Milestone 3 & 4 — Model Evaluation and RUL Labelling

Date: October 27, 2025

Author: Chivukula Sai Rithivik

Milestone 3 — Model Evaluation and Performance Analysis

This section summarizes the model evaluation results for the LSTM-based Remaining Useful Life (RUL) prediction system, applied to all four subsets (FD001–FD004) of the NASA CMAPSS dataset. The goal was to analyse model accuracy, residual patterns, and prediction bias using consistent evaluation metrics and visualization outputs

Evaluation Pipeline Each dataset followed an identical evaluation flow:

1. Load test sequences (`.npz`) and trained LSTM model (`.h5`).
2. Predict RUL for each test sample.
3. Compute RMSE, MAE, and R^2 metrics.
4. Generate plots showing predicted vs actual RUL, residual distribution, and residual trends.
5. Save outputs and bias analysis reports for all datasets (FD001–FD004)

Key Findings

The model achieved strong overall performance, with R^2 scores of **0.78** and **0.86**, showing high predictive accuracy.

Predicted vs Actual plots demonstrated a close match, with only slight deviations observed as the system approached the end of its life cycle.

In the **mixed-condition datasets (FD002 and FD004)**, prediction variability was slightly higher — likely due to the more diverse operating environments.

Overall, the evaluation confirmed that the **LSTM model effectively captured temporal patterns and tracked equipment degradation** with high reliability.

Milestone 4 — RUL Labelling and Alert Generation

The purpose of this milestone was to convert the predicted Remaining Useful Life (RUL) values into interpretable condition labels, enabling early-warning and maintenance planning mechanisms. The labelling logic was based on pre-defined thresholds applied uniformly across all FD subsets.

Labelling Logic Each predicted RUL value was assigned a condition category as follows: -

NORMAL: RUL > 50 cycles - WARNING: 20 < RUL ≤ 50 cycles - CRITICAL: RUL ≤ 20 cycles This classification provides an interpretable maintenance indicator for each sample, reflecting the health state of the engine component under observation. The labelling process automatically saves a CSV file for each dataset, containing the actual and predicted RUL values along with their respective condition labels

Outcome and Utility

The introduction of RUL-based labelling successfully bridges predictive modelling with actionable insights. It allows maintenance teams to categorize operational states, identify engines nearing end-of-life, and schedule inspections in advance. This alert system forms the basis for real-time health monitoring in subsequent milestones.

Examples of saved files:

bias_analysis_fd1.csv

metrics_fd1.csv

pred_vs_actual_fd1.png

residual_dist_fd1.png

residual_trend_fd1.png

rul_labels_fd1.csv

In this same format the files are saved for all the Datasets.

Combined Conclusion Milestones 3 and 4

collectively deliver a complete evaluation and interpretation framework for RUL prediction. The evaluation phase established quantitative model credibility, while the labelling phase translated model output into an intuitive health-state system.