Towards Reliable RUL Prediction: Impact of Feature Selection on Degradation Modelling

Oluwasegun Oluwole Gbore[1], Mehak Shafiq[2], Amit Kumar Jain[3*], and Don McGlinchey[4]

[1,2,4]*Department of Mechanical Engineering, School of Computing, Engineering and Built Environment, Glasgow Caledonian University, Glasgow, G4 0BA, UK*

[3]*Department of Mechanical Engineering, Birla Institute of Technology and Science Pilani, Pilani, Rajasthan, 333031, India.*

gboreoluwasegun@gmail.com
mehak.shafiq@gcu.ac.uk
**amit.jain@pilani.bits-pilani.ac.in*
d.mcglinchey@gcu.ac.uk

ABSTRACT

This study investigates feature selection techniques for predicting the Remaining Useful Life (RUL) of aircraft engines, addressing the persistent challenge of inaccurate predictions due to suboptimal feature selection. In this context, a robust methodology was developed to select optimal features for enhancing the model's predictive power. Using inferential statistical methods for analysing operation data from aircraft engines, the study involved data pre-processing to test its feasibility, feature engineering to minimise data variability, backward elimination for linear regression, random forest and gradient boosting for effective feature selection. The models' performance was evaluated for predictive accuracy and reliability using various performance metrics. Findings show that the random forest model with an R-squared value of 0.86 surpassed linear regression (0.76) and gradient boosting (0.73). It further highlighted that the integration of advanced feature selection techniques in non-linear modelling substantially improved the prediction accuracy of RUL while also capturing the essential degradation patterns typical in aircraft engines, as depicted in the Partial Dependence Plots (PDPs). All the three models highlighted the critical importance of the 'time' (current age) feature in predicting RUL, accounting for more than half of the model's predictive power. The findings of this work not only supported some initial hypotheses regarding sensor relationships and operational settings' effects but also unveiled complex interactions previously unrecognized. By identifying and eliminating redundant sensors though a systematic approach of feature selection, this study significantly contributes to the field of predictive maintenance for aircraft engines in enhancing the robustness of predictive models.

1. INTRODUCTION

Maintenance costs have historically been a significant portion of the overall expenses in various industries, ranging from 15% in food-related sectors to as high as 60% in heavy industries like iron and steel, pulp and paper. According to surveys, nearly one-third of the maintenance costs in various industries are wasted because of the improper and unnecessary maintenance practices (Mobley, 2002). To combat this, predictive maintenance has emerged as a transformative approach. Using data analytics, predictive maintenance not only helps in reducing unforeseen downtime and equipment failures but also improves maintenance schedules, thus enhancing the machine's performance (Theissler et al., 2021; Karuppiah et al., 2021). Predictive maintenance practices have been significantly improved with the advancement in sensor technology and data analysis, as these can efficiently gather massive raw data from sensors and convert it into meaningful information to avoid any disruptions and raise the system's security and efficiency (Lee et al., 2019; Zhang et al., 2019; Çınar et al., 2020; Javaid et al., 2021; Chinta et al., 2023).

The estimation of the machinery's RUL remains a cornerstone within the predictive maintenance strategy. In the aviation industry, the reliable and accurate prediction of the aircraft engines' RUL is crucial to ensure safety and reduce operational costs. However, this remains a daunting task because of the highly variable operating conditions of the aircraft engines. Also, the stochastic nature of degradation

https://doi.org/10.36001/IJPHM.2025.v16i2.4296

processes and different failure mechanisms within the engine further complicate the accurate prediction of RUL (Li et al., 2018; Adryan & Sastra, 2021; Zhao et al., 2023).

Researchers have approached RUL prediction from various perspectives and different methodologies have been implemented. For instance, similarity methods have been used that compare previous degradation patterns with current data, while sensor-based degradation models predict RUL based on the identification of the instances when a machine is likely to cross the predefined operating limits (Natsumeda, 2023; Shaheen et al., 2023). Some researchers have adopted physics-based models that predict RUL by assuming the finite crack at the initial stage and linking the extent of damage to the growth of crack using stain life methodology and fracture mechanics (Rial et al., 2014). Another common approach is the energy-based fatigue life prediction that calculates life by dividing the total monotone strain energy by strain energy per cycle. However, both physics-based and energy-based models have limitations in their practical applications, as the former cannot fully capture the machine's behaviour under variable conditions due to sophisticated mechanical structures, whereas the latter involves assumptions and requires more experimental data for reliable estimations. This demands a shift and search for more sophisticated and state-of-art machine-learning techniques for better prognostics (Lei et al., 2018; Hu et al., 2019; Liu et al., 2020).

Data-driven approaches—utilising probabilistic models, artificial intelligence, and stochastic methods—offer an alternative by utilizing operational data for RUL prediction. These methods include statistical analysis of life cycle information and advanced forecasting techniques to identify anomalies and trends with better accuracy by analysing historical and real-time data (Muneer et al., 2021; Victoria & Priyardarshini, 2023). But the efficacy of different methods used varies according to the conditions and datasets. For instance, deep neural networks offer better accuracy but at the cost of interpretability, while linear models like regression offer ease of interpretation and might suffice for systems with linear degradation patterns but are unable to capture non-linear degradation profiles, which are typical in mechanical systems like aircraft engines (Zhang et al., 2017; Melis & Jaakkola, 2018; Salahuddin et al., 2022; Mishra et al., 2023; Talaat et al., 2023). In such cases, more advanced models, like random forest and gradient boosting, perform well. Some studies encourage the use of adaptive modelling techniques, which assist in striking a balance between interpretability and accuracy over the system's lifecycle because these techniques dynamically select models according to the system environment. (Angelov et al., 2010). On top of that, using a hybrid approach by leveraging the strengths of both simpler and complex models a robust framework can be developed for effective RUL prediction where simple models can be used for initial screening and complex ones for detailed analysis (Liao & Köttig, 2014; Liao & Köttig, 2016; Li et al., 2022; Gardner-Frolick et al., 2022).

Regardless of the modelling approach, feature engineering plays a critical role in improving the model's performance. Well-engineered features can significantly enhance the prediction accuracy of both simple and complex models. Hence, the effective feature selection techniques not only determine the most influential parameters of engine degradation, but they also minimise the excessive maintenance costs and operational risks (Hong et al., 2020; Xiong & Wang, 2022; Kartal & Altunkaynak, 2024; Hayajneh et al., 2024). Despite its critical importance, many organisations struggle with implementing proper feature selection techniques, which result in inefficient predictive maintenance practices (Rong et al., 2019). Moreover, the adaptation of feature selection techniques to dynamic operational data, which typically varies non-linearly with the change in operating conditions, has been less emphasized in the current literature. Majority of the research provides generalised insights without digging into the unique datasets and operational characteristics of complex industrial machinery such as aircraft engines. This oversight results in suboptimal modelling results, hence jeopardizing the reliability and accuracy of RUL predictions.

To address the aforementioned research gaps, this work aims to enhance the accuracy and reliability of RUL predictions for aircraft engines by identifying and implementing optimal feature selection techniques in predictive modelling. In this regard, for the first time, the performance of backward elimination, random forest, and gradient boosting models is evaluated in identifying the essential features affecting the degradation processes in aircraft engines. This not only will help in eliminating redundant and irrelevant features but will also enhance the robustness of the predictive models. Consequently, this will pave the way for better industry practices and ensure greater reliability in the aerospace sector.

This study utilizes the FD001 subset of the C-MAPSS dataset, a widely recognized benchmark in predictive maintenance research. The dataset comprises simulated sensor and operational data from 100 aircraft engines, each progressing through a full degradation cycle under consistent flight conditions. For each engine, 21 sensor readings and 3 operational settings were recorded across multiple time steps (Saxena & Goebel, 2008). The dataset enables RUL estimation by providing time-series data up to the point of failure for each engine. The process followed in this study is shown in Figure 1, which involves several sequential steps. The data is rigorously pre-processed, which includes converting text files into Excel format for better visualization and ensuring it is clean and ready for analysis. Then, the correlation coefficient analysis and Variance Inflation Factor (VIF) are applied to determine its feasibility. Further, the predictive model is developed using the multiple linear

regression method to establish a relationship between the RUL and operational parameters of aircraft engines. Logarithmic transformation is also applied to the data in order to improve the model performance. Furthermore, the performance of three feature selection techniques, named backward elimination, random forest, and gradient boosting, in enhancing the accuracy of the predictive model are compared using the same performance metrics as shown in Figure 1. For this work, the code is written in Python and run in Google Colab.

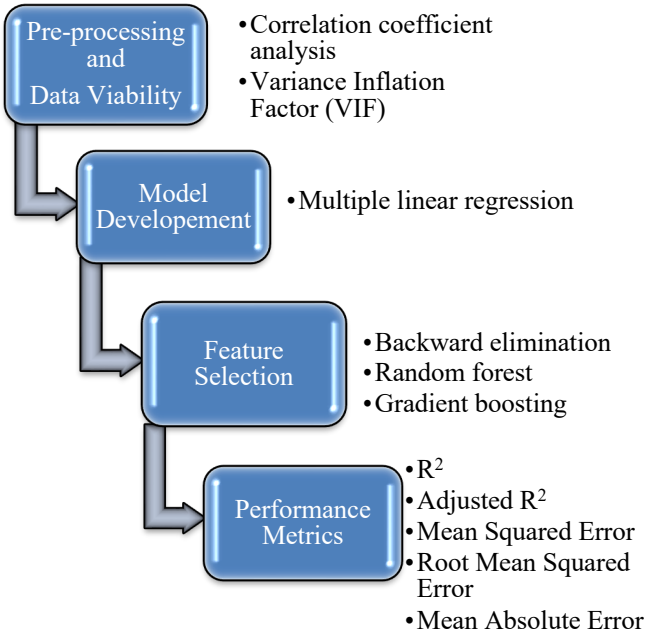

Figure 1. Sequential steps undertaken in current study

2. METHODOLOGY

2.1. Data Feasibility Assessment

The viability and integrity of the data are assessed using two key statistical methods. The first method, Pearson's correlation coefficient, is used to examine the linear relationships between variables, while the second method, VIF, is applied to further scrutinize the features (independent variables) for multicollinearity.

2.1.1. Pearson Correlation Coefficient

Pearson's correlation coefficient quantifies the degree to which pairs of variables are linearly connected, providing insights into potential model predictors and highlighting any signs of multicollinearity. For each pair of features, the Pearson correlation coefficient (r) is calculated using Eq. (1), which provides a value between -1 and +1 that indicates both the strength and direction of the relationship. Coefficients close to +1 and -1 indicate strong positive and negative linear relationships, respectively, whereas coefficients around 0 imply little to no linear relationships. Although, features exhibiting strong correlations with RUL are remarked as important predictors, the relationships between features are scrutinized for indicators of multicollinearity. When multicollinearity is identified, one of the features is removed, preferably the one with the lowest correlation with the RUL.

The mathematical expression of Pearson correlation coefficient (r) described for two features X and Y with n number of observations each is given as follows (Yu & Hutson, 2022):

$$r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})\,(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})}} \quad (1)$$

Where X_i and Y_i denote the individual sample points and $\bar{X}$ and $\bar{Y}$ are the mean values of X and Y, respectively, and are given as:

$$\bar{X} \; and \; \bar{Y} = \frac{\sum_{i=1}^{n} X_i}{n} \; and \; \frac{\sum_{i=1}^{n} Y_i}{n} \quad (2)$$

n represents the number of observations.

2.1.2. Variance Inflation Factor (VIF)

VIF serves as a metric quantifying the degree to which the variance of an estimated regression coefficient inflates as a result of multicollinearity in the model. As predictor correlations increase, VIF values rise, indicating increased multicollinearity and potential issues in model estimation. A VIF near 1 indicates minimal inflation, while values surpassing a threshold (typically 5 or 10) imply problematic multicollinearity, warranting further investigation and possible variable exclusion from the model.
Features with lower variance (i.e., lower VIFs) may have limited predictive power as they exhibit minimal change across the dataset. Addressing high VIFs leads to more stable and interpretable models, which improve predictive accuracy and facilitate informed decision-making in model development and refinement. The following formula is used to calculate the variance inflation factor for each predictor variable in a regression model (Cheng et al., 2022):

$$VIF = \frac{1}{1-R_i^2} \quad (3)$$

where $$R_i^2 = 1 - \frac{\sum_{j=1}^{n}(Y_{ij}-\hat{Y}_{ij})^2}{\sum_{j=1}^{n}(Y_{ij}-\bar{Y}_{ij})^2} = 1 - \frac{SSE}{SST} \quad (4)$$

In the equations above, R_i^2 is the coefficient of determination of predictor i on all the other predictors, whereas Y_{ij}, $\hat{Y}_{ij}$, and $\bar{Y}_i$ are the actual predictor, predicted predictor of regression model, and mean value of predictor i for the number of observations j (Garcia et al., 2014).

2.2. Logarithmic Transformation

Logarithmic transformation, which is typically performed using the natural logarithm, is used in this study, as it is an effective data transformation method for resolving scale differences within a dataset. Applying this transformation can help create a more uniform scale across the data by transferring skewed distributions into more symmetrical ones. This is how it eliminates the influence of outliers by compressing the data scale and helps in preventing the model's training process from being distorted (Shachar et al., 2018; Kuhn and Johnson, 2018).

As machinery and components approach the end of their useful life, degradation patterns often become increasingly nonlinear and less predictable. This shift in behavior, particularly in the later stages, poses challenges for accurate RUL prediction (Cui et al., 2023), and requires models that can accommodate dynamic and irregular degradation trajectories. This process of linearization can be extremely useful for predictive models by facilitating the application of linear models and improving the interpretability of the result when there are multiplicative relationships between the features and RUL (Babu et al., 2016).

The mathematical form of the transformation process for each value x_i within each feature X is given as below (Shim, Bonifay & Wiedermann, 2022):

$$\acute{x}_i = \log_b(x_i) \tag{5}$$

$$log\ Y_i = \beta_0 + \beta_1 log\ X_1 + \beta_2\ log\ log\ X_2 + \cdots + \beta_n log\ X_n + \epsilon \tag{6}$$

where, x_i and $\acute{x}_i$ are the original and transformed feature values, respectively.

2.3. Model Development

Multiple Linear Regression (MLR) model is based on the principle that there exists a linear relationship between the dependent variable (Y) and multiple independent variables ($X_1, X_2, \ldots\ldots, X_n$). The general form of the MLR model is given as (Maulud & Abdulazeez, 2020):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \tag{7}$$

In the above equation, $\beta_1, \beta_2, \ldots, \beta_n$ represent the coefficients of the independent variables and β_0 is the y-intercept.

First of all, the dependent variable (i.e., RUL) is defined, and independent variables (i.e., features) are selected based on the knowledge base of the system and expected results. Then, coefficients ($\beta_1, \beta_2, \ldots, \beta_n$) are estimated using the Ordinary Least Squares (OLS) method, which is widely accepted because it is simple to use, interpretable, and efficient under common conditions. The OLS estimate is made to find the line (or hyperplane in multiple dimensions) that best fits a set of data points, thereby reducing the sum of the squared differences between the predicted and observed values of the dependent variable by the linear model. This estimation leads to a set of normal equations as given below, which are solved to find the coefficient estimates.

$$SSR = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 \tag{8}$$

where SSR is the sum of the squared residuals and Y_i and $\hat{Y}$ represent the observed and predicted values of the dependent variable for the i^{th} observation, respectively.

Several key assumptions underlie the MLR model, and their validity must be assessed to ensure reliable results; the relationship between features and RUL is linear, the residuals (errors) are independent of each other, normally distributed and not too highly correlated, the variance of residuals is constant across all levels of the independent variables.

2.4. Feature Selection

With the aim of developing strong predictive model, the selection of relevant features is critical in improving the model performance and interpretability. Among the available techniques of feature selection, backward elimination, random forest (Tree-Based) method, and gradient boosting method stand out for their distinct approaches and advantages in refining the feature space.

2.4.1. Backward Elimination

Backward elimination process aims to refine the predictive model by retaining only those variables that significantly contribute to explaining the variance in the dependent variable, thereby enhancing the model's simplicity and interpretability. The process begins with the full regression model (i.e. Equation 7).

Each predictor's coefficient $\hat{\beta}_i$ undergoes a t-test to assess its statistical significance, where the null hypothesis $\left(H_0: \beta_i = 0\right)$ implies no effect on Y. The t-test is computed as (Fashoto et al., 2021):

$$t = \frac{\hat{\beta}_i - 0}{SE(\hat{\beta}_i)} \tag{9}$$

where $\hat{\beta}_i$ is the estimated coefficient for the predictor variable X_i, 0 is the hypothesized value of the coefficient under the null hypothesis usually testing if $\left(\hat{\beta}_i = 0\right)$,

indicating no effect, $SE(\hat{\beta}_i)$ is the standard error of the estimated coefficient $\hat{\beta}_i$. The standard error $SE(\hat{\beta}_i)$ measures the average difference between the estimated $\hat{\beta}_i$ and the actual (unknown) coefficient β_i , given the data. It is calculated from the data and depends on the variability of the Y values and the distribution of the X_i values.

The p-value is determined for each coefficient, reflecting the probability of observing the given data under H_0 . The predictors with highest p-value exceeding a predefined threshold ($\alpha = 0.05$) are considered the least significance and sequentially removed, thereby reducing the number of predictors and reformulating the model. The reduced model is then fitted using Ordinary Least Squares (OLS), updating the estimates of the coefficients for the remaining predictors. This process is repeated iteratively until all the remaining predictors have p-values below the threshold. The final optimised model retains only statistically significant features, resulting in greater interpretability and prediction power.

2.4.2. Random Forest

A random forest regressor is an ensemble learning technique in which multiple decision trees are generated during the training process, and predictions are combined by averaging of regression tasks. The number of trees in the forest and the seed used by the random number generator for reproducibility are key initialising parameters.

During training, random forest uses bootstrap samples (given in Equation 10) to build individual trees.

$$D_i = \{(x_1^i, y_1^i), (x_2^i, y_2^i), \dots, (x_n^i, y_n^i)\} \quad (10)$$

Each decision tree in a random forest splits the data based on a feature and split point that minimizes the variance within each child node compared to the parent node. The variance reduction (VR) for a feature at a split is calculated as:

$$VR = Var(Parent) - \left(\frac{N_{left}}{N_{parent}} \times Var(left) + \frac{N_{right}}{N_{parent}} \times Var(right)\right) \quad (11)$$

Where $Var(.)$ is the variance of the target variable in the node, N_{parent} , N_{left} and N_{right} are the number of samples in the parent, and right and left child nodes, respectively.

After training, the random forest model provides the importance of each feature. This importance is a measure of how much each feature contributes to reducing the variance (uncertainty) in the predictions across all trees in the forest. So that features with higher importance scores are considered more valuable for predicting the target variable (RUL).

2.4.3. Gradient Boosting

Gradient boosting is a powerful ensemble technique that develops models in a sequential manner, with each new model incrementally corrects previous models' errors, using decision trees as the base learners. It is widely used for both classification and regression tasks. When it comes to feature selection, gradient boosting evaluates feature importance as a by-product of the model training process, making it an effective method for identifying significant features in a dataset.

This method involves three key components: a loss function (L), to measure the model fit, shallow decision trees (i.e., few levels deep) as weak learner, and an additive model, where each tree address the residual errors of its predecessors. An initial guess for the model is made, usually a constant value such as the mean or median of the target variable. The pseudo-residuals are computed as negative gradient of the loss function with respect to the model predictions. For observation (i) at iteration (m), this is given as:

$$r_{im} = -\left[\frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F(x_i)}\right]_{F(x_i)=F_{m-1}(x_i)} \quad (12)$$

where $L(y_i, F(x_i))$ is the loss function evaluated for the ith data point. L measures discrepancy between observed target y_i and the model's prediction $F(x_i)$. $F_{m-1}(x_i)$ represents the accumulated predictions up to the previous step $(m-1)$. r_{im} represents the negative gradient of the loss function by the model's predictions

The decision tree $h_m(x)$ is fitted to these pseudo-residuals, the multiplier Y_m is optimized for the tree's predictions to minimize the loss when added to the current model. This is done by:

$$Y_m = argmin_Y \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + Yh_m(x_i)) \quad (13)$$

The model is then updated with the new tree:

$$F_m(x) = F_{m-1}(x) + VY_m h_m(x) \quad (14)$$

The final model after M iterations become:

$$F_M(x) = F_0(x) + \sum_{m=1}^{M} VY_m h_m(x) \quad (15)$$

where $h_m(x)$ represents the corrections to be applied to the previous ensemble's predictions, V is the learning rate, a parameter that scales down the contribution of each tree to prevent overfitting, N is the total number of data points in the dataset, M is the total number of iterations or trees in the ensemble.

Feature importance is typically assessed based on the improvement in model performance related to each feature. Two common measures are; total gain, which is the sum of the reduction in loss provided by splits over a given feature across all trees, and the frequency of use, which is how often a feature is used to split across all trees.

Mathematically, the importance of a feature (*j*) can be expressed as the sum of the gains (*ΔGain*) of all splits that use *j*, normalized by the total number of splits across all trees:

$$I(j) = \frac{\sum_{all\ splits\ using\ j} \Delta Gains(j)}{\sum_{all\ splits} \Delta Gains} \quad (16)$$

After training process, gradient boosting ranks the features based on their importance scores. Low-importance score features can be eliminated from the model as they are less useful in making predictions. As such, besides being a powerful predictive modelling technique, gradient boosting also performs feature selection given its inherent computation of feature importance.

2.5. Performance Metrics

In the realm of feature selection, the evaluation of model performance involves identifying features that minimise errors and optimise model accuracy in predicting RUL of aircraft engines. In this study, the following six performance metrics are used for the evaluation of linear regression model using different techniques of feature selection.

Coefficient of determination (R^2): It represents the proportion of the variance explained by the model; values near 1 indicate better fit, which means more variance is explained by the model (García et al., 2014).

$$R^2 = 1 - \frac{SSR}{SST} \quad (17)$$

SSR and SST are the sum of squares of residuals and the total sum of squares, respectively.

Adjusted R^2: It refines R^2 and offers a more precise metric for feature selection by accounting for the number of predictors in the model (Fashoto et al., 2021).

$$Adjusted\ R^2 = 1 - \frac{(1-R^2)(n-1)}{n-p-1} \quad (18)$$

Here n and p are the sample size and the number of independent variables, respectively.

Mean Squared Error (MSE): It calculates the average squared difference between the estimated and actual values, with values near zero are better.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad (19)$$

Where y_i and $\hat{y}_i$ represent the actual value (from the dataset) and predicted value (by the model) respectively, and *N* is the total number of observations in the dataset.

Root Mean Squared Error (RMSE): This is a commonly used metric for measuring the model's error in predicting quantitative data and represents the square root of MSE

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \quad (20)$$

Mean Absolute Error (MAE): It calculates the average of the absolute difference between actual and estimated values for all observations. MEA, unlike MSE and RMSE, provides a linear score, which means all individual differences are equally weighted, and lower values of MAE indicate better model performance.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \quad (21)$$

Partial Dependence Plots (PDPs): It demonstrates the effect of a single or two features, X_k, on the model prediction after considering the average effects of all other features. The partial dependence function, $f_{Xk}(x)$, is as follows (Inglis et al., 2022):

$$f_{Xk}(x) = \frac{1}{N} \sum_{i=1}^{N} \hat{f}(x, X_i, -k) \quad (22)$$

where $\hat{f}$ is the prediction function of the trained model, x is a value of feature X_k, and $(X_i, -k)$ represents all features for the i^{th} observation except X_k which is set to x.

PDPs are especially useful for models like gradient boosting and random forest because they display complex, non-linear feature interactions, helping to elucidate the "black box" nature of such models.

3. RESULTS AND ANALYSIS

A viability check using the correlation analysis and VIF was performed to identify redundancy and multicollinearity among features. This process is crucial for ensuring that each variable in a predictive model contributes unique, valuable information and that coefficient estimates are not distorted by high inter-variable correlation, thereby preserving the model's reliability and interpretability.

Figure 2 illustrates the correlation heat map which shows the pairwise correlation between different sensor measurements and operational settings of the aircraft engine. High correlation coefficients, represented by darker shades of red,

suggest strong positive linear relationships, whereas darker shades of blue indicate strong negative relationships. Near-zero values, shown in lighter shades, imply a weak or no linear correlation.

The heat map provides crucial insights about the relationships between aircraft engines, highlighting high correlations among certain operational settings and sensor measurements. Operational settings 1 and 2 show a high correlation coefficient (0.944), suggesting they are closely linked, possibly controlled together or respond to similar engine states. Conversely, these two settings have an inverse relationship with operational setting 3, as evidenced by the negative correlation coefficient (-0.237 with setting 1 and -0.385 with setting 2). This may reflect a compensatory mechanism within the engine's operation, where adjustments to settings 1 and 2 are counterbalanced by changes in the setting 3.

On the other hand, sensor measurements 5 and 6 exhibit a very strong positive correlation (0.996 and 0.996, respectively), which may reflect sensors measuring related engine parameters, perhaps due to physical proximity or

shared functional pathways. One of the most significant negative correlation between 'time' and 'RUL' (-0.788) validates the expected trend of decreasing RUL with increase in time due to wear and degradation over time. The high degree of correlation among sensors suggests some level of redundancy. Including all these sensors in a predictive model could unnecessarily complicate the model without adding value, potentially making it prone to overfitting. Meanwhile, sensors with low correlations may be valuable for detecting specific types of engine faults or performance issues that are not apparent from more correlated measurements. Thus, the correlation matrix not only indicates which variables to include or exclude from the model, but also provide insights into the complex interplay of engine operations and sensor responses.

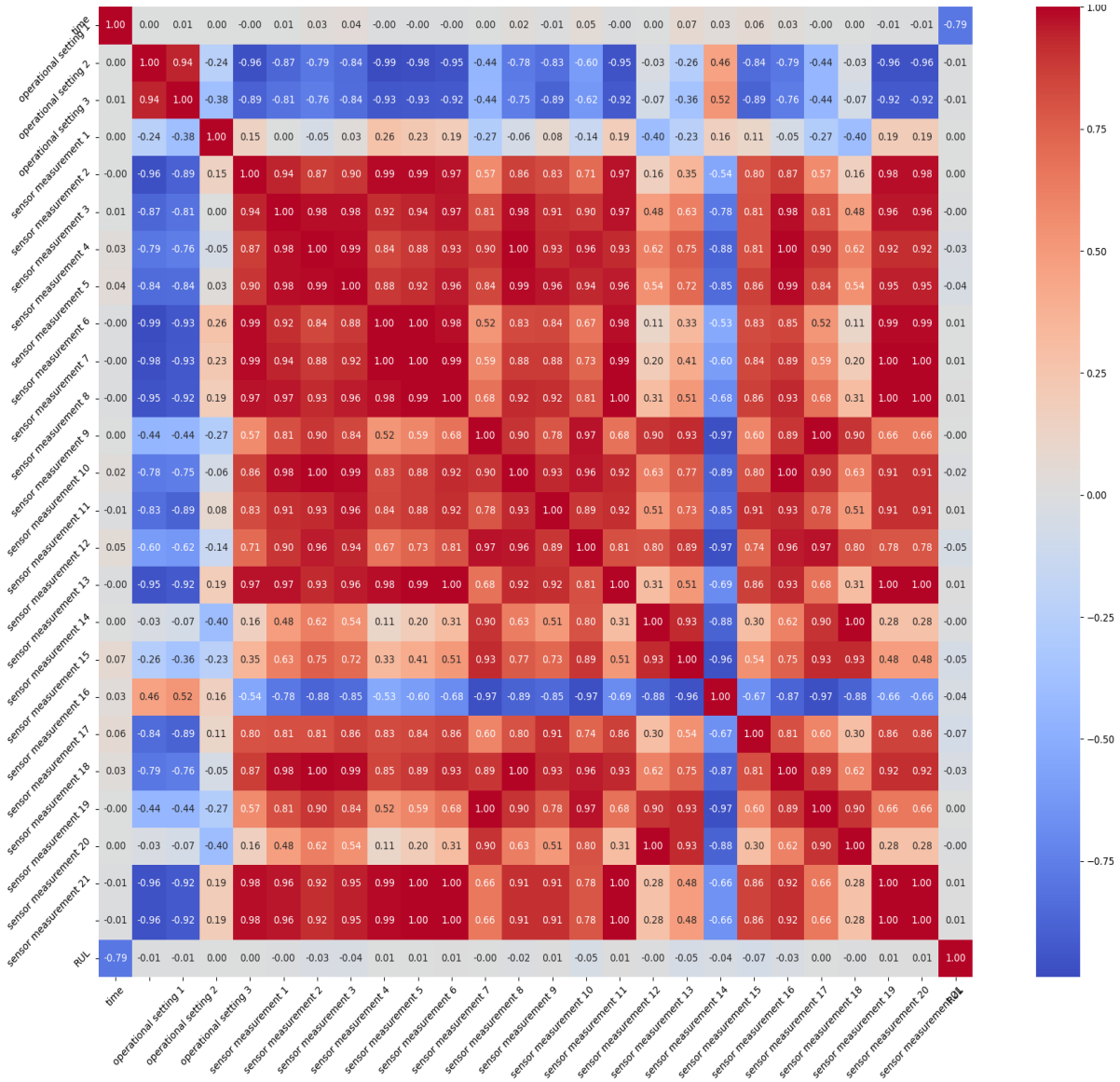

Figure 2. Correlation Heat Map of the Aircraft Engine's Sensor features

Table 1 presents the VIF, providing insights into multicollinearity among features. As can be seen that several features have VIFs much greater than 1, indicating that these variables have some level of linear correlation with other predictor variables in the dataset. Notably, the 'const' feature has an extremely high VIF, signaling it is highly linearly dependent on other features. This is because of the model intercept being included in the VIF calculation, which is not typically necessary since it is not a predictor. Some sensor measurements, such as 1, 5, 18, and 19, exhibit very high VIF values (in the range of millions and even billions), suggesting severe multicollinearity. This could distort the coefficient estimates, undermine the statistical significance of the independent variables and affect the outcome and predictive power of the regression model. The presence of high VIF values is a strong indication that feature engineering is needed.

Techniques like feature selection (choosing a subset of relevant predictors) and feature extraction (creating new features from the original ones while preserving essential information) could be used. Given the observed high VIFs, transforming or reducing these features becomes crucial to mitigate multicollinearity.

The logarithmic transformation was systematically applied across the dataset to the features to normalize the scale and distribution of the features, thereby preparing the dataset for more effective analysis. Table 2 demonstrates the raw and transformed data for some features as an example.

After performing the data viability checks, the multiple linear regression model was developed as described in section 2.3, resulting in an R-squared value of 0.693 and adjusted R-squared of 0.692.

Serial Number	Features	Description	VIF
1	Time	Operational cycles	2.34
2	Operational Setting 1	Altitude (ft)	3.02 e07
3	Operational Setting 2	Mach number	2.30 e05
4	Operational Setting 3	Throttle resolver angle (°C)	1.13 e07
5	Sensor Measurement 1	Total temperature at fan inlet (°R)	2.16 e10
6	Sensor Measurement 2	Total temperature at LPC outlet (°R)	1.43 e04
7	Sensor Measurement 3	Total temperature at HPC outlet (°R)	6.67 e02
8	Sensor Measurement 4	Total temperature at LPT outlet (°R)	7.53 e02
9	Sensor Measurement 5	Pressure at fan inlet (psia)	2.74 e08
10	Sensor Measurement 6	Total pressure in bypass-duct (psia)	2.31 e06
11	Sensor Measurement 7	Total pressure at HPC outlet(psia)	1.15 e05
12	Sensor Measurement 8	Physical fan speed (rpm)	1.25 e07
13	Sensor Measurement 9	Physical core speed (rpm)	4.37 e03
14	Sensor Measurement 10	Engine pressure ratio (P50/P2)	4.81 e03
15	Sensor Measurement 11	Static pressure at HPC outlet (psia)	7.73 e02
16	Sensor Measurement 12	Ratio of fuel flow to Ps30 (pps/psi)	1.71 e05
17	Sensor Measurement 13	Corrected fan speed (rpm)	8.59 e06
18	Sensor Measurement 14	Corrected core speed (rpm)	3.51 e02
19	Sensor Measurement 15	Bypass Ratio	1.09 e03
20	Sensor Measurement 16	Burner fuel-air ratio	9.97
21	Sensor Measurement 17	Bleed Enthalpy (BTU/lbm)	7.98 e02
22	Sensor Measurement 18	Demanded fan speed (rpm)	1.30 e11
23	Sensor Measurement 19	Demanded corrected fan speed (rpm)	8.94 e10
24	Sensor Measurement 20	HPT coolant bleed (lbm/s)	9.05 e03
25	Sensor Measurement 21	LPT coolant bleed (lbm/s)	9.09 e03

Table 1. Variance inflation score of features

3.1. Model Comparison: Feature Selection Techniques

Following the logarithmic transformation of the dataset to mitigate the effects of large-variance numerical features, the performance of backward elimination, gradient boosting and random forest regression models is compared to identify the most effective approach for RUL prediction.

The application of the backward elimination technique resulted in a more streamlined and efficient model by removing five features, including various sensor measurements and operational settings, due to their high p-values, indicating a lack of statistical significance in predicting RUL. Following this, the relative significance of each feature in making accurate predictions of RUL is evaluated using random forest and gradient boosting regression models, as illustrated in Figures 3 and 4. Notably, the 'time' feature stands out in both models with an importance score of about 0.543 and 0.866 in random forest and gradient boosting, respectively. This shows that time, which represents the machinery's usage duration, is the main predictive factor for RUL. Sensor measurement 11 ranked as the next most important feature (0.095) in random forest, while gradient boosting identified sensor measurement 15 as the second most significant feature (0.145). However, the importance score of these second features is substantially less than time, indicating a steep drop-off in importance between the top feature and subsequent ones, contributing incrementally less to the model's ability to predict RUL. Both models show a similar trend, with feature importance diminishing progressively down the list. The bar graphs present a clear picture that only a handful of features

significantly drive the model's predictions, with gradient boosting requiring fewer features than random forest to achieve comparable performance.

Time (Cycles)	Operational Setting 1 (ft)	Operational Setting 2	Operational Setting 3 (°C)	Sensor Measurement 1 (°R)	Sensor Measurement 2 (°R)	Sensor Measurement 3 (°R)
Raw Data						
1	10.01	0.25	20.00	489.05	604.13	1499.45
2	0.01	0.01	100.00	518.67	642.13	1584.55
3	34.99	0.84	60.00	449.44	555.42	1368.17
4	20.01	0.70	0.00	491.19	607.03	1488.44
5	42.01	0.84	40.00	445.00	549.52	1354.48
Transformed Data						
0.69	2.40	0.22	3.04	6.19	6.41	7.31
1.10	0.01	0.01	4.61	6.25	6.47	7.37
1.38	3.58	0.61	4.11	6.11	6.32	7.22
1.61	3.04	0.53	0.01	6.20	6.41	7.31
1.79	3.76	0.61	3.71	6.10	6.31	7.21

Table 2. Logarithmic transformation of the data

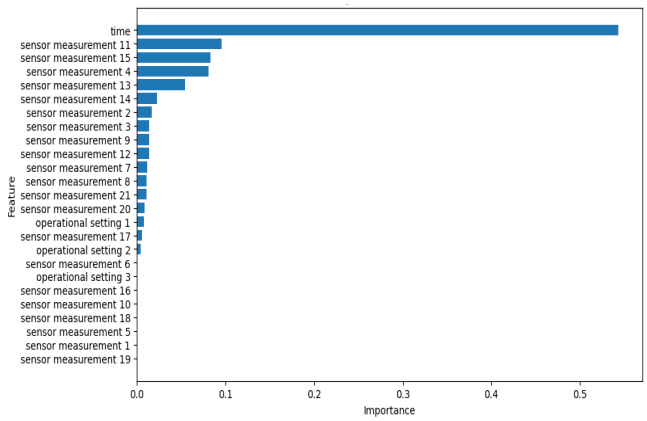

Figure 3. Feature importance of the random forest model

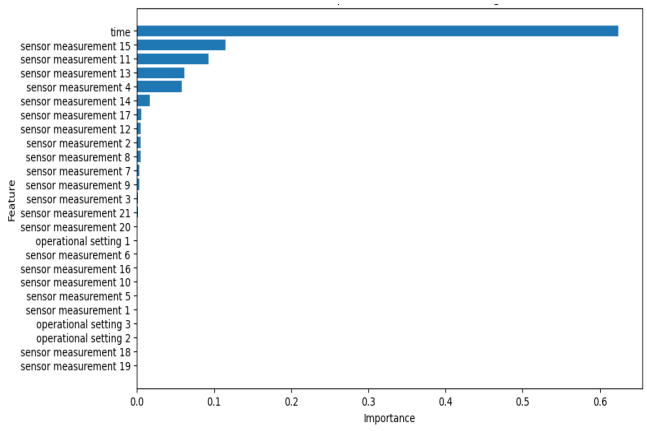

Figure 4. Feature importance of the gradient boosting model

Figures 5 and 6 show the scatter plots comparing the actual with predicted RUL using random forest and gradient boosting methods, respectively. As can be seen, most points in both plots cluster around the ideal dashed line, indicating generally accurate predictions. However, there is noticeable dispersion, especially at lower RUL values in random forest and at higher RUL values in gradient boosting. This variability suggests that while both models perform well overall, the accuracy of random forest decreases for predictions involving lower RUL. On the other hand, gradient boosting model predictions vary more widely for engines with longer expected life.

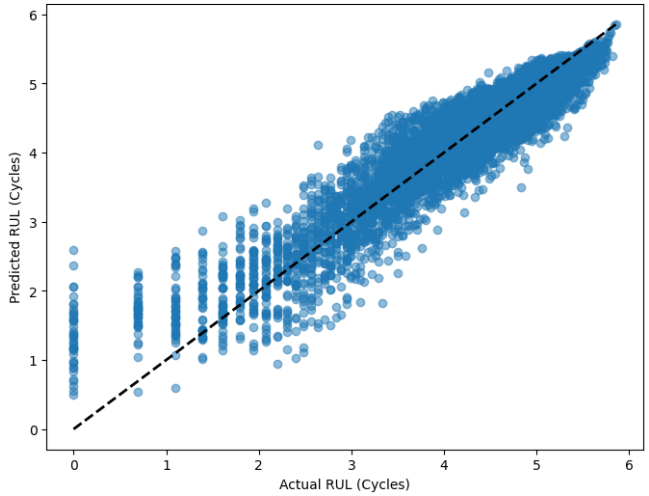

Figure 5. Actual vs predicted RUL plot of the random forest model

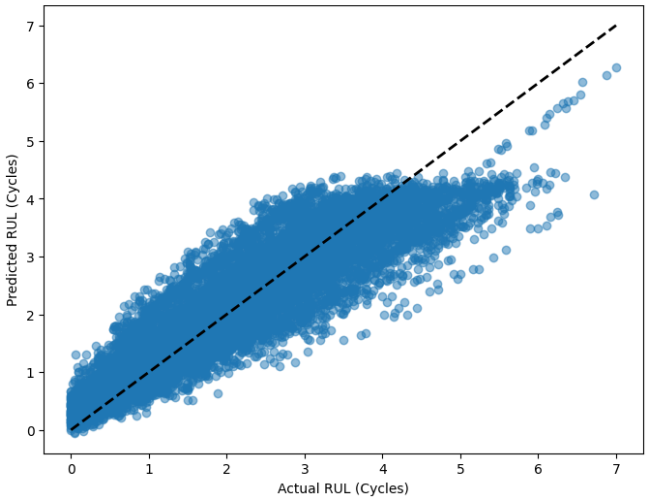

Figure 6. Actual vs predicted RUL plot of the gradient boosting model

The performance of the optimised models is further quantified through several metrics presented in Table 3. Backward elimination achieved an R-squared value of 0.762, indicating that 76.20% of the variance in the RUL was explained by the model, with relatively low values of RMSE and MAE reflecting the model's predictions are fairly close to the actual values. In contrast, random forest outperformed the other models with the highest R-squared of 0.864 and lower values of error metrics (i.e., MSE: 0.128, RMSE: 0.357, and MAE: 0.271), indicating superior predictive capability. Gradient Boosting, while achieving a reasonably good R-squared of 0.733, resulted in relatively high values of MSE, RMSE, and MAE which are 1248.52, 35.33, and 26.95, This indicates that the average errors are somewhat significant; but, without additional context such as the scale and range of RUL values, it is difficult to assess the severity of these errors. It is worth noting that the difference between the R-squared and Adjusted R-squared values in all the three models is very minimal, implying that the number of features included in the models is appropriate and does not introduce unnecessary complexity.

Metric	Score		
	Backward elimination	**Random forest**	**Gradient boosting**
R-squared	0.762	0.864	0.733
Adjusted R^2	0.761	0.864	0.732
Mean Squared Error (MSE)	0.225	0.128	1248.52
Root Mean Squared Error (RMSE)	0.473	0.357	35.334
Mean Absolute Error (MAE)	0.358	0.271	26.956

Table 3. Performance comparison of feature selection techniques

3.2. Partial Dependence Plots (PDPs)

Partial Dependence Plots (PDPs) are used to visualize the isolated impact of individual features on a model's prediction, by averaging out the effects of all other variables. In this study, PDPs were generated for both the Random Forest and Gradient Boosting regression models to explore how selected features influence the prediction of Remaining Useful Life (RUL) in aircraft engines. To illustrate the models' interpretability and feature sensitivity, three representative features were selected: time (representing operational age), Mach number (operational setting 2), and demanded corrected fan speed (sensor measurement 19). These features were chosen to reflect both dominant and marginal contributors to model predictions. Figure 7 presents a side-by-side comparison of how each model responds to changes in these inputs, offering insight into the internal decision-making processes and highlighting the contrasting ways these models interpret engine degradation.

Both models exhibit a clear negative relationship between time and predicted Remaining Useful Life (RUL), which aligns with the intuitive expectation that engine health

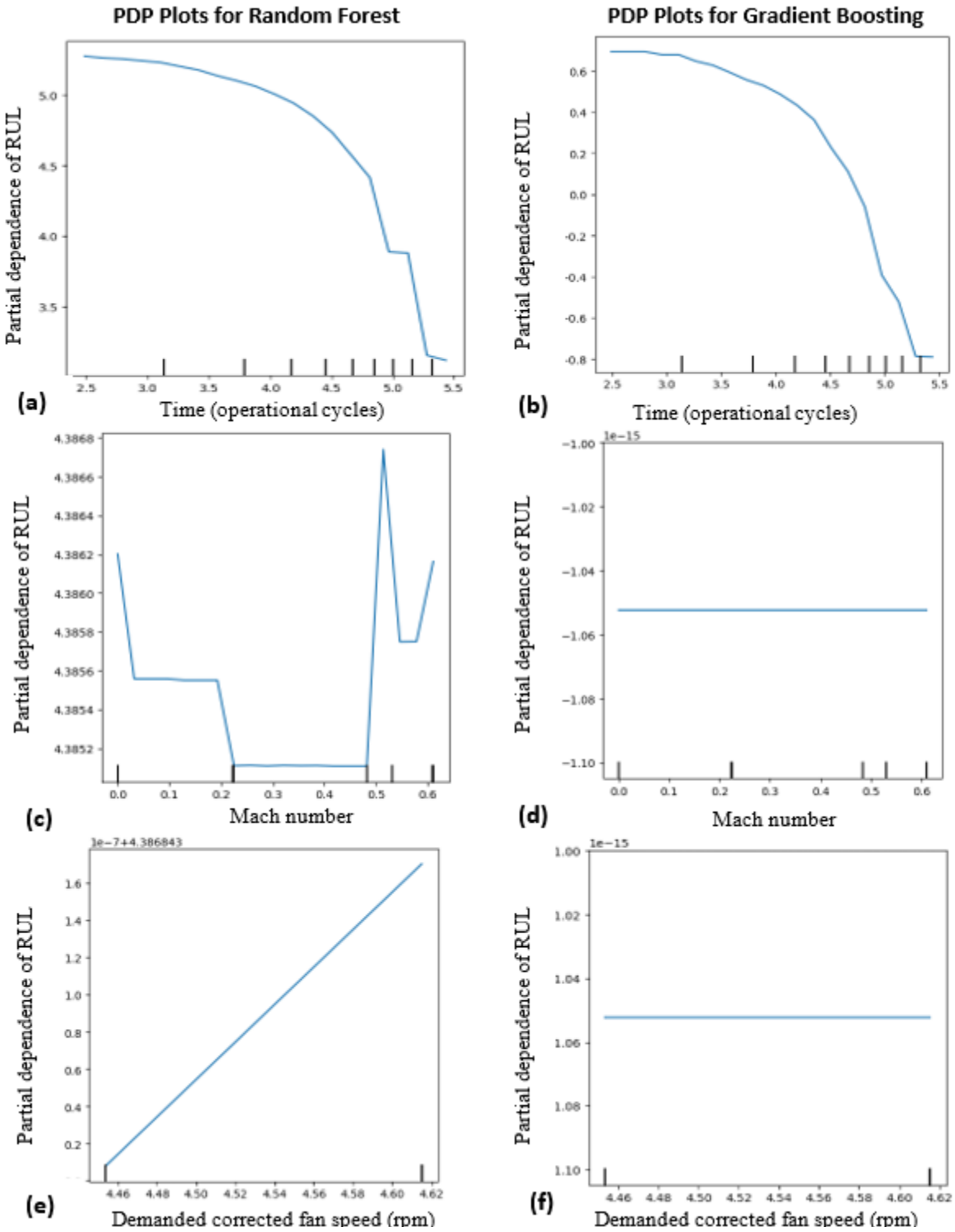

Figure 7. Comparison of the PDP plots of random forest and gradient boosting models illustrating the effect of some operational features on RUL

declines as operational cycles accumulate. In Figure 7a, the decrease in partial dependence of RUL was initially moderate but becomes significantly sharper around the 5-cycle mark, suggesting the model detects a distinct degradation phase, a critical point at which wear accelerates. This behavior reflects Random Forest's ability to capture nonlinear, threshold-based changes in the data. In contrast, Figure 7b shows a smooth and gradual decline throughout the time range, indicating a consistent interpretation of wear progression. This suggests that Gradient Boosting is better suited for modeling continuous degradation trends, while Random Forest is more responsive to sudden drops in performance. Overall, both models confirm time as a dominant predictor of RUL, though they capture its effects in structurally different ways.

When examining Mach number and demanded corrected fan speed, both models exhibit low sensitivity to these features, albeit in model-specific ways. The Random Forest PDP for Mach number in Figure 7c displays minor, inconsistent fluctuations in predicted RUL, suggesting a weak or noisy influence. For fan speed, the Random Forest PDP in Figure 7e shows a subtle positive linear trend, implying that slightly higher fan speeds may correspond to longer predicted RUL. However, given the narrow range of this, this trend may be due to localized data behavior. Furthermore, the slight variation in this PDP plot suggests that this variable has relatively little influence on the model's prediction. In contrast, the Gradient Boosting PDPs for both features (Figures 7d and 7f are nearly flat, indicating that the model considers these variables uninformative for RUL prediction. This consistent insensitivity highlights Gradient Boosting's ability to filter out irrelevant features, reinforcing its strength in maintaining a focused and generalizable prediction structure. In addition, negative values observed in particle dependence of RUL for the Gradient Boosting Regression model (Figures 7b, 7d, and 7f) suggest that the dependence of predicted RUL on the feature of interest is lower as compared to the model's average prediction. This behavior is a common characteristic of PDPs and does not imply a negative lifespan. Instead, it reflects a relative decline in the model's expected RUL.

Overall, the PDP analysis highlights the complementary strengths of the Random Forest and Gradient Boosting models in predicting Remaining Useful Life (RUL) for aircraft engines. The Random Forest model demonstrates strong responsiveness to critical degradation phases, particularly through its sharp transitions in response to time, suggesting it is well-suited for identifying early signs of failure or sudden wear. However, its interpretation of less relevant features, such as Mach number and fan speed, can be inconsistent or overly sensitive. In contrast, the Gradient Boosting model captures smooth and gradual degradation patterns, showing greater stability and robustness in its predictions. Its insensitivity to weak features like Mach number and fan speed reflects effective internal regularization, making it more reliable in later stages of degradation when wear progresses slowly and predictably. Based on the feature importance analysis (Figures 3 and 4), which identifies the most revealing features; Figures 5 and 6, which present the comparison between actual and predictive power of the models; and the Partial Dependence Plots in Figure 7, Gradient Boosting appears to provide smooth and consistent responses to gradual degradation, while Random Forest demonstrates greater sensitivity to abrupt changes. Therefore, selecting the appropriate model thus depends on the specific degradation profile and operational context of the engine system.

To contextualize the effectiveness of the proposed approach, the obtained results were compared with prior studies that employed the same FD001 subset of the C-MAPSS dataset. The Random Forest model in this study achieved an R^2 value of 0.864, which aligns with performance levels reported in the literature. For instance, Muneer et al. (2021) implemented a deep learning-based framework combining convolutional and recurrent layers (CNN-LSTM) and reported R^2 values of approximately 0.860 on FD001. Their study emphasized the importance of capturing temporal dependencies to improve RUL prediction. In another effort, Li et al. (2018) applied ensemble learning techniques, combining the outputs of multiple models to improve robustness, and achieved an R^2 score of 0.865. Their findings highlighted that ensemble methods can effectively reduce variance and increase prediction reliability through model diversity. This aligns with the principle demonstrated in our study, where a single tree-based model, when paired with careful feature selection and natural logarithmic transformation, was able to achieve a comparable level of accuracy. Although a direct benchmarking exercise was not conducted, these comparisons suggest that properly tuned, interpretable models like Random Forest can perform competitively with more complex architectures in the context of RUL prediction.

4. CONCLUSION

This study significantly contributes to the field of predictive maintenance by optimizing the performance of RUL prediction models for aircraft engines. Using multiple linear regression for model development, the main aim of the study was to enhance the model's prediction accuracy through optimal feature selection, addressing the challenge of inefficient RUL estimations. For this purpose, three techniques were investigated: backward elimination, random forest, and gradient boosting, and their performance was compared using various metrics. Results indicated that certain features were more critical and demonstrated a much stronger correlation with RUL compared to others, indicating their greater importance in the predictive models.

The key contributions of this study are as follows:

- The random forest model performed best with a higher R-squared value (0.864) and lower error metrics compared to backward elimination for linear regression and gradient boosting. This implies that complex models subjected to non-linear transformations can substantially improve the accuracy of predictive model. Furthermore, negative correlations between operational parameters suggested a compensating mechanism within engine operations. This insight can lead to more targeted interventions to optimize engine performance.
- The study identified significant sensor redundancy, particularly among certain sensors with nearly perfect correlations. This finding is crucial for optimizing sensor deployment and reducing computational costs without compromising model accuracy. However, in safety-critical applications such as aircraft, this kind of redundancy may be intentionally added to enhance reliability and fault tolerance; therefore, any optimization needs to be carried out cautiously for robust system performance.
- The Partial Dependence Plots (PDPs) for both Random Forest and Gradient Boosting models highlight how each model interprets the influence of selected features on RUL prediction. While time emerged as a strong and consistent predictor in both models, Mach number and demanded corrected fan speed showed negligible or inconsistent influence, particularly in Gradient Boosting.
- The analysis reinforces the theoretical expectation that RUL declines with time, validating time as a critical predictor. The Random Forest model captured abrupt degradation patterns, while Gradient Boosting offered smoother predictions, together providing deeper insight into how sensor data can inform predictive maintenance strategies. The minimal impact of Mach number and fan speed also underscores the importance of feature selection in improving model interpretability and efficiency.

The empirical evidence from the study can inform the development of refined maintenance protocols that reflect the true drivers of engine degradation, improving overall maintenance efficacy and reliability. This ensures enhanced operational efficiency and reliability, laying a strong foundation for future research and practical applications in predictive maintenance.

5. FUTURE RECOMMENDATION

- In this study, a natural logarithmic transformation was applied to address data skewness and reduce variability among features, contributing to improved model performance. While effective, alternative transformation techniques such as the square root or Box-Cox methods may further enhance the robustness and accuracy of predictive models. Future research could explore and compare these approaches to assess their impact on feature distribution and model outcomes.
- Future research may also include a comparison of the outcomes of natural logarithmic transformation with other bases, such as base-10 logarithm, because examining alternative logarithmic bases could provide useful insights into the model's sensitivity to transformation choices.
- Moreover, as the current work focused exclusively on sensor-based degradation data, the influence of maintenance activities on RUL prediction could not be explicitly evaluated due to the absence of maintenance records in the dataset. Incorporating such records in future studies could provide valuable insights into how scheduled and unscheduled interventions shape degradation trajectories and affect the reliability of predictive maintenance strategies.

REFERENCES

Adryan, F.A., & Sastra, K.W. (2021). Predictive Maintenance for Aircraft Engine Using Machine Learning: Trends and Challenges. *Avia*, 3(1). doi:https://doi.org/10.47355/avia.v3i1.45.

Angelov, P., Filev, D.P., & Kasabov, N. (2010). Evolving Intelligent Systems. *John Wiley & Sons*.

Babu, G.S., Zhao, P., & Li, X.-L. (2016). Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life. *Database Systems for Advanced Applications*, 9642, pp. 214–228. doi:https://doi.org/10.1007/978-3-319-32025-0_14.

Cheng, J., Sun, J., Yao, K., Xu, M., & Cao, Y. (2022). A variable selection method based on mutual information and variance inflation factor. Spectrochimica Acta Part A: *Molecular and Biomolecular Spectroscopy*, 268, p. 120652. doi:https://doi.org/10.1016/j.saa.2021.120652.

Chinta, V.S., Kethi Reddi, S., & Yarramsetty, N. (2023). Optimal feature selection on Serial Cascaded deep learning for predictive maintenance system in automotive industry with fused optimization algorithm. *Advanced Engineering Informatics*, [online] 57, p. 102105. doi:https://doi.org/10.1016/j.aei.2023.102105.

Çınar, Z. M., Nuhu, A. A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0. *Sustainability*, 12(19), 8211. https://doi.org/10.3390/su12198211

Cui, X., Lu, J. and Han, Y., 2023. Remaining Useful Life Prediction for Two-Phase Nonlinear Degrading Systems

with Three-Source Variability. Sensors, 24(1), p.165. https://doi.org/10.3390/s24010165

Darlington, R.B., & Hayes, A.F. (2017). Regression analysis and linear models: concepts, applications, and implementation. *New York: The Guilford Press*.

Fashoto, S., Mbunge, E., Ogunleye, G., & Van den Burg, J. (2021). Implementation of Machine Learning for Predicting Maize Crop Yields using Multiple Linear Regression and Backward Elimination. Malaysian *Journal of Computing*, 6(1), p. 679. doi:https://doi.org/10.24191/mjoc.v6i1.8822.

García, C.B., García, J., López Martín, M.M., & Salmerón, R. (2014). Collinearity: revisiting the variance inflation factor in ridge regression. *Journal of Applied Statistics*, 42(3), pp. 648–661. doi:https://doi.org/10.1080/02664763.2014.980789.

Gardner-Frolick, R., Boyd, D., & Giang, A. (2022). Selecting Data Analytic and Modeling Methods to Support Air Pollution and Environmental Justice Investigations: A Critical Review and Guidance Framework. *Environmental Science & Technology*, 56(5), pp. 2843–2860. doi:https://doi.org/10.1021/acs.est.1c01739.

Hayajneh, A.M., Alasali, F., Salama, A., & Holderbaum, W. (2024). Intelligent Solar Forecasts: Modern Machine Learning Models and TinyML Role for Improved Solar Energy Yield Predictions. *IEEE Access*, [online] 12, pp. 10846–10864. doi:https://doi.org/10.1109/ACCESS.2024.3354703.

Hong, C.W., Lee, C., Lee, K., Ko, M.-S., Kim, D.E., & Hur, K. (2020). Remaining Useful Life Prognosis for Turbofan Engine Using Explainable Deep Neural Networks with Dimensionality Reduction. *Sensors*, 20(22), p. 6626. doi:https://doi.org/10.3390/s20226626.

Hu, Y., Liu, S., Lu, H., & Zhang, H. (2019). Remaining Useful Life Model and Assessment of Mechanical Products: A Brief Review and a Note on the State Space Model Method. *Chinese Journal of Mechanical Engineering*, 32(1). doi:https://doi.org/10.1186/s10033-019-0317-y.

Inglis, A., Parnell, A., & Hurley, C.B. (2022). Visualizing Variable Importance and Variable Interaction Effects in Machine Learning Models. *Journal of Computational and Graphical Statistics*, 31(3), pp.766–778. doi:https://doi.org/10.1080/10618600.2021.2007935.

Javaid, M., Haleem, A., Singh, R.P., Rab, S., & Suman, R. (2021). Significance of Sensors for Industry 4.0: Roles, Capabilities, and Applications. *Sensors International*, [online] 2, p. 100110. doi:https://doi.org/10.1016/j.sintl.2021.100110.

Kartal, E., & Altunkaynak, A. (2024). Empirical-Singular-Wavelet Based Machine Learning Models for Sea Level Forecasting in the Bosphorus Strait: A Performance Analysis. *Ocean modelling (Oxford. Print)*, 188, pp. 102324–102324. doi:https://doi.org/10.1016/j.ocemod.2024.102324.

Karuppiah, K., Sankaranarayanan, B., & Ali, S.M. (2021). On sustainable predictive maintenance: Exploration of key barriers using an integrated approach. *Sustainable Production and Consumption*, 27, pp. 1537–1553. doi:https://doi.org/10.1016/j.spc.2021.03.023.

Kuhn, M., & Johnson, K. (2018). Max Kuhn and Kjell Johnson. Applied Predictive Modeling. *New York, Springer. Biometrics*, 74(1), pp. 383–383. doi:https://doi.org/10.1111/biom.12855.

Lee, S.M., Lee, D.H., & Kim, Y.S. (2019). The quality management ecosystem for predictive maintenance in the Industry 4.0 era. *International Journal of Quality Innovation*, [online] 5(1). doi:https://doi.org/10.1186/s40887-019-0029-5.

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, [online] 104, pp. 799–834. doi:https://doi.org/10.1016/j.ymssp.2017.11.016.

Li, Z., Goebel, K., & Wu, D. (2018). Degradation Modeling and Remaining Useful Life Prediction of Aircraft Engines Using Ensemble Learning. *Journal of Engineering for Gas Turbines and Power*, 141(4), 041008. https://doi.org/10.1115/1.4041674

Li, P., Zhang, Z., Grosu, R., Deng, Z., Hou, J., Rong, Y., & Wu, R. (2022). An end-to-end neural network framework for state-of-health estimation and remaining useful life prediction of electric vehicle lithium batteries. *Renewable and Sustainable Energy Reviews*, 156, p. 111843. doi:https://doi.org/10.1016/j.rser.2021.111843.

Liao, L., & Kottig, F. (2014). Review of Hybrid Prognostics Approaches for Remaining Useful Life Prediction of Engineered Systems, and an Application to Battery Life Prediction. *IEEE Transactions on Reliability*, 63(1), pp. 191–207. doi:https://doi.org/10.1109/tr.2014.2299152.

Liao, L., & Köttig, F. (2016). A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction. *Applied Soft Computing,* 44, pp. 191–199. doi:https://doi.org/10.1016/j.asoc.2016.03.013.

Liu, Q., Liang, J., & Ma, O. (2020). A physics-based and data-driven hybrid modeling method for accurately simulating complex contact phenomenon. *Multibody System Dynamics*, 50(1), pp. 97–117. doi:https://doi.org/10.1007/s11044-020-09746-w.

Maulud, D., & Abdulazeez, A.M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, 1(4), pp. 140–147. doi:https://doi.org/10.38094/jastt1457.

Melis, D.A., & Jaakkola, T. (2018). Towards Robust Interpretability with Self-Explaining Neural Networks. [online] *Neural Information Processing Systems*. Available at: https://proceedings.neurips.cc/paper/2018/hash/3e9f0fc9b2f89e043bc6233994dfcf76-Abstract.html [Accessed 9 Mar. 2024].

Mishra, D., Awasthi, U., Pattipati, K.R., & Bollas, G.M. (2023). Tool wear classification in precision machining using distance metrics and unsupervised machine learning. *Journal of intelligent manufacturing*. doi:https://doi.org/10.1007/s10845-023-02239-5.

Mobley, R.K. (2002). An Introduction to Predictive Maintenance. Elsevier.

Muneer, A., Taib, S.M., Naseer, S., Ali, R.F., & Aziz, I.A. (2021). Data-Driven Deep Learning-Based Attention Mechanism for Remaining Useful Life Prediction: Case Study Application to Turbofan Engine Analysis. *Electronics*, 10(20), p. 2453. doi:https://doi.org/10.3390/electronics10202453.

Natsumeda, M. (2023). e-RULENet: remaining useful life estimation with end-to-end learning from long run-to-failure data. *SICE Journal of Control, Measurement, and System Integration*, 16(1), pp. 164–171. doi:https://doi.org/10.1080/18824889.2023.2195982.

Raposo, F. (2016). Evaluation of analytical calibration based on least-squares linear regression for instrumental techniques: A tutorial review. *TrAC Trends in Analytical Chemistry*, 77, pp. 167–185. doi:https://doi.org/10.1016/j.trac.2015.12.006.

Rial, D., Kebir, H., Wintrebert, E., & Roelandt, J.-M. (2014). Multiaxial fatigue analysis of a metal flexible pipe. *Materials & Design (1980-2015)*, 54, pp. 796–804. doi:https://doi.org/10.1016/j.matdes.2013.08.105.

Rong, M., Gong, D., & Gao, X. (2019). Feature Selection and Its Use in Big Data: Challenges, Methods, and Trends. *IEEE Access*, 7, pp. 19709–19725. doi:https://doi.org/10.1109/access.2019.2894366.

Salahuddin, Z., Woodruff, H.C., Chatterjee, A., & Lambin, P. (2022). Transparency of deep neural networks for medical image analysis: A review of interpretability methods. *Computers in Biology and Medicine*, [online] 140, p. 105111. doi:https://doi.org/10.1016/j.compbiomed.2021.105111.

Saxena, A., and Goebel, K. (2008). PHM08 Challenge Data Set, NASA Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA.

Available at: https://data.nasa.gov/download/nk8v-ckry/application%2Fzip

Shachar, N., Mitelpunkt, A., Kozlovski, T., Galili, T., Frostig, T., Brill, B., Marcus-Kalish, M., & Benjamini, Y. (2018). The Importance of Nonlinear Transformations Use in Medical Data Analysis. *JMIR Medical Informatics*, 6(2), p. e27. doi:https://doi.org/10.2196/medinform.7992.

Shaheen, B., Kocsis, Á., & Németh, I. (2023). Data-driven failure prediction and RUL estimation of mechanical components using accumulative artificial neural networks. *Engineering Applications of Artificial Intelligence*, 119, p. 105749. doi:https://doi.org/10.1016/j.engappai.2022.105749.

Shim, H., Bonifay, W., & Wiedermann, W. (2022). Parsimonious asymmetric item response theory modeling with the complementary log-log link. *Behavior Research Methods*. doi:https://doi.org/10.3758/s13428-022-01824-5.

Talaat, F.M., Aljadani, A., Alharthi, B., Farsi, M., Badawy, M., & Elhosseini, M.A. (2023). A Mathematical Model for Customer Segmentation Leveraging Deep Learning, Explainable AI, and RFM Analysis in Targeted Marketing. *Mathematics*, 11(18), pp. 3930–3930. doi:https://doi.org/10.3390/math11183930.

Theissler, A., Pérez-Velázquez, J., Kettelgerdes, M., & Elger, G. (2021). Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. *Reliability Engineering & System Safety*, 215, p. 107864. doi:https://doi.org/10.1016/j.ress.2021.107864.

Victoria, M., & E. Priyadarshini (2023). Aeroengine gas trajectory prediction using time-series analysis auto regressive integrated moving average. *Aircraft Engineering and Aerospace Technology*. doi:https://doi.org/10.1108/aeat-01-2023-0018.

Xiong, M., & Wang, H. (2022). Digital twin applications in aviation industry: A review. *The International Journal of Advanced Manufacturing Technology*, 121. doi:https://doi.org/10.1007/s00170-022-09717-9.

Yu, H., & Hutson, A.D. (2022). Inferential procedures based on the weighted Pearson correlation coefficient test statistic. *Journal of Applied Statistics*, pp. 1–16. doi:https://doi.org/10.1080/02664763.2022.2137477.

Zhang, W., Yang, D., & Wang, H. (2019). Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. *IEEE Systems Journal*, [online] 13(3), pp. 2213–2227. doi:https://doi.org/10.1109/JSYST.2019.2905565.

Zhang, Y., Xin, Y., Li, Q., Ma, J., Li, S., Lv, X., & Lv, W. (2017). Empirical study of seven data mining algorithms on different characteristics of datasets for biomedical classification applications. *BioMedical Engineering OnLine*, 16(1). doi:https://doi.org/10.1186/s12938-017-0416-x.

Zhao, J., Han, X., Ouyang, M., & Burke, A.F. (2023). Specialized deep neural networks for battery health prognostics: Opportunities and challenges. *Journal of Energy Chemistry*, [online] 87, pp.416–438. doi:https://doi.org/10.1016/j.jechem.2023.08.047.