machines

MDPI

Article

Remaining Useful Life Prediction of Aeroengines Based on Multi-Head Attention Mechanism

Lei Nie, Shiyi Xu *, Lvfan Zhang, Yehan Yin, Zhengqiong Dong and Xiangdong Zhou

Hubei Key Laboratory of Modern Manufacturing Quantity Engineering, Hubei University of Technology, Wuhan 430068, China; leinie@hbut.edu.cn (L.N.); zhanglvfan@163.com (L.Z.); tina011015@163.com (Y.Y.); dongzhq@hbut.edu.cn (Z.D.); xdz@hbut.edu.cn (X.Z.)

* Correspondence: xushiyi97@163.com

Abstract: Aeroengines are the core components of an aircraft; therefore, their health determines flight safety. Currently, owing to their complex structure and problems associated with their various detection parameters, predicting the remaining useful life (RUL) of aeroengines is very important to ensure their safety and reliability. In this paper, we propose a new hybrid method based on convolutional neural networks (CNN), timing convolutional neural networks (TCN), and the multi-head attention mechanism. Firstly, an CNN-TCN model is established for multi-dimensional features, in which two layers of the CNN extract features of multi-dimensional input data, and the TCN process the timing features. Subsequently, the outputs of multiple CNN-TCNs are weighted using the multi-head attention mechanism, and the results are stitched together. Next, we compare the root mean square error (RMSE) and scores of various RUL prediction methods to show the superiority of the proposed method. The results showed that compared with previous research results, the RMSE and Score of FD001 decreased by 10.87% and 42.57%, respectively, whereas those of FD003 decreased by 14.13% and 58.15%, respectively.

Keywords: convolutional neural network; temporal convolutional network; multi-head attention; aeroengine; remaining useful life

check for updates

Citation: Nie, L.; Xu, S.; Zhang, L.; Yin, Y.; Dong, Z.; Zhou, X. Remaining Useful Life Prediction of Aeroengines Based on Multi-Head Attention Mechanism. *Machines* **2022**, *10*, 552. https://doi.org/10.3390/machines10070552

Academic Editor: Hongrui Cao

Received: 14 June 2022
Accepted: 5 July 2022
Published: 8 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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

As a core component of the aircraft, the health of the aeroengine determines the flight safety [1]. Therefore, the prediction of the remaining useful life (RUL) of aeroengines is crucial, as it could help engineers in making reasonable maintenance decisions, reducing the cost of airline operations, and improving flight quality [2–4].

Currently, according to the prediction principle, there are two main categories of RUL prediction: physical failure, data-driven [5,6] and hybrid models [7,8]. The physical failure model-based approach combines a priori knowledge of the composition, mechanical dynamics principles, and degradation mechanisms of the equipment with sensor monitoring data to construct the RUL physical prediction model. Although this method can achieve high prediction accuracy, it is less versatile, and the modelling process is complex. The data-driven approach extracts useful information from sensor monitoring parameters and uses data analysis to mine valid features to characterize the health of an aeroengine. This technique achieves less accurate results than the previous approach but is easier to use and has better flexibility. The hybrid approach does not avoid the problem that physical failure models are difficult to obtain. Therefore, the data-driven RUL prediction method has been extensively studied in complex devices modeling.

The data-driven RUL prediction method commonly applies two prediction schemes: (1) data fusion is used to map multi-sensor monitoring data to a one-dimensional (1D) health indicator (HI), and then the HI is used for RUL prediction; (2) multisensor monitoring data are directly used to predict RUL. Zhou et al. [9] extracted a new HI from the

operating parameters of lithium-ion batteries for degradation modelling and RUL prediction. Yang et al. [10] proposed a dynamic HI smoothing approach to smoothen the current HI value against the previously predicted value. Lee et al. [11] defined an HI for the filter and then used a recurrent neural network (RNN) algorithm to predict the HI value from the degradation point to the end-of-life to generate the RUL. Gou et al. [12] proposed a RNN-based HI (RNN-HI) for RUL prediction of bearing. These methods fused multiple sensor data to construct a composite HI. The single channel network model predicts RUL using the constructed HI to characterize the degradation process of mechanical equipment.

Based on the abovementioned theory, an attempt can be made to construct a multidimensional HI using multi-dimensional features to characterize the degradation process of equipment. Ansari et al. [13] constructed a multi-channel artificial neural network (ANN) for extracting multiple features of batteries, and their proposed model showed strong versatility. Peng et al. [14] proposed a prediction model based on the idea of classification and parallel processing. Zhao et al. [15] used a two-channel hybrid model to predict the RUL of aeroengines, which demonstrated better performance than the conventional prediction models. Li et al. [16] experimentally concluded that a dual-path directed acyclic graph predicts better than a single-path convolutional neural network (CNN) or a long short-term memory. Based on the summary of the findings and approaches used in literature, we consider multidimensional features as multidimensional HI. The nonlinear mapping capability of a multichannel network structure was used to establish relationships between multidimensional features and RUL.

However, due to the high dimensionality of currently used monitoring parameters and prediction models that do not adequately extract valid information from monitoring data, predicting the health and safety of an aeroengine is difficult. We propose a new RUL prediction method that combines CNN-TCN and a multi-head attention mechanism based on research related to network structures with multiple channels. The method uses CNNs to mine temporal features, a TCN to improve the computing efficiency of the network while ensuring the integrity of long time sequences, and a self-attention mechanism to focus on useful information. Different sensor monitoring parameters were modelled separately to enable parallel processing of different sensor data, maximizing data integrity while improving the computational efficiency of the network. The proposed method was validated using NASA's C-MAPSS data, and the experimental results show a significant improvement in the RUL prediction of aeroengines.

2. Theoretical Basis

2.1. Convolutional Neural Network

CNN is a deep learning method with a strong generalization ability, and it has achieved favorable results in processing multi-array signals such as image, time series, and audio signals [16–19]. In this study, CNN was selected to uncover channel and spatial features that can effectively characterize the degradation process of aeroengines.

As shown in Figure 1, CNN is usually set alternately at the convolution and pooling layers. We used two CNNs to process the multisource sensor signals of an aeroengine in this paper. Moreover, each layer of the CNN has several convolution kernels of a consistent size that traverse the input multidimensional features in chronological order to create a high-dimensional feature space. Then, different feature spaces are combined to produce inputs to the next network [20].

The convolution layer is the core of the CNN and mainly comprises convolution kernels, which mainly extract features. The convolution layer realizes local sensing and weight sharing, reducing the complexity of model and computation cost. If $x_{n,l}$ represents the nth feature map of layer l, output $z_{n,l}$ of layer l can be calculated as follows:

$$z_{n,l} = k_{n,l} * x_{l-1} + b_{n,l} = \sum_{c=1}^{C} k_{c,n,l} * x_{c,l-1} + b_{n,l} \quad (1)$$

where $*$ is the convolution operator, $k_{n,l}$ is the nth weight of layer l, $b_{n,l}$ is the bias, and C is the number of input channels.

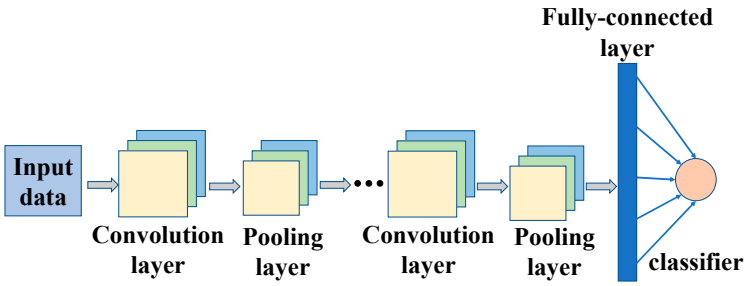


Figure 1. CNN structure.

The activation function ReLU is used to nonlinearly transform the output of the activation layer to improve the applicability of the network, and it is calculated as follows:

$$S_{n,l} = ReLU(z_{n,l}) = \max\{0, z_{n,l}\} \tag{2}$$

where $S_{n,l}$ represents the output of the activation function.

As a common layer after the convolution layer, the pooling layer reduces network parameters to improve the computational efficiency. In this study, we selected the max pooling layer, which can be calculated as

$$p_{n,l} = \max_{(n-1)V+1 \le t \le nV} \{S_{n,l}\} \tag{3}$$

where V is the parameter size of the pooled area and $p_{n,l}$ is the output of the pooling layer.

2.2. Temporal Convolutional Network

For the multi-dimensional sensor long-time-series signal of an aeroengine, the conventional CNNs are limited by the depth of the network. In addition, CNNs cannot effectively process the time series. RNNs can capture the latent temporal patterns but face difficulty in avoiding gradient disappearance or explosion. To address these problems, a temporal convolutional network (TCN) is used, which is a sequential prediction model characterized by layered stacks of dilated causal convolution (DCC) with residual connections (RC) [21,22]. The RC is a constituent unit of the TCN and is illustrated in Figure 2.

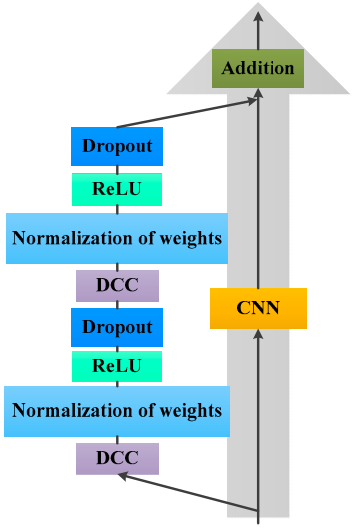


Figure 2. TCN residual block structure.

Causal convolution in TCN avoids information disclosure and enhances the memory of past information on the network. Causal convolution ensures that when processing time-series data, the output of time t is only convolved with the convolution of time t and earlier elements in the previous layer. However, the deeper the network depth, the more past information is memorized, and the increase of network depth will affect the efficiency of model training. To solve to solve this problem, we introduce the dilated convolution.

Dilated convolution can be sampled at input intervals during convolution. While ensuring that TCN has a wider field of view and receives more historical data, dilated convolution avoids the problems caused by extremely deep networks. The dilated convolution operation F on element s of the sequence is defined as:

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

where d is the dilation factor, k is the filter size, a filter f: $\{0, 1, \ldots, k-1\}$, $s - d \cdot i$ accounts for the direction of the past, X is the input and F is the output.

A residual block is a key structure of the TCN, which is defined as follows:

$$o = Activation(X + F(X)) \tag{5}$$

2.3. Multi-Head Attention

In the multi-dimensional long-time-series prediction of aeroengines, some features are independent of each other. Therefore, we utilized a multi-headed attention mechanism to separately process different sensor monitoring parameters.

The self-attention mechanism is similar to that in TCN, enabling parallel computation. It filters essential messages from the input features and assigns them different weights according to their importance [23]. Hence, the model focuses on the message with a larger weight so that we can quickly capture the degradation signal of the aeroengines. We used the scaled dot-product attention mechanism, a commonly used method, which first obtains the corresponding weights of the query and key matrixes through point multiplication, then the Softmax function is used for normalization. Finally, the Attention is obtained by weighted summation, as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

$$Q = A_f W^Q \tag{7}$$

$$K = A_f W^K \tag{8}$$

$$V = A_f W^V \tag{9}$$

where Q is a query matrix, K is a key matrix, V is a values matrix, and A_f is an input matrix. Further, W^Q, W^K, and W^V represent the weight matrixes of Q, K, and V, respectively, and d_k is the dimension of Q, K, and V.

The self-attention mechanism focuses on the details of the input message according to the target. The multi-head attention mechanism is based on the combination of several self-attention mechanisms. For multiple sensor signals of aircraft engines, the multi-head attention mechanism is utilized to achieve simultaneous attention to different parameters, and finally the obtained results are spliced to obtain the final attention, as follows:

$$Multihead(Q, K, V) = Concat(head_1, head_2, \cdots, head_n)W \tag{10}$$

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{11}$$

where W_i^Q, W_i^K, and W_i^V represent the weight matrices of Q, K, and V in the ith attention head respectively, W represents the weight matrix of the multi-head attention mechanism, and the output of the multi-attention mechanism is spliced by the merge layer.

3. Proposed Methodology

The existing RUL prediction studies based on aeroengines use both CNNs and RNNs as data-driven prediction methods. However, CNNs cannot effectively process timing signals, and RNNs cannot avoid problems related to long-term dependence. Based on the related research, we propose a multi-head attention model based on CNN-TCN to predict the RUL of aeroengines; it contains two CNN layers and a TCN layer. The two CNN layers feature a multi-source sensor signal, and then the extracted features are input into the TCN for processing. Subsequently, the multi-dimensional features of the aeroengines are processed separately using the multi-head attention mechanism, which ensures the integrity of the input data and focuses on the most weighted message. The RUL prediction process of aeroengines based on the CNN-TCN and multi-head attention mechanism is shown in Figure 3.

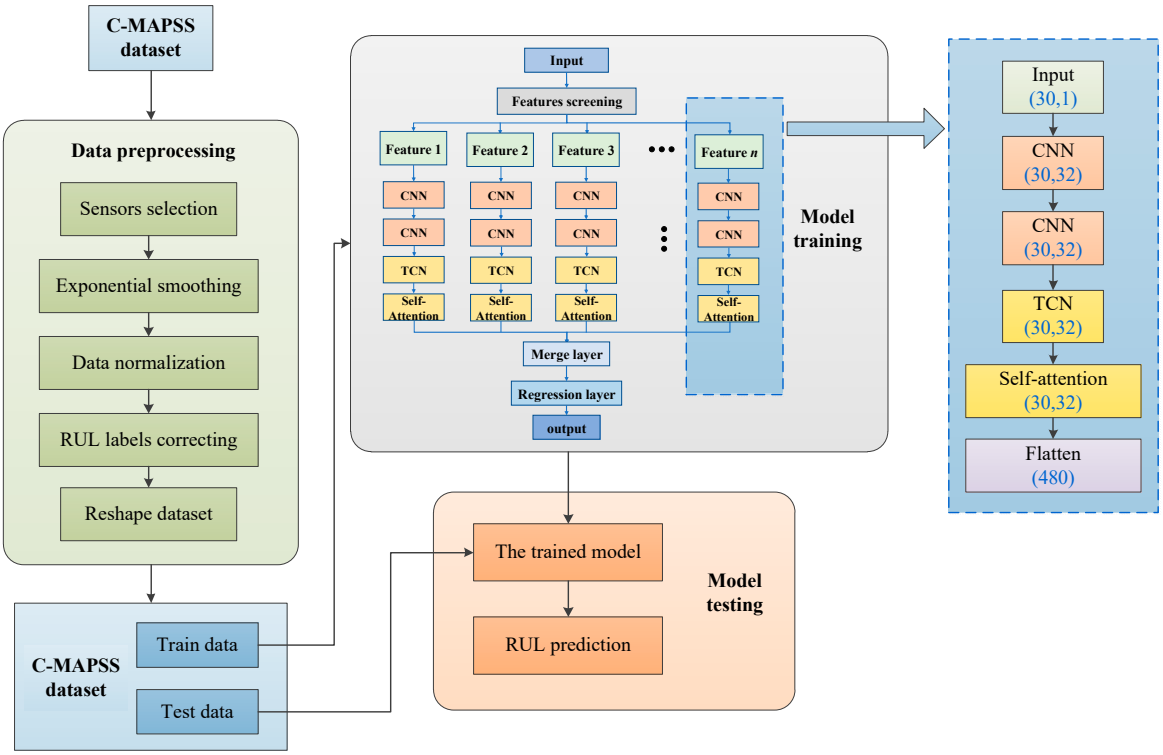


Figure 3. RUL prediction process of aeroengines.

1. **Data preprocessing**: According to the existing research experience [24–26], from among the 21 sensors, 14 sensors with large changes are selected. The 14 selected sensors are processed by exponential smoothing (ES) to remove environmental noise and retain the original degradation messages. Then, the sensors used as features are normalized to remove dimensional interference. A sliding window is introduced for secondary processing of preprocessed data. As the length of the sliding window increases, more data information is collected. However, this may cause the short-term state change to be ignored. Therefore, this article selects sliding windows with lengths of 30 and 40.
2. **Model construction**: The RUL tag on the divided training dataset is used as the input of the prediction model to train the model. We constructed a 14-channel CNN-TCN network for separate modeling of different features to enable parallel processing of

different data. Subsequently, we chose the concatenate function in the Merge layer to stitch the multidimensional data, and two dense layers were utilized to regress.

3. **RUL prediction**: The trained model is called to make predictions about the test set, and the prediction results are compared with the true values.

4. Experiment

4.1. Dataset Description

The experimental data are derived from the degradation data of NASA C-MAPSS turbofan engines, including FD001~FD004 subsets [27]. Each dataset contains three files: the training set, test set, and RUL true values. The proposed method was evaluated on the FD001 and FD003 in the C-MAPSS datasets. The detailed description of the datasets is presented in Tables 1 and 2.

Table 1. Composition of the C-MAPSS dataset.

Dataset	FD001	FD003
Training engines	100	100
Testing engines	100	100
Sensor measurements	12	12
Operation conditions	1	1
Fault modes	1	2

Table 2. C-MAPSS outputs to measure system response.

No.	Symbol	Description	Units
1	T2	Total temperature at fan inlet	(°)
2	T24	Total temperature at LPC outlet	(°)
3	T30	Total temperature at HPC outlet	(°)
4	T50	Total temperature at LPT outlet	(°)
5	P2	Pressure at fan inlet	Pa
6	P15	Total pressure in bypass-duct	Pa
7	P30	Total pressure at HPC outlet	Pa
8	Nf	Physical fan speed	r/min
9	Nc	Physical core speed	r/min
10	epr	Engine pressure ratio (P50/P2)	-
11	Ps30	Static pressure at HPC outlet	Pa
12	Phi	Ratio of fuel flow to Ps30	pps/psi
13	NRf	Corrected fan speed	r/min
14	NRc	Corrected core speed	r/min
15	BPR	Bypass Ratio	-
16	FarB	Burner fuel-air ratio	-
17	htBleed	Bleed Enthalpy	-
18	Nf_dmd	Demanded fan speed	r/min
19	PCNfR_dmd	Demanded corrected fan speed	r/min
20	W31	HPT coolant bleed	lbm/s
21	W32	LPT coolant bleed	lbm/s

4.2. Sensors Selection

Some sensor parameters of aeroengines that are not related to the degradation process during operation should be rejected. The accurate selection of the number of features that are highly correlated with the lifecycle can improve the efficiency of network training. For subsets FD001 and FD003, a few of the 21 sensor parameters remain essentially constant throughout the lifecycle. Hence, sensors 1, 5, 6, 10, 16, 18, and 19 were discarded [24–26].

4.3. Exponential Smoothing

Exponential smoothing is a time-series forecasting method that evolved from the moving average method. This method makes better use of the utility of recent observations

on the predicted values than those commonly used methods such as moving average, spline, etc. In addition, the weights on the observations are scalable. In the C-MAPSS dataset, the correlation between the current value and the surrounding values decreases with the number of cycle steps. Therefore, it is feasible to use the ES method to smooth the original sensor parameters of 100 aeroengines. This method is a special weighted moving average method, where the current value can be regarded as a weighted average of the current actual value and the previous moment value [28]. The calculation is as follows:

$$\begin{cases} S_t = \alpha \cdot y_t + (1-\alpha) \cdot S_{t-1}, t \geq 2 \\ S_1 = y_1 \qquad\qquad\quad , t = 1 \end{cases} \tag{12}$$

where S_t is the observed value at t, S_{t-1} is the observed value at $t-1$, y_t is the true value at t, and α represents the smoothing constant, which ranges from 0 to 1.

The value of α in ES determines the degree of smoothing. The higher the value of α, the greater the impact the recent data information has on the forecast. Conversely, the data tend to be flat. When the sensor monitoring data fluctuate but do not significantly change over time, α can be valued between 0.1 and 0.5. Therefore, 14 sensor detection parameters of 100 aeroengines were preprocessed with α of 0.1, 0.3, and 0.5. Two sensor monitoring data were randomly selected from FD001 and FD003, and the processed data are shown in Figure 4.

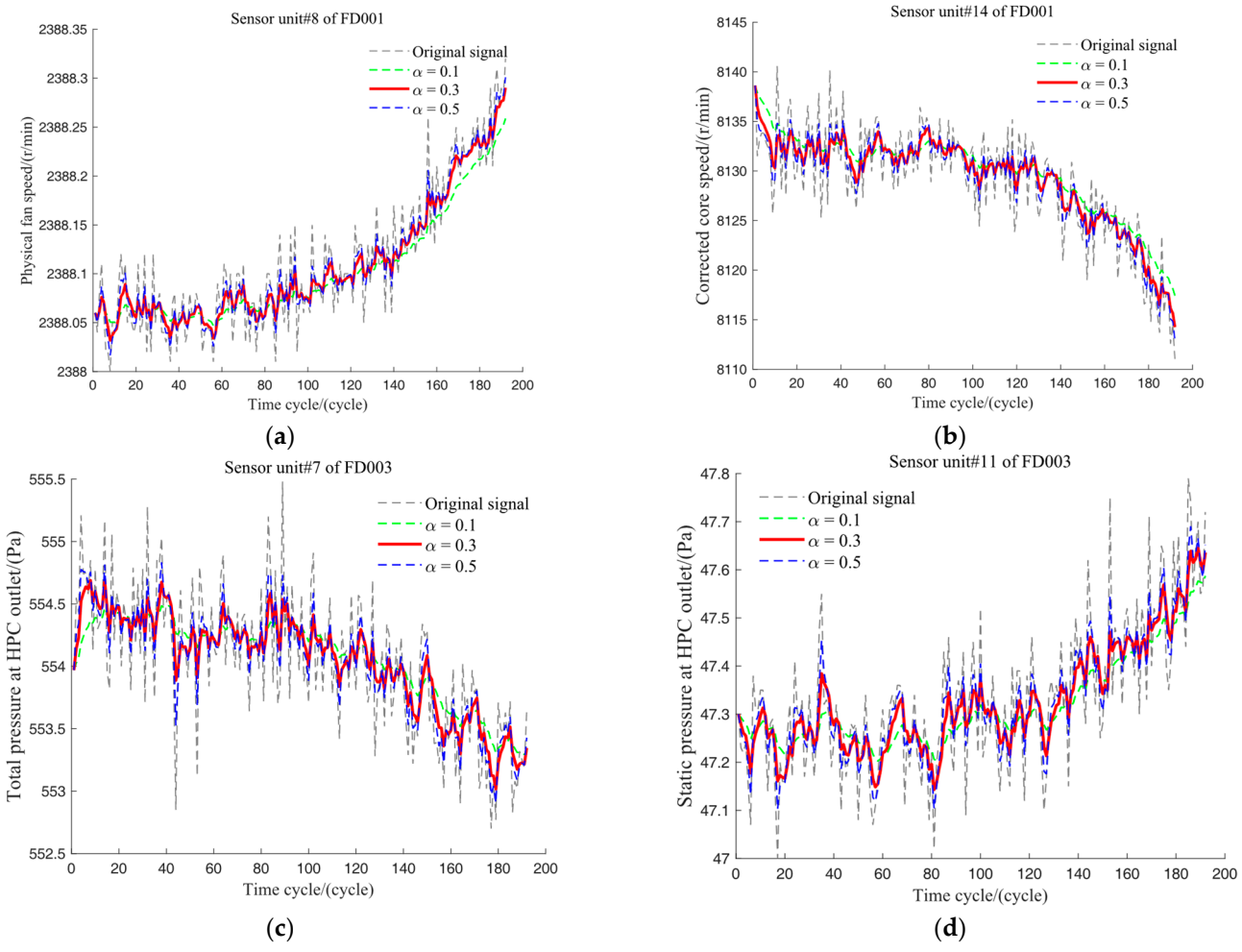


Figure 4. (**a**,**b**) Performance of ES on sensors #8 and #14 of FD001; (**c**,**d**) Performance of ES on sensors #7 and #11 of FD003.

As shown Figure 4, when α = 0.1, the curve fluctuation after ES is small; however, it is impossible to make a good trend fit for the second half of the cycle period. When α = 0.5, the curve after ES fits the cycle period well; however, the rejection of ambient noise is not complete. In contrast, when α = 0.3, not only does the degradation curve of the cycle period fit well but the interference of environmental noise is also avoided to a great extent. In summary, the 14 sensor detection parameters of 100 aeroengines in FD001 and FD003 were smoothed using α of 0.3.

4.4. Data Normalization

The efficient use of data is important for improving the training efficiency and RUL prediction accuracy. As the data collected come from different types of sensors, the data must be preprocessed. For the datasets FD001 and FD003, the MinMaxScaler is used to scale data for each sensor signal. Given n as the time cycle, the raw sensor data are denoted as $X = [x_1, x_2, x_3, \ldots\ x_n]$, with each sensor calculated as

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{13}$$

4.5. RUL Target Function

Owing to fatigue damage, friction damage, or fracture in the process of operation, the RUL of the components will inevitably decrease with time. In the early stages of aircraft engine operation, the wear on mechanical parts is negligible. Hence, the aircraft engine is assumed to be in a healthy state. With an increase in the working time, the wear of components cannot be ignored when the critical degradation value, R, is reached, and the aircraft engine enters the degradation state. According to several studies, the commonly used R values for multivariate sensor data of aeroengines are 120, 125, 130, and 135 [29,30]; in this study, we set the R value as 130. The RUL target function for aeroengines is defined as follows:

$$RUL = \begin{cases} 130, x < a - 130 \\ a - x, x \geq 130 \end{cases} \tag{14}$$

where x is the number of cycles at the measured point, and a is the maximum number of cycles.

4.6. Metrics

The score [31–33] and root mean square error (*RMSE*) are the two commonly used evaluation metrics for the C-MAPSS dataset; these are defined as follows:

$$MSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {d_i}^2} \tag{15}$$

$$score = \begin{cases} \sum_{i=1}^{n} (e^{-\frac{d_i}{13}} - 1), d_i < 0 \\ \sum_{i=1}^{n} (e^{\frac{d_i}{10}} - 1), d_i \geq 0 \end{cases} \tag{16}$$

where d_i is the difference between the predicted and true RUL values.

The RMSE is used to measure the deviation of the predicted value from the true value. The smaller the RMSE value, the closer the true value is to the predicted value. In actual conditions, the positive or negative difference between the predicted and true values has a significant impact on the subsequent maintenance and work guarantee. Advanced prediction allows for timely repair before failure; however, premature prediction leads to unnecessary waste. Furthermore, lagging prediction leads to more significant consequences, and not predicting the failure in time creates a safety hazard. As the RMSE does not reflect the true magnitude between the predicted and true values, the introduction of a score increases the penalty for lagged predictions, with lower scores indicating better predictions.

5. Results and Analysis

5.1. Ablation Study

5.1.1. Time Window

For long time sequences, the window length is an important parameter, which is directly related to the final accuracy of the deep learning model [34,35]. We introduced a time window to reconstruct data for 14-dimensional sensor data and RUL labels. For an extremely small time window, the correlation between features and time cannot be captured. In contrast, an extremely large time window contains more useful information but tends to cause the network to ignore short-term changes in features. The operational flow of the time window is shown in Figure 5.

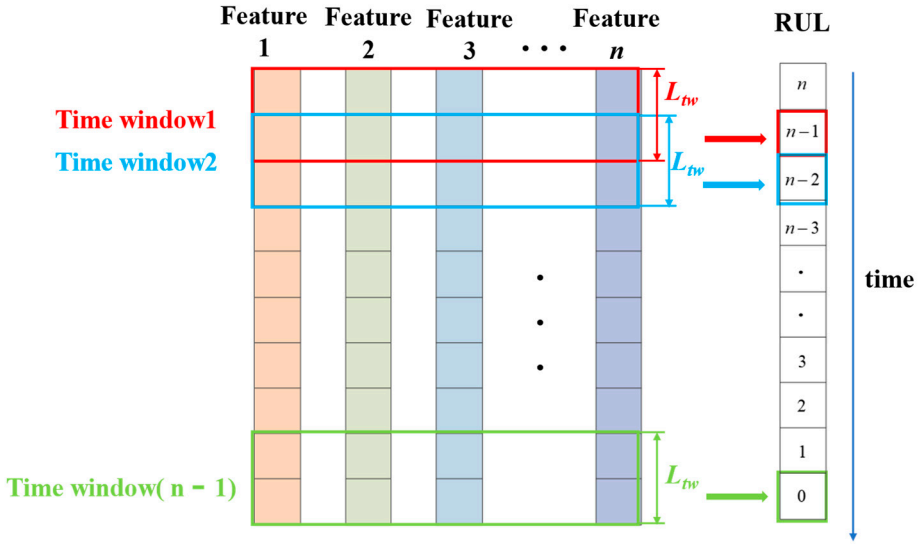


Figure 5. Time-window processing sequence.

To determine the optimal time-window length for FD001 and FD003, we experimented with different time-window lengths (time-window length is the number of cycles per interception) and their effects on the prediction results, as shown in Table 3. The results show that the choice of the time-window length plays a crucial role in the final accuracy of the deep learning model. The prediction effect improves with the increasing length of the time window at the beginning of the experiment, and the prediction performance mostly decreases gradually after the time-window length of $L_{tw} > 30$. Hence, for FD001 and FD003, we recommend $L_{tw} = 30$ and 40, respectively.

Table 3. Effect of different time-window lengths.

Time Window Length	FD001		FD003	
	RMSE	Score	RMSE	Score
$L_{tw} = 10$	17.12	226	16.45	635
$L_{tw} = 20$	14.73	115	15.22	522
$L_{tw} = 30$	11.07	62	11.25	126
$L_{tw} = 40$	13.78	117	10.39	95
$L_{tw} = 50$	13.90	129	11.81	220
$L_{tw} = 60$	14.54	91	13.59	273
$L_{tw} = 70$	14.42	103	13.00	191

5.1.2. Model Features

To explore the structure of the optimal model, this section describes the effect of different layers and hyperparameters on the prediction performance. Separate tests were performed on FD001 and FD003, and the results of the study are listed in Tables 4 and 5.

Table 4. Effect of different model layers.

Model	FD001		FD003	
	RMSE	Score	RMSE	Score
1-layer CNN + 1-layer TCN	12.04	91	12.50	187
1-layer CNN + 2-layer TCN	14.43	76	12.95	310
2-layer CNN + 2-layer TCN	14.91	119	13.47	274
2-layer CNN + 1-layer TCN	11.07	62	10.39	95
3-layer CNN + 1-layer TCN	14.23	88	12.49	158
4-layer CNN + 1-layer TCN	13.30	93	11.59	303
5-layer CNN + 1-layer TCN	15.34	109	13.42	156

Table 5. Effect of different model parameters.

Model	FD001		FD003	
	RMSE	Score	RMSE	Score
filters = 64, kernel size = 3	11.57	75	13.21	185
filters = 64, kernel size = 2	15.94	172	12.81	181
filters = 32, kernel size = 3	11.07	62	10.39	95
filters = 32, kernel size = 2	13.38	78	12.96	220

The comparison of the effects of different layers shows that the model with the combination of a 2-layer CNN and a 1-layer CNN has the best prediction effect. Compared with other structures, the optimal structure reduces the RMSE and Score of FD001 by 21.15% and 35.42% on average, and those of FD003 by 18.45% and 58.87% on average, respectively. Moreover, the effects of different network parameters were evaluated. The results show that the best prediction is achieved when filters = 32 and kernel size = 3. The final optimal model parameters obtained are shown in Table 6.

Table 6. Network parameters based on CNN-TCN and multi-head attention.

Type	Definition	Output
Input Layer	The input layer	(30, 1)
Conv1D	filters = 32, kernel size = 3	(30, 32)
Batch Norm	Batch Normalization	(30, 32)
Conv1D	filters = 32, kernel size = 3	(30, 32)
Batch Norm	Batch Normalization	(30, 32)
TCN	filters = 32, kernel size = 3	(30, 32)
Batch Norm	Batch Normalization	(30, 32)
SeqSelfAttention	Self-attentional layer	(30, 32)
MaxPooling1D	The pooling layer	(15, 32)
Flatten	Returns a 1D array	(480)
Concatenate	Merge the 14D channels	(6720)
Dense	Network structure is (6720, 50)	(50)
Dense	Network structure is (50, 1)	(1)

5.2. Comparison with the State-of-the-Art Models

The proposed network was trained with FD001 and FD003 in the C-MAPSS dataset and tested on all test aeroengines. To further analyze the proposed method and demonstrate its superiority, the predicted results of different sub-datasets are shown in Figure 6. In addition, Figure 7 shows the predicted and actual degradation processes for the two randomly selected engines from FD001 and FD003.

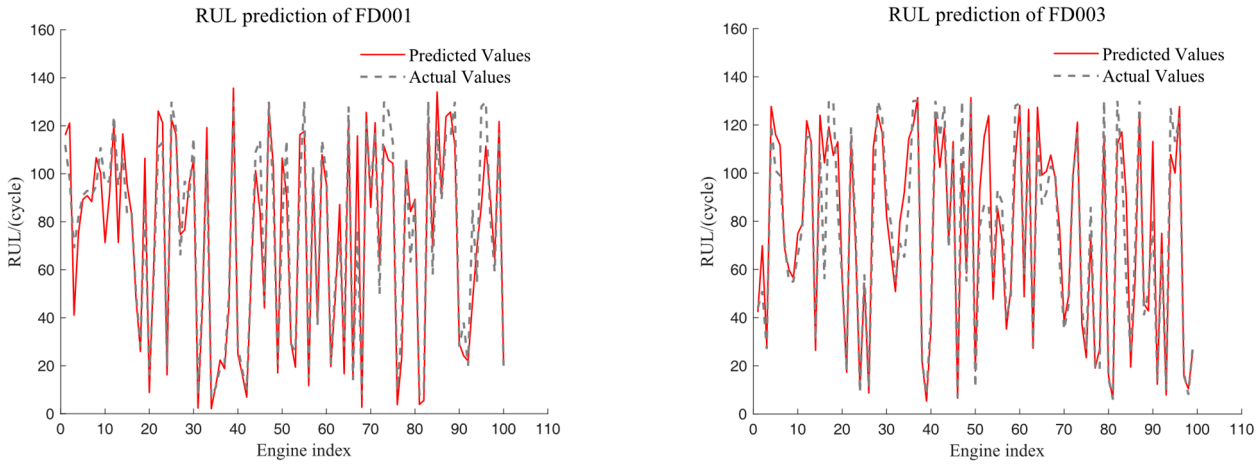


Figure 6. (**a**,**b**) Performance of the proposed model prediction on theFD001 and FD003 test sets.

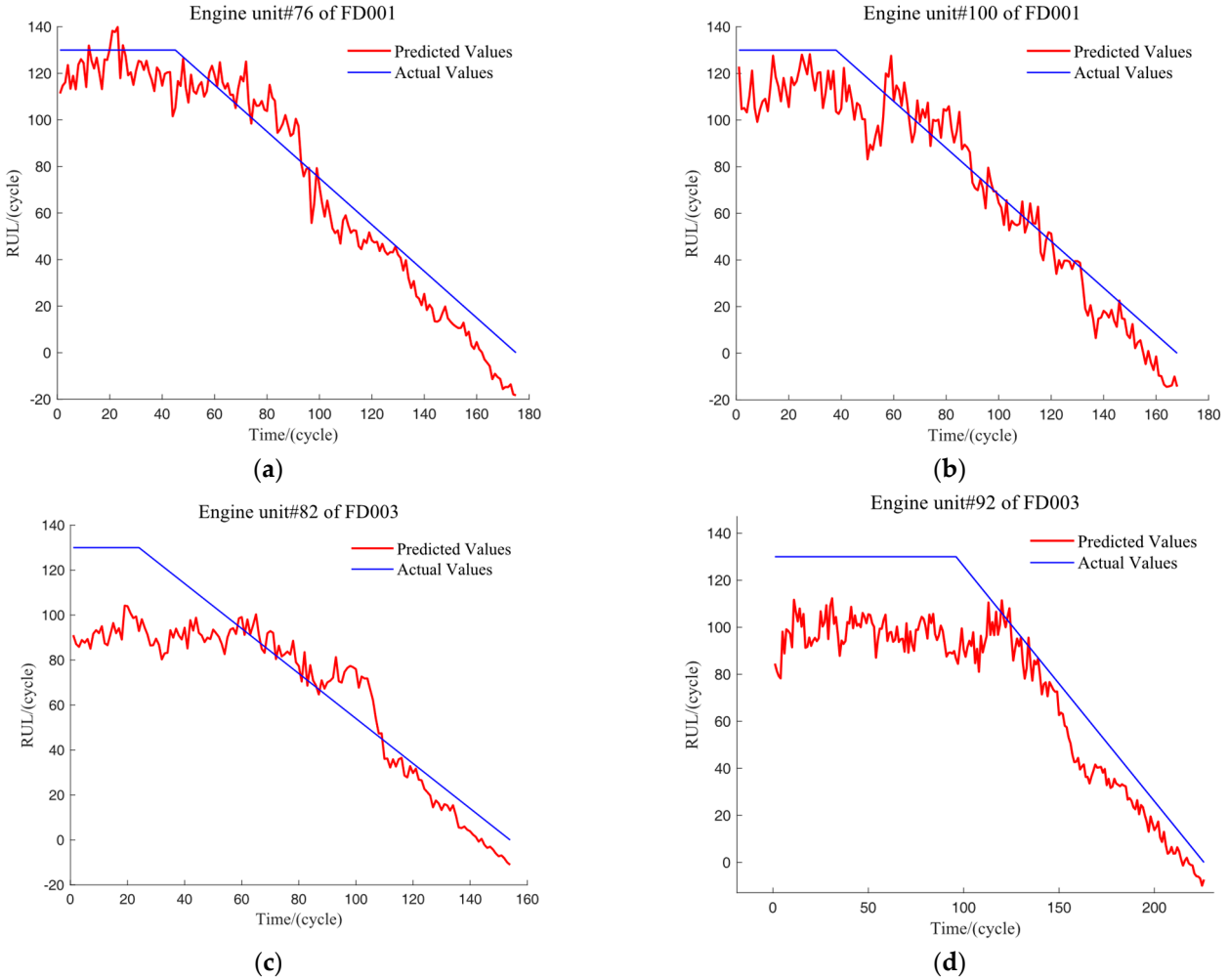


Figure 7. (**a**,**b**) RUL prediction results for engine units #76 and #100 of FD001; (**c**,**d**) RUL prediction results for engine units #82 and #92 of FD003.

The C-MAPSS dataset is a more widely used public dataset and has been used extensively in the research field. To verify the effectiveness of the proposed method, we compared the proposed method with the current mainstream machine learning methods

and composite methods proposed in previous studies. The comparison results are shown in Table 7. As observed, favorable prediction results were obtained on this dataset by different research methods, and the proposed prediction model showed a significant improvement in prediction performance compared with the state-of-the-art methods. Compared with the optimal method, the RMSE and Score of FD001 decreased by 10.87% and 42.57%, whereas those of FD003 decreased by 14.13% and 58.15%, respectively. Of the two performance evaluation metrics, the Score is significantly improved.

Table 7. Comparison of predicted results of different methods.

Model	FD001		FD003	
	RMSE	Score	RMSE	Score
DCNN [36]	12.61	273	12.64	284
RBPF [29]	15.94	383	16.17	375
BiLSTM+ED [37]	14.47	273	17.48	574
RBM+LSTM [31]	12.56	231	12.10	251
CNN+LSTM [15]	12.58	231	12.18	257
AGCNN [30]	12.42	226	13.39	227
Proposed model	11.07	62	10.39	95
Improved percent	10.87%	42.57%	14.13%	58.15%

6. Conclusions

This study introduced a novel model for predicting RUL based on CNN, TCN, and a multi-head attention mechanism; the model was proved suitable for long time series. The proposed method uses a two-layer CNN to process the input long time series to uncover features that can effectively characterize the degradation process of aeroengines. The introduction of TCN improves the gradient propagation capability and network computational efficiency while ensuring the extraction of time series feature information. The multi-head attention mechanism is used to increase the depth of a network in both vertical and horizontal directions. The multi-head structure maximizes the retention of useful information of the original sensor parameters. The proposed method was evaluated on FD001 and FD003 of the C-MPASS dataset, with the results showing improved accuracy. In this study, we tested the experimental data of aeroengines under a single operating condition; future studies must include the RUL prediction test for complex operating conditions.

Author Contributions: Conceptualization, L.N.; methodology, L.N. and S.X.; investigation, L.Z. and Y.Y.; software, S.X.; validation, Y.Y.; writing—original draft preparation, S.X.; writing—review and editing, L.N., S.X., L.Z., Y.Y., Z.D. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China Program (No. 51975191).

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the anonymous reviewers and the editor for their valuable and insightful suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ye, Z.; Zhang, Q.; Shao, S.; Zhao, Y.; Zhou, H.; Chen, C. Remaining Useful Life Prediction of Aeroengine Based on Ghost Approach. In Proceedings of the 2021 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD), Xi'an, China, 21–23 October 2021; pp. 1–6.
2. Kim, T.S.; Sohn, S.Y. Multitask learning for health condition identification and remaining useful life prediction: Deep convolutional neural network approach. *J. Intell. Manuf.* **2021**, *32*, 2169–2179. [CrossRef]

3. Liu, J.Q.; Lei, F.; Pan, C.L.; Hu, D.B.; Zuo, H.F. Prediction of remaining useful life of multi-stage aero-engine based on clustering and LSTM fusion. *Reliab. Eng. Syst. Saf.* **2021**, *214*, 107807. [CrossRef]
4. Wang, C.; Lu, N.; Cheng, Y.; Jiang, B. A Data-Driven Aero-Engine Degradation Prognostic Strategy. *IEEE Trans. Cybern.* **2021**, *51*, 1531–1541. [CrossRef] [PubMed]
5. Elsheikh, A.; Yacout, S.; Ouali, M.S. Bidirectional handshaking LSTM for remaining useful life prediction. *Neurocomputing* **2019**, *323*, 148–156. [CrossRef]
6. Wu, D.Z.; Jennings, C.; Terpenny, J.; Gao, R.X.; Kumara, S. A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. *J. Manuf. Sci. Eng. Trans. Asme* **2017**, *139*, 071018. [CrossRef]
7. Vachtsevanos, G.; Lewis, F.; Roemer, M.; Hess, A.; Wu, B. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*; Wiley: Hoboken, NJ, USA, 2006.
8. Kai, G.; Celaya, J.; Sankararaman, S.; Roychoudhury, I.; Saxena, A. *Prognostics: The Science of Making Predictions*; CreateSpace Independent Publishing Platform: Charleston, SC, USA, 2017.
9. Zhou, Y.P.; Huang, M.H.; Chen, Y.P.; Tao, Y. A novel health indicator for on-line lithium-ion batteries remaining useful life prediction. *J. Power Sources* **2016**, *321*, 1–10. [CrossRef]
10. Yang, F.; Habibullah, M.S.; Zhang, T.Y.; Xu, Z.; Lim, P.; Nadarajan, S. Health Index-Based Prognostics for Remaining Useful Life Predictions in Electrical Machines. *IEEE Trans. Ind. Electron.* **2016**, *63*, 2633–2644. [CrossRef]
11. Lee, S.; Lee, S.; Lee, K.; Lee, S.; Chung, J.; Kim, C.W.; Yoon, J. Data-driven health condition and RUL prognosis for liquid filtration systems. *J. Mech. Sci. Technol.* **2021**, *35*, 1597–1607. [CrossRef]
12. Guo, L.; Li, N.P.; Jia, F.; Lei, Y.G.; Lin, J. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* **2017**, *240*, 98–109. [CrossRef]
13. Ansari, S.; Ayob, A.; Lipu, M.S.H.; Hussain, A.; Saad, M.H.M. Multi-Channel Profile Based Artificial Neural Network Approach for Remaining Useful Life Prediction of Electric Vehicle Lithium-Ion Batteries. *Energies* **2021**, *14*, 7521. [CrossRef]
14. Peng, C.; Chen, Y.F.; Chen, Q.; Tang, Z.H.; Li, L.L.; Gui, W.H. A Remaining Useful Life Prognosis of Turbofan Engine Using Temporal and Spatial Feature Fusion. *Sensors* **2021**, *21*, 418. [CrossRef] [PubMed]
15. Zhao, C.Y.; Huang, X.Z.; Li, Y.X.; Iqbal, M.Y. A Double-Channel Hybrid Deep Neural Network Based on CNN and BiLSTM for Remaining Useful Life Prediction. *Sensors* **2020**, *20*, 7109. [CrossRef] [PubMed]
16. Li, J.L.; Li, X.Y.; He, D. A Directed Acyclic Graph Network Combined With CNN and LSTM for Remaining Useful Life Prediction. *IEEE Access* **2019**, *7*, 75464–75475. [CrossRef]
17. Nilwong, S.; Hossain, D.; Kaneko, S.; Capi, G. Deep Learning-Based Landmark Detection for Mobile Robot Outdoor Localization. *Machines* **2019**, *7*, 25. [CrossRef]
18. Pham, M.T.; Kim, J.M.; Kim, C.H. 2D CNN-Based Multi-Output Diagnosis for Compound Bearing Faults under Variable Rotational Speeds. *Machines* **2021**, *9*, 199. [CrossRef]
19. Yang, B.Y.; Liu, R.N.; Zio, E. Remaining Useful Life Prediction Based on a Double-Convolutional Neural Network Architecture. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9521–9530. [CrossRef]
20. Wang, B.; Lei, Y.; Li, N.; Yan, T. Deep separable convolutional network for remaining useful life prediction of machinery. *Mech. Syst. Signal Process.* **2019**, *134*, 106330. [CrossRef]
21. He, K.; Su, Z.; Tian, X.; Yu, H.; Luo, M. RUL Prediction of Wind Turbine Gearbox Bearings Based on Self-Calibration Temporal Convolutional Network. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3501912. [CrossRef]
22. Pan, M.; Hu, P.; Gao, R.; Liang, K. Multistep prediction of remaining useful life of proton exchange membrane fuel cell based on temporal convolutional network. *Int. J. Green Energy* **2022**. [CrossRef]
23. Zhang, Z.; Song, W.; Li, Q. Dual-Aspect Self-Attention Based on Transformer for Remaining Useful Life Prediction. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 2505711. [CrossRef]
24. Caceres, J.; Gonzalez, D.; Zhou, T.; Droguett, E.L. A probabilistic Bayesian recurrent neural network for remaining useful life prognostics considering epistemic and aleatory uncertainties. *Struct. Control. Health Monit.* **2021**, *28*, e2811. [CrossRef]
25. Shi, Z.; Chehade, A. A dual-LSTM framework combining change point detection and remaining useful life prediction. *Reliab. Eng. Syst. Saf.* **2021**, *205*, 107257. [CrossRef]
26. Xiong, M.L.; Wang, H.W.; Fu, Q.; Xu, Y. Digital twin-driven aero-engine intelligent predictive maintenance. *Int. J. Adv. Manuf. Technol.* **2021**, *114*, 3751–3761. [CrossRef]
27. Frederick, D.K.; Decastro, J.A.; Litt, J.S. *User's Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS)*; NASA: Washington, DC, USA, 2007.
28. De Marco, L.M.; Trierweiler, J.O.; Farenzena, M. Determination of Remaining Useful Life in Cyclic Processes. *Ind. Eng. Chem. Res.* **2019**, *58*, 22048–22063. [CrossRef]
29. Cai, H.S.; Feng, J.S.; Li, W.Z.; Hsu, Y.M.; Lee, J. Similarity-based Particle Filter for Remaining Useful Life prediction with enhanced performance. *Appl. Soft Comput.* **2020**, *94*, 106474. [CrossRef]
30. 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.* **2021**, *17*, 1197–1207. [CrossRef]
31. Ellefsen, A.L.; Bjørlykhaug, E.; Æsøy, V.; Ushakov, S.; Zhang, H.X. Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture. *Reliab. Eng. Syst. Saf.* **2019**, *183*, 240–251. [CrossRef]

32. Ding, H.; Yang, L.; Cheng, Z.; Yang, Z. A remaining useful life prediction method for bearing based on deep neural networks. *Measurement* **2021**, *172*, 108878. [CrossRef]
33. Xu, Q.; Chen, Z.; Wu, K.; Wang, C.; Wu, M.; Li, X. KDnet-RUL: A Knowledge Distillation Framework to Compress Deep Neural Networks for Machine Remaining Useful Life Prediction. *IEEE Trans. Ind. Electron.* **2022**, *69*, 2022–2032. [CrossRef]
34. Huang, C.G.; Huang, H.Z.; Li, Y.F. A Bidirectional LSTM Prognostics Method Under Multiple Operational Conditions. *IEEE Trans. Ind. Electron.* **2019**, *66*, 8792–8802. [CrossRef]
35. Zheng, S.; Ristovski, K.; Farahat, A.; Gupta, C. Long Short-Term Memory Network for Remaining Useful Life Estimation. In Proceedings of the IEEE International Conference on Prognostics and Health Management (ICPHM), Dallas, TX, USA, 19–21 June 2017; pp. 88–95.
36. 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]
37. Hou, G.S.; Xu, S.; Zhou, N.; Yang, L.; Fu, Q.H. Remaining Useful Life Estimation Using Deep Convolutional Generative Adversarial Networks Based on an Autoencoder Scheme. *Comput. Intell. Neurosci.* **2020**, *2020*, 9601389. [CrossRef] [PubMed]