



Article

# State of Health Estimation for Lithium-Ion Batteries Using an Explainable XGBoost Model with Parameter Optimization

**Zhenghao Xiao** <sup>1,2</sup>, **Bo Jiang** <sup>1,2,\*</sup>, **Jiangong Zhu** <sup>1,2</sup>, **Xuezhe Wei** <sup>1,2</sup> and **Haifeng Dai** <sup>1,2,\*</sup><sup>1</sup> School of Automotive Studies, Tongji University, Shanghai 201804, China; 2332900@tongji.edu.cn (Z.X.); zhujiangong@tongji.edu.cn (J.Z.); weixzh@tongji.edu.cn (X.W.)<sup>2</sup> Clean Energy Automotive Engineering Center, Tongji University, Shanghai 201804, China

\* Correspondence: jiangbo15@tongji.edu.cn (B.J.); tongjidai@tongji.edu.cn (H.D.)

**Abstract:** Accurate and reliable estimation of the state of health (SOH) of lithium-ion batteries is crucial for ensuring safety and preventing potential failures of power sources in electric vehicles. However, current data-driven SOH estimation methods face challenges related to adaptiveness and interpretability. This paper investigates an adaptive and explainable battery SOH estimation approach using the eXtreme Gradient Boosting (XGBoost) model. First, several battery health features extracted from various charging and relaxation processes are identified, and their correlation with battery aging is analyzed. Then, a SOH estimation method based on the XGBoost algorithm is established, and the model's hyper-parameters are tuned using the Bayesian optimization algorithm (BOA) to enhance the adaptiveness of the proposed estimation model. Additionally, the Tree SHapley Additive exPlanation (TreeSHAP) technique is employed to analyze the explainability of the estimation model and reveal the influence of different features on SOH evaluation. Experiments involving two types of batteries under various aging conditions are conducted to obtain battery cycling aging data for model training and validation. The quantitative results demonstrate that the proposed method achieves an estimation accuracy with a mean absolute error of less than 2.7% and a root mean squared error of less than 3.2%. Moreover, the proposed method shows superior estimation accuracy and performance compared to existing machine learning models.



**Citation:** Xiao, Z.; Jiang, B.; Zhu, J.; Wei, X.; Dai, H. State of Health Estimation for Lithium-Ion Batteries Using an Explainable XGBoost Model with Parameter Optimization.

*Batteries* **2024**, *10*, 394. <https://doi.org/10.3390/batteries10110394>

Academic Editors: Vaclav Knap, Daniel Auger and Abbas Fotoohi

Received: 31 August 2024

Revised: 3 November 2024

Accepted: 5 November 2024

Published: 7 November 2024



**Copyright:** © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

To cope with climate warming and environmental pollution, electric vehicles have been widely used around the world. The power battery is one of the core components of the electric vehicle power system. Lithium-ion batteries have become the preferred power energy system for electric vehicles because of their advantages in superior energy density, high efficiency, and long service life [1–3]. However, during storage and operation, the physical and chemical properties of the battery will change, resulting in a decrease in battery capacity and power performance [4]. Therefore, the accurate and reliable estimation of the battery State of Health (SOH), which is considered one critical technique in the Battery Management System (BMS), is used to monitor and control the battery aging state [5]. Currently, SOH estimation methods can be divided into two broad categories: model-based methods and data-driven methods [6,7].

According to the adopted battery model structure, model-based SOH estimation methods can be further divided into electrochemical model (EM)-based methods and equivalent circuit model (ECM)-based methods. The EM-based method estimates battery SOH by simulating the electrochemical reaction process inside the cell during operation [8]. Chen et al. [9] considered the double layer resistance increase, solid electrolyte interface (SEI) film growth, and cathode aggregate crack propagation as the main aging effects,

leading to the development of an electrochemical-thermal-aging coupled model, and the SOH estimation max error was less than 1%. The EM-based methods usually use complex partial differential equations to describe battery characteristics, requiring high memory and computing resources. Compared to EMs, the ECM has a simpler structure and greatly reduces the number of parameters that need to be learned over time, making it easier to perform online estimation. Huang et al. [10] established an improved equivalent circuit model (IECM) based on current-time data of the constant voltage charging process. They used the model parameters including two resistances and two capacitances of the second-order IECM as health indicators for SOH estimation. The maximum SOH error of this method was less than 2% at different temperatures and data lengths. The advantage of ECM-based methods is that it is simple and easy to perform online estimation. However, in practical applications, the uncertainty and unknown interference of the model inevitably affect the battery dynamic system, causing deviation in the estimated SOH [11].

The point of the data-driven SOH estimation method is to establish the association model between battery health indicators (HIs) and battery SOH through some regression analysis methods or machine learning approaches. Machine learning models do not require predefined mathematical models and can deal with nonlinear and complex problems [12]. Commonly used machine learning algorithms for SOH estimation include relevance vector machine (RVM) [13], support vector machine (SVM) [14,15], Gaussian process regression (GPR) [16,17], random forest regression [18], etc. Wu et al. [19] proposed a SOH estimation approach based on an optimized feature selection method using the particle swarm optimization algorithm. The optimal feature set obtained was used to estimate the battery SOH with the ridge regression algorithm. The proposed method achieved SOH estimation error similar to SVR and GPR while having a simpler structure and lower computation cost. Chen et al. [20] extracted the increase in mean ohmic resistance as a new HI and used a hierarchical extreme learning machine to estimate battery SOH. The accuracy and robustness were verified on the datasets of four batteries at three different temperatures with dynamic loading profiles. Li et al. [21] used the idea of ensemble learning to generate differential data samples and synthesize the output of a series of base learners, avoiding the regionality of data training samples. The SOH estimation results were compared with traditional machine learning methods including NN, GPR, kernel ridge regression (KRR) and SVM via NASA battery datasets, indicating superior accuracy and stability of the proposed method.

Recently, neural network techniques, mainly including convolutional neural networks (CNN) [22,23] and recurrent neural networks (RNN) [24], gradually demonstrated excellent performance in battery SOH estimation. Lin et al. [25] proposed a novel adaptive tunable hybrid radial basis function network. To increase the adaptability of the hybrid network, the structural parameters of the proposed hybrid network were adaptively modulated by Brownian motion modeling and particle filter. The experimental results demonstrated the accuracy and robustness of the model. Sheng et al. [23] proposed a deep convolutional neural network algorithm based on ensemble learning and transfer learning (DCNN-ETL), whose performance was more accurate and robust on smaller data sets and unseen data. Yao et al. [26] used CNN and a long short-term memory neural network (LSTM) to extract temporal features, and spatial features were obtained by the graph sample aggregate (GraphSAGE), which uncovered the deep information among HIs. And a graph neural network (GNN) was used to estimate battery SOH, whose feasibility was verified on MIT, NASA, and experimental datasets. Kheirkhah-Rad et al. [27] extracted a new HI, namely referenced-based charging time (RCT). Based on RCT, they implemented a deep feed-forward neural network to estimate battery SOH. The results on 17 cells showed a root mean square percentage error (RMSPE) of 0.43%. The performance of the data-driven estimation methods largely depends on the adaptability of data-driven models [28]. However, there are challenges in hyper-parameter optimization for data-driven methods. Specifically, potential challenges include search space and running time [29]. Especially when the algorithm has a large number of hyper-parameters, the search space will be

complex and the running time will be long. This results in the inefficiency of the hyper-parameter optimization process. Thus, a hyper-parameter optimization algorithm that can increase the efficiency of the hyper-parameter optimization process and ensure high performance of the machine learning methods is needed.

Although data-driven methods have achieved high estimation accuracy in battery SOH estimation, they still face significant challenges, including the issue that data-driven methods lack effective interpretability [30]. In most cases, the association model between the battery HIs and SOH is considered a “black box”, and the internal operations are usually unknown [31]. All the studies using data-driven methods mentioned above lack such interpretability. The explainability analysis of machine learning is an important technique to solve this problem. The explainability analysis includes global and local explainability analysis. The former analyzes the overall impact of the features on the result and can help to reveal the importance of each feature. The latter focuses on a single sample point and can explain the cause of potential outliers of results. There are few studies on the explainability of data-driven methods for SOH estimation. Wang et al. [32] proposed an explainability-driven model improvement framework for lithium-ion battery SOH estimation. The superiority and effectiveness of the proposed framework were validated on different datasets and different models. However, this study lacks local explainability analysis, which can help to explain the local outliers of estimation results. Von Bülow et al. [33] proposed a new method using Gaussian-filtered saliency maps to visualize battery operational states that are relevant to deep neural network (DNN) models. The results show that the proposed method can add transparency and interpretability to the SOH prediction results of the two state-of-the-art DNNs. However, this method can only provide qualitative explainability analysis, and the correlation values given by it have no clear physical meaning. To sum up, a quantitative explainability analysis method that covers both global and local explainability is needed to enhance the explainability and transparency of the data-driven SOH estimation model.

In order to solve the problems stated above, this study proposes an explainable XGBoost method for battery SOH estimation. Specifically, several HIs related to battery degradation were extracted from charging data, and an XGBoost model was established to estimate the battery SOH. To enhance the adaptability of the proposed SOH estimation method, the Bayesian optimization algorithm (BOA) [34] was used to optimize the model hyper-parameters. Moreover, global and local explainability analysis were conducted to interpret the estimation results based on the TreeSHAP method. The main contributions of this study are as follows:

1. To solve the challenge of determining the optimal values of the many hyper-parameters in the XGBoost model, the Bayesian optimization algorithm was employed to achieve parameter self-adjustment. This can realize the balance of optimization effect and efficiency and enhance the adaptability of the estimation model.
2. TreeSHAP was used for global and local explainability analysis, revealing the impact that HIs have on the battery SOH estimation and the cause of abnormal estimation results. The explainability analysis plays an important role in reducing the “black-box” characteristics of machine learning models and enhancing deep understanding;
3. The accuracy of the proposed battery SOH method was verified through two aging datasets from different commercial batteries, containing 15 cells with over 8000 total aging cycles experienced. The experimental results show that the proposed method achieves considerable accuracy for different cells under different aging conditions.

The remainder of this study is organized as follows. Section 2 describes the details of two used batteries and the schedules for aging experiments to construct the aging datasets. Section 3 describes the HIs extracted from the experimental data, and the correlation between HIs and battery aging is investigated based on the Pearson correlation coefficient. Section 4 shows the fundamental methodologies of the proposed battery SOH estimation method, including the introduction of XGBoost and the BOA for hyper-parameter optimization, followed by explainability analysis techniques. The performance of the proposed

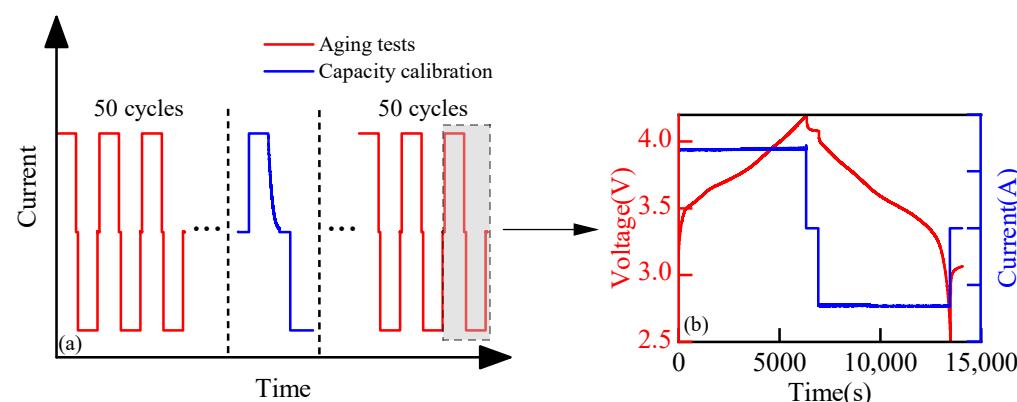
SOH estimation approach is discussed in Section 5. The limitations and outlook of this research are summarized in Section 6. Finally, Section 7 concludes this study.

## 2. Experimental Design and Battery Dataset

### 2.1. The Battery Cycling Aging Test

The battery-aging data for model training and testing were generated from the aging experiments using two 18650-type cylindrical batteries [35]. The cathode material of one battery was  $\text{LiNi}_{0.8}\text{Co}_{0.1}\text{Mn}_{0.1}\text{O}_2$  (noted as cell 1), and that of the other battery was  $\text{Li}(\text{NiCoAl})\text{O}_2$  (noted as cell 2). The nominal capacities of these two types of batteries were 2.9 Ah and 2.75 Ah, respectively. The charging and discharging cut-off voltages are 4.2 V and 2.5 V, respectively. For convenience, the following battery cyclic datasets of cells 1 and 2 are noted as Dataset I and Dataset II.

A battery test schedule was designed to investigate the influence of charging/discharging current and operating temperature on battery aging, which included two main steps, capacity calibration and an aging test, as shown in Figure 1a. The calibration procedure for both datasets adopted a 0.5 C constant current-constant voltage (CC-CV) charging and a 0.5 C discharging to obtain the residual capacity during the cycling aging, and a 1 h rest was set between charging and discharging. Capacity calibration was performed at 25 °C every 50 aging cycles.



**Figure 1.** The battery test schedule: (a) general schedule; (b) aging tests.

As mentioned above, operating current and temperature were the main influencing factors in this study. Hence, detailed test matrixes were designed for these two cells, as shown in Tables 1 and 2. The aging test procedure included four steps: constant current (CC) charging, rest, CC discharging, and rest, as shown in Figure 1b. For Dataset I, the charging rates of all cells were set as 0.5 C, while the discharging rates were set as shown in Table 1. The rest time after the charging process was 10 min. For Dataset II, the discharging rates of all cells were set as 1 C, while the charging rates were set as shown in Table 2. The rest time after the charging process was 30 min. During the aging cyclic tests, the batteries were placed in the thermal chamber, and the detailed temperature settings are also shown in Tables 1 and 2.

**Table 1.** Aging test conditions in Dataset I.

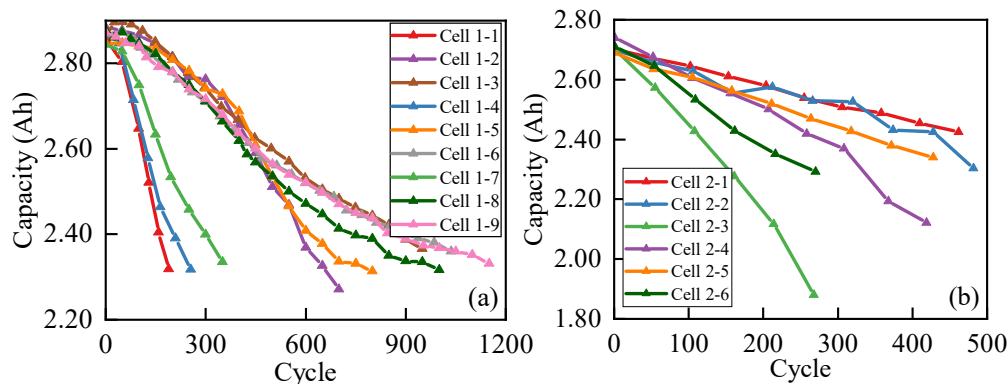
Discharging Current Rate	Temperature		
	10 °C	25 °C	40 °C
0.5 C	Cell 1-1	Cell 1-2	Cell 1-3
1 C	Cell 1-4	Cell 1-5	Cell 1-6
2 C	Cell 1-7	Cell 1-8	Cell 1-9

**Table 2.** Aging test conditions in Dataset II.

Charging Current Rate	Temperature	
	25 °C	40 °C
0.2 C	Cell 2-1	Cell 2-4
0.5 C	Cell 2-2	Cell 2-5
1 C	Cell 2-3	Cell 2-6

## 2.2. The Battery Aging Dataset

Based on the above test matrixes, the capacity fading of the used cells is shown in Figure 2. It can be seen in Figure 2a that the capacity fading rate of cells 1-1, 1-4 and 1-7 was significantly higher than that of other cells. Moreover, cells 1-2, 1-5 and 1-8 also had a higher capacity fading rate than cells 1-3, 1-6 and 1-9. This indicates that a lower temperature will cause faster capacity decay, especially for 10 °C, which will dramatically increase the aging rate of the battery. In addition, at 10 °C and 25 °C, the aging rate of the battery decreased with the increase in the discharge rate. The possible reason is that with the same charging current rate, a battery with a smaller discharging current rate works in a relatively wide SOC interval. The lower limit of the SOC cycle interval decreases as the discharging current rate decreases. Cycling in the low SOC range will lead to faster attenuation of battery capacity and will likely lead to nonlinear accelerated attenuation of capacity [36]. This law is consistent with the law obtained in Refs. [37,38]. At 45 °C, the discharge rate has no significant effect on the aging rate of the battery. As can be seen in Figure 2b, the aging rate of cells 2-3 and 2-6 is higher than that of other cells. This indicates that a high charging rate (1 C) will significantly increase the battery aging rate. The main reason is that when the battery is charged at a low temperature or a high rate, the cathode will precipitate lithium metal attached to the graphite surface, and will react with the electrolyte, further consuming available circulating lithium-ions, resulting in rapid battery capacity decay [39].

**Figure 2.** Capacity fading: (a) Dataset I; (b) Dataset II.

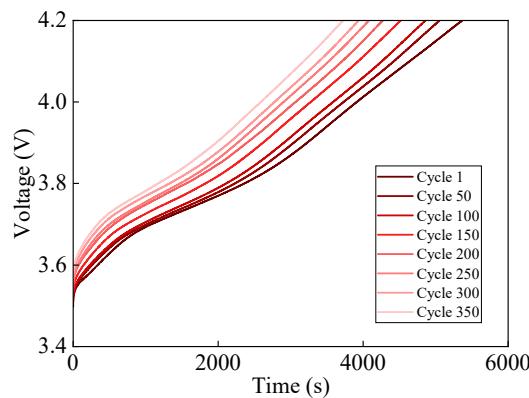
## 3. The Battery Health Indicator Extraction

For data-driven battery SOH estimation methods, the input of the data-driven model is the feature set extracted from the battery data, which has a significant impact on the output of the SOH estimation model. In this study, the feature is renamed as health indicator (HI). In other words, it is particularly critical to extract HIs that can effectively characterize the degree of battery aging. As described in Section 2, battery voltage and current data were obtained from battery-aging experiments. With the aging of the battery, the original charging and rest data also show a certain trend of variation. Therefore, based on the variation trend of original voltage and current data during battery aging, several HIs with a strong correlation with battery aging were extracted from the aging test data.

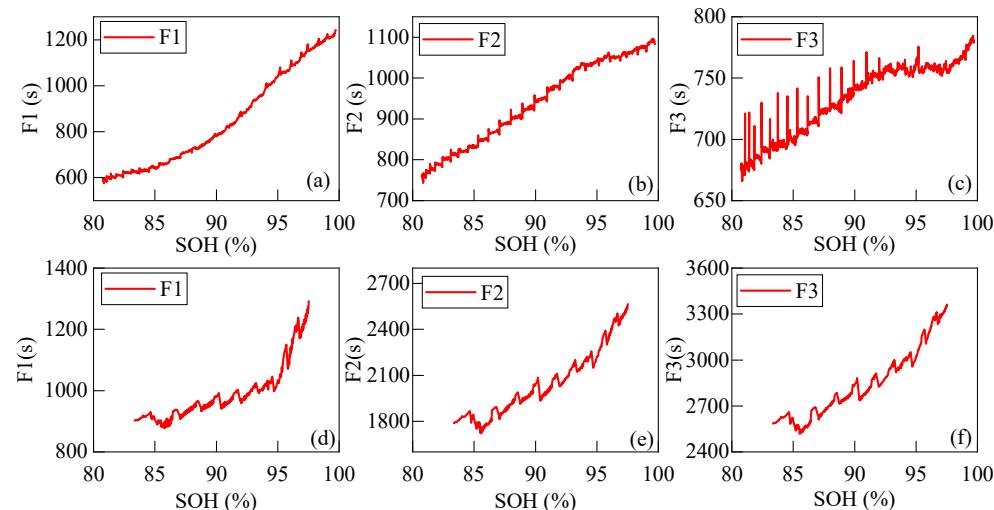
### 3.1. The Extracted Battery HIs

#### 3.1.1. Time During Equal Voltage Increase

Figure 3 shows the variation of the charging voltage–time curves during battery aging. It can be seen that with the aging of the battery, the slope of the voltage–time curve increases. As a result, the time during equal voltage increase (TEVI) decreases. Therefore, TEVI in 3.6–3.7 V, 3.7–3.8 V, and 3.8–3.9 V were extracted as HIs, named F1–F3. The variation in TEVI of cells 1–9 and 2–2 during battery aging is shown in Figure 4 as an example, which experienced the longest aging test in each dataset.



**Figure 3.** The variation of the charging voltage–time curves during battery aging.



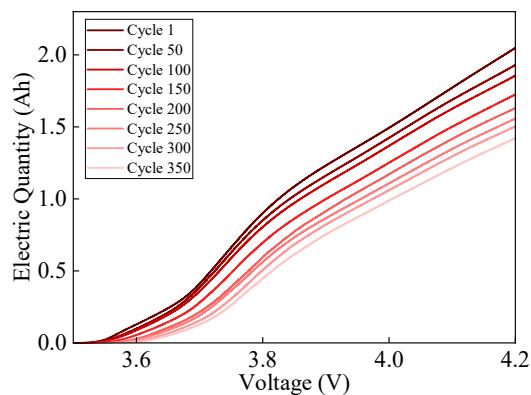
**Figure 4.** TEVI variation during battery aging: (a–c) cell 1–9; (d–f) cell 2–2.

#### 3.1.2. Electric Quantity During Equal Voltage Increase

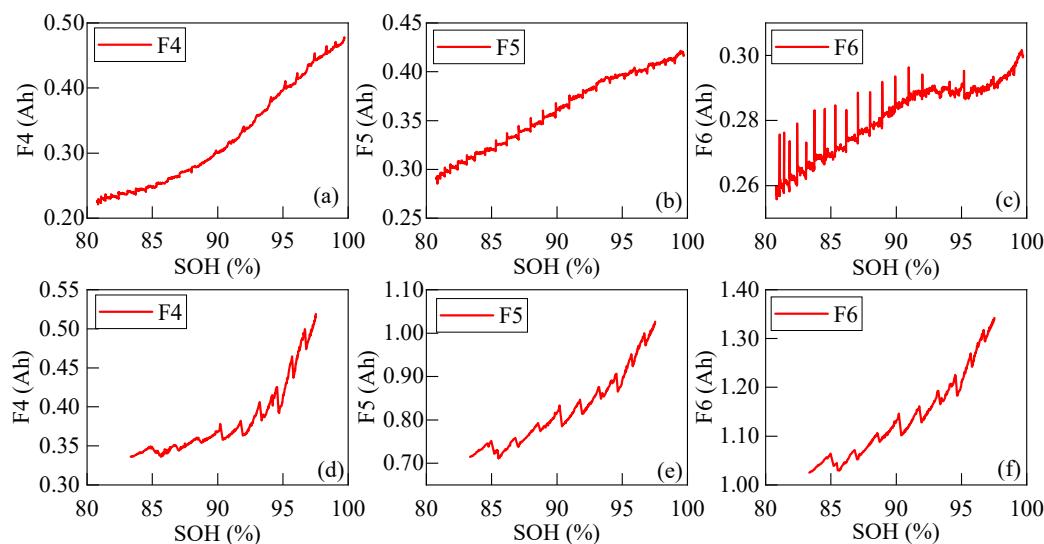
Figure 5 shows the variation of the charging electric quantity–voltage curves during battery aging. Similarly, the electric quantity during equal voltage increase (QEVI) shows a decreasing tendency with the aging of the battery. Therefore, QEVI in 3.6–3.7 V, 3.7–3.8 V, and 3.8–3.9 V are extracted as HIs, named F4–F6. The variation of QEVI of cells 1–9 and 2–2 during battery aging is shown in Figure 6 as an example.

#### 3.1.3. Voltage Drop During Rest

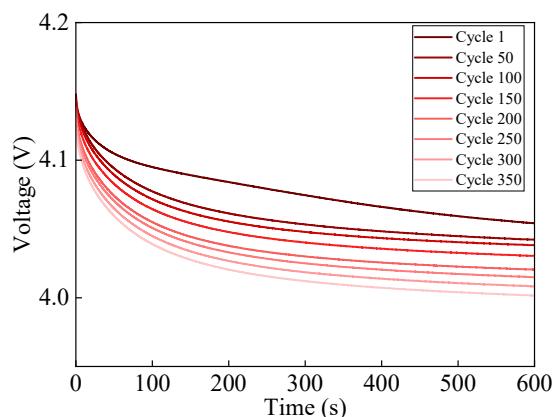
Figure 7 shows the variation in the charging voltage–time curves during the rest process after charging. It can be seen that as the battery ages, both the speed and the range of voltage drop increase. Therefore, voltage drop during rest (VDR) is extracted as an HI, namely F7. The variation of VDR of cells 1–9 and 2–2 is shown in Figure 8 as an example.



**Figure 5.** The variation of the charging electric quantity–voltage curves during battery aging.



**Figure 6.** QEVI variation during battery aging: (a–c) cell 1-9; (d–f) cell 2-2.

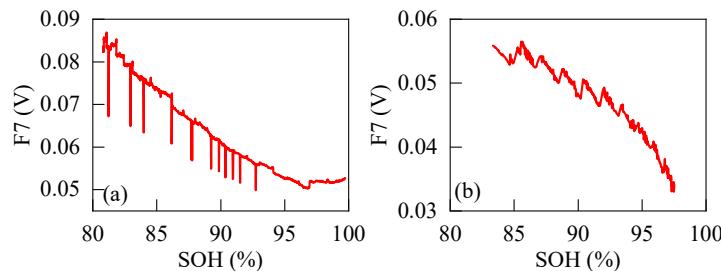


**Figure 7.** The variation in the rest voltage–time curve during battery aging.

### 3.1.4. Peak Value and Position of Charging IC Curve

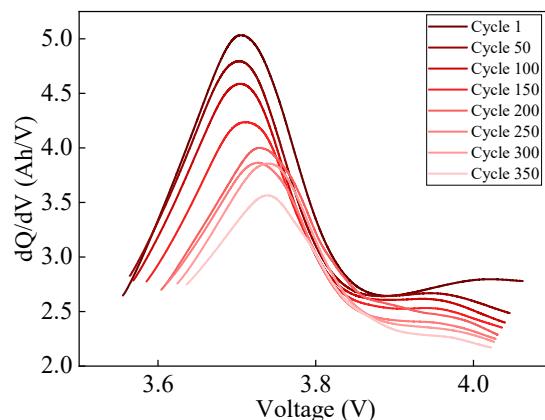
The peak in the IC curve has a unique shape, height, and position, which reflects the electrochemical reaction that occurs during the charging and discharging of lithium batteries [40], and the change in the peak may be related to the loss of active materials in lithium batteries [41]. The IC curve can transform the voltage plateau, at which the voltage curve rises slowly while reactions take place violently inside the battery, into the  $dQ/dV$  peak, which is easy to observe. Thus, it can reflect the distribution of charge (Q) in different

voltage intervals through characteristic peaks [42] as well as small changes that are not easy to observe in the voltage-time curve. It contains more intuitive aging characteristics and is therefore widely used for SOH estimation of lithium-ion batteries [43].

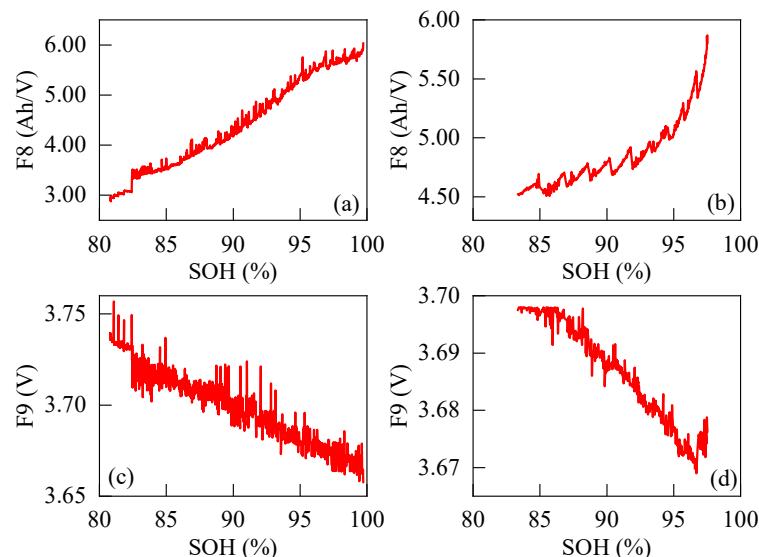


**Figure 8.** VDR variation during battery aging: (a) cell 1-9; (b) cell 2-2.

Figure 9 shows the variation of the IC curves during the aging of the battery. It can be seen that the peak of the IC curves shows a general tendency of moving to the lower right direction as the battery ages. In other words, the peak value decreases while the peak position increases. Therefore, the peak value of the IC curve (PVIC) and peak position of the IC curve (PPIC) are extracted as HIs, named F8 and F9. The variation of PVIC and PPIC of cells 1-9 and 2-2 during battery aging is shown in Figure 10.



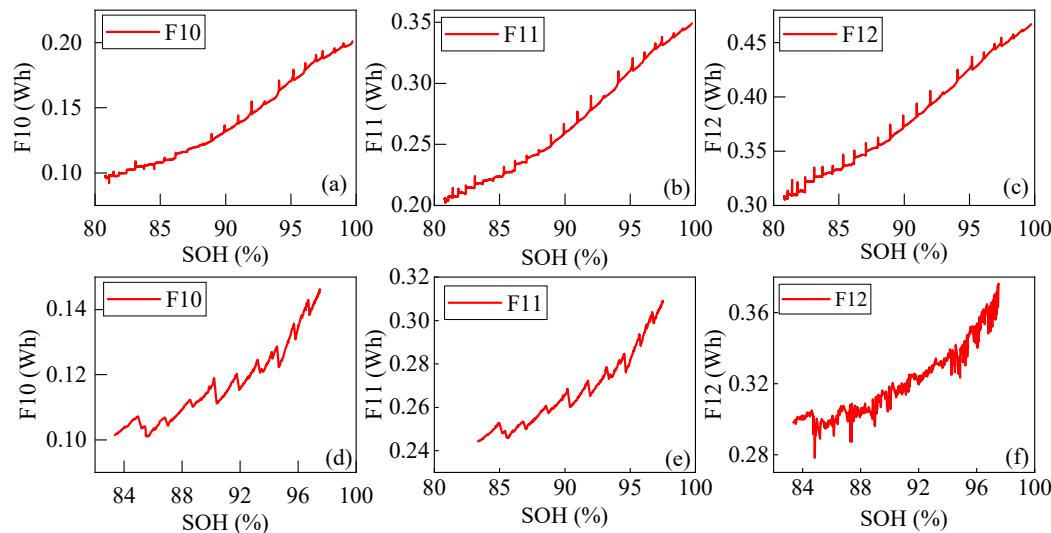
**Figure 9.** The variation of the IC curves during the aging of the battery.



**Figure 10.** PVIC and PPIC variation during battery aging: (a) PVIC of cell 1-9; (b) PVIC of cell 2-2; (c) PPIC of cell 1-9; (d) PPIC of cell 2-2.

### 3.1.5. Area Covered by Charging Q-V Curve

It can be seen in Figure 5 that the area covered by the charging Q-V curves (AQVs) shows a decreasing tendency with the aging of the battery. Therefore, AQVs in 3.6–3.8 V, 3.8–4.0 V, and 4.0–4.2 V are extracted as HIs, named F10–F12. The variation in AQVs of cells 1–9 and 2–2 during battery aging is shown in Figure 11 as an example.



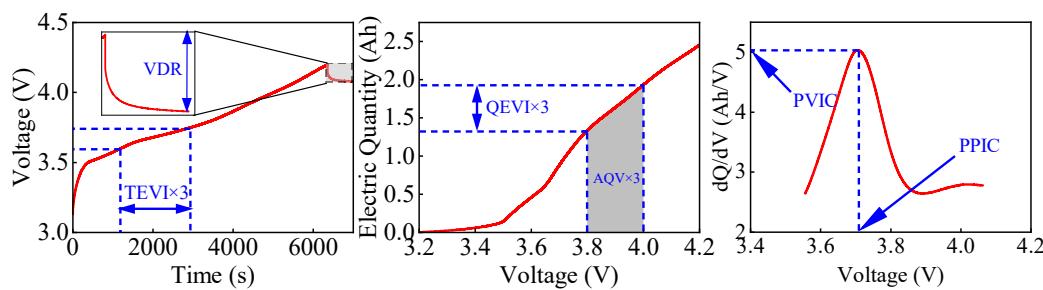
**Figure 11.** AQV variation during battery aging: (a–c) cell 1–9; (d–f) cell 2–2.

### 3.2. The Correlation Analysis Between Battery SOH and HIs

As stated above, 12 HIs are extracted from battery charging data, as shown in Figure 12. HIs include TEVI in three voltage intervals and VDR extracted from voltage–time curves, QEVI and AQV in three voltage intervals extracted from electric quantity–voltage curves, and PVIC and PPIC extracted from charging IC curves. The correlation between them and battery aging is analyzed based on the Pearson correlation coefficient (PCC), as shown in Equation (1).

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

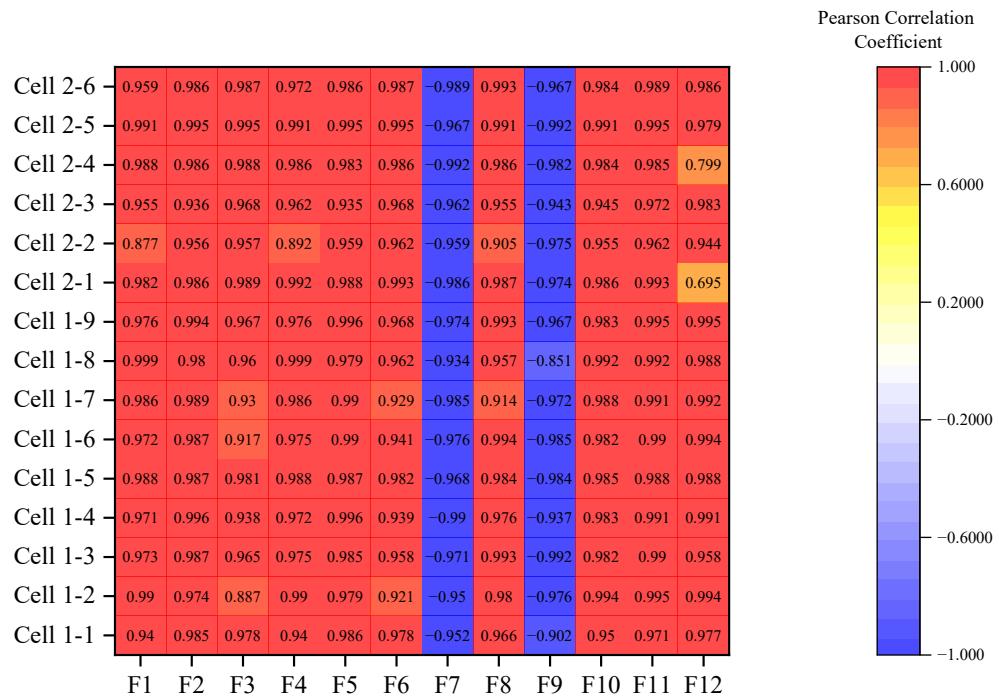
where  $x_i$  and  $y_i$  represent HI and SOH, respectively,  $\bar{x}$  and  $\bar{y}$  represent the mean value of HI and SOH, respectively.



**Figure 12.** HIs extracted from battery charging data.

The PCCs between 12 HIs and the degradation of the cells are shown in Figure 13. It can be seen that F7 and F9 are negatively correlated with battery aging while the other HIs have a positive correlation with battery aging. Except for the F12 of cells 2–1 and 2–4, the absolute PCC values of other HIs are higher than 0.85. This indicates that most HIs are

strongly correlated with battery aging. Furthermore, the PCC values of F10 and F11 on all cells are over 0.95, further indicating the strong correlation between HIs and battery aging.



**Figure 13.** PCCs between HIs and battery aging.

#### 4. SOH Estimation Method

This section presents the methodologies of the proposed SOH estimation method. A BOA-XGBoost algorithm is proposed, followed by the utilization of TreeSHAP for the explainability analysis of the proposed method.

##### 4.1. XGBoost for SOH Estimation

XGBoost [44] is a regression tree-based boosting algorithm based on the gradient boosting decision tree (GBDT) algorithm. Compared with GBDT, the performance and efficiency of XGBoost are greatly improved, and it is a representative algorithm in ensemble learning [45]. The modeling idea of the algorithm is as follows: give a generalized definition of the objective function, find a suitable regression tree in each iteration to fit the residual predicted in the last iteration, minimize the objective function, and approximate the estimated value to the real value [46].

Let  $m$  represents the number of regression trees.  $W$  is the set of all regression trees and  $f_m$  is a function in the function space  $W$ ; the prediction model of XGBoost is shown in Equation (2):

$$\hat{y}_i = \sum_{i=1}^m f_m(x_i), f_m \in W \quad (2)$$

where  $W = \{f(x) = \omega_{q(x)}\}$ ,  $\omega$  is the weight of a leaf node,  $q(x)$  is a leaf node, and the prediction value is the sum of prediction values of  $m$  regression trees. The objective function of XGBoost contains a loss function term and a regular term. Adding regular terms reduces the variance of the model and makes the model easier to learn through the training set to prevent overfitting. The objective function of the model is shown in Equation (3):

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{i=1}^m \Omega(f_m) \quad (3)$$

$$\Omega(f_m) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (4)$$

where  $l$  is the loss function used to measure the error between predicted value  $\hat{y}_i$  and real value  $y_i$ ;  $n$  is the number of samples;  $\Omega(f_m)$  is the regular term, as is shown in Equation (4). It penalizes the complexity of the model.  $f_m$  is the function of the  $m$ -th tree;  $\gamma$  is the complexity of a leaf node;  $T$  is the number of leaf nodes on a tree;  $\lambda$  is the parameter that measures punishment. The incremental model function of the tree is used to minimize the objective function, so when  $m$  trees are generated, the predicted value is

$$\hat{y}_i^{(m)} = \hat{y}_i^{(m-1)} + f_m(x_i) \quad (5)$$

where  $\hat{y}_i^{(m-1)}$  is the predicted value of  $m - 1$  trees, and  $f_m(x_i)$  is the model of the  $m$ -th tree. So, after  $m$  trees are generated, the objective function is

$$L^{(m)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(m-1)} + f_m(x_i)) + \Omega(f_m) \quad (6)$$

The second-order approximation is used to quickly optimize the objective function, and the objective function is expanded by second-order Taylor expansion, as shown in Equation (7):

$$L^{(m)} \approx \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(m-1)} + g_i f_m(x_i) + \frac{1}{2} h_i f_m^2(x_i)\right) + \Omega(f_m) \quad (7)$$

$$g_i = \partial \hat{y}_i^{(m-1)} l\left(y_i, \hat{y}_i^{(m-1)}\right), h_i = \partial^2 \hat{y}_i^{(m-1)} l\left(y_i, \hat{y}_i^{(m-1)}\right) \quad (8)$$

where  $g_i$  and  $h_i$  are the first and second derivatives of the loss function, respectively, as is shown in Equation (8). The objective function can be simplified as

$$L^{(m)} \approx \sum_{i=1}^n l\left(g_i f_m(x_i) + \frac{1}{2} h_i f_m^2(x_i)\right) + \Omega(f_m) \quad (9)$$

Define  $I_j = \{i | q(x_i) = j\}$  as the sample set of leaf nodes. The objective function after the conversion of the regular term can be written as

$$\begin{aligned} L^{(m)} &\approx \sum_{i=1}^n l\left(g_i f_m(x_i) + \frac{1}{2} h_i f_m^2(x_i)\right) + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \\ &= \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \end{aligned} \quad (10)$$

For the fixed structure  $q(x)$ , the optimal weight  $\omega_j^*$  of the leaf node  $j$  can be solved by setting the first derivative to zero, and the optimal solution can be substituted into the objective function to obtain the corresponding optimal target value  $L^{(m)}(q)$ , as is shown in Equation (11):

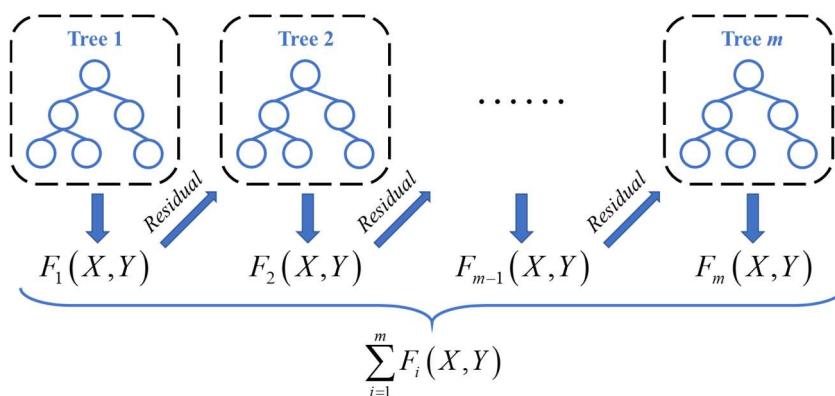
$$\omega_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, L^{(m)}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{\left( \sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (11)$$

The above formula is a score function to measure the quality of the tree structure, but enumerating all the tree structures usually results in an enormous computational cost and is unsuitable for practical use. Instead, a greedy algorithm is adopted to iteratively add branches to the tree starting from a leaf. Assuming that the instance set of the sample

divided into left and right nodes by the segmentation point is  $I_L$  and  $I_R$ , the calculation formula for the loss reduction after splitting is as follows

$$L_{split} = \frac{1}{2} \left[ \frac{\left( \sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left( \sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left( \sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (12)$$

Using the point of maximum loss reduction calculated by the above formula as the best segmentation point, the tree  $f_m(x)$  of different structures is continuously generated by searching all possible splitting ways. Similarly, other nodes recursively repeat this process until the maximum depth of the tree is reached or stop growing when the sum of sample weights is less than a set threshold to prevent overfitting. The training process of a tree is thus complete. The training process for other trees is the same. The 2.1.1 version of the xgboost python package is used. A diagram of the XGBoost algorithm is shown in Figure 14.



**Figure 14.** The diagram of the XGBoost.

#### 4.2. Optimization of Hyper-Parameters Based on Bayesian Optimization Algorithm

Hyper-parameters refer to the parameters that need to be manually set before the learning process begins in a machine learning algorithm, rather than the parameters obtained by the model through training. XGBoost algorithm has many hyper-parameters, including `n_estimators`, `learning_rate`, `max_depth`, `gamma`, `min_child_weight`, `reg_alpha`, `reg_lambda`, and `subsample`. At present, the main hyper-parameter adjustment methods include grid search algorithm (GSA), random search (RS), and Bayesian optimization algorithm (BOA).

GSA is an exhaustive search that lists all the possible combinations of hyper-parameters and then traverses each combination to find the best-performing combination. It can find the global optimal solution within a specified range, but it is inefficient, time-consuming, and requires high computing resources.

Unlike GSA's exhaustive search method, RS uses random numbers to approximate the optimal value. Instead of trying all combinations of hyper-parameters, RS only needs to try a relatively small number of combinations. Therefore, when the number and ranges of hyper-parameters are too large, RS can get the approximate optimal solution more efficiently, but there is a problem of poor precision.

For the above two methods, each search is independent of other searches; that is, each new search does not utilize the results of the previous search, resulting in a waste of computing resources. In contrast, each search of BOA refers to the previous search results. BOA uses the previous evaluation results to form a probability model that maps the hyper-parameters to the scoring probability of the objective function. This probability model is called a “proxy” for the objective function. The advantage of the proxy model is that it is easier to optimize than the objective function, and BOA selects the hyper-parameters that

perform best on the proxy function to evaluate the actual objective function. The specific steps are as follows:

- (1) Establish the proxy probability model of the objective function;
- (2) Find the hyper-parameter that performs best on the proxy model;
- (3) Apply the hyper-parameter to the objective function;
- (4) Update the proxy probability model that contains the results;
- (5) Repeat steps (2) to (4) until the set number of iterations is reached.

Compared with the GSA, BOA has fewer iterations and higher search efficiency. Compared with the RS, BOA makes it easier to obtain the optimal solution. Therefore, BOA performs better in terms of efficiency and accuracy for XGBoost, whose hyper-parameters are complex and need to be optimized across a large range. The 1.5.1 version of the scikit-learn python package is used.

#### 4.3. Explainability Analysis of Machine Learning Algorithms

More research has focused on eXplainable Artificial Intelligence (XAI) to improve the explainability of complex black-box machine learning models. XAI can be roughly divided into intrinsically transparent models and post hoc models that explain complex black-box models. Post hoc XAI models can be further classified as model-agnostic (that is, can be applied to any model) and model-specific techniques that are designed to explain a particular model [47].

SHapley Additive exPlanation (SHAP) is a post hoc XAI technique that explains machine learning model predictions using a game-theoretic approach [48]. SHAP describes the prediction of an observation as a sum of contributions from individual feature values. The average contribution margin for each feature is calculated for all possible additive sequences to obtain the Shapley value of the feature, as shown in Equation (13):

$$\Phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)] \quad (13)$$

where  $S$  represents any subset of features excluding the  $i$ -th feature,  $|S|$  is the size of the subset,  $M$  is the number of features, and the prediction function is represented as  $f_x$ . SHAP has two variations, the KernelSHAP and the TreeSHAP. The former is a model-agnostic algorithm, while the latter is specific for tree-based models. Previous study [49] has proved that the TreeSHAP algorithm is superior to traditional feature attribution methods in terms of consistency, missingness, and local accuracy. Moreover, the TreeSHAP is much more efficient than the KernelSHAP. As stated before, XGBoost is a tree-based algorithm. Therefore, TreeSHAP is used in this study for explainability analysis. The 0.46.0 version of the shap python package is used.

#### 4.4. The General Framework of the SOH Estimation Method

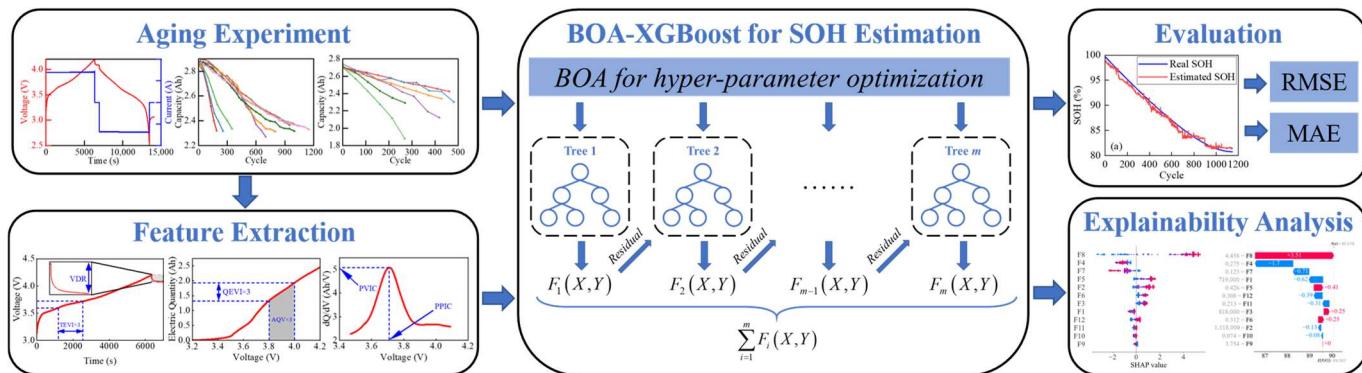
Concerning the evaluation of the SOH estimation performance, two accuracy metrics, including the mean absolute error (MAE) and the root-mean-square error (RMSE), are used in this study, as shown as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (14)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (15)$$

The general framework of the proposed SOH estimation method is shown in Figure 15. Firstly, two battery aging datasets are obtained from the aging cycle experiments of two kinds of batteries. From the datasets, 12 HIs highly correlated with battery degradation are extracted. Moreover, the battery SOH is estimated using XGBoost, whose

hyper-parameters are optimized based on BOA. The RMSE and MAE of estimation results are calculated to evaluate the proposed method. Besides, TreeSHAP is used for explainability analysis to quantificationally reveal the impact that HIs have on the SOH estimation results.

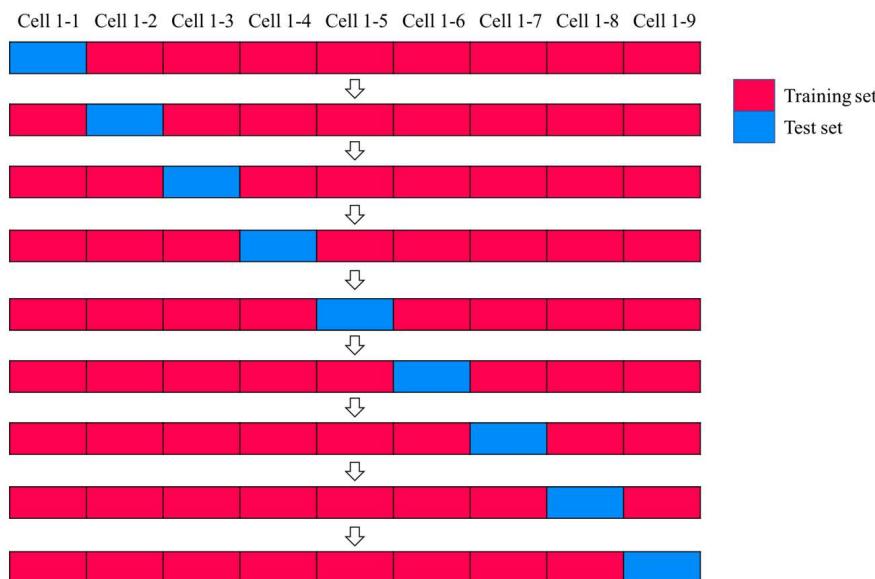


**Figure 15.** The general framework of the proposed SOH estimation method.

## 5. Results and Discussion

### 5.1. The Verification of SOH Estimation

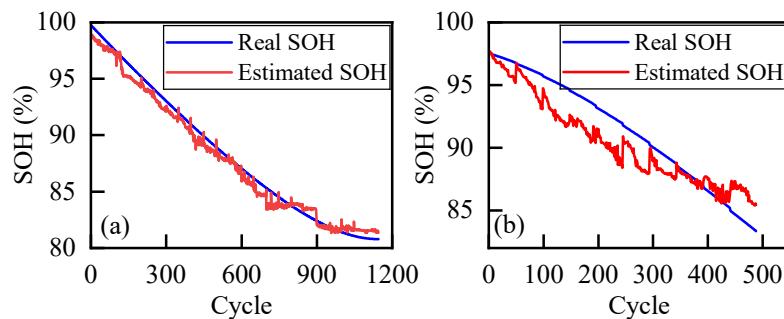
The training/test set-splitting strategy of this study is shown in Figure 16. For convenience, the figure only shows the strategy of Dataset I, but the training based on Dataset II utilizes the same strategy. In this study, for each time of model training, one cell in the training set is selected as the test set, and the other cells are used as the training set. Using the training set data as model input, BOA is used to optimize the hyper-parameters of the XGBoost model, and a set of optimal hyper-parameters is obtained. SOH estimation results are obtained on the test set using the optimal hyper-parameters.



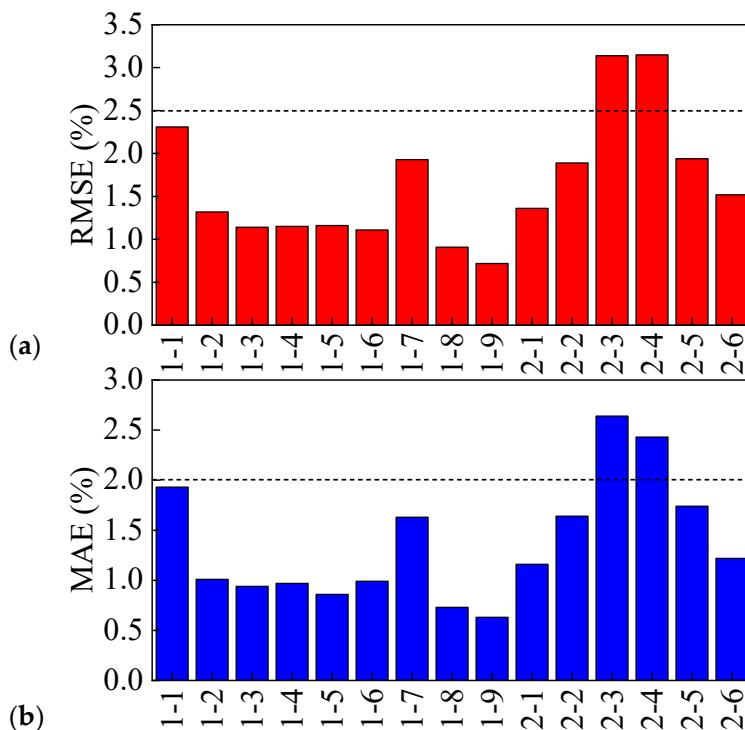
**Figure 16.** Training–test split of the dataset.

Based on the HIs extracted as stated above and the training–test splitting strategy, the SOH estimation results of cells 1–9 and 2–2 obtained by the aforementioned BOA-XGBoost algorithm are shown in Figure 17. It can be seen from Figure 17a that the SOH estimation results of cell 1–9 achieve high accuracy almost throughout the aging process, with a maximum error of less than 2%. In addition, the RMSE and MAE are 0.72% and 0.63%, respectively. As for cell 2–2, the SOH estimation results show some fluctuations during the

aging process, but still achieve high overall accuracy, with RMSE of 1.89% and MAE of 1.64%. The estimation metrics of all cells are shown in Figure 18. As can be seen from the figure, except for cells 2-3 and 2-4, the RMSE and MAE of SOH estimation for most cells is less than 2.5% and 2%, respectively. This proves the estimation accuracy of the proposed BOA-XGBoost algorithm.

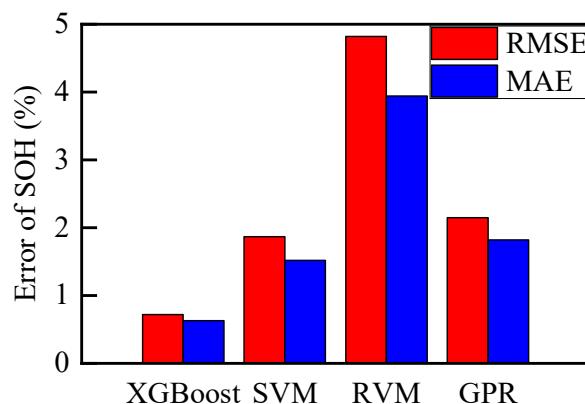


**Figure 17.** SOH estimation results using BOA-XGBoost: (a) cell 1-9; (b) cell 2-2.



**Figure 18.** SOH estimation error: (a) RMSE; (b) MAE.

In battery SOH estimation, the commonly used machine learning methods include SVM, RVM, and GPR. Taking the SOH estimation of cell 1-9 as an example, the estimation results of BOA-XGBoost, SVM, RVM, and GPR algorithm are compared. The SOH estimation errors of these four algorithms are shown in Figure 19. It can be seen that the RMSE of RVM is over 5% and that of SVM and GPR are around 2%. However, XGBoost has a RMSE of less than 1%. The comparison of MAE indicates a similar result. The proposed XGBoost method has superior performance in terms of accuracy compared to SVM, RVM, and GPR. Additionally, a comparative analysis of several previous studies is conducted, as is shown in Table 3. Methods that are compared include SVM, GPR, and CNN. The average RMSE of all 15 cells is used for the comparison. The comparative analysis result also indicates the proposed method's superior performance.



**Figure 19.** SOH estimation error comparison of XGBoost, SVM, RVM, and GPR.

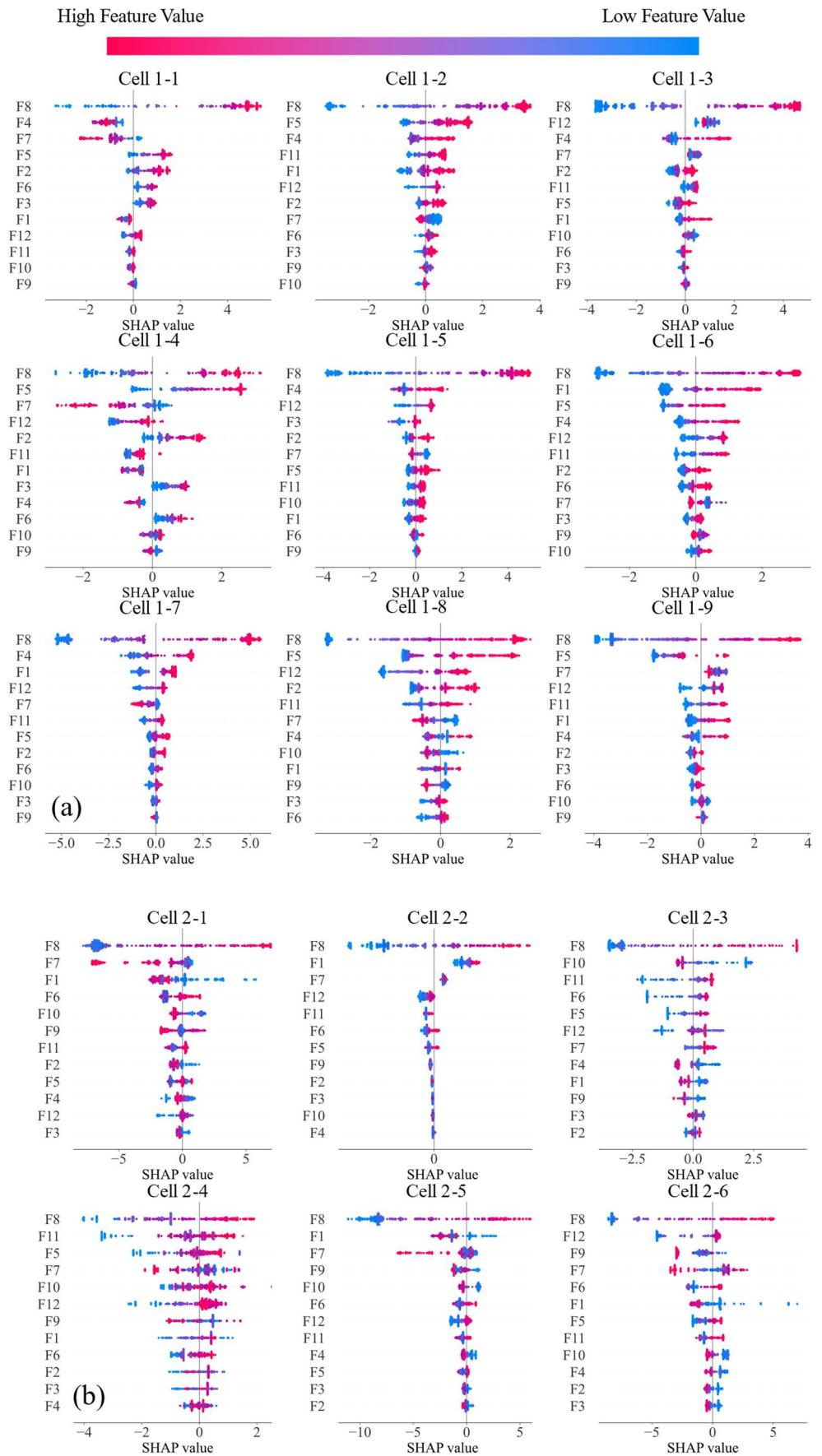
**Table 3.** Comparison of RMSE with previous studies.

Reference	Method	RMSE/%
Proposed method	BOA-XGBoost	1.65 *
[50]	SVM	2.76
[51]	GPR	2.92
[52]	CNN	2.93

\* The average RMSE of all cells.

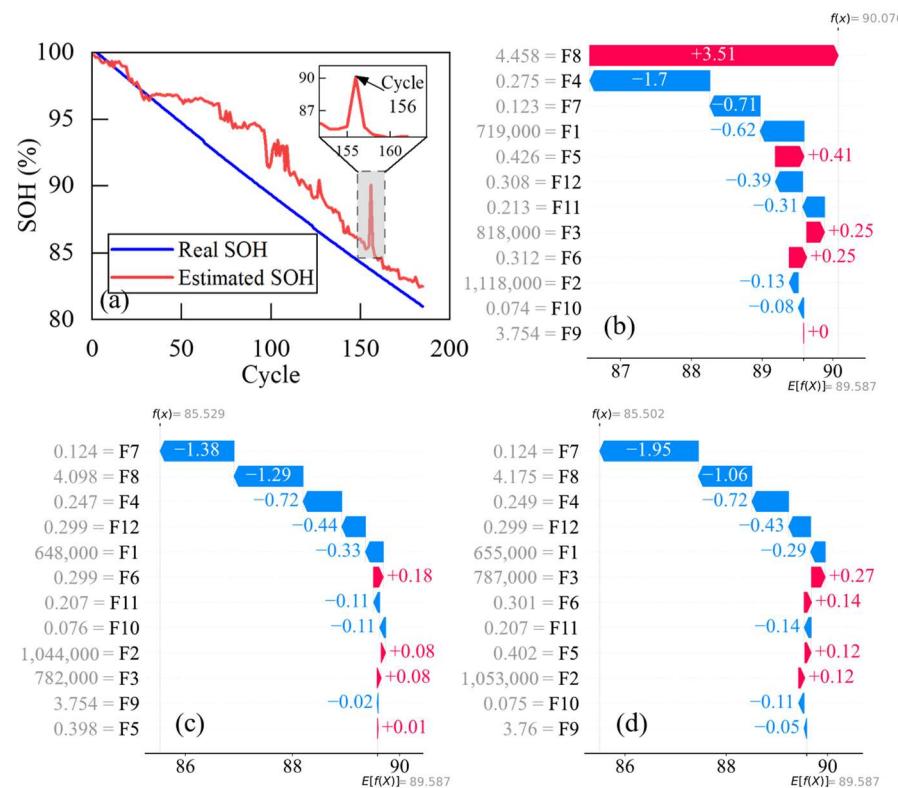
### 5.2. Explainability Analysis Based on TreeSHAP

The SHAP can be used for explainability analysis of the model under all samples and analysis of the predicted value at a single sample point, called global interpreter and local interpreter, respectively. The SHAP summary plots are shown in Figure 20, which shows the overview of the SHAP value for each HI. Each point in the figure represents a sample point. The color of the points indicates the value of HI. As shown in the color bar at the top of Figure 20, sample points with higher HI values are indicated by red points, while those with lower HI values are blue. In addition, the X-axis indicates the SHAP value. Sample points with higher SHAP values are located on the right side of the figure, while those with lower SHAP values are on the left. Therefore, the wider the points of a HI are distributed along the X-axis, the greater its influence on the SOH estimation result. HIs are arranged from top to bottom in Figure 20 by the average absolute value of the SHAP value. It can be seen that F8(PVIC) has the largest average absolute value of SHAP, which means that it has the greatest impact on the results of SOH estimation. Moreover, the red points of F8 are always located on the right side of the figure, while the blue points on the left, indicating that F8 has a positive contribution to SOH estimation results. In addition, for cell 2-3, it can be seen from Figure 20b that many points are stacked on the same horizontal coordinate. Although the HI values of different sample points vary, the SHAP values of these sample points are the same, which means that the contribution to the SOH estimation results is the same. This indicates that, when estimating the SOH of cell 2-3, the model does not capture changes in the HIs of cell 2-3 as well as other cells, resulting in a relatively high error. With the global explainability analysis based on TreeSHAP, the differences in the contributions of different HI to the SOH estimation can be quantitatively analyzed.

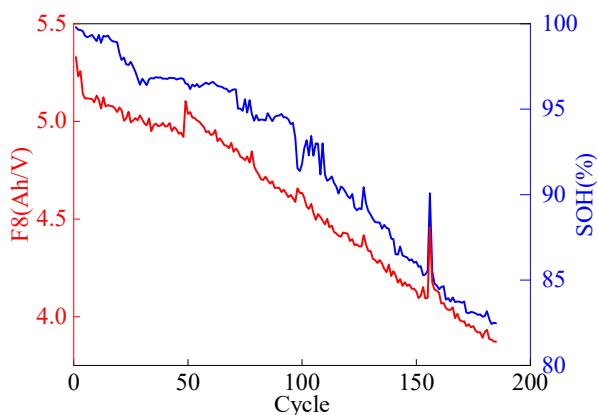


**Figure 20.** SHAP summary plots: **(a)** Dataset I; **(b)** Dataset II.

In the SOH estimation results obtained in the previous section, it can be found that there are some outliers. For outliers in SOH estimation, the local interpreter in SHAP can be used for explainability analysis. It can be found in Figure 21a that there is an obviously abnormal estimation value, cycle 156 in cell 1-1, with an estimated SOH of 90.076%. In contrast, the estimated SOHs of cycle 155 and 157 are 85.529% and 85.502%, respectively. In other words, there is an outlier at cycle 156 that is approximately 4.5% higher than the normal value. SHAP waterfall plots are used for local explainability analysis. They quantitatively explain the SOH estimation result at a specific sample point. Numerically, the waterfall plot splits the deviation between the estimated SOH at a particular sample point and the mean SOH for all sample points and shows the contribution value for each HI. In other words, the deviation equals the sum of the SHAP values of all HIs. The SHAP waterfall plots of cycles 156, 155, and 157 are shown in Figure 21b–d, respectively. On the left side of the figure are the values of each HI. The arrows and the numbers in/next to them represent the SHAP value of each HI. It can be found that the F8 values of cycles 155 and 157 are 4.098 and 4.175, respectively. However, cycle 156 has an abnormally high F8 value of 4.458. And the SHAP value of F8 of cycle 156 (+3.51) is also abnormally higher than that of cycle 155 (−1.29) and 157 (−1.06). In addition, the SHAP values of other HIs fluctuate within the normal range. Therefore, it can be judged that the SOH estimation outlier of cycle 156 is caused by the abnormal fluctuation of F8. Figure 22 shows the F8 value and SOH estimation result of cell 1-1. It can be verified that there is an abnormal fluctuation of the F8 value at cycle 156. This supports the result of the local interpreter. By using the local interpreter in SHAP to analyze the explainability of the sample point where the outlier is located, the quantitative analysis results of the influence of HIs on the estimated SOH can be obtained. It can reveal the causes of the outlier and enhance the explainability of the model.



**Figure 21.** SHAP analysis for abnormal estimation value: (a) outlier in cell 1-1; (b–d): SHAP waterfall plot of cycles 156, 155, and 157.



**Figure 22.** F8 and SOH estimation result of cell 1-1.

## 6. Limitations and Outlook

The proposed explainable battery SOH estimation method based on BOA-XGBoost and TreeSHAP realizes accurate SOH estimation on the datasets containing 15 cells under different aging conditions. Nevertheless, there are still some limitations of the proposed approach, which can also be considered as the research direction of future studies.

1. In this study, 12 HIs are extracted as feature datasets, but the number of features is relatively small. In particular, since the battery aging experiment in this study is a CC charge/discharge process, the extracted HIs are also limited to the CC charging and rest process. However, at present, more and more electric vehicles have begun to adopt fast charging technology. The charging current in the fast-charging scenario is often dynamic and time-varying. The feature extraction method in this study needs considerable improvement to adapt to dynamic charging conditions in fast-charging scenarios to meet with requirements of implementation on electric vehicles.
2. In this study, HI extraction, model optimization, and training are carried out offline. However, in practical applications, offline SOH estimation methods are difficult to deploy on electric vehicles. Therefore, future research should focus on online estimation methods to achieve higher practicability.
3. The research object of this paper is cells, but in practical application, lithium-ion batteries are mostly assembled into battery packs in electric vehicles. Potential problems for the battery packs consist of inconsistent capacity and characteristics between cells, leading to difficulty in applying the proposed method. Therefore, the proposed battery SOH estimation method should take battery inconsistency into account and improve performance in such scenarios.
4. The explainability analysis conducted in this study can explain the SOH estimation results by giving a quantitative analysis of the contribution of each HI. However, it is currently not possible to directly relate SOH estimation results to aging mechanisms using this explainability analysis method. Future research should focus on extending explainability analysis to the aging trajectories and mechanisms to further improve the explainability and transparency of data-driven methods. One possible solution is to combine explainability analysis methods with diagnostic methods for battery aging mechanisms. Specifically, the explainability analysis is used to reveal the trend of HIs. Thus, the aging diagnosis method is used to judge the aging mechanism of the battery. In this way, explainability analysis can support BMS to diagnose the aging mechanisms.

## 7. Conclusions

Battery SOH is one of the key parameters of the lithium-ion battery, which represents the remaining capacity of the battery. The accurate estimation of battery SOH is crucial to battery management and electric vehicles. This study proposes an adaptive and explainable

SOH estimation method based on BOA-XGBoost and TreeSHAP. 12 HIs highly correlated with the degradation of battery capacity are extracted from the charging curves. Using each cell in turn as the test set and the others as the training set, the BOA is used to optimize the hyper-parameters of the XGBoost model. The proposed method shows superior accuracy and robustness on the datasets containing 15 cells under different aging conditions, with a maximum RMSE of 3.15% and MAE of 2.64%. Moreover, based on TreeSHAP, the explainability analysis reveals the impact that HIs have on battery SOH estimation results, and the cause of abnormal estimated SOH is uncovered. This enhances the explainability and transparency of the XGBoost model.

**Author Contributions:** Conceptualization, Z.X., B.J. and H.D.; methodology, Z.X. and B.J.; software, Z.X., B.J. and J.Z.; validation, B.J. and X.W.; formal analysis, B.J. and X.W.; investigation, B.J. and H.D.; data curation, J.Z. and B.J.; writing—original draft preparation, Z.X., B.J. and X.W.; writing—review and editing, J.Z. and H.D.; visualization, B.J. and X.W.; supervision, J.Z. and H.D.; project administration, H.D.; funding acquisition, B.J. and H.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is financially supported by the National Natural Science Foundation of China (NSFC, Grant Nos. 52307248 and U20A20310) and Shanghai Rising-Star Program (Grant No. 22YF1450400).

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

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Berecibar, M.; Gandiaga, I.; Villarreal, I.; Omar, N.; Van Mierlo, J.; Van den Bossche, P. Critical review of state of health estimation methods of Li-ion batteries for real applications. *Renew. Sust. Energy Rev.* **2016**, *56*, 572–587. [[CrossRef](#)]
- Cadini, F.; Sbarufatti, C.; Cancelliere, F.; Giglio, M. State-of-life prognosis and diagnosis of lithium-ion batteries by data-driven particle filters. *Appl. Energy* **2019**, *235*, 661–672. [[CrossRef](#)]
- Zackrisson, M.; Fransson, K.; Hildenbrand, J.; Lampic, G.; O'Dwyer, C. Life cycle assessment of lithium-air battery cells. *J. Clean. Prod.* **2016**, *135*, 299–311. [[CrossRef](#)]
- Peng, Y.; Liu, D. Data-driven prognostics and health management: A review of recent advances. *Chin. J. Sci. Instrum.* **2014**, *35*, 481–495.
- Lipu, M.S.H.; Hannan, M.A.; Hussain, A.; Hoque, M.M.; Ker, P.J.; Saad, M.H.M.; Ayob, A. A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations. *J. Clean. Prod.* **2018**, *205*, 115–133. [[CrossRef](#)]
- Tagade, P.; Hariharan, K.S.; Ramachandran, S.; Khandelwal, A.; Naha, A.; Kolake, S.M.; Han, S.H. Deep Gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis. *J. Power Sources* **2020**, *445*, 14. [[CrossRef](#)]
- Sauer, D.U.; Wenzl, H. Comparison of different approaches for lifetime prediction of electrochemical systems—Using lead-acid batteries as example. *J. Power Sources* **2008**, *176*, 534–546. [[CrossRef](#)]
- Xue, C.Y.; Jiang, B.; Zhu, J.G.; Wei, X.Z.; Dai, H.F. An Enhanced Single-Particle Model Using a Physics-Informed Neural Network Considering Electrolyte Dynamics for Lithium-Ion Batteries. *Batteries* **2023**, *9*, 511. [[CrossRef](#)]
- Chen, S.Q.; Zhang, Q.; Wang, F.C.; Wang, D.F.; He, Z.Q. An electrochemical-thermal-aging effects coupled model for lithium-ion batteries performance simulation and state of health estimation. *Appl. Therm. Eng.* **2024**, *239*, 17. [[CrossRef](#)]
- Huang, P.; Gu, P.W.; Kang, Y.Z.; Zhang, Y.; Duan, B.; Zhang, C.H. The state of health estimation of lithium-ion batteries based on data-driven and model fusion method. *J. Clean. Prod.* **2022**, *366*, 12. [[CrossRef](#)]
- Zhang, X.; Wang, Y.J.; Liu, C.; Chen, Z.H. A novel approach of battery pack state of health estimation using artificial intelligence optimization algorithm. *J. Power Sources* **2018**, *376*, 191–199. [[CrossRef](#)]
- Qiao, D.D.; Wei, X.Z.; Jiang, B.; Fan, W.J.; Lai, X.; Zheng, Y.J.; Dai, H.F. Quantitative Diagnosis of Internal Short Circuit for Lithium-Ion Batteries Using Relaxation Voltage. *IEEE Trans. Ind. Electron.* **2024**, *10*, 13201–13210. [[CrossRef](#)]
- Hu, C.; Jain, G.; Schmidt, C.; Strief, C.; Sullivan, M.; ASME. Online Estimation of Lithium-Ion Battery Capacity Using Sparse Bayesian Learning. In Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Boston, MA, USA, 2–5 August 2015.
- Deng, Y.W.; Ying, H.J.; Jiaqiang, E.; Zhu, H.; Wei, K.X.; Chen, J.W.; Zhang, F.; Liao, G.L. Feature parameter extraction and intelligent estimation of the State-of-Health of lithium-ion batteries. *Energy* **2019**, *176*, 91–102. [[CrossRef](#)]
- Feng, X.N.; Weng, C.H.; He, X.M.; Wang, L.; Ren, D.S.; Lu, L.G.; Han, X.B.; Ouyang, M.G. Incremental Capacity Analysis on Commercial Lithium-Ion Batteries using Support Vector Regression: A Parametric Study. *Energies* **2018**, *11*, 2323. [[CrossRef](#)]
- Li, X.Y.; Yuan, C.G.; Li, X.H.; Wang, Z.P. State of health estimation for Li-Ion battery using incremental capacity analysis and Gaussian process regression. *Energy* **2020**, *190*, 11. [[CrossRef](#)]

17. Jiang, B.; Zhu, Y.L.; Zhu, J.G.; Wei, X.Z.; Dai, H.F. An adaptive capacity estimation approach for lithium-ion battery using 10-min relaxation voltage within high state of charge range. *Energy* **2023**, *263*, 11. [[CrossRef](#)]
18. Li, Y.; Zou, C.F.; 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](#)]
19. Wu, J.; Cui, X.C.; Zhang, H.; Lin, M.Q. Health Prognosis With Optimized Feature Selection for Lithium-Ion Battery in Electric Vehicle Applications. *IEEE Trans. Power Electron.* **2021**, *36*, 12646–12655. [[CrossRef](#)]
20. Chen, L.; Ding, Y.H.; Wang, H.M.; Wang, Y.J.; Liu, B.H.; Wu, S.X.; Li, H.; Pan, H.H. Online Estimating State of Health of Lithium-Ion Batteries Using Hierarchical Extreme Learning Machine. *IEEE Trans. Transp. Electrif.* **2022**, *8*, 965–975. [[CrossRef](#)]
21. Li, Y.Y.; Sheng, H.M.; Cheng, Y.H.; Kuang, H.J.; IEEE. Lithium-ion battery state of health monitoring based on ensemble learning. In Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Auckland, New Zealand, 20–23 May 2019; pp. 554–559.
22. Qian, C.; Xu, B.; Chang, L.; Sun, B.; Feng, Q.; Yang, D.; Ren, Y.; Wang, Z. Convolutional neural network based capacity estimation using random segments of the charging curves for lithium-ion batteries. *Energy* **2021**, *227*, 120333. [[CrossRef](#)]
23. Shen, S.; Sadoughi, M.; Li, M.; Wang, Z.; Hu, C. Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries. *Appl. Energy* **2020**, *260*, 114296. [[CrossRef](#)]
24. Li, W.; Sengupta, N.; Dechent, P.; Howey, D.; Annaswamy, A.; Sauer, D.U. Online capacity estimation of lithium-ion batteries with deep long short-term memory networks. *J. Power Sources* **2021**, *482*, 228863. [[CrossRef](#)]
25. Lin, M.Q.; Zeng, X.P.; Wu, J. State of health estimation of lithium-ion battery based on an adaptive tunable hybrid radial basis function network. *J. Power Sources* **2021**, *504*, 11. [[CrossRef](#)]
26. Yao, X.Y.; Chen, G.L.; Pecht, M.; Chen, B. A novel graph-based framework for state of health prediction of lithium-ion battery. *J. Energy Storage* **2023**, *58*, 11. [[CrossRef](#)]
27. Kheirkhah-Rad, E.; Parvareh, A.; Moeini-Aghetaie, M.; Dehghanian, P. A Data-Driven State-of-Health Estimation Model for Lithium-Ion Batteries Using Referenced-Based Charging Time. *IEEE Trans. Power Deliv.* **2023**, *38*, 3406–3416. [[CrossRef](#)]
28. Jiang, B.; Tao, S.Y.; Wang, X.Y.; Zhu, J.G.; Wei, X.Z.; Dai, H.F. Mechanics-based state of charge estimation for lithium-ion pouch battery using deep learning technique. *Energy* **2023**, *278*, 11. [[CrossRef](#)]
29. Bahri, M.; Salutari, F.; Putina, A.; Sozio, M. AutoML: State of the art focusing on anomaly detection, challenges, and research directions. *Int. J. Data Sci. Anal.* **2022**, *14*, 113–126. [[CrossRef](#)]
30. Lee, G.; Kim, J.; Lee, C. State-of-health estimation of Li-ion batteries in the early phases of qualification tests: An interpretable machine learning approach. *Expert Syst. Appl.* **2022**, *197*, 12. [[CrossRef](#)]
31. Deng, Z.; Hu, X.; Lin, X.; Xu, L.; Che, Y.; Hu, L. General Discharge Voltage Information Enabled Health Evaluation for Lithium-Ion Batteries. *IEEE/ASME Trans. Mechatron.* **2020**, *26*, 1295–1306. [[CrossRef](#)]
32. Wang, F.J.; Zhao, Z.B.; Zhai, Z.; Shang, Z.G.; Yan, R.Q.; Chen, X.F. Explainability-driven model improvement for SOH estimation of lithium-ion battery. *Reliab. Eng. Syst. Saf.* **2023**, *232*, 12. [[CrossRef](#)]
33. von Bülow, F.; Hahn, Y.; Meyes, R.; Meisen, T. Transparent and Interpretable State of Health Forecasting of Lithium-Ion Batteries with Deep Learning and Saliency Maps. *Int. J. Energy Res.* **2023**, *2023*, 23. [[CrossRef](#)]
34. Kobliha, M.; Schwarz, J.; Ocenásek, J. Bayesian optimization algorithms for dynamic problems. In *Applications of Evolutionary Computing, Proceedings*; Rothlauf, F., Ed.; Lecture Notes in Computer Science; Springer: Berlin, Germany, 2006; Volume 3907, pp. 800–804.
35. Jiang, B.; Dai, H.F.; Wei, X.Z.; Jiang, Z. Multi-Kernel Relevance Vector Machine With Parameter Optimization for Cycling Aging Prediction of Lithium-Ion Batteries. *IEEE J. Emerg. Sel. Top. Power Electron.* **2023**, *11*, 175–186. [[CrossRef](#)]
36. Zhu, J.G.; Knapp, M.; Sorensen, D.R.; Heere, M.; Darma, M.S.D.; Müller, M.; Mereacre, L.; Dai, H.F.; Senyshyn, A.; Wei, X.Z.; et al. Investigation of capacity fade for 18650-type lithium-ion batteries cycled in different state of charge (SoC) ranges. *J. Power Sources* **2021**, *489*, 12. [[CrossRef](#)]
37. Liu, S.J.; Winter, M.; Lewerenz, M.; Becker, J.; Sauer, D.U.; Ma, Z.Y.; Jiang, J.C. Analysis of cyclic aging performance of commercial  $\text{Li}_4\text{Ti}_5\text{O}_{12}$ -based batteries at room temperature. *Energy* **2019**, *173*, 1041–1053. [[CrossRef](#)]
38. Atalay, S.; Sheikh, M.; Mariani, A.; Merla, Y.; Bower, E.; Widanage, W.D. Theory of battery ageing in a lithium-ion battery: Capacity fade, nonlinear ageing and lifetime prediction. *J. Power Sources* **2020**, *478*, 8. [[CrossRef](#)]
39. Han, X.B.; Ouyang, M.G.; Lu, L.G.; Li, J.Q.; Zheng, Y.J.; Li, Z. A comparative study of commercial lithium ion battery cycle life in electrical vehicle: Aging mechanism identification. *J. Power Sources* **2014**, *251*, 38–54. [[CrossRef](#)]
40. Anseán, D.; García, V.M.; González, M.; Blanco-Viejo, C.; Viera, J.C.; Pulido, Y.F.; Sánchez, L. Lithium-Ion Battery Degradation Indicators Via Incremental Capacity Analysis. *IEEE Trans. Ind. Appl.* **2019**, *55*, 2992–3002. [[CrossRef](#)]
41. Birkl, C.R.; McTurk, E.; Zekoll, S.; Richter, F.H.; Roberts, M.R.; Bruce, P.G.; Howey, D.A. Degradation Diagnostics for Commercial Lithium-Ion Cells Tested at  $-10^{\circ}\text{C}$ . *J. Electrochem. Soc.* **2017**, *164*, A2644–A2653. [[CrossRef](#)]
42. Qiao, D.D.; Wei, X.Z.; Jiang, B.; Fan, W.J.; Gong, H.; Lai, X.; Zheng, Y.J.; Dai, H.F. Data-Driven Fault Diagnosis of Internal Short Circuit for Series-Connected Battery Packs Using Partial Voltage Curves. *IEEE Trans. Ind. Inform.* **2024**, *20*, 6751–6761. [[CrossRef](#)]
43. Fly, A.; Chen, R. Rate dependency of incremental capacity analysis ( $dQ/dV$ ) as a diagnostic tool for lithium -ion batteries. *J. Energy Storage* **2020**, *29*, 13. [[CrossRef](#)]

44. Chen, T.Q.; Guestrin, C.; Assoc Comp, M. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
45. Shi, Y.; Li, J.; Ren, J.; Zhang, K. Prediction of residual service life of lithium-ion battery using WOA-XGBoost. *Energy Storage Sci. Technol.* **2022**, *11*, 3354–3363.
46. Zhang, B.; Wei, Z.Y.; Ren, J.D.; Cheng, Y.Q.; Zheng, Z.Q. An Empirical Study on Predicting Blood Pressure Using Classification and Regression Trees. *IEEE Access* **2018**, *6*, 21758–21768. [[CrossRef](#)]
47. Harinarayan, R.R.A.; Shalinie, S.M. XFDDC: Explainable Fault Detection Diagnosis and Correction framework for chemical process systems. *Process Saf. Environ. Protect.* **2022**, *165*, 463–474. [[CrossRef](#)]
48. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
49. Molnar, C.; Casalicchio, G.; Bischl, B. Interpretable Machine Learning—A Brief History, State-of-the-Art and Challenges. In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), Electr Network, Ghent, Belgium, 14–18 September 2020; pp. 417–431.
50. Deng, Z.W.; Yang, L.; Cai, Y.S.; Deng, H.; Sun, L. Online available capacity prediction and state of charge estimation based on advanced data-driven algorithms for lithium iron phosphate battery. *Energy* **2016**, *112*, 469–480. [[CrossRef](#)]
51. Deng, Z.W.; Hu, X.S.; Lin, X.K.; Che, Y.H.; Xu, L.; Guo, W.C. Data-driven state of charge estimation for lithium-ion battery packs based on Gaussian process regression. *Energy* **2020**, *205*, 11. [[CrossRef](#)]
52. Li, Y.H.; Li, K.; Liu, X.; Zhang, L. Fast battery capacity estimation using convolutional neural networks. *Trans. Inst. Meas. Control* **2020**, *14*, 014233122096425. [[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.