Estimating the Remaining Useful Lifetime of a Turbofan Engine using Ensemble of Machine Learning Algorithms

Jackline Mwende Mutunga[1*], Dr-Ing James Kimotho[2] and Prof. Peter Muchiri[3]

Abstract—Predicting the Remaining Useful Lifetime (RUL) of a component or system is important for effective and efficient maintenance. Prognostics approaches, used in predicting the future reliability of a system by assessing the extent of degradation of the product from its expected normal operating conditions, can be classified into physics-based and data driven. The later has received huge attention from researchers as it does not require expertise knowledge of the system at hand. Ensemble technique, associated with aggregation of predictions produced by multiple learning algorithms to improve predictive performance, robustness and accuracy in prognostics is the area of interest for this research that ensembles various selected regression Machine Learning Algorithms (MLA). The effect of ensembling various MLAs and MLAs models built using similar data is presented and a comparative study done with using only a single model. A case study of a NASA turbofan engine degradation simulation data set is presented. Simple averaging approach is proposed in combining the output of different sub models both at the training and predictive stages. Of the selected MLAs, ensemble regression, the best performing, had a Mean Absolute Error (MAE) of 39.63 for a single model compared to 22.91 for an ensemble of various sub models. The numerical results indicate that the ensemble approach outperforms use of individual machine learning models.

Keywords— Ensemble technique, Machine Learning, Prognostics, Remaining Useful Lifetime.

.

I. INTRODUCTION

Maintenance is a very important aspect in most industries such as aerospace and manufacturing industries since unplanned breakdown could result in disasters and accidents with extensive losses, expensive downtime and high financial costs. Traditional approaches such as corrective and scheduled preventive maintenance are increasingly being replaced by condition-based maintenance (CBM) that is more efficient and reliable [1]. Reducing maintenance costs, maximizing operational availability and improving the system's reliability and safety are the major objectives of CBM.

Prognostics, the process of predicting the future reliability of a system by assessing the extent of degradation from its expected normal operating conditions has taken center stage in CBM by facilitating estimation of remaining useful lifetime of a system which is an important parameter in maintenance planning [2].

Turbofan engines gradually degrade until failure occurs or end of lifetime if no maintenance has been carried out. Reliable degradation assessment and RUL estimation make sense on both aviation safety and rational maintenance decisions being a critical part of the entire aircraft system[3].

Generally, prognostic methods can be categorized into: model-based, data-driven, and hybrid methods [4]. Model-based prognostics use models derived from first principles. The parameters are correlated to the material properties and stress levels, which are generally identified by using specific experiments, finite element analysis or other suitable techniques. Orchard et al. [5] used the particle filter which is a method that leverages degradation models based on the first principle of underlying degradation mechanisms. The accuracy of this approach depends on prior knowledge of physical behavior of the system. For complex systems, this is not always available, or it is too expensive to acquire limiting its application. Data-driven prognostics on the other hand utilizes models learned exclusively from data. Training data is used to design and train a predictive model while testing data is used to validate the predictive model [6]. Hybrid prognostics combines model-based and data-driven prognostics to maximize the advantages of both approaches while minimizing corresponding disadvantages.

Ensemble prognostics has demonstrated to improve prediction accuracy by combining multiple learning algorithms [7]. It is also associated with reduced uncertainty from data noise or model errors. Given that the test data length is short compared to the training data and long-term projections are necessary, the prognostic accuracy may fail to provide desirable results. Theory research and practice have indicated that combination forecast has better precision than individual forecast [8]. It is important for the models being ensembled to be diverse so that they can complement each other. Ensemble of the output from various models results in a more accurate output with tighter uncertainty bounds than the average output of any individual model alone [9]

Ensemble learning aggregates predictions produced by multiple learning algorithms to improve predictive performance [10] [11]. The approach trains diverse sub-models then merge the output with various strategies. These sub-models are built on three key components: a strategy to build diverse models; a strategy to construct accurate sub-models; a strategy to combine the outputs of the sub-models in a way such that the correct predictions are amplified, while incorrect ones are counteracted [12]. The strategies used in practice to create models could either be training the

J.M. Mutunga, Department of Mechanical Engineering, PAUSTI (+254O702217151; e-mail: jacquiemwende@gmail.com).
Dr.-Ing J.K. Kimotho, Department of Mechanical Engineering, JKUAT (e-mail: jkuria@eng. jkuat.ac.ke).

individual weak learner with different data sets or use of different model types or training methods [13].
Zhang et al. [14] applied ensemble learning method for predicting RUL of rolling bearings by combining two artificial neural networks using a simple weight-vector. Gaussian process regression was also combined with similarity-based regression to demonstrate increased prediction accuracy in employing ensemble approach [15]. Wu et al. [16] introduced a random forests-based prognostic approach to train a predictive model by aggregating a collection of regression trees and whose effectiveness was demonstrated using tool wear prediction. Kalman filter was used to combine prediction results produced by different Artificial Neural Networks by [17] using the prognostic data sets for the 2008 IEEE PHM Data Challenge Competition that demonstrated better performance compared to using each artificial neural network model. Vimala et al. [18] reviewed some of the best performing regression machine learning algorithms using the NASA turbofan dataset. Gaussian process regression and linear regression are some of the best performing algorithms amongst those reviewed.
In this paper, an ensemble learning-based fusion prognostics method is presented. Ensemble technique is applied at both the training and prediction stages using various similar datasets and various regression MLAs. Related research has focused only one ensemble strategy and not both to the best of our knowledge.
The rest of this paper is organized as follows. In section II describes the general methodology for generating multiple weak learners and fusion strategy. In Section III, the proposed method is implemented based on a NASA turbofan engine degradation simulation data set [19] and the results discussed. Section IV entails a discussion of obtained results while section V is the conclusion and areas for future work.

II. METHODOLOGY

This section presents the methodology that has been used. It includes a description of the dataset used, a list of algorithms used for prediction of the RUL; as illustrated in Fig.1 below, experimental set up and ensemble strategy used.

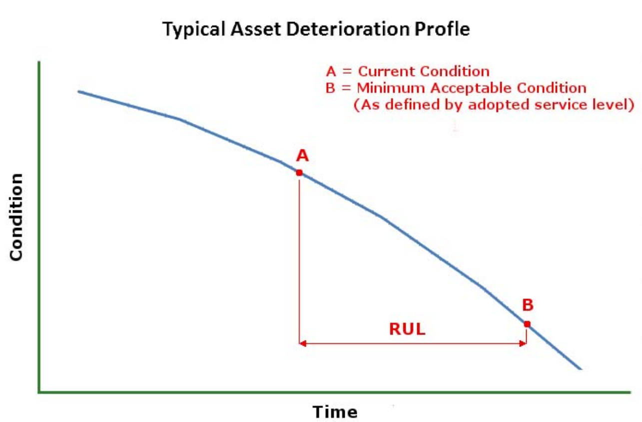

Fig.1 Deterioration profile of a system

A. Data Set

Data used is from the NASA data repository [19]. It consists of run-to-failure sensor measurements from degrading turbofan engines. Although the turbofan engines are similar, each engine starts with different degree of initial conditions and there are variations in the manufacturing process of the engines that are unknown to the user. Each engine's performance can be changed by adjusting three operational settings for the turbofan engines under consideration. Every engine has 21 sensors collecting different measurements related to the engine state at runtime with the data being of the time series nature as presented in the table below. At the start of the time series, the engine operates normally but after certain number of cycles, a fault is developed in the engine which then gradually fails.

TABLE I DATASET SCHEMA

Index	Data Fields	Type	Description
1	ID	Integer	Aircraft engine identifier
2	Cycle	Integer	Time in cycles
3	Setting 1	Double	Operational setting 1
4	Setting 2	Double	Operational setting 2
5	Setting 3	Double	Operational setting 3
6	Sensor 1	Double	Sensor measurement 1
7	Sensor 2	Double	Sensor measurement 2
8	Sensor 3	Double	Sensor measurement 3
…	…	…	…
26	Sensor 21	Double	Sensor measurement 21

The datasets are provided as text files for training, testing and measurement of accuracy. The training data entails the engine's run-to-failure data as the faults grow in magnitude until the system fails; the testing data entails the engine's operating data without failure events recorded since the data is truncated before failure threshold is reached; the ground truth data contains the information of true remaining cycles for each engine in the testing data. Four such datasets are provided. However, for this study only dataset 1 is used.

B. Machine Learning Algorithms Evaluated

The machine learning algorithms used in predicting the RUL are discussed below. Their selection is influenced by a review on regression MLAs by[18] and [20].

1. *Gaussian process regression:* It is a MLA that involves a gaussian process in lazy learning and a measure of the similarity between points to predict the value for an unseen point from training data. The prediction is not just an estimate for that point, but also has uncertainty information [21].
2. *Ensemble regression:* It combines a set of trained weak learner models and data on which these learners were trained. It can predict ensemble response for new data by aggregating predictions from its weak learners.
3. *Binary regression decision trees:* A tree is constructed as the predictive model with its branches illustrating the outcome of the decisions taken. The

observations about an item can then be converted to conclusions with the help of this decision tree[18].

4. *Gaussian kernel regression:* Kernel regression is an estimation technique to fit data that does not assume any underlying distribution to estimate the regression function n hence categorized as non-parametric technique. The idea of kernel regression is putting a set of identical weighted function called Kernel local to each observational data point.
5. *Support vector machine:* The numeric input variables in the data which are in the different columns for an n-dimensional space. A hyper plane is a line that splits the input variable space. A hyper plane is usually selected to bets separate the points in the input variable space by their class. This is implemented in practice using a kernel.
6. *Linear regression:* It is a basic algorithm used in predictive analysis. The relationship between the dependent variable; in this case the RUL, and the independent variables is explained using the regression estimates [22].

C. *Workflow*

The dataset was obtained from prognostic-data repository of Prognostics Centre of Excellence in NASA[19]. Data pre-processing was done to reduce noise levels and remove outliers. Feature selection was done using a monotonicity function while data smoothing was done using 'rloess' at intervals of 10 window-length. MATLAB was used to construct the models and test the results. The datasets were read from given files and the inbuilt algorithms applied using MATLAB. The results were stored in variables which were later used in performance analysis. From the results obtained, a graph was plotted to facilitate visualization of the results.Fig.2 illustrates this.

D. *Ensemble Strategy*

This paper proposes an ensemble approach that builds sub-models using different similar datasets and different MLAs. Data set 1 contains variables for 100 similar engines. These variables are used to develop 100 sub-models for each MLA being used. Given that the engines have different degree of initial conditions and there are variations in their manufacturing process creates diverse models which are desired in creating ensembles. Their outputs are ensembled using simple averaging. This strategy is illustrated in Fig.3. The input variables for each engine are represented by X_i while Z_i represents the average of the predicted output from all the sub-models.

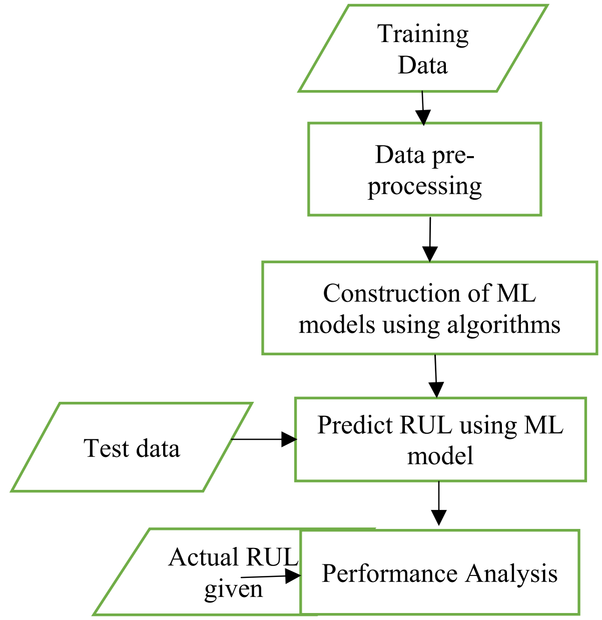

Fig.2 Flowchart for the Machine Learning model

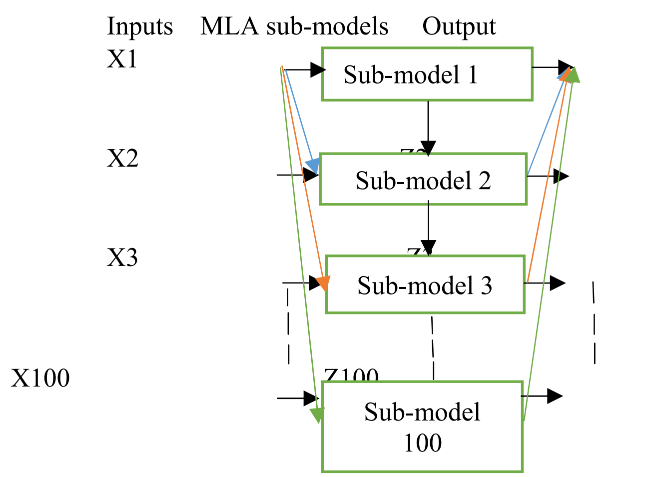

Fig 3. Paradigm of the ensemble approach used in this research to create sub-models using similar datasets.

III. RESULTS

Training and evaluation of the Machine Learning Models was carried out using dataset I that contains data from 100 similar engines. The data variables were classified according to the engine ID given in the dataset for both training and testing data. The 100 sub-models were developed as a result. The evaluation was done in two phases: phase one (single model) used one machine learning model, developed using training data from all the 100 engine units, to obtain the predicted output while phase two (ensemble model), developed by creating 100 sub-models using the training data, and its output obtained as the average of all the 100-machine learning sub-models. This was done for all the six regression MLA discussed above.

Performance evaluation was done using Mean Absolute Error (MAE) and Mean Squared Error (MSE) metrics and the results presented in Table II.

Mean Absolute Error(MAE) $=\frac{1}{n_{te}}\sum_{i=1}^{n_{te}}|T_i - P_i|$

Mean Squared Error (MSE)$=\frac{1}{n_{te}}\sum_{i=1}^{n_{te}}(T_i - P_i)^2$

Where; *T-True value*
P-predicted value
n_{te}-Number of test trajectories

Evaluation of the performance of the ensemble regression modelling was further carried out using training data from engine with ID 1to demonstrate failure propagation until the failure threshold is reached with the results presented in Table III and Fig.4.
Ensemble of the best two performing algorithms; ensemble regression and binary regression decision trees was done to further enhance the predicted results. The MSE for cases where single modelling and ensemble modelling were initially used were 2995 and 1029 respectively. Consequently, the MAE were 41.44 and 25.93 respectively.

TABLE II Mean Absolute Error and Mean Squared Error Values

PERFORMANCE EVALUATION WITH RUL AS TARGET				
Algorithm	**MAE**		**MSE**	
	Single model	**Ensemble model**	**Single model**	**Ensemble model**
Gaussian process regression	54.84	33.85	4906.7	1898.2
Ensemble regression	39.63	22.91	2840.2	875.2
Binary regression decision trees	45.89	29.51	3416.2	1261
Gaussian kernel regression	42.94	37.84	2886.9	2296
Support vector machine	51.92	32.39	4392.1	1640.2
Linear regression	54.86	33.87	4911.3	1900.3
Range for best 10 performing MLA	15.38-24.51		546.6-1078	

IV. DISCUSSION

For consistent comparison of data, the different algorithms were evaluated in the same way on the same data. For each RUL prediction, the results were compared with the actual RUL values available in the dataset provided. The ensemble regression had the best performance for both the metrics used and performed better using the ensemble model compared to the single model approach.
Ensemble regression model had good performance during the early life stage of the engine compared to just before the failure threshold was reached as indicated in Fig.4. It gave late RUL predictions which are undesired in maintenance.
An ensemble of the best performing two MLAs models was done; ensemble regression and binary regression decision trees illustrating the effectiveness of ensemble technique as the MAE and MSE obtained from the already sub-ensembled individual models gave results close to those of the best performing MLA model.

EVALUATION USING ENSEMBLE REGRESSION											
No of cycle	0	20	40	60	80	100	120	140	160	180	192
Actual RUL	192	172	152	132	112	92	72	52	32	12	0
Predicted RUL(single model)	192	171	145	131	111	84	73	55	36	14	1
Percentage error(one model)	0.0	-0.36	-4.6	-1.0	-0.75	-8.5	1.4	5.2	12	14.5	
Predicted RUL(ensemble model)	192	174	149	130	114	93	75	59	42	27	21
Percentage error(ensemble model)	0.0	1.20	-1.9	-1.6	1.6	0.6	3.5	14	30	125	

TABLE III Evaluation using Ensemble Regression

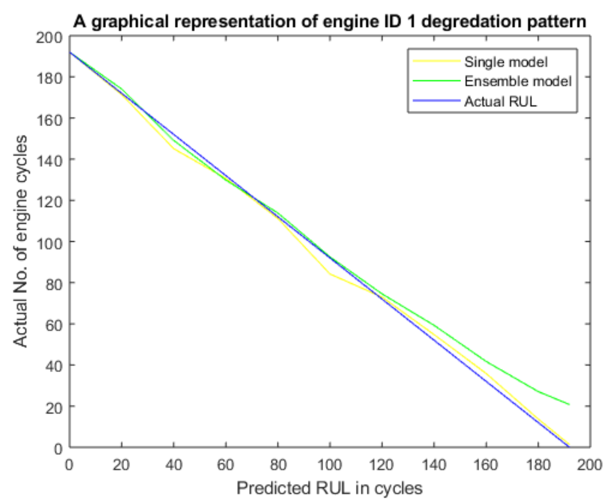

Fig 4. A graphical representation of engine 1 degradation pattern

V. CONCLUSION AND FUTURE SCOPE

In this paper, we have demonstrated the effect of using ensemble technique in RUL prediction. The ensemble models employed were built using similar datasets and combining outputs from best performing regression MLAs. The ensemble technique of combining the sub-models developed from similar datasets is 42% and 69% more effective when using MAE and MSE metrics respectively than single machine learning modeling as evident in the results obtained. Ensemble of the two best performing MLAs had a performance improvement of the weaker MLA by 18% and 12% for MSE and MAE respectively where sub-modelling was employed. Further research needs to be carried out on optimizing the number of sub-models used in developing the ensemble to reduce computational load and on incorporating uncertainties to improve on the accuracy of RUL prediction

VI. References

[1] Z.Zhao, L.Bin, X.Wang and W.Lu, "Remaining useful life prediction of aircraft engine based on degradation pattern learning," in *Reliability Engineering & System Safety*, 2017.

[2] Pecht, M., Prognostics and Health Management of Electronics, New York: Wiley-Interscience, 2008.

[3] 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," in *Eksploatacja i niezawodnosc – Maintenance and Reliability*, 2016.

[4] Liao L, and Kottig F., "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] Orchard M, and Vachtsevanos G., "A particle filtering approach for on-line failure prognosis in a planetary carrier plate.," in *Int J Fuzzy Logic Intel System*, 2007.

[6] 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.

[7] Hu C, Youn B, Wang P, and Yoon J, "Ensemble of data-driven prognostic algorithms for robust prediction of remianing useful life," in *Reliable Engineering System Safety*, 2012.

[8] J.M.Bates, and C.W.J.Granger, The combination of Forecasts, 2001.

[9] J.Sun, Z. Hongfu, Y.Haibin, and Michael Pecht, "Study of Ensmeble Learning-Based Fusion Prognostics," 2010.

[10] MP.Perrone ,and LN Cooper, "Ensemble methods for hybrid Neural Networks," in *When Networks disagree*, 1992.

[11] C.Zhang, P.Lim , A.Qin, and K.Tan, "Multiobjective deep belief networks ensemble for remianing useful life estimation in prognostics.," in *IEEE Trans Neural Network*, 2016.

[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.M.J.Tax, and R.P.W.Duin, "Experiments with Classifier Combining Rules," in *International Workshop Multiple Classifier Systems*, 2000.

[14] B.Zhang , L.Zhang , and J.Xu , "Remaining useful life prediction for rolling element beraing based on ensemble learning," 2013.

[15] Baraldi P, Mangili F, and Zio E, "A belief function theory based approach to combining different representation of uncertainity in prognostics," 2015.

[16] D.Wu , C.Jennings, J.Terpenny , and S.Kumara , "Cloud-based machine learning for predictive analytics: prediction of tool wear," in *IEEE International conference on big dara*, 2016.

[17] I, Peel, "Data driven prognostics using a Kalman filter ensemble of neural network models," in *International conference on prognostics and health management*, 2008.

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

[19] 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.

[20] T.Doan, and J.Kalita, "Selecting Machine Learning Algorithms using Regression Models," in *IEEE 15th International Conference on Data Mining Workshops*, 2015.

[21] G.Kopsiaftis, E.Protopapadakis and A.Voulodimos, "Gaussian Process Regression Tuned by Bayesian Optimization for Seawater Intrusion Prediction," in *Computational Intelligence and Neuroscience*, 2019.

[22] W. Ahmad, S. A. Khan and J.-M. Kim, "A hybrid prognostics technique for rolling element bearings using adaptive predictive models," in *IEEE Transactions on Industrial Electronics*, 2017.