Predicting the remaining useful life (RUL) of NASA's turbo engines using machine learning

Prajil Bhagat
2022359
prajil22359@iiitd.ac.in

Wasif Ali
2022583
wasif22583@iiitd.ac.in

Ashish Bargoti
2022114
ashish22114@iiitd.ac.in

Subham Maurya
2022510
subham22510@iiid.ac.in

Abstract

Engine deterioration has typically been associated with risks of dramatic losses, including the probable loss of a space mission and loss of life. The current practice of routine periodic checks of engines before each flight is not only extremely expensive but does not even avert sudden in-flight failure. This project intends to correct some of the flaws that have limited the application of methods to predict the RUL of parts of the engine to develop a machine learning model that will predictively point out with necessary accuracy the Remaining Useful Life of engine parts. The model provides precursory warnings related to possible failure by using historical data from the engine, thereby allowing for proactive maintenance strategies. Such an approach will increase the safety of operations by reducing the chances of critical engine failures as well as minimizing the costs required for maintenance, thereby providing a more efficient solution for engine maintenance. The source code for this project is available on GitHub at https://github.com/Subham-Maurya/TurboEngine-RUL-Estimation.

1. Introduction

The gradual degradation of engine components over time poses significant challenges in aviation, where engine failures can lead to mission-critical disruptions or even catastrophic outcomes. Current maintenance strategies involve routine inspections before each flight, which, while necessary, are both costly and inefficient. Despite these checks, unforeseen in-flight engine failures remain a possibility, presenting risks to both safety and operational continuity.

The central problem addressed in this project is the inability of existing maintenance practices to predict the exact Remaining Useful Life (RUL) of engine components. Without accurate RUL predictions, maintenance schedules are either overly conservative, leading to unnecessary costs, or inadequate, increasing the risk of unexpected failures.

Our goal is to develop a machine learning model that predicts the RUL of engine parts with high accuracy, enabling proactive maintenance. By leveraging historical engine performance data, the model aims to provide early warnings of potential failures, reducing downtime and maintenance costs, while significantly improving safety and reliability.

2. Literature Survey

2.1. Damage Propagation Modeling for Aircraft Engine Prognostics

Abhinav Saxena[1], Kai Goebel[2], Don Simon[3], Neil Eklund[4]

Discusses how damage propagation in aircraft gas turbine engines can be modeled for the purpose of estimating Remaining Useful Life (RUL). The model uses the C-MAPSS (Commercial Modular Aero-Propulsion System Simulation) software to simulate engine performance by varying flow and efficiency parameters under various operating conditions. Damage propagation is represented as an exponential deterioration in efficiency and flow, constrained by predefined thresholds. A health index is used to determine engine failure by monitoring multiple sensor outputs. This approach was used to generate data for a prognostics competition, aimed at developing algorithms for predicting failures in a data-driven manner. The competition focused on penalizing late failure predictions more heavily than early ones, with implications for system safety and maintenance scheduling. The paper emphasizes the significance of realistic degradation modeling in improving prognostic algorithms for complex systems such as aircraft engines.

2.2. Recurrent Neural Networks for Remaining Useful Life Estimation

Felix O. Heimes

Presents a solution to the 2008 IEEE Prognostics and Health Management challenge using a Recurrent Neural Network (RNN) to estimate the Remaining Useful Life (RUL) of complex systems. The RNN is trained with an Extended Kalman Filter (EKF) method and back-propagation through time to model system degradation. The solution demonstrated improved accuracy in predicting RUL by incorporating time-domain filtering and internal memory, addressing issues such as noise in sensor data and the temporal nature of the input. This approach placed second in the competition, achieving high accuracy with relatively low model complexity. The model's performance was further enhanced using evolutionary algorithms to optimize the neural network structure.

3. Dataset Overview

The dataset contains 26 columns, organized as follows:

- **Column 1:** Engine ID
- **Column 2:** Time in cycles
- **Columns 3-5:** Operational settings, representing varying engine conditions
- **Columns 6-26:** Sensor measurements from 21 different sensors

Each engine's data forms a multivariate time series, where engines start with different initial wear conditions and experience operational variability. Additionally, the dataset reflects real-world challenges by incorporating sensor noise, making the prediction of engine failure more realistic.

3.1. Experimental Scenario

The dataset is divided into training and test sets:

- **Training Data:** Engines are run until failure, providing complete life cycle data.
- **Test Data:** Time series data ends before failure, requiring the prediction of the Remaining Useful Life (RUL) for each engine.

The primary challenge is to predict RUL for the engines in the test set, while accounting for operational variability, sensor noise, and degradation patterns.

3.2. Data Preprocessing

To prepare the dataset for model training, several preprocessing steps were performed:

3.2.1 Data Cleaning and Feature Engineering

The dataset was meticulously examined for missing values and duplicates to ensure data quality and integrity. No missing values were detected, and any duplicate entries were removed.To enhance the predictive power of the model, additional features were derived from the raw sensor data for each engine. These included:

- Mean, standard deviation, and median of sensor values.
- Statistical features such as skewness to capture trends and variations in sensor measurements over time.

3.2.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to gain deeper insights into the dataset and prepare it for modeling. Several techniques were employed to understand the structure, distribution, and relationships within the data:

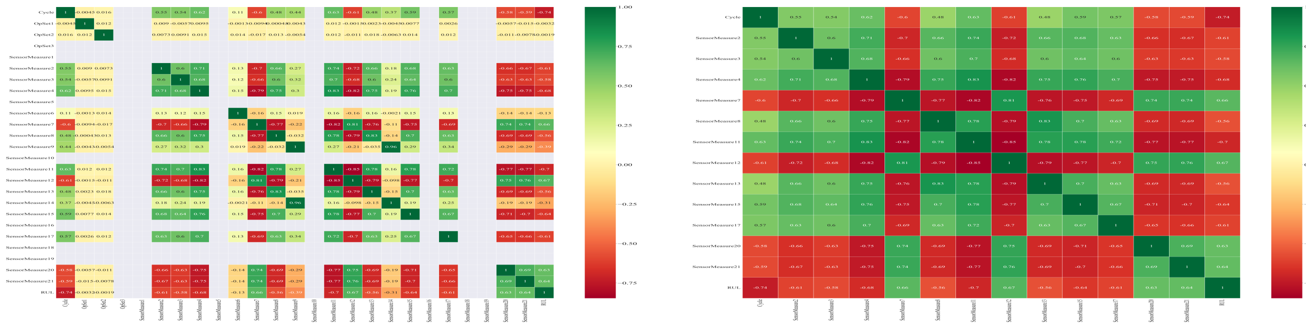
Descriptive Statistics The `describe()` function was used to generate descriptive statistics for all features in the dataset. This provided essential summaries such as the mean, standard deviation, minimum, and maximum values for each sensor and operational setting.

Constant Feature Removal Features with constant values across all records provide no useful information for the machine learning models.Any feature that exhibited the same value for all instances was removed from the dataset, as they do not contribute to learning patterns or differentiating between engine conditions.

Correlation Analysis A correlation heatmap was generated to analyze the relationships between the various features and the target variable, RUL.Only features with a correlation coefficient of at least 0.5 were retained, as they were considered significant for predicting RUL.

Outlier Detection The box plots helped in identifying such anomalies and allowed for a decision on whether to remove or retain them based on their potential impact on the model's performance.

These EDA steps ensured that the dataset was well-understood, cleaned, and optimized for the machine learning model development, laying a strong foundation for accurate RUL prediction.



(a) All Features Included

(b) With Selective Features

Figure 1. Correlation Matrix

3.3. Prediction Target

The target variable in this project is the Remaining Useful Life (RUL) of each engine, which represents the number of operational cycles remaining until failure. The task is to accurately predict RUL using the historical sensor and operational data, which will enable proactive maintenance and minimize the risk of unexpected engine failures.

4. Methodology

4.1. Problem Formulation

The primary objective of this study is the prediction of Remaining Useful Life (RUL) for NASA's turbo engines, a critical task for predictive maintenance and operational planning. RUL prediction involves estimating the number of operational cycles remaining before the engine reaches the end of its life.

To better understand and address the complexity of this task, the problem was approached in two phases:

- **Classification Phase:** The problem was initially simplified as a classification task, categorizing engine health into discrete labels.
- **Regression Phase:** The task was then addressed in its original form as a regression problem to predict the continuous RUL value.

This phased approach enabled progressive exploration of the data and model capabilities while ensuring actionable insights at each step.

4.2. Phase 1: Classification Task

To simplify the problem initially, the prediction of RUL was formulated as a classification task. The engine's condition was categorized into three discrete labels based on the *Life Ratio (LR)*:

$$LR = \frac{\text{Current Cycle}}{\text{End Cycle (EOL)}}$$

The classification labels were defined as follows:

- **Good Condition (Label 0):** $LR \leq 0.6$
  The engine is in optimal working condition, with minimal degradation.
- **Moderate Condition (Label 1):** $0.6 < LR \leq 0.8$
  The engine shows signs of degradation but still has significant operational life remaining.
- **Warning Condition (Label 2):** $LR > 0.8$
  The engine is near the end of its lifecycle and requires immediate maintenance.

4.2.1 Model Training: Random Forest Classifier

To classify the engine's condition, a *Random Forest Classifier* was employed due to its robustness and interpretability. The training process included the following steps:

- **Feature Selection:** Relevant sensor data and operational settings were identified using correlation analysis and domain knowledge.
- **Model Parameters:** The model was configured with optimal hyperparameters, including the number of estimators (trees), maximum depth, and minimum samples per leaf.
- **Training and Validation:** The data was split into 70% for training and 30% for validation. The model was trained to minimize classification errors.

4.2.2 Model Evaluation

The classifier's performance was evaluated on both validation and test datasets using metrics such as:

- **Accuracy:** The proportion of correctly classified instances.
- **Precision and Recall:** To assess the model's performance for each condition label.

- **Confusion Matrix:** Provided insights into misclassifications across the three labels.

Results: The Random Forest Classifier achieved high classification accuracy, effectively distinguishing between the three condition labels. However, while the classification framework provided actionable insights, it lacked the granularity needed for precise maintenance planning, prompting the transition to the regression task.

4.3. Phase 2: Regression Task

To address the original problem of predicting the RUL as a continuous value, a *Convolutional Neural Network (CNN)* was utilized. CNNs are particularly well-suited for this task due to their ability to extract meaningful features from sensor data.

4.3.1 CNN Architecture

The CNN was designed with the following layers:

- **Input Layer**: Accepts 2D time-series input of shape $(32 \times 13 \times 1)$.
- **Convolutional Layers**:
  - Layer 1: 64 filters, kernel size 3, ReLU activation.
  - Layer 2: 32 filters, kernel size 3, ReLU activation.
- **Flatten Layer**: Converts the multidimensional output into a 1D vector.
- **Fully Connected Layer**: Dense layer with a single neuron for linear regression (predicting RUL).

4.3.2 Model Training

- **Loss Function**: Mean Squared Error (MSE) to minimize prediction error.
- **Optimizer**: Adam optimizer for efficient gradient updates.
- **Training Process**: The model was trained on data from 100 engines, with 15 epochs per engine.

4.3.3 Enhanced Evaluation

The model was evaluated on five randomly selected engines, with the following additional analyses:

- **Statistical Summary**: Computed average RMSE and standard deviation across engines, along with RMSE distribution visualization.
- **Outlier Detection**: Identified engines with unusually high RMSE values (above mean + 2 standard deviations).
- **Error Analysis**: Examined cycles with high prediction errors for outliers.
- **Improvement Suggestions**: Highlighted potential enhancements in feature selection, data scaling, and model retraining.

Results: The CNN demonstrated strong predictive performance, effectively capturing the temporal patterns in sensor data. The model provided accurate and granular predictions of RUL, enabling more precise maintenance planning compared to the classification approach.

4.4. Comparative Insights: Classification vs. Regression

The phased approach of starting with classification and transitioning to regression provided valuable insights:

- **Classification (Random Forest):** Simplified the problem, offering actionable labels for quick maintenance decision-making but lacked the precision needed for long-term planning.
- **Regression (CNN):** Addressed the original RUL prediction problem with greater granularity and accuracy, enabling fine-tuned maintenance schedules.

By combining these two approaches, the study leveraged the strengths of both techniques, culminating in a robust methodology for predictive maintenance of turbo engines.

5. Results

This section presents the Random Forest classifier's performance on the test dataset,as well as the performance of CNN model on the dataset

5.1. Classification Task

5.1.1 Confusion Matrix

Figure 3 summarizes the model's performance, normalized by true labels:

- **Good Condition**: 94% accuracy, 5.5% misclassified as "Moderate."
- **Moderate Condition**: 76% accuracy, with 17% misclassified as "Good" and 7.8% as "Warning."
- **Warning Condition**: 88% accuracy, 12% misclassified as "Moderate."

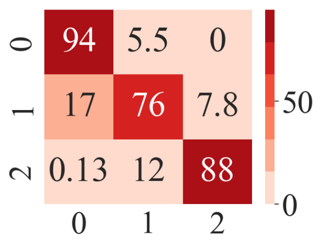


Figure 2. Confusion matrix of the Random Forest classifier on the test dataset (normalized by true labels).

5.1.2 Overall Accuracy

The model achieved 89.58% overall accuracy, performing well in identifying "Good" and "Warning" conditions but showing some overlap in "Moderate." Future improvements could involve feature engineering or hyperparameter tuning.

5.2. CNN Result

The CNN model shows a strong predictive performance for Engine 1, with an RMSE of 3.88, indicating a small average deviation of 3.88 cycles between predicted and actual RUL values. The plot (eg. Figure 3) reveals that the predicted and actual RUL values closely follow the same trend, confirming the model's accuracy. Overall, the model demonstrates good performance for this engine, with minimal error, suggesting it is reliable for RUL forecasting. Further testing across other engines also confirms its general effectiveness. The average RMSE across all engines was calculated as 6.1863, indicating a strong predictive capability of the model in estimating the RUL.

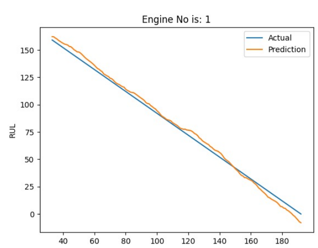


Figure 3. Remaining Useful Life (RUL) versus Cycle for Engine No. 1 found using CNN

6. Conclusion

6.1. Learnings and Key Findings

Through this project, we explored the prediction of the Remaining Useful Life (RUL) of NASA turbo engines using machine learning techniques, focusing primarily on Convolutional Neural Networks (CNNs). Key learnings include:

- Comprehensive application of data preprocessing techniques such as feature engineering, correlation analysis, and outlier detection
- Implementation and evaluation of a CNN model designed to capture temporal patterns in multivariate sensor data.

6.2. Challenges Faced

- Handling sensor noise and operational variability in the dataset, which required extensive exploratory data analysis and feature selection to ensure relevant and high-quality inputs.

6.3. Progress and Future Directions

- Exploring alternative deep learning architectures, such as Long Short-Term Memory (LSTM) networks, to further improve prediction accuracy.
- Incorporating additional real-world constraints and cost analyses to refine the practical applicability of the solution.

.

6.4. Contributions

Each member of the team contributed equally to the project, taking on the roles and responsibilities as outlined in the project proposal. All members worked collaboratively, ensuring that every aspect of the project, from data preprocessing to model training and evaluation, was completed in a timely and efficient manner. This shared commitment allowed us to achieve our goals and laid a solid foundation for the remaining work.

[1–4].

References

[1] Don Simon Neil Eklund Abhinav Saxena, Kai Goebel. Damage propagation modeling for aircraft engine prognostics, 2008. Damage propagation can be modeled within the modules of aircraft gas turbine engines. 5

[2] Felix O. Heimes. Recurrent neural networks for remaining useful life estimation, 2008. Solution to the 2008 IEEE Prognostics and Health Management challenge using a Recurrent Neural Network (RNN). 5

[3] Rafael Ramasso, Emmanuel Gouriveau. Remaining useful life estimation by classification of predictions: A case study on nasa engine degradation data, 2014. Predicting the RUL of engines based on a classification of predicted life intervals using NASA's engine degradation dataset. 5

[4] Ristovski Konstantin Farahat Amr Gupta Charu C Zheng, Shuai. Long short-term memory network for remaining useful life estimation, 2017. Explores the use of Long Short-Term Memory (LSTM) networks for RUL estimation of aircraft engines using NASA's C-MAPSS dataset. 5