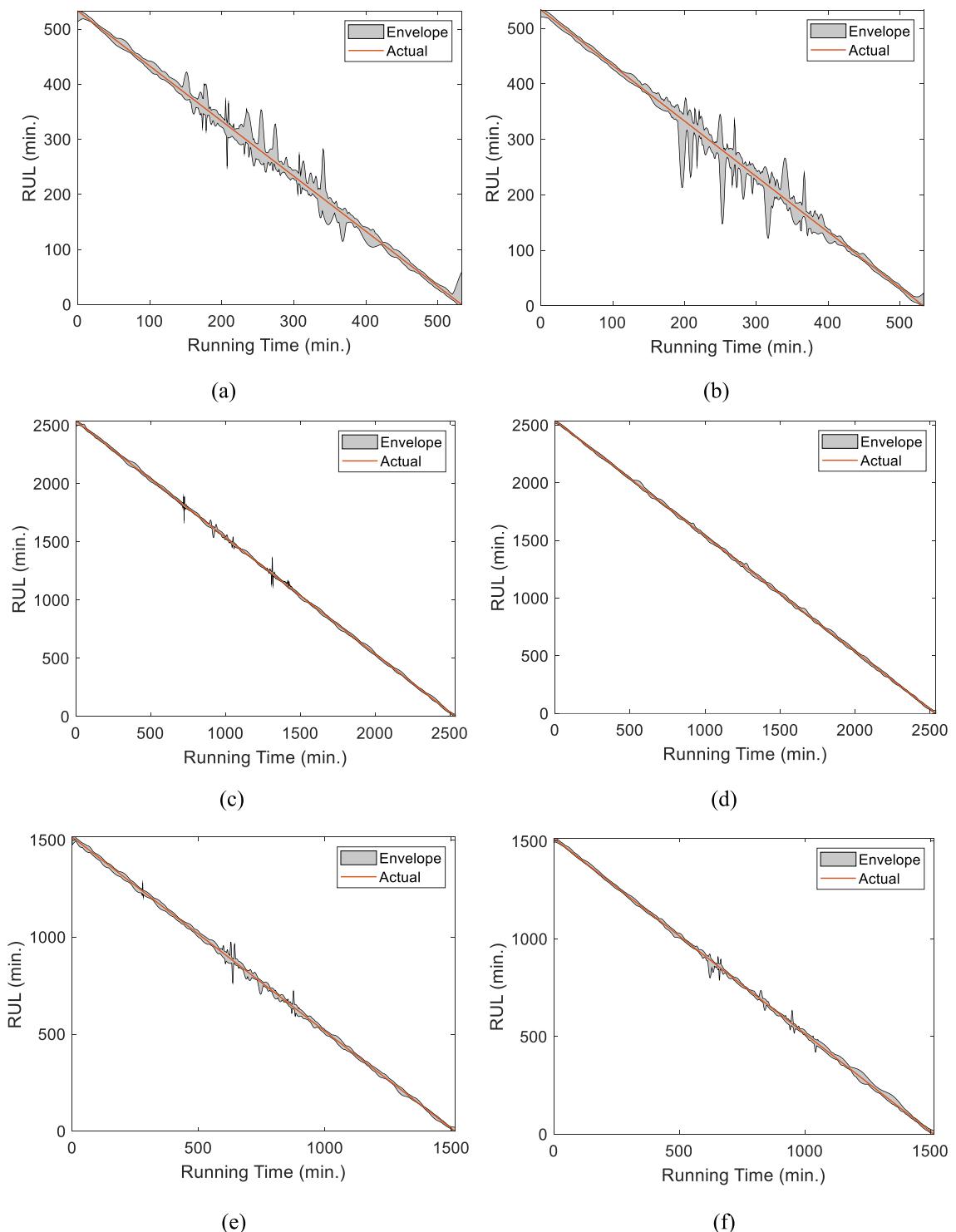
should be retrieved upon the entire time-series vibration signals, which may not be available in the *online/real-time* bearing RUL prediction task. To tackle this issue, we can periodically collect more data to carry out analysis and update the RUL prediction as we proceed. Another possible way of extending data for reliable RUL prediction is the adaptive Kalman filter, which is worthy of further investigation. Given all results shown above, the proof of concept using the proposed methodology that exploits the collective strength of suitable signal preprocessing and WNN has been conveyed, demonstrating the contribution of this research.



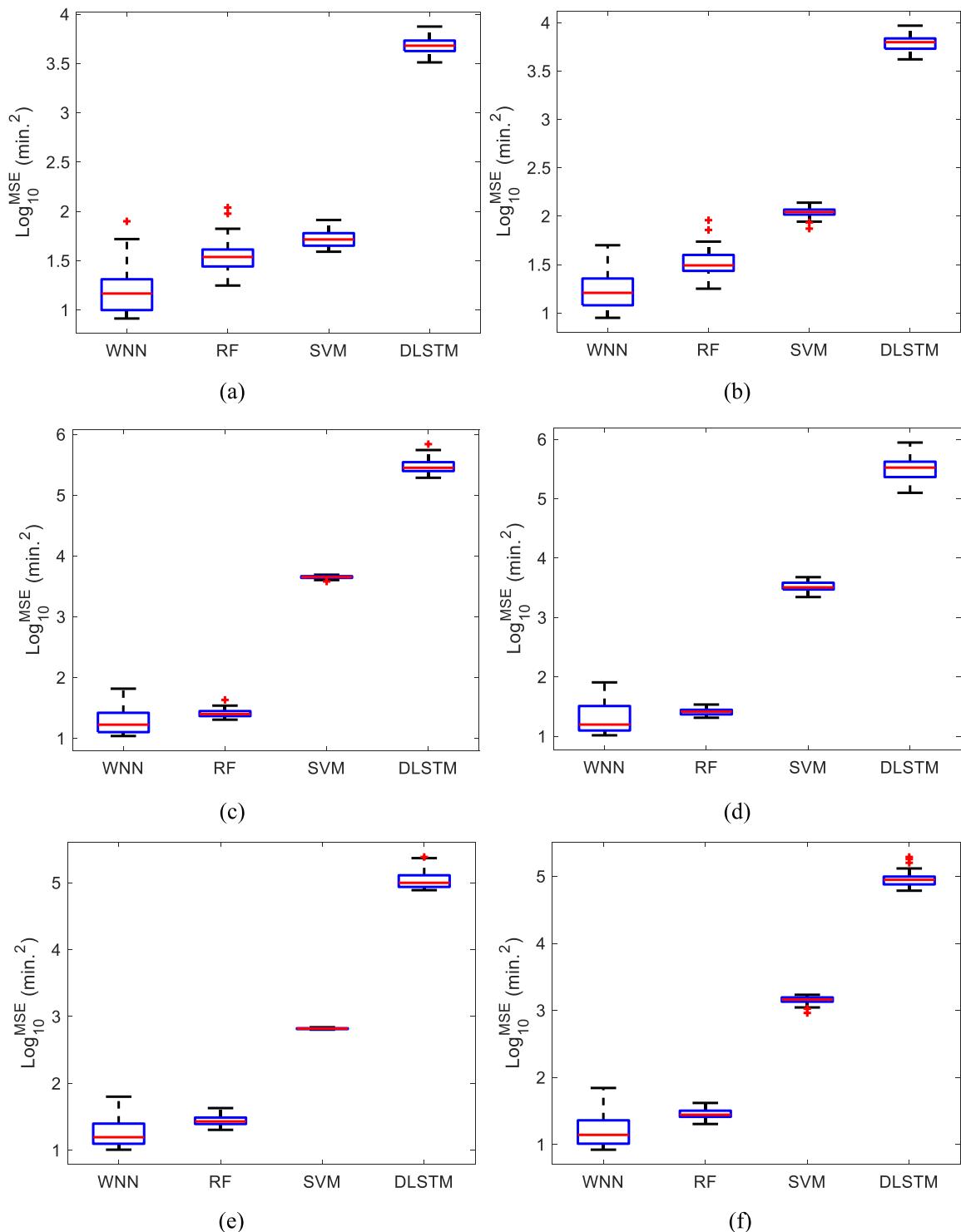
**Fig. 15.** RUL prediction envelopes (lower and upper bounds) (a) data collected from bearing 1; (b) data collected from bearing 2; (c) data collected from bearing 3; (d) data collected from bearing 4.



**Fig. 16.** Test rig for XJTU-SY bearing datasets [47].



**Fig. 17.** RUL prediction envelope (lower and upper bounds) (a) #8 dataset (horizontal vibration); (b) #8 dataset (vertical vibration); (c) #11 dataset (horizontal vibration); (d) #11 dataset (vertical vibration); (e) #14 dataset (horizontal vibration); (f) #14 dataset (vertical vibration).



**Fig. 18.** Comparison of prediction accuracy over 50 emulation runs among different methods (a) #8 dataset (horizontal vibration); (b) #8 dataset (vertical vibration); (c) #11 dataset (horizontal vibration); (d) #11 dataset (vertical vibration); (e) #14 dataset (horizontal vibration); (f) #14 dataset (vertical vibration).

## 4. Conclusion

In this research, a new bearing fault prognosis framework is synthesized, which mainly consists of the sequential time-domain signal preprocessing for bearing degradation feature extraction followed by the physics-based wavelet neural network (WNN) for bearing remaining useful life (RUL) prediction. Representative statistical metrics in the time domain are first extracted, followed by the empirical mode decomposition (EMD) to continuously highlight the essential degradation features of a bearing. Such signal preprocessing serves as a performance multiplier of WNN for subsequent bearing fault prognosis. With the degradation features as the input, the data correlation with respect to fault occurrence is adequately and effectively learned by the proposed WNN. The low-dimensional degradation features greatly facilitate the WNN model training with high computational efficiency. Compared to traditional neural networks, the wavelet function embedded in WNN allows the physical interpretation of the influence of degradation features on the bearing RUL. The B-spline mother wavelet particularly adopted in this research, is well suited to the EMD technique that is built upon the envelope interpolation using cubic spline. Different publicly accessible bearing datasets, i.e., NASA Ames Prognostics and XJTU-SY bearing datasets, are utilized to conduct the comprehensive case illustrations, and the efficacy of the proposed framework is holistically validated by comparing it with other benchmark methods.

## Funding

This research is supported by the National Science Foundation under grant CMMI – 2138522 and IIS – 1741174.

## Data availability

Data will be made available on request.

## References

- [1] I. El-Thalji, E. Jantunen, A summary of fault modelling and predictive health monitoring of rolling element bearings, *Mech. Syst. Signal Process.* 60–61 (2015) 252–272, doi:[10.1016/j.ymssp.2015.02.008](https://doi.org/10.1016/j.ymssp.2015.02.008).
- [2] Naipeng Li, Yaguo Lei, Z. Liu, J. Lin, A particle filtering-based approach for remaining useful life predication of rolling element bearings, in: Proceedings of the International Conference on Prognostics and Health Management, IEEE, 2014, pp. 1–8, doi:[10.1109/ICPHM.2014.7036367](https://doi.org/10.1109/ICPHM.2014.7036367).
- [3] L. Cui, X. Wang, H. Wang, H. Jiang, Remaining useful life prediction of rolling element bearings based on simulated performance degradation dictionary, *Mech. Mach. Theory* 153 (2020) 103967, doi:[10.1016/j.mechmachtheory.2020.103967](https://doi.org/10.1016/j.mechmachtheory.2020.103967).
- [4] D. An, N.H. Kim, J.H. Choi, Practical options for selecting data-driven or physics-based prognostics algorithms with reviews, *Reliab. Eng. Syst. Saf.* 133 (2015) 223–236, doi:[10.1016/j.ress.2014.09.014](https://doi.org/10.1016/j.ress.2014.09.014).
- [5] J. Singh, M. Azamfar, F. Li, J. Lee, A systematic review of machine learning algorithms for prognostics and health management of rolling element bearings: fundamentals, concepts and applications, *Meas. Sci. Technol.* 32 (2020) 012001, doi:[10.1088/1361-6501/ab8df9](https://doi.org/10.1088/1361-6501/ab8df9).
- [6] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak, Remaining useful life estimation based on nonlinear feature reduction and support vector regression, *Eng. Appl. Artif. Intell.* 26 (2013) 1751–1760, doi:[10.1016/j.engappai.2013.02.006](https://doi.org/10.1016/j.engappai.2013.02.006).
- [7] M. Zhao, B. Tang, Q. Tan, Bearing remaining useful life estimation based on time–frequency representation and supervised dimensionality reduction, *Measurement* 86 (2016) 41–55, doi:[10.1016/j.measurement.2015.11.047](https://doi.org/10.1016/j.measurement.2015.11.047).
- [8] T. Gao, Y. Li, X. Huang, C. Wang, Data-driven method for predicting remaining useful life of bearing based on Bayesian theory, *Sensors* 21 (2020) 182, doi:[10.3390/s21010182](https://doi.org/10.3390/s21010182).
- [9] J. Lyu, R. Ying, N. Lu, B. Zhang, Remaining useful life estimation with multiple local similarities, *Eng. Appl. Artif. Intell.* 95 (2020) 103849, doi:[10.1016/j.engappai.2020.103849](https://doi.org/10.1016/j.engappai.2020.103849).
- [10] J. Zhu, N. Chen, W. Peng, Estimation of bearing remaining useful life based on multiscale convolutional neural network, *IEEE Trans. Ind. Electron.* 66 (2019) 3208–3216, doi:[10.1109/TIE.2018.2844856](https://doi.org/10.1109/TIE.2018.2844856).
- [11] L. Ren, Y. Sun, J. Cui, L. Zhang, Bearing remaining useful life prediction based on deep autoencoder and deep neural networks, *J. Manuf. Syst.* 48 (2018) 71–77, doi:[10.1016/j.jmssy.2018.04.008](https://doi.org/10.1016/j.jmssy.2018.04.008).
- [12] X. Huang, P. Zhang, W. Shi, S. Dong, C. Wen, H. Lin, X. Chen, Frequency Hoyer attention based convolutional neural network for remaining useful life prediction of machinery, *Meas. Sci. Technol.* 32 (2021) 125108, doi:[10.1088/1361-6501/ac22f0](https://doi.org/10.1088/1361-6501/ac22f0).
- [13] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, J. Wang, Machine health monitoring using local feature-based gated recurrent unit networks, *IEEE Trans. Ind. Electron.* 65 (2018) 1539–1548, doi:[10.1109/TIE.2017.2733438](https://doi.org/10.1109/TIE.2017.2733438).
- [14] K. Zhou, E. Diehl, J. Tang, Deep convolutional generative adversarial network with semi-supervised learning enabled physics elucidation for extended gear fault diagnosis under data limitations, *Mech. Syst. Signal Process.* 185 (2023) 109772, doi:[10.1016/j.ymssp.2022.109772](https://doi.org/10.1016/j.ymssp.2022.109772).
- [15] A. Mosallam, K. Medjaher, N. Zerhouni, Nonparametric time series modelling for industrial prognostics and health management, *Int. J. Adv. Manuf. Technol.* 69 (2013) 1685–1699, doi:[10.1007/s00170-013-5065-z](https://doi.org/10.1007/s00170-013-5065-z).
- [16] M. Motahari-Nezhad, S.M. Jafari, Bearing remaining useful life prediction under starved lubricating condition using time domain acoustic emission signal processing, *Expert Syst. Appl.* 168 (2021) 114391, doi:[10.1016/j.eswa.2020.114391](https://doi.org/10.1016/j.eswa.2020.114391).
- [17] L. Liu, X. Song, K. Chen, B. Hou, X. Chai, H. Ning, An enhanced encoder–decoder framework for bearing remaining useful life prediction, *Measurement* 170 (2021) 108753, doi:[10.1016/j.measurement.2020.108753](https://doi.org/10.1016/j.measurement.2020.108753).
- [18] M. Qiu, W. Li, F. Jiang, Z. Zhu, Remaining useful life estimation for rolling bearing with SIOS-based indicator and particle filtering, *IEEE Access* 6 (2018) 24521–24532, doi:[10.1109/ACCESS.2018.2831455](https://doi.org/10.1109/ACCESS.2018.2831455).
- [19] L. Xu, P. Pennacchi, S. Chatterton, A new method for the estimation of bearing health state and remaining useful life based on the moving average cross-correlation of power spectral density, *Mech. Syst. Signal Process.* 139 (2020) 106617, doi:[10.1016/j.ymssp.2020.106617](https://doi.org/10.1016/j.ymssp.2020.106617).
- [20] C. Cheng, G. Ma, Y. Zhang, M. Sun, F. Teng, H. Ding, Y. Yuan, A deep learning-based remaining useful life prediction approach for bearings, *IEEE/ASME Trans. Mechatron.* 25 (2020) 1243–1254, doi:[10.1109/TMECH.2020.2971503](https://doi.org/10.1109/TMECH.2020.2971503).
- [21] M. Liang, K. Zhou, Probabilistic bearing fault diagnosis using Gaussian process with tailored feature extraction, *Int. J. Adv. Manuf. Technol.* 119 (2022) 2059–2076, doi:[10.1007/s00170-021-08392-6](https://doi.org/10.1007/s00170-021-08392-6).
- [22] K. Zhou, A deep long short-term memory network for bearing fault diagnosis under time-varying conditions, in: Proceedings of the 34th Conference on Mechanical Vibration and Sound, American Society of Mechanical Engineers, 2022, doi:[10.1115/DETC2022-88808](https://doi.org/10.1115/DETC2022-88808).

- [23] H. Tang, Y. Du, H.L. Dai, Rolling element bearing diagnosis based on probability box theory, *Appl. Math. Model.* 80 (2020) 944–960, doi:[10.1016/j.apm.2019.10.068](https://doi.org/10.1016/j.apm.2019.10.068).
- [24] L. Jiang, J. Xuan, T. Shi, Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis, *Mech. Syst. Signal Process.* 41 (2013) 113–126, doi:[10.1016/j.ymssp.2013.05.017](https://doi.org/10.1016/j.ymssp.2013.05.017).
- [25] J. Yu, Z. Guo, Remaining useful life prediction of planet bearings based on conditional deep recurrent generative adversarial network and action discovery, *J. Mech. Sci. Technol.* 35 (2021) 21–30, doi:[10.1007/s12206-020-1202-4](https://doi.org/10.1007/s12206-020-1202-4).
- [26] J. Ma, F. Xu, K. Huang, R. Huang, GMAR-GARCH model and its application in feature extraction for rolling bearing fault diagnosis, *Mech. Syst. Signal Process.* 93 (2017) 175–203, doi:[10.1016/j.ymssp.2017.01.043](https://doi.org/10.1016/j.ymssp.2017.01.043).
- [27] G. Georgoulas, T. Loutas, C.D. Stylios, V. Kostopoulos, Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition, *Mech. Syst. Signal Process.* 41 (2013) 510–525, doi:[10.1016/j.ymssp.2013.02.020](https://doi.org/10.1016/j.ymssp.2013.02.020).
- [28] C. Xu, S. Du, P. Gong, Z. Li, G. Chen, G. Song, An improved method for pipeline leakage localization with a single sensor based on modal acoustic emission and empirical mode decomposition with Hilbert transform, *IEEE Sens. J.* 20 (2020) 5480–5491, doi:[10.1109/JSEN.2020.2971854](https://doi.org/10.1109/JSEN.2020.2971854).
- [29] D. Yu, J. Cheng, Y. Yang, Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, *Mech. Syst. Signal Process.* 19 (2005) 259–270, doi:[10.1016/S0888-3270\(03\)00099-2](https://doi.org/10.1016/S0888-3270(03)00099-2).
- [30] B. Wei, B. Xie, H. Li, Z. Zhong, Y. You, An improved Hilbert–Huang transform method for modal parameter identification of a high arch dam, *Appl. Math. Model.* 91 (2021) 297–310, doi:[10.1016/j.apm.2020.09.048](https://doi.org/10.1016/j.apm.2020.09.048).
- [31] Y.H. Wang, C.H. Yeh, H.W.V. Young, K. Hu, M.T. Lo, On the computational complexity of the empirical mode decomposition algorithm, *Phys. A Stat. Mech. Appl.* 400 (2014) 159–167, doi:[10.1016/j.physa.2014.01.020](https://doi.org/10.1016/j.physa.2014.01.020).
- [32] Jun Zhang, G.G. Walter, Y. Miao, W.N.W. Lee, Wavelet neural networks for function learning, *IEEE Trans. Signal Process.* 43 (1995) 1485–1497, doi:[10.1109/78.388860](https://doi.org/10.1109/78.388860).
- [33] T. Li, Z. Zhao, C. Sun, L. Cheng, X. Chen, R. Yan, R.X. Gao, WaveletKernelNet: an interpretable deep neural network for industrial intelligent diagnosis, *IEEE Trans. Syst. Man Cybern. Syst.* 52 (2022) 2302–2312, doi:[10.1109/TSMC.2020.3048950](https://doi.org/10.1109/TSMC.2020.3048950).
- [34] B.R. Nayana, P. Geethanjali, Analysis of statistical time-domain features effectiveness in identification of bearing faults from vibration signal, *IEEE Sens. J.* 17 (2017) 5618–5625, doi:[10.1109/JSEN.2017.2727638](https://doi.org/10.1109/JSEN.2017.2727638).
- [35] E. Deléchelle, J. Lemoine, O. Niang, Empirical mode decomposition: an analytical approach for sifting process, *IEEE Signal Process. Lett.* 12 (2005) 764–767, doi:[10.1109/LSP.2005.856878](https://doi.org/10.1109/LSP.2005.856878).
- [36] A.K. Alexandridis, A.D. Zapranis, Wavelet neural networks: a practical guide, *Neural Netw.* 42 (2013) 1–27, doi:[10.1016/j.neunet.2013.01.008](https://doi.org/10.1016/j.neunet.2013.01.008).
- [37] H. Esen, F. Ozgen, M. Esen, A. Sengur, Artificial neural network and wavelet neural network approaches for modelling of a solar air heater, *Expert Syst. Appl.* 36 (2009) 11240–11248, doi:[10.1016/j.eswa.2009.02.073](https://doi.org/10.1016/j.eswa.2009.02.073).
- [38] Y. Chen, B. Yang, J. Dong, Time-series prediction using a local linear wavelet neural network, *Neurocomputing* 69 (2006) 449–465, doi:[10.1016/j.neucom.2005.02.006](https://doi.org/10.1016/j.neucom.2005.02.006).
- [39] N. Wang, H. Adeli, Self-constructing wavelet neural network algorithm for nonlinear control of large structures, *Eng. Appl. Artif. Intell.* 41 (2015) 249–258, doi:[10.1016/j.engappai.2015.01.018](https://doi.org/10.1016/j.engappai.2015.01.018).
- [40] M. Rana, I. Koprinska, Forecasting electricity load with advanced wavelet neural networks, *Neurocomputing* 182 (2016) 118–132, doi:[10.1016/j.neucom.2015.12.004](https://doi.org/10.1016/j.neucom.2015.12.004).
- [41] X. Li, J. Ouyang, T. Jiang, B. Yang, Integration modified wavelet neural networks for solving thin plate bending problem, *Appl. Math. Model.* 37 (2013) 2983–2994, doi:[10.1016/j.apm.2012.07.036](https://doi.org/10.1016/j.apm.2012.07.036).
- [42] J.W. Liu, F.L. Zuo, Y.X. Guo, T.Y. Li, J.M. Chen, Research on improved wavelet convolutional wavelet neural networks, *Appl. Intell.* 51 (2021) 4106–4126, doi:[10.1007/s10489-020-02015-5](https://doi.org/10.1007/s10489-020-02015-5).
- [43] M. Maynard, *Neural Networks: Introduction to Artificial Neurons, Backpropagation and Multilayer Feedforward Neural Networks With Real-World Applications*, Independently Published, 2020.
- [44] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, Springer, New York, 2014.
- [45] S. Ruder, An overview of gradient descent optimization algorithms, (2016). <http://arxiv.org/abs/1609.04747>.
- [46] J. Lee, J. Qiu, H. Yu, G. Lin, Bearing data set, (2007). <http://ti.arc.nasa.gov/project/prognostic-data-repository>.
- [47] B. Wang, Y. Lei, N. Li, N. Li, A hybrid prognostics approach for estimating remaining useful life of rolling element bearings, *IEEE Trans. Reliab.* 69 (2020) 401–412, doi:[10.1109/TR.2018.2882682](https://doi.org/10.1109/TR.2018.2882682).
- [48] Y. Lei, *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*, Elsevier Science, 2016.
- [49] X. Xu, Y. Lei, Z. Li, An incorrect data detection method for big data cleaning of machinery condition monitoring, *IEEE Trans. Ind. Electron.* 67 (2020) 2326–2336, doi:[10.1109/TIE.2019.2903774](https://doi.org/10.1109/TIE.2019.2903774).
- [50] H.R. Martin, F. Honarvar, Application of statistical moments to bearing failure detection, *Appl. Acoust.* 44 (1995) 67–77, doi:[10.1016/0003-682X\(94\)P4420-B](https://doi.org/10.1016/0003-682X(94)P4420-B).
- [51] L. Saidi, J. Ben Ali, E. Bechhoefer, M. Benbouzid, Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR, *Appl. Acoust.* 120 (2017) 1–8, doi:[10.1016/j.apacoust.2017.01.005](https://doi.org/10.1016/j.apacoust.2017.01.005).
- [52] N. Tandon, A comparison of some vibration parameters for the condition monitoring of rolling element bearings, *Measurement* 12 (1994) 285–289, doi:[10.1016/0263-2241\(94\)90033-7](https://doi.org/10.1016/0263-2241(94)90033-7).
- [53] M. Kuhn, K. Johnson, *Feature Engineering and Selection: A Practical Approach for Predictive Models*, CRC Press, 2019.
- [54] A. Teolis, *Computational Signal Processing With Wavelets*, Birkhäuser Boston, 2012.
- [55] G. Cheng, X. Wang, Y. He, Remaining useful life and state of health prediction for lithium batteries based on empirical mode decomposition and a long and short memory neural network, *Energy* 232 (2021) 121022, doi:[10.1016/j.energy.2021.121022](https://doi.org/10.1016/j.energy.2021.121022).