

# Remaining Useful Life Prediction Using Attention-LSTM Neural Network of Aircraft Engines

Marouane Dida<sup>1</sup>, Abdelhakim Cheriet<sup>2</sup>, Mourad Belhadj<sup>3</sup>

<sup>1,3</sup> Laboratoire d'Intelligence Artificielle et des Technologies de l'Information (LINATI), Kasdi Merbah University, Ouargla, Algeria  
Computer science and technology of Information Department, Kasdi Merbah University, Ouargla, Algeria  
*dida.marouane@univ-ouargla.dz*

<sup>2</sup> The National School of Artificial Intelligence, Algiers, Algeria  
*abdelhakim.cheriet@ensia.edu.dz*

<sup>3</sup> Les Systèmes experts, l'Imagerie et leurs applications dans l'Ingénierie (LESIA), Mohamed Khider University, Biskra, Algeria  
*belhadj.mourad@univ-ouargla.dz*

## ABSTRACT

Accurate prediction of the Remaining Useful Life (RUL) is essential for the effective implementation of Prognostics and Health Management (PHM) in aerospace, particularly in enhancing aero-engine reliability and forecasting potential failures to reduce maintenance costs and human-related risks.

The NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset, utilized in the 2021 PHM Data Challenge, serves as a widely recognized open-source benchmark, providing simulated turbofan engine data collected under realistic flight conditions. Previous deep learning approaches have leveraged this dataset to predict the remaining useful life of engine units.

However, data-driven methods for RUL prediction in aerospace often encounter challenges such as high model complexity, limited prediction accuracy, and reduced interpretability. To address these issues, this paper presents a novel hybrid framework that incorporates an attention mechanism to enhance aircraft engine RUL prognostics. Specifically, we employ a self-attention mechanism to effectively capture relationships and interactions among different features, enabling the transformation of high-dimensional feature spaces into lower-dimensional representations.

The proposed model, which integrates an LSTM network, demonstrates superior performance in predicting turbofan en-

gine RUL. Experimental results validate its effectiveness, achieving RMSE values of 12.33 and 11.76, along with score values of 200 and 212 on the FD001 and FD003 sub-datasets, respectively. These results surpass those of other state-of-the-art methods on the C-MAPSS dataset.

## 1. INTRODUCTION

Recent advancements in communication, sensor, and computing technologies have significantly improved access to sensor data from industrial machines, driving the Industry 4.0 revolution and fostering the rise of smart manufacturing (Tao, Qi, Liu, & Kusiak, 2018). The potential of predictive maintenance to enhance productivity and reduce maintenance costs has attracted substantial interest over the past decade, establishing it as a key focus within the modern industrial era (Lee, Ardakani, Yang, & Bagheri, 2015). Predicting the Remaining Useful Life (RUL) of a system is fundamental to all predictive maintenance applications (Lei et al., 2018).

Accurate RUL prediction enables proactive maintenance scheduling, effectively reducing costs while preventing system performance deterioration and catastrophic failures (Han et al., 2021). Deep learning methods have demonstrated superior performance in RUL prediction by autonomously extracting hierarchical representations from training samples, thereby eliminating the need for manual feature engineering and minimizing human resource consumption (Rohani Bastami, Aasi, & Arghand, 2019). The precise prediction of RUL, facilitated by the comprehensive utilization of extensive monitoring data, significantly enhances the safety and reliability of system operations.

Marouane Dida et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.  
<https://doi.org/10.36001/ijphm.2025.v16i2.4274>

Throughout its extended operational lifespan, an aero-engine undergoes gradual performance degradation due to factors such as fatigue, corrosion, and scale accumulation. While the specific variables influencing the degradation of rotating components vary, their combined impact alters the engine's efficiency and flow capacity. These performance indicators are collectively referred to as health parameters (Kurz & Brun, 2001). Although direct measurement of health parameters is often challenging, sensor data such as temperature, pressure, rotor speed, and fuel flow rate can be readily quantified and serves as the primary input for various advanced deep learning models used in RUL prediction. Researchers generally classify RUL prediction approaches into two main categories: physics-based and data-driven (H. Song, Liu, & Song, 2023)(Benkedjouh, Medjaher, Zerhouni, & Rechak, 2013). Due to the complexity of modeling requirements, physics-based approaches can be difficult to implement (An, Kim, & Choi, 2015)(Chao, Kulkarni, Goebel, & Fink, 2022). The rapid advancements in artificial intelligence have led to an increasing preference for data-driven methodologies, particularly deep learning techniques, in RUL prediction research (Tian, Yang, & Ju, 2023)(Xia, Song, Zheng, Pan, & Xi, 2020). Convolutional Neural Networks (CNNs) and their variants have proven effective in handling complex signals and disturbances, thereby improving the reliability of RUL forecasting (X. Li, Ding, & Sun, 2018).

However, CNN-based methodologies have inherent limitations. Their convolutional kernels restrict the range of information they can capture, limiting their ability to effectively learn long-term dependencies in the data. In contrast, Recurrent Neural Networks (RNNs) excel in processing sequential data due to their recurrent architecture, offering advantages in RUL prediction (Hu, Cheng, Wu, Zhu, & Shao, 2021). Although Long Short-Term Memory (LSTM) networks address the vanishing gradient problem encountered in standard RNNs, they still struggle to retain all crucial information across long sequences (T. Song, Liu, Wu, Jin, & Jiang, 2022). Additionally, the requirement for serial computation in these models significantly increases computational overhead (Peng, Jiao, Dong, & Pi, 2019). The limitations of conventional CNNs, RNNs, and their derivatives become more pronounced as the monitored sequence length grows.

While early methods primarily leveraged Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), recent research has shifted toward more sophisticated architectures. Attention-based models, such as the BiGRU-TSAM model proposed by Zhang et al. (J. Zhang et al., 2022), incorporate temporal self-attention mechanisms to selectively focus on time steps with greater prognostic relevance, enhancing interpretability and performance. Similarly, Liu et al. (C. Liu, Zhang, Yao, & Wu, 2021) introduced a dual attention-based temporal convolutional network (DATCN) for Remaining Useful Life (RUL) prediction, demonstrating its

effectiveness in capturing long-term dependencies while maintaining computational efficiency. In addition to performance gains, interpretability has emerged as a crucial concern in industrial AI applications.

In this study, a literature review is presented in Section 2, followed by an overview of the proposed RUL prediction model in Section 3. Section 4 examines the network layers and batch size of the AttnLSTM model to validate its performance using the NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset. In Section 5, the effectiveness and superiority of AttnLSTM are demonstrated through a comparative analysis with other deep learning techniques. Finally, this article concludes in Section 6.

## 2. LITERATURE REVIEW

Data-driven methodologies leverage historical data to evaluate current system health and predict the Remaining Useful Life (RUL) of components. Rather than directly modeling the complexities of wear, material degradation, and mechanical deterioration, these approaches employ artificial intelligence to identify failure patterns and extract critical features from large datasets.

Chen et al. (X. Chen, Shen, He, Sun, & Liu, 2013) introduced an innovative prognostics model that utilizes feature-based comparison and a multivariable support vector machine to estimate the RUL of mechanical components such as bearings, even when dealing with small sample sizes. Similarly, Elforjani et al. (Elforjani & Shanbr, 2017) investigated three supervised learning methods to establish correlations between acoustic emission features and RUL predictions.

Huang et al. (Huang, Liao, Zhang, & Li, 2018) developed a deep coupling convolutional neural network for intelligent diagnosis of compound faults, validating its effectiveness through gearbox malfunction assessments. Zhu et al. (Zhu, Chen, & Peng, 2018) proposed a deep feature learning approach that integrates time-frequency representation with a multi-scale convolutional neural network for RUL prediction, demonstrating its efficacy through experimental validation. Miao et al. (Miao, Li, Sun, & Liu, 2019) designed dual-task deep long short-term memory (LSTM) networks for degradation assessment and RUL prediction in aero-engines, successfully validating their model using the publicly available C-MAPSS lifetime dataset.

As an advanced subset of machine learning, deep learning enables the hierarchical modeling of complex processes and has been widely applied in system prognostics (Sun, Ma, Zhao, & Chen, 2018). Chemali et al. (Chemali, Kollmeyer, Preindl, Ahmed, & Emadi, 2017) designed an LSTM-based recurrent neural network (RNN) to enhance state-of-charge estimation for lithium-ion batteries.

The attention mechanism, inspired by human selective visual

attention, is extensively used in applications such as image captioning, machine translation, and video motion recognition. By allowing models to focus on the most relevant segments of sequential data, attention mechanisms significantly enhance predictive performance. Chen et al. (L. Chen et al., 2017) introduced a convolutional neural network (CNN) architecture incorporating both spatial and channel-wise attention mechanisms for image recognition, achieving outstanding results on benchmark datasets.

An end-to-end spatio-temporal attention model was developed by Song et al. (S. Song, Lan, Xing, Zeng, & Liu, 2017) to recognize and predict human actions in videos. Similarly, Zhai et al. (Zhai, Xiang, Zhang, Lv, & El Saddik, 2019) improved optical flow prediction by integrating a channel attention mechanism with dilated convolutional neural networks, demonstrating enhanced accuracy through experimental results. Additionally, Ran et al. (Ran, Shan, Fang, & Lin, 2019) achieved superior accuracy compared to baseline models by exploring attention-based LSTM architectures for travel time prediction.

Building upon these advancements, this study introduces an attention-based LSTM model specifically designed for predicting the remaining useful life (RUL) of aircraft engines.

### 3. THE PROPOSED METHODOLOGY

The Attention LSTM model, depicted in Figure 1, employs an LSTM layer to extract temporal features. Subsequently, an attention mechanism is integrated to identify and emphasize the most critical information within these extracted features, enhancing the accuracy of the prediction task. Each component of the model is discussed in detail below.

#### 3.1. Long Short-Term Memory Network (LSTM)

Standard RNNs suffer from several limitations. First, the backpropagation process can lead to vanishing or exploding gradients, hindering effective learning. Second, RNNs often exhibit "fading memory," where information from earlier time steps is lost as the sequence progresses. These issues arise when the number of time steps in the RNN network exceeds the capacity of its connections and layers. To overcome these challenges, LSTM networks were developed as an enhanced variant of standard RNNs.

The Long Short-Term Memory (LSTM) network, introduced by Hochreiter and Schmidhuber in 1997 (Hochreiter & Schmidhuber, 1997), employs a gating mechanism to regulate the flow of information, allowing it to effectively handle long-term dependencies in sequential data. Unlike conventional methods that process all input indiscriminately across time steps, LSTM utilizes specialized gates to selectively retain, update, or discard information. As illustrated in Figure 2, the memory cell consists of three key components: the forget

gate, the input gate, and the output gate, all of which work collaboratively to control information flow within the cell.

The proposed LSTM framework, illustrated in Figure 2, employs a sliding time-window methodology to process sequential input data. This approach segments the input into consecutive windows of fixed size, each containing a subset of features. These windows are then sequentially fed into the LSTM layers. By treating one input dimension as a pseudo-time axis, the architecture enables the LSTM layers to effectively capture temporal dependencies throughout the recurrent training process.

As depicted in Figure 2, the fundamental unit of an LSTM consists of a memory cell, an input gate ( $i_t$ ), an output gate ( $o_t$ ), and a forget gate ( $f_t$ ). At each time step  $t$ , the input  $x_t$  is selectively stored in the memory cell ( $C_t$ ) based on the behavior of the input gate. Simultaneously, the forget gate determines the extent to which the previous cell state ( $C_{t-1}$ ) is retained or discarded. Finally, the output gate regulates which portion of the current cell state ( $C_t$ ) contributes to the output ( $h_t$ ). The operational dynamics of the LSTM layers are governed by the following equations:

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

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \quad (2)$$

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \quad (3)$$

In this framework,  $f_t$ ,  $i_t$ , and  $o_t$  represent the forget gate, input gate, and output gate, respectively. The sigmoid activation function, denoted by  $\sigma$  and utilized within the LSTM cell (as illustrated in Figure 2), constrains the output values to the range  $[0, 1]$ . Additionally,  $W_i$ ,  $W_f$ , and  $W_o$  denote the weight matrices associated with the gates, while  $h_{t-1}$  represents the output from the previous time step,  $x_t$  is the current input, and  $b_i$ ,  $b_f$ , and  $b_o$  are the respective bias vectors for each gate.

The previous cell state,  $C_{t-1}$ , undergoes transformation into the updated cell state,  $C_t$ , as depicted in Figure 2. This transformation is governed by the following mathematical expression:

$$C_t = f_t \otimes C_{t-1} + i_t * (\tanh(W_c * [h_{t-1}, x_t] + b_c)) \quad (4)$$

Here,  $W_c$  represents the weight matrix,  $b_c$  denotes the bias vector, and  $C_{t-1}$  corresponds to the state of the previous cell.

The updated hidden state,  $h_t$ , is obtained by applying the hyperbolic tangent activation function to the current cell state, modulated by the output of the output gate (as shown in Fig-

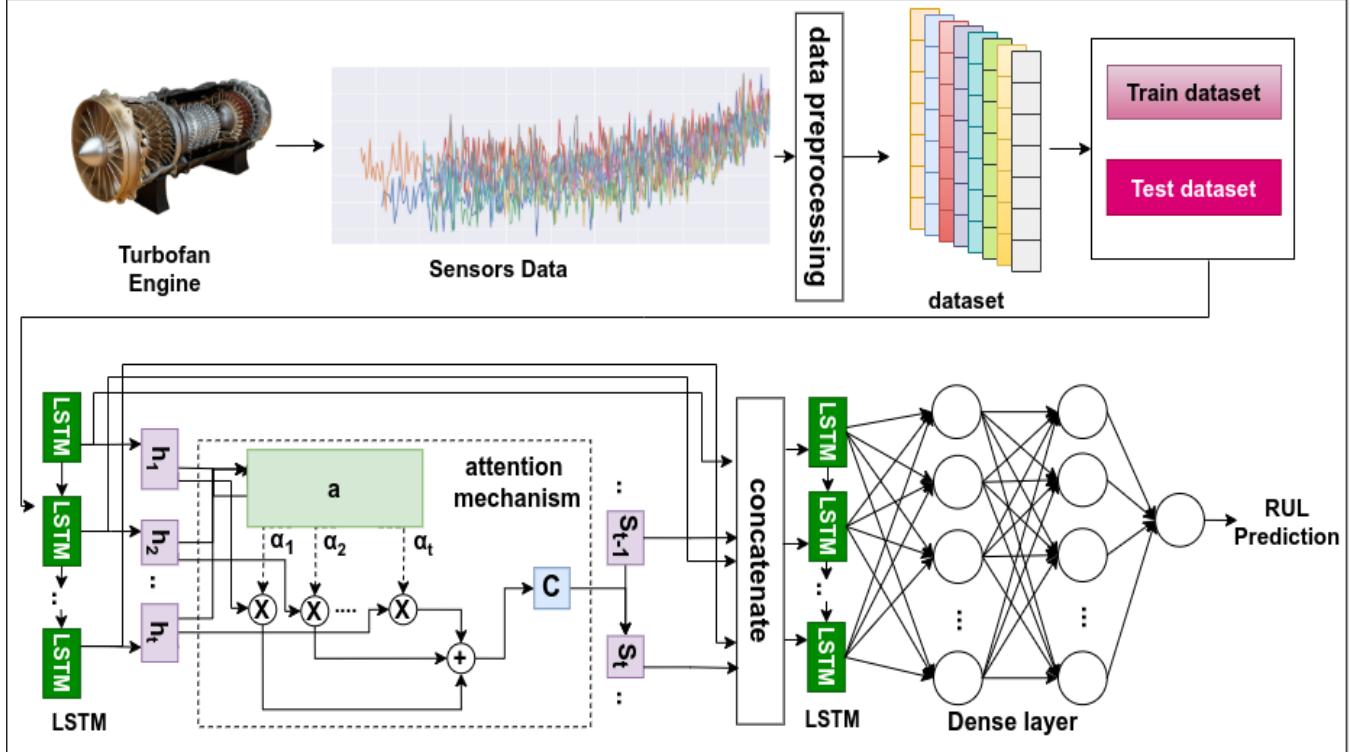


Figure 1. Architecture of the proposed method

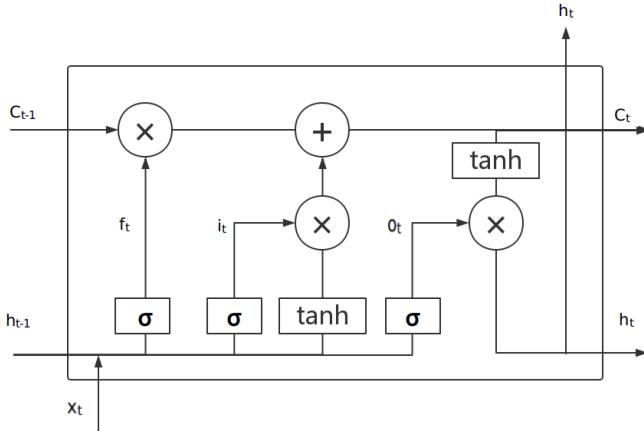


Figure 2. LSTM memory cell

ure 2). This process can be mathematically expressed as:

$$h_t = o_t \otimes \tanh(C_t) \quad (5)$$

### 3.2. Attention mechanisms

The Attention mechanism, a recent advancement, has transformed multiple fields where it has been applied (Vaswani, 2017). Originally introduced in natural language processing (NLP), it achieved remarkable success, leading to its widespread adoption across various domains.

To effectively capture temporal dependencies in time-series data, an attention mechanism is integrated, as illustrated in Figure 3. Here,  $S_t$  represents the LSTM hidden state at time step  $t$ . As information propagates through the LSTM cells, the sequence of hidden state vectors ( $h_t$ ) is processed through a learnable function ( $a$ ) to compute a set of attention weights ( $\alpha_i$ ). The context vector ( $C$ ) is then obtained by computing a weighted sum of the hidden state vectors ( $h_t$ ) using their corresponding attention weights ( $\alpha_i$ ) (Raffel & Ellis, 2015; Wu, Wang, Li, & Gao, 2018).

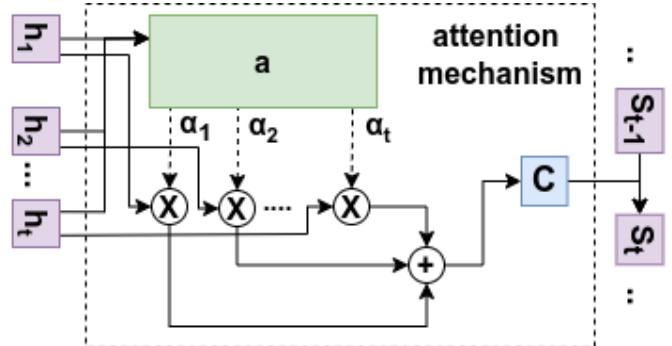


Figure 3. Attention mechanism

LSTM is utilized to effectively capture dependencies between past and future information. However, to mitigate potential information loss within the LSTM, an attention mechanism

is integrated. This mechanism enhances the model's ability to focus by assigning greater significance to critical information while reducing the influence of less relevant data. Notably, the attention mechanism operates at the neural unit level rather than directly on the input features, which is the rationale behind its naming convention. The functionality of this attention mechanism can be succinctly represented through the following three equations:

$$C = \sum_{i=1}^t \alpha_i h_i \quad (6)$$

$$\alpha_i = \exp(e_t) / \sum_{k=1}^t \exp(e_k) \quad (7)$$

$$e_i = a(h_i) = \sigma(Wh_i + b) \quad (8)$$

In this context,  $t$  represents the total number of time steps in the input sequence to the LSTM network. The vector  $h$  denotes the eigenvector output generated by the LSTM network. The scalar  $\alpha_i$  corresponds to the weight assigned to the vector  $h$ . The symbol  $\sigma$  represents the activation function, while  $W$  denotes the weight matrix that connects the input layer to the hidden layer.

### 3.3. Evaluation Metrics

To assess the performance of the models, we utilize two widely recognized evaluation metrics: the root mean squared error (RMSE) and the scoring algorithm (Score) proposed by the National Aeronautics and Space Administration (NASA) (Saxena, Goebel, Simon, & Eklund, 2008).

RMSE is a standard metric in regression tasks and is frequently employed to evaluate the accuracy of Remaining Useful Life (RUL) predictions. It is computed using the following formula:

$$RMSE = \sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (9)$$

In the equation above,  $n$  denotes the total number of predicted samples, while  $\hat{y}_i$  and  $y_i$  represent the predicted value and the actual Remaining Useful Life (RUL) for sample  $i$ , respectively.

While RMSE treats early and late predictions symmetrically, late predictions can have far more severe consequences in real-world applications, potentially leading to catastrophic failures. To address this limitation, Saxena et al. (Saxena et al., 2008) proposed a scoring function that has since become a widely accepted evaluation metric for industrial RUL prediction. This metric has also been integrated into the Inter-

national Conference on Prognostics and Health Management Data Challenge (X. Li et al., 2018). The present study adopts this scoring function, which is formally defined as follows:

$$Score = \begin{cases} \sum_{i=1}^n \exp^{\frac{\hat{y}_i - y_i}{10}} - 1 & : \hat{y}_i \geq y_i \\ \sum_{i=1}^n \exp^{\frac{y_i - \hat{y}_i}{13}} - 1 & : \hat{y}_i < y_i \end{cases} \quad (10)$$

In this equation,  $n$  represents the total number of test engines,  $\hat{y}_i$  denotes the predicted Remaining Useful Life (RUL), and  $y_i$  indicates the actual RUL for the final sample of engine  $i$ . Notably, the scoring function applies a more substantial penalty for late predictions.

## 4. EXPERIMENTAL SETUP

This section presents an in-depth analysis of the results obtained from the proposed RUL prediction model. The model's performance is evaluated by assessing the influence of various factors, such as the number of hidden layers and the length of the time window. Additionally, comparative analyses with other widely adopted neural network architectures are conducted to highlight the effectiveness of the proposed design.

All experiments were implemented in Python using the Keras and TensorFlow libraries. The computations were performed on a workstation equipped with a Ryzen 5 5600 CPU, 16GB of RAM, and an RTX 3060 GPU. To mitigate the effects of randomness, each experiment was repeated 10 times, and the results were averaged, with both mean values and standard deviations reported. The C-MAPSS dataset served as the benchmark for all RUL prediction experiments.

### 4.1. Dataset Description

The Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset is a widely utilized benchmark dataset developed by NASA for research in prognostics and health management (PHM) of aircraft engines. It has been referenced in over 28% of related studies (Ferreira & Gonçalves, 2022). This dataset provides simulated operational data of turbofan engines up to the point of failure, making it an essential resource for developing predictive maintenance models, particularly for Remaining Useful Life (RUL) prediction.

Table 1. C-MAPSS datasets for turbofan engines

| Column contents   | FD001 | FD002 | FD003 | FD004 |
|-------------------|-------|-------|-------|-------|
| training engines  | 100   | 260   | 100   | 249   |
| testing engines   | 100   | 259   | 100   | 248   |
| Condition number  | 1     | 6     | 1     | 6     |
| Fault mode number | 1     | 1     | 2     | 2     |

The C-MAPSS dataset consists of multiple subdatasets, each representing different operating conditions and fault modes, as detailed in Table 1. Each sample comprises 26 values, in-

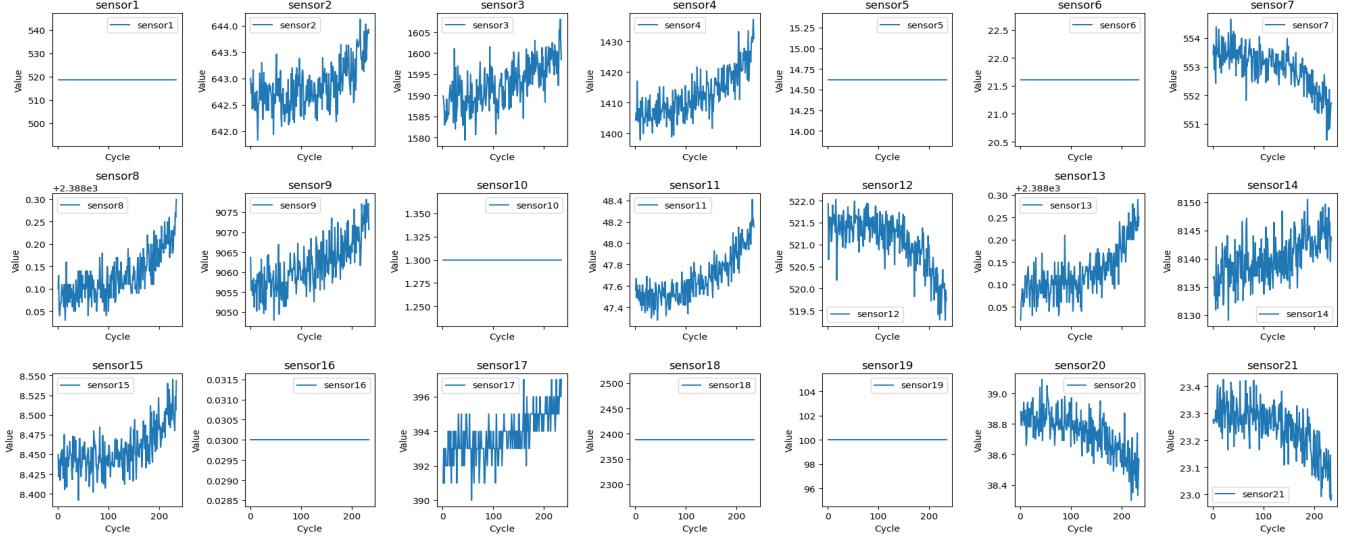


Figure 4. sensors data of the turbofan engines in FD001.

cluding the cycle number, unit ID, 21 sensor readings, and three operating conditions. The sensor data captures various parameters such as temperature, pressure, and rotational speed, as described in (Saxena et al., 2008). This dataset simulates aircraft engine performance under realistic flight conditions, accurately reflecting degradation patterns resulting from component wear and environmental fluctuations.

Certain sensors fail to provide meaningful data for monitoring engine health. For example, in the case of turbofan engine sensor data from FD001, as illustrated in Figure 4, the readings from sensors numbered 1, 5, 6, 10, 16, 18, and 19 remain constant throughout the cycles. Consequently, these sensors are excluded from the analysis, and the remaining 14 sensor readings are selected as raw features for data modeling, consistent with prior research (Z. Liu, Liu, Jia, Zhang, & Tan, 2021).

#### 4.2. Sliding Time Window Processing

Effective training of models on sequential sensor data requires comprehensive pre-processing. Directly training on raw data for regression or other predictive tasks may fail to capture crucial short-term dependencies due to the high dimensionality and temporal nature of sequential data. The Time Window Processing technique addresses this challenge by preserving associative relationships, enhancing feature extraction, and ultimately improving the algorithm's performance.

Figure 5 shows sliding time window of length three with a stride size of one is employed. This simplified example is provided for illustrative purposes only. The sliding window advances one step along the time axis to generate the next set of windowed data. This process continues iteratively until the sliding window reaches the end of the original sequence. The

Remaining Useful Life (RUL) associated with the final data point in each time window is assigned as the RUL for that window.

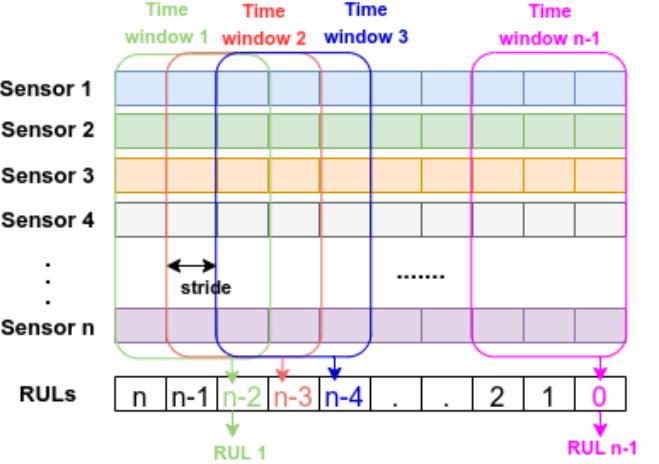


Figure 5. Sliding time window processing

A window size of 30 was chosen for this study. For sub-datasets FD001 and FD003, the window size was determined based on the engine with the shortest data sequence, ensuring uniformity across the training and testing datasets for each engine. The stride for the time windowing process was set to one. In the initial stages, RUL labels were assigned a constant value of 125. Engines in sub-datasets FD002 and FD004 that contained fewer than 30 samples were excluded from the study. Table 2 presents the dimensions and significance of the training and testing datasets after applying the time windowing procedure.

Table 2. C-MAPSS datasets for turbofan engines

| Dataset   | Train           | Test          |
|---|-----------------|---------------|
| Input (simple num, window size, engine feature) FD001 | (17731, 30, 14) | (100, 30, 14) |
| Output (simple num, RUL) FD001                        | (17731, 1)      | (100, 1)      |
| Input (simple num, window size, engine feature) FD003 | (21820, 30, 14) | (100, 30, 14) |
| Output (simple num, RUL) FD003                        | (21820, 1)      | (100, 1)      |

### 4.3. Data Normalization

Data preprocessing, including normalization and noise reduction, is a crucial step before model training. Normalization was applied to accelerate model convergence and prevent features with larger numerical values from dominating the training process, which could adversely affect performance. The normalization process is defined by Eq. (11).

$$x_{norm} = \frac{x - min(x)}{max(x) - min(x)} \quad (11)$$

The *max* and *min* functions represent the mathematical operations used to determine the highest and lowest values, respectively.

### 4.4. The Proposed Model Architecture

LSTM networks with attention mechanisms are employed for Remaining Useful Life (RUL) prediction due to their capability to effectively capture temporal dependencies in sequential data.

The LSTM architecture consists of two layers, each comprising 50 hidden units. The Adam optimizer is utilized with a learning rate of 0.001 and a batch size of 32. To mitigate overfitting during training, an early stopping mechanism is implemented, which halts the training process once the model begins to overfit the training data. The network architecture of the proposed model is illustrated in Figure 6.

The final values for hidden size, batch size, learning rate were selected based on a grid search and empirical validation on a held-out portion of the training data.

## 5. RESULTS

This section evaluates the performance of the proposed model using RMSE and scoring functions. The results are compared with those of existing models, and the findings are thoroughly analyzed.

### 5.1. Experimental Results

The training subsets comprise full run-to-failure data, while the testing subsets are truncated before the occurrence of any faults. Once trained, the AttnLSTM model was subsequently utilized to predict the Remaining Useful Life (RUL) using the C-MAPSS testing datasets.

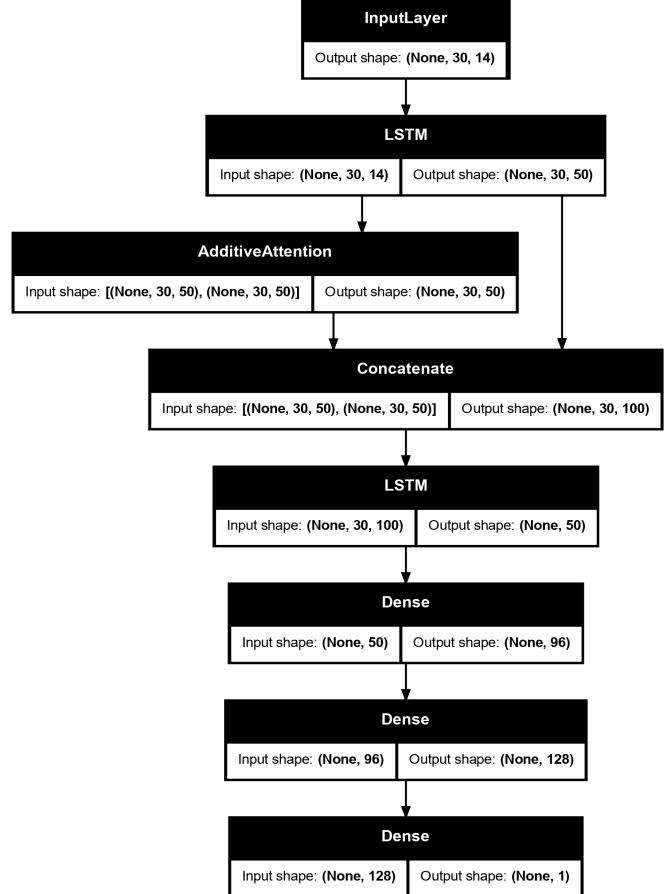


Figure 6. architecture of the model AM-LSTM

Figures 7 and 8 presents the RUL predictions for engines in the testing sub-datasets FD001 and FD003. In the figure, the x-axis represents the number of samples, while the y-axis denotes the remaining useful life. The green line corresponds to the actual RUL, whereas the orange line illustrates the predicted RUL.

The degree of alignment between the green line (actual RUL) and the orange line (predicted RUL) in Figures 7 and 8 serves as an indicator of the model's accuracy. The figure demonstrates that the AttnLSTM method yields more precise predictions, particularly as the system approaches failure. This improved accuracy empowers maintenance teams to schedule interventions proactively, thereby reducing the risk of unexpected system failures.



Figure 7. The results of the RUL predictions for the C-MAPSS sub-dataset FD001

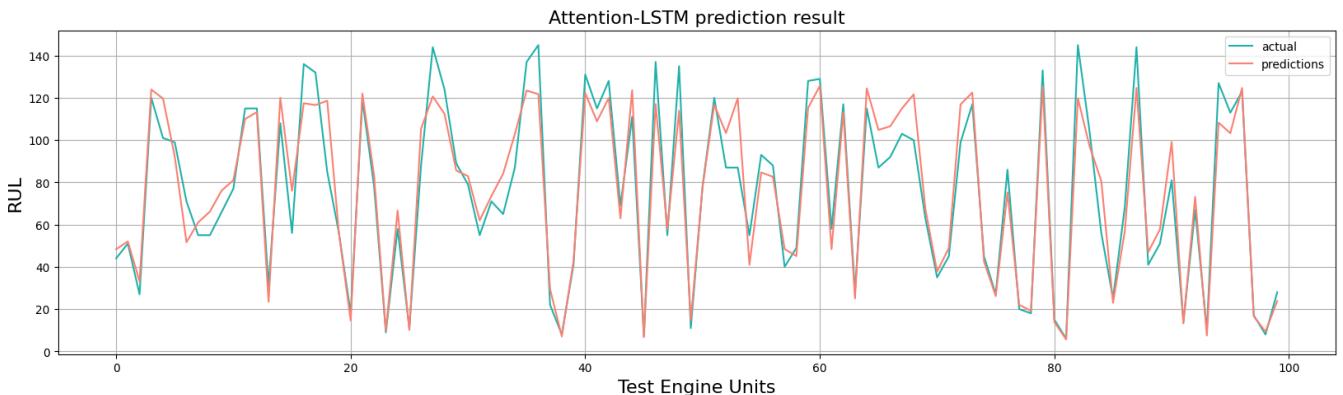


Figure 8. The results of the RUL predictions for the C-MAPSS sub-dataset FD003

Figures 9 and 10 illustrate the prediction performance for four representative engines (IDs: 31, 100, 78, and 92) from sub-datasets FD001 and FD003. The model successfully predicted the RUL for these engines, showcasing its capability to capture diverse degradation patterns, including gradual wear and early-stage performance decline. These results affirm the effectiveness of integrating domain knowledge into system reliability analysis, further reinforcing the model's practical applicability in predictive maintenance.

## 5.2. Comparisons With State-of-the-Art Methods

This study evaluates the proposed model against other state-of-the-art methods on the C-MAPSS dataset, using RMSE and Score as evaluation metrics. The best results for FD001 and FD003 are highlighted in bold in Table 3, demonstrating the superior performance of the proposed approach. All experiments were carried out on a system equipped with an NVIDIA RTX 3060 GPU, utilizing GPU acceleration to accelerate neural network training. Using more advanced GPUs could further reduce the training time of the proposed method.

The table also underscores the optimal performance of our approach on the FD001 and FD003 sub-datasets, reaffirming

its effectiveness in terms of score evaluation. The AttnLSTM model, based on the LSTM architecture, was specifically designed to accommodate the characteristics of multivariate time series sensor data. By incorporating an attention layer that employs dot product operations to compute feature similarity and assign appropriate weights, the model surpasses traditional fully connected layers in capturing degradation patterns. This attention mechanism improves the model's sensitivity to feature interactions and variations, allowing it to extract degradation information more effectively.

However, the spatial complexity of the algorithm poses a challenge, particularly for larger and higher-dimensional datasets. Future research should focus on optimizing the spatial complexity of attention-based algorithms.

Overall, this study underscores the potential of attention mechanisms for RUL prediction and suggests that such models hold great promise for addressing time-series problems.

## 5.3. Discussion

The experimental results presented in this study demonstrate that the proposed Attention-LSTM (AttnLSTM) architecture

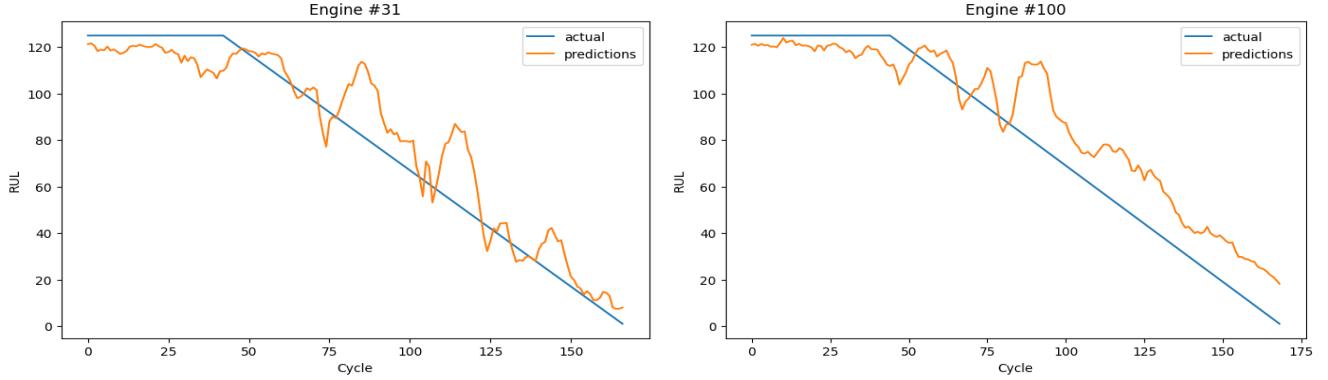


Figure 9. Prediction of test engines FD001.

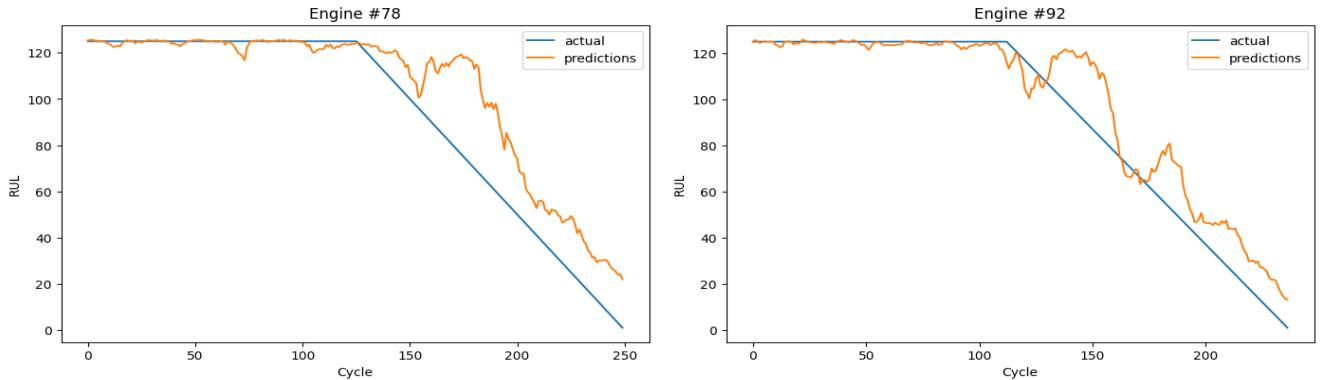


Figure 10. Prediction of test engines FD003.

Table 3. Comparison with the methods proposed in the literature on NASA C-MAPSS using score and RMSE.

| Models   | FD001        |            | FD003        |            |
|--|--------------|------------|--------------|------------|
|  | RMSE         | Score      | RMSE         | Score      |
| CNN(Sateesh Babu, Zhao, & Li, 2016)              | 18.45        | 1287       | 19.82        | 1596       |
| DCNN(X. Li et al., 2018)                         | 12.61        | 274        | 12.64        | 284        |
| LSTM(C. Chen, Shi, Lu, Zhu, & Jiang, 2022)       | 16.10        | 338        | 16.20        | 852        |
| AGCNN(H. Liu, Liu, Jia, & Lin, 2020)             | 12.40        | 226        | 13.40        | 227        |
| MODBNE(C. Zhang, Lim, Qin, & Tan, 2016)          | 15.00        | 334        | 12.50        | 422        |
| MLP(C. Zhang et al., 2016)                       | 16.80        | 561        | 18.50        | 480        |
| CNN-LSTM(Marouane, Belhadj, & Cheriet, 2024)     | 13.92        | 383        | 19.83        | 1099       |
| CATA-TCN(Lin et al., 2024)                       | 12.80        | 234        | 13.16        | 290        |
| AttnBiLSTM(Shah, Chadha, Schwung, & Ding, 2021)  | 15.87        | 473        | 15.10        | 676        |
| DBRNN-SLF(Hu et al., 2021)                       | 17.97        | 458        | 19.18        | 658        |
| AttnPINN(Liao, Chen, Wen, & Zhao, 2023)          | 16.89        | 523        | 17.75        | 1194       |
| SAM-CNN-LSTM(J. Li, Jia, Niu, Zhu, & Meng, 2023) | 12.60        | 261        | 13.80        | 253        |
| BiGRU-TSAM(J. Zhang et al., 2022)                | 12.56        | 213        | 12.45        | 232        |
| <b>AttnLSTM</b>                                  | <b>12.33</b> | <b>200</b> | <b>11.76</b> | <b>212</b> |

consistently outperforms baseline models and recent state-of-the-art approaches on the FD001 and FD003 subsets of the C-MAPSS benchmark. These gains are not coincidental but stem from a rigorously engineered architecture that captures both the temporal dynamics and contextual nuances inherent in degradation processes. At its core, the model integrates Long Short-Term Memory (LSTM) networks with

a learnable attention mechanism, enabling it to simultaneously model long-range dependencies and selectively emphasize the most informative temporal and feature-level inputs.

The LSTM component plays a foundational role by modeling non-stationary time-series signals generated by engines operating under diverse conditions. However, conventional

LSTMs treat all time steps and input features equally when summarizing sequential information, which is suboptimal in prognostics where degradation-relevant cues especially those closer to failure carry disproportionately greater predictive value. To overcome this limitation, the AttnLSTM introduces an attention mechanism positioned atop LSTM layers, which dynamically computes importance weights over both temporal and feature dimensions. This dual focus allows the model to amplify salient sensor readings and attenuate noise or redundant signals, functioning as an embedded form of adaptive feature selection.

Empirical correlation analysis revealed that sensor contributions to RUL prediction are not uniform, and the attention module capitalizes on this by learning to prioritize high-impact features during training. The resulting attention weights are not only functionally beneficial enhancing predictive accuracy and model stability but also interpretable, providing insight into which sensors and time steps are most critical under varying operational regimes. This interpretability bridges the gap between deep learning and domain-informed decision-making, offering actionable knowledge for engineers in sensor design, placement, and maintenance planning.

Crucially, the model maintains an efficient architecture two LSTM layers with 50 units each optimized using the Adam optimizer and regularized via early stopping. The input pipeline employs a dense sliding window (size 30, stride 1) and utilizes 14 preselected sensors based on prior studies, ensuring that the most informative variables are retained while minimizing overfitting. Despite its compactness, the model exhibits remarkable performance, achieving the lowest RMSE and NASA score metrics across multiple runs, with notably reduced variance indicating not only accuracy but robustness.

Furthermore, unlike heavier transformer-based approaches, AttnLSTM was designed with deployment in mind. Its computational size is small enough for real-time prediction in resource-constrained environments, such as aircraft maintenance systems. The transparency offered by attention visualization also meets a critical need in safety-critical domains, where explainability is paramount for operational trust and regulatory compliance. The proposed AttnLSTM model was evaluated for prediction efficiency on a standard high-performance computing platform (Ryzen 5600 processor, 16 GB RAM, and NVIDIA RTX 3060 GPU). The prediction time in engine cycles is approximately 3 ms per step, meeting the typical real-time requirements of predictive aviation health management systems. The model architecture contains approximately 61,000 trainable parameters, significantly fewer than transformer-based models, which typically exceed 10 million, reducing memory usage and latency. These properties confirm that AttnLSTM is suitable for immediate deployment in resource-constrained environments.

In summary, the proposed AttnLSTM model advances the

field of RUL prediction by offering a domain-aligned, interpretable, and computationally efficient architecture. Its performance stems from a thoughtful integration of temporal modeling, attention-based feature relevance estimation, and principled engineering design resulting in a scalable and practically deployable solution for predictive maintenance in real-world aerospace systems.

## 6. CONCLUSIONS AND FUTURE WORK

Predictive maintenance plays a crucial role in optimizing equipment maintenance schedules by minimizing unnecessary maintenance, reducing costs, and enhancing overall operational efficiency. A key objective of predictive maintenance is to accurately forecast equipment failures, particularly by estimating the Remaining Useful Life (RUL). By predicting potential failures in advance, maintenance planning can be significantly improved. This study introduces a novel deep learning-based model for predicting the RUL of turbofan jet engines.

The effectiveness of the proposed model was evaluated using the FD001 and FD003 sub-datasets from the C-MAPSS dataset. Compared to other state-of-the-art models, our approach demonstrated superior performance based on the evaluation score for these sub-datasets. These specific sub-datasets were chosen because they allow training the proposed model without necessitating structural modifications.

Future research will focus on extending the proposed model to achieve enhanced results on additional test datasets within the scope of this study. Additionally, upcoming work will explore integrating insights from other relevant literature to refine and further enhance the proposed model's architecture.

## REFERENCES

- An, D., Kim, N. H., & Choi, J.-H. (2015). Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliability Engineering & System Safety*, 133, 223–236.
- 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.
- Chao, M. A., Kulkarni, C., Goebel, K., & Fink, O. (2022). Fusing physics-based and deep learning models for prognostics. *Reliability Engineering & System Safety*, 217, 107961.
- Chemali, E., Kollmeyer, P. J., Preindl, M., Ahmed, R., & Emadi, A. (2017). Long short-term memory networks for accurate state-of-charge estimation of li-ion batteries. *IEEE Transactions on Industrial Electronics*, 65(8), 6730–6739.
- Chen, C., Shi, J., Lu, N., Zhu, Z. H., & Jiang, B. (2022).

- Data-driven predictive maintenance strategy considering the uncertainty in remaining useful life prediction. *Neurocomputing*, 494, 79–88.
- Chen, L., Zhang, H., Xiao, J., Nie, L., Shao, J., Liu, W., & Chua, T.-S. (2017). Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 5659–5667).
- Chen, X., Shen, Z., He, Z., Sun, C., & Liu, Z. (2013). Remaining life prognostics of rolling bearing based on relative features and multivariable support vector machine. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 227(12), 2849–2860.
- Elforjani, M., & Shanbr, S. (2017). Prognosis of bearing acoustic emission signals using supervised machine learning. *IEEE Transactions on industrial electronics*, 65(7), 5864–5871.
- Ferreira, C., & Gonçalves, G. (2022). Remaining useful life prediction and challenges: A literature review on the use of machine learning methods. *Journal of Manufacturing Systems*, 63, 550–562.
- Han, X., Wang, Z., Xie, M., He, Y., Li, Y., & Wang, W. (2021). Remaining useful life prediction and predictive maintenance strategies for multi-state manufacturing systems considering functional dependence. *Reliability Engineering & System Safety*, 210, 107560.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
- Hu, K., Cheng, Y., Wu, J., Zhu, H., & Shao, X. (2021). Deep bidirectional recurrent neural networks ensemble for remaining useful life prediction of aircraft engine. *IEEE Transactions on Cybernetics*, 53(4), 2531–2543.
- Huang, R., Liao, Y., Zhang, S., & Li, W. (2018). Deep de-coupling convolutional neural network for intelligent compound fault diagnosis. *Ieee Access*, 7, 1848–1858.
- Kurz, R., & Brun, K. (2001). Degradation in gas turbine systems. *J. Eng. Gas Turbines Power*, 123(1), 70–77.
- Lee, J., Ardakani, H. D., Yang, S., & Bagheri, B. (2015). Industrial big data analytics and cyber-physical systems for future maintenance & service innovation. *Procedia cirp*, 38, 3–7.
- Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to rul prediction. *Mechanical systems and signal processing*, 104, 799–834.
- Li, J., Jia, Y., Niu, M., Zhu, W., & Meng, F. (2023). Remaining useful life prediction of turbofan engines using cnn-lstm-sam approach. *IEEE Sensors Journal*, 23(9), 10241–10251.
- Li, X., Ding, Q., & Sun, J.-Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, 172, 1–11.
- Liao, X., Chen, S., Wen, P., & Zhao, S. (2023). Remaining useful life with self-attention assisted physics-informed neural network. *Advanced Engineering Informatics*, 58, 102195.
- Lin, L., Wu, J., Fu, S., Zhang, S., Tong, C., & Zu, L. (2024). Channel attention & temporal attention based temporal convolutional network: A dual attention framework for remaining useful life prediction of the aircraft engines. *Advanced Engineering Informatics*, 60, 102372.
- Liu, C., Zhang, L., Yao, R., & Wu, C. (2021). Dual attention-based temporal convolutional network for fault prognosis under time-varying operating conditions. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–10.
- Liu, H., Liu, Z., Jia, W., & Lin, X. (2020). Remaining useful life prediction using a novel feature-attention-based end-to-end approach. *IEEE Transactions on Industrial Informatics*, 17(2), 1197–1207.
- 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.
- Marouane, D., Belhadj, M., & Cheriet, A. (2024). Prediction of the remaining useful lifetime of turbofan aircraft engines using machine learning: A comparative study. In *2024 4th international conference on embedded & distributed systems (edis)* (pp. 137–142).
- Miao, H., Li, B., Sun, C., & Liu, J. (2019). Joint learning of degradation assessment and rul prediction for aero-engines via dual-task deep lstm networks. *IEEE Transactions on Industrial Informatics*, 15(9), 5023–5032.
- Peng, K., Jiao, R., Dong, J., & Pi, Y. (2019). A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter. *Neurocomputing*, 361, 19–28.
- Raffel, C., & Ellis, D. P. (2015). Feed-forward networks with attention can solve some long-term memory problems. *arXiv preprint arXiv:1512.08756*.
- Ran, X., Shan, Z., Fang, Y., & Lin, C. (2019). An lstm-based method with attention mechanism for travel time prediction. *Sensors*, 19(4), 861.
- Rohani Bastami, A., Aasi, A., & Arghand, H. A. (2019). Estimation of remaining useful life of rolling element bearings using wavelet packet decomposition and artificial neural network. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, 43, 233–245.
- Sateesh Babu, G., Zhao, P., & Li, X.-L. (2016). Deep convolutional neural network based regression approach for estimation of remaining useful life. In *Database systems for advanced applications: 21st international conference, dasfaa 2016, dallas, tx, usa, april 16-19, 2016, proceedings, part i* 21 (pp. 214–228).
- Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008).

- Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management* (pp. 1–9).
- Shah, S. R. B., Chadha, G. S., Schwung, A., & Ding, S. X. (2021). A sequence-to-sequence approach for remaining useful lifetime estimation using attention-augmented bidirectional lstm. *Intelligent Systems with Applications*, 10, 200049.
- Song, H., Liu, X., & Song, M. (2023). Comparative study of data-driven and model-driven approaches in prediction of nuclear power plants operating parameters. *Applied Energy*, 341, 121077.
- Song, S., Lan, C., Xing, J., Zeng, W., & Liu, J. (2017, Feb.). An end-to-end spatio-temporal attention model for human action recognition from skeleton data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).
- Song, T., Liu, C., Wu, R., Jin, Y., & Jiang, D. (2022). A hierarchical scheme for remaining useful life prediction with long short-term memory networks. *Neurocomputing*, 487, 22–33.
- Sun, C., Ma, M., Zhao, Z., & Chen, X. (2018). Sparse deep stacking network for fault diagnosis of motor. *IEEE Transactions on Industrial Informatics*, 14(7), 3261–3270.
- Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. *Journal of Manufacturing Systems*, 48, 157–169.
- Tian, H., Yang, L., & Ju, B. (2023). Spatial correlation and temporal attention-based lstm for remaining useful life prediction of turbofan engine. *Measurement*, 214, 112816.
- Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- Wu, L., Wang, Y., Li, X., & Gao, J. (2018). Deep attention-based spatially recursive networks for fine-grained visual recognition. *IEEE transactions on cybernetics*, 49(5), 1791–1802.
- Xia, T., Song, Y., Zheng, Y., Pan, E., & Xi, L. (2020). An ensemble framework based on convolutional bidirectional lstm with multiple time windows for remaining useful life estimation. *Computers in Industry*, 115, 103182.
- Zhai, M., Xiang, X., Zhang, R., Lv, N., & El Saddik, A. (2019). Optical flow estimation using channel attention mechanism and dilated convolutional neural networks. *Neurocomputing*, 368, 124–132.
- Zhang, C., Lim, P., Qin, A. K., & Tan, K. C. (2016). Multi-objective deep belief networks ensemble for remaining useful life estimation in prognostics. *IEEE transactions on neural networks and learning systems*, 28(10), 2306–2318.
- Zhang, J., Jiang, Y., Wu, S., Li, X., Luo, H., & Yin, S. (2022). Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism. *Reliability Engineering & System Safety*, 221, 108297.
- 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.