EV Battery Remaining Useful Life (RUL) — Project Report

Marcelo Silveira

Generated on 2025-09-21 13:26

Abstract

We predict Remaining Useful Life (RUL) of EV cells measured in diagnostic steps remaining until the 80% End-of-Life threshold. Using the Onori EV aging dataset, we build interpretable features, run grouped cross-validation by cell, and select a tuned Elastic Net model that achieves strong performance on cells with sufficient history.

1. Problem Definition & Success Criteria

- Task: supervised regression predicting RUL to 80% EOL.
- Motivation: schedule maintenance before capacity drops and quantify degradation.
- Success: generalize to unseen cells; evaluate with GroupKFold by cell using RMSE/MAE/R^2.

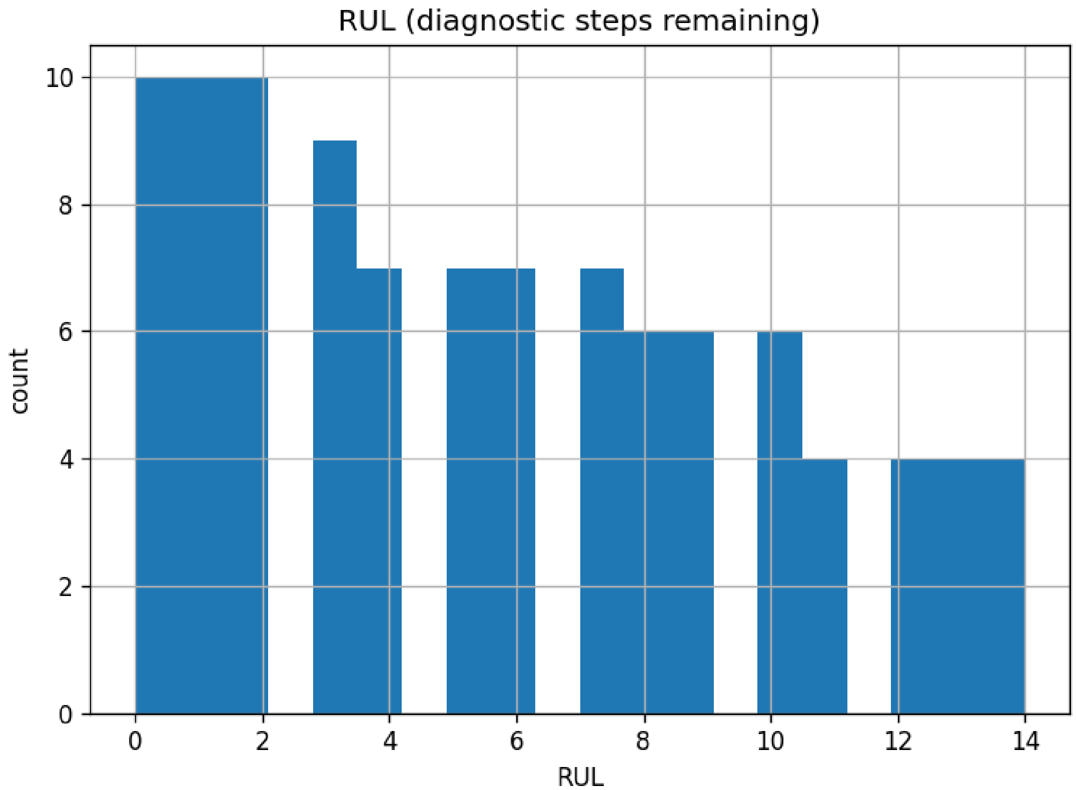
2. Data Provenance & Description

- Source: Onori EV aging diagnostic capacity tests (Diag_*/Capacity_test).
- Derived table: 101 diagnostics across 10 cells (raw Excel kept out of Git).
- EOL rule: 0.8 × initial discharge capacity per cell.
- Features: diagnostic index, capacity (Ah), fade fraction, local slope cap_slope_k3, plus parsed c_rate and temperature.

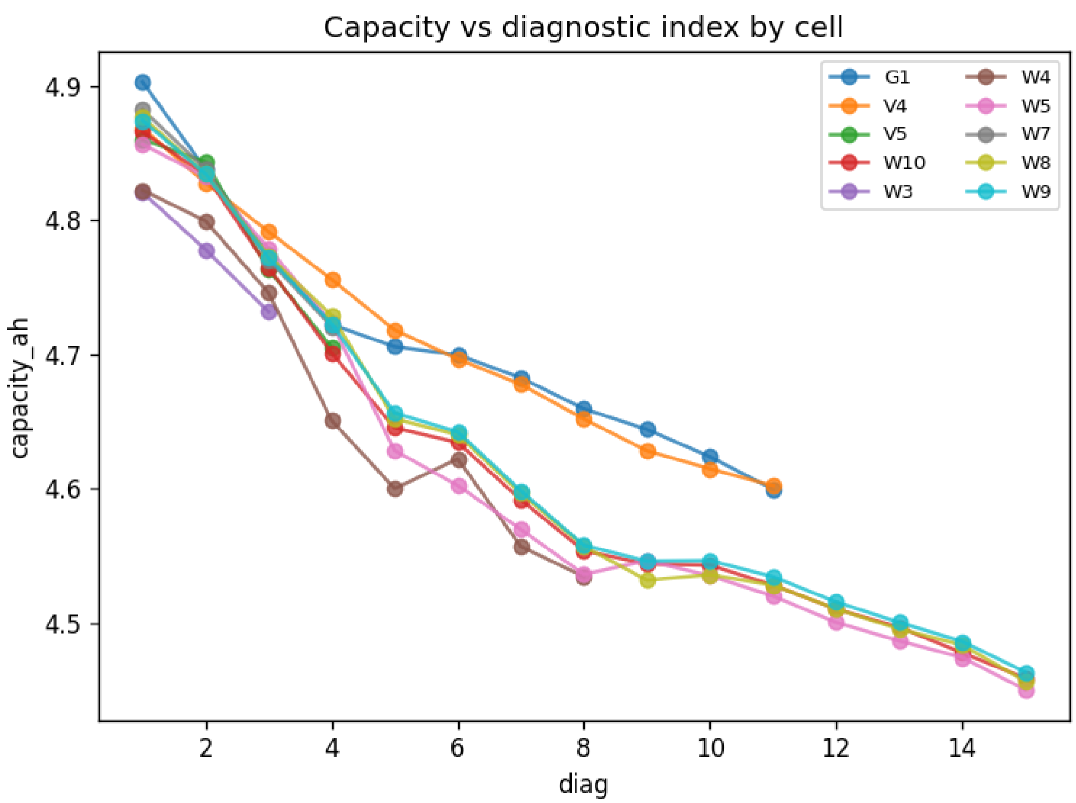
Derived dataset snapshot: 101 rows across 10 cells.

3. Cleaning & Exploratory Analysis

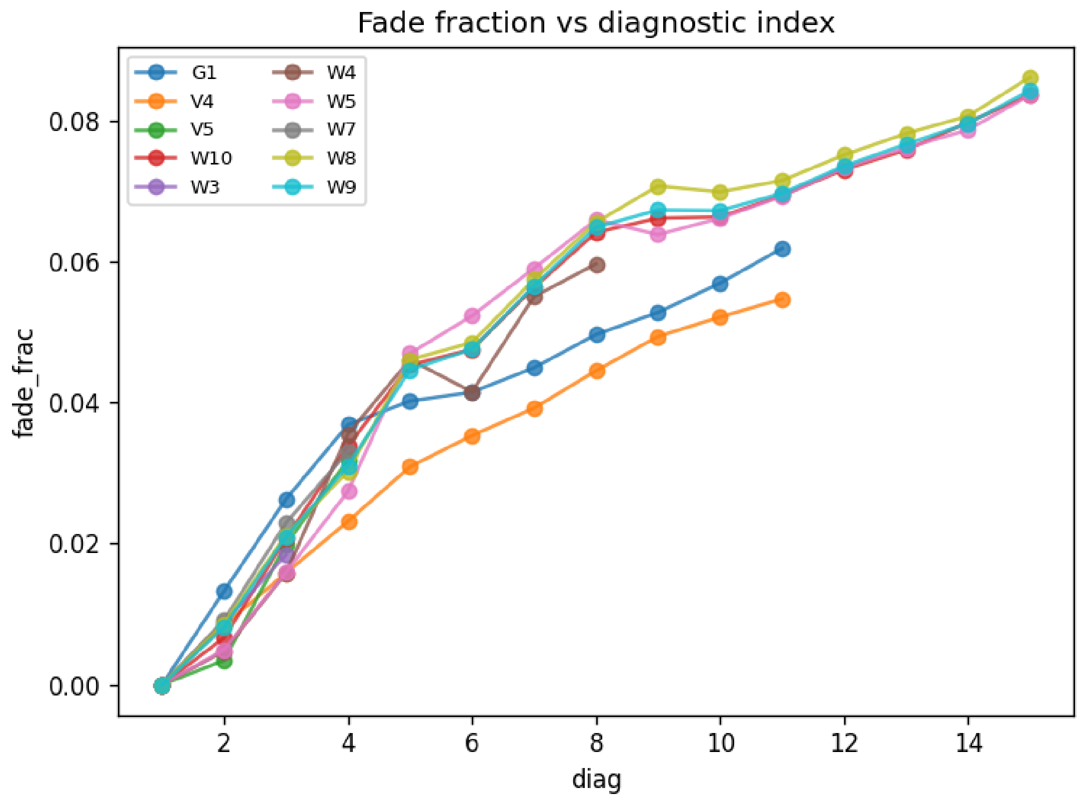
- Label: find first diagnostic with capacity ≤ 80% initial (diag_EOL); set RUL = diag_EOL − diag.
- Missing data: engineered features backfill neutral defaults (e.g., slopes=0, median temperature).
- Insights: RUL distribution is right-skewed; capacity declines monotonically but at varying rates across cell groups.



RUL distribution (diagnostic steps remaining).



Capacity vs diagnostic index by cell.



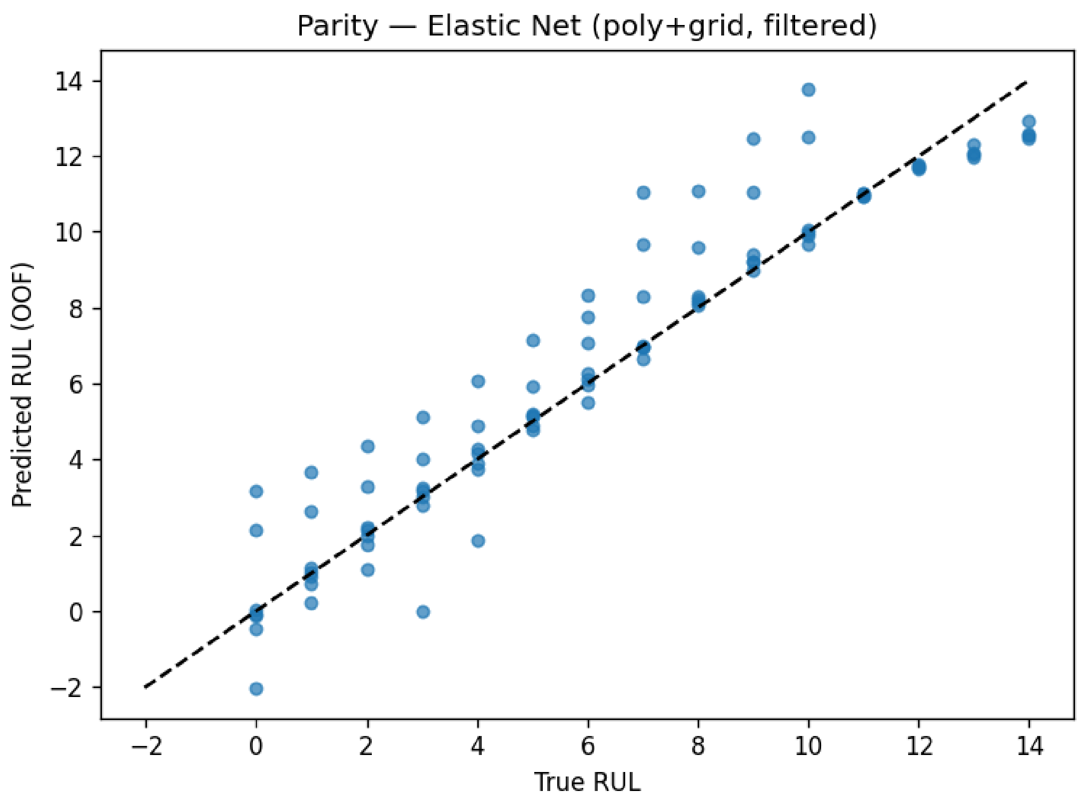
Fade fraction vs diagnostic index by cell.

4. Models & Cross-Validation

We compare interpretable and non-linear baselines using GroupKFold (≈leave-cell-out) to avoid leakage between diagnostics of the same cell.

- Baselines: mean predictor, diagnostic-index-only Linear Regression.
- Candidate models: Elastic Net, Random Forest, Histogram Gradient Boosting, SVR (RBF).
- Tuning: Elastic Net with polynomial features on cells with ≥5 diagnostics (OOF headline).

Model	RMSE	MAE	R^2
linear_diag_only	**3.375**	**2.774**	**0.349**
elastic_net	3.458	2.747	0.308
svr_rbf	3.634	3.098	0.236
random_forest	3.685	2.538	0.198
hist_gbm	3.773	3.012	0.111
baseline_mean	4.180	3.616	-0.017



Parity plot — Elastic Net (polynomial features, filtered ≥5 diags).

Cell	RMSE	MAE	R^2
G1	2.772	2.717	0.231
V4	1.579	1.490	0.751
W10	0.472	0.281	0.988
W4	2.201	1.849	0.077
W5	0.494	0.320	0.987
W8	0.531	0.376	0.985
W9	0.353	0.199	0.993

5. Results & Discussion

• Linear baselines already perform well on diagnostic index; Elastic Net adds stability via regularization.

• Tree ensembles and SVR did not outperform linear methods given the monotonic trends and small dataset.

• Tuned Elastic Net achieves ≈RMSE 1.35 steps, MAE 0.88, R^2 0.89 (out-of-fold).

• Coefficient inspection (see slides) matches battery intuition: capacity and fade terms dominate, diagnostic index negative.

6. Limitations & Future Work

- Only 10 cells; results may not generalize to other chemistries or temperatures.
- Some cells provide limited diagnostics (<5), constraining model training.
- Planned improvements: incorporate HPPC resistance deltas, quantile/PI models for conservative planning, expand datasets.

References

- Onori, S. et al. Lithium-ion battery aging dataset based on EV real-driving profiles, Stanford Bits & Watts, 2022.
- NASA Prognostics Center of Excellence. Li-ion battery aging datasets.

Repository: https://github.com/marbatis/electric-vehicle-telemetry

Reproduce: run notebooks/01_clean_eda.ipynb (EDA) and notebooks/02_models_cv.ipynb (CV + tuned OOF); both read tracked results/ and regenerate figures/leaderboard.