Toronto Metropolitan University

Department of Aerospace Engineering
Faculty of Engineering & Architectural Science

Semester (Term, Year)	Winter 2023
Course Code	AER850
Course Section	01
Course Title	Introduction to Machine Learning
Course Instructor	Dr. Krishna Kumar
Submission	Project
Submission No.	Project Phase III
Submission Due Date	04/12/2023
Title	Project Phase III
Submission Date	04/12/2023

Submission by (Name):	**Student ID (XXXX1234)**	**Signature**
Hussein Hamie	500876254	
Mishran Haque	500896460	

By signing the above you attest that you have contributed to this submission and confirm that all work you contributed to this submission is your own work. Any suspicion of copying or plagiarism in this work will result in an investigation of Academic Misconduct and may result in a "0" on the work, and "F" in the course, or possibly more severe penalties, as well as a Disciplinary Notice on your academic record under the Academic Integrity Policy 60, which can be found at www.torontomu.ca/senate/policies/

Table of Contents

List of Figures

List of Tables

I. Introduction

Turbofan Engines used in aircraft are prone to degradation over time, resulting in a decline in engine health and performance, which eventually leads to engine failure. To ensure the safe and efficient operation of aircraft engines, it is necessary to closely monitor engine degradation. Predictive Maintenance strategies can be used to anticipate the maintenance requirements of engines and detect any issues with the engine before they become critical. This reduces the probability of unscheduled downtimes and significantly reduces the chances of engine failure.

The objective of this project is to predict the remaining useful life (RUL) of turbofan engines. To achieve this, run-to-failure simulated data provided by NASA Ames is to be utilized, along with the Random Forest Regression model, and long short-term memory (LSTM) Neural Network model. The accuracies of these models are to be evaluated and compared to determine the optimal Machine Learning methodologies for predictive maintenance purposes.

II. Data Overview

The tool used to obtain the sensor measurements is called C-MAPSS. This tool is specifically used to simulate realistic commercial turbofan engines. The overall system simulates an engine model of 90,000 lbs. thrust class, with operational conditions at altitudes from sea level to 40,000 ft, Mach numbers from 0 to 0.9, and temperatures from -60 to 103 degrees Fahrenheit.

The following are the subsystems and or elements of C-MAPSS:

- Power Management System: Allows the engine to be operated at a wide range of thrust levels.
- Control System: Consists of a fan-speed controller, along with regulators and limiters. A high-limit regulator enables the engine to stay within the core speed limits, engine-pressure ratio, and High-Pressure Turbine exit temperatures. A limit regulator allows for the High-Pressure Compressor’s exit static pressure from dipping too low.
- Logic Structure: Integrate the control system’s components to emulate real engine controllers.

The diagram of the simulated engine is shown in Figure 1, and the integration of the various subsystems is shown in Figure 2.

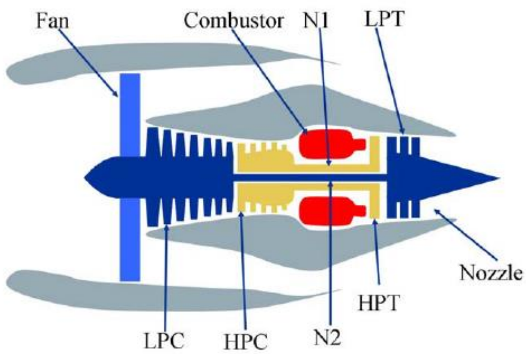

Figure 1: Diagram of the simulated engine [2]

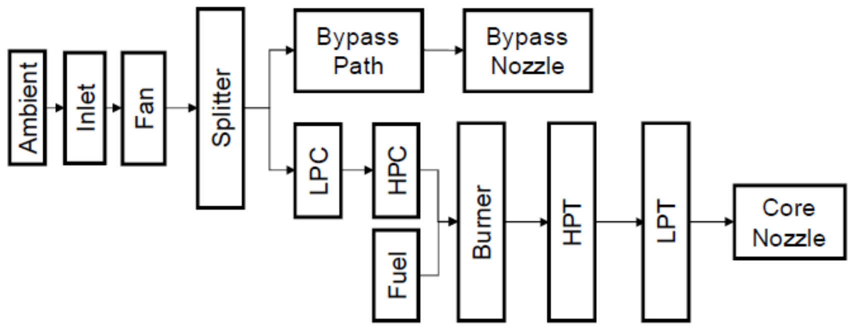

Figure 2: Diagram of integration of the separate subsystems and elements [2]

Table 1 displays the parameters/sensor data that C-MAPPS outputs. The displayed sensor data are acquired in the datasets attached to the lab and are used for the completion of the project.

Table 1: Sensor data

Symbol	**Description**	**Units**
T2	Total temperature at fan inlet	°R
T24	Total temperature at LPC outlet	°R
T30	Total temperature at HPC outlet	°R
T50	Total temperature at LPT outlet	°R
P2	Pressure at fan inlet	psia
P15	Total pressure in bypass-duct	psia
P30	Total pressure at HPC outlet	psia
Nf	Physical Fan speed	rpm
Nc	Physical Core speed	rpm
epr	Engine pressure ratio	N/A
Ps30	Static pressure at HPC outlet	pps/psi
phi	Ratio of fuel flow to static pressure at HPC outlet	rpm
NRf	Corrected fan speed	rpm
Nrc	Corrected core speed	N/A

BPR	Bypass ratio	N/A
farB	Burner fuel-air ratio	N/A
htBleed	Bleed Enthalpy	N/A
Nf_dmd	Demanded fan speed	rpm
PCNfR_dmd	Demanded corrected fan speed	rpm
W31	HPT coolant bleed	lbm/s
W32	LPT coolant bleed	lbm/s

The data sets used for the project consist of training and test subsets of multiple multivariate time series. Each of the separate data sets is obtained from a different unit of the same type of engine. Each separate unit starts with an unknown amount of initial wear and manufacturing variation. The engines used have three distinct operational settings, each having a significant impact on the performance and efficiency of the engine and are therefore indicated in the data. Additionally, due to factors such as environmental conditions, and electromagnetic interference, the sensor data used is contaminated with sensor noise impacting the accuracy and the precision of the actual measurements. In the training data sets, the engines operate until complete system failure due to developed faults. In the testing data sets, the time series ends at a point before reaching system failure.

Each training and testing data text file contains 26 columns of different variables, separated by spaces. The columns of the data correspond to the unit number of the engine, time in cycles, and three operational settings followed by sensor measurements. Each of the rows corresponds to the data taken during a single operational cycle of the engine (Refer to the data sets attached for more information on the data used for the project).

III. Methodology

A. Exploratory Data Analysis

To get a better understanding of the data being worked with, an exploratory analysis is done. Figure 3 shows a plot of the number of cycles before failure for each engine in the dataset from highest to lowest. This helps determine which engines have operated the longest prior to breaking down.

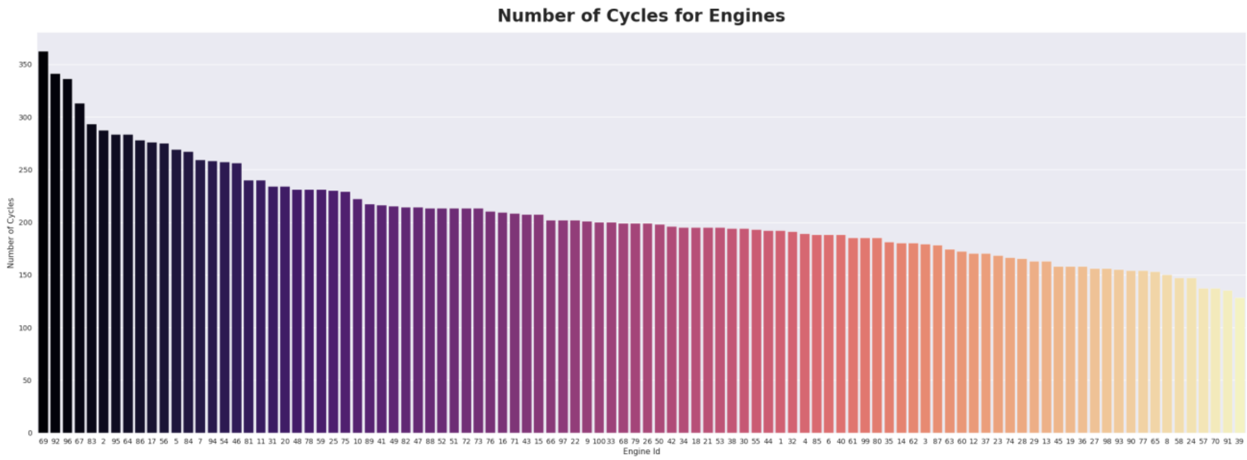

Figure 3: Number of cycles for engines

Figure 4 provides a visualization of the mean cycles of the engines before failure. From this figure it is seen that the mean number of cycles after which the turbofan engine fails is 206.

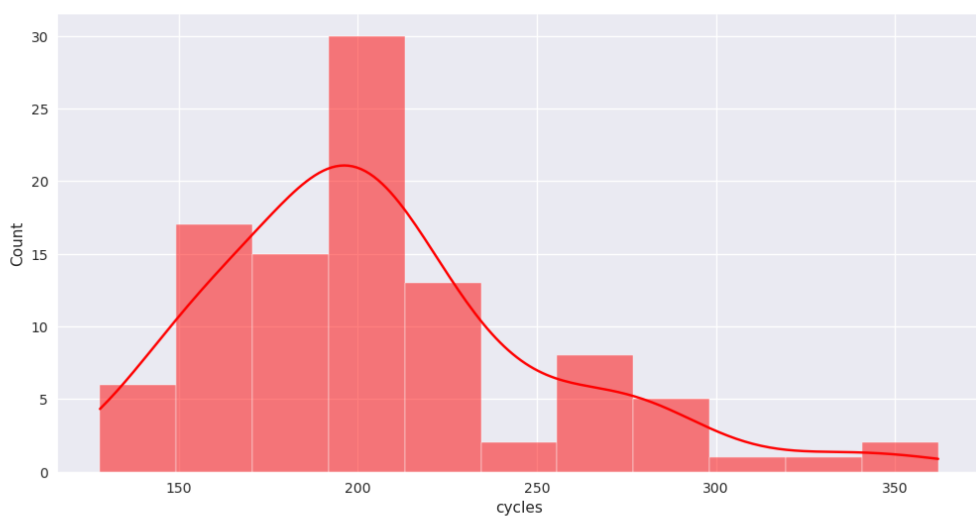

Figure 4: Mean cycles before engine failure

The computation of the standard deviation of each column is done, and the visualization is provided in Figure 5. This plot helps determine that the features with a standard deviation of 0 can be dropped since they do not have any significant impact on the training and testing data.

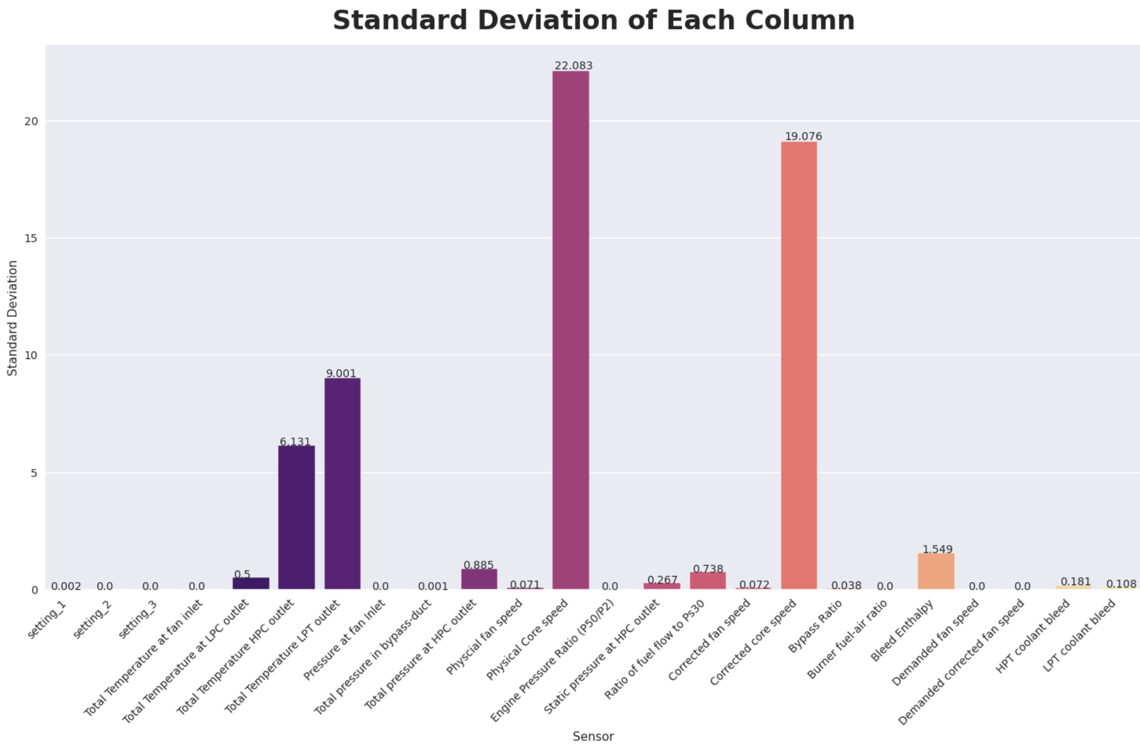

Figure 5: Standard deviation of each sensor data

Finally, a heatmap of the features is computed and shown in Figure 6. The heatmap allows for a visualization of the features of the dataset and how they are correlated to each other, along with the strength of their correlations. The information obtained from the heatmap is useful during the pre-processing of the data for the purposes of feature selection and dimensionality reduction (see the Data Pre-Processing section of the report).

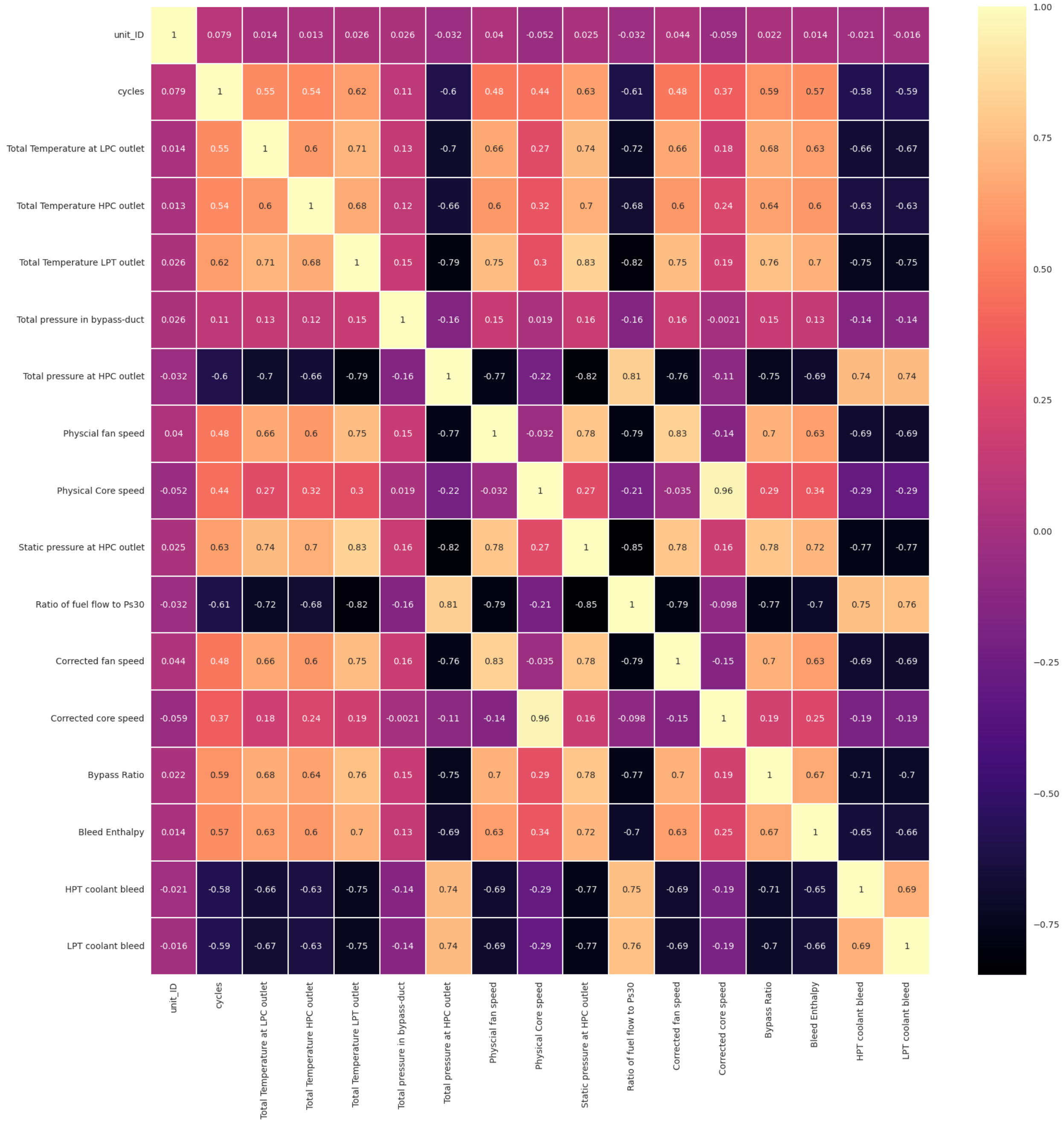

Figure 6: Heatmap of features

B. Data Pre-processing

The following steps are taken to process the data before moving on to model creation and training:

i) Any columns that do not contain features to be used as input variables are removed. These include columns that contain unit numbers, time cycles, and operational settings.

These columns do not contain features and their values are not required for training an ML model. This leaves the dataset containing only the 21 columns pertaining to sensor data.

ii) Any features (sensor data) that remain constant over the time cycles of the engine, are dropped. These features do not provide any meaningful input and only increase the dimensions and complexity of the model. In this project, the following features (sensor data) are dropped as they remain constant; ‘Total temperature at fan inlet’, ‘Total pressure in bypass-duct’, ‘Pressure at fan inlet’, ‘Engine pressure ratio’, ‘Burner fuel-air ratio’, ‘Demanded fan speed’, and ‘Demanded corrected fan speed’. This leaves the dataset containing only 14 feature columns, reducing it from the original 26 columns. Table 2 highlights the final selection of features after the irrelevant columns of the dataset are dropped.

Table 2: Selected features

Column Number	**Feature**
1	Total temperature at LPC outlet
2	Total temperature at HPC outlet
3	Total temperature at LPT outlet
4	Total pressure at HPC outlet
5	Physical fan speed
6	Physical core speed
7	Static pressure at HPC outlet
8	Ratio of fuel flow to static pressure at HPC outlet
9	Corrected fan speed
10	Corrected core speed
11	Bypass ratio
12	Bleed enthalpy
13	HPT coolant bleed
14	LPT coolant bleed

iii) The data is transformed using min-max scaling to translate each of the features, such that they are within the range between zero and one.

iv) For the LSTM model, the training and testing data is split into time windows with a shift of 1. This is done prior to reshaping the input data into a 3D tensor shape, where the first dimension of the tensor is the number of features, the second dimension is the time window (30 in this case), and the third dimension is the number of samples generated after the data is split into the time windows. The purpose of reshaping the data is so that the input of the multivariate data can be fed into a Neural Network model, as required by the Keras library [3].

v) Finally, the data is split into training and validation, with 20% of the data being used for validation.

C. Machine Learning Model

The RUL prediction for this project is done using two separate methodologies, and the results are compared. Firstly, a simple Random Forest Regressor is used for prediction, and the score of the regressor is evaluated. Subsequently, a more complex long short-term memory (LSTM) is used, and the Neural Network's accuracy is compared to the regression model.

The LSTM model is created using Keras' sequential class. Table 3 shows the details regarding the layers of the Neural Network model that is created, along with the units and activation functions for each layer.

Table 3: Neural Network model layers

Layer number	Layer type	Units	Activation function
1	LSTM	128	tanh
2	LSTM	64	tanh
3	LSTM	32	Tanh
4	Dense	96	relu
5	Dense	128	relu
6	Dense	1	-

The model uses MSE loss function and the Adam optimizer. The initial learning rate is set to 0.001, however, a learning rate scheduler is used so that during training epochs 0 to 5, the learning rate is set to 0.001, and after that, the rate is set to 0.0001. This is done to increase the accuracy of the model.

IV. Results & Discussion

Using the methodologies described in Section III of the report, the RUL of turbofan engines are predicted. Figure 7 shows a comparison between the true RUL and the predicted RUL obtained from using the Random Forest Regression model. The model prediction appears to be inline with the actual values with some noticeable drop off.

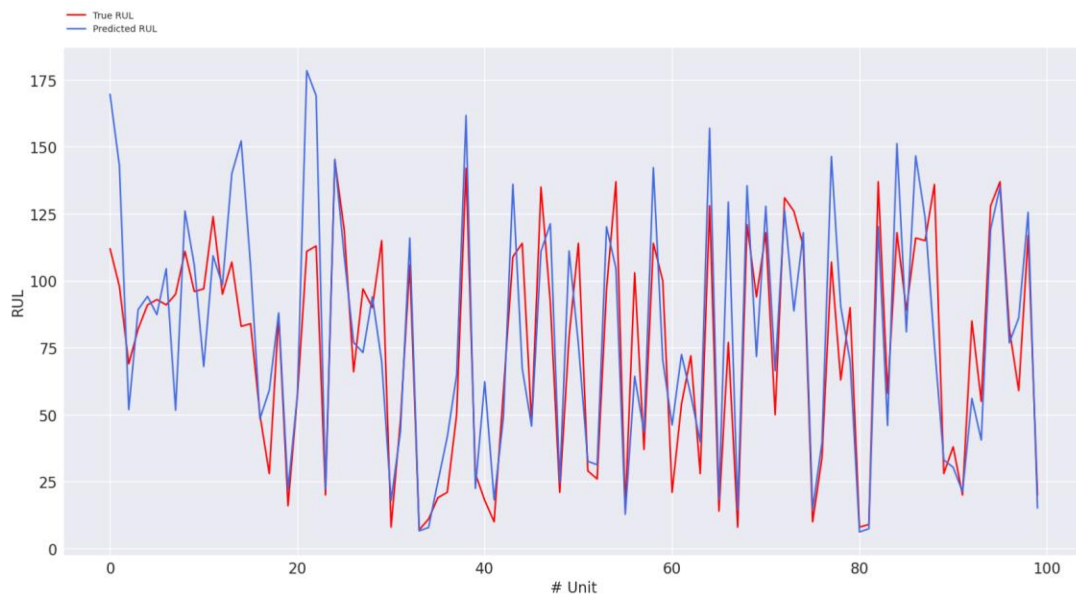

Figure 7: Comparison between true RUL and predicted RUL obtained using Random Forest Regression

Figure 8 shows the comparison between true RUL and predicted RUL that are obtained from using the LSTM model.

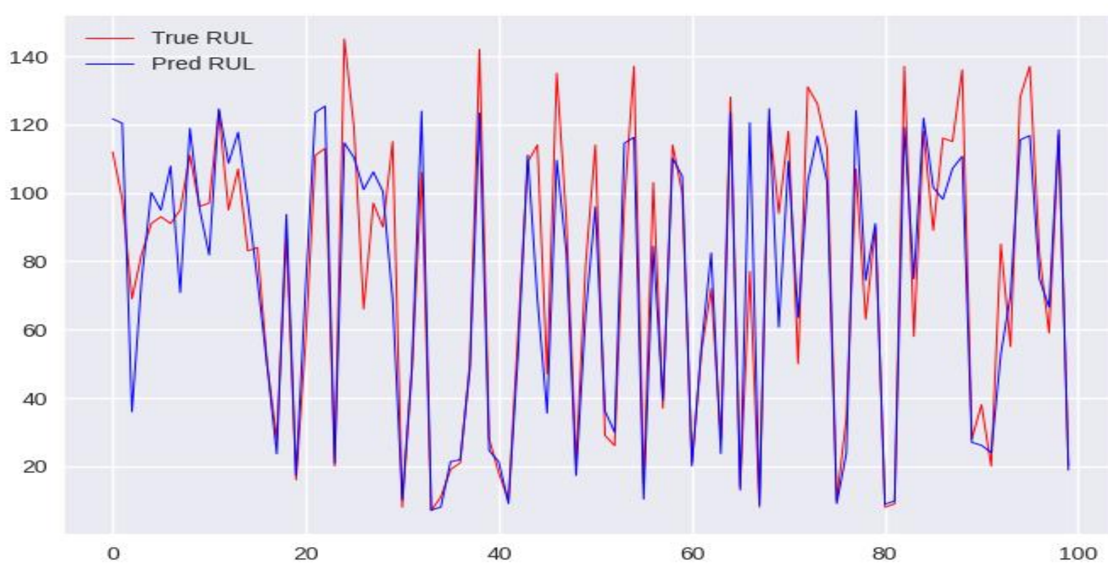

Figure 8: Comparison between true RUL and predicted RUL obtained using LSTM

The LTSM model benefits from having recurrent memory which is useful in sequential memory types such as the engine training data provided this is apparent in the plot as we can see that the prediction n more accurately maintains itself over longer spikes in the actual data.

Figure 9 gives a more thorough visualization by comparing both predictions in the same graph. This plot shows that the prediction using LSTM is a better fit than the one predicted using Random Forest Regression.

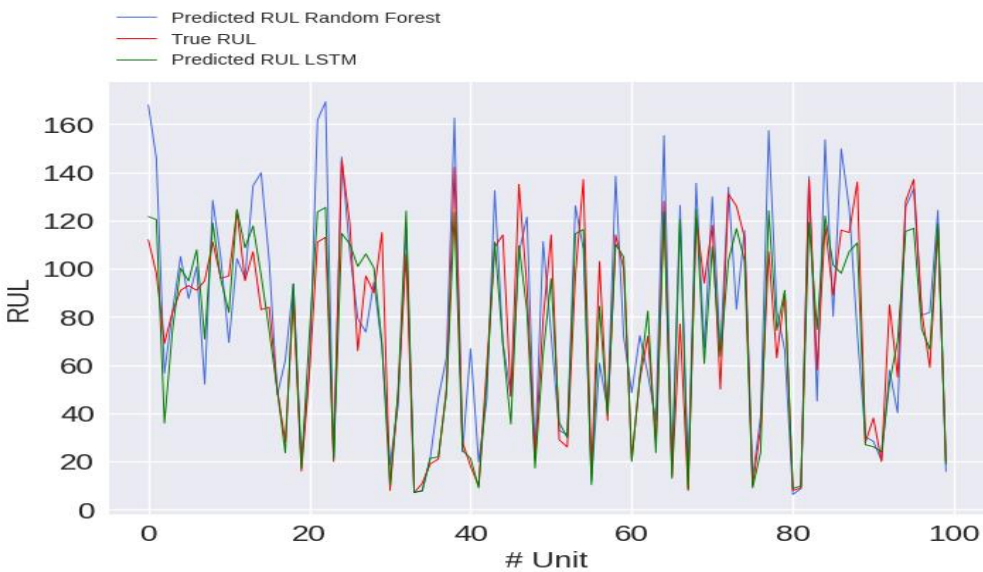

Figure 9: Prediction comparison between regression and LSTM models

To further prove the accuracy of the models, the RMSE of each model is determined using Keras' RMSE computation function. The RMSE of Random Forest and LSTM are 24.38 and 14.30 respectively.

Table 4: RMSE of Random Forest Regression and LSTM

Model	RMSE
Random Forest Regression	24.38
LSTM	14.30

V. Conclusions

The goal of this project is achieved through the successful prediction of the Remaining Useful Life (RUL) of turbofan engines from the NASA Ames C-MAPSS aircraft engine simulation dataset. The prediction was done through the use of both Regression and Neural Network models. More specifically, Random Forest Regressor and LSTM models were used, and their accuracies were compared. It was determined that the LSTM model provides a better prediction with a lower RMSE compared to the regression model. This analysis helped understand how to use Machine Learning techniques to tackle the problem of predictive maintenance.

References

[1] B² "NASA turbofan Jet Engine Data Set," *Kaggle*, 26-Jul-2019. [Online]. Available: https://www.kaggle.com/datasets/behrad3d/nasa-cmaps. [Accessed: 23-Feb-2023].

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

[3] U. Amin and K. D. Kumar, "Remaining useful life prediction of aircraft engines using hybrid model based on Artificial Intelligence Techniques," *2021 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 2021.

Appendix

*The Jupyter Notebook file and the source code for this project are attached separately.