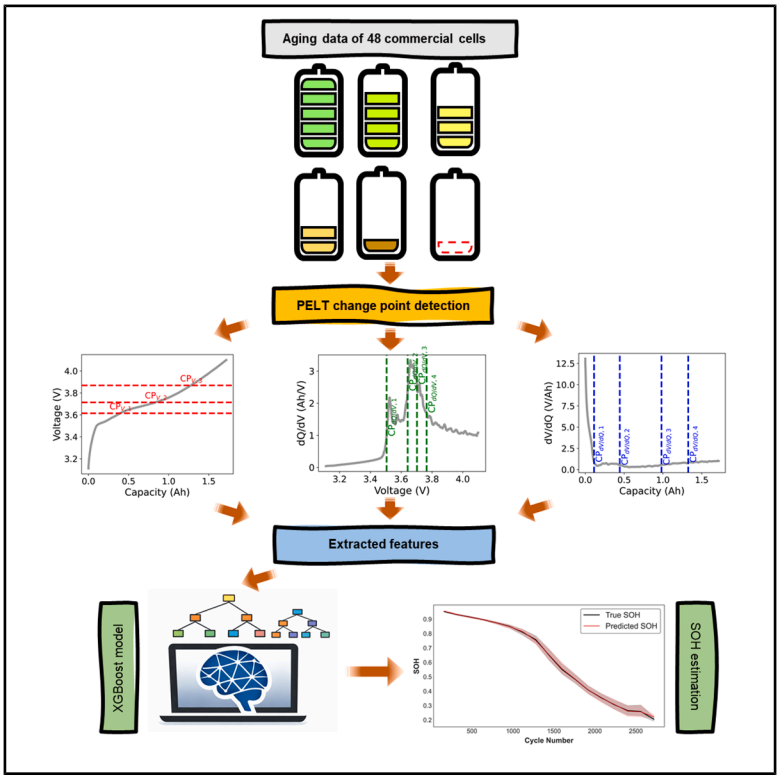
Cell Reports
Physical Science

Article

Changepoint detection as a light data-driven approach to battery state-of-health prediction

Graphical abstract



Authors

Hamid Hamed, Albin Conde Reis, Behnam Ghalami Choobar, Quanquan Pang, Rebecca Killick, Mohammadhossein Safari

Correspondence

momo.safari@uhasselt.be

In brief

Hamed et al. apply changepoint detection to extract features from voltage data and from derivatives of voltage and capacity in 48 NMC|graphite cells. They train an XGBoost model on these features to accurately predict battery SOH, demonstrating fast, interpretable, and transferable results suitable for real-time applications.

Highlights

- Aging analysis of 48 NMC|graphite battery cells
- Lightweight changepoint features enable fast and accurate battery SOH prediction
- Changepoint features correlate with lithium and electrode material loss

Hamed et al., 2026, Cell Reports Physical Science 7, 103157
March 18, 2026
https://doi.org/10.1016/j.xcrp.2026.103157

CellPress

Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

Changepoint detection as a light data-driven approach to battery state-of-health prediction

Hamid Hamed,[1,2,3] Albin Conde Reis,[1,2,3] Behnam Ghalami Choobar,[4] Quanquan Pang,[5] Rebecca Killick,[6] and Mohammadhossein Safari[1,2,3,7,*]

[1]Institute for Materials Research (IUMAT), UHasselt, Martelarenlaan 42, 3500 Hasselt, Belgium
[2]Energyville, Thor Park 8320, 3600 Genk, Belgium
[3]IMEC division IUMAT, 3590 Diepenbeek, Belgium
[4]Department of Chemical Engineering, University of Guilan, Rasht 41996-13776, Iran
[5]Beijing Key Laboratory for Theory and Technology of Advanced Battery Materials, School of Materials Science and Engineering, Peking University, Beijing 100871, China
[6]School of Mathematical Sciences, Lancaster University, Lancaster LA1 4YF, UK
[7]Lead contact
*Correspondence: momo.safari@uhasselt.be
https://doi.org/10.1016/j.xcrp.2026.103157

SUMMARY

Accurate prediction of battery state of health (SOH) remains challenging because degradation processes are highly sensitive to cell chemistry, manufacturing variability, and operating conditions, while available field data are often limited. Generalized and data-efficient modeling approaches are therefore required for reliable battery health assessment across different applications. Here, we report a data-driven feature extraction framework based on changepoint detection (CPD) to identify statistically meaningful transitions in battery aging data. The approach is applied to both capacity-check and regular aging cycles of $LiNi_xMn_yCo_zO_2$|graphite cells. The extracted features are used to train an extreme-gradient-boosting regressor, enabling accurate SOH estimation with root-mean-square errors of 0.013 and 0.023 for capacity-check and aging-cycle datasets, respectively. The features show strong correlation with lithium loss and active-material degradation, demonstrating that CPD provides a physics-aware and computationally efficient pathway for battery health prognosis.

INTRODUCTION

The data-driven models for state-of-health (SOH) prediction of batteries offer higher simplicity and lower computational efforts relative to the physics-based models,[1] eliminating the need for a detailed knowledge of the underlying physics and chemistry of the degradation mechanisms.[2] In this regard, machine learning (ML) increasingly gains more interest in the battery prognosis and diagnosis community.[3,4] For instance, neural networks (NNs) and support vector machines have been successfully employed to predict battery SOH and its remaining useful life (RUL).[5–7]

The data-driven models are usually trained by the raw time series of current, voltage, and temperature of the battery.[8–10] These input data, however, lack a direct and evident correlation with the capacity and power fade of the battery, which in turn hinders the performance, reliability, and physics-aware interpretability of the ML models made thereof. In this regard, feature engineering has emerged as an efficient approach where the raw data are transformed into some health indicators (HIs) before being fed to the ML models.[2,11,12] The purpose is to augment the raw data with some features that better reflect the underlying aging mechanisms of the battery, enhancing the SOH prediction accuracy and interpretability. A widely studied class of HIs is based on the incremental capacity (IC) and differential voltage (DV) analysis, in which the value and position of the peaks/valleys in the IC/DV curves have shown significant correlation with the battery SOH and RUL.[13,14] Various features have been reported in the recent literature based on time series of voltage and current,[15–18] as well as features derived from more specialized characterization techniques such as pulse-relaxation and impedance spectroscopy.[19,20]

The feature extraction from the time series data is particularly attractive because modern battery management systems (BMS) increasingly support real-time logging of voltage and current at the individual cell level.[21–23] Nevertheless, using field data introduces new challenges.[10] For example, charge and discharge data in the field rarely cover the full state-of-charge (SOC) window.[24] This limits the applicability of certain feature extraction methods—such as those based on IC and DV analysis—which require data across the full SOC range to be effective. Furthermore, IC and DV features perform best when low-rate, constant-current (CC) charge, or discharge profiles are available—conditions that are rarely met in field data, where the battery is subject to a variable C-rate with frequent instances of high current densities.[25] Another limitation of traditional feature extraction methods is their computational cost. In particular, IC and DV analysis demand extensive noise filtering before reliable

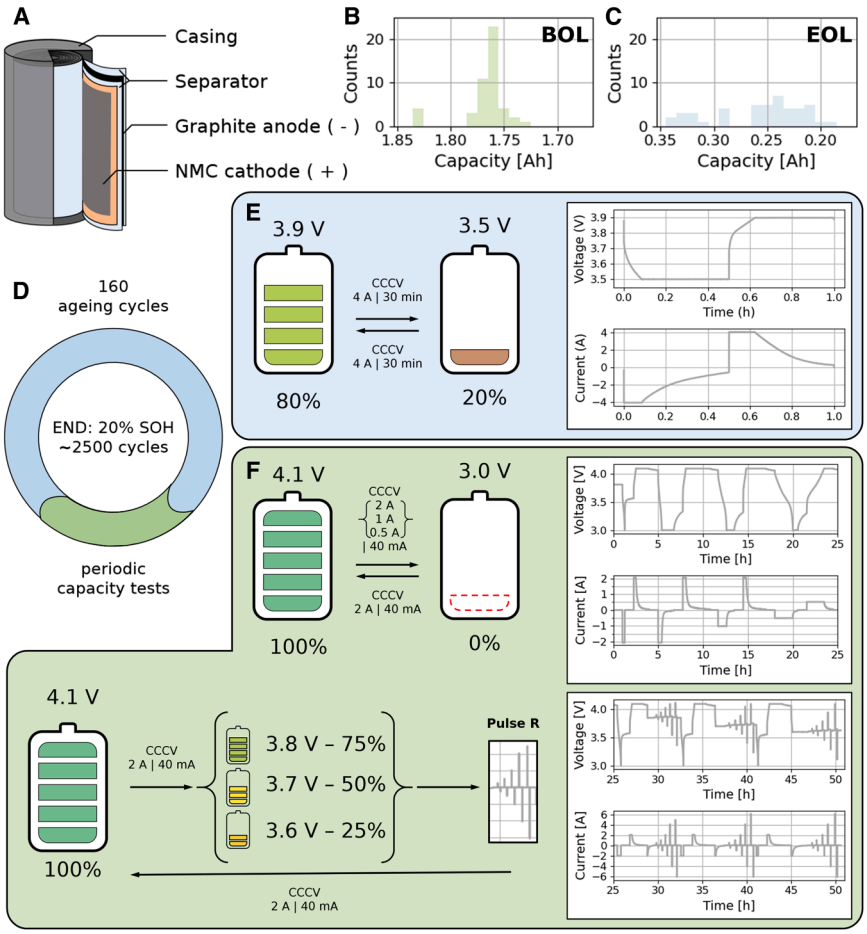


Figure 1. Dataset presentation

(A) NMC/graphite Sanyo/Panasonic UR18650E cylindrical cell cycled to collect the data.

(B) Capacity of the 48 cells at the EOL (end of life), i.e., after cycling.

(C) Capacity of the 48 cells at the BOL (beginning of life), i.e., before cycling.

(D) Cycling protocol overview. The cells were alternating a phase of 160 aging cycles with periodic capacity checks. The test ended when the aging cycling capacity reached 20% SOH (state of health) compared to the initial aging cycle capacity.

(E) Details of the aging cycling protocol: a CCCV discharge and a CCCV charge at 4 A, with a 60% DOD (depth of discharge) between 3.9 and 3.5 V.

(F) Details of the periodic capacity check protocol: a rate capability check followed by pulse resistance checks at different SOC.

features can be extracted from raw data, which adds to the processing complexity.[26]

In this work, we propose a feature extraction method based on the concept of "changepoint detection" (CPD). This method relies on the detection of the shifts in the statistical characteristics of a time series, mirroring the changes in the underlying processes within the system generating the data.[27–29] Unlike widely used feature extraction strategies that focus on geometric characteristics—such as knee points, curvature extrema, or peaks in IC and DV curves—our method does not rely on local extrema, inflection points, or predefined electrochemical markers. We demonstrate the effectiveness of the CPD for SOH prediction of lithium-ion batteries (LIBs) using a large experimental aging dataset from $LiNi_xMn_yCo_zO_2$ (NMC)|graphite cells. Our method performs well on both capacity check data—which include low-rate charge cycles over the full SOC range—and cycling data, where cells are cycled within a narrower SOC window at higher C-rates. We further employ a synthetic physics-generated aging dataset[30–32] to assess the physical interpretability of the CPD features and their potential in revealing the underlying aging mechanisms, including the loss of lithium inventory (LLI), and the loss of active material at the positive (LAM_{PE}) and negative (LAM_{NE}) electrodes.

RESULTS AND DISCUSSION

Aging dataset

We based our study on a public dataset provided by RWTH Aachen University.[33] It contains cycling data of 48 commercial cylindrical cells with a nominal capacity of 1.85 Ah. These cells are commercially available C-grade UR18650E Sanyo/Panasonic cells, composed of an NMC cathode and a graphite anode (see Figure 1A). Each cell is subjected to aging cycles, intermitted by periodic capacity checks every 160 cycles until the operating capacity (aging cycles) reaches 20% of its initial value.

The dataset is split into two subsets. The first subset is a regular cycling using a constant-current-constant-voltage (CCCV) protocol for both charge and discharge steps (see Figure 1E). During the CC phase, the current is set at ±4 A (c.a. 2.25 C) with the cut-off voltage limits of 3.9 and 3.5 V for charge and discharge, respectively, corresponding to 20% and 80% SOC, defined with respect to the initial capacity of the cell at the pristine state. During the constant-voltage (CV) step, the voltage is held at the cut-off values for so long that each CCCV step lasts exactly 30 min. Consequently, the duration of each aging cycle is 1 h, regardless of the SOH of the cell.

The periodic tests (see Figure 1F) are composed of rate-capability tests and pulse resistance tests. The former follows the same general CCCV protocol as the regular aging tests except that the depth-of-discharge (DOD) is set to 100%, i.e., on a voltage window spanning from 4.1 to 3.0 V, and the current rates during the successive CC steps are progressively lowered. In total, the rate-capability tests are composed of 4 cycles. The charge current is set at 2 A, while the discharge current for the three cycles in a row is 2, 1, and 0.5 A. The last cycle of the tests uses 0.5 A (c.a. 0.28 C) for the CC steps during both charge and discharge segments. The second part of the periodic tests is the pulse resistance tests, for which the data are not utilized in the present work.

We use two metrics for the definition of the SOH. The first metric is the capacity delivered at the CC step of the charge during the periodic capacity-check measurements at 0.5 A. These data are referred to as *capacity-check charge data* throughout this work. The second metric is the capacity that originates from the CC portion of the charge during regular cycle tests, referred to as *cycling charge data*. The exact definition of the SOH varies slightly depending on the metric used. For capacity-check charge data, SOH is defined as the ratio of the capacity to the nominal capacity of the cells (1.85 Ah), while for cycling charge data, the capacity is normalized to the first cycle for each battery cell.

CPD

To identify abrupt transitions in the battery degradation signals (e.g., voltage vs. capacity, DV, and IC curves), we employed a CPD method based on the *Pruned Exact Linear Time (PELT)* algorithm,[34] implemented via the Python ruptures package.[35] This approach segments the input signal into piecewise regions that are statistically distinct, enabling localized analysis of aging patterns across charge-discharge cycles.

The PELT algorithm minimizes a global cost function that balances segmentation fit against model complexity, effectively locating the optimal set of changepoints (CPs) in a time series. We used the *Radial Basis Function (RBF)* model as the cost function,[36] which is particularly sensitive to non-linear shifts in the signal, including changes in both the *mean*, *variance*, and higher-order structures. This makes it well-suited for electrochemical data where transitions may not manifest as simple step changes but instead as gradual or non-linear shifts in voltage or capacity behavior.

Formally, for a univariate signal $\boldsymbol{y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T\}$, we suppose there exists an unknown number, m, of ordered CPs at unknown locations, $0 = \tau_0, \tau_1, \ldots, \tau_m, \tau_{m+1} = T$, which split the data into independent segments. The PELT method identifies the optimal number and location of CPs by solving the following optimization exactly,

$$\min_{m,\{\tau_k\}} \left[\sum_{k=0}^{m} C\left(y_{\tau_k+1:\tau_{k+1}}\right) + \beta m \right],$$

to obtain the estimated locations $\widehat{\tau}_1, \ldots, \widehat{\tau}_{\widehat{m}}$ and number, $\widehat{m}$ of CPs. Here C is the RBF cost function computed for each segment, and β is a user-defined penalty term controlling the trade-off between the number of segments and model fit. In our analysis, a fixed penalty value was chosen to ensure consistent sensitivity across all datasets. See methods for further details on the RBF and PELT algorithms.

This method was applied independently to three signals.

- Voltage vs. capacity (V–Q),
- IC vs. voltage (dQ/dV–V),
- DV vs. capacity (dV/dQ–Q).

Prior to derivative-based analysis, a minimal preprocessing step was applied to mitigate the impact of experimental noise. Specifically, no filtering or smoothing was applied to the raw voltage or capacity signals. Smoothing was exclusively applied to the derivative vectors dV/dQ and dQ/dV, which are inherently more sensitive to measurement noise due to numerical differentiation. A moving-average (movmean) filter with a fixed window size of 10 data points was used for this purpose. In the present study, the experimental laboratory dataset benefits from highly controlled cycling conditions, fixed current protocols, and consistent measurement resolution, which together yield derivative curves with sufficiently strong signal structure after smoothing. However, the robustness to higher noise levels—particularly those encountered in field or onboard battery data—is a critical challenge and deserves dedicated attention in future works.

The CPD algorithm was applied separately to the capacity-check charge data and the cycling charge data, since, as discussed in the previous section, these datasets differ in both voltage range and charge current. For the capacity check data, a penalty term of 50 was selected, resulting in a maximum of 3 CPs for the voltage-capacity data, 5 CPs for the dQ/dV-voltage data, and 5 CPs for the dV/dQ-capacity data across all capacity-check cycles and battery cells (Figure S1). These numbers represent the maximum number of detected CPs per cycle, not a fixed count for every cycle. An example of the detected CPs from the capacity-check charge data is shown in the top row of Figure 2, corresponding to battery cell #10 and capacity-check charge cycle #480. For this case, 3 voltage-based, 4 dQ/dV-based, and 4 dV/dQ-based CPs were detected, as illustrated in Figures 2A, 2B, and 2C, respectively.

We also applied the CPD algorithm to the cycling charge data for all battery cells. Using a penalty term of 50 resulted in no detected CPs, which is expected due to the lower variation and limited information in the cycling charge data compared to the capacity-check data. Therefore, we reduced the penalty term to 10, which led to a maximum of 2 voltage-based, 1 dQ/dV-based, and 1 dV/dQ-based CP across the cycling charge data (Figure S2). Again, the number of detected CPs varies from cycle to cycle. An example from the cycling charge data is presented in the bottom row of Figure 2, corresponding to cell #20 and charge cycle #100. For this example, 1 voltage-based, 2 dQ/dV-based, and 1 dV/dQ-based CP were detected. In all plots of Figure 2, the detected CPs divide the data into segments, with the mean and variance of each segment displayed.

The number of detected CPs is sensitive to the parameters of the CPD model, particularly the penalty term. A preliminary sensitivity analysis was conducted to rationalize the choice of penalty terms in this study (Figure S3), while a dedicated optimization study is left for future works. A suitable penalty term is needed to reach a compromise between the number of detected CPs and stability in SOH prediction performance. For instance, although lower penalties can yield more CPs and, in the case of cycling data, lower RMSE, excessively large feature sets increase the risk of overfitting and computational cost. In this study, the detected CPs were used as features for our SOH estimation model. It is worth noting that CPD provides data segments with statistically distinct characteristics that could be comprehensively further analyzed to define aging-related features. Here, however, we focus solely on using the location of the CPs themselves as the features.

The CPD method results in a different number of detected CPs from cycle to cycle for both the capacity-check charge data and the cycling charge data. To ensure consistent interpretation of

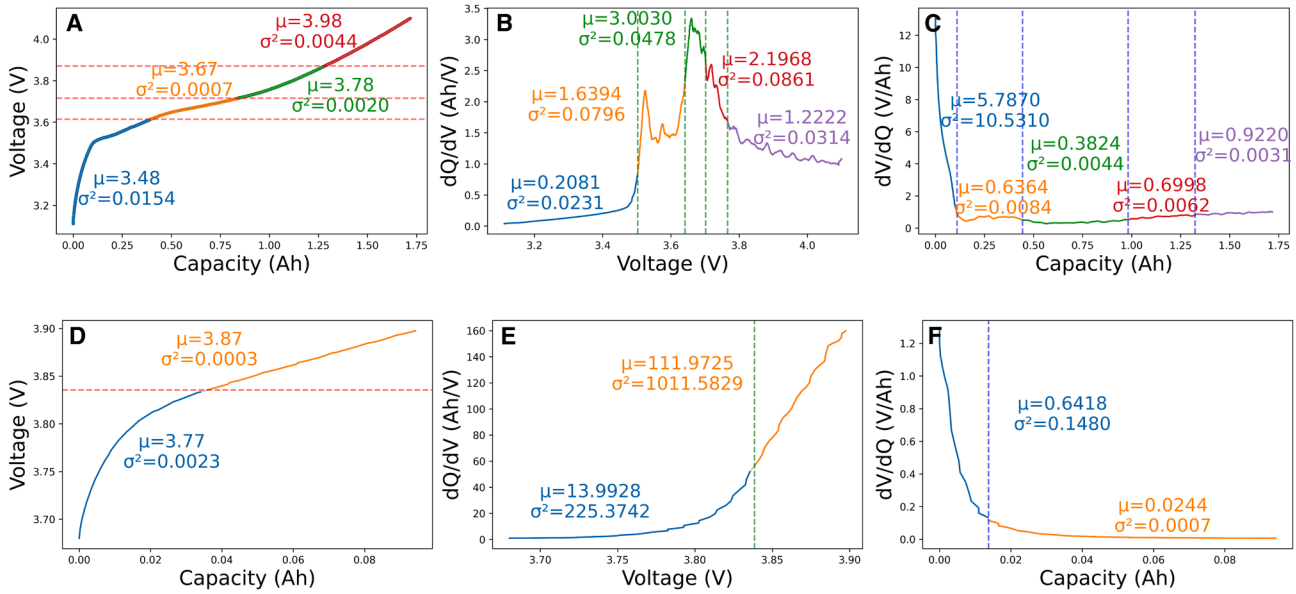


Figure 2. The identified CPs for two randomly selected battery cells

Top: capacity check data for cell number 10, cycle number 480, and (bottom) cycling data for cell number 20, cycle number 100; for (A) and (D) voltage vs. capacity; (B) and (E) incremental capacity; and (C) and (F) differential voltage data

the CPs across all cycles, a post-processing step was applied to reassign CP labels based on their numeric values rather than their original order of detection. All detected CP values were pooled and clustered using the KMeans algorithm into a predefined number of groups corresponding to the maximum number of CPs observed. The cluster centers were sorted in ascending order, and each CP was relabeled according to its cluster membership. This reassignment ensures that each CP label (e.g., cpdQdV,1, cpdQdV,2, …) consistently represents a specific transition region across all cells and cycles, facilitating meaningful comparison and analysis.

The extracted CPs are shown in Figures 3A and 3B for the capacity-check and cycling data, respectively. The capacity-check data exhibit a larger number of detected CPs, consistent with the wider voltage range. For capacity check, the number of voltage-based change points (Figure 3A) correlates with SOH: three change points for SOH >55%, two for 25% < SOH <55%, and one for SOH <25%. Near SOH ≈25%, either one or two CPs are observed. A comparable trend is found for the *dV/dQ*-based CPs from the cycling data (Figure 3B): no CPs are detected for SOH <35%, while SOH >50% consistently yields one CP, with less consistent behavior between 35% and 50%. In contrast, the *dQ/dV*- and *dV/dQ*-based CPs from the capacity-check data, as well as the voltage- and *dQ/dV*-based CPs from the cycling data, show no clear correlation with SOH.

To assess the statistical properties and relevance of the extracted CP features, a two-stage visualization approach was employed. First, the distribution of each feature was examined using kernel density estimation (KDE) plots. The results for the capacity-check charge data and the cycling charge data of all battery cells are shown in Figures 3C and 3D. As illustrated in these subplots, the detected voltage-based, dQ/dV-based, and dV/dQ-based CPs for both the capacity-check and cycling charge data resulted in distinct features, with minimal overlap between the detected CP groups.

In the capacity-check charge data (Figure 3C), the voltage-based CPs show mean transition voltages at 3.63 V (cpV,1), 3.74 V (cpV,2), and 3.89V (cpV,3). The dQ/dV-based CPs exhibit mean values ranging from 3.52 V (cpdQdV,1) to 3.90 V (cpdQdV,5), while the dV/dQ-based CPs show mean transitions between 0.12 Ah (cpdVdQ,1) and 1.27 Ah (cpdVdQ,5). In the cycling charge data (Figure 3D), the voltage-based CPs have mean values of 3.8347 V (cpV,1) and 3.8480 V (cpV,2). In the dQ/dV group, the mean values are 3.8363 V (cpdQdV,1) and 3.8528 V (cpdQdV,2), while the dV/dQ group has a single CP, cpdVdQ,1, with a mean value of 0.0129 Ah.

Following the distribution analysis, the relationship between each feature and the battery SOH was investigated through scatterplots. For each feature, SOH was plotted as a function of feature value. For the capacity-check charge data (Figure 3C), some CPs exhibit cluster-wise linear relationships with SOH—distinct clusters within a feature show approximately linear trends—while other CPs display non-linear or negligible associations. For instance, cpV,1 and cpdQdV,1 have high coefficients of determination (R^2 > 0.93), indicating strong and statistically significant correlations with SOH. In contrast, cpdQdV,2 has a negligible slope and wide confidence intervals encompassing zero, suggesting no meaningful trend. For the cycling charge data (Figure 3D), the relationships are generally more linear across all CPs. Features such as cpV,1 and cpdQdV,1 show moderate negative correlations with SOH (R^2 ≈ 0.39 and R^2 ≈ 0.56, respectively), while cpdVdQ,1 exhibits a strong positive correlation (R^2 ≈ 0.95).

ML approach

In this work, we employ the XGBoost (extreme gradient boosting) regression model to predict the SOH of LIB cells using features

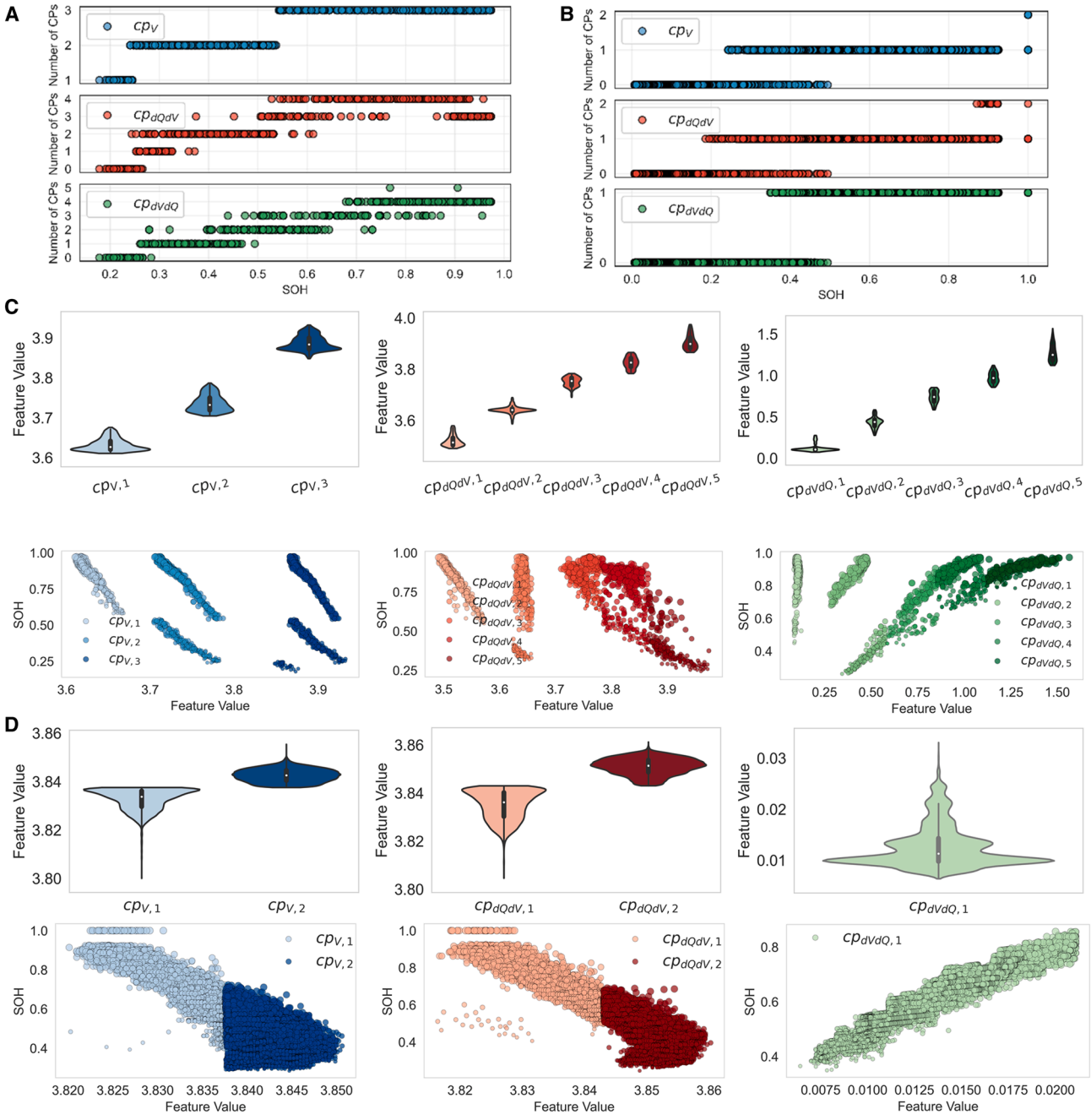


Figure 3. Statistical summary of detected changepoints in capacity check and cycling datasets

(A and B) Relationship between the number of detected changepoints and SOH

(C and D, top) Violin plots of voltage-, dQ/dV-, and dV/dQ-based changepoints. The violin width represents the kernel density estimate of the data, while the embedded boxplots indicate the median and interquartile range (25th–75th percentiles); whiskers extend to 1.5× of the interquartile range. Only observed data ranges are shown. (C and D, bottom) Scatterplots showing changepoint-SOH relationships.

derived from the CPD analysis of voltage-capacity, dQdV, and dVdQ curves. A key challenge in this approach is the variable number of CPs detected across cycles, which leads to non-uniform feature vectors under different aging conditions. Instead of discarding cycles or applying imputation, we leveraged XGBoost's inherent ability to handle such variability in the feature set natively.

XGBoost builds decision trees that learn optimal default directions for missing feature values during training. Specifically, when a feature is absent in a training sample, the algorithm learns which branch (left or right) minimizes the loss and uses that branch whenever the feature is missing in future samples. This enables the model to maintain predictive performance even

with sparse or incomplete data, without introducing bias through artificial imputation. This native handling of missing data is consistently applied in our implementation for both feature group comparisons and individual feature evaluations. Furthermore, the XGBoost model was optimized for its hyperparameters for the different cases. The optimized hyperparameter sets are reported in Table S1. This careful tuning ensures that the model achieves its best possible predictive performance for each feature group and individual feature evaluation. When comparing feature groups (i.e., voltage-based, dQ/dV-based, and dVdQ-based), the model is trained and evaluated on the full dataset, with missing feature values passed directly to XGBoost, to be internally managed. Similarly, in the evaluation of individual features within the best-performing group, the model is trained using only the single selected feature and all available cycles—again with missing values handled by the model without manual filtering or imputation.

The dataset is split into training and test subsets using a stratified sampling strategy based on the final cycle number of each cell, binned into quantiles. The split is performed on a per-cell basis, such that all cycles from a given battery cell are assigned exclusively to either the training or the testing set, thereby preventing information leakage while ensuring balanced representation of different aging stages across both subsets. Model performance is assessed using multiple regression metrics, including the coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE), providing a comprehensive view of both accuracy and variance.

In the first step, we utilized all extracted CPs—namely, those based on voltage, dQ/dV, and dV/dQ—as feature inputs to our XGBoost model. The estimated SOH values from XGBoost are plotted against the true SOH values of the test datasets in Figure 4A for the capacity-check charge data and Figure 4B for the cycling charge data.

For the capacity-check data, the model achieves an R^2 value of 0.998 and an RMSE of 0.013 on the test set. On the cycling charge data, the model shows an R^2 of 0.974 and an RMSE of 0.023. These results clearly demonstrate that the CPs extracted from the capacity-check charge data are richer in information relevant to battery degradation. Nevertheless, the model's performance on the cycling charge data remains highly promising, which highlights the practical utility of the CPD approach in real-world battery aging scenarios—particularly where capacity-check data are often unavailable.

It is important to note that during training on the cycling charge data, cycles for which no CPs were detected in any of the three feature groups were excluded from the model. This was not necessary for the capacity-check data, as each cycle in that dataset contained at least one detected CP. The predicted SOH values for five randomly selected cells from the test dataset are shown in Figures 4C and 4D, for the capacity-check charge data and the cycling charge data, respectively. The shaded areas around the predicted SOH curves represent a 5% uncertainty in the prediction. It is noteworthy that due to the large number of cycles in the cycling charge data, the SOH predictions in Figure 4D are displayed at intervals of every 50 cycles to enhance visual clarity of the model performance.

In the next step, we evaluated the performance of XGBoost using subsets of the full feature set. First, we compared the predictive power of voltage-based, dQ/dV-based, and dV/dQ-based features for both capacity check and cycling data. Furthermore, we investigated the effectiveness of individual CPs within the best-performing feature group, again for both capacity-check and cycling datasets.

Table 1 and Figure 4E compare the performance of XGBoost regression models trained using different CP feature groups from the capacity-check charge data. The model utilizing voltage-based CPs outperforms the others, achieving the lowest test MAE and RMSE values (0.013 and 0.016, respectively), as well as the highest test R^2 score (0.996). These results suggest that CPs in the voltage signal are more informative for predicting the target variable than features derived from differential capacity (dQ/dV) or DV (dV/dQ).

Table 1 and Figure 4E present the individual predictive power of selected voltage-based CP features from the capacity-check charge data. The feature cpV,1 significantly outperforms the others, achieving a test R^2 of 0.942 along with the lowest test MAE and RMSE values. In contrast, cpV,2 and cpV,3 exhibit substantially lower performance, with test R^2 values below 0.3. These results highlight that only a subset of the voltage-based CPs meaningfully contributes to model accuracy.

Table 1 compares the performance of XGBoost regression using different groups of CP features extracted from the cycling charge data. Among the feature groups, cp$dVdQ$ yields the best predictive accuracy, achieving a test R^2 of 0.958 and the lowest test MAE and RMSE values. In contrast, cpV performs poorly, with both training and test R^2 values around 0.7. These results indicate that the CPs derived from dV/dQ profiles are more informative for modeling cycling behavior.

Physics interpretability of the CPD method

The physics interpretability of the CPD features was put to test using a synthetic battery aging dataset reported by Dubarry et al.[25,31] This dataset provides more than 700,000 synthetic physics-based simulated charge/discharge cycles of an NMC|graphite cell. Each cycle represents a certain SOH caused by a unique combination of the aging progress in three different degradation tracks, namely LLI, LAM_{PE}, and LAM_{NE}. As such, this aging dataset provides a controlled environment to measure the correlations between the CPD features and the three different aging mechanisms. Such correlations are very helpful to develop degradation maps, such as those generated by Dubarry and Anseán for NMC-based cells.[25] For instance, in such a map, one can follow the evolution in the peak shapes and coordinates of the IC/DV plots in response to the progress in LLI, LAM_{PE}, and LAM_{NE} degradation modes.[25] Below, we showcase that the CPD features exhibit unique sensitivity to the LLI, LAM_{PE}, and LAM_{NE}, which enables the prediction of an underlying degradation mode based on the evolution patterns in the characteristics of the features.

We selected 10,000 representative cycles while preserving the overall SOH distribution (Figure 5A), because applying our methodology to the entire synthetic dataset would have been computationally prohibitive. The CPD was carried out using the same parameters as in the previous section, namely a RBF kernel

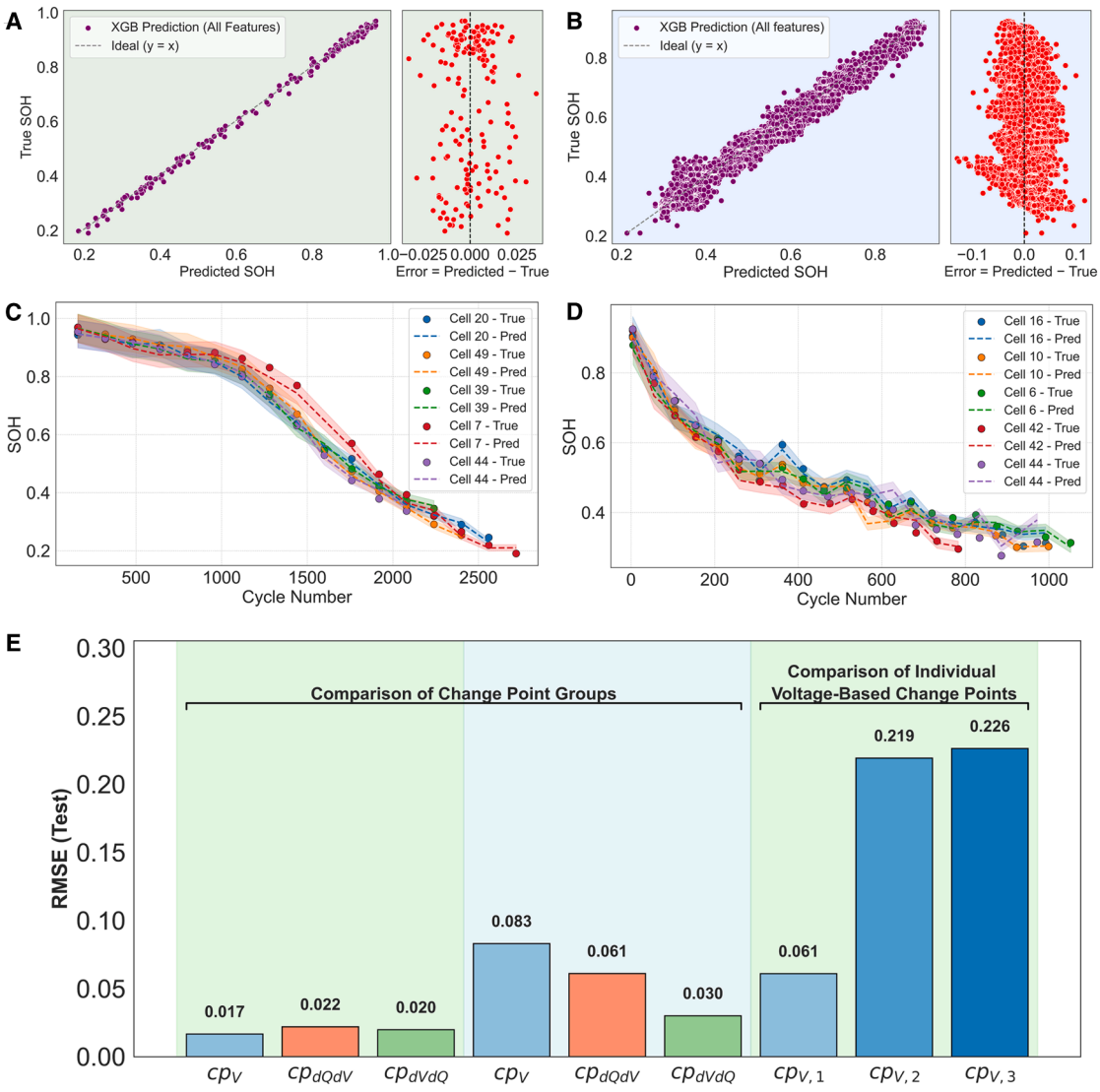


Figure 4. SOH prediction performance of XGBoost using change-point features

(A–D) Predicted vs. true SOH from the XGBoost model trained on all changepoints (voltage-, dQ/dV-, and dV/dQ-based) for (A) capacity check data and (B) cycling data. The red dashed line represents the ideal prediction (y = x). SOH trajectories for five randomly selected test cells for (C) capacity check data and (D) cycling data. Solid markers show true SOH; dashed lines indicate predictions, with shaded ±95% prediction bands.

(E) Performance comparison of XGBoost RMSE on test data for capacity check vs. cycling datasets, including individual voltage-based changepoints for the capacity check case.

and a penalty term of 50. The statistical summary of the extracted CP features shows that the detected voltage and derivative markers are well defined within narrow ranges, providing a consistent basis for correlation analysis (Figure 5B). The mean voltage CPs are concentrated between 3.5 and 4.0 V, with standard deviations in the order of 0.03–0.04 V, reflecting stable feature distributions across the cycles. For instance, cpV1 exhibits an average value of 3.56 V (σ = 0.033 V), while cpV2 and cpV3 have means of 3.78 and 4.01 V, respectively, each with comparably low variance (Figure 5B). Derivative-based features (*dQ*/*dV* and *dV*/*dQ*) displayed slightly higher variability but remained within physically reasonable ranges, suggesting robustness of the CPD process. Although missing data are present for certain features, with up to several thousand entries absent (e.g., more than 7,000 missing values for cpdQdV), the overall statistical structure is preserved (Figure 5B). The target degradation indicators, including LLI, cathode and anode loss, and SOH, show distributions consistent with expected battery aging behavior, providing a solid reference for correlation analysis.

To move beyond purely statistical associations and clarify *how* specific CPs reflect individual degradation mechanisms, we further analyzed controlled aging trajectories in which a single degradation mode dominates while the others remain negligible. Representative cycles corresponding to ~30% progression of

Table 1. Performance summary of XGBoost regression models using CP-based feature groups and individual voltage CPs

Dataset/feature type	Feature or group	MAE (training)	MAE (test)	MSE (test)	RMSE (test)	R^2 (training)	R^2 (test)
Capacity check (feature groups)	**cpV**	**0.002**	**0.013**	**0.0003**	**0.016**	**0.9999**	**0.996**
Capacity check (feature groups)	cp*dQdV*	0.002	0.016	0.0004	0.022	0.999	0.994
Capacity check (feature groups)	cp*dVdQ*	0.003	0.015	0.0004	0.0198	0.999	0.994
Capacity check (individual voltage CPs)	**cpV**,1	**0.033**	**0.042**	**0.004**	**0.061**	**0.949**	**0.942**
Capacity check (individual voltage CPs)	cpV,2	0.090	0.169	0.048	0.219	0.726	0.243
Capacity check (individual voltage CPs)	cpV,3	0.074	0.161	0.051	0.226	0.786	0.194
Cycling data (feature groups)	cpV	0.063	0.064	0.007	0.083	0.708	0.695
Cycling data (feature groups)	cp*dQdV*	0.043	0.048	0.004	0.061	0.842	0.826
Cycling data (feature groups)	**cp*dVdQ***	**0.023**	**0.025**	**0.001**	**0.030**	**0.960**	**0.958**

LLI, cathode loss, or anode loss were compared against an undegraded reference cycle. Figure S4 illustrates how the detected voltage- and derivative-based CPs shift under each isolated degradation mode, providing direct mechanistic insight into the origin of the observed correlations.

The correlation analysis reveals systematic and interpretable relationships between the detected CPs and the degradation modes (Figure 5C). For LLI, the strongest positive correlations were found with cpV1 (r = 0.464) and cpdQdV1 (r = 0.393), suggesting that early voltage and derivative-based CPs are closely linked to the depletion of lithium inventory. Strong negative correlations were observed for late-stage slope features such as cpdVdQ4 (r = −0.853) and cpdVdQ3 (r = −0.402), indicating that slope changes in the higher voltage regions of the voltage-capacity curve are inversely related to lithium loss. Cathode degradation exhibits a similar structure, with positive correlations for cpV1 (r = 0.393) and cpdVdQ1 (r = 0.338), while several high-voltage derivative features correlate negatively, including cpdQdV4 (r = −0.554) and cpdVdQ4 (r = −0.527). Anode loss is most strongly reflected in the voltage-based features, with cpV1 (r = 0.565), cpV3 (r = 0.529), and cpV2 (r = 0.417) showing clear positive associations. Negative correlations for anode loss are weaker and primarily found in derivative-based CPs such as cpdVdQ4 (r = −0.425) and cpdVdQ3 (r = −0.212). SOH displays the strongest correlations with the target features, with late-stage slope features cpdVdQ4 (r = 0.875) and cpdVdQ3 (r = 0.852) emerging as highly predictive indicators for global capacity fade, while the early stage features, such as cpV1 (r = −0.587) and cpdQdV1 (r = −0.430) show inverse correlations with the SOH. These trends are further visualized in Figure S5, which shows the continuous evolution of individual CP features as a function of each degradation mode and SOH across the full synthetic aging space.

The statistical and correlation patterns identified in our work are consistent with the degradation maps reported by Dubarry.[25] In their terminology, the relevant positions, peaks, or slopes in the voltage and derivative curves are referred to as features of interest (FOIs). These FOIs occupy the same mid-to-high voltage range (3.5–4.0 V) as our detected CPs and shift in predictable directions according to the degradation mechanisms. Particularly, Dubarry showed that LLI shifts certain low-voltage FOIs toward higher potentials, LAM_{PE} reduces or broadens peaks and can shift FOIs to lower voltages, and LAM_{NE} alters FOIs around graphitic staging transitions near 3.7–3.9 V, often changing relative intensities or separating nearby features. Our results align with this mapping: positive correlations of cpV1 (r = 0.464) and cpdQdV1 (r = 0.393) with LLI, coupled with strong negative correlations of late-slope features such as cpdVdQ,4 (r = −0.853), mirror the directional FOI shifts predicted by Dubarry for LLI-dominated degradation. Similarly, the positive correlation of cpV1 and cpdVdQ1 with cathode loss, and the negative correlations of higher-voltage derivative CPs, are consistent with Dubarry's description of LAM_{PE}-induced reductions in peak intensity and voltage shifts in positive-electrode FOIs. For anode loss, the dominant correlations with cpV1, cpV2, and cpV3 fall in the same voltage region where Dubarry's maps indicate the sensitivity of FOIs to LAM_{NE}, supporting the interpretation that graphitic staging transitions govern these relationships. Finally, the exceptionally strong correlations between SOH and late-curve slope features cpdVdQ4 and cpdVdQ3 align with Dubarry's observation that such FOIs respond simultaneously to multiple degradation mechanisms, making them integrative markers of overall capacity fade.

In conclusion, we introduced a CPD algorithm as an effective feature extraction methodology for the prediction of battery SOH. We applied this method to two types of experimental datasets from graphite/NMC cells differing in the cycling voltage window, namely the "cycling" (3.5–3.9 V) and "capacity check" (3–4.1 V). The detected CPs were categorized as voltage-based, dQdV-based, and dVdQ-based features. An ML model (XGBoost) was trained on 75% of the extracted features to estimate the SOH. A promising accuracy was achieved for the SOH estimation, with RMSE values of **0.013** for the capacity-check data and **0.023** for cycling data. The proposed lightweight data-driven model is computationally efficient: for instance, in the analyzed cycling dataset of 114,963 cycles across all battery cells, the CPD model extracted features in 1,765 s, corresponding to an average of ∼15 ms per cycle. This demonstrates that the model is fast, making its deployment feasible for onboard or real-time applications. As a preliminary concept, it could be integrated into embedded BMS or edge devices to process streaming voltage and current data and continuously update SOH estimates.

A key finding of this study is that many cycles from "cycling" dataset exhibited no detectable CPs across feature groups when using a standard penalty (e.g., 10), particularly at low SOH (SOH <0.35, see Figure S6), whereas "capacity-check" cycles consistently revealed CPs. This indicates that, as cells age, capacity checks become increasingly crucial for a reliable SOH

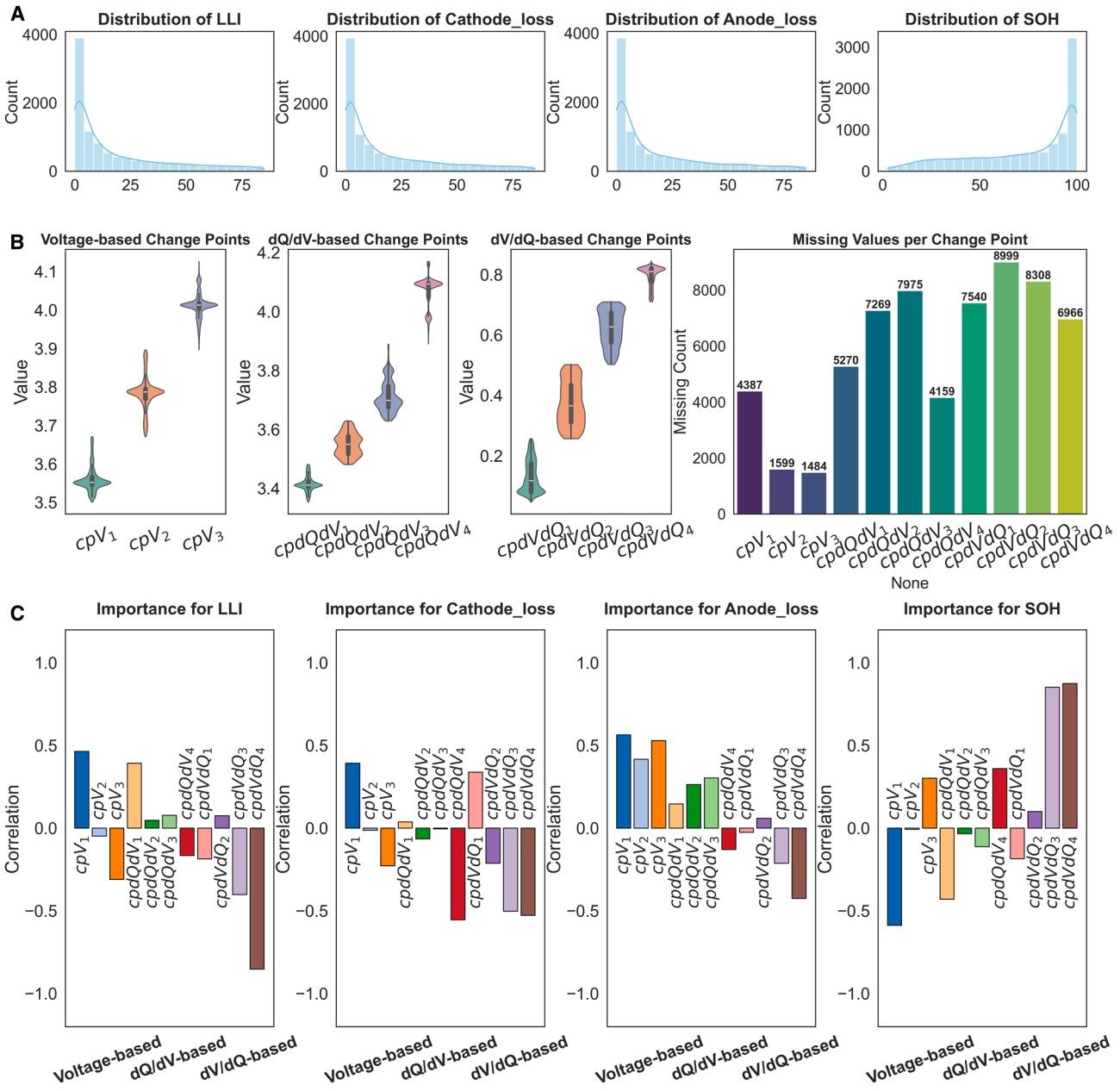


Figure 5. Summary of changepoint analysis for synthetic NMC battery data from the work of Dubarry and Beck

(A) Distributions of target variables, illustrating their range and density across all samples.

(B) Violin plots showing the distribution of voltage-based, dQ/dV-based, and dV/dQ-based change-point features across cycles. The violin width represents the kernel density estimate of the data within the observed range, while embedded boxplots indicate the median and interquartile range (25th–75th percentiles), with whiskers extending to 1.5× the interquartile range. A bar plot summarizes the number of missing values per feature, highlighting both the variability of each change point and the completeness of the dataset.

(C) Feature importance for each target, quantified as the Pearson correlation between features and targets. Bars are grouped by feature type, with annotations indicating individual feature names.

prediction. The analysis further showed that detection is sensitive to the choice of penalty: reducing the penalty (e.g., to 5) enabled CP identification in cycling data across nearly the full SOH range. This observation is consistent with the work of Haynes et al.,[37] which highlights that reliance on a single penalty value can be suboptimal. Accordingly, the absence of CPs under certain conditions may reflect suboptimal parameter selection, emphasizing the need for systematic optimization of penalty-kernel settings to ensure robust detection across the battery lifespan.

We further evaluated the performance of individual CP feature groups. For the capacity-check data, **voltage-based CPs** exhibited the highest predictive power, especially the feature

cpV,1, which emerged as the most informative. This indicates that voltage data alone, when collected under appropriate conditions, can provide critical insights into battery degradation—potentially reducing the need for more complex characterization techniques such as electrochemical impedance spectroscopy (EIS). Moreover, voltage-based features are advantageous over derivative-based features (dQ/dV or dV/dQ), as they are less sensitive to noise, device precision, and preprocessing complexities such as filtering and smoothing. In contrast, for **cycling data**, voltage-based CPs performed poorly. The best SOH predictions were obtained using **dV/dQ-based** CPs. This highlights the need to focus on voltage-capacity derivatives when working with high-rate, partial-window cycling data where subtle variations are better captured through differential analysis.

We investigated the physical interpretability of the CPD features using a large-scale synthetic aging dataset of the NMC/graphite cell. Our analysis showed that the detected CPs correlate systematically with the important category of losses within the cell, namely the loss of lithium and the active-material loss at the positive and negative electrodes. In particular, late-stage slope features exhibited the strongest correlations with SOH, demonstrating why these CPs serve as excellent predictors of global capacity fade in NMC-based batteries. This analysis highlights the added value of the physics-generated synthetic data for **transfer learning** approaches where models pre-trained with synthetic datasets are adapted to real-world experimental data to substantially enhance the robustness and accuracy of SOH estimation.

Here, we presented a proof of concept enabling the pursuit of several avenues in future works. First, the use of a comprehensive experimental aging dataset capturing cell-to-cell variations across a broad range of operating conditions would allow for a more thorough assessment of the approach. Second, the applicability of the proposed method across different battery chemistries, including lithium-iron-phosphate (LFP) cells with distinct open-circuit-voltage characteristics, calls for further study. Lastly, given the demonstrated computational efficiency of the method on aging data, future efforts will focus on validating the model under varying operational conditions, evaluating its robustness against measurement noise, and testing its integration within BMS architectures to assess real-time and onboard feasibility.

METHODS

Technical overview of the RBF-PELT algorithm

For a univariate signal $y = y_{1:T} = \{y_1, y_2, \ldots, y_T\}$, we suppose there exists an unknown number, m, of ordered CPs at unknown locations, $0 = \tau_0, \tau_1, \ldots, \tau_m, \tau_{m+1} = T$ that split the data into independent segments. That is, for distinct distributions F_i,

$$y_t \sim \begin{cases} F_1 & \text{if } 0 < t \leq \tau_1 \\ F_1 & \text{if } \tau_1 < t \leq \tau_2 \\ & \vdots \\ F_{m+1} & \text{if } \tau_m < t \leq T \end{cases}$$

To identify the existence and location of CPs, we need a measure of fit for a segment of data to an assumed distributional form. In our work, we use the RBF[36] with kernel,

$$K(y_s, y_t) = \exp\left(-\gamma \| y_s - y_t \|^2\right),$$

where γ is a bandwidth parameter estimated as the inverse of the median of all pairwise distances over s and t. The cost (measure of fit for a segment) is then defined via the feature map, $\Phi(y_t)$, for this kernel,

$$C(y_{s:t}) = \sum_{i=s}^{t} \| \Phi(y_i) - \widehat{\mu} \|_{\mathcal{H}}^2 .$$

Here $\widehat{\mu}$ is the fitted mean of the embedded signal $\{\Phi(y_t)\}_{s:t}$ and the norm is taken over an appropriate Hilbert space $\mathcal{H}$.

To identify CPs, we seek to minimize the total cost of fit across all segments. If we consider this directly, the optimal fit is a change at each location, as the cost would be equal to zero. We seek a parsimonious balance between the number of CPs and the overall fit to the data. As is common in statistical modeling, we introduce a penalty β to the existence of each CP to ensure parsimony. This results in seeking to optimize,

$$\min_{m,\{\tau_k\}} \left[\sum_{k=0}^{m} C\left(y_{\tau_k+1:\tau_{k+1}}\right) + \beta m \right], \qquad \text{(Equation 1)}$$

for $\{\tau\}$ and m.

Calculating the optimal of Equation 1 directly would require 2^T calculations. Fortunately, we can use dynamic programming to reformulate the problem as a sequential search from 1, …,T and focus on identifying the final CP only (with prior CPs identified earlier in the algorithm). Furthermore, pruning the search for the potential last CP locations brings us to an $\mathcal{O}(T)$ computational algorithm. This is the essence of the PELT algorithm, which optimizes Equation 1 exactly via the recursion,

$$F(1) = C(y_1);$$
$$F(t) = min_{s \in R_t}[F(s) + C(y_{(s+1):t})] + \beta.$$

Here, R_t is the pruned set of past indices, where a CP could potentially occur, updated according to

$$R_t = \{t-1\} \cup \{s \in R_{t-1} : F(s) + C(y_{(s+1):t}) < F(t) + \beta\},$$

Intuitively, if we have gone past a true CP location in the iteration, the "last" CP location cannot be prior to the true CP. Thus, R_t should only contain time points at, or since, the last true CP, rather than the entire past. If CPs occur regularly through the data, the computational gain from pruning can be substantial.

Technical overview of the XGBoost algorithm

XGBoost is an ensemble learning method based on gradient-boosted decision trees. It minimizes a regularized objective function that balances model fit and complexity:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t),$$

where.

- y_i: true target for sample i
- $\hat{y}_i^{(t-1)}$: prediction at iteration $t-1$
- f_t: decision tree added at iteration t
- $\Omega(f_t)$: regularization term to prevent overfitting

The algorithm uses a second-order Taylor expansion of the loss function to compute optimal splits, with gain:

$$\text{Gain} = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] - \gamma,$$

where

- G_L,G_R: sum of gradients on the left and right of the split
- H_L,H_R: sum of Hessians on the left and right of the split
- λ,γ: regularization terms

In this work, XGBoost was used in its regression form ("reg:squarederror"), leveraging its native handling of missing features and regularization to predict the SOH of lithium-ion cells.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Mohammadhossein Safari (momo.safari@uhasselt.be).

Materials availability

This study did not generate any unique material.

Data and code availability

- The aging data used in this study are publicly available from RWTH University under the DOI: https://doi.org/10.18154/RWTH-2021-04545
- The source code and processed data have been deposited on GitHub at https://github.com/HamidHamed1990/battery-changepoint-analysis. The associated DOI is https://doi.org/10.5281/zenodo.18451242.
- All data are available in the main text and supplemental information. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support from FWO-Vlaanderen. H.H. is a junior postdoctoral fellow (12A1R24N) of the Research Foundation Flanders (FWO-Vlaanderen).

AUTHOR CONTRIBUTIONS

Conceptualization, H.H., A.C.R., and M.S.; writing – original draft, H.H., A.C.R., and R.K.; writing – review & editing, H.H., B.G.C., Q.P., A.C.R., R.K., and M.S.; visualization, H.H., B.G.C., A.C.R., R.L., Q.P., and M.S.; supervision, R.K. and M.S.; and funding acquisition, H.H. and M.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrp.2026.103157.

Received: November 17, 2025
Revised: December 22, 2025
Accepted: February 4, 2026
Published: March 10, 2026

REFERENCES

1. Finegan, D.P., Zhu, J., Feng, X., Keyser, M., Ulmefors, M., Li, W., Bazant, M.Z., and Cooper, S.J. (2021). The application of data-driven methods and physics-based learning for improving battery safety. Joule *5*, 316–329.
2. Sulzer, V., Mohtat, P., Aitio, A., Lee, S., Yeh, Y.T., Steinbacher, F., Khan, M.U., Lee, J.W., Siegel, J.B., Stefanopoulou, A.G., and Howey, D.A. (2021). The challenge and opportunity of battery lifetime prediction from field data. Joule *5*, 1934–1955.
3. Roman, D., Saxena, S., Robu, V., Pecht, M., and Flynn, D. (2021). Machine learning pipeline for battery state-of-health estimation. Nat. Mach. Intell. *3*, 447–456.
4. Xie, Y., Wang, S., Zhang, G., Takyi-Aninakwa, P., Fernandez, C., and Blaabjerg, F. (2024). A review of data-driven whole-life state of health prediction for lithium-ion batteries: Data preprocessing, aging characteristics, algorithms, and future challenges. J. Energy Chem. *97*, 630–649.
5. Oji, T., Zhou, Y., Ci, S., Kang, F., Chen, X., and Liu, X. (2021). Data-driven methods for battery soh estimation: Survey and a critical analysis. IEEE Access *9*, 126903–126916.
6. Wei, J., Dong, G., and Chen, Z. (2018). Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression. IEEE Trans. Ind. Electron. *65*, 5634–5643.
7. Wang, F., Zhai, Z., Zhao, Z., Di, Y., and Chen, X. (2024). Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis. Nat. Commun. *15*, 4332.
8. Ren, L., Zhao, L., Hong, S., Zhao, S., Wang, H., and Zhang, L. (2018). Remaining useful life prediction for lithium-ion battery: A deep learning approach. IEEE Access *6*, 50587–50598.
9. Yang, D., Zhang, X., Pan, R., Wang, Y., and Chen, Z. (2018). A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve. J. Power Sources *384*, 387–395.
10. You, G.-w., Park, S., and Oh, D. (2016). Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach. Applied energy *176*, 92–103.
11. van Vlijmen, B., Lam, V.N., Asinger, P.A., Cui, X., Schaeffer, J., Geslin, A., Ganapathi, D., Sun, S., Herring, P.K., and Gopal, C.B. (2025). Aging matrix visualizes complexity of battery aging across hundreds of cycling protocols. Energy Environ. Sci. *18*, 6641–6654.
12. Xiong, R., Sun, Y., Wang, C., Tian, J., Chen, X., Li, H., and Zhang, Q. (2023). A data-driven method for extracting aging features to accurately predict the battery health. Energy Storage Mater. *57*, 460–470.
13. Ospina Agudelo, B., Zamboni, W., and Monmasson, E. (2021). Application domain extension of incremental capacity-based battery SoH indicators. Energy *234*, 121224.
14. Dubarry, M., and Beck, D. (2022). Perspective on mechanistic modeling of Li-ion batteries. Acc. Mater. Res. *3*, 843–853.
15. Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., et al. (2019). Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy *4*, 383–391.
16. Mayilvahanan, K.S., Takeuchi, K.J., Takeuchi, E.S., Marschilok, A.C., and West, A.C. (2022). Supervised learning of synthetic big data for Li-ion battery degradation diagnosis. Batter. Supercaps *5*, e202100166.
17. Li, X., Ju, L., Geng, G., and Jiang, Q. (2023). Data-driven state-of-health estimation for lithium-ion battery based on aging features. Energy *274*, 127378.
18. Chinomona, B., Chung, C., Chang, L.-K., Su, W.-C., and Tsai, M.-C. (2020). Long short-term memory approach to estimate battery remaining useful life using partial data. IEEE Access *8*, 165419–165431.
19. Sylvestrin, G.R., Maciel, J.N., Amorim, M.L.M., Carmo, J.P., Afonso, J.A., Lopes, S.F., and Ando Junior, O.H. (2025). State of the Art in Electric Batteries' State-of-Health (SoH) Estimation with Machine Learning: A Review. Energies *18*, 746.

20. Zhang, Y., Tang, Q., Zhang, Y., Wang, J., Stimming, U., and Lee, A.A. (2020). Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. Nat. Commun. *11*, 1706.

21. Kumar Thakur, A., Sathyamurthy, R., Velraj, R., Saidur, R., Pandey, A.K., Ma, Z., Singh, P., Hazra, S.K., Wafa Sharshir, S., Prabakaran, R., et al. (2023). A state-of-the art review on advancing battery thermal management systems for fast-charging. Appl. Therm. Eng. *226*, 120303.

22. Wang, Y., Zhang, X., Li, K., Zhao, G., and Chen, Z. (2023). Perspectives and Challenges for Future Lithium-Ion Battery Control and Management (Elsevier).

23. Suganya, R., Joseph, L.M.I.L., and Kollem, S. (2024). Understanding lithium-ion battery management systems in electric vehicles: Environmental and health impacts, comparative study, and future trends: A review. Results Eng. *24*, 103047.

24. Wei, Z., Ruan, H., Li, Y., Li, J., Zhang, C., and He, H. (2022). Multistage state of health estimation of lithium-ion battery with high tolerance to heavily partial charging. IEEE Trans. Power Electron. *37*, 7432–7442.

25. Dubarry, M., and Anseán, D. (2022). Best practices for incremental capacity analysis. Front. Energy Res. *10*, 1023555.

26. Li, Y., Abdel-Monem, M., Gopalakrishnan, R., Berecibar, M., Nanini-Maury, E., Omar, N., van den Bossche, P., and Van Mierlo, J. (2018). A quick on-line state of health estimation method for Li-ion battery with incremental capacity curves processed by Gaussian filter. J. Power Sources *373*, 40–53.

27. Lund, R.B., Beaulieu, C., Killick, R., Lu, Q., and Shi, X. (2023). Good practices and common pitfalls in climate time series changepoint techniques: A review. J. Clim. *36*, 8041–8057.

28. Killick, R., Eckley, I.A., Ewans, K., and Jonathan, P. (2010). Detection of changes in variance of oceanographic time-series using changepoint analysis. Ocean Engineering *37*, 1120–1126.

29. Adams, R.P., and MacKay, D.J. (2007). Bayesian online changepoint detection. arXiv. https://doi.org/10.48550/arXiv.0710.3742.

30. Li, R., Kirkaldy, N.D., Oehler, F.F., Marinescu, M., Offer, G.J., and O'Kane, S.E. (2025). The importance of degradation mode analysis in parameterising lifetime prediction models of lithium-ion battery degradation. Nat. Commun. *16*, 2776.

31. Dubarry, M., and Beck, D. (2020). Big data training data for artificial intelligence-based Li-ion diagnosis and prognosis. J. Power Sources *479*, 228806.

32. Baure, G., and Dubarry, M. (2019). Synthetic vs. real driving cycles: A comparison of electric vehicle battery degradation. Batteries *5*, 42.

33. Li, W., Sengupta, N., Dechent, P., Howey, D., Annaswamy, A., and Sauer, D.U. (2021). One-shot battery degradation trajectory prediction with deep learning. J. Power Sources *506*, 230024.

34. Killick, R., Fearnhead, P., and Eckley, I.A. (2012). Optimal detection of changepoints with a linear computational cost. J. Am. Stat. Assoc. *107*, 1590–1598.

35. Truong, C., Oudre, L., and Vayatis, N. (2020). Selective review of offline change point detection methods. Signal Process. *167*, 107299.

36. Arlot, S., Celisse, A., and Harchaoui, Z. (2019). A kernel multiple changepoint algorithm via model selection. Journal of machine learning research *20*, 1–56.

37. Haynes, K., Eckley, I.A., and Fearnhead, P. (2017). Computationally efficient changepoint detection for a range of penalties. J. Comput. Graph Stat. *26*, 134–143.