Predicting Aircraft Engine Failures using Artificial Intelligence

Asmae BENTALEB, Kaoutar TOUMLAL, Jaafar ABOUCHABAKA

Laboratory of Research in Informatics, Faculty of Science,

Ibn Tofail University, Kenitra, Morocco

Abstract—Nowadays, the aviation sector continues to develop especially with the emergence of new technologies, and solutions. Hence, there is an increasing demand for enhanced safety and operational efficiency in the aviation industry. As to guarantee this safety, the aircraft's engines must be monitored, controlled and maintained, however in an efficient way. Thus, the research community is working continuously in order to provide solutions that are efficient and cost effective. Artificial intelligence and more specifically machine learning models have been employed in this sense. Here comes the proposition of this article. It presents solutions implementing predictive maintenance using machine learning models. They help in predicting aircraft's failures. This is in order to avoid operations of unscheduled maintenance and disruptions of services.

Keywords—*Aircraft engine failures; machine learning; predictive maintenance; C-MAPSS; aviation safety*

I. INTRODUCTION

The aviation industry has entered a new phase of aircraft maintenance and reliability through the integration of cutting-edge technology and advanced data analysis. In aviation, the availability and well functioning of aircraft components have always been crucial. Aircraft systems and component availability is increased by making accurate failure predictions. The timing of maintenance operations is a crucial factor in determining the total cost of maintenance and overhaul for aircraft components, which account for a substantial amount of all operating expenses for aviation systems. In the aviation industry, there are three primary forms of equipment maintenance. Corrective maintenance deals with maintenance procedures and unplanned issues, such as machine and equipment breakdowns, that arise when using aircraft equipment. Preventive maintenance aims to reduce unplanned repairs through periodic maintenance, preventing equipment failures or machinery breakdowns. Tasks are planned to avoid unexpected downtime and breakdown events, minimizing the need for repair operations. Predictive maintenance, as its name implies, utilizes parameters measured during equipment operation to anticipate potential failures. Its goal is to intervene before faults occur, reducing unexpected failures by providing people working in maintenance, with more reliable scheduling options for preventive maintenance. Evaluating system reliability is crucial in selecting the appropriate maintenance strategy. With the emergence of artificial intelligence technologies, preventive maintenance has know interesting progress. Thanks to AI approaches, and its ability to analyse large historical data including aircraft components, engine performance, sensor readings, and maintenance records, preventive analytics can be implemented in order to predict issues before they happen. This reduces the risk of unplanned downtime and allow timely intervention. AI also helps in the efficient prioritisation of tasks based on their criticality, and optimises the resources allocation accordingly. Finally, AI technologies allow the monitoring of aircrafts in real time thanks to deploying sensor and other IoT devices on the aircrafts components in order to monitor their health and performance. In this article, one of the aspects of AI will be used to implement the predictive maintenance, which is the machine learning one. The choice of these models is based on their performance in the literature. They will be explored in analysing and exploring the extensive Commercial Modular Aero-Propulsion System Simulation (CMAPSS) dataset. The article suggests an approach that starts with an in depth exploration and preparation of the data which is the core module of machine learning and the decision making system. This includes using histograms to understand the distribution of relevant variables. This step offers insights into the statistical characteristics of the data and aids in identifying potential patterns and anomalies. This process involves selecting and engineering relevant attributes that provide a comprehensive view of engine health and potential failure scenarios. These meticulous steps serve as the foundation for constructing robust predictive models with the potential to redefine aviation maintenance practices, which is the next step of building machine learning models. This article elucidates the significance of these advancements, the methodologies deployed, the resulting insights, and their far-reaching implications for the aerospace industry in order to enhance the safety and efficiency of aircraft engines.

A. Problem Statement

Predictive maintenance in aviation is a key factor in ensuring flight safety and performance. Using advanced technologies such as analysis of sensor data, artificial intelligence, airlines can anticipate potential failures and take corrective action before problems become critical. This proactive approach helps minimize flight interruptions, reduce maintenance costs and optimize resources use. It is within this framework that this project is located.

Hence, here the main issue is how to allow the prediction of these engines' failures?

II. RELATED WORKS

Prognostics and health management are critical in today's industrial big data era because they improve the accuracy of failure predictions in the future, which reduces expenses associated with inventory, maintenance, and labor. The NASA Commercial Modular Aero-Propulsion System Simulation dataset, an open-source benchmark with simulated turbofan

engine units subjected to realistic flight circumstances, was used for the 2021 PHM Data Challenge. The goal of earlier deep learning applications in this field was to forecast how long engine units would stay useful. Nevertheless, the lack of identification of failure mode information in these methods, has limited their Interpretability and practical usefulness.

To overcome these constraints, a novel prognostics approach has been introduced, incorporating a tailored loss function. This approach aims to concurrently assess the remaining usable life, identify the probable failing component or components, and anticipate the current state of health. The suggested approach combines principal component analysis to orthogonalize statistical time-domain characteristics, which are then fed into supervised regressors like XGBoost, artificial neural networks, random forests and extreme random forests. Almong these approaches, ANN-Flux was considered to be the most effective, with AUROC and AUPR values higher than 0.95 for every classification assignment.

Furthermore, ANN-Flux demonstrates a remarkable 38% reduction in the root mean square error (RMSE) for remaining useful life compared to previous methodologies, utilizing the same test split of the dataset. Importantly, this improvement is achieved with significantly less computational cost, showcasing the potential of the proposed approach in advancing the field of prognostics and health management in industrial contexts [1].

This study describes the aviation industry, and how it involves a vast amount of information and maintenance data holding the potential to yield meaningful insights into forecasting future actions. This study seeks to introduce machine learning models that include feature selection and data elimination techniques for predicting aircraft systems failures. Over a two-year period, maintenance and failure data for aircraft equipments were systematically collected, identifying nine input variables and one output variable. A hybrid data preparation model is proposed to enhance the accuracy of failure count predictions in a two-stage process.

The first step uses ReliefF, a feature selection technique, to determine which factors have the greatest and least impact. To remove inconsistent or noisy data, a modified version of the K-means method is applied in the next step. Using Multilayer Perceptron (MLP) as an Artificial Neural Network (ANN), Support Vector Regression (SVR), and Linear Regression (LR) as machine learning techniques, the hybrid data preparation model's performance is evaluated on the maintenance dataset. The models' efficacy is measured using evaluation measures such as the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Correlation Coefficient (CC).

The findings indicate that the hybrid data preparation model successfully predicts the failure count of the equipment, showcasing its efficiency in enhancing the accuracy of forecasting in the aviation industry [2].

An essential facilitator of intelligent maintenance systems is the capability to foresee the remaining useful lifetime (RUL) of their components, known as prognostics. Constructing data-driven prognostics models necessitates datasets that encompass run-to-failure trajectories. Nevertheless, in many real-world applications, obtaining large, representative run-to-failure datasets proves challenging as failures are infrequent

in numerous safety-critical systems. In order to stimulate the advancement of prognostics methodologies, authors in [3] have formulated a new authentic dataset comprising run-to-failure trajectories for a fleet of aircraft engines operating under genuine flight conditions. This dataset was generated utilizing the Commercial Modular AeroPropulsion System Simulation (CMAPSS) model developed by NASA.

This dataset incorporates damage propagation modeling that adds two more levels of accuracy to the methodology developed in previous research. First of all, it takes into consideration actual flying circumstances as reported by a commercial aircraft. By connecting the degradation process to its operational history, it also improves the degradation modeling. This dataset is useful not just for predictive issues but also for providing health and fault class information. Because of this, it has two uses: it can be used for fault diagnostics as well as prognostics.

It is necessary to have datasets with run-to-failure trajectories available in order to generate data-driven prognostic models. The dataset offers a new realistic dataset of run-to-failure trajectories for a small fleet of aircraft engines under realistic flight conditions in order to aid in the development of these approaches. This synthetic dataset was created using damage propagation modeling, which adds two new levels of authenticity to the modeling approach utilized in earlier research. It starts by taking into account actual flight circumstances as reported by a commercial aircraft. By connecting the degradation process to the operating history, it further expands the degradation modeling. The dataset was created using the dynamic model of the Commercial Modular AeroPropulsion System Simulation (C-MAPSS) [4].

This study [5] outlines the methodology for modeling damage propagation within the components of aircraft gas turbine engines. The proposed approach involves generating response surfaces for all sensors through a thermo-dynamical simulation model that considers variations in flow and efficiency across the modules of interest. Specifically, an exponential rate of change for both flow and efficiency loss is applied to each dataset, starting from a randomly selected initial deterioration set point. The rate of change signifies an unspecified fault with a progressively deteriorating impact, with constraints on the upper threshold but otherwise random selection for fault rates.

Damage can continue to spread until a certain failure criterion is satisfied. At each instant in time, a health index is defined as the minimum of several overlaying operating margins; when the health index approaches zero, the failure criterion is met. The time series (cycles) of sensed measurements, usually from aircraft gas turbine engines, make up the model's output. The produced data are used by Prognostics and Health Management (PHM) as challenge data [6].

III. The Proposed Approach

The main objective of this work is to set up a system capable of predicting potential failures in the turbofan engine of aircraft. To do this, an accurate and reliable artificial intelligence model was created, using data analysis and predictive techniques to anticipate failures.

- Early detection of anomalies: The main objective is to use models and advanced AI technologies to detect

early signs of anomalies or failures in the engine before they become critical.

- Increasing aircraft reliability: By identifying potential failures in advance, this work aims to improving the flight reliability and safety.

Initially, a detailed description of the C-MAPSS dataset is provided, including an introduction to the input variables and dataset composition. The suggested approach uses the categorization method in order to predict the eventual failure components. Three steps sum up the proposed methodology: First, histograms 2) Equilibrium data 3) smoothing of data 4) extraction of features; 5) obtaining the final predictions by developing a supervised machine learning model.

A. Dataset Description

The N-CMAPSS dataset has 40 engine units altogether and is divided into four supplied subsets. An overview of the failure mechanisms found in each subset is given in Fig. 1. The dataset's overall goal is to predict the RUL till catastrophic failure. Engine units are normally rated between 60 and 100 cycles. The duration of each flight cycle varies, and it is distinguished by 18 time series signals: Four descriptors of the flight data that summarize the dynamic operating conditions and fourteen real-time sensor measurements. Each cycle comprises the following additional variables in addition to the time series signals that are helpful in understanding the context of a flight cycle: the unit number, cycle number, a binary health state variable hs (set to 0 for unhealthy status and 1 for healthy status), and a categorical flight class variable Fc that represents the flight' length that is (set to 1 for short flights, 2 for medium flights, and 3 for long flights). Here, the simulated engines are operated past their optimal state until they eventually shut down [7].

Subset Name	Units	Failure Mode	Fan Fail	HPC Fail	HPT Fail	LPT Fail
DS01	10	1	No	No	Yes	No
DS02	10	2	Yes	No	No	No
DS03	10	3	No	Yes	No	No
DS04	10	4	No	No	No	Yes

Fig. 1. The N-CMAPSS dataset failure mode description.

B. Data Preparation

1) Histograms: To analyze the distribution of values in the dataset, a histogram was created, as shown in the Fig. 2 below. This histogram provides a clear graphical representation of the frequencies of the different values of the variable to be studied. On the horizontal axis, the different ranges of values were placed, and on the vertical axis, the frequency or density of occurrences in each bin, was represented. The shape of the histogram gives us immediate insight into the

central tendency, the overall shape of the data distribution. This visualization allows us to identify patterns, thus providing important information for understanding the nature of this dataset.
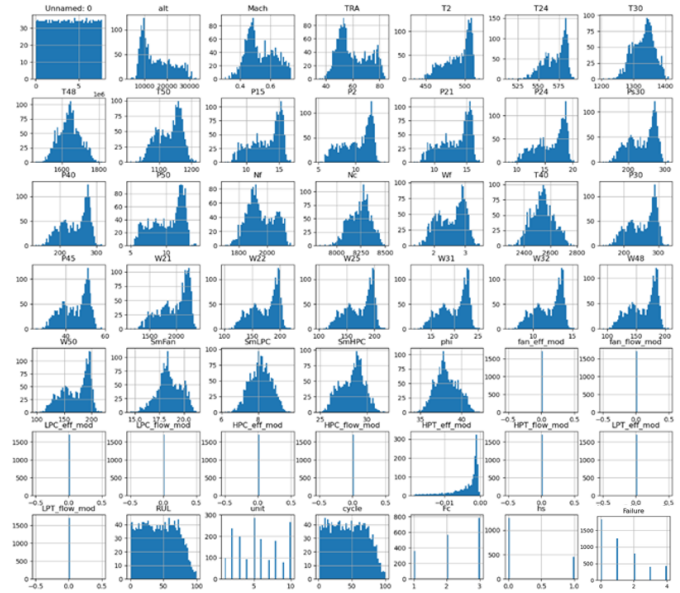

Fig. 2. The distribution of the N-CMAPSS dataset.

2) Data balancing: To address the issue of unbalanced data which is a challenging issue in machine learning, the Synthetic Minority Over-sampling Technique (SMOTE) has been opted in that sense. When one class significantly outnumbers the other, it can lead to biased model performance. SMOTE offers a solution by creating synthetic instances of the minority class. By interpolating between existing data points, SMOTE effectively increases the number of minority class samples, re-balancing the dataset. This approach not only mitigates the bias but also enables machine learning models to better recognize and generalize from the minority class. When coupled with other techniques or algorithms, SMOTE contributes to improved classification accuracy and, in turn, more robust and equitable model predictions [8] (Fig. 3 and 4).

3) Data smoothing: Using the Simple Moving Average (SMA), which is a straightforward yet effective technique widely employed in data analysis and time series forecasting. as shown in Fig. 5, SMA involves calculating the average of a fixed number of data points within a specified window or interval. This rolling average smooths out short-term fluctuations and emphasizes the overall trend in the data. SMA is particularly useful in identifying trends, cycles, and underlying patterns in time series data. Its simplicity and ease of implementation make it a popular choice for quick, preliminary analyses and trend detection. By reducing noise and highlighting long-term changes, SMA offers valuable insights for decision-making across various domains [9] (Fig. 5).

4) Data normalization: The Min-Max Scaler was used which is a widely used technique in data preprocessing, particularly in machine learning and statistics. This method
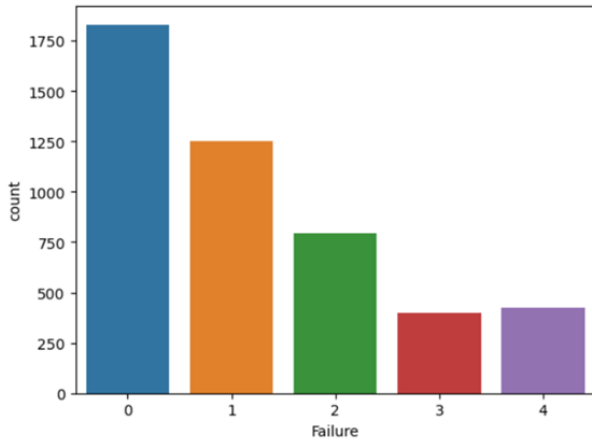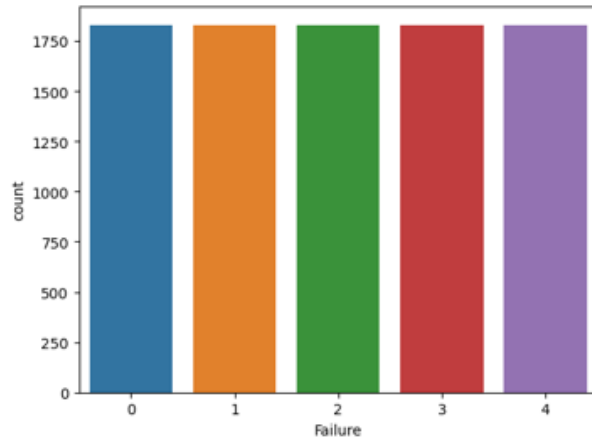
Fig. 3. The imbalanced dataset classes.

Fig. 5. The dataset before and after data smoothing.

Fig. 4. The balanced dataset classes after using smote method.

$$\text{Scaled Value } (x') = \frac{x - \min}{\max - \min}$$

units; however, the dataset has over 63 million timestamps and needs to be reduced for further data processing. Predictions were provided on a per-cycle basis, as in earlier studies.

In this article, the main focus is on the extraction of the cycle-wide statistical time domain feature to summarize the distribution for each time series using the mean.

D. Training the Models

In this article, four diverse machine learning models have been trained and developped including: Random forest [12], Support Vector Machines (SVM) [13], K-Nearest Neighbors (KNN) [14], and Gradient Boosting [15]. Each of these models represents a unique approach to solving predictive tasks, showcasing the versatility of machine learning techniques.

Here are the reasons why these models have been chosen. Decision Trees are known for their simplicity and interpretability. They are like a flowchart, making decisions by splitting data based on features. SVM, on the other hand, excels at finding optimal decision boundaries, making it valuable in tasks like classification and regression. K-Nearest Neighbors relies on the wisdom of the crowd, assigning data points to the most common class among their neighbors, while Gradient Boosting combines the wisdom of many weak learners to create a robust, ensemble model (Fig. 6).

E. Hyperparameters Finetuning

Hyperparameter adjustment is a pivotal phase in fine-tuning machine learning models, where the optimal set of hyperparameters is sought to achieve peak model performance. These hyperparameters, like learning rates, regularization terms, or tree depths, shape a model's behavior and its ability to generalize to new data. Two prominent methods for hyperparameter optimization are Grid Search and Random Search [16]

scales and transforms data to fit within a specified range, typically between 0 and 1, by subtracting the minimum value and dividing by the range of values within a feature. The Min-Max Scaler ensures that all features share a common scale, eliminating discrepancies in magnitudes and helping machine learning models perform optimally. This approach is valuable in scenarios where maintaining the original data distribution is not critical, and consistent feature scaling is more crucial [10].

C. Feature Extraction

Feature extraction is a fundamental process in data analysis and machine learning that involves transforming raw data into a more concise and informative representation. The objective is to retain essential information while reducing dimensionality and computational complexity. In essence, feature extraction selects the most relevant attributes or characteristics from the original dataset, thereby enhancing the efficiency and effectiveness of subsequent analysis or modeling [11].

Selection and extraction of features are required in order to lower the dataset's input dimensionality. Fig. 1 of the C-MAPSS dataset shows that there are only 40 turbofan engine
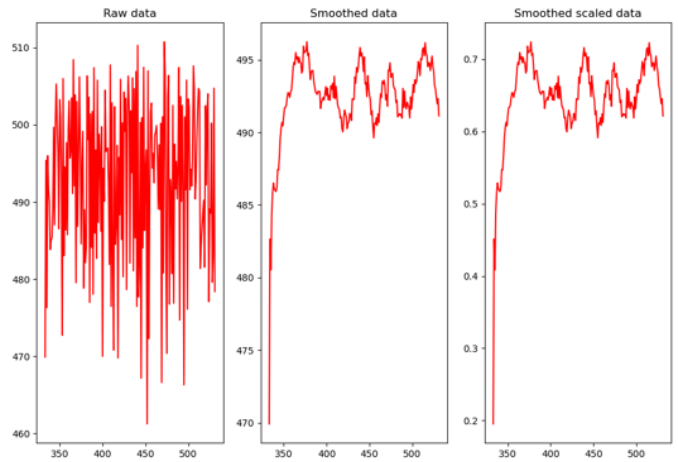
Modèle	Training accuraccy	Testing accuracy
Random forest	95.%	88.9%
KNN	82.8%	72.6%
SVM	51.8%	54.9%
GradientBoosting	87.3%	75.5%

Fig. 6. Models accuracy with the defaults parameters.

Grid Search systematically explores a predefined hyper-parameters space by evaluating models at various combinations of hyperparameters' values. It involves a structured and exhaustive search, where every possible combination is assessed. While Grid Search ensures thorough coverage of the hyperparameters space, it can be computationally expensive and impractical for large search spaces [17] (Fig. 7).

Model	Training accuraccy	Testing accuracy
Random forest	98.3%	91.6%
KNN	87%	75.1%
SVM	84.2%	80%
GradientBoosting	98.9%	95.9%

Fig. 7. Models accuracy with the Grid Search parameters.

In contrast, Random Search takes a more stochastic approach. It randomly samples hyperparameters values from specified distributions, which allows for a more efficient exploration of the hyperparameters space. Random Search can often yield excellent results with fewer iterations, making it a valuable alternative, particularly in scenarios where computational resources are limited [18].

Both methods serve as powerful tools for finding the optimal hyperparameters. The choice between Grid Search and Random Search depends on the nature of the problem, available resources, and the desired balance between thoroughness and efficiency in the hyperparameter optimization process (Fig. 8).

Model	Training accuraccy	Testing accuracy
Random forest	76.5%	68.2%
KNN	80.2%	70.8%
SVM	81.3%	82.8%
GradientBoosting	92.3%	80.1%

Fig. 8. Models accuracy with the Random Search parameters.

IV. RESULTS

The comparative analysis of machine learning models' performances on the engines' related data within was conducted with particular rigor to select the model best suited to the studied problem. The four examined models, namely Random Forest, kNN, SVM, and Gradient Boosting, underwent thorough evaluation, with meticulous optimization of their hyperparameters to maximize their performance.

The results of this evaluation revealed interesting achievements. As shown in the figure, the Gradient Boosting model emerged as the leader, displaying a remarkable accuracy of 90.9%. This outstanding result underscores the model's ability to effectively capture complex and non-linear relationships present in the dataset. The superior performance of Gradient Boosting compared to other models indicates its relevance and predictive power for the specific application.

The decision to select the Gradient Boosting model as the final model is supported by this superior performance. However, it is crucial to consider other aspects, such as the model's complexity and the resources required for its deployment. While Gradient Boosting is powerful, it may be more demanding in terms of training time and computational resources (Fig. 9).
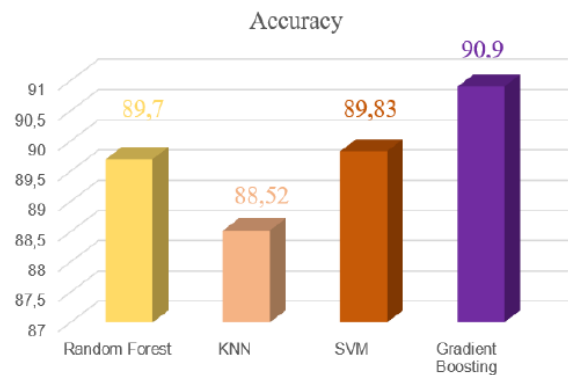

Fig. 9. The performance of the models.

As to position our results, gradient boosting algorithm in this article has given better accuracy value thanks to the used finetuning methods. If compared with the literature results [19], it gave better results. For example, in this article, gradient boosting has given less interesting results when applied to aircraft' dataset.

V. CONCLUSION

In conclusion, this article has performed a fascinating exploration of the world of machine learning in the context of predicting aircraft engine failures. A deep dive has been taken into the intricacies of four key machine learning models: Gradient Boosting, Decision Trees, Support Vector Machines (SVM) and K-Nearest Neighbors (KNN). These models serve as powerful tools in addressing the critical challenge of aviation engine reliability.

Throughout this journey, the inner workings of these models has been identified and their implication in the real-world applications in enhancing aircraft maintenance practices, has been witnessed. The significance of predictive maintenance in aviation safety and efficiency cannot be overstated, and machine learning provides a pathway to achieving this pivotal goal. Furthermore, the article has shed light on the importance

of hyperparameters' adjustment, as well as two powerful techniques, Grid Search and Random Search, that empower data scientists to fine-tune these models. This hyperparameters' optimization process represents a crucial step in ensuring the effectiveness of AI-driven predictive maintenance. In the realm of aviation, where safety is paramount, the fusion of data and artificial intelligence is transforming the landscape of engine failure prediction. The diversity of machine learning models, coupled with meticulous hyperparameters' tuning, have allowed the proactive detection and handling of potential engine issues, thus enhancing both the safety and efficiency of aircraft operations. Thus, the domain of Artificial intelligence is going to play a pivotal role in redefining the standards of excellence in aircraft engine maintenance.

VI. Data Availability

Concerning the availability of the C-MAPSS dataset, it can be downloaded from the Center of Excellence Data Repository of NASA; through the following link: https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository.

References

[1] X. H. Joseph Cohen and J. Ni, "Fault prognosis of turbofan engines: Eventual failure prediction and remaining useful life estimation." 2022.

[2] O. I. Kadir Celikmih and H. Uguz, "Fault prognosis of turbofan engines: Eventual failure prediction and remaining useful life estimation." 2020.

[3] K. G. Manuel Arias Chao, Chetan Kulkarni and O. Fink, "Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics." 2021.

[4] ——, "Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics." 2021.

[5] N. E. Abhinav Saxena, Don Simon, "Damage propagation modeling for aircraft engine run-to-failure simulation." 2021.

[6] K. Peters, "Predictive maintenance of turbofan engines," 2020.

[7] A. I. Izaak Stanton, Kamran Munir, "Predictive maintenance analytics and implementation for aircraft: Challenges and opportunities," 2022.

[8] C. Maklin, "Synthetic minority over-sampling technique (smote)," 2022.

[9] R. Dhir, "Data smoothing: Definition, uses, and methods," 2022.

[10] B. S. Dalwinder Singh, "Investigating the impact of data normalization on classification performance," 2020.

[11] Z. M. Wamidh K.Mutlag, Shaker K.Ali, "Feature extraction methods: A review," 2020.

[12] A. Shafi, "Random forest classification with scikit-learn," 2023.

[13] A. Navlani, "Support vector machines with scikit-learn tutorial," 2019.

[14] T. Srivastava, "A complete guide to k-nearest neighbors," 2023.

[15] P. Grover, "Gradient boosting from scratch," 2017.

[16] R. Adams, "Tunability: Importance of hyperparameters of machine learning algorithms," 2019.

[17] ——, "Grid search, random search, genetic algorithm: A big comparison for nas," 2019.

[18] J. Bergstra, "Random search for hyper-parameter optimization," 2022.

[19] S. Patil, A. Patil, V. Handikherkar, S. Desai, V. Phalle, and F. Kazi, "Remaining useful life (rul) prediction of rolling element bearing using random forest and gradient boosting technique," 11 2018, p. V013T05A019.