Journal of Modern Technology
https://journal.krrishpub.com/index.php/jmt/index

RESEARCH ARTICLE

Explainable Machine Learning Algorithms for State-of-Health Estimation and Remaining Useful Life Prediction of Lithium-Ion Batteries

Rajanand Patnaik Narasipuram [a,*], Aymen Flah [b,c,d], Alireza Hosseinpour [e]

[a] *Chief Engineer, Automotive & Mobility Group, Cyient Ltd., Hyderabad 500 032, India.*
[b] *Applied Science Research Center, Applied Science Private University, Amman, 11931, Jordan.*
[c] *National School of Engineering of Gabes, University of Gabes, Zrig Eddakhlania 6029, Gabès, Tunisia.*
[d] *College of Engineering, University of Business and Technology, Jeddah, 21448, Saudi Arabia.*
[e] *Associate Professor, Department of Electrical Engineering, University of Zabol., Zabol, 98615-538, Iran.*

Abstract

State-of-health (SOH) estimation and remaining useful life (RUL) prediction address related but distinct aspects of lithium-ion battery degradation. This study evaluates separate XGBoost regressors for SOH estimation and RUL prediction using four NASA Ames lithium-ion cells with complete capacity-fade trajectories. The reference capacity is defined as the median capacity of the first five discharges, while end of life (EOL) is defined as the first subsequent cycle at which capacity falls to or below 80% of this reference. At a fixed observation point of 600 s, the prediction features comprise electrical, thermal, cycle-related, initial-reference, and preceding electrochemical impedance spectroscopy (EIS) measurements. Full-discharge capacity and future EOL information are used only to construct retrospective target labels and are excluded from the corresponding predictors. A leave-one-cell-out evaluation yields 351 evaluated cycles. XGBoost achieves pooled SOH and RUL mean absolute errors (MAEs) of 1.157 percentage points and 17.948 cycles, respectively, while ridge regression achieves 1.195 percentage points and 17.611 cycles. The held-out B0006 cell exhibits a substantially higher XGBoost RUL MAE of 34.886 cycles, highlighting the limitations of cross-cell prognostic transfer despite low pooled SOH error. TreeSHAP analysis identifies the change in 600-s voltage relative to the cell's initial reference as the largest average contribution to SOH predictions. Removing this feature increases SOH root-mean-square error (RMSE) from 1.473 to 2.929 percentage points, while reducing RUL RMSE from 22.731 to 20.836 cycles. These findings demonstrate that feature importance and predictive utility differ between SOH estimation and RUL prediction.

Keywords: Lithium-ion batteries, State-of-health (SOH), Remaining useful life (RUL), XGBoost, Explainable machine learning.

Article information:
ISSN: 3049-3102 (Online)
Published by: **Krrish Scientific Publications Pvt. Ltd.**
DOI: https://doi.org/10.71426/jmt.v3.i2.pp494-503
Received: *02 Nov. 2026* | Revised: *18 Nov. 2026* | Accepted: *21 Nov. 2026* | Published: *22 Nov. 2026*
Copyright ©2026 Author(s).
This is an open-access article distributed under the terms of the Creative Commons Attribution–NonCommercial 4.0 International License (CC BY-NC 4.0).

1. Introduction

Battery state of health is often defined through the remaining deliverable capacity relative to a reference, whereas remaining useful life asks when a specified end-of-life (EOL) rule will first be reached [1], [2]. The two variables are connected but not interchangeable. A present capacity estimate can be accurate even when the future trajectory changes with load, temperature, or cell-specific degradation. In the NASA battery-monitoring formulation of Saha et al., inaccessible damage states and uncertain future usage make prognostics fundamentally conditional on both measurements and a failure definition [3]. A method that reports a low SOH error but does not examine its RUL errors on independent cells therefore leaves a central operational question unanswered.

Early-life capacity information can predict cycle lifetime under controlled protocols, as demonstrated by Severson et al. [4]. Their result also emphasizes the distinction between a life prediction made after a fixed early observation period

*Corresponding author
Email addresses: rajanand.patnaik@ieee.org, rajanand.ee@gmail.com (Rajanand Patnaik Narasipuram), flahaymening@yahoo.fr, aymen.flah@vsb.cz (Aymen Flah), a.hoseinpour@uoz.ac.ir (Alireza Hosseinpour).

and a rolling prediction made at each later cycle. Ng et al. review data-driven battery state inference across SOC and SOH tasks [5]; translating those models to prognosis requires an explicit horizon and a reproducible capacity threshold. In laboratory data, rows from successive cycles of one cell share electrode condition, temperature history, and instrumentation. Randomly splitting such rows estimates interpolation within the same specimen rather than transfer to an unseen specimen. A cell-disjoint evaluation is a more demanding, and often much smaller, experiment.

Future duty cycles are another source of uncertainty. Lu et al. modeled battery degradation under uncertain future conditions with a recurrent network [8]. A variant LSTM has also been used for joint SOH and RUL estimation [9]. These architectures can encode time dependence, but architecture names alone say little about whether future capacity values, test-cell trajectories, or post-EOL records entered the training procedure. The model's observation point matters as well: an estimator made after a complete discharge has information unavailable 600 s into that same discharge. The present work fixes a 600-s measurement boundary and reserves the final discharge capacity as a label.

Electrical, thermal, and impedance observations represent different aspects of cell operation. Impedance spectroscopy can reveal patterns associated with degradation [10], yet it is a laboratory procedure whose timing and availability must be stated if it becomes a predictor. A review of SOH and RUL methods documents how different targets, operating protocols, and EOL thresholds complicate comparison [11]. The public NASA PCoE recordings allow voltage, current, case temperature, periodic impedance parameters, and full-discharge capacity to be tied to the same sequence of operations. They also contain several limitations: four cells in the selected complete cohort, different voltage cutoffs at discharge, and room-temperature accelerated cycling. Treating these as a representative electric-vehicle fleet would be unwarranted.

Generalization remains difficult when target-cell labels are scarce. A 2023 deep-learning study examined SOH estimation without additional degradation experiments on a target cell [12]; its setting is distinct from an initially calibrated test cell. Here the first five full discharges of each cell provide only an initial-capacity and voltage reference, after which current-cycle predictors are confined to the first 600 s. We do not treat those five calibration cycles as evidence of an unknown-start, zero-calibration deployment. Nor do we assume that a single EOL crossing measured on a short laboratory history captures the uncertainty of future real-world use.

Interpretability can help audit a fitted model, provided the object of explanation is precisely identified. A review of explainable machine learning for lithium-ion batteries notes that feature rankings and local explanations have been applied to diagnosis and state estimation [6], [7], [13]. SHAP expresses a fitted prediction as a baseline and additive feature contributions under an explicit feature-coalition convention [14]. Correlated voltages at 120, 300, and 600 s can redistribute those contributions; an attribution is not an intervention on electrochemistry. In particular, high importance for an impedance channel would not prove that a new impedance sensor improves performance on another cell.

This paper contributes an executable, cell-held-out benchmark of two linked but separately trained targets. First, it defines a cell-specific 80% EOL rule from five observed calibration discharges, extracts 14 temporally available electrical, thermal, and preceding-impedance features, and audits every source file with SHA-256 checksums. Second, it compares fixed-configuration XGBoost and ridge regressors for both SOH and RUL over four leave-one-cell-out folds, reporting pooled and cell-specific errors in their proper units. Third, TreeSHAP explanations of the XGBoost fits are checked for additivity and compared with feature-removal experiments. The emphasis is on what these four physical cells do and do not support, especially the possibility that acceptable present-health estimation coexists with poor forecast transfer.

2. Literature review

2.1. From measured health to future life

Tree explanations provide a local-to-global aggregation procedure for ensembles [15]. Their efficient evaluation makes repeated cell-held-out interpretation practical, but the explanatory target is a fitted predictor, not the battery itself. XGBoost introduces shrinkage and regularization into boosted tree fitting [16]; with only a few independent cells, its ability to fit nonlinear cycle features can also fit cell-specific signatures. We therefore compare it with a deliberately simpler linear ridge model under the same feature definitions. A small advantage in a pooled cycle metric is not automatically evidence of a better prognostic model when its worst held-out cell remains poor.

Short-window features have a defensible role in health estimation. A linear method based on short-term battery features [17] and an online LSTM capacity estimator [18] offer contrasting ways to exploit temporal measurements [31]. The former has a more transparent mapping and lower data demand; the latter can capture nonlinear history at the price of a training and state-reset protocol. Our predictor fixes its observation window at 600 s so that all tested examples share the same amount of within-discharge information. This is a measured diagnostic point, not a claim about arbitrary drive-cycle telemetry. Voltage observations from a partial discharge can carry information about capacity, yet they are affected by rate, temperature, relaxation, and the starting charge state.

Impedance is similarly informative under controlled conditions but costly to interpret as an onboard sensor. Online impedance measurements have been connected to capacity estimates [19], and high-rate SOH estimation from selected impedance frequencies has been studied [20]. Field-data degradation diagnosis couples impedance-aware models to artificial intelligence [21]. These settings differ from the NASA archive's intermittent laboratory spectroscopy. We use only the latest EIS parameters acquired *before* the predicted discharge and record the number of discharge cycles since that observation. If an EIS measurement has not yet occurred, the feature is missing and is imputed from the other training cells. A retrospective nearest-EIS join would leak later tests; a causal join avoids that error but does not solve instrument availability in an EV.

A survey of machine learning for SOH and RUL [22] emphasizes that the future-life target depends on operating conditions. A pack-level prognostics study [23] makes a further distinction between a cell's measured condition and aggregation across cells in a pack. The four-cell NASA cohort tests neither pack imbalance nor thermal gradients. Its cells were subjected to different final discharge voltages (2.7, 2.5, 2.2, and 2.5 V), so absolute delivered ampere-hours are not directly comparable. Dividing each capacity by a cell-specific early reference reduces scale differences but does not undo the cutoffs' influence on the shape of capacity fade or the chosen EOL cycle.

A multiphysics-feature framework for SOH [24] illustrates how electrical and thermal measurements can be integrated. A sequence-to-sequence model forecasts future voltage-capacity curves [25], whereas a discharge-voltage approach predicts full capacity and resistance degradation curves from early data [26]. Their target objects and observation windows are different: forecasting a curve at a specified future cycle, estimating present capacity, and predicting the first threshold crossing are not equivalent evaluations. In particular, knowing the present capacity exactly still leaves the future degradation slope unresolved. We avoid conflating the two tasks by training a separate RUL regressor rather than deriving RUL from a possibly precise SOH point estimate without a future-trajectory model.

The threshold itself is part of the scientific outcome, rather than a harmless unit conversion. Changing the failure fraction changes which cycles have observable exact RUL, the training distribution, and the length of the remaining horizon. A comparison that reports one life error without specifying the capacity denominator, crossing rule, and censoring convention is therefore difficult to reproduce. A more ambitious model could forecast the capacity trajectory and derive several thresholds from it, but the trajectory forecast would still need validation on new cells. A direct RUL regressor has fewer intermediate assumptions while losing the explicit future-curve representation.

Energy-derived features have been proposed for battery SOH estimation [27]; their extraction can depend on how much of a charge or discharge is available. A full-cycle integral, a time-to-cutoff statistic, or any feature calculated after 600 s would make the stated partial-discharge task easier by using information from the future of the prediction point. Our compact feature family is restricted to three early voltage samples, corresponding temperature values, mean early current, a voltage slope, preceding impedance, and age. It makes the information boundary inspectable in both code and manuscript. The feature family is an engineering choice, not a proof of an optimal diagnostic window.

The unit of statistical independence deserves equal attention. A random cycle split can place almost identical local capacity states of one cell on both sides of a partition, while a cell split exposes differences in starting capacity and aging rate. With only four cells, leave-one-cell-out evaluation gives four genuine test trajectories but little basis for a narrow confidence interval. Pooling 351 cycles yields a descriptive error weighted by the lengths of these trajectories; it is not a 351-specimen trial. An EOL threshold defined by the first observed crossing can be affected by a capacity recovery or a single noisy discharge. Smoothing or requiring a sustained crossing would define a different target and must be fixed before comparison. Some public aging records are censored because the experiment stops before a threshold; they cannot be assigned an observed exact RUL without additional assumptions. All four selected NASA cells reach the specified 80%-of-initial reference in their measured histories.

TreeSHAP's additivity checks that the contributions reconstruct the tree output, but a feature with a large average absolute value may act as a proxy for a correlated input or a laboratory-specific protocol. Removing it and refitting can help measure dependence of predictive performance on that recorded channel under the same cohort. Even a consistent ablation benefit would remain associational, because the remaining tree structure and feature correlations change. We report distinct SHAP rankings for SOH and RUL, use the same fold-specific test predictions as in the metric tables, and include ablations that do not hide a contrary outcome.

2.2. Relation to battery materials and thermal-management reviews

Kamana et al. [28] synthesize battery thermal hazards, cooling strategies, and predictive-control opportunities for intelligent battery management; this motivates collecting temperature alongside electrical signals and eventually assessing prognosis under thermal excursions. Sravanthi and Chandra Sekhar [29] survey nano-metal-oxide materials and nanotechnology for electric-vehicle batteries, illustrating that materials and chemistry affect aging processes. Neither review reports the same four-cell, 600-s, calibrated SOH/RUL benchmark; consequently, numerical MAE values from the present work must not be compared as if those papers used identical inputs, threshold definitions, or test splits. Our contribution is therefore complementary: an auditable prediction protocol with explicit cross-cell failure cases, rather than a new cooling scheme or electrode material.

3. Methodology

Table 1 defines the principal abbreviations, quantities, and symbols used in the feature construction, regression objectives, and evaluation.

3.1. Source, reference, and temporal boundary

The source is the official NASA Ames PCoE Battery Data Set [3], specifically the nested `BatteryAgingARC-FY08Q4` archive described by the original NASA acquisition notes [3]. Cells B0005, B0006, and B0007 have 168 recorded discharges each, while B0018 has 132. Charge, discharge, and impedance operations occur in their recorded order. At approximately 24 °C ambient, the cells were charged with constant-current/constant-voltage control and discharged under approximately 2-A loading to different cutoffs.

For cell c, let $Q_{c,n}$ be the full-discharge capacity in ampere-hours at its nth discharge, and let the first five discharges defined as (1):

$$Q_{0,c} = \text{median}_{n=1,\ldots,5}\, Q_{c,n}, \qquad H_{c,n} = 100 Q_{c,n}/Q_{0,c} \quad (1)$$

Table 1: Definitions of abbreviations and mathematical symbols used in this study.

Abbreviation or symbol	Definition	Unit or interpretation
SOH; $H_{c,n}$	State of health at discharge n for cell c, calculated as $100Q_{c,n}/Q_{0,c}$	%
RUL; $L_{c,n}$	Remaining useful life, defined as $e_c - n$ up to the first EOL crossing	Discharge cycles
EOL; e_c	End of life; first observed discharge index after calibration with capacity no greater than $0.8Q_{0,c}$	Discharge index
$Q_{c,n}$; $Q_{0,c}$	Full-discharge capacity at index n; median capacity of first five discharges of cell c	Ah
c; n; k	Cell identifier; discharge index; index used over the five reference discharges	Identifiers
$V_{c,n}(t)$; $\Delta V_{c,n}$	Voltage at elapsed time t; voltage at 600 s minus reference median 600-s voltage	V
I; T; $s_{c,n}$	Measured current; case temperature; early voltage slope multiplied by 10^3	A; °C; mV/s
EIS; R_e; R_{ct}	Electrochemical impedance spectroscopy; estimated electrolyte and charge-transfer resistances	Ω
XGBoost; SHAP; TreeSHAP	Extreme gradient boosting; Shapley additive explanations; tree-ensemble SHAP implementation	Methods
$\mathbf{x}_i$; y_i; $\hat{y}_i$	Feature vector; true label; fitted prediction for observation i	Target-dependent
$\mathcal{T}$; N	Training observations in a fold; number of evaluated predictions	Counts
f_m; m; K_m; w_{mj}	Boosting-stage tree, boosting stage, terminal-leaf count, and leaf weight	Model parameters
λ; γ	L_2 leaf-weight and tree-complexity penalties in the written XGBoost objective	Regularization parameters
F; S; ϕ_0; ϕ_j	Feature index set, subset of features, SHAP baseline, and contribution of feature j	Prediction units
MAE; RMSE; P90; pp	Mean absolute error; root-mean-square error; 90th percentile absolute error; percentage points	SOH: pp; RUL: cycles

The median is fixed before predicting cycle 6 onward. We define an observed EOL index and a retrospective life target by Equation (2).

$$\begin{aligned} e_c &= \min\{n > 5 : Q_{c,n} \le 0.8Q_{0,c}\}, \\ L_{c,n} &= e_c - n, \qquad 6 \le n \le e_c. \end{aligned} \tag{2}$$

This first-crossing convention yields EOL indices 107, 61, 125, and 78 for B0005, B0006, B0007, and B0018, respectively. The SOH model never receives $Q_{c,n}$ from its target discharge; the RUL model never receives e_c or a future capacity as input. Because the five calibration capacities of each held-out cell are observed, the protocol estimates transfer to a newly aged but *initially calibrated* cell.

3.2. Multisource features and split

At the 600-s prediction instant of each subsequent discharge, 14 predictors are assembled. The electrical group contains discharge index, early voltage at 120, 300, and 600 s, voltage slope from 120 to 600 s, current averaged on a 10-s grid through 600 s, and the 600-s voltage minus the median 600-s voltage of the first five discharges. The thermal group contains case temperatures at 120 and 600 s and their rise. The EIS group contains the most recent preceding electrolyte resistance R_e, charge-transfer resistance R_{ct}, and elapsed discharge count since that measurement. The initial capacity $Q_{0,c}$ is the fourteenth predictor. The time-specific voltage difference and slope, for example, are (3):

$$\begin{aligned} \Delta V_{c,n} &= V_{c,n}(600) - \text{median}_{k\le 5}\, V_{c,k}(600), \\ s_{c,n} &= \frac{10^3[V_{c,n}(600) - V_{c,n}(120)]}{480}. \end{aligned} \tag{3}$$

All interpolation points lie within the current discharge; processing stops if a selected trace is shorter than 600 s. Impedance operations later in the file are never searched backward toward an earlier prediction. Forty-two evaluated cycles have no preceding EIS value; their three EIS-related fields receive medians fitted only on the other three training cells in each fold. The data contain 351 evaluated pre-EOL cycles: 102, 56, 120, and 73 from the four cells. No augmented or simulated rows are used.

To assess cross-cell transfer rather than interpolation between adjacent cycles, each of four leave-one-cell-out folds trains on three complete cell histories and predicts the fourth. Model settings, features, threshold, and evaluation window are fixed before the held-out cell is scored; no fold's test cells select a tree count. There is no validation search over many architectures. This removes sample adjacency across train and test, but the same laboratory, nominal cell family, and ambient protocol remain shared. When pooling predictions we weight each cycle equally; cell-specific errors accompany the pooled values because the number of independent test units is four.

The overall sequence from source verification and reference calibration to held-out evaluation is summarized in Figure 1.

The two early reference variables require five complete capacity tests and five corresponding 600-s voltage observations on each test cell. They are fixed before any later test-cycle predictions. This calibration requirement differs materially from an estimator initialized with no full-discharge measurements; its cost and possible drift should be included when comparing deployment protocols. The cycle counter begins with the publisher's first recorded discharge rather than an inferred manufacturing-age origin.

The measured initial capacities, EOL cycles, evaluated trajectory lengths, and discharge cutoff voltages for the four cells are summarized in Table 2.

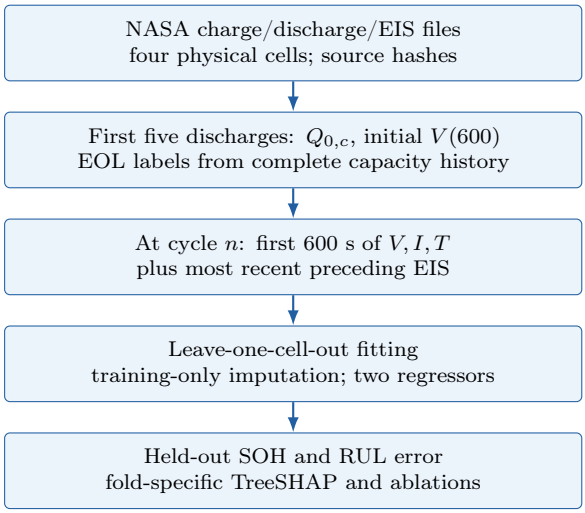

Figure 1: Temporal workflow.

Table 2: Measured-cell reference and evaluated trajectories

Held-out cell	$Q_{0,c}$ (Ah)	EOL cycle	Evaluated	Cutoff (V)
B0005	1.835	107	102	2.7
B0006	2.013	61	56	2.5
B0007	1.881	125	120	2.2
B0018	1.840	78	73	2.5

3.3. Regressors and model-specific explanation

Following the regularized tree-boosting formulation of Chen and Guestrin [16], two separate XGBoost regressors share the feature construction but have different labels, $H_{c,n}$ and $L_{c,n}$. Each has 200 trees of maximum depth 3, learning rate 0.04, minimum child weight 5, row subsampling 0.85, column subsampling 0.9, and L_2 penalty 10. These settings are fixed in the script, as is random seed 32. A ridge model, motivated by the regularized linear estimator of Hoerl and Kennard [30], with training-fold median imputation, training-fold standardization, and penalty 20 is a linear comparator. The XGBoost objective at boosting step m is

$$\mathcal{L}^{(m)} = \sum_{i\in\mathcal{T}} \left(y_i - \hat{y}_i^{(m-1)} - f_m(\mathbf{x}_i)\right)^2 + \gamma K_m + \frac{\lambda}{2}\sum_{j=1}^{K_m} w_{mj}^2, \quad (4)$$

In (4), the training set contains only the other three cells. The number of leaves is K_m, and w_{mj} denotes a leaf weight. The implementation uses squared-error regression and a penalty of 10; other splitting penalties retain their software defaults. Ridge and XGBoost fit SOH and RUL independently. Negative ridge life predictions are clipped to zero only at evaluation; fitted coefficients remain unchanged.

For a fitted tree ensemble and a held-out observation, TreeSHAP [14], [15] returns as in (5).

$$f(\mathbf{x}) = \phi_0 + \sum_{j=1}^{14} \phi_j(\mathbf{x}),$$
$$\phi_j = \sum_{S\subseteq F\setminus\{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \cdot [f_{S\cup\{j\}}(\mathbf{x}) - f_S(\mathbf{x})]. \quad (5)$$

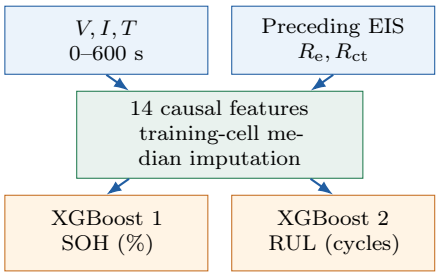

Figure 2: Two independent target heads use the same early-cycle measurements.

The software's native tree contribution routine is applied separately in every test fold. We verify its row-wise sum against the unrounded tree prediction within 10^{-4}. For each target, the global displayed importance is the cycle-weighted mean of held-out $|\phi_j|$ across the four folds. It is measured in SOH percentage points or remaining cycles, according to the regressor. It is not a percentage of error, and the two units must not be ranked in one list.

Figure 2 illustrates the shared feature inputs and the separate SOH and RUL regression models.

Algorithm 1: Cell-held-out health and life evaluation.

1. Verify four raw MATLAB files against the included SHA-256 manifest; preserve the recorded operation order.
2. For each cell, calculate Equations (1)–(2); use only cycles $6, \ldots, e_c$ and derive predictors by Equation (3) from the first 600 s and preceding EIS.
3. For each held-out cell and target, fit imputation and the two regressors on the other three cells only; freeze the fixed hyperparameters.
4. Predict all held-out cycles, clip negative RUL estimates to zero, and append their errors without treating the cycles as independent specimens.
5. Compute TreeSHAP for each XGBoost fold, check Equation (5), and repeat cell-held-out fitting for exploratory feature-removal diagnostics.

For actual labels y_i and predictions $\hat{y}_i$, the descriptive pooled metrics are written as (6):

$$\text{MAE} = \frac{1}{N}\sum_i |\hat{y}_i - y_i|, \qquad \text{RMSE} = \sqrt{\frac{1}{N}\sum_i (\hat{y}_i - y_i)^2}. \quad (6)$$

4. Results and discussion

4.1. SOH and RUL generalization

Table 3 reports predictions concatenated from four genuinely held-out cells. For SOH, XGBoost has 1.157-percentage-point MAE and 1.473-percentage-point RMSE, while ridge has 1.195 and 1.401. The nonlinear model's slightly smaller MAE but larger RMSE implies no uniform advantage; large errors on a limited number of cycles change the squared metric. Both models see the same 14 features and initial cell calibration. These errors are relative to the NASA full-discharge capacity divided by the cell's first-five-cycle reference, rather than an independent electrochemical capacity measurement.

The pooled SOH MAE values for ridge regression and XGBoost are also compared graphically in Figure 3.

Table 3: Pooled leave-one-cell-out errors, $N = 351$ cycles

Target	Model	MAE	RMSE	Bias	P90
SOH (pp)	Ridge	1.195	1.401	−0.234	2.076
SOH (pp)	XGBoost	1.157	1.473	0.373	2.234
RUL (cycles)	Ridge	17.611	23.830	5.866	44.520
RUL (cycles)	XGBoost	17.948	22.731	−3.758	38.524

4.2. Advantages and boundaries of the proposed evaluation

The principal advantages are methodological rather than an unconditional accuracy gain. First, the leave-one-cell-out split tests transfer to an unseen physical cell instead of mixing highly correlated cycles across train and test. Second, limiting electrical and thermal signals to 600 s, while joining only preceding EIS measurements, makes the information available at prediction time explicit. Third, independent SOH and RUL models expose a practical distinction: a small capacity-state error does not imply an accurate future threshold time. Fourth, training-only imputation, fixed settings, per-cell reporting, and feature-removal checks make the experiment easier to audit. These benefits are balanced by the need for five full-discharge calibration cycles, a standardized 600-s observation, and only four independent held-out batteries. The proposed approach is not established as superior to ridge regression on every metric.

4.3. Discussion relative to earlier studies

Severson et al. [4] investigated lifetime prediction from early-life cycling data, whereas the present study repeatedly predicts remaining cycles from each observed diagnostic discharge after initial calibration. Lu et al. [8] studied future degradation trajectories under uncertain conditions and a larger experimental cohort; our fixed-protocol four-cell exercise cannot verify such operating-condition robustness. Li et al. [9] used a recurrent architecture for health and life prediction, whereas the present comparison intentionally uses interpretable tabular regressors. The JMT review of thermal safety and predictive BMS control [28] identifies thermal-management issues that are not independently varied here. Similarly, the materials review [29] describes chemistry and electrode-design factors that may alter degradation but are not controlled by the present feature set. Because target definitions, cycle windows, cell populations, and validation schemes differ, a numerical ranking of the studies' prediction errors would be misleading. Instead, the key empirical finding here is the gap between pooled SOH error and B0006's 34.886-cycle RUL error, emphasizing that prognostic transfer needs separate evidence beyond an accurate present-health estimate.

The RUL target is much harder. XGBoost's pooled 17.948-cycle MAE and 22.731-cycle RMSE describe only the first threshold crossing for these four aging sequences. Ridge has 17.611 and 23.830 cycles. In B0005, XGBoost SOH MAE is 0.681 pp and RUL MAE is 9.147 cycles; in B0006 these rise to 2.377 pp and 34.886 cycles. The B0006 RUL prediction has a positive mean bias of 34.886 cycles, so the model repeatedly forecasts substantially more life than is observed. On B0007, by contrast, the RUL mean bias is −15.027 cycles. A pooled sign of −3.758 cycles

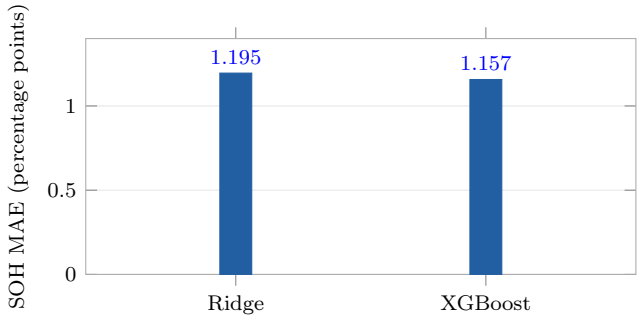

Figure 3: SOH model comparison from 351 cell-held-out cycles.

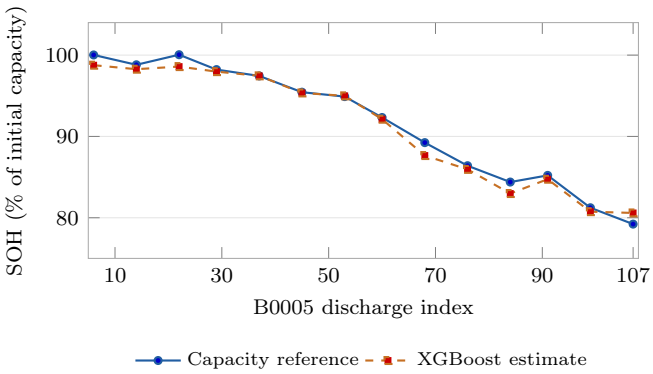

Figure 4: Held-out B0005 SOH trace, regularly subsampled for display; metrics use all 102 test cycles.

partly cancels these opposing cell errors and should not be used as an operational safety margin.

The ridge comparator also has a 49.568-cycle RUL MAE on B0006; XGBoost improves that particular held-out cell to 34.886 cycles but remains much too optimistic for a threshold-triggered decision. For B0007, XGBoost RUL MAE is 19.358 cycles. These variations are larger than the pooled ridge-versus-XGBoost difference, which is why selecting an architecture by the pooled score alone would miss the scientifically consequential failure mode.

Figure 4 illustrates B0005. The XGBoost prediction tracks much of the measured decline but misses local capacity recovery and several drops by about one percentage point. Only sparse displayed points appear in the figure; Table 3 uses every evaluated cycle. An attractive trace on B0005 must be read beside B0006, whose SOH error is more than three times higher. In particular, the first-five reference does not guarantee that subsequent voltage trajectories are comparable across cells with different discharge cutoffs. The number of test cycles for B0005, B0006, B0007, and B0018 is 102, 56, 120, and 73, so the pooled statistic gives the longest trajectory the greatest weight.

The B0006 RUL trace in Figure 5 makes a more consequential error visible. At cycle 6, observed RUL is 55 cycles, while the XGBoost estimate is about 102; even at the observed EOL cycle 61, it estimates about 18 cycles remaining. This is not a one-time plotting anomaly: all 56 of its evaluated cycles contribute to the 34.886-cycle MAE. A replacement decision tied to a fixed cycle budget could be delayed by this positive error. We do not convert that observation into a failure probability, because four cells

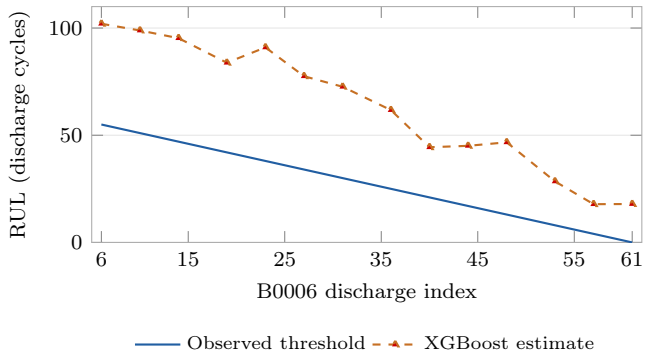

Figure 5: Held-out B0006 exposes persistent RUL overprediction; sparse display points do not alter the 56-cycle error calculation.

Table 4: Top ten SOH TreeSHAP features on held-out cells

Rank	Input feature	Mean $\lvert\phi\rvert$ (pp)
1	600-s voltage change from initial	3.773
2	Voltage at 600 s	0.998
3	Voltage slope, 120–600 s	0.512
4	Voltage at 300 s	0.358
5	Discharge index	0.219
6	Temperature rise, 120–600 s	0.131
7	Electrolyte resistance	0.111
8	Temperature at 120 s	0.060
9	Voltage at 120 s	0.058
10	Charge-transfer resistance	0.048

cannot estimate a calibrated upper tail. A defined EOL threshold is necessary for a reproducible RUL label, but it does not make a point forecast sufficiently reliable for a maintenance decision.

4.4. Interpretation and feature-removal checks

TreeSHAP for SOH assigns the greatest cycle-weighted mean absolute contribution to the difference between current and initial 600-s voltage (3.773 pp), followed by current 600-s voltage (0.998 pp) and early voltage slope (0.512 pp). Table 4 gives the top ten. The ranking is internally additive for each fitted tree regressor, but changes in these correlated voltage descriptors need not correspond to separable mechanisms of capacity loss. The initial-reference voltage is measured on the same cell, so it can compensate for between-cell offsets as well as track degradation. An attribution of 3.773 pp is a model-output contribution relative to a baseline, not an observed SOH error.

The eight largest SOH feature contributions are visualized in Figure 6, complementing the numerical ranking in Table 4.

For RUL, the largest mean absolute XGBoost contributions are voltage at 120 s (7.354 cycles), initial-reference voltage change (5.520), cycle index (4.451), voltage at 600 s (3.844), and charge-transfer resistance (3.141). These are explanations for the *RUL regressor*, not the SOH regressor, and their magnitudes have a different unit. They also do not resolve why B0006 is forecast late. Its error may reflect a faster observed capacity decline, differences in initial normalization, or model extrapolation across test cells. Different discharge cutoff settings may also alter the observed

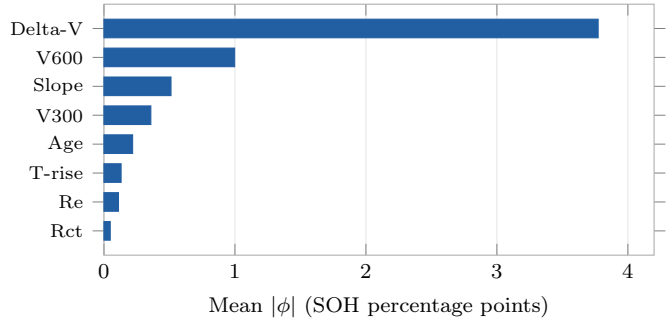

Figure 6: Eight largest SOH TreeSHAP features, aggregated over held-out predictions. Feature names correspond to Table 4; bar length measures tree output attribution.

capacity trajectory and the first-crossing EOL label. These mechanisms are confounded in four cells; SHAP attribution cannot identify a causal explanation.

A controlled refit provides a stricter diagnostic. Removing three impedance-related features raises XGBoost SOH RMSE from 1.473 to 1.480 pp and RUL RMSE from 22.731 to 22.978 cycles, a small change here. Removing the initial voltage-difference feature increases SOH RMSE to 2.929 pp, consistent with the large SOH SHAP contribution. Yet it *decreases* RUL RMSE to 20.836 cycles. Thus the feature that looks influential in both fitted models helps this SOH task but is associated with worse RUL transfer in this four-cell refit. An age-and-initial-capacity-only baseline has 5.326-pp SOH RMSE and 37.755-cycle RUL RMSE, confirming that the measured early-cycle signals carry substantial predictive information relative to that restricted baseline. These ablations are exploratory comparisons on the same four test cells; they do not independently validate a new feature choice.

5. Prognostic horizon and EOL sensitivity

5.1. Error as the threshold approaches

A pooled RUL MAE mixes long-range forecasts with predictions near a replacement threshold. We stratified the same 351 XGBoost held-out predictions by retrospectively observed remaining cycles, without using the strata to fit a model. For 187 observations with more than 40 cycles remaining, MAE is 22.360 cycles and signed bias is −13.677 cycles. For the 80 observations with 21–40 remaining, MAE is 15.678 cycles and bias is +6.490 cycles. For 84 observations with 0–20 remaining, MAE is 10.287 cycles and bias is +8.562 cycles. The absolute error decreases toward EOL, partly because the possible nonnegative RUL range becomes narrower. More concerning is the sign change: the average long-horizon estimate is conservative, whereas the average near-EOL estimate is optimistic. These averages still hide the large positive bias on B0006 and negative bias on B0007. No stratum supplies enough independent cells to estimate an operational probability of late replacement.

Ridge shows a different pattern: its MAE is 16.528 cycles above 40 remaining, 21.016 at 21–40, and 16.781 at 0–20. Its near-EOL bias is +13.511 cycles, larger than XGBoost's +8.562. The difference is descriptive, since the same cell histories create each horizon band and the bands are not randomized populations. A maintenance

controller would need to specify whether positive RUL error is more costly than negative error. MAE treats both signs equally, so a loss that explicitly penalizes optimism would answer a different decision problem. Such a loss would require enough new cells for an honest assessment after it is chosen.

5.2. Dependence on the capacity-failure rule

We repeated the XGBoost leave-one-cell-out RUL procedure with a first crossing at 85% and 75% of the same five-discharge reference. At 85%, EOL occurs at cycles 81, 47, 86, and 63 for B0005, B0006, B0007, and B0018. There are 257 eligible cycle predictions, with pooled RUL MAE 12.581 cycles and RMSE 16.638. At 75%, those EOL indices become 129, 74, 164, and 100; the 447 eligible predictions have MAE 23.184 cycles and RMSE 29.645. The original 80% rule falls between them at 351 predictions and 17.948-cycle MAE. All three thresholds are observed within these four measured files; no future trajectory is extrapolated merely to assign a label.

These numerical changes are not a controlled improvement from moving the threshold. Altering EOL changes each target, the length of each cell's evaluated record, the distribution seen by the fold-specific model, and the prevalence of long-range forecasts. B0006 remains difficult at all three thresholds: its XGBoost RUL MAEs are 32.614, 34.886, and 39.793 cycles for 85, 80, and 75%, respectively. Reporting the threshold sensitivity exposes the dependence of an attractive summary number on the endpoint definition. A cross-study comparison is only meaningful when the capacity reference, cutoff protocol, and EOL rule are harmonized. The supplementary CSV retains all per-cell scores and sample counts, so the pooled comparisons can be reweighted by cell if a different descriptive estimand is needed.

6. Sensitivity, validity, and reproducibility

6.1. What the two errors mean

The capacity target is measured at the end of a complete discharge, whereas the predictors stop at 600 s. This makes SOH estimation an early-discharge inference rather than simple extraction of the already known capacity, but it also requires a standardized discharge from a charged condition. An EV pack experiencing arbitrary partial charge and regenerative events does not present the same observation. The recorded ambient is around 24 °C; case temperature changes during a discharge, but there is no independently held-out cold or hot aging cohort in this evaluation. EIS parameters require a diagnostic measurement that an ordinary vehicle controller may not perform. The near-zero impedance ablation effect on SOH does not prove EIS is generally dispensable; another chemistry, rate, or temperature may have a different relationship.

Different discharge cutoff voltages complicate comparisons among the four cells. The 80% threshold in Equation (2) is relative to each cell's own initial measured capacity, not 80% of one universal rated capacity. B0006's initially measured 2.013 Ah and EOL at cycle 61 contrast with B0007's 1.881 Ah and EOL at cycle 125. Early-cycle voltage features can encode a combination of cell condition, cutoff procedure, and charging preparation. The split prevents the algorithm from seeing future rows of the test cell, but it cannot disentangle these sources of variation from four specimens. The first threshold crossing is observed in all four files; no survival method for censored cells is needed here. A larger cohort that includes non-crossing cells would require a censoring-aware endpoint and a different loss or evaluation protocol.

It is tempting to infer RUL from a precise SOH estimate by assuming a fixed degradation slope. The B0006 outcome demonstrates why such an assumption needs testing. Its held-out SOH XGBoost MAE is 2.377 pp, yet the direct RUL model overpredicts remaining life by 34.886 cycles on average. The capacity curve contains transient recoveries; neither a point health estimate nor a fold-averaged SHAP contribution gives a calibrated distribution of future slopes. A risk-sensitive prognosis would need independent-cell replication, a forecast distribution conditioned on future usage, and a prespecified acceptable rate of optimistic predictions. We intentionally avoid assigning nominal 95% intervals from 351 serially dependent points.

6.2. Reproduction and scope

The companion ZIP contains the four unchanged NASA MATLAB cell files, the acquisition README, a checksum manifest identifying the official nested archive, and an executable Python script. The script verifies each included source file, extracts discharge and preceding EIS operations, writes the 351-row cycle-level dataset, performs the four held-out folds, checks TreeSHAP additivity, and writes predictions and feature ablations. NumPy, SciPy, pandas, scikit-learn, and XGBoost versions are recorded in the result summary. The fold predictions and all plotted numerical coordinates can be recalculated without inventing a cell or retrieving a proprietary database. The complete 200-MB multi-cohort NASA wrapper archive is not duplicated; only the original four-cell nested cohort selected for this experiment is included.

The study has no temporal test on a new usage policy, no between-laboratory check, no pack replication, and no sensor-fault campaign. Its SOH and RUL metrics are therefore internal to a controlled 18650 laboratory cohort. The descriptive differences between ridge and XGBoost are too small, and the number of independent cells too low, to support a significance claim. Interpretable model output is useful for inspecting feature reliance but cannot replace experimental validation of why a cell aged at a particular rate.

7. Conclusion

Four independent held-out NASA cell trajectories provided a stringent, small-sample comparison of present SOH inference and first-threshold RUL prediction. With a five-discharge initial reference and 600-s early-discharge predictors, XGBoost produced pooled SOH MAE of 1.157 percentage points and RUL MAE of 17.948 cycles across 351 evaluated cycles. Ridge produced 1.195 percentage points and 17.611 cycles. The largest individual RUL failure was B0006, for which XGBoost overpredicted by 34.886 cycles on average. SOH TreeSHAP emphasized the change

in 600-s voltage from its initial reference; removing that feature worsened SOH RMSE from 1.473 to 2.929 percentage points, while reducing RUL RMSE from 22.731 to 20.836 cycles. The contrast demonstrates that an explanatory ranking and a transferable forecast answer different questions.

These results support a reproducible within-dataset diagnostic experiment, not a calibrated electric-vehicle prognostics guarantee. Future work should lock the feature and threshold definitions before testing more cells with common cutoff protocols, multiple temperatures and usage policies, explicit censored histories, and independently checked capacity. Cell-level uncertainty estimates and a prespecified penalty for optimistic RUL predictions are needed before using the forecasts for maintenance decisions.

Declarations and ethical statements

Conflict of interest: The authors declare that there is no conflict of interest.

Funding statement: The authors declare that no specific funding was received for this research.

Artificial Intelligence usage statement: During the preparation of this manuscript, the authors utilized ChatGPT solely for grammatical corrections and visualizations. The author carefully reviewed and revised the generated content and take full responsibility for the accuracy, integrity, and originality of the final manuscript.

Availability of data and materials: The data and/or materials that support the findings of this study are available from the corresponding author(s) upon reasonable request.

CRediT authorship contribution statement

Rajanand Patnaik Narasipuram: Conceptualization, Methodology & Writing – original draft. **Aymen Flah**: Conceptualization & Writing – review. **Alireza Hosseinpour**: Conceptualization & Writing – review.

Publisher's note

Krrish Scientific Publications Pvt. Ltd. and the ***Journal of Modern Technology*** remain neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Narasipuram RP, Pasha MM, Tabassum S, Tandon AS. The Electric Vehicle Surge: Effective Solutions for Charging Challenges with Advanced Converter Technologies. *Energy Engineering*. 2025 Jan 1;122(2):431–69. Available from: https://doi.org/10.32604/ee.2025.055134

[2] Mahmoudi C, Flah A, Sbita L. An overview of electric vehicle concept and power management strategies. In *2014 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM), Tunis, Tunisia*. 2014 Nov 3 (pp. 1-8). IEEE. Available from: https://doi.org/10.1109/cistem.2014.7077026

[3] Saha B, Goebel K, Poll S, Christophersen J. Prognostics Methods for Battery Health Monitoring Using a Bayesian Framework. *IEEE Transactions on Instrumentation and Measurement*. 2009;58(2):291-296. Available from: https://doi.org/10.1109/tim.2008.2005965

[4] Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Z, et al. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*. 2019;4(5):383-391. Available from: https://doi.org/10.1038/s41560-019-0356-8

[5] Ng MF, Zhao J, Yan Q, Conduit GJ, Seh ZW. Predicting the state of charge and health of batteries using data-driven machine learning. *Nature Machine Intelligence*. 2020;2(3):161-170. Available from: https://doi.org/10.1038/s42256-020-0156-7

[6] Arandhakar S, Nakka J, Krishna VBM. A comprehensive analysis and future prospects on battery energy storage systems for electric vehicle applications. *Energy Sources Part a Recovery Utilization and Environmental Effects*. 2024 Sep 18;46(1):13004–31. Available from: https://doi.org/10.1080/15567036.2024.2401118

[7] Rao VS, Sajja GS, Manur VB, Arandhakar S, Krishna VBM. An exploratory study on intelligent active cell balancing of electric vehicle battery management and performance using machine learning algorithms. *Results in Engineering*. 2025 Feb 28;25:104524. Available from: https://doi.org/10.1016/j.rineng.2025.104524

[8] Lu J, Xiong R, Tian J, Wang C, Hsu CW, Tsou NT, et al. Battery degradation prediction against uncertain future conditions with recurrent neural network enabled deep learning. *Energy Storage Materials*. 2022;50:139-151. Available from: https://doi.org/10.1016/j.ensm.2022.05.007

[9] Li P, Zhang Z, Xiong Q, Ding B, Hou J, Luo D, et al. State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network. *Journal of Power Sources*. 2020;459:228069. Available from: https://doi.org/10.1016/j.jpowsour.2020.228069

[10] Zhang Y, Tang Q, Zhang Y, Wang J, Stimming U, Lee AA. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. *Nature Communications*. 2020;11(1):1706. Available from: https://doi.org/10.1038/s41467-020-15235-7

[11] Ge MF, Liu Y, Jiang X, Liu J. A review on state of health estimations and remaining useful life prognostics of lithium-ion batteries. *Measurement*. 2021;174:109057. Available from: https://doi.org/10.1016/j.measurement.2021.109057

[12] Lu J, Xiong R, Tian J, Wang C, Sun F. Deep learning to estimate lithium-ion battery state of health without additional degradation experiments. *Nature Communications*. 2023;14(1):2760. Available from: https://doi.org/10.1038/s41467-023-38458-w

[13] Faraji Niri M, Aslansefat K, Haghi S, Hashemian M, Daub RD, Marco J. A Review of the Applications of Explainable Machine Learning for Lithium-Ion Batteries: From Production to State and Performance Estimation. *Energies*. 2023;16(17):6360. Available from: https://doi.org/10.3390/en16176360

[14] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*. 2017;30. Available from: https://doi.org/10.48550/arXiv.1705.07874

[15] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*. 2020;2(1):56-67. Available from: https://doi.org/10.1038/s42256-019-0138-9

[16] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2016. p. 785-794. Available from: https://doi.org/10.1145/2939672.2939785

[17] Ang EYM, Paw YC. Efficient linear predictive model with short term features for lithium-ion batteries state of health estimation.

Journal of Energy Storage. 2021;44:103409. Available from: https://doi.org/10.1016/j.est.2021.103409

[18] Li W, Sengupta N, Dechent P, Howey D, Annaswamy A, Sauer DU. Online capacity estimation of lithium-ion batteries with deep long short-term memory networks. *Journal of Power Sources*. 2021;482:228863. Available from: https://doi.org/10.1016/j.jpowsour.2020.228863

[19] Mc Carthy K, Gullapalli H, Kennedy T. Online state of health estimation of Li-ion polymer batteries using real time impedance measurements. *Applied Energy*. 2022;307:118210. Available from: https://doi.org/10.1016/j.apenergy.2021.118210

[20] Jiang S, Song Z. Estimating the State of Health of Lithium-Ion Batteries with a High Discharge Rate through Impedance. *Energies*. 2021;14(16):4833. Available from: https://doi.org/10.3390/en14164833

[21] Li W, Chen J, Quade K, Luder D, Gong J, Sauer DU. Battery degradation diagnosis with field data, impedance-based modeling and artificial intelligence. *Energy Storage Materials*. 2022;53:391-403. Available from: https://doi.org/10.1016/j.ensm.2022.08.021

[22] Rauf H, Khalid M, Arshad N. Machine learning in state of health and remaining useful life estimation: Theoretical and technological development in battery degradation modelling. *Renewable and Sustainable Energy Reviews*. 2022;156:111903. Available from: https://doi.org/10.1016/j.rser.2021.111903

[23] Chang C, Wu Y, Jiang J, Jiang Y, Tian A, Li T, et al. Prognostics of the state of health for lithium-ion battery packs in energy storage applications. *Energy*. 2022;239:122189. Available from: https://doi.org/10.1016/j.energy.2021.122189

[24] Son S, Jeong S, Kwak E, Kim JH, Oh KY. Integrated framework for SOH estimation of lithium-ion batteries using multiphysics features. *Energy*. 2022;238:121712. Available from: https://doi.org/10.1016/j.energy.2021.121712

[25] Tian J, Xiong R, Shen W, Lu J. Data-driven battery degradation prediction: Forecasting voltage-capacity curves using one-cycle data. *EcoMat*. 2022;4(5):e12213. Available from: https://doi.org/10.1002/eom2.12213

[26] Ibraheem R, Strange C, dos Reis GA. Capacity and Internal Resistance of lithium-ion batteries: Full degradation curve prediction from Voltage response at constant Current at discharge. *Journal of Power Sources*. 2023;556:232477. Available from: https://doi.org/10.1016/j.jpowsour.2022.232477

[27] Gong D, Gao Y, Kou Y, Wang Y. State of health estimation for lithium-ion battery based on energy features. *Energy*. 2022;257:124812. Available from: https://doi.org/10.1016/j.energy.2022.124812

[28] Kamana SB, Okedu KE, Soni BP, Nannam HC, Kumar CVB. Thermal Safety, Cooling Technologies, and Predictive Control for Electro-Thermal Management and Intelligent Battery Management Systems in Electric Vehicles. *Journal of Modern Technology*. 2026;3(1):406–419. Available from: https://doi.org/10.71426/jmt.v3.i1.pp406-419

[29] Sravanthi CL, Chandra Sekhar JN. Advances in Nano Metal Oxides and Nanotechnology for Lithium-Ion Batteries in EV: A Comprehensive Review. *Journal of Modern Technology*. 2025;2(1):254–263. Available from: https://doi.org/10.71426/jmt.v2.i1.pp254-263

[30] Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*. 1970;12(1):55–67. Available from: https://doi.org/10.1080/00401706.1970.10488634

[31] Abdussalam M, Amin W, Zhu Y, Huang Q, Soleimani A, Pinnarelli A, et al. Optimizing energy storage in microgrids under uncertainty: Balancing reliability, cost, and battery wear. *Energy Conversion and Management X*. 2026 Sep 25;32:102334. Available from: https://doi.org/10.1016/j.ecmx.2026.102334