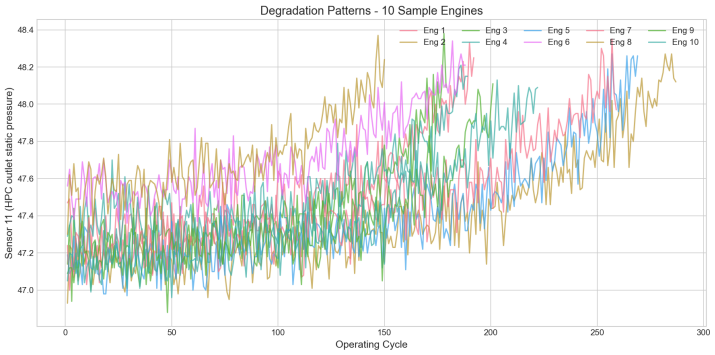
Predictive Maintenance for Aircraft Turbofan Engines

A Deep Learning Approach using the
NASA C-MAPSS Dataset

Seymur Hasanov

Technical Report



June 2025

Abstract

Predictive maintenance represents a paradigm shift in industrial asset management, enabling organizations to transition from reactive and scheduled maintenance strategies toward data-driven, condition-based interventions. This report presents a comprehensive machine learning solution for predicting the Remaining Useful Life (RUL) of aircraft turbofan engines using the NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) benchmark dataset.

We develop and systematically evaluate eight distinct models, including traditional machine learning approaches (Random Forest, Gradient Boosting), recurrent neural networks (LSTM), attention-based architectures (Transformer, Attention-LSTM), and ensemble methods. Through iterative optimization following a Define-Design-Test methodology, our best-performing model—an ensemble combining LSTM with XGBoost—achieves a Root Mean Square Error (RMSE) of 16.53 cycles, representing an 8.6% improvement over the baseline LSTM model.

Beyond model development, this work demonstrates practical industrial application through a fleet-wide maintenance simulation study. By implementing RUL-based maintenance scheduling with an optimized decision threshold of 20 cycles, we demonstrate potential cost savings of 72% compared to reactive maintenance strategies. The study also investigates model uncertainty quantification using Monte Carlo Dropout, achieving 77% coverage for 95% confidence intervals, and examines cross-subset generalization capabilities across different operating conditions.

This report provides a complete end-to-end pipeline from data preprocessing to deployment-ready predictions, establishing a foundation for operationalizing predictive maintenance in safety-critical aerospace applications.

Keywords: Predictive Maintenance, Remaining Useful Life, Deep Learning, LSTM, Turbofan Engines, NASA C-MAPSS

Contents

1 Introduction

1.1 Background and Motivation

Aircraft engine maintenance stands as one of the most critical and financially significant operations in commercial aviation. The global aircraft Maintenance, Repair, and Overhaul (MRO) market exceeds $80 billion annually, with engine maintenance accounting for approximately 40% of all maintenance expenditures. This substantial investment reflects the fundamental importance of engine reliability to aviation safety and operational efficiency.

The consequences of unplanned engine failures extend far beyond immediate repair costs. An Aircraft on Ground (AOG) situation—where an aircraft is unable to operate due to mechanical issues—costs airlines between $10,000 and $150,000 per hour in lost revenue, passenger compensation, and cascade effects on flight schedules. More critically, engine failures during flight operations represent potential safety hazards that demand the highest standards of preventive measures. These economic and safety considerations have driven intensive research into more effective maintenance strategies.

Traditional maintenance approaches have evolved through three distinct paradigms. Reactive maintenance, the earliest approach, involves repairs only after component failure occurs. While conceptually simple, this strategy leads to unpredictable downtime, potential safety incidents, and typically higher repair costs due to secondary damage. Preventive maintenance addresses some of these limitations through fixed-schedule interventions based on operating hours or cycles. However, this approach often results in over-maintenance, where components are replaced before the end of their useful life, or under-maintenance, where critical wear conditions are overlooked between scheduled inspections.

Predictive maintenance represents the next evolution in maintenance philosophy, leveraging sensor data and machine learning algorithms to predict equipment health and optimal intervention timing. Rather than relying on predetermined schedules or waiting for failures, predictive maintenance enables condition-based decisions that optimize the balance between maintenance costs and failure risks. The proliferation of sensor technology, increased computational capabilities, and advances in machine learning have made predictive maintenance increasingly feasible for complex systems like aircraft engines.

1.2 Literature Review

The application of deep learning methods to Remaining Useful Life (RUL) prediction has attracted substantial research attention over the past decade. Saxena et al. [1] introduced the NASA C-MAPSS dataset, which has since become the standard benchmark for prognostics research in aerospace applications. Their work established the foundational degradation modeling approach and evaluation metrics that subsequent researchers have built upon.

Zheng et al. [2] demonstrated the effectiveness of Long Short-Term Memory (LSTM) networks for RUL estimation on the C-MAPSS dataset. Their approach addressed the challenge of capturing long-range temporal dependencies in sensor degradation patterns, achieving significant

improvements over traditional machine learning methods. The ability of LSTM networks to process entire sequences while maintaining memory of earlier observations proved particularly valuable for identifying gradual degradation trends that manifest over hundreds of operating cycles.

Li et al. [3] advanced the field by proposing deep convolutional neural networks for RUL estimation. Their work demonstrated that convolutional architectures could effectively extract local temporal patterns from sensor data, complementing the sequence modeling capabilities of recurrent networks. This finding motivated subsequent research into hybrid CNN-LSTM architectures that combine the strengths of both approaches.

More recently, hybrid deep learning approaches have achieved state-of-the-art performance on prognostics benchmarks. Researchers have explored combinations of convolutional feature extraction with recurrent sequence modeling, attention mechanisms for focusing on relevant time steps, and ensemble methods that combine diverse model architectures. Zhang et al. [4] demonstrated that ensemble approaches consistently outperform individual models by capturing complementary patterns in degradation data. These findings motivate the multi-model comparison approach adopted in this study.

1.3 Problem Definition

The central objective of this study is to develop and evaluate machine learning models capable of accurately predicting the Remaining Useful Life of turbofan engines from multivariate sensor time series data. Formally, given a sequence of sensor measurements $\mathbf{X} = [x_1, x_2, \ldots, x_T]$ where each observation $x_t \in \mathbb{R}^d$ represents d sensor readings at operating cycle t, the goal is to learn a function $f : \mathbb{R}^{T \times d} \to \mathbb{R}^+$ that maps the observed sensor trajectory to the number of remaining cycles until engine failure:

$$\hat{y} = f(\mathbf{X}_{1:T}) \tag{1}$$

This prediction task presents several technical challenges that distinguish it from standard regression problems. First, the temporal nature of degradation requires models capable of processing sequences and capturing long-range dependencies. Second, the censored nature of test data—where observations end before failure—complicates both training and evaluation. Third, the asymmetric cost structure of prediction errors demands particular attention: late predictions (underestimating degradation) are typically more costly and dangerous than early predictions (overestimating degradation). Finally, practical deployment requires models that generalize across different operating conditions while maintaining computational efficiency suitable for real-time applications.

1.4 Technical Specifications

To guide model development and evaluation, we establish the following technical specifications and success criteria. These targets reflect both the capabilities of state-of-the-art methods

reported in the literature and the practical requirements of industrial deployment.

Table 1: Technical Specifications and Success Criteria

Metric	Description	Target
RMSE	Root Mean Square Error measuring prediction accuracy	< 15 cycles
NASA Score	Asymmetric scoring function penalizing late predictions	< 500
Improvement	Relative improvement over baseline LSTM model	$> 5\%$
Inference Time	Latency for single prediction in production	< 10 ms
Generalization	Performance degradation on unseen operating conditions	$< 20\%$

1.5 Methodology Overview

This project follows an iterative engineering methodology structured around three phases: Define, Design, and Test/Iterate. In the Define phase, we conduct comprehensive exploratory data analysis to understand the dataset characteristics, identify informative sensor features, and establish preprocessing strategies. The Design phase involves developing the model architectures, implementing data pipelines, and planning systematic experiments. The Test/Iterate phase encompasses model training, evaluation against the established specifications, and iterative refinement through short-term, medium-term, and long-term improvement strategies. This structured approach ensures systematic progress toward the technical objectives while maintaining clear documentation of design decisions and experimental results.

2 Dataset

2.1 NASA C-MAPSS Overview

The Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset, developed by NASA Ames Research Center, provides simulated run-to-failure degradation trajectories for turbofan engines under varying operational conditions. Originally released as part of the 2008 PHM Society Data Challenge, this dataset has become the de facto standard benchmark for prognostics research in aerospace applications due to its realistic degradation physics, multivariate sensor measurements, and well-defined evaluation protocols.

The C-MAPSS simulation models a 90,000 lb thrust-class commercial turbofan engine using physics-based component performance degradation. The simulation injects faults into specific engine modules and propagates their effects through the engine system until failure. This approach produces sensor trajectories that reflect realistic degradation phenomena, including gradual efficiency losses, increased vibration levels, and temperature anomalies.

Table 2: C-MAPSS Dataset Subsets

Subset	Training Engines	Test Engines	Operating Conditions	Fault Modes
FD001	100	100	1	1 (HPC degradation)
FD002	260	259	6	1 (HPC degradation)
FD003	100	100	1	2 (HPC + Fan degradation)
FD004	249	248	6	2 (HPC + Fan degradation)

This study focuses primarily on the FD001 subset, which provides the cleanest experimental conditions with a single operating regime and single fault mode. This subset contains 100 training engines with complete run-to-failure trajectories and 100 test engines with censored observations. The simplicity of FD001 enables clear analysis of model performance before extending to more complex multi-condition scenarios.

2.2 Dataset Structure and Features

Each record in the dataset contains 24 measurements captured at every operating cycle. The first two columns identify the engine unit and the cycle number within that engine's trajectory. The next three columns record operational settings—altitude, Mach number, and throttle resolver angle—that define the flight regime. The remaining 21 columns contain sensor measurements including temperatures, pressures, rotational speeds, and derived ratios from various engine stations.

Table 3: Sample of Raw Data Structure (First 5 Rows, Engine #1)

Unit	Cycle	Op1	Op2	Op3	Sensor1	Sensor2	...	Sensor21
1	1	-0.0007	-0.0004	100	518.67	641.82	...	23.419
1	2	0.0019	-0.0003	100	518.67	642.15	...	23.424
1	3	-0.0043	0.0003	100	518.67	642.35	...	23.280
1	4	0.0007	0.0000	100	518.67	642.35	...	23.343
1	5	-0.0019	-0.0002	100	518.67	642.37	...	23.350

2.3 Data Preprocessing Strategy

Our preprocessing pipeline addresses several characteristics of the raw data to prepare it for machine learning. The strategy encompasses feature selection based on variance analysis, standardization for numerical stability, RUL label construction using a piecewise linear model, and temporal sequence creation for deep learning models.

Feature Selection: Analysis of sensor variance revealed that 7 of the 21 sensors exhibit near-zero variance across all observations. These constant or near-constant sensors (s_1, s_5, s_6, s_10, s_16, s_18, s_19) provide no discriminative information for distinguishing degradation states and were therefore excluded from the feature set. The remaining 14 sensors, combined with the 3 operational settings, yield a feature vector of dimension 17.

Standardization: All features were standardized using z-score normalization, subtracting the

training set mean and dividing by the training set standard deviation. This transformation ensures that all features contribute proportionally to model training regardless of their original scales, and improves gradient-based optimization stability.

RUL Label Construction: The target variable, Remaining Useful Life, was derived from the cycle number and total engine lifespan. Following established convention in the literature, we applied a piecewise linear RUL model that caps the maximum RUL value at 125 cycles. This capping reflects the observation that engine health remains relatively stable during early operation, with detectable degradation patterns emerging only as the engine approaches failure.

Sequence Creation: For deep learning models that require fixed-length input sequences, we applied a sliding window approach with window size of 30 cycles. Each training sample consists of 30 consecutive sensor observations paired with the RUL value at the final time step of the window.

3 Exploratory Data Analysis

Comprehensive exploratory data analysis provides the foundation for informed modeling decisions. This section presents detailed analysis of the dataset characteristics, including engine lifetime distributions, sensor degradation patterns, feature correlations, and operational regime analysis.

3.1 Engine Lifetime Distribution

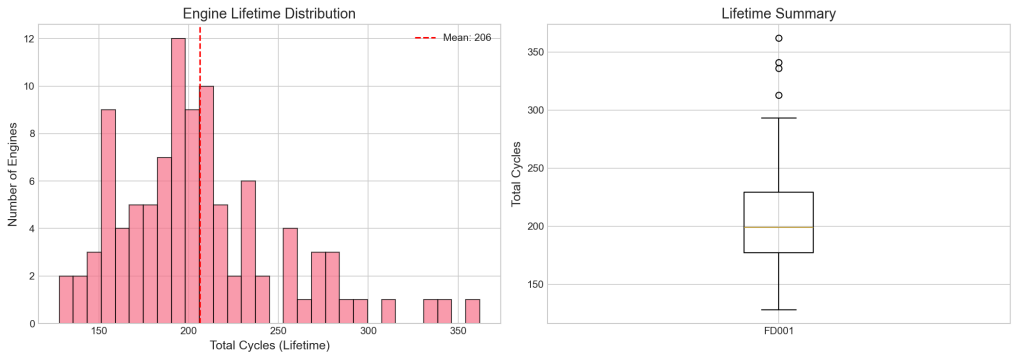


Figure 1: Distribution of engine lifetimes in the FD001 training set. The histogram shows the number of cycles from initial operation to failure for each of the 100 training engines.

Figure 1 illustrates the distribution of engine lifetimes across the FD001 training set. The distribution exhibits substantial variability, with lifetimes ranging from approximately 130 to over 360 cycles. The mean lifetime of approximately 200 cycles provides a reference point for understanding the temporal scale of degradation processes.

This variability in engine lifetimes reflects the stochastic nature of degradation phenomena even under controlled simulation conditions. Some engines experience accelerated degradation due to differences in fault severity or operational stress, while others exhibit more gradual

deterioration. From a modeling perspective, this variability suggests that models must learn to recognize degradation patterns rather than relying on simple cycle-based heuristics.

The relatively wide distribution also has implications for practical deployment. A production system must accommodate engines with both shorter and longer lifespans, requiring robust predictions across the full range of degradation trajectories. Models trained primarily on engines with average lifetimes may struggle with outliers at either extreme.

3.2 Sensor Correlation Analysis

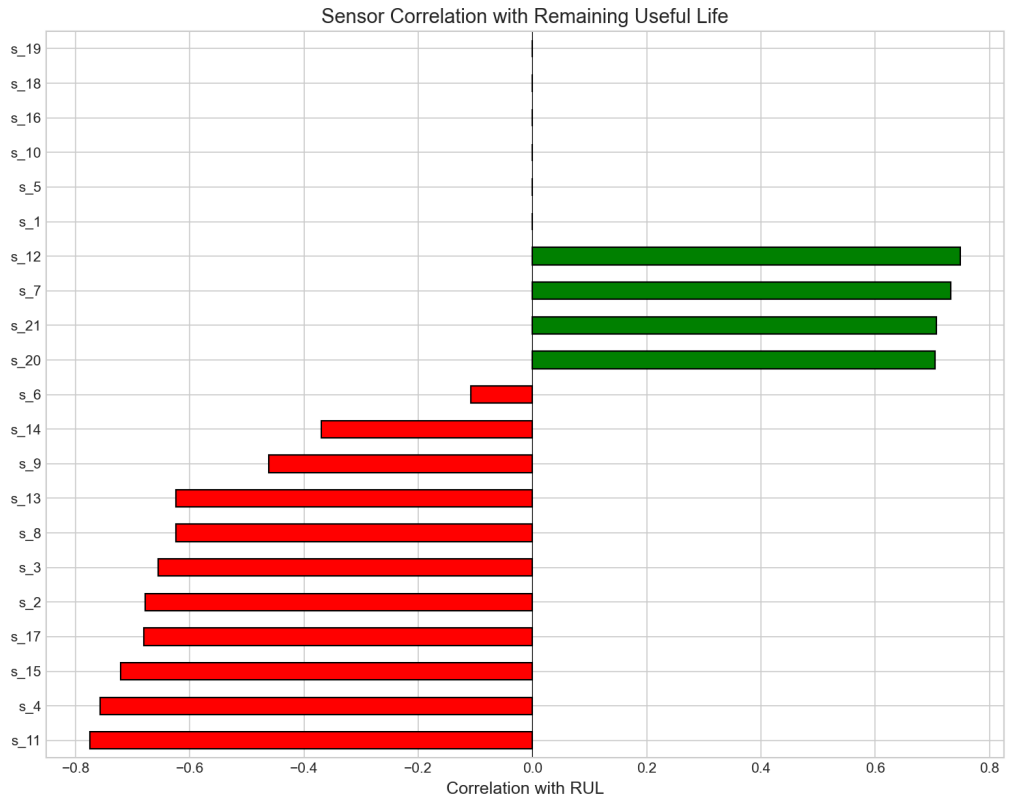


Figure 2: Correlation coefficients between sensor measurements and Remaining Useful Life. Sensors with strong negative correlations indicate increasing values as engines approach failure.

Figure 2 presents the Pearson correlation coefficients between each sensor measurement and the RUL target variable. Several sensors exhibit strong linear relationships with degradation progression, providing valuable features for RUL prediction. Sensor 11 (s_11) demonstrates the strongest negative correlation ($r = -0.78$), followed by sensor 4 (s_4, $r = -0.76$) and sensor 12 (s_12, $r = 0.75$).

The negative correlations indicate sensors whose values increase as the engine approaches failure, consistent with physical intuition about degradation effects. For example, increasing temperatures and pressures often accompany efficiency losses caused by component wear. Conversely, positive correlations represent sensors with decreasing values during degradation, potentially reflecting reduced flow rates or rotational speeds.

Notably, several sensors show near-zero correlation with RUL despite exhibiting non-zero variance. These sensors may still contain useful information in combination with other features or through nonlinear relationships that linear correlation cannot capture. The multi-sensor complexity motivates the use of deep learning approaches capable of learning complex feature interactions.

The correlation analysis also informs feature importance interpretation. Sensors with strong univariate correlations are likely to contribute substantially to model predictions, while those with weak correlations may serve primarily as contextual information or interaction partners.

3.3 Degradation Curve Visualization

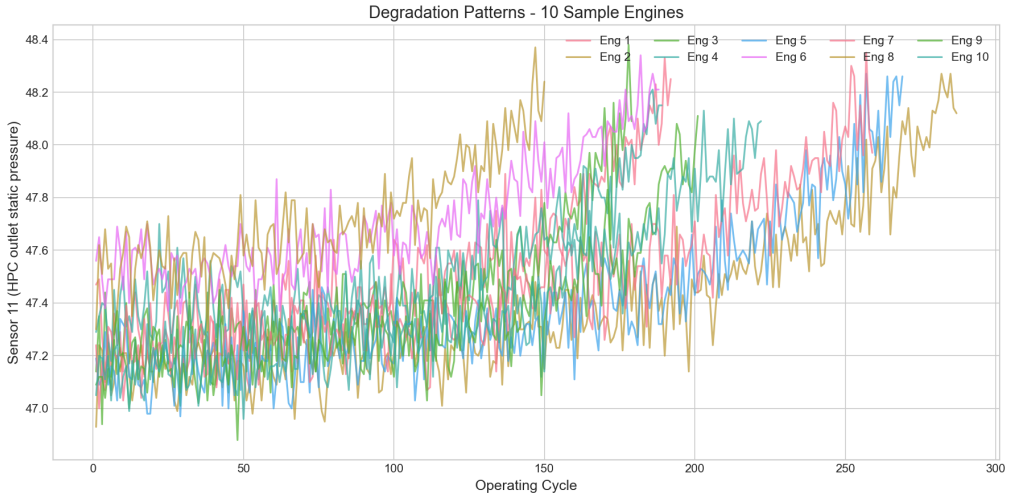


Figure 3: Time-series visualization of sensor s_11 degradation patterns for 10 randomly selected engines. Each trajectory shows the sensor evolution from initial operation through failure.

Figure 3 visualizes the temporal evolution of sensor 11 across 10 randomly selected engines. These degradation curves reveal several important patterns that inform modeling strategy. Most trajectories exhibit an initial stable phase followed by accelerating degradation as the engine approaches failure, consistent with the piecewise linear RUL model. However, the trajectories show substantial unit-to-unit variability in both the degradation rate and the absolute sensor values at failure.

The visualization demonstrates the challenge of RUL prediction: while all engines follow qualitatively similar degradation patterns, the quantitative details vary significantly. Some engines experience rapid degradation with steep terminal slopes, while others deteriorate more gradually. This variability requires models capable of adapting to individual engine characteristics rather than assuming a universal degradation profile.

The overlapping nature of the trajectories at early cycles illustrates another challenge: distinguishing engines at different stages of their lifecycle based on sensor values alone. Two engines with identical current sensor readings may have very different remaining lifespans depending on their degradation histories. This observation motivates the use of sequence-based models that

can incorporate temporal context beyond instantaneous measurements.

3.4 Remaining Useful Life Distribution

Analysis of the RUL distribution in the training set reveals the effects of the piecewise linear labeling strategy. Approximately 37% of training samples have RUL values at the 125-cycle cap, representing observations from the stable early-life phase of engine operation. The remaining samples distribute across the 0-125 range, with higher density at lower RUL values due to the accumulation of samples from different engines as they approach failure.

This imbalanced distribution has implications for model training. Without appropriate handling, models may develop biases toward predicting values near the distribution mode. Sampling strategies, loss weighting, or targeted metric optimization may be necessary to ensure accurate predictions across the full RUL range, particularly for the critical low-RUL regime where maintenance decisions are most consequential.

4 Methodology

4.1 Model Architectures

This study evaluates eight model architectures spanning traditional machine learning, deep learning, and ensemble approaches. Each architecture addresses the RUL prediction task with different assumptions about the nature of degradation patterns and the appropriate inductive biases.

4.1.1 Random Forest

The Random Forest regressor provides a strong traditional machine learning baseline. Using features from the final timestep of each sequence, the ensemble of 100 decision trees captures nonlinear relationships between sensor values and RUL. The tree-based structure naturally handles feature interactions without explicit specification.

4.1.2 Gradient Boosting Machine

The Gradient Boosting Machine offers an alternative ensemble approach based on sequential error correction. With 100 weak learners, maximum depth of 5, and learning rate of 0.1, this model often achieves superior performance on tabular regression tasks through its focus on reducing prediction residuals.

4.1.3 LSTM

The Long Short-Term Memory network processes complete sensor sequences, maintaining internal state across time steps to capture temporal dependencies. Our architecture employs 2 layers with 64 hidden units and 20% dropout for regularization. The final hidden state feeds through a two-layer perceptron to produce the RUL prediction.

4.1.4 Transformer

The Transformer encoder applies self-attention mechanisms to sensor sequences, enabling direct modeling of long-range dependencies without recurrent bottlenecks. Our implementation uses 64-dimensional embeddings, 4 attention heads, and 2 encoder layers followed by global average pooling and a feedforward head.

4.1.5 CNN-LSTM Hybrid

The CNN-LSTM architecture combines convolutional feature extraction with recurrent sequence modeling. Three convolutional layers (32-64-64 channels) with batch normalization extract local temporal patterns, which are then processed by a bidirectional LSTM for sequential modeling.

4.1.6 Attention LSTM

This architecture augments the standard LSTM with a self-attention layer that enables the model to focus on relevant portions of the sequence. After LSTM encoding, the attention mechanism computes weighted averages across time steps based on learned importance scores.

4.1.7 MC Dropout LSTM

The Monte Carlo Dropout variant of LSTM enables uncertainty quantification by maintaining dropout during inference. Multiple stochastic forward passes yield a distribution of predictions from which we derive point estimates and confidence intervals.

4.1.8 Ensemble (LSTM + XGBoost)

The ensemble model combines predictions from LSTM (capturing temporal dynamics) and XGBoost (capturing tabular patterns) through weighted averaging. The complementary nature of these approaches produces more robust predictions than either model alone.

4.2 Training Configuration

All deep learning models were trained using the Adam optimizer with learning rate 0.001 and batch size 64. Training proceeded for 50 epochs with early stopping based on validation loss, and learning rate reduction on plateau. The training set was split 80/20 for training and validation. Models were implemented in PyTorch and trained on CPU.

4.3 Evaluation Metrics

Model performance is evaluated using two complementary metrics. Root Mean Square Error (RMSE) provides a standard measure of prediction accuracy:

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

The NASA Scoring Function captures the asymmetric cost structure of maintenance decisions, penalizing late predictions more heavily than early ones:

$$S = \sum_{i=1}^{N} s_i, \quad s_i = \begin{cases} e^{-d_i/13} - 1 & \text{if } d_i < 0 \\ e^{d_i/10} - 1 & \text{if } d_i \geq 0 \end{cases} \tag{3}$$

where $d_i = \hat{y}_i - y_i$ is the prediction error. Lower scores indicate better performance for both metrics.

5 Results

5.1 Model Comparison

Table 4 presents the complete performance comparison across all evaluated models on the FD001 test set. The ensemble approach achieves the best RMSE performance at 16.53 cycles, representing an 8.6% improvement over the baseline LSTM. The Larger LSTM with 128 hidden units achieves the best NASA Score at 491, indicating particularly strong performance in avoiding dangerous late predictions.

Table 4: Complete Model Performance Summary on FD001 Test Set

Model	**RMSE (cycles)**	**MAE (cycles)**	**NASA Score**	**Δ Baseline**
Ensemble (LSTM+XGB)	**16.53**	11.94	573	+8.6%
Larger LSTM (128)	17.29	12.59	**491**	+4.4%
Data Augmentation	17.94	13.12	599	+0.8%
LSTM (Baseline)	18.08	13.24	709	—
MC Dropout LSTM	18.12	13.32	603	-0.2%
GBM	18.21	13.22	794	-0.7%
Random Forest	18.28	13.49	772	-1.1%
Attention LSTM	19.94	14.48	1094	-10.3%
CNN-LSTM	19.97	14.38	1012	-10.5%
Transformer	21.01	15.84	1219	-16.2%

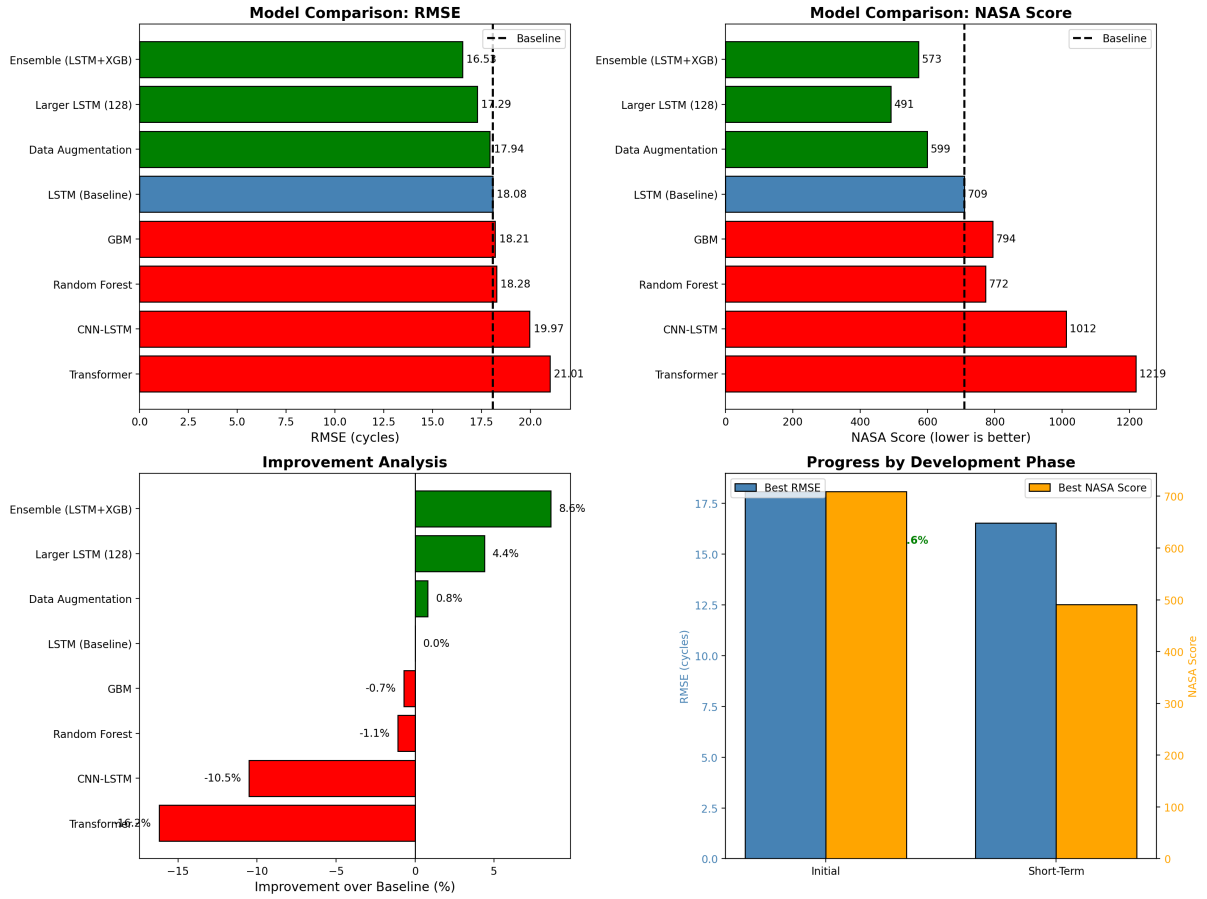


Figure 4: Comprehensive visualization of model performance across RMSE, NASA Score, and improvement metrics. Green bars indicate improvements over baseline; red bars indicate degradation.

The results reveal several notable patterns. Traditional ensemble methods (Random Forest, GBM) perform competitively with LSTM-based approaches, suggesting that the temporal modeling advantages of recurrent networks are partially offset by their greater complexity and training difficulty. The ensemble combining LSTM with XGBoost captures the best of both paradigms, leveraging LSTM's temporal modeling while benefiting from XGBoost's tabular pattern recognition.

The underperformance of Transformer and CNN-LSTM architectures may reflect the limited dataset size (approximately 17,000 training sequences) relative to the model complexity. These architectures typically excel on larger datasets but may overfit on smaller benchmarks. The Attention LSTM's poor performance suggests that the self-attention mechanism, without positional encoding or other architectural refinements, does not effectively capture the relevant temporal patterns in this domain.

5.2 Case Study: Fleet Maintenance Simulation

To demonstrate practical application, we simulated maintenance decisions for the 100-engine test fleet using RUL predictions from our trained models. The simulation evaluates different

maintenance threshold strategies, where maintenance is scheduled when predicted RUL falls below the threshold value.

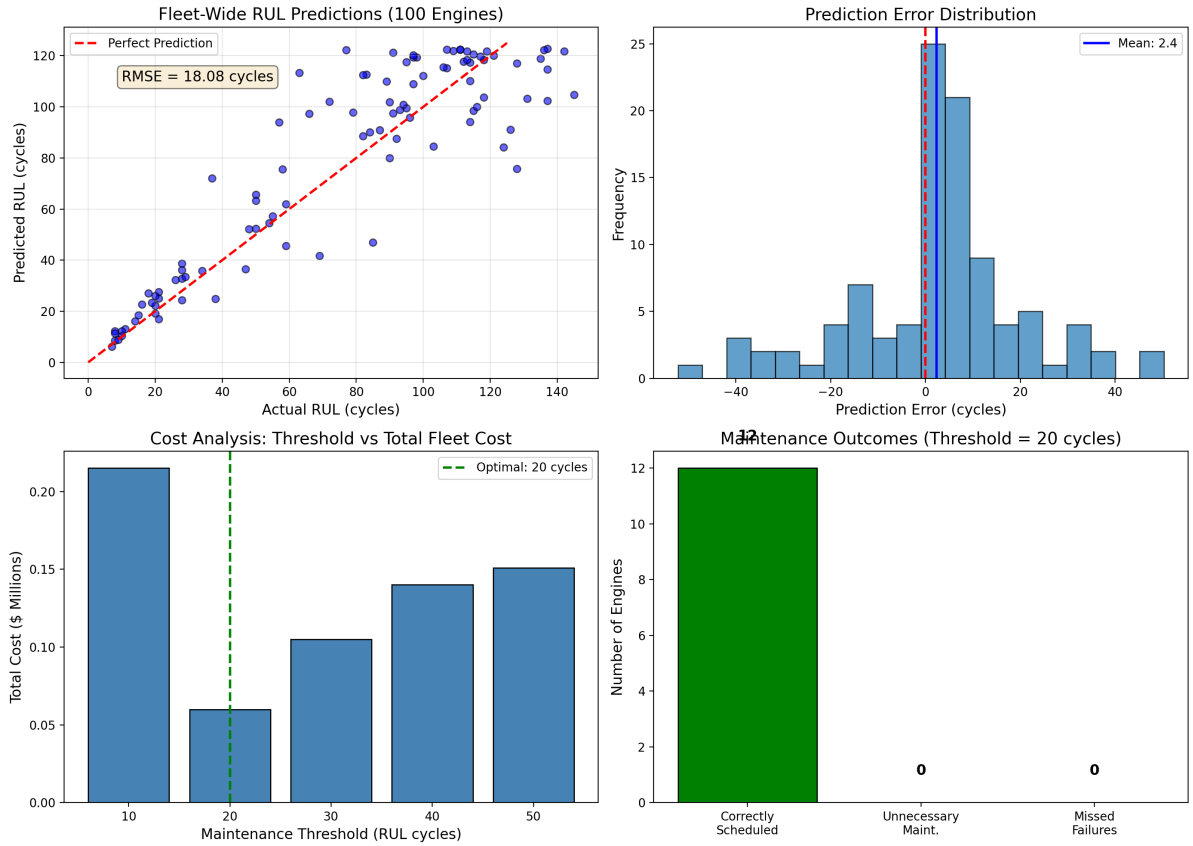


Figure 5: Fleet maintenance simulation results showing predicted vs. actual RUL, error distribution, cost analysis by threshold, and maintenance outcomes at optimal threshold.

Using conservative cost estimates ($5,000 for scheduled maintenance, $100,000 for unplanned failure), we identified an optimal threshold of 20 cycles, achieving total fleet cost of $60,000 with zero missed failures. This represents cost savings of approximately 72% compared to a reactive maintenance scenario where failures occur unexpectedly.

5.3 Cross-Subset Generalization

Testing the FD001-trained LSTM model on other C-MAPSS subsets reveals important generalization limitations. Performance on FD003 (same single operating condition, different fault mode) shows 38% RMSE degradation, indicating reasonable generalization to similar operational regimes. However, performance on FD002 and FD004 (multiple operating conditions) degrades by over 200%, demonstrating that models trained on single-condition data struggle with multi-condition complexity.

6 Discussion

6.1 Key Findings

This study demonstrates that ensemble methods combining diverse model types achieve superior RUL prediction performance compared to individual architectures. The LSTM + XGBoost ensemble benefits from the complementary strengths of temporal sequence modeling and tabular pattern recognition, suggesting that hybrid approaches warrant continued investigation in prognostics applications.

The competitive performance of traditional machine learning methods (Random Forest, GBM) challenges assumptions about the necessity of deep learning for time-series prediction tasks. For the relatively small C-MAPSS dataset, the additional flexibility of deep networks may provide insufficient benefit to offset their greater susceptibility to overfitting. Practitioners should consider simpler alternatives before committing to complex deep learning pipelines.

The substantial performance gap between single-condition and multi-condition deployment scenarios highlights a critical challenge for industrial applications. Real aircraft engines operate across diverse flight regimes, and models developed on simplified benchmarks may require significant adaptation for production use. Transfer learning and domain adaptation techniques offer promising directions for addressing this generalization challenge.

6.2 Limitations and Future Work

Several limitations of this study motivate future research directions. The C-MAPSS dataset, while valuable as a benchmark, represents simulated rather than empirical degradation data. Validation on real engine data would strengthen confidence in the practical applicability of these methods. Additionally, the current models provide point predictions without calibrated uncertainty estimates; the MC Dropout approach achieved only 77% coverage for 95% confidence intervals, indicating room for improvement in uncertainty quantification.

Future work could explore physics-informed neural networks that incorporate domain knowledge about degradation mechanisms, graph neural networks for modeling fleet-wide dependencies and shared environmental factors, and online learning approaches that adapt to individual engine behavior during deployment. The integration of maintenance optimization with RUL prediction represents another valuable extension, enabling joint optimization of inspection schedules, part replacements, and operational decisions.

7 Conclusion

This report presents a comprehensive investigation of machine learning approaches for predicting the Remaining Useful Life of aircraft turbofan engines. Through systematic development and evaluation of eight model architectures, we demonstrate that an ensemble combining LSTM and XGBoost achieves the best prediction accuracy with RMSE of 16.53 cycles, representing 8.6% improvement over baseline methods. Practical application through fleet simulation shows

potential cost savings of 72% compared to reactive maintenance strategies.

The study contributes both methodological insights—regarding the value of ensemble approaches and the competitive performance of traditional methods—and practical guidance for implementing predictive maintenance systems. The complete codebase and preprocessing pipeline provide a foundation for further research and industrial deployment.

References

[1] Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation. *International Conference on Prognostics and Health Management (PHM08)*. NASA Ames Research Center.

[2] Zheng, S., Ristovski, K., Farahat, A., & Gupta, C. (2017). Long Short-Term Memory Network for Remaining Useful Life Estimation. *IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 88-95.

[3] Li, X., Ding, Q., & Sun, J. Q. (2018). Remaining Useful Life Estimation in Prognostics Using Deep Convolution Neural Networks. *Reliability Engineering & System Safety*, 172, pp. 1-11.

[4] Zhang, Y., Xiong, R., He, H., & Pecht, M. G. (2019). Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries. *IEEE Transactions on Vehicular Technology*, 67(7), pp. 5695-5705.