


# A Hybrid Prognostic Approach Based on Deep Learning for the Degradation Prediction of Machinery

 Ahmet Kara<sup>1</sup>

<sup>1</sup>Corresponding Author Department of Industrial Engineering, Hitit University, Turkey;  
ahmetkara@hitit.edu.tr

Received 09 April 2021; Revised 27 June 2021; Accepted 07 July 2021; Published online 31 August 2021

## Abstract

Remaining useful life (RUL) prediction is of great significance for prognostic and health management (PHM) as it can achieve more reliable and effective maintenance strategies. With the advances in the field of deep learning, data-driven methods have provided promising prognostic prediction results. Hence, this research presents a data-driven prognostic approach based on deep learning models for predicting the RUL of mechanical systems effectively. Multiple separable convolution layers, a bidirectional Long Short-Term Memory (LSTM) layer, and fully-connected layers (FCL) are included in the proposed network, named the SC-BLSTM, to accomplish more accurate prognostic prediction from the raw degradation data acquired by different sensors. The proposed SC-BLSTM approach aims to learn complex and nonlinear features from the input data and capture temporal dependencies from the learned features. The presented approach in this research is tested and verified on the degradation data of turbofan engines (C-MAPSS dataset) from NASA. The result demonstrated that the SC-BLSTM is able to achieve more effective RUL prediction compared with some existing prognostic models.

**Keywords:** Remaining useful life, deep learning, separable convolutions, bidirectional LSTM, prognostics

## 1. Introduction

Prognostics and Health Management (PHM) is crucially significant for providing productivity and availability of modern mechanical systems. While conventional procedures like corrective maintenance and preventive maintenance have limited ability to decrease the overall cost and avoid unexpected failures, PHM provides the requirements related to more effective maintenance decision-making [1]. The fundamental objectives of PHM technology consist of avoiding unexpected system failures, minimizing the overall costs, and enhancing the reliability of the entire system. The remaining useful life (RUL) prediction of machinery has recently attracted more and more attention in PHM technologies, which is one of the most critical and challenging tasks [2].

Generally, prognostic approaches such as RUL estimation can be grouped into three categorized: model-based, data-driven, and hybrid methods [3]. Model-based techniques typically use historical degradation information to develop a mathematical model. They are less capable of predicting RUL effectively in complicated and noisy environments [4]. Conversely, data-driven methods aim to capture the degradation processing of complex mechanical systems using the past degradation data. With the recent advancements of sensor technology, data-driven models have been extensively utilized to estimate RUL effectively in the literature. Khazaei et al. [5] developed a new data-driven approach based on the artificial neural network (ANN) for RUL prediction of mechanical systems. Wang and Mamo [6] combined the gradient boosted regression (GBR) method and the artificial bee colony (ABC) algorithm to estimate the degradation progression of prismatic cells. As the last group, hybrid methods, which consist of the integration of model-based and data-driven methods, focus on taking advantage of the strengths of these models. Xue et al. [7] developed a prognostic approach that integrates unscented kalman filter and support vector regression (SVR) tuned by a genetic algorithm for RUL prediction. In this article, a novel data-driven technique is proposed for accurate and effective RUL estimation of mechanical systems.

Recently, several prognostic approaches based on deep neural networks have been extensively introduced in order to capture complex and nonlinear patterns from degradation data, and these



approaches have significantly increased the efficiency and reliability of RUL prediction. Convolutional neural network (CNN) [8], long-short term memory (LSTM) [9], gated recurrent unit (GRU) [10], and some hybrid methods [11], [12] have been implemented to take advantage of their superiority. For instance, Chen et al. [13] built an encoder-decoder structure consisting of CNN, bidirectional GRU, and attention mechanism for bearing RUL prediction. Wang et al. [14] utilized multiple separable CNN and fully-connected layers (FCL) to extract high-level features from raw degradation data with the aim of estimating the RUL of machinery. A deep learning-based prognostic method with partial observations addressed by Li et al. [15] was applied to realize more accurate and effective RUL estimation using an image dataset of cutting wheel.

Li et al. [16] presented a multi-scale deep CNN structure to learn complex features from turbofan engine degradation data, and their approach provided an effective RUL prediction performance. A hybrid deep learning framework that combines LSTM network and fully-connected layer introduced by Xia et al. [17] was handled to achieve RUL prediction of mechanical systems effectively. Yang et al. [18] presented a deep learning-based approach, which is combined CNN, attention mechanism and multiple bidirectional GRU, to investigate the interpretability of the neural network. Besides, an ensemble approach including CNN and bidirectional LSTM with multiple time windows developed by Tangbin Xia et al. [19] was utilized to learn the degradation trend of machinery.

In this paper, an integrated deep learning approach, called the SC-BLSTM, including multiple separable CNNs, a bidirectional LSTM and fully-connected layers, is addressed to effectively accomplish the RUL prediction of machinery. Multiple separable CNNs with different kernel sizes are applied to obtain discriminative information from the raw degradation measurements collected with multiple sensors. The second component of the proposed network comprising a bidirectional LSTM layer is employed to capture high-level representations from the learned features. Finally, fully-connected layers are applied to estimate the RUL of machinery. Experimental results demonstrated that the SC-BLSTM approach performs better performance compared with the existing related models. The primary contributions of this research are listed as follows:

- A data-driven based on multiple separable CNNs, bidirectional LSTM, and FCL is proposed for the RUL estimation of machinery so as to automatically reveal the sophisticated and nonlinear patterns from the degradation process.
- The hyperparameters of the proposed method, including the dropout rate, the number of filters in each separable convolutional layer, the number of units and batch size, are tuned by a grid search algorithm.
- Experimental consequences on the C-MAPSS dataset indicated that the presented method is capable of providing more effective RUL prediction compared with the existing relevant studies.

The rest of this research is designed as follows. Section 2 gives details about the components of the presented approach. Section 3 explains the SC-BLSTM framework in detail. The experimental settings are introduced in Section 4. The experimental outcomes are offered to verify the effectiveness and superiority of the SC-BLSTM model. In Section 5, this paper is summarized and concluded.

## 2. Technical Background

In this section, it will be presented details of the background on the basic components of the network consisting of depthwise separable convolution and bidirectional LSTM.

### 2.1 Depthwise separable convolution

The depthwise separable convolution, which is also named separable convolution, aims to effectively extract temporal and cross-channel relationships from different sensor data. Depthwise separable convolutions have been widely applied in the field of image classification, as they reduce computational time and the number of network parameters and avoid redundant learning correlations [20]. In contrast to traditional convolution, depthwise separable convolution comprises two sections, including

depthwise convolution and pointwise convolution, as seen in Figure 1. In the first step, a single convolution kernel is implemented to each input channel by depthwise convolution to capture temporal patterns separately. The number of input channels remains the same following the operation of depthwise convolution. This operation can be formulated as in Equations 1.

$$DepthConv(W_{depth}, x)_{(i, j)} = \sum_{h, l}^{H, L} W_{(h, l)} * x_{(i+h, j+l)} \quad (1)$$

Where  $W_{depth}$  represents the weight matrix of  $DepthConv$ .  $H$  and  $L$  denote the height and width of the input data. The  $*$  mark is the convolution operator.  $(i, j)$  represents the coordinates of the output features. As the second step, pointwise convolutions apply a  $1 \times 1$  convolution operation by combining the outputs of the first step. This step aims to effectively extract the cross-channel patterns from different data [21]. Accordingly, pointwise convolution operation can be expressed as:

$$PointConv(W_{point}, x)_{(i, j)} = \sum_k^K W_k * x_{(i, j)} \quad (2)$$

Where  $W_{point}$  denote the weight matrix.  $k$  is the size of convolution kernel. In general, the overall operation of depthwise separable convolution can be calculated by

$$SepConv(W_{depth}, W_{point}, x)_{(i, j)} = PointConv(W_{point}, x)_{(i, j)} * DepthConv(W_{depth}, x)_{(i, j)} \quad (3)$$

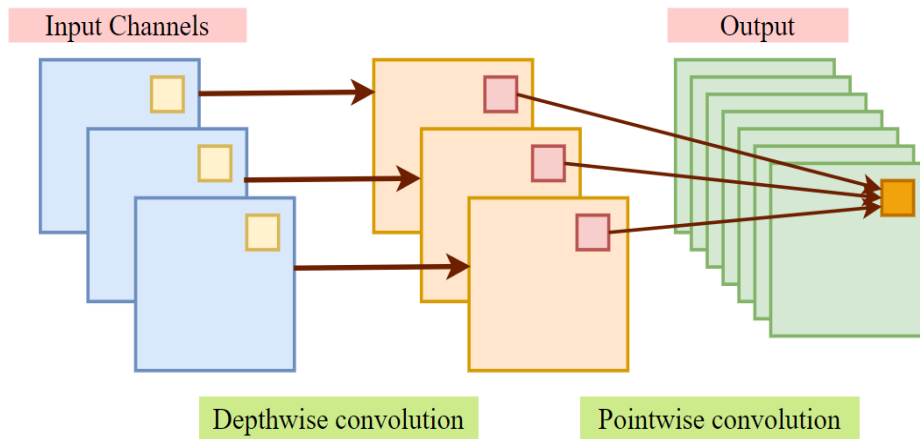


Figure 1 Depthwise separable convolution.

## 2.2 Bidirectional LSTM

In the SC-BLSTM model, bidirectional LSTM (BLSTM) is used to obtain temporal correlations between the features learned by separable CNN. BLSTM is widely employed in different time series prediction problems and indicates remarkable performance in RUL estimation applications [22]. The conventional LSTM layers consider the sequences of past information in a single direction. On the other hand, BLSTM consists of two different LSTM layers, which is utilized to consider past and future information. As is illustrated in Figure 2, the output of the BLSTM is obtained by combining the outputs of two layers. The overall process of BLSTM can be formulated as follows.

$$\vec{h}_t = \delta(W_{\vec{h}, \vec{h}} \cdot \vec{h}_{t-1} + W_{\vec{h}, x} \cdot x_t + b_{\vec{h}}) \quad (4)$$

$$h_t = \delta(W_{h, h} \cdot h_{t+1} + W_{h, x} \cdot x_t + b_h) \quad (5)$$

$$h_t = \theta(W_{h, h} \cdot h_t + W_{h, \vec{h}} \cdot \vec{h}_t) \quad (6)$$

Where  $\vec{h}_t$  denotes the forward LSTM layer,  $\overleftarrow{h}_t$  represents the backward LSTM layer.  $\delta(\cdot)$  represents the LSTM operation.  $\theta(\cdot)$  denotes the activation function.

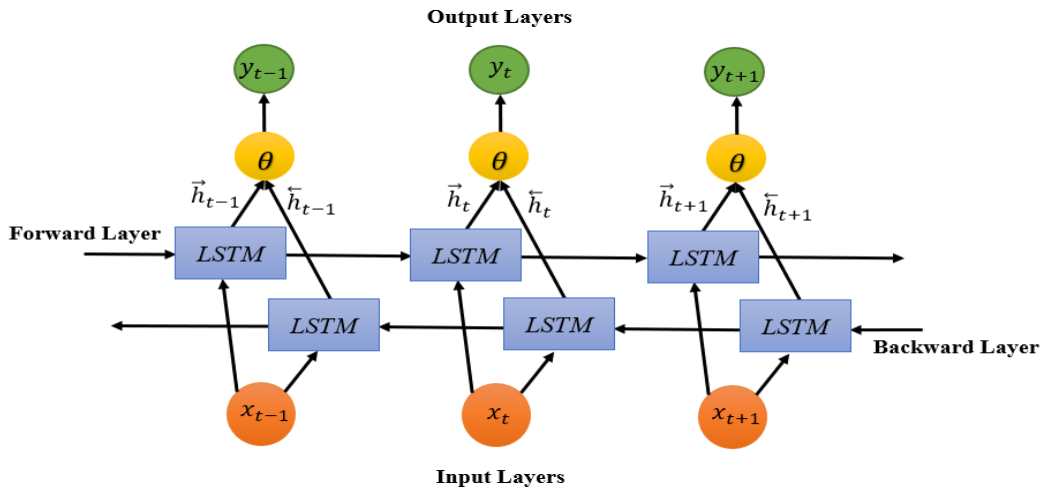


Figure 2 The Structure of Bidirectional LSTM.

### 3. The Proposed SC-BLSTM Model

Figure 3 illustrates the presented SC-BLSTM approach's architecture. First, two 1D separable CNN is built to automatically discover superior representations from the raw sensor measurement. The size of input data is  $tw \times ft$ , where  $tw$  denote the time window size and  $ft$  is the number of the predetermined features. In each separable CNN, the size of the kernels is different from each other. After learning the discriminative information by two separable CNNs, a BLSTM layer is used to effectively capture temporal dependencies from the extracted features. Then, a fully-connected layer is employed to map the learned features. In the last layer, a fully-connected layer with a single neuron is utilized as the output layer of the proposed network to perform RUL prediction.

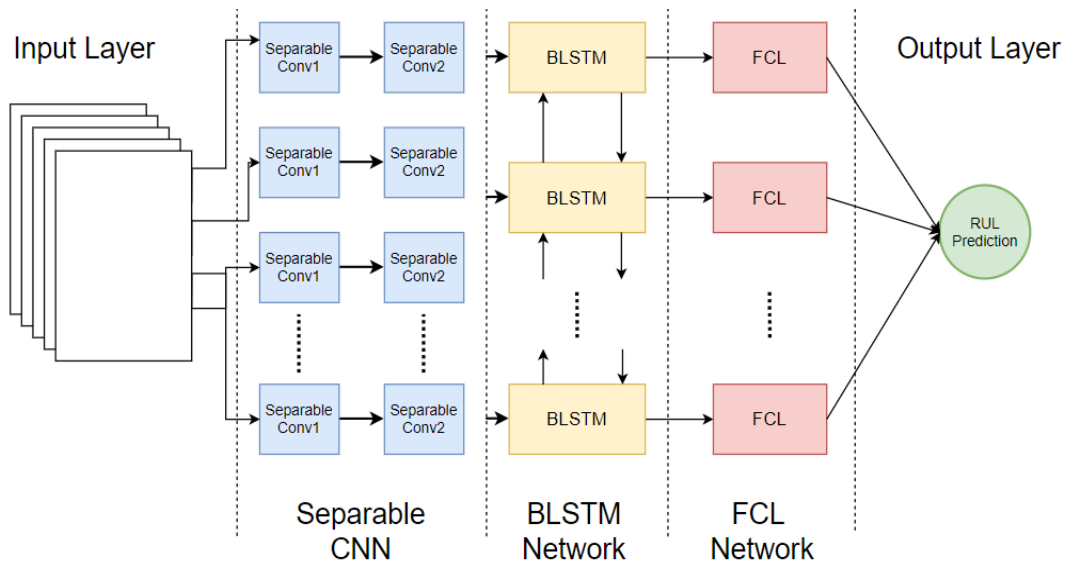


Figure 3 The Framework of the SC-BLSTM Model.

As depicted in Figure 4, the SC-BLSTM method comprises three steps; data preprocessing step, training step and testing step. In data preprocessing step, the raw degradation data is processed to perform effective prognostic prediction with the SC-BLSTM method. Training step aims to construct the optimized network using training dataset. Finally, the trained SC-BLSTM method is applied in testing step to achieve promising prognostic predictions using the testing dataset. To prevent the overfitting problem in the training step, the dropout operation is adopted after the BLSTM layer and the first FCL.

Besides, the mean square error (MSE) is applied as the loss function of the SC-BLSTM approach, which is formulated as follows.

$$MSE = \frac{1}{N_s} \sum_{i=1}^{N_s} (o_i - p_i)^2 \tag{7}$$

Where  $N_s$  is the number of the samples,  $o_i$  denotes the observed RUL values, and  $p_i$  represents the predicted RUL. In the proposed approach, the Adam algorithm is adopted to optimize the weight and biases of the network and enhance the prediction performance.

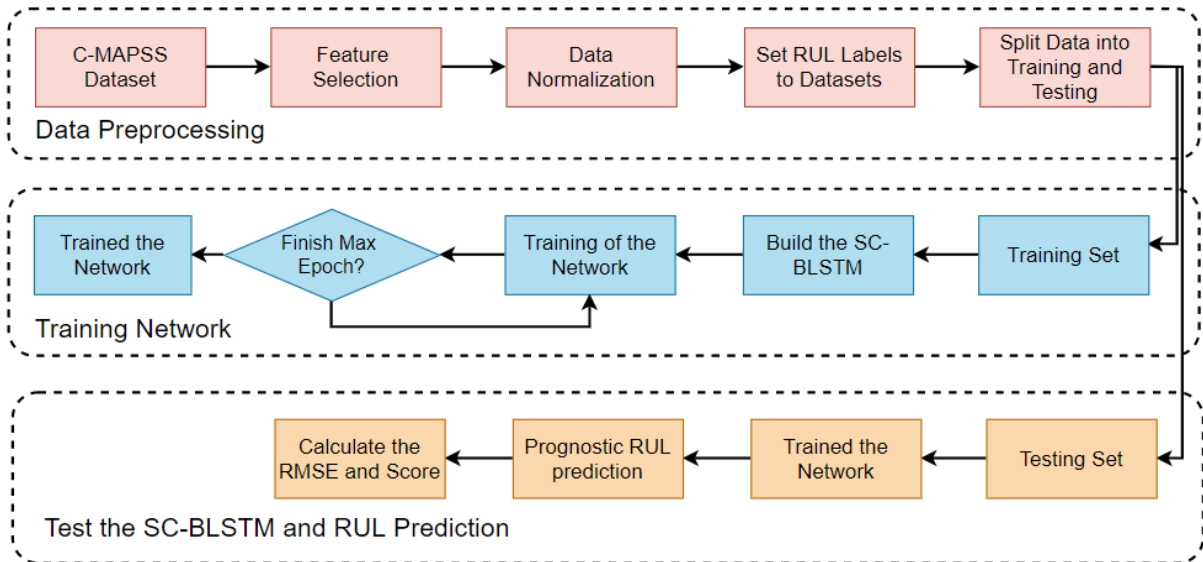


Figure 4 Implementation process of the SC-BLSTM prognostic approach.

#### 4. Experimental Study

This section shows the RUL estimation performance of the proposed SC-BLSTM approach and comparisons with the existing works. The assessment of the prognostic prediction performance is performed on the popular turbofan engine C-MAPSS dataset.

Table 1 The details of the turbofan engine dataset

Data	FD001	FD002	FD003	FD004
Units for training	100	260	100	249
Units for testing	100	259	100	248
Operating conditions	1	6	1	6
Fault modes	1	1	2	2

##### 4.1 C-MAPSS dataset

The Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset [23] introduced by NASA has been extensively utilized for the evaluations of RUL prediction. The C-MAPSS dataset is related to the aircraft turbofan engines degradation. This dataset is categorized into four subsets with varying operating conditions and fault modes, as depicted in Table 1, and each subset contains a training dataset and a testing dataset. All subsets include measurements collected by 21 sensors, engine number, time step, and three operational situations. The main aim is to predict the RUL value of each turbofan engine unit in the testing dataset. In this research, the maximum values of the training labels are clipped to no more than 125 with the aim of decreasing the prediction error when the actual RUL is greater than 125. This situation, which is a rectified RUL procedure, has been reported to remarkably enhance overall prognostic RUL prediction performance in the literature [17], [19]. The detailed definitions regarding the C-MAPSS are addressed in [23]. In this paper, FD003 and FD004 are used in experimental studies.

## 4.2 Experimental setting

In this article, the grid search algorithm is used to tune the hyperparameters of the proposed SC-BLSTM model for the purpose of increasing RUL estimation accuracy. The hyperparameters that need to be optimized consist of the number of filters in two separable CNN, the rate of dropout, and the number of units in BLSTM and FCL. To refrain from a large grid search space due to the computational time, the number of units and filters vary from 8, 16, 24, 32, to 48. Besides, the rate of dropout varies from 0.2, 0.3, to 0.5. Moreover, it should be noticed that 5-fold cross-validation is adopted in the training phase of each subset. The hyperparameters values tuned by the grid search algorithm and the other parameters values of the proposed network are reported in Table 2.

In the C-MAPSS dataset, several sensors have constant measurements that do not offer valuable information for RUL prediction. For this reason, these sensors measurements are removed from the input data. In this study, it is preferred 14 sensor measurements consisting of 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20, and 21 as the raw input data. Besides, the data gathered by the different sensors are standardized to be in the range [0, 1] utilizing the Min-Max scaling technique by Equation 8.

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (8)$$

For the purpose of appraising the performance of the suggested SC-BLSTM method, two metrics comprising the root mean square error (RMSE) and the scoring function have been adopted in this paper. The scoring function has been widely employed by various researchers in works regarding the C-MAPSS. These metrics are calculated as follows.

$$Scoring = \left\{ \sum_{i=1}^N \exp \exp \left( -\frac{d_i}{13} \right) - 1, \quad d_i < 0 \quad \sum_{i=1}^N \exp \exp \left( \frac{d_i}{10} \right) - 1, \quad d_i \geq 0 \right. \quad (9)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N d_i^2} \quad (10)$$

Where N represents the total number of the samples and  $d_i = RUL_i - RUL'_i$  is the prediction error for the  $i$  data sample. In addition, the testing dataset is fed into the proposed SC-BLSTM approach trained using the training dataset to estimate the RUL of the turbofan engine, and the prediction accuracy of the SC-BLSTM model is acquired.

Table 2 Parameter settings of the experimental methods

Parameter	Value (FD003 / FD004)	Parameter	Value (FD003 / FD004)
Number of filters	8-48 / 16-16	Activation function	Tanh / Tanh
Kernel sizes	3-5 / 3-5	Time window size	30 / 30
Neurons in BLSTM	48 / 8	Panding	same / same
Neurons in FCL	16/ 20	L2 regularization	1e-4 / 1e-4
Dropout rate	0.4 / 0.3	Learning rate	0.001 / 0.001
Batch size	100 / 100	$R_{early}$	125
Number of epochs	150 / 150	Optimizer	Adam

Finally, a specific seed value is adopted to provide reproducible results from the proposed network. All experiments are conducted on a personal computer with Intel Core i7-9750H CPU, and 16 GB of RAM using the Tensorflow 2.5.0 and sci-kit learn 0.24.

## 4.3 Results analysis and discussions

In this section, the prognostic outcomes achieved by the SC-LSTM approach for RUL estimation utilizing the C-MAPSS dataset are analyzed and compared with the results of the existing related methods in the literature.

The RUL estimation results of the test engines according to the last recorded life cycles from the subset FD003 and FD004 are illustrated in Figure 5. The test engines are sorted by the actual RUL values for better visualization. It can be obviously observed that the prognostic prediction values of the SC-LSTM approach are capable of following the real RUL values of the testing engine units in FD003 and FD004 subsets generally. In particular, it has been seen that the prediction error tends to be more in cases where the real RUL values of the test engines are higher. Moreover, the prognostic prediction effectiveness of the presented SC-BLSTM model on the FD003 sub-dataset is superior compared with the prediction results of the FD004 sub-dataset. The reason is that the FD004 subset consists of more operational situations and fault modes. Generally, it can be observed that the presented SC-BLSTM approach has the ability to learn the degradation progression of the mechanical systems.

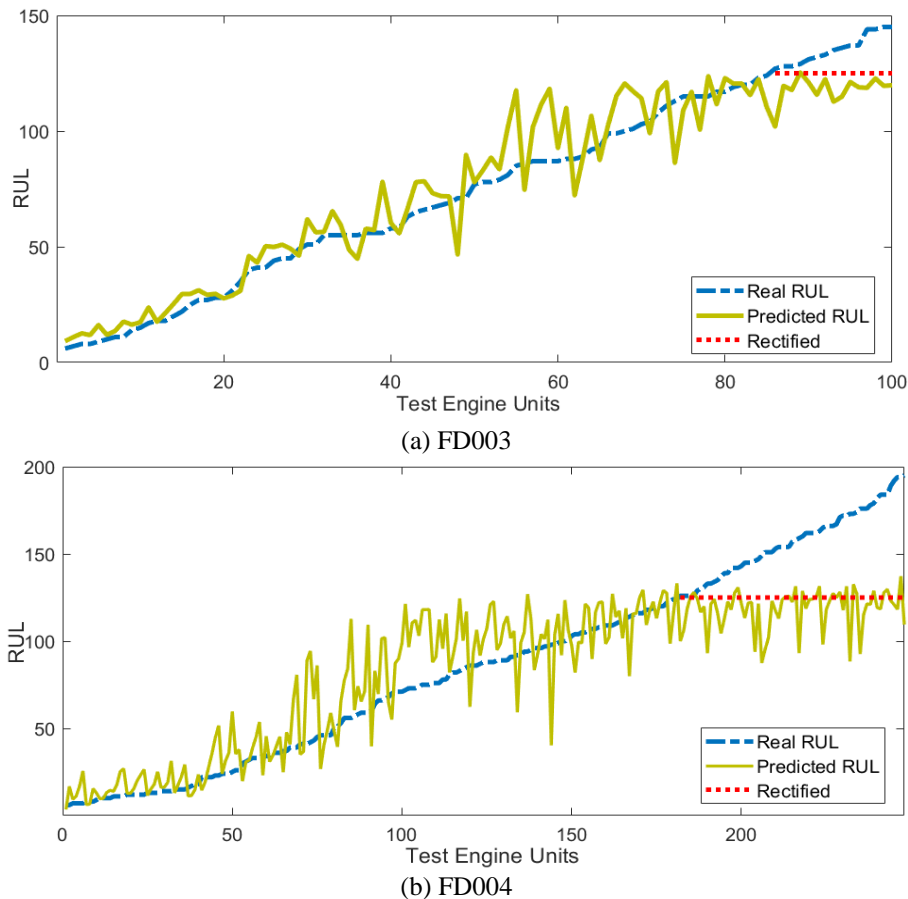


Figure 5 Comparison of prediction results for test engines sorted RUL in FD003 and FD004.

In order to investigate the proposed SC-BLSTM prediction performance for the life-time of a turbofan engine unit, the life-time prediction results of four different units in FD003 and FD004 sub-datasets are demonstrated in Figure 6. The numbers of randomly selected units from the test engines in the FD003 and FD004 sub-datasets are 39, 94, 40, and 68, respectively. It has been observed that the SC-BLSTM approach can mostly perform a remarkable RUL estimation over the degradation progression in all four examples. Furthermore, the developed SC-BLSTM approach tends to predict RUL values in the early stages to be adjacent to the rectified  $R_{early}$  as illustrated in all subplots. Afterwards, it can be said that RUL estimates tend to decline linearly until the end of their life cycle of test engine units. In the last periods of the degradation processes of 39, 94, and 40 engine units, the RUL prediction errors have a small value. This shows that the performance of the RUL prediction improves when the testing turbofan engines are close to failure. The prognostic efficiency in the last periods of the mechanical systems is important to make effective maintenance decisions, ensure system reliability and availability, and decrease the overall cost. The proposed SC-BLSTM model is able to achieve more robust and effective prognostic prediction in the last stages.

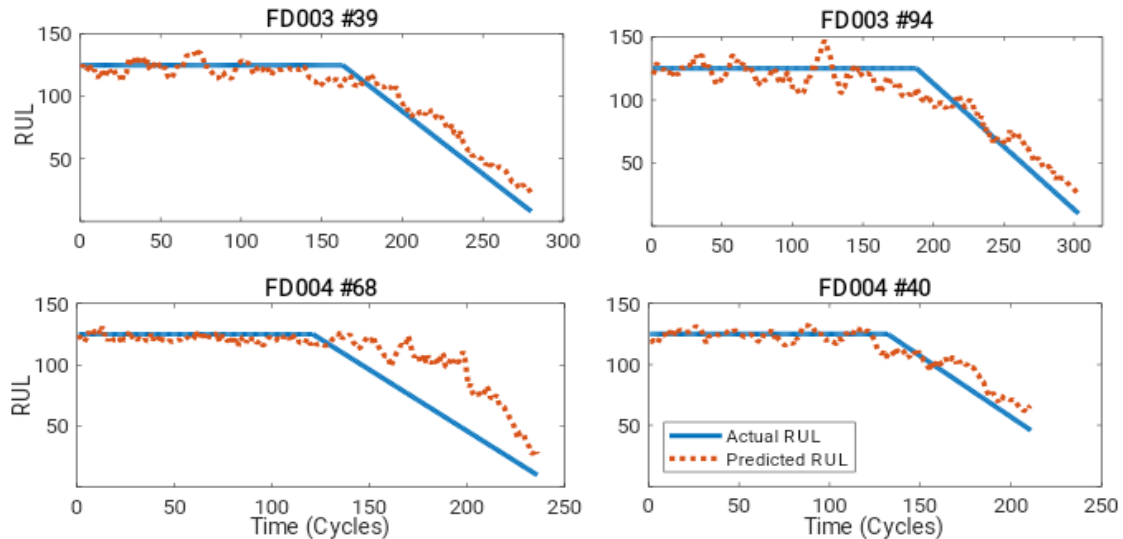


Figure 6 Examples of RUL estimations for the test engines units in FD003 and FD004.

In this research, the C-MAPSS dataset utilized to evaluate the performance of techniques developed for prognostic analysis is discussed, and the prognostic accuracy of the SC-BLSTM is compared with the existing data-driven approaches. According to RMSE metric, Table 3 reports a comparison of the SC-BLSTM with the other state-of-the-art approaches consisting of Extreme learning machine (ELM) [24], Multilayer Perceptron (MLP) [24], Support Vector Regression (SVR) [25], Relevance Vector Regression (RVR) [25], Deep belief network (DBN) [24], CNN [25], and MODBNE [24]. From Table 3, it can be said that the presented SC-LSTM approach has provided a promising prognostic accuracy compared with the results of the other benchmark approaches in terms of RMSE metric. Compared to the MODBNE approach with the second-best RMSE values on FD003 and FD004 subsets, the proposed SC-BLSTM method approximately achieves an enhancement of 1.28% and 1.40%, respectively.

Table 3 RUL prediction results of various models.

Related Approaches	RMSE	
	FD003	FD004
ELM [24]	18.90	38.43
MLP [24]	18.47	30.96
RVR [25]	22.37	34.34
SVR [25]	21.05	45.35
DBN [24]	14.71	29.88
CNN [25]	19.82	29.16
MODBNE [24]	12.51	28.66
Proposed SC-BLSTM	<b>12.35</b>	<b>28.26</b>

Table 4 illustrates the prediction performance of the SC-BLSTM method related to the scoring metric on the sub-dataset FD003 and FD004. It can be said that the prognostic prediction errors of the SC-BLSTM approach perform generally smaller results compared with the other state-of-the-art techniques in terms of the scoring metric. More specifically, the SC-BLSTM approach obtains the best prediction performance on the subset FD003. On the other hand, it performs the second-best on the subset FD004. In summary, the proposed SC-BLSTM approach has achieved a promising prognostic performance in mechanical systems taking into consideration all experimental consequences.

## 5. Conclusions

In this article, a deep learning-based prognostics approach, named SC-BLSTM, is addressed for the RUL prediction of mechanical systems. The presented SC-BLSTM employs the integration of the two separable CNN, a BLSTM layer, and fully-connected layers to achieve more accurate and reliable RUL



Table 4 A comparison of the Scoring metrics obtained from different approaches.

Related Approaches	Scoring	
	FD003	FD004
ELM [24]	573.78	10444.35
MLP [24]	479.85	121414.47
RVR [25]	1431.60	26509.00
SVR [25]	1598.30	371140.00
DBN [24]	442.43	7954.51
CNN [25]	1596.20	7886.40
MODBNE [24]	421.91	<b>6557.62</b>
Proposed SC-BLSTM	<b>222.24</b>	6849.18

prediction from multivariate degradation data. Firstly, two separable CNN modules are constructed to automatically learn the complex and nonlinear characteristics from the raw degradation data. Then, a BLSTM layer is leveraged to effectively capture temporal dependencies from the inputs. Finally, the extracted high-level features are fed into the fully-connected layers to achieve RUL prediction. In the SC-BLSTM approach, the dropout procedure is applied to overcome the overfitting in the training processing.

The effectiveness and superiority of the SC-BLSTM method are experimentally verified using the C-MAPSS turbofan engine dataset. The comparison between the proposed method and the existing related studies reveals that the SC-BLSTM method offers promising solutions in real-life PHM applications. More specifically, it is observed that RMSE metrics of the SC-BLSTM method on the FD003 subset have improved 6.04%, 37.69%, and 1.28%, respectively, compared with the other DBN, CNN, and MODBNE approaches. Besides, the reduction on the FD004 subset in terms of RMSE is 5.42%, 3.09%, 1.40%, respectively. It should be mentioned that although the SC-LSTM method performs the best RUL accuracy in the FD003 subset in terms of the scoring criteria, our method provides the second-best result in the FD004 subset. In future research, it will be focused on developing a hybrid structure based on transfer learning and deep learning methodologies to accomplish more effective RUL prediction of mechanical systems.

## References

- [1] H. Zhang, Q. Zhang, S. Shao, T. Niu, and X. Yang, "Attention-Based LSTM Network for Rotatory Machine Remaining Useful Life Prediction," *IEEE Access*, vol. 8, pp. 132188–132199, 2020, doi: 10.1109/ACCESS.2020.3010066.
- [2] B. Wang, Y. Lei, N. Li, and W. Wang, "Multiscale Convolutional Attention Network for Predicting Remaining Useful Life of Machinery," *IEEE Trans. Ind. Electron.*, vol. 68, no. 8, pp. 7496–7504, Aug. 2021, doi: 10.1109/TIE.2020.3003649.
- [3] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliab. Eng. Syst. Saf.*, vol. 172, pp. 1–11, Apr. 2018, doi: 10.1016/j.res.2017.11.021.
- [4] Z. Chen, M. Wu, R. Zhao, F. Guretno, R. Yan, and X. Li, "Machine Remaining Useful Life Prediction via an Attention-Based Deep Learning Approach," *IEEE Trans. Ind. Electron.*, vol. 68, no. 3, pp. 2521–2531, Mar. 2021, doi: 10.1109/TIE.2020.2972443.
- [5] M. Khazaei, A. Banakar, B. Ghobadian, M. A. Mirsalim, and S. Minaei, "Remaining useful life (RUL) prediction of internal combustion engine timing belt based on vibration signals and artificial neural network," *Neural Comput. Appl.*, Nov. 2020, doi: 10.1007/s00521-020-05520-3.
- [6] F.-K. Wang and T. Mamo, "Gradient boosted regression model for the degradation analysis of prismatic cells," *Comput. Ind. Eng.*, vol. 144, p. 106494, Jun. 2020, doi: 10.1016/j.cie.2020.106494.
- [7] Z. Xue, Y. Zhang, C. Cheng, and G. Ma, "Remaining useful life prediction of lithium-ion batteries with adaptive unscented kalman filter and optimized support vector regression,"

- Neurocomputing*, vol. 376, pp. 95–102, Feb. 2020, doi: 10.1016/j.neucom.2019.09.074.
- [8] R. Wang and N. Chen, “Defect pattern recognition on wafers using convolutional neural networks,” *Qual. Reliab. Eng. Int.*, vol. 36, no. 4, pp. 1245–1257, Jun. 2020, doi: 10.1002/qre.2627.
- [9] S. Xiang, Y. Qin, C. Zhu, Y. Wang, and H. Chen, “Long short-term memory neural network with weight amplification and its application into gear remaining useful life prediction,” *Eng. Appl. Artif. Intell.*, vol. 91, p. 103587, May 2020, doi: 10.1016/j.engappai.2020.103587.
- [10] L. Ren, X. Cheng, X. Wang, J. Cui, and L. Zhang, “Multi-scale Dense Gate Recurrent Unit Networks for bearing remaining useful life prediction,” *Futur. Gener. Comput. Syst.*, vol. 94, pp. 601–609, May 2019, doi: 10.1016/j.future.2018.12.009.
- [11] A. Al-Dulaimi, S. Zabihi, A. Asif, and A. Mohammadi, “A multimodal and hybrid deep neural network model for Remaining Useful Life estimation,” *Comput. Ind.*, vol. 108, pp. 186–196, Jun. 2019, doi: 10.1016/j.compind.2019.02.004.
- [12] A. Khorram, M. Khalooei, and M. Rezaghi, “End-to-end CNN + LSTM deep learning approach for bearing fault diagnosis,” *Appl. Intell.*, vol. 51, no. 2, pp. 736–751, Feb. 2021, doi: 10.1007/s10489-020-01859-1.
- [13] Y. Chen, G. Peng, Z. Zhu, and S. Li, “A novel deep learning method based on attention mechanism for bearing remaining useful life prediction,” *Appl. Soft Comput.*, vol. 86, p. 105919, Jan. 2020, doi: 10.1016/j.asoc.2019.105919.
- [14] B. Wang, Y. Lei, N. Li, and T. Yan, “Deep separable convolutional network for remaining useful life prediction of machinery,” *Mech. Syst. Signal Process.*, vol. 134, p. 106330, Dec. 2019, doi: 10.1016/j.ymsp.2019.106330.
- [15] X. Li, X. Jia, Y. Wang, S. Yang, H. Zhao, and J. Lee, “Industrial Remaining Useful Life Prediction by Partial Observation Using Deep Learning With Supervised Attention,” *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 5, pp. 2241–2251, Oct. 2020, doi: 10.1109/TMECH.2020.2992331.
- [16] H. Li, W. Zhao, Y. Zhang, and E. Zio, “Remaining useful life prediction using multi-scale deep convolutional neural network,” *Appl. Soft Comput.*, vol. 89, p. 106113, Apr. 2020, doi: 10.1016/j.asoc.2020.106113.
- [17] M. Xia, X. Zheng, M. Imran, and M. Shoaib, “Data-driven prognosis method using hybrid deep recurrent neural network,” *Appl. Soft Comput.*, vol. 93, p. 106351, Aug. 2020, doi: 10.1016/j.asoc.2020.106351.
- [18] Z. Yang, J. Zhang, Z. Zhao, Z. Zhai, and X. Chen, “Interpreting network knowledge with attention mechanism for bearing fault diagnosis,” *Appl. Soft Comput.*, vol. 97, p. 106829, Dec. 2020, doi: 10.1016/j.asoc.2020.106829.
- [19] T. Xia, Y. Song, Y. Zheng, E. Pan, and L. Xi, “An ensemble framework based on convolutional bi-directional LSTM with multiple time windows for remaining useful life estimation,” *Comput. Ind.*, vol. 115, p. 103182, Feb. 2020, doi: 10.1016/j.compind.2019.103182.
- [20] R. Shang, J. He, J. Wang, K. Xu, L. Jiao, and R. Stolkin, “Dense connection and depthwise separable convolution based CNN for polarimetric SAR image classification,” *Knowledge-Based Syst.*, vol. 194, p. 105542, Apr. 2020, doi: 10.1016/j.knosys.2020.105542.
- [21] G. Huang, Y. Zhang, and J. Ou, “Transfer remaining useful life estimation of bearing using depth-wise separable convolution recurrent network,” *Measurement*, vol. 176, p. 109090, May 2021, doi: 10.1016/j.measurement.2021.109090.
- [22] A. Elsheikh, S. Yacout, and M.-S. Ouali, “Bidirectional handshaking LSTM for remaining useful life prediction,” *Neurocomputing*, vol. 323, pp. 148–156, Jan. 2019, doi: 10.1016/j.neucom.2018.09.076.
- [23] 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*, Oct. 2008, pp. 1–9, doi: 10.1109/PHM.2008.4711414.
- [24] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, “Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 10, pp. 2306–2318, Oct. 2017, doi: 10.1109/TNNLS.2016.2582798.

- [25] G. Sateesh Babu, P. Zhao, and X.-L. Li, “Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, pp. 214–228.