energies

MDPI

Article

A GBRT-Based State-of-Health Estimation Method for Lithium-Ion Batteries

Chun Chang [1], Yedong He [1], Yutong Wu [1], Yuanzhong Xu [1] and Jiuchun Jiang [1,2,*]

[1] Hubei Key Laboratory for High-Efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan 430068, China
[2] Zhuhai Campus, Beijing Institute of Technology, Zhuhai 519088, China
* Correspondence: jcjianghubt@163.com

Abstract

Lithium-ion batteries are widely applied in transportation, communication, and other fields. Nevertheless, during prolonged cycling operation, internal electrochemical reactions inevitably lead to the degradation of the state-of-health (SOH). To ensure the reliability and safety of lithium-ion batteries, accurate SOH estimation is of critical importance. Nevertheless, under practical operating conditions, obtaining fully recorded charge–discharge data is often impractical. Motivated by the practical charging behaviors of lithium-ion batteries, this paper proposes a practical SOH estimation method based on incremental capacity analysis, dynamic time warping (DTW), and gradient-boosting regression trees (GBRTs). Three health indicators—interval incremental capacity features, local capacity–voltage curve similarity, and segmented voltage curve similarity—are extracted. The proposed method requires only 0.13 V and 0.07 V voltage windows on the Oxford and CALCE datasets. The effectiveness of the proposed model is verified across both public datasets and laboratory test data. Experimental results demonstrate RMSE values of approximately 2.5% and 2.0%, respectively. Compared with mainstream SOH estimation algorithms, the proposed approach delivers comparable accuracy while achieving training time reductions of up to 57.6% and 91.9% relative to GPR and SVM, making it suitable for real-time battery management systems.

Keywords: lithium-ion batteries; state of health; health indicators; machine learning

Check for updates

Received: 28 December 2025
Revised: 22 January 2026
Accepted: 25 January 2026
Published: 27 January 2026

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Lithium-ion batteries exhibit advantages including high energy density, strong low-temperature adaptability, high operational reliability, and extended cycle life [1]; ensuring safe and reliable operation often requires health monitoring [2,3]. During long-term operation, irreversible electrochemical degradation leads to a continuous decline in the state-of-health (SOH), which directly affects the available capacity and operational safety. Accurate SOH estimation is therefore a fundamental requirement for battery management systems (BMSs) to enable reliable operation, timely maintenance, and early warning of safety risks [4,5]. The battery health state can generally be directly measured through a complete diagnostic cycle under low-current conditions [6]. However, in practical applications, batteries are typically operated under irregular charging behaviors and varying operating conditions, making it difficult to obtain complete charge–discharge data. This mismatch between the data requirements of many existing SOH estimation methods and the realities of practical battery usage poses a critical challenge for the development of reliable and

deployable SOH estimation techniques. Therefore, accurately predicting battery SOH in narrow-data-window conditions is of great importance.

Regarding lithium battery state-of-health (SOH) estimation, existing research predominantly employs model-based and data-driven approaches [7–9]. Model-based methods utilize lithium battery material properties, operating environmental conditions, and aging/degradation mechanisms to construct battery models. These methods estimate battery capacity by simulating actual operating conditions through variations in-model parameters,. Specific approaches include electrochemical models [10–12], degradation mechanism models [13], empirical models [7] and equivalent circuit models [14–16]. Model-based methods integrate physical and electrochemical knowledge to analyze the underlying degradation mechanisms of lithium-ion batteries (LIBs) or to interpret current, voltage, and electrochemical impedance spectroscopy (EIS) measurements. Mathematical or physical equations are then formulated to describe the general degradation behavior of LIBs. Reference [17] proposes a lithium-ion battery capacity estimation method based on a particle filter, in which an electrochemical model is integrated into the particle filtering framework. Model parameters reflecting battery degradation are treated as state variables, and measurement and process model equations are formulated to fit the capacity evolution process and known state variables, thereby constructing a capacity estimation model. Reference [18] analyzes the relationship between capacity loss and cycle number, as well as temperature, fitting a power–law relationship between capacity loss and cycling time and an Arrhenius relationship between capacity loss and temperature. Based on this analysis, a lithium-ion battery capacity estimation method combining a semi-empirical model with conventional calibration is proposed. Although model-based methods achieve considerable success in LIB diagnosis, constructing accurate mathematical or physical models remains challenging, as it requires comprehensive knowledge of the internal states and degradation mechanisms of lithium-ion batteries. In contrast, data-driven methods do not require explicit modeling of the complex internal electrochemical mechanisms of batteries. Instead, they directly extract and analyze information related to battery health from operational monitoring data collected during battery aging and cycling. Various machine learning and deep learning algorithms are then employed for SOH estimation [19], including support vector machines [20], neural networks [21,22], and gaussian process regression [23].

Data-driven lithium-ion battery SOH estimation methods generally begin with the analysis of battery operational data to extract highly sensitive health indicators that can effectively characterize battery aging and degradation. The effectiveness of health indicator extraction has a significant impact on both the estimation accuracy and practical applicability of the SOH estimation method. Reference [24] extracts complete voltage, current, and power curves during lithium battery charging and discharging. It utilizes the Douglas–Peucker algorithm to extract health indicators, which are then input into an SVM model optimized by the Differential Brainstorming with Optimization (DBSO) algorithm for lithium battery SOH estimation. Reference [25] extracts health indicators such as voltage, current, and time from constant-current and constant-voltage charging phases during a complete lithium battery charge cycle. It employs a whale optimization algorithm-enhanced Elman model for SOH estimation. Reference [26] proposes a data-driven state-of-health (SOH) estimation method based on multiple charge–discharge features. In this method, several health indicators related to voltage, time, and capacity are extracted from both charging and discharging processes and are fused to construct a supervised regression model for SOH estimation. Reference [27] adopts a data-driven online capacity estimation approach based on random forest regression and incremental capacity analysis is employed for the feature selection. The usually rely on broad voltage ranges and relatively high-dimensional feature sets. References [28,29] investigated SOH

estimation using gaussian process regression (GPR) based on charging-curve features. The model achieves accurate SOH estimation by analyzing the characteristics of the voltage curve during constant current charging. More recently, interpretable and data-efficient machine learning frameworks have attracted increasing attention. Reference [30] proposed an interpretable machine learning-based battery diagnosis approach, demonstrating that carefully designed health indicators can enable accurate health assessment under practical constraints. Reference [31] provided a comprehensive review of machine learning applications in lithium-ion battery research, highlighting both the potential and the challenges of data-driven health monitoring under diverse operating conditions.

Despite the progress in data-driven and model-based SOH estimation, most existing methods either rely on full charge–discharge curves that are rarely available under practical operating conditions, or employ complex models with a high computational cost and limited interpretability. In addition, the selection health indicators are often heuristic. To address these issues, this paper proposes an SOH estimation framework that integrates incremental capacity analysis, curve similarity metrics, and gradient-boosting regression trees. The main work includes the following:

(1) Statistically analyzing real charging behavior to identify feasible and frequently occurring voltage windows for health indicator extraction.
(2) Objectively selecting SOH-sensitive voltage windows through Pearson correlation analysis; a compact yet informative set of health indicators is constructed, enabling accurate SOH estimation using only very narrow voltage segments.
(3) The GBRT algorithm model established in this paper can achieve the estimation accuracy of complex algorithm models and the training time is relatively short.

The organization of this paper is as follows. Section 2 introduces the data and the methodology for extracting health indicators. Section 3 presents the SOH estimation methodology. Section 4 discusses the estimation results, and Section 5 concludes the paper.

2. Extraction of Health Indicators for Lithium-Ion Batteries

2.1. Incremental Capacity–Based Health Indicator Extraction Method

The current mainstream approach for estimating the state-of-health (SOH) of lithium-ion batteries utilizes operational data collected during charging and discharging processes, primarily including voltage, current, and temperature. These data are used to extract health indicators that characterize battery aging and degradation, which are subsequently employed for algorithmic SOH estimation. In practical applications, lithium-ion battery charge–discharge data are influenced by user behavior and rarely contain complete charging or discharging cycles. Therefore, it is essential to analyze real-world operational data to identify the most commonly occurring charge and discharge intervals during actual usage. Extracting health indicators within these representative intervals enables the development of SOH estimation algorithms that more accurately reflect practical operating conditions. In real-world operational data, the discharging process is highly affected by user behavior, whereas the charging process is primarily governed by predefined charging control strategies implemented in the battery management system (BMS), resulting in more regular and structured charging data. Accordingly, this study analyzes the distributions of state of charge (SOC) and voltage at the start and end of the charging process.

Since the subsequent analysis focuses on individual battery cells, the voltage of a single cell within the battery pack is selected as the representative voltage for the current state. At the beginning of charging, the highest cell voltage within the pack is selected, while all other cells exhibit lower voltages. This ensures that the charging process of each cell includes the charging data segment corresponding to the voltage level. Similarly, at the end of charging, the lowest cell voltage within the pack is selected as the representative value.

(1) State analysis at the start of charging:

The distribution of battery states at the start of charging is shown in Figure 1. The central region presents the initial state points of each charging segment. The upper histogram illustrates the SOC distribution, while the right histogram shows the voltage distribution. To quantitatively characterize these distributions, the median values of SOC and voltage are calculated. The results indicate that the median SOC at the start of charging is 0.43 and the median cell voltage is 3.680 V. Selecting the median values ensures enough charging segments for subsequent analysis, covers a wide range of practical charging conditions, and thereby improves the robustness and accuracy of SOH prediction.

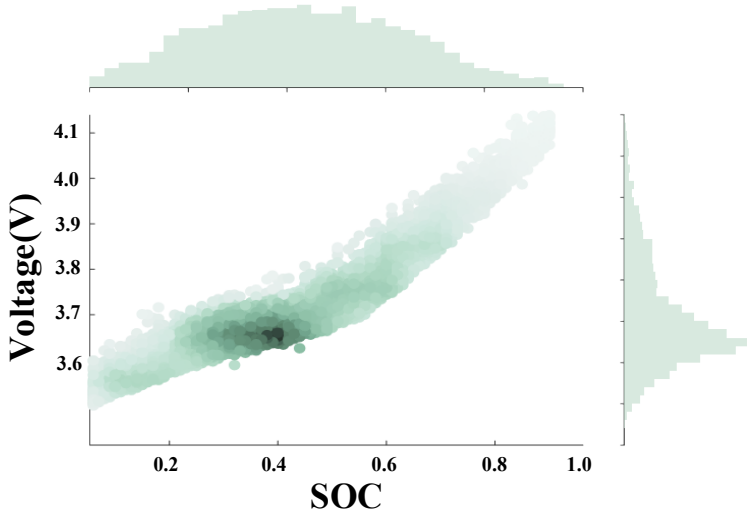


Figure 1. State distribution at the start of charging.

(2) State analysis at the end of charging:

The battery state distribution at the end of charging is shown in Figure 2. It can be observed that, at the end of the charging process, the battery SOC is mainly concentrated in the range of 0.98–1.00, while the battery voltage is primarily distributed between 4.20 V and 4.26 V. By calculating the median values of SOC and voltage, the median SOC at the end of charging is determined to be 0.97, and the median battery voltage is 4.200 V.

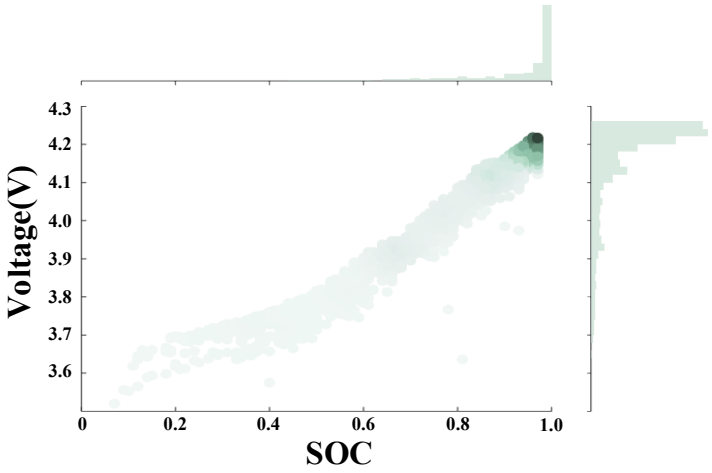


Figure 2. State distribution at the end of charging.

In summary, the charging segments most frequently observed during practical operation fall within the SOC range of 0.43–0.97. Accordingly, the subsequent analysis focuses on charging data within this SOC interval.

(3) Incremental capacity analysis of lithium-ion batteries:

The incremental capacity indicators reflect changes in electrochemical reaction kinetics and loss of active material. Degradation phenomena such as SEI growth, lithium inventory loss, and electrode structural changes directly alter the shape and magnitude of incremental capacity within specific voltage ranges.

This analysis, as well as the selection of the voltage window and the derivation of the health indicator, were all conducted entirely using the training data. The laboratory data provide complete charge–discharge cycles, enabling accurate and reliable determination of battery capacity. Furthermore, the high resolution of the data allows for detailed examination of the charge–discharge curves, facilitating the identification of sensitive health indicators that correlate closely with capacity variations.

This study first conducts incremental capacity analysis on the Oxford dataset, which includes a total of eight Kokam (The lithium-ion cells used were sourced from Kokam Co., Ltd. SLPB533459H4, Siheung-si, Gyeonggi-do, Republic of Korea) 740 mAh lithium-ion pouch cells [32]. Considering subsequent practical applications, and given that conventional electric vehicles are equipped with a limited number of temperature sensors rather than individual sensors for each cell, the temperature of each cell cannot be directly monitored. Therefore, only the voltage, current, and time data are used for the analysis of the subsequent dataset.

The analysis first performs incremental capacity analysis on the selected charging segment of this study (SOC in the range of 0.43–0.97, voltage in the range of 3.81–4.16 V) at 0.01 V intervals, as shown in Figure 3. The correlation between incremental capacity at each voltage interval and SOH is then calculated using the Pearson correlation coefficient, with the results presented in Figure 4a.

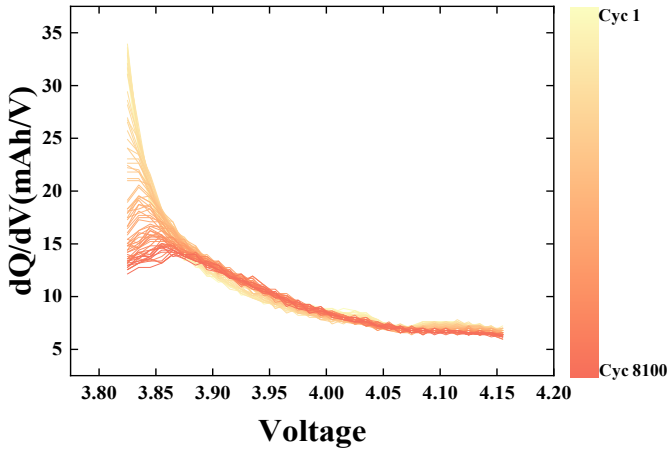


Figure 3. Results of incremental capacity analysis.

The region where the absolute value of the Pearson correlation coefficient exceeds 0.8 can be divided into three voltage intervals: 3.81–3.84 V, 3.93–3.96 V, and 4.07–4.14 V. Taking the QV curve as an example, these three voltage intervals are compared with the complete charging process as shown in Figure 4b. The gray curve represents the full QV curve of a single charging cycle. The green section indicates the practical health indicators extraction interval selected using lithium battery usage data analysis, while the red section highlights the three high-correlation health indicator extraction intervals identified in this section. It is evident that the health indicator extraction method designed in this paper requires only a minimal portion of the charging curve (a total voltage range of 0.13 V) on the Oxford dataset to achieve high-sensitivity health indicator extraction. Simultaneously, the voltage

range capacity increment across these three voltage intervals is calculated as the first-class health indicator.

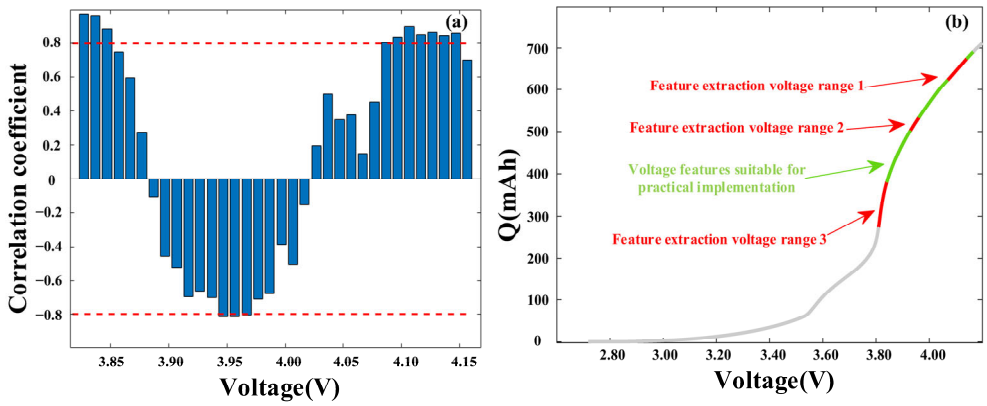


Figure 4. (**a**) Pearson correlation coefficients between the incremental capacity of each voltage interval and SOH. (**b**) Voltage interval comparison chart.

The second battery dataset used in this study is the cycle test dataset from the Center for Advanced Life Cycle Engineering at the University of Maryland (CALCE dataset). The CALCE dataset employed here comprises six lithium-ion batteries with $LiCoO_2$ cathodes and a nominal capacity of 1.35 Ah [33].

Incremental capacity analysis is first performed on the selected charging segment of this study (SOC in the range of 0.43–0.97, voltage in the range of 3.98–4.20 V) at 0.01 V intervals. Since the battery current is adjusted when the voltage reaches 4.20 V in this dataset, to avoid polarization effects caused by the current change, 4.20 V is not used as the endpoint for segment selection; instead, 4.19 V is chosen as the starting point for the subsequent charging segment. The Pearson correlation coefficient between the change in incremental capacity in each voltage interval and the battery SOH is then calculated. As shown in Figure 5a, the voltage interval where the absolute value of the Pearson correlation coefficient exceeds 0.8 is 4.10–4.17 V. A comparison with the complete charging curve is presented in Figure 5b. It is evident that, for the CALCE dataset, a voltage interval of only 0.07 V is sufficient for extracting the health indicators. The capacity increment within this voltage interval is calculated and used as the first-class health indicator.

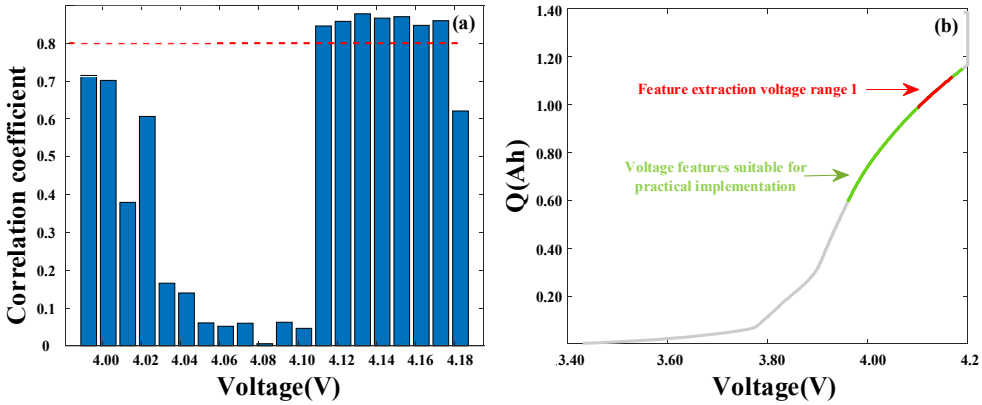


Figure 5. (**a**) Pearson correlation coefficients between the incremental capacity of each voltage interval and SOH. (**b**) Voltage interval comparison chart.

2.2. Health Indicator Extraction Method Based on DTW Distance

The local capacitance–voltage (Q–V) indicators quantify deviations in local Q–V relationships caused by aging-induced polarization increase and reaction heterogeneity. Reduced similarity indicates shifts in equilibrium potentials and increased internal re-

sistance. The segment voltage indicators quantify deviations in local voltage–capacity relationships caused by aging-induced polarization increase. Reduced similarity indicates shifts in equilibrium potentials and increased internal resistance.

This section extracts two health indicators: local Q–V curve similarity and segment voltage curve similarity. Both indicators are calculated using the dynamic time warping (DTW) distance.

Due to the effects of long-term battery aging, battery capacity gradually decreases. Consequently, under constant-current charging conditions, voltage curves with the same initial voltage and the same number of sampling points correspond to different voltage intervals across different charge–discharge cycles; that is, while the initial voltage remains the same, the cutoff voltage varies. Voltage curves from smaller voltage ranges should therefore be compared with corresponding portions of curves from larger voltage ranges when calculating similarity. Traditional curve similarity methods, such as Euclidean distance, cannot scale these sequential voltage signals, resulting in coarse similarity assessments. Therefore, the dynamic time warping (DTW) algorithm is employed for curve similarity computation.

Taking two voltage signals V_1 and V_2 with extracted lengths m as an example, the specific calculation process is as follows:

1. Matrix Construction: First, construct an $m \times m$ matrix corresponding to the two voltage signals of length m. The element (i, j) in the matrix represents the distance $d(V_{1,i}, V_{2,j})$ between points $V_{1,i}$ and $V_{2,j}$. Here, the Euclidean distance is employed. Each element within the matrix can also be interpreted as the distortion between the two points. By tracing a path from the bottom-left to the top-right of the matrix, we identify the points where the two voltage signals can be aligned.
2. Path Selection: The selected path within the matrix is defined as W, representing the mapping relationship between V_1 and V_2. This path selection must satisfy three constraints: (1) Boundary condition: $w_1 = (1, 1)$, indicating that the warping path begins at the lower-left corner and terminates at the upper-right corner. (2) Continuity condition: If $w_{k-1} = (a', b')$, the subsequent point $w_k = (a, b)$ must satisfy $a - a' \leq 1$ and $b - b' \leq 1$. This prohibits skipping points within the matrix for subsequent matching; each point can only align with its adjacent points. This ensures every coordinate in V1 and V2 appears in W. (3) Monotonicity: If $w_{k-1} = (a', b')$, then the next path point $w_k = (a, b)$ must satisfy $(a - a') \geq 0$ and $(b - b') \geq 0$. This constrains path selection to be time monotonic. Combining these three constraints, each selected path point in the matrix can only be chosen in one of three directions. Specifically, if the path has already passed through path point (i, j), the next selectable path point can only be one of the following three: $(i + 1, j)$, $(i, j + 1)$, or $(i + 1, j + 1)$.
3. DTW Distance Calculation: Based on the path selection constraints described above, all possible selection scenarios are traversed to calculate the warping cost. The minimum value among these costs is then taken as the DTW distance between the two voltage signals.

2.3. Construction of Health Indicator Parameter Set for Lithium-Ion Batteries

In summary, the process of incremental capacity analysis and health indicator parameter set construction first employs the SOC range commonly observed during actual lithium battery operation to perform incremental capacity analysis, thereby identifying voltage intervals for extracting highly correlated health indicators. Within these extracted voltage intervals, the capacity increment is obtained. Subsequently, local Q–V curve similarity and segment voltage curve similarity are calculated based on the DTW distance. The three types of health indicators obtained through this procedure collectively form the health indicator parameter set.

3. GBRT-Based Lithium Battery SOH Estimation Method

3.1. Principles of the GBRT Algorithm

Gradient-boost regression tree (GBRT) is an iterative classification tree algorithm composed of multiple regression trees, where the conclusions of all trees are aggregated to produce the final answer. Due to its strong generalization capabilities, it has garnered significant attention in recent years [34].

The core of the GBRT algorithm model lies in the fact that each regression tree is not trained on the same dataset, but instead learns from the residuals of the previous regression trees. This enables the model to progressively reduce estimation errors through multiple regression trees.

The GBRT algorithm model construction process is shown in Figure 6, with the specific steps as follows:

1. Train a regression tree using the input dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_n, y_n)\}$
2. Calculate the gradient for the health indicator x_i, used in the m-th regression tree.

$$r_{m,i} = -[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}]_{f(x)=f_{m-1}(x)} \quad (1)$$

3. Based on the computed $r_{m,i}$ values, a new regression tree is fitted to obtain the leaf node region $R_{m,j}$ for the m-th regression tree, where $j = 1, 2, \ldots, J$.
4. For each node region $j = 1, 2, \ldots, J$, compute the output value $R_{m,j}$ on each region.

$$c_{m,j} = \arg\min \sum_{x_i \in R_{m,j}} L(y_i, f_{m-1}(x_i) + c) \quad (2)$$

Here, L denotes the loss function, which represents the difference between the model's computed value and the training label value.

5. Update to form a new regression tree.

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J} c_{m,j} \quad (3)$$

6. Combine the trained regression trees to obtain the GBRT algorithm model.

The GBRT model was implemented using decision trees as weak learners. The maximum tree depth was limited to 3–5 to avoid overfitting, the learning rate was set to 0.05–0.1, and the number of estimators ranged from 100 to 300. These hyperparameters were selected using a grid search combined with k-fold cross-validation on the training dataset, with RMSE as the optimization criterion. Compared to other regression algorithms, the GBRT algorithm model can flexibly handle various types of data, including continuous and discrete values. Compared to simpler algorithm models like decision trees, the GBRT algorithm model achieves higher prediction accuracy by leveraging the collaboration of multiple regression trees. Compared to complex algorithm models such as neural networks, the GBRT algorithm can significantly reduce the time required for model training while achieving comparable accuracy.

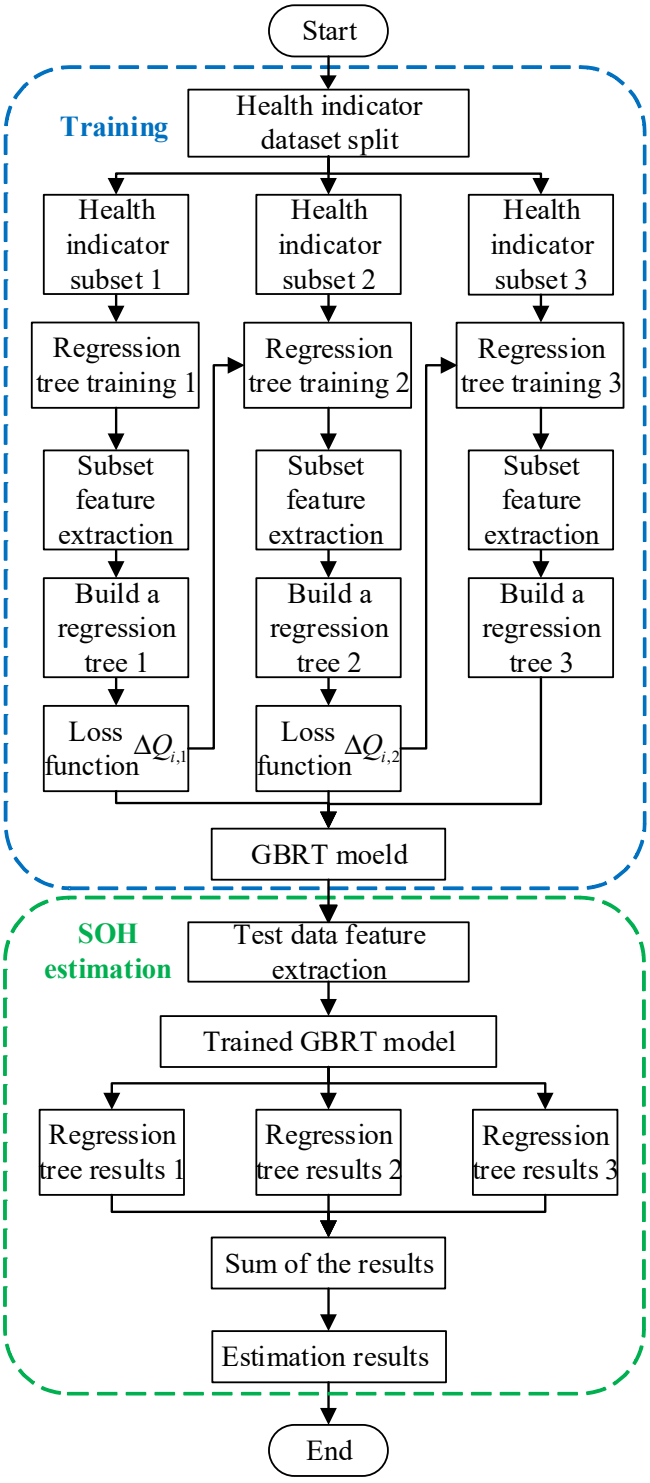


Figure 6. Flowchart of GBRT model construction and application.

3.2. SOH Estimation Model Method Design for Lithium-Ion Batteries

In this section, GBRT model structures are designed based on the health indicator dataset structures of the Oxford and CALCE datasets, as illustrated in Figure 7.

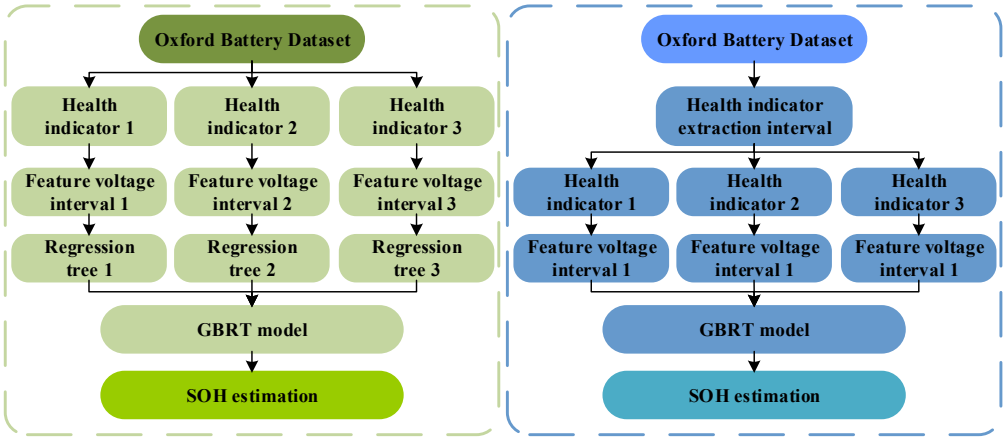


Figure 7. Schematic diagram of GBRT estimation model structure and estimation process.

(1) Structural design of the SOH estimation model for the Oxford dataset

The health indicators in the Oxford dataset are distributed across three voltage ranges: 3.81–3.84 V, 3.93–3.96 V, and 4.07–4.14 V. Therefore, based on the GBRT algorithm, one regression tree is trained for each voltage range, resulting in a total of three regression trees that collectively form the GBRT model.

(2) Structure design of the SOH estimation model for the CALCE dataset

A total of three health indicators are extracted from the CALCE dataset, all within the same voltage range. Therefore, a separate regression tree is trained for each of the three health indicators, which are then combined into a GBRT algorithm model.

4. Experimental Analysis and Discussion

4.1. Lithium Battery Health Indicator Extraction and Correlation Analysis

Correlation analysis is performed on the three types of health indicators extracted from the Oxford and CALCE datasets. Health indicators with an absolute Pearson correlation coefficient greater than 0.8 are selected for SOH estimation.

4.1.1. Health Indicator Extraction and Correlation Analysis of the Oxford Dataset

The Oxford dataset extracted three health indicator voltage intervals. The calculated parameters—capacity increment, local QV curve similarity, and segment voltage curve similarity—for these three health indicator voltage intervals, along with their correlation coefficients with battery SOH, are shown in Tables 1–3.

Table 1. Pearson correlation coefficients for capacity increments across voltage ranges.

Voltage Interval	Pearson Correlation Coefficient	*p*-Value
3.81~3.84 V	0.9876	0
3.93~3.96 V	−0.9012	0.1359×10^{-27}
4.07~4.14 V	0.9768	0

Table 2. Pearson correlation coefficients for local QV curves.

Voltage Interval	Pearson Correlation Coefficient	*p*-Value
3.81~3.84 V	−0.9479	1.6674×10^{-38}
3.93~3.96 V	−0.4415	6.5351×10^{-5}
4.07~4.14 V	−0.5805	3.8733×10^{-8}

Table 3. Pearson correlation coefficients for fragment voltage.

Voltage Interval	Pearson Correlation Coefficient	*p*-Value
3.81~3.84 V	−0.9720	2.8284×10^{-48}
3.93~3.96 V	−0.8236	6.6878×10^{-20}
4.07~4.14 V	−0.9530	3.9706×10^{-40}

Combining the three categories of health indicators extracted from the respective intervals described above, this dataset utilizes a total of seven lithium battery charging process health indicators, as shown in Table 4.

Table 4. Summary of health indicators in the Oxford dataset.

Voltage Interval	Health Indicator Name	Correlation Coefficient	Sign
3.81~3.84 V	Incremental capacity	0.9876	ΔQ_1
	Curve similarity	−0.9720	$d_{v,1}$
	QV curve similarity	−0.9479	$d_{Qv,1}$
3.93~3.96 V	Incremental capacity	−0.9012	ΔQ_2
	Curve similarity	−0.8236	$d_{v,2}$
4.07~4.14 V	Incremental capacity	0.9768	ΔQ_3
	Curve similarity	−0.9530	$d_{v,3}$

4.1.2. Health Indicator Extraction and Correlation Analysis of the CALCE and Oxford Dataset

The computational methods for extracting health indicators in the CALCE dataset are identical to those used in the Oxford dataset, with the results presented in Table 5.

Table 5. Summary of health indicators in the CALCE Dataset.

Health Indicator Name	Correlation Coefficient	Sign
Incremental capacity	0.9377	ΔQ
Curve similarity	−0.9097	d_{Qv}
QV curve similarity	−0.8019	d_v

4.2. SOH Estimation Experimental Results and Performance Analysis

4.2.1. SOH Estimation Results for the Oxford Dataset

The Oxford dataset comprises a total of eight batteries. Four batteries are randomly selected to form the model training set, designated Cell 1, Cell 2, Cell 3, and Cell 7. The remaining four batteries constitute the model test set, designated as Cell 4, Cell 5, Cell 6, and Cell 8.

The GBRT model's SOH estimation performance across four batteries is shown in Table 6. This study employs five performance metrics: Root Mean Square Error (*RMSE*), Mean Absolute Error (*MAE*), Mean Absolute Percentage Error (*MAPE*), Mean Square Error (*MSE*), and R^2. The corresponding formulas are presented in Equation (4).

Table 6. Summary of estimation errors for the Oxford dataset.

Cell Number	Cell 4	Cell 5	Cell 6	Cell 8
RMSE (%)	1.73	3.08	3.00	2.01
MAE (%)	0.62	0.90	1.05	0.75
MAPE (%)	0.69	0.98	1.15	0.88
MSE (%)	0.07	0.13	0.12	0.13
R^2	0.9831	0.9388	0.9414	0.9812

$$\begin{cases} RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \\ MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \\ MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \\ MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \\ R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\overline{y}_i - y_i)^2} \end{cases} \quad (4)$$

Here, $\hat{y}_i$ denotes the SOH estimate for the i-th cycle, y_i denotes the measured SOH value for the i-th cycle, and $\overline{y}_i$ denotes the average of the measured SOH values for the i-th cycle.

The SOH estimation results for four test cells in the Oxford dataset are shown in Figure 8. It can be observed that the GBRT model achieves accurate SOH estimation for all four cells, with RMSE values below 3.08% and R^2 values above 0.93. Considering the results across all four cells, the model performs slightly worse during the initial stages of battery operation compared to the mid-to-late stages. After the first few data points with relatively larger errors, the model provides accurate SOH estimation throughout subsequent charge–discharge cycles. The relatively larger errors observed in the first few data points are attributed to the early stage of battery life, during which degradation-induced electrochemical changes are weak and the incremental capacity features exhibit only limited variation across cycles.

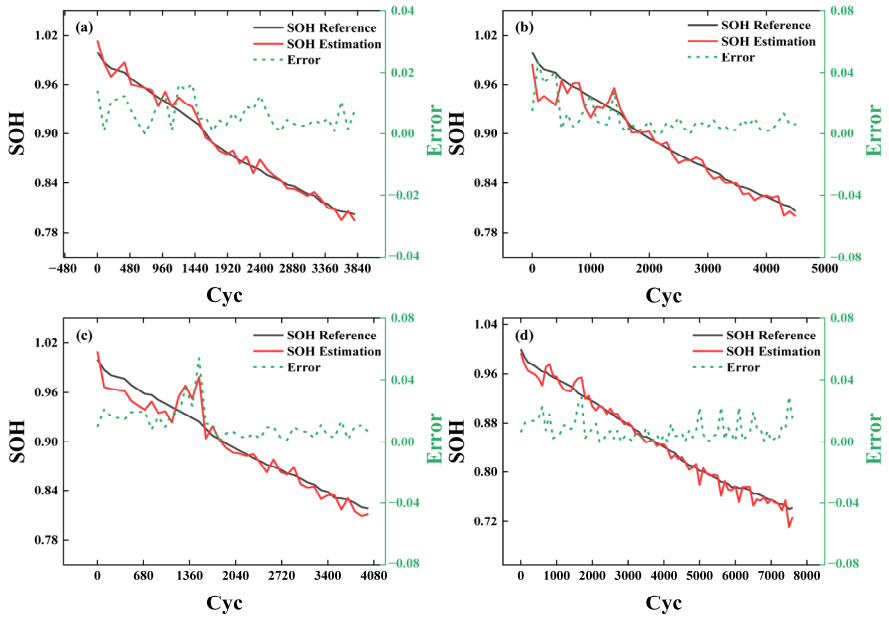


Figure 8. SOH estimation results and RMSE distribution for (**a**) Cell 4, (**b**) Cell 5, (**c**) Cell 6, and (**d**) Cell 8 in the Oxford dataset.

4.2.2. SOH Estimation Results for the CALCE Dataset

The CALCE dataset used in this paper comprises six batteries. Three batteries—Cell b, Cell g, and Cell h—were randomly selected as training data. The remaining three batteries—Cell d, Cell e, and Cell f—served as test data. The SOH estimation results for the three test batteries are shown in Figure 9 and Table 7. Considering that several estimation errors exhibit similar patterns, Table 7 only presents the RMSE and R^2 values.

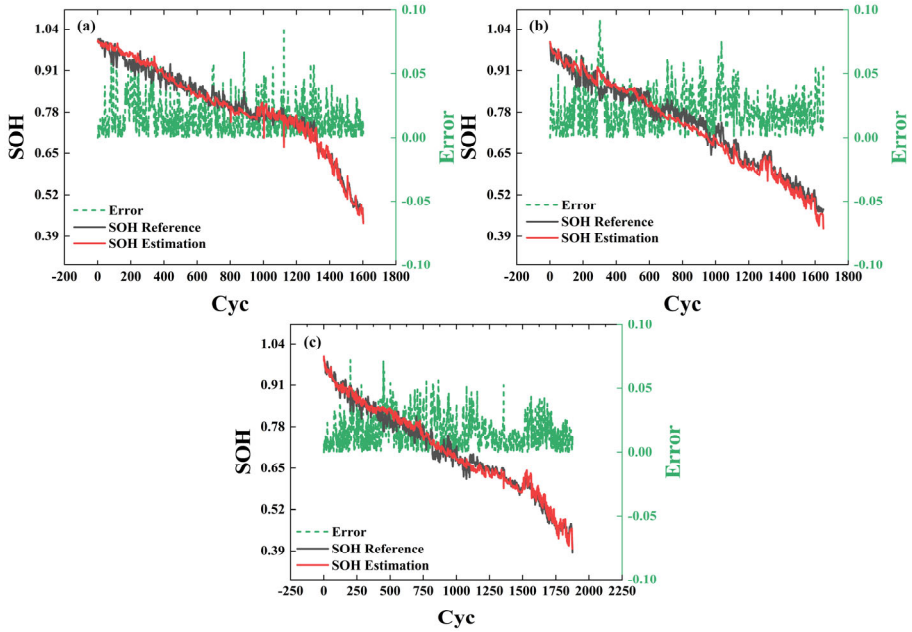


Figure 9. SOH estimation results and RMSE distribution for (**a**) Cell d, (**b**) Cell e, and (**c**) Cell f in the CALCE dataset.

Table 7. Summary of estimation errors for the CALCE dataset.

Cell Number	Cell d	Cell e	Cell f
RMSE (%)	2.08	2.64	2.02
R^2	0.9758	0.9666	0.9780

Due to the large number of sampling points in the CALCE dataset itself, which exhibits significant fluctuations and noise, coupled with the narrow extraction range for the health indicator (0.07 V voltage interval), training the GBRT model still achieves high SOH estimation accuracy. The RMSE for all three batteries remains around 0.02, with R^2 values exceeding 0.96. Cell d exhibits a pronounced capacity regeneration phenomenon between the 1000th and 1300th cycles. Within this interval, the battery SOH still correlates with the health indicators extracted in this study, and the model is able to accurately estimate SOH even in the capacity regeneration region. This behavior is mainly attributed to capacity regeneration phenomena and relaxation effects, which may temporarily mask long-term degradation trends. Under such conditions, the extracted health indicators exhibit reduced sensitivity to SOH, leading to increased estimation uncertainty. In noisy laboratory environments, measurement fluctuations further exacerbate this issue by introducing spurious variations into incremental capacity-based features.

4.2.3. Laboratory Data SOH Estimation Results

All laboratory data used in this study are obtained in the laboratory. The battery degradation data are collected from three lithium iron phosphate (LFP) batteries with a voltage range of 2–3.65 V and a nominal capacity of 10 Ah. To acquire degradation data, the batteries undergo cycling tests under three different operating conditions, as detailed in Table 8. Among the three batteries, two are randomly selected as training data, denoted as Cell exp1 and Cell exp2, while the remaining battery, Cell exp3, is used as test data.

Table 8. Laboratory condition design.

Cell	Cycling Condition			Number	Periodic Measurement
	Temp	Rate	Cyc		
exp1	45 °C	0.5 C	120	1	Periodic measurements were conducted every 10 cycles.
exp2	45 °C	0.75 C	120	1	
exp3	45 °C	1 C	120	1	

Incremental capacity analysis is first performed on the selected charging segment of this study (SOC in the range of 0.43–0.97, voltage in the range of 3.42–3.56 V) at 0.01 V intervals. The laboratory battery dataset is similar to the CALCE dataset described in Section 2.3. The Pearson correlation coefficient between the capacity change in each 0.01 V voltage interval and battery SOH is then calculated. As shown in Figure 10a, the voltage interval where the absolute Pearson correlation coefficient exceeds 0.8 is 3.44–3.56 V. A comparison with the complete charging curve is presented in Figure 10b. It is evident that, for the laboratory dataset, a voltage interval of only 0.12 V is sufficient for extracting the health indicators. The capacity increment within this voltage interval is calculated and used as the first-class health indicator.

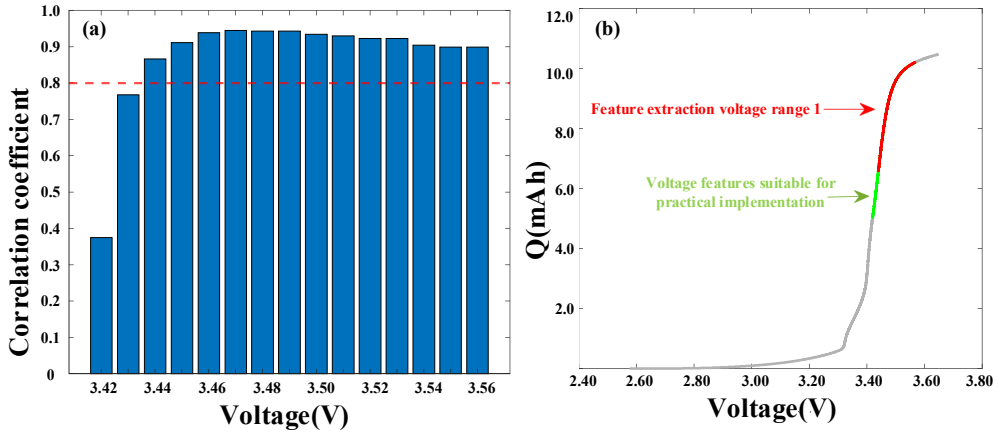


Figure 10. (**a**) Pearson correlation coefficients. (**b**) Voltage interval comparison chart.

The health indicators extracted from the laboratory data are identical to those extracted from the CALCE dataset, totaling three indicators within the same voltage range. Therefore, a separate regression tree was trained for each of the three health indicators, which were then combined into a GBRT algorithm model.

The SOH estimation results for the tested batteries are shown in Figure 10 and Table 9. Considering that several estimation errors exhibit similar patterns, Table 9 only uses the RMSE and R^2.

Table 9. Summary of estimation errors for laboratory datasets.

Cell Number	Cell exp1
RMSE (%)	1.98
R^2	0.8539

The SOH estimation results for the test cell in the laboratory dataset are shown in Figure 11. Due to the relatively high number of sampling points, larger noise, limited data volume, and the occurrence of capacity regeneration in the later cycles, the dataset is more challenging. Nevertheless, the GBRT model achieves accurate SOH estimation

for this test cell, with an RMSE below 2.0% and an R^2 above 0.83. Considering the SOH estimation results across all three datasets, the GBRT model performs slightly worse during the initial stages of battery operation compared to the mid-to-late stages. After the first few data points with relatively larger errors, the model provides accurate SOH estimation throughout subsequent charge–discharge cycles, demonstrating both its accuracy and practical applicability.

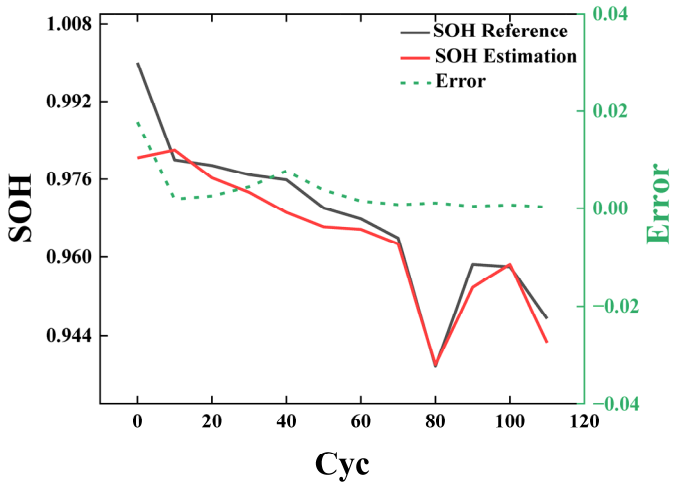


Figure 11. SOH estimation results for Cell exp 3 in the laboratory dataset.

4.3. Comparative Analysis with Other Methods

Taking the Oxford dataset as an example, this section compares the GBRT algorithm model used in this paper with commonly used SOH estimation algorithms, including linear regression (LR), a support vector machine (SVM), gradient-boosted trees (GBT), gaussian process regression (GPR), and back-propagation (BP) neural networks—a total of six algorithm models. A comparative analysis of the training time and prediction accuracy for each algorithm model is presented.

The training time of each SOH estimation model is recorded, as shown in Figure 12. The GBRT model, GBT model, and linear regression model have relatively simple structures and therefore require shorter training times; for the Oxford dataset with limited data volume, all three models complete training in less than one second. In contrast, the gaussian process regression model, BP neural network model, and SVM model have higher structural complexity, resulting in training times exceeding 2 s.

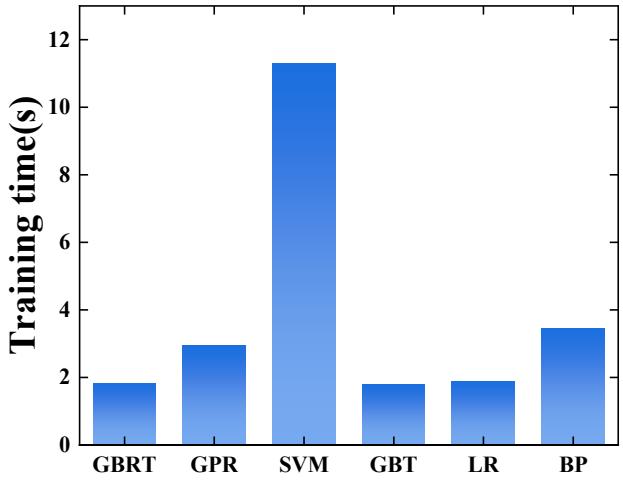


Figure 12. Model training time comparison.

The estimation performance of various SOH estimation models on four test batteries is shown in Figure 13. Taking the RMSE as an example, the estimation errors across the four batteries are illustrated in Figure 14. It can be observed that the GBRT model achieved estimation accuracy comparable to the optimal model on all four batteries.

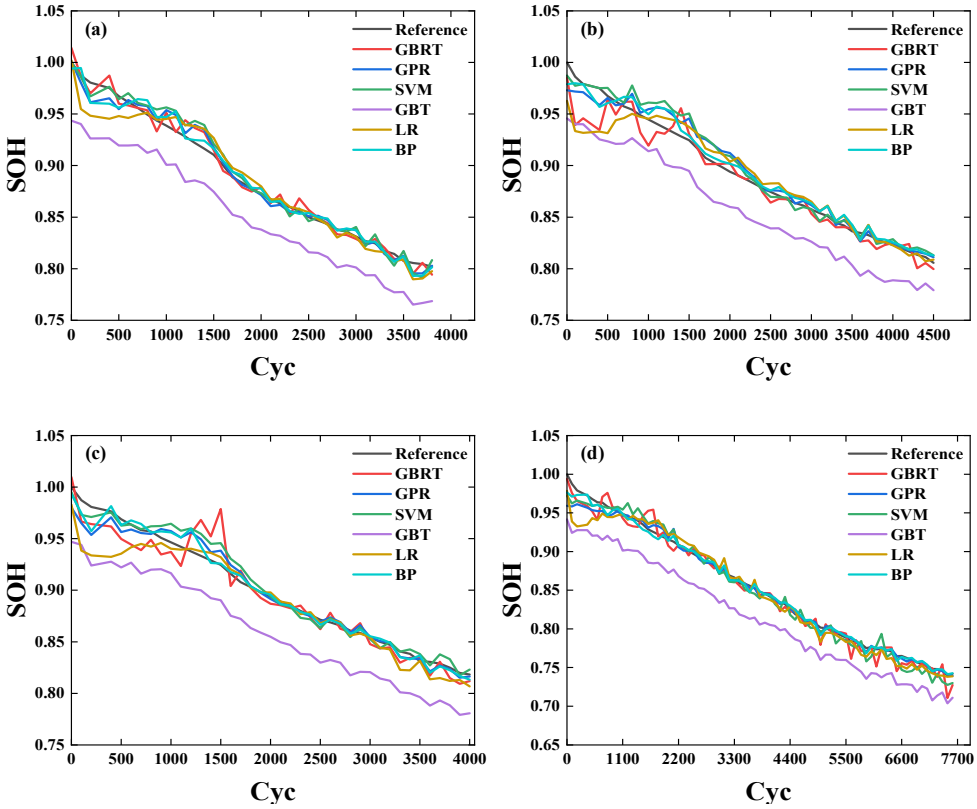


Figure 13. Model estimation results for (**a**) Cell 4, (**b**) Cell 5, (**c**) Cell 6, and (**d**) Cell 8 in the Oxford dataset.

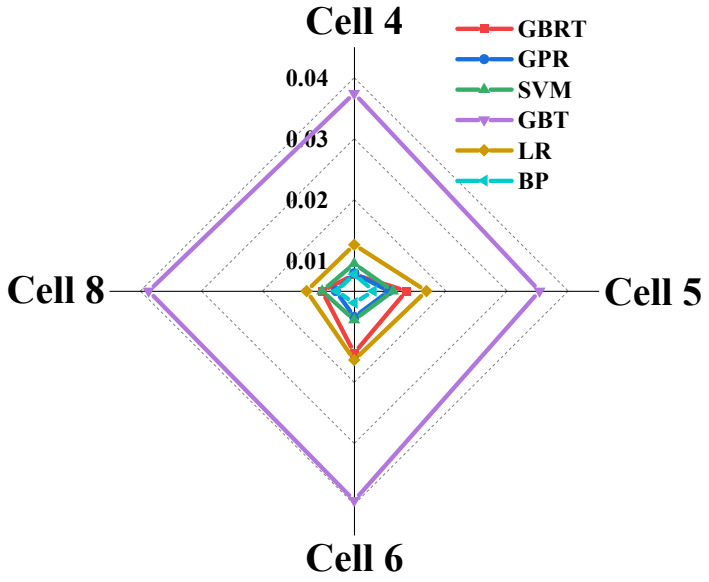


Figure 14. Radar chart of estimation errors for various SOH estimation models.

Considering both estimation accuracy and training time, the GBRT estimation model achieves a training duration comparable to that of the GBT and linear regression models. Compared with the GBT model, which has the shortest training time, the GBRT model increases training time by only 6.5%, while its SOH estimation accuracy (taking Cell 4 in

the Oxford dataset as an example) outperforms the best linear regression model among the two alternatives. Specifically, the RMSE is reduced by 0.0049, corresponding to a 38.9% decrease relative to the linear regression model, indicating that the GBRT model can achieve higher SOH estimation accuracy while maintaining rapid computation. Compared with the gaussian process regression (GPR), BP neural network, and SVM models, the GBRT model achieves the highest estimation accuracy, slightly outperforming the other three models. In terms of training time, however, the GBRT model is significantly faster than these alternatives: training time is reduced by 57.6% compared with the fastest GPR model, and by 91.9% compared with the slowest SVM model. Taking into account the minimal voltage intervals required for health indicator extraction and the relatively fast training time, the GBRT estimation model constructed in this study can provide rapid and accurate SOH estimation for lithium-ion batteries while maintaining high precision.

5. Conclusions

During long-term operation of lithium-ion batteries, rapid and accurate SOH estimation is crucial for ensuring battery safety and reliability, providing important guidance for the scientific maintenance and management of battery systems. To achieve accurate and interpretable SOH estimation, this study presents a practical and robust SOH estimation framework based on incremental capacity analysis, curve similarity, and GBRTs. By statistically identifying frequently occurring charging voltage regions and extracting SOH-sensitive health indicators from a narrow voltage window, the proposed method achieves high estimation accuracy with significantly reduced data and computational requirements. Validation on public datasets and laboratory data demonstrates strong robustness against measurement noise and capacity regeneration, highlighting its suitability for real-world applications. The main contributions of this study are as follows:

(1) To address the challenge of extracting highly sensitive health indicators during actual lithium-ion battery operation, The charging behavior of batteries in practical usage is first analyzed to identify feasible voltage intervals for health indicator extraction. Incremental capacity analysis was then used to determine the specific voltage intervals for extracting these indicators. Based on incremental capacity and DTW distance, three types of health indicators—interval capacity increment, local Q–V curve similarity, and segment voltage curve similarity—were extracted and their Pearson correlations analyzed on the Oxford and CALCE datasets. The health indicator extraction method proposed in this study requires only 0.13 V and 0.07 V voltage intervals on the Oxford and CALCE datasets, respectively, to construct the health indicator parameter sets.

(2) Using the extracted health indicators, a GBRT-based SOH estimation model was developed. On the Oxford dataset, the proposed method achieves an RMSE around 2.5%, with a minimum RMSE of 1.73%. On the CALCE dataset, the RMSE remains around 2.0. Laboratory validation further demonstrates that the proposed method maintains an RMSE below 5.0, confirming its robustness across different battery types and degradation patterns.

(3) Compared with mainstream SOH estimation models, including linear regression, SVM, GPR, and BP neural networks, the GBRT-based model achieves comparable or superior estimation accuracy while significantly reducing computational cost. Specifically, the training time of the GBRT model is reduced by 57.6% compared with GPR and by 91.9% compared with SVM, while increasing by only 6.5% relative to simple boosting trees. Meanwhile, the SOH estimation error is reduced by up to 38.9% compared with linear regression, demonstrating a favorable balance between accuracy and efficiency.

Future work will focus on adaptive voltage window selection and multi-feature fusion, enabling the method to further enhance robustness and applicability.

Author Contributions: C.C.: Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. Y.H.: Writing—review and editing, Writing—original draft, Software, Methodology, Investigation. Y.W.: Writing—review and editing. Y.X.: Methodology, Investigation, Funding acquisition. J.J.: Supervision, Resources, Methodology. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Hubei Province Science and Technology Innovation Base (Platform) Project (Grant No. 2025CSA053) and the National Natural Science Foundation of China (Grant No. 52377207).

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sciatti, F.; Tamburrano, P.; Distaso, E.; Amirante, R.; Radogna, A.; Grassi, G. A comprehensive overview of lithium-ion batteries for electric vehicles: Materials, performance, safety, recycling, and emerging technologies. *J. Energy Storage* **2026**, *144*, 119694. [CrossRef]
2. Citroni, R.; Mangini, F.; Frezza, F. Efficient Integration of Ultra-low Power Techniques and Energy Harvesting in Self-Sufficient Devices: A Comprehensive Overview of Current Progress and Future Directions. *Sensors* **2024**, *24*, 4471. [CrossRef] [PubMed]
3. Afaynou, I.; Faraji, H.; Choukairy, K.; Djebali, R.; Rezk, H. Comprehensive Analysis and Thermo-Economic Optimization of a Hybrid Phase Change Material-Based Heat Sink for Electronics Cooling. *Heat Transf.* **2025**, *54*, 3754–3774. [CrossRef]
4. Vega-Muratalla, V.O.; Ramírez-Márquez, C.; Lira-Barragán, L.F.; Ponce-Ortega, J.M. Review of lithium as a strategic resource for electric vehicle battery production: Availability, Extraction, and Future Prospects. *Resources* **2024**, *13*, 148. [CrossRef]
5. Zhang, H.; Li, Y.; Zheng, S.; Lu, Z.; Gui, X.; Xu, W.; Bian, J. Battery lifetime prediction across diverse ageing conditions with inter-cell deep learning. *Nat. Mach. Intell.* **2025**, *7*, 270–277. [CrossRef]
6. Barré, A.; Deguilhem, B.; Grolleau, S.; Gérard, M.; Suard, F.; Riu, D. A review on lithium-ion battery ageing mechanisms and estimations for automotive applications. *J. Power Sources* **2013**, *241*, 680–689. [CrossRef]
7. Lin, Y.H.; Tian, L.L.; Ding, Z.Q. Ensemble remaining useful life prediction for lithium-ion batteries using fusion of historical and real-time degradation data. *IEEE Trans. Veh. Technol.* **2023**, *72*, 5934–5947. [CrossRef]
8. Shrivastava, P.; Soon, T.K.; Idris, M.Y.I.B.; Mekhilef, S. Overview of model-based online state-of-charge estimation using Kalman filter family for lithium-ion batteries. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109233. [CrossRef]
9. Li, Y.; Liu, K.; Foley, A.M.; Zülke, A.; Berecibar, M.; Nanini-Maury, E.; Van Mierlo, J.; Hoster, H.E. Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109254. [CrossRef]
10. Moura, S.J.; Krstić, M.; Chaturvedi, N.A. Adaptive PDE observer for battery SOC/SOH estimation. *J. Dyn. Syst. Meas. Control.* **2012**, *45295*, 101–110. [CrossRef]
11. Xiong, R.; Li, L.; Li, Z.; Yu, Q.; Mu, H. An electrochemical model-based degradation state identification for lithium-ion batteries in all-climate electric vehicle applications. *Appl. Energy* **2018**, *219*, 264–275. [CrossRef]
12. Bole, B.; Kulkarni, C.S.; Daigle, M. Adaptation of an electrochemistry-based lithium-ion battery model to account for deterioration under randomized use. *Annu. Conf. PHM Soc.* **2014**, *6*, 2490. [CrossRef]
13. Jeon, D.H.; Kim, S.; Hempelmann, R. State-of-the-art review of degradation mechanisms of commercial lithium-ion batteries. *J. Power Sources* **2025**, *646*, 237242. [CrossRef]
14. Feng, T.; Yang, L.; Zhao, X.; Zhang, H.; Qiang, J. Online identification of lithium-ion battery parameters based on an improved equivalent-circuit model and its implementation on battery state-of-power prediction. *J. Power Sources* **2015**, *281*, 192–203. [CrossRef]
15. Andre, D.; Meiler, M.; Steiner, K.; Walz, H.; Soczka-Guth, T.; Sauer, D.U. Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. II: Modelling. *J. Power Sources* **2011**, *196*, 5349–5356. [CrossRef]
16. Hu, X.; Li, S.; Peng, H. A comparative study of equivalent circuit models for Li-ion batteries. *J. Power Sources* **2012**, *198*, 359–367. [CrossRef]
17. Lyu, C.; Lai, Q.; Ge, T.; Yu, H.; Wang, L.; Ma, N. Remaining useful life prediction of lead-acid batteries using an electrochemical model within a particle filtering framework. *Energy* **2017**, *120*, 975–984. [CrossRef]
18. Han, X.; Ouyang, M.; Lu, L.; Li, J. A comparative study of commercial lithium-ion battery cycle life in electric vehicle applications: Capacity loss estimation. *J. Power Sources* **2014**, *268*, 658–669. [CrossRef]

19. Lu, J.; Xiong, R.; Tian, J.; Wang, C.; Sun, F. Deep learning to estimate lithium-ion battery state of health without additional degradation experiments. *Nat. Commun.* **2023**, *14*, 2760. [CrossRef]
20. Yao, Q.; Song, X.; Xie, W. State-of-health estimation of lithium-ion batteries based on CNN–WNN–WLSTM. *Complex Intell. Syst.* **2024**, *10*, 2919–2936. [CrossRef]
21. Fan, M.; Geng, M.; Yang, K.; Zhang, M.; Liu, H. State of health estimation of lithium-ion batteries based on electrochemical impedance spectroscopy. *Energies* **2023**, *16*, 3393. [CrossRef]
22. Wang, J.; Deng, Z.; Yu, T.; Yoshida, A.; Xu, L.; Guan, G. State of health estimation based on modified Gaussian process regression for lithium-ion batteries. *J. Energy Storage* **2022**, *51*, 104512. [CrossRef]
23. Zhang, Y.; Liu, D.; Yu, J.; Peng, Y.; Peng, X. Remaining useful life prediction using EMA with weighted bagging Gaussian process regression. *Microelectron. Reliab.* **2017**, *75*, 253–263. [CrossRef]
24. Zhang, M.; Chen, W.; Yin, J.; Feng, T. Lithium battery health factor extraction based on improved Douglas–Peucker algorithm and SOH prediction using XGBoost. *Energies* **2022**, *15*, 5981. [CrossRef]
25. Ma, C.; Zhai, X.; Wang, Z.; Tian, M.; Yu, Q.; Liu, L.; Liu, H.; Wang, H.; Yang, X. State of health prediction for lithium-ion batteries using multi-view feature fusion and support vector regression ensemble. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 2269–2282. [CrossRef]
26. Agudelo, B.O.; Zamboni, W.; Postiglione, F.; Monmasson, E. Battery state-of-health estimation based on multiple charge and discharge features. *Energy* **2023**, *263*, 125637. [CrossRef]
27. Li, Y.; Zou, C.; Berecibar, M.; Nanini-Maury, E.; Chan, J.C.W.; Van den Bossche, P.; Van Mierlo, J.; Omar, N. Random forest regression for online capacity estimation of lithium-ion batteries. *Appl. Energy* **2018**, *232*, 197–210. [CrossRef]
28. Yang, D.; Zhang, X.; Pan, R.; Wang, Y.; Chen, Z. A novel Gaussian process regression model for state-of-health estimation of lithium-ion batteries using charging curves. *J. Power Sources* **2018**, *384*, 387–395. [CrossRef]
29. Li, X.; Wang, Z.; Yan, J. Prognostic health condition estimation for lithium-ion batteries using partial incremental capacity and Gaussian process regression. *J. Power Sources* **2019**, *421*, 56–67. [CrossRef]
30. He, Y.; Zeng, Q.; Tang, L.; Liu, F.; Li, Q.; Yin, Y.; Xu, S.; Deng, B. Unlocking Ultrafast Diagnosis of Retired Batteries via Interpretable Machine Learning and Optical Fiber Sensors. *J. Energy Storage* **2025**, *139*, 118725. [CrossRef]
31. Gao, T.; Wu, Y. Applications and Advances of Machine Learning in the Development of Solid-State Electrolytes for Lithium-Ion Batteries. *ACS Omega* **2025**, *10*, 60094–60109. [CrossRef]
32. Richardson, R.R.; Birkl, C.R.; Osborne, M.A.; Howey, D.A. Gaussian process regression for in situ capacity estimation of lithium-ion batteries. *IEEE Trans. Ind. Inform.* **2019**, *15*, 127–138. [CrossRef]
33. He, N.; Williard, N.; Osterman, M.; Pecht, M. Prognostics of lithium-ion batteries based on Dempster–Shafer theory and Bayesian Monte Carlo methods. *J. Power Sources* **2011**, *196*, 10314–10321. [CrossRef]
34. Liu, Z.; Gilbert, G.; Cepeda, J.M.; Lysdahl, A.O.K.; Piciullo, L.; Hefre, H.; Lacasse, S. Modelling shallow landslides using machine learning algorithms. *Geosci. Front.* **2021**, *12*, 385–393. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.