

Article

A Double-Channel Hybrid Deep Neural Network Based on CNN and BiLSTM for Remaining Useful Life Prediction

Chengying Zhao ¹, Xianzhen Huang ^{1,2,*}, Yuxiong Li ¹  and Muhammad Yousaf Iqbal ³

¹ School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China; zhaochengying0223@163.com (C.Z.); 13354260226@163.com (Y.L.)

² Key Laboratory of Vibration and Control of Aero Propulsion Systems Ministry of Education of China, Northeastern University, Shenyang 110819, China

³ School of Mechanical Engineering, Taiyuan University of Technology, Taiyuan 030024, China; a13840311401@163.com

* Correspondence: xzhhuang@mail.neu.edu.cn; Tel.: +86-024-8367-3817

Received: 17 November 2020; Accepted: 9 December 2020; Published: 11 December 2020



Abstract: In recent years, prognostic and health management (PHM) has played an important role in industrial engineering. Efficient remaining useful life (RUL) prediction can ensure the development of maintenance strategies and reduce industrial losses. Recently, data-driven based deep learning RUL prediction methods have attracted more attention. The convolution neural network (CNN) is a kind of deep neural network widely used in RUL prediction. It shows great potential for application in RUL prediction. A CNN is used to extract the features of time-series data according to the spatial feature method. This way of processing features without considering the time dimension will affect the prediction accuracy of the model. On the contrary, the commonly used long short-term memory (LSTM) network considers the timing of the data. However, compared with CNN, it lacks spatial data extraction capabilities. This paper proposes a double-channel hybrid prediction model based on the CNN and a bidirectional LSTM network to avoid those drawbacks. The sliding time window is used for data preprocessing, and an improved piece-wise linear function is used for model validating. The prediction model is evaluated using the C-MAPSS dataset provided by NASA. The predicted results show the proposed prediction model to have a better prediction performance compared with other state-of-the-art models.

Keywords: RUL prediction; double-channel hybrid model; convolution neural network; bidirectional long short-term memory network

1. Introduction

In recent years, mechanical equipment is becoming more and more complicated, and mechanical reliability is strictly required, so the method of prognostic and health management (PHM) has attracted more and more attention than before [1–6]. The prediction of remaining useful life (RUL) is an important part of PHM. Efficient RUL prediction can ensure the development of maintenance strategies and reduce industrial losses. The methods of RUL prediction can be roughly divided into three categories: (1) a physical model approach, (2) data-driven approach, and (3) hybrid method approach.

The physical model approach is to establish a mathematical degradation physical failure model based on the failure mechanism or performance degradation process of the system [7,8]. At present, this method is mainly used in systems with explicit models and requires researchers to have professional knowledge. Moreover, it is challenging to build dynamic mathematical and physical models accurately for multiple failure modes in the system.

The data-driven approach directly uses sensor monitoring data. First, the data are analyzed using signal processing and other technologies. Then, the features reflecting the system degradation and fault are extracted. Subsequently, the system degradation model is established. Finally, the RUL prediction is completed [9–11]. The data-driven method is mainly based on experimental data, and then the method generally summarizes the degradation laws to predict RUL. Data-driven methods can be divided into mathematical statistics-based methods and machine learning-based methods. There are many typical mathematical statistical models for RUL prediction, including the autoregressive moving average (ARMA) model [12], Wiener process model [13,14], Markov model [15], etc.

Compared with statistical models, machine learning-based methods have been widely used in recent years. Deep learning is an emerging method in the field of machine learning and has achieved fruitful results in the areas of image processing, speech recognition, and machine translation [16–18]. Deep learning can be regarded as a neural network that contains multiple hidden layers. It is an idea that simulates the human brain to learn layer by layer and cause it to realize to learn abstract features from the raw sensor data automatically. It has been widely used in RUL prediction. The neural network can directly establish the relationship between the degradation data of the system and the RUL. Gebraeel et al. [19] proposed an artificial neural network (ANN). Vibration information of bearings from the beginning to the time of failure was collected and was input into the neural network for training. The trained model was used to estimate the failure time of the bearing. Aydemir and Acar [20] proposed an anomaly-triggered RUL estimation method. The data of sensors are monitored, and when statistically significant changes are detected, these are taken as the starting point of degradation to trigger the data-driven RUL prediction model. Fu et al. [21] proposed an adaptive domain autoencoder and long short-term memory (DASE) model to predict the RUL of mechanical parts, which improved the adaptability of the model. Li et al. [22] proposed a neural network with the supervised attention mechanism to screen data with more significant degradation feature information for RUL prediction. Loannis et al. [23] proposed a method to update the RUL prediction of the rolling milling process in real time. The nonparametric statistical process is used to update the prediction. In addition to the above neural network prediction models, models based on self-organizing mapping (SOM) and the backpropagation (BP) neural network model [24], support vector machine (SVM) model [25,26], convolutional neural network (CNN) [27], recurrent neural network (RNN) [28,29], and other models are also applied to RUL prediction. At present, most models cannot fully extract the features of each dimension with multidimensional big data samples. This paper adopts a hybrid model to solve this gap.

The hybrid model method is to integrate the physical model and the data-driven approach. The advantages of the two methods are integrated to improve the prediction performance of the model. Maio et al. [30] combined relevance and exponential regression to realize the RUL prediction. Zheng et al. [31] proposed a hybrid model that combines Kalman filtering and relevance vector machine to predict the degradation trend of batteries.

As RUL prediction is carried out, the neural network does not need to extract the data feature manually. The neural network has strong feature extraction capability and reduces the problem of overreliance on expert experience and signal processing technology. It is suitable for the requirements of diversity, nonlinearity, and high-dimensional data analysis under the background of big data. When extracting features from time-series data, the CNN is used to extract the features of time-series data according to the spatial feature method. The paper in [32] directly used a deep CNN model to predict the RUL. This way of processing features without considering the time dimension will affect the prediction accuracy of the model. On the contrary, the commonly used LSTM network considers the timing of the data and the interdependence between time-series data. A bidirectional long short-term memory (BiLSTM) network is a combination of two LSTMs, i.e., forward and backward. Based on LSTM, BiLSTM can extract the feature of forward and backward simultaneously. Wang [33] proposed a deep BiLSTM network to predict the RUL. However, compared with CNN, it lacks spatial data extraction capabilities. To solve those gaps and extract deep features from the raw sensor data,

this paper proposes a double-channel hybrid prediction model based on CNN and a BiLSTM network. The parallel channels are independent of each other and do not affect each other in the training process, thus improving the prediction accuracy of the model. Based on the C-MAPSS dataset, the model proposed in this paper is compared with the RUL prediction model proposed in recent years, which shows that the double-channel hybrid model has better prediction capability.

The rest of the paper is organized as follows: In Section 2, it introduces the relevant literature review. In Section 3, the CNN and the LSTM network are introduced, and a double-channel hybrid model is proposed. Section 4 presents the data preprocessing method and model evaluation method. Section 5 shows the experimental results, and Section 6 gives the conclusions.

2. Literature Review

CNNs have been widely used in RUL prediction due to their excellent property of extracting spatial features from the raw data. Babu et al. [34] proposed a novel regression method based on a deep CNN to estimate RUL. The convolution and pooling filters are applied to multichannel sensor data along the time dimension. The features learned from the raw sensor signal are automatically combined systematically. Through the deep architecture, the model determines the high-level abstract representation of the raw sensor data. Moreover, feature learning and RUL estimation are mutually enhanced through supervised feedback. Ren et al. [35] proposed a method to predict the RUL of a bearing based on a deep CNN. For the first time, a smooth method was innovatively applied to deal with the problem of discontinuous prediction results. Experimental results showed that this method improved the accuracy of the RUL prediction of bearings. A CNN fails to consider the correlation between the time-series of the raw sensor data. For the far-away time data, it is easy to lose its characteristic information, resulting in poor long-term prediction ability and insufficient prediction accuracy. In this paper, bidirectional long short-term memory (BiLSTM) and the CNN are combined to solve this problem effectively.

LSTM networks can capture long-term dependencies between features. They can improve the accuracy of RUL prediction. Zhang et al. [36] used the LSTM neural network to predict the RUL of lithium-ion batteries. The elastic mean square backpropagation method is used to optimize the model. Wu et al. [37] proposed using a vanilla LSTM neural network to obtain better prediction accuracy under complex operation, working conditions, model degradation, and intense noise. Furthermore, a dynamic difference technique was proposed to extract interframe information to enhance the cognitive ability of the model degradation process.

3. Methodology

3.1. Convolutional Neural Network

The CNN was firstly proposed by Lecun et al. [38] for image processing. The CNN is a class of feed-forward neural network that contains a convolutional operation and deep structure. Compared with the early backpropagation (BP) neural network, the essential characteristics of CNN are local perception and parameter sharing. It can extract the spatial features from the raw input data and realizing a high-dimensional feature representation of raw data. Figure 1 shows the convolution and pooling operation. The convolutional layer uses convolutional kernels to convolve local regions of the input data to generate corresponding features. The parameter sharing of a CNN decreases the parameters in the calculation process of the neural network. This effectively avoids the phenomenon of having too many parameters, which leads to overfitting the neural network. In this way, the overfitting of the neural network is resolved. Moreover, the system memory occupied in the training process of the neural network significantly decreases. Formula (1) shows the convolution operation of the i -th convolution kernel.

$$y_i = \varphi(W_i X_{i-1} + b_i) \quad (1)$$

where X_{i-1} represents the input of the i -th layer CNN; W_i represents the i -th filter kernel; b_i and φ represent the bias term and the nonlinear activation function, respectively.

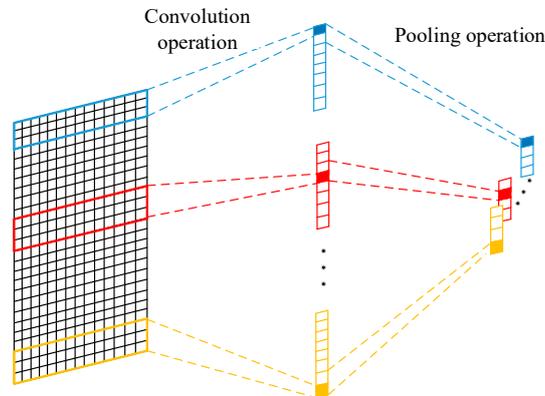


Figure 1. Convolution and pooling operation.

The final output of one convolution layer is to connect the results of the convolution operation of n convolution kernels together, which can be represented as:

$$Z = [y_1, y_2, \dots, y_n] \quad (2)$$

In this paper, the maximum pooling layer is used to downsample the output of the CNN and extract critical local features. Moreover, the parameters and amount of calculation of the model decreased. Consequently, the operation speed is improved, and the overfitting problem is avoided.

The 2-dimensional convolution network was used in this paper. It is a 1-dimensional operation on the 2-dimensional input data, and it was used to extract the spatial features of the time-series from the input sample.

3.2. Bidirectional Long Short-Term Memory Neural Network

In this paper, a BiLSTM was used to extract the long-term dependency characteristics of input sample data. The BiLSTM was composed of two independent LSTM neural networks, and it has two directions of forward and backward propagation. The LSTM network proposed by Hochreiter and Schmidhuber [39] is a unique recurrent neural network (RNN) neural network. It has a particular network structure consisting of an input gate structure and an output gate structure. LSTM affects the state of an RNN at each time through the gate structure. The gates only limit directions of the information flow. The activation function of the gate's structure is sigmoid. Figure 2 shows the cell of LSTM.

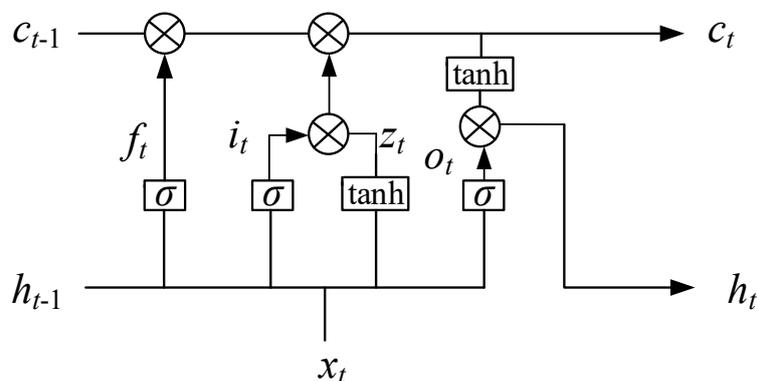


Figure 2. Long short-term memory cell.

The core of the LSTM structure is the forget gate and the input gate, which can save long-term memory effectively. The forget gate makes the neural network abandon the previous useless information by the current input x_t and the output h_t of the last moment. The network overcomes gradient disappearance or gradient explosion during training by importing a gate structure and memory cells. The input gate determines the information to be added to the state c_{t-1} to generate a new state c_t according to the parameters x_t and h_{t-1} . Through the forget gate and input gate structure, the LSTM can effectively determine which information is forgotten and which information is retained. Therefore, LSTM extracts useful information; the output gate determines the output of the h_t according to the output h_{t-1} at the moment of the latest state c_t and the current input x_t . The LSTM specific formula is defined as follows:

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (4)$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \quad (5)$$

$$z_t = \tan h(W_{zx}x_t + W_{zh}h_{t-1} + b_z) \quad (6)$$

$$c_t = f \cdot c_{t-1} + i \cdot z \quad (7)$$

$$h_t = o \cdot \tan h(c_t) \quad (8)$$

BiLSTM has two directions in propagation, and both directions are simultaneously transmitted to the output unit. This can capture past and future information, as shown in Figure 3. At each time t , the input is provided to both the forward and backward LSTM networks. The result of the BiLSTM output can be represented as:

$$h_t = \vec{h}_t \odot \overleftarrow{h}_t \quad (9)$$

where \vec{h}_t and \overleftarrow{h}_t represent the forward and backward results of BiLSTM, respectively.

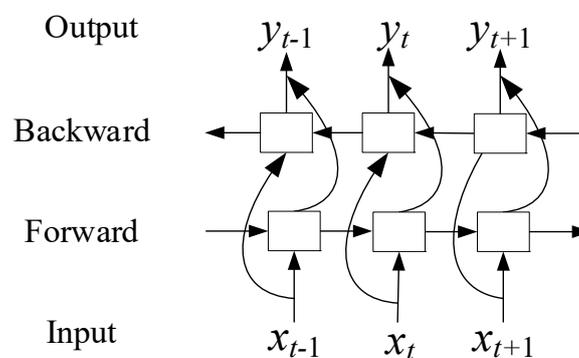


Figure 3. Bidirectional long short-term memory (BiLSTM) neural network.

3.3. The CNN-BiLSTM Neural Network Structure

In order to fully extract the spatial and temporal features of the input data for RUL prediction, this paper proposes a double-channel hybrid neural network model based on a CNN and a BiLSTM. The structure of the model is shown in Figure 4. The CNN and BiLSTM respectively extract features from the raw data, then concatenate them together and map them to a fully connected layer. The combination of CNN and BiLSTM increases the amount of data, which can then improve the accuracy and stability of prediction.

The input data samples adopt a 2-dimensional structure of $W_l \times n$. Here, W_l is the length of the time-series, and n is the number of features. In the process of data processing, a sliding time window was used to extract data samples, and the sliding step size of the window is 1. For different datasets of the C-MAPSS, different time-window lengths were used.

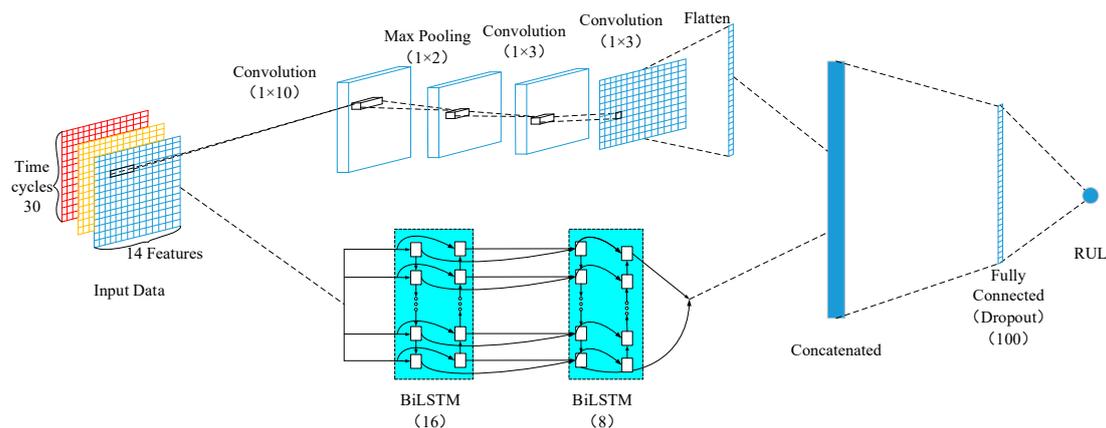


Figure 4. Convolutional neural network (CNN)-BiLSTM hybrid neural network.

The first channel of the hybrid model is composed of three layers of CNN, a layer of maximum pooling, and a layer of flatten. The CNN structure of the first channel was used to extract the spatial features from the sensor's raw data; however, the depth of time features extraction of the raw high-dimensional data is far from enough. Therefore, the BiLSTM structure was adopted to complete the extraction of the data time-series features and extract the long-term dependencies between data features. Meanwhile, the BiLSTM structure can avoid gradient disappearance and gradient explosion during training the model. The first convolutional layer used 10 filters of size (1×10) , and the maximum pooling filter of size (1×2) is adopted. The second layer of convolution used 10 filters with filter size of (1×3) . The last layer of the convolutional used a filter of size (1×3) . If the data dimension is too small, the first channel CNN will fail. In order to prevent failure, the stacked CNN adopted the zero-filling method. The convolutional data were finally input to the flatten layer. The purpose of the flatten layer is to convert the features extracted from the first channel convolution neural network into a 1-dimensional structure, which is convenient for concatenating with the features extracted from the BiLSTM path. The second channel stacked two layers of BiLSTM. Moreover, the cells of the two layers were set to 16 and 8, respectively. The input data were converted to 1-dimensional data before they were inputted into the BiLSTM path.

The feature data matrices extracted from the first channel and the second channel were concatenated, inputting to a fully connected layer containing 100 neurons. To prevent overfitting, the dropout technique was used. The dropout technology stops the hidden layer neurons with self-defined probability numbers from working in the forward propagation of the training process. In doing so, it improves the generalization ability of the model and enhances the robustness of the model. In this paper, the dropout rate is 0.5. Hence, a part of neurons in the network will be hidden with a self-defined probability of 0.5 in each iteration during the training process. A subnetwork will be obtained by randomly sampling the network parameters, and only the subnetwork will be updated in this iteration. All the above layers use "ReLU" as the activation function. Finally, a regression layer containing a neuron was used to predict the target RUL.

The BP algorithm was used to train the model, and the model weights and biases were updated continuously. The Adams optimizer algorithm was used to monitor the training error "MSE" in order to update the model parameters continually. The learning rate was set to 0.001. The mini-batch size was 512, and the maximum training epoch was set as 200.

4. Experiments

4.1. C-MAPSS Dataset

In this part, the C-MAPSS dataset was used to verify the predictive performance of the proposed double-channel hybrid deep neural network model. The C-MAPSS dataset has four subsets (FD001, FD002, FD003, and FD004), recording 21 different types of sensor data respectively. The monitoring

data of all sensors reflect the well-being status of the engine. These four subsets contain different operating conditions and fault modes. Each subset possesses its training dataset and test dataset. The train dataset records the data of all sensors in the whole life cycle from the start of the run to the time of the failure. The test dataset records all sensor data from the start of engine operation to a random cycle point in time. In datasets, the first column shows the serial number of the engine, with the second column representing the number of cycles of the engine, the third to fifth column showing different operating settings, and the sixth to the twenty-sixth column displaying the data recorded by 21 various sensors. The details of the four datasets are shown in Table 1.

Table 1. Introduction of the C-MAPSS dataset.

Dataset	C-MAPSS			
	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
Training samples	17731	48819	2Tab820	57522
Testing samples	100	259	100	248

4.2. Data Preprocessing

The 21 sensors in each subset record different data, and the data range between different sensors varies significantly. The unprocessed data affected the performance of predicting the engine RUL, so it is essential to standardize the data. Data normalization is an essential step in data processing. Additionally, the output value of Sensors 1, 5, 6, 10, 16, 18, and 19 are constant, without providing useful information for RUL prediction. Therefore, the other 14 sensors' data were selected for normalization processing for RUL prediction.

There are two ways to normalized the data, the Z-score normalization as given in Formula (10), and the min-max standardization as given in Formula (11). Min-max normalization is employed to process the data in this paper.

$$y_j^i = (x_j^i - \bar{x}^i) / \sigma^i \quad (10)$$

$$y_j^i = (x_j^i - x_{min}^i) / (x_{max}^i - x_{min}^i) \quad (11)$$

where x_j^i is the j -th output of the i -th sensor; \bar{x}^i is the mean value of all outputs of the i -th sensor; x_{min}^i is the minimum value of all output of the i -th sensor; and x_{max}^i is the maximum value of all output of the i -th sensor.

In the model of RUL prediction, the cycle number of multivariate time-series included in the input sample data has a direct impact on the prediction performance. The larger the cycle number of samples, the more data information is monitored by sensors, and the better the model's prediction effect. In this paper, a sliding time window was used to extract samples of standardized data. According to the minimum cycle number of all engines in the four datasets, the time-window lengths of the four datasets are 30, 21, 36, and 18 cycles, respectively, and the sliding step size is one cycle. Figure 5 shows that using a sliding time window to extract input data samples from a multivariate time-series normalized the data. The RUL of the last cycle of the time window was used as the label for the sample. Meanwhile, n is the number of engine cycles represented, l is the length of the time window, then $n - l + 1$ samples can be extracted from the multivariate time-series dataset of the engine.

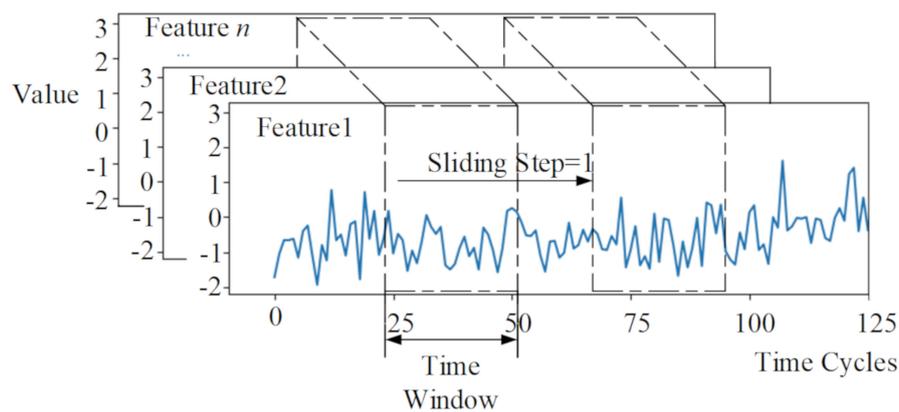


Figure 5. Sliding time window to extract data samples.

4.3. Evaluation Methods

In this paper, a scoring function and the root mean square error (RMSE) are used to evaluate the performance of the proposed model of RUL estimation. The detailed introduction is shown as follows.

The scoring function is as follows:

$$s_i = \begin{cases} e^{-\frac{h_i}{13}} - 1, & h_i < 0 \\ e^{\frac{h_i}{10}} - 1, & h_i > 0 \end{cases}, \text{ score} = \sum_{i=1}^N s_i \quad (12)$$

where N is the number of test samples; h_i means the error between the $RUL_{\text{prediction}}$ and RUL_{actual} ; and $h_i = RUL_{\text{prediction}} - RUL_{\text{actual}}$. The scoring function was proposed at the International Conference on Prognostics and Health Management Data Challenge (2008).

The root mean square error (RMSE) is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N h_i^2} \quad (13)$$

Figure 6 shows two indicators of the evaluation model, i.e., the RMSE and scoring function, giving different penalties for early and late prediction. Value of h ($h = 0$) means that the $RUL_{\text{prediction}}$ is equal to RUL_{actual} , and the two indicators receive a zero. When the error increases, the values of the RMSE and the score also increase correspondingly. RMSE can directly reflect the overall RUL prediction error and is also one of the commonly used indexes. For the scoring function, if the predicted value is less than the actual value, it is predicted early, and the penalty coefficient is small. In contrast, late prediction induced a large penalty coefficient. The advance prediction of the RUL can make maintenance plans and reduce losses in industries. RMSE has the same penalty for early and late prediction.

4.4. RUL Target Function

The RUL in the C-MAPSS dataset was considered to decrease linearly with time when predicting the RUL. In practice, the performance of the machine is stable at the beginning of work, and the degradation of the machine can be ignored. However, the RUL of mechanical parts decreases linearly with time due to various working conditions such as wear and vibration. A piece-wise linear RUL method is presented in [29]. Therefore, this study also used a piece-wise linear function to represent the RUL, as shown in Figure 7, and the maximum RUL was set to 125.

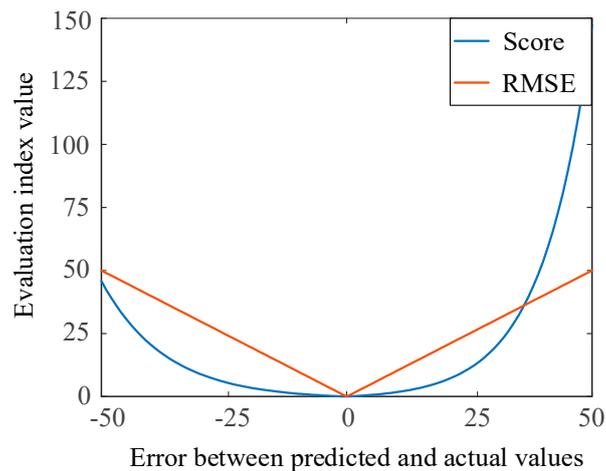


Figure 6. Comparison between two evaluation indicators.

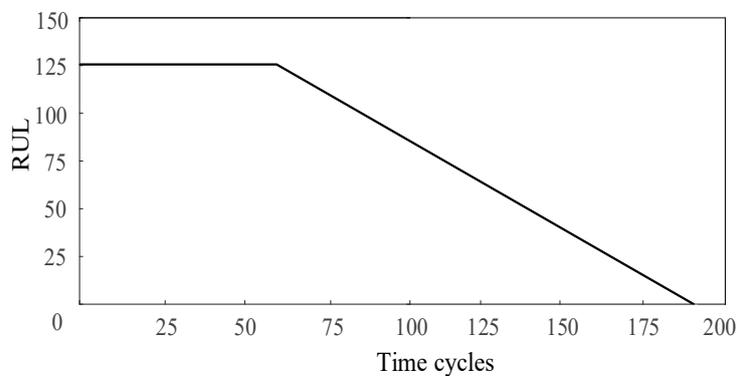


Figure 7. The piece-wise linear remaining useful life (RUL) target function.

5. Results and Analysis

In this section, the double-channel hybrid prediction model based on the CNN and BiLSTM is evaluated by the C-MAPSS datasets. For the four datasets of the C-MAPSS, a different sliding time-window length is used to process the data samples, and then we input the model for training. The model is evaluated by the RMSE and scoring function. The last time-window data of all engines in the four test datasets are selected to predict the RUL. The prediction results are shown in Figure 8. The solid black line indicates the actual RUL of the engine, and the solid red line represents the predicted RUL. The predicted values of the proposed model are very close to the actual values. Figure 9 shows the error frequency distribution between the predicted and actual RUL of the engine for the C-MAPSS datasets.

In previous work [32,34,40,41], different deep learning models were established, and predictable the performance through the C-MAPSS dataset was analyzed. Tables 2 and 3 show the RMSE and score comparison between the proposed hybrid network model and other reported models. The result indicated that the proposed model shows a smaller RMSE and score in the four datasets. The improved ratio is the degree to which the RMSE/score is improved compared with the other papers' models. Compared with other models, the apparent improvement of the score index shows that the hybrid network model can better realize RUL prediction ahead of time, thus realizing predictive maintenance of industrial production. The significant improvement of the RMSE index indicates that the model has a more accurate prediction ability.

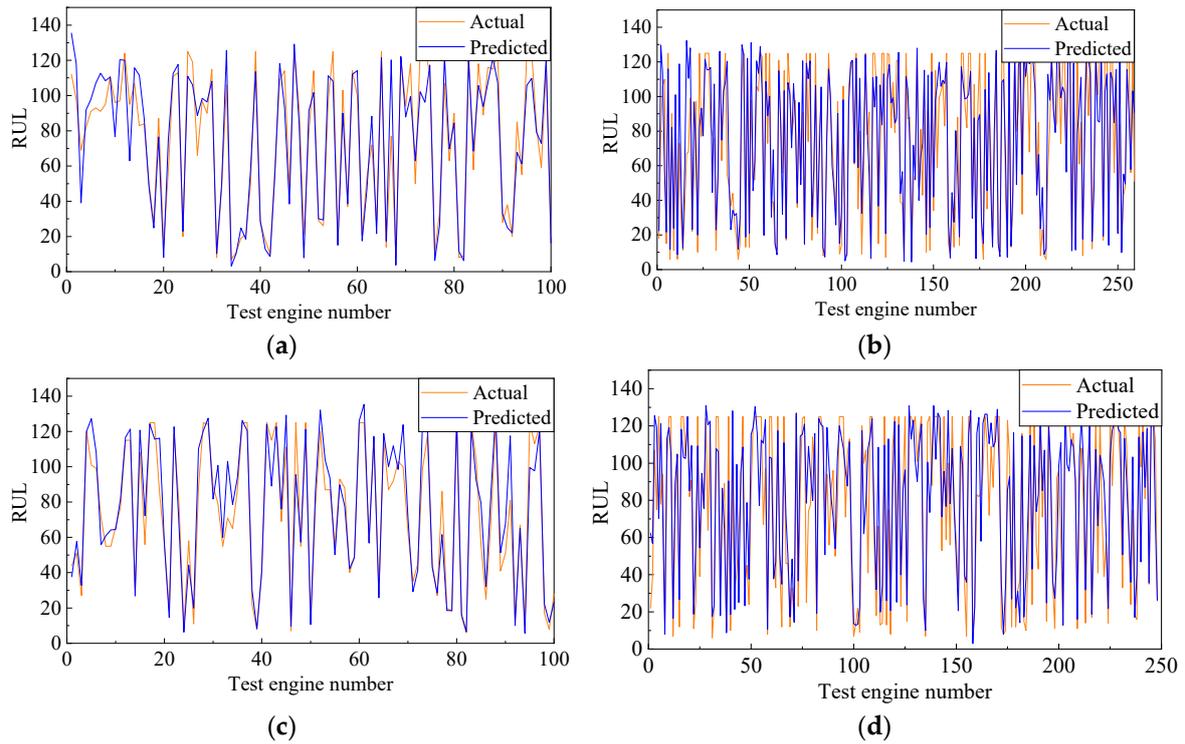


Figure 8. Predicted and actual values of (a) FD001 dataset, (b) FD002 dataset, (c) FD003 dataset, and (d) FD004 dataset.

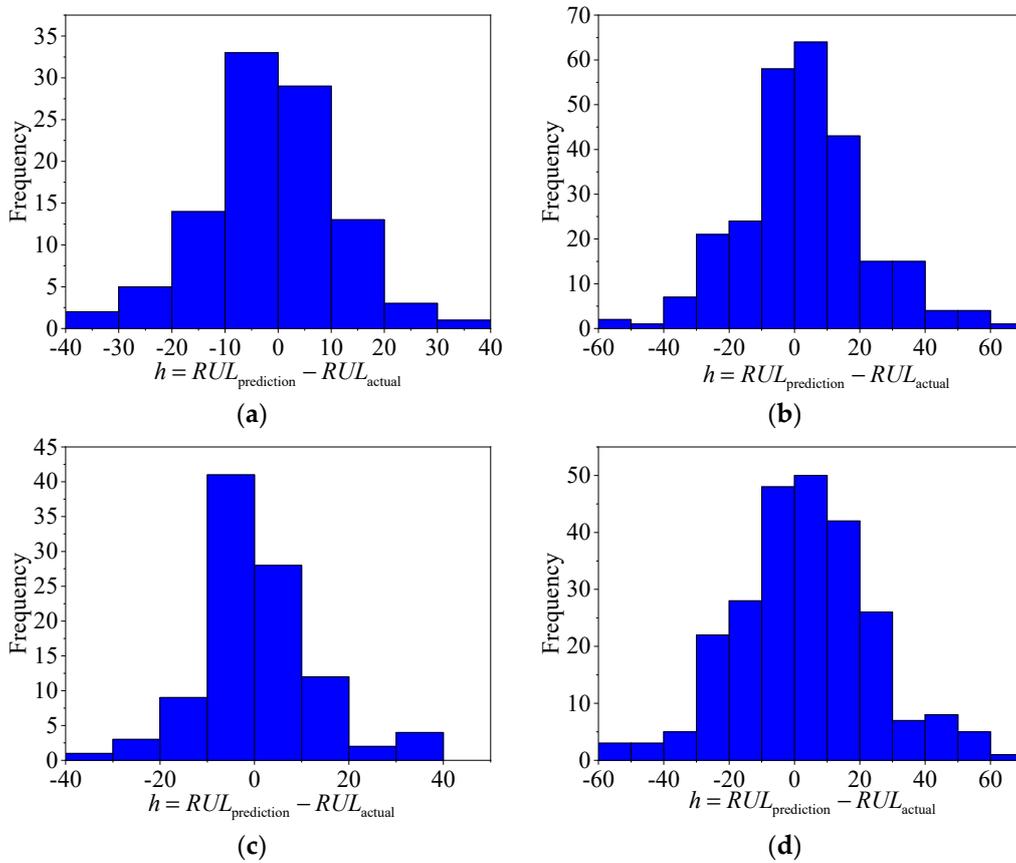


Figure 9. The frequency distribution of the prediction error of (a) FD001 dataset, (b) FD002 dataset, (c) FD003 dataset, and (d) FD004 dataset.

Table 2. RMSE comparison based on C-MAPSS dataset.

Method	C-MAPSS			
	FD001	FD002	FD003	FD004
MLP [34]	37.56	80.03	37.39	77.37
SVR [34]	20.96	42.00	21.05	45.35
RVR [34]	23.80	31.30	22.34	34.34
CNN [34]	18.45	30.29	19.82	29.16
MODBNE [40]	15.04	25.05	12.51	58.66
LSTM [41]	16.14	24.49	16.18	28.17
BiLSTM [33]	13.65	23.18	13.74	24.86
DCNN [32]	12.61	22.36	12.64	23.31
Proposed method	12.58	19.34	12.18	20.03
Improved ratio	2.38%	13.51%	3.64%	14.07%

Table 3. Score comparison based on C-MAPSS dataset.

Method	C-MAPSS			
	FD001	FD002	FD003	FD004
MLP [34]	1.80×10^4	7.80×10^6	1.74×10^4	5.62×10^6
SVR [34]	1.38×10^3	5.90×10^5	1.60×10^3	3.71×10^5
RVR [34]	1.50×10^3	1.74×10^4	1.43×10^3	2.65×10^6
CNN [34]	1.29×10^3	1.36×10^4	1.6×10^3	7.89×10^3
MODBNE [40]	3.34×10^2	5.59×10^3	4.21×10^2	6.56×10^3
LSTM [41]	3.38×10^2	4.45×10^3	8.52×10^2	5.55×10^3
BiLSTM [33]	2.95×10^2	4.13×10^3	3.17×10^2	5.43×10^3
DCNN [32]	2.74×10^2	1.04×10^4	2.84×10^2	1.25×10^4
Proposed method	2.31×10^2	2.65×10^3	2.57×10^2	3.40×10^3
Improved ratio	15.69%	35.84%	9.51%	38.74%

The prediction capability of the model proposed in this paper is significantly improved. Compared with the traditional MLP, SVR, and RVR models, the proposed model does not need artificial feature extraction. It can realize independent feature extraction in the time dimension and space dimension, with better applicability, smaller error, and better prediction performance. Compared with CNN, LSTM, and BiLSTM networks, the double-channel neural network extracts features independently of each other and then combines the features extracted by the two channels, making full use of the advantages of the CNN and BiLSTM networks, increasing the amount of data and improving the prediction accuracy.

The prediction results of the FD001 and FD003 datasets with smaller errors are dramatically better than the FD002 and FD004 datasets. Because the operating conditions of the FD002 and FD004 datasets are more complicated than other datasets, the operation mode and failure mode affected the working process of the system and increased the difficulty of prediction. The FD002 and FD004 datasets contain six different operating conditions. In contrast, FD001 and FD003 have only one operating condition. The scores predicted by the FD002 and FD004 datasets are large, caused by a large number of engines included in these two datasets.

Figure 10 shows the comparison of the predicted and actual RUL values for the four engines selected from the test datasets of FD001, FD002, FD003, and FD004 with the double-channel hybrid CNN-BiLSTM model, single-channel CNN model, and single-channel BiLSTM model. The 135th engine data in the FD004 data set has 435 cycles. In order to facilitate observation, a set of data is drawn from every four points in this paper. It can be seen from the figure that compared with the CNN and BiLSTM model, the prediction effect of most engines using the double-channel CNN-BiLSTM model in this paper is better than other methods. Moreover, the predicted RUL fluctuates around a constant value in the early periods. With the increase in the number of engine cycles, the actual and

predicted RUL values decrease linearly in line with the reality of the engine degradation phenomenon. However, the expected value is close to the real value in the last few cycles. With the increase of engine cycles, the model's ability to capture engine degradation is significantly improved. In the latter stage, the model's predictive ability becomes even better. The model proposed in this paper shows good predictive ability.

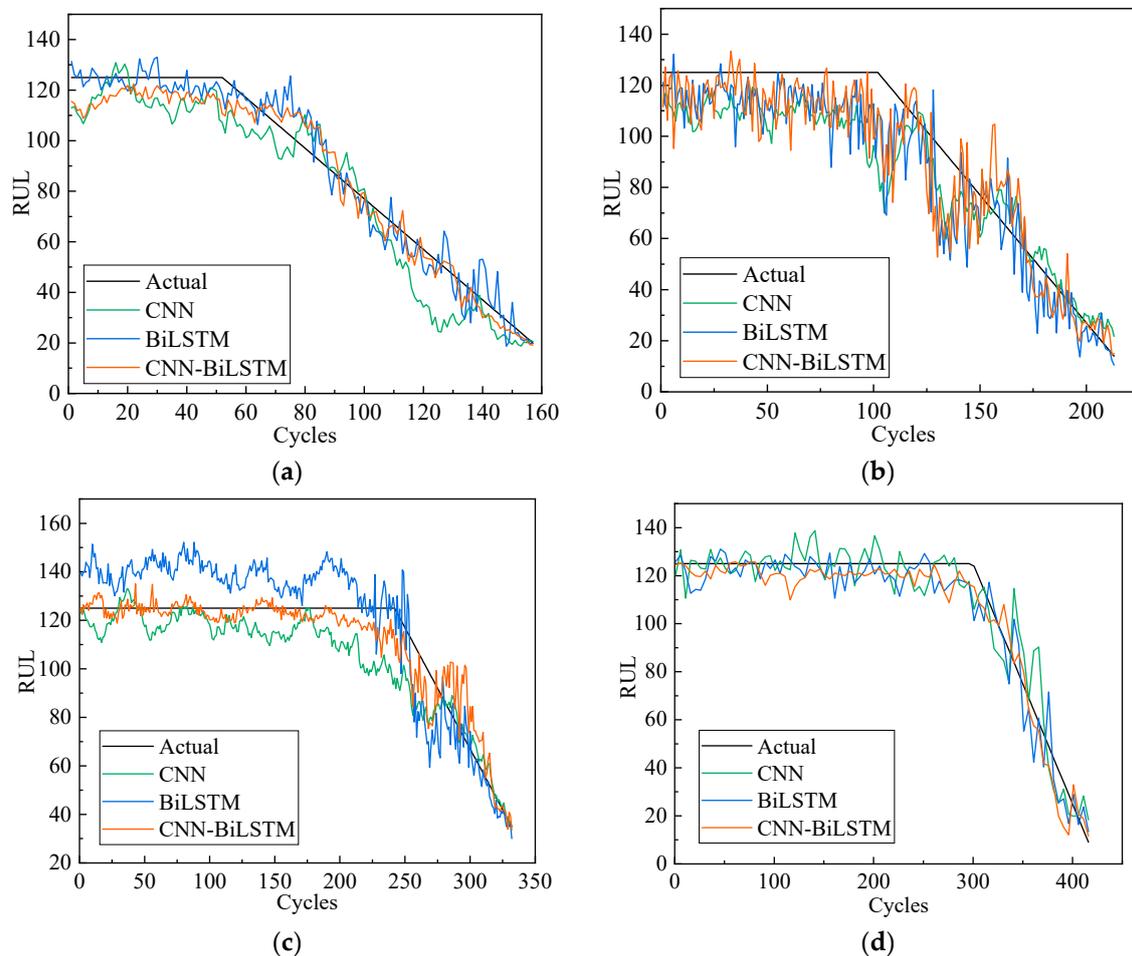


Figure 10. Different models predicting the whole life cycle of 4 engines with their predicted RUL and the actual RUL: (a) Unit #24 in FD001, (b) Unit #64 in FD002, (c) Unit #71 in FD003, (d) Unit #135 in FD004.

Figure 11 shows the effect of the sliding time-window length on the model's predictive performance when extracting samples in the FD001 dataset. In this paper, the length of 7 time-windows with different sizes was selected from the FD001 dataset, i.e., 2, 5, 10, 15, 20, 25, and 30. Other parameters of the model were set in the same way, and the box plot drawn by the RMSE and score was obtained after ten runs. Figure 10 shows the better performance of the model after training the data samples extracted with a large time-window length. In a certain range, the larger the window length, the more information contained in the extracted samples, which is more conducive to the extraction of data features by the model. The minimum cycle number of the engine in the dataset in the FD001 dataset is 31 cycles, and the time-window length is 30 cycles.

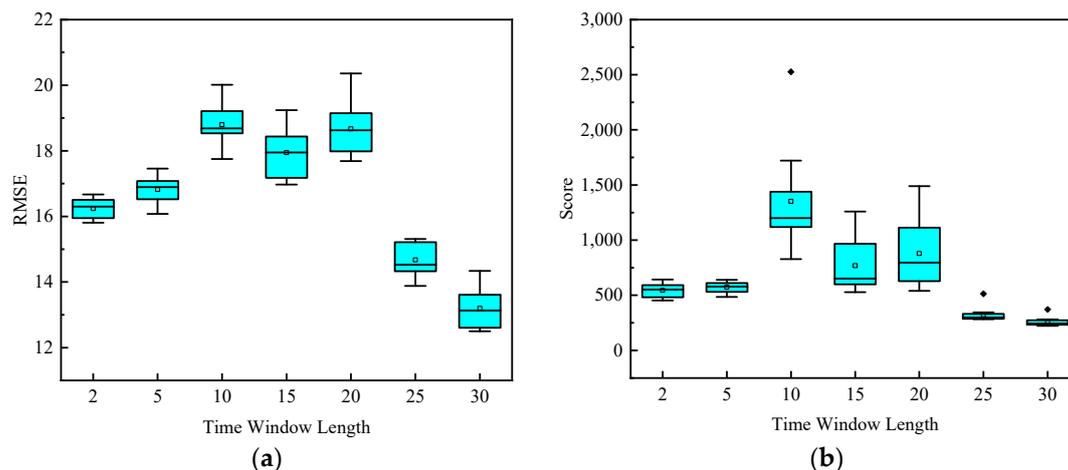


Figure 11. Box diagram of the evaluation index for the influence of the time-window length on model prediction performance: (a) RMSE (b) Score

Table 4 shows the performance comparison between the only CNN channel model, the only BiLSTM channel model, and the hybrid model on the FD001 dataset. The only CNN channel model is mainly composed of three convolution layers, a maxpooling layer, a fully-connection layer, and a regression layer. The only BiLSTM channel model is mainly composed of two BiLSTM network layers, a fully-connection layer and a regression layer. All parameter settings are consistent with those of the hybrid model proposed in this paper. It was observed that the performance of the proposed hybrid model is significantly better than that of the other two models. The hybrid model combines the advantages of the CNN and BiLSTM to extract the temporal and spatial features of the raw sensor data; the fully-connection layer increases the output of the nonlinearity. Finally, the regression layer was used to implement the RUL prediction.

Table 4. Performance comparison between single channel model and hybrid model.

Method	RMSE	Score
Only CNN channel	15.32	318
Only BiLSTM channel	18.86	793
Hybrid model	12.51	224

6. Conclusions

This paper proposed a double-channel hybrid prediction model based on a CNN and BiLSTM network. The sliding time-window was used for data preprocessing, and an improved piece-wise linear function was used for model validating. The prediction model was evaluated using the C-MAPSS dataset provided by NASA. This paper can be concluded as follows:

1. Compared with other state-of-the-art models, the hybrid model gets a better RMSE and score. The results show that the proposed double-channel hybrid model has a better predictive capability. The data of four engines in the dataset were extracted for prediction. The model can predict the degradation trend of engine performance.
2. Within a specific time-window length range, increasing the length of the time window will improve the RUL prediction accuracy of the double-channel hybrid model.
3. Compared with the prediction performance of the CNN channel model and the BiLSTM channel model proposed in this paper, the hybrid model exhibits better prediction capability.

Author Contributions: Funding acquisition, X.H.; investigation, C.Z.; methodology, C.Z.; project administration, C.Z.; resources, C.Z.; software, C.Z.; validation, Y.L.; writing—original draft, C.Z.; writing—review and editing, C.Z. and M.Y.I. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (51975110), Liaoning Revitalization Talents Program (XLYC1907171), and Fundamental Research Funds for the Central Universities (N2003005).

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

References

- Zhang, C.J.; Yao, X.F.; Zhang, J.M.; Jin, H. Tool condition monitoring and remaining useful life prognostic based on a wireless sensor in dry milling operations. *Sensors* **2016**, *16*, 795. [[CrossRef](#)] [[PubMed](#)]
- Lasheras, F.; Nieto, P.; de Cos Juez, F.; Bayón, R.; Suárez, V. A hybrid PCA-CART-MARS-based prognostic approach of the remaining useful life for aircraft engines. *Sensors* **2015**, *15*, 7062–7083. [[CrossRef](#)] [[PubMed](#)]
- Liu, Z.; Mei, W.J.; Zeng, X.P.; Yang, C.L.; Zhou, X.Y. Remaining useful life estimation of Insulated gate bipolar transistors (IGBTs) based on a novel volterra k-nearest neighbor optimally pruned extreme learning machine (VKOPP) model using degradation data. *Sensors* **2017**, *17*, 2524. [[CrossRef](#)] [[PubMed](#)]
- Zhang, W.; Li, X.; Jia, X.D.; Ma, H.; Luo, Z.; Li, X. Machinery fault diagnosis with imbalanced data using deep generative adversarial networks. *Measurement* **2019**, *152*, 107377. [[CrossRef](#)]
- Zhang, W.; Li, X.; Ding, Q. Deep residual learning-based fault diagnosis method for rotating machinery. *ISA Trans.* **2019**, *95*, 295–305. [[CrossRef](#)] [[PubMed](#)]
- Huang, X.Z.; Li, Y.X.; Zhang, Y.M.; Zhang, X.F. A new direct second-order reliability analysis method. *Appl. Math. Model.* **2017**, *55*, 68–80. [[CrossRef](#)]
- Hanachi, H.; Liu, J.; Banerjee, A.; Chen, Y.; Koul, A. A physics-based modeling approach for performance monitoring in gas turbine engines. *IEEE Trans. Reliab.* **2015**, *64*, 197–205. [[CrossRef](#)]
- Seyed, M.M.H.N.; Jin, X.; Ni, J. Physics-based Gaussian process for the health monitoring for a rolling bearing. *Acta Astronaut.* **2019**, *154*, 133–139.
- Si, X.S.; Zhang, Z.X.; Hu, C.H. An adaptive remaining useful life estimation approach with a recursive filter. In *Data-Driven Remaining Useful Life Prognosis Techniques*; Springer: Berlin/Heidelberg, Germany, 2017.
- Fan, Z.L.; Liu, G.B.; Si, X.S.; Zhang, Q. Degradation data-driven approach for remaining useful life estimation. *J. Syst. Eng. Electron.* **2013**, *24*, 173–182. [[CrossRef](#)]
- Peng, Y.F.; Cheng, J.S.; Liu, Y.F.; Li, X.J.; Peng, Z.H. An adaptive data-driven method for accurate prediction of remaining useful life of rolling bearings. *Front. Mech. Eng.* **2018**, *13*, 301–310. [[CrossRef](#)]
- Long, Y.W.; Luo, H.W.; Zhi, Y.; Wang, X.Q. Remaining useful life estimation of solder joints using an ARMA model optimized by genetic algorithm. In Proceedings of the 19th International Conference on Electronic Packaging Technology, Shanghai, China, 8–11 August 2018; pp. 1108–1111.
- Tang, S.J.; Yu, C.Q.; Wang, X.; Guo, X.S.; Si, X.S. Remaining useful life prediction of lithium-ion batteries based on the Wiener process with measurement error. *Energies* **2014**, *7*, 520–547. [[CrossRef](#)]
- Khanh Le, S.; Fouladirad, M.; Barros, A.; Levrat, E.; Lung, B. Remaining useful life estimation based on stochastic deterioration models: A comparative study. *Reliab. Eng. Syst. Safe* **2013**, *112*, 165–175.
- Liu, Y.; Zuo, M.J.; Li, Y.F.; Huang, H.Z. Dynamic reliability assessment for multi-state systems utilizing system-level inspection data. *IEEE Trans. Reliab.* **2015**, *64*, 1287–1299. [[CrossRef](#)]
- Graves, A.; Mohamed, A.; Hinton, G.E. Speech recognition with deep recurrent neural networks. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013; pp. 6645–6649.
- Zhang, K.; Zuo, W.M.; Chen, Y.J.; Meng, D.Y.; Zhang, L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [[CrossRef](#)] [[PubMed](#)]
- Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv* **2014**, arXiv:1409.1259.
- Gebraeel, N.; Lawley, M.; Liu, R.; Parmeshwaran, V. Residual life predictions from vibration-based degradation signals: A neural network approach. *IEEE Trans. Ind. Electron.* **2004**, *51*, 694–700. [[CrossRef](#)]
- Aydemir, G.; Acar, B. Anomaly monitoring improves remaining useful life estimation of industrial machinery—Science Direct. *J. Manuf. Syst.* **2020**, *56*, 463–469. [[CrossRef](#)]
- Fu, B.H.; Wu, Z.Y.; Guo, J.C.H. Remaining useful life prediction under multiple operation conditions based on domain adaptive sparse auto-encoder. In Proceedings of the IEEE International Conference on Prognostics and Health Management (ICPHM), Detroit, MI, USA, 8–10 June 2020.

22. Li, X.; Jia, X.D.; Wang, Y.L.; Yang, S.J.; Lee, J. Industrial remaining useful life prediction by partial observation using deep learning with supervised attention. *IEEE ASME Trans. Mechatron.* **2020**, *25*, 2241–2251. [[CrossRef](#)]
23. Loannidis, A.; Nikalaos, N.; Kosmas, A. Energy-based prognosis of the remaining useful life of the coating segments in hot rolling mill. *Appl. Sci.* **2020**, *10*, 6827.
24. Xi, L.F. Residual life predictions for ball bearing based on neural networks. *Chin. J. Mech. Eng.* **2007**, *43*, 137. [[CrossRef](#)]
25. Liu, Z.J.; Zuo, M.J.; Qin, Y. Remaining useful life prediction of rolling element bearings based on health state assessment. *Proc. Inst. Mech. Eng. Part C* **2016**, *230*, 314–330. [[CrossRef](#)]
26. Khelif, R.; Chebel-Morello, B.; Malinowski, S.; Laaj-ili, E.; Fnaiech, F.; Zerhouni, N. Direct remaining useful life estimation based on support vector regression. *IEEE Trans. Ind. Electron.* **2017**, *64*, 2276–2285. [[CrossRef](#)]
27. Lin, Z.B.; Gao, H.L.; Zhang, E.Q.; Cao, W.Q.; Li, K.S. Diamond-coated mechanical seal remaining useful life prediction based on convolution neural network. *Int. J. Pattern Recognit. Artif. Intell.* **2020**, *34*, 214–228. [[CrossRef](#)]
28. 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](#)]
29. Heimes, F.O. Recurrent neural networks for remaining useful life estimation. In Proceedings of the IEEE International Conference on Prognostics and Health Management PHM, Denver, CO, USA, 6–9 October 2008.
30. Maio, F.D.; Tsui, K.L.; Zio, E. Combining relevance vector machines and exponential regression for bearing residual life estimation. *Mech. Syst. Signal Process.* **2012**, *31*, 405–427. [[CrossRef](#)]
31. Zheng, X.J.; Wu, H.Y.; Chen, Y. Remaining useful life prediction of lithium-ion battery using a hybrid model-based filtering and data-driven approach. In Proceedings of the 11th Asian Control Conference (ASCC), Gold Coast, Australia, 17–20 December 2017; pp. 2698–2703.
32. Li, X.; Ding, Q.; Sun, J.Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliab. Eng. Syst. Safe* **2018**, *172*, 1–11. [[CrossRef](#)]
33. Wang, J.J.; Wen, G.L.; Yang, S.P.; Liu, Y.Q. Remaining useful life estimation in prognostics using deep bidirectional LSTM neural network. In Proceedings of the Prognostics and System Health Management Conference (PHM), Chongqing, China, 26–28 October 2018; pp. 1037–1042.
34. Babu, G.S.; Zhao, P.; Li, X. Deep convolutional neural network based regression approach for estimation of remaining useful life. In Proceedings of the International Conference on Database Systems for Advanced Applications, Dallas, TX, USA, 16–19 April 2016; pp. 214–228.
35. Ren, L.; Sun, Y.Q.; Wang, H.; Zhang, L. Prediction of bearing remaining useful life with deep convolution neural network. *IEEE Access* **2018**, *6*, 13041–13049. [[CrossRef](#)]
36. Zhang, Y.Z.; Xiong, R.; He, H.W.; Pecht, M. Long short-term memory recurrent neural network for remaining useful life prediction of lithiumion batteries. *IEEE Trans. Veh. Technol.* **2018**, *67*, 5695–5705. [[CrossRef](#)]
37. Wu, Y.T.; Yuan, M.; Dong, S.P.; Lin, L.; Liu, Y.Q. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing* **2018**, *275*, 167–179. [[CrossRef](#)]
38. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
39. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
40. Zhang, C.; Lim, P.; Qin, A.K.; Tan, K.C. Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. *IEEE Trans. Neural Netw.* **2017**, *28*, 2306–2318. [[CrossRef](#)] [[PubMed](#)]
41. Zheng, S.; Ristovski, K.; Farahat, A.K.; 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.

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



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).