Open Access Journal

Journal of Sustainable Research in Engineering Vol. 5 (2) 2019, 50-61

Journal homepage: http://sri.jkuat.ac.ke/ojs/index.php/sri

ISSN (Online): 2409-1243

Health-Index Based Prognostics for a Turbofan Engine using Ensemble of Machine Learning Algorithms

Jackline M Mutunga[1], James K. Kimotho[2] and Peter Muchiri[3]

[1]Department of Mechanical Engineering, Pan African University Institute for Basic Sciences, Technology and Innovation P. O. BOX 62000-00200 Nairobi, Kenya

[2]Department of Mechanical Engineering, Jomo Kenyatta University of Agriculture and Technology P.O BOX 62000-00200, Nairobi, Kenya

[3]Department of Mechanical Engineering, Dedan Kimathi University of Technology P.O BOX 22992-00100, Nairobi, Kenya

[*] *Corresponding Author - E-mail: jacquiemwende@gmail.com*

Abstract A turbofan engine is a critical component of the aircraft, and monitoring its performance is important to avoid catastrophic failures and expensive downtime. Technologies in condition monitoring have made this possible by using sensors to collect data regarding fault propagation in systems. Machine Learning Algorithms (MLA) are useful tools for data analytics modeling. They use features from datasets to detect patterns and build predictive models. The predictive models are then used with new data, to determine the future reliability of a system by assessing the extent of degradation from its expected normal operating conditions. This in turn facilitating determination of the system's Remaining Useful Life (RUL). Several prognostics approaches have been proposed to predict RUL for complex systems. There is a need to further increase their accuracy and robustness, with the aim of increasing reliability. This can be achieved by use of ensemble techniques.

Ensemble of predicting models developed using different MLAs or models developed using similar datasets are some of the ensemble techniques used in RUL modeling. Their results have demonstrated to achieve better performance compared to single modeling. This work aims at further increasing the prediction accuracy and robustness by combining these two ensemble techniques. A case study based on the National Aeronautics and Space Administration (NASA) turbofan engine degradation simulation dataset FD001 is presented. Evaluation results demonstrate that the developed ensemble model had better performance having a score value of 115. This is in comparison to the best approach in literature using similar dataset, where modeling was done using a single MLA and a score value of 231 was achieved. This illustrates the superiority of the developed prognostics approach having a diverse strategy in developing the RUL predicting model.

Keywords Ensemble, Health Index, Machine Learning, Prognostics, Remaining Useful Lifetime.

1. Introduction

Maintenance involves all activities undertaken to retain or restore a system to a given functional condition [1]. Condition-based maintenance (CBM) is a strategic maintenance approach that monitors the performance of a system using sensors for data acquisition. Analysis of this data influences maintenance actions to be undertaken with the aim of maximizing system life. CBM has proved to be a more efficient and reliable approach compared to the traditional corrective and preventive approaches, taking the center stage in maintenance [2].

CBM can be categorized into prognostics and diagnostics. Diagnostics involves detection, isolation and identification of faults in systems. Prognostics predicts the Remaining Useful Life (RUL) given the current health state of a system. Predicted remaining useful life of a system is an important parameter in maintenance planning [3].

Prognostic methods can be categorized into; physics-based, data-driven and hybrid methods [4]. Physics-based prognostics combines a physical damage model with measured data to predict future behavior of degradation or damage, and to predict the RUL. The modeling parameters are correlated to the material properties and stress levels, which are generally identifed by using specific experiments, finite element analysis or other suitable techniques. The accuracy of this approach depends on prior knowledge of the physical behavior of a system. For complex systems, this is not always available, or it is too expensive to acquire limiting its application. Data-driven prognostics utilizes models developed exclusively from data. Training data is used to design and train a predictive model while testing data is used to validate the model [5]. Hybrid prognostics combines physics-based and data-driven prognostics to maximize on the advantages of both approaches while minimizing corresponding disadvantages. The use of physics-based and hybrid approaches is limiting in practice because of unavailability of underlying physical knowledge in many practical systems [6]. As a result, data-driven approaches are preferred for prognostics as they don't require expert knowledge of the system.

Data-driven prognostics is classified into two approaches based on the training target. The first approach uses the RUL as the training target hence resulting in direct RUL predictions. The second approach uses Health Index (HI) as the training target which upon prediction, the HI is mapped on the RUL.

Turbofan engine is a critical system of the aircraft. It is composed of various sub-systems such as the fan, the compressor, combustor among others, linked together. If no maintenance intervention is carried out, a turbofan engine gradually degrades until end of life. Reliable degradation assessment and RUL estimation make sense on both aviation safety and rational maintenance decisions being a critical part of the aircraft system [7]. The complexity of this system influences selection of data-driven approach for prognostics.

Ensemble is a machine learning paradigm where multiple learners are trained to solve the same problem and their outputs combined. It has demonstrated to improve prediction accuracy by combining multiple learning algorithms' outputs [8, 9]. Ensemble of the outputs from various models results in output with tighter uncertainty bounds than the average output of any individual model [10, 11]. It is important for the models used in ensembling to be diverse so that they can complement each other. In contrast to ordinary machine learning approaches that learn one hypothesis from training data, ensemble methods construct a set of hypotheses and combine them [12]. This is an appealing strategy as it boosts weak learners to strong learners that can make more accurate predictions.

Ensemble models can be developed either sequentially or using a parallel approach. Sequential ensembles have the base learners generated sequentially with the aim of exploiting their dependency. Parallel ensembles aim at exploiting the independence between the base learners. Mislabeling of the weights assigned to base learners in sequential ensembling could result in a vast error margin. As a result, parallel approach is preferred since the error can be reduced by averaging.

Ensemble techniques can also be categorized based on the level of implementation of the fusion methodology; feature-level fusion and decision-level fusion [13, 14]. Feature-level fusion integrates feature information that results from independent analysis methods. Prior knowledge about the degradation mechanism and physical laws is usually implemented to create desired features. Decision-level fusion aggregates the outputs from various models developed using MLAs either by simple averaging or weighted averaging. Simple averaging assigns equal weights to all the outputs while weighted averaging assigns weights to outputs depending on their relevance. Weighted averaging is considered more superior as the final average reflects the importance of each output hence more descriptive compared to simple averaging.

The rest of this paper is organized as follows. Section 2 introduces recent and related work on the turbofan engine degradation dataset. Section 3 entails the methodology for generating multiple base learners and the fusion strategies used. Evaluation metrics used are also presented. In Section 4, the results based on the NASA turbofan engine degradation simulation dataset FD001 [15] are discussed. Section 5 is the conclusion and areas for future work.

2. Related Work

The turbofan engine degradation simulation has been extensively used to evaluate several data driven prognostics approaches. This section reviews some of the recent studies applied on this dataset.

Neural networks have widely found application in modeling of complex systems. This is because they independently establish the relationship between the input and output datasets. Zheng et al. [16] proposed used of Long Short-Term Memory Network. The developed approach was able to reveal hidden patterns in the turbofan engine dataset and achieved higher accuracy in comparison to traditional Recurrent Neural Networks (RNN). Andre Listou et al. [17] proposed the semi-supervised deep architecture approach to predict RUL of turbofan engines. The proposed approach was compared to supervised ensemble of convolutional and feed forward neural networks. It had better accuracy associated with decrease in score value from 274 to 231. Lin et al. [18] ensembled random forest and Extreme Learning Machine (ELM) algorithms. The ensemble technique contributed to error margin reduction from a Root Mean Square Error (RMSE) of 8.64 when only ELM was used, to a RMSE of 6.89.

Wang et al. [19] proposed the use of a linear regression model for calculating the HI from multi-dimensional sensor readings, and a similarity-based prognostics approach was then used to estimate the RUL. Riad et al. [20] improved on this by applying linear regression followed by a smoothing process with third-order polynomial curve fitting in calculating the HI. The calculated HI was used as the input to a multi-layer perceptron neural network to predict the RUL. The smoothing process improved the prediction accuracy. Yang et al. [6] further improved on this by combining outputs from various Neural Network models to demonstrate effectiveness of ensembling. Direct RUL prediction had a RMSE of 4.12 compared to 2.74 where HI-based prognostics was applied. This demonstrates that using HI as the target when training increases the accuracy of the predictions compared to direct RUL prediction.

Khelif et al. [21] modeled using SVM and obtained a competitive score value of 448 compared to neural networks that had a score value of 1046. This is attributed to the high classification accuracy when using a kernel function in modeling.

Li et al. [22] also demonstrated the effectiveness of ensemble technique in modeling. The MLAs ensembled included; random forests, classification and regression tree, RNN, autoregressive model, adaptive network-based fuzzy inference systems, relevance vector machine and elastic net. Particle swarm optimization and sequential quadratic optimization approaches were used to determine the optimal weights for each base learner. Model validation was done using dataset FD004. A score value of 26.382 was achieved representing high reliability of the model in RUL prediction.

Support Vector Machine and Extreme Learning Machine algorithms are used to develop the proposed ensemble model. This is because of their high classification accuracy and good generalization performance respectively. Their working principles are discussed below.

1) Extreme Learning Machine (ELM)

Extreme learning machine is a single-hidden layer feedforward neural network first proposed by Huang [23]. Its learning process is faster compared to other feedforward neural networks. Fig.1 illustrates the structure of ELM. It consists of input neurons, hidden neurons and output neurons. The input vector x_j is fed to the input neuron j, which links it to the hidden neuron l with an input weight a_l. The corresponding hidden neuron l is linked to the output neuron by an output weight β_l. The input weights and the hidden layer biases b_l are randomly generated while the output weight is obtained by a generalized inverse operation of the hidden layer output matrix [23].

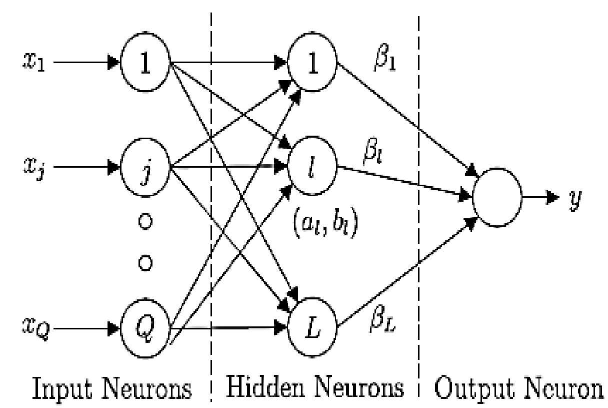

Fig. 1. Structure of ELM [23]

2) Support Vector Machine (SVM)

Support vector machine is a supervised MLA that solves binary data classification problems by finding a hyperplane that separates the data into classes [24]. Support vectors lie on the bounding planes which are parallel to the hyperplane. SVM aims at maximizing the distance between the two bounding planes by minimizing the vector orthogonal to the hyperplane [25].

A kernel function is used to map features into a higher dimensional feature space [26]. This allows for construction of a hyperplane in the higher dimensional feature space without explicitly performing calculations in the feature space for data that is not linearly separable [26]. The Radial basis function (RBF), a kernel function used for this research is given by Equation 5.

$$K(x_i, x_j) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right) \quad (1)$$

where γ is a tunable kernel parameter. C, a regularization parameter that controls trade-offs between maximizing the margin and minimizing misclassification is also tuned during training to provide better regularization. Large values of C lead to overfitting while small values lead to misclassification [26]. Optimization of these parameters is crucial on the accuracy of the predicted RUL.

Extensive research has been done on methods of estimating RUL of turbofan engines. However, the aspect of combining various ensembling techniques has not been fully explored. Modeling using more than one ensemble technique diversifies the developed model increasing its accuracy. This paper focuses on investigating the effect of modeling using two ensemble techniques; ensemble based on similar datasets with varying initial wear conditions and ensemble based on different machine learning algorithms. HI-based RUL prediction technique discussed in Section III is used.

3. Methodology

This section presents the methodology used as illustrated in Fig. 2. It contains two major phases; training phase and the testing phase. During the training phase, the raw data signals were processed to extract useful information regarding fault propagation. The machine learning algorithm then learned the relationship between the processed training data and the health index, defined as the target. As a result, a predictive model was developed. The testing or online prognostics phase involved use of processed test dataset as the input to the developed predictive model. The predicted health index was then mapped to RUL by extrapolation to the predefined failure threshold. Predictions from various models were then ensembled to obtain the final predicted RUL.

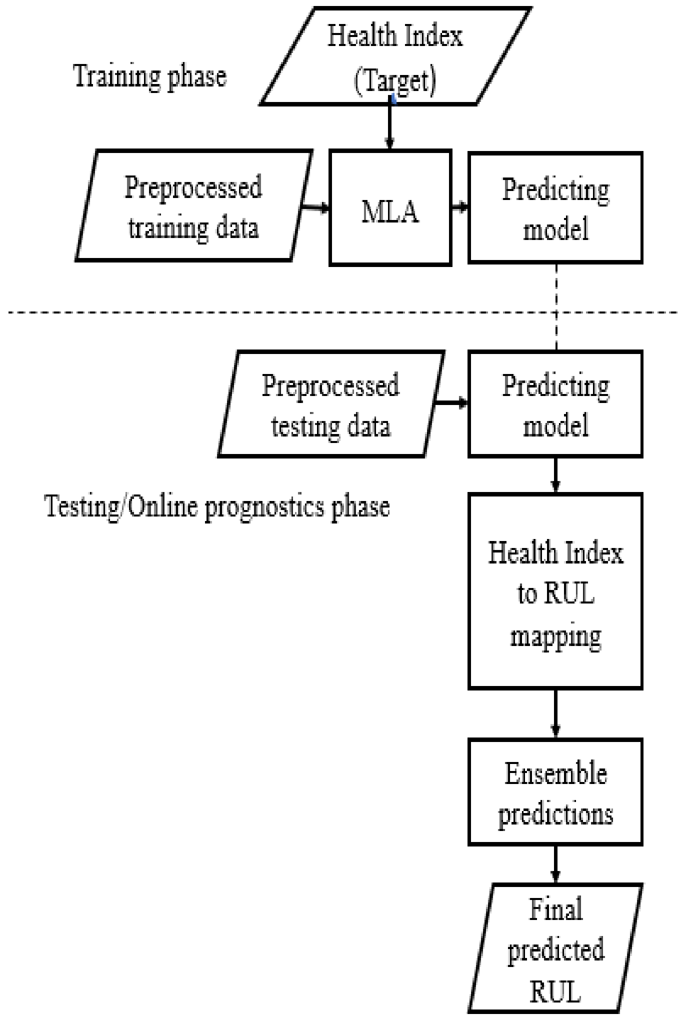

Fig. 2. Methodology implementation procedure

3.1 System and Data Description

This paper considered dataset FD001 that describes the degradation of a simulated turbofan engine monitored using multiple sensors. The simulation model of the turbofan engine was developed using Commercial Modular Aero-Propulsion System Simulation (CMAPSS), a simulation tool developed at NASA and widely used in engine health monitoring research for simulating realistic large commercial turbofan engines [15].

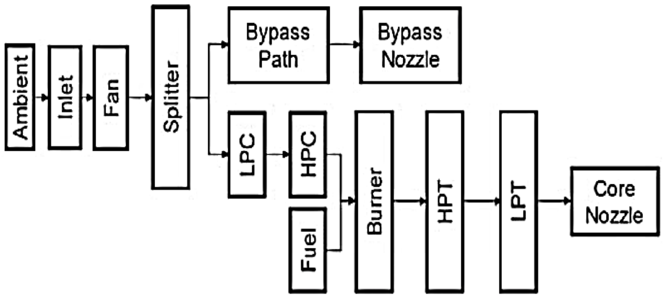

Fig. 3. Subroutines of a model and their interconnections for the turbofan engine simulated in CMAPSS [15]

A model of a 90,000 lb thrust engine [15] illustrated in Fig. 3, was developed and simulations run for operations at altitude range of between 0 and 42 000 ft., Mach number (speed) range of 0 to 0.84, and Throttle Resolver Angle (TRA) range of 20 to 100. The CMAPSS simulation model was embedded in MATLAB Simulink. The software had 14 inputs and generated 21 outputs that were available for analysis.

The simulation model was specifically used to characterize degradation in engine performance due to wear and tear based on the usage pattern of the engines. Unknown variance in the initial level of wear and random noise were introduced to represent system variability as is the case in real life application. Each engine started with different degrees of initial wear and manufacturing variation, which was unknown to the user, and was in operation until the failure threshold was reached. The failure threshold was used to define the end-of-life beyond which the unit was considered to have failed [15].

Dataset FD001 consisted of 100 training units, 100 testing units, and a file recording the actual RUL of the 100 testing units. Each training unit was run to failure, while the testing unit was stopped at some random point prior to its failure. Every engine had 21 sensors collecting different measurements related to the engine state at runtime with the data being of the time series nature as presented in Table I. At the start of the time series, the engine operated normally but after certain number of cycles, a fault developed in the engine which then gradually failed.

Table I
PMH08 challenge dataset parameters available to participants as sensor data [27]

Parameters	Symbol	Description	Unit
Unit	—	—	—
Time	—	—	t
Setting 1	—	Altitude	ft
Setting 2	—	Mach Number	M
Setting 3	—	Sea-level Temperature	°F
Sensor 1	T2	Total temperature at fan inlet	°R
Sensor 2	T24	Total temperature at LPC outlet	°R
Sensor 3	T30	Total temperature at HPC outlet	°R
Sensor 4	T50	Total temperature at LPT outlet	°R
Sensor 5	P2	Pressure at fan inlet	psia
Sensor 6	P15	Total pressure in bypass-duct	psia
Sensor 7	P30	Total pressure at HPC outlet	psia
Sensor 8	Nf	Physical fan speed	rpm
Sensor 9	Nc	Physical core speed	rpm
Sensor 10	epr	Engine pressure ratio	—
Sensor 11	Ps30	Static pressure at HPC outlet	psia
Sensor 12	phi	Ratio of fuel flow to Ps30	pps/psi
Sensor 13	NRf	Corrected fan speed	rpm
Sensor 14	NRc	Corrected core speed	rpm
Sensor 15	BPR	Bypass Ratio	—
Sensor 16	farB	Burner fuel-air ratio	—
Sensor 17	htBleed	Bleed Enthalpy	—
Sensor 18	Nf_dmd	Demanded fan speed	rpm
Sensor 19	PCNfR_dmd	Demanded corrected fan speed	rpm
Sensor 20	W31	HPT coolant bleed	lbm/s
Sensor 21	W32	LPT coolant bleed	lbm/s

LPC/HPC=Low/High Pressure Compressor - LPT/HPT= Low/High Pressure Turbine

3.2 Data Processing

Data processing techniques used included; data normalization, data smoothing and feature selection. The z-score of normalization, Equation 2, was adopted as it handles outliers well. Data normalization increased consistency and made the process of mapping inputs to the target during modeling more efficient.

$$z = \frac{x - \mu}{\sigma} \tag{2}$$

where x is the dataset before normalization, μ and σ are the mean and standard deviation for the corresponding variable respectively, and z is the normalized dataset.

Data smoothing was done to remove noise from the data allowing important patterns/trends to be revealed as illustrated in Fig.4. Local Regression Smoothing technique was used being more robust compared to the moving average approach.

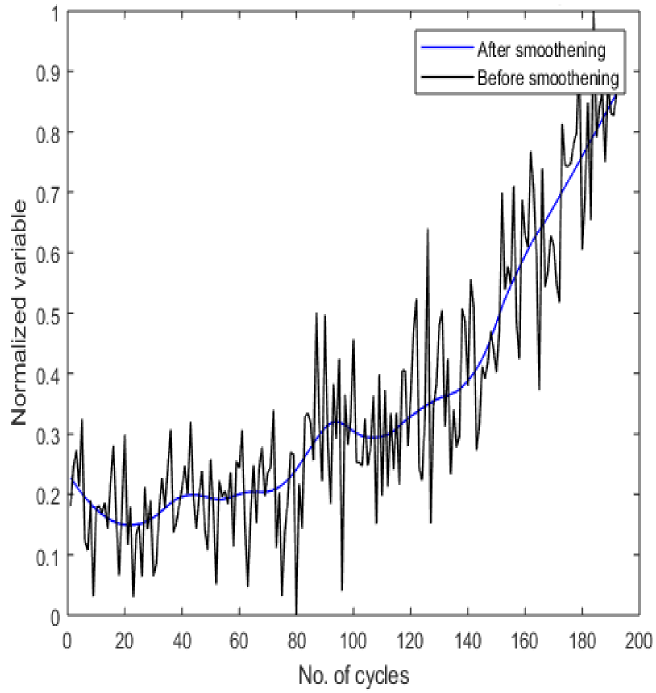

Fig. 4. Effect of data smoothing

Feature selection was done to identify a smaller subset from the main feature set. This aimed to reduce redundancy. A monotonicity function, Equation 3, was used to remove any features that had a low monotonic value as they did not represent a clear trend of fault propagation [28].

$$M = \left| \frac{no.of\ \frac{dx}{dt} > 0 - no.of\ \frac{dx}{dt} < 0}{n-1} \right| \qquad (3)$$

where n is the number of observations in a feature and dx/dt is the derivative of the feature variables with respect to the cycles for the engine. The value of M ranged between 0 and 1 with M=1 representing highly monotonic features and M=0 representing non-monotonic features.

From the given training datasets with 26 variables, those selected for training had a monotonic value greater than 0.25. The time variable together with sensors 2,3,4,7,8,9,11,12,13,14,15,17,20 and 21 where used as the training dataset. Fig. 5 compares the degradation trends of variables with high and low monotonic values.

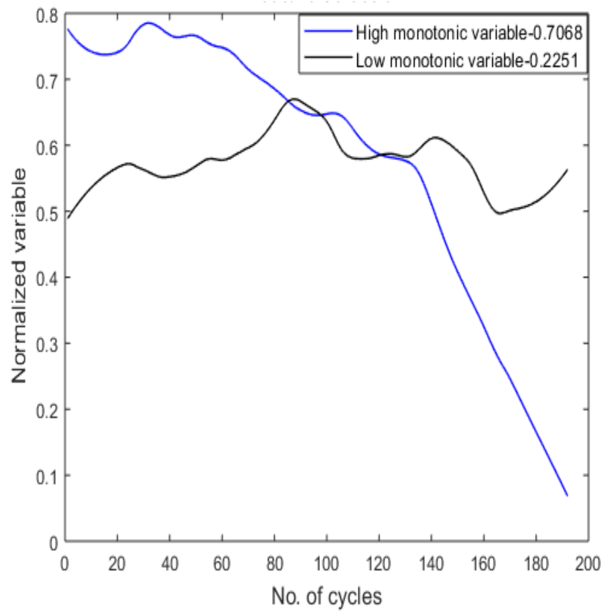

Fig. 5. Variance in monotonic value

3.3 RUL Modelling and Ensemble

Support Vector Machine and Extreme Machine Learning Algorithm were used for modeling. This is based on their high classification accuracy and good generalization capabilities respectively, upon review. Parameter optimization of the MLAs was done to yield a minimized generalization error. This maximized the margin over the hyperplane coefficients and minimized an estimate of the generalization error over the set of kernel parameters. Upon implementation, the optimal values of C and γ where found to be 1 and 8 respectively.

3.4 RUL Predicting Technique

A comparison of the direct RUL predicting technique and the HI-based RUL predicting technique was done to select the best performing technique. When modeling using the direct RUL predicting technique, the MLA learned the relationship between the processed training data and the RUL used as the target. The output from the model when test dataset was used as input to the model was the predicted RUL as illustrated in Fig. 6.

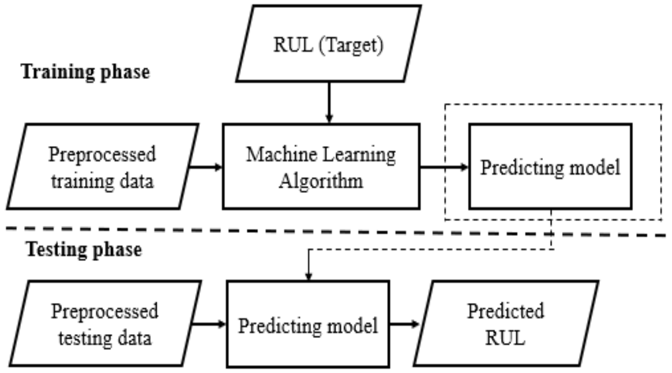

Fig. 6. Direct RUL prediction technique

HI-based RUL predicting technique, illustrated in Fig 7, used the health index as the target during the training phase. During the testing phase, the predicted HI was mapped to the corresponding RUL to establish its predicted end of life.

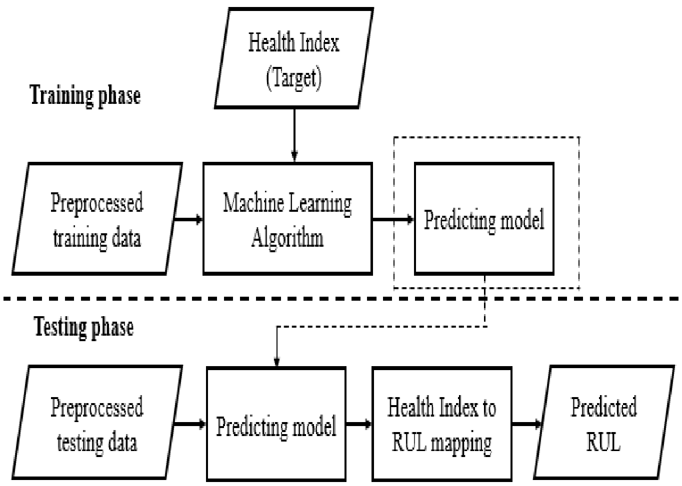

Fig. 7. HI-based RUL prediction technique

The HI was defined as exponential since most systems experience exponential decay to represent system degradation as shown in Fig. 8. The exponential curve fitting tool in MATLAB was employed to define the propagation of the HI. The generic form of an exponential model for 1-D data is given by Equation 4.

$$z = a.\exp(b.t + c) \tag{4}$$

where a, b and c are parameters learned during exponential curve fitting.

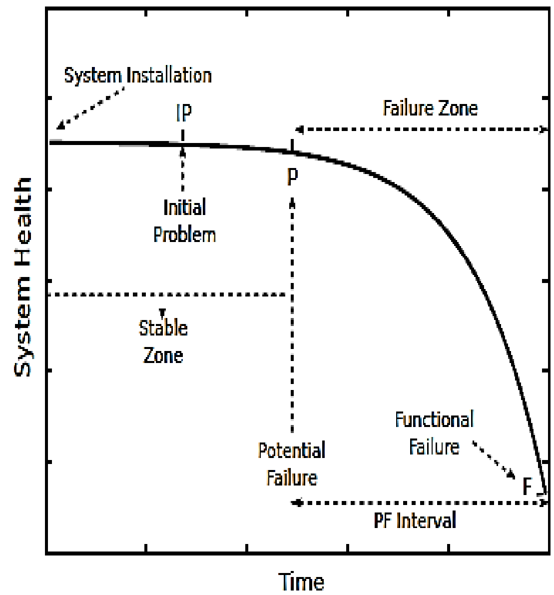

Fig. 8. System life model [29]

AutoRegressive modeling (AR) was used to map the predicted HI to the corresponding RUL. Given that the test dataset was truncated, it did not get to the set threshold where the engine was considered to have reached end of life. The modeling sequence represented by Equation 5 was established to each engine's predicted HI values to obtain the degradation trend.

$$\mathbf{HI_i} = \sum_{\mathbf{k=1}}^{\mathbf{m}} \mathbf{a_k x_{i-k}} + \mathbf{e_i} \quad \mathbf{i = 1, 2 \ldots n} \tag{5}$$

where a_k are the model parameters, m is the model order, e_i is the residual of the model and n is the number of data points in HI. The AR model parameters for a sample unit were determined using the Yule-Walker method [30] and results presented in Table II. Model order 2 was selected as the optimal order since the residual did not change significantly when the model order was increased afterwards.

TABLE II
AR model parameters for a sample unit

Model order	**Model parameters**	**Model residual**
1	[1, -0.98]	-0.9813
2	**[1, -0.98,9.11 e-15, 2.22e-05]**	**2.18e-05**
3	[1, -0.98,9.11 e-15, 2.22e-05]	2.22e-05
4	[1, -0.98,9.11 e-15, -2.68e-15, 2.27e-15]	2.26e-05
5	[1, -0.98,9.11 e-15, -2.68e-15, 2.46e-15,2.31e-05]	2.31e-05

The first section of Fig. 9 represents the predicted HI, with the second section being the corresponding extrapolation to the set threshold of 1. The RUL was then defined as the difference in number of cycles between the current time t_c and the time at end of life of the engine t_{EOL}.

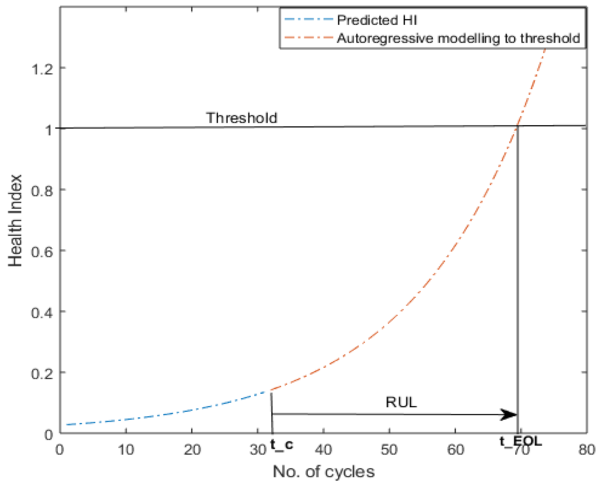

Fig. 9. Autoregressive modeling for turbofan engine

3.5 Ensemble Techniques

Ensembling was done to aggregate outputs of various base learner models. Where some predicting models had early predictions and others late predictions, this strategy optimized the RUL predictions. Ensemble based on similar datasets with varying initial wear conditions was implement first for each modeling MLA. Further ensemble based on different MLAs was then done.

Fig. 10 illustrates ensembling technique based on similar datasets with varying initial wear conditions. Since the training dataset had 100 similar units (engines), 100 predicting models were developed using a common MLA. During the testing phase, the test data for each unit was used as input to each of the developed models and the outputs ensembled. Simple averaging was used to obtain the final predicted RUL for this ensemble strategy

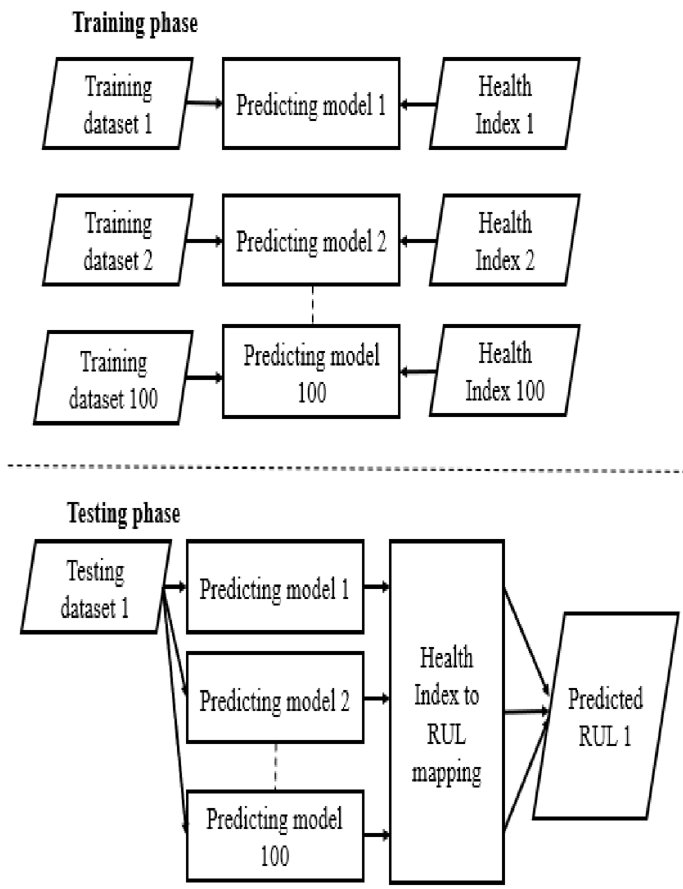

Fig. 10. Ensemble based on similar datasets with varying initial wear conditions

Further ensembling based on different MLAs illustrated in Fig. 11 used weighted averaging described by Equation 6, to aggregate the outputs.

$$RUL_{ens} = \frac{\sum_{i=1}^{n} w_i \; RUL_i}{\sum_{i=1}^{n} w_i} \tag{6}$$

where RUL_i is the RUL estimated by method i, w_i is the weight assigned to method *i* and *n* is the number of MLAs.

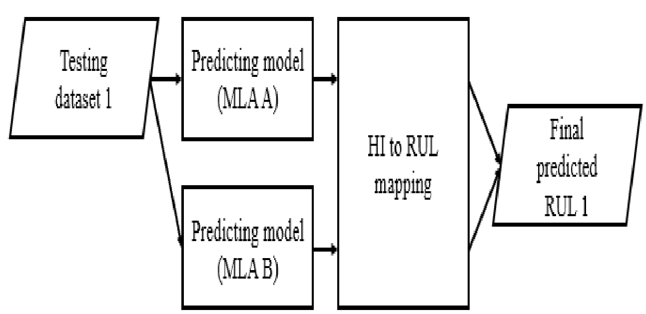

Fig. 11. Ensemble based on different MLAs

3.4 Performance Evaluation

The prognostics metrics selected for performance evaluation of the developed model were Mean Absolute Error, Mean Squared Error and the Score function. This was based on metrics used by other researchers to allow for ease of comparison of obtained results.

3.4.1 Mean Absolute Error (MAE)

For this research scope, error was defined as the difference between the Actual Time to Failure (ATTF) and Estimated Time to Failure (ETTF).

$$E = ATTF - ETTF \tag{7}$$

Absolute error;

$$|E| = |ATTF - ETTF| \tag{8}$$

The Mean Absolute Error (MAE) was calculated as the average of the absolute error terms for all the 100 units under consideration.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|E| \tag{9}$$

3.4.2 Mean Square Error (MSE)

With the ATTF and ETTF, the MSE was calculated using Equation 10. A higher MSE score represented a larger average prediction error.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}E^2 \tag{10}$$

3.4.3 Score function ([27])

The score function is a metric score of estimated calculations and a weighted sum of RUL errors. In either an early or late case, the penalty grows exponentially with increasing error. The asymmetric preference is controlled by user-defined acceptable early and late parameters, a_1 and a_2, in the scoring function described below.

$$S = s_1 + s_2 \tag{11}$$

$$s_1(E < 0) = \sum_{i=1}^{n} e^{-\left(\frac{E}{a_1}\right)} \tag{12}$$

$$s_2(E \geq 0) = \sum_{i=1}^{n} e^{\left(\frac{E}{a_2}\right)} \tag{13}$$

where S is the computed score, n is the number of predicted units, and E is the error term. In this work, values of a_1 and a_2 being 10 and 13 respectively were adopted based on the PMH guidelines so that late predictions are penalized more compared to early predictions.

4. Results and Discussion

For consistent comparison of results, the different MLAs were evaluated in the same way using the metrics presented in Section 3, on dataset FD001. For each RUL prediction, the results were compared with the actual RUL values provided. Further comparison was done with results obtained from other proposed prognostics approaches using the same dataset.

From Table III, the use of HI-based RUL prediction had better performance compared to the direct RUL predicting technique. When modeling using ELM, HI-based technique had a lower MAE of 31.16 compared to 44.59 when direct RUL prediction technique was used. The MSE reduced by 50% when modeling using ELM. The score value of SVM modeling also reduced from 1.0e+04 to 951. This is attributed to its normalized nature that was able to account for varying lifetimes of various engines during modeling.

TABLE III
Performance comparison of RUL predicting techniques

Performance metric	**MLA**	**Direct RUL predicting technique**	**HI-based RUL predicting technique**
MAE	SVM	30.84	28.65
	ELM	44.59	31.16
MSE	SVM	1449	1436
	ELM	3182	1603
Score function	SVM	1.0e+04	951
	ELM	1.06e+06	558

Ensemble based on similar dataset with varying initial wear condition had better performance compared to single modeling as presented in Table IV. For modeling using SVM, the MSE had a 47% decrease with the score value also reducing by 83%. This ensemble technique was able to model various degrading trends associated with the varying initial wear conditions resulting in increased accuracy of the RUL predictions.

For further ensembling based on different MLAs, the score function was used to assign weights to the predicting models having a better description of effects of early and late predictions. The model developed using ELM was assigned weight corresponding to 55% while the predicting model developed using SVM had a weight equivalent to 45%. The MAE and MSE values increased with comparisons to results obtained when only SVM was used. However, the final score value obtained of 115 was lower compared to that of either SVM or ELM. This increase in accuracy is attributed to the combined high classification accuracy of SVM and good generalization capabilities of ELM.

TABLE IV
Effect of ensemble technique based on HI-based RUL predicting technique

Performance metric	MLA	Single predicting model	Ensemble based on similar datasets	Further ensemble based on different MLAs
MAE	SVM	28.65	21.47	**22.39**
	ELM	31.16	23.81	
MSE	SVM	1436	762	**832**
	ELM	1603	939	
Score function	SVM	951	155	**115**
	ELM	558	129	

Ensemble based on different MLA aimed at maximizing the benefits associated with both MLAs. Fig. 12 represents an instance where SVM had better performance than ELM. This is attributed to the high classification accuracy of SVM outperforming the good generalization capabilities of ELM. The vice versa is presented in Fig. 13. Modeling using an ensemble of the two models demonstrated an accuracy of 100% for engine 64 as presented in Fig. 14. The high accuracy levels ensure effective maintenance strategy are adopted to avoid the consequences associated with unplanned engine failure.

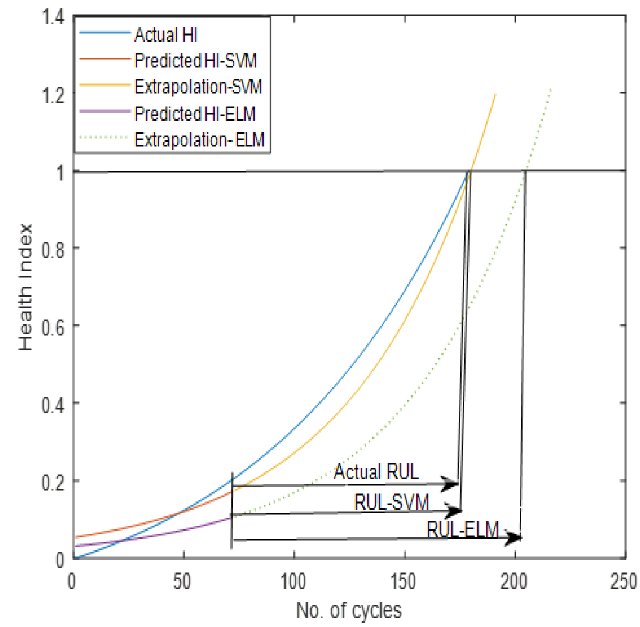

Fig. 12. RUL prediction for engine 78

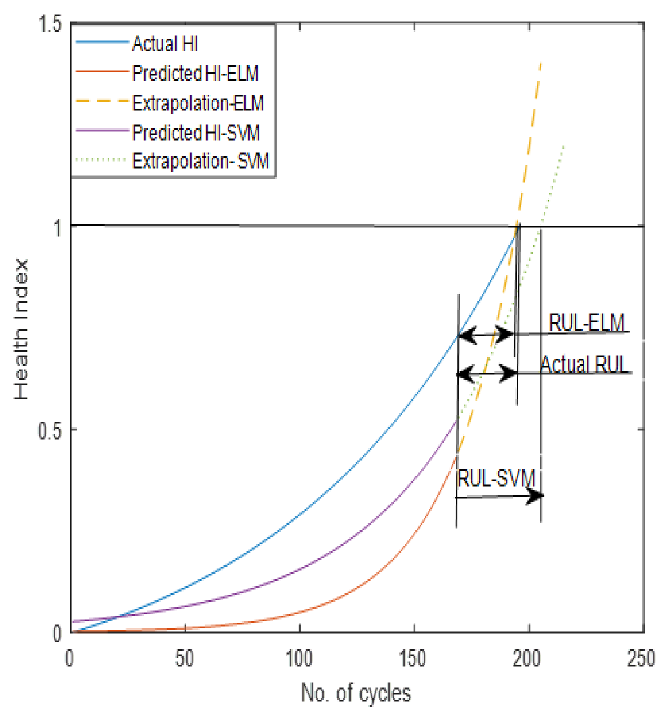

Fig. 13. RUL prediction for engine 64

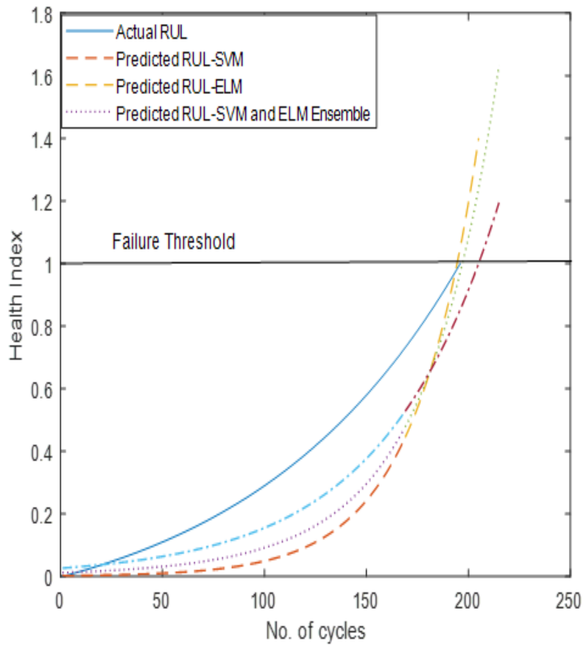

Fig. 14. Ensemble modeling for engine 64

TABLE V
Benchmarking with other approaches based on dataset FD001

MLA	Score value
ELM [31]	267
Light Gradient Boost [32]	250
Semi-supervised Deep Architecture [17]	231
Ensemble of SVM and ELM	**115**

Comparison of obtained results with the current best performing prognostics approaches in literature using the same dataset is presented in Table V. These approaches used a single MLA. Results obtained from the developed model for individual MLAs still outperformed these approaches. ELM and SVM had score values of 129 and 155 respectively. This is attributed to the high accuracy levels associated with ensemble based on similar datasets with varying initial wear conditions. Combining two ensemble approaches demonstrated to have even higher accuracy with a score value of 115 with comparison to the other approaches.

5. Conclusion

In this paper, the superiority of using ensemble approach at two stages; ensemble based on similar datasets with varying initial wear conditions and ensemble based on different MLAs has been demonstrated. The developed model was implemented using NASA turbofan engine dataset FD001. The results indicate an increase in accuracy compared to other proposed prognostics approaches in literature. The obtained score value of 115 is lower compared to other reviewed approaches with score values above 200, an indication of increased accuracy in RUL prediction. Future research will explore optimization of the number of models ensembled based on similar datasets with the aim of reducing computational load. The feasibility and effectiveness of the developed model to other applications will also be analyzed.

References

[1] J. Moubray, Reliability-centered Maintenance, New York: Industrial Press Inc., 1997.

[2] Z. Zhao, L. Bin, X. Wang and W. Lu, "Remaining useful life prediction of aircraft engine based on degradation pattern learning," *Reliability Engineering and System Safety, Elsevier,* vol. 164, pp. 74-83, 2017.

[3] M. Pecht , "Prognostics and Health Management of Electronics," in *Encyclopedia of Structural Health Monitoring*, John Wiley & Sons, Ltd, 2009.

[4] L. Liao and F. Kottig, "Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction.," in *IEEE Trans Reliab*, 2014.

[5] Z. Li, D. Wu, C. Hu, and J. Terpenny, "An ensemble learning-based prognostic approach with degredation-dependent weights for remaining useful life prediction," in *Reliability Engineering and Sysytem Safety*, 2018.

[6] Y. Feng, M. S. Habibullah, T. Zhang and Z. Xu, "Health Index-Based Prognostics for Remaining Useful Life Predictions in Electrical Machines," *IEEE Transactions on Industrial Electronics*, 2016.

[7] J. Shi., G. Wang, Y. Li, and X. Li, "Health index synthetization and remaining useful life estimation for turbofan engines based on run-to-failure datasets," *Maintenance and Reliability*, vol. 18(4), pp. 621-631, 2016.

[8] C. Hu, B. Youn, P. Wang , and J. Yoon , "Ensemble of data-driven prognostic algorithms for robust prediction of remianing useful life," *Reliability Engineering and System Safety, Elsevier*, pp. 120-135, 2012.

[9] J. M. Bates and C. W. J. Granger, "The combination of Forecasts," *Operational Research Society*, 2014.

[10] J. Sun, Z. Hongfu, Y. Haibin, and M. Pecht, "Study of Ensemble Learning-Based Fusion Prognostics," in *Prognostics and Health Management Conference*, 2010.

[11] M. P. Perrone ,and L. N. Cooper, "When Networks Disagree: Ensemble methods for hybrid Neural Networks," R.J. Mammone, 1993.

[12] J. Liu, V. Vitelli, R. Seraoui, and E. Zio, "Dynamic Weighted PSVR-Based Ensembles for Prognostics of Nuclear Components," in *European Conference of the prognostics and Health Managment Society*, 2014.

[13] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, pp. 6-23, 1997.

[14] A. J. Volponi, T. Brotherton and R. Luppold, "Development of an Information Fusion System for Engine Diagnostics and Health Management," in *AIAA 1st Intelligent Systems Technical Conference*, Chicago, Illinois, 2004.

[15] A. Saxena, K. Goebel, D. Simon and N. Eklund, "Damage Propagation Modeling for Aircraft Engine Run-to-failure Simulation," in *IEEE PHM'08 conference*, 2008.

[16] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long Short-Term Memory Network for Remaining Useful Life estimation," *IEEE International Conference on Prognostics and Health Management,* pp. 88-95, 2017.

[17] A. Listou Ellefsen, E. Bjørlykhaug, V. Æsøy, S. Ushakov and H. Zhang , "Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture," *Reliability Engineering and System Safety,* vol. 183, pp. 240-251, 2019.

[18] L. Lin, F. , Fang, X. Xie, and S. Zhong, "Random forests-based extreme learning machine ensemble for multi-regime time series prediction," *Expert Systems with Applications,* vol. 83, pp. 164-176, 2017.

[19] T. Wang, J. Yu, D. Sieggel and J. Lee, "A similarity-based prognostics approach for remaining useful life estimation of engineered systems," in *International Conference on Prognostics and Health Management*, 2008.

[20] A. Riad and H. Elminir, "Evaluation of Neural Networks in the subject of prognostics as compared to linear regression model," *Int. J. Eng. Technology,* 2010.

[21] R. Khelif, B. Chebel-Morello, S. Malinowski, E. Laajili, F. Fnaiech and N. Zerhouni, "Direct Remaining Useful Life Estimation Based on Support Vector Regression," *IEEE Transactions on Industrial Electronics,* vol. 64, no. 3, pp. 2276-2285, 2017.

[22] Z. Li, K. Goebel, and D. Wu, "Degradation Modeling and Remaining Useful Life Prediction of Aircraft Engines Using Ensemble Learning," *Journal of Engineering for Gas Turbines and Power,* vol. 141, no. 4, 2019.

[23] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Science Direct,* vol. 70, p. 489–501, 2006.

[24] C. Cortes & V. Vapnik, "Support vector networks," *Machine Learning,* vol. 20, p. 273, 1995.

[25] C.-C. Lin and C.-J. Chang , "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology,* vol. 2, 2013.

[26] B. F. Buxton and R. Burbidge , "An introduction to support vector machines for data mining," *Operational Research Society,* pp. 3-15, 2001.

[27] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, "Metrics for Evaluating Performance of Prognostic Techniques," 2008.

[28] J. Coble and J.W.Hines, "Identifying optimal prognostics parameters from data: A genetic algorithms approach.," in *Annual Conference of the Prognostics and Health Management Society*, 2009.

[29] K. Lewis, "Technical note - tn 016: 2017 maintenance requirements," Asset Standards Authority (ASA), Australia, 2017.

[30] J. Durbin, "The fitting of time series in models," *Review of the International Statistical Institute*, vol. 28, pp. 233-243, 1960.

[31] C. Zheng, W.Liu, B. Chen, D. Gao, Y. Cheng and J. Peng, "A Data-driven Approach for Remaining Useful Life Prediction of Aircraft Engines," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2018.

[32] F. Li, L. Zhang, B. Chen, D. Gao, Z. Huang and J. Peng, "A Light Gradient Boosting Machine for Remaining Useful Life Estimation of Aircraft Engines," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2018.

[33] C. Zhang, P. Lim , A. Qin, and K. Tan, "Multiobjective deep belief networks ensemble for remianing useful life estimation in prognostics.," *IEEE Transactions on Neural Networks and Learning Systems,* pp. 1-13, 2016.

[34] C. Croarkin and P. Tobias, NIST/SEMATECH e-Handbook of Statistical Methods, 2013.

[35] V. Mathew, T. Toby, V. Singh, B. Maheswara, and M. Kumar, "Prediction of remaining useful lifetime of turbofan engine using machine learning," in *IEEE International conference on circuits and systems*, 2017.

[36] O. Chapelle, V. Vapnik, O. Bousquet and S. Mukherjee, "Choosing Multiple Parameters for Support Vector Machines," *Machine Learning,* 2002.

[37] Y. Zhang, H. Zuo, and F. Bai, "Classification of fault location and performance degradation of a roller bearing," *Measurement,* vol. 46, p. 1178–1189, 2013.

[38] C. Chen, B. Zhang, G. Vachtsevanos and M. Orchard, "Machine Condition Prediction Based on Adaptive NeuroFuzzy and High-Order Particle Filtering," *IEEE Trans. Ind. Electron.,* vol. 58, 2011.

[39] C. Zhang, P. Lim, A. K. Qin and K. C. Tan, "Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics," *IEEE Transactions on Neural Networks and Learning Systems*, 2016.

[40] L. Lin, F. Wang, X. Xie and S. Zhong, "Random forests-based extreme learning machine ensemble for multi-regime time series prediction," *Expert Systems with Applications,* vol. 83, pp. 164-176, 2017.