PROGNOSIS APPROACH FOR PREDICTIVE MAINTENANCE OF NASA TURBOFAN ENGINES

Yeoh Li Phng[1]

[1]Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Arau, 02600, Malaysia
s211062164@studentmail.unimap.edu.my

ABSTRACT

The prognosis of the remaining useful life (RUL) of turbofan engine provides an important basis for predictive maintenance and remanufacturing and plays a major role in reducing failure rate and maintenance costs. This paper explores the application of machine learning, specifically Random Forest, in predicting the Remaining Useful Life (RUL) of NASA Turbofan Engines for the purpose of enhancing predictive maintenance strategies. Utilizing a dataset containing sensor readings and operational parameters, a Random Forest model is trained to forecast the RUL, providing valuable insights into the health and longevity of the engines. Additionally, the model is employed to identify the specific sensor(s) contributing to engine failure, aiding in pinpointing potential issues and streamlining maintenance efforts. The results showcase the effectiveness of Random Forest in predicting RUL and its potential in sensor-fault diagnosis for NASA Turbofan Engines.

KEYWORDS

Predictive Maintenance, Remaining Useful Life (RUL), NASA Turbofan Engines, Random Forest, Machine Learning, Maintenance Optimization, C-MAPSS, Piece-wise Linear Degradation.

1. INTRODUCTION

In the realm of aerospace engineering, the efficiency and reliability of propulsion systems, such as turbofan engines, play a pivotal role in ensuring the success of missions and the safety of aerospace vehicles. Failures in safety-critical systems, such as those found in aircraft engines or NASA turbo engine, can lead to substantial economic disruptions and carry significant social costs. Predicting the time until a system failure occurs is crucial for maintaining the functionality of these safety-critical systems and minimizing the impact on society. This challenge, known as prognostics, involves forecasting the Remaining Useful Life (RUL) of industrial assets, providing valuable insights for intelligent maintenance strategies.

Successfully implementing prognostic methods in real-life applications is essential for revolutionizing maintenance approaches, providing the capability to detect potential failures well in advance. This foresight not only enables timely interventions but also plays a pivotal role in reducing operational costs, minimizing machine downtime, and mitigating the risk of catastrophic consequences if systems are not maintained promptly [1]. The ongoing development and deployment of robust prognostic techniques significantly contribute to more efficient and cost-effective maintenance practices. These techniques are particularly crucial for ensuring the longevity and reliability of safety-critical systems, such as aircraft engines [2], exemplified by the intricate mechanisms of NASA Turbofan engines illustrated in Figure 1. By incorporating advanced prognostic methodologies, industries can proactively address maintenance needs, enhancing operational resilience and safeguarding against unforeseen challenges.

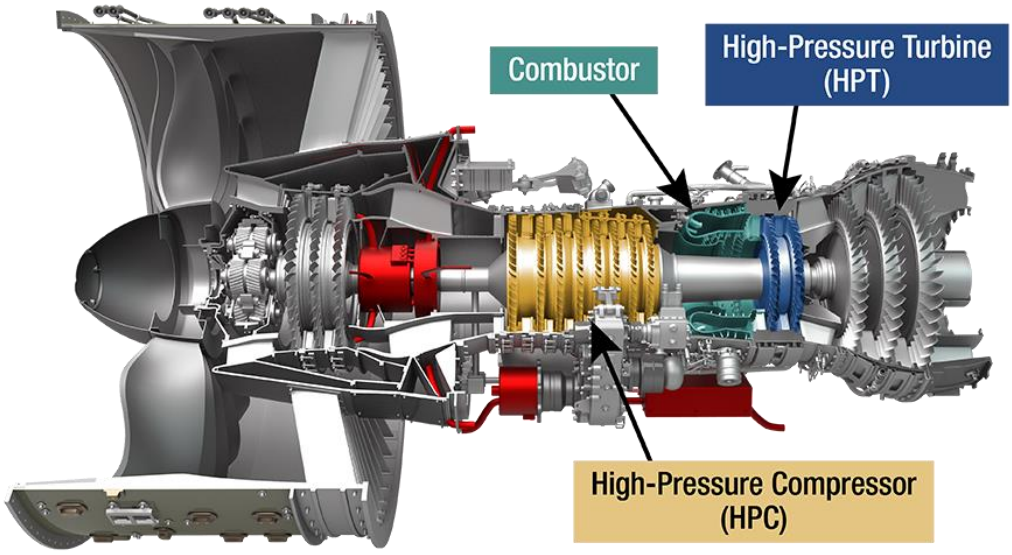

Figure 1: NASA Turbofan Engine [3]

In the expansive landscape of predictive maintenance, three predominant categories of approaches emerge, as depicted in Figure 2, each presenting distinctive methodologies for forecasting the Remaining Useful Life (RUL) of industrial assets. Knowledge-based approaches anchor themselves in domain expertise and explicit rules, harnessing the insights of experts to formulate predictions. Notably, fuzzy logic stands out as a representative of this category, leveraging linguistic variables and expert knowledge to intricately model degradation processes [1], [4]. This methodological diversity underscores the adaptability of predictive maintenance strategies, offering tailored solutions that align with the unique intricacies of different industrial contexts.

Besides, physics-based approaches leverage an understanding of the underlying system mechanics, employing techniques like Kalman filters or particle filters to model degradation based on known physical properties [4]. While theoretically robust, physics-based approaches face challenges in complex systems where the full extent of physical interactions may not be entirely known. The choice between these approaches depends on factors such as system complexity, data availability, and the depth of understanding of the underlying processes. Combining knowledge-based, data-driven, and physics-based methods can provide a holistic and robust solution to predictive maintenance challenges. In contrast, data-driven approaches, including machine learning (ML) methods, extract patterns and insights directly from historical data. ML algorithms, like Random Forest, have gained popularity due to their ability to handle complex, non-linear relationships within the data. Random Forest, a versatile and robust ML algorithm, belongs to the data-driven approach, where it learns from historical data without relying on explicit knowledge of the underlying system dynamics [5], [6].

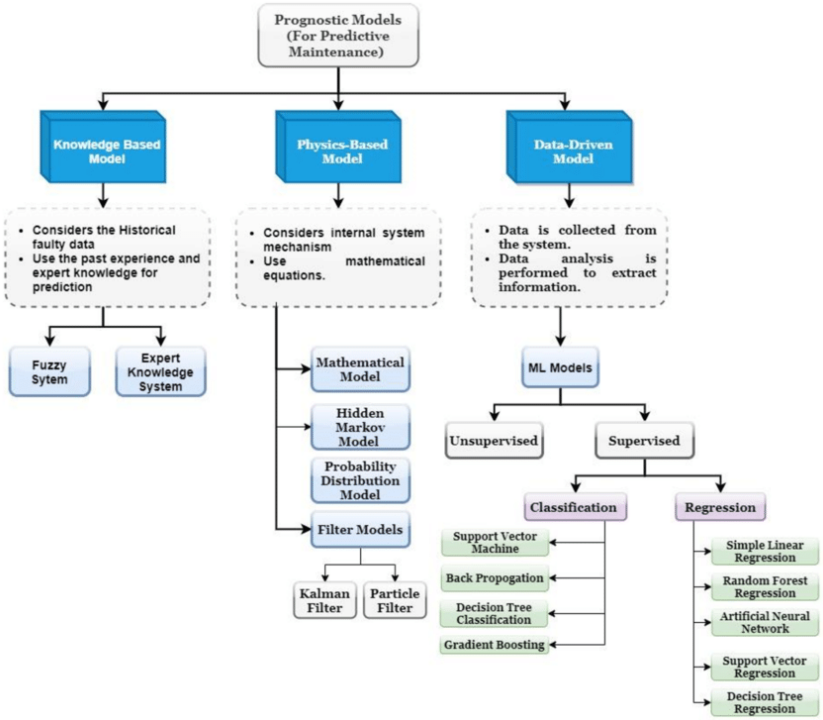

Figure 2: Prognostic Models for Predictive Maintenance [4]

In summary, this paper aiming to enhance the reliability and efficiency of predictive maintenance in the aerospace industry, adopts a data-driven approach without explicitly revealing the specific techniques used. By bridging the gap between data-driven methodologies and the intricacies of safety-critical systems, this project sets out to contribute to the advancement of predictive maintenance for NASA turbofan engines.

2. LITERATURE REVIEW

This review explores jet engine working principles, delving into NASA's Turbofan technology. It investigates AI applications, particularly in estimating remaining useful life (RUL), and emphasizes the use of Random Forest algorithms for predictive maintenance in complex systems, offering insights at the intersection of aerospace engineering and AI.

2.1. Jet Engines Working Principle

Jet engines, commonly referred to as gas turbines, operate on a fundamental principle shared across various types, such as turbofans [7]. The engine initiates its process by drawing in air at the front with a large fan. Following this, a compressor comes into play, raising the air pressure. This component, consisting of multiple blades attached to a shaft, spins at high speeds, compressing and squeezing the air. Once compressed, the air is mixed with fuel, and an electric spark ignites the mixture. The ensuing combustion produces burning gases that expand and are expelled through the nozzle at the engine's rear. This expulsion of gases generates a backward thrust, propelling both the engine and the aircraft forward. As the hot air proceeds toward the nozzle, it passes through another set of blades known as the turbine, which is connected to the same shaft as the compressor. The spinning of the turbine, extracting energy from the high-energy airflow, causes the compressor to spin as well [8].

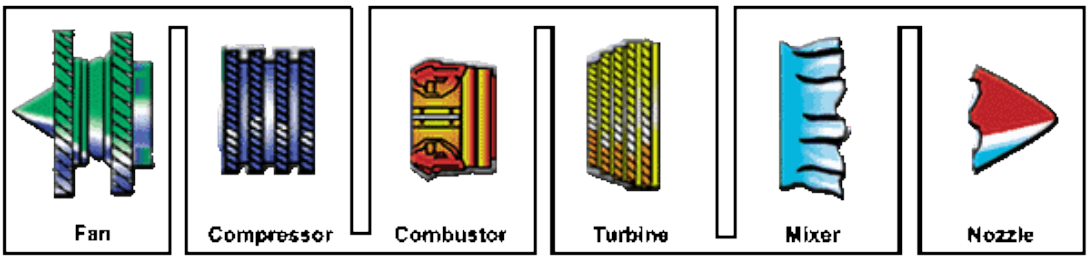

Figure 3: Parts of a Jet Engine [7]

A jet engine consists of several crucial components, each playing a specific role in its overall functionality and its can be shown in Figure 3. The fan, situated at the engine's front, is the initial component in a turbofan. This large spinning fan draws in significant quantities of air, most commonly made of titanium blades. The fan accelerates the air and splits it into two parts. One part continues through the core of the engine, where other components act upon it, while the second part bypasses the core, flowing through a duct to the back of the engine. This bypassed air contributes to the force propelling the airplane forward and also aids in quieting the engine.

The compressor, a key component in the engine core, is made up of fans with numerous blades attached to a shaft. Its role is to compress the incoming air into progressively smaller areas, resulting in an increase in air pressure and energy potential. The compressed air is then forced into the combustion chamber. The combustor is where the compressed air is mixed with fuel and ignited. The combustion process involves multiple nozzles spraying fuel into the airstream. The resulting mixture catches fire, producing a high-temperature, high-energy airflow. The combustor's interior often employs ceramic materials to create a heat-resistant chamber, as temperatures can reach up to 2700°. The high-energy airflow exiting the combustor then enters the turbine, causing the turbine blades to rotate. The turbines are connected by a shaft, which turns the blades in the compressor and spins the intake fan at the front. While this rotation extracts some energy from the high-energy flow, it is utilized to drive both the fan and the compressor [7].

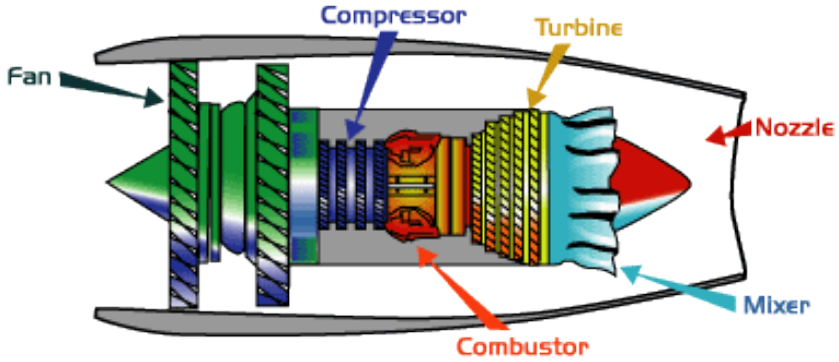

Figure 4: Jet Engines [7]

Finally, the nozzle serves as the exhaust duct of the engine, responsible for producing the thrust that propels the airplane forward. The airflow, now depleted of energy after passing through the turbine, combines with the colder air that bypassed the engine core. This combination produces an exhaust that results in a forward thrust and a complete jet engines can be shown in Figure 4. In some cases, a mixer may precede the nozzle, combining the high-temperature air from the engine core with the lower-temperature air bypassed in the fan. This mixer contributes to reducing engine noise [8].

2.2. NASA Turbofan Engine

A turbofan engine has a large fan at the front, which sucks in air. Most of the air flows around the outside of the engine, making it quieter and giving more thrust at low speeds. Most of today's airliners are powered by turbofans. In a turbojet all the air entering the intake passes through the gas generator, which is composed of the compressor, combustion chamber, and turbine. In a turbofan engine only a portion of the incoming air goes into the combustion chamber. The remainder passes through a fan, or low-pressure compressor, and is ejected directly as a "cold" jet or mixed with the gas-generator exhaust to produce a "hot" jet. The objective of this sort of bypass system is to increase thrust without increasing fuel consumption. It achieves this by increasing the total air-mass flow and reducing the velocity within the same total energy supply [8]. Figure 5 shows the turbofan engine.

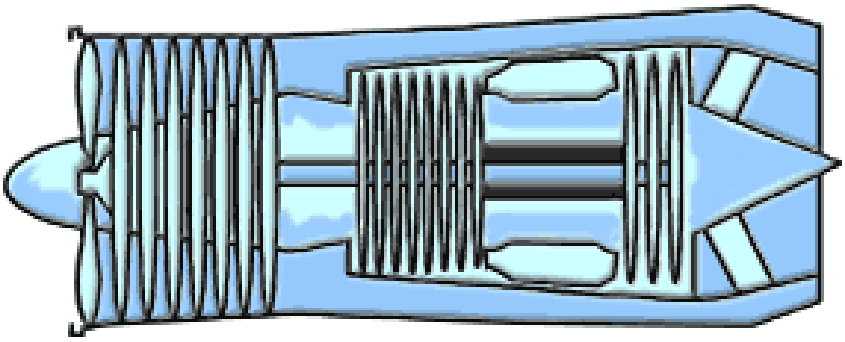

Figure 5: Turbofan Engine [8]

A turbofan engine represents a modern iteration of the fundamental gas turbine engine design. In essence, it comprises a core engine surrounded by a fan at the front and an additional turbine at the rear. The fan and fan turbine consist of numerous blades, akin to the core compressor and turbine, connected by an extra shaft. This additional turbomachinery, illustrated in green on the schematic, employs a two-spool engine configuration—designated as one "spool" for the fan and another "spool" for the core. Some advanced engines further incorporate additional spools for specific compressor sections, enhancing overall compressor efficiency [8].

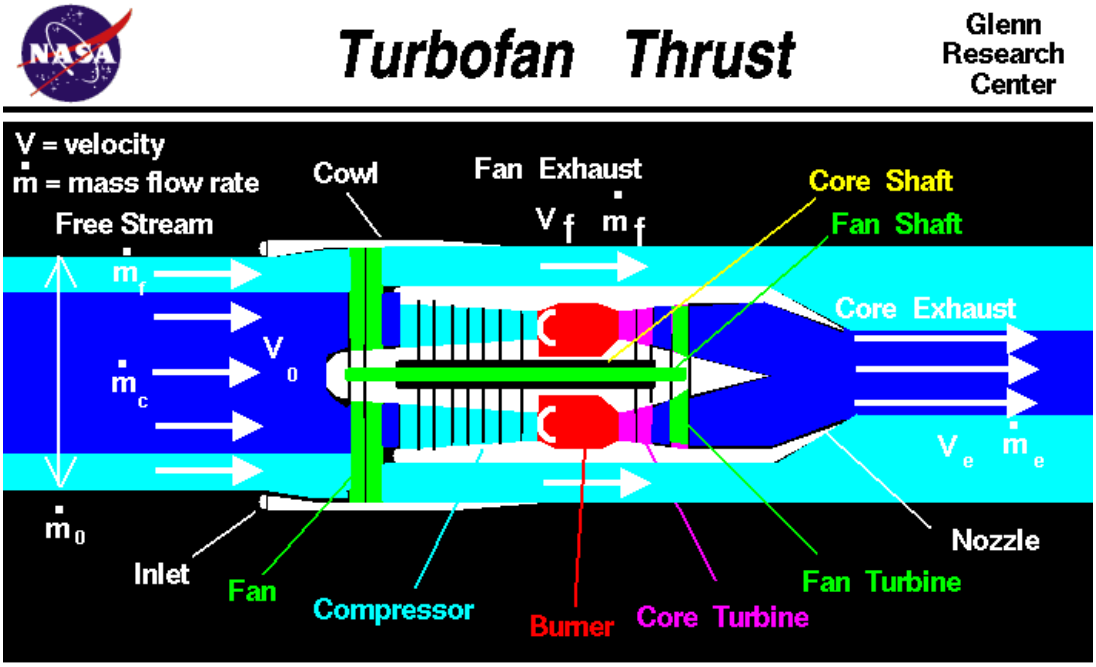

Figure 6: Turbofan Thrust [9]

The functioning of a turbofan engine can be shown in Figure 6, it involves the capture of incoming air through the engine inlet. A portion of this air, depicted in blue in the figure, traverses through the fan, proceeds into the core compressor, and then enters the burner, where it mixes with fuel, initiating combustion. The resultant hot exhaust passes through both the core and fan turbines before exiting through the nozzle, resembling the process in a basic turbojet. This airflow, identified as the core airflow (denoted by $\dot{m}_c$), represents one component of the overall thrust generation [9].

Simultaneously, the remaining portion of incoming air, shown in light blue in the figure, bypasses the engine and flows around it, akin to the airflow around a propeller. This bypassed air, directed through the fan, exhibits a slightly increased velocity from free stream and is termed the fan airflow or bypass airflow (denoted by $\dot{m}_f$). The ratio of fan airflow to core airflow is termed the bypass ratio (denoted as bpr), expressed as in (1)

$$bpr = \frac{\dot{m}_f}{\dot{m}_c} \tag{1}$$

The total mass flow rate through the engine inlet ($\dot{m}_0$) is the sum of the core and fan flows, given by (2)

$$\dot{m}_0 = \dot{m}_f + \dot{m}_c \tag{2}$$

Thrust generation in a turbofan results from both core and fan contributions. Denoting the exits of the core and fan as stations "e" and "f" respectively, and considering the free stream as station "0", the basic thrust equation for each stream contributes to the total thrust (F) expressed in (3):

$$F = \dot{m}V_f - \dot{m}_f V_0 + \dot{m}V_e - \dot{m}_c V_0 \tag{3}$$

Combining terms and utilizing the bypass ratio definition, the final thrust equation is obtained in (4):

$$F = \dot{m}V_e + bpr * \dot{m}_c V_f - \dot{m}V_0 \tag{4}$$

The efficiency of a turbofan engine stems from the fact that the addition of the fan incurs only a marginal change in the core's fuel flow rate. Consequently, a turbofan generates more thrust with nearly the same amount of fuel utilized by the core, rendering it highly fuel-efficient. High bypass ratio turbofans, in particular, approach the fuel efficiency of turboprops. Enclosed by the inlet and composed of multiple blades, the fan operates efficiently at higher speeds, distinguishing turbofans on high-speed transports from propellers employed on low-speed transports. Low bypass ratio turbofans, even more fuel-efficient than basic turbojets, find use in modern fighter planes, often equipped with afterburners to ensure efficient cruising and high thrust during combat scenarios. Despite flying at supersonic speeds, fighter planes employ inlets to slow down [9].

2.3. C-MAPSS NASA Datasets [10]

NASA C-MAPSS (Commercial Modular Aero-Propulsion System Simulation) turbofan engine degradation dataset [3] is derived from the C-MAPSS database originally created by NASA Army Research Lab. The primary control system of the turbofan engine consists of a fan controller, regulator, and limiter. The fan controller manages the normal operation of flight conditions, directing air into inner and outer culverts, as illustrated in Figure 8. A low-pressure compressor (LPC) and high-pressure compressor (HPC) supply compressed high-temperature, high-pressure gases to the combustor. The low-pressure turbine (LPT) can decelerate and pressurize air, enhancing the chemical energy conversion efficiency of aviation kerosene. High-

pressure turbines (HPT) generate mechanical energy by utilizing high-temperature and high-pressure gas to strike turbine blades. The low-pressure rotor (N1), high-pressure rotor (N2), and nozzle ensure the combustion efficiency of the engine. Figure 7 shows the C-MAPSS turbofan engine structure diagram [10].

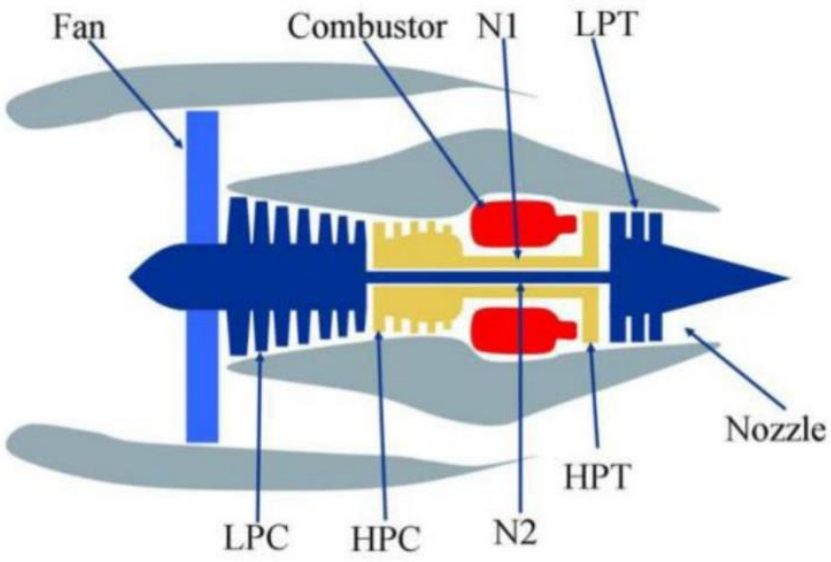

Figure 7: C-MAPSS turbofan engine structure [1]

The C-MAPSS database comprises four subsets of data labeled FD001 to FD004, generated from different time series and incorporating cumulative spatial complexity. Each data subset includes both a test data set and a training data set, and the number of engines varies within each subset. The engines in the dataset exhibit varying degrees of initial wear-and-tear, considered normal. Notably, there are three operating settings significantly influencing engine performance. Figure 8 shows the schematic representation of the C-MAPSS model [1].

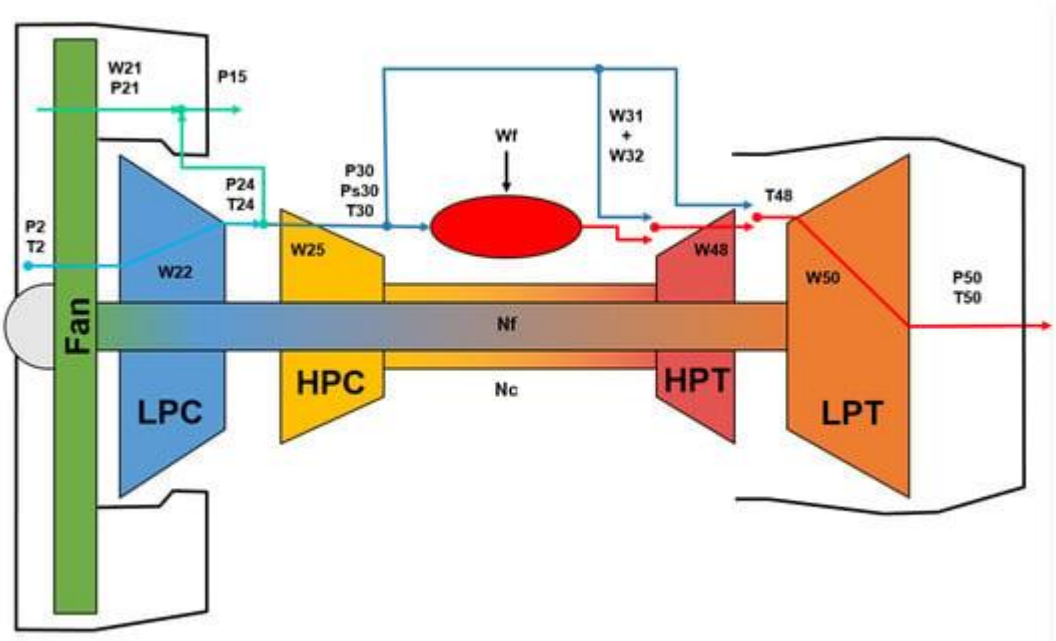

Figure 8: C-MAPSS schematic representation [9].

Table 1: Column contents of dataset [1]

Column Contents	
Serial Number	Variable Name
1	Unit Number
2	Time, in cycles
3	Operational Settings 1
4	Operational Settings 2
5	Operational Settings 3
6	Sensor Measurement 1
7	Sensor Measurement 2
8	…
9	Sensor Measurement 21

Each engine operates normally at the beginning of each time series and experiences failure at the end of the time series. In the training set, faults increase until the system fails, while in the test set, the time series concludes before system failure [11]. Each time series in the dataset consists of 21 sensor parameters and 3 other parameters that indicate the running state of the turbofan engine. The dataset is provided in the form of a compressed text file, where each row represents a snapshot taken during a single operation cycle, and each column represents a different variable. The specific contents and description of the dataset are outlined in Table 1 and 2. Table 1 shows the column contents of the datasets while Table 2 shows the data description of turbofan sensor.

Table 2: Data Description of turbofan engine sensors [1]

Turbofan engine sensors description		
Sensor Number	Sensor Description	Units
1	Fan inlet temp	°R
2	LPC outlet temp	°R
3	HPC outlet temp	°R
4	LPT outlet temp	°R
5	Fan inlet pressure	psia
6	bypass-duct pressure	psia
7	HPC outlet pressure	psia
8	Physical fan speed	rpm
9	Physical core speed	rpm
10	Engine Pressure ratio	--
11	HPC outlet Static pressure	psia
12	Ratio of fuel flow to Ps30	pps/psia
13	Corrected fan speed	rpm
14	Corrected core speed	rpm
15	Bypass ratio	--
16	Burner fuel-air ratio	--
17	Bleed Enthalpy	--
18	Required fan speed	rpm
19	Required fan conversion speed	rpm
20	High pressure turbine cool airflow	lb/s
21	Low pressure turbine cool airflow	lb/s

2.4. Artificial Intelligence in Predictive Maintenance

Artificial Intelligence (AI) stands at the forefront of revolutionizing Predictive Maintenance (PdM) strategies, introducing proactive measures to foresee potential failures and optimize asset management [13]. Within the realm of predictive maintenance, AI applications can be broadly classified into distinct types, each addressing specific facets of asset health management. This discussion will focus on comparing Failure Detection and Remaining Useful Life (RUL) Prediction, two fundamental AI-based approaches in predictive maintenance.

Failure Detection, also known as fault detection, constitutes a critical aspect of predictive maintenance. The primary objective is to identify anomalies or deviations from the normal behavior of a system that might indicate impending failures. AI models for failure detection leverage machine learning algorithms, anomaly detection models, and deep learning methods. These models analyze historical data, learning patterns to raise alerts when the system's behavior significantly deviates from the norm. The advantage lies in the early detection of anomalies, enabling timely intervention to reduce downtime and prevent catastrophic failures [12-14]. Figure 9 shows the example of failure detection.

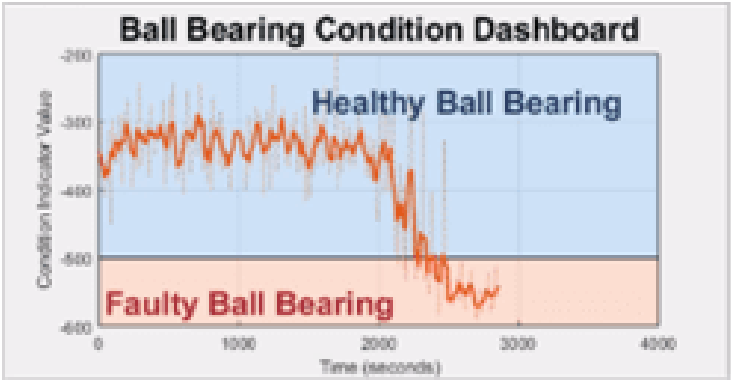

Figure 9: Failure Detection Example [12]

On the other hand, Remaining Useful Life (RUL) Prediction involves estimating the operational lifespan of an asset before a potential failure. This proactive approach empowers organizations to schedule maintenance activities optimally, minimizing downtime and reducing costs. AI techniques for RUL prediction encompass regression models, survival analysis, and machine learning algorithms like Random Forest, Support Vector Machines (SVM), or recurrent neural networks (RNNs). By analyzing historical data, these models forecast the remaining life of components, aiding in effective maintenance planning and resource optimization [13]. Figure 10 shows an example of RUL predictions.

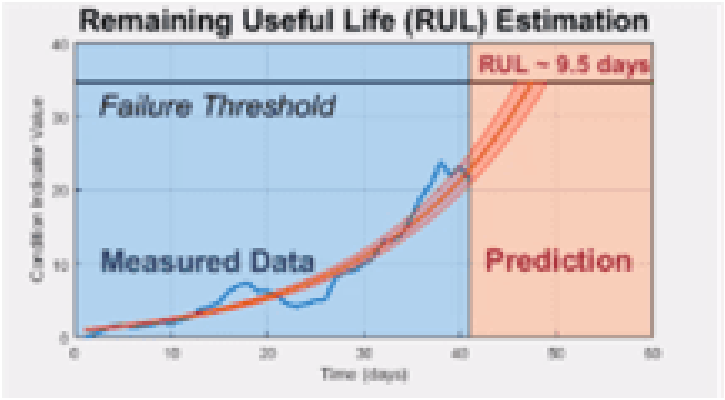

Figure 10: RUL Prediction Example [12]

Prognostics and Health Management (PHM) represent a holistic approach that combines both failure detection and RUL prediction. PHM systems continuously monitor the health of assets, predicting their future condition and providing a comprehensive understanding of asset states. Leveraging AI techniques such as sensor data fusion, model-based reasoning, and advanced machine learning algorithms, PHM ensures the continuous assessment of component health and predicts potential failures. This integrated approach enables organizations to maintain assets at peak performance levels, preventing failures and optimizing maintenance strategies.

Additionally, hybrid approaches offer a versatile solution by combining elements of failure detection and RUL prediction. These models integrate real-time monitoring for immediate fault detection with RUL prediction models for longer-term planning. By leveraging both rule-based systems and AI techniques, hybrid approaches provide a balanced strategy adaptable to diverse maintenance needs. The choice between these approaches depends on factors such as the criticality of assets, available data, and the desired maintenance strategy, allowing organizations to tailor their predictive maintenance efforts to specific operational requirements. In conclusion, AI in predictive maintenance offers a spectrum of approaches, each contributing to the enhancement of reliability and efficiency in industrial asset management [15].

2.5. Random Forest Algorithm

Random Forest, or RF for short, is the most used ensemble method. One powerful ensemble learning technique that stands out in the field of machine learning is Random Forest, which builds an ensemble of decision trees by combining random feature selection and bootstrap sampling. This approach, which was first presented by Leo Breiman [5], is very useful in applications like fraud detection because of its durability and excellent predicted accuracy. The diversity that Random Forest introduces is what makes it so effective. For example, every decision tree is trained using a different part of the dataset, and every split in the tree construction process takes a different subset of features into account [5], [6]. Figure 11 shows the decision tree of the random forest architecture.

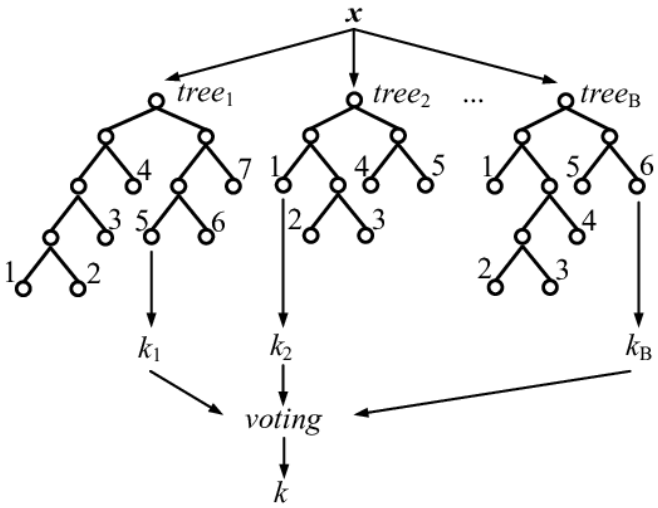

Figure 11 Random Forest Architecture [16]

This unpredictability of random decision trees reduces overfitting and improves the model's capacity to identify complex patterns in the data. Using majority voting for classification tasks and averaging for regression tasks, the ensemble of trees contributes to a strong and accurate final prediction during the prediction phase [5]. Reduced overfitting, resilience to noise and outliers, insights into feature relevance, and effective parallelization during training are a few noteworthy benefits. Random Forest's interpretability and adaptation to high-dimensional datasets are key

factors in its broad adoption in the predictive maintenance domain, facilitating the creation of trustworthy and efficient predictive maintenance systems [13], [15]. Figure 12 shows another example of the architecture structure of the random forest.

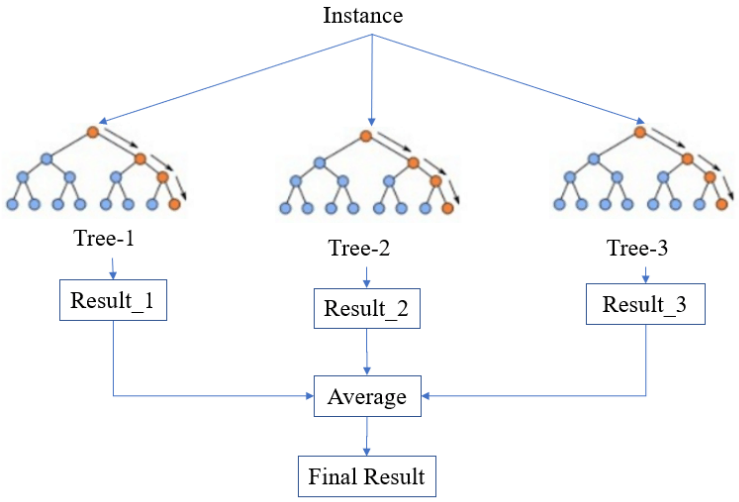

Figure 12 Architecture Structure of Random Forest [17]

3. METHODOLOGY

The predictive maintenance methodology for turbofan engines is a multi-step process designed to harness the power of machine learning for robust Remaining Useful Life (RUL) predictions. The foundation of this methodology lies in the meticulous collection and preprocessing of the dataset. The NASA C-MAPSS dataset [10], derived from the C-MAPSS database simulator by NASA Army Research Lab, encapsulates a wealth of information, including sensor measurements, operational settings, and RUL for each engine. Data preprocessing is a crucial initial step to ensure the dataset's quality and relevance. In the training stage of the model, the original turbofan engine data should be pre-processed, and the pre-processed data can be put into the model to obtain the parameters required by the model. This involves addressing missing values, eliminating features with constant sensor readings, normalizing the dataset to create a standardized foundation for subsequent analysis, feature selection, data standardization and normalization, setting the size of sliding window, and RUL label setting of training set and test set. The FD001 dataset contains 21 sensor features and 3 operating parameters (flight altitude, Mach number, and throttling parser Angle). The number of running cycles is also one of the features, so with a total of 25 features [10].

Once the dataset is preprocessed, the focus shifts to the generation of the target variable, RUL, which is pivotal for supervised machine learning. Two distinct approaches are explored for target generation: a linear degradation model and a piecewise linear degradation model. The linear degradation model assumes a consistent and gradual decline in engine health over time, while the piecewise linear model accommodates more complex degradation patterns. Piece-wise linear regression model can prevent the algorithm from overestimating RUL [13]. For an engine, equipment can be considered healthy during its initial period. The degradation process will be obvious before the whole equipment runs for a period of time or is used by a certain extent, that is, near the end of the life of the equipment. Set the normal working state of the device to a constant value, and the RUL of the device will drop linearly with time after a certain period of time. This model limits the maximum value of RUL, which is determined by the observed data in the degradation stage. The maximum RUL value of the data set observed from the degradation

phase of the experiment is set to 125, and the part exceeding 125 is uniformly specified as 125. When the critical period is reached, RUL decreases linearly as the running period increases. The choice between these models is influenced by the inherent characteristics of the dataset and the expected behavior of the engines under consideration. Figure 13 shows the piece-wise linear remaining useful life (RUL) target function.

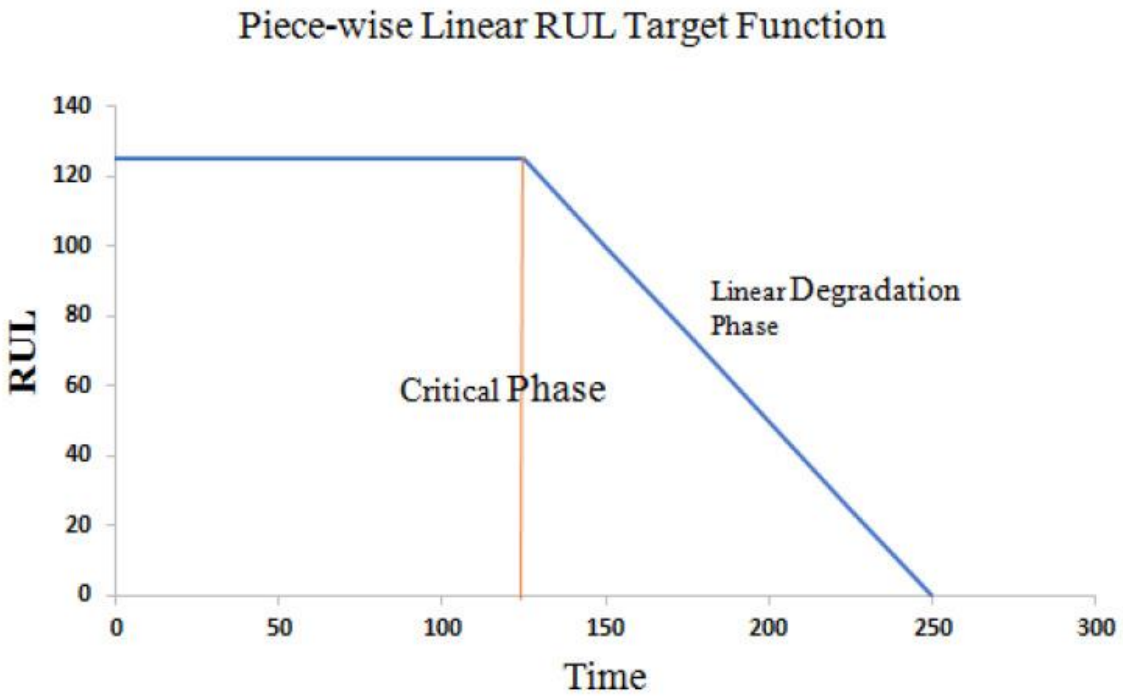

Figure 13: Piece-wise linear remaining useful life (RUL) target function [13]

Feature engineering follows, where the dataset is refined to include only relevant features for training the machine learning model. Box plots are employed to visualize sensor measurements, facilitating the identification and exclusion of sensors with constant values. This step ensures that the training data encompasses only those features that contribute meaningfully to the model's learning process. With the preprocessed and engineered data in place, the Random Forest Regression model is chosen for training. The Random Forest algorithm, known for its ability to handle complex relationships within the data, is well-suited for predicting RUL in turbofan engines. The model is trained on the preprocessed data, and hyperparameter optimization is performed using GridSearchCV to fine-tune the model's configuration for optimal performance. Following training, the model is evaluated using test data to assess its predictive capabilities. Performance metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared, are calculated to quantify the accuracy of the model in predicting RUL [15]. These metrics serve as benchmarks for assessing the model's effectiveness in capturing the health and degradation patterns of turbofan engines.

Results of the model evaluation are presented and discussed, comparing the true RUL values with the predicted RUL values. Various visualization tools, including scatter plots and line charts,

are employed to illustrate the model's performance visually. Feature importance analysis is conducted to gain insights into the contribution of different sensors to the predictions, providing a deeper understanding of the factors influencing engine health. In addition to the model evaluation, a sensitivity analysis is performed to identify the sensors that have the most significant impact on failure predictions. This analysis enhances interpretability and aids in understanding the critical factors leading to engine failures. In the final step of the methodology, the trained Random Forest Regression model is applied to make predictions on new, unseen data. The predictions are then evaluated using the same performance metrics as in the training and testing phases to ensure the model's consistency and generalizability. The methodology shown in Figure 14 shown a comprehensive overview of predictive maintenance for turbofan engines.

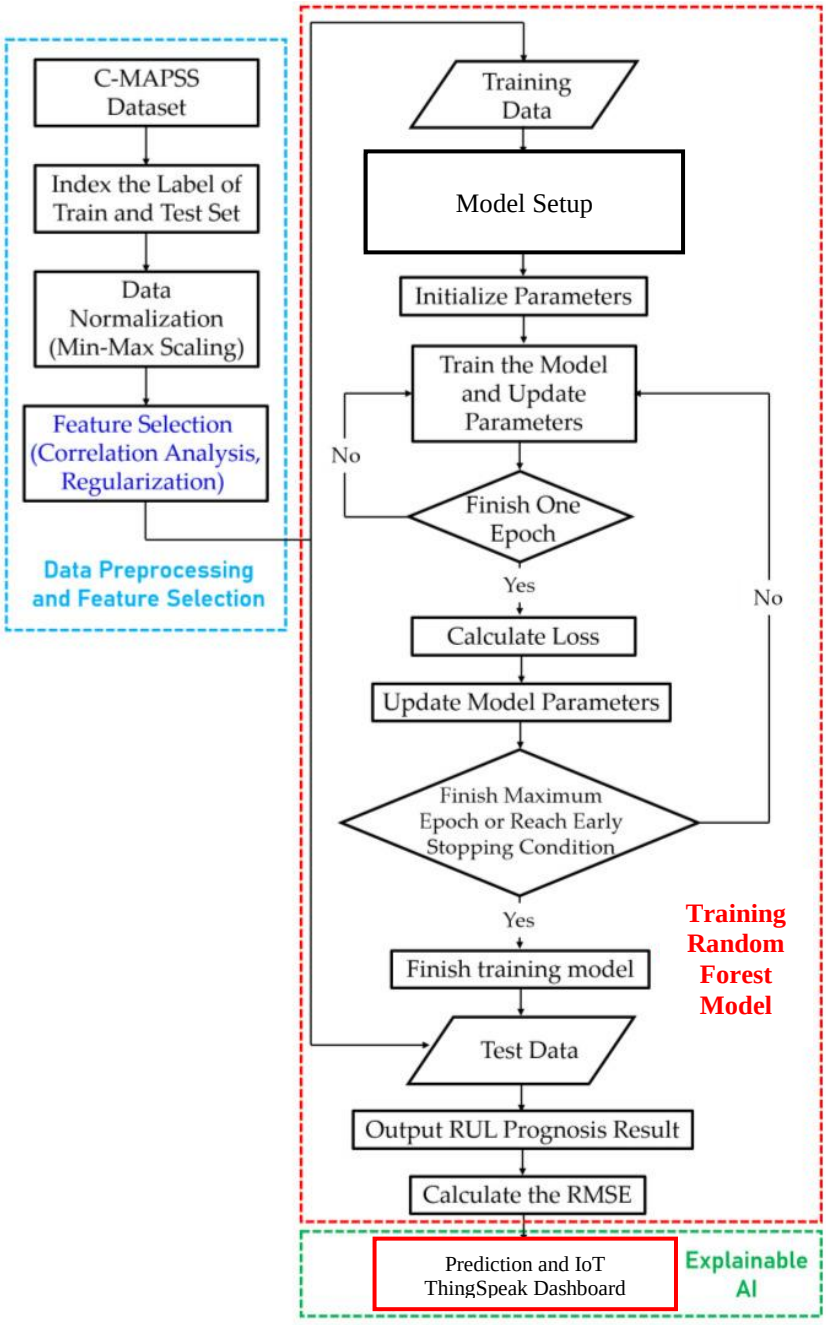

Figure 14: Predictive Maintenance Overview

4. EXPERIMENTAL RESULTS AND FINDINGS

In the early phase, the dataset is accessed via the NASA Open data portal from C-MAPSS Jet Engine Simulated database. The datasets consisted of 20,631 samples, each comprising 26 features, encompassing critical operational details and sensor measurements. The data loading and preprocessing phase, executed meticulously to ensure data integrity, facilitated a robust foundation for subsequent analyses. Noteworthy features 0-25 representing the 'unit_number,' 'time_in_cycles,' and operational settings 1 to operational settings 3, with sensor measurements ranging from 1 to 21. Figure 15 shows the first and last 5 data in the datasets. This comprehensive approach laid the groundwork for the subsequent application of degradation models to estimate the Remaining Useful Life (RUL) of Engine 2.

```
df.head()
```
✓ 0.1s

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0	1	1	-0.0007	-0.0004	100.0	518.67	641.82	1589.70	1400.60	14.62	21.61	554.36	2388.06	9046.19	1.3	47.47	521.66	2388.02	8138.62	8.4195	0.03	392	2388	100.0	39.06	23.4190
1	1	2	0.0019	-0.0003	100.0	518.67	642.15	1591.82	1403.14	14.62	21.61	553.75	2388.04	9044.07	1.3	47.49	522.28	2388.07	8131.49	8.4318	0.03	392	2388	100.0	39.00	23.4236
2	1	3	-0.0043	0.0003	100.0	518.67	642.35	1587.99	1404.20	14.62	21.61	554.26	2388.08	9052.94	1.3	47.27	522.42	2388.03	8133.23	8.4178	0.03	390	2388	100.0	38.95	23.3442
3	1	4	0.0007	0.0000	100.0	518.67	642.35	1582.79	1401.87	14.62	21.61	554.45	2388.11	9049.48	1.3	47.13	522.86	2388.08	8133.83	8.3682	0.03	392	2388	100.0	38.88	23.3739
4	1	5	-0.0019	-0.0002	100.0	518.67	642.37	1582.85	1406.22	14.62	21.61	554.00	2388.06	9055.15	1.3	47.28	522.19	2388.04	8133.80	8.4294	0.03	393	2388	100.0	38.90	23.4044

```
df.tail()
```
✓ 0.1s

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
20626	100	196	-0.0004	-0.0003	100.0	518.67	643.49	1597.98	1428.63	14.62	21.61	551.43	2388.19	9065.52	1.3	48.07	519.49	2388.26	8137.60	8.4956	0.03	397	2388	100.0	38.49	22.9735
20627	100	197	-0.0016	-0.0005	100.0	518.67	643.54	1604.50	1433.58	14.62	21.61	550.86	2388.23	9065.11	1.3	48.04	519.68	2388.22	8136.50	8.5139	0.03	395	2388	100.0	38.30	23.1594
20628	100	198	0.0004	0.0000	100.0	518.67	643.42	1602.46	1428.18	14.62	21.61	550.94	2388.24	9065.90	1.3	48.09	520.01	2388.24	8141.05	8.5646	0.03	398	2388	100.0	38.44	22.9333
20629	100	199	-0.0011	0.0003	100.0	518.67	643.23	1605.26	1426.53	14.62	21.61	550.68	2388.25	9073.72	1.3	48.39	519.67	2388.23	8139.29	8.5389	0.03	395	2388	100.0	38.29	23.0640
20630	100	200	-0.0032	-0.0005	100.0	518.67	643.85	1600.38	1432.14	14.62	21.61	550.79	2388.26	9061.48	1.3	48.20	519.30	2388.26	8137.33	8.5036	0.03	396	2388	100.0	38.37	23.0522

Figure 15: first and last 5 data

The application of a linear degradation model to Engine 2, executed to generate a visual representation of RUL over operational cycles. The resulting graph (Figure left) portrayed a discernible decline in RUL, signifying a consistent degradation pattern. Subsequently, a piecewise linear degradation model was introduced to capture nuanced degradation phases, revealing distinct patterns in the RUL over cycles (Figure Right). This comparative analysis provided insights into the engine's health and can be shown in Figure 16.

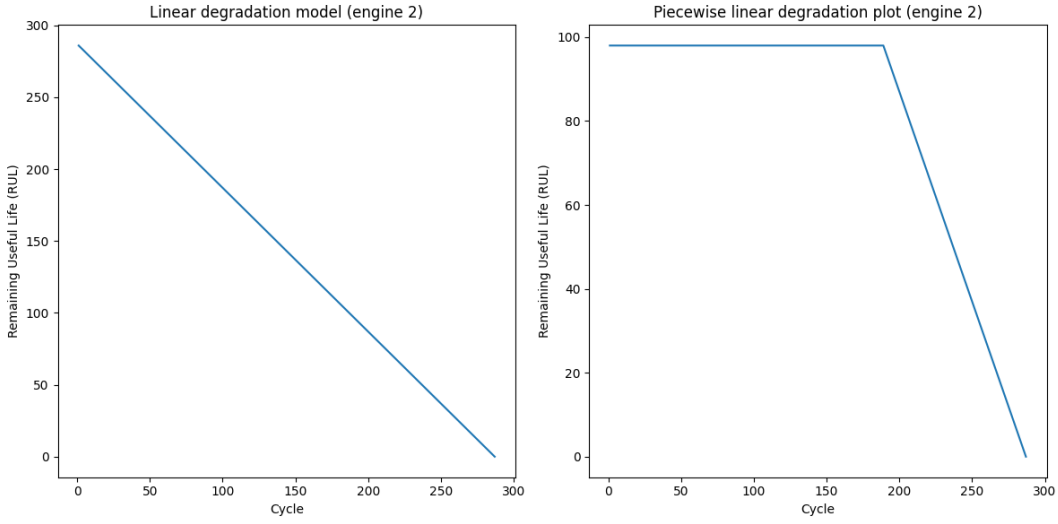

Figure 16: Linear Degradation model vs Piecewise Linear Degradation

To preprocess the training data for algorithm training, specific columns are unnecessary, including those containing engine and cycle information (columns 0 and 1) and operational settings (columns 2 to 4). This study opts not to utilize operational setting values for training, focusing solely on sensor measurements only. To gain a preliminary understanding of the data distribution within these sensor measurement columns, density plots are employed. The generated boxplots, displayed in Figure 17, provide a visual representation of the statistical summary.

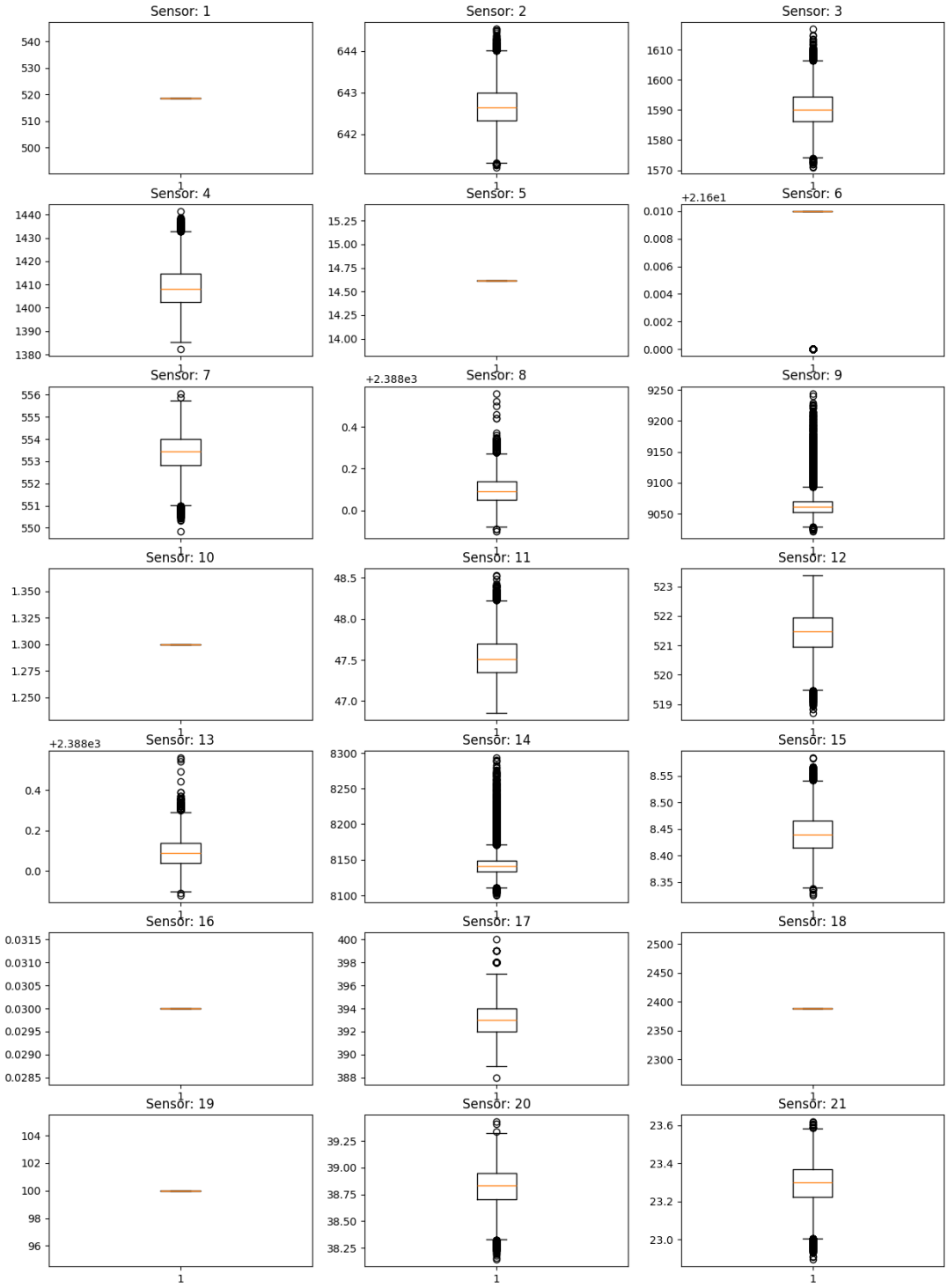

Figure 17: Statistical Summary of Sensor Measurement.

Upon analysis of the density plots, it is observed that columns 5, 9, 14, 20, 22, and 23 representing the Sensor 1, 5, 6, 10, 16, 18 and 19, exhibit constant values, suggesting limited variability across operational cycles. Additionally, suspicions arise regarding the diversity of values in Column 10. To further investigate, an examination of Column 10 is warranted. Columns with constant values are deemed uninformative for algorithm training, as they contribute minimal variation. Furthermore, during the normalization process, which involves subtracting the mean and dividing by the standard deviation, issues may arise if a column possesses a constant value, leading to a zero standard deviation. As a remedy, and in adherence to best practices, these columns containing constant values will be dropped from the dataset to enhance the algorithm's effectiveness and prevent normalization-related challenges.

After removing the unnecessary features, the next procedure is to preprocess data for multiple engines, involving the removal of unnecessary columns and the creation of batches for training and testing. The approach addresses challenges such as windowing, shift, and feature selection, crucial for training machine learning models on sequential data. The resulting processed datasets are shuffled for training, and the dimensions of the training and test data are printed for verification. This modular and systematic data preparation lays the foundation for subsequent model training and evaluation in the context of predicting remaining useful life for engine components.

Next, The Random Forest (RF) regression model is trained using the processed training data. The entire process, from training the model to obtaining predictions and calculating mean values, is executed efficiently in 10.6 seconds, showcasing the computational speed and effectiveness of the Random Forest algorithm in this predictive maintenance context. The predictive performance of the Random Forest regression model on the test dataset is evaluated through various metrics. The Root Mean Squared Error (RMSE) is calculated to be 19.02, providing a measure of the average magnitude of the model's prediction errors. The Mean Absolute Error (MAE) is computed as 14.29, representing the average absolute difference between the predicted and true remaining useful life values. Additionally, the R-squared value, a measure of how well the model explains the variance in the data, is determined to be 0.79. These metrics which can be shown in Figure 18 collectively indicate that the Random Forest model exhibits a reasonable predictive accuracy, capturing a substantial portion of the variability in the remaining useful life of the engines based on the given features. However, further analysis and potential model refinement could be explored to enhance predictive performance. Figure 19 and 20 shows the graph of True vs Predicted RUL, while showing the Feature Importance in Figure 21.

```
RMSE:  19.01973068765989
Mean Absolute Error:  14.288599999999999
R-squared:  0.7905168083153984
```

Figure 18: Performance Metrics of Random Forest (01)

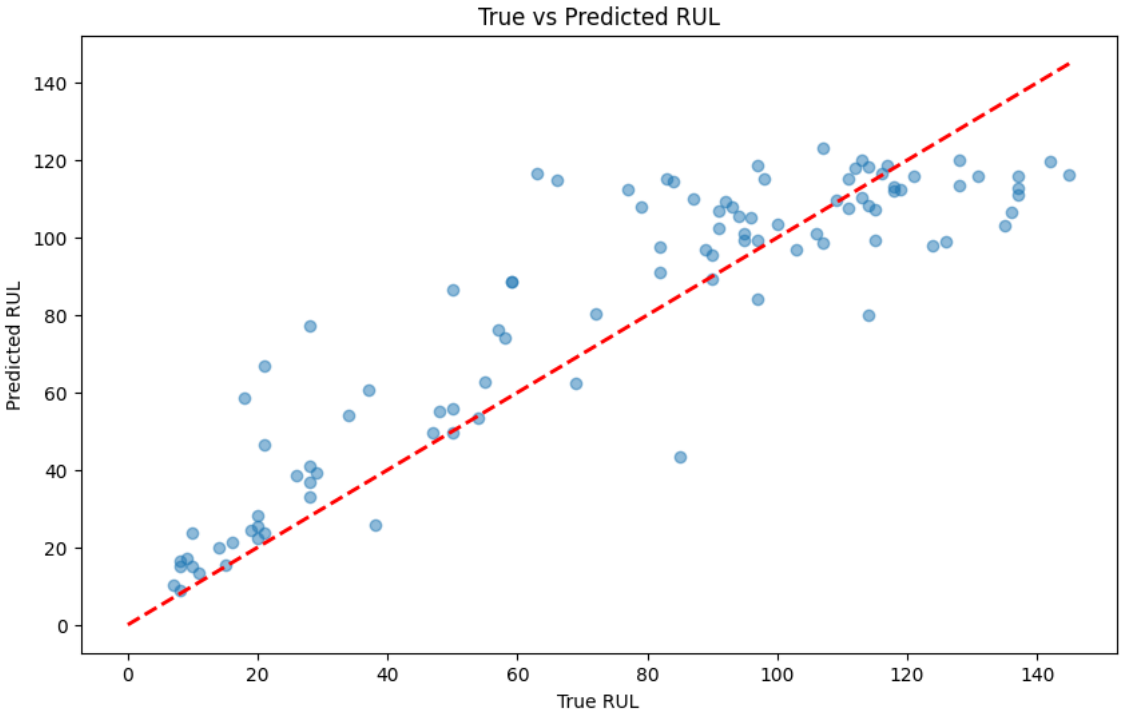

Figure 19: Performance Metrics of Random Forest (02)

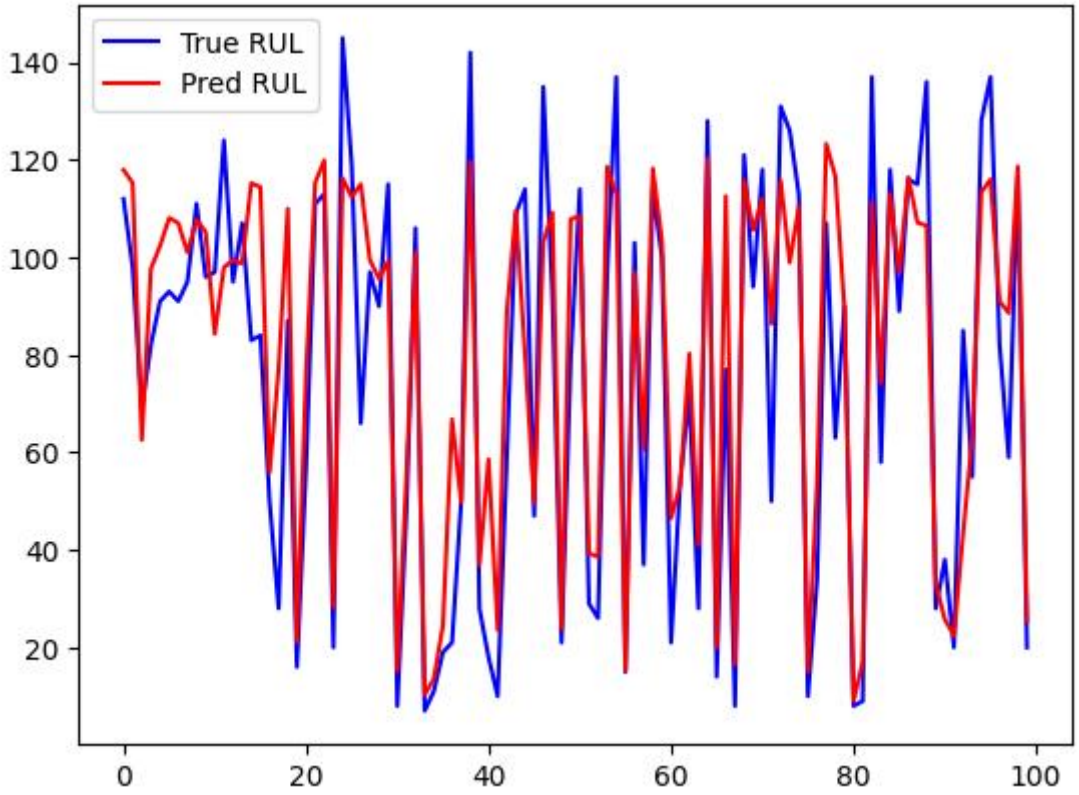

Figure 20: Performance Metrics of Random Forest (03)

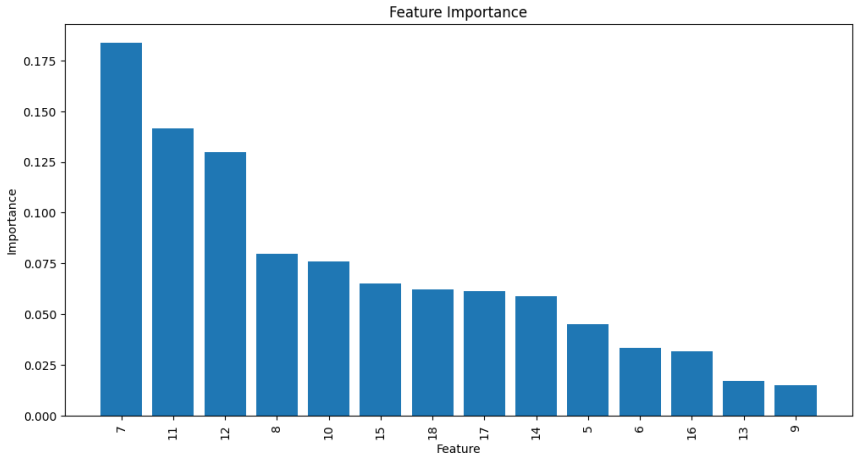

Figure 21: Feature Importance based on predictions

In the inference phase, a critical aspect of evaluating the model's performance involves randomly selecting a row from the processed test data, representing sensor values for a specific engine. This selected row is utilized to predict the Remaining Useful Life (RUL) using the trained Random Forest Regressor model. The model generates a continuous RUL prediction, including decimal points, reflecting its regression nature. Additionally, the failure status of the selected row is determined by comparing the rounded RUL prediction with a predefined failure threshold. If the rounded RUL falls below this threshold, the component is deemed to have failed; otherwise, it is considered operational.

For example, if the model predicts a RUL of 60, which falls below the set threshold of 75 (indicating that maintenance is required when the model predicts RUL below 75), the failure status will be marked as 'Fail.' Subsequently, the system identifies the sensor contributing most significantly to the failure prediction. This information is crucial for maintenance decisions, as it pinpoints the specific sensor values that have the major impact on the failure. This comprehensive insight empowers users to make informed decisions about maintenance or replacement of specific sensors, ultimately contributing to an increase in the Remaining Useful Life. Figure 22 illustrates the example of 'Fail' Health Status of sensors, showcasing the predictive maintenance information, while Figure 23 exemplifies the 'Pass' Health Status of sensors for operational components.

```
Randomly picked row index: 122
Remaining Useful Life (RUL) Prediction: 60.0
Failure Status: Fail
Sensor causing failure: 8
```

Figure 22: Fail Health Status

```
Randomly picked row index: 436
Remaining Useful Life (RUL) Prediction: 118.0
Failure Status: Pass
```

Figure 23: Pass Health Status

To enhance real-time monitoring and facilitate a comprehensive Internet of Things (IoT) ecosystem, the predictions were seamlessly integrated into a ThingSpeak dashboard displaying the real time information of the remaining useful life and the sensor values which causing the engine failures. This integration allows for continuous surveillance of critical turbofan engine parameters, enabling efficient predictive maintenance by identifying impending failures and their root causes. Figure 24 shows the IoT ThingSpeak Dashboard.

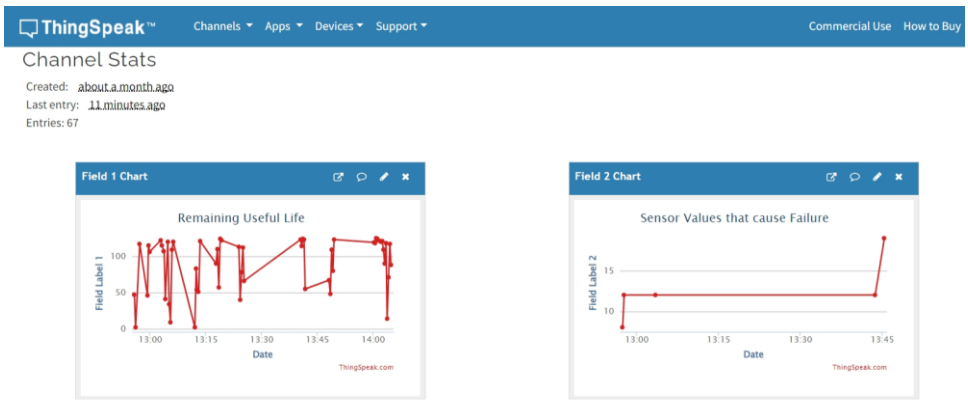

Figure 24: IoT ThingSpeak Dashboard

5. CONCLUSIONS

In conclusion, the developed predictive maintenance model, leveraging a Random Forest Regressor, exhibits promising capabilities in evaluating the Remaining Useful Life (RUL) of engines based on sensor measurements. Through meticulous preprocessing, which includes windowing and feature selection, the model is trained on a dataset encompassing operational data and corresponding RUL values. Notably, the model demonstrates efficiency, achieving a runtime of 10.6 seconds during training. Evaluation on test datasets yields notable performance metrics, including a Root Mean Squared Error (RMSE) of 19.02, Mean Absolute Error (MAE) of 14.29, and an R-squared value of 0.79. The model's predictive capacity is further explored by randomly selecting 15 data from the testing data, predicting RUL, determining failure status, and identifying the sensor contributing to failure. Importantly, the developed framework has successfully implemented a complete Internet of Things (IoT) system, where it will be integrating with ThingSpeak. This IoT system enables real-time monitoring, providing valuable insights into critical engine parameters and fostering proactive maintenance strategies. Future work may focus on refining the model, exploring advanced feature engineering techniques, and addressing potential post-processing considerations for RUL predictions.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my course lecturers, Dr. Ammar and Dr. Sukhairi, for their invaluable guidance and support throughout the "IoT and Data Analytics" course. Their expertise and commitment to fostering a conducive learning environment have been instrumental in my academic journey. I appreciate the opportunities provided to delve into the intricacies of data analytics and IoT technologies, culminating in the writing of this technical paper. This exposure has not only enhanced my understanding of the subject matter but has also allowed me to apply theoretical concepts to real-world scenarios. I am truly thankful for the enriching learning experience and the chance to contribute to the field through this paper.

REFERENCES

[1] C. Peng, Y. Chen, Q. Chen, Z. Tang, L. Li, and W. Gui, "A Remaining Useful Life Prognosis of Turbofan Engine Using Temporal and Spatial Feature Fusion," Sensors, vol. 21, no. 2, p. 418, Jan. 2021. [Online]. Available: https:// www.ncbi.nlm.nih.gov/pm c/articles/PMC7827555/.

[2] M. A. Chao, C. Kulkarni, K. Goebel, and O. Fink, "Aircraft Engine Run-to-Failure Dataset under Real Flight Conditions for Prognostics and Diagnostics," Data, vol. 6, no. 1, p. 5, Jan. 2021. [Online]. Available: https://doi.org/10.3390/data6010005.

[3] NASA, "HyTec Technical Portfolio," [Online]. Available: https://www.nasa.gov/directorates/armd/hytec-technical-portfolio/. [Accessed: Jan 27, 2024].

[4] S. Sayyad, S. K. V. C, A. Bongale, and A. Bongale, "Estimating Remaining Useful Life in Machines Using Artificial Intelligence: A Scoping Review," Symbiosis Institute of Technology, January 2021. [Online]. Available: https://www.researchgate.net/publication/348728876_Estimating_Remaining_Useful_Life_in_Machines_Using_Artificial_Intelligence_A_Scoping_Review. [Accessed: Jan. 27, 2024].

[5] L. Breiman, "Random forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324/METRICS

[6] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, Jun. 2016, doi: 10.1007/S11749-016-0481-7.

[7] NASA Glenn Research Center. "Aerospace Concepts - Jet Engines." https://www.grc.nasa.gov/www/k-12/UEET/StudentSite/engines.html [Accessed: Jan. 27, 2024].

[8] NASA Glenn Research Center. "How Jet Engines Work." https://www.grc.nasa.gov/www/k-12/airplane/Animation/turbtyp/etff.html [Accessed: Jan. 27, 2024].

[9] NASA Glenn Research Center. "Turbofan Engine." https://www.grc.nasa.gov/www/k-12/airplane/turbfan.html [Accessed: Jan. 27, 2024].

[10] NASA. "CMAPSS Jet Engine Simulated Data." https://data.nasa.gov/Aerospace/CMAPSS-Jet-Engine-Simulated-Data/ff5v-kuh6/about_data [Accessed: Jan. 27, 2024].

[11] Saxena A., Goebel K., Simon D., Eklund N. Damage propagation modeling for aircraft engine run-to-failure simulation; Proceedings of the 1st International Conference on Prognostics and Health Management (PHM08); Denver, CO, USA. 6–9 October 2008.

[12] MathWorks. "What Is Predictive Maintenance?" [Online]. Available: https://www.mathworks.com/help/predmaint/gs/what-is-predictive-maintenance.html [Accessed: Jan. 27, 2024].

[13] L. Viale, A. P. Daga, A. Fasana, and L. Garibaldi, "Least squares smoothed k-nearest neighbors online prediction of the remaining useful life of a NASA turbofan" Revised 12 January 2023, Accepted 24 January 2023, Available online 1 February 2023.

[14] U. Thakkar and H. Chaoui, "Remaining Useful Life Prediction of an Aircraft Turbofan Engine Using Deep Layer Recurrent Neural Networks,"Actuators, vol. 11, no. 3, p. 67, 2022. https://doi.org/10.3390/act11030067

[15] A. Muneer et al., "Deep-Learning Based Prognosis Approach for Remaining Useful Life Prediction of Turbofan Engine," Symmetry, vol. 13, no. 10, p. 1861, 2021. [Online]. Available: https://doi.org/10.3390/sym13101861

[16] "A general random forest architecture. | Download Scientific Diagram." Accessed: Dec. 06, 2023. [Online]. Available: https://www.researchgate.net/figure/A-general-random-forest-architecture_fig1_301411533

[17] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," *IEEE Access*, vol. 6, pp. 14277–14284, Feb. 2018, doi: 10.1109/ACCESS.2018.2806420.