WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 2, Number 3, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v2n3.17
Journal homepage: https://wepub.org/index.php/IJCSIT/index

# A Hybrid ST-Bilstm Based Prediction Method for Estimating the Life of Ball Mill Rolling Bearings

Jinjiang Zhao*, [1, 2, 3] and Guanwen Zhang[1, 2]

[1]College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou Gansu 730050, China

[2]Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou Gansu 730050, China

[3]Institute of Automation, Gansu Academy of Sciences, Lanzhou Gansu 730000, China

*Corresponding Author: 1769337516@163.com

## ABSTRACT

Aiming at the problems of low prediction accuracy and slow convergence speed of existing ball mill rolling bearing Remaining Useful Life (RUL) prediction models, an RUL prediction method using Temporal Convolutional Network (TCN) and Bidirectional Long and Short-term Memory (BiLSTM) is proposed. Firstly, the Savitzky-Golay (SG) filter is used to remove the interference of random noise in the original vibration signals; then the short-term features of the sequences are extracted using TCN and capture the long-term correlations in the data with the help of BiLSTM; finally, the attention mechanisms assign different weights to features in order to achieve the goal of being able to more accurately extract information about the health state of the device. Theoretical analysis and data validation demonstrate that the method improves the accuracy and adaptability of model series predictions, and the IEEE PHM 2012 dataset is used to verify the effectiveness of the proposed method.

## KEYWORDS

Rolling bearings; Temporal convolutional networks; Bidirectional long- and short-term memory networks; Attentional mechanisms; Remaining life prediction

## 1. INTRODUCTION

With the continuous progress of industrialization, the importance of the metallurgical industry is becoming more and more prominent, and it plays an increasingly important role in the industrial and economic development of the country [1]. In the metallurgical beneficiation process, ball mills are commonly used mechanical equipment, and the use of ball mills for material grinding is a crucial part of the whole smelting process. However, due to the long-term operation of metallurgical equipment in harsh industrial environments, the longer the use of ball mills, the more difficult it is to avoid wear, ageing, and failure of rolling bearings [2]. This will not only affect economic production efficiency, but in serious cases, it may even cause accidents, resulting in personnel and property losses. Therefore, it is a meaningful and extremely important work to predict the RUL of ball mill rolling bearings [3].

Currently, data-driven methods are primarily focused on machine learning and deep learning [4]. It is often necessary to involve relevant expertise in the feature engineering process when working with unstructured data, which limits the generalizability of machine learning [5]. With deep learning, many domains can benefit from reduced feature engineering requirements and more efficient solutions. With sufficient data, deep learning prediction techniques are currently frequently used in the field of

RUL prediction as they are more productive than conventional machine learning techniques [6]. In the field of RUL prediction, Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and their variation networks have found widespread use. For example, Ma et al. [7] can automatically learn time-frequency domain spatial information from vibration signals by introducing convolutional operations into the state transition process of RNNs, which contributes to the performance of rotating machinery life prediction. However, RNNs may experience gradient issues during model training; thus, researchers have extensively improved the RNN structure to address the issue and have produced a large number of models. The most well-known of them is the LSTM suggested by Hochreiter and Schmidhuber [8]. Yousuf et al. [9] proposed a regression model based on LSTM using the most basic electrical characteristics of the device to predict the RUL of a ring oscillator (RO) circuit. However, LSTM is a unidirectional RNN structure that can only process current time information based on past time information. In order to analyze past and future information simultaneously, Bidirectional-LSTM (Bi-LSTM) employs a bi-network structure to convey implicit layer features, which shows stronger performance in bearing RUL tests. For example, Zhao et al. [10] proposed a RUL prediction model based on BiLSTM with an attention mechanism, where the device state feature information is automatically extracted by BiLSTM, and then the features are given various weights using the attention method so that the health state information of the device can be extracted more accurately. The complicated chain structure of LSTM and Bi-LSTM results in longer training times and worse sensitivity to long time series, despite the fact that they can learn the dependencies of the data before and after time series data. In order to create prediction models that are more effective, some researchers have implemented bearing RUL prediction using CNN. Li et al. [11] used CNN to construct a prediction model and reconstructed LSTM for memorizing long-term degradation data of lithium batteries by inputting time-series features and non-time-series features of CNN spreading in order to increase RUL forecast accuracy. Despite the fact that CNNs are quite good at extracting locally relevant features, they lack sensitivity to temporal information and are prone to ignoring the back-and-forth correlation of temporal information in the data. Bai et al. [12] presented a TCN that solves this issue by fusing causal convolution with null convolution, which has achieved success in time series modeling applications and has similar temporal feature extraction capabilities to RNN. For mechanical prediction, Wang et al. [13] proposed a novel TCN that combines attention and soft thresholding techniques. This enables the network training process to concentrate on features that are more essential to RUL prediction by giving each feature map a specific threshold. However, a network structure based on unidirectional feature transfer is used by the majority of the current TCNs, which makes the TCNs only able to extract past temporal feature information and unable to obtain future temporal feature information.

In conclusion, even though deep learning techniques have produced positive outcomes in the area of bearing RUL prediction, the following problems still need to be resolved:

TCN cannot use knowledge about future time features and can only assess the state of the current bearing operation based on past time feature information;

Bi-LSTM can extract time-dependent features of vibration data, but it cannot extract long time series features of vibration data.

To address the issues mentioned above, this paper proposes a RUL prediction method for bearings based on ST-BiLSTM. The method first uses an SG filter for the noise reduction process. Compared with other noise reduction methods, the SG filter is used to ensure that the shape and width of the data are preserved, while the noise is filtered to increase the accuracy of the data prediction. Then, TCN is used to learn the features of its sequences in the temporal dimension, and BiLSTM is introduced to further learn the before and after conditions to extract more information from the sequences for prediction. Finally, the attention mechanism is introduced to highlight the high-impact features of the vector weights of the outputs at different moments in the TCN-BiLSTM network, improve the accuracy of RUL prediction, and speed up the convergence of the model.

## 2. THEORETICAL BACKGROUND

### 2.1. SG Filter

The SG filtering algorithm based on the principle of least squares is a polynomial smoothing algorithm that ensures that when filtering the noise, the data's form and width remain constant. To help obtain a more accurate prediction of the time series data, its smoothing and denoising of the original sequence are used, and a window size of represents the subsequence of G as the following equation:

$$\{g_{s-m},\cdots,g_s,\cdots,g_{s+m}\}, s\in[m+1,T-m] \tag{1}$$

As used to fit the data points in the window, the R-order polynomial $p(k)$ is defined as:

$$p(k)=\sum_{v=0}^{R} a_v k^v, k\in[-m,m] \tag{2}$$

where $a_v$ represents the SG filter's $v^{th}$ coefficient. The least squares method is then applied to reduce the error:

$$\varepsilon=\sum_{i=-m}^{m}(p(k)-x_{s+i})^2 \tag{3}$$

By calculating the $a_0$, we can then get the $p(0)$ that fits the window's centre point $x_s$ the best. Each point in $G$ becomes the window's centre point by translating the window, and after the smoothing procedure, the sequence $\overline{G}$ is obtained.

### 2.2. TCN

As depicted in Figure 1, the literature [12] established a TCN structure for each task. TCN is a time series improvement model based on one-dimensional CNN, which mainly consists of three parts: causal convolution, dilation convolution, and residual block. The process of causal convolution guarantees that the value at a given time t in the upper layer is solely influenced by the value at the preceding time t in the lower layer. Nevertheless, when expanding the perceptual realm of input through the incorporation of stacked layers or the inclusion of convolutional kernels, the computational burden of the network would inevitably escalate. Hence, the technique of incorporating empty values into the convolution kernel is suggested as a means to augment the receptive field, sometimes referred to as dilation convolution. The use of dilation convolution allows the upper nodes to have a larger receptive field in order to facilitate the introduction of additional historical information and mitigate excessive computational processes, it is imperative to implement certain measures. Residual blocks, on the other hand, the utilization of residual connections within network layers facilitates the process of feature extraction in sequences, while also mitigating the issues of gradient vanishing or explosion. In the context of a one-dimensional series of inputs $g\in R^n$ and a convolution kernel $f:\{0,\cdots,k-1\}\to R$, let us consider the null convolution operation F applied to an element $s$ of the sequence. This operation is defined as follows:

$$F(s)=(g*_d f)(s)=\sum_{i=0}^{k-1} f(i)\cdot g_{s-d\cdot i} \tag{4}$$

Where $*_d$ represents the convolution operator and $f(i)$ denotes the convolution kernel, $i=1,2,\cdots k-1$.

Currently, the data will undergo scanning by a one-dimensional convolutional kernel at each layer. The sequence data that has undergone cleaning is transferred through the network layer, resulting in the processed data being produced as the network's output. The duration of learning knowledge increases as the depth of the network increases. The TCN exhibits comparable predictive capabilities to the RNN in the context of time series forecasting. The utilisation of TCN, akin to CNN, necessitates the execution of both feed-forward and feedback computations. The process of feedforward computation involves extracting relevant feature information from the input data. This information is then utilized in feedback computation to correct faults within the network, resulting in an output that is more likely to closely approximate the true value. One distinction lies in the fact that TCNs are equipped with a sole one-dimensional convolutional kernel and are designed to handle one-dimensional input data.

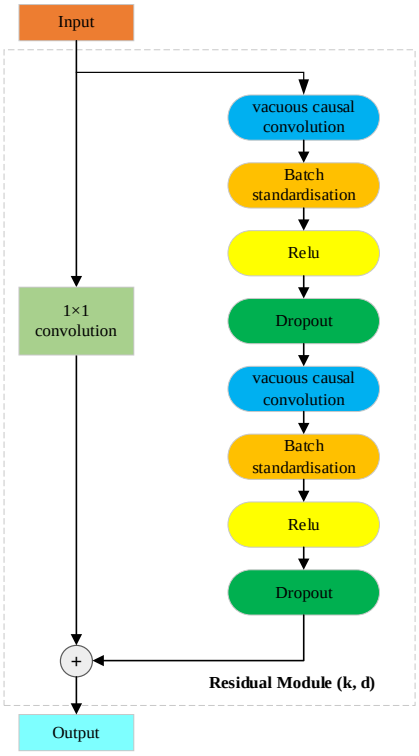



**Figure 1:** Schematic diagram of TCN structure

## 2.3. BiLSTM

In 1997, Schmidhuber et al. suggested Long Short-Term Memory (LSTM) as a solution to the problem of gradient explosion or cancellation that occurs in traditional recurrent neural networks [8], which realizes the selective transmission of information through the gating mechanism, and Figure 2 illustrates its interior architecture. According to the following calculating formula:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\hat{C}_t = \sigma(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{7}$$

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \tag{8}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = o_t * \tanh(C_t) \tag{10}$$

where $i_t$, $f_t$, and $o_t$ are the input threshold, the forgetting threshold, and the output threshold, respectively. The activation function $Sigmoid$ is denoted as $\sigma$. tanh is a function that activates hyperbolic tangents. The internal state of the cell at time $t-1$ is represented by $C_{t-1}$. The state of the cell at time $t$ is represented by candidate $\hat{C}_t$. The variable $C_t$ represents the internal state of the cell at a specific time point, denoted as $t$. $h_t$ is the external state of the cell at moment $t$. The weight matrices of the state unit, input gate, forgetting gate, and output gate at time $t$ are denoted as $W_c$, $W_i$, $W_f$, and $W_o$, respectively. The state unit, input gate, oblivion gate, and output gate are associated with four bias vectors denoted as $b_c$, $b_i$, $b_o$, and $b_f$. "$*$" denotes the vector product.

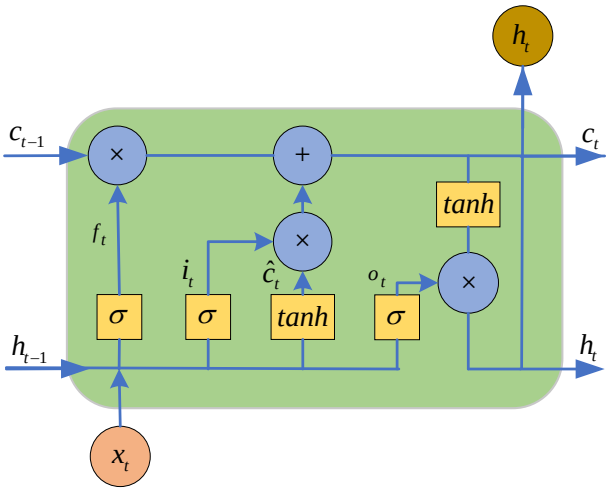



**Figure 2:** LSTM cell structure

Many studies have demonstrated that LSTMs can effectively deal with temporal relationships between inputs and outputs and learn data correlations for time series. However, LSTM can only use information from previous times to predict the state at the current time and cannot use information from future times. To address this problem, BiLSTM is constructed using two LSTMs in opposite directions, thus enabling the processing of bidirectional temporal information. In BiLSTM, forward LSTM computes the hidden layer information along the forward order of the temporal data and backward LSTM computes the hidden layer information along the reverse order of the temporal data, and then fuses the hidden layer information of the two LSTMs and obtains the output information of Bi-LSTM. BiLSTM can obtain bi-directional information from historical and future time data and has a more powerful feature learning performance than LSTM. The structure of the BiLSTM module is shown in Figure 3.

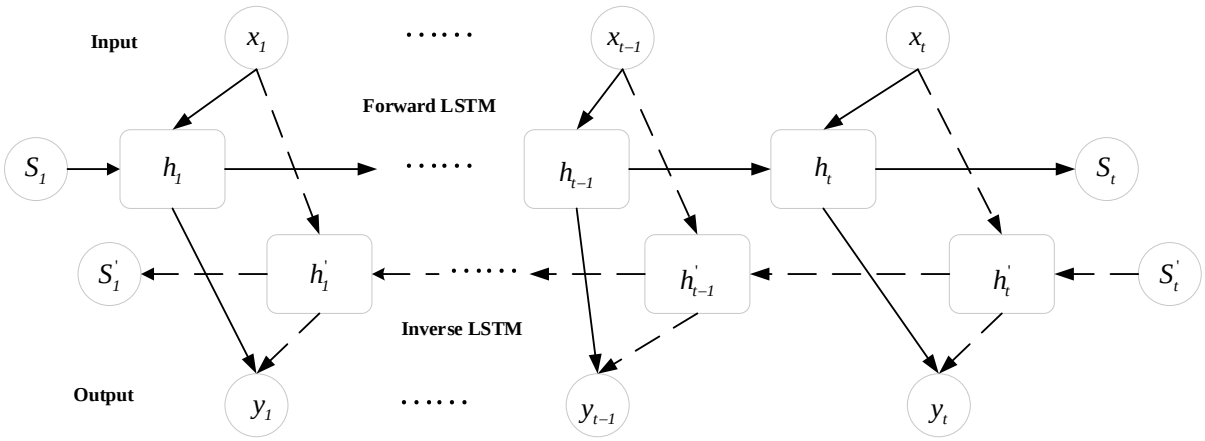



**Figure 3:** BiLSTM structure

### 2.4. Attention

The attention mechanism focuses on the distribution of input weights, as illustrated in Figure 4. The attention value is calculated for a task-specific query vector $q$ using an attention distribution that incorporates keywords and their associated importance values for each input.

For $N$ sets of input messages $G = [g_1, \cdots, g_n] \in R^{D \times N}$, let $Key = Value = G$, then the distribution of attention at the N-th position is

$$a_n = soft\max(s(key_n, q)) = soft\max(s(G_n, q)) = \frac{\exp(s(G_n, q))}{\sum_{j=1}^{N} \exp(s(G_n, q))} \tag{11}$$

where $s(G_n, q)$ is the attention rating function. This work employs $g^T q / \sqrt{N}$ to calculate the attention distribution for each input vector inside a specific task.

The information selection process is employed to encode the input information $G$, which is computed in the following manner:

$$att(q, G) = \sum_{i=1}^{N} \alpha_i G_i \tag{12}$$

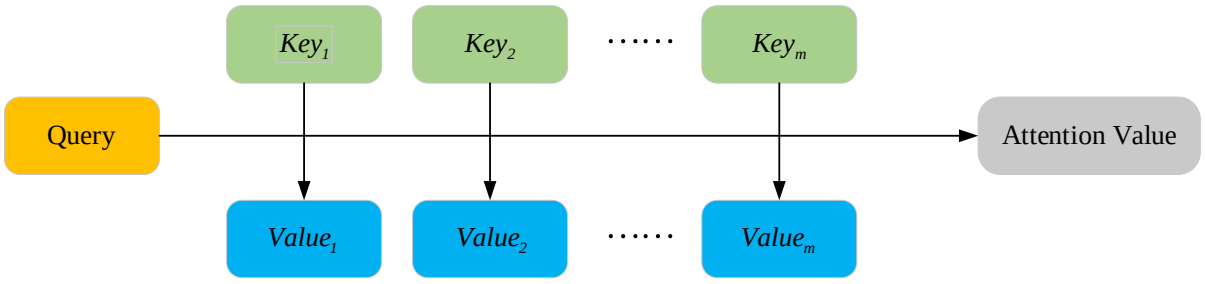



**Figure 4:** Attention mechanism schematic

## 3. ST-BILSTM-BASED REMAINING LIFE PREDICTION METHOD

The ST-BiLSTM-based network structure designed in this paper is shown in Figure 5. The advantage of TCN in dealing with time series problems is applied to the lifetime prediction to learn the sequence features; then the BiLSTM model is used to learn the before and after conditions to extract more information in the sequence for prediction; after that, the attention mechanism is introduced to dynamically adjust the weights of the attributes; finally, the output of the RUL is produced using the fully connected layer.

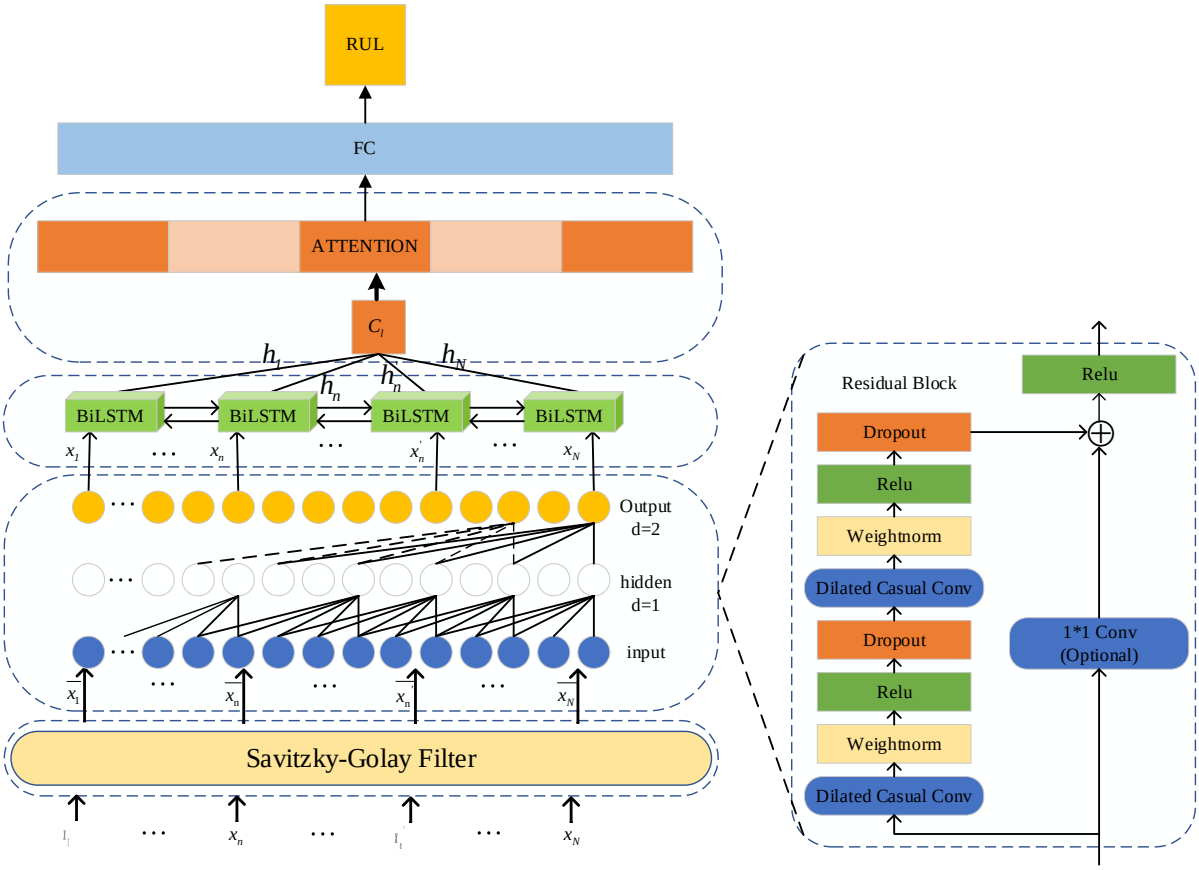



**Figure 5:** ST-BiLSTM-based prediction model graphs

The main steps of ST-BiLSTM-based prediction are shown in Figure 6.

Firstly, experimental preparation is carried out, and the original time domain vibration signal is noise reduced and smoothed using the SG filter, and the smooth time domain vibration signal after noise reduction is obtained (Figure 7 displays the SG filter's effect diagram); then the time domain vibration signal is normalized after noise reduction; immediately after that, the training set and test set are created from the obtained normalized data, and the remaining lifetime is normalized to between (0,1) as the network training label.

Secondly, model training is carried out by inputting TCN-BiLSTM-ATTENTION to the divided training set data, and finally, the output is performed using fully connected layers, and the trained model is saved. Specifically, TCN incorporates a causal convolution module and a dilation convolution module, and its use of causal convolution can successfully guarantee that the convolution process is carried out in the sequence specified by the historical timeline, thus guaranteeing the temporal order, while TCN can have a bigger perceptual field with fewer layers thanks to the use of null convolution, thus acquiring more historical data over time and guaranteeing the accuracy of the data. To achieve sequence feature extraction, the residual link between each network layer is used while avoiding the phenomenon of gradient disappearance or explosion; Two LSTM layers are stacked in opposite directions to create BiLSTM, so the input will be provided to both the forward and reverse layers of the LSTM, and the output will be determined by the two LSTM layers together, and its bidirectional cyclic structure increases the transfer of information from the past to the future, which enables BiLSTM to better examine the data's temporal features; the attention mechanism highlights the features that have a greater impact on the remaining life, increases the RUL prediction's accuracy and accelerates the convergence speed of the model.

Finally, the final RUL prediction is obtained by downscaling the fully connected layer.

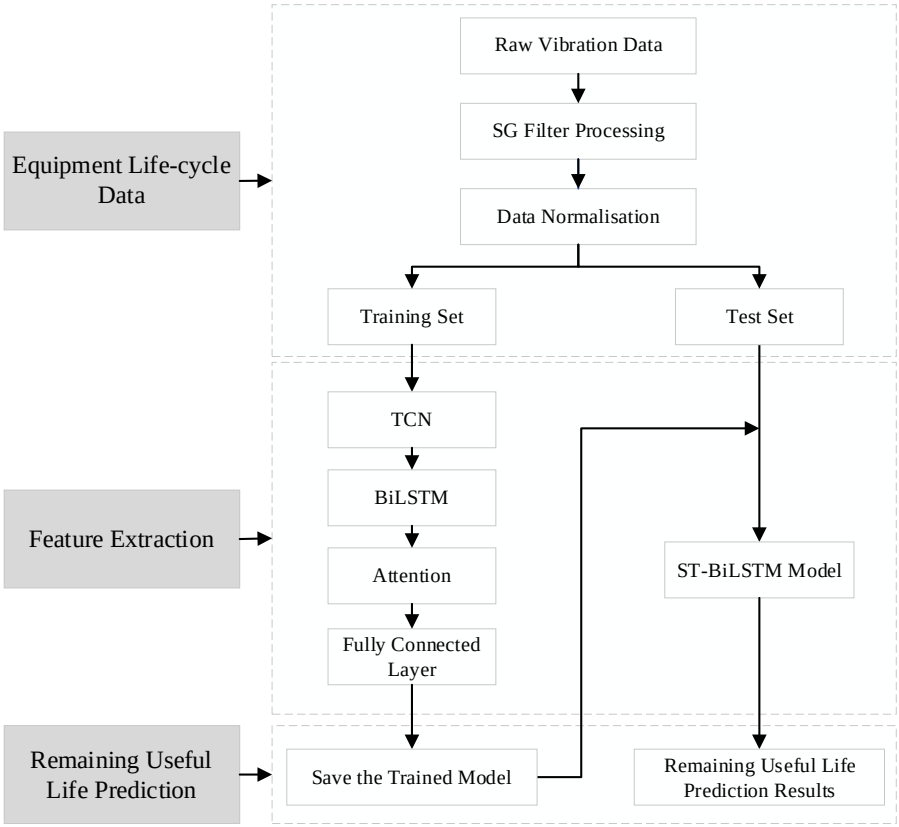
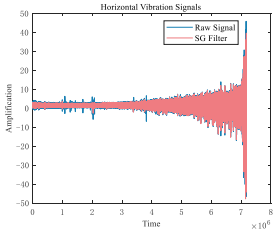
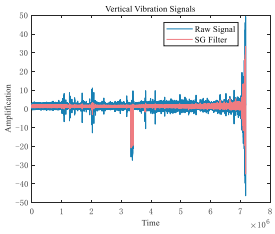
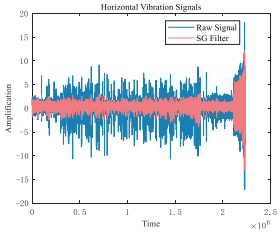


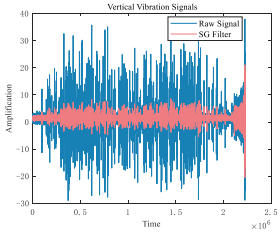

**Figure 6:** ST-BiLSTM-based prediction flowchart



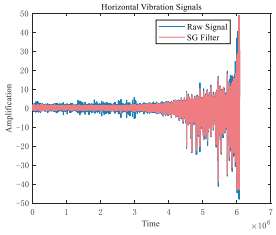



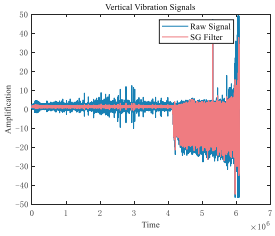

a. Bearing 1-1 noise reduction



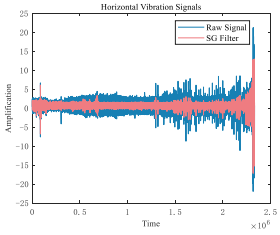



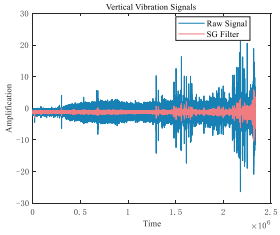

b. Bearing 1-2 noise reduction



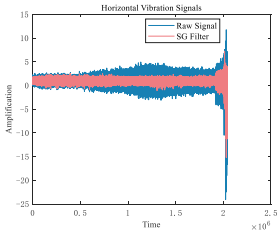



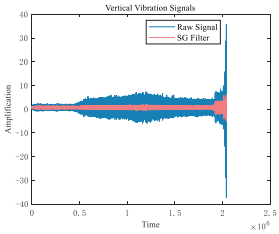

c. Bearing 1-3 noise reduction



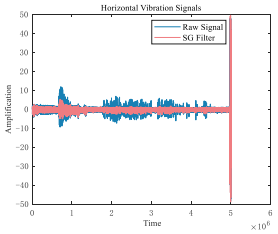



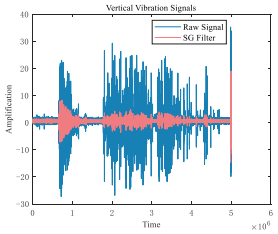

d. Bearing 2-1 noise reduction




e. Bearing 2-2 noise reduction




f. Bearing 2-3 noise reduction

**Figure 7:** SG noise reduction effect

# 4. EXPERIMENTAL VERIFICATION

## 4.1. Experimental Dataset

In order to confirm the effectiveness and superiority of the bearing RUL prediction approach suggested in this paper, we took advantage of the dataset made available by the IEEE 2012 PHM Data Challenge, which was obtained from the PRONOSTIA test bed. The data acquisition platform, shown in Figure 8, consists of sensors, asynchronous motors, and test bearings. Before testing began, all bearings were in good condition. On the bearing box, a total of two accelerometers were attached for the purpose of measuring horizontal and vertical vibrations. The data were captured every 10s at a sampling frequency of 25.6 kHz for a duration of 0.1s, which resulted in a collection of 2560 data points. When the bearing vibration signal's amplitude was greater than 20g ($1g = 9.8m/s^2$), the bearing was considered to have failed, and the experiment was stopped. Table 1 displays the operational conditions for this data set.

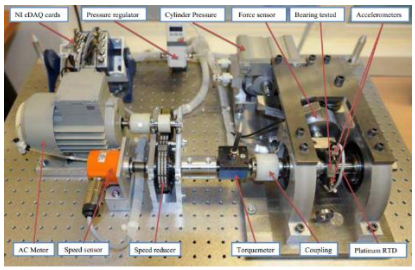



**Figure 8:** the PRONOTIA platform

**Table 1:** Operating conditions for the FEMTO-ST bearing dataset

| Operating Condition | Load(N) | Speed(rpm) | Bearing |
|---|---|---|---|
| Condition 1 | 4000 | 1800 | B1-1~B1-7 |
| Condition 2 | 4200 | 1650 | B2-1~B2-7 |
| Condition 3 | 5000 | 1500 | B3-1~B3-3 |

## 4.2. Data Preprocessing

The raw vibration signals were first processed for noise reduction, as shown in Figure 7. Since different operational settings may result in different sensor values and the obtained data represents different physical properties, in order to eliminate the effect of data irregularities on the outcomes of the predictions, before any training and testing, data normalization must be performed by averaging and calculating the standard deviation of the data measured by each sensor under each operating condition based on the classified operating conditions. After that, Z-score normalization is done for each data point, i.e., subtracting the mean value of the sensor under the condition and dividing it by the standard deviation of the sensor under the condition, which is calculated as shown in equation (13):

$$Z_i = \frac{X_i - X}{\delta} \tag{13}$$

where $X$ represents the mean of the sequence $X$ and $\delta$ represents the sequence's standard deviation $X$ .

## 4.3. Indicators for the Evaluation of Experiments

In order to assess its performance against other cutting-edge PHM algorithms. For the purpose of evaluating overall performance, we make use of the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and scoring mechanism. Mathematically, MAE and RMSE are defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Rul_i^{pre} - Rul_i^{act}\right| \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Rul_i^{pre} - Rul_i^{act})^2} \tag{15}$$

where $Rul_i^{act}$ is the true value of remaining life, $Rul_i^{pre}$ is the estimated value of remaining life, and n represents the total length of the prediction.

Reference [13] is where the scoring function's definition may be found. The scoring function designed in reference [13] differs from the scoring function in PHM 2012 [14] in that it considers the consequences of underestimation, overestimation, and the entire life-cycle operating phase. Equations (16) and (17) display the detailed descriptions:

$$er_i = Rul_i^{pre} - Rul_i^{act} \tag{16}$$

$$A_i = \begin{cases} \exp(-\ln(0.6)\cdot(\frac{er_i}{10})), er_i \leqslant 0 \\ \exp(\ln(0.6)\cdot(\frac{er_i}{40})), er_i > 0 \end{cases} \tag{17}$$

Equation (18) displays the evaluation function's score:

$$Score = \alpha\frac{1}{m}\sum_{i=1}^{m}A_i + \beta\frac{1}{n-m}\sum_{m+1}^{n}A_i \tag{18}$$

where $\alpha$ and $\beta$ are weights for early and late bearing stages, respectively, and $m$ represents the proportion of early phases.

## 4.4. Parameter Configuration of ST-BiLSTM

We need to set hyperparameters in ST-BiLSTM, which are detailed in Table 2. The determination of these hyperparameters involves the process of cross-validating several training datasets and taking into account the accuracy of the predictions. The trials were carried out within the context of deep learning using the TensorFlow framework. The computer system is equipped with an 11th Generation Intel (R) Core(TM) i7-11370H central processing unit and 16GB RAM.

**Table 2:** ST-BiLSTM network structure parameter settings

| Parameter name | Parameterization |
|---|---|
| Input | Input dimension (2560*2)<br>The number of output channels is 16<br>The convolution kernel size is 12<br>The step size is 4 |
| TCN | Dropout layer discard rate is 0.5<br>3 Residual Blocks<br>Global average pooling<br>Lambda Layer |
| BiLSTM | The number of LSTM cells is 16<br>The dropout layer discard rate is 0.2 |
| Attention | Permute Layer and Reshape Layer |
| FC | The quantity of output channels is 1<br>The activation function is Relu |
| Output | The output dimension is 1 |

## 4.5. Experimental Prediction Results and Discussion

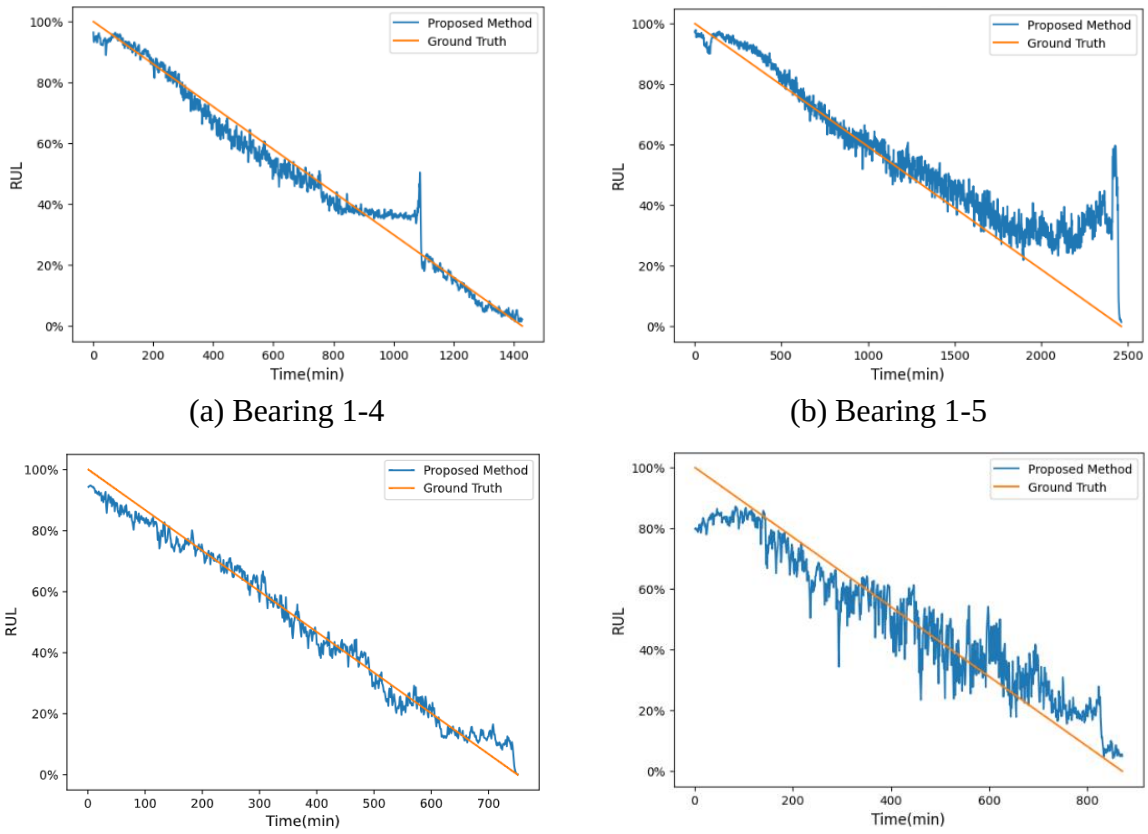
To comprehensively assess the performance of the prediction network, it is necessary to conduct a thorough evaluation. We selected the bearing vibration data under working conditions 1 and 2. As can be seen from Figure 7, both its horizontal and vertical vibration signals contain degradation information, so its horizontal and vertical vibration signals are spliced and processed to form a 3-dimensional matrix to mine its deep degradation information in each working condition. In the context of testing, a single set of bearing data is utilized, while the remaining bearing data is employed for training the network. Four different methodologies were chosen for the purpose of comparing their respective outcomes in predicting RUL with the findings obtained by the ST-BiLSTM model. The outcomes of the five methodologies' predictions are presented in Table 3.

**Table 3:** Predicted outcomes of various methodologies

| Test Set | TCN-SA | | | BiLSTM-SA | | | TCN-BiLSTM | | | CNN-BiLSTM | | | ST-BiLSTM | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE | RMSE | Score | MAE | RMSE | Score | MAE | RMSE | Score | MAE | RMSE | Score | MAE | RMSE | Score |
| B 1-1 | 12.9 | 15.2 | 0.82 | 12.2 | 16.3 | 0.57 | 14.5 | 18.0 | 0.82 | 19.4 | 25.1 | 0.78 | 8.72 | 11.5 | 0.86 |
| B 1-2 | 12.9 | 18.4 | 0.84 | 17.3 | 20.8 | 0.46 | 9.82 | 13.6 | 0.73 | 11.9 | 14.1 | 0.60 | 10.9 | 14.7 | 0.85 |
| B 1-3 | 10.3 | 13.2 | 0.73 | 24.1 | 30.6 | 0.38 | 16.9 | 23.2 | 0.80 | 9.00 | 10.5 | 0.68 | 6.70 | 8.34 | 0.83 |
| B 1-4 | 3.75 | 5.07 | 0.90 | 8.45 | 12.8 | 0.69 | 17.7 | 23.3 | 0.79 | 8.34 | 9.98 | 0.69 | 5.43 | 6.94 | 0.90 |
| B 1-5 | 7.38 | 9.23 | 0.74 | 14.5 | 18.6 | 0.51 | 11.3 | 14.1 | 0.72 | 7.32 | 10.3 | 0.74 | 16.0 | 21.2 | 0.81 |
| B 1-6 | 15.1 | 19.8 | 0.83 | 11.4 | 15.5 | 0.58 | 7.91 | 11.9 | 0.72 | 9.57 | 13.4 | 0.64 | 16.1 | 20.3 | 0.81 |
| B 1-7 | 15.4 | 19.8 | 0.82 | 19.6 | 24.4 | 0.51 | 18.6 | 21.6 | 0.80 | 21.7 | 23.4 | 0.76 | 13.2 | 18.0 | 0.83 |
| B 2-1 | 14.3 | 18.1 | 0.74 | 27.9 | 33.4 | 0.31 | 16.8 | 20.9 | 0.72 | 22.2 | 28.5 | 0.70 | 19.5 | 26.1 | 0.78 |
| B 2-2 | 17.9 | 23.0 | 0.77 | 21.0 | 25.9 | 0.41 | 11.9 | 16.7 | 0.78 | 17.9 | 21.6 | 0.75 | 4.12 | 5.29 | 0.87 |
| B 2-3 | 10.9 | 13.2 | 0.80 | 29.3 | 35.3 | 0.29 | 21.3 | 27.5 | 0.77 | 18.1 | 23.1 | 0.74 | 16.1 | 19.4 | 0.79 |
| B 2-4 | 9.11 | 13.0 | 0.83 | 33.5 | 37.8 | 0.22 | 22.1 | 27.8 | 0.72 | 19.6 | 26.7 | 0.67 | 4.66 | 5.93 | 0.91 |
| B 2-5 | 11.4 | 14.1 | 0.78 | 25.6 | 31.3 | 0.33 | 28.9 | 36.1 | 0.73 | 27.7 | 34.6 | 0.71 | 3.83 | 4.81 | 0.89 |
| B 2-6 | 13.7 | 18.5 | 0.80 | 25.1 | 29.8 | 0.31 | 23.2 | 30.6 | 0.76 | 6.57 | 8.71 | 0.73 | 8.10 | 10.7 | 0.84 |
| B 2-7 | 24.4 | 30.8 | 0.75 | 31.6 | 39.4 | 0.30 | 19.8 | 27.0 | 0.72 | 15.2 | 18.9 | 0.67 | 11.5 | 14.3 | 0.80 |

In this experiment, a total of 14 bearings were subjected to testing. As seen from the data presented in Table 3, the scoring values obtained from the method provided in this research exhibit higher magnitudes when compared to those obtained through the comparison method, except that the scores of tested bearings 1-6 and 2-3 are lower than those of TCN-SA, and the scores of bearings 1-4 are the same as those of TCN-SA. This implies that the ST-BiLSTM model is more effective in capturing information related to bearing degradation. As evidenced by its performance, we designed a network model of TCN-BiLSTM-Attention, where the advantages of TCN in handling the time series problem are applied to the lifetime prediction to learn the sequence features, the BiLSTM model is used to learn the before and after conditions to extract more information in the sequence for prediction, the attention mechanism is introduced to dynamically adjust the weights of the attributes, and finally the fully-connected layer is employed to generate the output representing the RUL. The low RUL prediction accuracy of BiLSTM-SA is due to the weak feature extraction function of BiLSTM data; thus, its RUL prediction accuracy is low. Compared with CNN-BiLSTM and TCN-BiLSTM, ST-BiLSTM has higher scores, which indicates the importance of the attention mechanism to dynamically adjust the attribute weights. While TCN-SA incorporates a self-attention mechanism, its limited perceptual field results in suboptimal prediction of RUL for orientation. In conclusion, the

enhanced acquisition of sequential characteristics is the underlying factor contributing to the greater predictive outcomes attained by the ST-BiLSTM through the implementation of the TCN-BiLSTM-Attention architecture.




(a) Bearing 1-4 (b) Bearing 1-5

(c) Bearing 2-4 (d) Bearing 2-6

**Figure 9:** Results of RUL visualization for test bearings in the PHM 2012 dataset

The prediction results for bearings 1-4, 1-5, 2-4, and 2-6 are illustrated in Figure 9. The effectiveness of our proposed ST-BiLSTM in capturing bearing deterioration information under various operating situations and accurately predicting the RUL may be observed in Figure 9. Specifically, in Figures 9(a) and 9(b), during a later phase of the prediction process, the ST-BiLSTM model is capable of acquiring data pertaining to the deterioration of the bearing and subsequently terminating the test bearing's lifespan prematurely. Early shutdown is of significant importance at industrial application sites because of its ability to prevent additional failures and accidents, thereby reducing the risk of injuries and fatalities. However, the performance of ST-BiLSTM is subpar in the later phases of predicting bearings 1-5. This could be attributed to the existence of many faults in the bearing and the lack of degradation information towards the conclusion of its lifespan.

## 5. CONCLUSIONS

This paper presents a novel approach for predicting RUL using the ST-BiLSTM model. Our method incorporates the TCN and BiLSTM models while also incorporating the attention mechanism and SG filtering algorithm. Specifically, in order to enhance prediction accuracy, we initially employ an SG filter to mitigate the presence of noise in the time series data, therefore reducing its interference; then the utilization of TCN is employed for the purpose of extracting probable features from the aforementioned data. Additionally, LSTM is utilized to effectively capture the long-term dependencies that are present within the series; finally, the attention mechanism is employed to allocate varying weights to the attributes in order to improve the accuracy of predictions. The conclusions drawn from the analysis of the PHM 2012 bearing degradation dataset are as follows:

1) The ST-BiLSTM-based equipment remaining service life prediction method utilizes an SG filter to ensure the consistent form and width of the data during the process of noise filtering. The TCN-BiLSTM network is capable of extracting information on the operating state features of the equipment and assigning different weights to the extracted feature information through the attention mechanism to better extract information regarding the condition of the equipment's health;
2) The incorporation of an attentional mechanism has been shown to enhance the precision of the deep neural network model's forecast of the RUL;
3) The ST-BiLSTM model obtained better evaluation indexes for life prediction and higher accuracy of life prediction compared with TCN-SA, BiLSTM-SA, TCN-BiLSTM, and CNN-BiLSTM methods.

## REFERENCES


[1] D K Nayak, D P Das, S K Behera, et al., "Monitoring the fill level of a ball mill using vibration sensing and artificial neural network," Neural Computing and Applications, vol. 32, pp. 1501-1511, 2020.

[2] K Sharma, D Goyal, and R Kanda., "Intelligent Fault Diagnosis of Bearings based on Convolutional Neural Network using Infrared Thermography," Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, vol. 236, no. 12, pp. 2439-2446, 2022.

[3] L Cui, X Wang, H Wang, et al., "Research on remaining useful life prediction of rolling element bearings based on time-varying Kalman filter," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 6, pp. 2858-2867, 2019.

[4] W Zhang, D Yang, and H Wang., "Data-driven methods for predictive maintenance of industrial equipment: A survey., IEEE systems journal, vol. 13, no. 3, pp. 2213-2227, 2019.

[5] F Li, L Zhang, B Chen, et al., " A light gradient boosting machine for remainning useful life estimation of aircraft engines," 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 3562-3567, 2018.

[6] D Wu, C Jennings., J Terpenny, et al. "A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests," Journal of Manufacturing Science and Engineering, vol. 139, no. 7, Article ID 071018, 2017.

[7] M Ma, and Z Mao., "Deep recurrent convolutional neural network for remaining useful life prediction," 2019 IEEE International Conference on Prognostics and Health Management (ICPHM). IEEE, pp. 1-4, 2019.

[8] A Graves, and A Graves., "Long short-term memory," Supervised sequence labelling with recurrent neural networks, pp. 37-45, 2012.

[9] S Yousuf, S A Khan, and S Khursheed., "Remaining useful life (RUL) regression using Long–Short Term Memory (LSTM) networks," Microelectronics Reliability, vol. 139, Article ID 114772, 2022.

[10] Z H Zhao, Q Li, and S P Yang, et al., "Prediction of remaining service life based on BiLSTM and attention mechanism," Journal of Vibration and Shock, vol. 41, no. 6, pp.44-50, 2022.

[11] D Li, and L Yang., "Remaining useful life prediction of lithium battery based on sequential CNN–LSTM method," Journal of Electrochemical Energy Conversion and Storage, vol. 18, no. 4, Article ID 041005, 2021.

[12] S Bai, J Z Kolter, and V Koltun., "An empirical evaluation of generic convolutional and recurrent networks for sequence modelling," arXiv, 2018.

[13] Y Wang, L Deng, L Zheng, et al., "Temporal convolutional network with soft thresholding and attention mechanism for machinery prognostics," Journal of Manufacturing Systems, vol. 60, pp. 512-526, 2021.

[14] P Nectoux, R Gouriveau, and K Medjaher, et al., "An experimental platform for bearings accelerated degradation tests," Proceedings of the IEEE International Conference on Prognostics and Health Management IEEE, pp. 23-25, Beijing, China, 2012.