

Research papers

## A feature extraction approach for state-of-health estimation of lithium-ion battery



Changhao Piao <sup>a,b,c</sup>, Rongli Sun <sup>a,c</sup>, Junsheng Chen <sup>b,c,\*</sup>, Mingjie Liu <sup>b,c</sup>, Zhen Wang <sup>d</sup>

<sup>a</sup> School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, 400065, China

<sup>b</sup> School of Automation, Chongqing University of Posts and Telecommunications, 400065, China

<sup>c</sup> Intelligent Energy Technology Research Center, Chongqing University of Posts and Telecommunications, 400065, China

<sup>d</sup> Continental Software Systems Development Center (Chongqing) Co., Ltd., 401120, China

ARTICLE INFO

**Keywords:**

Lithium-ion battery  
SOH estimation  
Health feature extraction  
Incremental capacity analysis

ABSTRACT

Accurate state-of-health (SOH) estimation is essential to ensure the reliable and safe usage of lithium-ion batteries (LIBs). A novel health feature extraction approach is proposed in this manuscript for battery SOH estimation. Firstly, the degradation data are collected from LIBs with two different life stages, and then the discrete incremental capacity (IC) curve is obtained under different constant voltage intervals  $\Delta V$ . The corresponding charging voltage range with obvious variation trend of IC is selected and divided into several subintervals with  $\Delta V$ . The average IC of each subinterval is obtained. Furthermore, the consistency between the average IC of each voltage subinterval and battery capacity is analyzed and evaluated based on raw discrete IC curve. The average IC with the most consistent in relation to battery capacity degeneration is selected as the health feature. The impact of varying  $\Delta V$  and  $\Delta t$  values on the feature is conducted based on Spearman correlation analysis, and the health feature with maximum Spearman correlation coefficient is used to build battery SOH estimation model. Finally, two SOH estimation models and comparative analysis of the performance between proposed health feature and other accepted features are utilized to verify the proposed health feature extraction approach. The results demonstrate that our extracted health feature effectively reveals the battery performance degeneration.

### 1. Introduction

LIBs have emerged as the predominant energy storage solution for portable electronic devices, electric vehicles (EVs) and energy storage systems. However, due to irreversible electrochemical changes and the impacts of operating conditions, the performance of LIBs deteriorates during repeated charging and discharging processes [1]. The decline of battery performance is a gradual and long-term process. The aging of LIBs is sensitive to many stress factors such as temperature, current-rate, and depth of discharging [2,3]. Severe performance degradation of battery can result in failure, and even fire or explosion. SOH is a key indicator utilized for quantitative characterization of battery performance, which is typically represented by the ratio between the current capacity and nominal capacity. Accurate battery SOH estimation reveals the extent of battery fading, and also provides users with valuable insights into battery maintenance, safety forecast and ladder utilization [4–6]. Additionally, the accuracy of the SOH estimation is crucial for intelligent battery management. Therefore, accuracy SOH estimation of

battery is essential to ensure the reliable and safe operation of LIBs.

Nevertheless, accurate estimation of battery SOH remains challenging because the actual available capacity of battery cannot be obtained directly. To address this issue, extensive studies have been conducted over the last decades and are continuing to develop different methodologies for SOH estimation. Generally, these approaches can be organized into three categories: 1) equivalent model-based methods, 2) data-driven methods, 3) feature analysis methods.

Equivalent model-based methods primarily employ state or parameter solvers to estimate the battery capacity through the analysis of the internal microscopic physical and electrochemical processes, especially the decay process of batteries. Ruan et al. [7] proposed the CV phase reconstruction method combining Q-V modeling and open-circuit voltage (OCV) estimation iteratively to estimate the battery SOH precisely. Sadabadi et al. [8] developed an equivalent circuit model (ECM) of a 12 V PbA battery to estimate the battery SOH, and they studied the evolution of ECM parameters. Liu et al. [9] extracted a joint estimation method for SOH, state of charge (SOC) and state of power (SOP) of

\* Corresponding author at: School of Automation, Chongqing University of Posts and Telecommunications, 400065, China.

E-mail address: [chenjunsheng@cqupt.edu.cn](mailto:chenjunsheng@cqupt.edu.cn) (J. Chen).

batteries based on the autoregressive equivalent circuit model (AR-ECM), and the square root unscented Kalman filter (SR-UKF) was used to realize the online estimation of SOH and SOC. Xu et al. [10] applied a minimalist electrochemical model to describe the distribution of the lithium content inside the battery relating the SOH to the capacity fading due to irreversible loss of Li. They identified the parameters by least square method and estimate battery SOH by Kalman filter. Mamun A. A. et al. [11] developed the polynomial and exponential function for the fusion model by using discharge voltage and internal resistance as aging characteristics, and then they utilized the particle swarm optimization (PSO) method to obtain the optimal coefficients for the regression model. Singh et al. [12] developed A model-based condition monitoring strategy in this paper for LIBs on the basis of an electrical circuit model incorporating hysteresis effect. However, the majority of these methods involve a significant number of nonlinear equations, resulting in increase for both solving complexity and time. Moreover, the internal parameters of the battery are difficult to accurately measure.

Data-driven methods utilize machine learning techniques to capture nonlinear characteristics for the purpose of establishing mapping relationships between health features and battery capacity with a large amount of historical data such as current, voltage and temperature. Gong et al. [13] proposed a data-driven model framework based on deep learning for estimating SOC and SOH, which mainly consisted of long short-term memory (LSTM) neural network and back propagation neural network. Lin et al. [14] developed a data-driven approach to estimate the SOH of LIBs with consideration of the battery's internal resistance, which integrated effectively the equivalent circuit model (ECM) and the data-driven method. Deng ZH. et al. [15] proposed a data-driven method based on the random partial charging process and sparsed Gaussian process regression (GPR) to ensure the healthy and sustainable development of EVs. Zhang et al. [16] presented a data-driven method for estimating the capacity of Li-ion battery based on the charge voltage and current curves based on the k-nearest neighbor (kNN) regression. But this method requires a lot of time to train and adjust the parameters. However, the main limitation of machine learning methods lies in their heavy reliance on high-quality datasets of battery usage behavior. Additionally, due to the long training time, these developed universal estimation model is mostly trained offline.

The feature analysis approaches focus on extracting the variables that are highly related to the battery capacity degradation from the voltage and capacity data in charging or discharging process to reflect the battery performance decrease. Jiang et al. [17] adopted IC analysis and IC peak area analysis, aging mechanisms in the batteries were studied. Chang CH. et al. [18] proposed an on-line method based on the fusion of IC and wavelet neural networks with genetic algorithm (GA-WNN) to estimate SOH under current discharge. They adopted IC analysis and peak position and peak height analysis. Zhang L. et al. [19] proposed the SOC based IC analysis methods, three feature points (FPs) were extracted from the SOC based IC curves, and then the relations between FPs and cell SOCs/capacities were quantified. The major limitation of the proposed method was that the SOC and capacity values could not be corrected until a BMS effectively captures FPs. Lewerenz et al. [20] used the differential voltage (DV) analysis, a trend of increasing homogeneity of lithium distribution can be measured with characteristic points. Zhang et al. [21] presented a model-free SOH calculation method by fusion of coulomb counting method and DV analysis, and it realized rapid online SOH calculation under constant current discharging stage. Yang et al. [22] reported a voltage reconstruction model, which not only accurately estimated the SOH but also quantitatively identified the aging modes. Merla et al. [23] present differential thermal voltammetry (DTV) as an in-situ SOH estimator for LIBs. The DTV technique was able to diagnose the battery fading without relying on supporting results from other methods nor previous cycling data. Feng et al. [24] proposed probability density function (PDF) for evaluating the battery SOH of electric storage batteries by analyzing the

charge/discharge (C/D) data. The PDF method extended the application of the IC/DV analysis method and the PDF algorithm was promising to be used in the online SOH evaluation of LIBs. Goh et al. [25] proposed a novel approach of health indicator (HI) extraction based on the U-chord curvature model, they split the discharge process into various phases based on the curvature of the discharge curve and extracted many HIs with a high correlation to battery SOH in the discharge platform stage of the discharge curve. Tian et al. [26] proposed a SOH estimation method based on differential temperature-incremental capacity-voltage (DT-IC-V) health features (HFs), and a set of DT-IC-V HFs were designed in a relatively small charging segment to reduce the difficulty of obtaining data in practice. Xiong et al. [27] proposed a novel method combining four algorithms, i.e. the correlation coefficient, least absolute shrinkage and selection operator regression, neighborhood component analysis, and ReliefF algorithm to select the most important features, which were derived from the measured and calculated parameters to estimate the battery SOH. However, the health feature extraction presented above methods is complex and time-consuming relatively. Therefore, to develop a straightforward and effective approach for extracting health feature is imperative.

To overcome the limitations, this manuscript presents a novel approach for extracting the health feature to accurately estimate battery SOH. The contributions of this manuscript are elaborated as follows:

- We extract a health feature that enables to reveal battery performance degeneration from original discrete IC curve without smoothing by analyzing the shift trend of IC curve under different cycle. Thus, it not only inherits the advantages of IC analysis method but also avoids the laborious data preprocessing procedure.
- We present a method for evaluating the consistency between the average IC of each voltage interval and battery capacity, thereby the average IC of corresponding interval with the most consistent in relation to battery capacity degeneration is determined and utilized to estimate battery SOH.
- We validate the effectiveness of extracted health feature in battery SOH estimation on two batteries with different life stages using SVR-based and BPNN-based SOH estimation models.

The remainder of this paper is organized as follows: Section 2 proposes the health feature extraction approach based on IC analysis. Section 3 utilizes SVR-based and BPNN-based SOH estimation model to validate the proposed health feature extraction approach. Section 4 discusses and evaluates the estimation results of the above models with extracted feature. Section 5 states the conclusions together with future works.

## 2. Health features extraction

In this chapter, we describe the aging experiment process for LIBs, and then present the health feature extraction approach. The accelerated decline tests for LIBs are conducted on four batteries of the same type (Nickel-Rich NMC-based) with two different life stages to acquire the cycling dataset used in this manuscript. The SOH degradation feature is then captured based on the IC analysis with the acquired data.

### 2.1. Data acquisition

Four LIBs with the same rated capacity of 36 Ah and rated voltage of 3.6 V are utilized for accelerated degradation tests. To facilitate the following description, the four batteries are labeled as A1, A2, B1, and B2 respectively. The specifications and cycle condition of these batteries are listed in Table 1.

Battery A1 and B1 in Group 1 have the same life stage, with initial test capacities of 34.4 Ah and 34.43 Ah respectively. The initial test capacities of battery A2 and B2 with the same life stage in Group 2 are 33.29 Ah and 33.31 Ah respectively. The batteries in Group1 undergo a

**Table 1**

The four batteries' specifications and cycle condition.

| Group                         | 1              |                | 2              |                |
|-------------------------------|----------------|----------------|----------------|----------------|
| Battery label                 | A1             | B1             | A2             | B2             |
| Chemical composition          | Ni-rich<br>NMC | Ni-rich<br>NMC | Ni-rich<br>NMC | Ni-rich<br>NMC |
| Rated Capacity (Ah)           | 36.0           | 36.0           | 36.0           | 36.0           |
| Rated Voltage (V)             | 3.6            | 3.6            | 3.6            | 3.6            |
| Manufacturer                  | LG Chem        | LG Chem        | LG Chem        | LG Chem        |
| Upper cut-off voltage (V)     | 4.15           | 4.15           | 4.15           | 4.15           |
| Lower cut-off voltage (V)     | 2.5            | 2.5            | 2.5            | 2.5            |
| Charging constant current (A) | 36             | 36             | 36             | 36             |
| Discharging current (A)       | 36             | 36             | 36             | 36             |
| Initial capacity (Ah)         | 34.40          | 34.43          | 33.29          | 33.31          |
| Temperature (°C)              | 25             | 25             | 25             | 25             |
| sampling period(s)            | 1              | 1              | 1              | 1              |
| Relaxation time(h)            | 0.5            | 0.5            | 0.5            | 0.5            |
| Total cycle number            | 480            | 480            | 320            | 320            |

total of 480 cycles, while the batteries in Group2 undergo a total of 320 cycles. The battery test system specifications are presented in **Table 2**. The accelerated degradation tests lasted more than ten months.

The voltage and current of cell in a complete cycling with 1C charging rate is shown in **Fig. 1(a)**, which includes five processes, i.e.: (a) CC charging (the cell was firstly charged at CC mode with one current rate until reaching the charging cut-off voltage of 4.15 V); (b) CV charging (it was charged in CV mode until the current dropped below 0.05C); (c) Relaxation after charging with 0.5 h; (d) CC discharging (the cell was discharged with one current rate until reaching the discharging cut-off voltage of 2.5 V); (e) Relaxation after discharging with 0.5 h. The CC discharging capacity is treated as the battery residual capacity during cycling.

**Fig. 1(b)** plots the capacity decline of four batteries with increasing cycles. As shown in **Fig. 1(b)**, After undergoing 480 cycles, the capacities of battery A1 and B1 have decreased to 27.38 Ah and 27.89 Ah, respectively. The capacities of battery A2 and B2 have degraded to 31.60 Ah and 31.25 Ah after 320 cycles. The capacity of A1 and B1 battery experience a sharp decline after reaching 400 cycles, while the capacities of battery A2 and B2 gradually decrease. The capacity degradations of batteries which are difficult to quantify directly are nonlinear. The nonlinearity of batteries capacity decline varies with different life stage. In view of this, we present a convenient approach to extract an effective and robust health feature that can accurately indicate the extent of battery fading.

## 2.2. Health feature extraction and analysis

### a. IC analysis

The capacity degradation is comparable between battery A1 and B1, while the capacity decay of battery A2 and B2 is similar. Hereinafter, we analyze and elaborate on battery A1 and A2. To analyze the discharging voltage characteristics of LIBs, the charging voltages of battery A1 and A2 are depicted in **Fig. 2**. The electrolytic reaction between the electrodes and electrolyte slows down, resulting in a rapid balance of positive and negative voltage in battery. Hence, the charging voltages reach the charging cut-off voltages in a shorter period with the battery cycle increasing. Besides, compared the battery A1 and battery A2, their voltage curves have obvious differences. The voltage curves of A1 change apparently with the number of cycles increases in **Fig. 2(a)**, while the curves of battery A2 just show subtle differences in **Fig. 2(b)**. The voltage in plateau regions of voltage curves increases slowly, while the reaction inside the battery is vigorously. Therefore, capturing universal health features still remains a challenging for battery SOH estimation.

Based on the data including the current and voltage collected during the charging process, **Fig. 3** plots the curves of the charging capacity versus battery terminal voltage. It can be obviously seen that the slopes of these curves vary across different voltage ranges and exhibit disparity in different cycles. Furthermore, the overall tendency of the curves moves downward as the cycles increase, although the tendency of battery A2 is moderate in **Fig. 3(b)**. The variations in slope across different voltage ranges and cycles have garnered the attention of numerous scholars. In this case, IC analysis method was proposed to gain profound insights into the variations in charging capacity at different voltages. The advantage of the IC analysis method lies in its ability to transform voltage plateau regions on voltage curves into distinct  $dQ/dV$  values on IC curves, which are easily observable and identifiable.

IC analysis method has been treated as an effective and significant way to estimate battery health conditions recently. IC analysis method focusses on analyzing the change rate of the charge capacity with voltage. The IC curve is plotted based on the voltage-capacity change rate ( $V \cdot dQ/dV$ ) which is obtained by taking the first derivative of the battery's voltage-capacity ( $V-Q$ ) curve during a CC charging or discharging process. Additionally, IC analysis can be explained from an electrochemical perspective. The IC is related to the ion insertion process of the chemical reactions inside the battery.

The corresponding capacity and voltage are acquired beforehand, and the derivation of the IC in charging process can be expressed as follows,

$$IC = \frac{dQ}{dV} = \frac{I \cdot dt}{dV} = I \frac{dt}{dV} \quad (1)$$

where  $I$  and  $t$  are the charging constant current and time.  $Q$  represents the battery capacity as the product of time and current.  $V$  refers to battery voltage.

Generally, the IC value can be calculated with constant voltage intervals (CVIs) or constant time intervals (CTIs). The CVIs are adopted to obtain the IC curves in this manuscript. Since the battery voltage and current are sampled with a constant sampling period 1 s (1 Sample/s), the discrete form of IC can be denoted by

$$\left. \frac{dQ}{dV} \right|_{k,dv} = \frac{1}{dv} \sum_{s=k}^{k+i-1} I(t_{s+1} - t_s) \quad (2)$$

where  $k$  represents the  $k$ -th point of the IC curve,  $dv$  is the constant voltage interval,  $s$  is the sampling point,  $i$  is the total number of sampling points contained in the corresponding voltage interval.

To compare the IC curves with different voltage intervals, we applied several constant voltage intervals to obtain the IC curves of battery A1 and A2. The obtained original IC curves of the A1 battery and A2 battery

**Table 2**

Battery test system specifications.

| Manufacturer | Neware technology                      |                                     |
|--------------|--|-------------------------------------|
| Voltage      | Measuring Range (Charge/<br>Discharge) | 5 mV–10 V                           |
|              | Minimum Discharge Voltage              | 0 V                                 |
|              | Accuracy                               | ±0.1 % of FS                        |
|              | Stability                              | ±0.1 % of FS                        |
| Current      | Measuring Range (Charge/<br>Discharge) | 500 mA–100A                         |
|              | Cut-off Current with CV                | 0.05C                               |
|              | Accuracy                               | ±0.1 % of FS                        |
|              | Stability                              | ±0.1 % of FS                        |
| Charge       | Mode                                   | CC, CV, CP                          |
|              | Cut-off conditions                     | Voltage, Current, Time,<br>Capacity |
| Discharge    | Mode                                   | CC, CP                              |
|              | Cut-off conditions                     | Voltage, Current, Time,<br>Capacity |

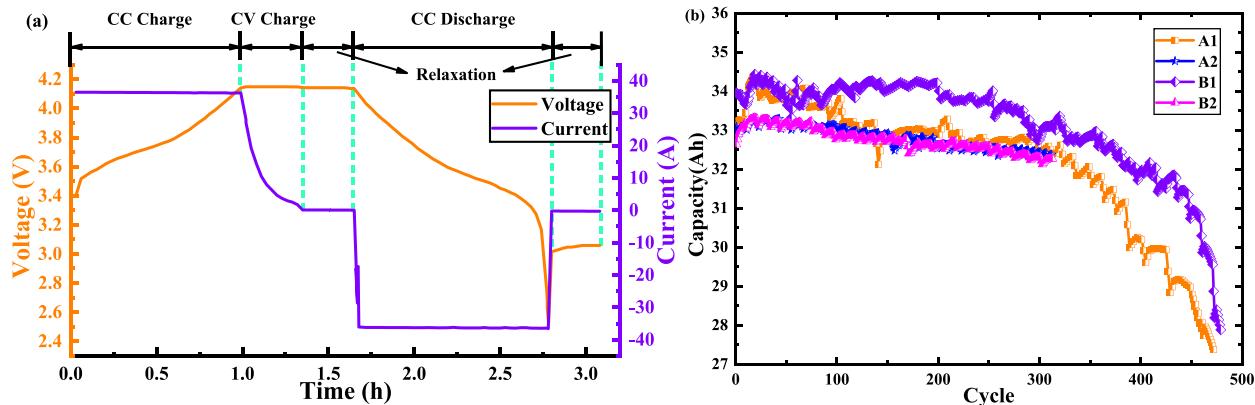


Fig. 1. The schedule of battery cycle test (a) and capacity degradations of the four batteries (b).

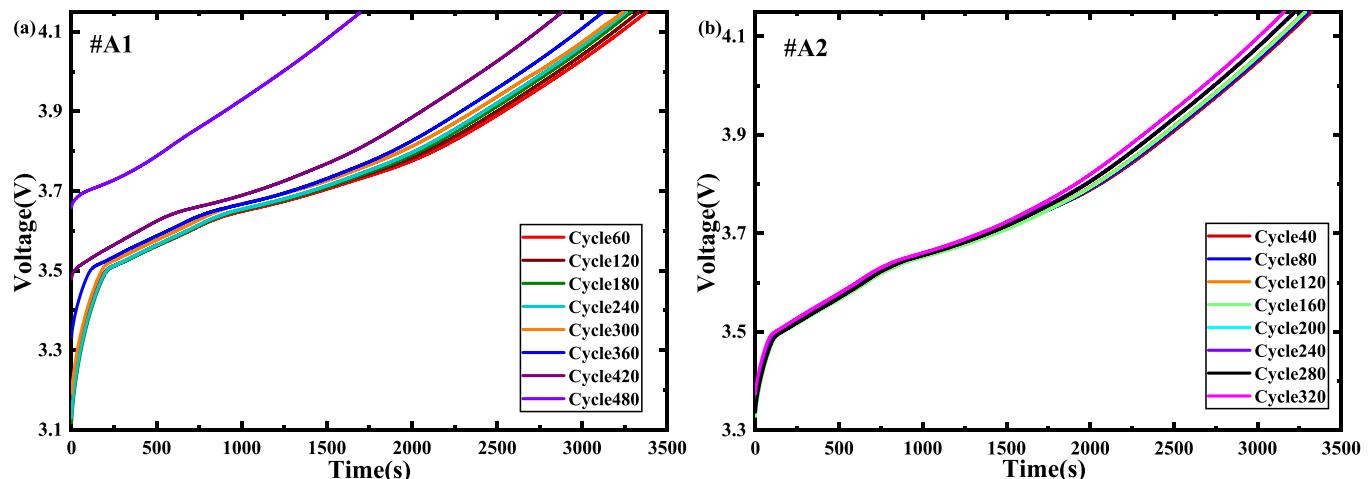


Fig. 2. The charging voltage curves of two batteries with different cycles.

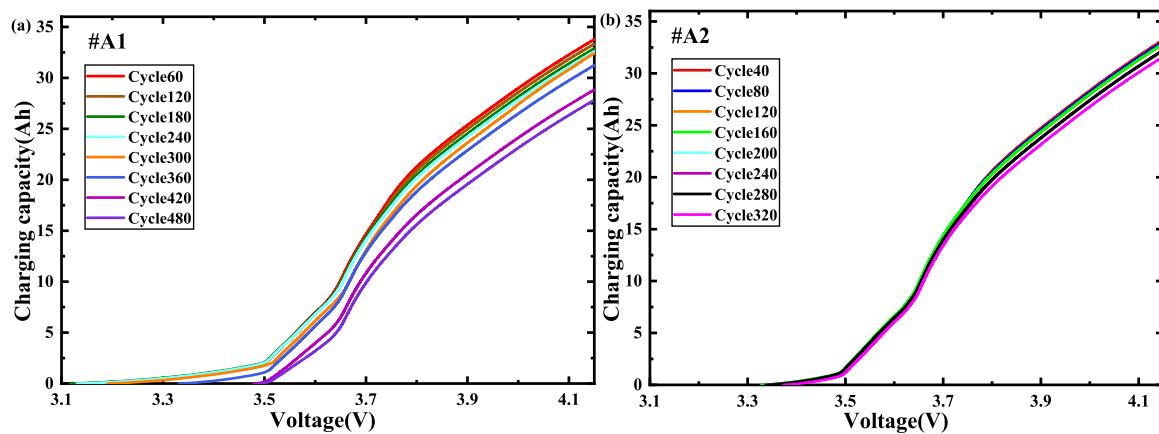
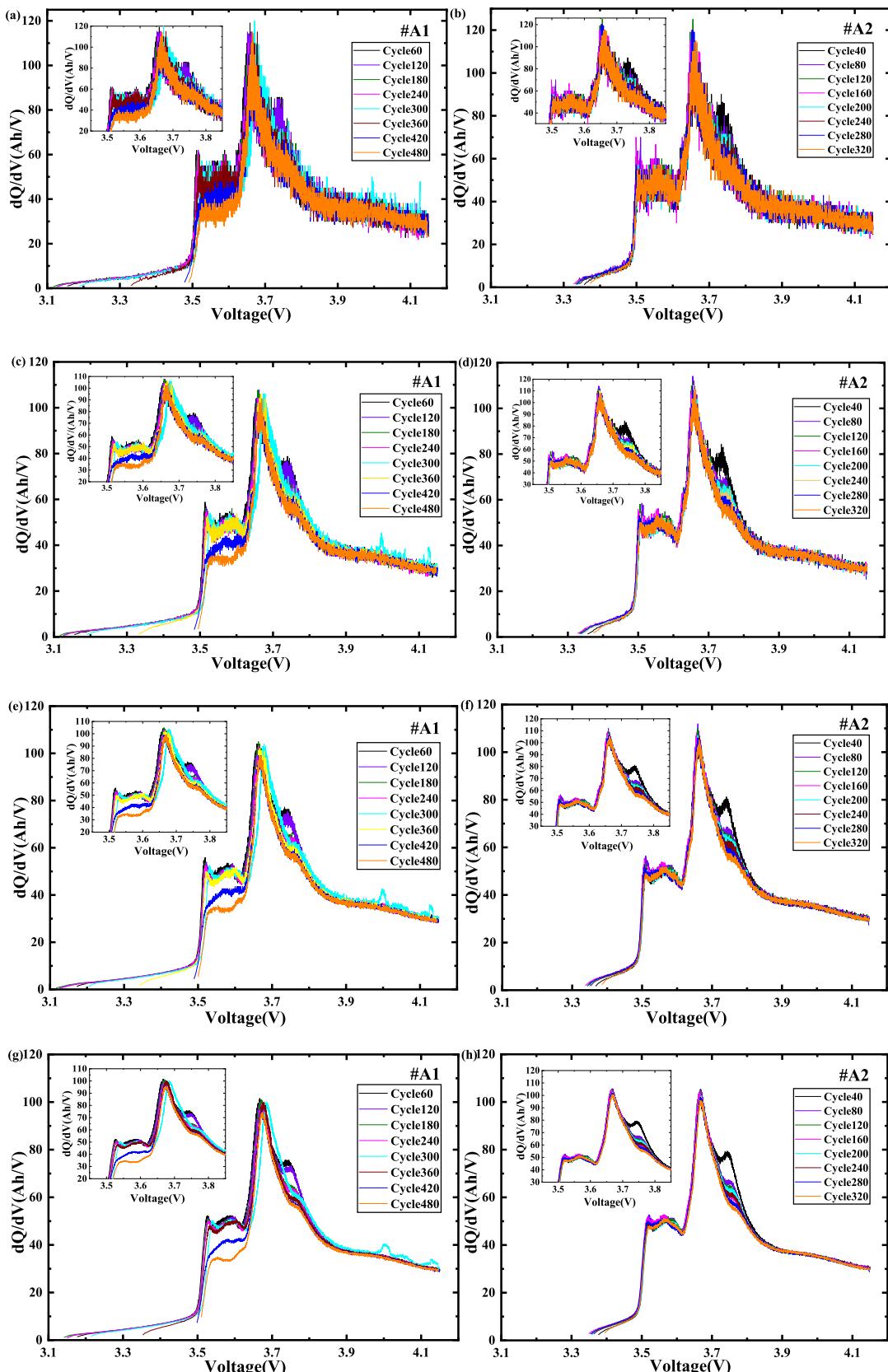


Fig. 3. Voltage-capacity curves. (a) V-Q curve of battery A1; (b) V-Q curve of battery A2.

with voltage intervals of 2 mV, 5 mV, 10 mV and 20 mV are presented in Fig. 4. As depicted in Fig. 4, it is obvious that with an increase of cycles, the IC curves demonstrate a rightward shift within certain voltage ranges or move downward within some ranges. These shift trends of IC curves with Battery A1 is particularly noticeable within the voltage interval [3.5, 3.65] and [3.7, 3.8] in Fig. 4(a), (c), (e) and (g). Battery A2 exhibits a moderate trend in its IC curves, while an obvious tendency within the voltage range of [3.7, 3.8] in Fig. 4(b), (d), (f) and (h). It also

means that the IC curve tend to move rightward or downward and the IC value corresponding to a certain voltage gradually decreases especially in part of the voltage range with the capacity degeneration. Therefore, a variable capable of reflecting the dynamic change in IC curve holds significant potential for accurately estimating battery SOH. Besides, according to the comparison of IC curves under different CVIs, it is evident that a small voltage interval may lead to some noise disturbances, while a large interval will submerge some significant features.



**Fig. 4.** Original IC curves of the battery A1 and A2 with different CVIs. (a) A1, CVI = 2 mV; (b) A2, CVI = 2 mV; (c) A1, CVI = 5 mV; (d) A2, CVI = 5 mV; (e) A1, CVI = 10 mV; (f) A2, CVI = 10 mV; (g) A1, CVI = 20 mV; (h) A2, CVI = 20 mV.

An appropriate voltage interval is essential to calculate IC value. The influence on health feature extraction of different voltage intervals will be provided in Subsection c. *Health feature analysis*.

### b. Health feature extraction approach

Some health features such as peak values, positions, and areas are extracted from smooth and denoised IC curves to establish battery capacity degradation models. However, these features are highly sensitive to filtering techniques. Even slight modifications on algorithm parameters can result in significant changes to feature value that ultimately impact the evaluation. To remedy this deficiency, a novel health feature is captured in this manuscript from original IC curves. Thus, it not only inherits the advantages of IC analysis method but also avoids the laborious data preprocessing procedure.

As mentioned before, the IC values of partial voltage ranges show a rightward or downward trend with the capacity degeneration. Therefore, we propose a feature extraction approach aiming to identify a variable that exhibits the most consistent trend with battery capacity degeneration from raw IC value according to the dynamic change in IC curve. The detailed process for health feature extraction is summarized as follows:

**Step 1:** We denote the two-dimensional data vector of voltage  $u$  and corresponding IC value  $dQ/dV$  (denoted by  $z$  for simplicity) in the  $l$ -th cycle obtained by Eq. (2) as  $\mathbf{X}_l = [u_{l,p}, z_{l,p}]^T$  ( $l = 1, 2, \dots, m; p = 1, 2, \dots, n$ ), where  $m$  is the total number of cycles for battery, and  $n$  is the number of sampling data in a certain cycle. It is noteworthy that the number of sampling data in each cycle is different due to the battery performance degradation.

**Step 2:** A voltage range  $[u_{lower}, u_{upper}]$  usually a middle segment is selected according to the trend of IC curves under different cycles. It should be noted that the selected voltage range should be covered by the charging voltage in each cycle. The selected voltage range  $[u_{lower}, u_{upper}]$  is divided into several subintervals  $[u_{lower} + (q - 1) \cdot \Delta v, u_{lower} + q \cdot \Delta v]$  ( $q = 1, 2, \dots, s$ ) with  $\Delta v$ , where  $s$  is the total number of subintervals which can be obtained by

$$s = \frac{u_{upper} - u_{lower}}{\Delta v} \quad (3)$$

It should be noted that the total number of subintervals is the same with different cycles because the selected voltage range  $[u_{lower}, u_{upper}]$  is contained in each cycle. And then the sampling voltage points within the subinterval are determined.

**Step 3:** To mitigate the impact of noise, the average IC value  $\bar{z}_{l,q}$  of the  $q$ -th subinterval in the  $l$ -th cycle is calculated using the corresponding original IC data of each sampling voltage within the subinterval.

**Step 4:** The consistency between the average IC of each voltage subinterval and battery capacity is described by

$$f(q) = \sum_{l=2}^m \text{sgn}((Q_l - Q_{l-1}) \cdot (\bar{z}_{l,q} - \bar{z}_{l-1,q})) \quad (4)$$

where,  $\text{sgn}(\cdot)$  is the signum;  $Q_l$  denotes the battery capacity of the  $l$ -th cycle. It is worth noting that when multiple maximums are detected simultaneously, we need to refer to the consistency of the adjacent subintervals and capacity to determine the appropriate subinterval.

**Step 5:** The  $q$ -th subinterval is obtained according to the principle that  $f(q)$  is maximal. Thereby, the average IC  $\bar{z}_{l,q}$  of the  $q$ -th subinterval with the most consistent in relation to battery capacity degeneration is extracted as the health feature for battery SOH estimation.

Apparently, a small  $\Delta v$  will cause the significant influence of noise interference on the average IC of the subinterval, while a large  $\Delta v$  may obscure some crucial information. An appropriate interval  $\Delta v$  is necessary to extract health feature.

For the sake of simplicity, we will present the feature extraction results for battery A1 in the following. The consistency between the average IC of each voltage subinterval and capacity under different  $\Delta v$  and  $\Delta v$  conditions for battery A1 is shown in Fig. 5. With the value of CVI  $\Delta v$  increases, the overlap of voltage ranges used by adjacent voltage points in the calculation of IC also increases. Thus, it leads to a decrease in discrimination for each subinterval and is not conducive to select an appropriate subinterval. As the value of interval  $\Delta v$  increases, the number of subintervals decreases and the impact of  $\Delta v$  becomes less.

The extracted health features under different  $\Delta v$  and  $\Delta v$  values and capacity are presented in Fig. 6. Most of the extracted health features exhibit a strong correlation with battery capacity, except for small values of  $\Delta v$  and  $\Delta v$  (such as 5 mV). The most probable reason is that the small  $\Delta v$  and  $\Delta v$  would lead to some noise disturbances. However, the health features obtained with different parameters still have differences. In the following subsection, we will quantify the correlation between health features and capacity to determine the most correlated feature for SOH estimation.

### c. Health feature analysis

As previously mentioned, an adaptive  $\Delta v$  and an appropriate  $\Delta v$  are crucial for ensuring a high-quality health feature. The correlation between features under different parameters ( $\Delta v$  and  $\Delta v$ ) and the battery capacity is discussed. Spearman correlation analysis is appropriate for monotonic variables, while Pearson correlation analysis is suitable for normally distributed variables. Considering the monotonicity of battery capacity and health feature, we employ Spearman correlation analysis to quantify the correlation between two variables. The detailed Spearman correlation coefficient is described as follows,

$$\gamma_s = 1 - \frac{6}{m(m^2 - 1)} \sum_{l=1}^m d_l^2 \quad (5)$$

where,  $m$  is the total number of battery cycles,  $d_l$  denotes the difference of ranks between  $\bar{z}_{l,q}$  and  $Q_l$ .

The detailed results of correlation analysis between the extracted health feature under different parameters and capacity for battery A1 are presented in Fig. 7. The extracted health features of battery A1 under  $\Delta v = 10$ ,  $\Delta v = 20$  and  $\Delta v = 5$ ,  $\Delta v = 20$  outperform that obtained by other parameters. Meanwhile, the health features obtained in these two cases are approximately completely correlation. Thus, this health feature is used to build battery SOH estimation model. Moreover, from Fig. 5, a larger  $\Delta v$  or a small  $\Delta v$  (such as  $\Delta v = 20$  mV or  $\Delta v = 5$  mV) will cause the extracted feature being insufficient to accurately reveal the battery capacity decay.

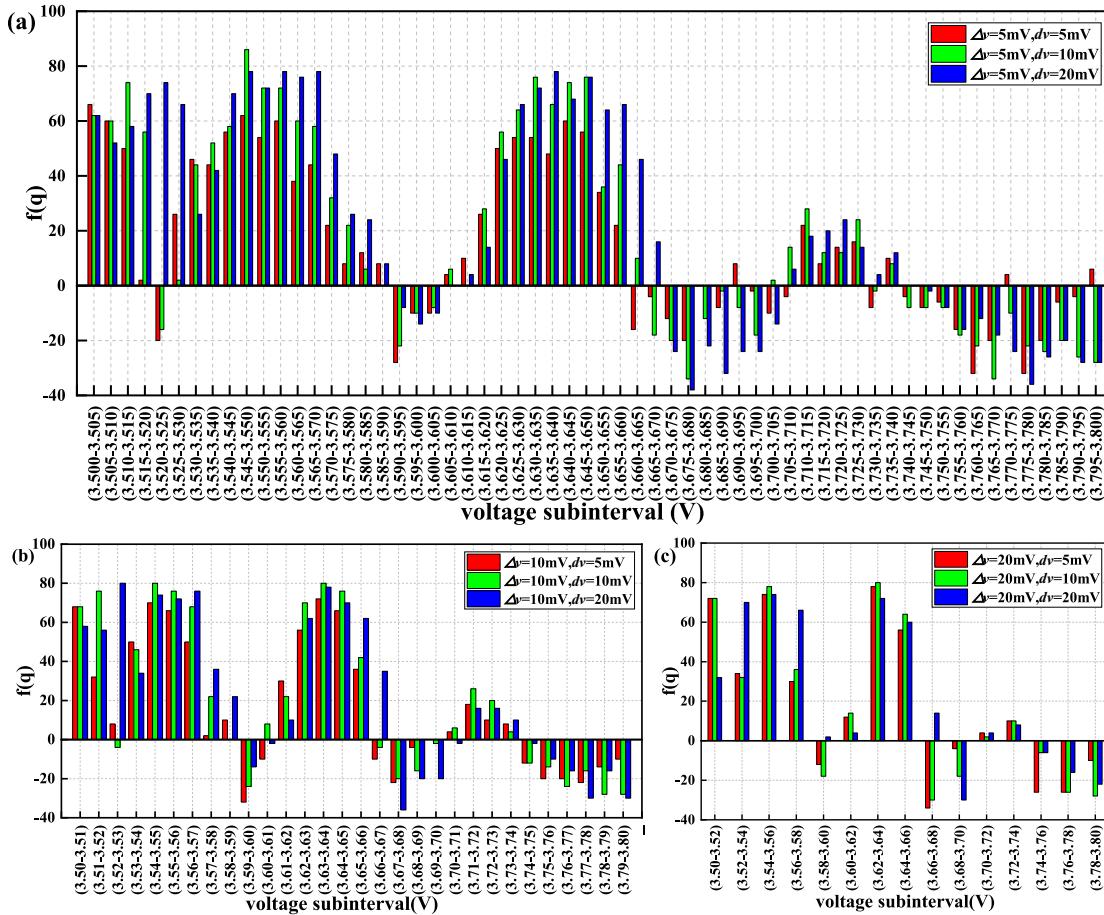
## 3. Battery SOH estimation modeling

The most relevant average IC under different  $\Delta v$  and  $\Delta v$  values is utilized as health feature to establish battery SOH estimation models. To validate the effectiveness of extracted health feature, we develop SOH estimation models based on SVR and BPNN algorithm in this part.

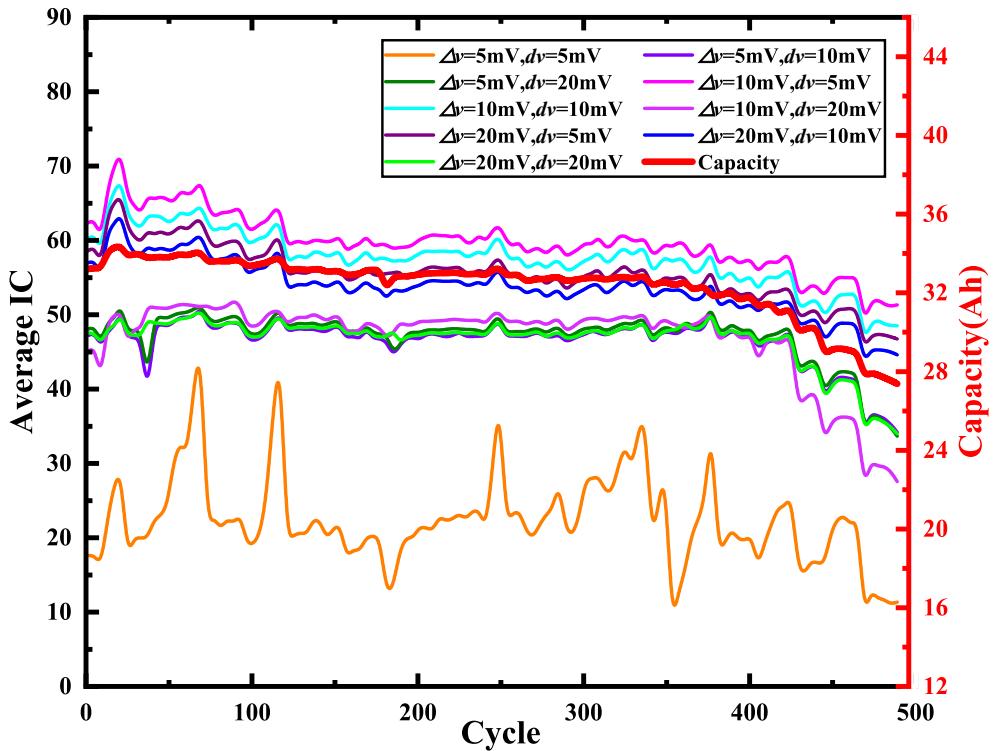
### 3.1. Principle of model algorithm

#### a. SVR algorithm

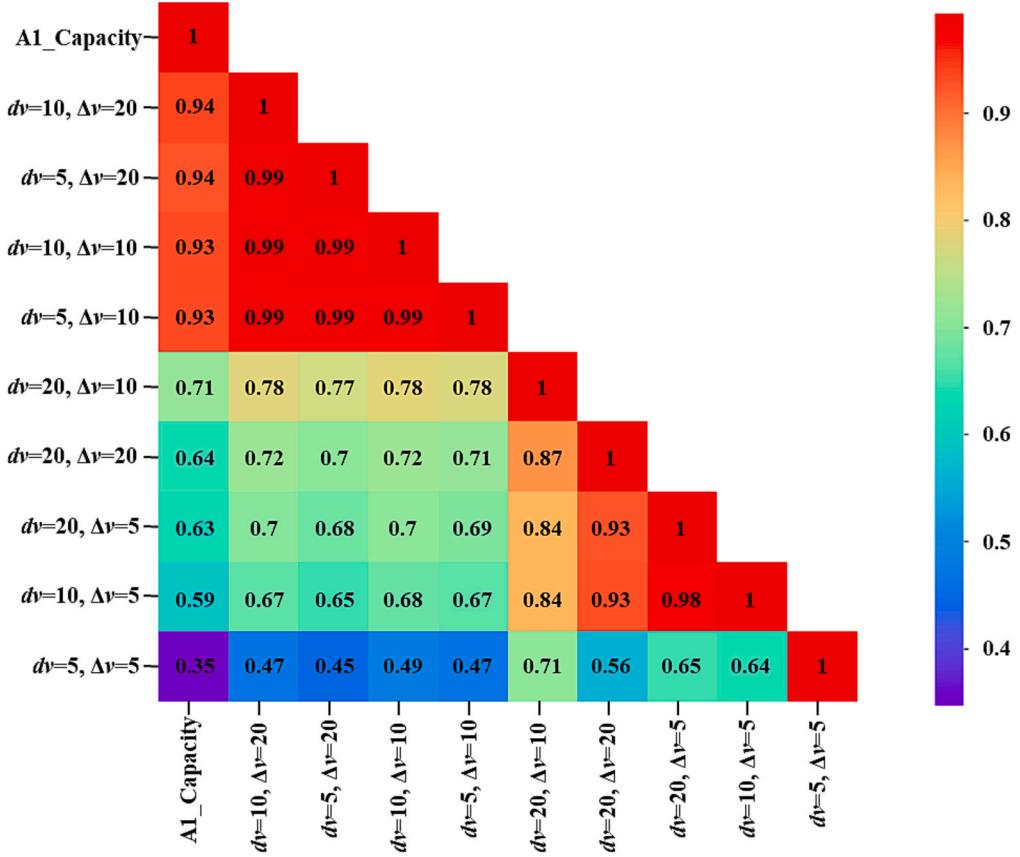
The Support Vector Machines (SVM) algorithm was originally proposed for solving binary classification problem. SVM effectively balances learning accuracy and adaptability based on VC dimension theory and structure minimization theory. SVR is derived from SVM to realize regression problems. The SVM algorithm aims to identify the hyperplane that maximizes the margin between support vectors or all data points and the classification plane, while SVR seeks a regression hyperplane that minimizes the distance between all data points and this plane. The



**Fig. 5.** The consistency between the average IC of each voltage subinterval and battery capacity with different  $dv$  and  $\Delta v$  values for battery A1. (a)  $\Delta v = 5 \text{ mV}$ ; (b)  $\Delta v = 10 \text{ mV}$ ; (c)  $\Delta v = 20 \text{ mV}$ .



**Fig. 6.** The extracted health features under different  $dv$  and  $\Delta v$  values for battery A1.



**Fig. 7.** The correlation between the extracted health features under different parameters and capacity of battery A1.

SVR algorithm distinguishes itself from traditional regression algorithms by not only considering the degree of fitting, but also taking into account its generalization ability. Therefore, it is a suitable choice for estimating battery SOH.

In SVR algorithm, the low-dimensional data is mapped into a high-dimensional space to convert nonlinear problem into linear one. Given the dataset  $D = \{(x_i, y_i)\} (i = 1, 2, \dots, n)$ , where  $x_i$  is the health feature, and  $y_i$  is the battery SOH. The primary objective of SVR is to identify a function that minimizes the maximum deviation from the target value within a tolerable error  $\epsilon$ . The function can be described as

$$f(x) = w \cdot \phi(x_i) + b \quad (6)$$

where, “.” denotes the dot product,  $w$  and  $b$  are the weight and bias parameter,  $\phi(\cdot)$  is the function that convert nonlinear problems into linear ones with high-dimensional mapping. To ensure the errors within tolerable error  $\epsilon$ , the main goal is to obtain suitable parameters  $w$ . The slack variables  $\xi_i^\vee$  and  $\xi_i^\wedge$  are introduced to transform the regression problem into a convex quadratic optimization problem as follows

$$\begin{aligned} \min & \left\{ \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n (\xi_i^\vee + \xi_i^\wedge) \right\} \\ \text{s.t.} & \begin{cases} y_i - (w \cdot \phi(x_i) + b) \leq \epsilon + \xi_i^\vee \\ w \cdot \phi(x_i) + b - y_i \leq \epsilon + \xi_i^\wedge \\ \xi_i^\vee, \xi_i^\wedge \geq 0 \end{cases} \end{aligned} \quad (7)$$

where,  $c$  is the penalty parameter, that regulates the extent of penalty imposed on observations lying outside the  $\epsilon$  boundary to prevent overfitting. The slack variables are utilized to quantify the extent to which the sample fails to satisfy the constraint.

Furthermore, the Lagrange multiplier is employed to transform the above optimization problem into its dual form,

$$\begin{aligned} \min_{\alpha^\wedge, \alpha^\vee} & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i^\wedge - \alpha_j^\vee) (K(x_i, x_j)) \\ & + \sum_{i=1}^n [(\epsilon - y_i) \alpha_i^\wedge + (\epsilon + y_i) \alpha_i^\vee] \\ \text{s.t.} & \begin{cases} \sum_{i=1}^n (\alpha_i^\wedge - \alpha_i^\vee) = 0 \\ 0 \leq \alpha_i^\wedge \leq c, i = 1, 2, \dots, n \\ 0 \leq \alpha_i^\vee \leq c, i = 1, 2, \dots, n \end{cases} \end{aligned} \quad (8)$$

where,  $\alpha_i^\wedge$  and  $\alpha_i^\vee$  are the Lagrange multiplier,  $K(\cdot)$  is the kernel function, and Gaussian kernel is used in our manuscript.

Finally, the regression function can be rewritten by

$$f(x) = \sum_{i=1}^n (\alpha_i^\wedge - \alpha_i^\vee) K(x_i, x) + b^* \quad (9)$$

where,  $\alpha^\wedge$ ,  $\alpha^\vee$  and  $b^*$  are optimal solutions of corresponding parameters.

### b. BPNN algorithm

BPNN is the fundamental and widely used neural network, which performs forward propagation, backpropagation, and weight modification between layers. It has achieved a high level of maturity in both network theory and performance. BPNN has notable advantages of prominent nonlinear mapping capability, low computational complexity, high parallelism and a flexible network structure that can be adjusted according to specific requirements. Therefore, BPNN is utilized for the battery SOH estimation in this manuscript.

The BPNN has three-layer named input layer, hidden layer and output layer. It consists of two processes of the forward computation of data stream and the back propagation of the errors. Backpropagation involves adjusting the network parameters by computing the error between the output layer and expected value, with the aim of minimizing this error. During the forward process, the input signals are processed layer by layer from the input to hidden layers until reaching the output layer. The state of neurons in each layer only affects those in the next one. If the predicted output of a BP neural network does not match expectations, backpropagation is used to dynamically adjust the weights and biases of each layer based on prediction errors. The forward propagation formula of BPNN is as follows

$$\mathbf{a}_m = \sigma(\mathbf{W}_m^T \mathbf{a}_{m-1} + \mathbf{b}_m) \quad (10)$$

where,  $\mathbf{a}_m$  is the output of the  $m$ -th layer,  $\mathbf{W}_m$  and  $\mathbf{b}_m$  represent the weight and bias matrix of the  $m$ -th layer.  $\sigma(\cdot)$  denotes the activation function.

The loss function is expressed by

$$J = \frac{1}{2} \|\mathbf{a}_o - \mathbf{y}\|_2^2 \quad (11)$$

where,  $\mathbf{a}_o$  is the output of the BPNN,  $\mathbf{y}$  is the label value, and  $\|\mathbf{P}\|_2$  is the L2 norm of vector  $\mathbf{P}$ .

### 3.2. Battery SOH estimation process

The battery SOH estimation with proposed health feature involves four steps, namely health feature extraction, offline model training, online SOH estimation, and results analysis. The specific scheme of the battery SOH estimation is shown in Fig. 8. The detailed process is summarized as follows:

**Step 1:** The discrete IC values are obtained with CVI of  $dv$  based on the collected cycling data including battery current and voltage. And then the preselected voltage range is selected and divided into several subintervals with  $\Delta v$ . Whereafter, the average IC of the corresponding voltage subinterval with the most consistent in relation to battery capacity degeneration is calculated. The average IC values for different parameters ( $dv$  and  $\Delta v$ ) are obtained. Moreover, the average IC of the corresponding voltage subinterval that exhibits the strongest correlation with capacity can be identified with Spearman correlation analysis and selected as the health feature.

**Step 2:** The extracted health feature is utilized to establish battery SOH estimation. The SVR and BPNN algorithms are employed to capture the mapping relationship between health feature and battery capacity. Specifically, the penalty parameter  $c$  and Gaussian kernel function radius  $g$  of SVR-based model are determined with the grid search method. The number of hidden nodes in the three-layer BPNN model is determined by evaluating the accuracy of estimation results using 3–10

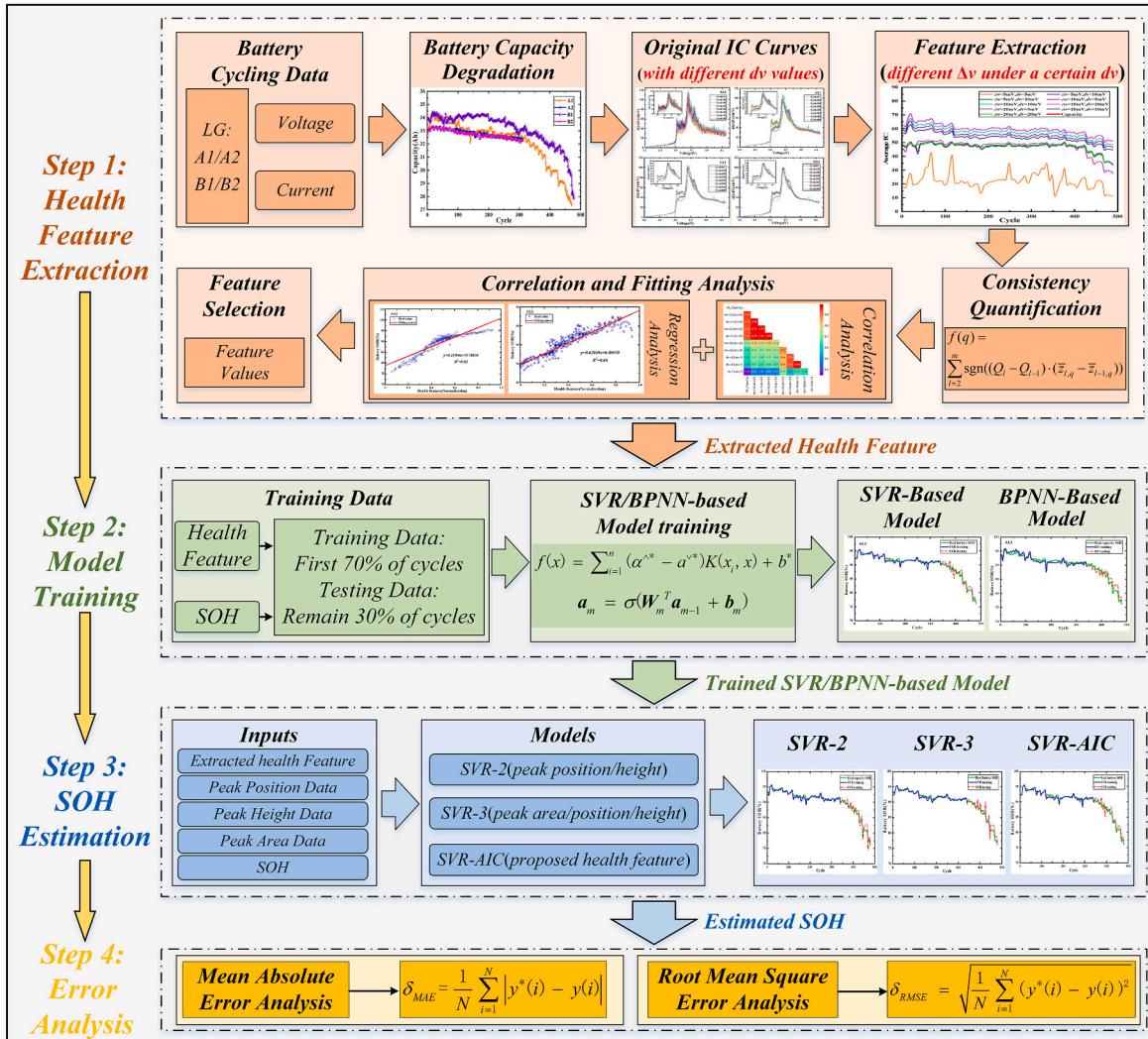


Fig. 8. The framework of the battery SOH estimation.

nodes, and the sigmoid function is selected as the activation function in this paper.

**Step 3:** The health feature of the test data is derived based on the determined voltage subinterval, and then feed into the trained SVR-based and BPNN-based models to estimate the battery SOH.

**Step 4:** Three error indicators named the absolute error ( $\Delta SOH$ ), the mean absolute error ( $MAE$ ) and root mean squared error ( $RMSE$ ) are employed to evaluate the accuracy of battery health estimation. The three error indicators are described by

$$\Delta SOH = y_i^* - y_i \quad (12)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i^* - y_i| \quad (13)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i^* - y_i)^2} \quad (14)$$

where  $N$  is the total number,  $y_i^*$  and  $y_i$  represent the estimated value and real value respectively.

#### 4. Results and discussion

Based on the features extracted and the established battery models, we utilize cells A1 and A2 to illustrate the effectiveness of the presented method.

##### 4.1. SOH estimation results and analysis

###### a. Regression analysis

As previously mentioned, the average IC of corresponding voltage subinterval presents a decreasing trend with the battery capacity degradation. Therefore, we conduct a regression analysis on the two variables of average IC and battery SOH in order to quantify their dependence.

Fig. 9 presents the results of linear regression analysis results for batteries A1 and A2. The results show that the goodness of fittings for two variables of both battery A1 and A2 exceeds 0.8, which demonstrates the feasibility of using the proposed health feature to reveal the battery SOH. The superior performance of Battery A2 compared to that of Battery A1 may be attributed to its lower capacity decline.

###### b. SVR-based and BPNN-based estimation results and analysis

Based on the health feature extracted approach proposed in Sub-section 2.2, the health feature of battery A1 is derived from the voltage subinterval of [3.62, 3.64], while health feature of battery A2 is

obtained from the subinterval of [3.76, 3.77]. During the model training process, we specify the different sizes of training data (first 60 % cycles data, first 70 % cycles data, and first 80 % cycles data) to train SVR-based and BPNN-based SOH estimation models. The models trained using the first 70 % cycles data outperform in both two batteries. Hence, we utilize the first 70 % cycles data as the training data and the remaining 30 % cycles data as the testing data.

The penalty parameter  $c$  and Gaussian kernel function radius  $g$  of the SVR-based model were determined through grid search, with values set as 0.02 and 10, respectively. Besides, other hyperparameters of SVR-based model for both battery A1 and A2 are set as follows: the tolerance for stopping criterion  $tol$  and the tolerable error  $\epsilon$  are set as 0.0001 and 0.01, respectively. The optimal number of hidden nodes for the three-layer BPNN-based SOH estimation models of battery A1 and A2 is both determined to be 8 after evaluating estimation accuracy using 3–10 nodes. The L-BFGS method has been chosen for parameter optimization due to its stable and faster to train and ease of convergence checking.

The SOH estimation results using SVR-based and BPNN-based model for two batteries are shown in Fig. 10. It is evident that SVR-based model and BPNN-based model both present prominent precisions.

Fig. 10(a), (b), (c) and (d) present the battery A1 SOH estimation and the absolute errors  $\Delta SOHs$ . The  $\Delta SOHs$  of SVR-based model and BPNN-based model are all within 3 %. The  $\Delta SOH$  gets larger when the number of cycles reaches 400. The results of battery A2 SOH estimation and  $\Delta SOHs$  are plotted in Fig. 10(e), (f), (g) and (h), the  $\Delta SOHs$  of SVR-based model and BPNN-based model are both within 2 %. Specifically, the significant deviations are mainly concentrated in the stage of rapid capacity degradation of battery.

The MAEs and RMSEs for two models of two batteries are given in Table 3. The SVR-based model outperforms the BPNN-based model in both battery A1 and A2 according to the comparison. The reason may be that SVR is suitable for learning with small samples. BPNN-based model may achieve high accuracy with the increase of data size.

##### 4.2. Performance evaluation and discussion

Furthermore, two representative feature extraction methods are used to evaluate the validity of the proposed health feature extraction approach. Reference [2] applies filter algorithms to smooth battery IC curves and advanced signal process approach to fit the peaks of IC curves. And then the peak area, position and height of smooth IC curve is treated as the significant features to establish the battery degradation model. Literature [18] extracts the peak position and height as the important health feature variables from IC curves using Pearson correlation coefficient. We apply the above two methods to obtain the health features, and then establish SOH estimation methods with SVR respectively. The hyperparameters of above two SVR models are set as in previous configurations. To facilitate explanation, the estimation model

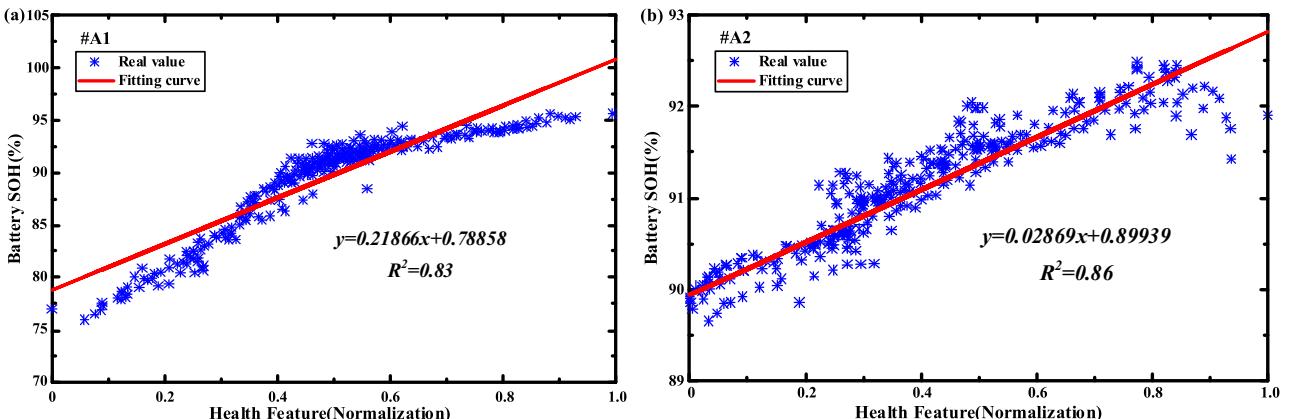
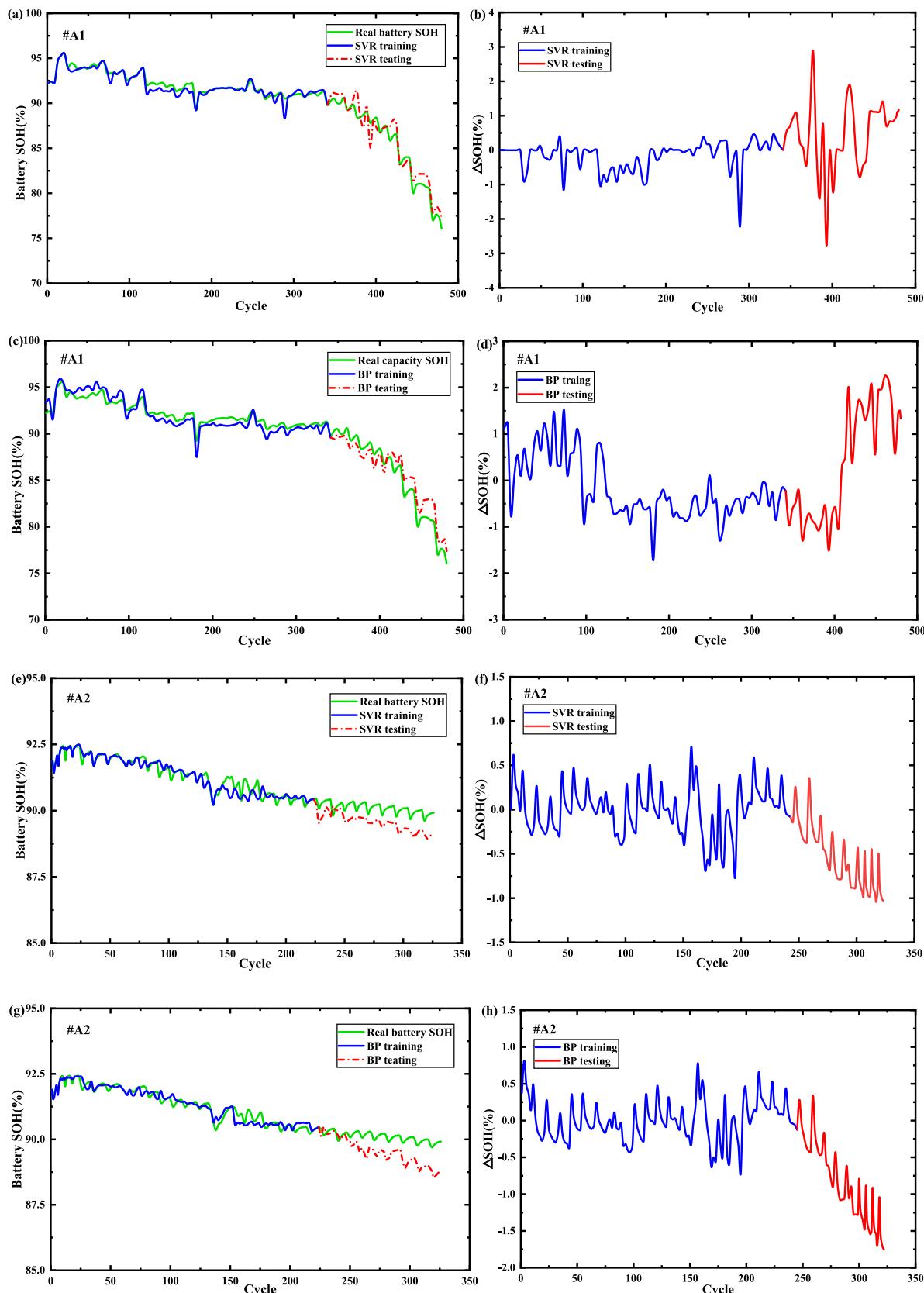


Fig. 9. The fitting results of health feature and battery SOH for two batteries with linear regression. (a) Battery A1; (b) Battery A2.



**Fig. 10.** The SOH estimation results for two batteries with SVR-based model and BPNN-based model. (a)–(d) Battery A1; (e)–(h) Battery A2.

**Table 3**

The performance evaluation of two SOH estimation models for two batteries.

| Battery    | SVR    |         | BPNN   |         |
|------------|--------|---------|--------|---------|
|            | MAE(%) | RMSE(%) | MAE(%) | RMSE(%) |
| Battery A1 | 0.889  | 1.145   | 1.162  | 1.331   |
| Battery A2 | 0.622  | 0.692   | 0.957  | 1.086   |

using peak area, position and height is labeled SVR-3, and the model with peak position and height is labeled SVR-2. Our method is denoted as SVR-AIC.

The Battery A1 SOH estimation results with above two health feature variables for are presented in Fig. 11. The performance evaluation of estimation methods with three models is shown in Table 4. The outcomes of estimation model using proposed feature in this manuscript are comparable to that of SVR-3, yet superior to result of SVR-2. It is worth noting that the proposed health feature method provides satisfactory results for the battery SOH estimation without smoothing the IC curve.

## 5. Conclusion

Accurate SOH estimation is essential to ensure the reliable and safe usage of LIBs. A novel health feature extraction approach is proposed based on IC analysis for battery SOH estimation in this manuscript. Firstly, the discrete IC curve of battery is obtained under different CVI of  $dv$  based on charging data. And then the charging voltage range with obvious variation trend of IC is selected and divided into several subintervals with  $\Delta v$ . Whereafter, the average IC of each voltage subinterval

**Table 4**

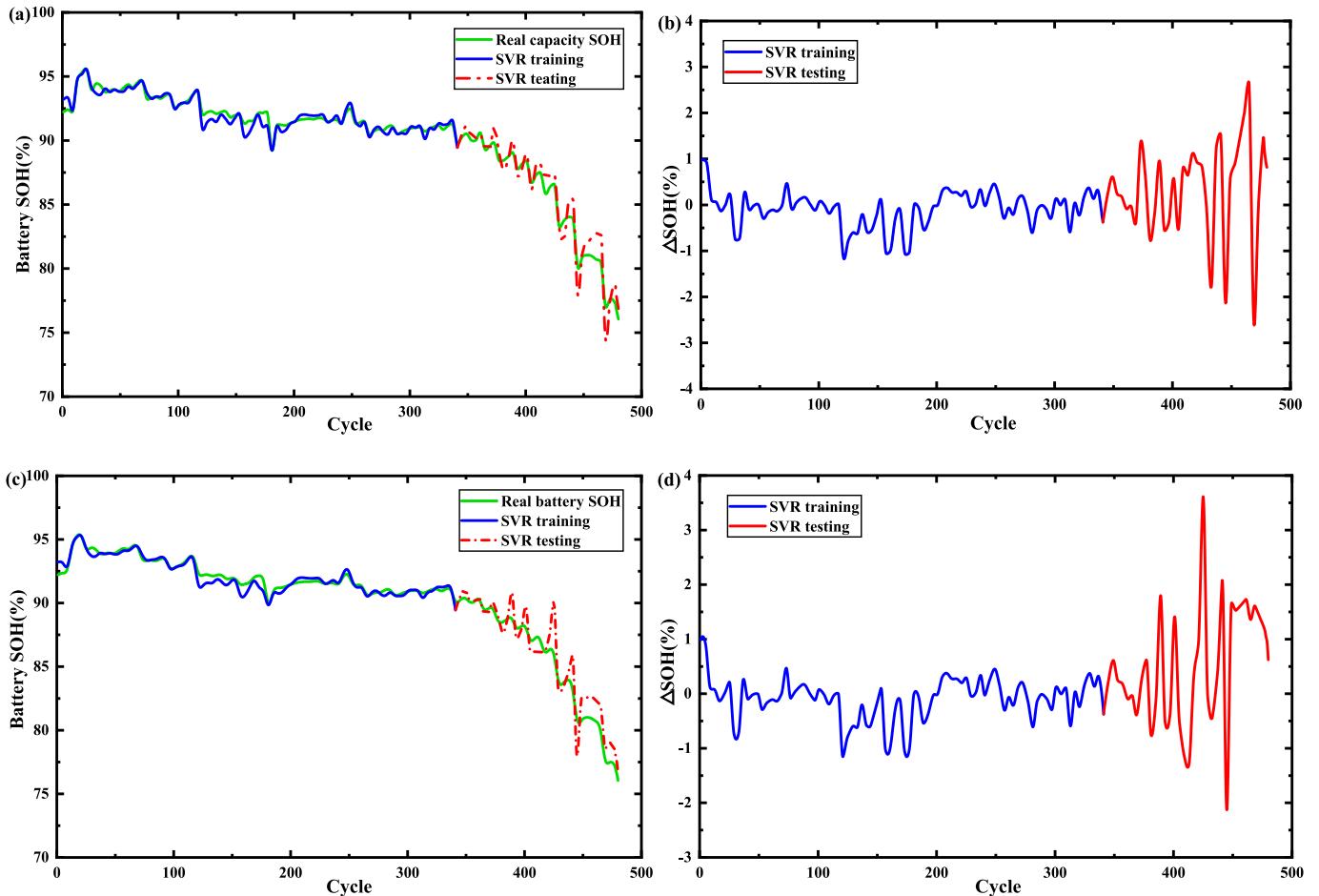
The performance evaluation of different SOH estimation models.

| Error indicators | SVR-2 | SVR-3 | SVR-AIC (Our method) |
|------------------|-------|-------|----------------------|
| MAE(%)           | 0.919 | 0.867 | 0.889                |
| RMSE(%)          | 1.250 | 1.155 | 1.145                |

is obtained. Furthermore, the average IC with the most consistent in relation to battery capacity degeneration is selected as the health feature. The correlation between features under different parameters ( $dv$  and  $\Delta v$ ) and the battery capacity is discussed. Finally, the most relevant average IC is selected as the feature to build SVR-based and BPNN-based battery SOH estimation models. The results demonstrate that the proposed health feature effectively reveals the battery performance degeneration. Additionally, this paper only utilizes the average IC of the extracted voltage subintervals to estimate battery SOH. Future work should prioritize the extraction and fusion of multiple health features.

## CRediT authorship contribution statement

**Changhao Piao:** Project administration, Funding acquisition. **Rongli Sun:** Writing – original draft, Methodology, Validation, Writing – review & editing. **Junsheng Chen:** Supervision, Conceptualization, Methodology, Writing – review & editing. **Mingjie Liu:** Supervision, Writing – review & editing. **Zhen Wang:** Data curation, Supervision.



**Fig. 11.** The SOH estimation results of battery A1 with different features. (a) SVR-3 estimation; (b) SVR-3 estimation errors; (c) SVR-2 estimation; (d) SVR-2 estimation errors.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The authors do not have permission to share data.

## Acknowledgments

This work is supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN202100620).

## References

- [1] X. Lu, Y. Qu, Y. Wang, et al., A comprehensive review on hybrid power system for PEMFC-HEV: issues and strategies, *Energy Convers. Manag.* 171 (2018) 1273–1291.
- [2] X. Li, C. Yuan, Z. Wang, State of health estimation for Li-ion battery via partial incremental capacity analysis based on support vector regression, *Energy* 203 (2020), 117852.
- [3] M. Berecibar, I. Gandiaga, I. Villarreal, et al., Critical review of state of health estimation methods of Li-ion batteries for real applications, *Renew. Sust. Energ. Rev.* 56 (2016) 572–587.
- [4] Y. Wang, R. Pan, C. Liu, et al., Power capability evaluation for lithium iron phosphate batteries based on multi-parameter constraints estimation, *J. Power Sources* 374 (2018) 12–23.
- [5] X. Li, Z. Wang, L. Zhang, Co-estimation of capacity and state-of-charge for lithium-ion batteries in electric vehicles, *Energy* 174 (2019) 33–44.
- [6] L. Zhou, L. He, Y. Zheng, et al., Massive battery pack data compression and reconstruction using a frequency division model in battery management systems, *J. Energy Storage* 28 (2020), 101252.
- [7] H. Ruan, H. He, Z. Wei, et al., State of health estimation of lithium-ion battery based on constant-voltage charging reconstruction, *IEEE J. Emerging Sel. Top. Power Electron.* 11 (4) (2021) 4393–4402.
- [8] K.K. Sadabadi, P. Ramesh, P. Tulpule, et al., Model-based state of health estimation of a lead-acid battery using step-response and emulated in-situ vehicle data, *J. Energy Storage* 30 (2021), 102353.
- [9] F. Liu, C. Shao, W. Su, et al., Online joint estimator of key states for battery based on a new equivalent circuit model, *J. Energy Storage* 52 (2022), 104780.
- [10] Z. Xu, J. Wang, P.D. Lund, Co-estimating the state of charge and health of lithium batteries through combining a minimalist electrochemical model and an equivalent circuit model, *Energy* 240 (2022), 122815.
- [11] I. Narayanan, D. Wang, et al., A stochastic optimal control approach for exploring tradeoffs between cost savings and battery aging in datacenter demand response, *IEEE Trans. Control Syst. Technol.* 26 (1) (2017) 360–367.
- [12] A. Singh, A. Izadian, S. Anwar, Model based condition monitoring in lithium-ion batteries, *J. Power Sources* 268 (2014) 459–468.
- [13] Q. Gong, P. Wang, Z. Cheng, A data-driven model framework based on deep learning for estimating the states of lithium-ion batteries, *J. Electrochim. Soc.* 169 (3) (2022), 030532.
- [14] M. Lin, C. Yan, W. Wang, et al., A data-driven approach for estimating state-of-health of lithium-ion batteries considering internal resistance, *Energy* 277 (2023), 127675.
- [15] Z. Deng, X. Hu, P. Li, et al., Data-driven battery state of health estimation based on random partial charging data, *IEEE Trans. Power Electron.* 37 (5) (2021) 5021–5031.
- [16] Y. Zhang, T. Wik, J. Bergström, et al., State of health estimation for lithium-ion batteries under arbitrary usage using data-driven multi-model fusion, *IEEE Trans. Transp. Electrific. (Early Access)* (2023). <https://doi.org/10.1109/TTE.2023.3267124>.
- [17] Y. Jiang, J. Jiang, C. Zhang, et al., State of health estimation of second-life LiFePO<sub>4</sub> batteries for energy storage applications, *J. Clean. Prod.* 205 (2018) 754–762.
- [18] C. Chang, Q. Wang, J. Jiang, et al., Lithium-ion battery state of health estimation using the incremental capacity and wavelet neural networks with genetic algorithm, *J. Energy Storage* 38 (2021), 102570.
- [19] L. Zheng, J. Zhu, D.C.C. Lu, et al., Incremental capacity analysis and differential voltage analysis based state of charge and capacity estimation for lithium-ion batteries, *Energy* 150 (2018) 759–769.
- [20] M. Lewerenz, A. Marongiu, A. Warnecke, et al., Differential voltage analysis as a tool for analyzing inhomogeneous aging: a case study for LiFePO<sub>4</sub> graphite cylindrical cells, *J. Power Sources* 368 (2017) 57–67.
- [21] S. Zhang, X. Guo, X. Dou, et al., A rapid online calculation method for state of health of lithium-ion battery based on coulomb counting method and differential voltage analysis, *J. Power Sources* 479 (2020), 228740.
- [22] S. Yang, C. Zhang, J. Jiang, et al., A voltage reconstruction model based on partial charging curve for state-of-health estimation of lithium-ion batteries, *J. Energy Storage* 35 (2021), 102271.
- [23] Y. Merla, B. Wu, V. Yuft, et al., Novel application of differential thermal voltammetry as an in-depth state-of-health diagnosis method for lithium-ion batteries, *J. Power Sources* 307 (2016) 308–319.
- [24] X. Feng, J. Li, M. Ouyang, et al., Using probability density function to evaluate the state of health of lithium-ion batteries, *J. Power Sources* 232 (2013) 209–218.
- [25] H.H. Goh, Z. Lan, D. Zhang, et al., Estimation of the state of health (SOH) of batteries using discrete curvature feature extraction, *J. Energy Storage* 50 (2022), 104646.
- [26] A. Tian, C. Yang, Y. Gao, et al., A state of health estimation method of lithium-ion batteries based on DT-IC-V health features extracted from partial charging segment, *Int. J. Green Energy* 20 (9) (2023) 997–1011.
- [27] R. Xiong, Y. Sun, C. Wang, et al., A data-driven method for extracting aging features to accurately predict the battery health, *Energy Storage Materials* 57 (2023) 460–470.