Tuesday, 30-05-2023


Mechanical department, IIT-D

ADVISOR

Professor Shaurya Shriyam
Proffessor Amit Gupta

STUDENTS

Devanshi Sawhney
Nayan Asati


# Federated Learning for Estimating SOC & SOH of Fleet of EVs

# 1. INTRODUCTION

- **Electric vehicles are becoming more visible and widely used.**
- **The Indian government is placing strong regulatory emphasis on electric vehicles.**
- **This has created an urgent need to accurately estimate the state-of-health (SOH) of lithium-ion battery packs.**
- **However, there are several challenges in achieving this objective:**
    - **Variability caused by different cell chemistries.**
    - **Variations due to manufacturing and batch processing.**
    - **Variances resulting from storage and usage patterns by customers.**

# SOH

- The state-of-health (SoH of a battery describes the difference between a battery being studied and a fresh battery and considers cell aging.
- It is defined as the ratio of the maximum battery charge to its rated capacity. It is expressed as a percentage as seen below.

$$\mathrm{SoH}/\% = 100\frac{Q_{\mathrm{max}}}{C_{\mathrm{r}}}$$

Where:

- Cr= The rated capacity
- Qmax= The maximum charge that can be stored in the battery.

# SOC

- The state of charge of a battery describes the difference between a fully charged battery and the same battery in use. It is associated with the remaining quantity of electricity available in the cell.
- It is defined as the ratio of the remaining charge in the battery, divided by the maximum charge that can be delivered by the battery. It is expressed as a percentage as below.

$$\mathrm{SoC}/\% = 100\frac{(Q_0 + Q)}{Q_{\max}} = \mathrm{SoC}_0/\% + 100\frac{Q}{Q_{\max}}$$

Where:

- Q0= Initial charge of the battery.
- Q= The quantity of electricity delivered by or supplied to, the battery. It follows the convention of the current: it is negative during the discharge and positive during the charge.
- Qmax= The maximum charge that can be stored in the battery.

## The battery of EVS are affected by various factors such as:

- **Distance travelled:** The more distance an EV covers, the more energy it consumes from the battery to power the electric motor.
- **Operating and ambient temperature:** High temperatures accelerate chemical reactions within the battery, increasing internal resistance and promoting faster degradation.
- **Charging frequency:** As an EV covers more distance, it requires more frequent charging sessions to replenish the battery. Repeated charging cycles can have a small impact on the overall health and lifespan of the battery, affecting its long-term performance.
- **Congestion** such as heavy traffic, impacts the battery of EVs due to increased energy demand per unit of distance traveled, frequent start-stop cycles, limited cooling opportunities, and potential thermal stress, indirectly affecting the battery life.
- **High discharge rates**, including rapid acceleration and aggressive driving, stress the battery, accelerating degradation and reducing its overall lifespan.

# Why State of Health Estimation??

- Electric vehicles are becoming more visible and widely used.
- Accurate state of health (SoH) estimation of an electric vehicle (EV) battery is essential for determining its overall health and remaining useful life.
- SoH estimation helps in predicting the battery's future performance and enables proactive maintenance and replacement strategies, ensuring optimal reliability and cost-effectiveness
- State of health and state of charge estimation contribute to optimizing battery performance and prolonging its lifespan, leading to reduced maintenance costs and enhanced overall customer satisfaction

# Different Techniques for Estimation SOH

## 1. Direct Measurement method

**A. Open Circuit Voltage (OCV) Measurement:**

**Simple method and low computational complexity**

**Limitation: Requires numerous measurements to establish the intrinsic relationship between SoH and OCV.**

**B. Ohmic Resistance Measurement**

**Calculates current battery SoH using the definition of the SoH internal resistance.**

**Limitation: Factors affecting accuracy: operation temperature, accumulation of error, and sensor noise. These factors can introduce substantial deformations to the aging curve, leading to differences between estimated SoH and practical SoH.**

# Different Techniques for Estimation SOH

**Model Based Approaches**

**A.Electrochemical Model (EM):**

**Derives battery performance degradation mechanism based on physical and chemical reactions inside the battery.**

**Involves solving a series of partial differential equations (PDE) for accurate SoH estimation.**

**Limitations: Complex Computational Process: Solving the PDE in the electrochemical model is challenging and computationally intensive. Difficulties in accurately capturing all hidden and intricate nonlinear degradation features hinder the construction of an accurate degradation model**

# Different Techniques for Estimation SOH

## Model Based Approaches

### B. Equivalent Circuit Model (ECM):

Ignores complex physical and chemical processes inside the battery. Simulates the battery's output effect using basic electronic components.

Frequently transforms the SoH estimation problem into parameter estimation using extended Kalman filtering (EKF) or particle filtering (PF) methods.

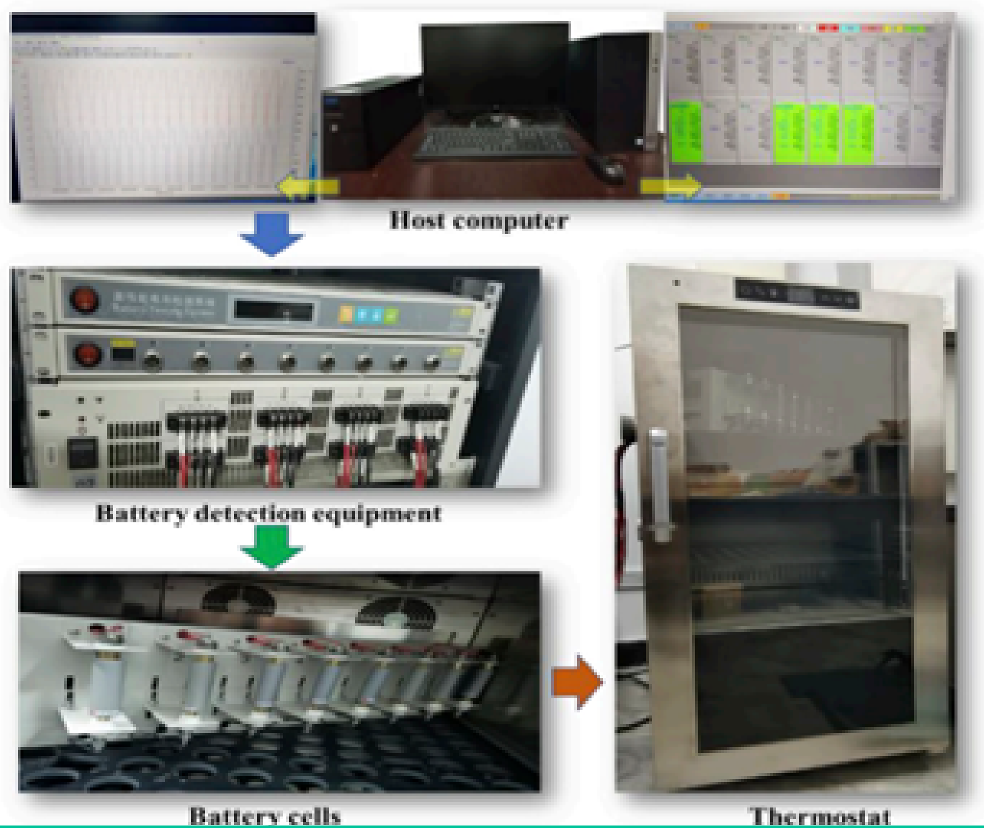
Limitations: The values of the model parameters are definite under fixed operation conditions, limiting the model's applicability. Equivalent circuit models overlook the intricate physical and chemical reactions happening inside the battery

# Data-sets for battery Aging

1. Battery Dataset from IEEE DataSet
2. Battery Degradation Dataset (Fixed Current Profiles & Arbitrary Uses Profiles) - Mendley Dataset
3. Li-ion Battery Aging Datasets - NASA Dataset
4. Lithium-ion battery aging dataset based on electric vehicle real-driving profiles

Link to Dataset comparing Website :-

https://medium.com/batterybits/comparison-of-open-datasets-for-lithium-ion-battery-testing-fd0de091ca2



Battery aging experimental setup

Batteries used:
2 in service batteries N1 & N2
1 retired battery R1

# Research Paper Review 1 -

1. First, the IC (Incremental Capacity) curves are drawn based on the voltage data of the constant current charging phase and denoised by the smoothing spline filter.
2. Then, the Pearson correlation coefficient method is used to select the critical health indicators from the features extracted from the IC curves.
3. Selection of HI:

   HIs selected are:

   N1: vpeak1, yvpeak2, and yvpeak4

   N2: yvpeak2, yvpeak3, and yvpeak4

   R1: vpeak1 and yvpeak1

## SOH Estimation

- **Generate Testing and Training sets**

Extract HIs using the ICA method based on the recorded voltage variation data for each constant current charging phase, and then constitute the feature matrix.

- **Parameter Identification Process**

Optimize the L2 regularization parameter and the shrinkage scale of the enhancement nodes of the BLS network using the PSO algorithm with the fitness function below based on the training set

- **Generate the estimated state of health**

with the established SOH estimation model using the particle swarm optimization broad learning system approach.

# Results

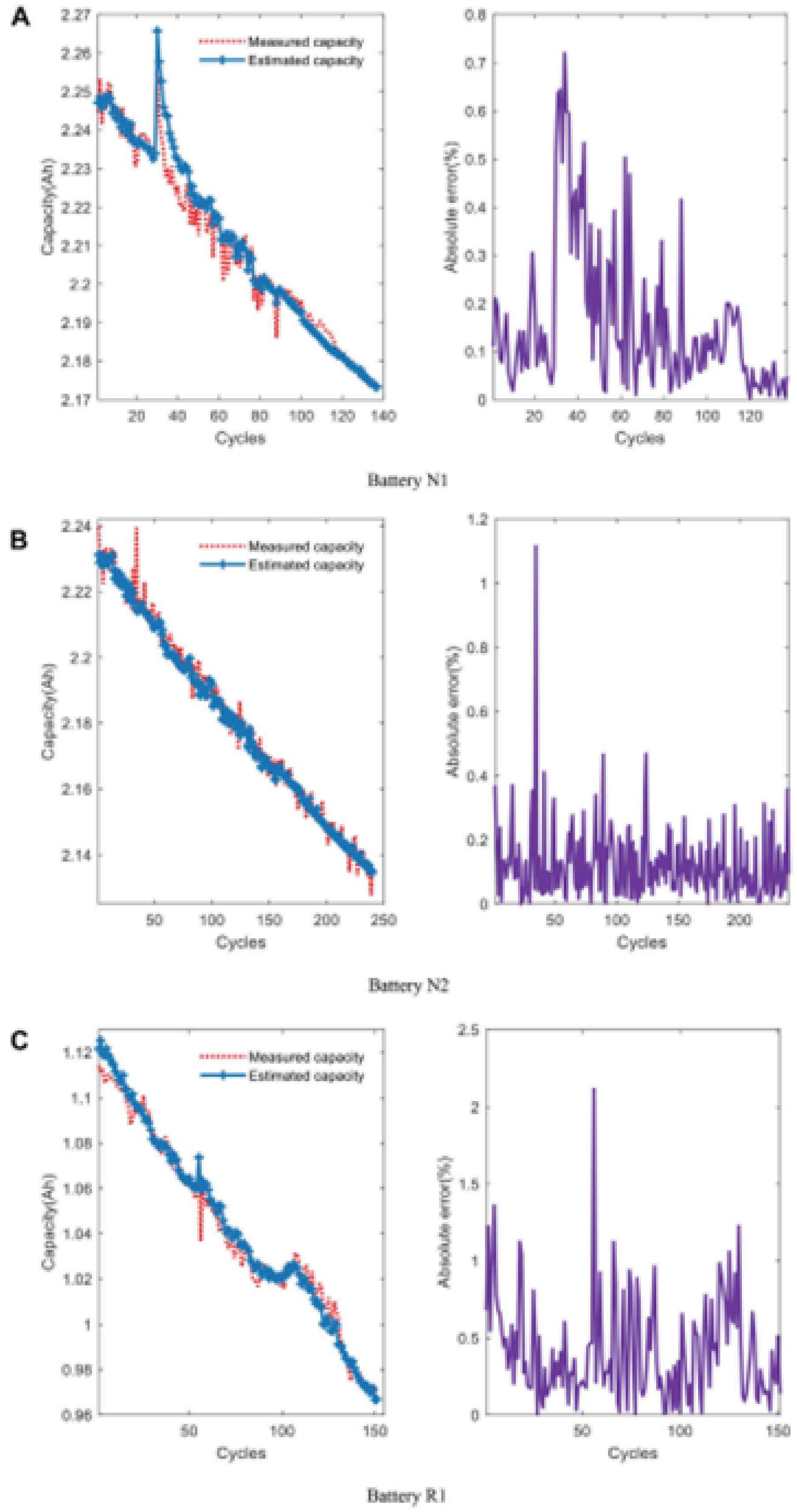



FIGURE 8
The results and absolute error of estimation experiment for batteries N1, N2, and R1.

TABLE 4 The statistical errors of SOH estimation results.

| Battery label | MSE (%) | MAPE (%) | RMSE (%) | $R^2$ | Time (s) |
|---|---|---|---|---|---|
| N1 | 0.33 | 0.1574 | 0.49 | 0.9562 | 1.07 |
| N2 | 0.32 | 0.1241 | 0.37 | 0.9830 | 1.15 |
| R1 | 0.48 | 0.4165 | 0.56 | 0.9824 | 1.26 |

# Research Paper Review 2 -

1. NASA battery dataset contains measurements of battery charge and discharge cycles, including voltage, current, temperature, and capacity.
2. The four most important HFs are selected, which are L1, CT1, CT, and T1.

**Table 1** Parameters and experimental conditions for selected batteries

| Battery number | Discharge current/A | Cut-off voltage/V | Temperature/°C | Rated Capacity/Ah | End-of-life capacity/Ah |
|---|---|---|---|---|---|
| B0005 | 2 | 2.7 | 24 | 2 | 1.40 |
| B0007 | 2 | 2.2 | 24 | 2 | 1.45 |
| B0029 | 4 | 2.0 | 43 | – | 1.65 |
| B0031 | 4 | 2.5 | 43 | – | 1.70 |

# SOH Estimation

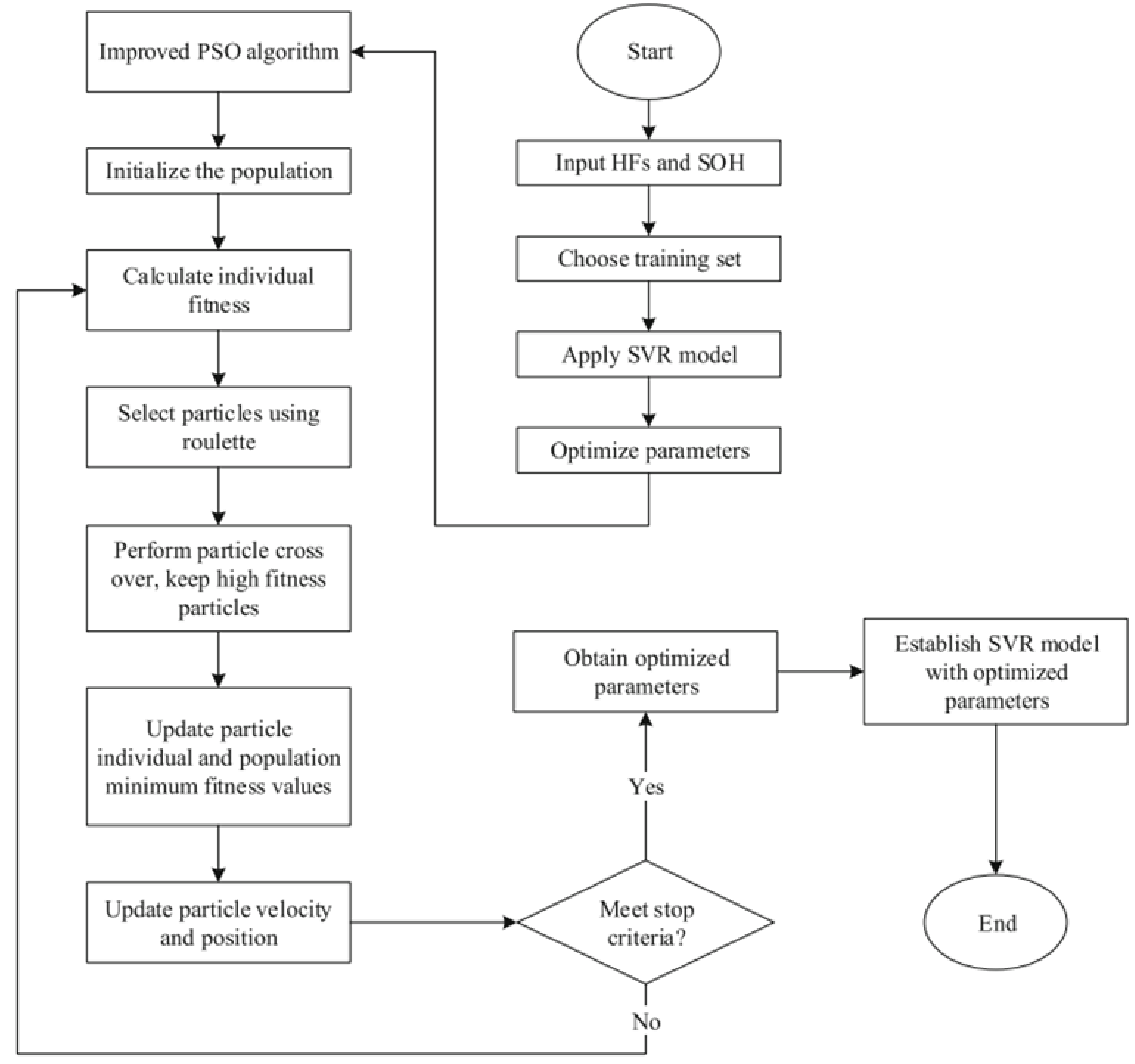



**Fig. 5** Flowchart of SOH estimation method based on GA-PSO-SVR

# Results:

Fig. 7 Comparison between the actual values and estimated values

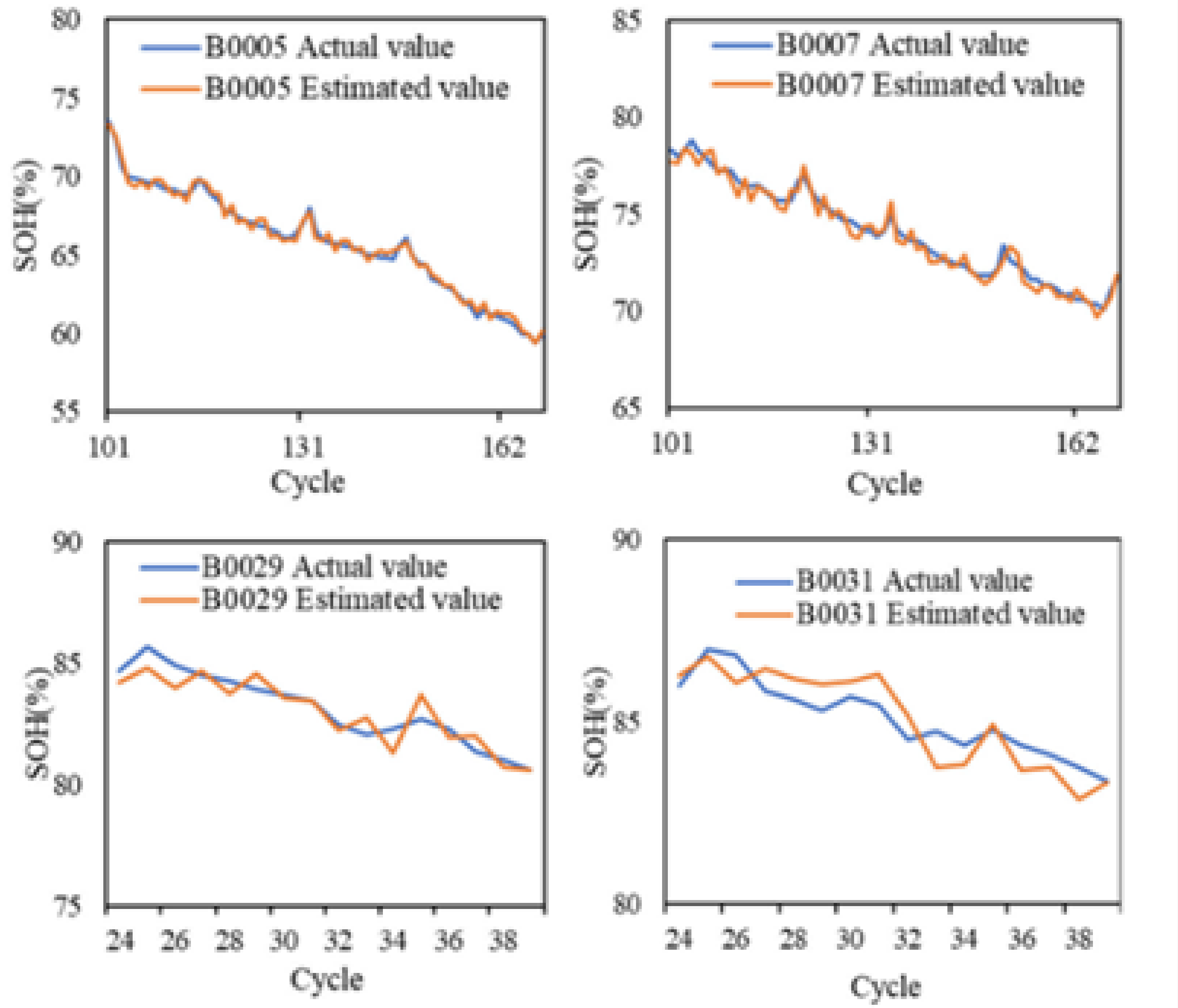



**Table 2** Optimal parameter values and SOH estimation results

| Parameter | B0005 | B0007 | B0029 | B0031 |
|---|---|---|---|---|
| $C$ | 0.4150 | 0.3847 | 0.5420 | 0.5810 |
| $\gamma$ | 1.3950 | 1.7854 | 0.8452 | 0.4984 |
| RMSE (%) | 0.28 | 0.31 | 0.49 | 0.52 |
| MAPE (%) | 0.41 | 0.47 | 0.65 | 0.72 |

# Research Paper Review 3 -

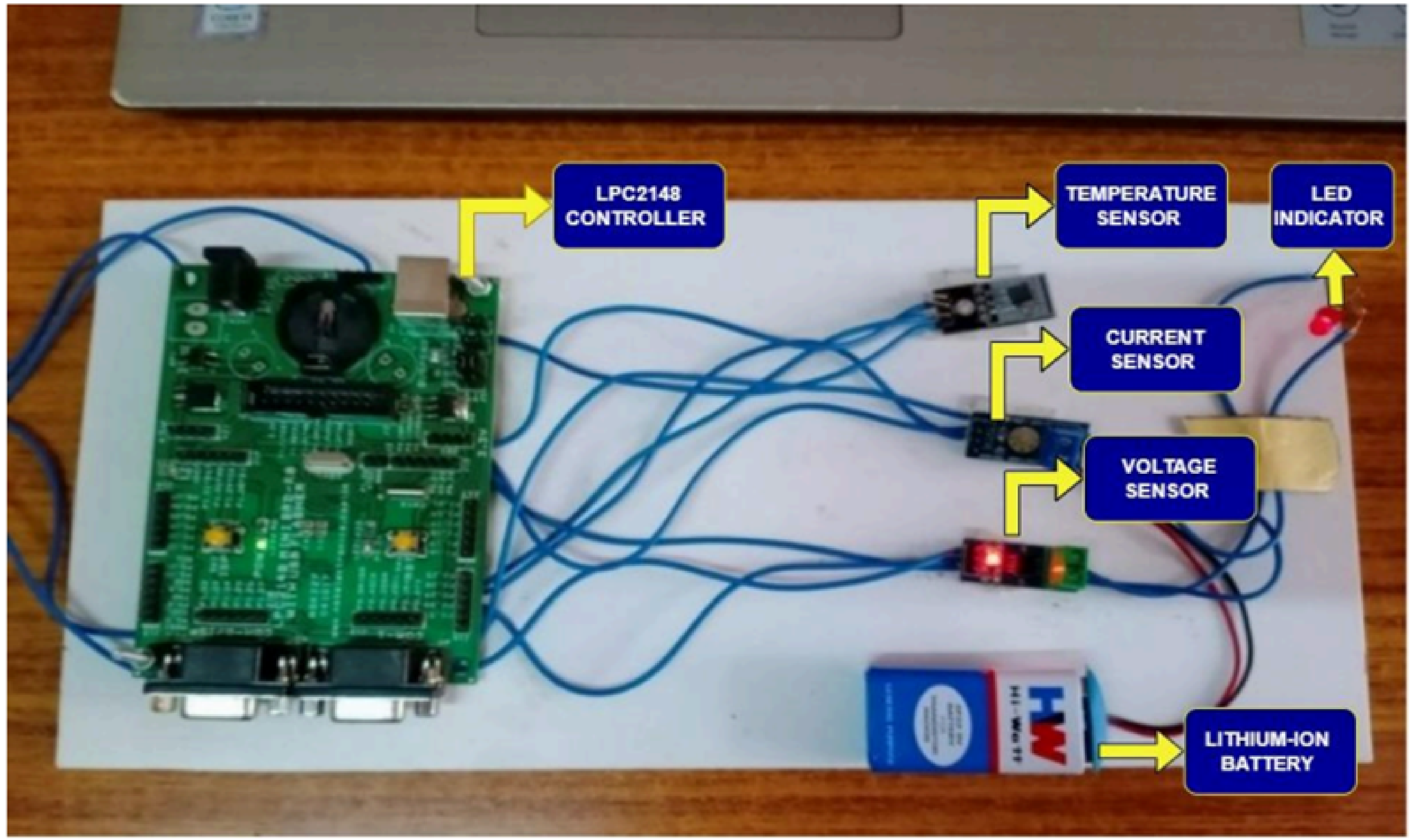



**Fig.3.** Experimental setup of the battery management system

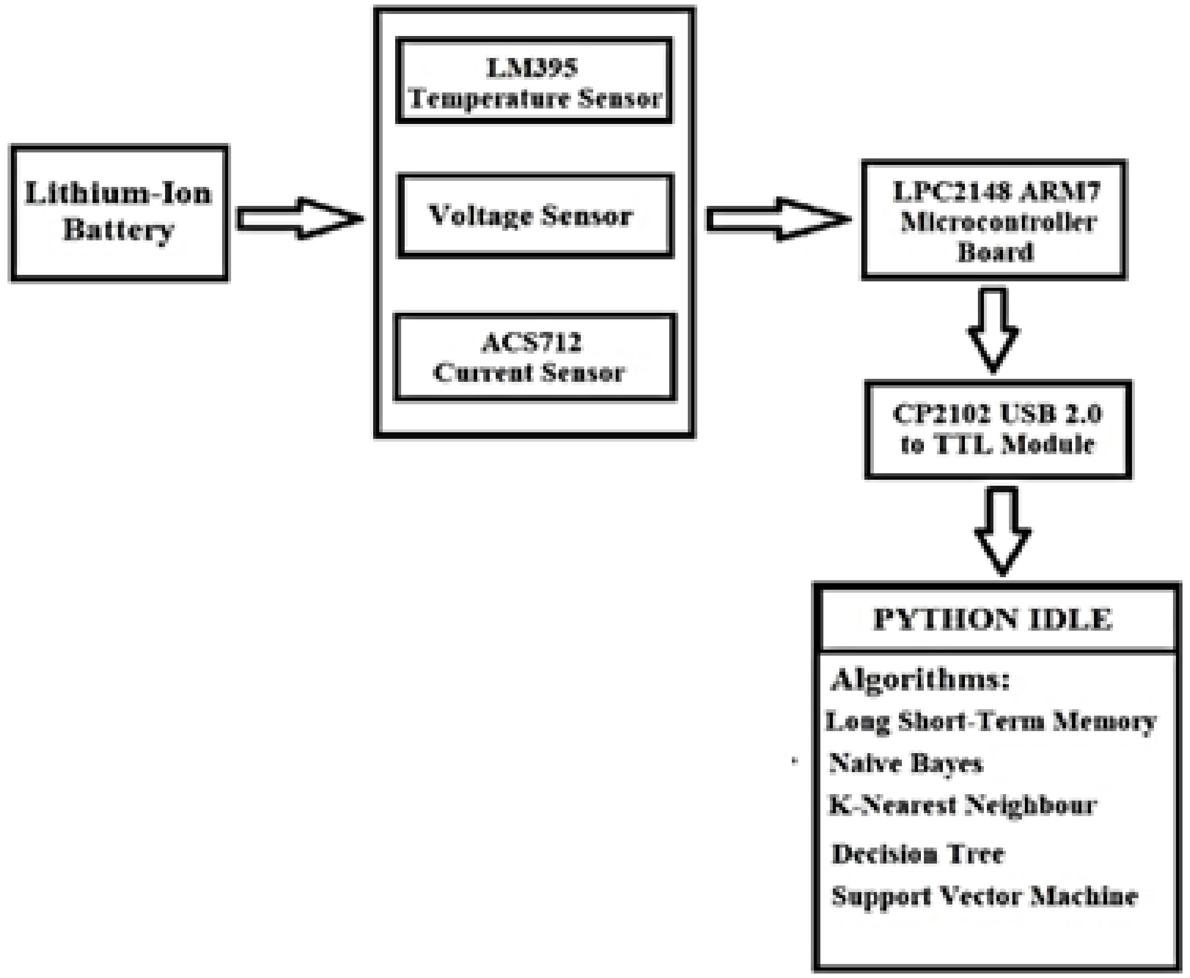



**Fig.2.** Block Diagram of the proposed model

The parameters such as voltage, current and temperature are taken from the sensors connected to the LPC2148 ARM board and the values are given as dataset to the Long Short-Term Memory (LSTM), Decision Tree (DT), K-Nearest Neighbors (KNN), Naïve Bayes (NB) and Support Vector Machine (SVM) Algorithms.

# Results:

| ALGORITHM | ACCURACY | PRECISION | RECALL | F1-SCORE |
|---|---|---|---|---|
| **Naïve Bayes** | **88%** | **95%** | **91%** | **93%** |
| Decision Tree | 72% | 84% | 80% | 82% |
| KNN | 80% | 90% | 86% | 88% |
| SVC | 76% | 79% | 95% | 86% |

**Table.1.** comparison of classification metrices

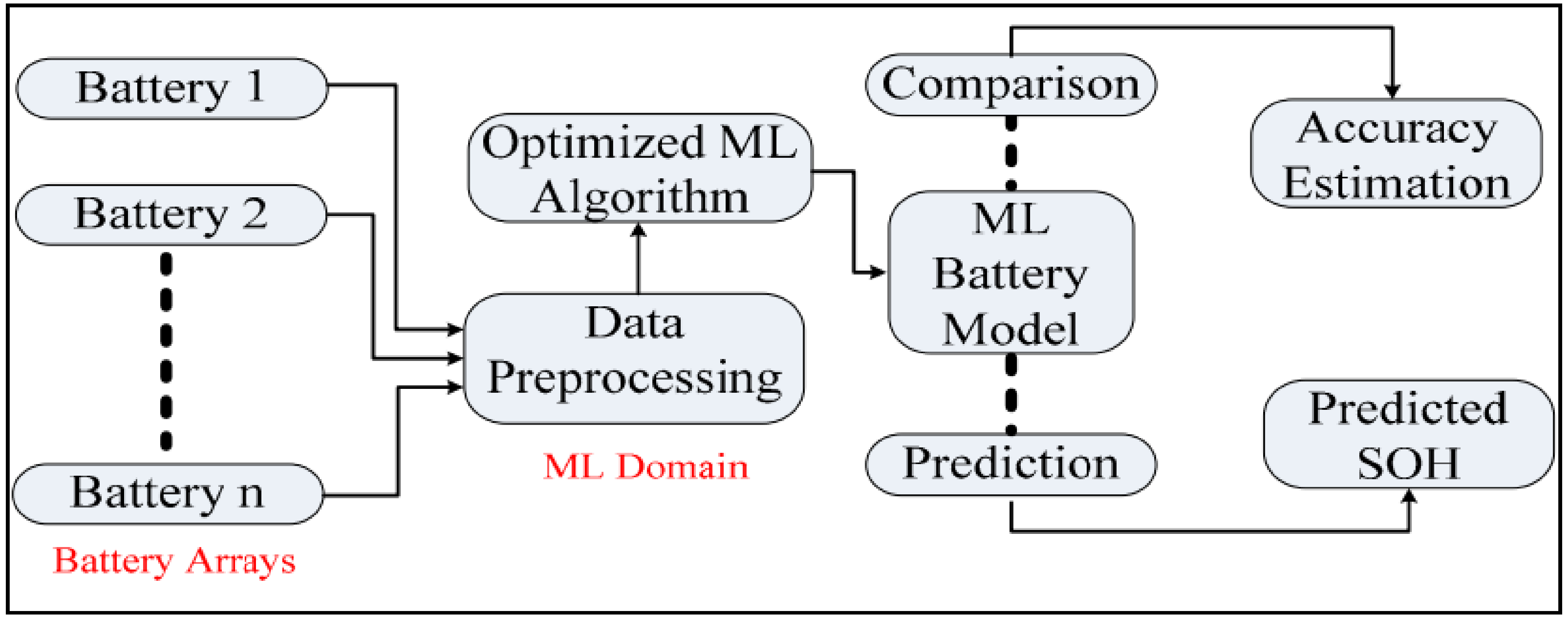
# Research Paper Review 4 -

This technology, however, is still fragile and is impeded by a host of aspects, including protection, cost, recycling, and infrastructure charges. Understanding battery aging mechanisms, as well as safety concerns,is critical for more accurate lifetime forecasts and increased battery efficiency. The most difficult task is to discover aging mechanisms.

The most common method is “coulomb counting,” which involves a simple integration of current over time to estimate SOH [15]. It necessitates periodic calibration, which cannot be accomplished in real time [16]. The aging estimation problem was modeled with Equation (1), with an input ‘u’ (state vector) and an output ‘y’ (voltage), both dependent on variables ‘x’

$$\left\{ \begin{array}{l} x' = Ax + Bu \\ y = Cx \end{array} \right\}$$

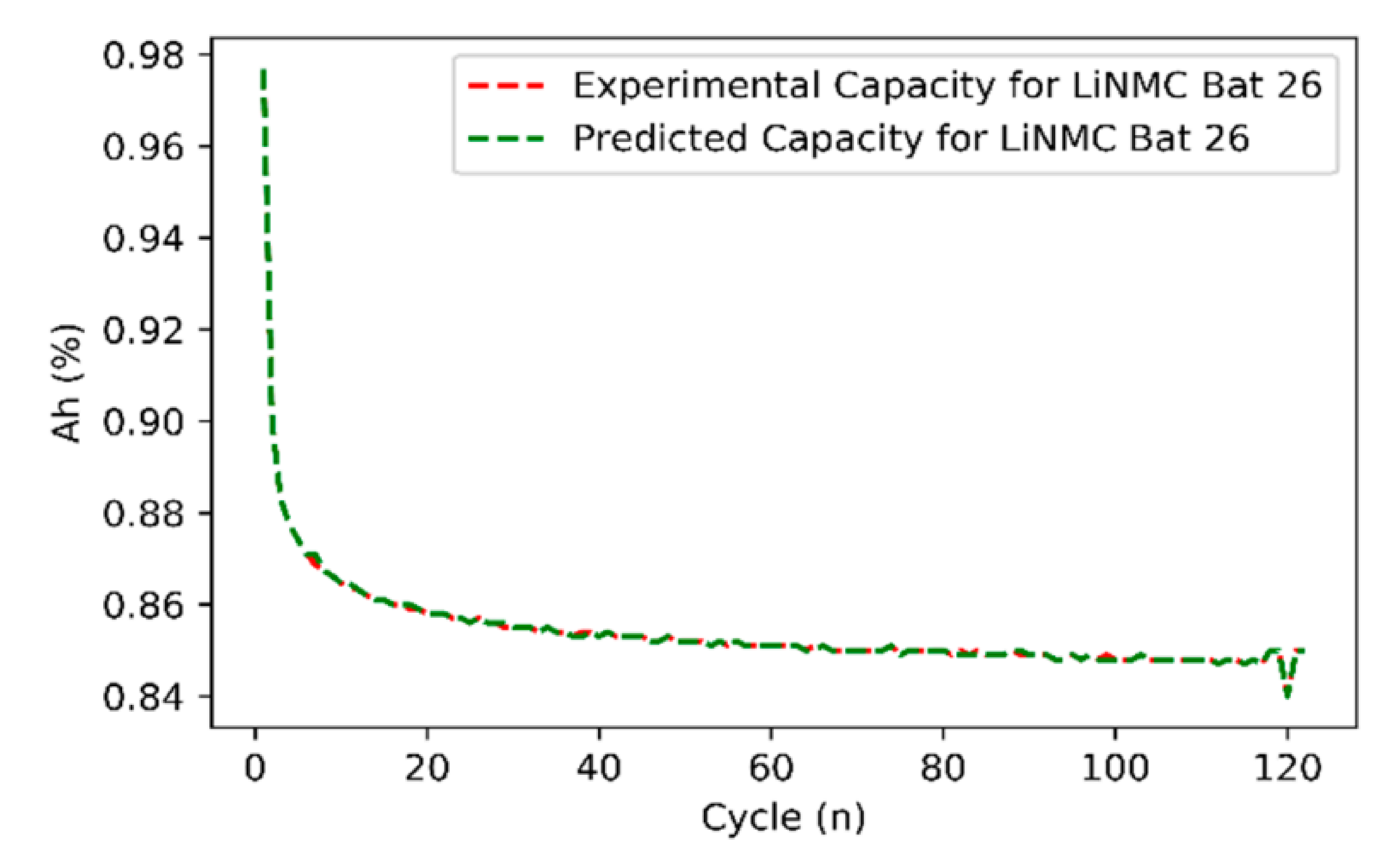
# Research Paper Review 4 -




**Algorithm 1**

**obtain** (battery data)
**calculate** (**SOH** from **Charge Capacity**)
**while** (data in each datasets) **do**
    **divide** (data)
    **invoke** (best ML algorithm)
    **create** (model), **verify** (model)
    **return** (model)
    **print** (results)
  **end while**
**end of execution**

# Research Paper Review 4 -




## Results

# Research Paper Review 5 -

The SOH is a crucial indicator to measure the degree of battery aging, which can show the health state of the battery. The SOH is defined according to the capacity and internal resistance.

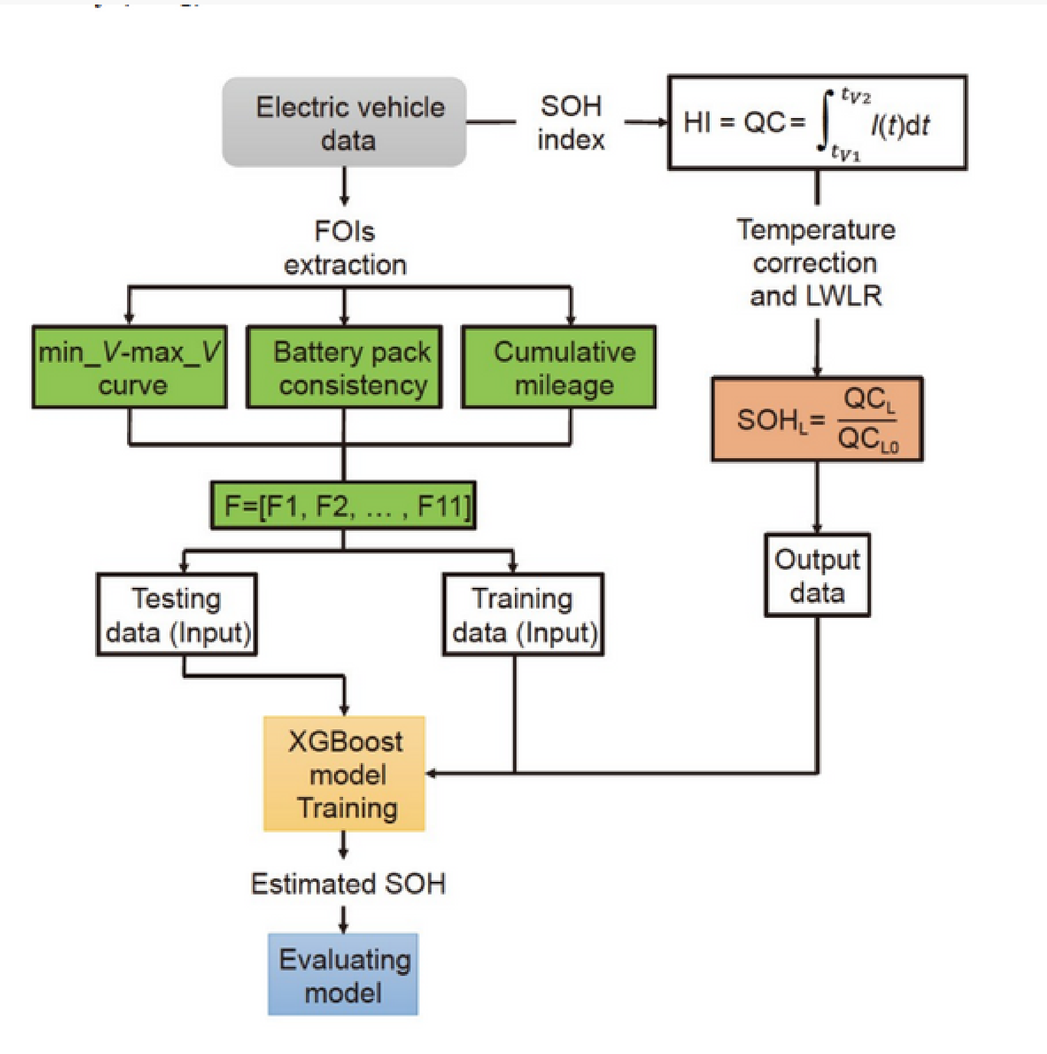
$$\mathrm{SOH} = \frac{Q_{\mathrm{cur}}}{Q_0} \times 100\%,$$

EVs are commonly not wholly discharged in practical applications, because of the driver's charging habits and battery protection. Accordingly, it is not feasible to di rectly compute the battery-aging evaluation index in eq. (1). Therefore, it is necessary to find an HI to indirectly reflect and evaluate battery aging.

# Research Paper Review 5 -

The HI is in line with the battery-degradation trend from a global point of view. This problem can make a data-driven algorithm model difficult to train in the absence of sufficient training data. To address this issue, the charged-capacity fading trend is characterized.

Health index based on LWLR



# Research Paper Review 7 -

- The data used throughout this work is composed of impedance and maximum capacity measurements taken from a set of BMW i3 battery modules.
- Impedance is obtained through EIS and capacity measurements are taken from CC CV charging cycles.

**Table 1.** Characteristics of Li-ion cells used.

| Chemistry | Nominal Capacity | Rated Voltage | Cut-Off Voltages Lower | Cut-Off Voltages Upper |
|---|---|---|---|---|
| NMC | 63 A h | 3.7 V | 3 V | 4.125 V |

- All experiments were performed at a constant temperature of 25 °C.
- Data spanning around 13 months were recorded.

# SOH Estimation

- A fully-connected feed-forward neural network is chosen as an estimator.

$$y = f(\alpha) = f\left(\sum_{i=1}^{n} w_i x_i + b\right) = f\left(\boldsymbol{w}\boldsymbol{x}^T + b\right)$$

- An Adam optimiser is used with mean-squared-error (MSE) as loss metric, along with early stopping and batch training to avoid over-fitting while training.
- The data split between training and validation is 70% and 30% of the samples

# Results

**Table 4.** Test results with impedance measurements at 10% SoC.

| Train | Test | Structure | RMSE | MAE | ME | $\mu_{res}$ | $\sigma_{res}$ |
|---|---|---|---|---|---|---|---|
| C1–C4 0–13 M | C5–C6 0–13 M | FNN (10-10-10, ReLU) | 2.8% | 1.7% | 10% | −1.5% | 2.3% |
| C1–C6 0–10 M | C1–C6 11–13 M | FNN (10-10-10, ReLU) | 2.4% | 1.7% | 5.4% | 1.7% | 1.2% |

**Table 5.** Test results with impedance measurements at 50% SoC.

| Train | Test | Structure | RMSE | MAE | ME | $\mu_{res}$ | $\sigma_{res}$ |
|---|---|---|---|---|---|---|---|
| C1–C4 0–13 M | C5–C6 0–13 M | FNN (10-10-10, ReLU) | 2% | 1.4% | 5.9% | 0% | 2% |
| C1–C6 0–10 M | C1–C6 11–13 M | FNN (10-10-10, ReLU) | 2.1% | 1.6% | 5.8% | 1.6% | 1.5% |

**Table 6.** Test results with impedance measurements at 90% SoC.

| Train | Test | Structure | RMSE | MAE | ME | $\mu_{res}$ | $\sigma_{res}$ |
|---|---|---|---|---|---|---|---|
| C1–C4 0–13 M | C5–C6 0–13 M | FNN (10-10-10, ReLU) | 1.5% | 0.9% | 6.3% | −0.2% | 1.5% |
| C1–C6 0–10 M | C1–C6 11–13 M | FNN (10-10-10, ReLU) | 2.5% | 1.9% | 6% | 1.8% | 1.9% |

**Table 7.** Test results with combined impedance measurements.

| Train | Test | Structure | RMSE | MAE | ME | $\mu_{res}$ | $\sigma_{res}$ |
|---|---|---|---|---|---|---|---|
| C1–C4 0–13 M | C5–C6 0–13 M | FNN (10-10-10, ReLU) | 1.9% | 1.1% | 6.3% | −0.2% | 2.1% |
| C1–C6 0–10 M | C1–C6 11–13 M | FNN (10-10-10, ReLU) | 3.1% | 2.7% | 5% | 2.5% | 1.8% |

# Research Paper Review 8 -

**Table 3**
Summary of ML SOC estimation algorithms.

| Algorithms | Refs. | Data profiles | Input and output | Hyperparameters | Performance metric (at 25 °C) |
|---|---|---|---|---|---|
| BPNN | Hannan et al. (2018) | CALCE dataset (DST, FUDS) | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Hidden neurons = 24<br>- Sigmoid function | -RMSE = 0.91% (FUDS)<br>-MAE =0.59% (FUDS) |
| RBFNN | Chang (2013) | Constant current discharge test | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Hidden neurons = 24 | -MAPE = 1.82% |
| ELM | Hossain Lipu et al. (2019) | BJDST, US06, DST, FUDS | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Hidden neuron s = 328/220/158/147 | -RMSE = 0.76% (BJDST)<br>-MAE = 0.55% (BJDST) |
| WNN | Cui et al. (2018) | NEDC, UDDS, UKBC | - Inputs = [V(t), I(t)]<br>- Output = [SOC(t)] | - Levenberg–Marquardt (L-M) algorithm<br>- Hidden neurons = 10 | -MAE = 0.59% (NEDC) |
| WNN | Xia et al. (2018) | NEDC, EUDC, UDDS, UKBC | - Inputs = [V(t), I(t)]<br>- Output = [SOC(t)] | - Levenberg–Marquardt (L-M) algorithm<br>- Hidden neurons = 10 | -MAE = 0.8% (UDDS) |
| DNN | He et al. (2014) | CALCE dataset (DST, FUDS, US06) | - Inputs = [ I(t), I(t–4), ..., I(t–4k), V(t), V(t–4), ..., V(t–4k), T(t), T(t–4), ..., T(t–4k)]<br>- Output = [SOC(t)] | - Two hidden layers<br>- Hidden neurons = 5 per hidden layer<br>- $k = 30$ | -RMSE = 1.4% (FUDS)<br>-RMSE = 1.5% (US06) |
| DNN | Chemali et al. (2018) | Panasonic NCR18650PF dataset (US06, HWFET) | - Inputs = [V(t), I(t), V_avg(t), I_avg(t)]<br>-Output = [SOC(t)] | - Four hidden layers<br>- Hidden neurons = 50 per hidden layer | - RMSE = 0.78% (US06)<br>- MAE = 0.61% (US06) |

# Research Paper Review 8 –

| | | | | | |
|---|---|---|---|---|---|
| DNN | How et al. (2020) | CALCE dataset (DST, FDUS, US06, BJDST) | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Four hidden layers<br>- Hidden neurons = 64 per hidden layer | - RMSE = 3.68% (DST, FDUS, US06, and BJDST) |
| NARX | Lipu et al. (2018) | FDUS, US06 | - Inputs = [V(t), V(t−1), V(t−2), V(t−3), I(t), I(t−1), I(t−2), I(t−3), T(t), SOC(t−1), SOC(t−2)]<br>- Output = [SOC(t)] | - Hidden neurons = 16<br>- Feedback delays = 2<br>- Input delays = 3 | - RMSE = 0.68% (FUDS)<br>- MAE = 0.48% (FUDS) |
| NARX | Wang et al. (2021b) | CC-CV, DST, UDDS, FUDS | - Inputs = [V(t), V(t−1), V(t−2), I(t), I(t−1), I(t−2), T(t), SOC(t−1), SOC(t−2)]<br>- Output = [SOC(t)] | - Window size: changes with the characteristics of data<br>- Hidden neurons = 12<br>- Feedback delays = 2<br>- Input delays = 2 | - RMSE = 0.48% (UDDS)<br>- MAE = 0.24% (UDDS) |
| LSTM | Ma et al. (2021) | Panasonic NCR18650PF dataset (US06, LA92, UDDS) | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t), SOE(t)] | - Number of LSTM neurons = 256<br>- Time step = 50 | - RMSE = 1.71% (UDDS)<br>- MAE = 1.39% (UDDS) |
| LSTM | Tian et al. (2020a) | CALCE dataset (DST, FDUS, US06) | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Number of LSTM neurons = 32<br>- Time step = 30 | - RMSE = 0.9% (FUDS)<br>- MaxE = 2.7% (FUDS) |
| BiLSTM | Bian et al. (2020) | Panasonic NCR18650PF & CALCE datasets | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Hidden neurons = 64<br>- Numbers of BLSTM layers = 2 | - MAE = 0.84% (FUDS)<br>- MaxE = 3.46% (FUDS) |
| GRU | Li et al. (2019c) | Panasonic NCR18650PF & CALCE & High-rate pulse discharge condition datasets | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Hidden neurons = 1000<br>- Time step = 100 | - MAE = 0.86% (FUDS)<br>- MaxE = 3.13% (FUDS) |

# Research Paper Review 8 -

**Table 3** (*continued*).

| | | | | | |
|---|---|---|---|---|---|
| GRU | Yang et al. (2019a) | DST, FUDS | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Hidden neurons = 150<br>- GRU layer = 1 | -RMSE = 1.05% (FUDS)<br>-MAE = 0.77% (FUDS) |
| GRU | Xiao et al. (2019) | CACLE dataset (CC-CV, DST, FUDS, US06) | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Hidden neurons = 260<br>- Optimizer = Nadam & AdaMax | -RMSE = 0.64% (FUDS)<br>-MAE = 0.49% (FUDS) |
| CNN+LSTM | Song et al. (2019) | Private dataset (DST, US06, FUDS) | - Inputs = [V(t), I(t), T(t), V_avg(t), I_avg(t)]<br>- Output = [SOC(t)] | - Neurons in LSTM layer = 300<br>- Filters in convolutional layer = 6<br>- Neurons in fully connected layer = 80; | -RMSE = 1.31%<br>-MAE = 0.92%<br>(DST, US06, and FUDS) |
| CNN+GRU | Huang et al. (2019) | DST, FUDS | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Two GRU layers<br>- Hidden neurons =150 & 80 per GRU layers<br>- Filters in convolutional layer = 8 | -RMSE = 1.54% (FUDS)<br>-MAE = 1.26% (FUDS) |
| CNN | Bhattacharjee et al. (2021) | Panasonic NCR18650PF and LG 18650HG2 datasets (UDDS, US06, HWFET, LA92) | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Number of convolutional layers = 2<br>- Number of filters in conv layer = 8 | -MAE = 0.8% (US06)<br>-MaxE = 4.72% (US06) |
| SVM | Alvarez Anton et al. (2013) | DST, CC-CV | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Number of support vectors = 903<br>- Penalty factor $\gamma = 0.125$<br>- Kernel width $\sigma = 1$ | - RMSE = 0.4% (DST)<br>- MaxE < 4% (DST) |
| SVM | Li et al. (2020c) | CC-CV | - Inputs = [V(t), I(t), T(t)]<br>- Output = [SOC(t)] | - Penalty factor $\gamma = 1.41$<br>- Kernel width $\sigma = 1.18$ | - AE = 1.2% |

# Research Paper Review 8 -

**Table 6**
Summary of ML SOH estimation algorithms.

| Algorithms | Refs. | Data profiles (Public/ Private) | Number of input features | Correlation analysis | Hyperparameters | Performance metric |
|---|---|---|---|---|---|---|
| BPNN | won You et al. (2016) | Private dataset (UDDS, US06, and NYCC) | 80 (groups 1&3) | None | - Hidden neurons = 80 | -RMSE: 0.66% |
| BPNN | Wu et al. (2016) | Private dataset (CC-CV test) | 11 (group 1) | None | - Hidden neurons = 40<br>- Hyperbolic tangent sigmoid function | -MAE = 29.42 (RUL estimation) |
| RBFNN | Mao et al. (2021a) | NASA dataset (5#, 6#, and 7#) | 3 (group 3) | Pearson correlation analysis | - Single hidden layer | -RMSE = 0.7%<br>-MAE = 0.53% |
| RBFNN | Lin et al. (2021) | NASA dataset (5#, 6#, and 7#) and CACLE dataset (35#, 36#, 37#, and 38#) | 3 (group 2) | Pearson correlation analysis | - Structural parameters are adaptively modulated by BM modeling and PF predictor | -RMSE = 0.7%<br>-MAE = 0.61% |
| ELM | Tian and Qin (2021) | NASA dataset | 3 (group 3) | Pearson correlation analysis | - Hidden neurons = 10<br>- Data block length = 50 | -MAE = 0.64% |
| ELM | Pan et al. (2018) | Private dataset (CC discharge test, NEDC, and pulsed discharge test) | 2 (group 2) | Pearson correlation analysis & Spearman correlation analysis | - Hidden neurons = 20<br>- Sigmoid function | -RMSE = 1.09%<br>-MAE = 1.72% |
| DNN | Zhang et al. (2019a) | NASA dataset (5#, 6#, 7#, and 18#) | 2 (group 3) | Spearman correlation analysis | - Two hidden layers<br>- Hidden neurons= 6 & 5 | -SOH error < 3%<br>-RMSE = 5.41 (RUL estimation) |
| DNN | Xia and Qahouq (2021) | Private dataset (CC-CV test) | 5 (group 3) | Grey relational analysis | - Two hidden layers<br>- Hidden neurons = 128 per layer | -RMSE = 1.576%<br>-MAPE = 5.022% |
| NARX | Khaleghi et al. (2021) | Private dataset | 5 (group 1) | None | - Hidden neurons = 10<br>- Feedback delays = 2 | -RMSE = 0.26% |

# Research Paper Review 8 -

| | | | | | | |
|---|---|---|---|---|---|---|
| NARX | Cui et al. (2021) | NASA dataset (5#, 6#, 7#, and 18#) | 8 (group 3) | None | - Hidden neurons = 50<br>- Feedback delays = 8 | -MAE = 0.72%<br>-MaxE = 4.69% |
| LSTM | Kaur et al. (2021) | Private dataset (CC charge and discharge test) | 3 (group 1) | None | - One hidden layer<br>- Hidden neurons = 50<br>- Sampling rate = 10 | -RMSE = 0.042<br>-MAE = 0.0216<br>(Capacity estimation) |
| LSTM | Kim et al. (2021) | CALCE, Cavendish Laboratory, and NASA datasets | 1 (group 2) | None | - Sliding windows = 100<br>- Dropout masks = 10 000 | -MAPE = 1.95%<br>(Capacity estimation) |
| GRU | Ungurean et al. (2020) | NASA dataset | 5 (group 2) | None | - Two hidden layers<br>- Hidden neurons = 50 per layer | -AE = 2.91% |
| GRU | Fan et al. (2020) | NASA and Oxford datasets | 3 (group 1) | None | - Hidden neurons = 256 (GRU)<br>- Convolution number = 64<br>- Size of each convolution layer = 32 × 1 | -MAE = 1.03%<br>-MaxE = 4.11% |

# Research Paper Review 8 -

**Table 6** (*continued*).

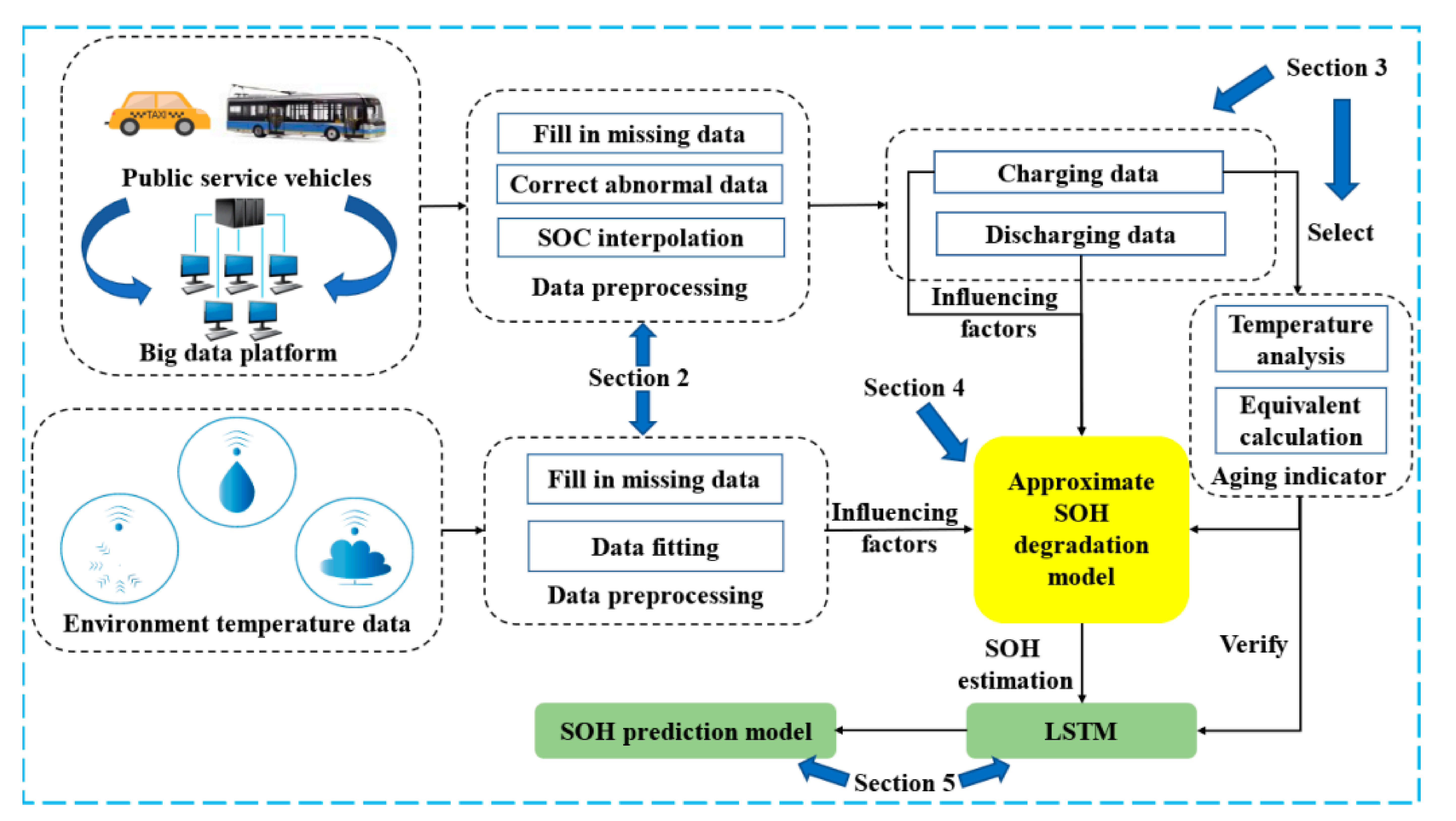
| | | | | | | |
|---|---|---|---|---|---|---|
| CNN | Shen et al. (2019) | NASA and private datasets | 3 (group 1) | None | - 5 convolution stages<br>- 3 fully connected stages<br>- ReLU activation function<br>- Optimizer: Stochastic gradient descent (SGD) | -RMSE = 1.477%<br>-MaxE = 9.479%<br>(Capacity estimation) |
| CNN | Zhou et al. (2020) | NASA dataset (5#, 6#, and 18#) and CALCE dataset | 1 (group 1) | None | – Number of convolution kernels = 256<br>- Size of the kernel = 3 × 1; | -RMSE = 1.1%<br>-MAE = 0.9% |
| SVM | Liu et al. (2020) | NASA dataset (5#, 6#, and 7#) and Private dataset (NCR18650PF) | 3 (group 3) | Grey relational analysis | –Mutation factor of the DE algorithms = 0.6 | -MAPE = 0.23% |
| SVM | Li et al. (2020b) | Private dataset (CC-CV test) | 3 (group 3) | Pearson correlation analysis | - Iteration = 100<br>- Standard deviation = 1e−4 | -RMSE = 0.94%<br>-MAE = 0.769% |
| SVM | Yang et al. (2018) | Private dataset (CC-CV test, HPPC test) | 4 (group 2) | none | - Number of particles in PSO = 30 | -RMSE = 1.92% |
| GPR | Li et al. (2020a) | NASA dataset (5#, 6#, 7#, and 18#) | 11 (group 3) | Pearson correlation analysis | - SE kernel function | -RMSE = 0.78%<br>-MAE = 0.4% |
| GPR | Jia et al. (2020) | NASA dataset (5#, 6#, 7#, and 18#) | 3 (group 3) | Grey relational analysis | - SE kernel function | -RMSE = 0.37% |

# Research Paper Review 9 -

- The vehicle data were derived from the New Energy Vehicle National Supervision Platform, and the operating data were collected from May 2019 to September 2020.
- The driving range was 4799 km~68,644.9 km.
- The vehicle studied in this paper is a pure electric bus (North BFC6109GBEV5 model), and the power battery pack was composed of 372 battery cells which had a rated capacity of 404 Ah and a rated pack voltage of 598.92 V.
- The environmental temperature data came from Airwise (hz.zc12369.com, accessed on 11 July 2021), which is a visual display platform of atmospheric environmental data.

$$\mathrm{SOC_{AI}}(k) = \mathrm{SOC_1} + \frac{i(k) * dt(k)}{\sum_{j=1}^{n} i(j) * dt(j)}$$

# SOH Estimation and prediction based on real world data



# Results

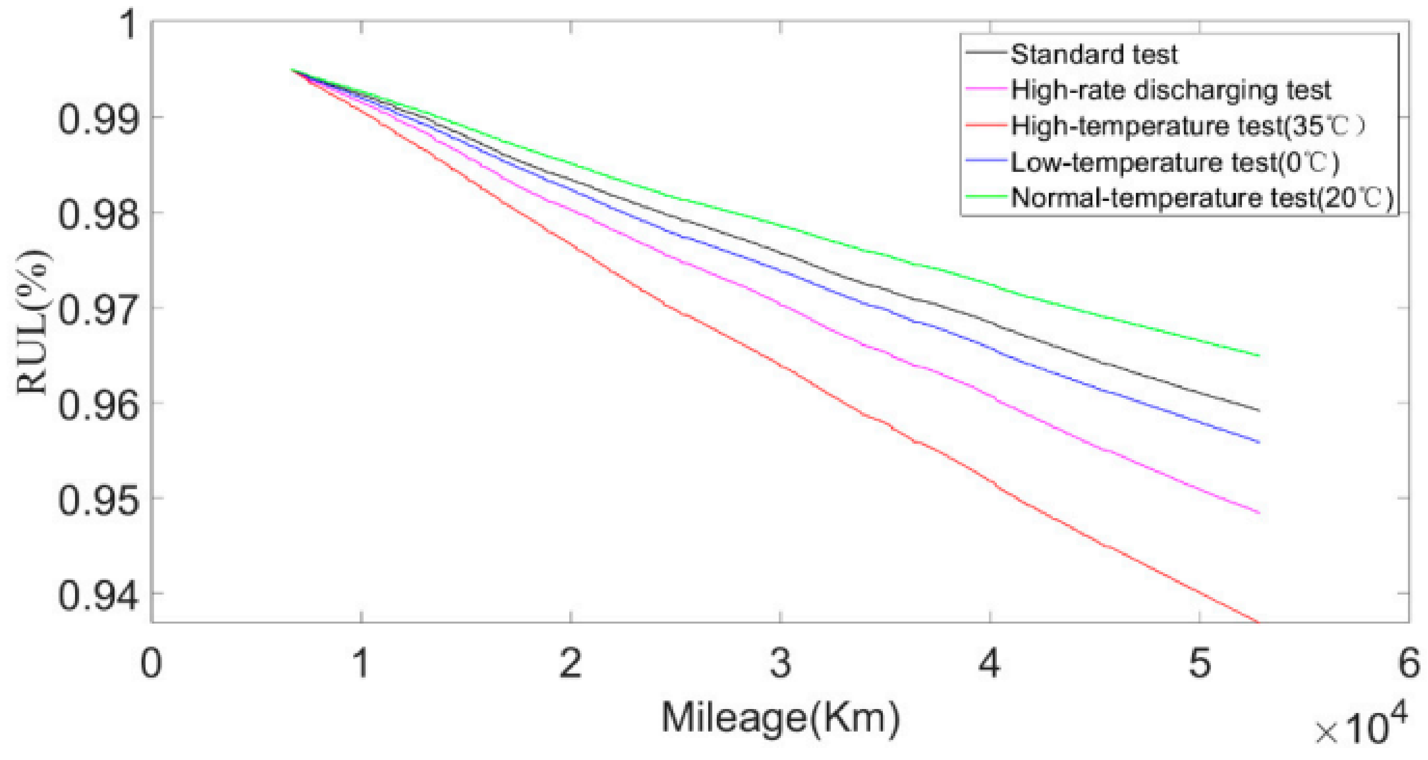


**Figure 25.** RUL calculation results of battery under different test environments.

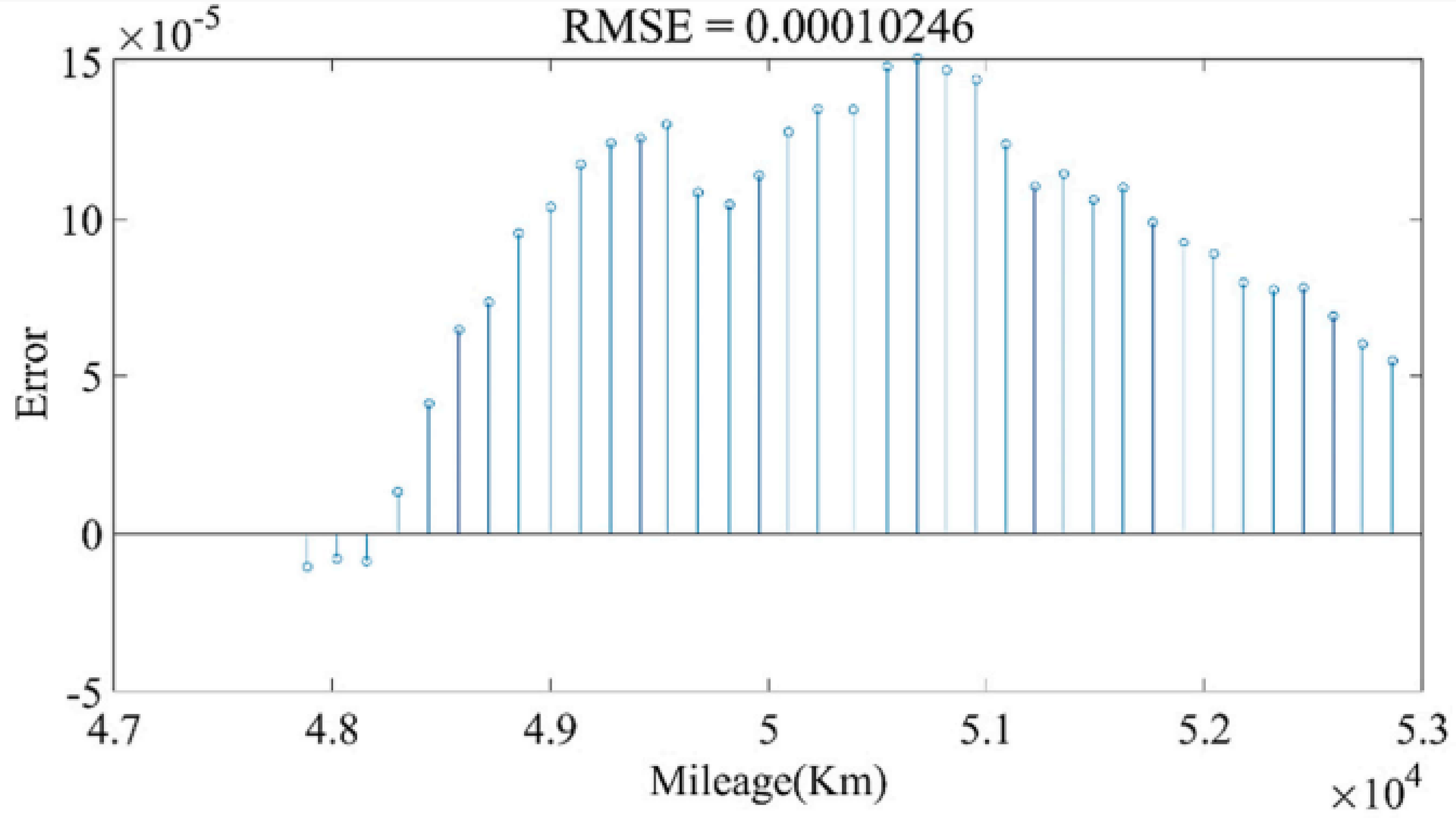


**Figure 28.** The RMSE of LSTM.

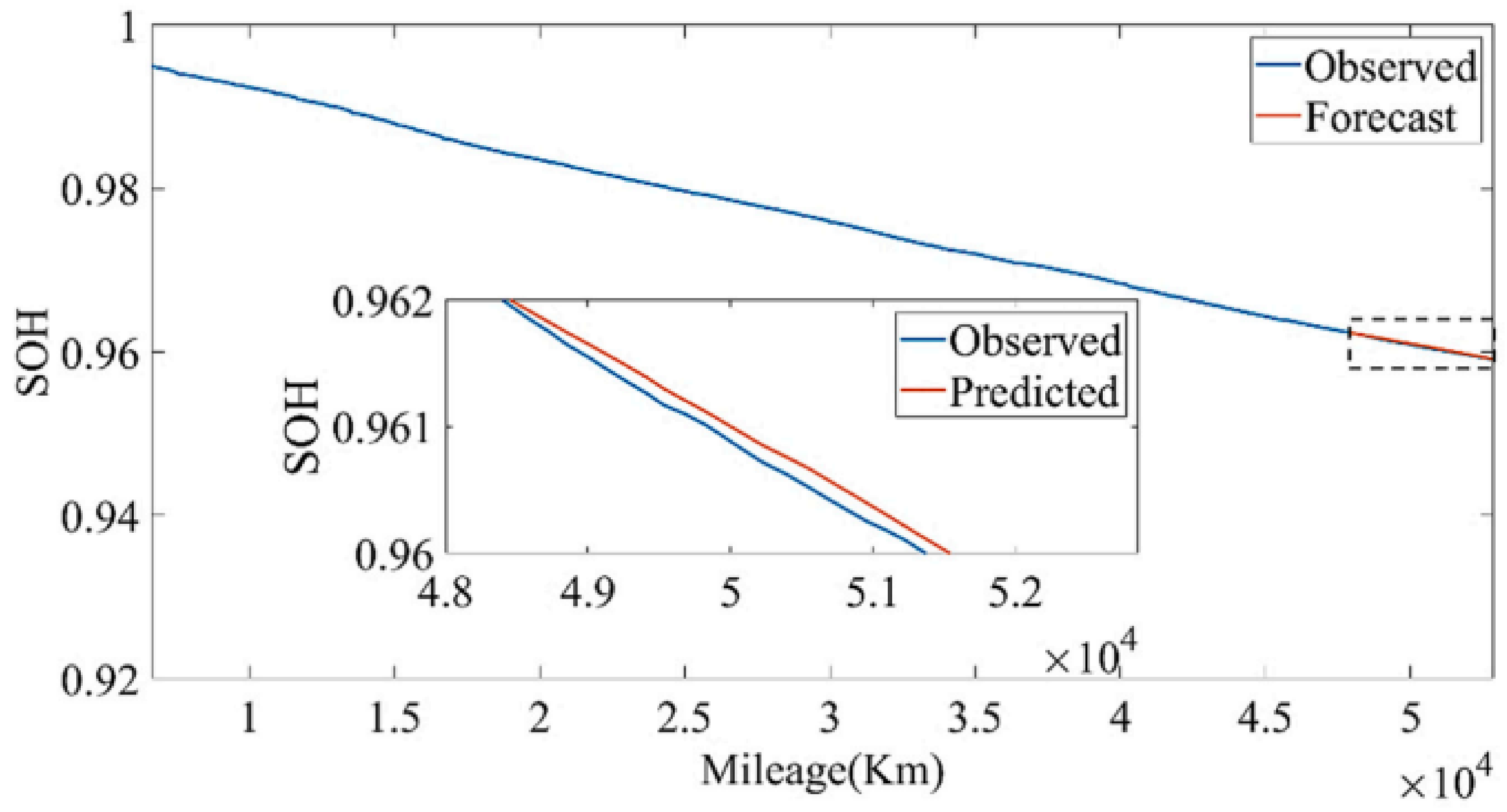


**Figure 27.** The SOH estimation result based on LSTM.

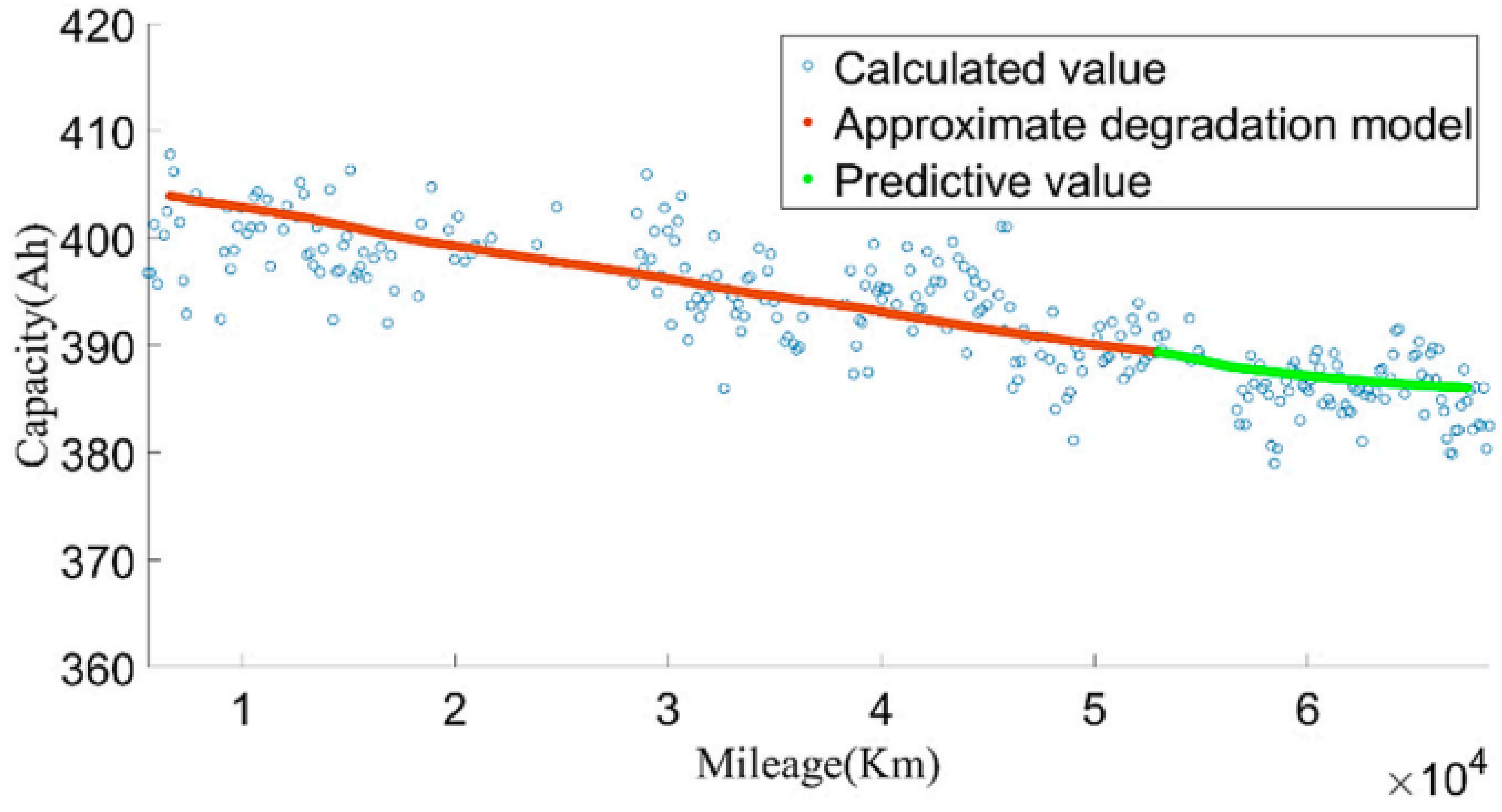


**Figure 29.** The SOH prediction results based on LSTM.

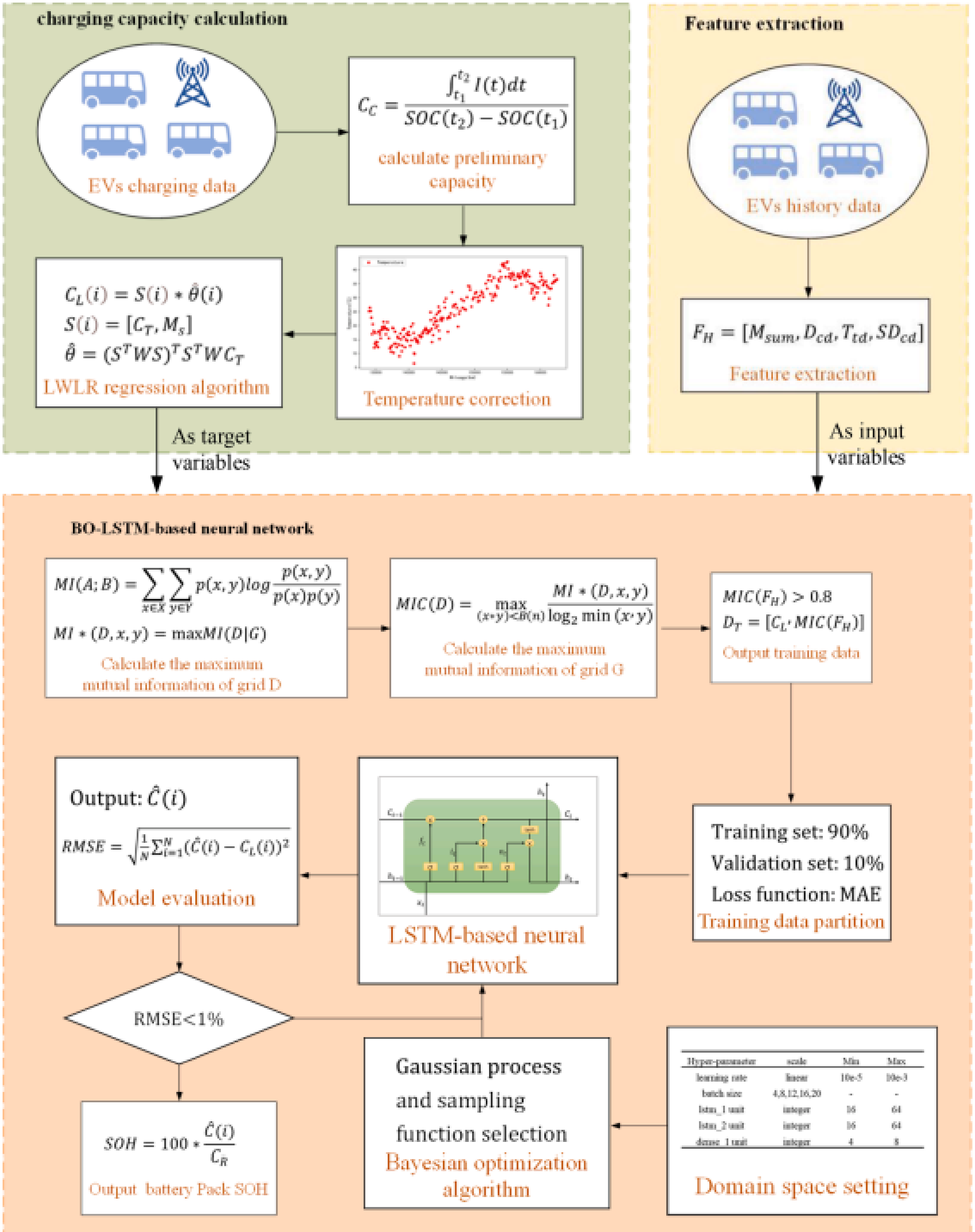
# Research Paper Review 10 -

- The charging capacity is calculated by historical charging data, and locally weighted linear regression (LWLR) algorithm is used to qualitatively characterize the capacity decline trend.
- The health features are extracted from historical operating data, maximal information coefficient (MIC) algorithm is used to measure the correlation between health features and capacity.
- Then, long and short-term memory (LSTM)-based neural network will further learn the nonlinear degradation relationship between capacity and health features.
- Bayesian optimization algorithm is used to ensure the generalization of the model when different electric vehicles produce different user behaviors.
- The estimation method is validated by the 300 days historical dataset from 100 vehicles with different driving behavior

Basic information of real vehicle data.

| Data sources | Number of EVs | Sampling days | Rated capacity of battery pack (Ah) |
|---|---|---|---|
| A | 100 | 300 | 280 |

# Flow Chart–
# SOH Estimation



# Results

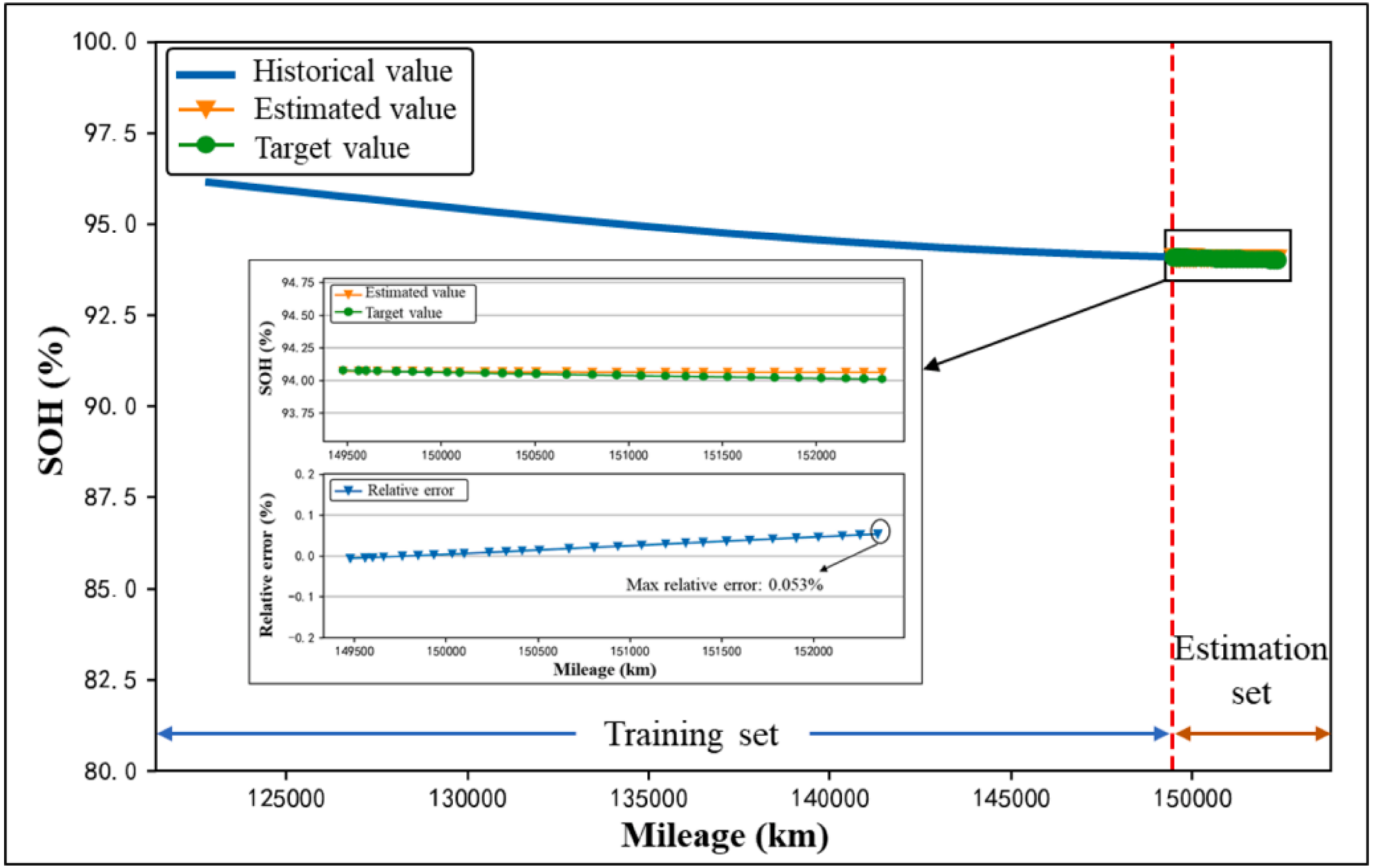



Fig. 6. Estimation result for EV in data source A.

# 4. FEDERATED LEARNING OVERVIEW

Federated learning offers advantages over traditional machine learning models when it comes to estimating the State of Health (SOH) of lithium-ion batteries in electric vehicles (EVs). Here are some key points highlighting the benefits of federated learning for this purpose:

- Privacy: Federated learning allows training a model using data distributed across multiple EVs without the need to share the raw data. This preserves the privacy of individual EV owners and their battery-related information.
- Data Diversity: By leveraging data from various EVs, federated learning can capture a more diverse range of usage patterns, driving conditions, and battery characteristics. This improves the robustness and generalization capabilities of the SOH estimation model.
- Real-time Updates: Federated learning enables continuous model improvement as new data becomes available from EVs. The model can be updated and refined without requiring the data to be sent to a central server, ensuring up-to-date and accurate SOH estimations.

# 4. FEDERATED LEARNING OVERVIEW

- Lower Communication Overhead: Instead of transmitting large amounts of raw data, federated learning only requires exchanging model updates, resulting in reduced communication bandwidth and energy consumption.
- Decentralized Computation: The computation and training of the SOH estimation model can be performed locally on each EV, reducing the dependence on centralized computational resources and allowing for faster and more efficient model training.
- Scalability: Federated learning is well-suited for large-scale deployment, as it can handle a large number of participating EVs simultaneously, enabling the creation of a comprehensive and diverse training dataset.
- Increased Reliability: With federated learning, the SOH estimation model becomes more resilient to failures or disconnections in individual EVs, as the model's parameters are collectively learned across multiple devices.


Thank You !!