Remaining Useful Life Estimation in Prognostics and Health Management using LSTM Neural Network and Vector Auto-Regression Models

Ritu Thombre*, Sanket Gajbhiye†, Meera Dhabu‡
Department of Computer Science & Engineering, Visvesvaraya National Institute of Technology
Nagpur, India
Email: *rituthombre99@gmail.com, †gajbhiye.sanket99@gmail.com, ‡meeradhabu@cse.vnit.ac.in

***Abstract*—Prognostics and health management (PHM) is an essential topic in the industry for asset health management and monitoring the operational availability of these assets by using their historical data, which consists of raw sensor measurements and operational settings. However, due to the lack of accurate models in most cases, there is a scope of improvement in the field of asset health management and PHM. This paper proposes a data-driven approach for prognostics using the combination of Vector Auto-Regression Model and Long Short Term Memory(LSTM) Neural Network. There is no need of any prior expertise on prognostics or signal processing due to ability of Machine Learning (ML) algorithms to efficiently model complex data, which facilitates the application of the proposed method. To demonstrate the effectiveness of the proposed approach, experiments are carried out on CMAPSS (Commercial Modular AeroPropulsion System Simulation) dataset from NASA which includes Run-to-Failure simulated data from turbofan jet engines. Experimental evaluation shows that high prognostic accuracy on the RUL estimation is achieved with proposed models.**

***Index Terms*—Prognostics and Health Management, Remaining Useful Life, Machine Learning, LSTM Neural Network, CMAPSS dataset, Vector Auto-Regression.**

I. INTRODUCTION

Prognostics and engineering maintenance plays a crucial role in many industrial areas such as aerospace, manufacturing, automotive, etc. Also, for big cyber-physical systems, like power grid, prognostics and health management (PHM) is a vital aspect and very challenging because of the added dimension of having equipment at distributed locations to maximize the operational availability, reduction of maintenance costs and improvement of system reliability and safety by monitoring the facility conditions.

Remaining useful life (RUL) is the length of time a machine is likely to operate before it requires repair or replacement. RUL can be estimated based on historical data, which consists of sensor measurements and operational settings which is very important for improving maintenance schedules to avoid engineering failures and save the resultant costs [1].

Over recent years, discovering the relationship between the monitored system historical data and determining the corresponding RUL has been receiving increasing attention in data-driven prognostics. A number of machine learning (ML) techniques have been proposed and developed to learn the mapping from the collected feature data to the associated RUL. The advantage of applying ML techniques for PHM is because of the ability of ML algorithms to efficiently model highly nonlinear, complex, multi-dimensional system without prior expertise using the system physical behaviour data.

The contribution of the proposed work in this paper is as follows:

1) Data-driven models are constructed without any prior expertise on prognostics or signal processing for predicting RUL.
2) Data-driven approach is proposed for prognostics using a combination of Vector Auto-Regression model and Long Short Term Memory (LSTM) neural network; thus, the proposed model expected to obtain higher prognostic accuracy as compared to using a single model.
3) The proposed models are capable of predicting RUL multiple cycles into the future.

To demonstrate and validate the effectiveness of the proposed work, the RUL for turbofan jet engines is estimated using the proposed models as a case study on the NASA CMAPSS (Commercial Modular Aero-Propulsion System Simulation [2]) dataset [3].

The rest of the paper is organized as follows. Related works are discussed in Section II. The proposed model are presented in Section III. Details about the dataset, pre-processing and performance measures are parts of Section IV. Results and analysis of the proposed models are discussed in Section V followed by Section VI which concludes the work presented.

II. RELATED WORK

Most of the prognostic and health management systems are designed using model-based methods [4], data-driven methods [5] and hybrid methods [6]. This paper mainly focuses on data-driven methods.

Data-driven approaches usually require historical data for training models and they do not rely on much prior expertise on prognostics and are easy to be generalized. Some of the data-driven approaches include the traditional multi-layer perceptron [7] approach for modeling the RUL of the laboratory-tested bearings which reported the prediction results superior to the reliability-based approaches. Neural network (NN) [8] approach for modeling RUL from degradation signals because

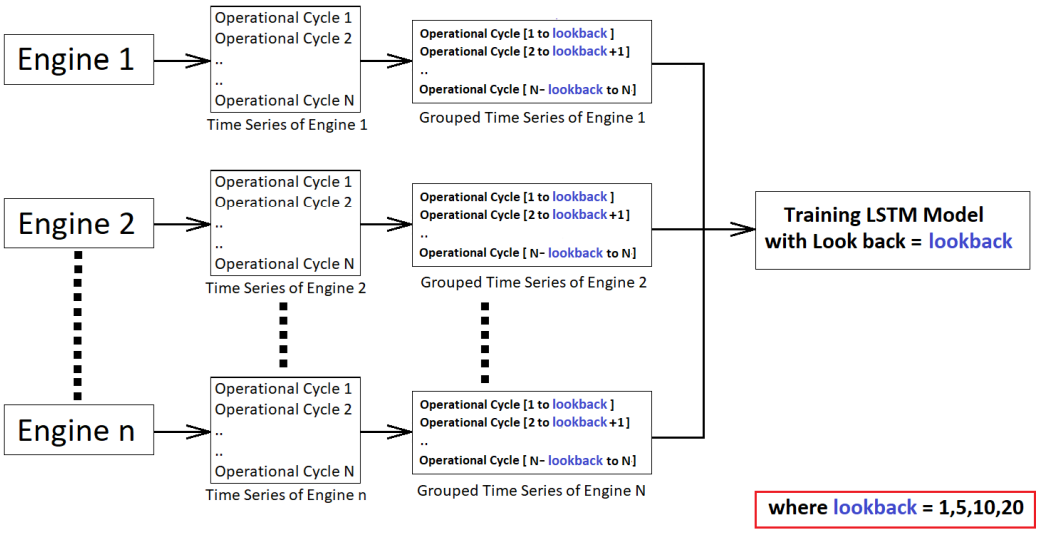

Fig. 1. Flow diagram for training LSTM Model.

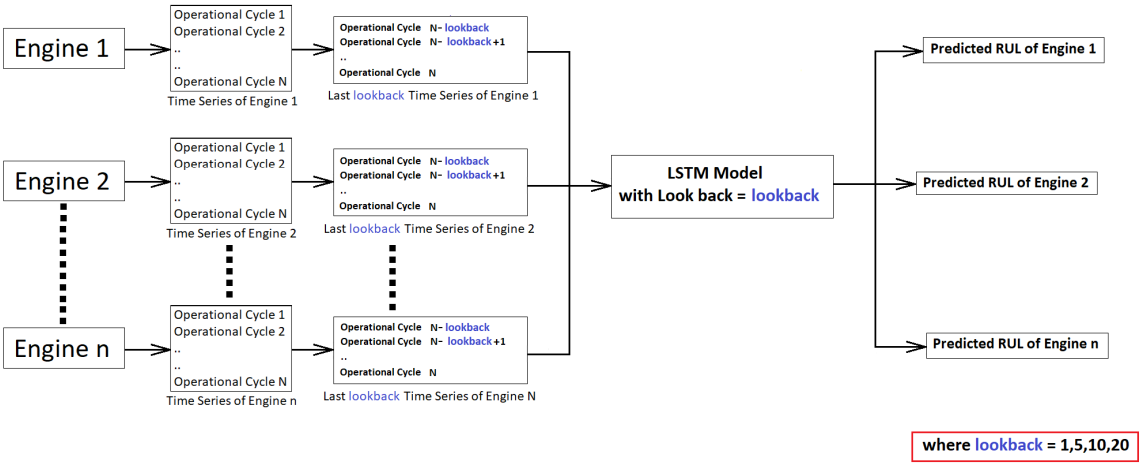

Fig. 2. Flow diagram of LSTM Model for RUL Prediction.

of vibrations and validation using real-world vibration monitoring data collected from pump bearings in the field which achieved more accurate RUL predictions. In [9] support vector machine (SVM) approach to predict RUL of bearings using the isometric feature mapping reduction technique (ISOMAP) is proposed. Support vector regression (SVR) and effectively modeled the evolution of the degradation, fault diagnosis and RUL estimation using LSTM neural network which diagnoses and predicted performance in the cases of complicated operations, hybrid faults and strong noises, is presented in [10]. Deep Convolution Neural Networks (DCNN) for RUL prediction is proposed in [11]. The hidden Markov model is proposed in [12] by developing a statistical modelling methodology for performing both diagnosis and prognosis in a unified framework based on segmental hidden semi-Markov models (HSMMs) for health monitoring of hydraulic pumps.

Most of the research work done towards developing data-driven models for RUL prediction mentioned above demonstrated RUL prediction only for one cycle into the future. Whereas the models proposed in this work are capable of predicting RUL multiple cycles into the future.

III. PROPOSED MODELS

A. Regression Model

Regression-based models are easy to implement [13] and provide a rough lower bound on the performance for models like LSTM neural network and VAR model which are later used in the proposed work to improve the performance.

Regression-based models implemented in this paper include linear regression, logistic regression and random forest regression and they are constructed as follows:

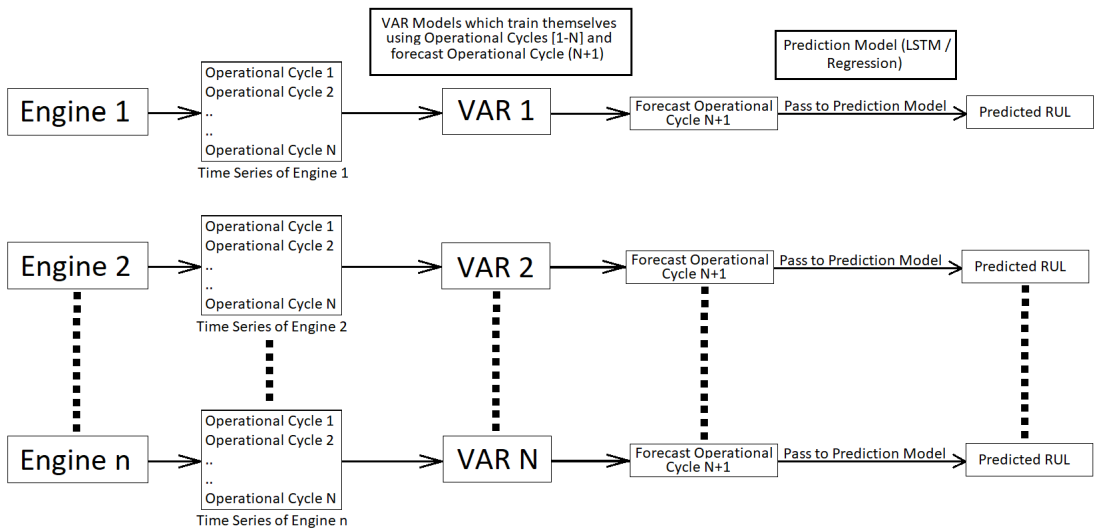

Fig. 3. Flow diagram of Vector Autoregression (VAR) model for RUL prediction.

TABLE I
LSTM MODEL ARCHITECTURE.

Layer	Input shape	Output shape
LSTM	Columns x Lookback	128
LSTM	128	64
Dense	64	16
Dense	16	1

1) Training data and test data are pre-processed as discussed in IV-B.
2) Regression models are trained using time-series of all the engines from training data all at once.
3) Time-series of each engine from the test data is passed to trained models one by one.
4) Final predicted RUL for an engine is taken as mean of all the predicted RULs of entire time-series of that engine.

B. LSTM (Long Short Term Memory) Model

LSTM [14] networks are well suited in time-series forecasting [15] which is an advantage over previously discussed regression models.

1) Model Architecture: LSTM Models used in the proposed work consists of two LSTM layers followed by two dense layers. Multiple LSTM layers are not added to prevent overfitting caused by larger-scale parameters in deeper LSTM networks [16].

The LSTM model architecture details are given in the Table I, where number of columns = number of features in the dataset, and lookback = 1,5,10,20.

Lookback of LSTM cannot be increased beyond 20 because lookback should be less than maximum number of cycles of any engine runs for, and it is observed that minimum of maximum cycles of any engine is 31. Also, increasing lookback leads to slower computation.

Following are the steps for RUL prediction using the LSTM model :

1) Training data and test data are passed through pre-processing modules discussed in IV-B.
2) Time-series of each engine in the training data is grouped according to the lookback and are then used to train the LSTM model.
3) Pre-processed time-series from the test data are grouped according to the corresponding lookback of LSTM.
4) RUL for each engine is predicted by using the last group of (1,5,10 or 20 which are lookback for LSTM) time-series of that engine.

Training of LSTM Model and RUL prediction using LSTM model is shown in Fig. 1 and Fig. 2 respectively.

C. VAR (Vector Autoregression) Model

Vector auto-regression (VAR) is a statistical model used to capture the relationship between multiple quantities as they change over time. VAR generalizes an autoregressive model for multivariate time-series and is used for data description, forecasting, structural inference, policy analysis, etc [17].

For the prediction module, RUL is required to be predicted for the immediate next cycle based on the given past time-series of every engine. Each engine has its own VAR model; therefore, the number of VAR models is equal to the number of engines present in the dataset.

Following are the steps for RUL prediction using the VAR model :

1) Training and test data are passed through pre-processing modules discussed in IV-B and pre-processed training data is used to train LSTM and regression models.

2) Time-series of each engine from the test data is passed to corresponding VAR model which forecasted future time-series one cycle ahead.
3) The forecasted time-series by VAR model is passed to trained regression or LSTM model which calculates RUL and compares it to actual RUL from the dataset.

VAR Model architecture is shown in Fig. 3.

IV. EXPERIMENTAL SETUP

A. Dataset

The NASA CMAPSS(Commercial Modular Aero-Propulsion System Simulation) dataset [3] is used to validate the effectiveness of the proposed models. This dataset consists of time-series of turbofan jet engines along with their remaining useful life. Each time-series consists of the engine ID, current operational cycle, three operational settings and 21 sensor measurements.

Engines start with unknown initial condition and evolve with the progression of their time-series. The engine develops a fault at some point during the series. In the training set, the fault grows in magnitude until system failure. In the test set, the time-series ends some time prior to system failure, and the goal is to estimate the number of operational cycle remaining (RUL) before the system failure.

B. Pre-processing

Pre-processing is divided into two phases, viz, pre-processing datasets and generating RULs for training data which will be used to train the models.

1) Pre-processing datasets: Following are the steps for involved in data pre-processing :

1) **Data Analysis** : To analyse the various statistics like minimum, maximum, standard deviation, etc.
2) **Normalization** : Done using MinMaxScaler.
3) **Noise removal** : PowerTransform is used to remove noise and also shift the distribution of dataset to Gaussian distribution.

It is observed that when both MinMaxScaler and PowerTransform, are used for data pre-processing, supervised machine learning leads to better results [18].

2) Generating RULs for training data: To obtain RUL for the training data :

- In the training set, every engine with a unique Machine ID runs for a particular number of cycles and the engine fails at the end of all these cycles.
- So RUL for every cycle of an engine is defined as: (total number of cycle Engine runs for) - (current operational cycle)

C. Performance Measures

Performance measures used to evaluate various models in this work are Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) as shown below: [19].

$$RMSE = \sqrt{(\frac{1}{n})\sum_{i=1}^{n}(y_i - x_i)^2} \tag{1}$$

$$MSE = (\frac{1}{n})\sum_{i=1}^{n}(y_i - x_i)^2 \tag{2}$$

$$MAE = (\frac{1}{n})\sum_{i=1}^{n}|(y_i - x_i)| \tag{3}$$

N = number of data points
y_i = actual observations of time-series
x_i = estimated time-series

V. RESULTS AND ANALYSIS

A. Regression Models

The results obtained by regression models are presented in Table II.

TABLE II
RESULTS OF VARIOUS REGRESSION MODELS.

Regression Model	MSE	RMSE	MAE
Random Forest	3179.90	56.39	49.34
Linear	2747.63	52.41	44.35
Logistic	2454.39	49.54	40.50

Regression models cannot learn from their past inputs. Since the dataset used for this experiment consists of time-series data, results obtained by regression models are not optimal enough.

However, the results obtained by regression models provide a boundary for the performance of LSTM and VAR models. Also, logistic regression in combination with VAR showed better results as compared to logistic regression alone.

B. LSTM Models

The results obtained by LSTM models are presented in Table III and comparison of predicted vs actual RUL for LSTM with lookback of 1,5,10,20 is shown in Fig 4, Fig 5, Fig 6, Fig 7 respectively.

TABLE III
RESULTS OBTAINED BY LSTM MODEL.

Lookback	MSE	RMSE	MAE
1	1082.23	32.89	26.39
5	1854.72	43.06	33.38
10	1694.33	41.16	32.39
20	1629.12	40.36	31.48

It is not possible to create a different model for every engine as LSTM requires a long historical database of measurements [20]. However, number of cycles every engine runs is very short (between 200-250) which is insufficient to train different LSTM model separately for every engine.

Lookback can not be increased beyond a certain value, and thereby LSTM models cannot take full advantage of the entire

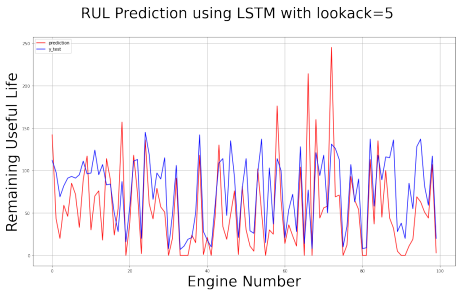

Fig. 4. LSTM results with Lookback=1.

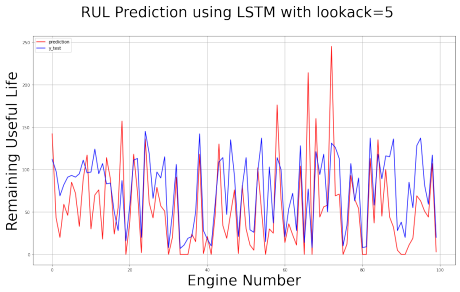

Fig. 5. LSTM results with Lookback=5.

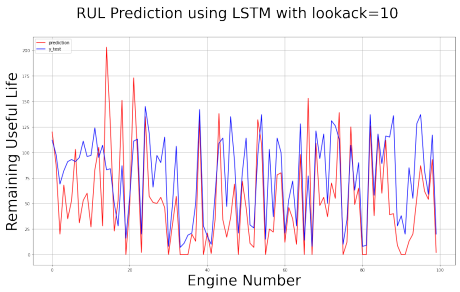

Fig. 6. LSTM results with Lookback=10.

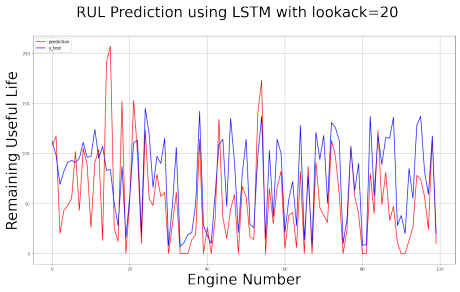

Fig. 7. LSTM results with Lookback=20.

time-series of an engine from the test dataset. RUL for an engine from the test dataset is only based on its last 1,5,10 or 20 time-series. This problem is resolved by using the VAR model which have modifiable order.

Performance of LSTM can be further improved by adding dropout [21] and employing feature optimization techniques such as genetic algorithm [22].

C. VAR Models

The results obtained using VAR models are presented in Table IV and comparison of predicted vs actual RUL for VAR with logistic regression and VAR with LSTM is shown in Fig 8 and 9 respectively.

TABLE IV
RESULTS OBTAINED BY VAR.

Model used to Predict RUL	MSE	RMSE	MAE
Logistic Regression	1262.74	35.53	28.68
LSTM(lookback=1)	861.63	29.35	22.91

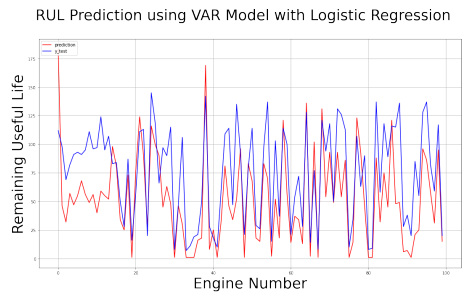

Fig. 8. VAR results with Logistic Regression for RUL Prediction.

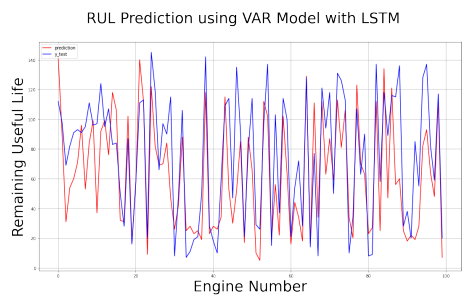

Fig. 9. VAR results with LSTM(lookback=1) for RUL Prediction.

Previously discussed LSTM models suffered the problem of increasing lookback beyond a certain value and thereby LSTM models cannot take full advantage of the entire time-series of an engine from the test dataset. This problem is solved by using autoregressive models which have dynamic order.

VAR model is used along with logistic regression and LSTM with lookback=1 since logistic regression showed better results compared to other regression models and LSTM with lookback=1 showed better results compared to LSTM with lookback=5,10,20.

Lookback=1 for LSTM is compatible with VAR for predicting RUL one cycle into the future based on the past time-series of an engine. LSTM models with greater lookback can be used with VAR if RUL is to be predicted further into the future.

As it can be clearly seen from the above results, VAR performed better as compared to LSTM and regression models alone because of the following reasons :

1) Unlike LSTM, VAR model only requires at least 50 but preferably more than 100 observations [23] for training

and are implemented individually for each engine, as each engine in the dataset has more than 150 observations.

2) Order of a particular VAR for a certain engine depends on the number of cycles that engine has been operating for, unlike LSTM where lookback can not be increased beyond 20.
3) Entire time-series of an engine from test dataset is used in estimating RUL unlike LSTM, where only last 1,5,10 or 20 time-series were used in RUL prediction.

VI. CONCLUSIONS

The main motive of this paper was to develop a robust data-driven approach in prognostics and health management for RUL prediction. Various machine learning and neural network models for RUL prediction are implemented and compared. Experiments are carried out on the popular CMAPSS dataset to show the effectiveness of the proposed models. The goal is to estimate the remaining useful life of turbofan-jet-engine units.

With data normalization using MinMaxScaler and noise removal using PowerTransform which shifts the distribution of the dataset to Gaussian distribution, good prognostic performance is achieved by combining VAR with LSTM for RUL prediction method.

Results of linear regression, logistic regression, random forest regression, LSTM with lookback of 1,5,10,20, VAR models in combination with logistic regression and LSTM (lookback=1) are compared. The results of VAR combined with LSTM (lookback=1) has shown better results. The proposed VAR with LSTM for RUL prediction model can also be used to predict RUL multiple cycles into the future by increasing the steps of VAR model. Comparison of performance measures of these models is shown in Fig. 10, Fig. 11 and Fig. 12.

While good experimental results have been obtained by the VAR with LSTM, further architecture optimization is still necessary. Efforts will be made to enhance model performance by using feature optimization techniques such as Genetic Algorithm [22], by adding dropout to LSTM models which had shown to increase model performance [21] and by using scoring function as a performance metric which has been proposed by many researchers [24], [25], [26]. In PHM late prediction are more dangerous than early prediction, which is because late prediction usually leads to more severe consequences in many fields such as aerospace industries. Scoring function takes this asymmetry of late vs. early predictions into account, and penalizes late predictions.

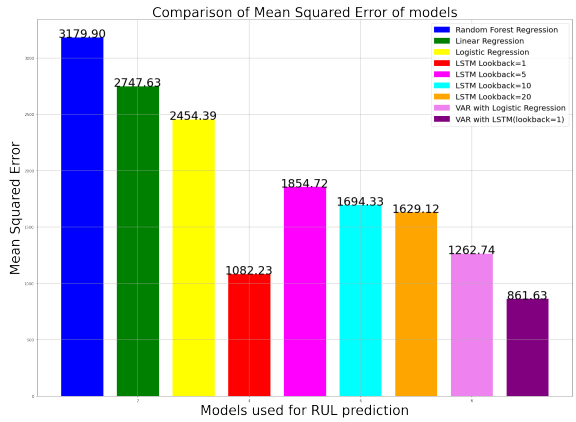

Fig. 10. Comparison of Mean Squared Error of various models.

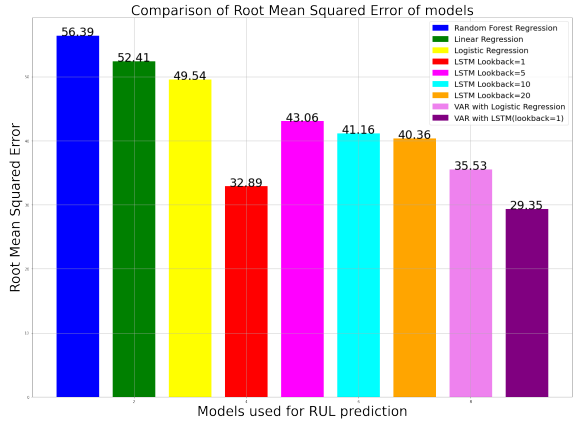

Fig. 11. Comparison of Root Mean Squared Error of various models.

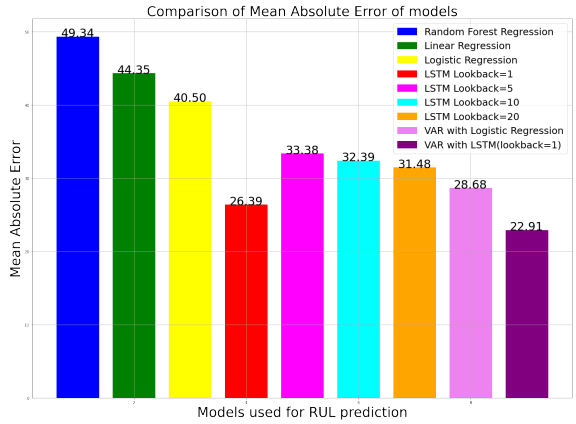

Fig. 12. Comparison of Mean Absolute Error of various models.

REFERENCES

[1] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—reviews, methodology and applications," *Mechanical Systems and Signal Processing*, vol. 42, no. 1, pp. 314–334, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327013002860

[2] D. K. Frederick, J. A. DeCastro, and J. S. Litt, "User's guide for the commercial modular aero-propulsion system simulation (c-mapss)," 2007.

[3] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *2008 International Conference on Prognostics and Health Management*, 2008, pp. 1–9.

[4] M. Pecht and J. Gu, "Physics-of-failure-based prognostics for electronic products," *Transactions of the Institute of Measurement and Control*, vol. 31, no. 3-4, pp. 309–322, 2009. [Online]. Available: https://doi.org/10.1177/0142331208092031

[5] F. O. Heimes, "Recurrent neural networks for remaining useful life

estimation," in *2008 International Conference on Prognostics and Health Management*, 2008, pp. 1–6.

[6] A. Heng, S. Zhang, A. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724–739, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327008001489

[7] R. Huang, L. Xi, X. Li, C. Richard Liu, H. Qiu, and J. Lee, "Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 193–207, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327005002165

[8] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: a neural network approach," *IEEE Transactions on Industrial Electronics*, vol. 51, no. 3, pp. 694–700, 2004.

[9] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Remaining useful life estimation based on nonlinear feature reduction and support vector regression," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 7, pp. 1751–1760, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197613000365

[10] M. Yuan, Y. Wu, and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using lstm neural network," in *2016 IEEE International Conference on Aircraft Utility Systems (AUS)*, 2016, pp. 135–140.

[11] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering System Safety*, vol. 172, pp. 1–11, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832017307779

[12] M. Dong and D. He, "A segmental hidden semi-markov model (hsmm)-based diagnostics and prognostics framework and methodology," *Mechanical Systems and Signal Processing*, vol. 21, no. 5, pp. 2248–2266, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327006002214

[13] K.-Y. Lee, K.-H. Kim, J.-J. Kang, S.-J. Choi, Y.-S. Im, Y.-D. Lee, and Y.-S. Lim, "Comparison and analysis of linear regression artificial neural network," *International Journal of Applied Engineering Research*, vol. 12, pp. 9820–9825, 01 2017.

[14] R. C. Staudemeyer and E. R. Morris, "Understanding lstm–a tutorial into long short-term memory recurrent neural networks," *arXiv preprint arXiv:1909.09586*, 2019.

[15] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A comparison of arima and lstm in forecasting time series," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018, pp. 1394–1401.

[16] K. Xu, H. Wang, and P. Tang, "Image captioning with deep lstm based on sequential residual," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 361–366.

[17] J. H. Stock and M. W. Watson, "Vector autoregressions," *Journal of Economic perspectives*, vol. 15, no. 4, pp. 101–115, 2001.

[18] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the influence of normalization/transformation process on the accuracy of supervised classification," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2020, pp. 729–735.

[19] C. J. Willmott, "Some comments on the evaluation of model performance," *Bulletin of the American Meteorological Society*, vol. 63, no. 11, pp. 1309 – 1313, 01 Nov. 1982. [Online]. Available: https://journals.ametsoc.org/view/journals/bams/63/11/1520-0477_1982_063_1309_scoteo_2_0_co_2.xml

[20] T. Boulmaiz, M. Guermoui, and H. Boutaghane, "Impact of training data size on the lstm performances for rainfall runoff modeling," *Modeling Earth Systems and Environment*, vol. 6, no. 4, pp. 2153–2164, Dec 2020. [Online]. Available: https://doi.org/10.1007/s40808-020-00830-w

[21] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," 2015.

[22] P. Ghamisi and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 309–313, 2015.

[23] G. E. P. Box and G. C. Tiao, "Intervention analysis with applications to economic and environmental problems," *Journal of the American Statistical Association*, vol. 70, no. 349, pp. 70–79, 1975. [Online]. Available: https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1975.10480264

[24] Z. Zhao, Bin Liang, X. Wang, and W. Lu, "Remaining useful life prediction of aircraft engine based on degradation pattern learning," *Reliability Engineering System Safety*, vol. 164, pp. 74–83, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832017302454

[25] J. B. Coble and J. W. Hines, "Prognostic algorithm categorization with phm challenge application," in *2008 International Conference on Prognostics and Health Management*, 2008, pp. 1–11.

[26] P. Wang, B. D. Youn, and C. Hu, "A generic probabilistic framework for structural health prognostics and uncertainty management," *Mechanical Systems and Signal Processing*, vol. 28, pp. 622–637, 2012.