Adaptive Res-LSTM Attention-based Remaining Useful Lifetime Prognosis of Rolling Bearings

Boubker Najdi[1], Mohammed Benbrahim[2], Mohammed Nabil Kabbaj[3]

[1,2,3] *Systems Engineering, Modeling and Analysis Laboratory, Faculty of Sciences, Sidi Mohamed Ben Abdellah University, Fez, Morocco.*

aboubakr.najdi@usmba.ac.ma

mohammed.benbrahim@usmba.ac.ma

n.kabbaj@usmba.ac.ma

ABSTRACT

Predicting the Remaining Useful Lifetime (RUL) of bearings is crucial for the maintenance and reliability of rotating machinery. This paper presents a novel approach utilizing PRONOSTIA and XJTU-SY datasets for RUL prediction. The proposed methodology leverages Synchrosqueezing Wavelet Transform (SSWT) and Random Projection (RP) to extract significant features from vibration signals. These features are then fed into a Residual Network (ResNet) combined with a temporal attention layer, followed by a Long Short-Term Memory (LSTM) model, referred to as the Adaptive Residual Attention LSTM (ARAL), to assess the Health Indicator (HI) of the bearings. Notably, an exponential data labeling technique is employed instead of traditional linear labeling, enhancing the robustness of the HI assessment. Following the HI estimation, the three-sigma method is applied to identify the degradation starting point. Subsequently, Gaussian Process Regression (GPR) is utilized to predict the RUL from this point forward. The proposed method demonstrates superior performance compared to existing techniques, providing more accurate and reliable RUL predictions. Experimental results show that this integrated approach effectively captures the complex degradation patterns of bearings, making it a valuable tool for prognostics and health management in industrial applications.

1. INTRODUCTION

Rolling bearings are critical components in mechanical equipment, with their degradation directly impacting overall stability and safety. Degradation in bearings is often caused by multiple mechanisms, including wear, fatigue, and microfractures in materials, which develop under fluctuating loads and unstable operating conditions. These phenomena can compound over time, leading to more severe damage and eventual failure. Early fault detection and accurate prediction of their RUL are essential for optimizing maintenance, preventing accidents, and enhancing efficiency. Manual monitoring in such complex conditions is ineffective. With the rise of artificial intelligence, Prognostics and Health Management (PHM) has emerged as a vital field, yielding significant research and practical applications (X. Li, Yu, Lei, Li, & Yang, 2024; Ding et al., 2023; Suh, Lukowicz, & Lee, 2022).

To estimate rolling bearing RUL, both model-based and data-driven methods are commonly employed. Model-based methods aim to develop a physical model describing the degradation trend of the mechanical system or component (Londhe, Arakere, & Subhash, 2017; Qin, Xiang, Chai, & Chen, 2020). However, creating such models for complex equipment under varying conditions is challenging and limits their applicability. Conversely, data-driven methods use machine learning and deep learning technologies to analyze measurement data and uncover the dynamic characteristics of mechanical systems (Que & Xu, 2019).

Data-driven prognostics methods generally follow a three-step framework: data acquisition, HI construction, and RUL prediction (Lei et al., 2018). The HI in bearings serves as an abstract measure of the overall health or degradation status of the bearing. It is derived from features extracted from vibration signals that reflect the bearing's condition. The accuracy of RUL prediction is highly dependent on the HI construction phase, making the extraction of meaningful features from measured signals a critical challenge. A well-constructed HI, based on high-quality features, can efficiently track bearing degradation (Zhang, Tang, Han, & Deng, 2017).

Historically, bearing fault diagnosis and RUL prediction have relied on signal processing techniques such as Wavelet Packet Decomposition (WPD) and Empirical Mode Decomposition (EMD), which effectively extract features from vibration signals. However, these methods often struggle with non-stationary

Boubker Najdi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

https://doi.org/10.36001/IJPHM.2025.v16i1.4171

signals and noise, limiting their accuracy and applicability.

With advancements in sensing technology and computer science, big data-driven deep learning has become a powerful tool for autonomously extracting valuable features from extensive datasets. This approach has garnered significant attention in rolling bearing RUL prediction. For instance, time-frequency features, reduced in dimensionality, have been employed as inputs to multi-scale Convolutional Neural Networks (CNN) for assessing the remaining life of rolling bearings (Zhu, Chen, & Peng, 2019a). Other studies have proposed LSTM networks utilizing frequency domain features for RUL prediction of gears (Yan, Qin, Xiang, Wang, & Chen, 2020). Additionally, multi-technique prognostic approaches integrating multiple machine learning and deep learning models have been developed for RUL prediction (L. Li, Xu, & Li, 2023), while 2D Time-Scale (TS) images generated by Continuous Wavelet Transform (CWT) have been used as inputs for 2D CNNs to develop health indicators for predicting the remaining life of bearings (Yoo & Baek, 2018). Despite these advancements, many deep learning models still struggle with the vanishing gradient problem, which hampers the training of deep networks and limits their ability to capture complex patterns in the data.

Attention mechanisms (Vaswani et al., 2023), which can associate different parts of a sequence to create a distinctive representation, have gained success in diverse applications such as natural language processing, computer vision, and RUL prediction (Deng et al., 2023). These mechanisms function by establishing relationships between features from different points in the sequence, assigning varying levels of importance to these features, which helps to highlight critical information while downplaying less relevant data. By integrating a temporal attention mechanism, it becomes possible to effectively extract significant degradation signals from sensor data, while minimizing the impact of irrelevant information.

Building on these advancements, the SSWT and RP have shown great promise in our previous research on bearing fault diagnosis by capturing Time-Frequency (TF) information and reducing feature dimensionality, leading to better accuracy (Najdi, Benbrahim, & Kabbaj, 2024). SSWT (Daubechies, Lu, & Wu, 2011) provides precise localization of signal features, enhancing the detection of subtle changes. RP reduces computational complexity while retaining essential information from high-dimensional data (Achlioptas, 2003), addressing key limitations of traditional methods in non-stationary environments.

To address existing deep learning limitations, this paper proposes ARAL, a novel approach integrating a ResNet-based CNN, a temporal attention layer, and an LSTM network. This combination enhances feature extraction and improves RUL prediction accuracy. ResNet CNNs mitigate the vanishing gradient problem (Shah, Kadam, Shah, & Shinde, 2016), enabling deeper network training and complex feature extraction from vibration signals. The attention mechanism highlights the most relevant input data, and the LSTM network captures temporal dependencies, effectively modeling the data's sequential nature.

The main contributions of this research are the following:

- **Enhanced Feature Extraction**: Leveraging SSWT and RP to effectively capture and reduce the dimensionality of significant features from vibration signals, addressing key limitations of traditional methods.
- **Novel Health Indicator**: Introducing ARAL, combining a ResNet-based CNN, a temporal attention layer, and an LSTM network to improve the accuracy of HI assessments and RUL predictions by mitigating the vanishing gradient problem and capturing complex temporal dependencies.
- **Exponential Data Labeling**: Utilizing an exponential data labeling technique for more robust HI assessment and applying GPR following a three-sigma method to predict RUL, demonstrating superior performance in capturing complex degradation patterns and providing reliable RUL predictions.

The rest of the paper is structured as follows: Section 2 presents the theoretical framework of the proposed RUL estimation method. Section 3 details the experimental validation and results analysis using two bearing datasets. Finally, Section 4 provides the conclusion.

2. THE PROPOSED METHOD FOR RUL PREDICTION

2.1. The SSWT-RP Signal Transformation

The SSWT (Daubechies et al., 2011) is an advanced TF analysis method that enhances the concentration of a signal's energy in the Time-Frequency plane while preserving the time resolution. Given a signal $s(t)$, the CWT of $s(t)$ at scale a and time t is defined as (Sadowsky, 1994):

$$W_\psi s(a,t) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(\tau)\psi^* \left(\frac{\tau - t}{a}\right) d\tau \qquad (1)$$

From the CWT, the instantaneous frequency (IF) $\omega_s(a,t)$ is derived as:

$$\omega_s(a,t) = \frac{1}{2\pi j} \frac{\partial}{\partial t} \left(\ln W_\psi s(a,t)\right) \qquad (2)$$

The SSWT then refines the CWT by reassigning its coefficients from the Time-Scale plane (t,a) to the Time-Frequency plane $(t, \omega_s(a,t))$, effectively mapping:

$$SSWT\, s(t,\xi) = \int_0^\infty W_\psi s(a,t)\delta\left(\xi - \omega_s(a,t)\right)\frac{da}{a^{5/2}} \quad (3)$$

where ξ represents the reassigned frequency. This process results in a TF matrix $SSWT\, s(t,\xi)$, which provides a clearer and more concentrated representation of the signal's energy distribution over time and frequency.

Transforming big signals to TF can result in some redundancy in the representation which will be cumbersome for processing. To reduce the dimensionality of the TF matrix while preserving its essential features, we apply RP. Let the TF representation obtained from SSWT be $X \in \mathbb{R}^{m\times n}$, where m is the number of time points and n is the number of frequency bins. In RP, the original high-dimensional data X is projected into a lower-dimensional space using a randomly generated matrix $R \in \mathbb{R}^{k\times n}$, where $k \ll n$:

$$Y = X \cdot R^T \quad (4)$$

Here, $Y \in \mathbb{R}^{m\times k}$ is the reduced-dimensionality representation of the TF matrix. The elements of R are chosen to be independent and identically distributed (i.i.d.) and scaled to unit length. This projection approximately preserves the distances between the original data points, as guaranteed by the Johnson-Lindenstrauss lemma (Johnson & Lindenstrauss, 1984):

$$(1-\epsilon)\|X_i - X_j\|^2 \le \|Y_i - Y_j\|^2 \le (1+\epsilon)\|X_i - X_j\|^2 \quad (5)$$

for any $0 < \epsilon < 1$ and for all pairs of data points X_i, X_j in the original high-dimensional space.

The resulting matrix Y retains the most significant information from the original TF representation but in a much lower-dimensional space, making it suitable for subsequent processing tasks such as classification or regression.

2.2. ARAL

The input matrix from the SSWT-RP method is denoted as $Y \in \mathbb{R}^{H\times W}$, where H is the height and W is the width. The convolutional layer applies a set of filters (kernels) to Y to produce feature maps, as shown in Figure 1. Let $K \in \mathbb{R}^{k_H\times k_W\times F}$ be the filter, where k_H and k_W are the height and width of the filter, and F is the number of filters. The output $Y' \in \mathbb{R}^{H'\times W'\times F}$ is computed as:

$$Y'_{h',w',f} = \sum_{i=1}^{k_H}\sum_{j=1}^{k_W} K_{i,j,f}\cdot Y_{h'+i-1,w'+j-1} \quad (6)$$

where H' and W' are the height and width of the output feature map from the convolutional layer, determined by the strides and padding.

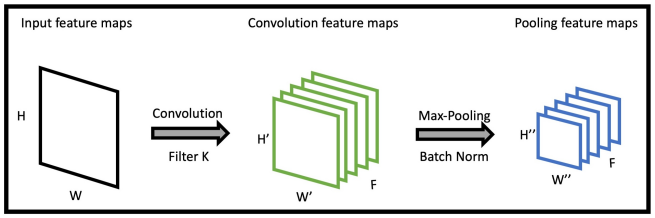


Figure 1. Convolution and pooling.

Next, Batch Normalization (BN) is applied to the output of the convolutional layer to achieve zero mean and unit variance, followed by scaling and shifting. Let $\hat{Y}$ be the normalized output:

$$\hat{Y}'_{h',w',f} = \gamma_f\left(\frac{Y'_{h',w',f} - \mu_f}{\sqrt{\sigma_f^2 + \epsilon}}\right) + \beta_f \quad (7)$$

where μ_f and σ_f^2 are the mean and variance of the feature map f, γ_f and β_f are learnable parameters, and ϵ is a small constant added for numerical stability.

The ReLU activation function introduces non-linearity:

$$\tilde{Y}'_{h',w',f} = \max(0, \hat{Y}'_{h',w',f}) \quad (8)$$

MaxPooling reduces the spatial dimensions (height and width) of the feature maps by taking the maximum value over a pooling window:

$$P_{h'',w'',f} = \max_{(i,j)\in \text{pool\_window}} \tilde{Y}'_{h'\cdot s_h+i,w'\cdot s_w+j,f} \quad (9)$$

where h'' and w'' are the new height and width after pooling, and s_h and s_w are the strides.

The output from the MaxPooling layer, P, is then fed into a series of Residual Blocks (ResBlock), as illustrated in Figure 2. Each ResBlock consists of two convolutional layers with an identity shortcut connection. For each block k:

$$U_k = \text{Conv}(P, F_k, (k_H, k_W), \text{stride}) \quad (10)$$
$$U_k = \text{BatchNormalization}(U_k) \quad (11)$$
$$U_k = \text{ReLU}(U_k) \quad (12)$$
$$U_k = \text{Conv}(U_k, F_k, (k_H, k_W), (1,1)) \quad (13)$$
$$U_k = \text{BatchNormalization}(U_k) \quad (14)$$

If the input dimensions differ from the output dimensions, we

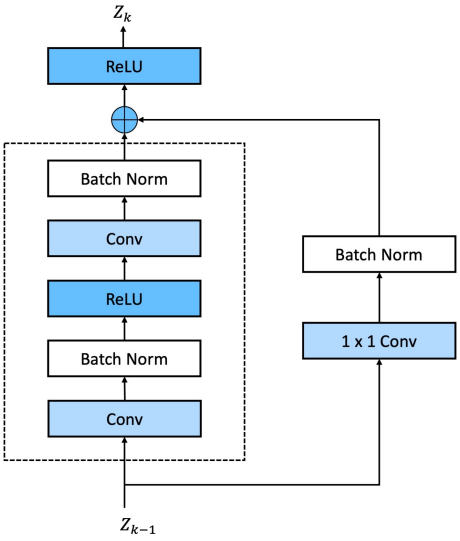


Figure 2. ResNet block.

perform a 1×1 convolution and BN on the identity mapping:

$$X_{\text{shortcut}} = \text{Conv}(P, F_k, (1,1), \text{stride}) \tag{15}$$

$$X_{\text{shortcut}} = \text{BatchNormalization}(X_{\text{shortcut}}) \tag{16}$$

Add the shortcut:

$$Z_k = \text{ReLU}(U_k + X_{\text{shortcut}}) \tag{17}$$

The final output Z_k, denoted as $Z_k \in \mathbb{R}^{H \times W \times F}$, of the ResBlock is then passed to the attention block, as illustrated in Figure 3.

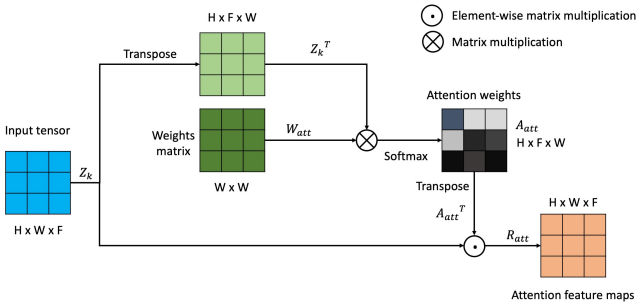


Figure 3. Temporal Attention block.

The attention mechanism starts by computing the attention weights. First, the input tensor Z_k is transposed and then multiplied by the weight matrix W_{att}. The attention mechanism assigns different weights to different time steps of the input tensor. The softmax function is applied to obtain the attention weights:

$$A_{\text{att},h,f,w} = \frac{\exp\left(\sum_j (Z_k)^T_{h,j,f} \cdot W_{\text{att},j,w}\right)}{\sum_k \exp\left(\sum_j (Z_k)^T_{h,j,f} \cdot W_{\text{att},j,k}\right)} \tag{18}$$

The final output of the attention layer is then calculated by performing an element-wise multiplication between the original input Z_k and the attention weights:

$$R_{\text{att}} = Z_k \odot A^T_{\text{att}} \tag{19}$$

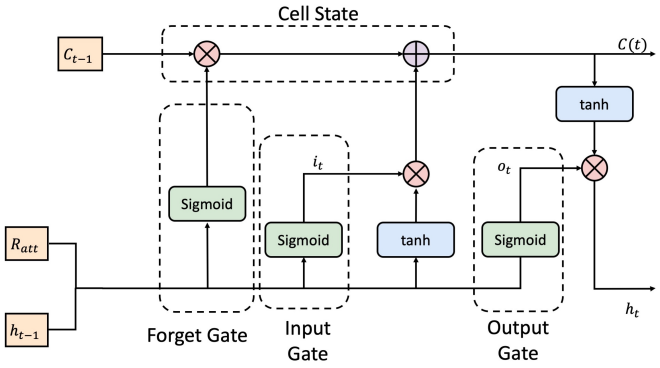


Figure 4. The LSTM architecture.

The LSTM layer processes the sequence data returned by the attention layer, capturing temporal dependencies, as represented in Figure 4. The LSTM equations are defined for each time step t as follows:

- Input gate: $i_t = \sigma(\mathbf{W}_i R_{\text{att},t} + \mathbf{U}_i h_{t-1} + \mathbf{b}_i)$
- Forget gate: $f_t = \sigma(\mathbf{W}_f R_{\text{att},t} + \mathbf{U}_f h_{t-1} + \mathbf{b}_f)$
- Cell state: $c_t = f_t \otimes c_{t-1} + i_t \otimes \tanh(\mathbf{W}_c R_{\text{att},t} + \mathbf{U}_c h_{t-1} + \mathbf{b}_c)$
- Output gate: $o_t = \sigma(\mathbf{W}_o R_{\text{att},t} + \mathbf{U}_o h_{t-1} + \mathbf{b}_o)$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function.

Finally, the output of the LSTM layer is calculated as follows:

$$h_t = o_t \otimes \tanh(c_t) \tag{20}$$

2.3. RUL prediction based on SSWT-RP-ARAL

In this section, we present the deep learning-based methodology for RUL prediction using SSWT, RP, and a combination of ResNets, Temporal Attention Block, and LSTM models. The procedure is shown in Figure 5 and comprises three stages: data preparation, HI construction, and RUL estimation. In the data preparation stage, vibration signals are transformed into TF domain features using SSWT, then dimensionality is reduced using RP. In the HI construction stage,

these features build the training and testing datasets. ResNet extracts deep features, the Temporal Attention Block enhances relevant temporal information, and LSTM captures temporal dependencies, training the model to estimate the HI values. In the RUL estimation stage, the three-sigma rule identifies the abnormal degradation point, and GPR predicts the future HI and degradation trends. The RUL is then calculated as the time difference between the HI reaching a threshold and the current time.

2.3.1. Health indicator constructed by SSWT-RP-ARAL

(i) Architecture of the ARAL-HI model

The architecture of the ARAL-HI model, presented in the Figure 6, consists of multiple stages that process the input data to construct health indicators. The initial stages involve the SSWT-RP transformed input Y, which is passed through the convolutional and ResBlocks, followed by the attention mechanism and LSTM layers. The final output of the LSTM layers serves as the input to a series of dense layers designed to further refine the health indicators.

After the LSTM layers, the output h_t is fed into one or more fully connected (dense) layers. These dense layers are equipped with dropout regularization to prevent overfitting. Let D be the dropout rate, and W_d and b_d be the weights and biases of the dense layers, respectively. The dense layer operation can be expressed as:

$$d_t = \text{Dropout}(\text{ReLU}(W_d h_t + b_d), D) \tag{21}$$

Following the dense layers, a final dense layer with a sigmoid activation function is applied to produce the HIs. The sigmoid layer converts the output into a value between 0 and 1, indicating the health state of the bearing. Let W_s and b_s be the weights and biases of the sigmoid layer. The health indicator HI_t is computed as:

$$HI_t = \sigma(W_s d_t + b_s) \tag{22}$$

The resulting HI_t represents the normalized health state at time step t, providing a comprehensive view of the bearing's condition over time.

(ii) Exponential Data labeling

Instead of using the conventional linear data labeling of the run-to-failure data, an exponential function is used to reflect the bearing defect progression over time. The exponential labeling function is defined as:

$$HealthLabel(t) = \frac{\exp(\alpha \cdot x) - 1}{\exp(\alpha) - 1} \tag{23}$$

where $x = \frac{t}{T}$, t is the current time step, T is the total time steps until failure, and α is the scaling factor. An example is shown in the Figure 7, where we construct the health label of the Bearing1_3. The scaling factor is chosen based on the best RMSE value of the training, with the scaling factors of 1.3 and 1.5 being tested, and 1.3 yielding the best results. Each time step t in the dataset is assigned a label $HealthLabel(t)$ using this exponential function, providing a more realistic representation of the bearing's degradation process.

(iii) Model Training

The network is trained using the labeled dataset, with the objective of minimizing the difference between the predicted health indicators and the true health labels. The training process includes forward propagation through the convolutional, residual, attention, LSTM, dense, and dropout layers, followed by a sigmoid activation function to produce the final HIs.

The loss function used to evaluate the model's performance is the RMSE, which is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{HI}_i - HI_i)^2} \tag{24}$$

where $\widehat{HI}_i$ is the predicted health indicator, HI_i is the true health indicator, and N is the total number of samples. The RMSE function measures the square root of the average squared differences between the predicted and true values.

The training process involves minimizing the RMSE by adjusting the model's parameters through backpropagation and Adam optimizer. This ensures that the predicted health indicators closely match the true health indicators, improving the model's prediction accuracy.

2.4. RUL estimation

The calculation of the RUL involves several key steps, each underpinned by important theoretical concepts:

2.4.1. Smoothing the HI Curve

Once the HI curve is obtained from the SSWT-RP-ARAL model, we first apply a smoothing technique to reduce noise and highlight the underlying degradation trend. The smoothing is done using a Moving Average (MA), which helps to smooth out short-term fluctuations and reveal the long-term trend in the HI data. The moving average is defined as:

$$\text{MA}(t) = \frac{1}{2k+1}\sum_{i=-k}^{k} HI(t+i) \tag{25}$$

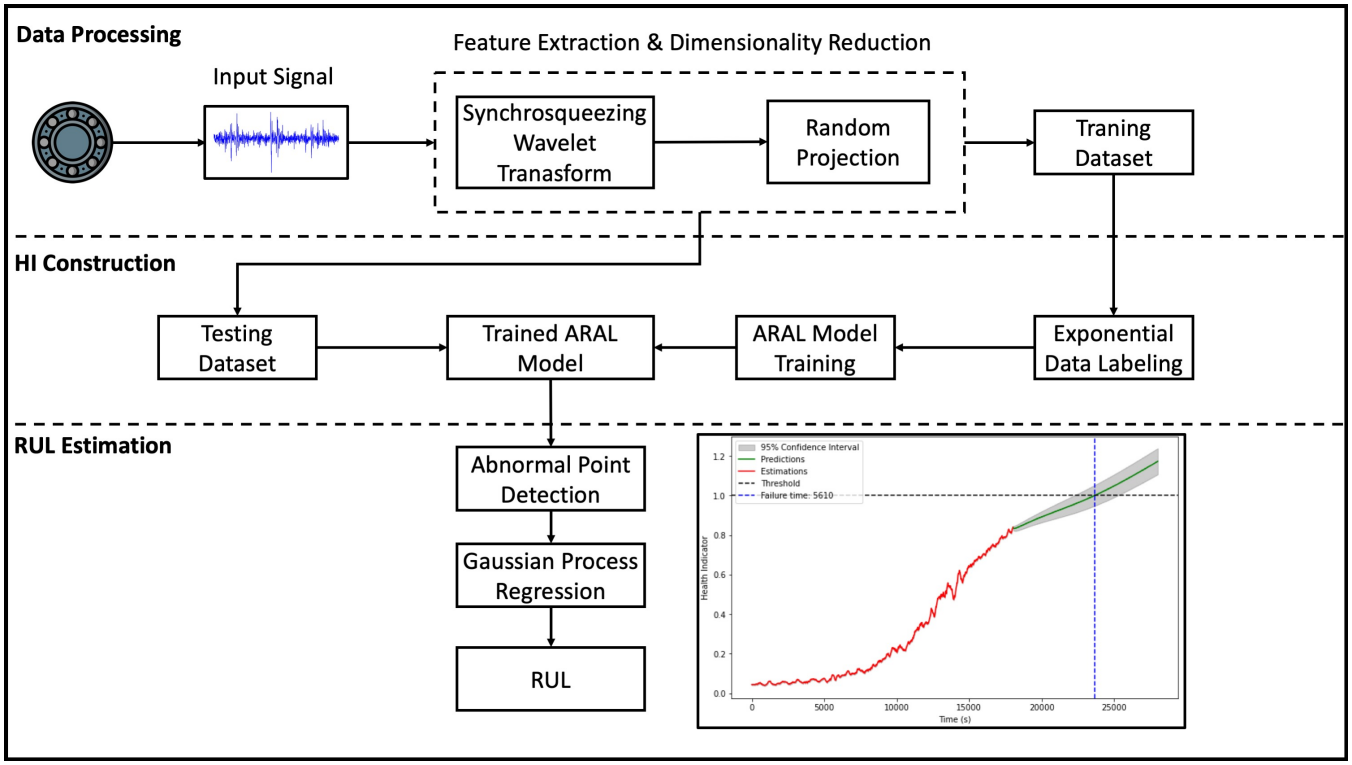


Figure 5. Framework of the proposed SSWT-RP-ARAL Method for RUL Prognosis.

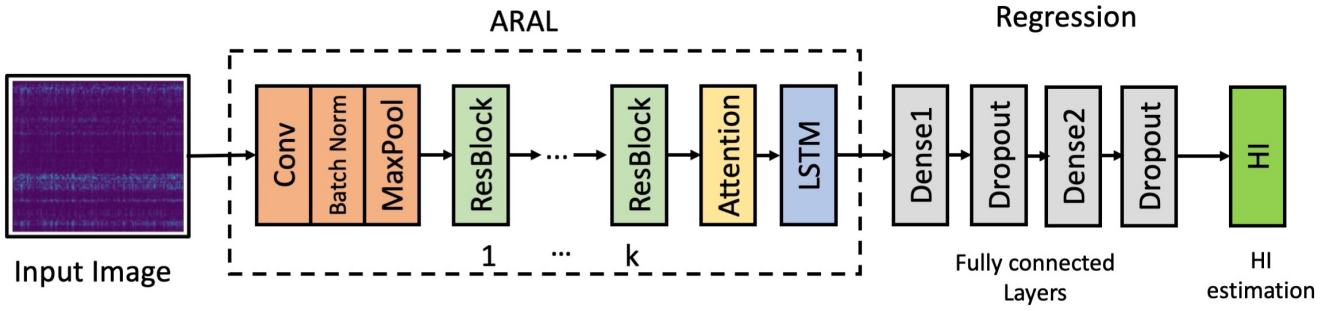


Figure 6. Architecture of the ARAL health assessment model.

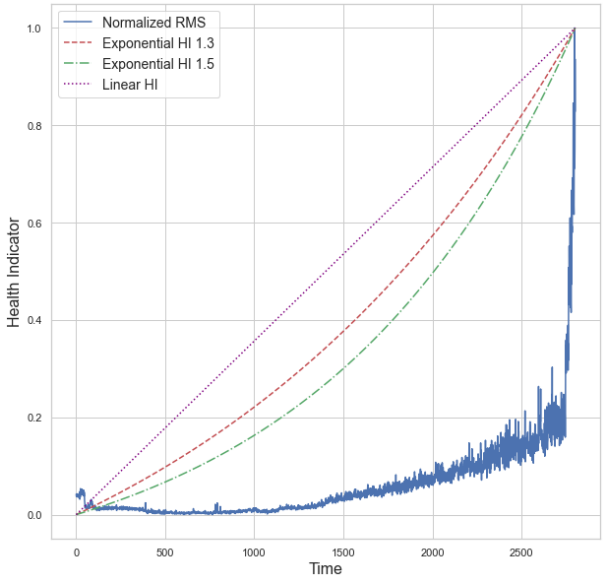


Figure 7. Exponential Labeling vs Linear Labeling.

where MA(t) is the moving average at time t, and k is the window size for the MA.

2.4.2. Detecting the Abnormal Point with 3-Sigma Method

To identify the point at which abnormal degradation starts, we use the 3-Sigma method. This statistical method detects points that deviate significantly from the mean of the smoothed HI curve. The mean μ and standard deviation σ of the smoothed HI data are calculated, and any point $HI(t)$ that satisfies the condition:

$$|HI(t) - \mu| > 3\sigma \tag{26}$$

is considered an abnormal point. The time corresponding to the first detected abnormal point is denoted as T_a. This point marks the beginning of significant degradation, and subsequent analysis focuses on the HI data from this point onward.

2.4.3. GPR model

To predict the future degradation and estimate the RUL, we fit a GPR model to the HI data from the abnormal point T_a to the current observation time T_s. GPR is a non-parametric, Bayesian approach that provides both a prediction and an uncertainty estimate. The GPR model is defined by a mean function $m(t)$ and a covariance function $k(t, t')$:

$$HI(t) \sim \mathcal{GP}(m(t), k(t, t')) \tag{27}$$

A combination of long and short term trend kernels allows the GPR model to capture various patterns and noise characteristics in the HI data.

2.4.4. RUL calculation

To calculate the RUL, we extrapolate the GPR model to predict the HI values until the HI reaches a failure threshold HI_{th}. The predicted time T_f at which the HI curve intersects the threshold is determined, and the RUL is calculated as the difference between T_f and the current observation time T_s:

$$RUL(T_s) = T_f - T_s \tag{28}$$

In practice, the RUL prediction process is detailed in Algorithm 1:

Algorithm 1 RUL Prediction via SSWT-RP-ARAL Method

Input: Vibration signals $s(t)$ for bearings $i = 1$ to N
Output: RUL estimation
Stage 1: Data Preparation
for $i = 1$ to N **do**
  Transform $s_i(t)$ to TF domain using SSWT, resulting in X_i
  Apply RP to X_i to obtain Y_i
end for
Stage 2: HI Construction and Training
for $i = 1$ to N **do**
  Label data with exponential function to create $HealthLabel(t)$
  Initialize $F_i = Y_i$
  for $k = 1$ to K **do**
    Pass F_i through ResNet Block k, updating F_i
  end for
  Apply Temporal Attention to F_i, yielding A_i
  Pass A_i through LSTM layers to obtain $h_{i,t}$
  Feed $h_{i,t}$ into Dense layers with Dropout, producing $d_{i,t}$
  Compute predicted $\widehat{HI}_{i,t}$ using Sigmoid activation on $d_{i,t}$
end for
Train model by minimizing RMSE between $\widehat{HI}_{i,t}$ and $HealthLabel(t)$ using Adam optimizer
Stage 3: RUL Estimation
for $i = 1$ to N **do**
  Smooth HI curve with MA to obtain MA(t)
  Detect first abnormal point $T_{a,i}$ using 3-Sigma method on MA(t)
  Fit GPR model on HI data from $T_{a,i}$ to current time T_s
  Calculate RUL as the time difference when MA(t) reaches failure threshold
end for
Return RUL for each bearing i

3. THE EXPERIMENTAL VALIDATION

To demonstrate the effectiveness and advantages of the proposed method for predicting the RUL of bearings, two renowned public datasets were utilized: the PRONOSTIA dataset (Nectoux et al., 2012) and the XJTU-SY dataset (Wang, Lei, Li, & Li, 2020a). The PRONOSTIA dataset, consisting of run-to-failure vibration signals from bearings, was employed to evaluate the performance of our method against several established techniques. Additionally, the XJTU-SY dataset, which provides comprehensive data on bearing degradation under varying operational conditions, was used to further validate our method's prediction capabilities. Benchmarking against various state-of-the-art techniques highlight-ed the predictive accuracy and robustness of our approach.

3.1. Case study 1: IEEE-PHM-2012-Challenge

3.1.1. Experimental setup and evaluation metrics

The PRONOSTIA dataset, part of the IEEE PHM 2012 Data Challenge (Nectoux et al., 2012), is used to validate the proposed RUL prediction method. This dataset comprises run-to-failure tests on bearings under controlled conditions, simulating real-world operational environments. Vibration signals were collected at a sampling frequency of 25.6 kHz in intervals of 0.1 seconds, recorded every 10 seconds. The dataset includes both horizontal and vertical vibration signals, though only horizontal signals were used in this study due to their higher relevance for bearing degradation (Pichler, Ooijevaar, Hesch, Kastl, & Hammer, 2020).

The dataset consists of 17 bearings tested under three different operating conditions, characterized by varying loads and rotational speeds. For instance, bearings under Condition 1 experienced a load of 4000 N and a speed of 1800 rpm. Table 1 provides a summary of the experimental setup.

Table 1. Summary of the PRONOSTIA dataset.

Condition	Load (N)	Speed (rpm)	Bearings
1	4000	1800	Learning: 1-1, 1-2, 2-1, 2-2, 3-1, 3-2 Testing: 1-3, 1-4, 1-5, 1-6, 1-7
2	4200	1650	Learning: 2-1, 2-2 Testing: 2-3, 2-4, 2-5, 2-6, 2-7
3	5000	1500	Learning: 3-1, 3-2 Testing: 3-3

To assess the effectiveness of our proposed method for RUL prediction, we evaluated its performance using testing bearings and compared the results with other studies. The evaluation metrics included the score metric, accuracy, and percent error. These metrics are those used in the IEEE PHM 2012 Prognostic Challenge.

The percent error for the i-th bearing is calculated as follows:

$$\mathrm{Er}_i = 100 \times \frac{\mathrm{ActRUL}_i - \hat{\mathrm{RUL}}_i}{\mathrm{ActRUL}_i} \tag{29}$$

where ActRUL_i represents the actual RUL, and $\hat{\mathrm{RUL}}_i$ denotes the predicted RUL of the i-th testing bearing.

A distinct scoring function to handle overestimations and underestimations differently is used. The scoring function is defined as:

$$Acc_i = \begin{cases} \exp\left(-\ln(0.5) \times \frac{\mathrm{Er}_i}{5}\right) & \text{if } \mathrm{Er}_i \leq 0 \\ \exp\left(+\ln(0.5) \times \frac{\mathrm{Er}_i}{20}\right) & \text{if } \mathrm{Er}_i > 0 \end{cases} \tag{30}$$

In this function, a perfect prediction (i.e., zero percent error) yields a score of 1. As the percent error increases, the score decreases. Notably, predictions that are later than the actual failure time (i.e., $\mathrm{Er}_i \leq 0$) result in a more significant penalty compared to early predictions.

The overall performance of the RUL prediction is represented by the average score across all testing bearings, calculated as:

$$\text{Score} = \frac{1}{11} \sum_{i=1}^{11} Acc_i \tag{31}$$

3.1.2. Analysis and results

(i) Data preparation and processing

Each sample from the PRONOSTIA dataset, consisting of 2,560 data points, undergoes a transformative process to prepare it for input into the ARAL-HI model. Initially, the data is normalized and converted into a 2560×128 time-frequency image using the SSWT, capturing intricate signal details. To reduce computational complexity, these high-dimensional images are then compressed into 128×128 images through RP, preserving essential features. These processed images are subsequently fed into the ARAL-HI model, which performs regression based on the exponential labeling strategy to accurately assess the degradation of the bearings. Bearing1_3 is taken as an example in Figure 8 to show the transformations of the time signal through his full life cycle.

(ii) Configuration of the ARAL-HI model

The proposed HI model in this study was implemented using Python 3.11 and TensorFlow 2.0, and trained on a high-performance computation server with 2 Tesla V100s-PCI 32Gb GPUs. Extensive hyperparameter tuning and optimization were performed using grid search to ensure optimal performance. The model was trained over 100 epochs with a batch size of 256, using the Adam optimizer with a learning rate of 3.8e-4, striking a balance between convergence speed and stability, and resulting in high accuracy for health assessment tasks. The details of the configuration are presented in Table 2

Table 2. The ARAL model configuration.

Layer No.	Type	Filter Size	Activation	Output Shape
1	InputLayer	-	-	(128, 128, 1)
2	Conv2D	(7, 7)	ReLU	(64, 64, 32)
3	MaxPooling2D	(3, 3)	-	(32, 32, 32)
4	ResBlock	(3, 3)	ReLU	(32, 32, 64)
5	ResBlock	(3, 3)	ReLU	(32, 32, 64)
6	ResBlock	(3, 3)	ReLU	(16, 16, 128)
7	ResBlock	(3, 3)	ReLU	(16, 16, 128)
8	ResBlock	(3, 3)	ReLU	(8, 8, 256)
9	ResBlock	(3, 3)	ReLU	(8, 8, 256)
10	Reshape	-	-	(8, 2048)
11	Attention Mechanism	-	Softmax	(8, 2048)
12	LSTM	256 units	-	(256)
13	Flatten	-	-	256
14	Dense	-	ReLU	2560
15	Dropout	-	-	2560
16	Dense	-	ReLU	512
17	Dropout	-	-	512
18	Dense	-	Sigmoid	1

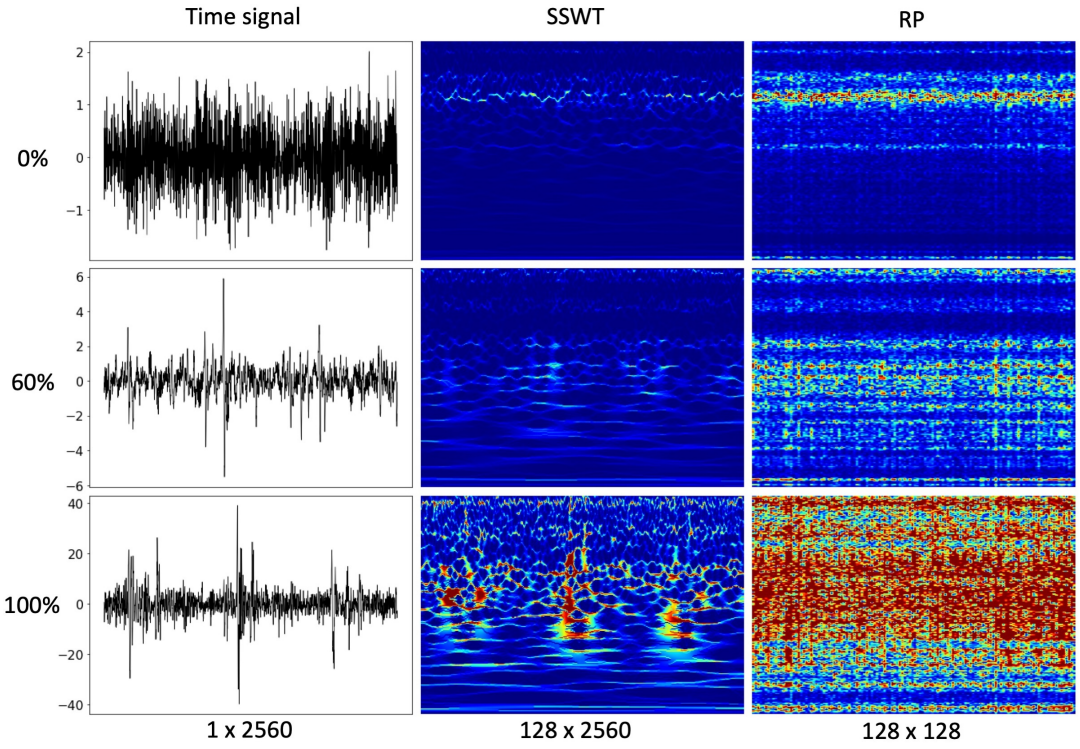


Figure 8. Snapshots of the fault progression of bearing1_3.

(iii) The Influence of the Data labeling

Figure 9 illustrates a comparison of RMSE values for the linear and exponential labeling approaches across the three conditions.

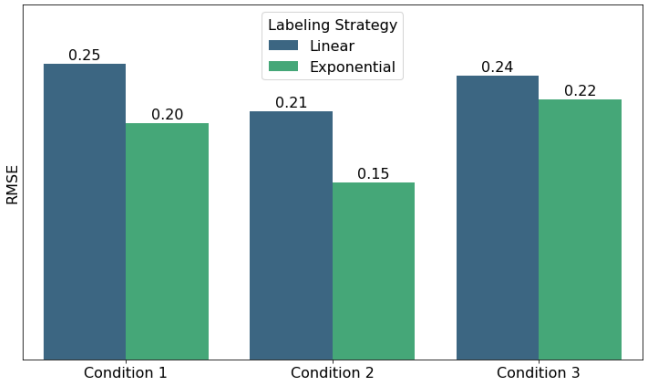


Figure 9. RMSE values for Linear and Exponential Labeling across three conditions.

To assess the impact of different labeling strategies on the prediction performance, we conducted ablation studies focusing on the influence of exponential labeling versus conventional linear labeling of the run-to-failure data. In the conventional linear labeling approach, the HI decreases linearly as the bearing progresses towards failure. Conversely, in the exponential labeling approach, the HI follows an exponential decay pattern, which is designed to provide a more sensitive response to early-stage degradation. We tested these approaches under the three different conditions of the PRONOSTIA dataset. The results indicated that the exponential labeling approach significantly improved the RMSE of the predictions compared to the linear labeling method. Specifically, the exponential model yielded lower RMSE values across all three conditions, facilitating the model's training process and better reflecting the degradation process of the bearing.

(iv) Number of ResBlocks Selection

We investigated the effect of varying the number of ResBlocks on the model's performance. Specifically, we tested configurations with 2, 4, 6, 8, and 10 ResBlocks. The results, measured in terms of RMSE, indicated that the optimal number of ResBlocks was 6, which yielded the lowest RMSE value. This suggests that having 6 ResBlocks balances model complexity and performance, effectively capturing the necessary features without overfitting.

Figure 10 illustrates the RMSE values for different numbers of ResBlocks.

(v) Effects of fitting interval selection

When using the full HI data to train and fit our GPR model, it frequently fails to converge due to an abnormal large step phenomenon present in the data. The step-like nature of the

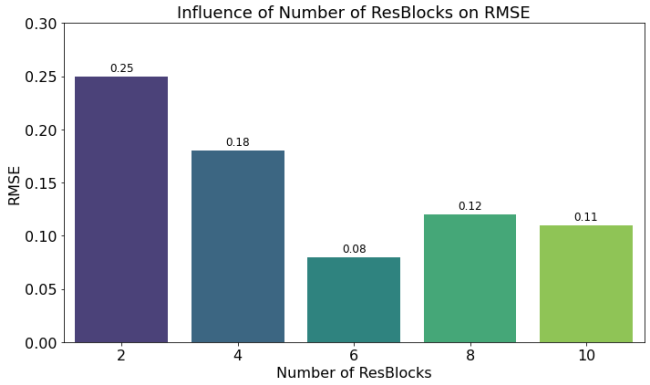


Figure 10. RMSE values for different numbers of ResBlocks.

HI curve leads to significant errors, making accurate RUL estimation challenging and, in many cases, causing the model to never converge to the RUL threshold. However, by incorporating the start fitting point using the 3-sigma method, the model is able to converge and estimate the RUL of the bearing. This improvement is illustrated in Figure 11, which demonstrates the effect using Bearing 1_4 and 2_3 as examples.

(vi) RUL results

Table 3 presents the RUL prediction results, indicating the error and the accuracy for each bearing. These results are used to benchmark the proposed method against other state-of-the-art techniques. Figure 12 illustrates the RUL prediction process for several test bearings.

Table 3. RUL prediction results of ARAL on PRONOSTIA dataset.

Bearing Number	Current Time (s)	Actual RUL (s)	Predicted RUL (s)	Error	Accuracy
1_3	18010	5730	5610	2.0942	0.9300
1_4	11380	339	320	5.6047	0.8235
1_5	23010	1610	1290	19.8758	0.5022
1_6	23010	1460	1510	-3.4247	0.9336
1_7	15010	7570	7430	1.8494	0.9379
2_3	12010	7530	7880	-4.6481	0.9063
2_4	6110	1390	1390	0.0000	1.0000
2_5	20010	3090	2640	14.5631	0.6037
2_6	5710	1290	1290	0.0000	1.0000
2_7	1710	580	600	-3.4483	0.9333
3_3	3510	820	760	7.3171	0.7760

3.1.3. Comparison and Analysis of RUL Prediction Methods

This section presents a comparative analysis between the proposed method and seven leading techniques to highlight its effectiveness and superiority. The results of the experiments, illustrated by the box plots in Figure 13 and detailed in Table 4, provide a comprehensive summary of the performance of the eight RUL prediction methods on the PRONOSTIA dataset.

Among the seven comparative methods:

1. (Liu, Song, Yang, Hao, & Peng, 2018) utilized health indicators derived from the Hilbert–Huang transform (HHT) to predict the RUL of bearings;
2. (Wang, Lei, Li, & Li, 2020b) introduced an approach combining an exponential degradation model with a relevance vector machine (RVM) to estimate the RUL, referred to as EMD-RVM;
3. (She & Jia, 2021) developed a bidirectional gated recurrent unit (Bi-GRU) model, enhanced with bootstrap techniques, for RUL prediction;
4. (L. Li et al., 2023) proposed a multi-technique fusion method for health prognostics, integrating relevance vector machine with a deep separable convolutional gated recurrent network (DSGGRN);
5. (Zhuang, Jia, Cao, & Zhao, 2022) implemented a semi-supervised approach, utilizing a double attention guided assessment method (SDAGA), for estimating the RUL;
6. (Hong, Zhou, Zio, & Hong, 2014) employed wavelet packet-empirical mode decomposition (WP-EMD) for feature extraction, followed by self-organizing mapping (SOM) for assessing performance degradation, named WP-EMD-SOM;
7. (Chen, Jin, Kong, Wang, & Xu, 2023) introduced a Global and Local Information Integrated Network (GLIIN) for predicting the remaining useful life of bearings.

Based on the error (Er) formula, if $Er < 0$, it means the predicted RUL is longer than the actual RUL, which is detrimental to equipment safety. The proposed method has three cases of negative errors, similar to HHT and DSGGRN. In contrast, methods like Bi-GRU and GLIIN exhibit a higher frequency of negative errors, with 4-6 occurrences. This suggests that the proposed method matches the safety assurance levels of the top-performing methods, and maintains a lower rate of unsafe predictions.

The box plot of errors indicates the distribution and spread of the error percentages for each method. The reference line at Error = 0 signifies the ideal scenario where predicted RUL matches the actual RUL perfectly. Methods like HHT, EDM-RVM, Bi-GRU, show a wide spread in error distribution, indicating variability in their predictions. DSGGRN and SDAGA display more compact error distributions, suggesting higher consistency and accuracy. WD-EMD and GLIIN have significant error spreads, demonstrating occasional large deviations from the actual RUL. The proposed method exhibits a relatively narrow error distribution, with most errors close to zero, indicating high accuracy and consistency.

The box plot of accuracies indicates the distribution and spread of the accuracy scores for each method. The reference

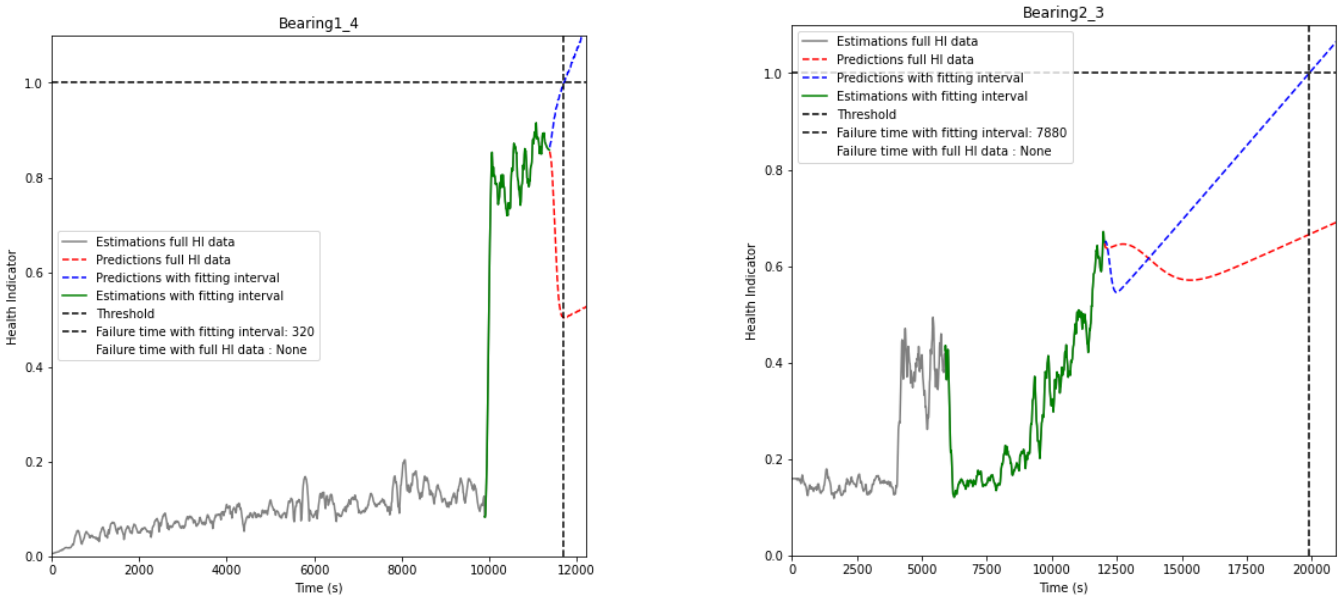


Figure 11. Effect of fitting interval selection

Table 4. State-of-art methods' results comparison with the proposed method.

Dataset	HHT (Er%)	EDM-RVM (Er%)	Bi-GRU (Er%)	DSGGRN (Er%)	SDAGA (Er%)	WP-EMD-SOM (Er%)	GLIIN (Er%)	Proposed (Er%)
Bearing 1_3	2.58	-1.05	-4.36	6.32	3.66	-1.04	0.87	2.09
Bearing 1_4	-9.14	-17.99	70.50	-7.67	2.65	-20.94	-131.91	5.60
Bearing 1_5	-0.99	21.74	6.21	14.90	-4.35	-278.26	-109.29	19.88
Bearing 1_6	6.03	6.16	-4.11	13.42	-0.68	19.18	5.27	-3.42
Bearing 1_7	-0.70	7.79	18.63	-0.96	10.57	-7.13	-10.31	1.85
Bearing 2_3	55.44	43.03	17.40	19.31	0.93	10.49	47.97	-4.65
Bearing 2_4	15.56	1.44	-1.44	4.53	-5.76	51.80	4.24	0.00
Bearing 2_5	49.19	18.77	5.18	-2.69	-2.27	28.80	0.88	14.56
Bearing 2_6	38.53	2.33	16.25	3.02	-5.43	-20.93	2.34	0.00
Bearing 2_7	5.17	-3.45	10.34	2.24	56.90	44.83	-138.40	-3.45
Bearing 3_3	2.56	13.41	6.10	-14.15	2.44	-3.66	0.98	7.32
Score	0.6101	0.5941	0.6190	0.6832	0.6928	0.3550	0.5413	0.7582

line at Accuracy = 0.6 marks the threshold for good prediction performance. HHT, EDM-RVM, and Bi-GRU show wide distributions around this threshold. For instance, the interquartile range (IQR) for HHT spans from approximately 0.4 to 0.75, suggesting variability in performance. The median accuracy for these methods is near or above 0.6, but the spread indicates that their predictions are not consistently reliable. DSGGRN and SDAGA have more compact distributions, with the majority of their accuracy scores above 0.6. The IQR for DSGGRN is approximately 0.55 to 0.85, and for SDAGA, it is about 0.65 to 0.85. These narrow IQRs reflect higher consistency and reliability in their predictions. WD-EMD and GLIIN exhibit significant variability, with fewer accuracy scores consistently exceeding 0.6. The IQR for WD-EMD is wide, from approximately 0.3 to 0.75, indicating less reliable performance. Similarly, GLIIN has a broad IQR ranging from about 0.4 to 0.7, suggesting variability and occasional poor predictions. The proposed method demonstrates a consistently high performance, with a compact distribution well above the 0.6 threshold. The IQR for the proposed method spans from approximately 0.75 to 0.85, and the median accuracy is around 0.8. This indicates superior stability and effectiveness. Over 75% of the cases achieve a score above 0.6, underscoring the method's robustness and reliability.

Among the eight methods, the proposed method has the lowest number of cases where the predicted RUL is larger than

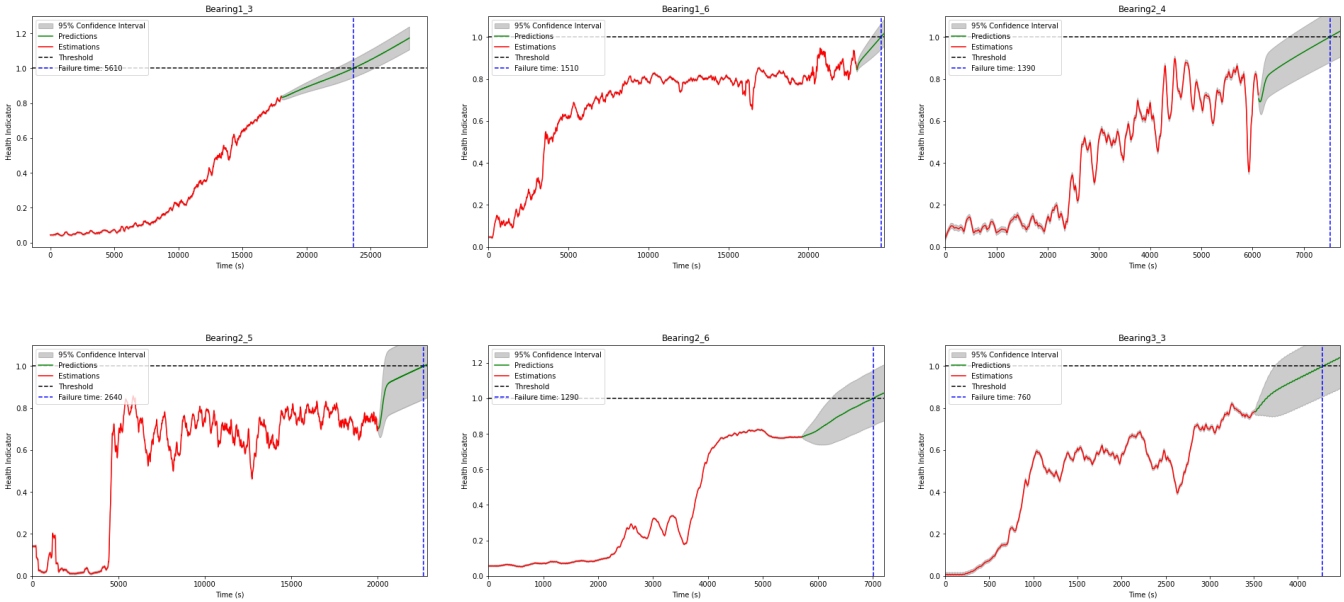


Figure 12. RUL prediction for some bearings

the actual RUL, which is crucial for ensuring equipment safety. The score of the proposed method is 0.7582, higher than the other methods. This improvement in score highlights the superior overall prediction performance of the proposed method. The box plots further illustrate the proposed method's stability and accuracy, with the least variability in errors and the highest consistency in achieving good accuracy scores. The proposed method demonstrates the best balance between low prediction errors and high accuracy, making it the most reliable method among those compared.

However, despite its strong overall performance, it is important to acknowledge certain limitations. A more detailed analysis reveals that for bearings with RULs under 1000 seconds (such as bearings 1_4, 2_7, and 3_3), the SDAGA method outperforms the proposed method in 2 out of the 3 instances. This indicates that in cases of shorter RUL predictions, our method may exhibit slight underperformance compared to SDAGA. While our method provides higher overall accuracy and stability, these observations suggest that further refinement might be needed to handle fast-degrading components more effectively.

This insight is critical for future users to understand the specific contexts in which the proposed method is most effective and where alternative methods may be more appropriate. Despite this limitation, the proposed method remains highly effective for general use, offering the best balance between low prediction errors and high accuracy across a wide range of bearings, as shown by the overall results.

3.2. Case study 2: XJTU-SY datasets

3.2.1. Experimental setup and evaluation metrics

The XJTU-SY rolling bearing dataset, developed by (Wang et al., 2020a), provides comprehensive vibration data for 15 rolling element bearings (model LDK UER204) under accelerated degradation conditions. The testbed consists of an AC induction motor, speed controller, support shaft, hydraulic loading system, and two heavy-duty roller bearings. Vibration signals are collected using two PCB352C33 accelerometers in both vertical and horizontal directions with a sampling frequency of 25.6 kHz, a sampling interval of 1 minute, and a duration of 1.28 seconds per sample. The dataset captures the full life cycle of the bearings across three operating conditions, with five bearings tested under each condition. These conditions involve varying radial forces and rotational speeds, simulating different stress levels on the bearings, as detailed in Table 5. The dataset documents two primary degradation trends—slow and abrupt—providing valuable data for training and testing predictive maintenance models.

For the assessment of predictive performance on this dataset, we use two key metrics: RMSE and Score (Saxena, Goebel, Simon, & Eklund, 2008), defined as:

$$\text{RMSE} = \sqrt{\frac{1}{Q}\sum_{i=1}^{Q} d_i^2} \tag{32}$$

where Q is the number of testing samples, and $d_i = \text{PreRUL}_i - \text{ActRUL}_i$ represents the error between the predicted RUL and the actual RUL for the i-th sample.

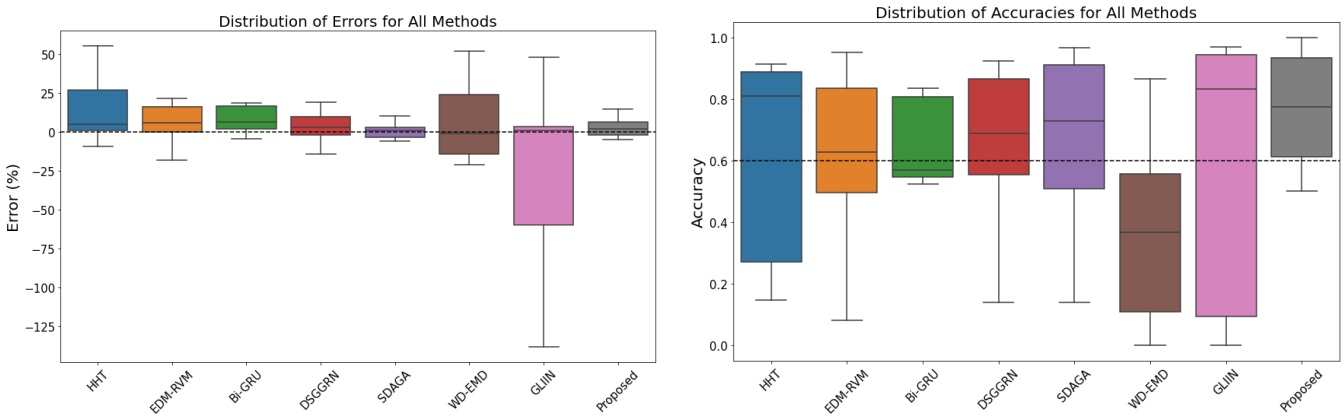


Figure 13. Distribution of RUL prediction errors and accuracies for the compared methods.

$$\text{Score} = \sum_{i=1}^{Q} \begin{cases} \exp\left(-\frac{d_i}{13}\right) - 1 & \text{if } d_i < 0 \\ \exp\left(\frac{d_i}{10}\right) - 1 & \text{if } d_i \geq 0 \end{cases} \tag{33}$$

where d_i is as defined above.

Table 5. Operating conditions and corresponding bearing datasets of XJTU-SY.

Condition	Radial Force (kN)	Shaft Speed (rpm)	Bearings
1	12	2100	Learning: 1_1, 1_2 1_3, 1_4 Testing: 1_5
2	11	2250	Learning: 2_1, 2_2 2_3, 2_4 Testing: 2_5
3	10	2400	Learning: 3_1, 3_2 3_3, 3_4 Testing: 3_5

3.2.2. Analysis and results

(i) Data preparation and processing

For the XJTU-SY dataset, each file contains 32400 data points, from which 2560 data points are randomly selected for simplicity. The selected data is then processed similarly to the PRONOSTIA dataset, undergoing normalization, transformation into 2560×128 TF images via SSWT, and dimensionality reduction to 128×128 images through RP. These images are subsequently fed into the ARAL-HI model for regression using the exponential labeling strategy to assess bearing degradation. The same model configuration for ARAL, presented in Table 2, is used in this experiment for model training and HI construction.

(ii) Experiment results

To ensure the robustness of our RUL predictions and reduce the impact of randomness, each experiment was repeated 20 times for the test bearings. We specifically focused on calculating the score and RMSE from the midpoint to the end of each bearing's lifetime, as predictions during this period are more reliable and meaningful due to the clearer degradation patterns (Singleton, Strangas, & Aviyente, 2015). By averaging the results over these 10 repetitions, we were able to obtain stable and accurate assessments of the model's performance, reflecting its true predictive capability during the critical later stages of bearing life. The RUL prediction results are presented in Table 6.

3.2.3. Comparison and Analysis of RUL Prediction Methods

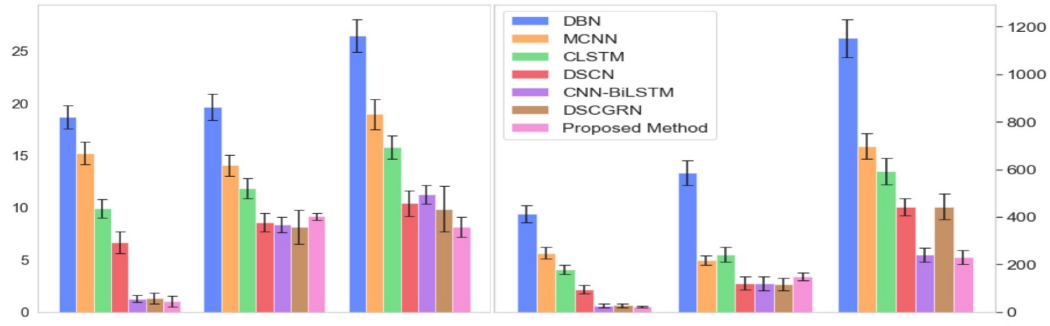
The Figure 14 provides a comparative analysis of various methods for predicting the RUL of bearings 1_5, 2_5, and 3_5, using both RMSE and SCORE metrics. Both metrics assess prediction accuracy, but with different emphases: RMSE maintains consistency across early and late predictions, while SCORE imposes tougher penalties on late predictions, as shown in the histograms. The methods compared include DBN (Deutsch & He, 2018), MCNN (Zhu, Chen, & Peng, 2019b), CLSTM (Zhao, Yan, Wang, & Mao, 2017), DSCN (Wang, Lei, Li, & Yan, 2019), CNN-BiLSTM (Cheng, Hu, Wu, Zhu, & Shao, 2021), DSCGRN (L. Li et al., 2023), and the Proposed Method.

For Bearing 1_5, the Proposed Method exhibits the lowest RMSE (1.02 ± 0.50) and SCORE (22.23 ± 3.10), indicating superior prediction accuracy. In contrast, for Bearing 2_5, DSCGRN outperforms the Proposed Method, achieving a lower RMSE (8.13 ± 1.64) and SCORE (116.29 ± 25.45). This suggests that DSCGRN is better at managing late prediction errors, which are more heavily penalized by the SCORE metric.

The results for Bearing 3_5 further emphasize the strength of the Proposed Method, which again achieves the lowest

Table 6. RUL prediction results for bearings 1_5, 2_5, and 3_5.

Bearing Number	Full Lifetime (min)	Starting Time (min)	RMSE (mean ± std)	Score (mean ± std)
1_5	52	26	1.02 ± 0.50	22.23 ± 3.10
2_5	339	170	9.15 ± 0.34	149.02 ± 15.50
3_5	114	57	8.14 ± 0.96	230.97 ± 30.20



Method	RMSE (mean ± std)			Score (mean ± std)		
	Bearing 1_5	Bearing 2_5	Bearing 3_5	Bearing 1_5	Bearing 2_5	Bearing 3_5
DBN	18.68 ± 1.13	19.62 ± 1.28	26.48 ± 1.56	413.97 ± 36.26	585.02 ± 51.27	1151.95 ± 79.70
MCNN	15.22 ± 1.07	14.04 ± 0.97	18.93 ± 1.43	247.76 ± 23.81	217.38 ± 20.83	696.99 ± 53.25
CLSTM	9.89 ± 0.89	11.84 ± 0.94	15.78 ± 1.12	178.94 ± 20.37	241.06 ± 31.52	592.15 ± 54.56
DSCN	6.67 ± 1.04	8.58 ± 0.92	10.41 ± 1.21	94.91 ± 17.05	121.03 ± 26.54	441.20 ± 36.10
CNN-BiLSTM	1.27 ± 0.32	8.37 ± 0.76	11.27 ± 0.88	26.75 ± 7.25	119.74 ± 28.69	241.06 ± 30.16
DSCGRN	1.33 ± 0.52	8.13 ± 1.64	9.87 ± 2.20	28.26 ± 8.14	116.29 ± 25.45	443.83 ± 55.23
Proposed Method	1.02 ± 0.50	9.15 ± 0.34	8.14 ± 0.96	22.23 ± 3.10	149.02 ± 15.50	230.97 ± 30.20

Figure 14. Performance comparison of different prognosis methods on the XJTU-SY dataset.

RMSE (8.14 ± 0.96) and SCORE (230.97 ± 30.20). Overall, the Proposed Method demonstrates robust performance across different bearings, excelling particularly in minimizing prediction errors. However, the slightly better performance of DSCGRN in Bearing 2_5 highlights that certain methods may be more effective under specific conditions, particularly in avoiding late RUL predictions.

4. Conclusion & Future Work

This paper introduced ARAL, an advanced method for predicting the RUL of rolling bearings by integrating SSWT, RP, ResNet, a Temporal Attention Mechanism, and LSTM models. The method effectively captures critical time-frequency features, addressing challenges like the vanishing gradient problem and enhancing prediction accuracy. Experimental results on the PRONOSTIA and XJTU-SY datasets demonstrate that ARAL outperforms existing techniques in terms of accuracy and reliability, achieving lower RMSE and higher prediction scores across various test cases. The use of exponential data labeling and GPR further improves RUL predictions, particularly in accurately capturing degradation trends and reducing prediction errors, making ARAL a robust tool for industrial prognostics and health management.

While ARAL shows strong overall performance, there are some limitations related to its real-time application. The primary challenge lies in the execution time required during the GPR fitting phase, particularly when processing long HI curves. This step can significantly slow down the prediction process, which could impact ARAL's usability in time-sensitive industrial environments. Optimizing the GPR process through model simplification or faster approximation techniques will be a key focus for future work to ensure that ARAL meets real-time performance requirements without compromising accuracy.

Additionally, ARAL demonstrates slight underperformance in predicting rapidly degrading components, such as bearings with shorter RULs. Addressing this limitation through further model refinements will also be explored. Future work will involve extending ARAL to other machinery types, integrating additional data sources, and improving the method's robustness and computational efficiency for broader industrial applicability.

References

Achlioptas, D. (2003). Database-friendly random projections: Johnson-lindenstrauss with binary coins. *Jour-*

nal of Computer and System Sciences, *66*(4), 671-687.

Chen, Z., Jin, X., Kong, Z., Wang, F., & Xu, Z. (2023). Global and local information integrated network for remaining useful life prediction. *Engineering Applications of Artificial Intelligence*, *126*, 106956. doi: https://doi.org/10.1016/j.engappai.2023.106956

Cheng, Y., Hu, K., Wu, J., Zhu, H., & Shao, X. (2021). A convolutional neural network based degradation indicator construction and health prognosis using bidirectional long short-term memory network for rolling bearings. *Advanced Engineering Informatics*, *48*, 101247.

Daubechies, I., Lu, J., & Wu, H.-T. (2011). Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Applied and Computational Harmonic Analysis*, *30*(2), 243-261.

Deng, Y., Guo, C., Zhang, Z., Zou, L., Liu, X., & Lin, S. (2023). An attention-based method for remaining useful life prediction of rotating machinery. *Applied Sciences*, *13*(4).

Deutsch, J., & He, D. (2018). Using deep learning-based approach to predict remaining useful life of rotating components. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *48*(1), 11-20. doi: 10.1109/TSMC.2017.2697842

Ding, Y., Jia, M., Cao, Y., Ding, P., Zhao, X., & Lee, C.-G. (2023). Domain generalization via adversarial out-domain augmentation for remaining useful life prediction of bearings under unseen conditions. *Knowledge-Based Systems*, *261*, 110199. doi: https://doi.org/10.1016/j.knosys.2022.110199

Hong, S., Zhou, Z., Zio, E., & Hong, K. (2014). Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method. *Digital Signal Processing*, *27*, 159-166. doi: https://doi.org/10.1016/j.dsp.2013.12.010

Johnson, W., & Lindenstrauss, J. (1984, 01). Extensions of lipschitz maps into a hilbert space. *Contemporary Mathematics*, *26*, 189-206. doi: 10.1090/conm/026/737400

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to rul prediction. *Mechanical Systems and Signal Processing*, *104*, 799-834. doi: https://doi.org/10.1016/j.ymssp.2017.11.016

Li, L., Xu, J., & Li, J. (2023). Estimating remaining useful life of rotating machinery using relevance vector machine and deep learning network. *Engineering Failure Analysis*, *146*, 107125. doi: https://doi.org/10.1016/j.engfailanal.2023.107125

Li, X., Yu, S., Lei, Y., Li, N., & Yang, B. (2024). Intelligent machinery fault diagnosis with event-based camera. *IEEE Transactions on Industrial Informatics*, *20*(1), 380-389. doi: 10.1109/TII.2023.3262854

Liu, X., Song, P., Yang, C., Hao, C., & Peng, W. (2018, 07). Prognostics and health management of bearings based on logarithmic linear recursive least-squares and recursive maximum likelihood estimation. *IEEE Transactions on Industrial Electronics*, *PP*, 1-1. doi: 10.1109/TIE.2017.2733469

Londhe, N. D., Arakere, N. K., & Subhash, G. (2017, 09). Extended Hertz Theory of Contact Mechanics for Case-Hardened Steels With Implications for Bearing Fatigue Life. *Journal of Tribology*, *140*(2), 021401. doi: 10.1115/1.4037359

Najdi, B., Benbrahim, M., & Kabbaj, M. N. (2024, March). Bearing fault diagnosis under varying work conditions based on synchrosqueezing transform, random projection, and convolutional neural networks. *Int. J. Progn. Health Manag.*, *15*(1). doi: 10.36001/IJPHM.2024.v15i1.3799

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., & Varnier, C. (2012, 06). Pronostia: An experimental platform for bearings accelerated degradation tests. In (p. 1-8).

Pichler, K., Ooijevaar, T., Hesch, C., Kastl, C., & Hammer, F. (2020, 05). Data-driven vibration-based bearing fault diagnosis using non-steady-state training data. *Journal of Sensors and Sensor Systems*, *9*, 143-155.

Qin, Y., Xiang, S., Chai, Y., & Chen, H. (2020). Macroscopic–microscopic attention in lstm networks based on fusion features for gear remaining life prediction. *IEEE Transactions on Industrial Electronics*, *67*(12), 10865-10875. doi: 10.1109/TIE.2019.2959492

Que, Z., & Xu, Z. (2019). A data-driven health prognostics approach for steam turbines based on xgboost and dtw. *IEEE Access*, *7*, 93131-93138. doi: 10.1109/ACCESS.2019.2927488

Sadowsky, J. S. (1994). The continuous wavelet transform: a tool for signal investigation and understanding..

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management* (p. 1-9). doi: 10.1109/PHM.2008.4711414

Shah, A., Kadam, E., Shah, H., & Shinde, S. (2016, 04). Deep residual networks with exponential linear unit.

She, D., & Jia, M. (2021). A bigru method for remaining useful life prediction of machinery. *Measurement*, *167*, 108277. doi: https://doi.org/10.1016/j.measurement.2020.108277

Singleton, R. K., Strangas, E. G., & Aviyente, S. (2015). Extended kalman filtering for remaining-useful-life estimation of bearings. *IEEE Transactions on Industrial Electronics*, *62*(3), 1781 – 1790.

Suh, S., Lukowicz, P., & Lee, Y. O. (2022). Generalized multiscale feature extraction for remaining useful life prediction of bearings with generative adversarial networks. *Knowledge-Based Systems*, *237*, 107866. doi:

https://doi.org/10.1016/j.knosys.2021.107866

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2023). *Attention is all you need.* Retrieved from https://arxiv.org/abs/1706.03762

Wang, B., Lei, Y., Li, N., & Li, N. (2020a). A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*, *69*(1), 401-412. doi: 10.1109/TR.2018.2882682

Wang, B., Lei, Y., Li, N., & Li, N. (2020b). A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*, *69*(1), 401-412. doi: 10.1109/TR.2018.2882682

Wang, B., Lei, Y., Li, N., & Yan, T. (2019). Deep separable convolutional network for remaining useful life prediction of machinery. *Mechanical Systems and Signal Processing*, *134*, 106330.

Yan, H., Qin, Y., Xiang, S., Wang, Y., & Chen, H. (2020). Long-term gear life prediction based on ordered neurons lstm neural networks. *Measurement*, *165*, 108205. doi: https://doi.org/10.1016/j.measurement.2020.108205

Yoo, Y., & Baek, J.-G. (2018). A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network. *Applied Sciences*, *8*(7). doi: 10.3390/app8071102

Zhang, Y., Tang, B., Han, Y., & Deng, L. (2017, feb). Bearing performance degradation assessment based on time-frequency code features and som network. *Measurement Science and Technology*, *28*(4), 045601. doi: 10.1088/1361-6501/aa56c9

Zhao, R., Yan, R., Wang, J., & Mao, K. (2017). Learning to monitor machine health with convolutional bi-directional lstm networks. *Sensors*, *17*(2).

Zhu, J., Chen, N., & Peng, W. (2019a). Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Transactions on Industrial Electronics*, *66*(4), 3208-3216. doi: 10.1109/TIE.2018.2844856

Zhu, J., Chen, N., & Peng, W. (2019b). Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Transactions on Industrial Electronics*, *66*(4), 3208-3216. doi: 10.1109/TIE.2018.2844856

Zhuang, J., Jia, M., Cao, Y., & Zhao, X. (2022). Semi-supervised double attention guided assessment approach for remaining useful life of rotating machinery. *Reliability Engineering & System Safety*, *226*, 108685. doi: https://doi.org/10.1016/j.ress.2022.108685

BIOGRAPHIES

Boubker Najdi's academic journey began in 2018 when he earned a B.Eng. degree in Mechatronics and Embedded Systems from the Faculty of Sciences, Fez. He continued his educational pursuits by obtaining an M.Eng. in Smart Industry from the University of Sidi Mohamed Ben Abdellah, Fez, Morocco, in 2020.

Currently, he is a Data Scientist and a Ph.D. student at the Faculty of Sciences, Fez. His areas of research expertise encompass a wide range of topics, including robotics, artificial intelligence, and predictive maintenance.

For inquiries or contact, you can reach him at the following email address: aboubakr.najdi@usmba.ac.ma.

Mohammed Benbrahim received the B.Eng. degree in electromechanical engineering from the Higher National School of Mines in 1997 and the M.Sc. and Ph.D. degrees in automatic and industrial informatics from the Mohammadia School of Engineers in 2000 and 2007, respectively.

Currently, he is a Full Professor at the Department of Physics, the director of the LIMAS laboratory and the program coordinator for the Master's program in Smart Industry at the Faculty of Sciences, Sidi Mohamed Ben Abdellah University. His research interests include robotics, automatic control, intelligent systems, predictive maintenance, modeling, and optimization. He can be contacted at email: mohammed.benbrahim@usmba.ac.ma.

Mohammed Nabil Kabbaj is a Full Professor at the Faculty of Sciences, University of Fez, where he is the Vice Dean in charge of Research and the Program Coordinator of the Mechatronics and Embedded Systems Bachelor's program. His research interests include control engineering, fault detection, and diagnosis of complex systems. Before joining the University of Fez, he received a Ph.D. degree from the University of Perpignan in 2004 and was a postdoctoral researcher at LAAS-CNRS in Toulouse. He can be contacted at email: n.kabbaj@usmba.ac.ma.