Multi-resolution LSTM-based Prediction Model for Remaining Useful Life of Aero-engine

Tiantian Xu, Guangjie Han, *Fellow, IEEE,* Hongbo Zhu, *Member, IEEE,* Tarik Taleb, *Senior Member, IEEE* Jinlin Peng

***Abstract*—Aircraft is an important means of travel and the most convenient and fast vehicle in long-distance transportation. The aircraft engine is one of the most critical parts of an aircraft, and its reliability and safety are extremely important. In this paper, we consider that the operating conditions of aero-engines are complex and changeable, and a multi-resolution long short-term memory (MR-LSTM) model is proposed. The model can effectively predict the remaining useful life (RUL) of an aero-engine, which is a priority issue within the Prognostics and Health Management (PHM) framework - and thus it can support maintenance decisions. Sequences with multiple temporal resolutions are generated by a reconstruction of the decomposed wavelets. A two-layer LSTM model is then designed: 1) the first layer LSTM is used to learn attention at different time resolutions as well as to generate an integrated historical representation; 2) the second layer LSTM is used to learn the long and short-term time dependencies in the integrated historical representation. Experimental evaluations using the C-MAPSS datasets (FD002 and FD004) and the N-CMAPSS dataset showed that compared to other state-of-the-art RUL prediction methods, the FD002 sub-dataset showed a 12.1% reduction in RMSE and a 3.8% reduction in Score; the FD004 sub-dataset showed a 21.8% reduction in RMSE and a decreased by 62.1%; the RMSE of the N-CMAPSS dataset decreased by at most 25.8% and the Score decreased by at most 35.1%.**

***Index Terms*—aero-engines, remaining useful life, discrete wavelet transform, long and short-term memory networks, attention mechanism.**

I. Introduction

As an essential means of transportation for modern civilization, airplanes have profoundly changed people's lives. The aircraft engine is the most critical component of the aircraft, and its reliability and safety are extremely important. However, the structure of an aero-engine is very complex, with many parts and complex environmental conditions. Many parts work under high temperatures, high pressure, high-speed rotation, strong vibration, and other harsh environments under high load and thermal shock. Therefore, it is easy to fail, and has many failure modes, multi-mode compound failure, and other significant characteristics [1]. With the rapid development of the aviation industry, the aero-engine has developed from the initial piston engine to today's turboshaft engine and turbofan engine, etc. While the performance of the aero-engine has been greatly improved, its structure has also become more and more complex, and the failure of some important components has become more and more frequent. As a result, more and more attention has been paid to the aero-engine safety assessment and maintenance decisions [2].

To improve the safety and economy of aircraft engines, developed countries, such as the USA, have focused on engine health management technology. After making full use of the latest research results in the fields of information technology and artificial intelligence, a health management technology, namely Prognostics and Health Management (PHM) [3], has been proposed. PHM consists of two main components, namely fault prediction and health management. Failure prediction refers to determining the time of failure based on the current state of the equipment and estimating the remaining useful life (RUL) of the equipment before failure occurs [4]. Health management is the rational scheduling of aero-engine maintenance based on the information obtained from the failure prediction, such as engine life, to ensure safe and healthy aircraft flight while minimizing the economic costs associated with aero-engine maintenance. RUL prediction is an important part of the prediction module, which is usually some kind of algorithm or model, combined with the current condition of the engine and the operating environment, to make an accurate and rapid prediction of the aero-engine failure and RUL, and then feed the predicted information into the ground maintenance support system, to assist the ground staff to grasp the health status of the equipment dynamically in real time.

The current state of research on RUL forecasting is divided into physical model-based approaches [5] and data-driven approaches [4], [6]–[12]. Physical model-based approaches require accurate modeling of complex system degradation but require prior knowledge about the physical system. The data-driven approach is based on historical condition monitoring data collected by sensors and uses machine learning techniques to learn data features to predict RUL. Modern aero engines have extensive sensing capabilities, and a large amount of sensor condition data can be monitored in real-time and over

Tiantian Xu is with the School of Software, Dalian University of Technology, Dalian, 116024, China (E-mails: tiantianxu0609@gmail.com).

Guangjie Han is with the Department of Internet of Things Engineering, Hohai University, Changzhou, 213022, China (e-mail: hanguangjie@gmail.com) (*Corresponding author: Guangjie Han.*)

Hongbo Zhu is with the School of Information Science and Engineering, Shenyang Ligong University, Shenyang, 110159, China (E-mails: hombochu@sina.com).

Tarik Taleb is with the Center of Wireless Communications, the University of Oulu, 90570, Finland (E-mails: Tarik.Taleb@oulu.fi).

Jinlin Peng is with the Artificial Intelligence Research Center, National Innovation Institute of Defense Technology, Beijing, P.R. China (E-mails: peng_jinlin@126.com).

time. As a result, these data-driven approaches have become increasingly mainstream in recent years.

The development of the data-driven RUL prediction methods has been rapid, with various Neural Network (NN)-based RUL prediction algorithms being mainstream [4], [6]–[12]. Although the data-driven RUL prediction methods do not require a deep understanding of engineering principles and models, NN-based RUL prediction methods still have some problems. The estimation error is greatly affected by different flight conditions, different operating conditions, and failure modes of the engine. However, the existing studies have not considered the effect of operating conditions.

The engine operating conditions are complex and changeable, and the measured values of performance parameters often change with the engine operating conditions, so it is impossible to judge the engine performance state directly with the measured values of performance parameters, and then estimate the RUL of the engine. In other words, the RUL of the engine is affected by both engine performance degradation and engine operating conditions, and the effect of performance degradation is often masked by the effect of operating conditions. How to effectively incorporate the effects of operating conditions into the model is the key to improving the accuracy of performance parameter prediction. The time window for each condition is different as each condition occurs for a different length of time. One working condition corresponds to a period of performance degradation; two adjacent working conditions will similarly produce a period of performance degradation. In addition, the degradation of one operating condition may affect the performance degradation curve of the next operating condition. Therefore, in the feature extraction stage, multi-time window reconstruction data needs to be considered, which is equivalent to a sliding window acting on sequences of different temporal resolutions.

To build a model that makes full use of the multi-time resolution information, a multi-resolution long short-term memory (MR-LSTM) model is proposed. The decomposition and reconstruction of the discrete wavelet transform (DWT) method enable not only the generation of sequences with multiple temporal resolutions but also the removal of noise from the signal. A two-layer LSTM model is designed after obtaining the multi-temporal resolution sequences using the DWT: the first layer LSTM is used to learn attention at different temporal resolutions as well as generate an integrated representation; the second layer LSTM is used to learn the long-term and short-term time dependence in the integrated representation. We use the C-MAPSS dataset and the N-CMAPSS dataset to evaluate the proposed RUL prediction method and compare it with several state-of-the-art RUL prediction methods to verify its effectiveness. The following is a summary of the major contributions of this paper.

- The temporal characteristics of aero-engine sensor data in multiple dimensions are analyzed, and it is demonstrated that the sensor signal data exhibit various patterns at different time resolutions. In addition, the DWT can retain well the peak and abrupt parts of the useful signal required in the original signal, with some noise immunity.
- We created MR-LSTM, an attentive LSTM model with two layers to reweight temporal resolution features and infer temporal dependencies, respectively. The proposed model is trained using multiple temporal resolution series and learns combined features from different resolution layers.
- The proposed MR-LSTM model was evaluated using C-MAPSS dataset and N-CAMPSS dataset, and compared with the latest methods. The experimental results demonstrate that the proposed model can provide more reasonable prediction results for engineering practice.

The rest of this article is organized as follows. Sec. II introduces the related work. Sec. III introduces the MR-LSTM-based RUL prediction model. Sec. IV presents the procedure of the proposed model and evaluates the performance of the proposed model. In Sec. V, the article is concluded.

II. Related Studies

We provide some deep-learning strategies for machine RUL prediction in this section. In the realm of RUL prediction, deep learning has emerged as a promising technique. It is effective in learning features from raw data without the need for professional diagnostic knowledge. In a multi-layer framework, deep learning can also accurately establish correlations between signal data and RUL. As a result, research into deep learning-based RUL prediction methods is becoming increasingly active, with an emphasis on improving prediction accuracies, such as the merging of deep learning models with current models, structure, and parameter optimization [13], [14]. While much of deep learning's success has been centered on classification issues, it has also shown to be an effective method for tackling prediction challenges. Jayasinghe *et al.* [15] combines temporal convolutional and LSTM layers with data augmentation to predict the RUL of mechanical devices. Liu et al. [8] proposed an engine RUL prediction method based on a feature attention mechanism that dynamically assigns weights to different input features. Then the features with different weights are input to the BGRU to learn the long and short-term dependencies. Finally, they are fed into a multilayer FCN to predict RUL.

In RUL prediction, a data-driven time series analysis approach uses artificial intelligence to extract dynamic patterns hidden in the data to predict future data points [16], [17]. Chen *et al.* [11] proposed a feature fusion framework to fuse manual features (mean and trend coefficients from linear regression) with features automatically learned by LSTM networks to improve RUL prediction performance. Huang*et al.* [12] proposed a prediction model (BLSTM). It integrates engine performance data with operating condition data in an integrated framework as the primary and secondary inputs to the network. However, it does not take into account that the measured engine performance parameters are affected by the engine operating conditions and directly feed the state parameters into the deep learning network.

Most prediction methods only consider the use of raw sensory data to predict the RUL of an engineering system or input the condition parameters directly into a deep learning

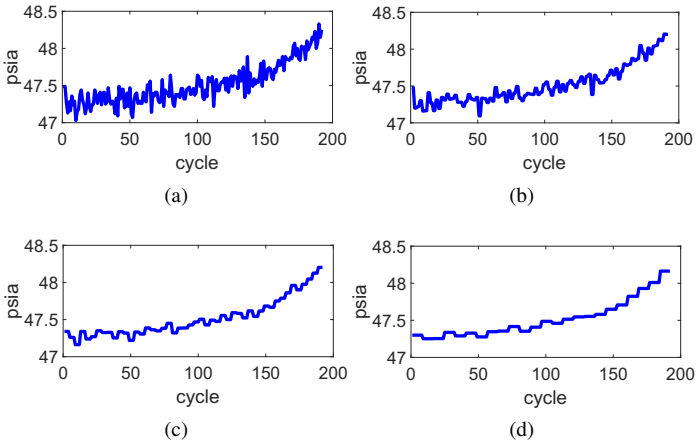


Fig. 1: (a) is the static pressure HPC outlet on the original time resolution; (b) (c) (d) is the static pressure HPC outlet after using the three-level DWT.

network to extract the features hidden in the condition signal. As the operating conditions of an aero-engine are complex and variable, and the RUL is influenced by the operating conditions, the effects of the operating conditions need to be taken into account when predicting the RUL. How to effectively incorporate the effects of operating conditions into the model is the key to improving the accuracy of performance parameter prediction. As each working condition occurs for a different length of time, signal data need to be reconstructed with multiple time windows, which is equivalent to a sliding window acting on sequences with different time resolutions. Therefore, we propose an MR-LSTM model that learns multi-temporal resolution representations from the rich information available under multiple operating conditions. In the MR-LSTM model, information at multi-resolutions is considered simultaneously, and each different time resolution is learned internally using an attention mechanism, subsequently generating an integrated representation to model both long-term and short-term time dependencies, achieving highly accurate predictions of RUL with some generalization capability.

III. Methodology

A. Multi-temporal resolution analysis

The DWT is widely used in signal processing and is well known for its properties such as preserving the peak and abrupt parts of the useful signal and having some noise immunity. This meets the requirements of multi-time window reconstruction of aero-engine signals. Therefore we use the DWT for multi-time window reconstruction of the signal. We show the result of DWT on the bypass ratio in Fig.1, and the bypass ratio sequence shows different patterns at different temporal resolutions. The large resolution shows the overall trend of the series data, while the small resolution provides more detailed information about the data.

B. Discrete wavelet transform

The low-frequency components of many signals are significant and frequently contain the signal's features, whilst the high-frequency components provide details or differences in the signal. We decompose and reconstruct sensor signal data based on the DWT, using binary scaling and displacement. To obtain a multi-temporal resolution series, DWT decomposes the signal into a set of fundamental functions. These basic functions are called wavelet mother functions, and the wavelet mother function $\psi_(t)$ is scaled and translated with a scaling factor of a and a translation factor of b. The wavelet mother functions $\psi_(t)$.

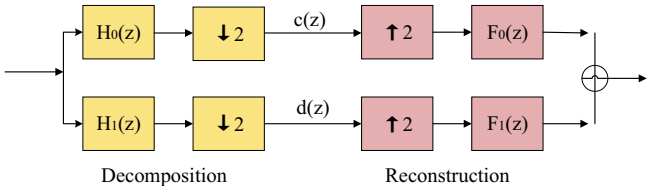


Fig. 2: Schematic diagram of wavelet transform decomposition and reconstruction.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \tag{1}$$

The DWT, by its very nature, does not require knowledge of the exact structure of the scale and wavelet functions, and the decomposition and reconstruction of the signal can be achieved based on the coefficients. Thus, the original signal is passed through two complementary filters to produce two signals, the approximate details of which are obtained by calculating the DWT coefficients [18].

The sensor signal is first discrete sampled to obtain $x(z)$, then the wavelet decomposition and reconstruction of the signal can be achieved in the form of subband filtering, the structure of the decomposition and reconstruction is shown in Fig. 2.

In the figure, $F_0(z)$ and $F_1(z)$ are the corresponding filter coefficients of the low-pass filter and high-pass filter respectively, and $H_0(z)$ and $H_1(z)$ are the corresponding filter coefficients of the mirror filter of the low-pass filter and high-pass filter respectively, satisfying $H_0(z) = F_0(-z)$ and $H_1(z) = F_1(-z)$. The signal is decomposed as follows: on the one hand, the signal $x(z)$ is downsampled after the low-pass filter, resulting in an average signal $c(z)$ with halved scale and resolution, i.e. the low-frequency component. On the other hand, the signal $x(z)$ is downsampled by a high-pass filter to obtain a detailed signal $d(z)$ with half the scale and resolution, i.e. the high-frequency component. The signal decomposition is given by

$$c_{j+1}(z) = \sum_{m\in z} c_j(m)F_0(m-2X), \tag{2}$$

$$d_{j+1}(z) = \sum_{m\in z} c_j(m)F_1(m-2X), \tag{3}$$

The reconstruction process is as follows: the signal is stretched by inserting zero values between two samples on average, i.e. upsampling, and then passing through a low-pass filter to obtain a large time window, low-resolution approximation, i.e. low-pass output; the detailed signal is

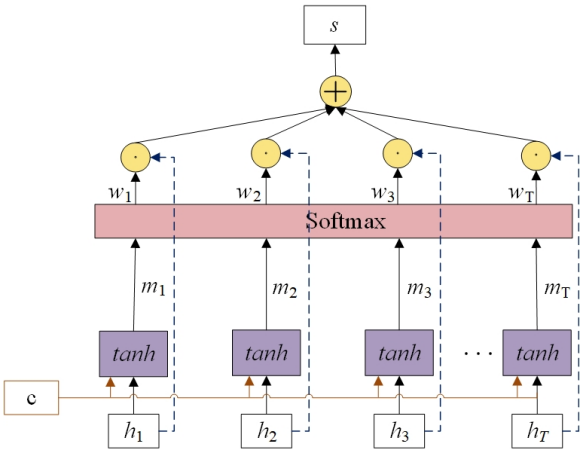


Fig. 3: The process of attention mechanism.

upsampled and then passed through a high-pass filter to obtain a high-pass output, and the two are added together to obtain the reconstructed signal $\hat{x}(z)$. The equation for signal reconstruction is

$$c_j(z) = \sum_{m \in z} c_{j+1}(m)F_0(z-2m) + d_{j+1}(m)F_1(z-2m). \quad (4)$$

The correlation between the signal and the filter is the essence of signal decomposition based on the wavelet transform, whereas the reconstruction is the convolution of the decomposed signal with the mirror filter bank. DWT extracts multi-temporal resolution features in this way to obtain more information, which is further used to construct regression models.

C. Attention mechanism

Attentional mechanisms, inspired by the brain's ability to resolve information overload, have been widely used in deep learning, especially for natural language processing and image processing [19]. The basic idea beneath the attention mechanism is to generate an attention weight matrix based on the input and weight of each input feature. Due to the differences in the importance contained in feature information at different temporal resolutions, we introduce the soft attention mechanism into the MR-LSTM model, weighting the features at different temporal resolutions one by one for scoring, and using the normalized average scoring as the weight parameter of the features, effectively achieving the combination of coarse and fine-grained features, the process of which is shown in Fig. 3.

As shown in Fig. 3, the attention model takes as input T states $h_1, h_2, ..., h_T$ and context c, and then computes $m_1, m_2, ..., m_T$ with a tanh layer. The value of m_i is determined by the relevance of the context vectors c and h_i:

$$m_i = \tanh(W_{cm}c + (W_{hm}h_i), \quad (5)$$

W_{cm} and W_{hm} are the weight matrices of the connection context c and the hidden layer state h_i, respectively. Using the attention mechanism, it is not necessary to encode the full input as a fixed-length vector and, using the softmax function, to project m in the direction of learning: $w_i \propto exp(< w_m, m_i >), \sum_i w_i h_i = 1$. The final output feature s of the attention mechanism is a weighting of all input states:

$$s = \sum_i w_i h_i, \quad (6)$$

where the value of the weight w_i is determined by the importance of the association between c and h_i.

D. The proposed MR-LSTM model

We propose a two-layer LSTM model based on an attention mechanism to make full use of multi-temporal resolution information. After decomposing the RPLR into multi-resolution sub-waves using DWT in the literature [20], several independent echo state networks are used to predict the sub-waves separately. The predictions of the different time-scale subwaveforms are then combined to achieve overall prediction for the aero-engines RUL. However, the importance of subwaves with different resolutions is different. Our proposed MR-LSTM model just solves this problem. After decomposing the original data into multi-resolution data using DWT, we integrate the multi-resolution data into a two-layer LSTM based on an attention mechanism. Instead of simply stacking LSTMs layer by layer, the MR-LSTM considers information at different scales and introduces an attention mechanism to understand the importance of each different resolution.

Fig.4 shows our proposed MR-LSTM-based RUL prediction model. First, the raw engine sensor data are fed to DWT for multi-temporal resolution conversion. The features of the multi-resolution sequence are then learned in the first layer of the LSTM, and the learned hidden layer features learn the importance of different resolution features under the attention mechanism. Then, the learned sequence features are combined with the weights generated by the attention model to obtain a composite representation of all temporal resolutions at time t. Afterward, the integrated representation of the previous time is fed into the second layer of LSTM to learn its long-and short-term dependencies. Finally, the fully connected layer is used for RUL prediction.

The details of the MR-LSTM are as follows: First, the input data $X = (x_1, x_2, ..., x_t, ...), x_t \in R^d, t \in (1, T)$ collected by the sensor is decomposed by DWT multi-temporal resolution as a sequence $R_1 = (x_1^1, x_2^1, ..., x_T^1), R_2 = (x_1^2, x_2^2, ..., x_T^2), ..., R_k = (x_1^k, x_2^k, ..., x_t^k), ..., k \in (1, K)$, where K denotes the scale factor, T is the time step, and x_T^k is the kth level input at time t. Then, we take $(x_1^k, x_2^k, ..., x_t^k, ...)$ as the input of the first level LSTM and get the hidden state as $(h_1^k, h_2^k, ..., h_t^k, ...)$. h_t^k denotes the kth level hidden layer feature at time t. As the following equation:

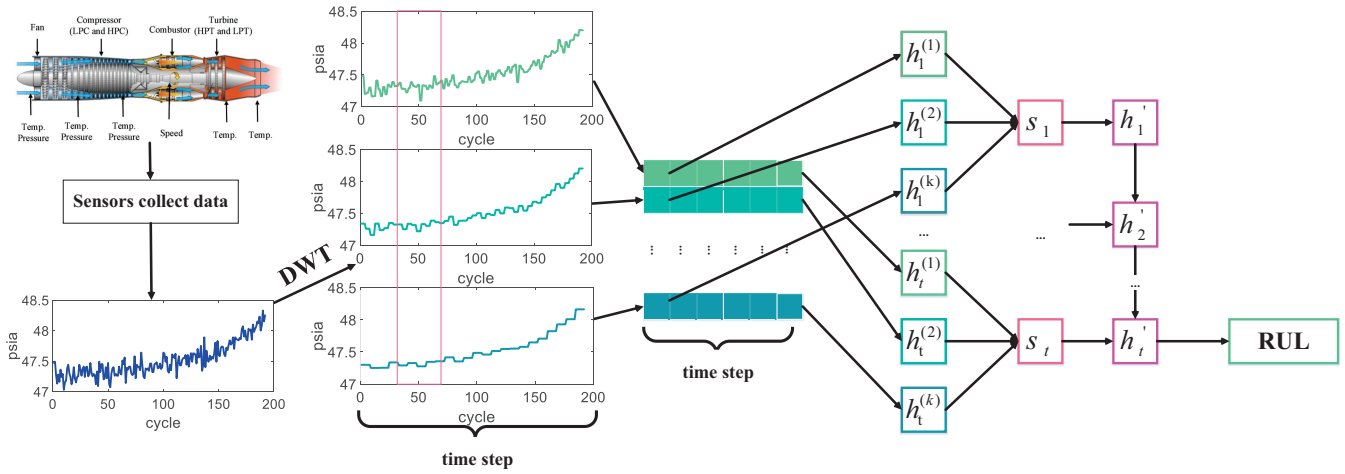


Fig. 4: Schematic representing the MR-LSTM pipeline: The raw sensor data is input to the LSTM for feature learning ($h_t^{(k)}$) after multi-time resolution transformation by DWT, and then $h_t^{(k)}$ is fed in the attention model to learn the importance of features at different time resolutions (s_t). After that, s_t is input to the second layer LSTM to learn long-term and short-term dependencies ($h_t^{'}$), and finally get the predicted value of RUL.

$$
\begin{aligned}
g_k &= \phi(W_{xc}x_t^k + W_{hc}h_t^{k-1} + b_c),\\
i_k &= \sigma(W_{xi}x_t^k + W_{hi}h_t^{k-1} + W_{ci}c_{k-1} + b_i),\\
f_k &= \sigma(W_{xf}x_t^k + W_{hf}h_t^{k-1} + W_{cf}c_{k-1} + b_f),\\
c_k &= f_k \odot c_{k-1} + i_k \odot g_k,\\
o_k &= \sigma(W_{xo}x_t^k + W_{ho}h_t^{k-1} + W_{co}c_{k-1} + b_o),\\
h_t^k &= o_k \odot \sigma(c_k),
\end{aligned}
\tag{7}
$$

where the symbols f, i, c, and o denote the forgetting gate, the input gate, the storage cell, and the output gate, respectively. W_{xi}, W_{hi}, W_{ci}, W_{xf}, W_{hf}, W_{cf}, W_{xc}, W_{hc}, W_{xo}, W_{ho} and W_{co} are the weights, b_i, b_f, b_c and b_o are the bias cells, $\sigma(.)$ and $tanh$ are the sigmoid function and tanh function.

Since the hidden features at different temporal resolutions contribute differently to the aero-engine RUL, the hidden features that contribute more to the RUL should be given greater weights. We use the attention mechanism to assign different weights w_t^k for different temporal resolutions. as the following equation:

$$u_t^k = \tanh(W_w h_t^k + b_w), \tag{8}$$

$$w_t^k = \frac{exp(u_t^k u_w)}{\sum_i^K u_t^k u_w}, \tag{9}$$

where u_t^k represents the hidden feature of h_t^k passing through the next layer of MLP. u_w represents the context vector, which is initialized as a random vector. Then the integral state s_t obtained based on the attention weight w_t^k is:

$$s_t = \sum_{k=1}^{K} w_t^k h_t^k, \tag{10}$$

We use s_t as input to the second layer of LSTM units to learn both long-term and short-term time dependence:

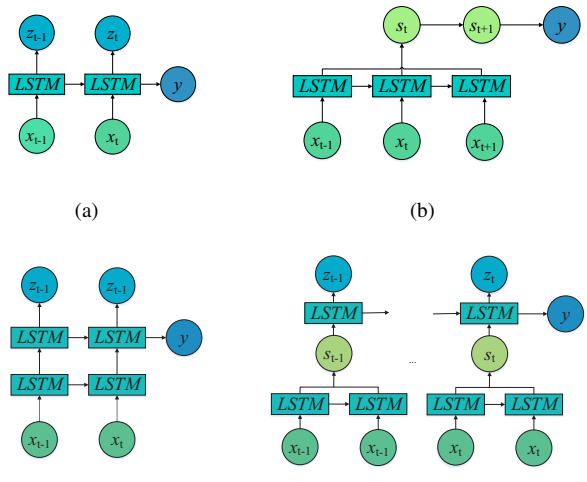


Fig. 5: Comparison of four LSTM architectures: (a) one-layer LSTM; (b) one-layer attention LSTM; (c) stacked LSTM; (d) our proposed MR-LSTM.

$$
\begin{aligned}
\begin{bmatrix} i_t \\ f_t \\ o_t \\ g_t \end{bmatrix} &= \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ tanh \end{bmatrix} W \cdot \begin{bmatrix} h'_{t-1} \\ s_t \end{bmatrix}\\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t,\\
h'_t &= o_t \odot tanh(c_t),
\end{aligned}
\tag{11}
$$

where w is the parameter vector for the second layer LSTM. Equation system 11 is the same calculation as the first layer LSTM. The final regression layer performs the prediction. The loss function is set to mean square error (MSE) loss.

We compared our proposed MR-LSTM model with a single-

layer LSTM as well as a stacked LSTM architecture. Fig.5a and Fig.5b are single-layer LSTMs, where Fig.5a is without the attention mechanism and Fig.5b is a single-layer LSTM with the attention mechanism. Fig.5c is stacked LSTMs that learn the LSTM units of the previous layer to obtain multi-time resolution information. Fig.5d is our proposed attentive LSTM model with two layers (MR-LSTM) that uses the attention mechanism to weigh the features of the hidden layers at different temporal resolutions and then learns a combined multi-temporal resolution representation at each time step.

IV. Simulation Experiments

A. Data Description

We validated the performance of the experimental algorithm using the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset. This dataset was provided by the NASA Data Centre [21]. The training set contains all the data for all samples from the initial state to the occurrence of failure, while the test set only provides data for the first part of the life cycle, all that needs to be done is to predict the RUL value of the equipment after the first part of the life cycle of the test set. The overall file is divided into four sub-datasets, FD001, FD002, FD003, and FD004, each of which is collected under different operating conditions and failure types (Tab. I).

TABLE I: Details about C-MAPSS dataset.

Data	FD001	FD002	FD003	FD004
Train trajectories	100	260	100	249
Test trajectories	100	259	100	248
Operating conditions	1	6	1	6
Fault modes	1	1	2	2

Each engine has varying degrees of initial wear or manufacturing variation, which are unknown and considered healthy, and each monitoring sample consists of 3 dimensions of operational information and 21 dimensions of sensor information. 3 operational parameters are flight altitude, Mach, and throttle lever solver angle. 21 sensor monitoring parameters are specific to the study module, obtained through the engine simulation model, and include noise.

With many internal components, diverse degradation patterns, and complex operating environments, different operating parameters may accelerate or decelerate the degradation process. A subset of data containing six different operating conditions, FD002 and FD004, were selected for the experiment, where the parameters for these six conditions are shown in Tab. II, where H is the height, Ma is the Mach number and TRA is the throttle lever angle.

Another dataset that has been used for evaluation is the N-CMAPSS dataset, which is a dataset generated from the operation-to-failure trajectories of a small group of aircraft engines under realistic flight conditions [22]. N-CMAPSS is divided into eight subsets according to different failure modes. Different subsets degrade for different reasons. Flights are classified into three categories according to their length: one for 1-3 hours flights, two for 3-5 hours flights, and three for longer flights. A detailed overview of the N-CMAPSS dataset is shown in Tab. III.

TABLE II: The parameters for these six conditions.

Condition	1	2	3	4	5	6
H	0	10	20	25	35	42
Ma	0	0.25	0.7	0.62	0.84	0.84
TRA	100	100	100	60	100	100

TABLE III: Details about N-CMAPSS dataset.

Name	Units	Flight classes	Failure modes
DS01	10	$1, 2, 3$	1
DS02	10	$1, 2, 3$	2
DS03	10	$1, 2, 3$	1
DS04	10	$2, 3$	1
DS05	10	$1, 2, 3$	1
DS06	10	$1, 2, 3$	1
DS07	10	$1, 2, 3$	1

B. Data pre-processing

As not all of the characteristic parameters provide valuable information in failure prediction, the trends of the 21 sensor parameters in the C-MAPSS dataset were observed and analyzed. Fig. 6 shows the trends of the 21 sensor data in the FD002 dataset over time, with the x-axis representing the number of cycles and the y-axis representing the sensor parameter values. Some of these variables (sensors 1, 5, 6, 10, 16, 18, 19) have stable and constant values and can be considered insensitive to performance degradation, and are not strongly associated with degradation trends. Therefore, these variables will not be used in the calculations in this paper. All the remaining 14 variables (sensors 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20, 21) were selected for modeling [23]. For the N-CAMPSS data set, the variables fan_flow_mod, fan_eff_mod, LPC_flow_mod, LPC_eff_mod, HPC_flow_mod, HPC_eff_mod, HPT_flow_mod, LPT_eff_mod, LPT_flow_mod have values of 0, so these variables are removed from the dataset.

As the individual attributes have different magnitudes, it is necessary to statute all the values of the attributes into an identical value space. Therefore, the data is normalized uniformly. We use the common min-max norm method to normalize the smoothed data, $max(k)$ and $min(k)$ are the maximum and minimum values on attribute k, respectively, and d denotes the value of sample i on attribute k. The normalization formula is:

$$\tilde{d}_{ik} = 2\frac{d_{ik} - \min(k)}{\max(k) - \min(k)} - 1. \tag{12}$$

C. Degradation model

Because precise knowledge of the target RUL needed to train the model is rarely accessible in PHM applications, it is

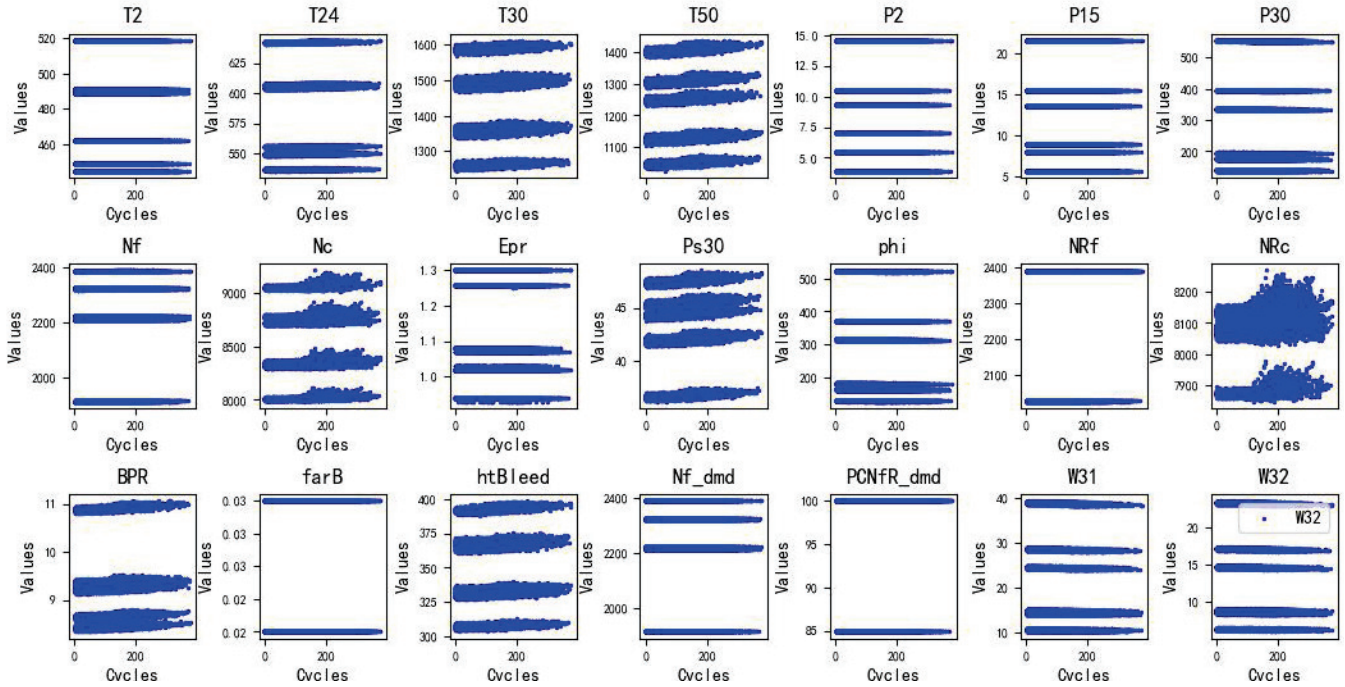


Fig. 6: Degradation of signals from all the 21 sensors in the data set FD002.

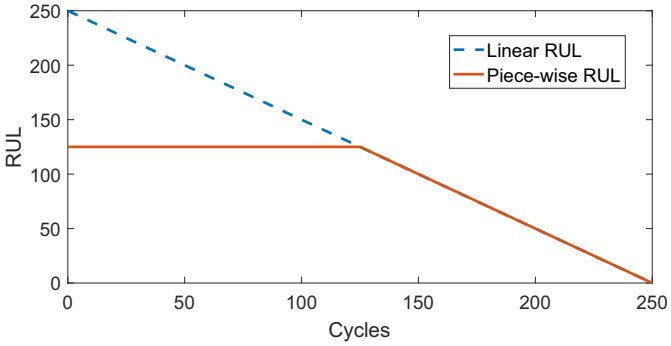


Fig. 7: Piece-wise linear RUL target function.

anticipated to use a physical model [6]. Because the equipment is healthy at first, and the degradation rises as the equipment approaches its end of life, we employed a piece-wise linear degradation model to establish the target RUL [6], [24], [25]. This strategy has been proven to be successful. The piece-wise linear function used for the C-MAPSS dataset is given in Fig. 7; it has a constant RUL stage of 125. N-CMAPSS has a constant RUL stage of 65, which is chosen according to the degradation of the data.

D. Evaluation Metric

In order to ensure the validity of the prediction results and to assess the prediction accuracy. We use a number of common evaluation methods, thus completing a comprehensive evaluation of prediction performance. The evaluation indicators are as follows:

(1) Root Mean Squared Error

Root Mean Squared Error (RMSE) is a measure of error that uses the sum of the squares of the point-by-point differences between the original and predicted values, divided by the total number of predictions. As can be seen, the RMSE is sensitive to data with very large or small deviations by squaring the difference to amplify the weight of the sample points with large deviations, and can effectively reflect the accuracy of the prediction.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2}. \tag{13}$$

(2) Penalty scores

PHM08 gives a penalty score (Score) for RUL prediction [26], which gives a greater penalty score for lagging predictions that exceed the actual life, and this scoring method allows the selection of methods with better forward prediction capability.

$$Score = \begin{cases} \sum_{i=1}^{N}\left(e^{-\frac{p_i - y_i}{13}} - 1\right), & if \quad p_i < y_i \\ \sum_{i=1}^{N}\left(e^{\frac{p_i - y_i}{10}} - 1\right), & if \quad p_i \geq y_i. \end{cases} \tag{14}$$

where n is the total number of observations in the test dataset, p_i is the predicted value and y_i is the actual value. All evaluation metrics are assessed as the lower the value, the better the prediction.

E. Experimental setup

The MR-LSTM model is a two-layer LSTM model with 50 hidden units per layer as feature extractors and two fully connected layers as regression modules, each followed by a dropout to prevent overfitting. We used the grid search method

TABLE IV: Default hyperparameters for MR-LSTM models.

Hyperparameter	Description	Value
Batch size	The number of samples taken in the training set for each training session	64
Dropout rate	The probability of randomly discarding neurons in the LSTM network	0.001
Epoch	The number of training sessions using all the samples in the training set	50
Ir of Adam	The learning rate of the optimization algorithm	0.3

to determine the hyperparameters of the model optimizer. Due to the high computational cost in a grid search, a part of the hyperparameters of the MRLSTM model was obtained empirically. Considering that prediction results can be influenced by many factors, the results of each experiment are averaged over 10 replicate experiments. We use the mean squared error (MSE) as the loss function, by minimizing the loss function so that the difference between the true and predicted values of the RUL becomes progressively smaller. Adam is used to optimizing the model parameters for each layer, and relu was the activation function. The empirically obtained default values of the hyperparameters are given in Table IV.

The parameters DWT scale factor, time window, and the number of hidden cells were chosen automatically. DWT scale factor = [2,3,4], time window = [30,60,90], and the number of hidden layer cells = [40,50,60] were set as a dictionary of parameters and the dictionary was passed to the grid search function. The optimal hyperparameters for the DWT grid search were 3, the time window was 60, and the number of hidden cells was 50. All methods are implemented in Python, with traditional machine learning models based on the scikit-learn package and neural networks implemented using TensorFlow. the decomposition and reconstruction of DWT use PyWavelet as the toolkit.

F. Experiment Results on CMAPSS

(1)Performance comparison using different scale factors and sliding windows

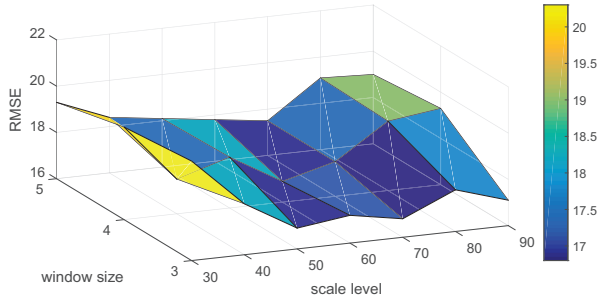
The size of the sliding window and the DWT scale has a significant impact on the RUL prediction performance. Fig. 8 illustrates the RMSE error of RUL on two sub-datasets using different DWT scale values and sliding window sizes.

As the size of the sliding window increases and the amount of information subsequently becomes larger, the tendency for the RMSE to decrease becomes apparent. In addition, the model performance is improved by increasing the scale factor. However, too large a sliding window size and scale factor will result in too much time granularity being integrated into the model, which will lead to over-fitting. Based on the experimental results, for the FD002 dataset, we choose a time window of 70 and a scale factor of 5. For the FD004 dataset, we choose a time window of 90 and a scale factor of 3.

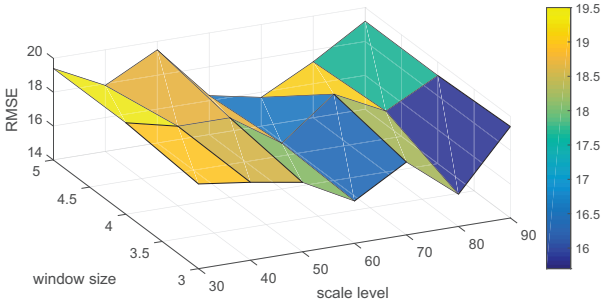
(2)Ablation study of the proposed approach

In Table V, the proposed MR-LSTM model (two-layer LSTM with attention and DWT) is compared with several LSTM variants using the FD004 dataset.

From the experimental results in Table V, the two-layer LSTM outperforms the single-layer LSTM, which indicates that considering temporal correlation is useful for prediction.



(a) FD002



(b) FD004

Fig. 8: RMSE errors for different sliding window sizes and time scales.

TABLE V: Performance comparison between MR-LSTM and other variants of LSTM methods.

Basenet	DWT	SM	CA	SA	RMSE	Score
LSTM-1					22.46	5061.39
LSTM-1	✓				20.07	2467.22
LSTM-1		✓			19.52	2207.76
LSTM-1	✓	✓			17.88	2003.91
LSTM-2					18.08	1972.09
LSTM-2	✓				17.56	1893.60
LSTM-2		✓			16.97	1883.62
LSTM-2	✓		✓		18.58	2237.45
LSTM-2	✓			✓	20.03	4586.21
LSTM-2	✓	✓			16.81	1785.33

LSTM-1 represents one-layer LSTM, LSTM-2 represents two-layer LSTM; SM: Self-attention; CA: Channel Attention; SA: Spatial Attention.

And the single-layer LSTM is compared with the single-layer LSTM using attention to conclude that attention has a positive effect on prediction. The experiments were conducted using self-attention, channel-attention, and spatial attention [27] with a two-layer LSTM backbone, and the experimental results obtained from self-attention outperformed the other two attention mechanisms. This is because the MR-LSTM model

only uses the attention mechanism to assign different weights to the hidden layer features at different scales, and does not need to use such complex mechanisms as the channel attention mechanism (what to pay attention to) and the spatial attention mechanism (where to pay attention to). The MR-LSTM model has the best RUL prediction performance. This is due to the fact that the MR-LSTM model introduces DWT to generate a large amount of information at different temporal resolutions. The hidden layer features at different temporal resolutions contribute differently to the RUL of the aero-engine, and the hidden layer features that contribute more to the RUL should be given more weight. The attention mechanism assigns adaptive weights to features at different temporal resolutions, which effectively improves the performance of RUL prediction and makes the model more robust.

(3) Performance comparison using different multi-time window reconstruction methods

To verify the effectiveness of DWT for multi-time window reconstruction, we designed to use different methods to compare with MR-LSTM. The details of the compared methods are extracted using CNN methods with different convolutional kernel sizes for multi-temporal resolution features and then fed into our proposed attention LSTM model with two layers for prediction. The experimental results are shown in Table VI.

TABLE VI: Performance comparison of MR-LSTM(DWT) with other multi-temporal resolution methods.

Methods	FD002		FD004	
	RMSE	Score	RMSE	Score
CNN-LSTM	19.35	3747	22.22	4844
MR-LSTM(DWT)	15.71	1434.27	16.81	1785.33

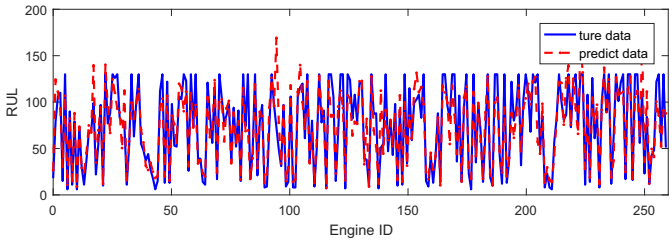
The results of Table VI show that the results of MR-LSTM are better. The reason is that the collected aero-engine signal data is noisy. The DWT method used in the MR-LSTM model, it is not only possible to generate sequences with multi-temporal resolutions but also to remove noise from the signal. In contrast, CNN only uses three convolutional block structures of different sizes to extract features at different temporal resolutions.

(4) Compare to state-of-the-arts

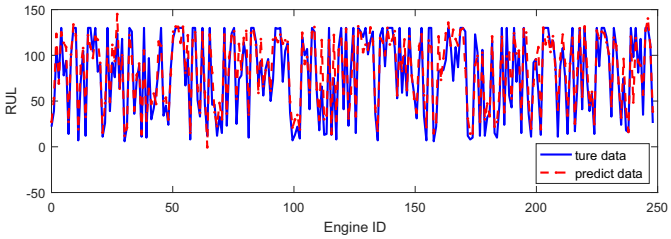
In this section, the proposed MR-LSTM model is compared with the RUL prediction models used in other papers. These comparison methods can be divided into two types: LSTM-based methods and CNN-based methods. Table VII summarizes the performance comparison results.

Our proposed MR-LSTM model has powerful sequential modeling capability compared with CNN-based methods (e.g., TCMN [15], AGCNN [8]), resulting in better RUL prediction performance. In addition, compared with mainstream LSTM-based methods (e.g., BLSTM [12], AtLSTM [11], DtLSTM [10], Transformer [28], GAN [29], etc.), the MR-LSTM model considers the effects of multiple conditions on RUL prediction performance, and thus achieves good prediction results on both data subsets.

To show the prediction results more clearly, we compare the actual values with the predicted values obtained using the MR-LSTM model. Fig. 9 shows the comparison results of the



(a) FD002



(b) FD004

Fig. 9: Comparison between the ground truth RUL and the RUL predicted by the proposed model (MR-LSTM) on the two sub-datasets.

actual and predicted values for all engine units in the two sub-datasets (FD002 and FD004).

Both Fig. 9 and Fig. 10 show shows a better fit of the predicted RUL to the true RUL on both sub-datasets, illustrating the validity of the MR-LSTM model for RUL prediction. In addition, FD002 outperforms FD004 in terms of prediction performance as FD002 is relatively simple and has only one failure mode.

In addition, we randomly select an engine unit from the two sub-datasets (FD002 and FD004), and compare its true value and predicted value. The AtLSTM method is an RUL prediction method that combines an attention mechanism and a feature fusion framework [11]. We show the comparison between the predicted and true values of the proposed MR-LSTM model and the AtLSTM model, and the results are shown in Fig 10.

As shown in Fig 10, the RUL predicted by the MR-LSTM model is closer to the true value compared to the AtLSTM model on these two sub-data. This is because the AtLSTM model does not take into account that the engine degradation is greatly influenced by the parameters of multiple operating conditions, resulting in its predicted values deviating more from the actual values, and our proposed MR-LSTM model can achieve stable prediction of RUL under multiple operating conditions.

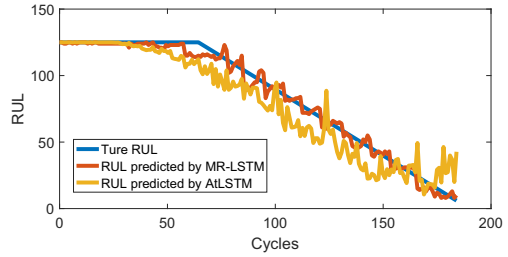
G. Results on the N-CAMPSS dataset

There are few and limited datasets representing aero engines. In this section, we use the N-CMAPSS dataset to validate the effectiveness of the MR-LSTM model. We use 70% of each subset for training and 30% for testing. We use the Transformer method [30] to compare with our proposed MR-LSTM method. We tested the RMSE errors of the proposed models MR-LSTM and Transformer models under seven sub-

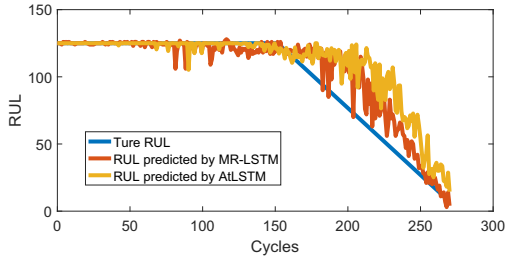
TABLE VII: Performance comparison between the MR-LSTM and other approaches found in the literature.

Data	Criterion	CNN		LSTM					
		TCMN [15]	AGCNN [8]	Transformer [28]	AtLSTM [11]	BLSTM [12]	GAN [29]	DtLSTM [10]	**MR-LSTM**
FD002	RMSE	20.45	19.43	22.81	21.51	21.11	20.06	17.87	15.71
	Score	3100.00	1492.00	-	4205.91	4793.00	4085.00	-	1434.27
FD004	RMSE	21.03	21.50	24.86	27.08	26.61	20.88	21.81	16.81
	Score	4000.00	3392.00	-	5649.14	4971.00	3872.00	-	1785.33

The experimental results of the comparison method are from the references, '-' means that this result does not exist in the references.



(a) FD002–7 unit



(b) FD004–12 unit

Fig. 10: Comparison of the ground truth RUL and the predicted RUL in a unit of the two sub-datasets.

datasets of the N-CMAPSS dataset, and the results are shown in Table VIII.

TABLE VIII: Performance comparison between the MR-LSTM and other approaches found in the literature on the N-CAMPSS dataset.

Data	Transfomer	MR-LSTM
DS01	4.57	3.39
DS02	5.34	4.52
DS03	6.73	6.27
DS04	9.25	8.36
DS05	5.18	7.91
DS06	4.08	8.31
DS07	6.71	7.29

As shown in Table VIII, the RMSEs obtained using the MR-LSTM model are less than 10 in all seven subsets of N-CMAPSS, indicating the high accuracy of our method. Moreover, the MR-LSTM model proposed in this paper obtains mostly better results than the Transformer method. Therefore, our proposed MR-LSTM still performs well under realistic flight conditions.

V. Conclusion

In this paper, we presented a unique multi-resolution LSTM model (MR-LSTM), which employs the DWT to extract plentiful information of various temporal resolutions from time sequences. Following that, the input from numerous temporal resolutions is combined into a two-layer LSTM model based on an attention mechanism, with the first layer learning attention from various temporal resolutions to generate an integrated representation of multiple temporal resolutions. The LSTM cells in the second layer are designed to make use of the temporal dependency among the input representations from the previous layer. Experiments were conducted with NASA's C-MAPSS dataset and the newly published N-CMAPSS dataset. The experimental results show that the proposed MR-LSTM model provides an effective RUL prediction for aero-engines. There is still some limitation in the proposed approach that should be addressed. Further work may investigate the integration of the knowledge distillation framework into the MR-LSTM approach for real-time aero-engine RUL prediction.

References

[1] H. Li, H. Z. Huang, Y. F. Li, J. Zhou, and J. Mi, "Physics of failure-based reliability prediction of turbine blades using multi-source information fusion," *Applied Soft Computing*, vol. 72, pp. 624–635, Nov. 2018.

[2] C. Álvarez, M. López-Campos, R. Stegmaier, F. Mancilla-David, R. Schurch, and A. Angulo, "A condition-based maintenance model including resource constraints on the number of inspections," *IEEE Transactions on Reliability*, vol. 69, no. 3, pp. 1165–1176, Dec. 2020.

[3] X. Jia, M. Zhao, Y. Di, Q. Yang, and J. Lee, "Assessment of data suitability for machine prognosis using maximum mean discrepancy," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5872–5881, Nov. 2018.

[4] Y. Gao, Y. Wen, and J. Wu, "A neural network-based joint prognostic model for data fusion and remaining useful life prediction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 117–127, Mar. 2021.

[5] Q. Zhai and Z.-S. Ye, "Rul prediction of deteriorating products using an adaptive wiener process model," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 2911–2921, Dec. 2017.

[6] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2306–2318, Oct. 2017.

[7] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, Apr. 2018.

[8] H. Liu, Z. Liu, W. Jia, and X. Lin, "Remaining useful life prediction using a novel feature-attention-based end-to-end approach," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 1197–1207, Jul. 2021.

[9] R. Liu, B. Yang, and A. G. Hauptmann, "Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 87–96, Jan. 2020.

[10] H. Miao, B. Li, C. Sun, and J. Liu, "Joint learning of degradation assessment and rul prediction for aero-engines via dual-task deep lstm networks," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5023–5032, Jun. 2019.

[11] Z. Chen, M. Wu, R. Zhao, F. Guretno, and X. Li, "Machine remaining useful life prediction via an attention based deep learning approach," *IEEE Transactions on Industrial Electronics*, vol. PP, no. 99, pp. 1–1, Jun. 2020.

[12] C. Huang, H. Huang, and Y. Li, "A bidirectional lstm prognostics method under multiple operational conditions," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 11, pp. 8792–8802, Jan. 2019.

[13] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2727–2736, Mar. 2018.

[14] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[15] L. Jayasinghe, T. Samarasinghe, C. Yuenv, J. C. Ni Low, and S. Sam Ge, "Temporal convolutional memory networks for remaining useful life estimation of industrial machinery," in *2019 IEEE International Conference on Industrial Technology (ICIT)*, 2019, pp. 915–920.

[16] P. Li, Z. Zhang, R. Grosu, Z. Deng, J. Hou, Y. Rong, and R. Wu, "An end-to-end neural network framework for state-of-health estimation and remaining useful life prediction of electric vehicle lithium batteries," *Renewable and Sustainable Energy Reviews*, vol. 156, no. 1, p. 111843, Mar. 2022.

[17] Z. Deng, L. Xu, H. Liu, X. Hu, Z. Duan, and Y. Xu, "Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles," *Applied Energy*, vol. 339, no. 1, p. 120954, Jun. 2023.

[18] M. J. Shensa *et al.*, "The discrete wavelet transform: wedding the a trous and mallat algorithms," *IEEE Transactions on signal processing*, vol. 40, no. 10, pp. 2464–2482, 1992.

[19] C. Yan, Y. Tu, X. Wang, Y. Zhang, X. Hao, Y. Zhang, and Q. Dai, "Stat: Spatial-temporal attention mechanism for video captioning," *IEEE transactions on multimedia*, vol. 22, no. 1, pp. 229–241, Jun. 2019.

[20] Z. Hua, Z. Zheng, E. Pahon, M.-C. Péra, and F. Gao, "Multi-timescale lifespan prediction for pemfc systems under dynamic operating conditions," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 1, pp. 345–355, Mar. 2022.

[21] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *2008 International Conference on Prognostics and Health Management*, Denver, CO, USA, 2008, pp. 1–9.

[22] M. Arias Chao, C. Kulkarni, K. Goebel, and O. Fink, "Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics," *Data*, vol. 6, no. 1, p. 5, Jan. 2021.

[23] T. Xu, G. Han, L. Gou, M. Martinez-Garcia, D. Shao, B. Luo, and Z. Yin, "Sgbrt: An edge-intelligence based remaining useful life prediction model for aero-engine monitoring system," *IEEE Transactions on Network Science and Engineering*, 2022.

[24] X. Li, Q. Ding, and J. Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering System Safety*, vol. 172, no. 10, pp. 1–11, 2018.

[25] Y. Wu, M. Yuan, S. Dong, L. Lin, and Y. Liu, "Remaining useful life estimation of engineered systems using vanilla lstm neural networks," *NEUROCOMPUTING*, vol. 275, no. 10, pp. 167–179, 2018.

[26] T. Wang, Jianbo Yu, D. Siegel, and J. Lee, "A similarity-based prognostics approach for remaining useful life estimation of engineered systems," in *2008 International Conference on Prognostics and Health Management*, Denver, CO, USA, 2008, pp. 1–6.

[27] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham: Springer International Publishing, 2018, pp. 3–19.

[28] Y. Mo, Q. Wu, X. Li, and B. Huang, "Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit," *Journal of Intelligent Manufacturing*, vol. 32, no. 10, p. 1–10, Oct. 2021.

[29] S. Behera and R. Misra, "Generative adversarial networks based remaining useful life estimation for iiot," *Computers & Electrical Engineering*, vol. 92, p. 107195, May. 2021.

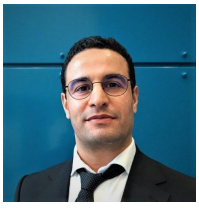
[30] X. Li, J. Li, L. Zuo, L. Zhu, and H. T. Shen, "Domain adaptive remaining useful life prediction with transformer," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, Aug. 2022.

Tiantian Xu is currently pursuing the Ph.D. degree in the School of Software at Dalian University of Technology, China. She received the M.S. degree in software engineering at Tianjin normal university of science and technology, in 2019. Her current research interests include fault detection of Industrial Internet of Things.

Guangjie Han [*corresponding author*] (Fellow, IEEE) is currently a Professor with the Department of Internet of Things Engineering, Hohai University, Changzhou, China. He received his Ph.D. degree from Northeastern University, Shenyang, China, in 2004. In February 2008, he finished his work as a Postdoctoral Researcher with the Department of Computer Science, Chonnam National University, Gwangju, Korea. From October 2010 to October 2011, he was a Visiting Research Scholar with Osaka University, Suita, Japan. From January 2017 to February 2017, he was a Visiting Professor with City University of Hong Kong, China. From July 2017 to July 2020, he was a Distinguished Professor with Dalian University of Technology, China. His current research interests include Internet of Things, Industrial Internet, Machine Learning and Artificial Intelligence, Mobile Computing, Security and Privacy. Dr. Han has over 500 peer-reviewed journal and conference papers, in addition to 160 granted and pending patents. Currently, his H-index is 64 and i10-index is 270 in Google Citation (Google Scholar). The total citation count of his papers raises above 14800+ times.

Dr. Han is a Fellow of the UK Institution of Engineering and Technology (FIET). He has served on the Editorial Boards of up to 10 international journals, including the IEEE TII, IEEE TCCN, IEEE Systems, etc. He has guest-edited several special issues in IEEE Journals and Magazines, including the IEEE JSAC, IEEE Communications, IEEE Wireless Communications, Computer Networks, etc. Dr. Han has also served as chair of organizing and technical committees in many international conferences. He has been awarded 2020 IEEE Systems Journal Annual Best Paper Award and the 2017-2019 IEEE ACCESS Outstanding Associate Editor Award. He is a Fellow of IEEE.

Hongbo Zhu (S'19-M'20) received the B.Sc., M.Eng., Ph.D. degrees from Northeastern University, Shenyang, China, in 2009, 2012, and 2020, respectively. He is an associate professor with School of Information Science and Engineering, Shenyang Ligong University. His research interests include medical image computing and deep learning.



Tarik Taleb (Senior Member, IEEE) received the B.E. degree (with distinction) in information engineering and the M.Sc. and Ph.D. degrees in information sciences from Tohoku University, Sendai, Japan, in 2001, 2003, and 2005, respectively. He is currently a Professor at the Center of Wireless Communications, the University of Oulu, Finland. He is the founder and the Director of the MOSA!C Lab, Espoo, Finland. He was an Assistant Professor with the Graduate School of Information Sciences, Tohoku University, in a laboratory fully funded by KDDI until 2009. He was a Senior Researcher and a 3GPP Standards Expert with NEC Europe Ltd., Heidelberg, Germany. He was then leading the NEC Europe Labs Team, involved with research and development projects on carrier cloud platforms, an important vision of 5G systems. From 2005 to 2006, he was a Research Fellow with the Intelligent Cosmos Research Institute, Sendai. He has also been directly engaged in the development and standardization of the Evolved Packet System as a member of the 3GPP System Architecture Working Group. His current research interests include architectural enhancements to mobile core networks (particularly 3GPP's), network softwarization and slicing, mobile cloud networking, network function virtualization, software defined networking, mobile multimedia streaming, intervehicular communications, and social media networking.

Jinlin Peng received the B.S. degree in electromechanical engineering from Beijing Institute of Technology, Beijing, China, and the Ph.D. degree in electrical engineering from the University of Leeds, UK. He was formerly a Post-Doctoral Researcher with the Department of Electronic Engineering, Tsinghua University, Beijing, China. He is now an associate professor in Artificial Intelligence Research Center, National Innovation Institute of Defense Technology, China. His main research interests include wireless network protocols, ad hoc networks, signal processing in wireless communications and machine learning.