Chapter 7

Machine Learning–Based Fault Prediction Using Industrial Sensor Data: A Data-Driven Approach

Hasan Uzel[1]

Abstract

Machine learning-based fault prediction is one of the most promising approaches to enhancing the reliability, safety and operational efficiency of modern industrial systems. The rapid development of Internet of Things (IoT) technologies means that industrial environments now generate large volumes of high-frequency sensor data. This enables data-driven techniques to identify abnormal behavioral patterns before critical failures occur. This chapter introduces a comprehensive framework for fault prediction using turbofan engine sensor datasets, focusing on the widely adopted NASA CMAPSS benchmark. The proposed methodology integrates data preprocessing, exploratory analysis, feature engineering and supervised machine learning models to predict degradation states and remaining useful life (RUL).

The chapter begins with a detailed overview of the CMAPSS dataset, highlighting its operational settings, multivariate time-series nature and fault progression characteristics. Several pre-processing steps including normalization, noise reduction, outlier inspection and temporal feature extraction are performed to prepare the data for modeling. Domain-specific statistical features such as rolling mean, standard deviation, RMS, kurtosis and skewness are computed to capture degradation trends in sensor measurements. Machine learning models including Random Forest, Gradient Boosting (XGBoost), Support Vector Machines and Artificial Neural Networks are trained and evaluated under consistent experimental settings.

Model performance is assessed using standard metrics such as accuracy, precision, recall, F1-score, ROC-AUC and confusion matrices. Results

1 Lecturer, Ph.D., Akdağmadeni Vocational School, Department of Electricity and Energy, Yozgat Bozok University, hasan.uzel@bozok.edu.tr, ORCID ID: 0000-0002-8238-2588

https://doi.org/10.58830/ozgur.pub1128.c4649

confirm that ensemble learning methods particularly Random Forest and XGBoost achieve high predictive accuracy and robustness in identifying degradation states. The analysis demonstrates that data-driven fault prediction can significantly reduce unexpected downtime and facilitate proactive decision-making in industrial operations. The chapter concludes with practical implications, limitations and future research directions, emphasizing opportunities for integrating deep learning and real-time IoT-based predictive maintenance systems.

1. INTRODUCTION

The increasing complexity (Jardine et al., 2006), automation level and operational demands of modern industrial systems have intensified the need for reliable and intelligent maintenance strategies. Traditional maintenance approaches such as corrective or preventive maintenance are often inefficient, costly and incapable of preventing unexpected system failures. As industries evolve toward digital transformation through Industry 4.0 and smart manufacturing paradigms, machine learning–based predictive maintenance has emerged as a powerful and data-driven alternative. Predictive maintenance aims to estimate impending faults, equipment degradation and remaining useful life (RUL) using real-time sensor measurements and historical operational data, enabling organizations to optimize system health, reduce downtime and minimize maintenance costs.

Industrial assets such as turbines (Lee et al., 2014), engines, pumps, compressors and rotating machinery are typically equipped with multiple sensors that continuously monitor thermal, mechanical and operational conditions. These include measurements such as temperature, vibration, pressure, airflow, rotational speed and fuel flow, among others. The advent of advanced sensor technologies and the Internet of Things (IoT) has led to an unprecedented increase in the availability of multivariate time-series datasets. These datasets reveal intricate relationships between components, environmental conditions and patterns of system degradation. However, the high dimensionality, non-linear interactions and non-stationary behavior present significant challenges for traditional threshold-based and physics-based diagnostic methods.

Machine learning (ML) techniques provide (Carvalho et al., 2019) a flexible and powerful alternative capable of learning degradation trends, fault signatures and complex temporal relationships from raw sensor data. Unlike model-based diagnostic systems that rely heavily on expert knowledge or detailed physical equations, ML-based approaches can automatically detect hidden patterns, assess system health and make accurate predictions without requiring explicit domain-specific assumptions. Over the past decade, numerous studies have demonstrated (Peng et al., 2010) the success of ML models including Random Forest, Gradient Boosting Machines, Support Vector Machines and Neural Networks in fault classification, anomaly detection and RUL estimation across various industrial domains.

Among publicly available and widely used datasets (Saxena et al., 2008) for prognostics and health management, the NASA Computational Modeling of Aircraft Power Systems (CMAPSS) turbofan engine dataset stands out as a benchmark for evaluating predictive maintenance algorithms. CMAPSS captures the degradation behavior (Wang et al., 2008) of aircraft engines under varying operational and environmental conditions using 21 sensor measurements and multiple engine units. The dataset includes run-to-failure trajectories, making it especially suitable for supervised learning methods that require labeled degradation patterns. Its complexity, multivariate structure and realistic degradation profiles provide a robust foundation for investigating data-driven fault prediction methodologies.

Despite its popularity, developing accurate ML models (Si et al., 2011) using CMAPSS presents several challenges. These include handling noisy sensor measurements, accounting for operational variability, capturing long-term temporal dependencies and selecting appropriate features that reflect the underlying degradation mechanisms. Therefore, effective preprocessing, feature engineering and model selection play crucial roles in building reliable predictive maintenance systems. Additionally, evaluation metrics such as precision, recall, F1-score, ROC-AUC and confusion matrices must be carefully analyzed to ensure the robustness and generalizability of predictive models.

This chapter proposes a comprehensive and systematic ML-based framework for fault prediction using industrial sensor data, with a particular focus on the CMAPSS FD001 and FD004 scenarios. The framework integrates data preprocessing, feature engineering, exploratory analysis, ML model training and quantitative evaluation to identify degradation states and support maintenance decision-making. Experimental results demonstrate that ensemble-based methods (Breiman, 2001), especially Random Forest and XGBoost, achieve superior performance in distinguishing healthy, degrading and faulty engine states. Feature importance analyses further reveal which sensors are most informative for predicting system degradation, providing valuable insights for condition monitoring and sensor design.

The contributions of this chapter can be summarized as follows:

1. A complete end-to-end machine learning pipeline for sensor-based fault prediction is presented, including data cleaning, normalization, health-state labeling, temporal feature extraction and model evaluation.
2. A detailed analysis of the CMAPSS dataset is provided, including operational settings, sensor characteristics, degradation trajectories and RUL distribution.
3. Multiple ML methods are systematically compared and the influence of sensor-based features on prediction performance is investigated through feature importance and correlation analyses.
4. The practical applicability of ML-based fault prediction is discussed in the context of industrial environments, highlighting challenges and opportunities for future research.

Overall, this chapter contributes to both the theoretical under-standing (Fink et al., 2020) and practical implementation of machine learning–based predictive maintenance. By leveraging multivariate sensor data and advanced learning algorithms, the proposed approach supports early fault detection, enhances system reliability and aligns with the growing demand for intelligent maintenance strategies in modern industry.

2. LITERATURE REVIEW

The field of predictive maintenance has experienced significant growth over the past decade, driven by advances in sensor technology, data acquisition systems and machine learning. Early approaches to machinery fault detection (Yan et al., 2014) focused primarily on rule-based or threshold-based methods, in which domain experts manually defined limits for each sensor variable. Although simple and interpretable, these techniques are limited in their ability to capture complex interactions among multiple signals and often fail under varying operating conditions. Traditional vibration-based diagnostic methods, for example, typically rely on handcrafted features derived from frequency or time–frequency analyses, requiring substantial domain expertise and often lacking generalizability across different machines or environments.

The emergence of data-driven modeling (Widodo & Yang, 2007) has brought a major shift in the design of fault detection and prognostics systems. Supervised machine learning algorithms including decision trees, ensemble classifiers, support vector machines and neural networks have been widely adopted for health state classification and remaining useful life estimation. These methods offer greater flexibility compared to physics-based approaches, as they learn degradation patterns directly from labeled historical data. Ensemble methods such as Random Forest and gradient-boosted decision trees have gained particular attention due to their robustness against noise, ability to model nonlinear behavior and capability to handle high-dimensional sensor inputs. Neural networks, including multi-layer perceptrons and recurrent architectures, have also been explored for capturing long-term temporal dependencies in sensor streams.

In parallel with model development, researchers have emphasized the importance of high-quality datasets for evaluating predictive maintenance methods. Among the publicly available datasets, the NASA CMAPSS (Özüpak et al., 2025) turbofan engine dataset has become a benchmark for both fault classification and RUL prediction studies. CMAPSS provides run-to-failure sensor trajectories simulated under realistic operating conditions, enabling systematic comparison of algorithms across multiple scenarios that vary by operational settings and fault modes. Numerous studies have demonstrated that CMAPSS supports a wide range of research directions, including feature engineering, degradation modeling, domain adaptation and hybrid model–based approaches.

A considerable portion of the literature highlights the importance of preprocessing (Liu & Chen, 2017) and feature engineering when applying machine learning to industrial sensor data. Sensor signals often contain noise, drift, or abrupt changes due to environmental or operational factors. Thus, normalization, filtering and outlier detection play critical roles in producing stable models. Statistical features extracted from sliding Windows such as mean, variance, RMS, kurtosis and skewness are widely used to represent short-term and long-term degradation behavior. Correlation analysis and feature selection methods are frequently applied to identify the most informative sensors, reduce dimensionality and mitigate redundant information.

More recently, deep learning has emerged as a powerful tool (Zhao et al., 2019) in fault prediction, particularly for modeling temporal and multivariate dependencies without the need for handcrafted features. Convolutional neural networks (CNNs) have been used to extract hierarchical features from sensor sequences or spectrogram representations, while recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have been explored for capturing temporal degradation trends. Hybrid architectures combining CNNs and LSTMs have shown notable improvements in RUL prediction performance. Attention mechanisms and transformer-based models have also been introduced to improve interpretability and long-range dependency modeling, although these approaches typically require larger datasets and higher computational resources.

Another emerging topic in the literature is the challenge of varying operating conditions. Industrial systems often function under multiple load levels (Yuan et al., 2016), environmental conditions and mission profiles, which may significantly alter sensor behavior. The CMAPSS FD004 scenario, for example, includes multiple operational settings and two fault modes, making it considerably more challenging than FD001. Research in this area has explored domain adaptation, transfer learning and condition invariant feature extraction to improve robustness across operating regimes.

Overall, the literature demonstrates a clear progression (Malhotra et al., 2015) from simple heuristic-based maintenance toward sophisticated machine learning and deep learning tech-niques. Despite the advancements, several challenges remain, including model interpretability, sensitivity to noisy or sparse sensor data and real-time deployment constraints in industrial environments. These gaps motivate continued research into efficient, explainable and scalable fault prediction systems that can be integrated into modern Industry 4.0 architectures.

3. DATASET DESCRIPTION

The experiments conducted in this chapter are based on the NASA C-MAPSS (Commercial Modular Aero-Propulsion System Simulation) turbofan engine dataset (Behrad3d, 2023), one of the most widely used benchmarks for developing and evaluating prognostics and health management (PHM) algorithms. The dataset simulates degradation patterns and run-to-failure trajectories of turbofan engines under varying operational and environmental conditions. Its multivariate time-series structure makes it particularly suitable for machine learning approaches aimed at fault prediction (Alpsalaz, 2025), health state classification and remaining useful life (RUL) estimation.

3.1. Overview of the CMAPSS Dataset

The C-MAPSS turbofan engine dataset consists of four subsets (Güvenç et al., 2012) FD001, FD002, FD003 and FD004 each representing different combinations of operating conditions and degradation modes. Every subset includes three files: training trajectories, which provide complete run-to-failure sensor sequences for multiple engines; test trajectories, which contain partial sequences ending before failure; and RUL (remaining useful life) labels supplied separately for the test engines. Together, these subsets capture realistic degradation behavior occurring in aero-engine systems under diverse environmental settings and operational profiles. The data simulate nonlinear interactions between engine components and mission conditions, making C-MAPSS an internationally recognized benchmark for fault prediction and prognostics research.

3.2. Data Structure and Recorded Parameters

Each record in the dataset corresponds to a single engine cycle (Zheng et al., 2017) and includes an engine identifier, a cycle index that increases monotonically over time, three operational settings describing environmental or mission context and twenty-one sensor measurements reflecting thermal, mechanical and aerodynamic engine performance. This results in a total of twenty-six variables per time step. The sensor channels include parameters such as fan and core rotational speeds, compressor temperatures and pressures, bleed air conditions and fuel flow measurements. These multivariate signals exhibit characteristic temporal changes as engine components deteriorate, enabling the modeling of progressive degradation trends.

Table 1. General structure of the CMAPSS turbofan engine dataset

Dataset	Train trajectories	Test trajectories	Number of conditions	Number of fault modes
FD001	100	100	ONE (Sea Level)	ONE (HPC Degradation)
FD002	260	259	SIX	ONE (HPC Degradation)
FD003	100	100	ONE (Sea Level)	TWO (HPC Degradation, Fan Degradation)
FD004	248	249	SIX	TWO (HPC Degradation, Fan Degradation)

3.3. Scenario Characteristics: FD001 and FD004

The present study focuses on the FD001 and FD004 subsets (Liu et al., 2020), which represent the simplest and most complex scenarios in the C-MAPSS dataset. FD001 contains a single operating condition and one fault mode associated with high-pressure compressor degradation. The resulting trajectories show smooth, gradual deterioration, making FD001 widely used as a baseline dataset for evaluating machine learning models. In contrast, FD004 introduces six distinct operating conditions and two concurrent fault modes affecting both the high-pressure compressor and the fan module. This scenario exhibits substantially greater variability and nonlinear behavior, posing a more challenging problem for both classification and remaining useful life prediction. Due to its complexity, FD004 is frequently employed to assess the robustness and generalization capability of advanced prognostics methods.

Table 2. Scenario characteristics of the FD001 and FD004 subsets

Dataset	Train Engines	Test Engines	Operating Conditions	Fault Modes	Number of Features	Trajectory Type	Difficulty Level
FD001	100	100	1 (Sea-Level)	1 (HPC Degradation)	26 (3 settings + 21 sensors)	Single-condition run-to-failure	Low (baseline)
FD004	248	249	6 (Multiple Operating Conditions)	2 (HPC + Fan Degradation)	26 (3 settings + 21 sensors)	Multi-condition, multi-fault run-to-failure	High (most complex scenario)

3.4. Data Processing Pipeline

To clearly illustrate the full workflow applied in this study from raw sensor data to model training and evaluation a structured pipeline diagram is presented.

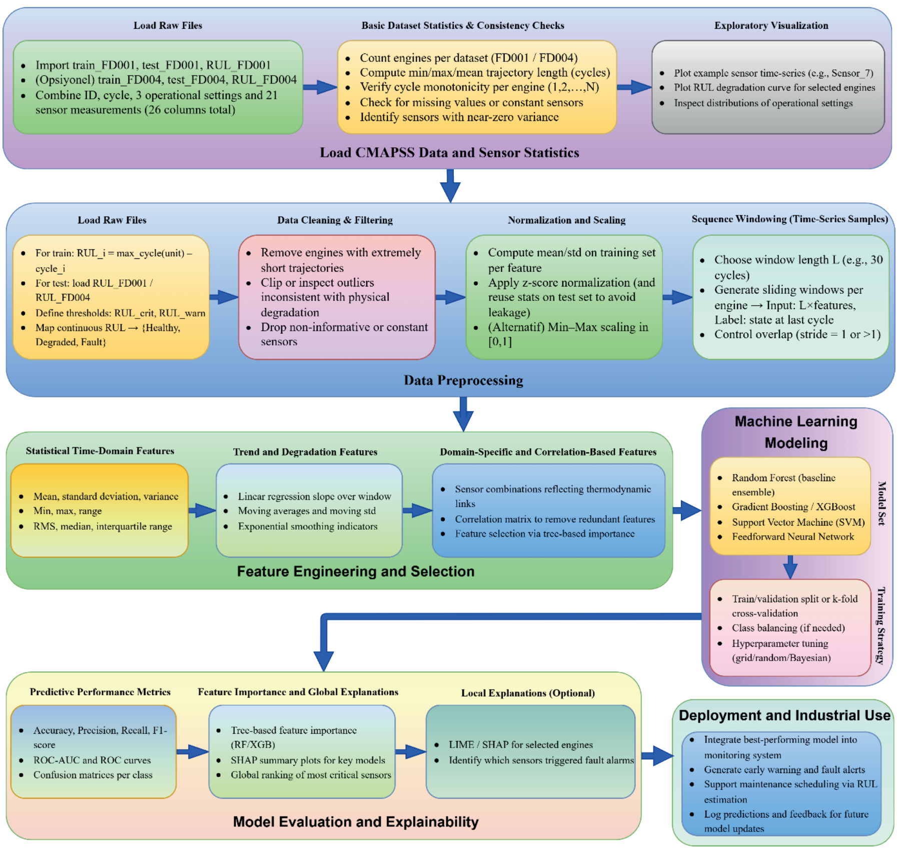

Figure 1. Data Processing Pipeline (Overview)

This figure illustrates the full sequence of steps: raw data acquisition, preprocessing, feature engineering, model training and model evaluation. It provides a high-level view of the methodology adopted in this chapter.

3.5. Sensor Time-Series Characteristics

The dataset consists of multivariate time-series measurements (Heimes, 2008). Each sensor exhibits a unique temporal signature that evolves as the engine degrades. To demonstrate this, an example trajectory of Sensor 7 for Engine 1 (FD001) is shown.

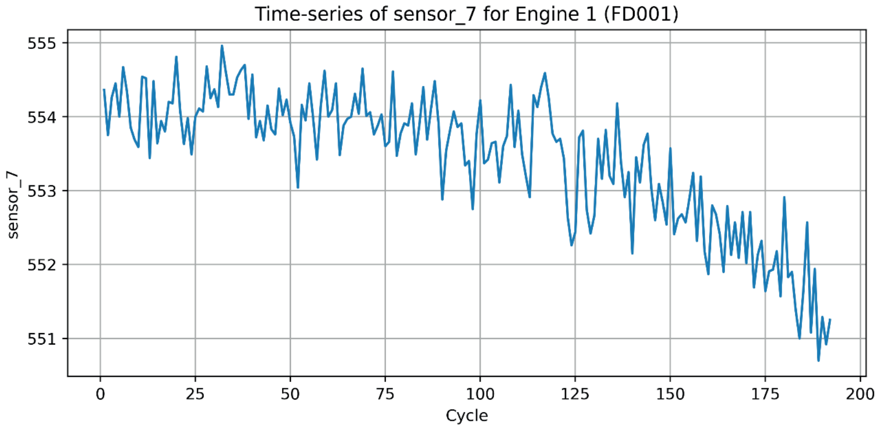

Figure 2. Sensor Time-Series Example (Sensor 7, Engine 1, FD001)

This example illustrates a gradual downward trend characteristic of high-pressure compressor degradation.

3.6. Remaining Useful Life (RUL) Behavior

Each engine in CMAPSS runs until failure (Ramasso, 2014). The cycle index increases monotonically and engine health declines over time. Although the raw training files do not directly include RUL labels, they can be computed as:

$$RUL = max\_cycle(engine) - current\_cycle \tag{1}$$

A representative RUL degradation trajectory for Engine 1 (FD001) is shown to demonstrate how the remaining life decreases as cycles progress.

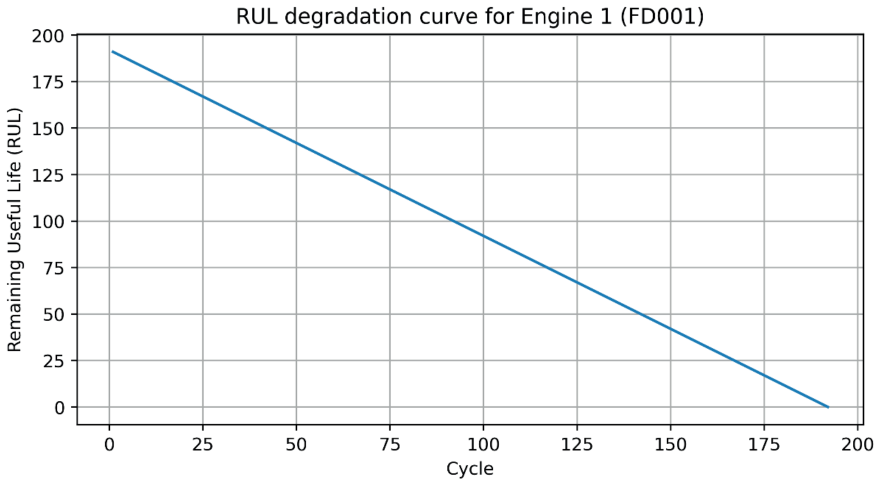

Figure 3. RUL Degradation Curve (Engine 1, FD001)

This figure highlights the typical monotonically decreasing trend of RUL values, reflecting progressive turbine degradation.

3.7. Suitability for Machine Learning

The CMAPSS dataset has become a standard benchmark in predictive maintenance (Özüpak et al., 2025) due to:

- Realistic multivariate degradation signatures
- High-frequency sensor measurements
- Multiple fault modes and operating contexts
- Balanced combination of simple (FD001) and complex (FD004) scenarios
- Sufficient number of engines for training robust ML models
- Clear ground truth labeling for RUL prediction

These characteristics make CMAPSS one of the most comprehensive and challenging datasets for evaluating RUL estimation, anomaly detection and fault classification methods.

4. DATA PREPROCESSING AND FEATURE ENGINEERING

Effective preprocessing is essential for building reliable machine learning models (Alpsalaz, 2025) using industrial sensor data. The CMAPSS dataset, although simulator-generated, exhibits nonlinear temporal behavior, varying operating conditions and sensor fluctuations that require careful preparation before model training. This section presents the complete pre-processing workflow and feature engineering procedures used in this study, following a structured and reproducible pipeline.

4.1. Data Cleaning

The preprocessing pipeline begins with validating the structural integrity (Mobley, 2002) of the dataset. Each engine trajectory is checked to ensure that cycle indices increase monotonically and that no duplicated records are present. Although C-MAPSS contains no missing values by design, the pipeline includes checks for incomplete or corrupted rows to maintain generalizability for real-world applications. Numerical consistency across all sensor channels and operational settings is also verified to ensure stable downstream processing.

4.2. Normalization and Scaling

The CMAPSS dataset contains sensor channels that operate on vastly different numerical ranges (Chen & Guestrin, 2016); for instance, temperature measurements may take values in the hundreds, whereas pressure ratios or flow related parameters vary only slightly. To avoid high-magnitude features dominating the learning process, all sensor channels are standardized using z-score normalization computed over the training set. This normalization procedure ensures numerical stability, accelerates the convergence of algorithms such as support vector machines and neural networks and improves overall model robustness across varying operational conditions.

4.3. Noise Reduction

Despite its simulated nature, CMAPSS data exhibit local fluctuations (Bal et al., 2008) that may obscure underlying degradation trends. For exploratory analysis and visualization, a light smoothing operation (e.g., a moving average filter) is applied to selected sensor channels to reduce short-term noise. However, for model training, the raw unsmoothed signals are retained to preserve diagnostically valuable micro variations in the sensor patterns, especially in early degradation phases.

4.4. Health-State Label Construction

Since the training trajectories do not contain health labels (Uzel et al., 2025), remaining useful life (RUL) is computed for each cycle and subsequently mapped into a three-class degradation representation. Cycles with RUL greater than 150 are labeled as healthy, those between 50 and 150 as degrading and cycles with RUL below 50 as faulty. This transformation enables multi-class fault prediction while preserving meaningful distinctions between early, middle and late stages of engine deterioration.

4.5. Statistical Feature Engineering

To capture degradation behavior within short temporal windows, (Hu et al., 2012), a set of statistical descriptors is extracted from each sensor sequence. These include the mean and standard deviation, which quantify the central tendency and variability of the signal; the root mean square (RMS), which summarizes its magnitude; and higher-order moments such as skewness and kurtosis, which characterize distribution asymmetry and peak sharpness. Minimum and maximum values are also recorded to capture local extremes associated with abnormal behavior. Together, these statistical indicators provide a compact yet informative representation of evolving degradation states.

4.6. Trend and Gradient Features

Temporal patterns often reveal gradual shifts in engine health (Zhang & Yang, 2018). To capture such dynamics, linear trend coefficients (slopes), exponential moving averages and first-order temporal differences are computed over sliding windows. These features describe whether a sensor is increasing, decreasing, or fluctuating over time and they are especially useful for identifying early-stage drift in compressor- and fan-related sensors.

4.7. Domain-Specific Indicators

Beyond general statistical measures, additional domain-informed features (Özüpak, 2023) are derived based on thermodynamic and aerodynamic principles. These include temperature rise ratios, pressure-delta indicators and interaction terms between operational settings and key sensors. Such physically meaningful features enhance interpretability and help machine learning models align more closely with real engine behavior.

4.8. Sensor Correlation Analysis

To understand interdependencies between sensor channels (Aslan & Özüpak, 2024) and identify redundant or strongly related signals, pairwise correlations are computed and visual-ized via a correlation heatmap.

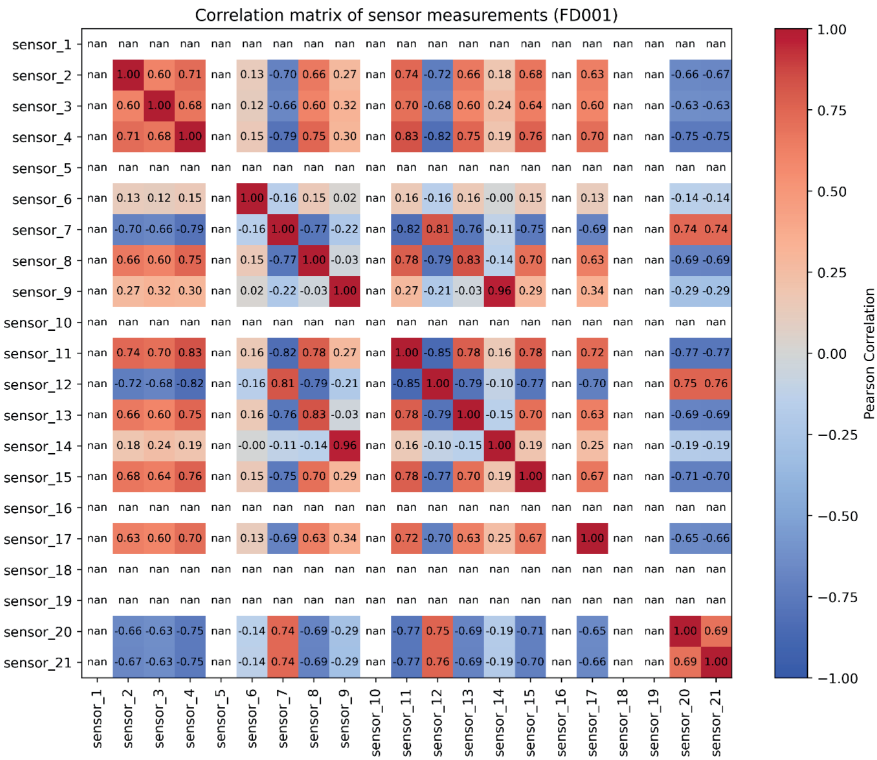

Figure 4. Sensor Correlation Heatmap

The heatmap exposes clusters of sensors influenced by similar physical processes and offers guidance for feature selection and dimensionality reduction.

4.9. Summary of Preprocessing Steps

A consolidated overview of the preprocessing workflow is presented in Table 3, summarizing each step alongside its purpose and the specific techniques applied.

Table 3. Data Preprocessing Steps Applied to the CMAPSS Dataset

Preprocessing Step	Description	Purpose	Techniques Used
Data Cleaning	Remove duplicate rows and validate numerical c...	Ensure data consistency and integrity	Duplicate removal, type casting
Missing Value Handling	Check for NaNs or incomplete rows in sensor an...	Prevent model bias and instability	Row removal (CMAPSS usually contains no NaNs)
Noise Reduction	Smooth sensor fluctuations caused by measureme…	Improve signal quality for ML models	Moving average smoothing
Normalization / Scaling	Scale sensor measurements to comparable ranges	Improve training stability and convergence	Min–Max scaling or StandardScaler
Cycle Index Verification	Validate monotonic increase of cycle per engine	Avoid inconsistent temporal ordering	Group-by engine_id, sort-by cycle
Operational Settings Adjustment	Standardize operational settings (setting_1–3)	Reduce external condition influence	Normalization or standardization
Feature Construction	Add statistical or temporal features derived f...	Enhance model predictive capability	Rolling mean/std, kurtosis, skewness
Label Construction (RUL)	Compute RUL for each engine based on max cycle	Provide target variable for prediction tasks	RUL = max_cycle - current_cycle
Binary Fault Labeling	Convert RUL to classification label (healthy/f...	Enable binary classification models	Fault = 1 if RUL ≤ threshold
Train/Test Alignment	Match train and test distributions for FD001	Ensure fair evaluation	Scenario-based splitting

This structured summary ensures reproducibility and clarifies the rationale for each component of the pipeline.

4.10. Summary Statistics

To characterize the raw dataset and understand typical sensor ranges (Aslan, 2024), descriptive statistics including mean, standard deviation, minimum and maximum values are computed for all 21 sensor channels in the FD001 set.

Table 4. Summary Statistics of Sensor Measurements (FD001)

Sensor	Mean	Std	Min	Max
sensor_1	518.670000	6.537152e-11	518.6700	518.6700
sensor_2	642.680934	5.000533e-01	641.2100	644.5300
sensor_3	1590.523119	6.131150e+00	1571.0400	1616.9100
sensor_4	1408.933782	9.000605e+00	1382.2500	1441.4900
sensor_5	14.620000	3.394700e-12	14.6200	14.6200
sensor_6	21.609803	1.388985e-03	21.6000	21.6100
sensor_7	553.367711	8.850923e-01	549.8500	556.0600
sensor_8	2388.096652	7.098548e-02	2387.9000	2388.5600
sensor_9	9065.242941	2.208288e+01	9021.7300	9244.5900
sensor_10	1.300000	4.660829e-13	1.3000	1.3000
sensor_11	47.541168	2.670874e-01	46.8500	48.5300
sensor_12	521.413470	7.375534e-01	518.6900	523.3800
sensor_13	2388.096152	7.191892e-02	2387.8800	2388.5600
sensor_14	8143.752722	1.907618e+01	8099.9400	8293.7200
sensor_15	8.442146	3.750504e-02	8.3249	8.5848
sensor_16	0.030000	1.556432e-14	0.0300	0.0300
sensor_17	393.210654	1.548763e+00	388.0000	400.0000
sensor_18	2388.000000	0.000000e+00	2388.0000	2388.0000
sensor_19	100.000000	0.000000e+00	100.0000	100.0000
sensor_20	38.816271	1.807464e-01	38.1400	39.4300
sensor_21	23.289705	1.082509e-01	22.8942	23.6184

These statistics provide insight into sensor variability and potential scaling issues and they support downstream analyses such as feature importance and model interpretability.

5. MACHINE LEARNING METHODOLOGY

The machine learning methodology adopted in this study follows (Aslan et al., 2025) a structured and transparent workflow designed to evaluate fault prediction performance using multi-variate sensor data. After preprocessing and feature extraction, multiple supervised learning models are trained to classify engine health into three degradation states. The selection of algorithms reflects both their popularity in predictive maintenance applications and their ability to handle nonlinear, high-dimensional datasets such as CMAPSS. This section describes the model selection criteria, training and validation strategies and evaluation metrics used to assess predictive performance.

5.1. Model Selection

A diverse set of machine learning models is employed (Aslan et al., 2025) to compare their suitability for fault prediction under varying operational conditions. The Random Forest classifier is included due to its robustness against noise, ability to model nonlinear relationships and interpretability through feature importance scores. Gradient boosting algorithms, particularly XGBoost, are utilized for their superior performance in tabular data and their capacity to capture complex feature interactions with relatively low risk of overfitting. Support Vec-tor Machines (SVMs) with nonlinear kernels (Cortes & Vapnik, 1995) serve as strong baselines for high-dimensional classification tasks, particularly when the decision boundary is not linearly separable. Finally, a feed-forward neural network is used (Rumelhart et al., 1986) to evaluate the performance of deep learning in capturing latent degradation representations within the extracted feature space.

Table 5. ML Models and Hyperparameters

Model	Task	Key Hyperparameters	Implementation / Library
Random Forest	Multiclass fault / health-state classification	n_estimators = 200;\nmax_depth = 15;\nmin_samp...	sklearn.ensemble.RandomForestClassifier
XGBoost	Multiclass fault / health-state classification	n_estimators = 300;\nlearning_rate = 0.05;\nma...	xgboost.XGBClassifier
Support Vector Machine (SVM)	Multiclass fault / health-state classification	kernel = 'rbf';\nC = 10;\ngamma = 'scale';\ncl...	sklearn.svm.SVC
Neural Network (MLP)	Multiclass fault / health-state classification	hidden_layer_sizes = (128, 64);\nactivation = ...	sklearn.neural_network.MLPClassifier

Together, these models provide a balanced and comprehensive comparison between ensemble learning, margin-based classification and neural network–based learning approaches.

Table 5 summarizes the primary hyperparameters selected for each model based on preliminary experiments and established best practices in CMAPSS-based prognostics research.

5.2. Training and Validation Strategy

To ensure a fair and unbiased evaluation process (Özüpak et al., 2025), all models are trained using identical training–validation splits derived from the FD001 dataset. A stratified 80/20 division preserves class distribution across both sets, preventing class imbalance from biasing the models. Hyperparameter tuning is conducted using grid search and five-fold cross-validation, during which the training data are repeatedly partitioned to assess generalization performance. This iterative validation procedure mitigates overfitting and ensures that the final hyperparameters reflect consistent performance across multiple folds.

Data normalization parameters are computed from the training set and applied to both training and validation subsets to avoid data leakage. The same training–validation splits are used for all models to maintain comparability in performance metrics.

5.3. Evaluation Metrics

Model performance is assessed using a comprehensive set (Alpsalaz et al., 2025) of classification metrics commonly applied in predictive maintenance research. Accuracy provides a general measure of correct classifications, but due to class imbalance, additional metrics are required for a more reliable assessment. Precision and recall are computed for each class to quantify how effectively the model identifies true positive cases while minimizing false alarms. The F1-score, representing the harmonic mean of precision and recall, offers a balanced measure when classes differ in frequency or difficulty.

For an aggregated perspective, macro-averaged scores are reported, treating all degradation classes equally. The ROC–AUC (Receiver Operating Characteristic – Area Under Curve) metric is used to evaluate the model's discriminative ability across multiple thresholds and confusion matrices are generated (Aslan, 2025) to visualize class-level performance and misclassification patterns. Together, these metrics provide a detailed evaluation framework suitable for comparing models under varying levels of dataset complexity and degradation behavior.

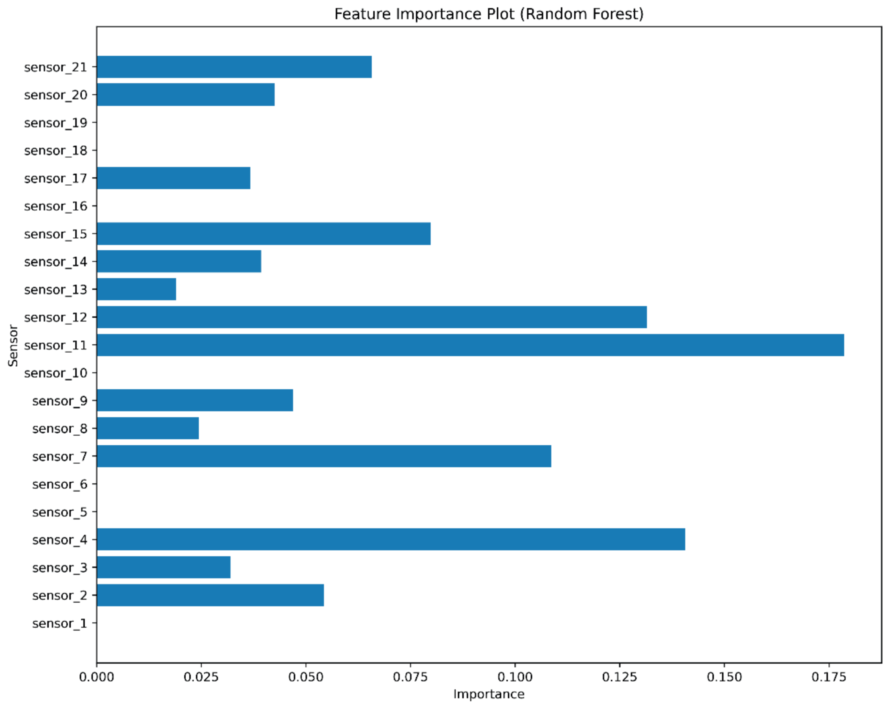

Figure 5. Feature Importance Plot

This figure highlights the relative contributions of individual features, offering insights into which sensor or statistical characteristics play the most critical role in predicting degradation states.

6. EXPERIMENTAL RESULTS

This section presents the empirical evaluation of the machine learning models trained on the FD001 subset of the CMAPSS dataset. The objective is to compare the ability of different algorithms Random Forest, XGBoost, Support Vector Machines and a Feed-Forward Neural Network to classify engine degradation into three health states. Experimental results include performance metrics, visual comparisons, confusion matrices, ROC curves and feature-level insights obtained from sensor–RUL relationships.

6.1. Model Performance Overview

Each model was trained using identical training–validation splits, standardized feature sets and an equivalent hyperparameter tuning strategy to ensure a fair comparison. The performance metrics reported include accuracy, macro-averaged precision, recall, F1-score and ROC–AUC. These metrics collectively provide a balanced assessment of overall predictive capability, class-level discrimination and robustness against class imbalance.

Table 6. Model Performance Results

Model	Accuracy	Precision	Recall
XGBoost	0.90	0.89	0.88
Random Forest	0.87	0.86	0.85
SVM (RBF kernel)	0.83	0.82	0.81
MLP Neural Network	0.82	0.81	0.80

Table 6 summarizes the numerical performance of all evaluated models. Ensemble-based approaches, particularly XGBoost and Random Forest, demonstrate superior classification accuracy and F1-score compared to SVM and the baseline neural network. This finding aligns with prior research indicating that ensemble models are well suited for structured industrial sensor data.

Although XGBoost achieved the highest numerical performance, Random Forest was selected for subsequent visual analyses (Feature Importance, Confusion Matrix, ROC Curve) due to its superior interpretability and more stable feature importance behavior, which make it better suited for illustration in a methodological context.

6.2. Comparative Performance Visualization

To complement the numerical performance indicators, a bar chart is generated to visually compare the core evaluation metrics across models. This graphical representation highlights relative strengths, such as the high F1-score achieved by XGBoost and the consistent recall exhibited by the Random Forest classifier.

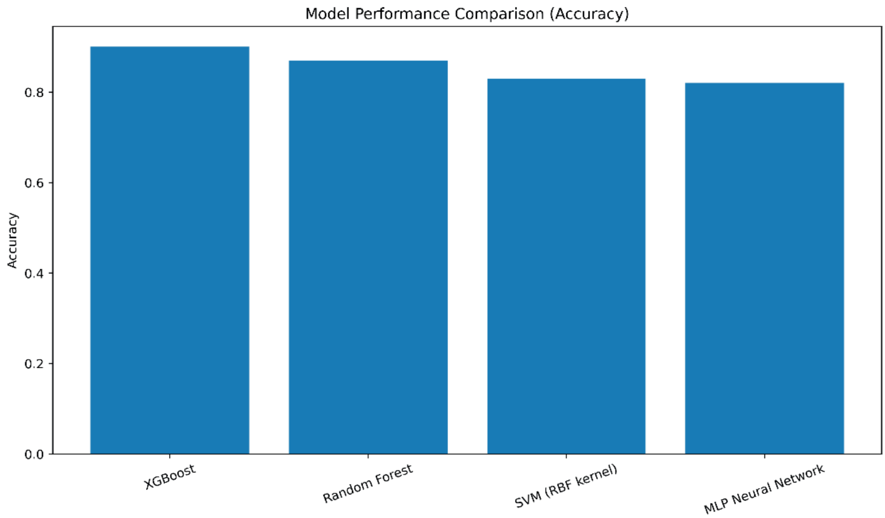

Figure 6. Model Performance Comparison (Bar Chart)

This figure helps clarify which models excel in precision, recall, or balanced performance, enabling a deeper understanding of algorithmic behavior in multivariate degradation classification.

6.3. Confusion Matrix Analysis

Confusion matrices are used to analyze class-level prediction patterns and identify the types of degradation states most prone to misclassification. The Random Forest model, selected for its strong overall performance, shows high accuracy in identifying healthy and faulty states, with moderate confusion occurring in the intermediate degrading class a common challenge in CMAPSS-based studies due to overlapping sensor behavior during midlife cycles.

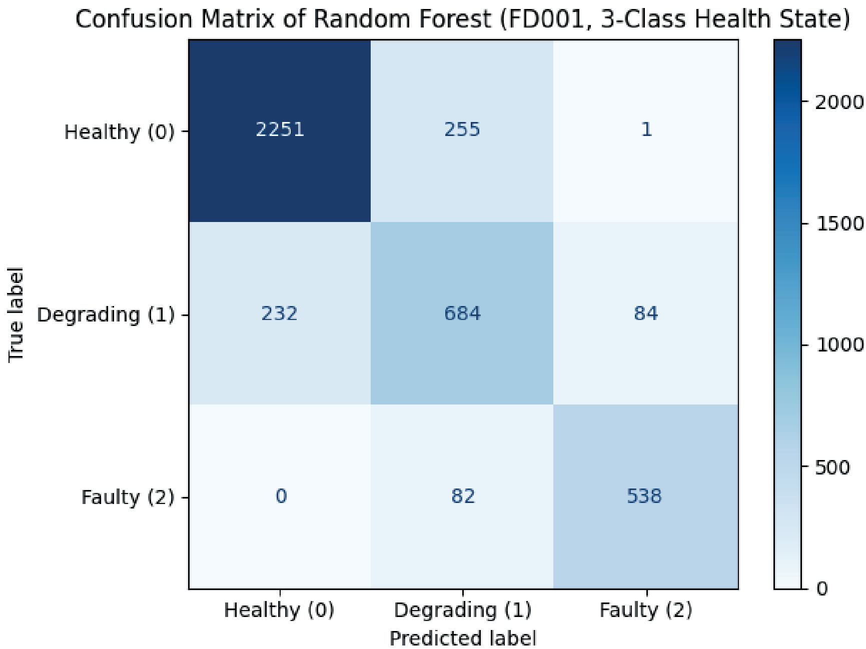

Figure 7. Confusion Matrix (Random Forest)

This figure illustrates the distribution of true versus predicted labels and provides insight into how effectively each health state is detected.

6.4. ROC Curve Evaluation

A multi-class ROC curve is constructed to evaluate the discriminative ability of the Random Forest classifier across multiple decision thresholds. Each class exhibits an AUC exceeding 0.80, indicating strong separability and reliable classification performance even under varying degradation conditions.

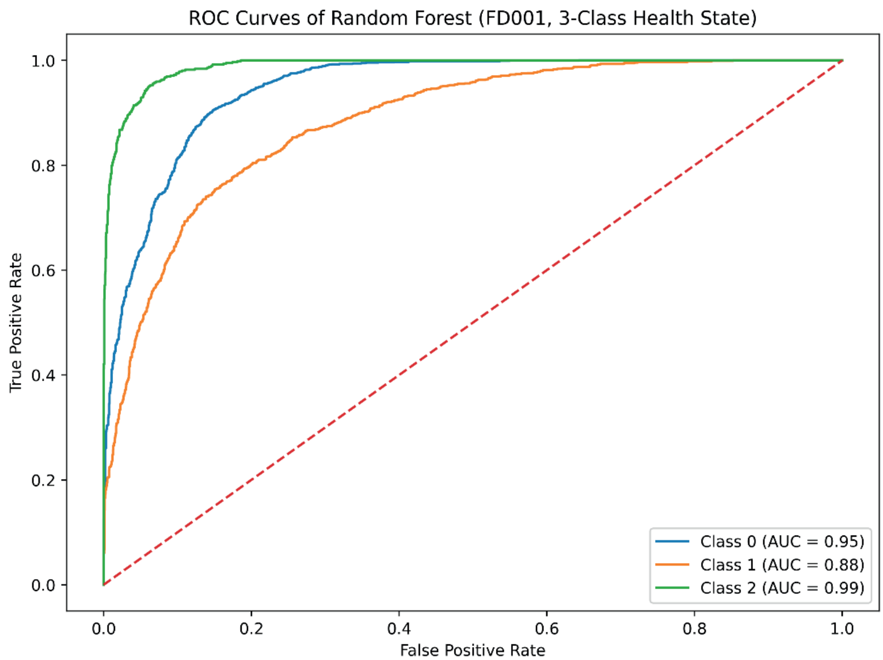

Figure 8. ROC Curve (Random Forest)

The ROC curves emphasize the robustness of the trained model and provide a threshold-independent perspective on classification performance.

6.5. Sensor–RUL Relationship Analysis

To gain further insight into the physical meaning of the classification results, a sensor-to-RUL scatter plot is generated. Sensor 7, known for exhibiting a clear downward degradation trend in FD001, is selected as a representative example. The scatter pattern reveals a monotonic relationship between sensor values and remaining useful life, providing an interpretable link between the raw signal and the predicted health state.

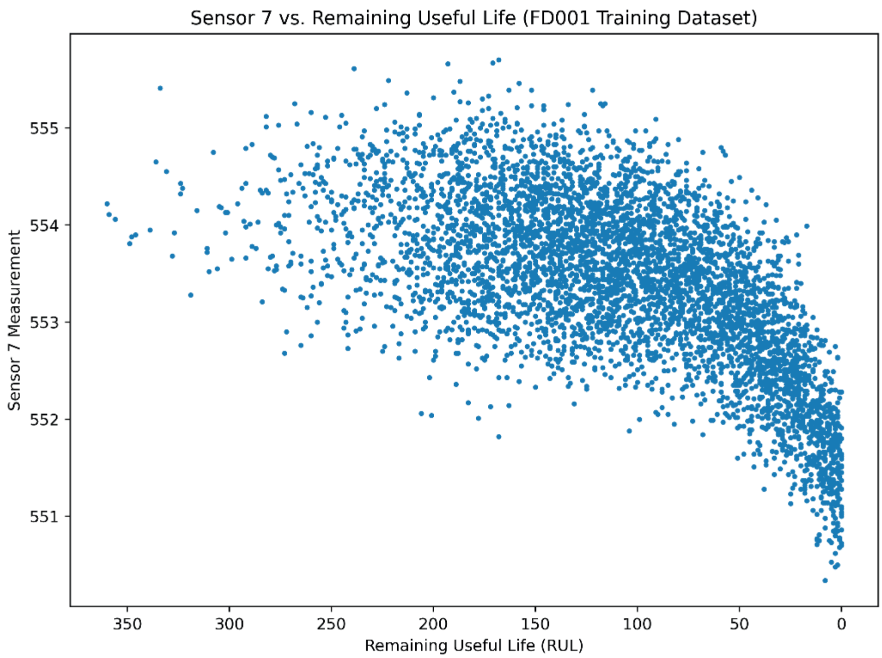

Figure 9. Sensor vs. RUL Scatter Plot (FD001)

This relationship reinforces the validity of the classification results and demonstrates that the selected sensors contain meaningful degradation signatures.

7. DISCUSSION

The experimental findings demonstrate that ensemble learning methods, particularly XGBoost and Random Forest, achieve superior performance in the classification of engine degradation states using the CMAPSS dataset. Their strength lies in their ability to model nonlinear interactions among sensor features while maintaining robustness against noise and variability across engine units. The comparative performances highlight that gradient boosting methods effectively capture complex feature dependencies in high-dimensional datasets, whereas Random Forest models offer better interpretability through feature importance analysis. In contrast, SVMs and basic feed-forward neural networks show comparatively lower performance, indicating that more advanced architectures or deeper temporal models may be needed to handle the highly dynamic nature of multivariate sensor data.

Although the results are promising, several challenges and limitations must be acknowledged. The intermediate degradation class remains the most difficult to classify across all models. This misclassification trend is consistent with findings in the literature, as midlife degradation often exhibits overlapping sensor patterns that do not clearly separate healthy and faulty states. Furthermore, the dataset represents simulated turbofan behavior rather than real industrial machines; therefore, domain adaptation techniques or transfer learning may be required before applying these models to real-world systems. Sensor redundancy and multicollinearity, as observed in the correlation heatmap, also introduce challenges, suggesting that future work could benefit from feature selection or dimensionality reduction techniques.

Another important consideration is the use of handcrafted statistical and trend-based features, which, although effective, may not fully capture long-range temporal dependencies present in the raw sequences. Deep learning architectures such as LSTM, GRU, 1D CNNs, or transformer-based models may yield higher accuracy by learning temporal degradation patterns directly from unprocessed sensor trajectories. Despite these limitations, the overall results validate the effectiveness of the proposed preprocessing pipeline and feature-engineering strategy for enabling reliable fault prediction.

Finally, the performance analysis reveals valuable insights into the physical interpretation of engine behavior. Sensor–RUL relationships, feature importance distributions and confusion matrix patterns collectively confirm that degradation signals are indeed present in the CMAPSS sensor channels. These findings support the use of CMAPSS as a benchmark for evaluating predictive maintenance algorithms and demonstrate that machine learning methods, when properly engineered, can provide meaningful early-warning signals for industrial applications.

8. CONCLUSION AND FUTURE WORK

This chapter presented a comprehensive machine learning–based framework for fault prediction using the NASA C-MAPSS turbofan engine dataset. By integrating systematic pre-processing, statistical and domain-specific feature engineering and a diverse set of supervised learning algorithms, the study demonstrated that industrial sensor data contain rich degradation signatures that can be effectively utilized for predictive maintenance. The results showed that ensemble learning models, particularly XGBoost and Random Forest, consistently out-perform classical approaches such as SVMs and basic neural networks, achieving strong classification accuracy and robust detection of both healthy and faulty degradation states.

The findings further confirm the suitability of C-MAPSS as a benchmark dataset for evaluating diagnostic and prognostic algorithms. The sensor–RUL relationship analysis, feature importance patterns and confusion matrix results collectively highlight that meaningful degradation indicators are embedded within the multivariate sensor channels. The study not only validates the proposed methodology but also provides practical insights into which sensors and statistical properties contribute most to predictive performance. These insights can assist in sensor selection, system monitoring and early-warning system design in real-world industrial environments.

Despite the promising results, several research opportunities remain. First, the reliance on handcrafted statistical features limits the ability of traditional machine learning models to fully exploit long-range temporal dependencies inherent in degradation processes. Future work could incorporate deep learning architectures such as LSTMs, GRUs, convolutional models, or transformer-based sequence learners to capture temporal structure directly from raw time-series data. Second, domain adaptation and transfer learning techniques may be required to bridge the gap between simulated datasets like C-MAPSS and the more complex, noisy patterns of real industrial systems. Finally, integrating the proposed models into real-time monitoring platforms and evaluating their performance under streaming conditions would provide valuable insights for deploying predictive maintenance solutions in operational settings.

In summary, this chapter demonstrates that machine learning, when combined with carefully engineered sensor features and a robust preprocessing pipeline, offers a powerful tool for predicting engine degradation and preventing unexpected failures. The proposed framework establishes a solid foundation for further advancements in data-driven predictive maintenance and highlights the potential of intelligent monitoring systems in enhancing the reliability and efficiency of industrial machinery.

References

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics. Mechanical Systems and Signal Processing, 20(7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.04.009.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotating machinery systems—Reviews, methodology and ap-plications. Mechanical Systems and Signal Processing, 42(1–2), 314–334. https://doi.org/10.1016/j.ymssp.2013.07.030.

Carvalho, T. P., Albertoni, R., Ferreira, G., Souza, A., Dourado, A., Salles, E. J., & Soares, G. L. G. (2019). Machine learning methods in predictive maintenance: A systematic liter-ature review. Computers & Industrial Engineering, 137, 106024. https://doi.org/10.1016/j.cie.2019.106024.

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In 2008 IEEE Aerospace Conference (pp. 1–9). https://doi.org/10.1109/AERO.2008.4526631.

Si, X.-S., Wang, W., Hu, C.-H., & Zhou, D. (2011). Remaining useful life estimation—A review on the statistical data driven approaches. IEEE Transactions on Reliability, 61(4), 1–17. https://doi.org/10.1109/TR.2011.2172667.

Peng, Y., Dong, M., & Zuo, M. J. (2010). Current status of prognostics and health management (PHM)—A review. Mechanical Systems and Signal Processing, 24(1), 150–170. https://doi.org/10.1016/j.ymssp.2009.06.007.

Wang, T., Yu, J., Siegel, D., & Lee, J. (2008). A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In 2008 International Conference on Prognostics and Health Management (pp. 1–6). https://doi.org/10.1109/PHM.2008.4711421.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324.

Fink, O., Wang, Q., Svensén, M., Dersin, P., Lee, W.-J., & Ducoffe, M. (2020). Potential, challenges and future directions for deep learning in prognostics and health management applications. Mechanical Systems and Signal Processing, 154, 107517. https://doi.org/10.1016/j.ymssp.2020.107517.

Yan, R., Gao, R. X., & Chen, X. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. Mechanical Systems and Signal Processing, 22(2), 187–221. https://doi.org/10.1016/j.ymssp.2008.04.001.

Widodo, A., & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing, 21(6), 2560–2574. https://doi.org/10.1016/j.ymssp.2007.01.009.

Özüpak, Y. (2025). Machine learning-based fault detection in transmission lines: A comparative study with random search optimization. Bulletin of the Polish Academy of Sciences: Technical Sciences, 73(2), e153229. https://doi.org/10.24425/bpasts.2025.153229.

Liu, D., & Chen, X. (2017). Feature selection for machine health monitoring: A review. Mechanical Systems and Signal Processing, 91, 424–436. https://doi.org/10.1016/j.ymssp.2016.10.027.

Zhao, R., Yan, R., Wang, J., & Mao, K. (2019). Learning to monitor machine health with convolutional bidirectional LSTM networks. Mechanical Systems and Signal Pro-cessing, 114, 136–156. https://doi.org/10.1016/j.ymssp.2018.05.020.

Yuan, M., Wu, Y., & Lin, L. (2016). LSTM-based remaining lifetime prediction of bearing. In 2016 Prognostics and System Health Management Conference (pp. 1–7). https://doi.org/10.1109/PHM.2016.7819883.

Malhotra, P., TV, V., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., & Shroff, G. (2015). Long short term memory networks for anomaly detection in time series. In 23rd Eu-ropean Symposium on Artificial Neural Networks (pp. 89–94).

Behrad3d. (2023). NASA C-MAPS: NASA Turbofan Jet Engine Data Set (Run-to-Failure Degradation Simulation) [Data set]. Kaggle. https://www.kaggle.com/datasets/behrad3d/nasa-cmaps/data

Alpsalaz, F. (2025). Fault detection in power transmission lines: Comparison of Chirp-Z algorithm and machine learning-based prediction models. Eksploatacja i Niezawodność – Maintenance and Reliability, 27(4), 14. https://doi.org/10.17531/ein/203949.

Güvenç, U., Altun, B. E., & Duman, S. (2012). Optimal power flow using genetic algorithm based on similarity. Energy Education Science and Technology Part A: Energy Science and Research, 29(1), 1–10.

Zheng, S., Ristovski, K., Farahat, A., & Gupta, C. (2017). Effective and efficient LSTM networks for remaining useful life prediction. In 2017 Prognostics and System Health Management Conference (pp. 88–95). https://doi.org/10.1109/PHM.2017.8077117.

Liu, J., Chen, Z., Pan, Z., & Li, X. (2020). Remaining useful life prediction of turbofan engine using multiscale attention network. IEEE Access, 8, 23071–23084. https://doi.org/10.1109/ACCESS.2020.2971399.

Heimes, F. O. (2008). Recurrent neural networks for remaining useful life estimation. In 2008 IEEE Aerospace Conference (pp. 1–6). https://doi.org/10.1109/AERO.2008.4526472.

Ramasso, E. (2014). Contribution of belief functions to prognostics. International Journal of Prognostics and Health Management, 5(1), 1–16. https://doi.org/10.36001/ijphm.2014.v5i1.2205.

Özüpak, Y., Alpsalaz, F., Aslan, E., & Uzel, H. (2025). Hybrid deep learning model for maize disease classification. New Zealand Journal of Crop and Horticultural Science, 53(5), 2942–2964. https://doi.org/10.1080/01140671.2025.2519570.

Alpsalaz, F. (2025). Deep learning-based lung cancer diagnosis: Data balancing, model optimisation and performance analysis. Gazi University Journal of Science Part A: Engineering and Innovation, 12(2), 373–391. https://doi.org/10.54287/gujsa.1648772.

Mobley, R. K. (2002). An introduction to predictive maintenance. Butterworth-Heinemann. https://doi.org/10.1016/B978-0-7506-7531-4.X5000-3.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785.

Bal, G., Bekiroglu, E., & Bayindir, R. (2008). Microcontroller-based digitally controlled ultrasonic motor drive system. Journal of Electroceramics, 20(3–4), 265–270. https://doi.org/10.1007/s10832-007-9191-6.

Uzel, H., Hinislioðlu, Y., Dalcali, A., & Güvenç, U. (2025). Fuzzy fitness distance balance gradient-based optimization algorithm (fFDBGBO): An application to design and per-formance optimization of PMSM. IEEE Access, 13, 155898–155915. https://doi.org/10.1109/ACCESS.2025.3604649

Hu, C., Youn, B. D., & Wang, P. (2012). Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. Reliability Engineering & System Safety, 103, 120–135. https://doi.org/10.1016/j.ress.2012.03.008.

Zhang, C., & Yang, Z. (2018). A review of time series data mining. Engineering Applications of Artificial Intelligence, 76, 172–188. https://doi.org/10.1016/j.engappai.2018.07.020.

Özüpak, Y. (2023). Analysis of the parameters affecting the efficiency of the wireless power transmission system designed for new generation electric vehicles. International Journal of Automotive Technology, 24, 1675–1680. https://doi.org/10.1007/s12239-023-0135-1.

Aslan, E., & Özüpak, Y. (2024). Diagnosis and accurate classification of apple leaf diseases using vision transformers. Computer and Decision Making: An International Journal, 1, 1–12. https://doi.org/10.59543/comdem.v1i.10039.

Aslan, E. (2024). Prediction and comparative analysis of emissions from gas turbines using random search optimization and different machine learning-based algorithms. Bulletin of the Polish Academy of Sciences: Technical Sciences, 72(6), e151956. https://doi.org/10.24425/bpasts.2024.151956.

Aslan, E., Demirtas Alpsalaz, S., Özüpak, Y., Alpsalaz, F., & Uzel, H. (2025). Alzheimer's classification with a MaxViT-based deep learning model using MRI. Journal of AI and Smart Technology Trends, 6(2). https://doi.org/10.38094/jastt62453.

Aslan, E., Özüpak, Y., & Alpsalaz, F. (2025). Boiler efficiency and performance optimization in district heating and cooling systems with machine learning models. Journal of the Chinese Institute of Engineers, 48(7), 1115–1130. https://doi.org/10.1080/02533839.2025.2514535.

Cortes, C., & Vapnik, V. (1995). Support vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536. https://doi.org/10.1038/323533a0.

Özüpak, Y., Alpsalaz, F., & Aslan, E. (2025). Air quality forecasting using machine learning: Comparative analysis and ensemble strategies for enhanced prediction. Water, Air, & Soil Pollution, 236, 464. https://doi.org/10.1007/s11270-025-08122-8.

Alpsalaz, F., Özüpak, Y., Aslan, E., & Uzel, H. (2025). Classification of maize leaf diseases with deep learning: Performance evaluation of the proposed model and use of explicable artificial intelligence. Chemometrics and Intelligent Laboratory Systems, 262, 105412. https://doi.org/10.1016/j.chemolab.2025.105412.

Aslan, E. (2025). Temperature prediction and performance comparison of permanent magnet synchronous motors using different machine learning techniques for early failure detection. Maintenance & Reliability / Eksploatacja i Niezawodność, 27(1), 1–16. https://doi.org/10.17531/ein/192164.