

# The Noisy Multipath Parallel Hybrid Model for Remaining Useful Life Estimation (NMPM)

Ali Al-Dulaimi<sup>1</sup>, Arash Mohammadi<sup>2</sup>, and Amir Asif<sup>3</sup>,

<sup>1,3</sup> *Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada H3J 1P8*  
*ald\_a@encs.concordia.com*  
*amir.asif@encs.concordia.com*

<sup>2</sup> *Concordia Institute for Information System Engineering (CIISE), Concordia University, Canada*  
*arashmoh@encs.concordia.com*

## ABSTRACT

Recently, there has been an increasing surge of interest on development of parallel-hybrid models of different Deep Neural Networks (DNNs) architectures for Remaining Useful Life (RUL) estimation. In this regard, the paper introduces, for the first time in the literature, a new parallel-hybrid DNN-based framework for RUL estimation, referred to as the Noisy Multipath Parallel Hybrid Model for Remaining Useful Life Estimation (NMPM). The proposed NMPM framework comprises of three parallel paths, the first one utilizes a noisy Bidirectional Long-Short Term Memory (BLSTM) that used for extracting temporal features and learning the dependencies of sequence data in two directions, forward and backward. The second parallel path, employs noisy MultiLayer Perceptron (MLP) that consists of three layers to extract different class of features. The third parallel path utilizes noisy Convolutional Neural Networks (CNN) to extract a complementary class of features. The concatenated output of the three parallel paths is then fed into a Noisy Fusion Center (NFC) to predict the RLU. The proposed NMPM has been trained based on a noisy training mechanism to enhance its generalization behavior, as well as strengthen the model's overall accuracy and robustness. The NMPM framework is tested and evaluated using CMAPSS dataset provided by NASA illustrating superior performance in comparison to its state-of-the-art counterparts.

## 1. INTRODUCTION

Prognostics and Health Management (PHM), along with Deep Neural Networks (DNNs) have been revolutionizing maintenance by enabling predictive analytics to predict the likelihood of future failures and provide early warnings by

Ali Al-Dulaimi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

determining the failure patterns and factors that could affect industrial operations. Remaining useful life (RUL) is the key metric for predictive maintenance solutions. In order to build an effective maintenance strategy, maximize machine uptime, and minimize maintenance costs, an accurate RUL prediction is a substantial task. Therefore, the existing RUL prediction solutions need to be continually developed and strengthened. There exist three classes of RUL prognostic solutions, i.e., physics based, data driven, and hybrid prognostic methods (Kan, Tan, & Mathew, 2015). Among them, the hybrid solutions are the most promising approaches for remaining useful life (RUL) estimation. The hybrid method itself, has many types and styles (Xia et al., 2018). The focus of this paper is the hybrid methods that used DNN architectures. The hybrid model are explicitly designed to avoid the weaknesses and limitations of the individual underlying approaches, by combining the different features and advantages of each approach, resulting in an improved prognostic outcome. This main advantage has led to gain a lot of attention from many researchers.

**Prior Work:** (Zhao, Yan, Wang, & Mao, 2017) proposed a hybrid deep learning approach by using series integration of CNN architecture and a BLSTM, followed by fusion center of fully connected layers for the prediction task. (Jayasinghe, Samarasinghe, Yuen, Low, & Ge, 2018) utilized the integration of LSTM layers and temporal convolution layers along with data augmentation to predict the RUL values. (Song, Shi, Chen, Huang, & Xia, 2018) designed a hybrid method for improving the RUL prediction accuracy of turbofan engines, by combining an auto-encoder as a feature extractor, and bidirectional LSTM (BLSTM) in order to capture the bidirectional long-range dependencies of features. (Hinchi & Tkiouat, 2018) introduced a series hybrid method through the integration of the LSTMs and the CNNs layers for bearing RUL estimation. Also (Zhao, Wang, et al., 2017) designed a hybrid model by using the Gated Recurrent Unit (GRU) and

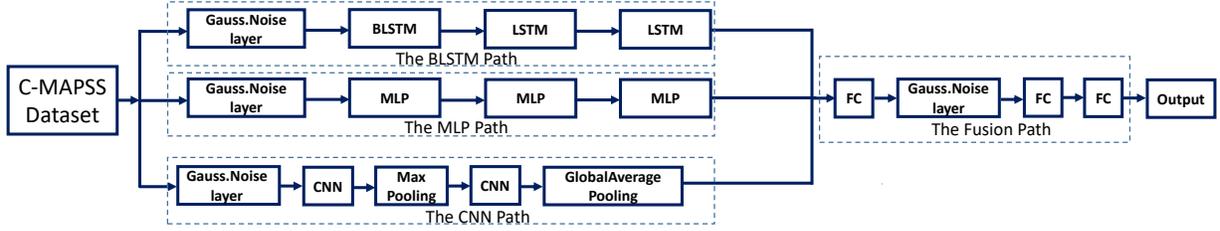


Figure 1. The proposed NMPM solution.

the Bi-directional GRU (BGRU) architectures for machine health monitoring task. (Al-Dulaimi, Zabihi, Asif, & Mohammadi, 2019) introduced a parallel hybrid DNN, named as (HDNN). It designed by using two parallel paths ( CNN and LSTM) followed by a fusion center to combine their outputs and then estimate the RUL. Moreover, (J. Li, Li, & He, 2019) Used CNN and LSTM in parallel paths, and then another LSTM to fuse the extracted features. Then a fully connected neural network for mapping these features to RUL. (Kong, Cui, Xia, & Lv, 2019) proposed a hybrid DNN framework that employed the CNN and LSTM integrated with health indicator (HI), for extracting different classes of features to effectively prognostication. (Yu, Kim, & Mechefske, 2019) has employed Restricted Boltzmann Machine (RBM) and vanilla LSTM to form a semi-supervised model for exploring the impact of unsupervised pre-training in residual lifetime estimation. Finally, we have proposed a new parallel hybrid deep neural network (NBLSTM) (Al-Dulaimi, Zabihi, Asif, & Mohammed, 2019), which Integrates the Bi-directional LSTM (BLSTM) and LSTM, in one path, parallel with another path based on CNN layers, then the output of both paths combined to form the RUL. Adding noise layers is an effective technique that has been adopted to improve the performance, generalization and robustness of the approach.

**Contributions:** This paper proposes a noisy multi-path parallel hybrid model for RUL prediction, referred to as the (NMPM). The (NMPM) method integrates three parallel paths followed by another one to combine the outputs and form the target result. The key contributions of this paper are: (i) A new multi-path parallel hybrid DNN framework that integrates three neural networks architectures (BLSTM and LSTM, MLP, and CNN) is proposed for estimating remaining useful life of complex systems. (ii) To the best of our knowledge, the NMPM is the first multi-path parallel hybrid DNN model for RUL estimation. (iii) The efficiency of the proposed NMPM solution is then verified and evaluated using the C-MAPSS dataset provided by NASA (Saxena & Goebel, 2008) and the achieved outcomes were the best among the existing methods.

The paper is organised in the following way: Section 2 to describe the dataset. Section 3 to develop and describe the proposed NMPM model. Section 4 presents the experimental

results. And finally, Section 5 will be the last section in the paper, that presents the conclusion.

## 2. DATA

In this section, a general overview of the dataset used to verify the efficacy of the proposed solution is given.

### 2.1. The NASA (C-MAPSS) Dataset

The C-MAPSS is the most popular simulated dataset for RUL prediction, which was produced on the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) (Saxena & Goebel, 2008). This software simulates several scenarios of the degradation behavior and the faults impact of the five main modules (High Pressure Compressor (HPC), High Pressure Turbine, Low Pressure Compressor, Low Pressure Turbine, and Fan) for the turbofan engine, over different sets of operating conditions and failure modes. According to these conditions and failure modes, the dataset can be grouped into 4 sub-datasets, which are FD001 to FD004 as shown in Table 1, that are composed of multivariate temporal data collected from 21 sensors. Each sub-dataset contains training sets including run to fail records of different engines, in addition to testing sets that include a number of instances of incomplete data ending before failure, the RULs of which are to be calculated using the proposed prediction models. Each sub-dataset has been considered as a matrix of size  $R \times 26$ , where  $R$  corresponds to the total length of the trajectories. Each record contains 26 columns each one of them corresponds to a particular variable. The first two fields are for the engine ID and the cycle index, respectively, then the next three columns represent the operating conditions, followed by the readings of 21 sensors. A trajectory in a specific dataset reflects a particular engine's lifetime. Each engine has different initial condition, yet, it considered to start in the healthy state, but the last cycle is classified as the failure of the system. During testing, the trajectories terminate at a specific time prior to failure to estimate the RUL of each engine. The true RUL measurements for test trajectories are available for verification purposes. For this study only (FD002 and FD004) sub-datasets have been considered. More details on the C-MAPSS dataset can be found in (Saxena, Goebel, Simon, & Eklund, 2008).

Dataset	C-MAPSS			
	FD001	FD002	FD003	FD004
Train Trajectories	100	260	100	249
Test Trajectories	100	259	100	248
Conditions	1	6	1	6
Fault Modes	1	1	2	2

Table 1. C-MAPSS Dataset Details (Saxena & Goebel, 2008).

## 2.2. Data Preparation

### 2.2.1. Data Normalization

As multiple sensors were used to collect the data and that under different operational conditions, in such cases the reading values will be in different scales, therefore the normalization is an essential step before feeding to the deep learning model. The min-max normalization is employed to unify the values to be within a specific range and to ensure impartial involvement from the readings of each sensor, and given by

$$\bar{x}_i = \frac{2(x_i - \min x_i)}{\max x_i - \min x_i} - 1, \quad (1)$$

where  $\bar{x}_i$  and  $x_i$ , respectively denote the normalized data and time sequence for the  $i_{th}$  sensor. The normalized data is within  $[-1, 1]$ . In this study, only 14 sensors out of 21 sensors, were selected as in reference (C. Zhang, Lim, Qin, & Tan, 2016), where, not all of the sensors measurements are useful and informative.

### 2.2.2. The piecewise linear target RUL

The piece-wise linear degradation is adopted in this paper, as it is the most popular and successful strategy for the utilized dataset (Al-Dulaimi, Zabihi, Asif, & Mohammadi, 2019), as it is the logical choice to have constant value of RUL when the system is healthy (new or maintained) and degradation is negligible, and after the initial period, the RUL decreases linearly with time and the degradation is increasing as in Fig. 2. In this study, the point of inflection when the degradation starts, has been chosen to be a value of 125.

## 2.3. Performance evaluation indicators

In order to evaluate, analyze, and compare performance of the proposed NMPM model with others, two commonly used performance evaluation indicators, i.e., Scoring function, and Root Mean Square Error (RMSE), are adopted.

- (i) *Scoring Function*: It is a function defined during the PHM data challenge competition (Saxena et al., 2008) in 2008 by the PHM community, that is given by

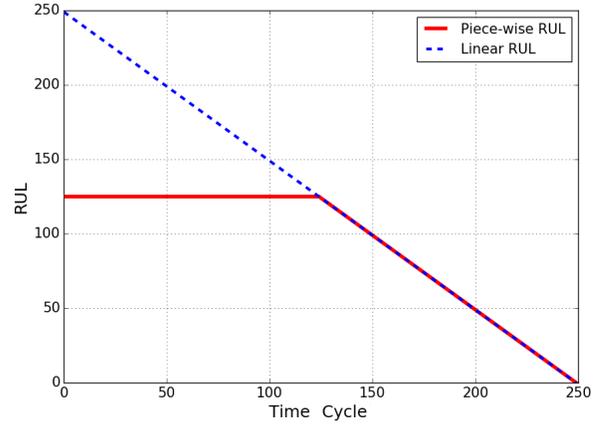


Figure 2. Representation of piece-wise approach.

$$S = \sum_{i=1}^{M_{te}} s_i, \quad (2)$$

$$\text{where } s_i = \begin{cases} e^{-\frac{h_i}{13}} - 1 & \text{for } h_i < 0 \\ e^{\frac{h_i}{10}} - 1 & \text{for } h_i \geq 0, \end{cases}$$

where  $S$  is the computed score,  $M_{te}$  is the total number of testing data samples, and  $h_i = \overline{RUL}_i - RUL_i$  (estimated RUL - true RUL, with respect to the  $i_{th}$  data point).

- (ii) *The RMSE*: A typical widely used performance indicator. That is given by

$$\text{RMSE} = \sqrt{\frac{1}{M_{te}} \sum_{i=1}^{M_{te}} h_i^2}. \quad (3)$$

The lower the assessment metrics, the better the proposed method perform.

## 3. THE PROPOSED FRAMEWORK

The key idea behind the proposed model is to extract as much information (features) as possible from the available datasets to better characterise the underlying problem. Our intuition is based on the following statement: “the more features you receive, the better the results you achieve (Brownlee, 2018)”. Furthermore, collecting more features would enhance the process of learning data attributes as, consequently, more information will be available to form the target values. To this end, an integrated idea has been used by utilizing three successful classes of DNNs, which are BLSTM, MLP, and CNN, to form a multipath parallel hybrid model. The proposed method, however, is not limited to these three techniques, and one can use another DNN model within the proposed parallel

hybrid structure, in case the new model can outperform any of the selected architectures. The outputs of the three paths are then combined to predict the RUL using fully connected neural networks.

### 3.1. Framework description

The main components of the proposed NMPM are listed below, as presented in Fig. 1

#### 3.1.1. The Bidirectional Long Short Term Memory (BLSTM) Path

The first path in the proposed NMPM model, has been designed based on one of the most successful DNNs architectures which is the BLSTM, in addition to the LSTM. The LSTM/BLSTM networks are the most efficient recurrent neural network models used in practice, they are specifically proposed to address the vanishing gradient problem, using their ability to capture time-dependent relationships through a gating system to provide a memory-based structure (Bhardwaj, Di, & Wei, 2018). The gating system is designed based on three gates (i) An input gate  $i^t \in \mathbb{R}^{M \times 1}$ , that controls the cell state updating procedure based on  $h^{t-1}$  and  $x^t$ , where  $M$  is the number of nodes,  $h^{t-1}$  denotes the LSTM output at time  $(t-1)$ , and  $x^t$  represents the input to LSTM at time  $t$ . (ii) A forget gate denoted by  $f^t \in \mathbb{R}^{M \times 1}$ , that decides on the contents to be maintained or forgotten. (ii) A forget gate denoted by  $f^t \in \mathbb{R}^{M \times 1}$ , that decides on the contents to be maintained or forgotten, and (iii) An output gate denoted by  $o^t \in \mathbb{R}^{M \times 1}$ , that computes the next value of the hidden state.

The BLSTM is the modified version of the LSTM architecture, that designed to capture the temporal dependencies between extracted features and take full advantage of the input data, as it used two LSTM layers to be applied in both directions of the hidden sequences (Schuster & Paliwal, 1997), i.e., forward  $\vec{h}^t$  and backward  $\overleftarrow{h}^t$ , which are then joined to calculate the output sequence. Fig. 3 presents the block diagram of the BLSTM network. At each time step  $t$ , the BLSTM model calculates both directions ( $\vec{h}^t$ , &  $\overleftarrow{h}^t$ ) separately, and then concatenates the outputs to form the BLSTM output denoted by  $h_{bi}^t$ . The corresponding hidden layer functions of the BLSTM architecture, are defined as

$$\vec{i}^t = \sigma(\vec{W}_i \vec{x}^t + \vec{U}_i \vec{h}^{t-1} + \vec{b}_i), \quad (4)$$

$$\vec{o}^t = \sigma(\vec{W}_o \vec{x}^t + \vec{U}_o \vec{h}^{t-1} + \vec{b}_o), \quad (5)$$

$$\vec{f}^t = \sigma(\vec{W}_f \vec{x}^t + \vec{U}_f \vec{h}^{t-1} + \vec{b}_f), \quad (6)$$

$$\vec{a}^t = \tanh(\vec{W}_c \vec{x}^t + \vec{U}_c \vec{h}^{t-1} + \vec{b}_c), \quad (7)$$

$$\vec{c}^t = \vec{f}^t \circ \vec{c}^{t-1} + \vec{i}^t \circ \vec{a}^t, \quad (8)$$

$$\vec{h}^t = \vec{o}^t \circ \tanh(\vec{c}^t), \quad (9)$$

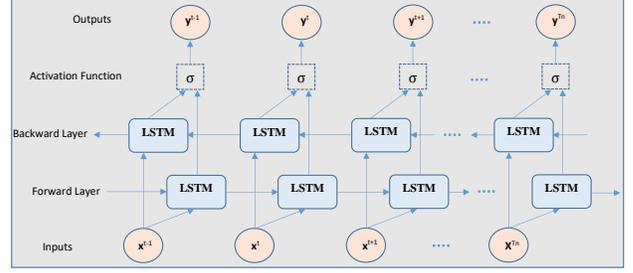


Figure 3. Block diagram of BLSTM.

and

$$\overleftarrow{i}^t = \sigma(\overleftarrow{W}_i \overleftarrow{x}^t + \overleftarrow{U}_i \overleftarrow{h}^{t+1} + \overleftarrow{b}_i), \quad (10)$$

$$\overleftarrow{o}^t = \sigma(\overleftarrow{W}_o \overleftarrow{x}^t + \overleftarrow{U}_o \overleftarrow{h}^{t+1} + \overleftarrow{b}_o), \quad (11)$$

$$\overleftarrow{f}^t = \sigma(\overleftarrow{W}_f \overleftarrow{x}^t + \overleftarrow{U}_f \overleftarrow{h}^{t+1} + \overleftarrow{b}_f), \quad (12)$$

$$\overleftarrow{a}^t = \tanh(\overleftarrow{W}_c \overleftarrow{x}^t + \overleftarrow{U}_c \overleftarrow{h}^{t+1} + \overleftarrow{b}_c), \quad (13)$$

$$\overleftarrow{c}^t = \overleftarrow{f}^t \circ \overleftarrow{c}^{t+1} + \overleftarrow{i}^t \circ \overleftarrow{a}^t, \quad (14)$$

$$\overleftarrow{h}^t = \overleftarrow{o}^t \circ \tanh(\overleftarrow{c}^t). \quad (15)$$

Then, the concatenated output vector  $h_{bi}^t$  is given by

$$h_{bi}^t = \vec{h}^t \oplus \overleftarrow{h}^t. \quad (16)$$

where terms  $W_i$ ,  $W_o$ ,  $W_f$ , and  $W_c \in \mathbb{R}^{M \times L}$  together with terms  $U_i$ ,  $U_o$ ,  $U_f$ ,  $U_c \in \mathbb{R}^{M \times M}$  constitute the weight matrices; Terms  $b_i$ ,  $b_o$ ,  $b_f$ ,  $b_c \in \mathbb{R}^{M \times 1}$  represent biases; Term  $\sigma(\cdot)$  denotes the sigmoid non-linear function; Operator “ $\circ$ ” represents an entry-wise product operation, which is performed by two vectors element-wise multiplication,  $c^t$  is the cell state at time  $t$ , and; finally,  $\tanh(\cdot)$  denotes the activation function. The BLSTM and LSTM have already shown outstanding results on a variety of issues (You, Jin, Wang, Fang, & Luo, 2016; S. Wang & Jiang, 2015; Sun, Su, Liu, & Wang, 2016; Graves & Schmidhuber, 2005; Graves, Jaitly, & Mohamed, 2013), significantly, in machine health monitoring (Al-Dulaimi, Zabihi, Asif, & Mohammed, 2019). This path consists of four layers, the first layer is a Gaussian noise layer that has zero mean and (0.01) standard deviation, followed by BLSTM layer defined by 10 cell structure with return sequences. The next two layers are LSTM layers defined by 10, and 21 cell structures, respectively. All cells in the LSTM layer have the same configuration in terms of parameter values and structure.

#### 3.1.2. The Multilayer Perceptron (MLP) Path

MLP is among the most widely used neural networks, as it forms the basis for all neural networks (Trenn, 2008). It is a feed forward artificial neural network architecture, that made of three main parts: an input layer, an intermediate layer (one or more) and output layer, where each layer is fully connected to the following layer of nodes, in other words, this multi-

layered perceptron consists of interconnected neurons that transmit information among themselves, similar to the human brain (Trenn, 2008). MLP often applied to supervised learning models, which employ back-propagation method to train the network. It has been used in a broad range of fields, like image identification, stock analysis, and election voting predictions (Trenn, 2008), in addition to the impressive success in PHM applications (X. Li, Ding, & Sun, 2018). This path based on the following settings: Three fully connected layers defined by (30, 27, and 10) neurons, respectively. The first layer of which followed by a noisy layer that has zero-mean Gaussian noise and (0.01) standard deviation. A dropout of (0.2) has been used, and the activation function was the rectified linear unit (ReLU), used for all layers.

### 3.1.3. The Convolutional Neural Networks (CNN) Path

The outstanding capability to identify spatial and temporal dependencies, made the CNN as one of the most powerful feature extraction tools. It has been effectively utilized in a broad range of applications, for example, computer vision, biomedical, speech recognition, and Remaining useful life estimation (X. Li et al., 2018; Babu, Zhao, & Li, 2016; Ren, Sun, Wang, & Zhang, 2018), to name but a few. The CNN has two main parts, i.e., feature extraction part and the classification part (W. Zhang, Peng, & Li, 2017). This path of the proposed NMPM model has been designed based on the following layers: two CNN layers, one max pooling layer, one Gaussian noise layer, and one layer of global average pooling. The path starts with a Gaussian noise layer that has zero-mean and (0.01) standard deviation, followed by the first CNN layer that has 10 filters of size  $(11 \times 1)$ , and then one max pooling layer that has  $(2 \times 1)$  filter, followed by the second CNN layer that has 100 filters of size  $(11 \times 1)$ . The last layer of this path is a global average pooling layer. The rectified linear unit (ReLU) was the activation function for all the CNN layers. In this paper, the input data to the CNN path of the proposed model NMPM is a 2D structure of the time sequence and the number of selected features. More details about the CNN layers can be found in (Al-Dulaimi, Zabihi, Asif, & Mohammadi, 2019).

**Max Pooling:** It is a sub-sampling technique employed for the size reduction of the feature maps through selecting the maximum value for each patch, while preserving the important information. Hence reducing the number of model parameters and simplifying the computational complexity of the network. In addition to control the overfitting problem.

**Global Average Pooling:** It is an operation that involves calculating the average value of all the elements in the feature map. The overfitting is usually avoided at this layer, since the global average pooling has no parameter to optimize, which result in speeding up the training of the model (Lin, Chen, & Yan, 2013).

### 3.1.4. The Fusion Path

The Fusion path acts as ‘‘Fusion Centre’’ that combines the output features formed by the three parallel paths to make the final prediction. Three different fully connected layers are used to build the fusion centre. The first layer has 103 neurons followed by a Gaussian noise layer that has zero mean and (0.01) standard deviation, then a dropout of 0.3, followed by the second and the third layers that have 107, and 1 neurons, respectively. The rectified linear unit (ReLU) was the activation function for all the layers.

### 3.2. Noisy Training

The training is a key process that involves finding patterns that map the input data attributes to the targeted values we want to predict (Goodfellow, Bengio, & Courville, 2016). To evaluate how well the proposed algorithm models the given data and find those patterns, the cost function is used, which is an indicator to show the error between the predicted results and the real values. Thus, the training process aims at minimizing the cost function by finding the optimal parameters (weights and biases). The proposed NMPM model, has used the mean squared error (MSE) as the cost function, and it is given by

$$\text{MSE} = \frac{1}{M_{tr}} \sum_{i=1}^{M_{tr}} h_i^2, \quad (17)$$

where  $M_{tr}$  denotes the number of training samples, and  $h_i$  represents the error between the true and the estimated RUL, i.e.,  $(h_i = \overline{RUL}_i - RUL_i)$ . Minimizing the cost function leads to make the estimated RUL as close as possible to the actual RUL. Adaptive moment estimation (Adam) (Kingma & Ba, 2014) is utilized for minimizing loss function. Each training set has been divided randomly into 85% and 15% for testing and validation, respectively. The Grid search (Bergstra, Bardenet, Bengio, & Kégl, 2011) approach along with several experiments have been conducted on the train dataset, and hyperparameters that achieved the best validation prediction performance were considered.

For better generalization and faster learning in the proposed model, a noisy training has been adopted by injecting Gaussian noise layers in each path of this approach (Yin et al., 2015), this technique was an efficient way of avoiding the overfitting issue, because the model learns the main concepts underlying the problem, instead of only memorizes the full dataset. Values of the hyperparameters such as the number of RNN layers, the cells in each layer, the CNN layers, the FC layers, batch size, standard deviation values and dropout rate, are selected using the grid search technique, as it is the most commonly used technique aimed at defining the appropriate set of hyperparameters for a particular model (Liashchynskiy & Liashchynskiy, 2019).

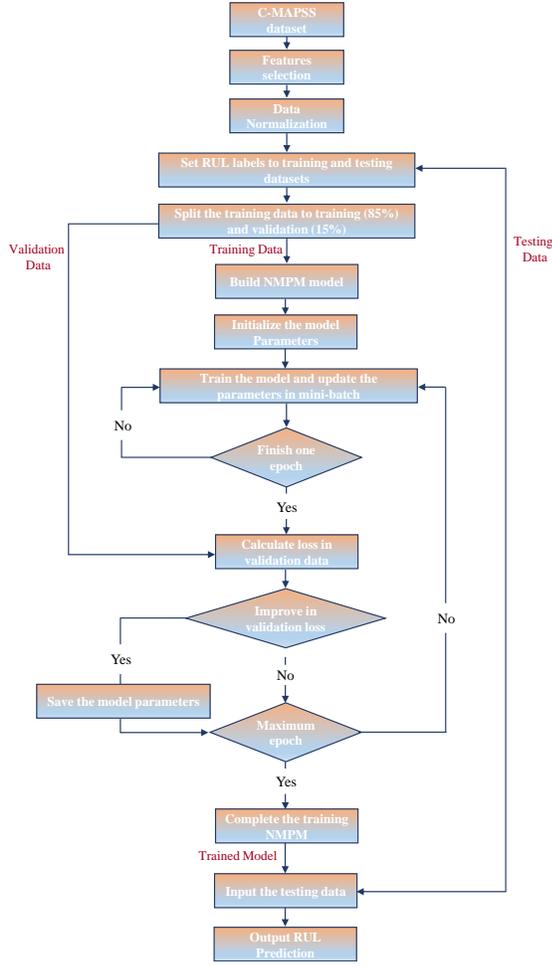


Figure 4. Flowchart of the proposed NMPM.

The batch size of 512 is used in addition to the mini batch gradient descent approach (Brownlee, 2017). The early stopping (Famouri, Taheri, & Azimifar, 2015) and dropout (Valchanov, 2018) techniques are used to mitigate the overfitting problems. The sliding window strategy is adopted with 30 and 15 window sizes, and step size of 1. Fig. 4 describes the complete procedures of the proposed NMPM approach.

#### 4. EXPERIMENTAL RESULTS

In this section, the proposed NMPM framework is evaluated and tested using the benchmarking NASA's C-MAPSS dataset. The impacts of various factors on the performance of the NMPM solution are investigated, including the noisy training, multiple parallel paths, and time window size. As well as, comparisons with other existing solutions were conducted and the results were reported.

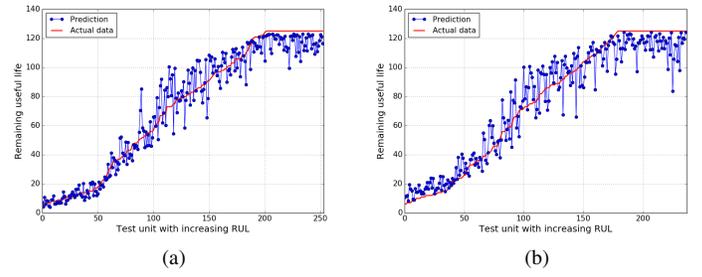


Figure 5. Predictions for last recorded measurement sample associated with different test engine in: (a) FD002 with 259 test engines, and; (b) FD004 with 248 test engines.

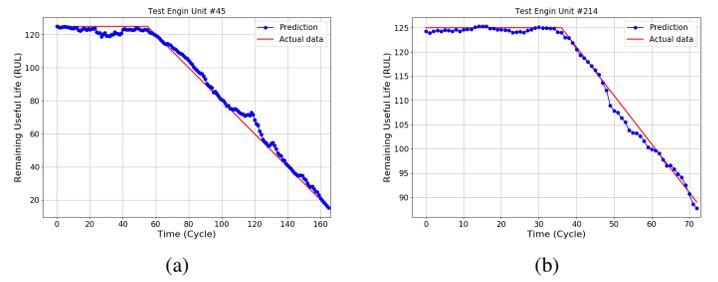


Figure 6. Estimated lifetime RUL predictions for randomly selected test engine units in: (a) Test engine 45 in FD002, and; (b) Test engine 214 in FD004.

#### 4.1. The RUL Estimation Results

Figs 5(a) and (b) exhibit the RUL prognostics performance of all turbofans testing samples from sub-datasets FD002 and FD004, sorted in ascending order (the lowest to highest RUL), where the prediction results for both sub-datasets related to the last recorded measurement sample of all the datasets. It is observed that the prediction RUL values is closer to the real values (ground truth), and the prediction accuracy for engines with a smaller RUL can be observed to be significantly higher when the engine units are close to the end of their lives (small RUL values), which is expected, as the engine operates normally in the early stages of its life when the degradation is negligible, and then the RUL decreases linearly with time when the system approaches its “end-of-life”, and the degradation is more critical. This trend is important and beneficial to health monitoring of equipment since an accurate prediction is more crucial for making decisions in the later period. Figs 6(a) and (b) illustrate RUL estimation results of engine units (45, and 214) selected at random from FD002 and FD004 sub-datasets, respectively. It is noted that the predicted RUL values precisely follow the actual values, which is pointing to the prediction quality of the proposed model.

#### 4.2. The Effects of Noisy Training and Multiple Parallel Paths

Table 2 shows the results of the proposed network with noise (NMPM) and without adding the noise (WNMPM). It demonstrates and proves that adopting the used technique of adding noise layer, has impressively improved the performance of the proposed model, and that due to preventing the model from memorizing all training examples, which in turn leads to boost the exploration performance of reinforcement learning algorithms, and then enhance the robustness and the generalization of the proposed model.

Additionally, the use of multiple parallel paths based on different DNNs architectures increases model efficiency, because it is essential to get as many informative features as possible from the available datasets in order to develop an effective solution. Moreover, the proposed model has been implemented twice and in each time, we removed one of the parallel paths (BLSTM path or MLP path) following the same settings of the proposed NMPM, and we found that using the three paths (as in our model) improves the results more than 65% in terms of score function and more than 29% in terms of RMSE.

TW=30		NMPM	WNMPM
FD002	Score	843.044	956.854
	RMSE	14.176	14.979
FD004	Score	991.214	1278.648
	RMSE	16.035	16.253

Table 2. The Effects of Noisy Training.

#### 4.3. The Effects of Different Time Window Size

The key point of a high quality prediction model is to extract more informative features, thus using larger size of time window ends with more precise RUL estimation. Table 3 Presents the impact of time-window size on the efficiency of the NMPM model. Where, a smaller window size of 15 is utilized, however, the results are still better than most of those reported in the literature for window size of 30.

#### 4.4. Comparison with Existing Solutions

To illustrate and evaluate the efficiency of the proposed NMPM solution, the prediction results of this model are compared with six different published studies, which are: NBLSTM (Al-Dulaimi, Zabihi, Asif, & Mohammed, 2019), HDNN (Al-Dulaimi, Zabihi, Asif, & Mohammadi, 2019), DSCN (B. Wang, Lei, Li, & Yan, 2019), DAG (J. Li et al., 2019), CapsNet (Ruiz-Tagle Palazuelos, Droguett, & Pascual, 2019), and CNNTW (Yang, Zhao, Jiang, Sun, & Mei, 2019), these studies represent the latest and the most successful solutions. The comparison has been conducted based on the most

Metrics	FD002	FD004
Score	1165.93	1901.148
RMSE	16.528	18.488

Table 3. The outcomes for the window size of 15.

TW=30		NMPM	NBLSTM	HDNN	DSCN	DAG	CapsNet	CNNTW
FD002	Score	843.044	1056.629	1282.424	367.56	2730	1229.72	2494.35
	RMSE	14.176	15.038	15.24	20.47	20.34	16.30	19.58
FD004	Score	991.214	1357.20	1527.425	168.45	33702	2625.64	4523.32
	RMSE	16.035	17.752	18.156	22.64	22.43	18.96	22.12

Table 4. Performance Comparison of different solutions with the proposed (NMPM).

complex scenarios within the C-MAPSS datasets, which are (FD002 and FD004). Due to this complexity (the number of operation conditions and fault modes), for these two situations, most of the existing methods fail to deliver competitive results. Whereas, the proposed NMPM model achieves impressive results.

Table 4 presents the performance comparison results of the proposed NMPM and the other six approaches. It can be clearly seen that the proposed NMPM solution has the highest prediction accuracy across all methods by achieving the lowest score values and a lowest RMSE values, which implies that the proposed NMPM is performing significantly better in turbofan engine RUL prediction. Here, two points can be highlighted: (i) achieving lower score values represents earlier prediction of RUL, which promotes more efficiency, effectiveness and safety in real life PHM applications, where, the score function measure has a higher penalty for late estimation) (Saxena et al., 2008), (ii) the model has achieved these outstanding outcomes with only 56000 parameters.

The results in terms of the score values have been improved, compared to the best results (NBLSTM) available in the literature as follows, 20.21%, and 27% for FD002, and FD004, respectively. While, in terms of the RMSE values, the proposed NMPM model achieved 5.73%, and 9.7% improvements for FD002, and FD004, respectively.

## 5. CONCLUSION

The paper proposed a novel framework for RUL estimation, i.e., the *Noisy Multipath Parallel Hybrid Model for Remaining Useful Life Estimation (NMPM)*. The noisy training has been adopted for better generalization and faster learning in the proposed model. The proposed NMPM network has been designed by utilizing three parallel paths each one of them based on different neural network architectures (BLSTM with LSTM, MLP, and CNN), the output features of these parallel paths will be integrated by multi-layered neural networks that act as the fusion center to estimate the RUL values. To show the effectiveness of the proposed NMPM solution, different experiments are performed using the most

complex sub-datasets (FD002 and FD004) of (C-MAPSS) datasets provided by NASA. The model is evaluated and compared with several methods in terms of the scoring function and RMSE value, and the results showed the prominent superiority of the proposed NMPM method. Although, the results were superior against state-of-the-art, the proposed model can be improved by adopting more flexible methods in selecting the point of inflection (when the degradation starts). Furthermore, one can pursue incorporation of approaches other than the piece-wise linear degradation approximation method.

## REFERENCES

- Al-Dulaimi, A., Zabihi, S., Asif, A., & Mohammadi, A. (2019). A multimodal and hybrid deep neural network model for remaining useful life estimation. *Computers in Industry*, 108, 186–196.
- Al-Dulaimi, A., Zabihi, S., Asif, A., & Mohammed, A. (2019). Nblstm: Noisy and hybrid cnn and blstm-based deep architecture for remaining useful life estimation. *Journal of Computing and Information Science in Engineering*, 1–14.
- Babu, G. S., Zhao, P., & Li, X.-L. (2016). Deep convolutional neural network based regression approach for estimation of remaining useful life. In *International conference on database systems for advanced applications* (pp. 214–228).
- Bergstra, J. S., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. In *Advances in neural information processing systems* (pp. 2546–2554).
- Bhardwaj, A., Di, W., & Wei, J. (Eds.). (2018). *Deep learning essentials: Your hands-on guide to the fundamentals of deep learning and neural network modeling*. Packt Publishing Ltd.
- Brownlee, J. (2017). A gentle introduction to mini-batch gradient descent and how to configure batch size. Available from: *ML Mastery*; <https://machinelearningmastery.com/gentle-introduction-mini-batch-gradient-descent-configure-batch-size/> (accessed 23 July 2017).
- Brownlee, J. (Ed.). (2018). *Better deep learning train faster, reduce overfitting, and make better predictions*. Machine learning mastery.
- Famouri, M., Taheri, M., & Azimifar, Z. (2015). Fast linear svm validation based on early stopping in iterative learning. *International Journal of Pattern Recognition and Artificial Intelligence*, 29(08), 1551013.
- Goodfellow, I., Bengio, Y., & Courville, A. (Eds.). (2016). *Deep learning*. MIT press.
- Graves, A., Jaitly, N., & Mohamed, A.-r. (2013). Hybrid speech recognition with deep bidirectional lstm. , 273–278.
- Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6), 602–610.
- Hinchi, A. Z., & Tkiouat, M. (2018). Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network. *Procedia Computer Science*, 127, 123–132.
- Jayasinghe, L., Samarasinghe, T., Yuen, C., Low, J. C. N., & Ge, S. S. (2018). Temporal convolutional memory networks for remaining useful life estimation of industrial machinery. *arXiv preprint arXiv:1810.05644*.
- Kan, M. S., Tan, A. C., & Mathew, J. (2015). A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mechanical Systems and Signal Processing*, 62, 1–20.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kong, Z., Cui, Y., Xia, Z., & Lv, H. (2019). Convolution and long short-term memory hybrid deep neural networks for remaining useful life prognostics. *Applied Sciences*, 9(19), 4156.
- Li, J., Li, X., & He, D. (2019). A directed acyclic graph network combined with cnn and lstm for remaining useful life prediction. *IEEE Access*, 7, 75464–75475.
- Li, X., Ding, Q., & Sun, J.-Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, 172, 1–11.
- Liashchynskiy, P., & Liashchynskiy, P. (2019). Grid search, random search, genetic algorithm: A big comparison for nas. *arXiv:1912.06059v1 [cs.LG]*.
- Lin, M., Chen, Q., & Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.
- Ren, L., Sun, Y., Wang, H., & Zhang, L. (2018). Prediction of bearing remaining useful life with deep convolution neural network. *IEEE Access*, 6, 13041–13049.
- Ruiz-Tagle Palazuelos, A., Droguett, E. L., & Pascual, R. (2019). A novel deep capsule neural network for remaining useful life estimation. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 1748006X19866546.
- Saxena, A., & Goebel, K. (2008). C-mapss data set.
- Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management* (pp. 1–9).
- Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673–2681.
- Song, Y., Shi, G., Chen, L., Huang, X., & Xia, T. (2018). Remaining useful life prediction of turbofan engine using hybrid model based on autoencoder and bidirectional long short-term memory. *Journal of Shanghai Jiaotong University (Science)*, 23(1), 85–94.

- Sun, L., Su, T., Liu, C., & Wang, R. (2016). Deep lstm networks for online chinese handwriting recognition. In *2016 15th international conference on frontiers in handwriting recognition (icfhr)* (pp. 271–276).
- Trenn, S. (2008). Multilayer perceptrons: Approximation order and necessary number of hidden units. *IEEE transactions on neural networks*, 19(5), 836–844.
- Valchanov, I. (2018). Machine learning: An overview. Available from: *Oracle Blogs* ; <https://blogs.oracle.com/datascience/machine-learning%3a-an-overview/>;(accessed 25 January 2018).
- Wang, B., Lei, Y., Li, N., & Yan, T. (2019). Deep separable convolutional network for remaining useful life prediction of machinery. *Mechanical Systems and Signal Processing*, 134, 106330.
- Wang, S., & Jiang, J. (2015). Learning natural language inference with lstm. *arXiv preprint arXiv:1512.08849*.
- Xia, T., Dong, Y., Xiao, L., Du, S., Pan, E., & Xi, L. (2018). Recent advances in prognostics and health management for advanced manufacturing paradigms. *Reliability Engineering & System Safety*, 178, 255–268.
- Yang, H., Zhao, F., Jiang, G., Sun, Z., & Mei, X. (2019). A novel deep learning approach for machinery prognostics based on time windows. *Applied Sciences*, 9(22), 4813.
- Yin, S., Liu, C., Zhang, Z., Lin, Y., Wang, D., Tejedor, J., ... Li, Y. (2015). Noisy training for deep neural networks in speech recognition. *EURASIP Journal on Audio, Speech, and Music Processing*, 2015(1), 2.
- You, Q., Jin, H., Wang, Z., Fang, C., & Luo, J. (2016). Image captioning with semantic attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4651–4659).
- Yu, W., Kim, I. Y., & Mechefske, C. (2019). Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme. *Mechanical Systems and Signal Processing*, 129, 764–780.
- Zhang, C., Lim, P., Qin, A. K., & Tan, K. C. (2016). Multi-objective deep belief networks ensemble for remaining useful life estimation in prognostics. *IEEE transactions on neural networks and learning systems*, 28(10), 2306–2318.
- Zhang, W., Peng, G., & Li, C. (2017). Rolling element bearings fault intelligent diagnosis based on convolutional neural networks using raw sensing signal. , 77–84.
- Zhao, R., Wang, D., Yan, R., Mao, K., Shen, F., & Wang, J. (2017). Machine health monitoring using local feature-based gated recurrent unit networks. *IEEE Transactions on Industrial Electronics*, 65(2), 1539–1548.
- Zhao, R., Yan, R., Wang, J., & Mao, K. (2017). Learning to monitor machine health with convolutional bi-directional lstm networks. *Sensors*, 17(2), 273.

## BIOGRAPHIES



**Ali Al-Dulaimi** received the B.Eng. and the M.Sc. degrees in electronics & communications engineering from Baghdad University, Baghdad, Iraq, in 1998, and 2001. Then MEng. degree in quality systems engineering from Concordia University, Montreal, Canada, in 2016, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering. His current research interests include machine learning and its applications in prognostic health management, and stochastic filtering & forecasting with applications to fault diagnosis and failure prognosis of dynamical system.



**Arash Mohammadi** received the B.Sc. degree from the University of Tehran, Tehran, Iran, in 2005, the M.Sc. degree from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2007, and the Ph.D. degree from York University, Toronto, ON, Canada, in 2013. He is currently an Associate Professor with Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada. Prior to joining Concordia University and for 2 years, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. Dr. Mohammadi is a registered professional engineer in Ontario. He is Director-Membership Developments of IEEE Signal Processing Society (SPS); General Co-Chair of “2021 IEEE International Conference on Autonomous Systems (ICAS),” and; Guest Editor for IEEE Signal Processing Magazine (SPM) Special Issue on “Signal Processing for Neurorehabilitation and Assistive Technologies”. He was Co-Chair of “Symposium on Advanced Bio-Signal Processing and Machine Learning for Assistive and Neuro-Rehabilitation Systems” as part of 2019 IEEE GlobalSIP, and “Symposium on Advanced Bio-Signal Processing and Machine Learning for Medical Cyber-Physical Systems,” as a part of IEEE GlobalSIP’18; The Organizing Chair of 2018 IEEE Signal Processing Society Video and Image Processing (VIP) Cup, and the Lead Guest Editor for IEEE Transactions on Signal & Information Processing over Networks Special Issue on “Distributed Signal Processing for Security and Privacy in Networked Cyber-Physical Systems”. He is recipient of several distinguishing awards including the Eshrat Arjomandi Award for outstanding Ph.D. dissertation from Electrical Engineering and Computer Science Department, York University, in 2013; Concordia President’s Excellence in Teaching Award in 2018, and; 2019 Gina Cody School of Engineering and Computer Science’s Research and Teaching awards in the new scholar category.



**Amir Asif** received the M.Sc. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University (CMU), Pittsburgh, PA in 1993 and 1996, respectively. He is a professor of Electrical Engineering and Computer Science at York University, where he is serving as Vice President, Research and Innovation. Previously, he was a professor of electrical and computer engineering and Dean of Engineering and Computer Science at Concordia University from 2014 to 2020, professor of electrical en-

gineering and computer science at York University, Toronto, ON, Canada, from 2002 to 2014 and on the faculty of CMU, where he was a Research Engineer from 1997 to 1999. Asif works in the area of statistical signal processing and communications. His current projects include distributed agent networks (autonomy and consensus in complex and contested environments); medical imaging (ultrasound elastography, brain computer interfaces), data science (graph signal processing in social networks); and health management of mission critical systems. He has authored over 175 technical contributions, including invited ones, published in international journals and conference proceedings, and a textbook "Continuous and Discrete Time Signals and Systems" published by the Cambridge University Press. Asif has served on the editorial boards of numerous journals and international con-

ferences, including Associate Editor for IEEE Transactions of Signal Processing (2014-18), IEEE Signal Processing letters (2002-2006, 2009-2013). He has organized a number of IEEE conferences on signal processing theory and applications. He has received several distinguishing awards including the York University Faculty of Graduate Studies Teaching Award in 2008; York University Wide Teaching Award (Full-Time Senior Faculty Category) in 2008 from York University; the FSE Teaching Excellence Award (Senior Faculty Category) from York's Faculty of Science and Engineering in 2006 and in 2004; and the CSE Mildred Baptist Teaching Excellence Award from York's Department of Computer Science and Engineering in 2006 and in 2003. He is a member of the Professional Engineering Society of Ontario and a senior member of IEEE.