Systems Assessment and Engineering Management

Journal Homepage: sciencesforce.com/saem

Syst. Assess. Eng. Manage. Vol. 1 (2024) 8–25

Paper Type: Original Article

A Data-driven Deep Learning Approach for Remaining Useful Life of Rolling Bearings

Ahmed Darwish [1,*]

[1] Department of Computer Science, Faculty of Computer and Informatics, Zagazig University, Zagazig, 44519, Egypt; adarwish@fci.zu.edu.eg.

Received: 30 Oct 2023 **Revised**: 11 Jan 2024 **Accepted**: 09 Feb 2024 **Published**: 13 Feb 2024

Abstract

The bearing is a commonly used rotating element, and its condition significantly impacts the operation and maintenance of machinery. Therefore, accurately predicting the Remaining Useful Life (RUL) of bearings holds great importance. Deep learning has made significant progress in RUL prediction. This study presents a Deep Learning (DL) model incorporating a Convolution Neural Network (CNN), Long Short-Term Memory (LSTM), and attention mechanism to enhance RUL prediction accuracy for rolling bearings. Initially, time domain input data is processed by the CNN for feature extraction. Subsequently, two LSTM layers are utilized to capture intricate temporal relationships and create more abstract data representations, followed by the incorporation of an attention mechanism to align input and output sequences based on the content or semantics of the input sequence. Ultimately, the final predictions are made through a Fully Connected (FC) layer. The effectiveness of the proposed model is evaluated using the IEEE PHM 2012 Challenge dataset, and its performance is compared to various deep learning models to showcase its efficacy. Experimental results indicate that the suggested CNN-ALSTM model is a reliable choice for predicting the RUL of rolling bearings, outperforming all other models considered.

Keywords: Remaining Useful Life; Long Short-Term Memory Network; Attention Mechanism; Convolution Neural Network; Rolling Bearings.

1 | Introduction

Rolling bearings are crucial and costly rotating elements found in various industrial machinery systems, such as high-speed railways, induction motors, and wind turbine drivetrain systems [1]. The advancement in manufacturing intelligence has led to an increased need for system reliability. Being a critical component of rotating machinery, the condition of rolling bearings significantly impacts equipment reliability. Failure of rolling bearings can result in significant financial losses and serious injuries [2; 3]. Research conducted by Rai and Upadhyay [4] indicates that around half of motor failures are attributed to faults in rolling bearings. Accurate prediction of remaining useful life (RUL) is essential to ensure the safe and steady operation of the system [5], enhance equipment reliability [6], and lower operation and maintenance expenses [7]. Bearings, as a key element supporting the operation of rotating machinery, play a crucial role in maintaining equipment stability. The condition and efficiency of bearings directly influence the safety and dependability of mechanical equipment. Therefore, it is imperative to thoroughly assess the reliability of bearings and enhance the methods for predicting the RUL of bearings [8]. Numerous scholarly research methodologies have been developed in

Corresponding Author: adarwish@fci.zu.edu.eg

https://doi.org/10.61356/j.saem.2024.1251

Licensee **Systems Assessment and Engineering Management**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

recent times for the examination of rolling bearings and the estimation of RUL. These methodologies are generally classified into two main groups: model-based techniques and data-driven approaches [9].

Model-based techniques commonly assess the deterioration trend of a specific component by employing mathematical or physical models, often represented as a set of ordinary and partial differential equations. Examples include the particle filter algorithm [10], Kalman filter algorithm [11], Weibull process model [12], Gamma process model [13], Paris-Erdogan model [14], Wiener process model [15], and Levy process model [16]. In a previous study [17], a method using an improved particle filter was introduced for predicting the Remaining Useful Life (RUL) of bearings. Furthermore, other researchers have utilized the proportional hazard model for RUL prediction, as seen in the works of Liao et al [18]. and Tian et al [19]. Chen and colleagues [20] presented an improved approach to forecast Remaining Useful Life (RUL) by employing a particle filter. This technique was created by combining the linear optimization resampling particle filter (LORPF) with the sliding-window gray model (SGM). Gai et al. [21] introduced a technique for estimating the fatigue lifespan by leveraging contact stress in accordance with the fatigue theoretical design approach. This method enables the determination of the peak contact stress given the specified radial and axial forces acting on the bearing, facilitating the derivation of the contact fatigue lifespan by referencing the contact fatigue life curve. Nevertheless, this empirical understanding is frequently based on universal principles, making it challenging to address various diverse degradation patterns that could arise. Wang et al. [22] introduced an enhanced exponential model (EM) for the estimation of the Remaining Useful Life (RUL) of bearings. In a separate study, Kumar et al. [23] put forward a new health deterioration metric for machinery using the concept of Kullback-Leibler divergence. Furthermore, model-based approaches often rely on a singular model for analyzing distinct datasets. These models commonly employ a series of equations to elucidate the failure mechanisms of bearings and deliver precise predictions. Nevertheless, as equipment becomes more intricate and intelligent, the failure mechanisms of bearings manifest varied and indirect traits, resulting in a rise in the quantity of parameters in the physical models and complicating the task of constructing precise physical models. Serious inaccuracies in RUL prediction arise from the constraints of model-based approaches when faced with dynamic environments and stochastic disturbances. Data-driven methodologies, in contrast to model-based techniques, offer an alternative for examining issues by establishing connections between monitoring indicators and RUL estimates without relying on universally adaptable models for intricate systems beyond individual cases.

The second strategy entails the utilization of data-driven methodologies. In contrast to model-centric approaches, data-driven techniques do not necessitate the development of a sophisticated electrochemical model for rolling bearings. Instead, they primarily extract implicit information from capacity degradation data of rolling bearings to facilitate the RUL. Data-centric methodologies include Machine Learning (ML) and Deep Learning (DL) techniques. Machine learning methods have emerged as a potent tool in various domains, granting computers the capacity to learn from data without explicit programming, continually enhancing their performance. The capability of machine learning to derive insights from data and execute tasks autonomously is reshaping our lifestyles, professions, and interactions with technology. With further advancements in this domain, we can anticipate even more profound impacts on our global society. Machine learning techniques can utilize extensive sets of sensor data, operational parameters, and historical maintenance records. This data-driven approach empowers machine learning models to comprehend intricate relationships among various factors influencing the condition and degradation of machinery. several initial machine learning (ML) algorithms such as support vector machine (SVR) [24], random forest (RF) [25], Gaussian process regression [26], hidden Markov model (HMM) [27], and artificial neural networks (ANN) [28] have been employed in predicting RUL for machinery, yielding significant results. Berghout et al. [29] outlined the essential procedures for predicting the remaining operational lifespan using machine learning and methodically discussed the potential opportunities and obstacles that lie ahead. Centered on the concepts of monotonicity and trend, Javed et al. [30] identified features from raw vibration data to forecast the remaining lifespan of cutting tools and bearings using extreme learning machines. Mejia et al. [28] extracted the wavelet packet decomposition coefficients from the raw sensor data and employed a mixture of Gaussians Hidden Markov

Model (HMM) to assess the current health condition of the equipment. The model was also used to predict the RUL value along with the associated level of certainty. Machine learning algorithms have progressed in the prediction of RUL for rolling bearings; nonetheless, numerous of these methods require substantial feature engineering to recognize crucial characteristics. In addition, the moderate intricacy of these models limits their capacity to effectively capture data and exhibit robust generalization capabilities.

Deep learning, a significant component of machine learning, has brought about significant transformations in numerous facets of our everyday lives through the utilization of artificial neural networks consisting of multiple layers to handle information in a manner that emulates the cognitive processes of the human brain. The progress in deep learning methodologies is largely credited to their adaptability, as they obviate the necessity for manual feature engineering by independently deriving feature representations. commonly used deep learning architectures such as Convolutional Neural Networks (CNNs) [31-33] and long short-term memory (LSTM) [34-36] are frequently employed in forecasting the RUL of Rolling Bearings. Wang et al. [33] introduced the spatiotemporal non-negative projected convolutional network (SNPCN) approach as a means to identify the deterioration characteristics present in neighboring matrices, leveraging a three-dimensional convolutional neural network (3DCNN). They subsequently confirmed the efficacy of this model on the PRONOSTIA platform. Wang et al. [37] introduced a deep separable convolutional network (DSCN) for predicting the Remaining Useful Life (RUL) of rolling bearings. This network utilizes the condition monitoring data from different sensors as its direct input. Yang et al. [38] developed a dual-CNN model for predicting RUL. The first CNN is responsible for pinpointing the initial failure point, while the second CNN is utilized for forecasting the RUL value. Zhou et al. [3] developed an unsupervised health indicator (HI) using a Gaussian mixture model (GMM) and Kullback-Leibler divergence (KLD). By integrating this with a GRU network, they conducted predictions on time series data. Their findings suggest that the utilization of this intelligent prediction approach holds promise for applications within the realm of engineering. An et al. [39] outlined a CNN that utilized a combination of stacked bi-directional and uni-directional LSTM networks for the purpose of predicting the RUL of tools. Xia et al. [40] utilized a CNN for the extraction of specific degradation features from the tool's sensor data. Subsequently, they identified the temporal relationships among these features using bi-directional long short-term memory (BLSTM) networks. Finally, the Remaining Useful Life (RUL) of the tools was forecasted through a fully connected layer.

The aforementioned deep learning-based methods for RUL prediction have significantly advanced the field of mechanical RUL forecasting. Nevertheless, these methods are constrained by certain drawbacks: firstly, while CNNs are effective at extracting features, their receptive field is restricted by the convolution kernels' volume when used in isolation. Secondly, recurrent networks require computations to progress from the beginning to the end. As the sequence length increases, the computation time of these networks grows notably, and the vanishing gradient issue makes it challenging for the model to learn long-term dependencies effectively.

This study introduces a new data-driven methodology that integrates the CNN network, LSTM network, and attention mechanism (CNN-ALSTM) to enhance the accuracy of RUL prediction for rolling bearings. The time-domain input data is fed into the CNN network for feature extraction. An attention mechanism has been incorporated within the LSTM architecture to assign weight values to the extracted features, thus highlighting important information and improving the RUL prediction of the model. The effectiveness of the CNN-ALSTM model is evaluated on the widely used IEEE PHM 2012 Challenge Dataset [41] for forecasting the RUL of Rolling Bearings. Our experimental results indicate that the proposed method reduces uncertainty in multi-step prediction tasks and outperforms other existing models in terms of accuracy. The major contributions of this paper are listed as follows:

- A novel method is suggested for predicting the RUL of Rolling Bearings. This approach entails employing CNN for extracting features from aging data, then utilizing LSTM to capture detailed temporal patterns from the extracted features, and finally reconstructing the output state through the application of layers of attention mechanism.

- An attention mechanism was created to improve the accuracy of predicting the RUL by selectively filtering input features and assigning higher significance to key attributes.
- The validation of the model was carried out using the IEEE PHM 2012 Challenge Dataset, showcasing its superior performance in predicting RUL in comparison to alternative models.

The subsequent sections of this paper are structured as follows. The materials and methods are outlined in Section 2. Subsequently, the proposed model will be discussed in Section 3. This will be followed by the presentation of experiments and results in Section 4. Followed by applications in section 5. Lastly, the conclusions will be presented in Section 6.

2 | Materials and Methods

2.1 | IEEE PHM 2012 Challenge Dataset

This portion of the study utilizes the IEEE PHM 2012 dataset to verify the proposed deep model. The dataset was gathered through a laboratory experimental setup called PRONOSTIA, designed to hasten the deterioration of bearings under both steady and changing operational conditions. It also gathers real-time health monitoring information such as vibration, temperature, rotational speed, and load force [41]. The experimental setup is capable of simulating the degradation of bearings in a short period, providing authentic experimental data, and outlining the deterioration process of the ball bearing over its lifespan. Figure 1 displays a depiction of the machinery. The vibration signals, both horizontal and vertical, are sampled at a frequency of 25.6 kHz, resulting in 25600 samples being captured per second. For this particular investigation, a total of 1560 samples have been selected, as illustrated in Figure 2. These samples provide data on three working conditions: 1800 rpm and 4000 N, 1650 revolutions per minute and 4200 N, and 1500 rpm along with 5000 N. Further specifics can be found in Table 1. Figure 3 depicts the vibration signals of both horizontal and vertical acceleration for Dataset Bearing1_3.

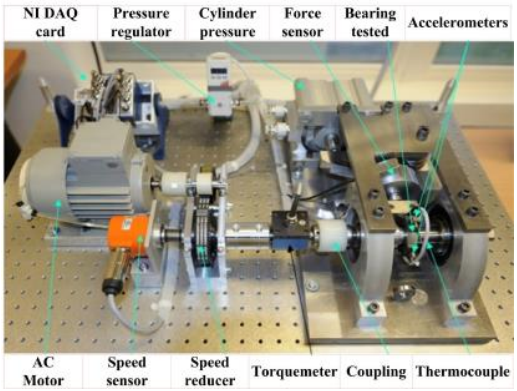


Figure 1. PRONOSTIA experimental platform [41].

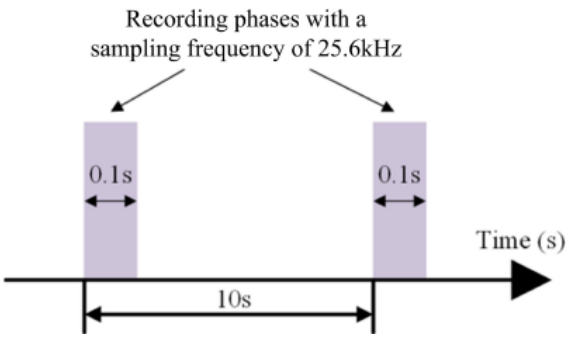
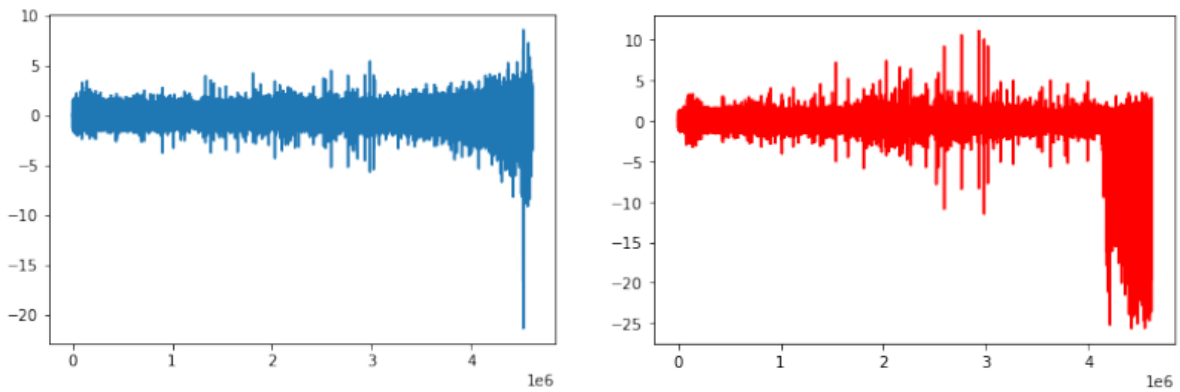


Figure 2. Illustration of acquisition parameters for vibration signals.



(a) (b)

Figure 3. 1d horizontal (top plot) and vertical (bottom plot) vibration signals.

Table 1. Data on three working conditions.

Datasets	**Operating conditions**		
	Condition_1	Condition_2	Condition_3
Rotating speed	2100	2250	2400
Radial force	1200	1100	1000
Learning sets	Bearing 1_1	Bearing 2_1	Bearing 3_1
	Bearing 1_2	Bearing 2_2	Bearing 3_2
Test sets	Bearing 1_3	Bearing 2_3	Bearing 3_3
	Bearing 1_4	Bearing 2_4	Bearing 3_4
	Bearing 1_5	Bearing 2_5	Bearing 3_5

2.2 | Convolutional Neural Network (CNN)

Convolutional Neural Networks are unlikely to possess the ability to automatically detect significant attributes from raw time series information, eliminating the requirement for manual feature extraction. This eliminates the requirement for domain-specific knowledge in selecting features and empowers the model to expose patterns that might not be discernible to humans. Similar to their application in image data, CNNs can identify low-level features like local patterns and trends in initial layers and high-level features such as temporal patterns and complex dependencies in deeper layers. This hierarchical feature learning ability enables CNNs to capture both local and global characteristics of time series data, including local patterns and dependencies. By utilizing convolutional filters with small receptive fields, CNNs can identify temporal patterns like recurring motifs, spikes, or abrupt changes in the time series. This capacity to capture local patterns is crucial for recognizing relevant features that could indicate significant events or anomalies in the data. CNNs employ parameter sharing, where the same set of weights (filters) is used across various temporal positions in the time series. This sharing of parameters reduces the number of trainable parameters in the network, resulting in more effective learning and enhanced generalization. Moreover, techniques such as pooling and dropout regularization can enhance the network's resilience to noise and overfitting. In certain scenarios, CNNs may offer quicker training on time series data due to their ability to process data simultaneously, leveraging GPU capabilities for efficient feature extraction.

2.3 | Long Short-Term Memory (LSTM)

The LSTM presents a new version of the recurrent neural network (RNN) structure, specifically crafted to address the issue of disappearing gradients in traditional RNNs, focusing on handling the challenges posed by long-term dependencies in predictive tasks. This innovative design features a complex memory cell arrangement that distinguishes itself by its capacity to effectively retain and utilize information across

extended sequences, making it well-suited for tasks requiring the prediction of long-term dependencies, such as Remaining Useful Life (RUL) prediction in Rolling Bearings. The LSTM cell structure is depicted in Figure 4. In the LSTM model, the forget gate f_t, input gate i_t, and output gate o_t form the three fundamental components that manage information flow and control interactions within the network. The forget gate f_t decides which information to exclude from the previous cell state. By considering the current input and previous hidden state, the forget gate f_t produces an output value between 0 (complete forgetfulness) and 1 (full retention). The mathematical representation of f_t is described by Eq. (1). The input gate i_t is responsible for selecting new data to retain in the cell state. By analyzing the current input and preceding hidden state, it generates an output between 0 and 1, along with a new candidate value for inclusion in the cell state. The mathematical representation of i_t is determined by Eq. (2). The output gate o_t controls the information to be transmitted as the hidden state of the current LSTM cell, based on the current input and previous hidden state. Its output ranges from 0 to 1. The mathematical representation of o_t is provided by Eq. (3). The candidate value c'_t signifies new data that may be added to the cell state at the current time step (t), generated by the input gate considering the current input and previous hidden state. The mathematical representation of c'_t is given by Eq. (4). Following this, the c_t value representing the unit state at time t is determined through Eq. (5). Subsequently, the h_t value representing the hidden state at time t is ascertained using mathematical Eq. (6).

$$f_t = \sigma\left(W_f\, x_t + U_f\, h_{t-1} + b_f\right) \tag{1}$$

$$i_t = \sigma(W_i\, x_t + U_i\, h_{t-1} + b_i) \tag{2}$$

$$o_t = \sigma(W_o\, x_t + U_o\, h_{t-1} + b_o) \tag{3}$$

$$c'_t = tanh(W_a\, x_t + U_a\, h_{t-1} + b_a) \tag{4}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t \tag{5}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{6}$$

Where the symbol σ denotes the sigmoid function, t represents the time step, x_t signifies the input feature at time t, h_{t-1} denotes the output hidden state from the previous time sample, the parameters $W_f, W_i, W_o, W_a, U_f, U_i, U_o, U_a, b_f, b_i, b_o, b_a$ are optimized during the training process.

2.4 | Scaled Dot-Product Attention

In the realm of time series analysis, the most recent data holds significant importance in accurately predicting future values. Scaled dot-product attention is a technique that assists in emphasizing these recent segments of the LSTM output by assigning greater significance to their respective key vectors during the scoring process. Scaled dot-product attention serves as a fundamental component utilized in various attention mechanisms. Integrating it post-LSTM enables the model to concentrate on specific segments of the LSTM output sequence that are deemed most pertinent for prediction purposes. The LSTM output undergoes projection into three distinct vector spaces - Query, Key, and Value - through linear transformations. The query vector (Q) denotes the focal point of the model at each time step. The key vector (K) embodies the available information at each time step within the sequence, while the value vector (V) contains the factual data content for each time step. A score is computed by conducting a scaled dot product between the query vector (Q) and each key vector (K) throughout the sequence, as depicted in Eq. (7). Subsequently, a softmax function is applied to these scores, transforming them into a probability distribution denoted as the attention weights (A) in Eq. (8). These weights signify the relative significance of each time step in the sequence concerning the current prediction (based on the query vector). Ultimately, the attention weights (A) are utilized to assign weight to the corresponding value vectors (V) by Eq. (9). This process generates a context

vector summarizing the most pertinent information from the entire sequence based on the current focus (query).

$$attention_{scores} = \frac{Q \ . \ K^T}{\sqrt{d_k}} \tag{7}$$

$$A = softmax(attention_{scores}) \tag{8}$$

$$context_{vector} = A \ . \ V \tag{9}$$

Where d_k is the dimensionality of the key vectors, and $\sqrt{d_k}$ represents a scaling factor to stabilize the gradients.

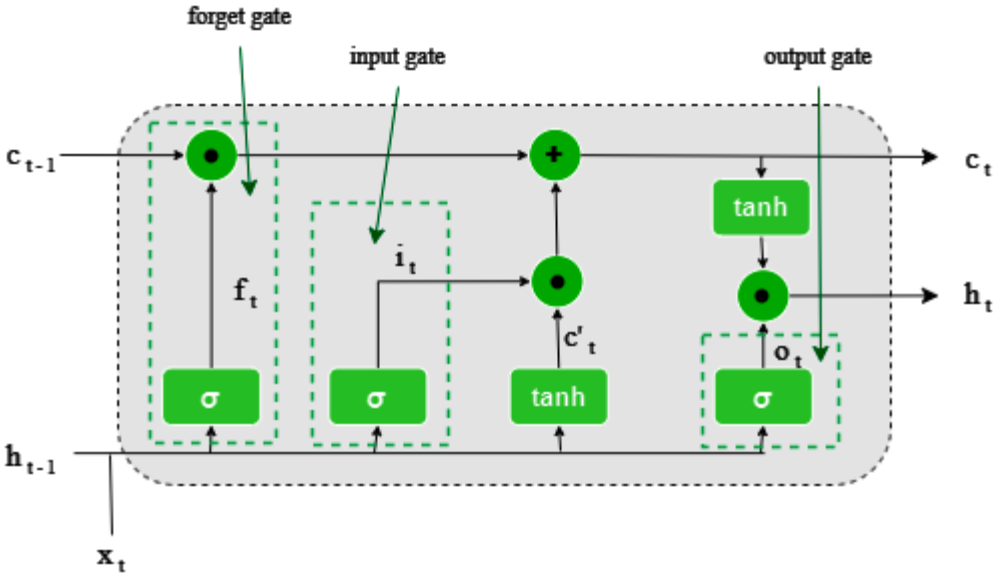


Figure 4. LSTM cell architecture.

3 | The Proposed Model

The estimation of Remaining Useful Life (RUL) for Rolling Bearings is viewed as a supervised regression challenge, which involves utilizing data sourced from the PRONOSTIA platform to train and assess various deep-learning models. This study presents a new Deep Learning model named CNN-ALSTM, which integrates CNN, LSTM, and an attention mechanism to predict the RUL of Rolling Bearings, as depicted in Figure 5. The CNN component is employed for extracting features from the time domain input data. It comprises three convolution blocks, each consisting of a convolution layer, a RELU activation function layer, and a dropout layer to mitigate overfitting, as demonstrated in Figure 6. The output from the CNN is then passed to the LSTM, which comprises two LSTM layers. Each LSTM layer in the sequence processes the input sequentially, transmitting information through memory cells and gates. The output of one LSTM layer acts as the input for the next layer, facilitating the model in capturing hierarchical representations of the input data. The inclusion of multiple LSTM layers enables the model to capture complex temporal relationships and develop more abstract data representations. In this work, a network architecture involving a CNN for feature extraction and two LSTM layers is employed to analyze the time sequence information extracted iteratively by the CNN, allowing for a comprehensive integration of the input data samples. Following the final LSTM layer, an attention mechanism layer is introduced to calculate the neuron weights of the hidden state layer. This layer plays a crucial role in computing the output through the LSTM, thereby assigning weight coefficients and reconstructing data to identify the essential aspects of the extracted features. The processed data from the hidden layer is merged and forwarded to the fully connected (FC) layer for the final RUL prediction. Lastly, Algorithm 1 outlines the pseudocode of the proposed model.

As delineated in Algorithm 1, the proposed framework undergoes a sequence of procedures to handle the input data. Initially, the input data goes through preprocessing before being forwarded to the input layer. This layer then transmits the data to a network structure comprising three convolution blocks aimed at extracting features. Each convolution block comprises a convolution layer with 32 filters, a RELU activation function layer, and a dropout layer with a dropout rate of 0.5. Subsequently, the output from the CNN network is input into the LSTM network, which features two LSTM layers with 128 neurons each, in addition to a Tanh activation function. This framework is crafted to examine the temporal sequence information iteratively, facilitating a thorough assimilation of the input data samples. Following this, the network's output is channeled to the Scaled dot-product attention mechanism to compute the weights of the neurons within the hidden state layer. The resultant attention output is then funneled into a fully connected layer with a single neuron to predict the remaining useful life of rolling bearings.

Algorithm 1 Pseudo-code of CNN-ALSTM

Input: **Input data (*D*), batch size (*Bs*), maximum epoch (*T*), and learning rate (*lr*)**

Output: ***loss (MSE),RMSE***

Conducting the preprocessing step
Input: **Construct an** input layer **to receive the input data**
/* Feature extraction based on the CNN */
/*First Convolution block*/
x: **Create a** Conv layer **with 32 filters and a kernel size of 2 to take the data from the** input layer.
x: **Add** RELU **activation function layer to** x.
x: **Add a** Dropout **layer with a dropout rate of 0.5 to** x.
/*Second Convolution block*/
x: **Create a** Conv layer **with 32 filters and a kernel size of 2 to take the data from the** First Convolution block.
x: **Add** RELU **activation function layer to** x.
x: **Add a** Dropout **layer with a dropout rate of 0.5 to** x.
/*Third Convolution block*/
x: **Create a** Conv layer **with 32 filters and a kernel size of 2 to take the data from the** Second Convolution block.
x: **Add** RELU **activation function layer to** x.
x: **Add a** Dropout **layer with a dropout rate of 0.5 to** x.
/* LSTM network*/
x: **Add an** LSTM **layer with 128 units to** *x*.
x: **Add an** LSTM **layer with 128 units to** *x*.
attention_output: **Add** Scaled dot-product attention **to** *x*.
/* Prediction Block */
x: **Add a** Linear layer **with** 1 **node to** *x*.
/* Optimization process */
$N = Size(D)/Bs$ **/* Estimate the number of batches */**
$\boldsymbol{t = 0}$**, Current epoch**
while $\boldsymbol{t < T}$
    $\boldsymbol{i = 0}$**, the current batch size.**
    while $\boldsymbol{i < N}$
        Compute the Score function **using the** $\boldsymbol{ith}$ batch.
        Update the weights based on the Adam to optimize the score **function.**
        $\boldsymbol{i = i + 1}$
    end while.
    $\boldsymbol{t = t + 1}$
end while

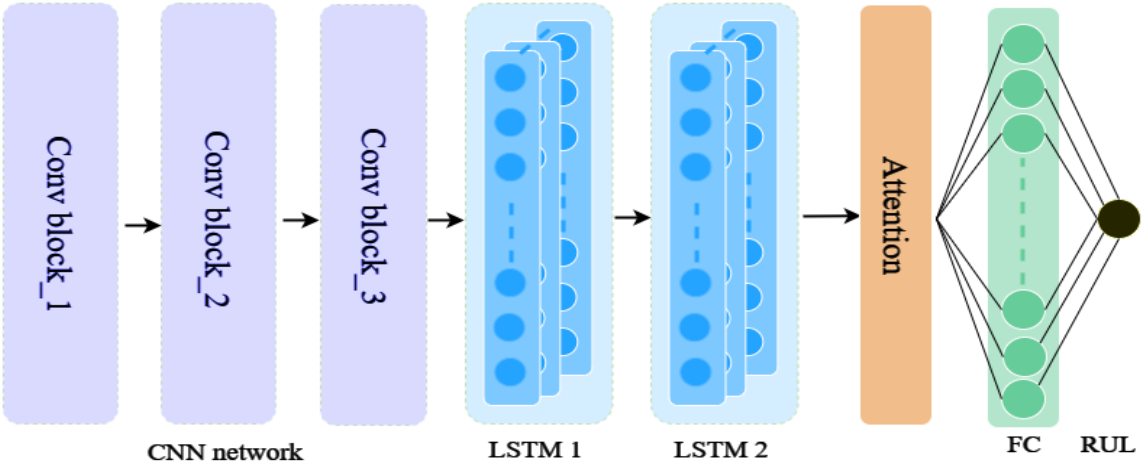


Figure 5. Flowchart of the proposed CNN-ALSTM.

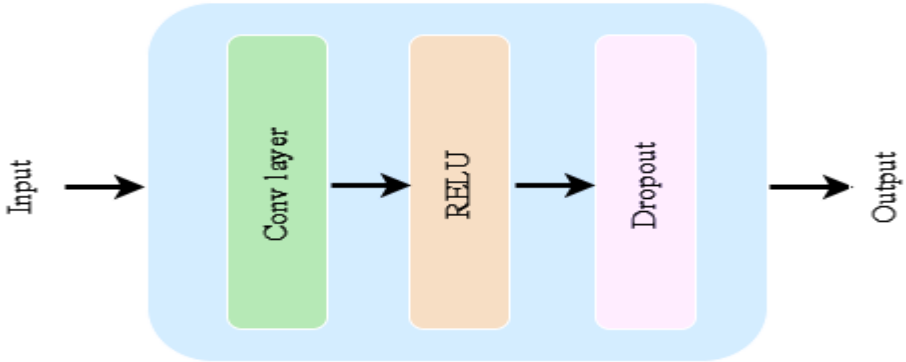


Figure 6. Conv block architecture.

4 | Experiments and Results

4.1 | Data Preprocessing

Within the domain of data processing, the dataset employed in the IEEE PHM 2012 Challenge encompasses a range of attributes with different magnitudes, which can potentially compromise the efficacy of Deep Learning (DL) models during the training phase. It is imperative that the inputs to the deep learning model exhibit uniformity, as any notable disparity could have detrimental effects on the model's performance. As a result, vibration signals gathered at each time interval should undergo normalization before their utilization in the model. One common technique for this is z-score normalization, also referred to as standardization, which serves to standardize the values of these attributes. This method of scaling holds significant importance in the realm of machine learning, especially when confronted with features that possess diverse scales or units. By adjusting the data to a standardized range, this technique ensures that the features are rescaled to yield a mean of 0 and a standard deviation of 1. The mathematical formula for min-max scaling is presented as Eq. (10). where $\mu_j,\ and\ \sigma_j$ donate to the mean and standard deviation of the jth feature, respectively.

$$x'_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j} \tag{10}$$

In the context of monitoring time series problems, it is imperative to consider the significance of historical data. The utilization of time window embedding has been shown to significantly enhance the effectiveness of the model [42] by incorporating valuable temporal information. Nonetheless, the incorporation of time window embedding results in an exponential growth in model size as it scales proportionally with the length of the sequence data. The implementation of the sliding window approach is a fundamental technique utilized in various applications of signal processing and machine learning, particularly when dealing with sequential data like time series. This approach involves breaking down a continuous stream of data points into smaller, partially overlapping (or non-overlapping) segments for subsequent analysis. During this procedure, a sliding window is utilized to segment the input data. Assuming d represents the initial value of the sliding window, and the total number of samples in the training data is n, the vector $x^{(i)}$ within the sliding window can be described as follows:

$$x^{(i)} = [x_i, x_{i+1}, x_{i+2}, \dots, x_{i+d-1}], \qquad i = 1 \rightarrow n - d \tag{11}$$

After establishing the most suitable embedding dimension, the training set $Data_{train}$ can be structured accordingly. These collections are depicted as:

$$Data_{train} = \left[\left(x^{(1)}, x_{d+1}\right), \dots, \left(x^{(i)}, x_{d+i}\right), \dots, \left(x^{(n-d)}, x_n\right)\right], i = 1 \rightarrow n - d \tag{12}$$

Where x_{d+i} is the label for input data, and the testing set follows the same technique.

4.2 | Evaluation Metrics

In this article, the Adam [43] optimization algorithm and mean square error (MSE) loss, calculated using Eq. (13), are utilized to optimize the network parameters. The introduction of root mean square error (RMSE) is used as an evaluation metric for the proposed model [44]. This is done by comparing RMSE values between the actual and predicted labels of all instances in the dataset. The RMSE value is computed mathematically as outlined in Eq. (14), where N represents the number of samples, and y_i and y'_i represent the true and predicted labels of the ith sample, respectively. The goal is to minimize both MSE and RMSE to improve the accuracy of predicting the Remaining Useful Life (RUL) of Rolling Bearings.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - y'_i\right)^2 \tag{13}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - y'_i\right)^2} \tag{14}$$

4.3 | Hyperparameter Tuning

The CNN_ALSTM model introduced in this study incorporates various hyper-parameters, such as the number of LSTM layers, the number of filters, kernel size, learning rate, and window size, etc., that need precise calibration to enhance efficiency and minimize the RMSE. Consequently, a sequence of experiments is conducted to investigate different configurations for each parameter to pinpoint the optimal values that result in a significant improvement in the model's performance, as illustrated in Table 2. Notably, the model's efficacy is influenced by the quantity of hidden units within each LSTM layer. Therefore, multiple experiments were executed to identify the most suitable number of hidden dimensions for the LSTM layer, ranging from 64, 128, 256, to 512. The impact of the number of hidden dimensions is depicted in Figure 7. The findings of these experiments reveal that the ideal number of hidden dimensions for the LSTM layer is 128. Likewise, a series of experiments were performed using window sizes of 8, 16, 32, and 64 to ascertain the optimal window size. The outcomes demonstrate that a window size of 32 is the most appropriate for the specific dataset. The effect of the window size is visualized in Figure 8. The learning rate plays a critical role in the

training process, affecting the convergence speed, model effectiveness, and training process stability. The experiments indicate that a learning rate of 0.001 produces the best performance. The impact of different learning rate values is illustrated in Figure 9.

Table 2. Experimental analysis of the influence of parameters on prediction results.

Bearings		1_3	1_4	1_5	2_3	2_4	2_5	3_3
Hidden dim of LSTM	64	0.083	0.111	0.155	0.213	0.179	0.131	0.221
	128	0.043	0.086	0.108	0.170	0.146	0.094	0.188
	256	0.062	0.095	0.121	0.182	0.163	0.102	0.197
	512	0.071	0.098	0.141	0.189	0.174	0.105	0.201
Window size	8	0.071	0.107	0.145	0.198	0.187	0.124	0.215
	16	0.057	0.098	0.121	0.185	0.160	0.110	0.203
	32	0.043	0.086	0.108	0.170	0.146	0.094	0.188
	64	0.052	0.095	0.119	0.181	0.154	0.101	0.195
No. of LSTM	1	0.114	0.194	0.210	0.278	0.204	0.164	0.235
	2	0.043	0.086	0.108	0.170	0.146	0.094	0.188
	3	0.081	0.103	0.147	0.193	0.184	0.115	0.202
Learning rate	0.0001	0.152	0.138	0.209	0.274	0.203	0.198	0.244
	0.001	0.043	0.086	0.108	0.170	0.146	0.094	0.188
	0.002	0.067	0.099	0.127	0.192	0.164	0.102	0.204
	0.01	0.197	0.254	0.386	0.399	0.272	0.216	0.315
Kernel size	1	0.154	0.214	0.295	0.304	0.284	0.201	0.311
	2	0.043	0.086	0.108	0.170	0.146	0.094	0.188
	3	0.092	0.136	0.199	0.251	0.207	0.182	0.261
No. of filters	8	0.102	0.182	0.193	0.214	0.201	0.189	0.257
	16	0.084	0.105	0.164	0.191	0.184	0.127	0.211
	32	0.043	0.086	0.108	0.170	0.146	0.094	0.188
	64	0.092	0.151	0.172	0.199	0.193	0.135	0.224

Table 3. The CNN_ALSTM hyperparameters.

Parameter	value
No. of LSTM layer	2
No. of Conv blocks	3
Hidden dim of LSTM	128
No. of filters	32
Kernal size	2
Window size	32
Dropout rate	0.5
Learning rate	0.001
Max no. of epoch	1000
Loss	MSE
Optimizer	Adam

The number of filters present in the layers of CNN plays a crucial role as a hyperparameter that affects the network's capacity, feature representation, generalization ability, and computational efficiency. It is imperative to meticulously adjust the number of filters according to the task's complexity, the attributes of the input data, and the computational resources available to achieve optimal results. Subsequently, a series of experiments were conducted to identify the ideal number of filters for the CNN layer, with variations such as 8, 16, 32, and 64. The results of these experiments demonstrate that the most appropriate number of filters for the

CNN layer is 32. The impact of different numbers of filters is illustrated in Figure 10. Similarly, the size of the kernel in a CNN layer is a critical hyperparameter that has a direct influence on the network's capacity to extract features from input data. Consequently, multiple experiments were carried out to establish the optimal kernel size for the CNN layer, with options including 1, 2, and 3. The findings of these experiments reveal that the most suitable kernel size for the CNN layer is 2. The impact of different kernel size values is illustrated in Figure 11. Despite the advantages of deeper networks, such architectures may lead to overfitting and heightened computational complexity, necessitating various experiments to ascertain the best number of Long Short-Term Memory (LSTM) layers. Empirical evidence suggests that a network comprising 2 LSTM layers demonstrates superior performance, as depicted in Table 2. The impact of different numbers of LSTM layers is illustrated in Figure 12. The final hyperparameters of the proposed model are detailed in Table 3.

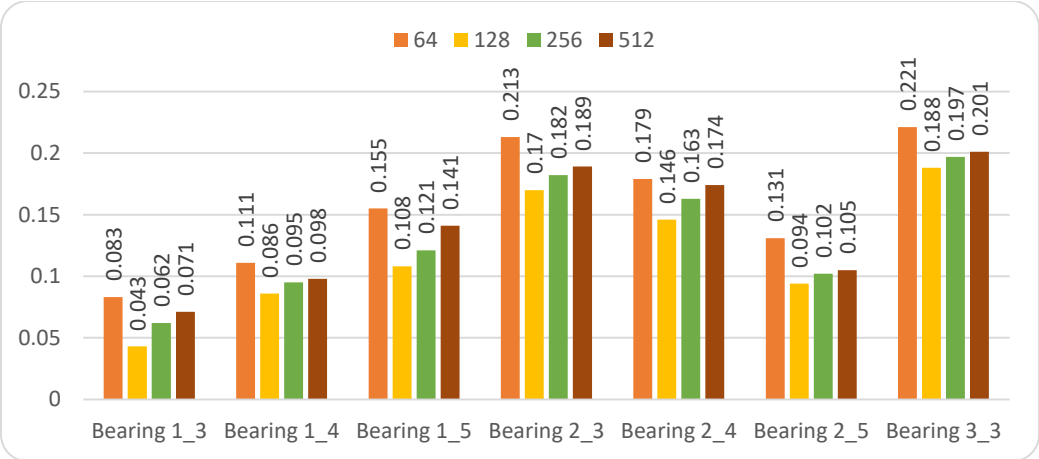


Figure 7. The representation of the number of hidden dimensions acquired through experiments.

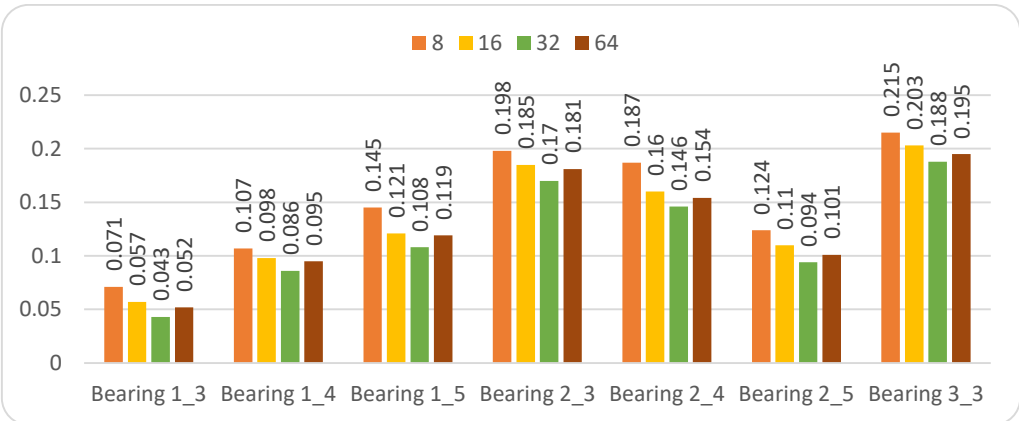


Figure 8. The representation of window size values acquired through experiments.

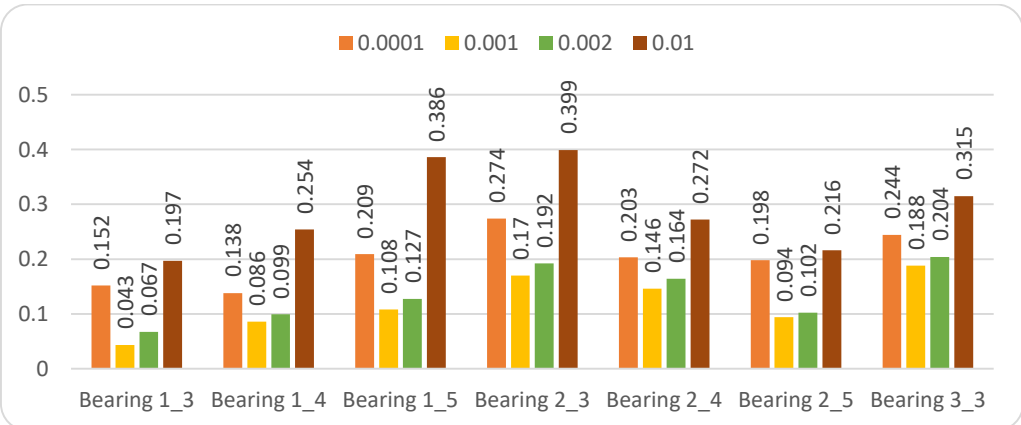


Figure 9. The representation of learning rate values acquired through experiments.

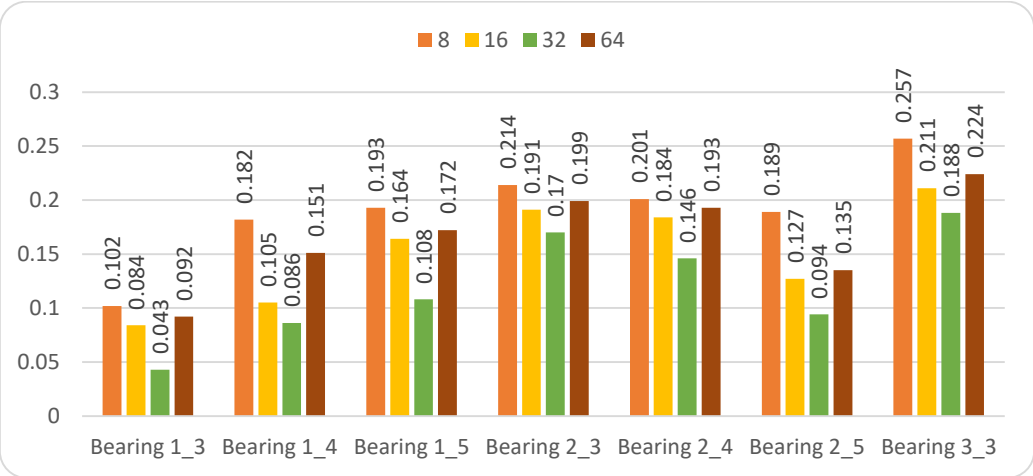


Figure 10. The representation of the number of filter values acquired through experiments.

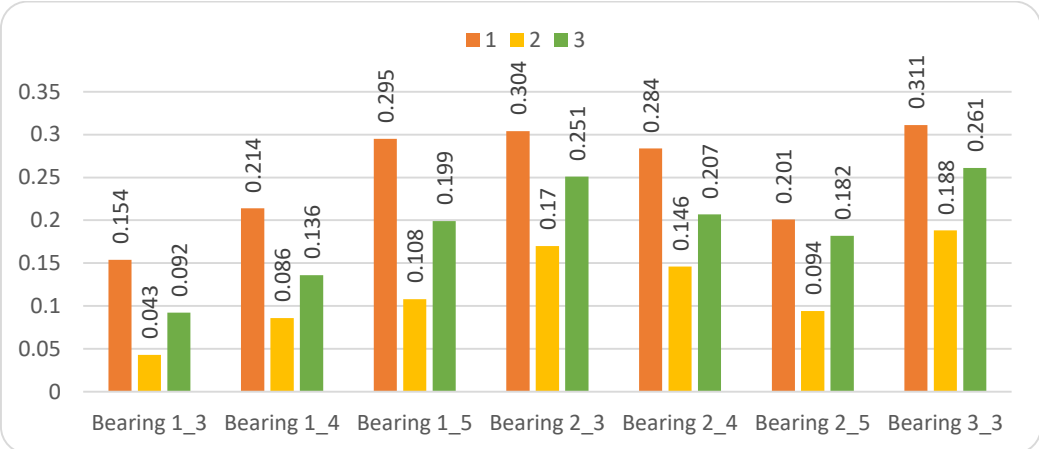


Figure 11. The representation of kernel size values acquired through experiments.

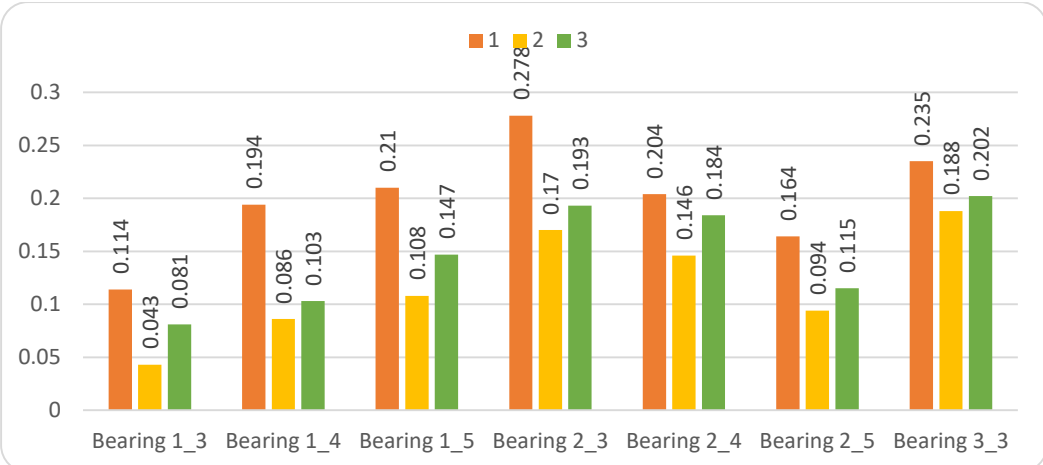


Figure 12. The representation of the number of LSTM layers acquired through experiments.

4.4 | Ablation Experiments

In the executed ablation study, an assessment is conducted on the effectiveness of a composite model framework integrating the CNN, LSTM, and Scaled dot-product attention mechanism models. The main objective is to analyze the impact of each component on predicting the RUL of Rolling Bearings and to evaluate the overall improvement in performance resulting from their combination. The evaluation will be based on the RMSE metrics. An examination is performed to ascertain the influence of the attention

mechanism on the model's performance. Findings are displayed in a table revealing forecast errors for CNN, LSTM, CNN+LSTM, CNN+Attention, LSTM+Attention, and CNN+LSTM+Attention (CNN_ALSTM). All outcomes are delineated in Table 4 and depicted in Figure 13. The integration of CNN excels in autonomously acquiring pertinent features directly from raw data, rendering them essential tools for diverse feature extraction tasks. Their capacity to capture local patterns and global structures in data has contributed to their widespread acceptance and success in various applications. The integration of LSTM is crucial in time series scenarios due to its capability to capture long-term dependencies, manage sequences of varying lengths, learn intricate temporal patterns, and effectively handle noisy and incomplete data. Their efficacy in modeling sequential data renders them valuable for a variety of time series forecasting tasks. The integration of the attention mechanism embedding leads to a noteworthy decrease in errors. Experimental results indicate that incorporating an attention mechanism significantly enhances the model's predictive performance. Furthermore, the method's performance improves gradually as the attention mechanism is reinforced, providing additional evidence of its efficacy.

Table 4. The results of the ablation study.

Bearing	**1_3**	**1_4**	**1_5**	**2.3**	**2_4**	**2_5**	**3_3**
CNN	0.124	0.134	0.187	0.341	0.332	0.162	0.384
LSTM	0.103	0.112	0.146	0.255	0.241	0.196	0.298
CNN+LSTM	0.084	0.105	0.131	0.205	0.188	0.125	0.231
CNN+Attention	0.081	0.094	0.129	0.194	0.178	0.109	0.208
LSTM+Attention	0.062	0.089	0.121	0.185	0.159	0.102	0.195
CNN+LSTM+Attention(CNN_ALSTM)	**0.043**	**0.086**	**0.108**	**0.170**	**0.146**	**0.094**	**0.188**

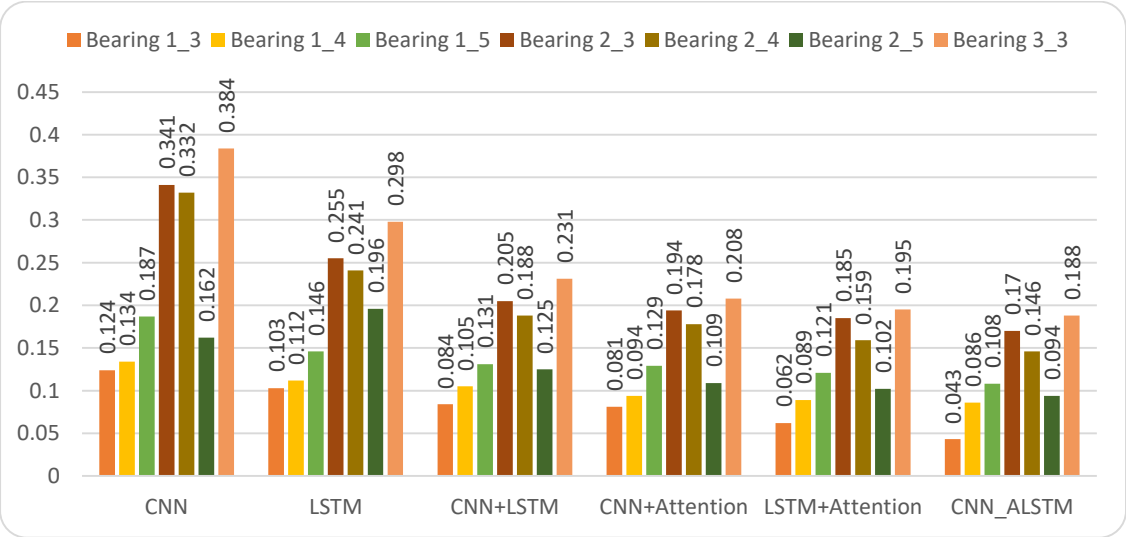


Figure 13. The representation of RMSE values acquired through ablation experiments.

5 | Application

This section presents the results achieved by the proposed At-LSTM model and various competing models for Bearing 1_3, Bearing 1_4, Bearing 1_5, Bearing 2_3, Bearing 2_4, Bearing 2_5, and Bearing 2_3 sourced from the IEEE PHM 2012 Challenge dataset. These results are evaluated using the RMSE metric to demonstrate the models' effectiveness in reducing the discrepancy between the predicted and target RUL. The outcomes of CNN_ALSTM Bearing 1_3, Bearing 1_4, Bearing 1_5, Bearing 2_3, Bearing 2_4, Bearing 2_5, and Bearing 2_3 sourced from the IEEE PHM 2012 Challenge dataset are compared with five competing models to demonstrate its effectiveness and efficiency. A comprehensive comparison is conducted between the outcomes of CNN_ALSTM and several rival models, including SVM [45], LSTM [46], ALSTM, CONVLSTM [47], and TCN, to showcase its superior performance. These results are displayed in the RMSE values as indicated in the Table 5.

Table 5. The precise specifications of the chosen Bearings from the IEEE PHM 2012 Challenge dataset.

Bearing	1_3	1_4	1_5	2.3	2_4	2_5	3_3
SVM	0.11	0.177	0.186	0.298	0.178	0.165	0.268
LSTM	0.083	0.138	0.15	0.169	0.165	0.145	0.23
ALSTM	0.065	0.122	0.126	0.139	0.161	0.139	0.221
CONVLSTM	0.055	0.103	0.131	0.224	0.154	0.134	0.209
TCN	0.046	0.088	0.114	0.174	0.149	0.095	0.193
Proposed method	**0.043**	**0.086**	**0.108**	**0.170**	**0.146**	**0.094**	**0.188**

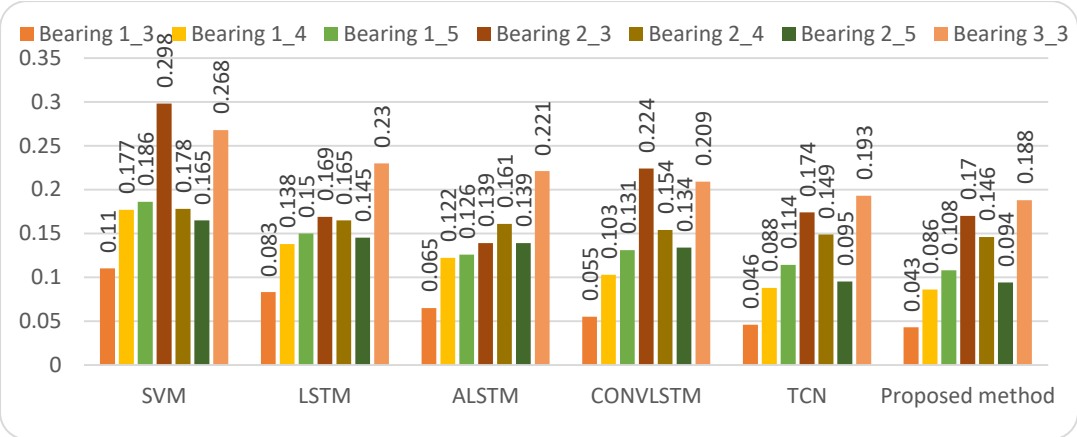


Figure 14. illustrates the representation of RMSE values acquired from different models.

The superior results are emphasized in bold font. CNN_ALSTM outperforms other methodologies with the RMSE metric. The data presented in the table indicates that CNN_ALSTM has the potential to surpass all the models under consideration in terms of RMSE for the specified bearings, achieving RMSE values of 0.043, 0.086, 0.108, 0.170, 0.146, 0.094, and 0.188 for Bearing 1_3, Bearing 1_4, Bearing 1_5, Bearing 2_3, Bearing 2_4, Bearing 2_5, and Bearing 2_3, respectively. This performance significantly exceeds that of the comparison methods under similar conditions. Upon comparison of our findings with the best outcomes achieved by the various models mentioned, our proposed model demonstrates a reduction in RMSE by 6.5%, 2.3%, 5.3%, 2.3%, 2%, 1.1%, and 2.6% for the IEEE PHM 2012 Challenge dataset for Bearing 1_3, Bearing 1_4, Bearing 1_5, Bearing 2_3, Bearing 2_4, Bearing 2_5, and Bearing 2_3, respectively. This suggested model is considered a robust solution for addressing this problem due to its ability to excel in the RMSE metric, which places equal emphasis on early and late predictions. To visually demonstrate the superiority of our model, Figure 14 is provided to showcase the RMSE values obtained by various algorithms for Bearing 1_3, Bearing 1_4, Bearing 1_5, Bearing 2_3, Bearing 2_4, Bearing 2_5, and Bearing 2_3.

6 | Conclusions

Bearings are commonly utilized in rotating machinery, and accurately predicting the RUL is crucial for making informed maintenance decisions to prevent unexpected downtime and ensure the safety of machinery. A new method known as CNN-ALSTM is proposed for forecasting the RUL of rolling bearings. This model combines CNN, LSTM, and a Scaled dot-product attention mechanism, with CNN used for extracting features from time domain input data. LSTMs are employed to capture and preserve patterns over long sequences, making them suitable for capturing complex temporal relationships in time series data. Additionally, an attention mechanism is integrated to align input and output sequences by considering the context or significance of the input sequence. The final RUL predictions are generated through a Fully Connected (FC) layer. Our experimental analysis utilized the IEEE PHM 2012 Challenge dataset. Comparison with the top-performing models referenced in literature showed that our proposed model achieved a reduction in RMSE of 6.5%, 2.3%, 5.3%, 2.3%, 2%, 1.1%, and 2.6% for the respective bearings (Bearing 1_3,

Bearing 1_4, Bearing 1_5, Bearing 2_3, Bearing 2_4, Bearing 2_5, and Bearing 2_3) in the IEEE PHM 2012 Challenge dataset.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Qian, Y., Yan, R., & Hu, S. (2014). Bearing degradation evaluation using recurrence quantification analysis and Kalman filter. IEEE Transactions on Instrumentation and Measurement, 63(11), 2599-2610. https://doi.org/10.1109/TIM.2014.2313034

[2] Liu, Z., Liu, H., Jia, W., Zhang, D., & Tan, J. (2021). A multi-head neural network with unsymmetrical constraints for remaining useful life prediction. Advanced Engineering Informatics, 50, 101396. https://doi.org/10.1016/j.aei.2021.101396

[3] Zhou, J., Qin, Y., Chen, D., Liu, F., & Qian, Q. (2022). Remaining useful life prediction of bearings by a new reinforced memory GRU network. Advanced Engineering Informatics, 53, 101682. https://doi.org/10.1016/j.aei.2022.101682

[4] Rai, A., & Upadhyay, S. H. (2017). The use of MD-CUMSUM and NARX neural network for anticipating the remaining useful life of bearings. Measurement, 111, 397-410. https://doi.org/10.1016/j.measurement.2017.07.030

[5] Canito, A., Fernandes, M., Mourinho, J., Tosun, S., Kaya, K., Turupcu, A., . . . Marreiros, G. (2021). Flexible architecture for data-driven predictive maintenance with support for offline and online machine learning techniques. Paper presented at the IECON 2021–47th Annual Conference of the IEEE Industrial Electronics Society.

[6] Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. Mechanical systems and signal processing, 42(1-2), 314-334. https://doi.org/10.1016/j.ymssp.2013.06.004

[7] Shen, F., & Yan, R. (2021). A new intermediate-domain SVM-based transfer model for rolling bearing RUL prediction. IEEE/ASME Transactions on Mechatronics, 27(3), 1357-1369. https://doi.org/10.1109/tmech.2021.3094986

[8] Gao, S., Yu, Y., & Zhang, Y. (2022). Reliability assessment and prediction of rolling bearings based on hybrid noise reduction and BOA-MKRVM. Engineering Applications of Artificial Intelligence, 116, 105391. https://doi.org/10.1016/j.engappai.2022.105391

[9] Zuo, T., Zhang, K., Zheng, Q., Li, X., Li, Z., Ding, G., & Zhao, M. (2023). A hybrid attention-based multi-wavelet coefficient fusion method in RUL prognosis of rolling bearings. Reliability Engineering & System Safety, 237, 109337. https://doi.org/10.1016/j.ress.2023.109337

[10] Li, S., Fang, H., & Shi, B. (2021). Remaining useful life estimation of lithium-ion battery based on interacting multiple model particle filter and support vector regression. Reliability Engineering & System Safety, 210, 107542. https://doi.org/10.1016/j.ress.2021.107542

[11] Cui, L., Wang, X., Wang, H., & Ma, J. (2019). Research on remaining useful life prediction of rolling element bearings based on time-varying Kalman filter. IEEE Transactions on Instrumentation and Measurement, 69(6), 2858-2867. https://doi.org/10.1109/TIM.2019.2924509

[12] Kundu, P., Darpe, A. K., & Kulkarni, M. S. (2019). Weibull accelerated failure time regression model for remaining useful life prediction of bearing working under multiple operating conditions. Mechanical systems and signal processing, 134, 106302. https://doi.org/10.1016/j.ymssp.2019.106302

[13] Wang, H., Liao, H., Ma, X., & Bao, R. (2021). Remaining useful life prediction and optimal maintenance time determination for a single unit using isotonic regression and gamma process model. Reliability Engineering & System Safety, 210, 107504. https://doi.org/10.1016/j.ress.2021.107504

[14] Haile, M. A., Riddick, J. C., & Assefa, A. H. (2016). Robust particle filters for fatigue crack growth estimation in rotorcraft structures. IEEE Transactions on Reliability, 65(3), 1438-1448. https://doi.org/10.1109/TR.2016.2590258

[15] Djeziri, M., Benmoussa, S., Mouchaweh, M. S., & Lughofer, E. (2020). Fault diagnosis and prognosis based on physical knowledge and reliability data: Application to MOS Field-Effect Transistor. Microelectronics Reliability, 110, 113682. https://doi.org/10.1016/j.microrel.2020.113682

[16] Wang, P., Gao, R. X., & Woyczynski, W. A. (2020). Lévy process-based stochastic modeling for machine performance degradation prognosis. IEEE Transactions on Industrial Electronics, 68(12), 12760-12770. https://doi.org/10.1109/TIE.2020.3047037

[17] Qian, Y., & Yan, R. (2015). Remaining useful life prediction of rolling bearings using an enhanced particle filter. IEEE Transactions on Instrumentation and Measurement, 64(10), 2696-2707. https://doi.org/10.1109/TIM.2015.2427891

[18] Liao, H., Zhao, W., & Guo, H. (2006). Predicting remaining useful life of an individual unit using proportional hazards model and logistic regression model. Paper presented at the RAMS'06. Annual Reliability and Maintainability Symposium, 2006.

[19] Tian, Z., & Liao, H. (2011). Condition based maintenance optimization for multi-component systems using proportional hazards model. Reliability Engineering & System Safety, 96(5), 581-589. https://doi.org/10.1016/j.ress.2010.12.023

[20] Chen, L., An, J., Wang, H., Zhang, M., & Pan, H. (2020). Remaining useful life prediction for lithium-ion battery by combining an improved particle filter with sliding-window gray model. Energy Reports, 6, 2086-2093. https://doi.org/10.1016/j.egyr.2020.07.026

[21] Cai, S., & Zhang, G. (2017). Fatigue life prediction of high-speed railway bearing based on contact stress. Paper presented at the 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM).

[22] Wang, G., & Xiang, J. (2021). Remain useful life prediction of rolling bearings based on exponential model optimized by gradient method. Measurement, 176, 109161. https://doi.org/10.1016/j.measurement.2021.109161

[23] Kumaraswamidhas, L., & Laha, S. (2021). Bearing degradation assessment and remaining useful life estimation based on Kullback-Leibler divergence and Gaussian processes regression. Measurement, 174, 108948. https://doi.org/10.1016/j.measurement.2020.108948.

[24] Benkedjouh, T., Medjaher, K., Zerhouni, N., & Rechak, S. (2013). Remaining useful life estimation based on nonlinear feature reduction and support vector regression. Engineering Applications of Artificial Intelligence, 26(7), 1751-1760. https://doi.org/10.1016/j.engappai.2013.02.006

[25] Kundu, P., Darpe, A. K., & Kulkarni, M. S. (2020). An ensemble decision tree methodology for remaining useful life prediction of spur gears under natural pitting progression. Structural Health Monitoring, 19(3), 854-872. https://doi.org/10.1177/14759217198657

[26] Aye, S. A., & Heyns, P. (2017). An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission. Mechanical systems and signal processing, 84, 485-498. https://doi.org/10.1016/j.ymssp.2016.07.039

[27] Shin, S.-H., Kim, S., & Seo, Y.-H. (2018). Development of a fault monitoring technique for wind turbines using a hidden markov model. Sensors, 18(6), 1790. https://doi.org/10.3390/s18061790

[28] Ali, J. B., Chebel-Morello, B., Saidi, L., Malinowski, S., & Fnaiech, F. (2015). Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. Mechanical systems and signal processing, 56, 150-172. https://doi.org/10.1016/j.ymssp.2014.10.014

[29] Berghout, T., & Benbouzid, M. (2022). A systematic guide for predicting remaining useful life with machine learning. Electronics, 11(7), 1125. https://doi.org/10.3390/electronics11071125

[30] Javed, K., Gouriveau, R., Zerhouni, N., & Nectoux, P. (2014). Enabling health monitoring approach based on vibration data for accurate prognostics. IEEE Transactions on Industrial Electronics, 62(1), 647-656. https://doi.org/10.1109/TIE.2014.2327917

[31] Li, X., Zhang, W., & Ding, Q. (2019). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliability Engineering & System Safety, 182, 208-218. https://doi.org/10.1016/j.ress.2018.11.011

[32] Song, Y., Gao, S., Li, Y., Jia, L., Li, Q., & Pang, F. (2020). Distributed attention-based temporal convolutional network for remaining useful life prediction. IEEE Internet of Things Journal, 8(12), 9594-9602. https://doi.org/10.1109/JIOT.2020.3004452

[33] Wang, X., Wang, T., Ming, A., Zhang, W., Li, A., & Chu, F. (2021). Spatiotemporal non-negative projected convolutional network with bidirectional NMF and 3DCNN for remaining useful life estimation of bearings. Neurocomputing, 450, 294-310. https://doi.org/10.1016/j.neucom.2021.04.048

[34] Ma, M., & Mao, Z. (2020). Deep-convolution-based LSTM network for remaining useful life prediction. IEEE Transactions on Industrial Informatics, 17(3), 1658-1667. https://doi.org/10.1109/TII.2020.2991796

[35] Sun, J., Zhang, X., & Wang, J. (2023). Lightweight bidirectional long short-term memory based on automated model pruning with application to bearing remaining useful life prediction. Engineering Applications of Artificial Intelligence, 118, 105662. https://doi.org/10.1016/j.engappai.2022.105662

[36] Chen, X., & Liu, Z. (2022). A long short-term memory neural network based Wiener process model for remaining useful life prediction. Reliability Engineering & System Safety, 226, 108651. https://doi.org/10.1016/j.ress.2022.108651

[37] 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. https://doi.org/10.1016/j.ymssp.2019.106330

[38] Zhu, J., Chen, N., & Peng, W. (2018). Estimation of bearing remaining useful life based on multiscale convolutional neural network. IEEE Transactions on Industrial Electronics, 66(4), 3208-3216. https://doi.org/10.1109/TIE.2018.2844856

[39] An, Q., Tao, Z., Xu, X., El Mansori, M., & Chen, M. (2020). A data-driven model for milling tool remaining useful life prediction with convolutional and stacked LSTM network. Measurement, 154, 107461. https://doi.org/10.1016/j.measurement.2019.107461

[40] Xia, T., Song, Y., Zheng, Y., Pan, E., & Xi, L. (2020). An ensemble framework based on convolutional bi-directional LSTM with multiple time windows for remaining useful life estimation. Computers in Industry, 115, 103182. https://doi.org/10.1016/j.compind.2019.103182

[41] Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., & Varnier, C. (2012). PRONOSTIA: An experimental platform for bearings accelerated degradation tests. Paper presented at the IEEE International Conference on Prognostics and Health Management, PHM'12. https://hal.science/hal-00719503/

[42] Zhang, Z., Song, W., & Li, Q. (2022). Dual-aspect self-attention based on transformer for remaining useful life prediction. IEEE Transactions on Instrumentation and Measurement, 71, 1-11. https://doi.org/10.1109/TIM.2022.3160561

[43] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

[44] Chang, Y., Li, F., Chen, J., Liu, Y., & Li, Z. (2022). Efficient temporal flow Transformer accompanied with multi-head probsparse self-attention mechanism for remaining useful life prognostics. Reliability Engineering & System Safety, 226, 108701. https://doi.org/10.1016/j.ress.2022.108701

[45] Soualhi, A., Medjaher, K., & Zerhouni, N. (2014). Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. IEEE Transactions on Instrumentation and Measurement, 64(1), 52-62. https://doi.org/10.1109/TIM.2014.2330494

[46] Zhang, H., Xi, X., & Pan, R. (2023). A two-stage data-driven approach to remaining useful life prediction via long short-term memory networks. Reliability Engineering & System Safety, 237, 109332. https://doi.org/10.1016/j.ress.2023.109332

[47] Wan, S., Li, X., Zhang, Y., Liu, S., Hong, J., & Wang, D. (2022). Bearing remaining useful life prediction with convolutional long short-term memory fusion networks. Reliability Engineering & System Safety, 224, 108528. https://doi.org/10.1016/j.ress.2022.108528

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.