



ARTICLE

# A Joint Estimation Method of SOC and SOH for Lithium-ion Battery Considering Cyber-Attacks Based on GA-BP

Tianqing Yuan<sup>1,2</sup>, Na Li<sup>1,2</sup>, Hao Sun<sup>3</sup> and Sen Tan<sup>4,\*</sup>

<sup>1</sup>Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin, 132012, China

<sup>2</sup>Department of Electrical Engineering, Northeast Electric Power University, Jilin, 132012, China

<sup>3</sup>College of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin, 132022, China

<sup>4</sup>Center for Research on Microgrids, Department of Energy, Aalborg University, Aalborg, 9220, Denmark

\*Corresponding Author: Sen Tan. Email: sta@energy.aau.dk

Received: 13 July 2024 Accepted: 14 August 2024 Published: 12 September 2024

## ABSTRACT

To improve the estimation accuracy of state of charge (SOC) and state of health (SOH) for lithium-ion batteries, in this paper, a joint estimation method of SOC and SOH at charging cut-off voltage based on genetic algorithm (GA) combined with back propagation (BP) neural network is proposed, the research addresses the issue of data manipulation resulting from cyber-attacks. Firstly, anomalous data stemming from cyber-attacks are identified and eliminated using the isolated forest algorithm, followed by data restoration. Secondly, the incremental capacity (IC) curve is derived from the restored data using the Kalman filtering algorithm, with the peak of the IC curve (ICP) and its corresponding voltage serving as the health factor (HF). Thirdly, the GA-BP neural network is applied to map the relationship between HF, constant current charging time, and SOH, facilitating the estimation of SOH based on HF. Finally, SOC estimation at the charging cut-off voltage is calculated by inputting the SOH estimation value into the trained model to determine the constant current charging time, and by updating the maximum available capacity. Experiments show that the root mean squared error of the joint estimation results does not exceed 1%, which proves that the proposed method can estimate the SOC and SOH accurately and stably even in the presence of false data injection attacks.

## KEYWORDS

Lithium-ion batteries; state of charge; state of health; cyber-attacks; genetic algorithm; back propagation neural network

## 1 Introduction

Characterized by low self-discharge and high output voltage, the lithium-ion battery is a green energy storage device, which is highly valued in many fields. To ensure optimal performance of vehicle power batteries, a battery management system (BMS) is configured [1].

Additionally, due to the continuous development of network technology, BMS has become a cyber-physical system whose safety directly impacts the vehicle's driving status. However, this also



means the BMS faces the threat of cyber-attacks [2,3]. Once the system is subjected to false data injection attacks (FDIAs) [4,5], the estimation accuracy of the battery's SOC and SOH will be affected. Hence, detecting and repairing FDIAs is significant for accurately estimating SOC and SOH. In Reference [6], Gallo et al. propose a distributed monitoring approach to detect false data, but it has low sensitivity and slow detection speed. In Reference [7], an FDIAs detection approach that combines model predictive control and artificial neural networks is proposed, but this method is prone to localized convergence during online identification. The isolated forest algorithm can detect anomalous data quickly and is well suited to deal with large-scale datasets, so this algorithm is used to detect and remove FDIAs widely [8].

The common methods for estimating SOC include the ampere-hour integration method, the open circuit voltage method (OCV), the filtering method, and the data-driven method. Among these, the ampere-hour integration method is simple in principle and computationally efficient, but it is susceptible to interference, and has high initial value requirements [9]. The OCV method requires long-term static processing of the battery and is inadequate for online applications [10]. The filtering method is generally combined with the equivalent circuit model (ECM) of the battery [11], offering strong self-correction ability as well as high accuracy, but the parameter calculation of ECM is cumbersome. Unlike methods relying on the intrinsic mechanism of the battery, the data-driven method employs a machine learning model to construct a SOC estimation model using extensive training data [12]. This method is easy to implement and offers high prediction accuracy, but it demands substantial computational resources and training data. The continuous rise of big data offers effective solutions to the aforementioned challenges.

Currently, the main SOH estimation methods include direct measurement method, empirical recession models method, and data-driven method. The direct measurement method assesses the parameters directly impacting battery life and constructs a function mapping to determine the SOH. While its principle is straightforward, its practical application is constrained by demanding accuracy requirements and computational complexity [13]. The empirical recession model utilizes functions to align with the battery's capacity decline trend [14], it demonstrates strong robustness but faces challenges in adapting to diverse capacity decay trends due to individual variations. The data-driven method reduces the process of physical modeling and parameter identification. Its accuracy largely hinges on the selection of HF and algorithms [15].

From the concept of SOC and SOH, it is evident that the common parameter between them is the current maximum available capacity, which decreases with an increased number of cycles. Therefore, conducting a joint estimation of SOC and SOH is essential. In Reference [16], Yin et al. propose a joint estimation method of SOC and SOH using the particle filtering algorithm. Li et al. [17] employ a combination of gate recurrent units and convolutional neural networks to estimate both parameters. However, the accuracy of filtering methods is influenced by model accuracy, and the deep learning method has high requirements on training data and hardware equipment, posing challenges for rapid updates of SOC and SOH. GA-BP neural network can effectively avoid the above problems, with high prediction accuracy and strong generalization ability, so it is used to construct the estimation model widely. Considering that the critical point of constant current charging and constant voltage charging is the charging cut-off voltage, and the SOC value is already at a very high level during constant voltage charging, the battery's capacity decline becomes more pronounced with the growth of charging time. Therefore, this paper investigates the joint estimation of SOC and SOH specifically at the charging cut-off voltage.

From the above introduction, it can be seen that the isolated forest algorithm can detect and eliminate FDIAs quickly, and the data-driven method has a good potential for application in the field of lithium-ion battery state estimation. Therefore, in this paper, the isolated forest algorithm is used to detect the abnormal data, after which the GA-BP neural network is used to build the estimation model. By combining the global search ability of GA and the nonlinear fitting ability of BP neural network, the neural network model can be effectively optimized to improve the accuracy and generalization ability of prediction. Experimental validation demonstrates that this method can estimate the SOC and SOH accurately and stably even in the presence of cyber-attacks.

## 2 Data Processing

### 2.1 SOC and SOH

SOC is defined as the ratio of the battery's current capacity to maximum available capacity, the SOC is calculated at charging cut-off voltage in the paper. The SOH is defined as the ratio of the current maximum available capacity to the rated capacity. The formulas are as follows:

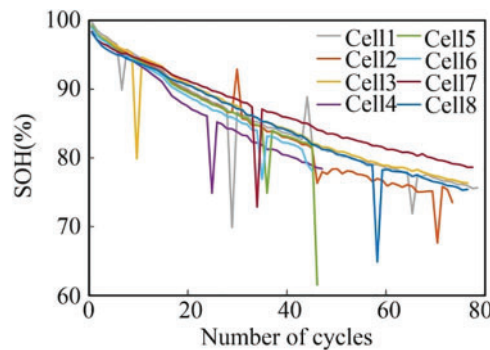
$$\begin{cases} SOC = \frac{IT}{C_m} \times 100\% \\ SOH = \frac{C_m}{C_o} \times 100\% \end{cases} \quad (1)$$

where:  $I$  represents the current value of constant current charging;  $T$  is the constant current charging time;  $C_m$  is the current maximum available capacity;  $C_o$  indicates the rated capacity.

### 2.2 Datasets

To validate the method proposed in this paper, the dataset comprises normal data points and anomalies. The normal points include data from the following lithium-ion battery datasets, while anomalies consist of artificially altered data aimed at simulating the FDIAs.

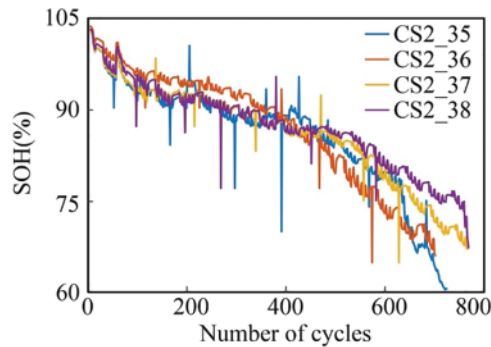
Eight batteries from the Oxford University battery dataset are selected as experimental data [18], these batteries are experimented at 40°C, with a rated capacity of 740 mA·h and a nominal voltage of 4.2 V. The experimental procedure is as follows: firstly, charge at a constant current of 1.48 A; secondly, discharge under simulated urban driving dynamic conditions; finally, repeat the above steps, and every 100 cycles, discharge occurs at a constant current until the voltage reaches 2.7 V. Fig. 1 depicts the SOH change curves of these eight batteries post-simulated cyber-attack.



**Figure 1:** SOH curves for simulated cyber-attacks on the Oxford University battery dataset

As shown in Fig. 1, each battery exhibits significant variability in SOH change curves, but all of them have a tendency to decrease SOH with an increased number of cycles. Additionally, the curves display abnormal fluctuations attributed to the cyber-attack.

Four batteries from the Center for Advanced Life Cycle Engineering (CALCE) of Maryland University battery dataset were selected as experimental data [19]. These batteries are rated at 1.1 A·h and are experimented at an ambient temperature of 1°C. The aging experimental process adopts a constant current charging plus constant voltage charging mode. In this mode, the charging stage was conducted at a constant current of 0.5 C with a charging cut-off voltage set to 4.2 V, and the charging was terminated when the current decreased to 0.05 A; the discharging stage was carried out at a constant current of 1C until the voltage dropped to 2.7 V, and then the aforementioned steps were repeated. The SOH change curves for batteries CS2\_35, CS2\_36, CS2\_37, and CS2\_38 after the simulated cyber-attack are depicted in Fig. 2.



**Figure 2:** SOH curves for simulated cyber-attacks on the CALCE battery dataset

### 2.3 Detection and Interpolation of Abnormal Data

In this paper, the isolated forest algorithm is employed for monitoring and removing anomalous data. The isolated forest algorithm is an unsupervised method for detecting anomalous data, its main idea is to map data into a binary tree structure through constantly randomized divisions and assess the isolation degree of each data based on its path length within the tree [20]. The steps for detecting anomalous data based on isolated forest algorithm are outlined below:

- (1) Import the sample dataset;
- (2) Select a factor and threshold randomly from the dataset for binary fission;
- (3) Repeat Step 2 until the tree reaches the specified height;
- (4) Calculate the length of the path from the sample data point to the root node;
- (5) Assess outliers based on the anomaly evaluation coefficients of the data points as follows:

$$S(x, n) = 2^{\left\{ \frac{E[h(x)]}{c(n)} \right\}} \quad (2)$$

where:  $n$  represents the total number of samples in the dataset;  $c(n)$  denotes the average path length of the binary tree;  $h(x)$  indicates the path length of the tested sample in a binary tree;  $E[h(x)]$  represents the average path length of the tested sample across all binary trees. If  $S(x, n)$  tends to approach 1, there is a higher probability of it being an anomaly; if all  $S(x, n)$  is less than or distributed around 0.5, the data is confirmed to be normal.

As abnormal data are excluded from the dataset, the measurement data of this node are missing, considering that the state estimation of lithium-ion battery is a time series problem, it is necessary to supplement the missing data. This paper employs the polynomial interpolation method to supplement the excluded anomalies. The method aims to make the sum of squares of the errors ( $I$ ) between the fitted function and the actual curves corresponding to each point  $(t_i, y_i)$  ( $i = 0, 1, 2, \dots, n$ ) reach a minimum value. The formula is as follows:

$$\frac{\partial I}{\partial a_j} = 2 \sum_{i=0}^m \left( \sum_{k=0}^n a_k t_i^k - y_i \right) \cdot t_i^j = 0 \quad (3)$$

where:  $j = 0, 1, 2, \dots, n$ , the optimal fitted curve is derived by solving  $a_0, a_1, a_2, \dots, a_n$ .

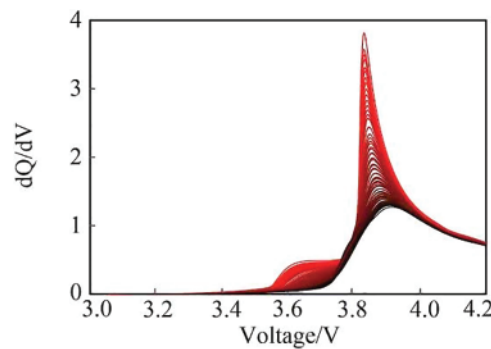
## 2.4 Health Factor Extraction

Health factor extraction is performed on the data following abnormal data repair, and the extracted HF should have a high correlation with variables, be easy to extract, and conform to practical applications. Battery charging and discharging curves change significantly with an aging degree, in the actual situation, the battery is strongly affected by the random load discharge during the discharging process. Conversely, charging processes are stable, resulting in smoother voltage curves conducive to HF extraction. Consequently, this study employs incremental capacity analysis (ICA) [21] to numerically differentiate the charging curves and approximate corresponding  $dQ/dV$  curves, the calculation formula is shown in Eq. (5). This method can better reflect the changes in the charging voltage curve caused by the aging.

$$\frac{dQ}{dV}(k) \approx \frac{Q(k) - Q(k-L)}{V(k) - V(k-L)} \quad (4)$$

where:  $Q$  is the capacity;  $L$  is the differential step;  $V$  is the voltage;  $k$  is the sampling moment.

Due to the low accuracy of current voltage and current sensors, causing the problem of noise interference when calculating the difference between capacity and voltage, the Kalman filter (KF) algorithm is used for smoothing [22]. One of the battery Cell1 IC curves is shown in Fig. 3.



**Figure 3:** IC curves of the Cell1 under full cycles

As shown in Fig. 3, the curve from bright to dark in the graph indicates the more serious the aging of the battery is, the lower the ICP value will be, and the higher the voltage corresponding to ICP ( $U_p$ ) will be. It can be seen that ICP and  $U_p$  have a good correlation with the SOH.

To quantify the correlation between ICP and SOH,  $U_p$  and SOH, this section evaluates them using Pearson and Spearman correlation coefficients expressed as:

$$Pearson = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \quad (5)$$

$$Spearman = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \quad (6)$$

where:  $X$  is the health factor;  $Y$  is SOH;  $x_i$  and  $y_i$  are sample data, respectively.

The correlation between two variables can be measured by Pearson and Spearman correlation coefficients. A higher absolute value indicates a stronger correlation between two variables. If the absolute value is 1, the variables are perfectly correlated. [Tables 1](#) and [2](#) present the Pearson and Spearman correlation coefficients for ICP with SOH and  $U_p$  with SOH, respectively. It can be seen that the correlation coefficients for each battery almost all exceed 0.95, indicating a strong correlation between the selected HFs and SOH.

**Table 1:** Coefficient between ICP and SOH

Battery number	Pearson	Spearman
Cell1	0.96468	0.99825
Cell2	0.95414	0.99546
Cell3	0.96889	0.99902
Cell4	0.97604	0.99931
Cell5	0.90186	0.99914
Cell6	0.97364	0.99908
Cell7	0.97131	0.99937
Cell8	0.96778	0.99880
CS2_35	0.95028	0.98789
CS2_36	0.96539	0.99508
CS2_37	0.96196	0.99164
CS2_38	0.96119	0.99090

**Table 2:** Coefficient between  $U_p$  and SOH

Battery number	Pearson	Spearman
Cell1	-0.97752	-0.98876
Cell2	-0.97821	-0.98988
Cell3	-0.98421	-0.99327
Cell4	-0.98573	-0.99326
Cell5	-0.98128	-0.99482
Cell6	-0.98327	-0.99117

(Continued)

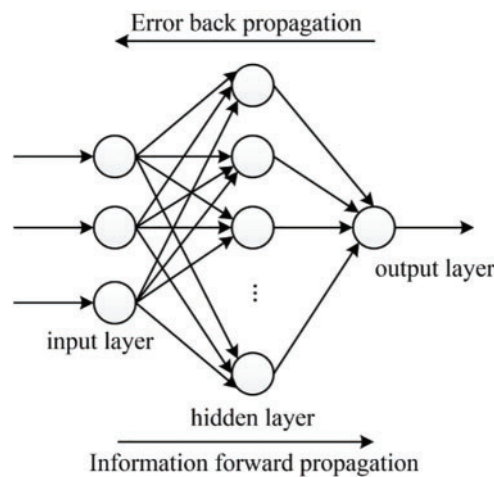
**Table 2 (continued)**

Battery number	Pearson	Spearman
Cell7	-0.98450	-0.99525
Cell8	-0.98299	-0.99070
CS2_35	-0.98153	-0.97835
CS2_36	-0.98077	-0.98169
CS2_37	-0.97962	-0.97376
CS2_38	-0.90031	-0.85771

### 3 Joint Estimation of SOC and SOH

#### 3.1 BP Neural Network

The structure of the BP neural network is depicted in Fig. 4 [23]. BP neural network comprises a three-layer structure consisting of input, hidden, and output layers. Neurons are present in each layer, with thresholds on them, and neurons in each layer are connected by weights. The process of neural network training can be described as follows: a sample is brought into the network model and transmitted to the neurons of the input layer, then it is processed sequentially through the hidden and output layers to produce the final output. This process known as forward propagation involves continuous layer-by-layer advancement and modification of the weights. If discrepancies occur between the output result and the actual value, an error is propagated in the reverse direction, similar to the forward propagation process. These steps are repeated until the error is minimized, ensuring sample outputs fall within acceptable error limits.



**Figure 4:** The BP neural network structure

To summarize, the process of training a BP neural network involves iteratively adjusting weights and thresholds to minimize errors. The weight update formula is as follows:

$$\Delta\omega = \eta \frac{\partial E}{\partial \omega} \quad (7)$$

where:  $\eta$  is the learning rate;  $\omega$  is the weight;  $E$  is the error.

Define the input information as  $S = (s_1, s_2, s_3, \dots, s_n)^T$ , the output information as  $C = (c_1, c_2, c_3, \dots, c_m)^T$ , the input layer has  $n$  nodes, the hidden layer has  $h$  nodes, and the output layer has  $m$  nodes. The output expressions of the hidden layer and the output layer are as follows:

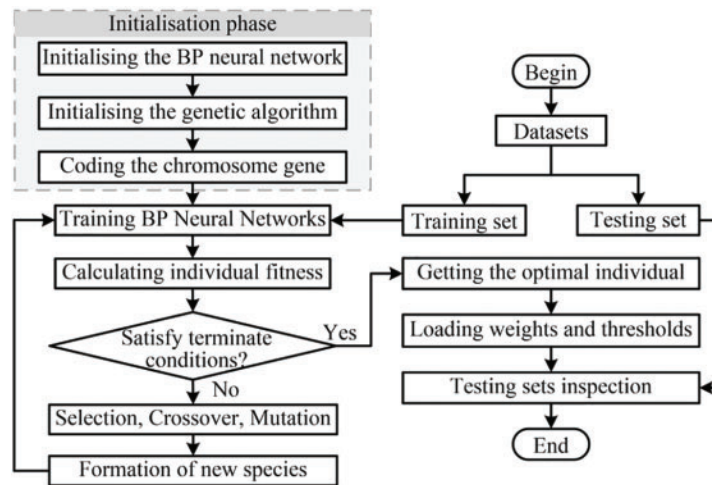
$$P_j = g\left(\sum_{i=1}^n \alpha_{ij}x_i\right), j = 1, 2, 3, \dots, h \quad (8)$$

$$O_f = g\left(\sum_{j=1}^h \beta_{jf}P_j\right), f = 1, 2, 3, \dots, m \quad (9)$$

where:  $P_j$  represents the output of the hidden layer,  $x_i$  denotes the output of the input layer,  $\alpha_{ij}$  signifies the weight between the input and hidden layers,  $O_f$  donates the output of the output layer,  $\beta_{jf}$  donates the weight between the hidden and output layers, and  $g$  denotes the activation function.

### 3.2 GA-BP Neural Network

The BP neural network exhibits some randomness in selecting initial weights and thresholds, potentially reducing the estimation accuracy. Therefore, this paper combines the GA with BP neural network, which can effectively overcome the shortcomings of the respective algorithms, such as slow convergence, falling into the local minima, high time cost, and so on. Thus, the neural network model can be optimized to improve the accuracy and training speed of regression prediction [24]. The flowchart of the GA-BP neural network is depicted in Fig. 5.



**Figure 5:** Flowchart of the GA-BP neural network algorithm

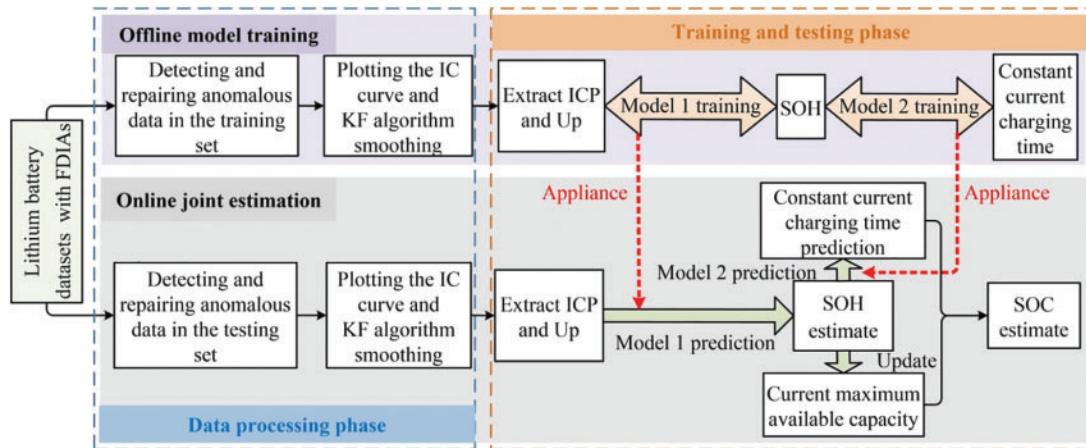
The implementation process of the GA-BP neural network is described as follows: firstly, delineate the training set and testing set. Secondly, initialize the BP neural network determine its parameters, and then initialize the parameters of the genetic algorithm, including the initial population size and the maximum number of genetic generations. Thirdly, encode the chromosomes gene. Subsequently, train the BP neural network using the training set, calculate individual fitness, and evaluate whether termination conditions are satisfied; if not, carry out the next operations with selection, crossover, and mutation to update the genetic coding of the population and then return to the step of training BP neural network to repeat the above process until the termination conditions are met. Finally, assign the



optimal chromosome coding to the initial weights and thresholds of the BP neural network, train the BP neural network using the optimized weights and thresholds, and verify its performance with the testing set.

### 3.3 Joint Estimation of SOC and SOH Based on GA-BP Neural Network

The joint estimation based on the GA-BP neural network is illustrated in Fig. 6.



**Figure 6:** Flowchart of the joint estimation of the SOC and the SOH

The joint estimation of SOC and SOH of lithium-ion batteries based on the GA-BP neural network is divided into two phases: offline training and online testing. The offline phase primarily involves training the GA-BP neural network. Initially, anomalous data in the training set are repaired. Subsequently, the KF algorithm is applied to extract smooth IC curves, with ICP and  $U_p$  serving as the health factors. Two BP neural networks are used to establish Model 1 and Model 2, respectively, where HF is the input and SOH is the output, which are brought into Model 1 for training; SOH is the input and constant current charging time is the output for Model 2 training. Theoretically, the constant current charging time can be measured online. Considering that electric vehicle users have a certain randomness during the startup charging phase, which makes it difficult to measure. Therefore, it is necessary to obtain the constant current charging time by constructing Model 2. Table 3 presents Pearson and Spearman correlation coefficients demonstrating a strong correlation, with coefficients exceeding 0.99 across all cells, underscoring the robust relationship between constant current charging time and SOH.

**Table 3:** Coefficient between constant current charging time and SOH

Battery number	Pearson	Spearman	Battery number	Pearson	Spearman
Cell1	0.99982	0.99992	Cell7	0.99985	0.99984
Cell2	0.99856	0.99858	Cell8	0.99983	0.99991
Cell3	0.99979	0.99985	CS2_35	0.99671	0.99408
Cell4	0.99989	0.99988	CS2_36	0.99761	0.99718
Cell5	0.99974	1	CS2_37	0.99684	0.99526
Cell6	0.99968	1	CS2_38	0.97839	0.97023

The online process focuses on estimating SOH and real-time correcting the SOC value at the charging cut-off voltage. Firstly, repairing abnormal data in the testing set, extracting health factors, and inputting them into trained Model 1 to obtain the SOH estimation value; secondly, the SOH estimation value is brought into the trained Model 2 to obtain the predicted value of the constant current charging time, and update the current maximum available capacity according to the SOH estimation value; finally, calculating the SOC estimation value under the charging cut-off voltage according to Eq. (1). In summary, the proposed method realizes the multi-step mapping from HF to the SOC estimation value, completing the joint estimation of SOC and SOH.

The GA-BP neural networks designed in this paper all adopt a 3-layer structure. Model 1 has 2 nodes and Model 2 has 1 node in the input layer; they all have 8 nodes in the hidden layer, and 1 node in the output layer. The maximum number of genetic generations is set to 100, the number of populations is 30, and the fitness function is the error of the BP neural network. Other designs of the GA-BP neural network are detailed below:

(1) Data preprocessing. A normalization method is used to preprocess the data before experimenting, the expressions for normalization and inverse normalization are as follows:

$$\hat{x} = \frac{2(x - x_{\min})}{x_{\max} - x_{\min}} - 1 \quad (10)$$

$$x = \frac{(x_{\max} - x_{\min})(\hat{x} + 1)}{2} + x_{\min} \quad (11)$$

where:  $x$  is the sample data value,  $x_{\max}$  is the largest data value and  $x_{\min}$  is the smallest data value.

(2) Selection of loss function. The model employs the mean square error (MSE) function as its loss function, expressed as follows:

$$MSE = \frac{1}{N} \sum_{k=1}^N (y_k - \hat{y}_k)^2 \quad (12)$$

where:  $y_k$  and  $\hat{y}_k$  denote the true and estimated values of SOH and SOC of the battery, respectively;  $N$  denotes the number of samples in the training set.

(3) Parameter optimizer selection. Adam is employed as the parameter optimization algorithm for the model. The initial learning rate is set to 0.001 and the number of iterations for training is set to 500.

#### 4 Experimental Results and Analysis

To validate the feasibility and accuracy of the paper's method, this section performs the joint estimation of SOC and SOH based on the aforementioned battery datasets. Evaluation metrics, namely root mean squared error (RMSE), mean absolute error (MAE), and maximum error (MAX), are employed to assess performance. These metrics are calculated as follows:

$$RMSE = \sqrt{\frac{1}{N} \left[ \sum_{i=1}^N (x_i - \hat{x}_i)^2 \right]} \quad (13)$$

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

$$MAX = \max (|x_i - \hat{x}_i|) \tag{15}$$

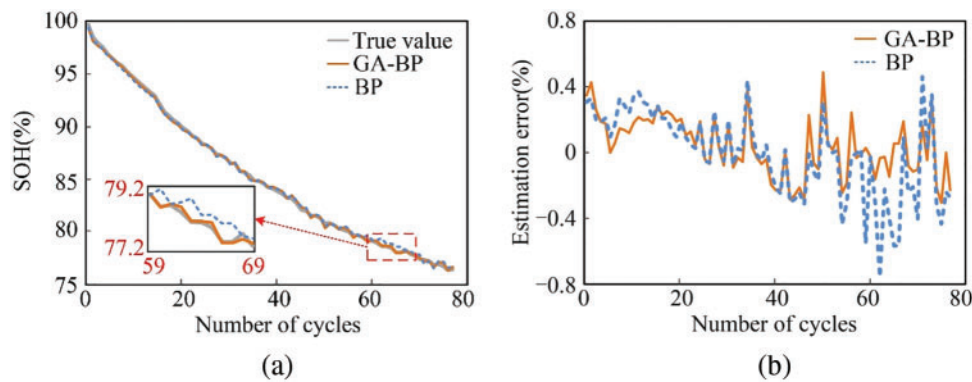
The battery state estimation model is required to maintain independence between training and testing datasets. The division results are summarized in Table 4.

**Table 4:** Testing set and training set division situation

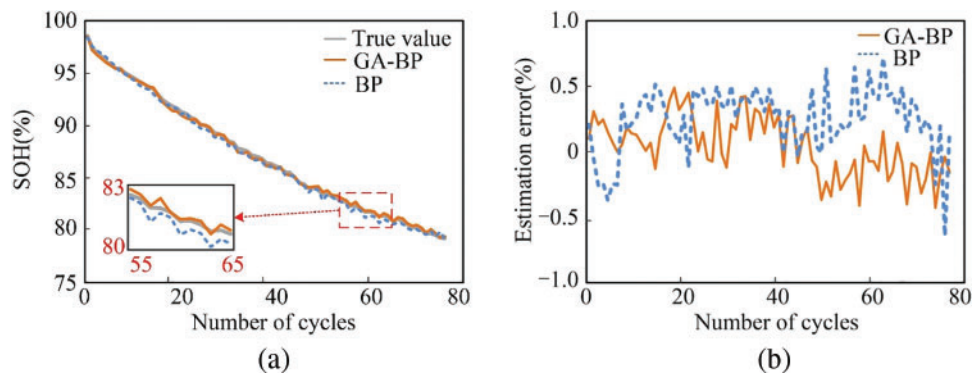
Dataset	Training set	Testing set
Oxford	Cell2, Cell3, Cell4, Cell5, Cell6, Cell8	Cell1, Cell7
CALCE	CS2_35, CS2_36, CS2_38	CS2_37

### 4.1 Results of SOH Estimation

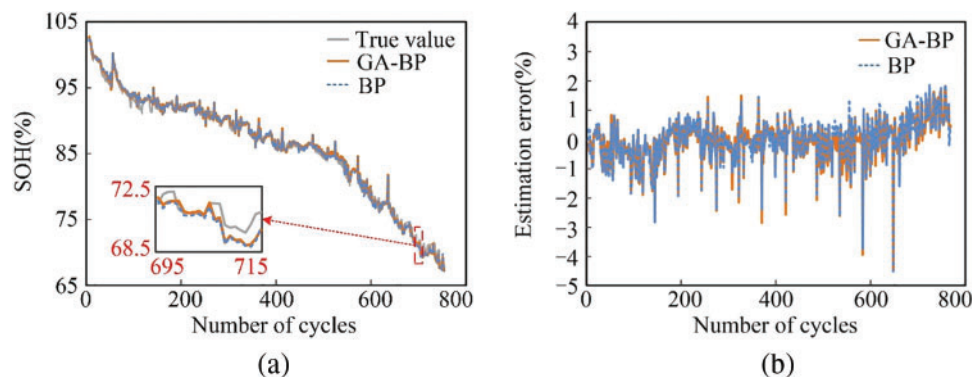
This section aims to demonstrate the superiority of the GA-BP neural network in estimating the SOH of lithium-ion batteries. The SOH estimation model based on the BP neural network is trained and tested using the Oxford and CALCE battery datasets, and the estimation results are compared with those obtained using the method proposed in this paper. Figs. 7 to 9 illustrate the SOH estimation curves and error curves for Cell1, Cell7, and CS2\_37, respectively.



**Figure 7:** The SOH estimation results of Cell1. (a) SOH. (b) Estimation error curve



**Figure 8:** The SOH estimation results of Cell7. (a) SOH. (b) Estimation error curve



**Figure 9:** The SOH estimation results of CS2\_37. (a) SOH. (b) Estimation error curve

Evaluation of the Oxford battery dataset (Figs. 7 and 8) reveals that both methods exhibit accurate SOH estimation curves. The BP neural network achieves an estimation error typically within  $\pm 0.6\%$ , whereas the GA-BP neural network achieves an error typically within  $\pm 0.4\%$ , with a maximum of  $\pm 0.5\%$ . This demonstrates superior SOH estimation accuracy with the GA-BP approach. Analysis of Table 5 indicates that this paper's algorithm achieves RMSE and MAE values within 0.3% and the MAX value not exceeding 0.7%, demonstrating relatively small errors. It shows that the HF selected in this paper can not only reveal the aging information of the battery well but also effectively deal with the capacity decline changes due to battery inconsistency. Furthermore, it underscores the GA-BP neural network's robust mapping capability for precise SOH estimation of lithium-ion batteries.

**Table 5:** Results of the SOH estimation for each method

	Battery number	RMSE (%)	MAE (%)	MAX (%)
BP	Cell1	0.220	0.169	0.670
	Cell7	0.289	0.227	0.749
	CS2_37	0.914	0.733	4.253
GA-BP	Cell1	0.170	0.135	0.423
	Cell7	0.207	0.167	0.624
	CS2_37	0.897	0.724	3.726

Evaluation of the CALCE dataset in Fig. 9 reveals superior accuracy and better followability of the GA-BP neural network compared to the BP neural network algorithm. Additionally, as shown in Table 5, both MAE and RMSE are maintained below 1%, and the MAX not exceeding 4%, which indicates the stable and reliable SOH estimation by the GA-BP neural network algorithm. These findings show that the proposed method has good generalizability and can be applied to SOH estimation for a variety of lithium-ion batteries.

#### 4.2 Results of SOC Estimation

Based on the accurate and reliable estimation of SOH, the joint SOC and SOH estimation is conducted. Figs. 10 to 12 depict the SOC estimation results using both methods on the testing data.

Specifically, the SOC estimation curves and error curves are shown for cells Cell1, Cell7, and CS2\_37, respectively. Table 6 presents the MAE, RMSE, and MAX values obtained by both algorithms.

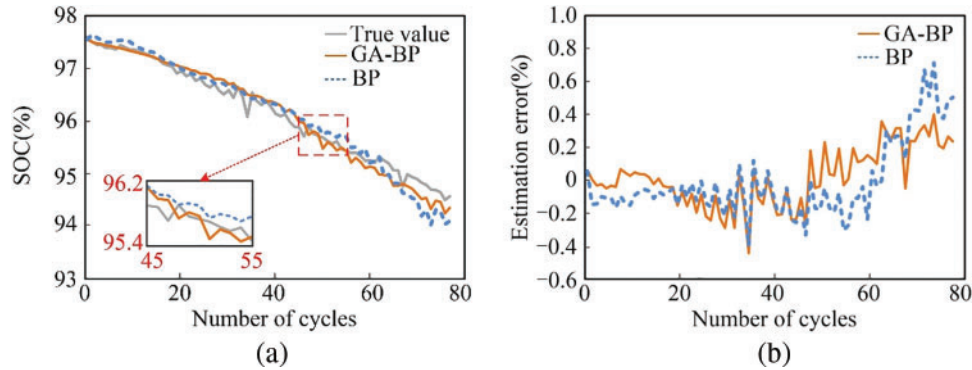


Figure 10: The SOC estimation results of Cell1. (a) SOC. (b) Estimation error curve

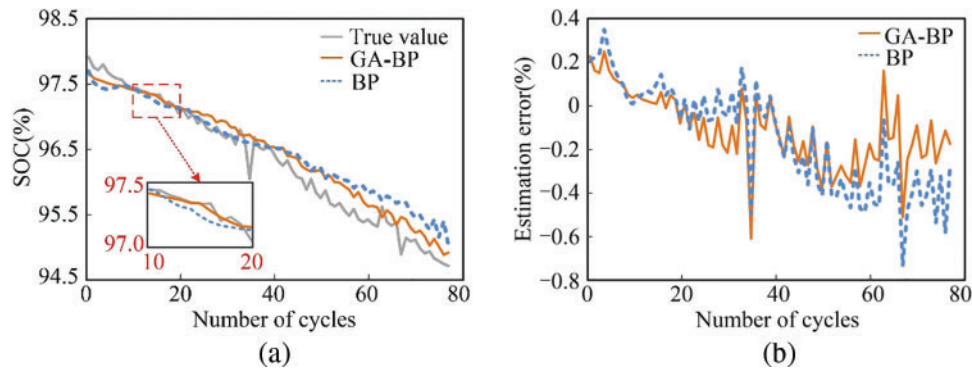


Figure 11: The SOC estimation results of Cell7. (a) SOC. (b) Estimation error curve

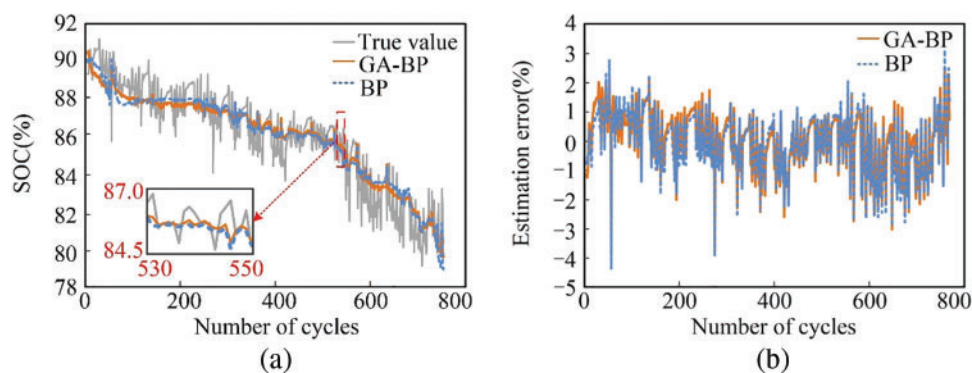


Figure 12: The SOC estimation results of CS2\_37. (a) SOC. (b) Estimation error curve

Figs. 10 and 11 depict the estimation results of both methods based on the Oxford battery dataset. The estimation curve of this paper’s method demonstrates superior performance, with errors typically within  $\pm 0.4\%$  and the maximum not exceeding  $\pm 0.6\%$ . Analysis of Table 6 reveals that the MAE and RMSE obtained by this paper’s method are smaller compared to those derived from the BP

neural network, with MAE and RMSE typically restricted to 0.3%, and the MAX not exceeding 0.5%, reflecting the advantages of this paper's method in estimating SOC and underscoring the GA-BP neural network's robust mapping capability.

**Table 6:** Results of the SOC estimation for each method

	Battery number	RMSE (%)	MAE (%)	MAX (%)
BP	Cell1	0.262	0.208	0.763
	Cell7	0.315	0.279	0.656
	CS2_37	0.657	0.485	4.251
GA-BP	Cell1	0.182	0.150	0.472
	Cell7	0.212	0.177	0.440
	CS2_37	0.653	0.467	4.257

Examination of the SOC estimation results from the CALCE dataset in Fig. 12 shows that although the SOC fluctuation of battery CS2\_37 at the charging cut-off voltage is drastic, the paper's method effectively tracks the true values. Analysis of Table 6 indicates that errors are maintained at a low level, with MAE and RMSE restricted to 0.7%, and the MAX not exceeding 5%, demonstrating superior performance. These findings reveal that the paper's method achieves high accuracy and broad applicability, facilitating real-time correction of SOC under the corresponding charging cut-off voltage of different types of lithium-ion batteries.

## 5 Conclusion

In this paper, a joint estimation method of SOC and SOH for lithium-ion batteries at charging cut-off voltage based on GA-BP neural network is proposed, this method addresses data tampering due to cyber-attacks on BMS. Firstly, FDIAs are detected and eliminated using the isolated forest algorithm, and then normal battery data can be reconstructed through polynomial interpolation method to supplement the eliminated data; secondly, the smooth IC curve is extracted using the KF algorithm from the repaired data, with ICP and  $U_p$  are taken as the HF; thirdly, mapping relationships between HF and SOH as well as between SOH and constant current charging time are established using the GA-BP neural network, facilitating the estimation of SOH through bringing HF into the trained model; finally, the SOH estimation value is used to determine constant current charging time, update the maximum available capacity, and compute SOC estimation, thereby achieving joint SOC and SOH estimation. Experimental results based on the Oxford and CALCE battery datasets demonstrate that despite BMS suffering from cyber-attacks, the paper's method maintains high accuracy. It proves effective across various battery types, ensuring reliable SOC and SOH estimates. Furthermore, this method considers the correlation among state parameters during battery usage, achieving long-term stable prediction of each parameter within a unified framework. However, this paper does not validate the stability and accuracy of the method in estimating battery state under complex operating conditions. Future research could extend this method to the state estimation of battery packs to assess its applicability in complex operating conditions.

**Acknowledgement:** We are grateful for the valuable comments provided by the anonymous reviewers, which have helped us to improve the content of this paper and make it even better.

**Funding Statement:** This research was funded by the Scientific Research Project of the Education Department of Jilin Province (No. JJKH20230121KJ).

**Author Contributions:** The authors confirm contribution to the paper as follows: study conception and design: Tianqing Yuan; data collection: Na Li; analysis and interpretation of results: Hao Sun; draft manuscript preparation: Sen Tan. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] M. Gholizadeh and A. Yazdizadeh, "Systematic mixed adaptive observer and EKF approach to estimate SOC and SOH of lithium-ion battery," *IET Electr. Syst. Transp.*, vol. 10, no. 2, pp. 135–143, 2020. doi: [10.1049/iet-est.2019.0033](https://doi.org/10.1049/iet-est.2019.0033).
- [2] W. Duo, M. Zhou, and A. Abusorrah, "A survey of cyber attacks on cyber physical systems: Recent advances and challenges," *IEEE/CAA J. Autom. Sin.*, vol. 9, no. 5, pp. 784–800, May 2022. doi: [10.1109/JAS.2022.105548](https://doi.org/10.1109/JAS.2022.105548).
- [3] Y. C. Li and Q. H. Liu, "A comprehensive review study of cyber-attacks and cyber security; emerging trends and recent developments," *Energy Rep.*, vol. 7, no. 8, pp. 8176–8186, Nov. 2021. doi: [10.1016/j.egypr.2021.08.126](https://doi.org/10.1016/j.egypr.2021.08.126).
- [4] S. Tan, P. L. Xie, J. M. Guerrero, and J. C. Vasquez, "Lyapunov-based resilient cooperative control for DC microgrid clusters against false data injection cyber-attacks," *IEEE Trans. Smart Grid*, vol. 15, no. 3, pp. 3208–3222, May 2024. doi: [10.1109/TSG.2023.3332946](https://doi.org/10.1109/TSG.2023.3332946).
- [5] J. Bi, F. Luo, G. Liang, X. Yang, S. He and Z. Y. Dong, "Impact assessment and defense for smart grids with FDIA against AMI," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 2, pp. 578–591, Mar.–Apr. 2023. doi: [10.1109/TNSE.2022.3197682](https://doi.org/10.1109/TNSE.2022.3197682).
- [6] A. J. Gallo, M. S. Turan, F. Boem, and T. Parisini, "A distributed cyber-attack detection scheme with application to DC microgrids," *IEEE Trans. Automat. Contr.*, vol. 65, no. 9, pp. 3800–3815, 2020. doi: [10.1109/TAC.2020.2982577](https://doi.org/10.1109/TAC.2020.2982577).
- [7] M. R. Habibi, H. R. Baghaee, F. Blaabjerg, and T. Dragicevic, "Secure MPC/ANN-based false data injection cyber-attack detection and mitigation in DC microgrids," *IEEE Syst. J.*, vol. 16, no. 1, pp. 1487–1498, 2022. doi: [10.1109/JSYST.2021.3086145](https://doi.org/10.1109/JSYST.2021.3086145).
- [8] H. Xu, G. Pang, Y. Wang, and Y. Wang, "Deep isolation forest for anomaly detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12591–12604, Dec. 2023. doi: [10.1109/TKDE.2023.3270293](https://doi.org/10.1109/TKDE.2023.3270293).
- [9] X. Y. Li, W. J. Zhu, Y. C. Liu, and C. X. Cai, "Improved Ah integration method considering the operation state of battery," *Battery Bimonthly*, vol. 53, no. 5, pp. 509–513, 2023. doi: [10.19535/j.1001-1579.2023.05.009](https://doi.org/10.19535/j.1001-1579.2023.05.009).
- [10] P. Wang, X. Y. Peng, Z. Cheng, and J. A. Zhang, "A multi-time scale joint state estimation method for lithium-ion batteries based on data-driven model fusion," (in Chinese), *Automot. Eng.*, vol. 44, no. 3, pp. 362–371+378, 2022. doi: [10.19562/j.chinasae.qcgc.2022.03.007](https://doi.org/10.19562/j.chinasae.qcgc.2022.03.007).
- [11] X. L. Dong, C. P. Zhang, and J. C. Jiang, "Evaluation of SOC estimation method based on EKF/AEKF under noise interference," *Energy Proc.*, vol. 152, pp. 520–525, 2018. doi: [10.1016/j.egypro.2018.09.204](https://doi.org/10.1016/j.egypro.2018.09.204).

- [12] M. Jiao, D. Q. Wang, and J. L. Qiu, "A GRU-RNN based momentum optimized algorithm for SOC estimation," *J. Power Sources*, vol. 459, no. 4, 2020, Art. no. 228051. doi: [10.1016/j.jpowsour.2020.228051](https://doi.org/10.1016/j.jpowsour.2020.228051).
- [13] P. Wang, Q. R. Gong, J. A. Zhang, and Z. Cheng, "An online state of health prediction method for lithium batteries based on combination of data-driven and empirical model," (in Chinese), *Trans. China Electrotech. Soc.*, vol. 36, no. 24, pp. 1–12, 2021.
- [14] Q. N. Han, F. Jiang, and Z. Cheng, "The state of health estimation framework for lithium-ion batteries based on health feature extraction and construction of mixed model," *J. Electrochem. Soc.*, vol. 168, no. 7, 2021, Art. no. 070509. doi: [10.1149/1945-7111/ac0e4c](https://doi.org/10.1149/1945-7111/ac0e4c).
- [15] J. P. Tian, R. Xiong, and W. X. Shen, "State-of-health estimation based on differential temperature for lithium ion batteries," *IEEE Trans. Power Electron.*, vol. 35, no. 10, pp. 10363–10373, 2020. doi: [10.1109/TPEL.2020.2978493](https://doi.org/10.1109/TPEL.2020.2978493).
- [16] X. H. Yin, Y. C. Song, W. Liu, and D. T. Liu, "Multi-scale state joint estimation for lithium-ion battery," *Chin. J. Sci. Instrum.*, vol. 39, no. 8, pp. 118–126, 2018.
- [17] C. Li, F. Xiao, Y. Fan, X. Tang, and G. Yang, "Joint estimation of the state of charge and the state of health based on deep learning for lithium-ion batteries," *Proc. CSEE*, vol. 41, no. 2, pp. 681–692, 2021.
- [18] C. Birkl, *Oxford Battery Degradation Dataset 1*. Oxford: University of Oxford, 2017.
- [19] C. Chen, J. Wei, and Z. Li, "Remaining useful life prediction for lithium-ion batteries based on a hybrid deep learning model," *Processes*, vol. 11, no. 8, 2023, Art. no. 2333. doi: [10.3390/pr11082333](https://doi.org/10.3390/pr11082333).
- [20] S. Hariri, M. C. Kind, and R. Brunner, "Extended isolation forest," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 4, pp. 1479–1489, 2019. doi: [10.1109/TKDE.2019.2947676](https://doi.org/10.1109/TKDE.2019.2947676).
- [21] J. T. He, Z. B. Wei, and X. L. Bian, "State-of-health estimation of lithium-ion batteries using incremental capacity analysis based on voltage capacity model," *IEEE Trans. Transp. Electrific.*, vol. 6, no. 2, pp. 417–426, 2020. doi: [10.1109/TTE.2020.2994543](https://doi.org/10.1109/TTE.2020.2994543).
- [22] J. A. Zhang, P. Wang, and Z. Cheng, "SOH estimation method for Li-ion battery based on ICA and Box-Cox transform," (in Chinese), *Proc. CSU-EPSSA*, vol. 34, no. 2, pp. 9–15, 2022.
- [23] Z. A. Yu, H. H. Shao, and K. Y. Chen, "SOC estimation of lithium ion battery based on IGWO-BP neural network," *J. Power Supply*, vol. 47, no. 9, pp. 1153–1157, 2023.
- [24] Q. Wu, R. Xu, Q. Yang, and L. Xu, "Lithium battery capacity estimation method based on PCA and GA-BP neural network," *Electron. Meas. Technol.*, vol. 45, no. 6, pp. 66–71, 2022.