Sustainable Machine Intelligence Journal

Journal Homepage: sciencesforce.com/smij

Sustain. Mach. Intell. J. Vol. 11 (2025) 80–94

Paper Type: Original Article

An Attention-Based Deep Learning Model for Remaining Useful Life Prediction of Aero-Engine

Ahmed Darwish [1,*], Karam M. Sallam [2,3] and Ibrahim Alrashdi [4]

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

[2] School of IT and Systems, Faculty of Science and Technology, University of Canberra, Canberra, Australia; karam.sallam@canberra.edu.au.

[3] Department of Computer Science, University of Sharjah, Sharjah, United Arab Emirates; Ksallam@sharjah.ac.ae.

[4] Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakaka, 2014, Saudi Arabia; irrashdi@ju.edu.sa.

Received: 28 Nov 2024 **Revised:** 28 Jan 2025 **Accepted:** 12 May 2025 **Published:** 16 May 2025

Abstract

The prognosis of remaining useful life (RUL) for the aero-engine plays an indispensable role in enhancing the safety of aircraft and diminishing maintenance expenses. Several deep learning (DL) methods have been recently employed to handle this prediction problem due to their adaptable architectures and superior performance in handling the nonlinear characteristics of prediction problems. However, those models still need further improvement to better predict the status of aircraft engines. Therefore, this paper presents a new DL model, dubbed attention residual block-adaptive long short-term memory (ARB-ALSTM), based on integrating the attention mechanism with the long short-term memory network (LSTM) model and residual block to better comprehend the characteristics of this problem, thereby aiding in achieving better prediction. In a more general sense, the attention residual block is responsible for understanding the input dataset and extracting the most effective features, which significantly affect the model's accuracy. Then, those extracted features are given to an LSTM with an adaptive attention mechanism to effectively capture and analyze long-term dependent information. This proposed model is evaluated using the NASA CMAPSS dataset and compared to several DL models to showcase its effectiveness. The experimental findings reveal that ARB-ALSTM is a strong alternative for predicting the RUL of aircraft engines because it could achieve better outcomes than all the compared models.

Keywords: Prognostics and Health Management; Aero-engine Prognosis; Remaining Useful Life; Attention Mechanism.

1 | Introduction

Prognostics and Health Management (PHM) is an essential part of smart manufacturing [1]. PHM is a computational framework that delves into the domain of physical knowledge pertaining to the operation and maintenance of structures, systems, and components (SSCs). Predicting RUL is a significant aspect of PHM that has attracted several researchers over the last few years to present a more effective approach [2]. In literature, RUL prediction techniques have been classified into three categories: knowledge-based methods, data-driven methods, and physical model-based methods [3]. In the physics-based method, previous physical knowledge is used to create physical models for showing the mechanical equipment's degradation process [4,

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

https://doi.org/10.61356/SMIJ.2025.11554

5]. However, this method performs poorly when used to create physical models for several complicated systems [6]. In addition, developing high-fidelity physical models that use trial and error to match experimental outcomes might take several years. To overcome those challenges, knowledge-based methods involve establishing a correlation between an observed operational condition of machinery and a pre-established degraded knowledge database to deduce the RUL [7]. However, knowledge bases in degradation states are often built by experts in a certain field, relying on established rules, known facts, or their personal experiences gained over time by operating equipment [8].

Compared to knowledge-based and physical model-based approaches, deep learning (DL) models are more capable of generalization and do not require expert knowledge. In addition, those models have a high ability to reveal the non-linear characteristics of the gathered sensor data. Some machine learning (ML) and DL techniques proposed for predicting the RUL are briefly reviewed in the rest of this section. In [9], an ML approach based on hybridizing several support vector regression sub-models is proposed for estimating the aircraft engines' RUL. Furthermore, the Bayesian optimization algorithm was used to tune the hyper-parameters of those sub-models to maximize their performance when applied to the remaining useful life (RUL). This approach was assessed using the C-MAPSS dataset and compared to some ML methods to show its effectiveness. The experimental findings show that it could achieve superior performance in 30% of the training samples.

In [10], an ensemble learning algorithm was proposed for tackling this prediction problem. This algorithm was based on integrating several base learners, such as classification and regression tree, autoregressive model, random forest (RF), recurrent neural networks (RNN), elastic net, and relevance vector machine. To estimate the near-optimal weights for those learners, sequential quadratic optimization and particle swarm optimization were employed in this study. This ensemble algorithm was assessed using the C-MAPSS dataset to evaluate its effectiveness. In addition, it was compared to all base learners to show that ensemble learning could significantly improve performance when applied to estimate the RUL. Mo et al. [11] presented a DL model based on the transformer encoder to capture long- and short-term dependencies in the gathered sensor data. This model was further improved by the gated convolutional unit to combine the local contexts at each time step. To show its effectiveness and efficiency in solving the C-MAPSS, it was compared to several DL models in terms of the root mean square error. This comparison showed that this model was the best for two sub-datasets out of four datasets in C-MAPSS. Costa [12] proposed a variational encoding model for the aero engines RUL prediction. This model is composed of two components: a regression model and a recurrent encoder. The recurrent encoder is responsible for regularizing the latent space to generate a self-explanatory map that better detects the degradation level of aero-engines. This model was assessed using C-MAPSS datasets and compared to several DL models to assess its effectiveness and efficiency. The experimental outcomes show that this model could perform better than all competitors for two out of four C-MAPSS datasets.

Boujamza [13] proposed an LSTM approach integrated with the attention mechanism to focus on the most informative information and skip less useful ones. This study used the C-MAPSS datasets to evaluate the performance of the proposed model and compared its performance to several DL models to assess its effectiveness. The experimental findings revealed the effectiveness of this model. In [14], a new DL approach based on the LSTM integrated with the attention mechanism was presented to detect the aero-engines RUL. Duan et al. [15] proposed a bidirectional gated recurrent unit (BiGRU)--based DL model for tackling this prediction problem. This model was further improved using two mechanisms: (I) an attention mechanism to assign weights to each timestep of information, and (II) a skip connection to enhance BiGRU encoding performance. Two common datasets, namely C-MAPSS and milling, were used to evaluate this model's performance. Compared to some recent DL models, this model could have comparable performance when employed to detect the RUL of aero-engines.

In [16], a newly proposed attention mechanism dubbed the global attention mechanism, was proposed to diminish the shortcomings of the dual attention mechanism in a new attempt to accurately estimate the aero-

engines RUL. This mechanism was hybridized with the temporal convolutional network and self-attention mechanism to present a new DL model for better predicting the RUL. The C-MAPSS datasets were utilized to observe the effectiveness of this model. Compared to some state-of-the-art methods, this model was the best. There are several other ML and DL models presented in the literature for estimating the RUL, some of which are multi-objective deep belief networks [17], LSTM improved using multi-scale sequence [18], physics-informed neural network [19], artificial neural network (ANN) [20], support vector machines (SVM) [21, 22], RF [23], and LSTM [24].

To predict the RUL of aero-engines more accurately, this paper proposes a new DL model, namely ARB-ALSTM, based on combining the attention mechanism with a residual block and an LSTM model, to better comprehend the nonlinear characteristics of this problem and improve future predictions. The attention residual block is in charge of comprehending the input dataset and extracting the most useful features, which have a big influence on the performance of the proposed model. After that, an LSTM equipped with an adaptive attention mechanism receives the extracted features to efficiently deal with long-term dependencies and focus on the most informative information. To showcase the efficacy of this proposed model, it is assessed using the CMAPSS datasets and compared with many DL models. Herein, the proposed model is assessed using the CMAPSS datasets and compared to several DL models to showcase its efficacy. The results of the conducted experiments show that ARB-ALSTM is a robust alternative for predicting aero engines RUL since it could get better results than any of the compared models. This paper's main contributions are shown in the following list:

- Proposing a new RUL prediction method for accurately predicting the RUL of aero-engines. This method is based on integrating the attention mechanism with a residual block and an LSTM model to better understand the nonlinear characteristics of this problem.
- The attention-based residual block based on a scaled dot product attention mechanism is proposed to filter input features, focusing on the most informative features and skipping the less informative ones.
- Validating this model's performance using the C-MAPSS datasets and comparing it to several DL models to show its effectiveness.
- The experimental findings show that ARB-ALSTM is a robust alternative method for estimating aero engines RUL more accurately.

This paper's remaining sections are arranged as follows: Methods and materials are covered in Section 2; the proposed model is explained in Section 3; experimental settings are offered in Section 4; findings and discussion are presented in Section 5; and the conclusion and future work are presented in Section 6.

2 | Methods and Materials

2.1 | C-MAPSS Dataset

The NASA C-MAPSS aircraft engine degradation dataset is a common dataset that has been widely used in the literature for testing the performance of the newly proposed DL models for estimating the RUL of the aero-engines [25]. This dataset is separated into four sub-datasets: FD003, FD002, FD001, and FD004, with each category simulated under various operational settings and fault modes (see Table 1). Each sub-dataset is partitioned into training and testing datasets to train the newly proposed models and test their generalization capabilities. Each instance in this dataset is represented by 26 numerical features, including the serial number, degradation cycle, three operational settings, and twenty-one sensors with different measurements. The characteristics of those sensors are described in [25]. Only the sensor signals that generate a decreasing or increasing pattern could be used to observe the RUL of the engines, while those with constant patterns over some time could not show the status of those engines. In Figure 1, we show the pattern of signals obtained by various sensors used in the FD001 training. This figure shows that some sensors, such as 1, 5, 6, 10, 16,

18, and 19, have constant patterns, thereby making them irrelevant to predicting the RUL of the engines, and the remaining sensors, including 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20, and 21, have either decreasing or increasing patterns [17].

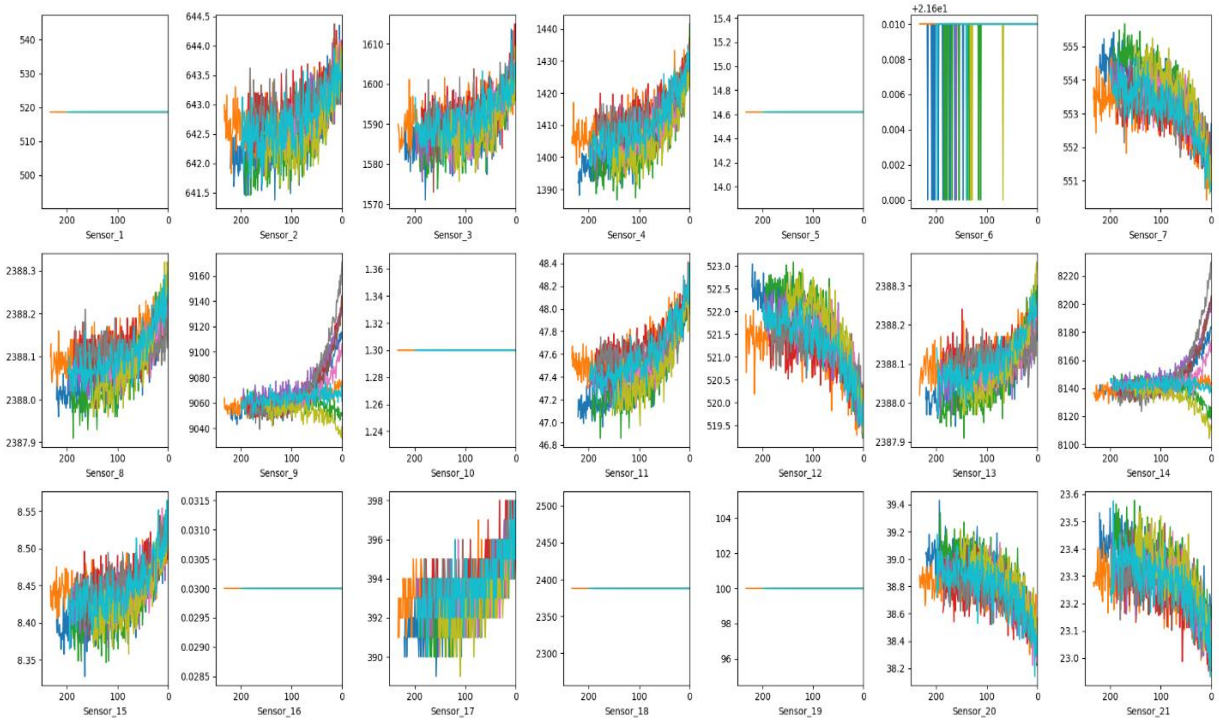

Figure 1. Visualization of the data collected by 21 sensors from the FD001 training set.

Table 1. Description of C-MAPSS.

Dataset	C-MAPSS			
	FD001	FD002	FD003	FD004
Operation condition	1	6	1	6
Test units	100	259	100	248
Max cycles	362	378	525	543
Train units	100	260	100	249
Fault mode	1	1	2	2
Min cycles	128	128	145	128
Avg cycle	206	207	247	246

2.2 |LSTM Layer

LSTM represents a new variant of RNN, which is presented to deal with the vanishing gradient problem to handle long-term dependency prediction. This model presents a more intricate memory cell that enables it to preserve information across extended sequences, rendering it highly appropriate for long-term dependency tasks, like the RUL estimation for aero-engines. The input gate i_t, forget gate f_t, and output gate o_t Are three components of LSTM, which are mathematically defined as follows:

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

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

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

Where σ indicates the sigmoid function, t indicates the time step, x_t indicates the input feature at time t, h_{t-1} indicates the output hidden state from the previous time sample, $W_f, W_i, W_o, U_f, U_i, U_o, b_f, b_i, b_o$ are weights

that are optimized during the training process. In addition, the candidate memory unit that represents the information added to the cell state at a particular time step is mathematically described as follows:

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

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

$$h_t = o_t \cdot \tanh(c_t) \quad (6)$$

where W_a, U_a, b_a are weights that are optimized during the training process. The unit state c_t at time t could be estimated using (5), and the hidden state h_t at time t could be computed according to (6). Figure 2 depicts the architecture of an LSTM cell.

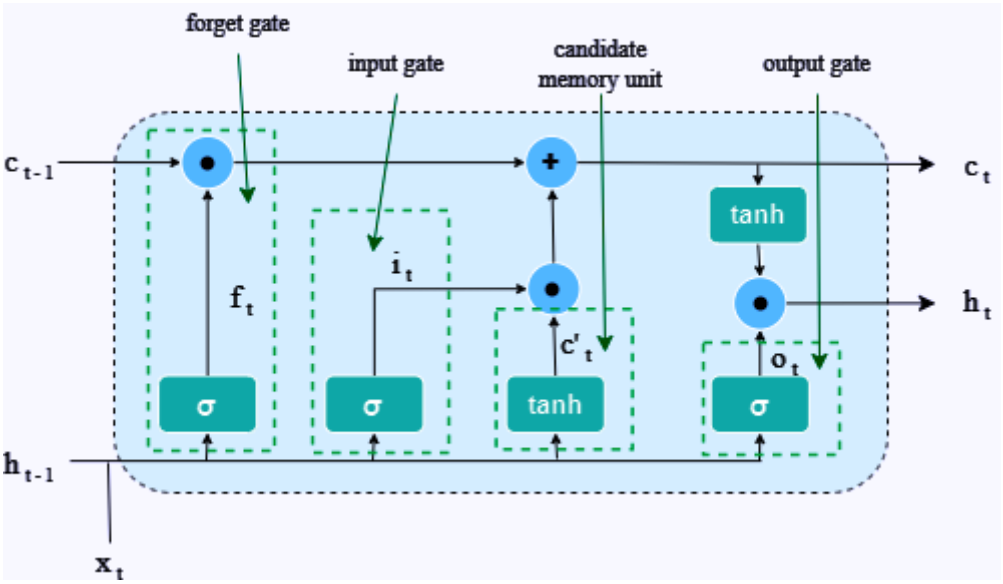

Figure 2. Architecture of an LSTM cell.

2.3 | Attention Residual Block (ARB)

The Attention Residual Block (ARB) is a constituent of neural network architecture that amalgamates both attention mechanisms and residual connections. The ARB is responsible for capturing long-term dependency and paying attention to the most informative features in the input sequence, thereby making it perform well for capturing long-term dependency in time-series predictions. This block consists of four components: a convolution layer, a dropout, an attention mechanism, and a fully connected layer. The convolution layer is in charge of identifying the most effective features, which affect the model's performance while solving classification or prediction problems. The dropout layer is integrated to prevent overfitting by randomly eliminating some trainable parameters during the training process. The scaled dot-product attention mechanism is used for better-identifying correlations between distant elements within a given sequence. This is achieved through the computation of attention scores by measuring the similarity between the query and key vectors. This mechanism gives the model the ability to focus on the most informative information throughout the input sequence. The dense layer plays a pivotal role in converting the output of the attention mechanism into a relevant representation to be further processed.

2.4 | Adaptive Attention Mechanism

The traditional method of assigning attention weights relies on the similarity between query and key vectors. However, the adaptive attention mechanism considers both the context and the input data, allowing the model to adjust the attention weights dynamically and selectively to focus on different segments of the input sequence. The adaptive attention mechanism is integrated after the LSTM layer. LSTM layers are proficient at capturing the long-term dependencies in input sequences. Nevertheless, they might encounter difficulties in capturing distant relationships or sufficiently focusing on significant segments of the input sequence.

Therefore, the adaptive attention mechanism is effectively integrated into an LSTM cell to aid in better learning temporal information and focusing on the most informative information. The output of this mechanism is calculated according to the following formulas:

$$Q = X \times W_q \tag{7}$$

$$K = X \times W_k \tag{8}$$

$$attn\_scores = softmax(tanh(Q + K) \times V) \tag{9}$$

$$attn_{output} = \sum_{i=1}^{N} (X_i \times attn\_scores_i) \tag{10}$$

where W_q, and W_k are trainable parameters, X stands for the input tensor, and N donates the number of samples.

3 | The Proposed DL Model

The estimation of the remaining useful life for an aero-engine is regarded as a supervised regression task, which is based on data collected from various sensors to be used for training and testing various DL models. In this paper, a new DL model based on LSTM, attention residual block, and adaptive attention mechanism is presented to detect the aero-engines RUL; this model is called ARB-ALSTM. This model receives the input sensor data and assigns it to two parallel paths. The first path includes an LSTM for learning temporal information from the input data. Meanwhile, the second parallel path includes an attention residual block that is responsible for extracting the most effective features to be submitted to the scaled dot product attention mechanism for focusing on the most informative features. Afterward, the output of the first path is given as an input to an adaptive attention mechanism to better select the informative features and skip the less beneficial ones. The outputs of both the first and second paths are concatenated and given as input to an LSTM to better learn temporal information. The output of this LSTM is given to a fully connected layer to perform the prediction step. Finally, Algorithm 1 presents the pseudocode of the proposed model, and Figure 3 depicts the flowchart of its architecture.

As described in Algorithm 1, the proposed model receives the input data to be preprocessed to get rid of several issues, like outliers and dominant features. Then, the input layer receives this data and submits it to two parallel paths: the first path contains an LSTM with 128 neurons and a Tanh activation function for learning temporal information from the input data, and the second includes an attention residual block (ARB) for extracting the most effective features and focusing on the most informative features. The ARB block receives the input sequences and passes them to a convolutional layer with 32 filters and a kernel size of 2 to generate a group of feature maps, which aid in extracting the most effective features from the input data. Those feature maps are then submitted to a dropout layer with a dropout rate of 0.6 to alleviate the number of trainable parameters in an attempt to loosen the overfitting problem and improve the proposed model's generalization capability. The output of this layer is given to the scaled-dot product attention mechanism to focus on the most informative features. Those features are then sent to a fully connected (FC) layer with 32 neurons and the ReLU activation function to pay attention to the most effective features among them. Afterward, an LSTM layer with 128 neurons and Tanh activation functions receives the output of the ARB block to further learn temporal information. That temporal information is then submitted to an adaptive attention mechanism to further focus on the most informative features.

The outputs of those two paths are concatenated using a concatenate layer and sent to an additional LSTM layer with 64 units and the Tanh activation function, which is followed by a dropout layer with a probability of 0.6 in a new attempt to enhance the proposed model's performance. Finally, the output of this layer is fed into three FC layers: the first two layers with 32 and 8 neurons, respectively, are used to better represent the input sequences, and the third layer with 1 neuron predicts the RUL of aero-engines.

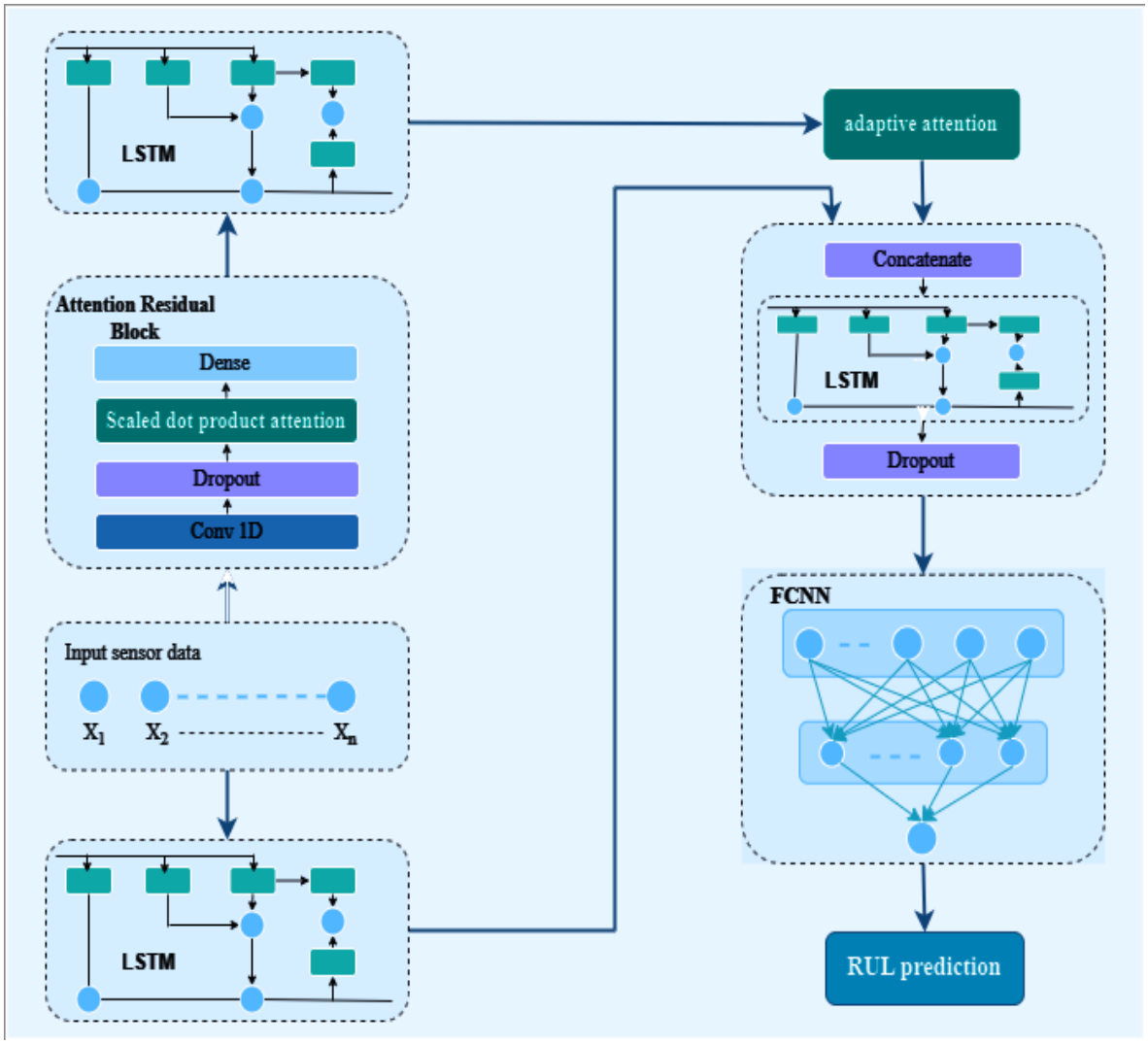

Figure 3. Flowchart of the proposed ARB-ALSTM.

Algorithm 1 Pseudo-code of ARB-ALSTM
Input: Input data (*D*), batch size (*Bs*), maximum epoch (*T*), and learning rate (*lr*)
Output: *loss* (***Score***), *RMSE*
Conducting the preprocessing step
/* Create the proposed DeepChurn model */
Input: Construct an input layer to receive the input data
/* Feature extraction and temporal learning based on the ARB block and LSTM */
/* First Parallel Path*/
P1: Create an LSTM layer with 128 units and a *Tanh* activation function to take the data from the input layer
/* Second Parallel Path*/
P2: Create a Conv-1D layer with 32 filters and a kernel size of 2 to take the data from the input layer
P2: Add a Dropout layer with a dropout rate of 0.6 to P2
P2: Add Scaled dot product attention mechanism to P2
P2: Add a dense layer with 32 nodes and *ReLU* activation function to P2
P2: Add an LSTM layer with 128 units and *Tanh* activation function to P2
P2: Add adaptive attention with 32 units to P2
/* Concatenation stage */
x: Concatenate ([P1, P2])
x: Add an LSTM layer with 64 units and *Tanh* activation function to *x*
x: Add a Dropout layer with a dropout rate of 0.6 to x
/* Prediction Block */
x: Add a dense layer with 32 nodes and *ReLU* to *x*

x: Add a dense layer with 8 nodes and *ReLU* to x
x: Add a dense layer with 1 node to x
/* Optimization process */
$N = Size(D)/Bs$ /* Estimate the number of batches */
$\boldsymbol{t} = \mathbf{0}$, Current epoch
while $\boldsymbol{t} < \boldsymbol{T}$
$\boldsymbol{i} = \mathbf{0}$, the current batch size
while $\boldsymbol{i} < \boldsymbol{N}$
Compute the Score function using the $\boldsymbol{ith}$ batch
Update the weights based on the Adam to optimize the score function
$\boldsymbol{i} = \boldsymbol{i} + \mathbf{1}$
end while
$\boldsymbol{t} = \boldsymbol{t} + \mathbf{1}$
end while

4 | Experimental Settings

4.1 | Data Preprocessing

4.1.1 | Data Normalization

The C-MAPSS dataset has several features with different scales, which negatively affect the performance of DL models during the training process. Therefore, those features need to be normalized to get rid of the potential distortions and biases, thereby helping the DL models become more accurate. Although several normalization methods, such as z-score normalization, min-max scaling, decimal scaling normalization, log scaling normalization, and robust scaling normalization, have been presented in the literature to perform this process, z-score normalization has been widely used due to its simplicity and effectiveness [26, 27]. Therefore, this method is considered in our study to normalize the C-MAPSS dataset for alleviating the effect of outliers and dominant features. The mathematical model of the z-score normalization method is as follows:

$$x'_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j} \tag{11}$$

where σ_j, and μ_j refer to the standard deviation and mean of the jth feature, respectively.

4.1.2 | Sliding Window Technique

In this study, the sliding window approach is utilized to segment the normalized aero-engine data into data samples under a fixed-size time window, as depicted in Figure 4. This window is shifted a time step over the normalized aero-engine data to generate the second data sample. This process is continued until the end of the data is reached. The sliding window's step size is set to 1 in our experiments because that helps better capture patterns and details in the time series [16]. According to [18], the sliding window size might vary from dataset to dataset because that might aid in presenting more accurate models. Since this study assesses the proposed model's performance using two sub-datasets, FD001 and FD003, from the C-MAPSS dataset, the sliding window sizes for them, according to [16, 18], are set to 31 and 60, respectively.

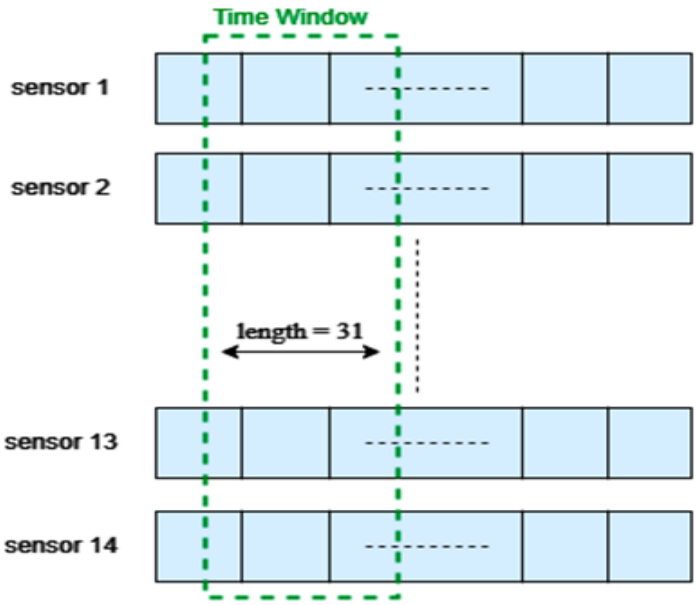

Figure 4. Sliding time window on the selected sensors.

4.2 | Evaluation Metrics

Two well-known performance metrics, known as the scoring function and root mean square error (RMSE), are utilized to assess the proposed model's performance when applied to forecast the aero-engines RUL. The first metric measures the root of the mean square error between the real and estimated labels of all samples in the dataset. This metric could be estimated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - y'_i\right)^2} \tag{12}$$

Where N represents the number of samples, and y_i and y'_i represents the true and predicted labels of the ith sample, respectively. Contrary to the RMSE metric, the scoring function gives higher weight to late predictions than early predictions, as defined in the following formula [28]:

$$Score = \begin{cases} \sum_{i=1}^{N}\left(e^{-\left(\frac{(y_i - y'_i)}{13}\right)} - 1\right), & d_i < 0 \\ \sum_{i=1}^{N}\left(e^{-\left(\frac{(y_i - y'_i)}{10}\right)} - 1\right), & d_i \geq 0 \end{cases} \tag{13}$$

Both RMSE and score values need to be minimized as much as possible to achieve better predictions for the RUL of aero-engines.

4.3 | Hyperparameter Tuning

The proposed ARB-ALSTM has some hyperparameters, such as batch size, dropout rate, number of filters, kernel size, learning rate, and number of neurons for each LSTM, that need to be accurately estimated to maximize its performance. Therefore, in this study, several experiments are conducted under various values for each parameter to estimate the most effective value, which substantially enhances the performance of the proposed model. For example, for the learning rate, several experiments are performed under 0.0001, 0.001, 0.002, 0.003, 0.005, and 0.01, and the results of those experiments show that the best learning rates for FD001 and FD003 are 0.002 and 0.003, respectively. Likewise, for the batch size, several experiments under 16, 32, 64, 80, 96, 128, and 256 are conducted to estimate the best values for FD001 and FD003. The experimental

findings show that the batch size of 96 is the best for these two sub-datasets. Finally, the experiments conducted under various values, including 0, 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6, for the dropout parameter show that 0.6 is the most effective value for both FD001 and FD003. Table 2 presents the final hyperparameter values of the proposed model when applied to solve both FD001 and FD003.

The proposed model's performance is influenced by the number of hidden units in each LSTM layer; therefore, several experiments were conducted to determine the optimal number of hidden units for each LSTM layer. According to Tables 3-5, the best numbers of hidden units for three LSTM layers used in the proposed model are 256, 128, and 64 in order of existing within the architecture of this model. According to Table 6, the ideal number of filters that yield the most favorable RMSE values is 32 for both the FD001 and FD003 datasets. Furthermore, many experiments were conducted to determine the optimal kernel size. According to the findings presented in Table 7, a kernel size of 2 yields the most favorable outcome because it achieves the lowest RMSE value for both the FD001 and FD003 datasets.

Table 2. The hyper-parameters of the ARB-ALSTM model for both FD001 and FD003.

	Window size	Learning rate	Batch size	Epoch	Attention size	Dropout rate	Optimizer	Loss
FD001	31	0.002	96	70	32	0.6	Adam	Score
FD003	30	0.003	96	60	32	0.6	Adam	Score

Table 3. The RMSE of the experiments under various hidden units for the first LSTM in the proposed model.

	16	32	64	128	156
FD001	16.25	15.57	14.03	13.1	**13.03**
FD003	14.69	14.64	14.33	12.48	**12.43**

Table 4. The RMSE of the experiments under various hidden units for the second LSTM in the proposed model.

	16	32	64	128	156
FD001	14.78	15.03	13.96	**13.03**	14.36
FD003	14.98	15.83	14.45	**12.43**	13.38

Table 5. The RMSE of the experiments under various hidden units for the second LSTM in the proposed model.

	16	32	64	128
FD001	14.36	17.66	**13.03**	14.36
FD003	15.03	12.93	**12.43**	13.08

Table 6. The RMSE of the experiments under various numbers of filters.

	16	32	64	128
FD001	16.42	**13.03**	15.52	15.24
FD003	15.33	**12.43**	14.96	16.07

Table 7. The RMSE of the experiments under various kernel sizes.

	1	2	3	4
FD001	13.73	**13.03**	14.22	15.6
FD003	12.87	**12.43**	12.68	16.18

5 | Results and Discussion

This section displays the outcomes obtained by the proposed ARB-ALSTM and some rival models for FD001 and FD003. Those outcomes are represented in the score and RMSE metrics to illustrate the ability of models

to minimize the difference between the desired and predicted RUL. In brief, in this section, we start by comparing the RUL labels estimated by the proposed model with actual RUL labels to illustrate how far it could minimize the difference between them. Afterward, the outcomes of ARB-ALSTM are extensively compared to some rival models, such as BiGRU-AS [15], Att-LSTM [13], RVE [12], TaNet [29], PINNs [19], Attention-LSTM [14], KGHM [30], BayesLSTM [31], and CP-LSTM [24], to show its superiority.

5.1 | Comparison between Estimated and Actual RUL

This section compares the actual RUL labels with those estimated by the proposed ARB-ALSTM for both FD001 and FD003 to illustrate how close the detected RUL is to the actual RUL. On each sub-dataset, the proposed model is trained for 70 or 60 epochs according to the tackled sub-dataset. Then, it is tested on the testing dataset to check its generalization capability. Its detected RUL value for each sample in the testing dataset is depicted with the actual RUL value in Figures 5 and 6 for FD001 and FD003, respectively. Inspecting those figures shows that the estimated RUL values are significantly close to the actual RUL values for the majority of the test instances in both FD001 and FD003. To quantitatively show the difference between estimated and desired RUL, Figures 7 and 8 are presented to show the error between them for each test instance in two considered sub-datasets. From those figures, for the FD001 sub-dataset, the error values are ranged in the interval $[-30, 40]$, while, for the FD003 sub-dataset, they are located in the range of -30 and 30. From the above analysis, it is concluded that the proposed ARB-ALSTM has high stability since it could achieve similar performance on two different datasets, in addition to having robust performance for minimizing the difference between the estimated and desired RUL.

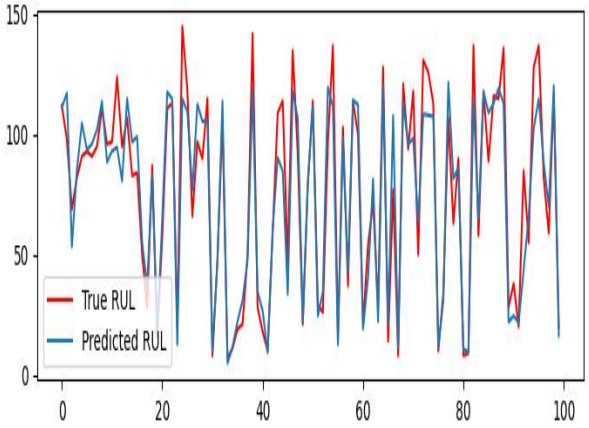

Figure 5. Depiction of the true and estimated RUL for FD001.

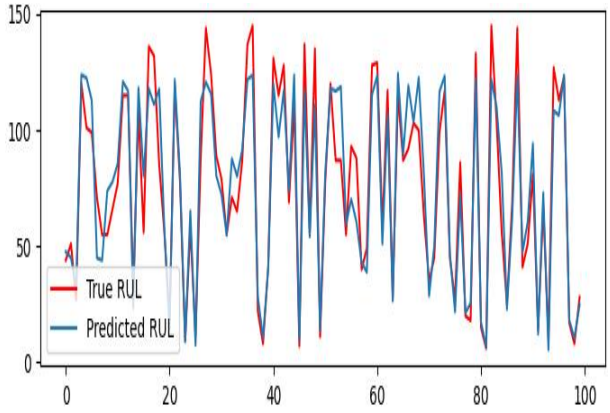

Figure 6. Depiction of the true and estimated RUL for FD003.

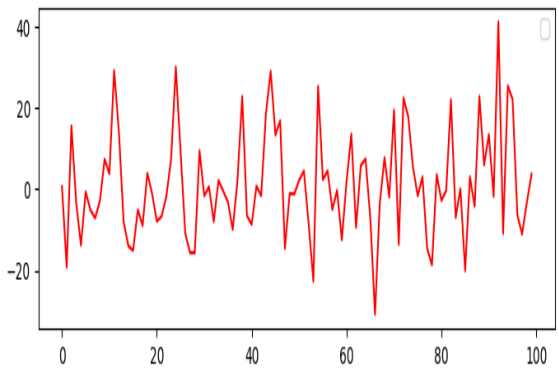

Figure 7. Error-values between the estimated and actual RUL for FD001.

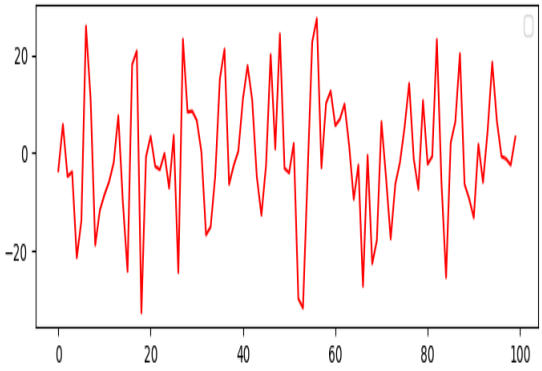

Figure 8. Error-values between the estimated and actual RUL for FD003.

5.2 | Comparison between ARE-ALSTM and Rival Models

In this section, the outcomes of ARB-ALSTM for both FD001 and FD003 are compared to those of nine rival models to show their effectiveness and efficiency. Those outcomes are represented in the score and RMSE values as reported in Table 8. This table shows that ARB-ALSTM could be better than all the compared models for the RMSE on two considered datasets, where it could achieve RMSE values of 13.03 and 12.43 for FD001 and FD003, respectively. However, it could not be the best for the scoring metric for FD003 because some models, such as CP-LSTM and Att-LSTM, could achieve a lower score value. Despite that, the proposed model is considered a strong alternative for tackling this problem because it could be the best under the RMSE metric, which gives equal weight to both late and early predictions. On the contrary, the scoring metric gives higher weight to late predictions than early predictions, thereby giving the model preference for late predictions. To graphically show the proposed model's superiority, Figure 9 is presented to show the RMSE values obtained by various algorithms for the two considered sub-datasets.

Table 8. RMSE and score values were obtained by the proposed and rival models for both FD001 and FD003.

	FD001		**FD003**	
	RMSE	score	RMSE	score
BiGRU-AS	13.86	284	15.53	428
Att-LSTM	13.95	320	12.72	223
RVE	13.42	323.82	12.51	256.36
DSAN	13.4	242	15.12	497
PINNs	16.89	523	17.52	1194
Attention-LSTM	15.45	455.92	14.67	473.97
KGHM	13.18	251	13.54	333
BayesLSTM	13.26	265.76	12.68	242.91
CP-LSTM	13.59	224.88	12.94	207.1
Proposed method	**13.03**	**222.58**	**12.43**	233.72

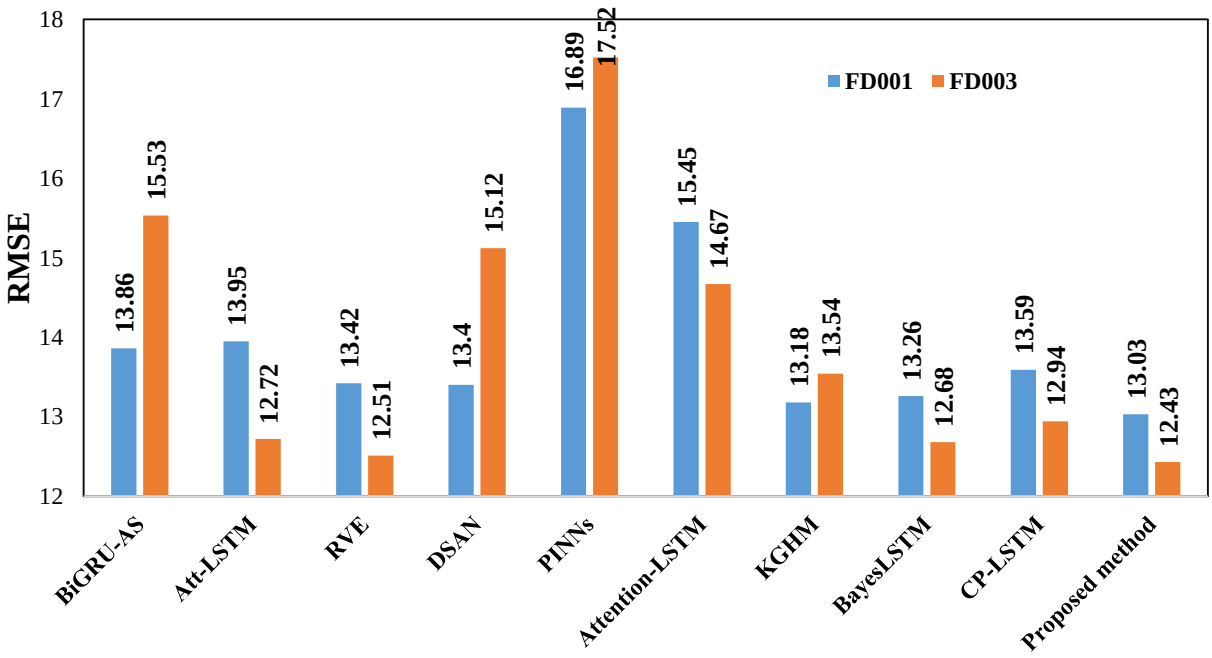

Figure 9. Depiction of RMSE values obtained by various models.

5.3 | Ablation Analysis

In this section, several experiments are performed to validate the impacts of adaptive attention and ARB mechanisms on the performance of the proposed ARB-ALSTM. These experiments encompass an LSTM devoid of any attention mechanism, an LSTM equipped with an adaptive attention mechanism (AA-TCN), and an LSTM integrated with an ARB mechanism (ARB-LSTM). The results of those experiments are displayed in Table 9. Inspecting this table illustrates that ARB-ALSTM displays superior performance for the score and RMSE metrics on both the FD003 and FD001 sub-datasets. This shows the significant effect of the attention mechanisms used on the performance of the LSTM model.

Table 9. Influence of various components used in the proposed ARB-ALSTM on its performance.

	FD001		**FD003**	
	RMSE	score	RMSE	score
LSTM	15.71	435.71	15.53	428
LSTM + AA	15.21	321.48	12.72	223
LSTM + ARB	16.34	337.1	12.51	256.36
LSTM+AA+ARB (ARB-ALSTM)	**13.03**	**222.58**	**12.43**	**233.72**

6 | Conclusion and Future Work

This paper offers a novel DL model, namely ARB-ALSTM, based on integrating the attention mechanism with an LSTM model and a residual block to better understand the features of the RUL prediction of aero-engines and improve future predictions. In a broader sense, the attention residual block is in charge of analyzing the input dataset and extracting the most effective features, which have a substantial impact on the model's accuracy. Those extracted features are then fed into an LSTM with an adaptive attention mechanism for successfully capturing and analyzing long-term dependent data. To validate its performance, this proposed model is assessed using the CMAPSS dataset and contrasted with several DL models, such as BiGRU-AS, Att-LSTM, RVE, DSAN, PINNs, Attention-LSTM, KGHM, BayesLSTM, and CP-LSTM. Experimental results indicate that ARB-ALSTM is a powerful alternative for forecasting the RUL of aircraft engines because it can obtain better outcomes than all the examined models. In future work, we will investigate the performance of the proposed model for the lithium-ion battery life prediction.

Funding

This research was conducted without external funding support.

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 no conflict of interest.

Ethical Approval

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

References

[1] Vogl, G.W., B.A. Weiss, and M. Helu, A review of diagnostic and prognostic capabilities and best practices for manufacturing. Journal of Intelligent Manufacturing, 2019. 30: p. 79-95.
[2] Cao, G. Remaining Useful Life prediction of Aircraft Engines Using DCNN-BiLSTM with K-means Feature Selection. in International Symposium on Artificial Intelligence and Robotics. 2023. Springer.
[3] Vrignat, P., F. Kratz, and M. Avila, Sustainable manufacturing, maintenance policies, prognostics and health management: A literature review. Reliability Engineering & System Safety, 2022. 218: p. 108140.
[4] Wei, J., G. Dong, and Z. Chen, Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression. IEEE Transactions on Industrial Electronics, 2017. 65(7): p. 5634-5643.
[5] Cui, L., et al., Research on remaining useful life prediction of rolling element bearings based on time-varying Kalman filter. IEEE Transactions on Instrumentation and Measurement, 2019. 69(6): p. 2858-2867.
[6] Sun, J., X. Zhang, and J. Wang, Lightweight bidirectional long short-term memory based on automated model pruning with application to bearing remaining useful life prediction. Engineering Applications of Artificial Intelligence, 2023. 118: p. 105662.
[7] Wang, Y., et al. Research on Fault Diagnosis System Based on Aeroengine Knowledge Base. in 2022 Global Reliability and Prognostics and Health Management (PHM-Yantai). 2022. IEEE.
[8] Djeziri, M.A., S. Benmoussa, and M.E. Benbouzid, Data-driven approach augmented in simulation for robust fault prognosis. Engineering Applications of Artificial Intelligence, 2019. 86: p. 154-164.
[9] Li, Y., et al., A novel remaining useful life prediction method based on multi-support vector regression fusion and adaptive weight updating. ISA transactions, 2022. 131: p. 444-459.
[10] Li, Z., K. Goebel, and D. Wu, Degradation modeling and remaining useful life prediction of aircraft engines using ensemble learning. Journal of Engineering for Gas Turbines and Power, 2019. 141(4): p. 041008.
[11] Mo, Y., et al., Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit. Journal of Intelligent Manufacturing, 2021. 32(7): p. 1997-2006.
[12] Costa, N. and L. Sánchez, Variational encoding approach for interpretable assessment of remaining useful life estimation. Reliability Engineering & System Safety, 2022. 222: p. 108353.
[13] Boujamza, A. and S.L. Elhaq, Attention-based LSTM for remaining useful life estimation of aircraft engines. IFAC-PapersOnLine, 2022. 55(12): p. 450-455.
[14] Cheng, X., et al., RUL Prediction Method for Electrical Connectors with Intermittent Faults Based on an Attention-LSTM Model. IEEE Transactions on Components, Packaging and Manufacturing Technology, 2023.
[15] Duan, Y., et al., A BiGRU autoencoder remaining useful life prediction scheme with attention mechanism and skip connection. IEEE Sensors Journal, 2021. 21(9): p. 10905-10914.
[16] Xu, Z., et al., Global attention mechanism based deep learning for remaining useful life prediction of aero-engine. Measurement, 2023. 217: p. 113098.
[17] Zhang, C., et al., Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. IEEE transactions on neural networks and learning systems, 2016. 28(10): p. 2306-2318.
[18] Lin, R., et al., Remaining useful life prediction in prognostics using multi-scale sequence and Long Short-Term Memory network⋆. Journal of computational science, 2022. 57: p. 101508.
[19] Liao, X., et al., Remaining useful life with self-attention assisted physics-informed neural network. Advanced Engineering Informatics, 2023. 58: p. 102195.
[20] Li, X., et al., An enhanced selective ensemble deep learning method for rolling bearing fault diagnosis with beetle antennae search algorithm. Mechanical Systems and Signal Processing, 2020. 142: p. 106752.
[21] Yan, M., et al., Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model. ISA transactions, 2020. 98: p. 471-482.
[22] Zhao, R., et al., Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 2019. 115: p. 213-237.
[23] Soni, P., et al. Multiclass classification for predicting Remaining Useful Life (RUL) of the turbofan engine. in 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). 2021. IEEE.
[24] Arunan, A., et al., A change point detection integrated remaining useful life estimation model under variable operating conditions. Control Engineering Practice, 2024. 144: p. 105840.
[25] Saxena, A., et al. Damage propagation modeling for aircraft engine run-to-failure simulation. in 2008 international conference on prognostics and health management. 2008. IEEE.
[26] Zheng, S., et al. Long short-term memory network for remaining useful life estimation. in 2017 IEEE international conference on prognostics and health management (ICPHM). 2017. IEEE.
[27] Singh, D. and B. Singh, Investigating the impact of data normalization on classification performance. Applied Soft Computing, 2020. 97: p. 105524.
[28] Sateesh Babu, G., P. Zhao, and X.-L. Li. 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. 2016. Springer.

[29] Fan, L., Y. Chai, and X. Chen, Trend attention fully convolutional network for remaining useful life estimation. Reliability Engineering & System Safety, 2022. 225: p. 108590.
[30] Li, Y., et al., Remaining useful life prediction of aero-engine enabled by fusing knowledge and deep learning models. Reliability Engineering & System Safety, 2023. 229: p. 108869.
[31] Xiang, F., et al., Bayesian gated-transformer model for risk-aware prediction of aero-engine remaining useful life. Expert Systems with Applications, 2024. 238: p. 121859.