MDPI

Article

Flexible Deep Learning-Based State of Health Estimation of Lithium-Ion Batteries with Features Extracted from Partial Charging Curves

Rucong Lai [1], Xiaoyu Li [2] and Jie Wang [2,3],*

1   Institute of Applied Physics and Materials Engineering, University of Macau, Macao 999078, China; yc17838@um.edu.mo
2   Key Laboratory of Optoelectronic Devices and Systems, Ministry of Education and Guangdong Province, College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China; xiaoyu070220202@szu.edu.cn
3   Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen 518060, China
*   Correspondence: jiewang0806@szu.edu.cn

Abstract: The state of health is a crucial state that suggests the capacity of lithium-ion batteries to store and restitute energy at a certain power level, which should be carefully monitored in the battery management system. However, the state of health of batteries is unmeasurable and, currently, it is usually estimated within a specific area of the whole charging data, which is very limited in practical application because of the incomplete and random charging behaviors of users. In this paper, we intend to estimate the state of health of batteries with flexible partial charging curves and normal multi-layer perceptron based on the degradation data of eight 0.74 Ah batteries. To make the estimation more adaptive and flexible, we extract several features from partial charging curves. Analysis of the relationship between extracted features and the state of health shows that the extracted features are useful in estimation. As the length of the partial charging curve increases, the extracted features still function well, and the root mean square error of the test set is lower than 1.5%. Further validation on the other two types of batteries reveals that the proposed method achieves high accuracy even with different sampling and working conditions. The proposed method offers an easy-to-implement way to achieve an accurate estimation of a battery's state of health.

Keywords: state of health; lithium-ion batteries; partial charging curves; deep learning

1. Introduction

1.1. Literature Review

Lithium-ion batteries (LIBs) have revolutionized our daily lives, enabling deeper penetration of energy storage in power systems, and are the technology of choice for electric vehicles [1]. Due to the complex electrochemical process inside batteries [2], monitoring different battery states using measurable signals is impractical to ensure the safety of wide-range battery applications. Particularly, the state of health (SOH), which is defined as the ratio between the current maximum capacity and the initial maximum capacity, reveals the aging status of LIBs and helps optimize the present working conditions [3]. Since SOH cannot be measured directly by common sensors, its estimation based on measurable variables, such as current, voltage, and temperature, has attracted extensive attention in both the academic and industrial communities. At present, two commonly used methods for SOH are model-based methods and data-driven methods.

Model-based methods have a longstanding tradition in the realm of SOH estimation, evident through instances such as electrochemical models and equivalent circuit models (ECMs). Electrochemical models characterize the electrochemical behavior of batteries [4], particularly using electrochemical impedance spectroscopy (EIS), which applies a small

amplitude alternating current (AC) signal to the battery and measures the resulting voltage across different frequencies. Marvin et al. [5] proposed a method that combines fractional order impedance modeling and short-term relaxation effects with EIS characterization for rapid SOH determination. Jiang et al. [6] extracted features from EIS incorporating Gaussian process regression (GPR) to estimate SOH. Impedance variations resulting from battery degradation can serve as indicators of its health. Nonetheless, a collection of parameters within electrochemical models poses challenges in terms of measurement or identification, particularly for commercial batteries [7]. Through the integration of SOH as a model parameter, ECMs are able to estimate SOH by minimizing the error in voltage simulation [8–10]. ECMs integrated with nonlinear filters are widely used because of their self-correction capability and insensitiveness to initial states [11]. Besides, SOH is coupled with other states in practical situations. Liu et al. [12] proposed a joint estimator based on ECM, which improved the SOH estimation performance. However, ECM has the problem of poor model generalization towards the wider range of working conditions [13]. Additionally, the existing ECM methods are reported to be inapplicable under high current rates or low-temperature conditions [14]. To circumvent the challenges faced by model-based methods, data-driven approaches have been introduced to the field of SOH estimation.

Data-driven methods disregard the internal physical models of the battery and instead estimate SOH solely based on collected data. As an illustration, Li et al. [15] employed incremental capacity (IC) analysis, derived from the charging voltage profile, to forecast SOH. They applied Gaussian Process Regression (GPR) to smoothen the IC curve and facilitate feature extraction. Additionally, in studies such as [16,17], the differential voltage (DV) curve has been utilized to extract inflection points that exhibit a clear connection with battery degradation. Commonly, features including peak height [18], position [19], and area [20] have been discerned as IC curves or DV curves, which are derived from constant-current charging profiles or discharging profiles. The above methods make full use of the charging curve or discharging curve to extract features, while in the study [21], voltage relaxation features are used to estimate the SOH of commercial batteries with traditional machine learning methods. These data-driven techniques exhibit notable accuracy in estimation, and the extracted features have demonstrated their utility. However, acquiring these features in practical scenarios can be challenging, despite the potential convenience of cloud storage for dataset processing [22]. Moreover, these features are extracted locally and might magnify the noise present in charging/discharging profiles. Hence, more flexible ways for SOH estimation are further required. Richardson et al. [23] endeavored to estimate SOH by directly utilizing charging data of a fixed length, which was sampled at a specified initial voltage. Tian et al. [24] used a flexible partial voltage-capacity curve to estimate the whole charging curve as well as the maximum capacity. This method indicated that deep learning methods can autonomously extract features from raw data and then map these feature vectors to corresponding SOH values. Nevertheless, since measured signals may be recorded using frequency range rather than voltage span, a joint estimator using the one-dimensional convolutional network to estimate electrode aging states and state of charge (SOC) has been proposed which used a partial time-voltage curve [25] as input. These methods significantly reduced the data processing time and offered versatile approaches to estimation.

1.2. Motivation and Contribution

A fundamental challenge in the data-driven methods for SOH estimation pertains to obtaining readily accessible features. Although some studies have extracted health-related features to achieve accurate SOH estimation [15,16,21], most require a pre-set voltage or voltage-relaxation stage. For instance, Zhu et al. [21] investigated the substantial linear correlation between the maximum voltage during the voltage-relaxation phase and SOH in commercial batteries. They employed diverse machine learning pipelines for SOH estimation. Extracted features from IC/DV analysis [15,16] have also demonstrated

their efficacy in data-driven SOH estimation. However, the aforementioned approaches necessitate a comprehensive charging-discharging profile before feature extraction. Such an assumption also did not consider the influence of user habits [26] and might not be adaptable to real-world scenarios. Therefore, it is urgent to develop a flexible SOH estimator with easily accessible features such as partial charging curves. Previous studies [24,25] have noticed this limitation and pioneered the SOH estimation with a piece of raw voltage curve during the constant current (CC) charging phase. The framework is easy to implement and does not require additional dataset processes. Across various battery datasets, the proposed method has consistently achieved remarkable estimation outcomes, but with different input lengths. As an illustration, the Oxford dataset [27] necessitated 400 data points, whereas other datasets only required 40 points. This variation arises due to the distinct time intervals at which these datasets were recorded. Differences in input length may change the structure of the deep learning framework and are not able to transfer when the collecting time interval is changed. Furthermore, extensive raw datasets often encompass noise that can impact the ultimate estimation performance. To mitigate this, segments of the voltage curve could be extracted into features of fixed lengths. By employing these standardized, length-based, aging-related features, the structure of data-driven methodologies can be streamlined, enabling the application of diverse machine learning techniques.

Partial voltage curves from the CC charge stage are varied with different voltage ranges; hence, the extracted features need to represent the local as well as the global representation of pieces of the voltage curve. Previous studies [15,16] used statistical methods to find related aging features, but most of them belong to local features such as maximum voltages or peaks of curves. Therefore, it is urgent to extract global representation from partial voltage curves without information loss. With these improvements, the structure of deep learning methods could be simplified. The contribution of this paper could be listed as follows:

(1) The proposed method employs a straightforward multi-perceptron layer (MLP) for accomplishing SOH estimation via randomly sampled short-term charging voltage curves. This approach extracts several features from a statistical standpoint and does not require a designated voltage range for feature extraction. As a result, it offers exceptional flexibility concerning input ranges and a consistent number of features for the MLP model.

(2) The introduced approach extracts features both locally and globally from randomly sampled charging voltage curves. This is achieved by identifying specific points of curves and employing curve fitting parameters. The method employs multiple parameters to characterize the curves, which offer diverse insights compared to other proposed feature extraction techniques. Experimental findings indicate that incorporating these local and global features enhances estimation results when compared to utilizing raw datasets alone. The paper also underscores the significance of the introduced features.

(3) The validity of the proposed method is demonstrated across three distinct battery datasets, encompassing varying operational conditions and sampling frequencies. Despite the extended curve lengths due to feature extraction, the number of inputs for the MLP remains constant. Notably, results indicate enhanced accuracy as curve lengths increase. Furthermore, the MLP exhibits superior generalization capabilities in comparison to other machine learning techniques.
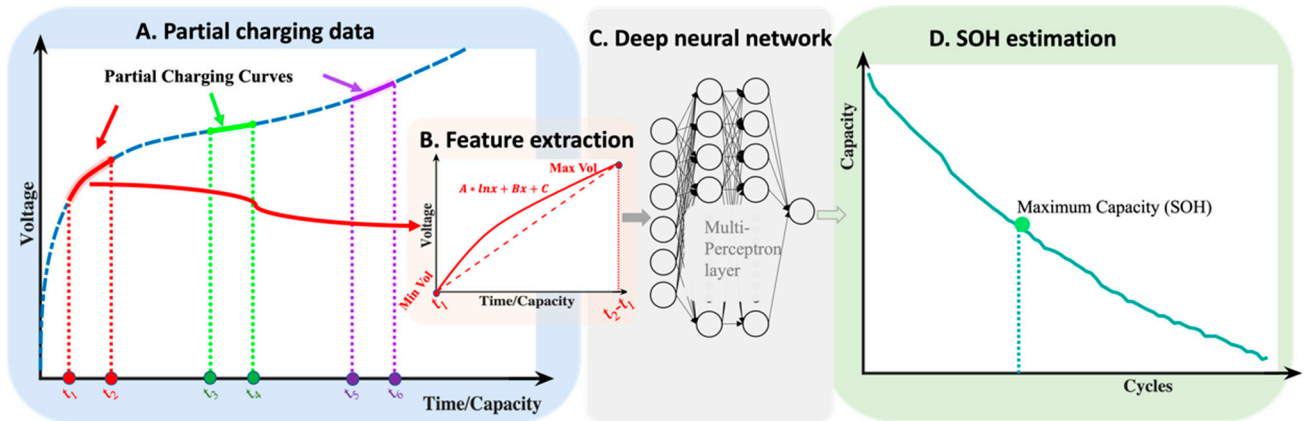
2. Method

2.1. Problem Definition

The whole framework of the proposed method is shown in Figure 1, which contains four sections, including partial charging data extraction, feature extraction, deep neural network, and SOH estimation. At first, the battery SOH is generally defined as:

$$\text{SOH} = \frac{q}{Q} \tag{1}$$

where q is the maximum capacity at the current cycle and Q is the initial maximum capacity.

Figure 1. Framework of the proposed method: (**A**) Random partial charging data of CC stage extraction. (**B**) Global and local feature extraction of partial charging data. (**C**) Deep learning structure. (**D**) SOH estimation of the current cycle.

A charging curve can be characterized by the voltage $V(t)$ that rises from the lower voltage threshold to the upper voltage limit with increased charging time t. $V(t)$ is then stated as $V(t) = [V(0), V(\Delta t), \ldots, V(N\Delta t)]$, where Δt represents the time interval for data collection and N denotes the total number of points after V reaches the upper voltage limit. In real-world scenarios, it is not typical for an electric vehicle to be fully charged from a completely depleted state [28]; instead, partial charging is a more common occurrence. To mimic the incomplete charge in real scenarios, flexible partial charging curves are extracted to estimate the SOH of the current cycle in this article. Given the fixed period T, the number of points of a partial charging curve is $M = T/\Delta t (M < N)$, so a portion of a charging curve could be represented by $H = [V(0), V(\Delta t), \ldots, V(M\Delta t)]$, where the time at the starting point is 0.

The SOH estimation problem can be formulated as follows: Given a battery that has been cycled for n times, flexible charging data is firstly extracted, feature-engineered, and then used to estimate the corresponding SOH, which can be written as:

$$\text{SOH}_{n,estimate} = f_{MLP}(H_{n,\ 0:M\Delta t}) \tag{2}$$

where $H_{0:M\Delta t}$ is a random piece of charging voltage data at the nth cycle covering M points and f_{MLP} denotes the deep learning method that is used to map the nonlinear mapping between the partial charging curve and the SOH of the nth cycle.

2.2. Feature Extraction and Analysis

Upon obtaining voltage and time samples from partial charging curves of the CC phase, certain studies [24,25] input the raw samples into estimators to forecast battery states, thereby alleviating data processing demands. Nonetheless, it is worth noting that recording frequencies for battery operating conditions can vary. For instance, the Oxford battery dataset records the voltage during the CC phase every second, while the CALCE dataset [29,30] does so every 30 s, and the NASA dataset [31] records it every 3 s. This variation may change the estimator's structure and heavy computations are introduced with long input sequences. Since the partial charging curves are stable, several features are extracted to represent the random partial charging curves. Previous studies [32,33] consider using summarizing statistics to illustrate the shape and position change of the voltage curve, such as maximum voltage, minimum voltage, and variance of voltage. These features could locally denote the statics of voltage curves with information loss, so global representations of voltage curves are urgent to explore.

A piece of charging data from the Oxford dataset is shown in Figure 1B, which exhibits a high degree of smoothness without noticeable jitter, approaching linearity while retaining a subtle curvature. Referring to the previous studies [21,34], the minimum voltage, maximum voltage, and dQ/dV are extracted from the partial charging curve. Apart from these local features, global features that show the curvature of the curve are introduced with the following equation:

$$A * lnx + B * x + c \tag{3}$$

To denote the curvature of the curve, the *ln* function is introduced, and the left items are for linear expression. Through clarifying the parameters, *A*, *B*, and *C*, the representation of the partial charging data is given. The trends of the mentioned features of the Oxford dataset, along with battery aging, are shown in Figure 2. Partial charging curves that start from 200 s, 300 s, 1200 s, and 2200 s are drawn, respectively, with the same sequence length of 300 s. Minimum voltage, maximum voltage, and parameter *C* rise with increased cycles due to the phenomenon of faster convergence to upper limit voltage because of the battery cell aging [33]. Parameter *C* has the same trend as the minimum voltage because *C* denotes the normalized minimum voltage when realizing the curve fitting. To ensure unbiased estimation and the corresponding scale, both the minimum voltage feature and the C have been retained. Figure 2a,b,f also shows that charged curves that started from 2200 s have steep trends. In previous research [15,33], feature dQ/dV has an obvious drop in specific voltage ranges; thus, these specific voltage ranges are picked to estimate SOH. In this paper, feature dQ/dV is extracted from every piece of charging data, with other features used as additional information when feature dQ/dV is not obvious (shown as a red triangle in Figure 2c). Curvature item *A* and linear item *B* provide information at different dimensions, which have no linear relationship with the above features. These two features fluctuate violently at low-capacity states and supply extra knowledge for SOH estimation at low-energy charging modes. The above six features ensure sufficient information for SOH estimation at flexible charging curves and fixed input numbers for estimators.
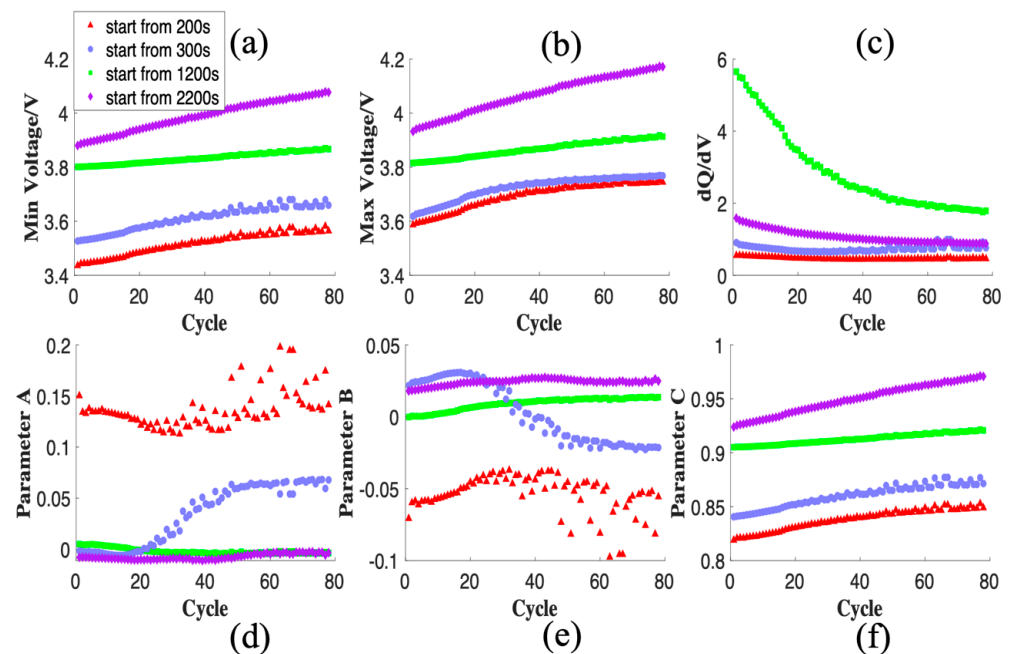
Figure 2. Feature trend of partial charging curves with different start charging moments affected by battery degradation. (**a**) Minimum voltage of the partial charging curve. (**b**) Maximum voltage of the partial charging curve. (**c**) Normalized IC curve of the partial charging curve. (**d**–**f**) Parameters of curve fitting for the partial charging curve.

2.3. Muti-Layer Perceptron

Deep learning methods have great generalization and performance in SOH estimation [35]. Since the partial curves are wrapped into six features, deep learning methods such as complex-designed convolutional neural networks (CNN) or long short-term memory networks (LSTM) are unnecessary in this study. The proposed estimator in this study is a multi-layer perceptron network (MLP). A MLP [36] is a modern feedforward artificial network consisting of fully connected neurons with a nonlinear kind of activation function, as shown in Figure 3. MLPs form the foundation for all neural networks and still greatly improve the power of deep learning when applied to classification and regression problems. The mathematical equation for one layer of MLP is:

$$h_j = f(\sum_{i=1}^n w_{ij} x_i + b_j) \tag{4}$$

where h_j is the output of the jth hidden neuron, x_i is the ith input feature, w_{ij} represents the weight of the connection between the ith input and jth hidden neuron, and $f(*)$ is the activation function expressing nonlinearity, allowing the layer to capture complex patterns in the data. By computing the weighted sum of the hidden layers and suitable activation function, the difference between predicted SOH and actual SOH is minimized, along with techniques such as gradient descent and backpropagation. Other potential methods such as CNN or RNN may require raw data sequences as inputs to make estimation, which means the length of sequences is important and must be carefully selected. In this paper, we intend to introduce a method that does not consider the length of the partial charging curves and propose a novel feature extraction method for partial charging curves. After feature extraction, the number of inputs is fixed, and MLP is totally enough to capture the inner relationship between the inputs and SOH.
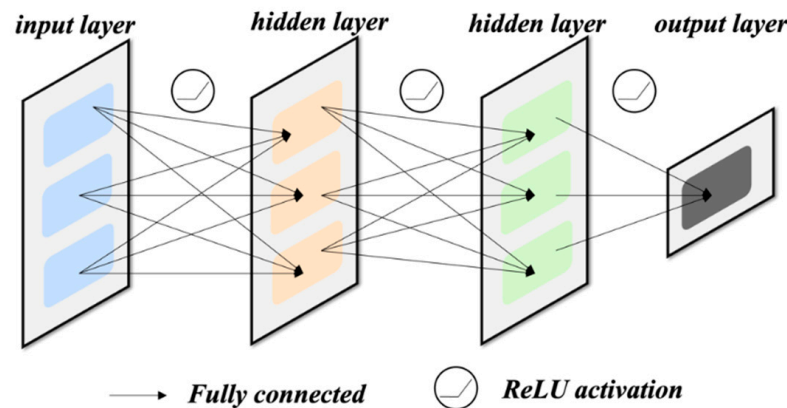

Figure 3. Multi-layer perceptron network.

The activation function adopted in this study is rectified linear unit (ReLU) [37], which can alleviate the problem of gradient vanishing through the derivative of either 1 or 0 when in the backpropagation stage. The mathematical expression is:

$$y_{ReLU} = \max(0, \ x) \tag{5}$$

A commonly used batch normalization layer which could avoid interval covariate shift is not used here, since the input number is small and the distribution of the input should remain varying to ensure the adaption of different voltage ranges.

2.4. Training Details

The proposed model aims to estimate SOH based on extracting features from the flexible partial curves with fixed sequences. To achieve this goal, the training details are concluded as follows.

The input partial charging data in the training dataset are represented by six statistical features with the above description. The first point of each partial charging dataset is treated as the voltage at zero seconds and the number of final layer neurons is one, as estimated by SOH. It is unnecessary to normalize the features between [0, 1] since the features' range is under 5, except feature dQ/dV, which is normalized by dividing the nominal capacity. The loss function utilized by backpropagation in this model is a mean square error (MSE), which could be formulated as:

$$\text{MSE} = \frac{1}{n}\sum(y_{label} - y_{estimate})^2 \tag{6}$$

where y_{labe} and $y_{estimate}$ are the ground truth SOH and estimated SOH, respectively. The process of backpropagation is carried out by the Adam optimizer [38]. The maximum number of training epochs is set to 1000 and the batch size is set to 400. In the Oxford dataset, there are 8 cells, each comprising around 80 cycles using a 1-C charging rate (approximately 3600 s' charging time). Cells 1 to 4 are utilized as the training dataset, while Cells 5 to 6 and Cells 7 to 8 are allocated for the validation and test datasets, respectively. The size of the training dataset depends on the length of the input series. The step size of partial charging curves is akin to the data sampling rate, which is 1 s in the Oxford dataset. With these configurations, millions of partial charging curves can be extracted. The trained MLP is saved and evaluated after one epoch's training, and the minimum error results on the validation dataset are selected into the eventual model. The proposed model is developed and evaluated with the Pytorch package and Python 3.9. A desktop equipped with an Ubuntu 18 system and NVIDIA GeForce 1080 graphical process unit (GPU) is used for the whole working hardware.

3. Results and Discussion

3.1. Ablation Study

The battery aging dataset to validate the proposed method is the Oxford dataset [27], which has been widely adopted for the evaluation of battery degradation diagnosis [39]. In this dataset, eight commercial batteries with 0.74-Ah nominal capacity are repetitively charging and discharging using an electrical vehicle (EV) driving profile, simulating battery aging in EV working conditions. All battery tests are conducted at 40 °C in a thermal chamber and the current, terminal voltage, and surface temperature are recorded with a sample period of 1 s. The 1C charging data are used to validate the proposed method. Due to the CC charging and constant temperature, only the voltage charging data are used as the input of the proposed method, ignoring the current and temperature. The data of the first four batteries are used as the training dataset, while the data of Cell No. 5 and No. 6 are used for the validation dataset. The length of the partial voltage sequence is 300, which means the random partial charging curve is collected in 5 min. The remaining two batteries are used as a test dataset. To observe the performance of the proposed method, the root mean square error (RMSE) metric is utilized to evaluate the overall estimation error, which is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(y_{label} - y_{estimate})^2} \tag{7}$$

To prove the effectiveness of the proposed features and model, ablation experiments are carried out, as shown in Table 1. At first, extreme gradient boosting (XGBoost) [40] is utilized as the baseline method to evaluate the efficiency of the features. XGBoost is a powerful and popular machine learning method for both regression and classification tasks. It can provide feature importance scores and incorporate regularization techniques, and is known for its efficiency and predictive accuracy. The estimation results of the validation dataset and test dataset are 1.87% and 4.26%, respectively, when the raw partial voltage curves are treated as inputs. After mapping the raw partial charging sequence into minimum voltage, maximum voltage, and dQ/dV, in these local features, the estimation results are getting worse, suggesting that local feature extraction from the partial voltage curve causes the

information loss in SOH estimation. However, the estimation performance improves with curve fitting parameters, which have globally represented the input sequences. Global features play a vital role in partial curve SOH estimation, with nearly 1.4% enhancement in the test dataset. The estimation results are further improved by using the MLP model. An interesting observation is that while the validation result of the MLP is higher than that of XGBoost, the test result is significantly better. This could be due to MLP having better generalization ability than XGBoost while the overfitting phenomenon happened using the XGBoost method. Additionally, the gap between the validation error and the test error is smaller. This suggests that the accuracy and generalization performance of the MLP are exceptional. Table 2 displays the contribution rates of features to SOH estimation. Parameter A holds the highest contribution at 19.63%, followed by maximum voltage at 16.75%. Conversely, minimum voltage contributes the least at 8.24%. Overall, the importance distribution appears balanced, indicating that SOH estimation does not heavily rely on specific features. Instead, each feature plays a distinct role in the estimation process.

Table 1. Ablation study estimation results.

Component/RMSE (%)	Validation	Test
XGBoost (raw data)	1.87	4.26
XGBoost (local features)	3.72	5.74
XGBoost (local and global feature)	0.84	3.38
MLP (local and global feature)	1.32	1.87

Table 2. Contributed importance of features for SOH estimation with XGBoost.

Features	min vol	max vol	*dQ/dV*	A	B	C
Importance (%)	8.24	16.75	12.03	19.63	11.27	13.06

3.2. Simultaneous Estimation of SOH

SOH estimation is of paramount importance for proactive maintenance scheduling when a battery's health deteriorates, and it plays a crucial role in precise Ampere-hour counting calibration, ensuring reliable State of Charge (SOC) estimation during battery cycling [7,41]. The proposed method offers a practical advantage, as it only requires a segment of charging data collected within a small timeframe, readily attainable in real-world scenarios. In this study, the window size for partial charging data can vary, with a moving step set to one. While these settings introduce some level of noise into the estimation, they also significantly increase the size of the training dataset, enabling comprehensive training of the data-driven model.

Beyond its accurate SOH estimation capabilities, the proposed method meets the demands of real-time estimation. It takes an average of only 0.40 milliseconds to perform one estimation using partial charging data. Consequently, the proposed method can be efficiently employed within battery management systems (BMS). Table 3 and Figure 4 present the estimation results and errors for SOH using random charging voltage curves collected over 300 s. Cells 5 and 6 represent the validation dataset, while Cells 7 and 8 serve as the test dataset. In Table 3, the Root Mean Square Error (RMSE) for these four cells is below 2%, affirming that feature extraction from a 300 s segment of charging data can simultaneously yield highly accurate SOH estimates during battery degradation. Figure 4 displays the distribution of SOH estimation errors in a plot format, including the ratio of outlier estimation errors. A box plot provides a visual summary of key statistical properties, such as the median, quartiles, and potential outliers. From Figure 4a, it is evident that median estimation errors are below 2%, and the majority of discrepancies between estimated maximum capacity and actual maximum capacity are within 5%. Figure 4b shows the ratio of outliers whose RMSE is beyond 5%, and it is noticeable that outliers account for less than 3%. Validation dataset results outperform test results, as the best

models were selected based on validation performance. In most estimation scenarios, the results closely align with the ground truth, with approximately 98% of estimations falling within 5% RMSE. In practical scenarios, those outliers could be significantly reduced by considering the previous estimations. If the present estimation results deviate significantly from the previous ones, they could be treated as outliers and eliminated from the estimation process. This assumption could be validated in future studies.

Table 3. RMSE of the estimation results for the validation dataset and test dataset.

| Cell Index	Cell 5	Cell 6	Cell 7	Cell 8
RMSE (%)	0.99	1.17	1.65	1.46
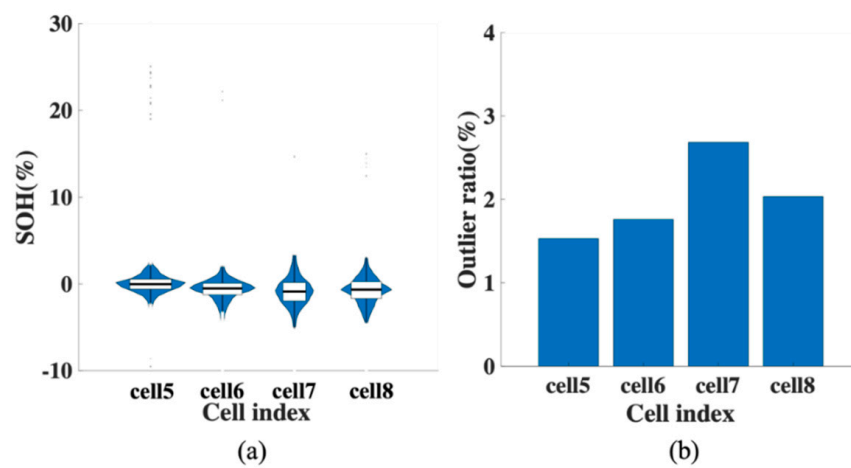
Figure 4. SOH estimation results: (**a**) SOH estimation error for validation dataset and test dataset. (**b**) Outliers ratio.

3.3. Influence of Length of Input Series

The preceding experimental results substantiate that SOH estimation can be effectively achieved using only six features, which capture both the local and global characteristics of the original partial charging curve within a specific length range. However, further investigation is needed to understand the efficacy of these features when applied to input sequences of different lengths.

To assess the impact of input sequence length, the length is systematically adjusted from 300 s to 900 s in increments of 150 s. Then, features are extracted, MLP models are trained, and evaluations are conducted accordingly. It is intuitive that as the input length increases, there may be an initial reduction in information loss, since the entire curves are condensed into six features. The comparative results of the test set are presented in Figure 5 and Table 4. Notably, the input dimension of the MLP remains constant throughout, ensuring that the estimation time remains unchanged despite varying input sequence lengths. Table 4 reveals that the average RMSE on the test set initially decreases with longer input lengths. However, when the input length surpasses 600, the reduction in average RMSE slows down, and it stabilizes when the input length reaches 750 and 900. Figure 5 illustrates the same trend, with an increasing number of samples closely aligned with the ground truth SOH values. Moreover, both the range and number of outliers diminish with longer input sequences. For instance, while the average RMSE remains the same when the input length is set at 750 and 900, the 900-length input exhibits fewer and smaller outliers. These findings indicate an overall enhancement in estimation results as the length of the input sequence is extended.
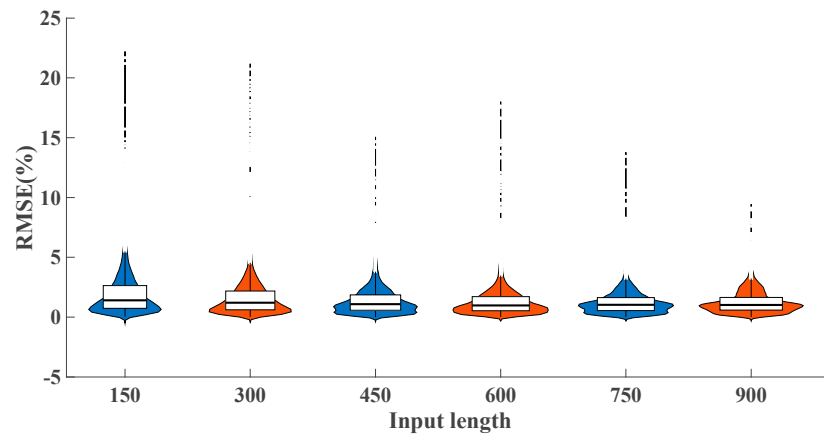
Figure 5. RMSE distribution of test dataset with different input lengths.

Table 4. Average RMSE of the test dataset.

Input Length	150	300	450	600	750	900
RMSE (%)	1.91	1.59	1.37	1.27	1.24	1.24

3.4. A Comparison with Classic Machine Learning Methods

To gauge the performance of the MLP in SOH estimation with limited inputs, three classic machine learning methods commonly used in regression tasks have been developed and compared: Random Forest (RF), Logistic Regression (LR), and XGBoost. As the feature extraction process condenses the input data into six features, these machine learning methods can be seamlessly applied without the need for additional modifications or custom designs. Furthermore, to validate that the extracted features effectively encapsulate the information contained in the original partial curves, the estimation results of LSTM with raw partial charging curves as inputs are included in this section. The detailed setups of these methods could be referred to [42]. The results are visualized in Figure 6. Comparing the performance of LR, RF, XGBoost, LSTM, and MLP, both LSTM and MLP stand out with the smallest errors, approximately 1.34% and 1.59%, respectively. While the LSTM estimator exhibits slightly better performance than MLP when utilizing the original partial charging curves, it comes at the cost of increased model complexity and computational time due to the significantly larger number of inputs. Nevertheless, the relatively small gap between the LSTM estimator and MLP estimator underscores that, even with a straightforward estimator structure, feature extraction effectively maps partial charging curves to their corresponding SOH values.
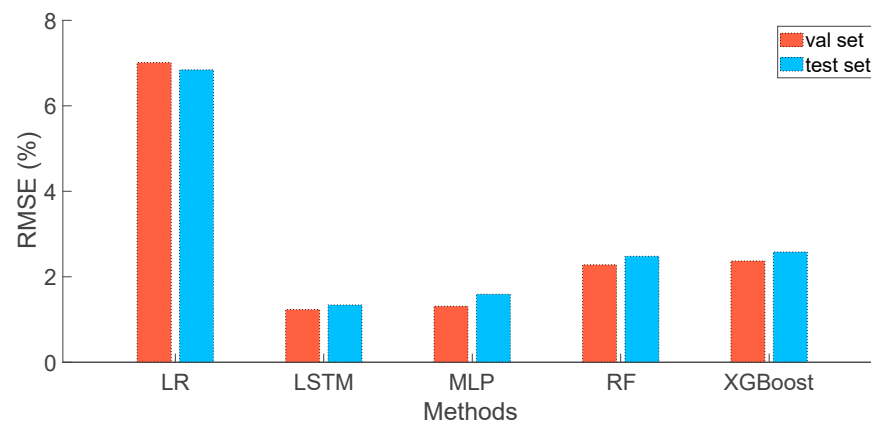

Figure 6. The average RMSE of different methods.

3.5. Validation of Different Batteries and Working Conditions

One of the significant advantages of deep learning methods is their capacity for generalization across various datasets and working conditions. To assess the method's performance on different datasets, the proposed approach is applied to estimate SOH for two additional battery types: the well-established NASA dataset [31] and the CALCE dataset [30]. Notably, these datasets feature different sampling intervals, with a 30-s interval in the CALCE dataset and a 3-s interval in the NASA dataset. This presents a considerable challenge for SOH estimation when utilizing raw partial charging curves as inputs, given the reduced information content. However, our proposed method addresses this challenge by employing local and global feature extraction, effectively capturing crucial details within the curves. This preprocessing aligns the input data before feeding it into the MLP-based estimation process. For reference, Table 5 provides a brief overview of the CALCE and NASA datasets, while the estimation results are illustrated in Figure 7. It is worth noting that these datasets differ in nominal capacity, constant current charging values, and sampling periods from the Oxford dataset. The window size for input series in the CALCE dataset is set to 20 points, while in the NASA dataset, it is 100 points. The total time window for the input series is under 10 min. The average Root Mean Square Error (RMSE) for these two datasets is less than 5%, ensuring reliable SOH estimation. This underscores the robustness of the proposed approach against sparse sampling, promising reduced computation costs and data storage requirements.

Table 5. Primary specifications of the CALCE and NASA datasets.

Dataset	Cathode Materials	Nominal Capacity (Ah)	Current (A)	Sample Period (s)
CALCE	$LiCoO_2$	1.1	0.55	30
NASA	Not reported	2.0	1.50	3
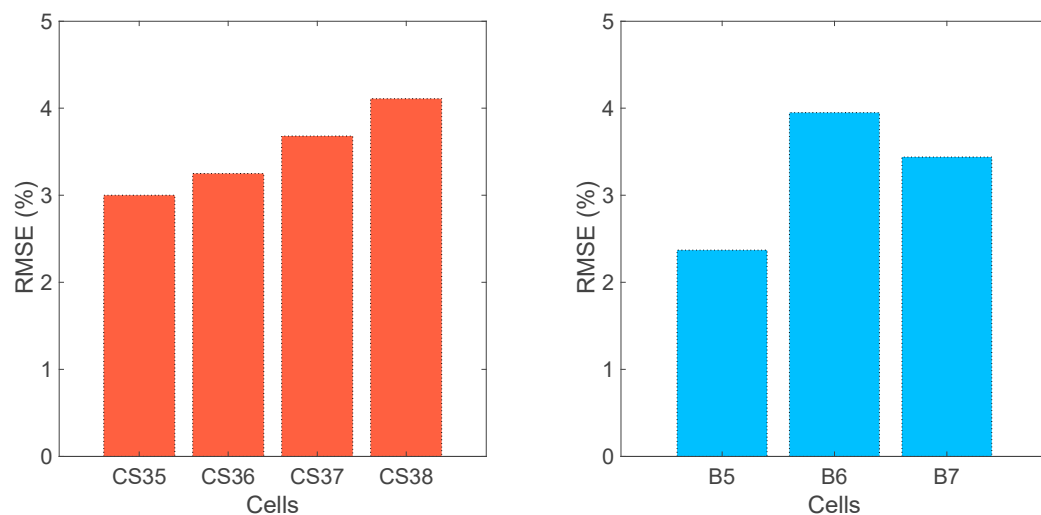
Figure 7. The average RMSE of the CALCE dataset (**left**) and NASA dataset (**right**).

4. Conclusions

Accurate SOH estimation for batteries is of paramount importance in battery management systems, particularly for enhancing the performance and longevity of traction batteries used in electric vehicles. SOH estimation based on partial charging curve data presents a formidable challenge due to varying sampling intervals. In this study, charging data within a small time window to estimate the current-cycle SOH is leveraged. To mitigate the impact of different sampling frequencies and input sequence lengths, both local and global features are extracted that describe the charging curves comprehensively, enabling precise SOH estimation. A multi-layer perceptron is employed to achieve generalized estimation by mapping these fixed features to SOH values. An ablation study

was conducted to compare model performance with and without the proposed features, demonstrating that these features encapsulate essential information required for accurate estimation. The validation results, based on three distinct datasets, affirm the effectiveness of our proposed methods. In the context of a 300 s time window, the root mean square error of SOH estimation remains below 1.5% across more than 400 thousand samples, spanning the degradation of 0.74 Ah batteries. Importantly, the estimation time for a single evaluation does not increase with a longer window size of charging data, thanks to the fixed number of features. In conclusion, the proposed methods exhibit high scalability and adaptability to varying input data lengths, battery chemistries, and operational conditions.

It is worth mentioning that partial charging curves may have the same trend by setting the start point to 0 s, even though they come from different cycles of the batteries. This causes outlier errors during the estimation. In the future, more physical parameters, such as strain, temperature, and current, will be introduced to distinguish cycles' differences and eliminate estimation errors. By extracting additional features with these physical parameters, the partial charging curves would provide more dimensional information to aid in estimation. Additionally, it is recommended to gather more experimental data consisting of different load curves at various temperatures and C-rates to train the proposed method, thereby improving its generalization ability.

Author Contributions: Conceptualization, R.L., X.L. and J.W.; Formal analysis, R.L. and J.W.; Methodology, X.L. and J.W.; Software, R.L.; Validation, R.L. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available on request from the corresponding author on reasonable request.

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

References

1. Xie, J.; Lu, Y.-C. A retrospective on lithium-ion batteries. *Nat. Commun.* **2020**, *11*, 2499. [CrossRef] [PubMed]
2. Han, X.; Lu, L.; Zheng, Y.; Feng, X.; Li, Z.; Li, J.; Ouyang, M. A review on the key issues of the lithium ion battery degradation among the whole life cycle. *eTransportation* **2019**, *1*, 100005. [CrossRef]
3. Vennam, G. A survey on lithium-ion battery internal and external degradation modeling and state of health estimation. *J. Energy Storage* **2022**, *52*, 104720. [CrossRef]
4. Zhang, L.; Hu, X.; Wang, Z.; Sun, F.; Dorrell, D.G. A review of supercapacitor modeling, estimation, and applications: A control/management perspective. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1868–1878. [CrossRef]
5. Messing, M.; Shoa, T.; Habibi, S. Estimating battery state of health using electrochemical impedance spectroscopy and the relaxation effect. *J. Energy Storage* **2021**, *43*, 103210. [CrossRef]
6. Jiang, B.; Zhu, J.; Wang, X.; Wei, X.; Shang, W.; Dai, H. A comparative study of different features extracted from electrochemical impedance spectroscopy in state of health estimation for lithium-ion batteries. *Appl. Energy* **2022**, *322*, 119502. [CrossRef]
7. Xu, L.; Lin, X.; Xie, Y.; Hu, X. Enabling high-fidelity electrochemical P2D modeling of lithium-ion batteries via fast and non-destructive parameter identification. *Energy Storage Mater.* **2022**, *45*, 952–968. [CrossRef]
8. Chen, C.; Xiong, R.; Shen, W. A lithium-ion battery-in-the-loop approach to test and validate multiscale dual h infinity filters for state-of-charge and capacity estimation. *IEEE Trans. Power Electron.* **2018**, *33*, 332–342. [CrossRef]
9. Zheng, L.; Zhang, L.; Zhu, J.; Wang, G.; Jiang, J. Co-estimation of state-of-charge, capacity and resistance for lithium-ion batteries based on a high-fidelity electrochemical model. *Appl. Energy* **2016**, *180*, 424–434. [CrossRef]
10. Topan, P.A.; Ramadan, M.N.; Fathoni, G.; Cahyadi, A.I.; Wahyunggoro, O. State of Charge (SOC) and State of Health (SOH) estimation on lithium polymer battery via Kalman filter. In Proceedings of the 2016 2nd International Conference on Science and Technology-Computer (ICST), Yogyakarta, Indonesia, 27–28 October 2016; pp. 93–96. [CrossRef]
11. Cheng, Z.; Yang, L.; Sun, X. Estimation of SOC and SOH of lithium-ion batteries based on adaptive square-root traceless kalman filtering algorithm. *Chin. J. Electr. Eng.* **2008**, *38*, 2384–2393.
12. Liu, F.; Shao, C.; Su, W.; Liu, Y. Online joint estimator of key states for battery based on a new equivalent circuit model. *J. Energy Storage* **2022**, *52*, 104780. [CrossRef]
13. Tran, M.-K.; Mathew, M.; Janhunen, S.; Panchal, S.; Raahemifar, K.; Fraser, R.; Fowler, M. A comprehensive equivalent circuit model for lithium-ion batteries, incorporating the effects of state of health, state of charge, and temperature on model parameters. *J. Energy Storage* **2021**, *43*, 103252. [CrossRef]

14. Li, Y.; Vilathgamuwa, M.; Farrell, T.; Choi, S.; Tran, N.; Teague, J. A Physics-Based Distributed-Parameter Equivalent Circuit Model for Lithium-Ion Batteries. *Electrochim. Acta* **2019**, *299*, 451–469. [CrossRef]

15. Li, X.; Yuan, C.; Li, X.; Wang, Z. State of health estimation for Li-Ion battery using incremental capacity analysis and Gaussian process regression. *Energy* **2020**, *190*, 116467. [CrossRef]

16. Wang, L.; Zhao, X.; Liu, L.; Pan, C. State of health estimation of battery modules via differential voltage analysis with local data symmetry method. *Electrochim. Acta* **2017**, *256*, 81–89. [CrossRef]

17. Wang, L.; Pan, C.; Liu, L.; Cheng, Y.; Zhao, X. On-board state of health estimation of LiFePO4 battery pack through differential voltage analysis. *Appl. Energy* **2016**, *168*, 465–472. [CrossRef]

18. Weng, C.; Cui, Y.; Sun, J.; Peng, H. On-board state of health monitoring of lithiumion batteries using incremental capacity analysis with support vector regression. *J. Power Sources* **2013**, *235*, 36–44. [CrossRef]

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

20. Tang, X.; Zou, C.; Yao, K.; Chen, G.; Liu, B.; He, Z.; Gao, F. A fast estimation algorithm for lithium-ion battery state of health. *J. Power Sources* **2018**, *396*, 453–458. [CrossRef]

21. Zhu, J.; Wang, Y.; Huang, Y.; Bhushan Gopaluni, R.; Cao, Y.; Heere, M.; Mühlbauer, M.J.; Mereacre, L.; Dai, H.; Liu, X.; et al. Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation. *Nat. Commun.* **2022**, *13*, 2261. [CrossRef]

22. Xiong, R.; Li, L.; Tian, J. Towards a smarter battery management system: A critical review on battery state of health monitoring methods. *J. Power Sources* **2018**, *405*, 18–29. [CrossRef]

23. 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]

24. Tian, J.; Xiong, R.; Shen, W.; Lu, J.; Yang, X.-G. Deep neural network battery charging curve prediction using 30 points collected in 10 min. *Joule* **2021**, *5*, 1521–1534. [CrossRef]

25. Tian, J.; Xiong, R.; Shen, W.; Lu, J.; Sun, F. Flexible battery state of health and state of charge estimation using partial charging data and deep learning. *Energy Storage Mater.* **2022**, *51*, 372–381. [CrossRef]

26. Zhao, Y.; Wang, Z.; Shen, Z.; Sun, F. Assessment of battery utilization and energy consumption in the large-scale development of urban electric vehicles. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2017318118. [CrossRef]

27. Birkl, C. Oxford Battery Degradation Dataset 1. 2017. Available online: https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac (accessed on 10 May 2023).

28. 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]

29. Xing, Y.; Ma, E.W.M.; Tsui, K.L.; Pecht, M. An ensemble model for predicting the remaining useful performance of lithium-ion batteries. *Microelectron. Reliab.* **2013**, *53*, 811–820. [CrossRef]

30. He, W.; Williard, N.; Osterman, M.; Pecht, M. Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method. *J. Power Sources* **2011**, *196*, 10314–10321. [CrossRef]

31. Bole, B.; Kulkarni, C.; Daigle, M. Randomized Battery Usage Data Set. NASA AMES Progn. Data Repos, 70. 2014. Available online: https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognosticdata-repository/#batteryrnddischarge (accessed on 10 May 2023).

32. Roman, D.; Saxena, S.; Robu, V.; Pecht, M.; Flynn, D. Machine learning pipeline for battery state-of-health estimation. *Nat. Mach. Intell.* **2021**, *3*, 447–456. [CrossRef]

33. 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]

34. Chang, C.; Wang, Q.; Jiang, J.; Wu, T. Lithium-ion battery state of health estimation using the incremental capacity and wavelet neural networks with genetic algorithm. *J. Energy Storage* **2021**, *38*, 102570. [CrossRef]

35. Hossain Lipu, M.S.; Ansari, S.; Miah, M.d.S.; Meraj, S.T.; Hasan, K.; Shihavuddin, A.S.M.; Hannan, M.A.; Muttaqi, K.M.; Hussain, A. Deep learning enabled state of charge, state of health and remaining useful life estimation for smart battery management system: Methods, implementations, issues and prospects. *J. Energy Storage* **2022**, *55*, 105752. [CrossRef]

36. Tolstikhin, I.O.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J.; et al. MLP-Mixer: An all-MLP Architecture for Vision. In Proceedings of the Advances in Neural Information Processing Systems 34 (NeurIPS 2021), Online, 6–14 December 2021; pp. 24261–24272.

37. Agarap, A.F. Deep Learning Using Rectified Linear Units (ReLU). *arXiv* **2018**, arXiv:1803.08375.

38. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.

39. 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] [PubMed]

40. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Zhou, T.; Xgboost: Extreme Gradient Boosting. R Package Version 0.6-4. 2017. Available online: https://CRAN.R-project.org/package=xgboost (accessed on 10 May 2023).

41. Zheng, Y.; Ouyang, M.; Han, X.; Lu, L.; Li, J. Investigating the error sources of the online state of charge estimation methods for lithium-ion batteries in electric vehicles. *J. Power Sources* **2018**, *377*, 161–188. [CrossRef]

42. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.