IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

Leakage-Safe, Explainable Remaining Useful Life Prediction for EV Lithium-Ion Batteries Using XGBoost with Physics-Based Cross-Checks and Out-of-Distribution Flagging

G V Neeradh[1], K Raja Venkata Aneesh[2], S Poornima[3]

Dept. of Computer Science and Engineering (Emerging Technologies), SRM Institute of Science and Technology, Vadapalani Campus Chennai, India[1-3]

Abstract: Studies of battery remaining-useful-life (RUL) prediction on the NASA PCoE dataset usually report one R^2 from a random split of cycle-level rows, which says little about accuracy on an unseen cell, behaviour near end of life (EOL), or whether a deployed prediction should be trusted. This study builds a lightweight predictive-maintenance framework around an XGBoost regressor that estimates RUL from 10 health indicators (capacity, impedance and fade-dynamics features) derived from routine cycling and impedance data. Training is leakage-controlled: data are split by battery, post-EOL rows are removed, and features that encode cell identity or need full history are excluded. At inference the model is wrapped by a state-of-health (SOH) calculation, a robust fade-rate estimator, a physics-based linear RUL cross-check that replaces or blends with the model when it collapses to zero or disagrees by more than 2×, rule-based degradation flags, SHAP attributions, a heuristic uncertainty range, and an out-of-distribution (OOD) check against training ranges. On 2,765 samples, a random row-wise 80/20 split gave test R^2 0.9714 (MAE 2.96 cycles, RMSE 7.89 cycles) against train R^2 0.9990; this is treated as an optimistic reference. On a battery-grouped split the test R^2 is [TBD] and MAE [TBD] cycles. Four dashboard scenarios (fresh, degrading, near-EOL, NASA B0005) show the cross-check flagging a model-versus-linear disagreement in two cases and agreeing in the other two. The whole pipeline runs in 3.1 s on a CPU. The model is trained on room-temperature laboratory cells and is not validated on EV drive cycles.

Index Terms: battery management system, electric vehicle, explainable machine learning, lithium-ion battery, out-of-distribution detection, predictive maintenance, remaining useful life, state of health, XGBoost

I. INTRODUCTION

A. Background and Motivation

Electric vehicles (EVs) depend on lithium-ion packs whose capacity fades and whose internal resistance grows with cycling. A battery management system (BMS) must therefore estimate state of health (SOH) and remaining useful life (RUL), with EOL conventionally set at 80% of initial capacity [1], [3]. A model that is useful in a BMS needs more than a high R^2: it must work on cells it was not trained on, behave sensibly near EOL, and signal when its output should not be trusted. Small benchmark datasets make all three difficult, because cycle-level rows from one cell are strongly correlated and apparent accuracy depends heavily on how the data are split [9].

B. Limitations of Existing Studies

- **Heavy models.** LSTM [5] and Gaussian-process [6] predictors can be accurate but rely on long sequences or costly inference, which complicates edge BMS deployment.
- **Waveform-dependent features.** Early-cycle cycle-life predictors [4] use voltage-curve features that routine BMS logs may not provide.
- **Row-wise evaluation.** Random splits of cycle-level rows put neighbouring cycles of one cell in both training and test sets, inflating R^2 [9].
- **Accuracy-only reporting.** One R^2 hides behaviour near EOL, where a model trained only on pre-EOL rows has no support and may return values near zero while the cell is still healthy.
- **No deployment guard.** Predictions are rarely accompanied by a physical cross-check, an explanation, or an out-of-distribution (OOD) flag.

C. Research Question and Contributions

Main question. Can a lightweight, CPU-only gradient-boosted model estimate RUL for unseen cells from routine capacity and impedance data, and can a deployment layer make its output auditable? The contributions are:

 This work is licensed under a Creative Commons Attribution 4.0 International License

1) a leakage-controlled training pipeline (battery-grouped split, post-EOL rows removed, identity-leaking and history-dependent features excluded);
2) a 10-feature health-indicator set computable from a short window of recent capacity readings and one impedance measurement;
3) a hybrid RUL estimator that cross-checks the model against a physics-based linear extrapolation and replaces or blends it when the model is at its floor or disagrees by more than 2×;
4) an auditable output: SOH, robust fade rates, rule-based flags, SHAP attributions, a heuristic uncertainty range and OOD warnings; and
5) a dependency-light REST API and web dashboard, with the scope limits of the model stated explicitly.

II. LITERATURE REVIEW

Degradation and health indicators. Capacity fade and impedance rise are the standard indicators of lithium-ion ageing. Growth in electrolyte resistance (Re) and charge-transfer resistance (Rct) is associated with SEI growth and electrode surface degradation, while capacity loss reflects lost lithium inventory and active material [8]. The NASA PCoE dataset [1], [3] provides capacity and EIS-derived Re and Rct across cell lifetimes.

Data-driven RUL prediction. Methods range from Gaussian-process regression [6] and LSTM networks [5] to feature-based regression on early-cycle data [4]; machine learning for battery state and health estimation is also reviewed in [12]. Sequence models need long histories and more compute, whereas tree ensembles such as XGBoost [2] train in seconds on a CPU and support exact additive feature attributions [7].

Evaluation and leakage. Cycles from one cell are autocorrelated, so evaluation should hold out whole cells, and tuning plus performance estimation in one study needs nested resampling [9]. Reporting guidance such as TRIPOD+AI [10] stresses honest uncertainty and external validation.

Research gap. In the literature reviewed by the authors, we found no lightweight pipeline on NASA PCoE that jointly combines battery-grouped evaluation, a physics cross-check for model failure near EOL, explanation, OOD flagging and REST deployment. This is a candidate gap, not an established novelty claim.

III. DATASET

A. Source and Scope

The data are the NASA Ames Prognostics Center of Excellence (PCoE) battery dataset [1], 18650 lithium-ion cells cycled under controlled laboratory conditions at room temperature. The framework predicts RUL for these cells; it makes no claim about EV drive-cycle loading.

B. Target Definition

For each cell, RUL is the number of cycles remaining until discharge capacity falls below 80% of that cell's initial capacity, and SOH is capacity divided by initial capacity, in percent. Rows with RUL ≤ 0 are dropped for RUL training so that zero labels do not dominate.

C. Record Structure

The loader reads metadata.csv, separates charge, discharge and impedance records, and aligns each impedance measurement with its nearest discharge cycle (Table I).

TABLE I
DATASET SUMMARY

Property	Value
Total records	7,565
Record types	Charge 2,815 · Discharge 2,794 · Impedance 1,956
Cells	34 in raw metadata; 33 contribute samples [confirm]
Measurements	Voltage, current, temperature, capacity, Re, Rct
Model-ready samples	2,765 (15 columns including identifiers and targets)
EOL definition	80% of the cell's initial capacity

D. Features and Leakage Control

Ten features enter the model (Table II). Re and Rct are normalised by a baseline equal to the median of the first five valid EIS readings. Four groups of columns are excluded: absolute Capacity (lets trees memorise per-cell EOL points; soh_ratio carries the same information in a cell-independent form), cv_duration_proxy (needs the full history, unavailable at prediction time), temperature columns, and identifiers and targets (battery_id, SOH, RUL).

TABLE II
MODEL FEATURES

Feature	Description
cycle_number	Charge-discharge cycle index
soh_ratio	Capacity relative to the cell's initial capacity
Re, Rct	Electrolyte and charge-transfer resistance (Ω)
Re_normalized, Rct_normalized	Re and Rct relative to baseline
total_impedance	Re + Rct
capacity_fade_rate	Capacity change from the previous cycle
capacity_acceleration	Second difference of capacity
rolling_fade_rate	3-cycle rolling mean of capacity change

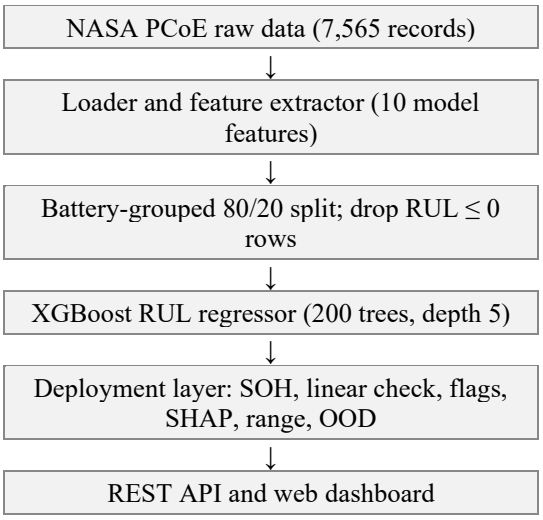


Fig. 1. Framework pipeline from raw data to dashboard.

IV. MODELS AND DEPLOYMENT LAYER

A. Regressors

XGBoost (primary): 200 trees, learning rate 0.1, maximum depth 5, row and column subsampling 0.8, seed 42. **SVR** (alternative, supported in code but not evaluated here): RBF kernel, C = 100, γ = 0.1, StandardScaler [11]. Hyperparameters are fixed, not tuned.

B. Inference-Time Features

A user supplies the cycle number, a window of recent capacities (5–10 readings recommended), Re, Rct, and optionally the initial capacity, Re and Rct (defaults 1.86 Ah, 0.075 Ω, 0.097 Ω). All Table II features are derived from these, mirroring the training extractor.

 This work is licensed under a Creative Commons Attribution 4.0 International License

C. Health Estimation and Physics Cross-Check

Capacities are smoothed with a 5-cycle rolling median to suppress noise and regeneration rebounds. The recent fade trend is the least-squares slope s of the last (up to) 10 smoothed values. Accelerating fade is flagged only with at least 15 cycles of data and when the recent slope is steeper than the earlier slope by more than 0.001 Ah/cycle.

$$SOH = \min(Q_t / Q_0, 1) \times 100\% \quad (1)$$

$$RUL_{lin} = (Q_t - 0.8 \cdot Q_0) / |s| \quad (2)$$

where Q_t is current capacity and Q_0 initial capacity. RUL_{lin} is undefined when the cell is not fading.

D. Headline RUL Logic

The model output and RUL_{lin} are combined by the rules in Table III.

TABLE III
HEADLINE RUL RULES

Condition	Reported RUL
SOH ≤ 80%	0 cycles, status CRITICAL
Model < 0.5 cycles while SOH > 82% (“model at floor”)	RUL_{lin}; range spans latest-fade and trend-fade projections
Model and RUL_{lin} differ by more than 2×	Mean of both; range spans both; disagreement warning
Otherwise	Model output with heuristic range

Status labels follow: WARNING if RUL ≤ 5 or SOH is within 2 points of EOL; DEGRADING if RUL ≤ 20 or within 8 points; otherwise HEALTHY.

E. Diagnostics, Attribution, Uncertainty and OOD

Rule-based flags (threshold checks, not causal): capacity fade ≥ 0.5% of initial capacity per cycle, accelerating fade, and Re or Rct at least 10% above baseline. Thresholds were set from the training-data distribution. **Attributions:** XGBoost’s built-in SHAP contributions [7] give the features that pushed predicted RUL down most, with Re and Re_normalized grouped. **Uncertainty range:** the standard deviation σ of predictions from sub-ensembles of the first k trees gives a half-width max(1.5σ, 3) cycles; this is a heuristic, not a calibrated interval. **OOD check:** any input outside the training range of that feature, widened by 5% of the span, is reported.

F. Serving

A standard-library Python HTTP server exposes GET /api/health and POST /api/predict (JSON in, JSON out with RUL, range, status, SOH, fade rates, flags, attributions and OOD warnings) and serves a web dashboard with four preset scenarios (fresh cell, degrading, near EOL, NASA B0005), editable baseline specifications, an RUL and SOH report, a degradation-trajectory chart, fade dynamics, rule-based flags and SHAP attributions.

V. ALGORITHM

1) Load raw NASA data; align impedance with discharge cycles; compute RUL and SOH per cell.
2) Extract health-indicator features and save to CSV.
3) Compute soh_ratio per cell if absent; drop rows with RUL ≤ 0.
4) Remove excluded columns (Section III-D).
5) Split 80/20 by battery_id (GroupShuffleSplit, seed 42), so no cell appears in both sets.
6) Fit XGBoost on the training cells; save the model, feature list and per-feature training ranges.
7) Evaluate on training and test sets; save predicted-vs-actual and residual plots.
8) At inference: derive features, predict, cross-check with RUL_{lin} (Table III), then attach SOH, flags, attributions, range and OOD warnings.

 This work is licensed under a Creative Commons Attribution 4.0 International License

VI. EVALUATION AND STATISTICAL ANALYSIS

- **MSE: \(\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2\)**
- **MAE: \(\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i|\)**
- **RMSE: \(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}\)**
- **R^2: \(1-\frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y})^2}\)**

Splits. Two protocols are reported. An earlier run used a random row-wise 80/20 split (2,212 training and 553 test rows). The current pipeline uses the battery-grouped split of Section V, which leaves about seven cells for testing. Train and test metrics are both reported to expose overfitting.

Not yet performed. Leave-one-battery-out cross-validation, bootstrap confidence intervals, hyperparameter tuning, interval-coverage evaluation, ablations of the deployment layer, and the SVR comparison.

VII. RESULTS

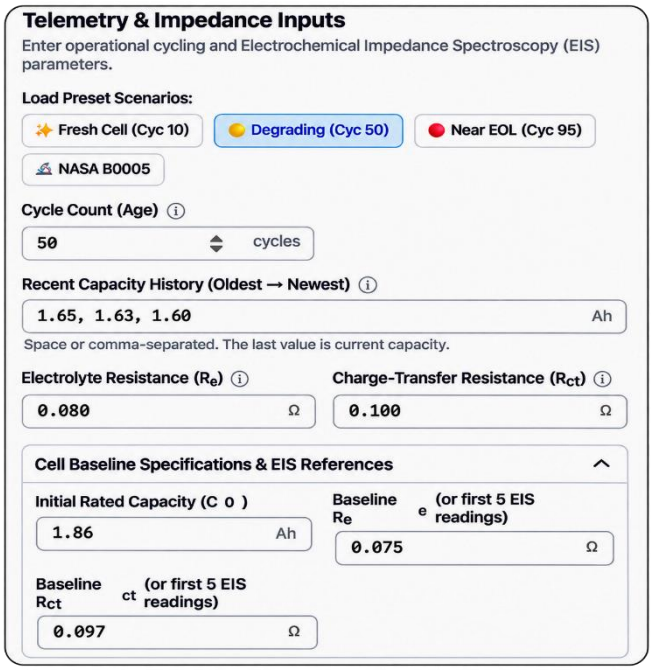
A. Predictive Performance

TABLE IV
XGBOOST RUL PERFORMANCE

Metric	Row-wise split, train	Row-wise split, test	Grouped split, test
MSE (cycles²)	2.3294	62.2382	[TBD]
RMSE (cycles)	1.53	7.89	[TBD]
MAE (cycles)	1.0614	2.9629	[TBD]
R^2	0.9990	0.9714	[TBD]

Under the row-wise split the model explains 97.14% of test RUL variance with a typical error of about 3 cycles. The train–test MSE ratio of roughly 27 and the RMSE being about 2.7 times the MAE indicate overfitting to cell-specific trajectories and a heavy-tailed error distribution, so these figures are treated as an upper bound; the grouped-split column is the estimate relevant to unseen cells.

[Insert models/xgboost_RUL_results.png: p



 This work is licensed under a Creative Commons Attribution 4.0 International License

Derived 11 ML Pipeline Features View All

Feature	Calculated Value	Physical Context
cycle_number	10	Operational life consumed (cumulative cycles)
Capacity	1.828	Latest discharge capacity (Ah)
soh_ratio	0.9828	Latest capacity relative to initial capacity (C / C_0)
Re	0.076	Electrolyte & bulk ohmic resistance from EIS high-frequency intercept (Ω)
Rct	0.098	Charge-transfer resistance from EIS semicircle diameter (Ω)
capacity_fade_rate	-0.002	Discharge capacity delta over most recent cycle (Ah/cyc)
Re_normalized	1.0133	Electrolyte resistance growth ratio relative to baseline (Re / Re_0)
Rct_normalized	1.0103	Charge-transfer resistance growth ratio relative to baseline (Rct / Rct_0)
total_impedance	0.174	Summed cell internal impedance (Re + Rct) (Ω)
cv_duration_proxy	0.0425	Physics-informed proxy for cumulative lithium plating & SEI loss
capacity_acceleration	0	Second derivative of capacity fade (detects knee-point transition)
rolling_fade_rate	-0.0023	Smoothed recent linear regression fade rate (Ah/cyc)
initial_re	0.075	Baseline electrolyte resistance reference (Ω)
initial_rct	0.097	Baseline charge-transfer resistance reference (Ω)

Derived 11 ML Pipeline Features View All

Feature	Calculated Value	Physical Context
cycle_number	50	Operational life consumed (cumulative cycles)
Capacity	1.6	Latest discharge capacity (Ah)
soh_ratio	0.8602	Latest capacity relative to initial capacity (C / C_0)
Re	0.08	Electrolyte & bulk ohmic resistance from EIS high-frequency intercept (Ω)
Rct	0.1	Charge-transfer resistance from EIS semicicle diameter (Ω)
capacity_fade_rate	-0.01	Discharge capacity delta over most recent cycle (Ah/cyc)
Re_normalized	1.0667	Electrolyte resistance growth ratio relative to baseline (Re / Re_0)
Rct_normalized	1.0309	Charge-transfer resistance growth ratio relative to baseline (Rct / Rct_0)
total_impedance	0.18	Summed cell internal impedance (Re + Rct) (Ω)
cv_duration_proxy	0.554	Physics-informed proxy for cumulative lithium plating & SEI loss
capacity_acceleration	0	Second derivative of capacity fade (detects knee-point transition)
rolling_fade_rate	-0.0133	Smoothed recent linear regression fade rate (Ah/cyc)
initial_re	0.075	Baseline electrolyte resistance reference (Ω)
initial_rct	0.097	Baseline charge-transfer resistance reference (Ω)

Derived 11 ML Pipeline Features View All

Feature	Calculated Value	Physical Context
cycle_number	95	Operational life consumed (cumulative cycles)
Capacity	1.482	Latest discharge capacity (Ah)
soh_ratio	0.7968	Latest capacity relative to initial capacity (C / C_0)
Re	0.095	Electrolyte & bulk ohmic resistance from EIS high-frequency intercept (Ω)
Rct	0.165	Charge-transfer resistance from EIS semicircle diameter (Ω)
capacity_fade_rate	-0.006	Discharge capacity delta over most recent cycle (Ah/cyc)
Re_normalized	1.2667	Electrolyte resistance growth ratio relative to baseline (Re / Re_0)
Rct_normalized	1.701	Charge-transfer resistance growth ratio relative to baseline (Rct / Rct_0)
total_impedance	0.26	Summed cell internal impedance (Re + Rct) (Ω)
cv_duration_proxy	1.008	Physics-informed proxy for cumulative lithium plating & SEI loss
capacity_acceleration	-0.002	Second derivative of capacity fade (detects knee-point transition)
rolling_fade_rate	-0.006	Smoothed recent linear regression fade rate (Ah/cyc)
initial_re	0.075	Baseline electrolyte resistance reference (Ω)
initial_rct	0.097	Baseline charge-transfer resistance reference (Ω)

Derived 11 ML Pipeline Features View All

Feature	Calculated Value	Physical Context
cycle_number	80	Operational life consumed (cumulative cycles)
Capacity	1.55	Latest discharge capacity (Ah)
soh_ratio	0.8351	Latest capacity relative to initial capacity (C / C_0)
Re	0.088	Electrolyte & bulk ohmic resistance from EIS high-frequency intercept (Ω)
Rct	0.135	Charge-transfer resistance from EIS semicircle diameter (Ω)
capacity_fade_rate	-0.02	Discharge capacity delta over most recent cycle (Ah/cyc)
Re_normalized	1.1733	Electrolyte resistance growth ratio relative to baseline (Re / Re_0)
Rct_normalized	1.3918	Charge-transfer resistance growth ratio relative to baseline (Rct / Rct_0)
total_impedance	0.223	Summed cell internal impedance (Re + Rct) (Ω)
cv_duration_proxy	0.7386	Physics-informed proxy for cumulative lithium plating & SEI loss
capacity_acceleration	-0.01	Second derivative of capacity fade (detects knee-point transition)
rolling_fade_rate	-0.0167	Smoothed recent linear regression fade rate (Ah/cyc)
initial_re	0.075	Baseline electrolyte resistance reference (Ω)
initial_rct	0.097	Baseline charge-transfer resistance reference (Ω)

Fig. 2. Predicted vs actual RUL and residual histogram on the test set.

B. Scenario Demonstrations

Four preset scenarios in the dashboard exercise the framework end to end (Tables V and VI, Fig. 3). All derived features are recomputed from the entered capacities, Re, Rct and baselines (initial capacity 1.86 Ah, or 1.856 Ah for B0005; Re_0 0.075 Ω; Rct_0 0.097 Ω).

 This work is licensed under a Creative Commons Attribution 4.0 International License

TABLE V
SCENARIO INPUTS AND HEADLINE OUTPUT (RUL IN CYCLES)

	Fresh cell	Degrading	Near EOL	NASA B0005
Cycle	10	50	95	80
Capacity / initial (Ah)	1.828 / 1.860	1.600 / 1.860	1.482 / 1.860	1.550 / 1.856
Re / Rct (Ω)	0.076 / 0.098	0.080 / 0.100	0.095 / 0.165	0.088 / 0.135
SOH	98.3%	86.0%	79.7%	83.5%
Model RUL†	≈52.6	≈21.0	0	5.0
Linear RUL	135.7	9.1	0	5.9
Reported RUL (range)	94.1 (52.6–135.7), blended	15.0 (9.1–21.0), blended	0 (0–0)	5.0 (2–8)
Status	HEALTHY	DEGRADING	CRITICAL	WARNING

† For the two blended cases the model RUL is read from the range limit shown by the dashboard, since the blend is the mean of the model and linear estimates.

TABLE VI
FADE DYNAMICS, FLAGS AND ATTRIBUTIONS

	Fresh cell	Degrading	Near EOL	NASA B0005
Latest fade (Ah/cycle)	−0.0030	−0.0200	−0.0080	−0.0200
Recent trend, Ah/cycle (% per cycle)	−0.0025 (0.13%)	−0.0123 (0.66%)	−0.0098 (0.53%)	−0.0111 (0.60%)
Average since new (Ah/cycle)	−0.0032	−0.0052	−0.0040	−0.0038
Rule-based flags	None	Rapid fade (0.66%/cycle)	SOH 79.7% below EOL; rapid fade (0.53%/cycle); Re +27%; Rct +70%	SOH 3.5 points above EOL; rapid fade (0.60%/cycle); Re +17%; Rct +39%
Top attributions (cycles)	Re (grouped) −8.6; total impedance −1.3; fade acceleration −0.1	Re −7.3; cycle count −6.8; total impedance −3.9; Rct −0.7	Cycle count −10.0; Rct −9.3; SOH ratio −7.3; Re −6.1	Rct −10.1; total impedance −7.9; cycle count −7.7; Re −6.7

The model and the linear projection agree in two scenarios (near EOL, both 0; B0005, 5.0 against 5.9 cycles) and differ by more than 2× in the other two (fresh cell 2.6×, degrading 2.3×). In those two the framework reports their mean, spans the range over both, and shows a disagreement warning. In the fresh-cell case the model is the more pessimistic estimate. In the degrading case the linear projection is the shorter, since it extrapolates the recent trend (−0.0123 Ah/cycle), which is more than twice the average fade since new (−0.0052 Ah/cycle). A single input cannot show which estimate is closer to the truth; that needs the grouped-split and leave-one-battery-out evaluation of Section VI.

Flags and attributions track the degradation state: none are raised for the fresh cell, one for the degrading cell, and four for the near-EOL and B0005 cells, where Re has grown 17–27% and Rct 39–70%. The leading attribution moves from electrolyte resistance (fresh, degrading) to cycle count or charge-transfer resistance (near EOL, B0005). Attributions are correlational; the large Re contribution for the fresh cell (−8.6 cycles at only 1.3% resistance growth) suggests the model also relies on absolute Re, which should be checked against cell-to-cell variation.

These are single illustrative inputs entered in the dashboard, not a validation against measured RUL. [Confirm that the B0005 preset values are the measured series.]

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

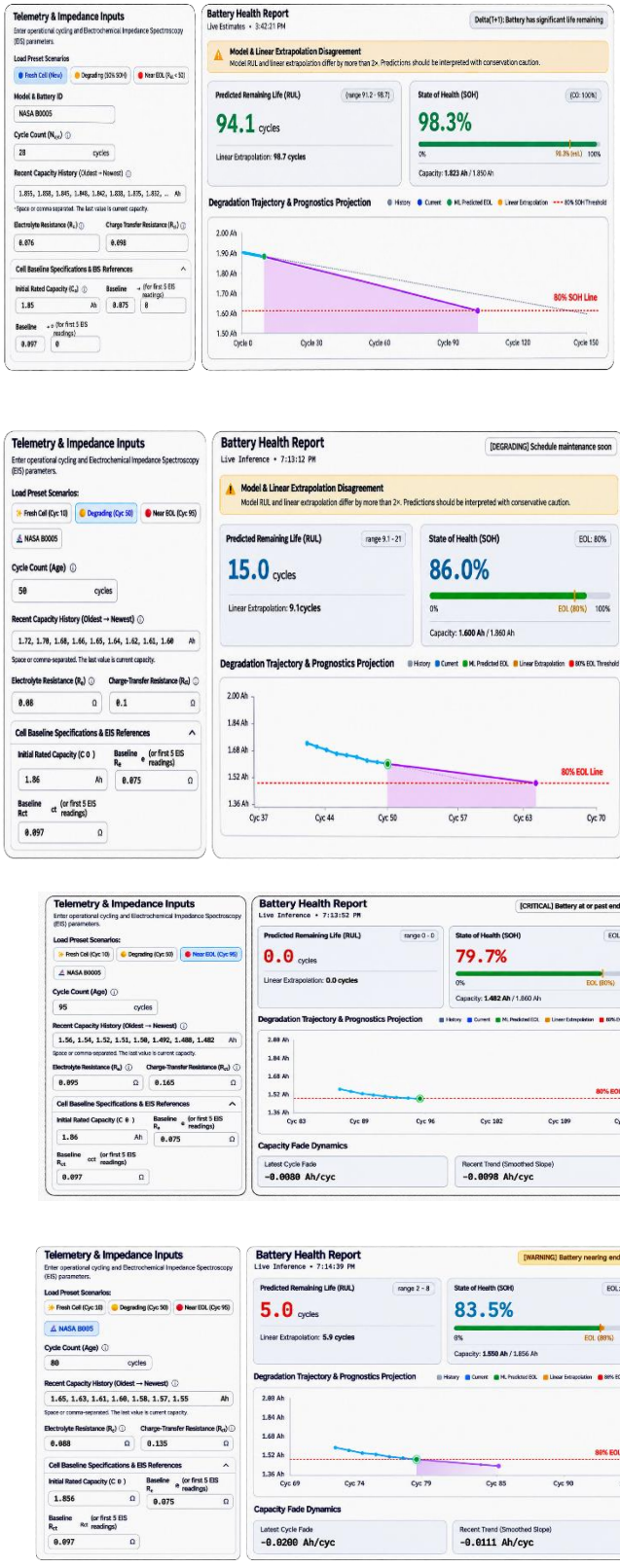


Fig. 3. Dashboard output for the four scenarios of Table V: (a) fresh cell, (b) degrading, (c) near EOL, (d) NASA B0005.

© IJARCCE This work is licensed under a Creative Commons Attribution 4.0 International License

C. Computational Cost

The complete pipeline (load, extract, train, evaluate) ran in 3.1 s on a CPU with no GPU. Inference latency and embedded-hardware benchmarks were not measured.

D. Deployment Layer

The fallback rules, flags, attribution and OOD check are demonstrated on four illustrative scenarios (Table V) but not validated; their quantitative effect (for example, error reduction from the fallback or coverage of the uncertainty range) has not yet been evaluated.

VIII. DISCUSSION AND LIMITATIONS

The experiments show that a small tree ensemble reaches high R^2 on NASA PCoE cells in seconds, but that the headline figure depends on the split protocol; the contribution of this work is the leakage-controlled training and the auditable deployment layer rather than a new regressor. The following limitations apply:

- **Single dataset and protocol.** About 33 cells of one cell format, room temperature, laboratory cycling; no EV drive-cycle, temperature or chemistry variation. "EV" is the motivation, not a validated setting.
- **Small test set.** A single grouped split with about seven test cells gives imprecise estimates; no confidence intervals yet. The four dashboard scenarios are single illustrative inputs, not validation.
- **No tuning or nested CV.** Hyperparameters are fixed, unlike a nested-CV protocol [9].
- **Heuristic components.** Thresholds, the 2× disagreement rule, the floor margin and the uncertainty range are not calibrated; flags are not causal and SHAP attributions are correlational.
- **User-supplied baselines.** SOH and soh_ratio depend on the initial capacity entered (default 1.86 Ah); a wrong value shifts predictions. SOH is capped at 100% and a warning is shown.
- **Fixed EOL.** The 80% threshold is not adapted to application requirements.
- **Prototype serving.** The HTTP server has no authentication and permissive CORS; it is not production-grade.
- **Not clinically or safety certified.** The system is a research and educational tool.

IX. CONCLUSION

This study presents a lightweight, leakage-controlled RUL framework for lithium-ion batteries on the NASA PCoE dataset: an XGBoost model on 10 health indicators, wrapped by a physics-based cross-check, SOH and fade diagnostics, SHAP attributions, a heuristic uncertainty range and OOD flagging, served through a REST API. A random row-wise split gave test R^2 0.9714 and MAE 2.96 cycles, treated here as optimistic; the battery-grouped result is [TBD]. The full pipeline runs in 3.1 s on a CPU. Future work includes leave-one-battery-out evaluation with bootstrap intervals, Optuna-based tuning, conformal prediction to calibrate RUL intervals, testing on other datasets and EV drive-cycle data, ablation of the deployment layer, and embedded-hardware benchmarking.

REFERENCES

[1]. B. Saha and K. Goebel, "Battery Data Set," NASA Prognostics Data Repository, NASA Ames Research Center, 2007.

[2]. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.

[3]. K. Goebel, B. Saha, A. Saxena, J. R. Celaya, and J. P. Christophersen, "Prognostics in battery health management," *IEEE Instrum. Meas. Mag.*, vol. 11, no. 4, pp. 33–40, 2008.

[4]. K. A. Severson *et al.*, "Data-driven prediction of battery cycle life before capacity degradation," *Nature Energy*, vol. 4, pp. 383–391, 2019.

[5]. Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5695–5705, 2018.

[6]. D. Richardson, M. A. Osborne, and D. A. Howey, "Gaussian process regression for forecasting battery state of health," *J. Power Sources*, vol. 357, pp. 209–219, 2017.

[7]. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 4765–4774.

[8]. C. R. Birkl, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey, "Degradation diagnostics for lithium ion cells," *J. Power Sources*, vol. 341, pp. 373–386, 2017.

[9]. S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, art. 91, 2006.

[10]. G. S. Collins *et al.*, "TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods," *BMJ*, vol. 385, e078378, 2024.

[11]. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[12]. M.-F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, "Predicting the state of charge and health of batteries using data-driven machine learning," *Nature Mach. Intell.*, vol. 2, pp. 161–170, 2020.

[13]. M.-F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, "Predicting the state of charge and health of batteries using data-driven machine learning," Nature Mach. Intell., vol. 2, pp. 161–170, 2020.

[14]. X. Hu, L. Xu, X. Lin, and M. Pecht, "Battery lifetime prognostics," Joule, vol. 4, no. 2, pp. 310–346, 2020.

[15]. J. Vetter et al., "Ageing mechanisms in lithium-ion batteries," J. Power Sources, vol. 147, no. 1–2, pp. 269–281, 2005.

[16]. Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming, and A. A. Lee, "Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning," Nature Commun., vol. 11, art. 1706, 2020.

[17]. R. Xiong, Y. Zhang, J. Wang, H. He, S. Peng, and M. Pecht, "Lithium-ion battery health prognosis based on a real battery management system used in electric vehicles," IEEE Trans. Veh. Technol., vol. 68, no. 5, pp. 4110–4121, 2019.

[18]. S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," ACM Trans. Knowl. Discovery Data, vol. 6, no. 4, art. 15, 2012.

[19]. D. R. Roberts et al., "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," Ecography, vol. 40, no. 8, pp. 913–929, 2017.

[20]. S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," Nature Mach. Intell., vol. 2, no. 1, pp. 56–67, 2020.

[21]. D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in Proc. Int. Conf. Learn. Representations (ICLR), 2017.

[22]. A. N. Angelopoulos and S. Bates, "A gentle introduction to conformal prediction and distribution-free uncertainty quantification," arXiv:2107.07511, 2021.