

Article

Aero-Engine Remain Useful Life Estimation via Scope-Coordinated Attention-Based Network

Zheng Liao ¹, Sijie Liu ², Jin Li ², Shuai Ma ^{3,4} and Gang Li ^{1,*}

¹ School of Information Engineering, Chang'an University, Xi'an 710064, China; 2017124047@chd.edu.cn

² School of Power and Energy, Northwestern Polytechnical University, Xi'an 710072, China; sijieliu_123@mail.nwpu.edu.cn (S.L.); jin_li@mail.nwpu.edu.cn (J.L.)

³ Shenyang Engine Research Institute, Shenyang 110066, China; shma@mail.nwpu.edu.cn

⁴ National Key Lab of Aerospace Power System and Plasma Technology, Shenyang Engine Research Institute, Shenyang 110066, China

* Correspondence: gangli@chd.edu.cn; Tel.: +86-029-82334097

Abstract: Research on the assessment of the remaining useful life (RUL) has garnered significant attention because of its critical relevance in prognostics and health management (PHM) across various sectors. Recently, data-driven methodologies have become increasingly important for RUL prediction. However, these methods often struggle to capture long-term dependencies and possess a limited receptive field, restricting their effectiveness in various RUL prediction scenarios. To address these limitations, this study proposes a novel approach called the scope-coordinated attention-based (SCAB) network for RUL prediction. The initial design features a novel multichannel feature integration block, which aims to effectively capture and integrate essential information from raw sensor data. Additionally, it is designed to expand the receptive field by capturing rich and diverse features for improved representation. Subsequently, a dual-attention block refines information and further expands the receptive field in both the channel and spatial domain. Moreover, a feature pyramid block with adaptive self-attention is developed to effectively capture long-term dependencies, further enhancing the information's detail and features by the multiscale feature fusion mechanism. The efficacy of the proposed SCAB model for RUL estimation was validated using the C-MAPSS public dataset. In comparison experiments, the SCAB model outperformed other methods in the FD002 subset while demonstrating excellent performances in FD001, FD003, and FD004. These results confirm that the SCAB model exhibits robust and superior performance in RUL prediction across various aeroengine scenarios.

Keywords: remaining-useful-life prediction; deep learning; prognostics health management; attention mechanism



Academic Editors: Dimitri Mavris and Olivia J. Pinon Fischer

Received: 6 February 2025

Revised: 12 March 2025

Accepted: 17 March 2025

Published: 19 March 2025

Citation: Liao, Z.; Liu, S.; Li, J.; Ma, S.; Li, G. Aero-Engine Remain Useful Life Estimation via Scope-Coordinated Attention-Based Network. *Aerospace* **2025**, *12*, 259. <https://doi.org/10.3390/aerospace12030259>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Prognostics and health management (PHM) integrate data, knowledge, and computation to monitor SSCs, enabling anomaly detection, fault diagnosis, and degradation prediction for enhanced efficiency, reliability, and safety. By supporting condition-based and predictive maintenance, just-in-time, just-right maintenance is ensured through PHM, optimizing production profits while minimizing costs and losses [1]. PHM is essential for aeroengine safety, efficiency, and cost optimization, enabling real-time monitoring and predictive maintenance. A key aspect of PHM, remaining-useful-life (RUL) prediction provides critical insights by estimating the time to failure, facilitating proactive maintenance,

failure prevention, and extended engine lifespan. Advancing RUL prediction methods is crucial for decision makers in manufacturing, transportation, and aerospace, ensuring reliable and cost-effective aeroengine operations [2].

RUL estimation aims to predict the remaining operational life or time until failure using both historical and real-time sensor data. Existing RUL prediction methods are generally classified into model-based and data-driven approaches. The model-based approach utilizes physics-of-failure (PoF) models, combining mathematical and physical representations to describe damage progression in machinery [3]. However, this approach is limited in accurately modeling the dynamics of mechanical systems or components [4]. Achieving accurate modeling of complex systems is often challenging, even for experts in the field. Moreover, the efficacy and adaptability of model-based solutions are often constrained by the variability in mechanical system behaviors across different applications [5]. On the other hand, data-driven approaches bypass the need for intricate prior knowledge by leveraging sensor data to assess equipment conditions and directly predict the RUL, making them more flexible and widely applicable.

Data-driven RUL prediction methods develop degradation models by capturing the nonlinear relationship between sensor data and RUL labels. These methods are categorized into traditional machine-learning and deep-learning-based approaches. The former includes techniques such as autoregressive (ARM) models [6], support vector machines (SVMs) [7], and extreme learning machines (ELMs) [8]. Although widely applied, traditional machine-learning methods rely on manual feature engineering and possess limited scalability, restricting their accuracy in handling large datasets. In contrast, deep-learning methods streamline feature extraction, offer robust nonlinear fitting, and have gained increasing popularity in RUL estimation.

With the advancement of deep learning, CNNs and RNNs have become dominant in RUL prediction. Because RUL estimation is fundamentally a time-series forecasting problem, RNN-based models have been extensively utilized. For instance, Zheng et al. [9] proposed an LSTM approach for RUL estimation by fully leveraging sequence information and revealing complex patterns in sensor data. In a study by Wang et al. [10], a variation of the LSTM model, known as bidirectional LSTM (Bi-LSTM), was employed to assess aeroengine deterioration and forecast the RUL. To further enhance feature extraction, hybrid deep-learning architectures have been explored. Xia et al. [11] integrated separate gatings for temporal modeling and adaptive feature weighting and a global temporal convolutional network for sequence fusion. Yu et al. introduced a Bi-LSTM-based autoencoder to convert high-dimensional data to low-dimensional representations, which are then used to construct a one-dimensional health index (HI) for RUL assessment [12]. Xiang et al. [13] designed multidifferential LSTMs and CNNs to enhance the processing ability of traditional LSTMs and CNNs and then proposed a spatiotemporally multidifferential model by combining these networks for the prediction of the RULs of equipment. Ta et al. [14] developed a 1D CNN model for SA-based concrete stress monitoring, integrating an impedance measurement model, stress-monitoring approaches, and K-fold cross-validation to assess accuracy under noise and at untrained stress levels. Meanwhile, da Costa et al. [15] proposed a data-driven domain adaptation approach using an LSTM and a domain adversarial neural network (DANN) to adapt RUL estimates to a target domain with only sensor data.

Beyond hybrid modules, some RUL estimation models integrate multiscale and attention-based mechanisms to enhance predictive performance. Chen et al. [16] proposed an LSTM network for extracting sequential features, an attention mechanism for weighting key features and time steps, and a feature fusion framework that integrates handcrafted and learned features to enhance RUL prediction. Zhao et al. [17] introduced GAM-CapsNet, a capsule neural network with a gated attention mechanism (GAM), to

enhance noise resistance and prioritize key features, while capsules with a Bayesian layer improve feature extraction and uncertainty quantification in RUL prediction. Li et al. [18] implemented an approach integrating multiscale feature extraction, a concatenation block, and a GRU-based fusion block for enhanced temporal learning, while leveraging Mish activation for improved performance. Tian et al. [19] propose a method integrating spatial correlation attention for variable importance and temporal attention to enhance LSTM's sequential learning for RUL prediction. Self-attention has recently emerged as a powerful attention mechanism designed to effectively capture long-term dependencies by computing relationships between all the elements in a sequence, enabling more context-aware feature learning [20]. The self-attention-based transformer model was initially developed for natural language processing (NLP) [21] and has since been widely applied to computer vision tasks [22,23], demonstrating effectiveness across various domains. In RUL estimation, transformers have been integrated with CNNs, LSTMs, and other attention-based modules to enhance predictive accuracy. Zhang et al. [24] propose an approach assigning self-learned weights to time instances based on significance, with a theoretically grounded update process and consistent interpretability across training runs for RUL prediction. Zhang et al. [25] developed a self-attention-based RUL prediction model with a parallel encoder–decoder structure, effectively handling long sequences while preventing interference between the sensor and temporal features.

However, there are still limitations that need to be addressed. First, in complex and dynamic RUL prediction scenarios, traditional CNNs rely on local features, while transformer models with fixed-length inputs struggle to capture subtle long-term dependencies effectively. Second, the traditional transformer primarily focuses on temporal information but fails to establish strong connections between input variables, limiting its receptive field. To address these issues, this paper proposes an advanced attention-based hybrid technique that employs a multichannel feature fusion mechanism to extract rich and diverse features, effectively expanding the receptive field. The attention mechanism focuses on key features in both local and global domains, enhancing feature representation while capturing subtle environmental changes in aeroengine scenarios and effectively establishing long-term dependencies. Additionally, a multiscale mechanism is implemented to enhance feature extraction, improve input details, and strengthen the model's stability, ensuring greater robustness and reliability in RUL prediction.

In this paper, an innovative approach, called the scope-coordinated attention-based (SCAB) network, is presented for estimating the remaining useful life (RUL) of aeroengines. Building on prior research, the SCAB network employs a novel multichannel feature integration block (MCFIB) that includes a depthwise separable convolution-enhanced residual block that captures and preserves the essential information inherent in the original data, an improved spatial pyramid attention (SPA) block for extracting structured and regularized information, and an improved adaptive spatial feature fusion (ASFF) block for extracting multiscale information. Additionally, this architecture incorporates an improved dual-attention network (DANet) that effectively focuses on the most significant features across both the spatial and channel domains. The network concludes with an advanced self-attention-based feature pyramid (SAFP) model that fuses critical information across the entire run-to-failure timeline. The proposed methodology was validated using aeroengine data, and a comparative analysis against other leading methods demonstrated the superiority of the SCAB network for advanced predictive maintenance. The main contributions of this paper are as follows:

- A novel MCFIB, including three channels, is proposed to enhance features extracted from raw sensor data and expand the receptive field by extracting rich and diverse features;

- The improved DANet module features an improved dual-attention network for spatial and channel enhancement. The spatial attention captures the semantic interdependencies between any two locations in the spatial dimension, while the channel attention emphasizes the interdependent channel mapping in the channel dimension;
- The SAFF module utilizes an advanced self-attention-based feature pyramid block, which emphasizes relevant information and models the long-term dependency in time series through self-attention. Additionally, a feature-fusion pyramid-based pooling mechanism is used to improve the model's performance under various operating conditions;
- The performance of the proposed technique was evaluated using aeroengine datasets. The experimental results indicate that the method has significant potential to enhance RUL prediction performance.

The remainder of this paper is organized as follows: Section 2 describes the proposed approach for estimating the RUL. The details of the aeroengines' RUL benchmark datasets are presented in Section 3. Section 4 presents an analysis of the aeroengines' RUL benchmark datasets using the proposed method. The paper concludes in Section 5.

2. Proposed Approach

2.1. Overall Architecture of the Proposed Method

The overall design of the proposed method, as illustrated in Figure 1, integrates an MCFIB, which extracts three key features: essential information from the original data using a depthwise separable convolution-enhanced residual block, structured and regularized information from an attention-based SPA block, and multiscale information from an ASFF block to achieve a broader receptive field. These integrated features are then processed through an improved dual-attention module, which refines critical information and expands the receptive field across both spatial and channel domains, ensuring more effective feature representation. Additionally, the SAFF module enhances input details and strengthens the model's stability through a pyramid-based pooling mechanism, which aggregates multiscale contextual information to maintain robustness in diverse RUL prediction scenarios. The self-attention mechanism within SAFF establishes long-term dependencies, allowing the model to effectively track progressive degradation patterns over extended time periods. The following subsections provide a detailed overview of the individual components. To manage the computational complexity, our model includes one MCFIB module, one dual-attention layer, and one SAFF module. The specific details are provided in Table 1, where N denotes the number of samples, and L denotes the sequence length of the sliding time window.

Table 1. Architectural overview and data handling of each module in the SCAB network.

Module	Block	Input Size	Output Size
MCFIB	DWGLU&RESBlock	$(N, L, 14)$	$(N, L, 128)$
	CA&SA&SPA	$(N, L, 14)$	$(N, L, 128)$
	ASFF	$(N, L, 14)$	$(N, L, 128)$
	Projection_out	$(N, L, 128)$	$(N, L, 128)$
DANet	Pam	$(N, L, 128)$	$(N, L, 128)$
	Cam	$(N, L, 128)$	$(N, L, 128)$
SAFF	SAFF	$(N, L, 128)$	$(N, L, 128)$
FC Layer		$(N, L, 128)$	$(N, L, 1)$

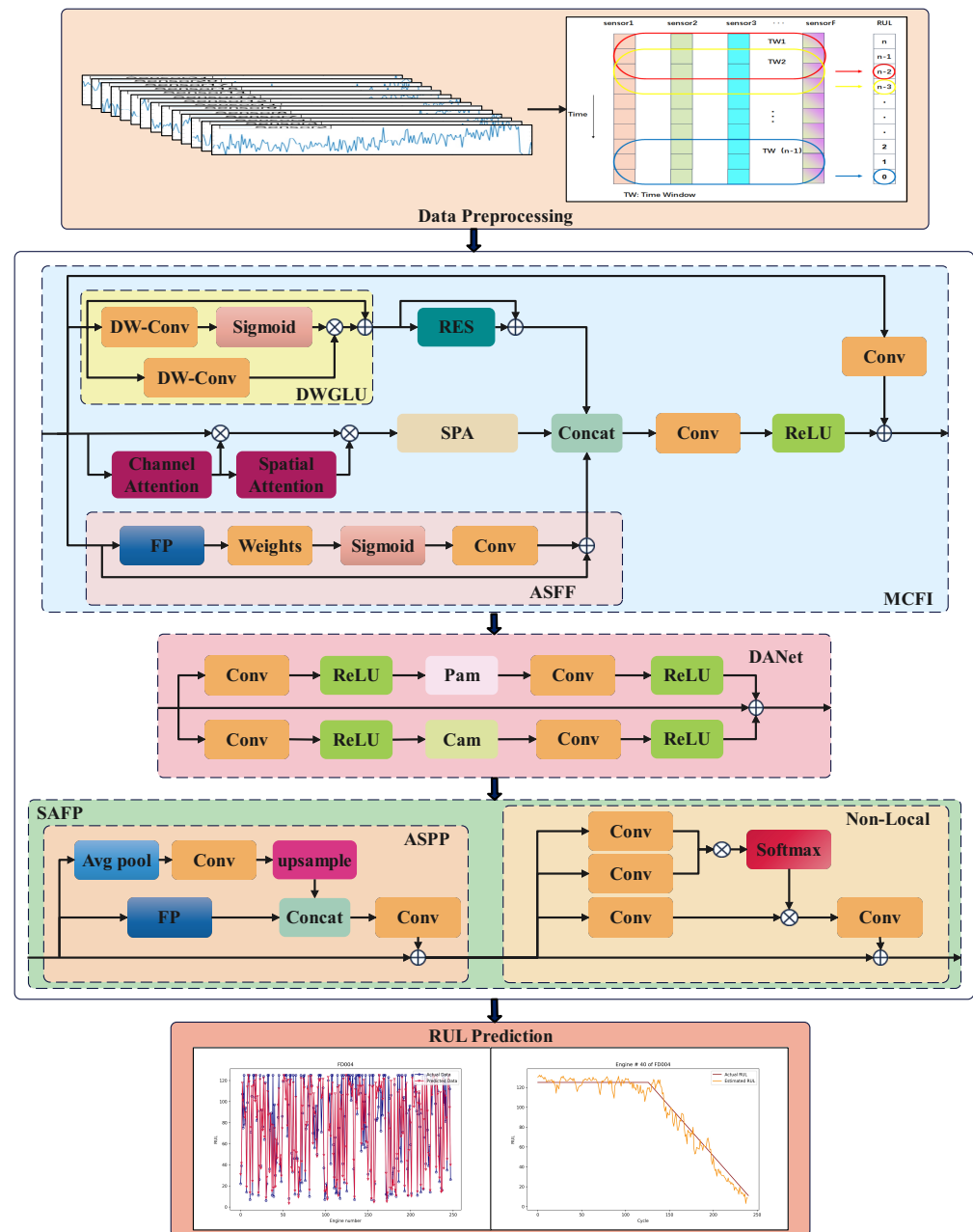


Figure 1. The overall architecture of the proposed method.

2.2. Multichannel Feature Integration Block

2.2.1. Depthwise Separable Convolution-Enhanced Residual Block

The bottleneck layer of the ResNet introduces a novel gated linear unit that can better convey the original information of the RUL to prevent overfitting by preserving essential degradation patterns and mitigating the vanishing gradient [26]. The bottleneck layer used in deep ResNet architectures (e.g., ResNet-50, ResNet-101, and ResNet-152) optimizes efficiency by reducing the computational cost and parameters while preserving representational power. It consists of three layers: a 1×1 convolution for feature compression, a 3×3 convolution for transformation, and another 1×1 convolution for expansion, minimizing the number of expensive operations on high-dimensional feature maps. The output of the bottleneck block can be expressed as follows:

$$F_r(x) = U(x) \cdot \sigma(V(x) \cdot \sigma(W(x) \cdot x)) + x \quad (1)$$

where $W(1 \times 1)$ reduces dimensions, $V(3 \times 3)$ extracts features, and $U(1 \times 1)$ restores the original dimensions, while the identity shortcut ensures the gradient flow.

Because the direct addition of input and output to the residual block can cause feature distortion, especially when input and output dimensions do not match, a lightweight depthwise gated linear unit (DWGLU) is introduced to process the input information, addressing dimensional mismatch while maintaining feature integrity [27]. The DWGLU enhances the neural network's efficiency by introducing a learnable gating mechanism that dynamically filters relevant features, improving both computational efficiency and long-term information retention. The DWGLU splits the input into a depthwise convolutional transformation and a sigmoid-based gate, formulated as follows:

$$\text{DWGLU} = (\text{DWConv}_1(x)) \odot \sigma(\text{DWConv}_2(x)) + x \quad (2)$$

where $x \in \mathbb{R}^{N \times L \times C}$ is the input, DWConv_1 and DWConv_2 are depthwise separable convolutions, σ is the sigmoid function, and $\odot$ denotes element-wise multiplication.

2.2.2. Attention-Based SPA Block

The improved spatial pyramid attention (SPA) effectively extracts the structural and regularization information from the original data by leveraging the key features identified through the channel attention and spatial attention [28]. SPA is an advanced attention mechanism designed to improve feature representation by capturing the spatial context at multiple scales. It combines the concept of spatial pyramids, which capture contextual information at different resolutions, with the ability to apply attention to important regions within the feature map. This mechanism enhances the model's focus on critical spatial features while reducing the impacts of irrelevant or noisy areas. Mathematically, SPA can be expressed as follows:

$$\begin{aligned} S(x) &= C(\text{AvgPool}(x, 1), \text{AvgPool}(x, 1), \text{AvgPool}(x, 1)) \\ F_s(x) &= \sigma(FC(FC(S(x)))) \end{aligned} \quad (3)$$

where C indicates the concat operation, P represents the average pooling function, σ indicates the sigmoid activation function, and FC stands for the fully connected layer.

SPA is a variant of the squeeze-and-excitation network (SENet) [29], where the core idea of SENet is to learn feature weights dynamically based on the loss function by assigning higher weights to effective feature maps while suppressing irrelevant or less informative features, ultimately improving the model's performance. Building on this principle, a self-attention mechanism is employed to enhance feature selection by integrating channel attention and spatial attention, emphasizing critical features within global and local contexts. This mechanism effectively captures long-term dependencies by leveraging both global and local domain features, enabling the extraction of crucial information over extended sequences while preserving structured relationships in the data. Mathematically, it can be formulated as follows:

$$\begin{aligned} f_1(x) &= CA(x) \cdot x \\ f_2(x) &= \lambda SA(f_1(x)) \cdot f_1(x) \end{aligned} \quad (4)$$

where λ denotes the sigmoid function, and CA and SA represent, respectively, the channel attention maps and spatial attention. The structure of the CA&SA module is shown in Figure 2.

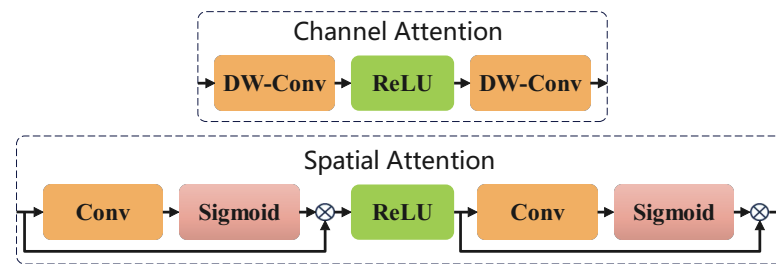


Figure 2. CA&SA module.

2.2.3. Adaptive Spatial Feature Fusion Block

The improved adaptive spatial feature fusion (ASFF) is a dynamic integration mechanism that improves multiscale learning by adaptively adjusting fusion weights across spatial resolutions, which can effectively obtain multiscale information in the RUL, ensuring efficient and context-aware feature representation [30]. By learning weight distributions for each feature map, ASFF emphasizes salient spatial information while suppressing irrelevant features, enabling the model to balance the local details and global context. Additionally, ASFF can effectively extract multiscale information from raw RUL data, refining feature selection to enhance fine-grained recognition and structural understanding. Mathematically, ASFF is formulated as follows:

$$\begin{aligned}
 F_1(x) &= F_1(x) \\
 F_2(x) &= \text{Conv}_1(x) \\
 F_3(x) &= \text{MaxPool}(\text{Conv}_1(x)) \\
 F_a(x) &= w_1 \cdot F_1(\mathcal{FC}_\infty(x)) + w_2 \cdot F_2(\mathcal{FC}_\infty(x)) + w_3 \cdot F_3(\mathcal{FC}_\infty(x))
 \end{aligned} \tag{5}$$

where w_i represents the dynamically learned weight for each feature scale (F_i), $w_1 + w_2 + w_3 = 1$, (where $w_i \in [0, 1]$), Conv_1 represents a convolutional operation with kernel size = 3 and stride = 2, and MaxPool shows a max pooling operation with kernel size = 3 and stride = 2.

$$F_o(x) = W_0(\text{Conv}(\text{C}(F_r(x) + F_s(x) + F_a(x)))) + x \tag{6}$$

where W_0 represents the ReLU activation function, and C indicates the concat operation. Thus, the MCFIB model enhances RUL prediction by extracting essential features, structural and regularization information, and multiscale information from raw data, resulting in a broader receptive field and more robust, accurate predictions of the RUL. The final result can be formulated using Equation (6).

2.3. Dual-Attention Module

The dual-attention network (DANet) is designed to refine the output from the previous step and enhance feature representation, utilizing both channel attention and position attention [31]. The final feature representation in the DANet is computed as a weighted sum of all the channel features combined with the original features, ensuring a more comprehensive feature integration. Unlike CNNs, which are restricted by local receptive fields, and traditional transformers, which primarily model temporal dependencies but struggle to establish strong interactions between input variables, thus limiting their receptive fields, the DANet employs spatial and channel attentions to dynamically emphasize the key features. This adaptive mechanism not only overcomes the limitations of fixed receptive fields but also effectively captures long-range semantic dependencies between feature maps, ultimately enhancing feature discriminability and representation quality. By capturing these dependencies, the DANet strengthens feature representation, making it

particularly advantageous for RUL prediction by prioritizing critical degradation patterns. This improved feature refinement further enhances the overall predictive accuracy and robustness of the algorithm.

The channel attention mechanism in the DANet dynamically enhances the representation of feature maps by learning global channel dependencies, ensuring that the network focuses on the most informative channels while reducing the impact of irrelevant ones. Instead of treating all the channels equally, this mechanism assigns higher attention weights to channels that contribute significantly to feature extraction, improving model's efficiency and discriminative power. This process is mathematically formulated as follows:

$$C(F) = \frac{\exp(R(F) \cdot T(R(F)))}{\sum_{i=1}^C \exp(R(F) \cdot T(R(F)))} \quad (7)$$

$$Cam = \beta \sum_{i=1}^C (T(C(F))F) + F$$

where R indicates the reshape function, T indicates the transpose function, C is the dimension of the model, and β represents a weight that is gradually learned from 0.

The position attention mechanism in the DANet improves feature modeling by learning spatial dependencies across an entire input, allowing the network to capture long-range positional relationships that traditional convolutional layers struggle to process. Unlike the standard spatial attention, which focuses on local regions, the position attention models global interactions between different spatial locations, enhancing contextual understanding. The mechanism is represented as follows:

$$P(F) = \frac{\exp(R(Conv(F)) \cdot T(R(Conv(F))))}{\sum_{i=1}^N \exp(R(Conv(F)) \cdot T(R(Conv(F))))} \quad (8)$$

$$Pam(F) = \alpha \sum_{i=1}^N (T(P(F))R(Conv(F))) + F$$

where R indicates the reshape function, T indicates the transpose function, N is the input sequence length, and α stands for initialization to 0 and gradually learning to assign more weights.

The DANet integrates both channel attention and position attention to refine feature representation across multiple dimensions, ensuring a holistic understanding of both channel-wise and spatial dependencies. By applying independent attention mechanisms to feature channels and spatial positions, the DANet can effectively strengthen the features extracted in the previous step to obtain a more accurate feature map. The final output is computed as follows:

$$F_{DA} = Pam + Cam \quad (9)$$

Using channel-based feature refinement and global spatial dependencies, the DANet provides a powerful framework for effectively strengthening the features extracted in the previous layers and ensuring a more accurate and contextually aware feature map, which leads to better performance in RUL prediction.

2.4. Self-Attention-Based Feature Pyramid Block

The feature pyramid mechanism improves RUL prediction models by effectively fusing multiscale information, allowing for the prediction of targets at different scales. Additionally, the integration of a self-attention mechanism enables the model to capture long-term dependencies, enhancing the accuracy and robustness of RUL predictions over time. Atrous spatial pyramid pooling (ASPP) [32], a variant of the feature pyramid structure,

enhances RUL prediction models by capturing multiscale contextual information through multiple branches of dilated convolutions with different dilation rates. These dilated convolutions effectively extract features across multiple scales, addressing the limitations of fixed receptive fields in CNNs. Each branch, with its unique dilation rate, captures features at varying scales, thereby expanding the receptive field and improving the model's ability to process multiscale information. The ASPP operation can be mathematically expressed as follows:

$$M_A(q) = \sum_{i=1}^N w_i \cdot \text{Conv}_{d_i}(q) + \text{AvgPool}(q) \quad (10)$$

where $\text{Conv}_{d_i}(q)$ represents the dilated convolution applied to the input feature map (q) with dilation rate d_i ; w_i are the learnable weights for each scale branch; $\text{AvgPool}(q)$ represents the global average pooling of the input (q), capturing global contextual features; and M_A is the final fused feature map after applying all the dilated convolutions and global pooling. By fusing multiscale features, ASPP enhances the ability to predict the RUL more accurately. This increases the performance and stability of the RUL prediction model, making it robust under varying conditions. Multiscale fusion through ASPP also prevents overfitting by capturing diverse spatial contexts.

Because of sparse sampling, ASPP lacks a correlation between pieces of long-distance information; the introduction of a non-local (NL) block helps to bridge this gap by computing correlations across all the positions in the feature map [33]. The NL module enhances feature extraction by capturing the long-range dependencies and global context, enabling the model to establish spatial interactions beyond local pixel relationships. By directly computing global interactions, the NL module eliminates the sequential constraints and vanishing gradients commonly observed in LSTMs while also mitigating the limitations of transformers, which struggle to capture subtle long-term dependencies because of fixed-length input constraints. This is achieved by computing the output at each position as a weighted sum of all the positions, with the weight determined by the similarity between feature points, expressed as follows:

$$M_N(r) = \frac{1}{C(i)} \sum_j f(r_i, r_j) g(r_j) \quad (11)$$

where y_i is the output feature at position i , r_i and r_j represent the input features at positions i and j , and $f(r_i, r_j)$ and $g(r_j)$ are the similarity function and transformation function, respectively. By incorporating the NL module with ASPP, the model effectively improves its ability to acquire the global context, compensating for the sparsity of information in long-range dilated convolutions and ultimately boosting performance in RUL predictions that require both fine details and global feature understanding.

The SAFF block enhances RUL prediction by capturing the multiscale context and long-range dependencies, extracting local features while modeling global correlations across spatial positions. This dual capability enables the model to track intricate degradation trends, which are crucial for system wear and failure prediction. By strengthening the model's ability to analyze complex degradation patterns, the SAFF block ensures more accurate and stable RUL estimations, effectively balancing fine details and long-term dependencies.

3. Experimental Details

3.1. Dataset Description

In this study, the C-MAPSS dataset, a benchmark turbofan engine dataset developed by NASA, is used to validate the efficacy of the proposed approach [34]. This dataset simulates

real turbofan engine deterioration, providing a realistic testbed for RUL prediction. A diagram of the simulated turbofan engine is shown in Figure 3.

The C-MAPSS dataset includes four subsets—FD001, FD002, FD003, and FD004—each varying in working conditions, fault modes, lifespans, and the number of engines, as outlined in Table 2. These differences provide a comprehensive basis for evaluating the proposed model’s robustness and generalization across diverse degradation patterns and operational environments.

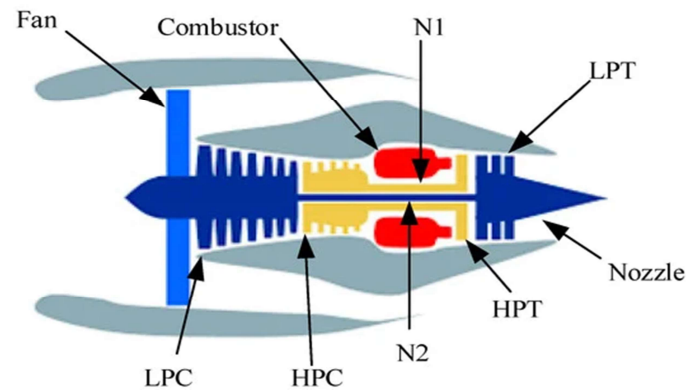


Figure 3. Simplified diagram of an aeroengine simulated in C-MAPSS [34].

Table 2. The detailed description of each subset.

Subsets	Training Engines	Testing Engines	Operating Conditions	Fault Types	Max Life Cycles
FD001	100	100	1	1	362
FD002	260	259	6	1	378
FD003	100	100	1	2	512
FD004	249	248	6	2	128

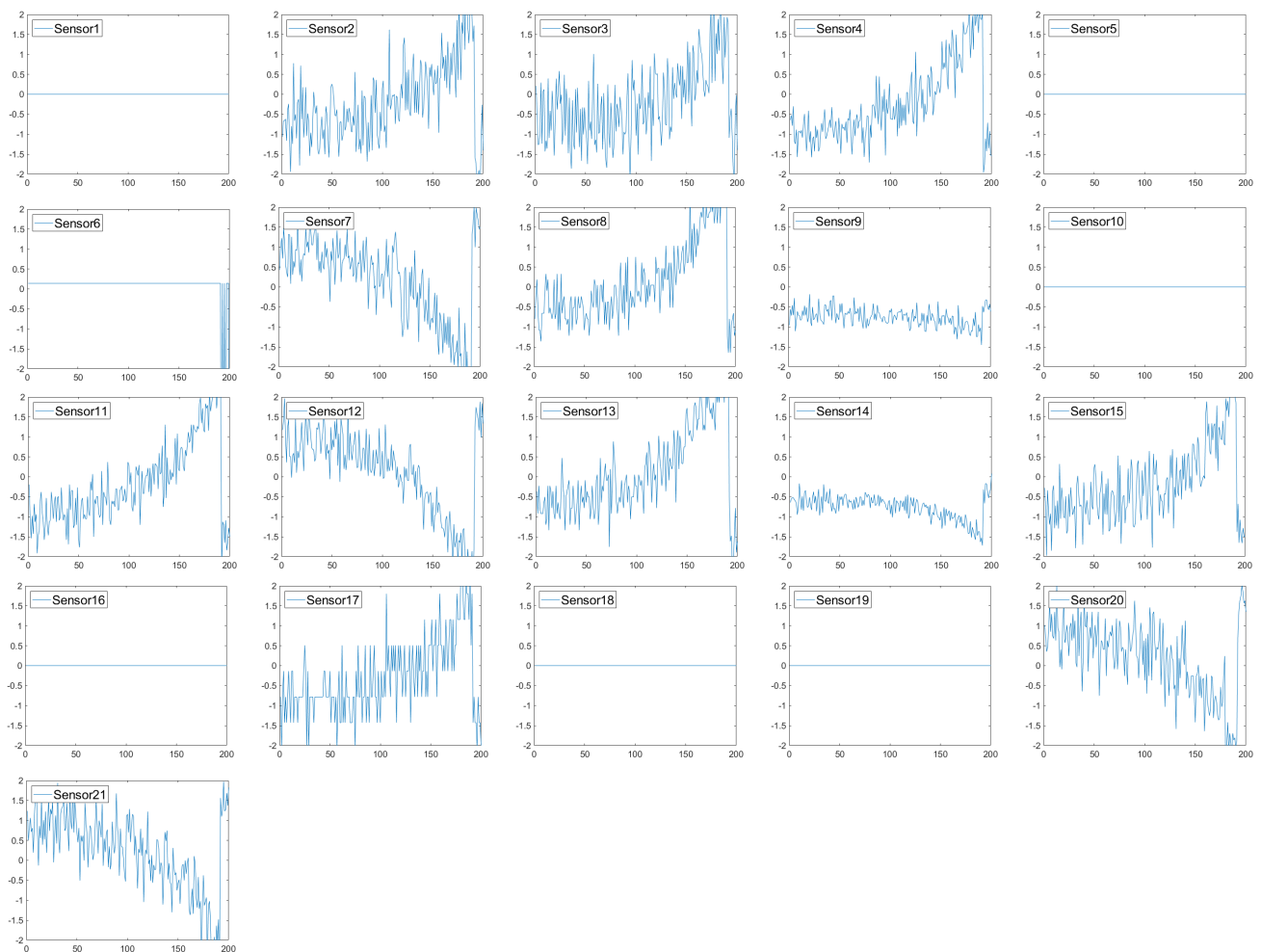
Each C-MAPSS subset consists of training and testing datasets, containing multivariate time series for multiple engine units, where an “engine unit” represents a group of aeroengines of the same type. The training set provides time-series data from the beginning of an engine’s operational cycle until its failure cycle, while the test set includes time series that end just before system failure. Additionally, variations in the initial wear and output conditions are present in the dataset, making it essential to account for differences at the start of each engine’s operation. Each subset includes 26 columns. The first column represents the engine’s numerical identifier, while the second column denotes the cycle index. The subsequent three columns detail the operational configurations of the aeroengines.

3.2. Data Preprocessing

Initially, the sensor selection is guided by prior research, focusing on relevant features that exhibit clear degradation patterns leading up to failure [35]. Figure 4 presents the sensor values from the FD001 subset, where a visual analysis reveals that some sensors remain stable throughout the degradation phase, making them less informative for RUL prediction. As a result, S1, S5, S6, S10, S16, S18, and S19 were excluded, while the remaining 14 sensors were retained as model inputs. Table 3 provides the detailed specifications for the selected sensors.

Table 3. Descriptions of the sensors.

No. of Sensors	Symbols	Description	Unit(s)
2	T24	Total temperature at LPC outlet	°R
3	T30	Total temperature at HPC outlet	°R
4	T50	Total temperature at LPT outlet	°R
7	P30	Total pressure at HPC outlet	psia
8	Nf	Physical fan speed	rpm
9	Nc	Physical core speed	rpm
11	Ps30	Static pressure at HPC outlet	psia
12	phi	Ratio of fuel flow to Ps30	pps/psi
13	NRf	Corrected fan speed	rpm
14	NRc	Corrected core speed	rpm
15	BPR	Bypass ratio	-
17	htBleed	Bleed enthalpy	-
20	W31	HPT coolant bleed	lbm/s
21	w32	LPT coolant bleed	lbm/s

**Figure 4.** Simplified diagram of a simulation in C-MAPSS [34].

The model's training parameters should be unaffected by the varying ranges of each sensor. Min–max normalization, also known as feature scaling or min–max scaling, is a data preprocessing technique used to standardize feature values to a uniform scale, typically $[0, 1]$ or $[-1, 1]$. This methodology is highly advantageous in machine learning and data mining as it addresses the issue of disparate ranges or units among features. Standardizing features to a uniform scale can enhance the efficacy of algorithms. In this study, the min–

max normalization technique was used to standardize the raw data [36], as implemented using Equation (12):

$$F(x) = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (12)$$

where x is the original data value; x_{min} and x_{max} denote the minimum and maximum values of the features, respectively; and F represents the normalized value. Given that the same sensor type may produce varying readings under different conditions, min–max normalization is applied under each specific operating condition [37].

Understanding the relationships between neighboring time intervals is crucial in time-series analysis. The sliding window approach captures these dependencies by segmenting data into fixed time windows, where each time point's input corresponds to an RUL value. The stride of the sliding window is set at one, maximizing the number of samples for training. Figure 5 illustrates this process. The window size must be shorter than the minimum life cycle, ensuring effective data representation. Through experimentation, a window length of 25 steps was chosen for the optimal performance.

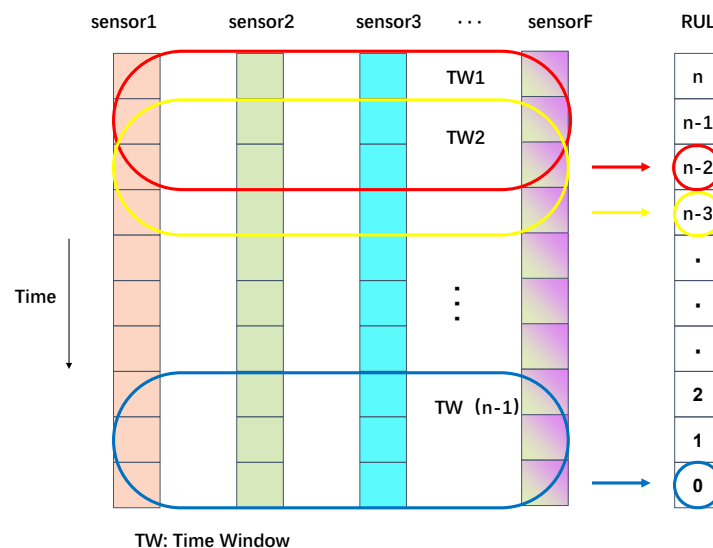


Figure 5. Brief explanation of sliding window handling.

Following previous studies [35,38], the RUL's upper limit is set at 125. Initially, the RUL label remains constant before transitioning to linear degradation until system failure. This assumption allows for a structured degradation model, making RUL estimation more reliable.

3.3. Implementation Details

The experiments were conducted on a workstation with specific hardware and software configurations. The system is powered by an Intel Core i7-14700KF CPU (3.40 GHz) and features an NVIDIA GeForce RTX 4070 GPU, providing the computational power required for model training. Additionally, the workstation is equipped with 32 GB of RAM, ensuring efficient data processing and training performances.

The proposed model was trained and optimized using the AdaX optimization algorithm, a novel adaptive learning rate method introduced in [39]. AdaX combines the advantages of AdaGrad and Adam, offering improved convergence speed and enhanced generalization, making it well suited for RUL prediction tasks. The initial learning rate was set at 0.001 and decreased by 25% after 18 epochs.

3.4. Overall Structure

Figure 6 presents the flowchart of the proposed architecture, structured into three main phases: data preparation, model construction, and RUL prediction. The model construction and RUL prediction phases incorporate procedures similar to those of data preparation, including sensor selection and data normalization. In the model construction phase, the SCAB model is developed and trained using the training dataset. Once trained, the model is applied to the testing datasets to perform RUL prediction, ensuring an accurate estimation of the remaining useful life.

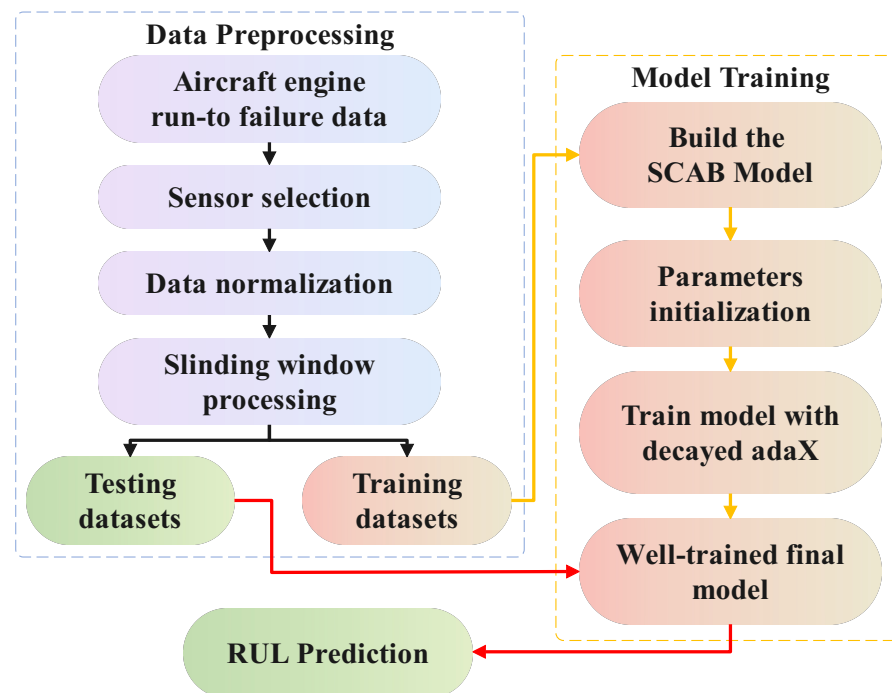


Figure 6. Flowchart of RUL prediction.

3.5. Evaluation Metrics

Two commonly employed metrics are used to evaluate the effectiveness of the proposed models for RUL outcomes. The first metric is the root-mean-square error (RMSE), while the second is known as the score.

- **RMSE:** The RMSE is commonly used to evaluate prediction models for RUL estimation, as it uniformly penalizes both premature and delayed predictions. The RMSE is defined using Equation (13) as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - y_i)^2} \quad (13)$$

where N is the total number of prediction samples, P_i represents the predicted value, and y_i denotes the true value;

- **Score:** The score was introduced at the International Conference on Prognostics and Health Management (PHM) as a method for evaluating models in data challenges [34]. The score is formulated using Equation (14) as follows:

$$Score = \begin{cases} \sum_{i=1}^N \exp^{-(h_i/13)} - 1, & \text{for } h_i < 0 \\ \sum_{i=1}^N \exp^{(h_i/10)} - 1, & \text{for } h_i \geq 0 \end{cases} \quad (14)$$

where h_i represents the prediction error for each instance.

4. Experimental Result Analysis and Discussion

4.1. Comparison with State-of-the-Art Methods

In this section, the proposed method is compared with several state-of-the-art approaches. Table 4 presents the RUL estimation results across the four subsets of the C-MAPSS dataset, alongside findings from 19 prior studies. These previous studies can be broadly categorized into three groups: (1) DL-based methods [10,40–45], (2) transformer-based methods [24,46–49], and (3) attention-based methods [17,19,36,38,50–52].

Table 4. Performance comparison between the proposed method and other approaches found in the literature.

Method	Year	FD001		FD002		FD003		FD004	
		RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
LSTM [40]	2018	16.14	339.00	24.49	4450.00	16.18	852.00	28.17	5550.00
Bi-LSTM [10]	2018	13.65	295.00	23.18	4130.00	13.74	317.00	24.68	5430.00
CNN-LSTM [41]	2019	14.40	290	27.23	9869	14.32	316	26.69	6594
HDNN [42]	2019	13.02	245	15.24	1282	12.22	288	18.16	1527
AGCNN [38]	2020	12.42	225.51	19.42	1492.00	13.39	227.09	21.50	3392.00
MS-DCNN [43]	2020	11.44	196.22	19.35	3747.00	11.67	241.87	22.22	4844.00
AEQRNN [36]	2021	‡	‡	19.10	3220	‡	‡	20.67	4594
GCU-Transformer [46]	2021	11.27	‡	22.81	‡	11.42	‡	24.86	‡
Double-Attention Network [47]	2022	12.25	198	17.08	1575	13.39	290	15.66	1741
GAM-CapsNet [17]	2022	12.42	273	17.38	1872	11.73	243	19.83	2490
BiGRU-TSAM [24]	2022	12.59	213.35	18.94	2264.13	12.45	232.86	20.47	3610.34
LSTM-Based Model [45]	2022	7.78	100	17.64	1440	8.03	104	17.63	2390
SCTA-LSTM [19]	2023	12.10	207	16.90	1267	12.14	248	21.93	3310
NT-TCN [44]	2023	<u>11.18</u>	203.94	16.53	1350.39	12.02	364.46	19.7	5137.81
STRUL [50]	2023	12.85	224	19.24	1950	13.74	252	22.34	3080
DLformer [48]	2024	‡	‡	15.93	1283.63	‡	‡	15.86	1601.45
BTCAN [51]	2024	14.46	309	19.88	2800	12.79	298	22.03	4224
DVGTFformer [49]	2024	11.33	<u>179.75</u>	14.28	797.26	11.89	254.55	15.5	1107.5
DDCA-TCN [52]	2024	12.14	237.6	<u>12.85</u>	<u>760.5</u>	11.37	225.5	13.86	903.7
Proposed Method	2024	11.68	209.67	12.79	749.22	<u>11.31</u>	<u>220.36</u>	<u>14.62</u>	<u>961.73</u>

Table 4 presents the updated comparative analysis of the SCAB model against prior methods in the C-MAPSS benchmark dataset. The last row highlights the SCAB model's performance, demonstrating its strong predictive capabilities. Notably, the SCAB model surpasses the other methods in the FD002 subset, achieving a 1.48% improvement in the score metric. Although its performances in FD001 and FD003 are slightly below those of the best method, the SCAB model achieves substantial improvements of 47.97% in FD002 and

59.76% in FD004 compared to the performances of the same method. Additionally, despite a slightly lower performance in FD004, the SCAB model achieves an 11.75% improvement over the same method in FD001.

For the RMSE metric, the SCAB model demonstrates a 0.46% improvement in FD002. Although it does not achieve the best performance in FD004, it still improves by 3.78% over the same method in FD001. Furthermore, although the SCAB model's performances in FD001 and FD003 are slightly lower than those of the best method, it outperforms the same method in FD002 and FD004 by 27.49% and 17.07%, respectively. These findings reinforce the SCAB model's robustness across different subsets, ensuring consistently high performance and adaptability to varying degradation scenarios.

To comprehensively validate the RUL estimation capability of the proposed method, Figure 7 illustrates a comparison between the predicted and actual RULs across the FD001, FD002, FD003, and FD004 sub-datasets.

- As shown in Figure 7a, the RUL estimation results for the FD001 dataset demonstrate that the SCAB model generally predicts the RUL accurately, especially when the actual RUL is low, with predictions closely aligning with the true values. However, in areas where the actual RUL values are higher, the predicted values are often underestimated. This discrepancy could be because of the rapid degradation of the engine. Additionally, multiple indicators may enter the high-risk stage more quickly, leading to a misalignment between the model's predictions and the actual degradation process. Moreover, specific intervals, such as 0–18 and 85–95, show higher prediction errors;
- Figure 7b shows the RUL estimation results for the FD002 dataset. In this dataset, the model tends to underestimate the RUL, with most predictions being more accurate at lower RUL values. This pattern can be attributed to the engine degradation characteristics, where the engine enters the degradation state later than expected, particularly in complex environments. Additionally, the inherent complexity of the FD002 dataset and its nonlinear fluctuations may make it more difficult for the model to handle long-term predictions, as the system is subjected to more dynamic and intricate conditions over time;
- As shown in Figure 7c, the RUL estimation results for FD003 are similar to those of FD001, showing better accuracy at lower actual RULs and larger deviations at higher RULs, likely because of the similar characteristics of the datasets. Unlike in FD001, some segments in FD003, such as 60–75, show an overestimation in the RUL, indicating a horizontal offset between the predicted and actual values;
- As depicted in Figure 7d, the RUL estimation for the FD004 dataset mirrors that of FD002, with predictions generally closely aligning with the true values at low RULs and lower at high RULs, especially in the intervals of 40–50 and 160–170, where the discrepancy between the predicted and actual values is pronounced.

Overall, the performance of the SCAB model across multiple sub-datasets confirms its superior predictive capabilities, particularly for lower actual RULs, where its accuracy is notably higher. However, the model exhibits some errors in predicting long-life engines, highlighting a limitation in detecting engine degradation in complex, long-duration environments. This indicates a need for improvement in capturing subtle degradation patterns that occur over extended periods. These findings suggest that although the SCAB model demonstrates robust predictive abilities across various subsets, further optimization and adjustments are required to enhance its ability to handle complex degradation scenarios.

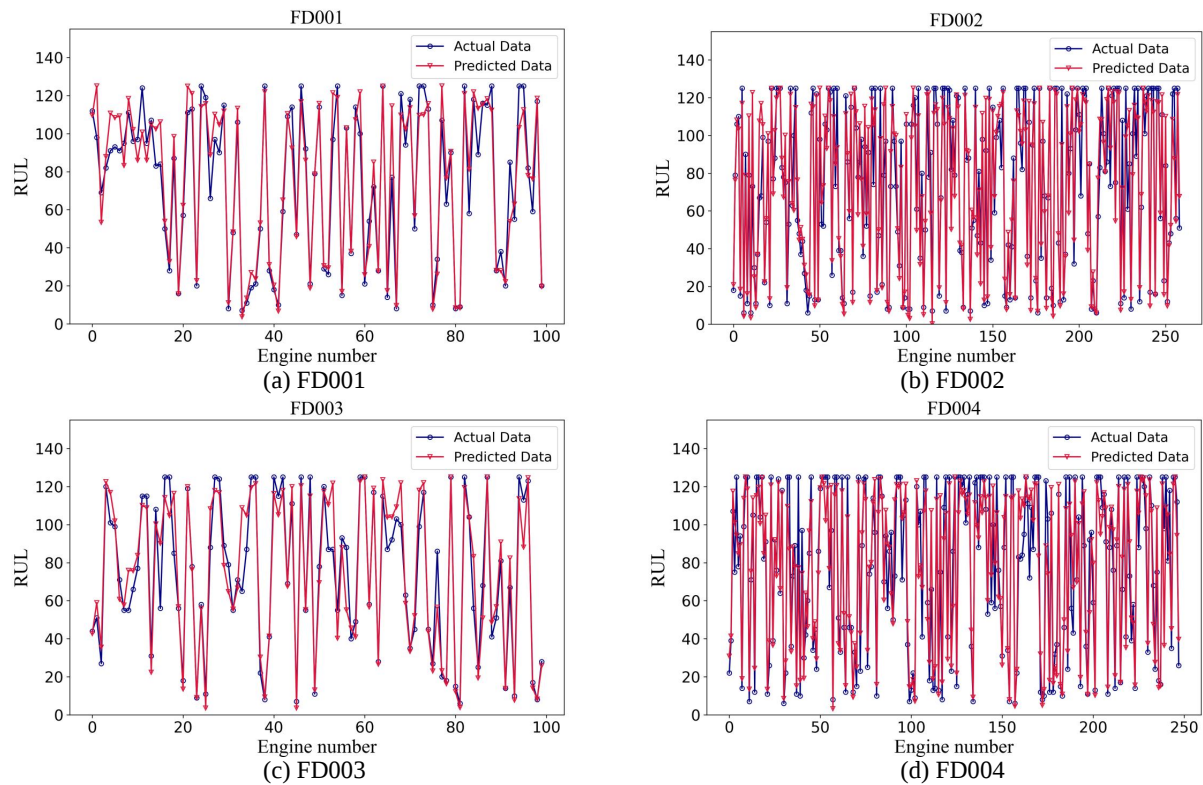


Figure 7. Prediction of each engine in four subsets.

To enhance the clarity and interpretability of the prediction results, the observed values were compared with the predicted ones. By randomly selecting specific engines from each sub-dataset and juxtaposing the actual RUL with the predicted RUL for each cycle, the model's predictive accuracy is clearly demonstrated. As shown in Figure 8, predictions for randomly selected engines from the four different subsets are displayed. The orange curve represents the predicted RUL values, while the red curve denotes the actual RUL values. Specifically, Figure 8a presents the prediction results for engine unit 34 from the FD001 test set, Figure 8b shows the results for engine unit 43 from the FD002 test set, Figure 8c illustrates the results for engine unit 46 from the FD003 test set, and Figure 8d depicts the degradation trajectory prediction for engine unit 40 from the FD004 test set.

From Figure 8, it is worth noting that in the FD001 and FD003 datasets, when the engine enters the degradation stage, the RUL predicted by the model shows obvious oscillations. After a period of time, the predicted results stabilize and become closer to the actual RUL values. In contrast, for the FD002 and FD004 datasets, the predicted RUL value begins to approach the actual value after a brief fluctuation once the engine starts to degrade. This pattern suggests that in the C-MAPSS dataset, when the engine begins to degrade, the predicted value initially deviates from the true value because of changes in the situation. However, after some time, the model stabilizes, and the predicted value aligns more closely with the actual value. This behavior clearly indicates that the SCAB network adapts well to different degradation scenarios and provides reliable predictions across various datasets.

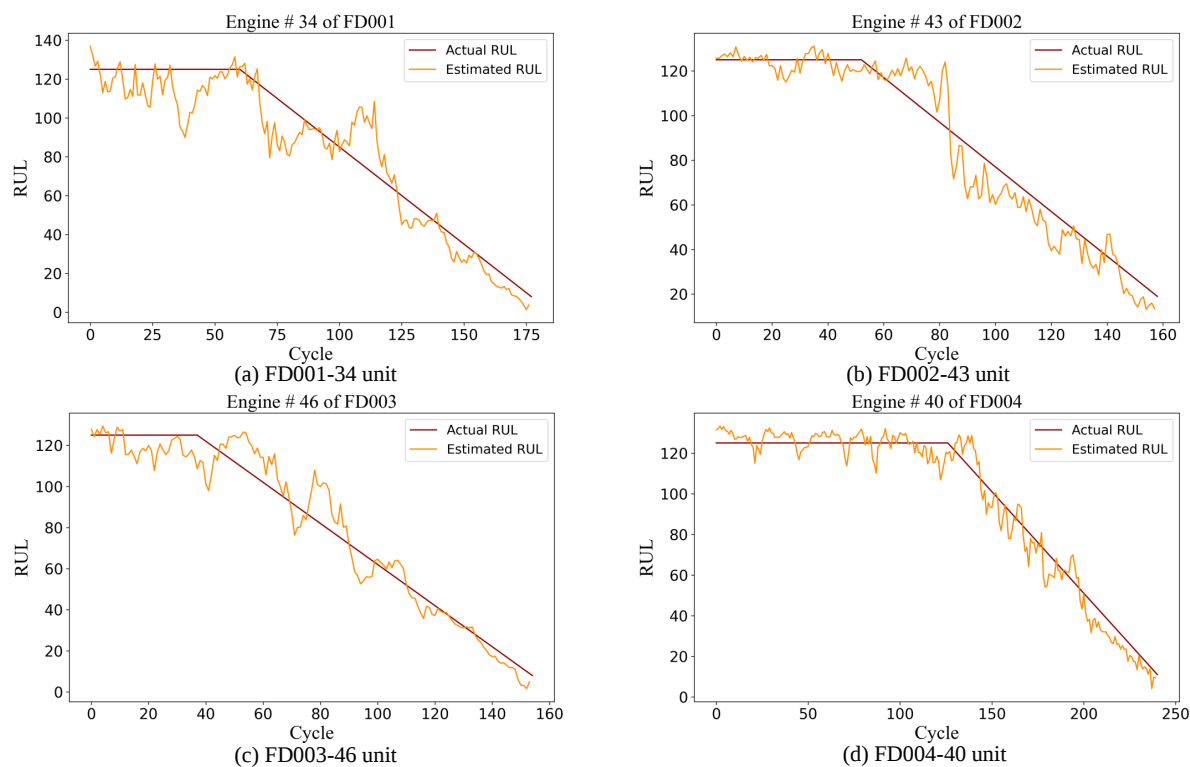


Figure 8. Full life cycle prediction of a randomly selected engine in each subset.

4.2. Ablation Study of the Proposed Method

To comprehensively evaluate the efficacy of each internal module within the proposed method, this section presents five ablation experiments, with the benchmark model (the MCFIB model), Experiment No. 1, serving as the baseline for comparison. Experiment No. 2 builds upon Model No. 1 by incorporating the ASPP module, resulting in the MCFIB + ASPP model. Experiment No. 3 builds upon Model No. 2 by replacing the ASPP module with the SAFF module, creating a model that combines the MCFIB and the SAFF, to assess the performance of the SAFF module. Experiment No. 4 builds upon Model No. 3 by replacing the MCFIB module with the DANet module, resulting in the DA + SAFF, to assess the performance of the DANet module. Finally, Experiment No. 5 presents the full SCAB model, which integrates all the modules, including the MCFIB, the DANet, and the SAFF.

As shown in Table 5, the experimental results are as follows:

- The proposed MCFIB module, serving as the benchmark model, has a certain impact on RUL predictions, but its performance remains limited. This indicates that the hierarchical and rich features extracted from multiple channels require further processing to achieve the optimal results;
- In comparing Experiment 2 and Experiment 1, the integration of ASPP brings significant improvements to the FD004 dataset, reducing the RMSE by 5.35% and improving the score by 24.45%. However, it leads to a decline in performance in the simpler FD003 dataset, suggesting that although ASPP effectively enhances feature extraction in complex scenarios with varying fault modes, it may cause overfitting in simpler environments;
- Further analysis in Experiment 3 and Experiment 2, where ASPP is replaced with SAFF, reveals significant enhancements in the model's performance. Specifically, SAFF reduces the RMSE by 9.29%, 7.74%, and 28.98%, and improves the score by 26.61%, 26.63%, and 76.34% in FD001, FD002, and FD003, respectively. This demonstrates that SAFF can effectively strengthen hierarchical features, achieving good performance

- and stability across both simple and complex situations while maintaining strong generalization ability;
- When comparing Experiment 4 and Experiment 3, replacing the MCFIB with the DANet results in substantial improvements in FD003, with the RMSE reduced by 18.85% and the score increased by 41.61%. However, the performance slightly declines in the other three datasets, indicating that the DANet improves the overall algorithmic stability;
 - The final model, Experiment 5, which integrates all the proposed modules, outperforms the other ablation results in most datasets. Notably, it achieves consistent improvements in both the score and RMSE metrics in FD001, FD002, and FD003, highlighting its superior performance.

Table 5. Ablation study of the proposed architecture in all the datasets.

No.	Model Structure	FD001		FD002		FD003		FD004	
		RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
No. 1	MCFIB	13.25	335.81	14.86	1091.38	20.24	1513.16	16.07	1360.41
No. 2	MCFIB+ASPP	13.56	305.71	14.34	1026.61	19.49	1643.98	15.21	1027.69
No. 3	MCFIB+SAFP	12.3	224.35	13.23	753.22	13.84	388.86	14.63	960.62
No. 4	DA+SAFP	12.19	243.35	13.57	767.96	11.33	227.03	14.95	981.33
No. 5	MCFIB+DA+SAFP	11.68	209.67	12.79	749.22	11.31	220.36	14.62	961.73

These ablation experiments validate the effectiveness of each component in the model, demonstrating the advanced capabilities, stability, and practical utility of the SCAB model in aerospace engines' RUL estimations, emphasizing the model's value in real-world applications.

5. Conclusions

In this study, the scope-coordinated attention-based (SCAB) network is proposed for remaining-useful-life (RUL) estimation in aeroengines. This model leverages a multichannel network to effectively extract features from raw data, incorporating various types of information, such as raw data, structured and regularized information, and multiscale data. This enables the model to expand its receptive field and capture rich, diverse features. To further enhance feature extraction, a DANet is introduced to refine the features extracted in the previous step by capturing the significance of the information in both the channel and spatial domains. Additionally, an adaptive self-attention-enhanced feature pyramid network (SAFP) is employed to expand multiscale information and establish long-term dependencies, improving the model's ability to learn complex relationships and accurately predict outcomes over extended periods. Notably, the proposed modules are implemented within an end-to-end architecture, enabling the model to directly process sequential raw sensor data and output RUL predictions without manual intervention.

The SCAB network's effectiveness has been thoroughly validated in the C-MAPSS dataset, with comparative analysis against state-of-the-art (SOTA) approaches confirming its superior performance. The SCAB network consistently outperforms the other methods in the FD002 subset. Although its performances in FD001 and FD003 are slightly below those of the best methods, the SCAB network demonstrates significant advantages in FD002 and FD004 compared to the performances of the same methods. Additionally, in FD004, the SCAB network exhibits remarkable performance, surpassing those of all the other methods except for the best method. These results highlight the SCAB network's strong stability and

reliability in RUL prediction across diverse aeroengine scenarios. Ablation studies further validated the necessity and effectiveness of each module.

The prediction performance of the proposed method may be affected when the method is applied to RUL predictions under unknown operating conditions. To address this, we plan to integrate domain adaptation techniques to enhance the model's transferability across various operating environments and devices. Furthermore, we recognize the importance of explainable AI in improving model interpretability. Future work will focus on incorporating explainability methods to enhance transparency and usability in predictive maintenance. Additionally, we will continue to optimize the proposed method and explore its real-world applications to ensure its practical effectiveness.

Author Contributions: Conceptualization, validation, and writing—original draft preparation, Z.L.; data preprocessing, S.L.; visualization, J.L. and S.M.; supervision and project administration, G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study are openly available at the following URL/DOI: <https://c3.ndc.nasa.gov/dashlink/resources/139>.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could appear to have influenced the work reported in this document.

References

1. Zio, E. Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. *Reliab. Eng. Syst. Saf.* **2022**, *218*, 108119. [CrossRef]
2. Rezazadeh, N.; Perfetto, D.; De Luca, A.; Caputo, F. Ensemble learning for estimating remaining useful life: Incorporating linear, KNN, and Gaussian process regression. In Proceedings of the International Workshop on Autonomous Remanufacturing, Caserta, Italy, 18–19 October 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 201–212.
3. Choi, J.H.; An, D.W.; Gang, J.H. A survey on prognostics and comparison study on the model-based prognostics. *J. Inst. Control Robot. Syst.* **2011**, *17*, 1095–1100. [CrossRef]
4. Qian, Y.; Yan, R.; Gao, R.X. A multi-time scale approach to remaining useful life prediction in rolling bearing. *Mech. Syst. Signal Process.* **2017**, *83*, 549–567. [CrossRef]
5. Rathore, M.M.; Shah, S.A.; Shukla, D.; Bentafat, E.; Bakiras, S. The role of AI, machine learning, and big data in digital twinning: A systematic literature review, challenges, and opportunities. *IEEE Access* **2021**, *9*, 32030–32052. [CrossRef]
6. Li, X.; Ding, Q.; Sun, J.Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliab. Eng. Syst. Saf.* **2018**, *172*, 1–11. [CrossRef]
7. Chen, Z.; Cao, S.; Mao, Z. Remaining Useful Life Estimation of Aircraft Engines Using a Modified Similarity and Supporting Vector Machine (SVM) Approach. *Energies* **2017**, *11*, 28. [CrossRef]
8. Javed, K.; Gouriveau, R.; Zerhouni, N. A New Multivariate Approach for Prognostics Based on Extreme Learning Machine and Fuzzy Clustering. *IEEE Trans. Cybern.* **2015**, *45*, 1. [CrossRef]
9. Zheng, S.; Ristovski, K.; Farahat, A.; Gupta, C. Long Short-Term Memory Network for Remaining Useful Life estimation. In Proceedings of the 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), Dallas, TX, USA, 19–21 June 2017.
10. Wang, J.; Wen, G.; Yang, S.; Liu, Y. Remaining useful life estimation in prognostics using deep bidirectional LSTM neural network. In Proceedings of the 2018 Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, China, 26–28 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1037–1042.
11. Xia, P.; Huang, Y.; Qin, C.; Xiao, D.; Gong, L.; Liu, C.; Du, W. Adaptive feature utilization with separate gating mechanism and global temporal convolutional network for remaining useful life prediction. *IEEE Sensors J.* **2023**, *23*, 21408–21420. [CrossRef]
12. Yu, W.; Kim, I.Y.; Mechefske, C. Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme. *Mech. Syst. Signal Process.* **2019**, *129*, 764–780. [CrossRef]
13. Xiang, S.; Qin, Y.; Luo, J.; Pu, H. Spatiotemporally Multidifferential Processing Deep Neural Network and its Application to Equipment Remaining Useful Life Prediction. *IEEE Trans. Ind. Inform.* **2022**, *18*, 7230–7239. [CrossRef]
14. Ta, Q.B.; Pham, Q.Q.; Pham, N.L.; Huynh, T.C.; Kim, J.T. Smart Aggregate-Based Concrete Stress Monitoring via 1D CNN Deep Learning of Raw Impedance Signals. *Struct. Control Health Monit.* **2024**, *2024*, 5822653. [CrossRef]

15. da Costa, P.R.d.O.; Akçay, A.; Zhang, Y.; Kaymak, U. Remaining useful lifetime prediction via deep domain adaptation. *Reliab. Eng. Syst. Saf.* **2020**, *195*, 106682. [\[CrossRef\]](#)
16. Chen, Z.; Wu, M.; Zhao, R.; Guretno, F.; Yan, R.; Li, X. Machine remaining useful life prediction via an attention-based deep learning approach. *IEEE Trans. Ind. Electron.* **2020**, *68*, 2521–2531. [\[CrossRef\]](#)
17. Zhao, C.; Huang, X.; Li, Y.; Li, S. A novel remaining useful life prediction method based on gated attention mechanism capsule neural network. *Measurement* **2022**, *189*, 110637. [\[CrossRef\]](#)
18. Li, X.; Jiang, H.; Liu, Y.; Wang, T.; Li, Z. An integrated deep multiscale feature fusion network for aeroengine remaining useful life prediction with multisensor data. *Knowl.-Based Syst.* **2022**, *235*, 107652. [\[CrossRef\]](#)
19. Tian, H.; Yang, L.; Ju, B. Spatial correlation and temporal attention-based LSTM for remaining useful life prediction of turbofan engine. *Measurement* **2023**, *214*, 112816. [\[CrossRef\]](#)
20. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [\[CrossRef\]](#)
21. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
22. Fang, Y.; Liao, B.; Wang, X.; Fang, J.; Qi, J.; Wu, R.; Niu, J.; Liu, W. You only look at one sequence: Rethinking transformer in vision through object detection. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 26183–26197.
23. Liu, Z.; Luo, S.; Li, W.; Lu, J.; Wu, Y.; Sun, S.; Li, C.; Yang, L. Convtransformer: A convolutional transformer network for video frame synthesis. *arXiv* **2020**, arXiv:2011.10185.
24. Zhang, J.; Jiang, Y.; Wu, S.; Li, X.; Luo, H.; Yin, S. Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism. *Reliab. Eng. Syst. Saf.* **2022**, *221*, 108297. [\[CrossRef\]](#)
25. Zhang, Z.; Song, W.; Li, Q. Dual-Aspect Self-Attention Based on Transformer for Remaining Useful Life Prediction. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [\[CrossRef\]](#)
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
27. Shi, D. TransNeXt: Robust Foveal Visual Perception for Vision Transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 17773–17783.
28. Guo, J.; Ma, X.; Sansom, A.; McGuire, M.; Kalaani, A.; Chen, Q.; Tang, S.; Yang, Q.; Fu, S. Spanet: Spatial pyramid attention network for enhanced image recognition. In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME), London, UK, 6–10 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
29. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
30. Liu, S.; Huang, D.; Wang, Y. Learning spatial fusion for single-shot object detection. *arXiv* **2019**, arXiv:1911.09516.
31. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
32. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [\[CrossRef\]](#)
33. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
34. Saxena, A.; Goebel, K.; Simon, D.; Eklund, N. Damage propagation modeling for aircraft engine run-to-failure simulation. In Proceedings of the 2008 International Conference on Prognostics and Health Management, Denver, CO, USA, 6–9 October 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 1–9.
35. Siahpour, S.; Li, X.; Lee, J. A novel transfer learning approach in remaining useful life prediction for incomplete dataset. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [\[CrossRef\]](#)
36. Cheng, Y.; Hu, K.; Wu, J.; Zhu, H.; Shao, X. Autoencoder quasi-recurrent neural networks for remaining useful life prediction of engineering systems. *IEEE/ASME Trans. Mechatron.* **2021**, *27*, 1081–1092. [\[CrossRef\]](#)
37. Ragab, M.; Chen, Z.; Wu, M.; Foo, C.S.; Kwok, C.K.; Yan, R.; Li, X. Contrastive adversarial domain adaptation for machine remaining useful life prediction. *IEEE Trans. Ind. Inform.* **2020**, *17*, 5239–5249. [\[CrossRef\]](#)
38. Liu, H.; Liu, Z.; Jia, W.; Lin, X. Remaining useful life prediction using a novel feature-attention-based end-to-end approach. *IEEE Trans. Ind. Inform.* **2020**, *17*, 1197–1207. [\[CrossRef\]](#)
39. Li, W.; Zhang, Z.; Wang, X.; Luo, P. Adax: Adaptive gradient descent with exponential long term memory. *arXiv* **2020**, arXiv:2004.09740.
40. Zhang, J.; Wang, P.; Yan, R.; Gao, R.X. Long short-term memory for machine remaining life prediction. *J. Manuf. Syst.* **2018**, *48*, 78–86. [\[CrossRef\]](#)
41. Wu, Z.; Yu, S.; Zhu, X.; Ji, Y.; Pecht, M. A weighted deep domain adaptation method for industrial fault prognostics according to prior distribution of complex working conditions. *IEEE Access* **2019**, *7*, 139802–139814. [\[CrossRef\]](#)

42. Al-Dulaimi, A.; Zabihi, S.; Asif, A.; Mohammadi, A. A multimodal and hybrid deep neural network model for remaining useful life estimation. *Comput. Ind.* **2019**, *108*, 186–196. [[CrossRef](#)]
43. Li, H.; Zhao, W.; Zhang, Y.; Zio, E. Remaining useful life prediction using multi-scale deep convolutional neural network. *Appl. Soft Comput.* **2020**, *89*, 106113. [[CrossRef](#)]
44. Liu, B.; Xu, J.; Sun, C.; Cui, X.; Xie, X.; Zhi, H. Remaining useful life prediction combining temporal convolutional network with nonlinear target function. *Meas. Sci. Technol.* **2022**, *34*, 034005. [[CrossRef](#)]
45. Asif, O.; Haider, S.A.; Naqvi, S.R.; Zaki, J.F.; Kwak, K.S.; Islam, S.R. A deep learning model for remaining useful life prediction of aircraft turbofan engine on C-MAPSS dataset. *IEEE Access* **2022**, *10*, 95425–95440. [[CrossRef](#)]
46. Mo, Y.; Wu, Q.; Li, X.; Huang, B. Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit. *J. Intell. Manuf.* **2021**, *32*, 1997–2006. [[CrossRef](#)]
47. Liu, L.; Song, X.; Zhou, Z. Aircraft engine remaining useful life estimation via a double attention-based data-driven architecture. *Reliab. Eng. Syst. Saf.* **2022**, *221*, 108330. [[CrossRef](#)]
48. Ren, L.; Wang, H.; Huang, G. DLformer: A Dynamic Length Transformer-Based Network for Efficient Feature Representation in Remaining Useful Life Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 5942–5952. [[CrossRef](#)]
49. Wang, L.; Cao, H.; Ye, Z.; Xu, H.; Yan, J. DVGFormer: A dual-view graph Transformer to fuse multi-sensor signals for remaining useful life prediction. *Mech. Syst. Signal Process.* **2024**, *207*, 110935. [[CrossRef](#)]
50. Wang, T.; Li, X.; Wang, W.; Du, J.; Yang, X. A spatiotemporal feature learning-based RUL estimation method for predictive maintenance. *Measurement* **2023**, *214*, 112824. [[CrossRef](#)]
51. Ren, L.; Li, S.; Laili, Y.; Zhang, L. BTCAN: A Binary Trend-Aware Network for Industrial Edge Intelligence and Application in Aero-Engine RUL Prediction. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–10. [[CrossRef](#)]
52. Zhu, Y.; Liu, Z. A Dual-Dimension Convolutional-Attention Module for Remaining Useful Life Prediction of Aeroengines. *Aerospace* **2024**, *11*, 809. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.