

Offline and Online Blended Machine Learning for Lithium-Ion Battery Health State Estimation

Chengqi She, Yang Li, *Member, IEEE*, Changfu Zou, *Member, IEEE*, Torsten Wik, *Member, IEEE*, Zhenpo Wang, and Fengchun Sun

Abstract—This article proposes an adaptive state of health (SOH) estimation method for lithium-ion batteries using machine learning. Practical problems with feature extraction, cell inconsistency, and online implementability are specifically solved using a proposed individualized estimation scheme blending offline model migration with online ensemble learning. First, based on the data of pseudo-open-circuit voltage measured over the battery lifespan, a systematic comparison of different incremental capacity features is conducted to identify a suitable SOH indicator. Next, a pool of candidate models, composed of slope-bias correction (SBC) and radial basis function neural networks (RBFNNs), are trained offline. For online operation, the prediction errors due to cell inconsistency in the target new cell are then mitigated by a proposed modified random forest regression (mRFR) based ensemble learning process with high adaptability. The results show that compared to prevailing methods, the proposed SBC-RBFNN-mRFR-based scheme can achieve considerably improved SOH estimation accuracy (15%) with only a small amount of early-age data and online measurements are needed for practical operation. Furthermore, the applicability of the proposed SBC-RBFNN-mRFR algorithms to real-world operation is validated using measured data from electric vehicles, and it is shown that a 38% improvement in estimation accuracy can be achieved.

Index Terms—Lithium-ion batteries, state of health estimation, incremental capacity analysis, modified random forest regression, online machine learning

I. INTRODUCTION

LITHIUM-ION (Li-ion) batteries have been reckoned as the backbone of electric vehicles (EVs) and key components of modern grid systems due to their salient merits of high energy and power densities, low self-discharge rate, and ever-declining costs in recent years [1]. However, the energy storage capacity and power capability of Li-ion batteries can gradually reduce caused by various aging mechanisms, leading to limited service life and degraded system performance over

time [2]. To ensure the safe, reliable, and efficient use of Li-ion batteries, the indicator of battery health, namely the state of health (SOH), must be precisely monitored and predicted, which forms a fundamental functionality of Li-ion battery management systems (BMSs) [3].

The SOH is a figure of merit defined by comparing an aged battery parameter to its pristine value at the beginning of life, commonly using battery capacity or internal resistance [4]. Many battery SOH estimation algorithms have been developed in the literature, and these algorithms can be generally classified into physics-based methods [5], [6], empirical methods [7], [8], and data-driven methods [9], [10]. Physics-based methods describe battery electrochemical and thermal dynamics that closely relate to battery aging based on partial differential equations (PDEs) for SOH estimation, but it requires high computation for online implementation. Empirical-based methods depict battery dynamics by employing phenomenological equivalent circuit models and monitor the capacity and/or internal resistance for battery SOH estimation, or using static functions to characterize the degradation trends. However, the accuracy of these methods can only be guaranteed under the similar operating conditions under which the models are developed. The weak extrapolability of these empirical models leads to the requirement of extensive experimental tests in order to cover full operating ranges for practical use. In the past five years, the data-driven methods have received rapidly growing research attention, because their model-free and easy-to-implement natures are in favor of real-world applications in view of recent advances in big data and artificial intelligence [11]–[13]. Among many data-driven techniques, incremental capacity analysis (ICA) is one of the most extensively investigated methods. In ICA, the incremental capacity (IC) curves are generated based on the capacity–voltage relationship obtained from constant-current charging/discharging, i.e.,

$$IC_t = \frac{dQ_t}{dV_t} = \frac{\Delta Q_t}{\Delta V_t} = \frac{Q_k - Q_{k-1}}{V_k - V_{k-1}} \quad (1)$$

where Q and V denote the measured battery capacity and voltage, respectively, t is time, and subscripts $k-1$ and k represent two consecutive discrete time instants. The movement of the IC curve is closely related to the phase transitions and phase equilibria during the lithiation and de-lithiation processes inside the Li-ion cells [14]. Thus, features of interest (FOIs) can be reasonably extracted from the IC curves to reveal the hidden relationships between the direct measurements to SOH [15]. The extracted FOIs are used next as the input of a

This work was supported by the China Scholarship Council (Contract 201906030171), the Ministry of Science and Technology of the People's Republic of China (Grant No. 2019YFE0107900), and the Swedish Energy Agency within the framework of Vehicle Strategic Research and Innovation Program (Grant No. 50187-1). Chengqi She and Yang Li contributed equally to this work. (Corresponding author: Changfu Zou.)

Chengqi She is with the National Engineering Laboratory for Electric Vehicles, Beijing Institute of Technology, Beijing, 100081, China, and also with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden (e-mail: shechengqi@bit.edu.cn).

Yang Li, Changfu Zou, and Torsten Wik are with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden (e-mail: yangli@ieee.org; tw@chalmers.se; changfu.zou@chalmers.se).

Zhenpo Wang and Fengchun Sun are with the National Engineering Laboratory for Electric Vehicles, Beijing Institute of Technology, Beijing, 100081, China (e-mail: wangzhenpo@bit.edu.cn; sunfch@bit.edu.cn).

predictive model for developing a SOH estimation algorithm.

Most FOIs are constructed by investigating the positions and amplitudes of the points of interest (POIs) on the IC curve. For instance, Li *et al.* compared the estimation accuracy of the IC features generated from different POIs [16]. The main POIs that have been identified as qualified for battery SOH estimation are the peak and valley points, especially in high-voltage regions [17], [18], and it usually requires the battery to cycle under a full charging/discharging protocol in the lifelong operation. Unfortunately, such a cycling process is rarely experienced in real-world applications, and in many cases, these POIs are inappropriate to be used for full life-cycle prediction. For example, the peak points on the IC curves may decrease and vanish towards the battery's end of life. As a result, the effectiveness of the SOH estimation algorithms based on these features would be significantly reduced for aged batteries. To address this problem, Li *et al.* focused on the datasets with a selected state of charge (SOC) range during actual charging processes and a model was proposed to adopt multipoint features as the model inputs [19]. Indeed, by selecting several points distributed in a region with a drastically changed IC curve, the risk of losing a single point feature as the battery ages can be lowered. Nevertheless, both features extracted from a single point and multiple points are prone to measurement noises. In contrast, using area features of the IC curves can effectively reduce the sensitivity to these noises and mitigate the influence of applied filtering algorithms, thereby achieving high accuracy for battery SOH estimation [20]. The theoretical backgrounds for all the IC features mentioned above have been explored in the literature, and it shows that features are associated with certain electrochemical reactions and degradation mechanisms inside batteries [21]–[23].

Indeed, the works mentioned above with different FOIs are valuable and have actively contributed to the advancements in Li-ion battery health management. Nevertheless, since the existing methods based on different FOIs were validated with a specific dataset for battery cells with different types of chemistry and specifications, for practical applications, it is unclear how different features can affect the estimation results unless a systematic comparison is conducted in advance. Since the datasets previously used have been collected from different sources that often differ from each other in experimental settings, it is not only unfair but also unreasonable to conclude one is superior to another directly. A comprehensive comparative study of the IC features before model development can potentially remove the need of a labor-intensive feature selection process, but unfortunately, there is a lack of such a study in the existing works.

Most of the algorithms based on IC features are investigated and validated at the battery cell level. A common problem found in the above-mentioned SOH estimation algorithms is that the effects of cell inconsistency are overlooked. In fact, cell inconsistency is inevitable due to manufacturing tolerance: even the same type of batteries from the same manufacturing batch will exhibit different characteristics. The inconsistency can be magnified with the cell being used due to unbalanced aging trajectories caused by thermal, electrical, and mechanical nonuniformity in the battery packs. Hence, the

suitability for extrapolating the predictive model derived by fitting the data from one cell or one pack to other individuals is not guaranteed.

The model migration method, originally proposed for modeling similar injection molding processes by Lu *et al.* [24], can be an effective tool to deal with the problems caused by unavoidable cell-to-cell variations. In the model migration method, a base model is first developed from a sufficiently large amount of data, and the developed base model is then migrated as a new model to describe a similar process where only a limited amount of new data are available, and this concept has been introduced into the battery management community recently. For example, to predict the aging trajectory and remaining useful life, the Bayesian Monte Carlo method [25] and artificial neural networks (ANN) [26] were used to estimate the migration parameters for the new model online. It should be noted that, in [25] and [26], batteries were cycling under a periodic current profile in a well-controlled laboratory environment, where experienced cycles were well-defined. The cycle number, which correlates closely with the SOH, can be used as model input to predict future battery aging behaviors. Unfortunately, in real-world EV and smart grid applications, the operating profiles of different battery cells or the same battery among different cycles can be very different. Therefore, it is difficult to give an appropriate and universal definition for the cycles over the battery lifespan for practical use.

In view of the above, involving the cell inconsistency and abandoning the features like cycle number are considered not just beneficial but usually imperative in many circumstances. This motivates us to develop a combined offline-online SOH estimation scheme with high adaptability to cell inconsistency and low requirements on operational data based on model migration and machine learning. Specifically, suitable IC features are first selected and extracted to provide the information for training a pool of predictive models fused with slope-bias correction (SBC) and radial basis function neural networks (RBFNNs). A comparative study of the three most commonly used FOIs is carried out, and it is found that the peak value in the high-voltage region of IC curves is the most effective feature to indicate battery SOH. This feature is used as the input of the proposed SBC-RBFNN models, where highly nonlinear relationships of the corrected parameters in the SBC method are identified by RBFNNs. To handle the cell inconsistency between the candidate offline models and the new target cell, a modified random forest regression (mRFR) based online ensemble learning is developed for the first time to realize individualized estimation by incorporating the adaptively generated weighting into the offline SBC-RBFNN models. The applicability and the effectiveness of our proposed SBC-RBFNN-mRFR-based method are validated through comparative studies with several state-of-the-art SOH estimation techniques using the data collected from laboratory tests. Finally, the potential to apply the proposed blended offline/online machine learning framework to real-world cases is exhibited, where the input features can be generally selected using system-level measurements during EV operation. An overview of the proposed method is depicted in Fig. 1.

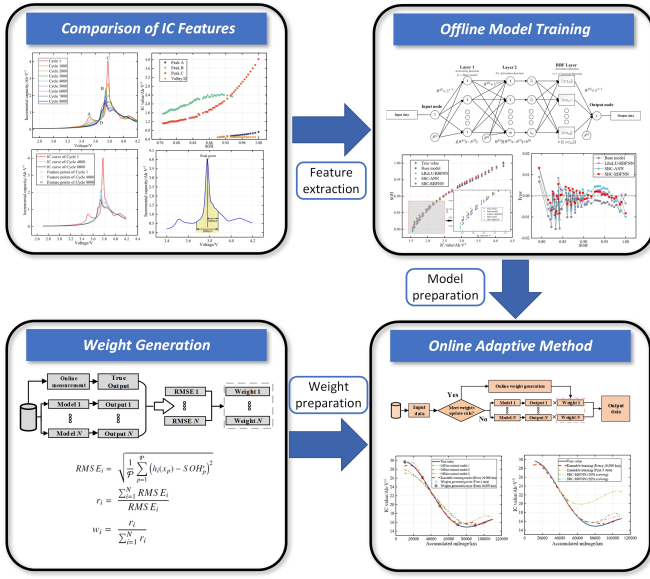


Fig. 1. Overview of the proposed method.

II. DATASET DESCRIPTION AND DATA PREPROCESSING

A. Description of Dataset

The Oxford Battery Degradation Dataset is first used in this work. The dataset contains measurements of battery aging data from eight commercial Kokam pouch cells of 740-mAh nominal capacity, with a graphite-based negative electrode and a lithium cobalt oxide/lithium nickel cobalt oxide positive electrode [27]. The cells were tested in a thermal chamber at 40 °C and repeatedly exposed to a 1C or C/25 constant-current-constant-voltage charging profile, followed by a driving cycle discharging profile obtained from the Urban Artemis profile. Here, the 1C current rate is defined as the current through a battery divided by the theoretical current draw under which the battery would deliver its nominal rated capacity in 1 h. Characterization measurements with a sampling frequency of 1 Hz were taken every 100 driving cycles. More detailed descriptions of this dataset can be found in [10], [27], and we denote the dataset as the *laboratory dataset*. Furthermore, a real-world dataset containing various operating information of EVs will be used for testing the proposed algorithm in Section V-C, where the description of the dataset will be provided.

B. Data Preprocessing

The terminal voltage of a Li-ion battery cell is a sum of the OCV, hysteresis voltage, and overpotentials due to polarization and the internal (ohmic) resistance, all affected by cell temperature. Usually, when the current rates applied to the battery are very low, both the hysteresis voltage and the polarization overpotentials are negligible [28], [29], and the effect of temperature change due to self-heating can also be ignored [30]. Hence, the voltage data measured under very low current rates in the laboratory dataset usually contain the most pertinent knowledge of the OCV, and we denote them as the pseudo-open-circuit voltage (pseudo-OCV) data, e.g., the

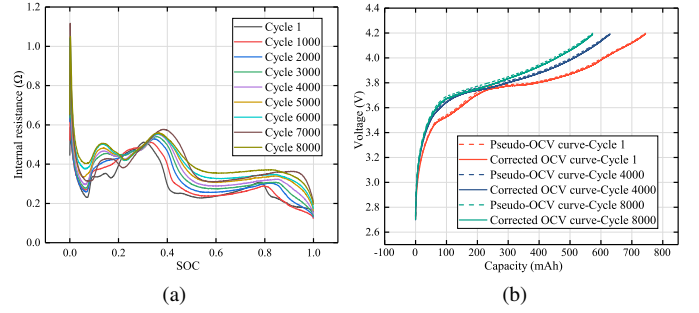


Fig. 2. (a) Comparison of internal resistance versus SOC relationships at different cycle numbers. (b) Comparison of original and corrected OCV curves at different cycle numbers.

C/25 constant current in the characterization processes in the Oxford Battery Degradation Dataset.

Although close to the true OCV data, the pseudo-OCV data are disturbed by measurement errors and affected by the inconsistent variation of the internal resistance. Directly using the pseudo-OCV data for feature extraction can thus lead to inaccurate SOH estimation results, and it is beneficial to approximate the true OCV curves before conducting feature extraction. To explain this, considering under the condition of low current rates, the terminal voltage during charging and discharging can be expressed as

$$V_{ch} = OCV_{ch} + I_{ch}R \quad (2)$$

$$V_{dc} = OCV_{dc} - I_{dc}R \quad (3)$$

where the symbols V , I , and R represent the terminal voltage (pseudo-OCV), the current magnitude, and the internal resistance, respectively, and the subscripts ch and dc denote the charging and discharging processes, respectively.

Since in this dataset the batteries were fully charged and fully discharged in each characterization, the SOC and the SOH can be defined and calculated by

$$SOC_t = Q_t/Q_n \quad (4)$$

$$SOH_n = Q_n/Q_0 \quad (5)$$

where Q_t represents the Coulomb-counting capacity at time t of the n th characterization, Q_n denotes the charging capacity of the n th characterization, and Q_0 is the cell capacity at the beginning of life.

Considering the consecutive charging/discharging process with the same current rate and ambient temperature, at the same SOC level $SOC_{ch} = SOC_{dc}$, we have $OCV_{ch} = OCV_{dc}$. Applying this condition to (2) and (3) yields the expression of the internal resistance

$$R = \frac{V_{ch} - V_{dc}}{I_{ch} + I_{dc}} = \frac{V_{ch} - V_{dc}}{2I} \quad (6)$$

where $I_{ch} = I_{dc} = I$.

Next, according to (2), the true OCV can be approximated by the following corrected OCV, i.e.,

$$OCV_{corrected} = V_{ch} - I_{ch}R. \quad (7)$$

Based on (6), an example of calculated internal resistances

versus SOC relationships at different cycles are obtained and shown in Fig. 2(a). Based on (7), an example of corrected OCV curves versus cell capacity relationships at different cycles are obtained and plotted along with the corresponding pseudo-OCV in Fig. 2(b).

It should be noted that directly using the corrected-OCV to calculate the IC curves may have some numerical problems due to the presence of measurement noises from voltage and current sensors. For example, during the charging process, the denominators in (1) should all be positive according to the characteristics of the battery. However, when the measured voltages vary disorderly due to measurement noises, the denominators can become zero or go negative, leading to erroneous numerical results. To tackle this problem, we follow the method proposed in [12], [16] by sequentially applying the support vector regression and the Gaussian window filter, so that the impact of the measurement noises can be effectively removed. With these techniques, the filtered IC curves are obtained using all the cycling data of the eight cells in the laboratory dataset, and these filtered IC curves will be used next for a comparative study. An example with nine IC curves at different cycle numbers of a cell is shown in Fig. 3.

III. COMPARISON OF IC FEATURES

In this section, a comparison of the three most adopted types of IC features in the literature, including single point features, multipoint features, and area features, is performed using the preprocessed IC data described in Section II-B. For each type of feature, a predictive model $y = h(x)$ is first established using the data from one of the eight cells in the laboratory dataset, to describe the relationship between the selected feature(s) (denoted by x) and the battery SOH (denoted by y). The rest cells form the testing group to evaluate the accuracy of the methods based on different features. Such a process was repeated several times by randomly selecting a cell to construct the predictive model.

To evaluate the accuracy of the model, the root-mean-square error (RMSE) is used as the criterion, defined as

$$RMSE = \sqrt{\frac{1}{M} \sum_{k=1}^M (h(x_k) - y_k)^2} \quad (8)$$

where k and M denote the index and the total number of data samples, respectively.

A. Single Point Features

As mentioned earlier, the peak and valley points on the IC curves represent the phase equilibria and phase transitions of the battery, and they contain the most apparent features of IC curves [31]. In this work, three peak points (A, B, and C) and one valley point (D), indicated in Fig. 3, are selected as the POIs, since their features are easy to extract as battery SOH indicators. For each of these points, the horizontal axis is related to the phase status, while the vertical axis denotes the transformed occasion of the battery phase. Based on the comparison and analysis in Appendix, the peak point C has been found to be qualified for battery health indication.

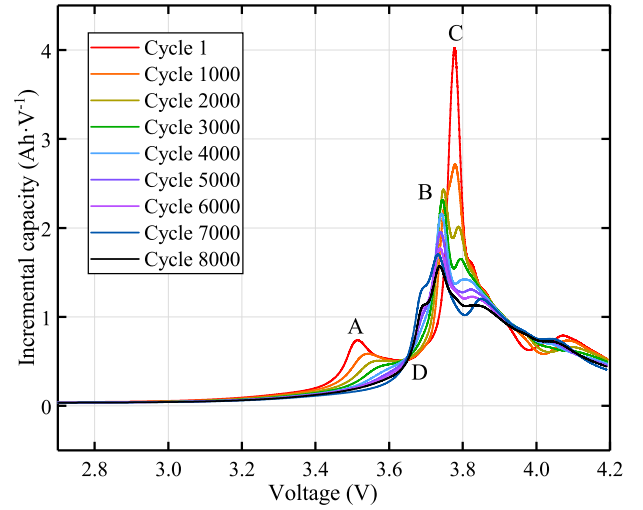


Fig. 3. Filtered IC curves at different cycle numbers with three peak points (A, B, and C) and one valley point (D).

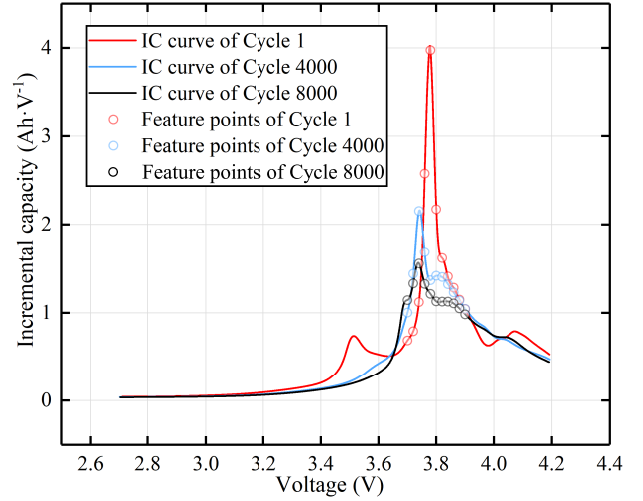


Fig. 4. Multipoint features: Eleven feature points (marked by circles) on the IC curves at different cycle numbers.

B. Multipoint Features

The IC peak or valley based on a single point may vanish before the battery reaches the end of life, e.g., 80% of SOH. In this situation, multipoint features can be used to deal with the problem [19]. To illustrate, the IC curves at three cycle numbers in Fig. 3 are plotted again in Fig. 4. It is observed from Fig. 4 that these IC curves have the most drastic changes when the voltage varies from 3.70 V to 3.90 V. Hence, eleven evenly spaced points are chosen for investigation in this region in this work, as suggested by [19]. Unlike the single point feature discussed in Section III-A, it is difficult to obtain the relationship between SOH and multipoint features intuitively, and thus the Pearson correlation analysis is carried out to identify the correlation between the IC values of these points and the battery SOH. Detailed analysis and results are given in Appendix. It is found that the average predictive RMSEs of the model based on the multipoint features are slightly larger than the result based on single point features.

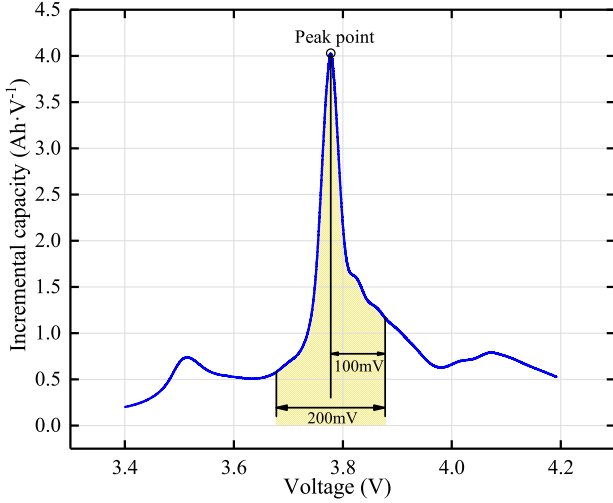


Fig. 5. Area feature obtained from the IC curve with a voltage interval of 200 mV.

TABLE I
COMPARISON OF RMSES BASED ON METHODS WITH DIFFERENT IC FEATURES

| | Single point feature | Multipoint feature | Area feature (20 mV) |
|--------------|----------------------|--------------------|----------------------|
| Average RMSE | 0.00804 | 0.01335 | 0.00805 |

C. Area Feature

Area features, physically representing the charging/discharging capacity of the cell, can be obtained by integrating the IC curve preferably over the voltage interval with drastic changes in IC values. As an example, Fig. 5 shows how the area is obtained when the voltage interval is 200 mV, centered at the peak point, for extracting the IC area feature. According to the analysis and result given in Appendix, there is a declining trend in the RMSE as the voltage interval becomes narrower.

We notice that since the IC value of the midpoint of the interval can be considered the charging capacity with a voltage interval close to zero, it can be inferred that the IC value of peak point C should be superior to all the area features in the present investigation. This is confirmed in Table I, where the average predictive RMSEs based on the above three IC feature selection methods are compared based on the investigated datasets using the narrowest interval of 20 mV. The results also show that the single point feature achieves the highest predictive accuracy among these three types of features, and the RMSE of the area feature is very close to that of the single point feature. Hence, the IC value of the highest peak point in the high-voltage region will be adopted in this study.

IV. DEVELOPMENT OF CELL-INCONSISTENCY-CONSCIOUS BATTERY SOH ESTIMATION METHOD

Directly using a single point feature for SOH estimation, however, can lead to significant estimation errors when cell inconsistency is non-trivial. To illustrate, Fig. 6 plots the relationships between the IC value of the peak point C and

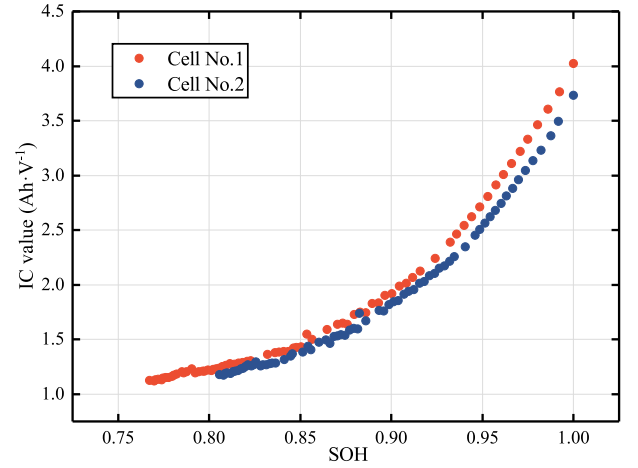


Fig. 6. Variations of IC values at different SOH for two cells with the same specification and cycled in the same environment.

SOH for two battery cells in the laboratory dataset. Although these two cells were cycled under the same condition, their IC profiles with respect to SOH are unsurprisingly different, and the level of difference varies over time. Specifically, as the IC value reduces, there is an initial decrease trend in the SOH difference between the two cells, while such a difference increases as the IC value further reduces. This means that a large estimation error can be introduced if the predictive model trained using a cell is directly applied to another by ignoring the inherent cell-to-cell inconsistency. Particular attention needs to be paid to the period close to the end of life of the batteries, where an accurate SOH estimation is extremely important to avoid battery failures. This problem will also be amplified in real-world situations where a higher level of cell inconsistency often exists.

Therefore, mitigating the impacts of model differences on the predictive results is a key task for achieving accurate and reliable battery SOH estimation. We propose to solve this problem in two steps using a scheme combining offline and online machine learning. First, an offline approach based on SBC-RBFNN is proposed to effectively train N individual models with high variability. Next, the N SBC-RBFNN models are properly blended in an efficient ensemble learning framework for online adaptive estimation. The two steps will be detailed in the following two subsections, respectively. The efficacy of the proposed framework will be verified in Section V by comparing it with several prevailing data-driven methods using the laboratory data, and its applicability to real-world operation will also be examined.

A. Development of Offline Models Based on SBC-RBFNN

The individual candidate models to be developed should cover the entire space of the possible cell variation and be simple to train. Since model migration can properly balance the individuality and similarity in generating these models, it is well-suited for our present investigation. Several model migration methods have been proposed in the literature to reliably deal with the predictive performance of similar processes

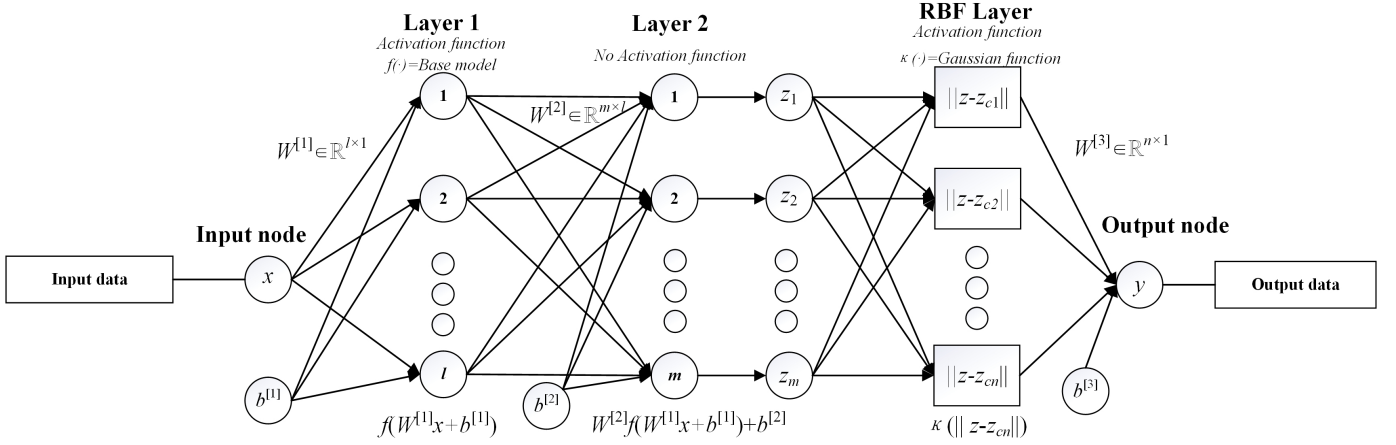


Fig. 7. Structure of the proposed offline SBC-RBFNN model.

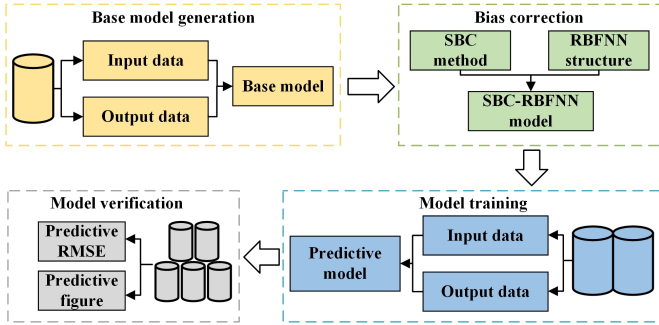


Fig. 8. Flowchart of the offline model training process.

and save experimental resources. Amongst them, the SBC is considered to be one of the most effective approaches yet simple to implement [24]–[26]. In the SBC, the inconsistency between the base model $f(\cdot)$ and the new model is completely parameterized by the input slope $W^{[1]}$ and bias $b^{[1]}$ as well as the output slope $W^{[2]}$ and bias $b^{[2]}$. The general equation $g(x)$ for the SBC model is given by

$$g(x) = W^{[2]}f(W^{[1]}x + b^{[1]}) + b^{[2]}. \quad (9)$$

One of the core contributions of this work, i.e., the proposed three-layer SBC-RBFNN framework, is schematized in Fig. 7, where Layer 1 and Layer 2 are described by the SBC model (9), and an RBFNN model is connected to the output of the SBC structure for its excellent capability to capture the underlying nonlinear relationship. In an RBFNN, the activation function can map the input feature to a high dimensional space and transform the nonlinear relationship into a linear one. This can make the complicated connection between inputs and outputs easily observed and distinguishable [32], [33] and it is thus well-suited to the problem at hand.

In Fig. 7, the numbers of nodes of the three layers in the SBC-RBFNN are denoted by l , m , and n , respectively. The IC values of peak point C are the inputs of the base model (Layer 1), corrected by the SBC function (Layer 2). The outputs of Layer 2 can be written as a vector $g(x) = z = [z_1, z_2, \dots, z_m]^T \in \mathbb{R}^m$, which acts as the input vector of the

RBF layer (Layer 3). As the model output, the SOH values are predicted by the RBFNN model, in which the Gaussian function is used as the transfer function, i.e.,

$$\kappa(z, z_{cj}) = \exp\left(-\frac{\|z - z_{cj}\|^2}{2\sigma_j^2}\right) \quad (10)$$

where $\kappa(\cdot, \cdot)$ represents the kernel function, $j \in \{1, 2, \dots, n\}$ is the index of the node in the RBF layer, and z_{cj} and σ_j are the RBF center vector and the standard deviation of the Gaussian function of the j th node, respectively. The RBFNN function can thus be written as

$$y = \sum_{j=1}^n w_j^{[3]} \cdot \kappa(z, z_{cj}) + b^{[3]} \quad (11)$$

where we denote $W^{[3]} = [w_1^{[3]}, w_2^{[3]}, \dots, w_n^{[3]}]^T$ and $b^{[3]}$ as the weights and bias of the RBF layer, respectively.

The offline SBC-RBFNN model structure is obtained by combining (9)–(11), denoted by $y = h(x)$. To obtain the parameters for one individual model, including the weights $W^{[1]}$, $W^{[2]}$, and $W^{[3]}$, the biases $b^{[1]}$, $b^{[2]}$, and $b^{[3]}$, as well as the center vector $[z_{c1}, z_{c2}, \dots, z_{cn}]$, the datasets are randomly split into the base model group, training group, and testing group, and they are used for base model generation, SBC-RBFNN model training, and model verification, respectively. The RMSE between the model and the expected output, calculated by (8), is used as the loss function in the training process. Once the predictive model is established, the remaining laboratory datasets of battery cells are used to verify and test the model accuracy. Fig. 8 presents the flowchart of the entire modeling, training, and verification process. For demonstrative purposes, in the figure, one of the eight cells is randomly chosen to generate the base model, and two other cells are selected to train the SBC-RBFNN model, and the remaining five cells in the laboratory datasets are used as the test group to verify the trained model. Such a process is repeated by N times with the data groups obtained by randomly splitting the datasets.

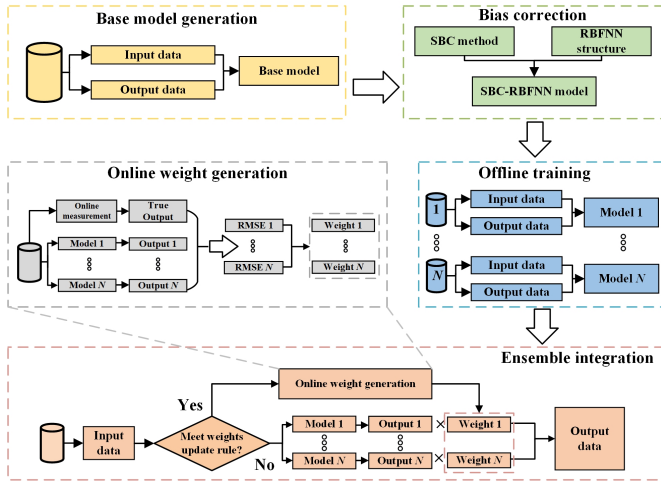


Fig. 9. Flowchart of the online ensemble learning process.

B. Online Model Adaptation Using Modified Random Forest Regression

In the previous subsection, the N SBC-RBFNN models offline trained from historical data are expected to provide sufficient information to cover the most representative aging behaviors of all individual cells, and they form a pool of candidate models for individualized estimation. In order to estimate the SOH for a new target cell, of which only limited early-age data and online measurements are available, it is essential to develop an algorithm to exploit these N offline models for online use. A general and simple strategy is to find a proper blending scheme by weighted averaging, i.e.,

$$\bar{h}_k = \sum_{i=1}^N w_i h_i(x_k) \quad (12)$$

where x_k and \bar{h}_k represent the blended model input and model output at time instant k , respectively, $h_i(\cdot)$ is the i th offline SBC-RBFNN model established based on the method described in Section IV-A, and our objective is to find proper weights $\{w_1, w_2, \dots, w_N\}$ for online operation.

We notice that for offline applications, a similar task usually can be readily addressed by using the random forest regression (RFR), a well-developed ensemble learning method [34]. In the RFR, a set of decision trees are planted and the prediction results from all individual trees are averaged to obtain the aggregated prediction output [35]. The RFR is superior to the methods based on a single decision tree since the overfitting issue can be overcome without trading away the accuracy of the prediction [36]. However, since the RFR method is conventionally designed for offline applications, the output of the RFR model is obtained simply by averaging the individual models trained from a large amount of offline data using ensemble voting. In other words, the same and fixed weight w_i is placed to each of the offline contributing models. Nevertheless, during online use for battery SOH estimation with high cell inconsistency, it is counterintuitive and injudicious to trust some offline models that behave very differently from the considered cell. Such a method is also inappropriate for the

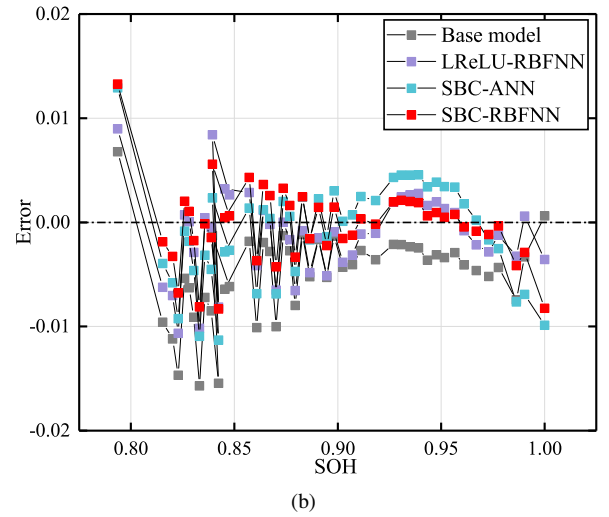
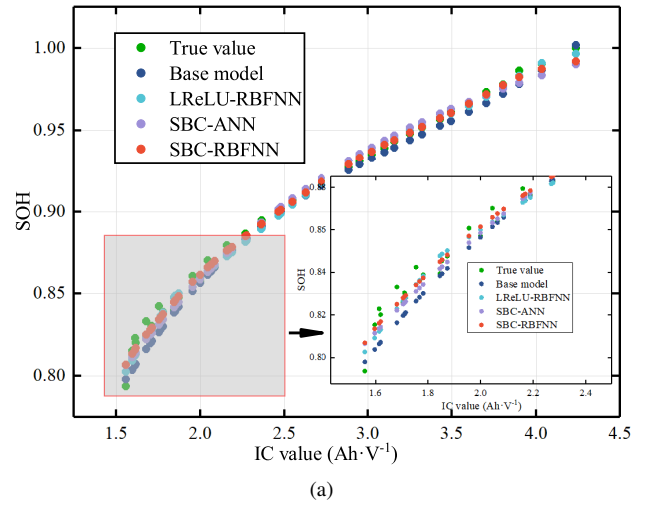


Fig. 10. Offline model predictive results of No. 6 cell in Case 1: (a) SOH. (b) SOH estimation error.

present study since for online estimation usually only limited data are available.

Here, to solve the problem for improving estimation accuracy in the presence of limited online data, a modified RFR (mRFR) based online ensemble learning algorithm is proposed to adaptively generate and adjust the weights w_i of each offline model, with the structure illustrated in Fig. 9. A similar “bootstrap” technique in the conventional RFR is adopted to randomly sample from the dataset with replacement, resulting in N different and uncorrelated decision trees, namely the offline models. The way to plant trees is different from the conventional RFR and has been described in Section IV-A. Once the N offline SBC-RBFNN models with corresponding model parameters have been determined, based on the real-time measurements that can reflect the user’s real-world behaviors, the weights w_i are next generated and updated online according to the following steps:

First, for i th candidate offline model, its RMSE over a selected time horizon \mathcal{P} in one online adaptive event is

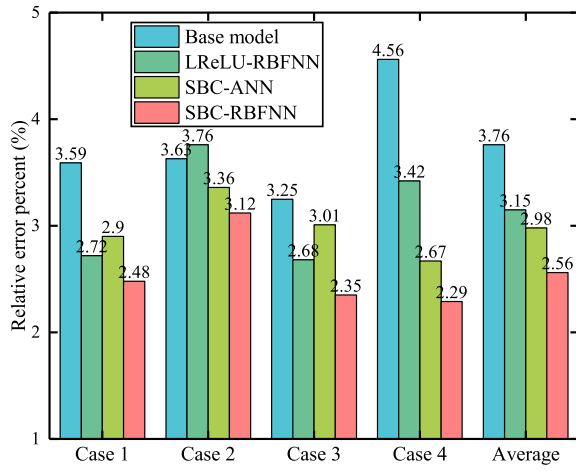


Fig. 11. Comparison of average REPs between different offline models based on the laboratory dataset.

calculated:

$$RMSE_i = \sqrt{\frac{1}{P} \sum_{p=1}^P (h_i(x_p) - h'_p)^2} \quad (13)$$

where $h_i(x_p)$ and h'_p are the estimated output from the i th model and the true output data (i.e., true SOH), respectively, both obtained at time step p . To calibrate the true SOH in terms of the battery capacity, a full discharge and charge process is needed. This can be readily triggered in experimental studies, but rarely occurs for real-world EV battery systems, which is also the reason to estimate SOH. A practical scenario is to generate SOH measurements when EVs do regular maintenance in the service center, which can be every six months or 10,000 km, for instance. In that case, the online adaption is then activated in a much slower timescale than the offline models.

Next, we consider the importance of each model inversely proportional to its calculated RMSE, and the weighting factor for the i th model is given by

$$r_i = \frac{\sum_{i=1}^N RMSE_i}{RMSE_i} \quad (14)$$

Finally, normalizing r_i yields the weight for the i th offline model, i.e.,

$$w_i = \frac{r_i}{\sum_{i=1}^N r_i} \quad (15)$$

with which the estimated SOH can be obtained based on (12). When the sampling time interval of p is larger than that of k , the weights w_i will be held constant until the next SOH measurement comes at $p + 1$.

The length of the horizon P in (13) affects the performance of the proposed mRFR ensemble learning method for battery SOH estimation. In this work, we propose two strategies (weight update rules) to determine P so that the weights w_i can be adaptively updated based on the user's behaviors. The first strategy is to adjust the weights according to the first several estimation errors from the offline models. This method is inspired by the pre-market qualification test of real-

world EVs, from which the early-stage datasets of true battery SOH can be obtained. The second strategy is established by considering that each EV will have regular maintenance in the service center once a stipulated amount of mileage is reached.

The introduced randomness in selecting the offline models can minimize the risk that all individual models constantly overestimate or underestimate battery SOH in the same direction. In the conventional RFR, a very large number of trees and deep layers within each tree are often needed to achieve high modeling accuracy but at the expense of demanding computational power and a long training period. In the proposed mRFR, all the trees are planted by the model-migration-aided RBFNN to map the IC peak to SOH. In practice, different and representative trees can be meticulously selected from the pool, thereby significantly mitigating the computational burden for online SOH estimation. In addition, the designed online adaptive estimator is able to assimilate the ever-changing user's behaviors and operating environments, and has the potential to provide high-fidelity SOH information.

V. RESULTS AND DISCUSSION

To validate the proposed combined offline-online battery SOH estimation algorithm based on SBC-RBFNN and mRFR, in this section, comparative studies with several prevailing approaches are provided.

A. Verification of the Offline Model

For the proposed offline SBC-RBFNN model, first, a cubic polynomial is found to be a good candidate to describe the base model in this study by trial and error. Backpropagation is chosen for the training of the SBC-RBFNN, where the gradient descent method is used to solve the fitting problem and the learning rate is set as 0.01 based on extensive training and testing. The maximum number of iterations for training is set to 1000 and the stop criterion is that the RMSE for the training set drops below 0.0001. The numbers of nodes in the three hidden layers are set to $l = 5$, $m = 5$, and $n = 25$, respectively, determined by trial-and-error. It is worth noting that the number of RBF layer nodes n is set much smaller than the size of training samples to avoid the problems of overfitting and poor generalization [37], [38]. Limiting the complexity of the network by reducing the number of possible centers has the added advantage of producing parsimonious networks [39].

Three benchmark algorithms are designed to verify the stability and effectiveness of the proposed offline model and they are described as follows.

Benchmark 1 directly uses the base model generated from one cell or vehicle dataset to estimate the battery SOH of others. Such a method is widely used in previous research works such as [16], [22].

Benchmark 2 is a Leaky Rectified Linear Unit (LReLU) based RBFNN model. Due to its capability to deal with the negative part of datasets, the LReLU is superior to the traditional ReLU function widely used in ANN [40], [41]. The structure and the configuration of the LReLU-RBFNN model are similar to that of the SBC-RBFNN model as depicted in

Fig. 7. The only difference is that the activation function in Layer 1 is replaced by the LReLU function, given as

$$y_k = \begin{cases} x_k, & \text{if } x_k \geq 0 \\ \frac{x_k}{\alpha_k}, & \text{if } x_k < 0 \end{cases} \quad (16)$$

where $\alpha_k \in (1, +\infty)$ is a fixed parameter. The major difference between the proposed SBC-RBFNN and the LReLU-RBFNN is that the latter lacks prior knowledge about the research subject while such information is available for the proposed SBC-RBFNN by model migration.

Benchmark 3 is an SBC-ANN model proposed in [26], which is used to verify the superiority of the RBFNN. The training method, learning rate, and nodes in Layer 1 and Layer 2 of the SBC-ANN are both set the same as the SBC-RBFNN for a fair comparison.

The laboratory datasets are used to examine the performance of the proposed model following the methodology in Section IV-A. Table II presents the verification results for all four dataset combinations, where the predictive RMSEs are calculated by (8). In Table II, it is clearly shown that the predictive accuracy can be significantly improved by using model migration: The SBC-RBFNN model can better rebuild the nonlinear relationships with a smaller predictive error compared to the SBC-ANN model. Furthermore, although trained by more datasets, the LReLU-RBFNN model still generates more significant predictive errors than the proposed SBC-RBFNN model. In fact, the LReLU-RBFNN even performs worse than the SBC-ANN model in terms of the average RMSE. This result exhibits the importance of the experience and information buried in the base model. The graphical results in Fig. 10 also show the superiority of the SBC-RBFNN model, especially in the region below 90% SOH, which indicates that the proposed method is more reliable than the two prevailing methods under comparison.

The relative error percentage (REP) is next used to signify the influences of estimated errors on the overall process of battery aging. The REP for the i th model, E_{ri} , is defined by

$$E_{ri} = \frac{E_i}{S_i} \times 100\% \quad (17)$$

where E_i and S_i represent the predictive RMSE and the range of the SOH for the i th cell, respectively. Comparisons of the calculated REPs are illustrated in Fig. 11 in terms of the average relative errors over the battery lifetime. It can be seen from Fig. 11 that by using the SBC-RBFNN method, the predictive accuracy is 32% higher than the direct estimated results using the base model, while the latter is the most commonly used method in the literature. In addition, by incorporating the model migration, we observe an improvement of about 19% and 14% in the predictive accuracy, compared to using the LReLU-RBFNN and the ANN-based methods, respectively.

B. Verification of the Online Adaptive Method Based on Modified Random Forest Regression

The performance of the mRFR-based online adaptive method for individualized battery SOH estimation will be examined in this subsection. The core step of the proposed

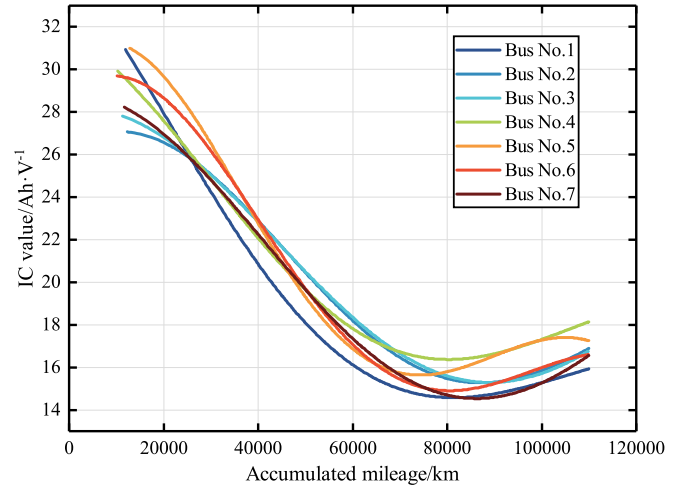


Fig. 12. IC Values of peak points versus accumulated mileages of seven selected electrified buses in the real-world dataset.

method is the weight generation (13)–(15), where the ensemble weights w_i are determined by the predictive errors of each offline model. Two offline models are trained individually using the datasets of two cells. Datasets of another three cells are extracted to produce the base model for the benchmark method, and the offline SBC-RBFNN models for ensemble learning to realize a fair comparison. The two weight update rules as described in Section IV-B are examined. In the first rule, the predictive RMSEs of the first five data are chosen to generate w_i . In the second rule, two kinds of updating intervals are chosen. The estimated results using the offline SBC-RBFNN will be used as the benchmark in this subsection. The benchmark method is configured according to the suggestion given in [26], where the first 30% and 50% of the data are used to train the individual prediction model. The predictive accuracy is tested using the remaining data in the dataset.

The numerical results are compared in Table III based on the laboratory datasets, where REP_A is the average REP of the test cells calculated by (17). Considering that the data sizes for each of the eight cells are different, the training data percentage of the online ensemble learning method for each case is provided in the table. It shows that in order to improve the predictive accuracy in terms of the average RMSE, the size of the training dataset has to be increased for the offline SBC-RBFNN model. In contrast, increasing the data size for training is not necessary for our proposed ensemble learning method: Only a small amount of data are needed to achieve about 90% improvement in the estimation accuracy compared to using the offline SBC-RBFNN model trained by the first 30% of data. Compared to the offline model trained with the first 50% of data, the percentage of the improvement on the performance is still high up to 80%. It is worth noting that the estimation accuracy using the two weight generation methods is very close, and both methods are very effective. In this case, with the high quality of the laboratory dataset, only a few measurements at the early operating stage are needed to determine the individual battery aging pattern.

TABLE II
RMSES OF PREDICTED SOH WITH DIFFERENT OFFLINE MODELS BASED ON THE LABORATORY DATASET

| Base Model | Training Cells | Model Type | Test Cells | | | | | Average |
|------------|---------------------|-------------|------------|---------|---------|---------|---------|----------------|
| | Case 1 | | No.1 | No.4 | No.5 | No.6 | No.7 | |
| No.2 | No.2, No.3 and No.8 | Base Model | 0.00498 | 0.00583 | 0.00620 | 0.00650 | 0.01215 | 0.00713 |
| | | LReLU-RBFNN | 0.00583 | 0.00458 | 0.00379 | 0.00399 | 0.00921 | 0.00548 |
| No.2 | No.3 and No.8 | SBC-ANN | 0.00807 | 0.00527 | 0.00387 | 0.00474 | 0.00759 | 0.00591 |
| | | SBC-RBFNN | 0.00716 | 0.00421 | 0.00293 | 0.00374 | 0.00727 | 0.00506 |
| | Case 2 | | No.1 | No.2 | No.5 | No.7 | No.8 | |
| No.4 | No.3, No.4 and No.6 | Base Model | 0.00820 | 0.00985 | 0.00421 | 0.00906 | 0.00734 | 0.00773 |
| | | LReLU-RBFNN | 0.00938 | 0.01017 | 0.00372 | 0.00896 | 0.00797 | 0.00804 |
| No.4 | No.3 and No.6 | SBC-ANN | 0.00611 | 0.00870 | 0.00380 | 0.00897 | 0.00801 | 0.00712 |
| | | SBC-RBFNN | 0.00613 | 0.00785 | 0.00322 | 0.00889 | 0.00697 | 0.00661 |
| | Case 3 | | No.3 | No.4 | No.5 | No.6 | No.8 | |
| No.1 | No.1, No.2 and No.7 | Base Model | 0.00493 | 0.00576 | 0.00633 | 0.00670 | 0.00972 | 0.00669 |
| | | LReLU-RBFNN | 0.00692 | 0.00606 | 0.00293 | 0.00396 | 0.00828 | 0.00563 |
| No.1 | No.2 and No.7 | SBC-ANN | 0.00690 | 0.00649 | 0.00539 | 0.00574 | 0.00640 | 0.00618 |
| | | SBC-RBFNN | 0.00559 | 0.00471 | 0.00305 | 0.00368 | 0.00762 | 0.00493 |
| | Case 4 | | No.2 | No.3 | No.4 | No.5 | No.6 | |
| No.7 | No.7, No.1 and No.8 | Base model | 0.01458 | 0.01098 | 0.00897 | 0.00631 | 0.00712 | 0.00959 |
| | | LReLU-RBFNN | 0.01453 | 0.00935 | 0.00558 | 0.00321 | 0.00412 | 0.00736 |
| No.7 | No.1 and No.8 | SBC-ANN | 0.00834 | 0.00508 | 0.00527 | 0.00422 | 0.00498 | 0.00558 |
| | | SBC-RBFNN | 0.00797 | 0.00470 | 0.00422 | 0.00330 | 0.00393 | 0.00482 |

C. Testing with Real-World EV Dataset

To verify the effectiveness of the proposed SBC-RBFNN-mRFR algorithm for real-world operation, a very large database from the Open Lab of the National Big Data Alliance of New Energy Vehicle of China was used in our study. The whole database contains information such as the position, accumulated mileage, states of drivetrain, and operating information of the battery systems of the EVs in China, recorded with a sampling interval of 10 s. Based on the information about position and temperature, seven EVs were identified as electrified city buses and selected as the data sources for our investigation. The dataset used for model training, testing, and verification can be the accumulated mileage, battery voltage, battery current, temperature, etc., and specifications and data preprocessing methods are detailed in [12].

Thanks to the relatively fixed and stable operational environment, electrified city buses usually have similar driving and charging patterns, especially when they are running on the same route. Consequently, their batteries should share very similar aging behaviors. Nevertheless, according to the IC value versus accumulated mileage relationship shown in Fig. 12 (see Ref. [12] in detail), there are obvious differences between the IC values of the analyzed buses. It can be seen that the difference between the IC value versus accumulated mileage curves in Fig. 12 is most likely the consequence of the inherent battery inconsistencies.

Note that the main purpose of this subsection is to examine the applicability and generality of the proposed model structure as shown in Fig. 7 to Fig. 9. Since the true SOH information is usually not known from the real-world dataset, for demonstrative purposes, the input and the output are replaced by the accumulated mileage and the IC value, respectively.

For the real-world operations, a similar training and testing configuration to the case of laboratory test for online operation, and the results using different algorithms are shown in Table IV. The datasets of the seven electrical city buses are randomly divided into three groups: One is to generate the base model, three are used to train three offline models, and the remaining three form the test group to evaluate the predictive performance of different online adaptive methods. All the above dataset combinations are generated randomly, and three different combinations of the dataset groups for training and testing are used to verify the performance of the online adaptive model. To illustrate, Fig. 13 shows the results of vehicle No. 6 in Case 1. Fig. 13(a) shows the process of ensemble integration. The data points of online weight update rules are also marked in Fig. 13(a) by circles with different colors, and it shows that ensemble learning can effectively enhance the predictive performance than only using one of the offline models.

The estimated IC values versus accumulated mileage relationships with different online adaptive methods are shown in Fig. 13(b). Here, the green and the red curves represent two well-established online adaptive methods in the literature, in which the offline model are trained by the first 30% and first 50% of the data, respectively. The results clearly show that by using these two methods, there are significant deviations from the ground truth, while the proposed ensemble-learning-based method has shown excellent predictive performance. In particular, with the proposed ensemble learning method based on the first weight generation strategy, 43%, 40%, and 58% predictive improvements can be achieved for the three models using only five recorded datasets, whereas the first 50% of the data have to be used for the SBC-RBFNN model training.

TABLE III
RMSES OF PREDICTED SOH WITH DIFFERENT ONLINE ADAPTIVE METHODS BASED ON THE LABORATORY DATASET

| Methods | Base Model | Offline Model | Online Adaptive | Test Cells | | | | | Average | REP _A |
|-----------|---------------------|---------------|-----------------------|------------|---------|---------|---------|---------|----------------|------------------|
| | | Case 1 | | No.1 | No.4 | No.5 | No.6 | No.7 | | |
| SBC-RBFNN | No.2, No.3 and No.8 | | First 30% of the data | 0.02608 | 0.03323 | 0.04294 | 0.06725 | 0.04321 | 0.04255 | 21.03% |
| | | | First 50% of the data | 0.02084 | 0.01701 | 0.01905 | 0.03417 | 0.02843 | 0.02390 | 11.81% |
| mRFR | No.2 | No.3 and No.8 | First 5 data (9%) | 0.00548 | 0.00451 | 0.00292 | 0.00393 | 0.00685 | 0.00474 | 2.34% |
| | | | Every 10 cycles (10%) | 0.00596 | 0.00422 | 0.00281 | 0.00383 | 0.00744 | 0.00485 | 2.39% |
| | | | Every 5 cycles (20%) | 0.00582 | 0.00406 | 0.00282 | 0.00371 | 0.00736 | 0.00475 | 2.35% |
| | | Case 2 | | No.1 | No.2 | No.5 | No.7 | No.8 | | |
| SBC-RBFNN | No.3, No.4 and No.6 | | First 30% of the data | 0.06054 | 0.04565 | 0.04184 | 0.06345 | 0.07366 | 0.05703 | 27.02% |
| | | | First 50% of the data | 0.03307 | 0.03211 | 0.02489 | 0.02871 | 0.04746 | 0.03325 | 15.75% |
| mRFR | No.4 | No.3 and No.6 | First 5 data (7%) | 0.00412 | 0.00455 | 0.00358 | 0.00814 | 0.00601 | 0.00528 | 2.50% |
| | | | Every 10 cycles (11%) | 0.00564 | 0.00544 | 0.00352 | 0.00815 | 0.00574 | 0.00570 | 2.70% |
| | | | Every 5 cycles (22%) | 0.00446 | 0.00530 | 0.00341 | 0.00802 | 0.00567 | 0.00537 | 2.54% |
| | | Case 3 | | No.3 | No.4 | No.5 | No.6 | No.8 | | |
| SBC-RBFNN | No.1, No.2 and No.7 | | First 30% of the data | 0.04937 | 0.04338 | 0.04704 | 0.03540 | 0.04041 | 0.04312 | 20.93% |
| | | | First 50% of the data | 0.02983 | 0.02306 | 0.01173 | 0.02419 | 0.03261 | 0.02428 | 11.79% |
| mRFR | No.1 | No.2 and No.7 | First 5 data (9%) | 0.00474 | 0.00539 | 0.00311 | 0.00420 | 0.00564 | 0.00462 | 2.24% |
| | | | Every 10 cycles (10%) | 0.00473 | 0.00380 | 0.00305 | 0.00401 | 0.00489 | 0.00409 | 1.99% |
| | | | Every 5 cycles (20%) | 0.00446 | 0.00350 | 0.00292 | 0.00351 | 0.00425 | 0.00373 | 1.81% |
| | | Case 4 | | No.2 | No.3 | No.4 | No.5 | No.6 | | |
| SBC-RBFNN | No.2, No.3 and No.8 | | First 30% of the data | 0.04932 | 0.05399 | 0.03572 | 0.04574 | 0.05480 | 0.04792 | 23.14% |
| | | | First 50% of the data | 0.02878 | 0.02459 | 0.02619 | 0.01488 | 0.01808 | 0.02250 | 10.86% |
| mRFR | No.7 | No.1 and No.8 | First 5 data (9%) | 0.00368 | 0.00364 | 0.00411 | 0.00525 | 0.00511 | 0.00436 | 2.10% |
| | | | Every 10 cycles (10%) | 0.00403 | 0.00403 | 0.00448 | 0.00509 | 0.00491 | 0.00451 | 2.18% |
| | | | Every 5 cycles (20%) | 0.00374 | 0.00363 | 0.00428 | 0.00500 | 0.00484 | 0.00430 | 2.08% |

On the other hand, with the second weight generation method (generated by every 10,000 km), 28%, 24%, and 24% predictive improvements can be achieved, respectively. Thus, for an EV in daily operation, high SOH estimation accuracy can be achieved with only a small amount of collected data. Hence, the results in Fig. 13 and Table IV verify the effectiveness of our proposed online adaptive method. REP_A in Table IV is the average REP of the tested vehicles, calculated by (17) with S_i denoting the range of IC value for the i th vehicle. Note that in Fig. 13, the increase in the IC values after the accumulated mileage of 80,000 km is due to increased ambient temperature [12].

The mRFR-based online adaptive method is further compared with the offline method using SBC-RBFNN and the numerical results are summarized in Table V. It shows that the average predictive RMSE of the mRFR-based method is smaller than the offline SBC-RBFNN models. A 15% and a 38% improvement are realized by incorporating the mRFR-based ensemble learning algorithm after fusing limited real-world measurements. We would like to point out that the divergent results between the two data sources are the consequence of inherent inconsistency in cell characteristics: The level of inconsistency for the cells in real-world EVs is generally higher than that of the laboratory-used cells. Hence, we conclude that the proposed online adaptive method shall lead to a more significant improvement in battery SOH estimation accuracy for practical applications compared to that in laboratory environments.

VI. CONCLUSIONS

In this study, an online adaptive ensemble learning scheme based on a combination of offline model training and online weight generation has been proposed to deal with the divergence problem in the battery SOH estimation caused by inherent inconsistency between individual research objects. The effectiveness of the IC features for battery SOH estimation was discussed based on high-quality laboratory datasets. According to a comprehensive comparative study, the height of the IC peak in the high-voltage region was found to be the best and selected as the feature for battery aging prediction. Then, a pre-trained model based on a combination of the SBC method with an RBFNN structure was developed. Four different dataset settings were chosen to test the stability of the SBC-RBFNN model among several existing offline training methods. Both the results based on the laboratory and real-world datasets showed that the proposed SBC-RBFNN model can significantly improve the predictive accuracy. An online adaptive scheme was next constructed by synthesizing the offline models and an ensemble integration process using a proposed modified random forest regression (mRFR) method. The weights for each offline model were generated by online measuring a few data from fixed cycle or mileage interval, which markedly reduced the requirement of datasets compared with the online adaptive methods in previous research. The effectiveness of the proposed SBC-RBFNN-mRFR model is validated using both the datasets from laboratory tests and real-world measurements. The proposed offline online blended

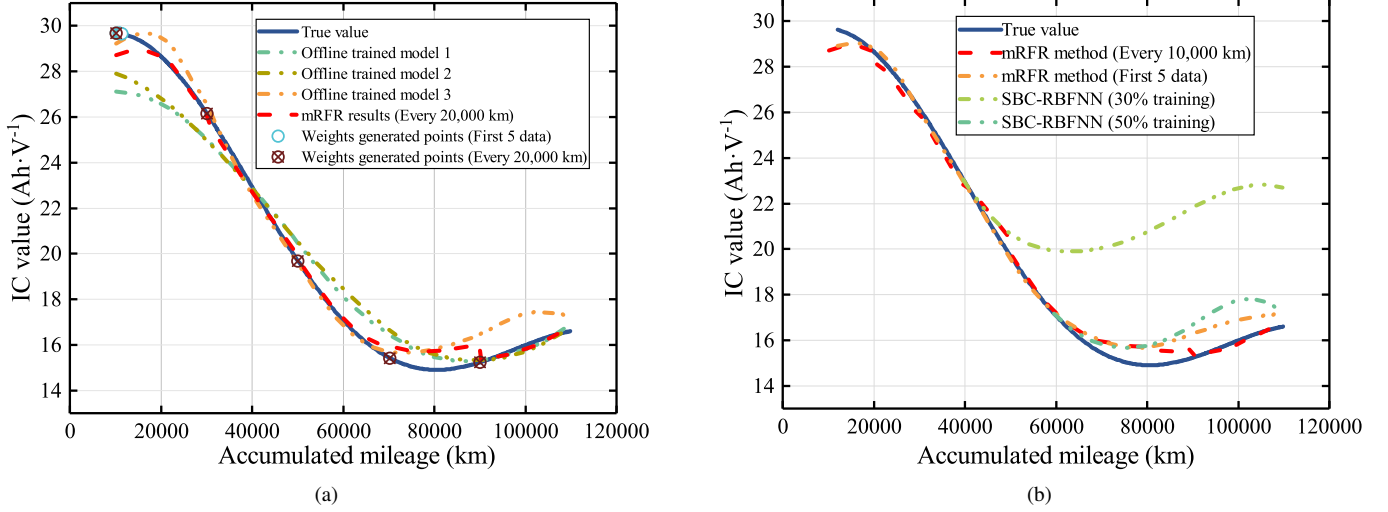





Fig. 13. (a) Two weight update rules of mRFR method for vehicle No. 6 in Case 1. (b) Comparison of the results using different online adaptive methods for vehicle No. 6 in Case 1. The increase in the IC values after the accumulated mileage of 80,000 km is due to increased ambient temperature [12].

TABLE IV
RMSES OF PREDICTED IC-VALUES WITH DIFFERENT ONLINE ADAPTIVE METHODS BASED ON THE REAL-WORLD DATASET

| Methods | Base Model | Offline Model | Online Adaptive | Test Vehicles | | | Average | REP _A |
|-----------|-------------------------------|---|------------------------|---------------|---------|---------|----------------|------------------|
| Case 1 | | | | No.4 | No.6 | No.7 | | |
| SBC-RBFNN | No.1, No.2, No.3, and No.5 |  | First 30% of the data | 4.04963 | 4.73145 | 4.33798 | 4.37302 | 31.20% |
| | | | First 50% of the data | 0.96629 | 1.13031 | 1.46371 | 1.18677 | 8.47% |
| mRFR | No.1 | No.2, No.3, and No.5 | First 5 data (0.8%) | 0.73058 | 0.58166 | 0.70287 | 0.67170 | 4.79% |
| | | | Every 20,000 km (0.8%) | 0.74749 | 0.37846 | 0.69259 | 0.60618 | 4.33% |
| | | | Every 10,000 km (1.6%) | 0.50626 | 0.37745 | 0.56829 | 0.48400 | 3.45% |
| Case 2 | | | | No.1 | No.5 | No.6 | | |
| SBC-RBFNN | No.2, No.3, No.4, and No.7 |  | First 30% of the data | 3.94096 | 3.11355 | 4.52841 | 3.86097 | 24.91% |
| | | | First 50% of the data | 1.11187 | 0.95490 | 1.33212 | 1.13297 | 7.31% |
| mRFR | No.3 | No.2, No.4, and No.7 | First 5 data (0.9%) | 1.29242 | 0.22556 | 0.52426 | 0.68075 | 4.39% |
| | | | Every 20,000 km (0.9%) | 1.18973 | 0.30125 | 0.44315 | 0.64471 | 4.16% |
| | | | Every 10,000 km (1.8%) | 1.17398 | 0.15583 | 0.21898 | 0.51626 | 3.33% |
| Case 3 | | | | No.2 | No.3 | No.7 | | |
| SBC-RBFNN | No.1, No.4, No.5, and No.6 |  | First 30% of the data | 4.76323 | 4.87612 | 4.85396 | 4.83110 | 38.11% |
| | | | First 50% of the data | 1.97669 | 2.17609 | 1.91992 | 2.02423 | 15.97% |
| mRFR | No.5 | No.1, No.4, and No.6 | First 5 data (0.8%) | 0.83303 | 0.85655 | 0.83189 | 0.84049 | 6.63% |
| | | | Every 20,000 km (0.8%) | 0.73653 | 0.78523 | 0.64587 | 0.72254 | 5.70% |
| | | | Every 10,000 km (1.7%) | 0.64041 | 0.65612 | 0.61877 | 0.63843 | 5.04% |

machine learning method will be extended to consider other factors such as temperature in our future work.

APPENDIX

For demonstration purposes, cells No. 1, 3, 7, and 8 in the laboratory dataset were used to train the predictive model to compare different IC features.

Single point features: The obtained relationships between the IC values of four POIs and battery SOH are shown in Fig. 14(a). It can be observed that the variation trends of the features do not follow the same pattern. Specifically, the peak point A and the valley point D disappear when the SOH drops to 0.9, the peak point B is almost invisible until the battery SOH falls to 0.92, while the peak point C covers the entire battery life with the SOH ranging from 1 to 0.8. Hence, only

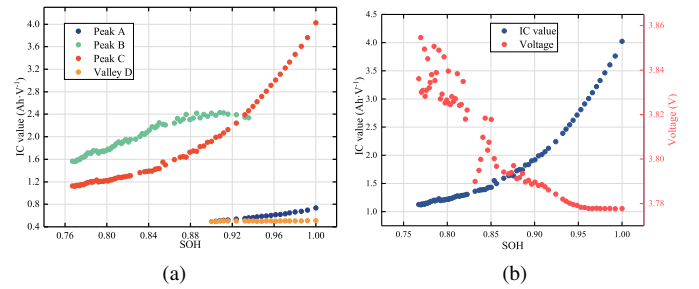


Fig. 14. Single point features. (a) IC value versus battery SOH relationship of the four points of interest. (b) IC value and voltage versus battery SOH relationships of the peak point C.

TABLE V
COMPARISON OF RMSES BETWEEN OFFLINE MODELS AND ONLINE ADAPTIVE MODELS

| Data resource | SBC-RBFNN (Offline Model) | | | | | SBC-RBFNN (Online Adaptive Model) | | | | | | |
|---------------|---------------------------|---------|---------|---------|----------------|-----------------------------------|---------|---------|---------|--------|----------------|------------------|
| | Case 1 | Case 2 | Case 3 | Case 4 | Average | REP _A | Case 1 | Case 2 | Case 3 | Case 4 | Average | REP _A |
| Laboratory | 0.00506 | 0.00661 | 0.00493 | 0.00482 | 0.00536 | 2.56% | 0.00475 | 0.00537 | 0.00373 | 0.0043 | 0.00454 | 2.18% |
| Real-world | 0.70880 | 1.08944 | 0.84616 | — | 0.88147 | 6.26% | 0.48400 | 0.51626 | 0.63843 | — | 0.54623 | 3.94% |

TABLE VI
RMSES OF PREDICTED SOH BASED ON THE SINGLE POINT FEATURE USING DIFFERENT TRAINING CELLS

| Training Cell | No.1 | No.2 | No.3 | No.4 | No.5 | No.6 | No.7 | No.8 | Average |
|---------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| No.1 | — | 0.00469 | 0.00685 | 0.00475 | 0.00671 | 0.00518 | 0.01348 | 0.01161 | 0.00761 |
| No.3 | 0.00688 | 0.00496 | — | 0.00601 | 0.00613 | 0.00399 | 0.00956 | 0.00688 | 0.00634 |
| No.7 | 0.01579 | 0.01172 | 0.01045 | 0.01249 | 0.00979 | 0.00789 | — | 0.00568 | 0.01054 |
| No.8 | 0.01161 | 0.00828 | 0.00689 | 0.00967 | 0.00745 | 0.00539 | 0.00431 | — | 0.00766 |

TABLE VII
CORRELATION COEFFICIENTS FOR THE EIGHT CELLS BASED ON MULTIPOINT FEATURES

| Point Cell | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 |
|---------------|----------------|---------|---------|---------------|---------------|---------------|---------------|--------|--------|--------|--------|
| No.1 | -0.9499 | -0.9008 | 0.0616 | 0.9876 | 0.9299 | 0.9854 | 0.9464 | 0.9064 | 0.8848 | 0.7957 | 0.7305 |
| No.2 | -0.9103 | -0.8082 | 0.1849 | 0.9854 | 0.9222 | 0.9794 | 0.9567 | 0.9186 | 0.8904 | 0.8190 | 0.7652 |
| No.3 | -0.9538 | -0.9124 | 0.0290 | 0.9886 | 0.9342 | 0.9824 | 0.9495 | 0.9149 | 0.8963 | 0.8331 | 0.7877 |
| No.4 | -0.9498 | -0.9582 | -0.4031 | 0.9826 | 0.9532 | 0.9848 | 0.9477 | 0.8893 | 0.8787 | 0.6781 | 0.5148 |
| No.5 | -0.8206 | -0.8091 | -0.4262 | 0.9240 | 0.9265 | 0.9658 | 0.9364 | 0.8535 | 0.8123 | 0.6293 | 0.5219 |
| No.6 | -0.9302 | -0.9502 | -0.5196 | 0.9766 | 0.9560 | 0.9768 | 0.9352 | 0.8548 | 0.8706 | 0.6444 | 0.4941 |
| No.7 | -0.9414 | -0.9292 | -0.0012 | 0.9840 | 0.9293 | 0.9713 | 0.9296 | 0.8816 | 0.8925 | 0.8270 | 0.8184 |
| No.8 | -0.9447 | -0.8799 | 0.3092 | 0.9846 | 0.9209 | 0.9800 | 0.9500 | 0.9178 | 0.9014 | 0.8513 | 0.8250 |
| Average | -0.9251 | -0.8935 | -0.0957 | 0.9767 | 0.9340 | 0.9782 | 0.9440 | 0.8921 | 0.8784 | 0.7597 | 0.6822 |

TABLE VIII
RMSES OF PREDICTED SOH BASED ON THE MULTIPOINT FEATURE USING DIFFERENT TRAINING CELLS

| Training Cell | No.1 | No.2 | No.3 | No.4 | No.5 | No.6 | No.7 | No.8 | Average |
|---------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| No.1 | — | 0.00972 | 0.00715 | 0.00808 | 0.01459 | 0.00692 | 0.01287 | 0.01048 | 0.00997 |
| No.3 | 0.01870 | 0.01757 | — | 0.02421 | 0.01235 | 0.01745 | 0.01004 | 0.00831 | 0.01552 |
| No.7 | 0.02659 | 0.02074 | 0.01604 | 0.02301 | 0.01125 | 0.01506 | — | 0.01122 | 0.01770 |
| No.8 | 0.01574 | 0.01160 | 0.00834 | 0.01505 | 0.00557 | 0.00801 | 0.00726 | — | 0.01022 |

TABLE IX
RMSES OF PREDICTED SOH BASED ON AREA FEATURE USING DIFFERENT TRAINING CELLS WITH DIFFERENT VOLTAGE INTERVALS

| Training Cell | 20 mV | 50 mV | 100 mV | 150 mV | 200 mV |
|---------------|----------------|---------|---------|---------|---------|
| No.1 | 0.00756 | 0.00782 | 0.00844 | 0.00916 | 0.00996 |
| No.3 | 0.00641 | 0.00667 | 0.00817 | 0.00786 | 0.00823 |
| No.7 | 0.01058 | 0.01094 | 0.01214 | 0.01318 | 0.01825 |
| No.8 | 0.00765 | 0.00795 | 0.00867 | 0.00949 | 0.01136 |
| Average | 0.00805 | 0.00835 | 0.00936 | 0.00992 | 0.01195 |

the peak point C is qualified for battery health indication. For comparison, the IC values and the voltages of peak point C are plotted versus the SOH in Fig. 14(b). It is clearly shown that the IC value bears a direct relation to the SOH, and the corresponding mathematical description can be much simpler than that of the voltage versus SOH relationship. Here, by trial and error, a sixth-degree polynomial function is found

sufficiently accurate to represent the IC values versus SOH relationship as shown in Fig. 14(b), and it will be used as our SOH predictive model. The calculated RMSEs are given in Table VI. It shows that the average RMSE is below 1% when the model is trained with cell No. 1, No. 3, or No. 8, while it is slightly over 1% when cell No. 7 is used for model training. The accuracy is generally acceptable for battery SOH estimation. The results also show that, although these battery cells are of the same specification and were cycled under the same condition, significant inconsistency was developed in terms of the decreasing trend of the peak points on the IC curves. For example, when cell No. 1 is used for model training, the RMSE of cell No. 7 is about 1.5 times higher than that of cell No. 6, indicating that there are considerable differences between them.

Multipoint feature: Based on the calculated correlation coefficients summarized in Table VII, the five points with the strongest correlations (highlighted with bold font) are selected

as battery aging features. Next, the IC values of feature points are used as the input of an ANN model for predicting the battery SOH. Such a method has shown to be effective in the previous study [42], while its high accuracy is also verified here with the calculated RMSEs given in Table VIII.

Area feature: Based on the suggestion given in [20], five voltage intervals five voltage intervals centered at the peak point C are selected for extracting the IC area feature. The interval lengths are 20, 50, 100, 150, and 200 mV, respectively. Again, a sixth-degree polynomial is chosen by trial and error as the fitting function. Table IX lists the predictive RMSEs of the five voltage intervals, where it can be seen that there is a reduction in the RMSE as the voltage interval becomes narrower.

REFERENCES

- [1] A. Manthiram, "A reflection on lithium-ion battery cathode chemistry," *Nature Commun.*, vol. 11, no. 1, p. 1550, 2020.
- [2] Y. Li, M. Vilathgamuwa, S. S. Choi, B. Xiong, J. Tang, Y. Su, and Y. Wang, "Design of minimum cost degradation-conscious lithium-ion battery energy storage system to achieve renewable power dispatchability," *Appl. Energy*, vol. 260, p. 114282, 2020.
- [3] S.-L. Wang, W. Tang, C. Fernandez, C.-M. Yu, C.-Y. Zou, and X.-Q. Zhang, "A novel endurance prediction method of series connected lithium-ion batteries based on the voltage change rate and iterative calculation," *J. Cleaner Prod.*, vol. 210, pp. 43–54, Feb. 2019.
- [4] A. Guha and A. Patra, "State of health estimation of lithium-ion batteries using capacity fade and internal resistance growth models," *IEEE Trans. Transp. Electrification*, vol. 4, no. 1, pp. 135–146, 2018.
- [5] Y. Li, Z. Wei, B. Xiong, and D. M. Vilathgamuwa, "Adaptive ensemble-based electrochemical-thermal-degradation state estimation of lithium-ion batteries," *IEEE Trans. Ind. Electron.*, 2021.
- [6] C. Zou, C. Manzie, D. Nešić, and A. G. Kallapur, "Multi-time-scale observer design for state-of-charge and state-of-health of a lithium-ion battery," *J. Power Sources*, vol. 335, pp. 121–130, 2016.
- [7] Z. Chen, C. C. Mi, Y. Fu, J. Xu, and X. Gong, "Online battery state of health estimation based on genetic algorithm for electric and hybrid vehicle applications," *J. Power Sources*, vol. 240, pp. 184–192, Oct. 2013.
- [8] Z. Chen, Q. Xue, R. Xiao, Y. Liu, and J. Shen, "State of health estimation for lithium-ion batteries based on fusion of autoregressive moving average model and Elman neural network," *IEEE Access*, vol. 7, pp. 102 662–102 678, 2019.
- [9] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, "Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods," *J. Power Sources*, vol. 239, pp. 680–688, Oct. 2013.
- [10] R. R. Richardson, C. R. Birkl, M. A. Osborne, and D. A. Howey, "Gaussian process regression for *insitu* capacity estimation of lithium-ion batteries," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 127–138, Jan. 2019.
- [11] G.-W. You, S. Park, and D. Oh, "Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach," *Appl. Energy*, vol. 176, pp. 92–103, Aug. 2016.
- [12] C. She, Z. Wang, F. Sun, P. Liu, and L. Zhang, "Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3345–3354, May 2020.
- [13] G. Crocioni, D. Pau, J.-M. Delorme, and G. Gruosso, "Li-ion batteries parameter estimation with tiny neural networks embedded on intelligent iot microcontrollers," *IEEE Access*, vol. 8, pp. 122 135–122 146, 2020.
- [14] M. Dubarry, B. Y. Liaw, M.-S. Chen, S.-S. Chyan, K.-C. Han, W.-T. Sie, and S.-H. Wu, "Identifying battery aging mechanisms in large format Li ion cells," *J. Power Sources*, vol. 196, no. 7, pp. 3420–3425, Apr. 2011.
- [15] M. Bercibar, I. Gandiaga, I. Villarreal, N. Omar, J. V. Mierlo, and P. V. den Bossche, "Critical review of state of health estimation methods of Li-ion batteries for real applications," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 572–587, Apr. 2016.
- [16] Y. Li, M. Abdel-Monem, R. Gopalakrishnan, M. Bercibar, E. Nanini-Maury, N. Omar, P. van den Bossche, and J. V. Mierlo, "A quick on-line state of health estimation method for Li-ion battery with incremental capacity curves processed by Gaussian filter," *J. Power Sources*, vol. 373, pp. 40–53, Jan. 2018.
- [17] X. Li, J. Jiang, L. Y. Wang, D. Chen, Y. Zhang, and C. Zhang, "A capacity model based on charging process for state of health estimation of lithium ion batteries," *Appl. Energy*, vol. 177, pp. 537–543, Sep. 2016.
- [18] Y. Gao, J. Jiang, C. Zhang, W. Zhang, Z. Ma, and Y. Jiang, "Lithium-ion battery aging mechanisms and life model under different charging stresses," *J. Power Sources*, vol. 356, pp. 103–114, Jul. 2017.
- [19] X. Li, Z. Wang, and J. Yan, "Prognostic health condition for lithium battery using the partial incremental capacity and Gaussian process regression," *J. Power Sources*, vol. 421, pp. 56–67, May 2019.
- [20] X. Tang, C. Zou, K. Yao, G. Chen, B. Liu, Z. He, and F. Gao, "A fast estimation algorithm for lithium-ion battery state of health," *J. Power Sources*, vol. 396, pp. 453–458, Aug. 2018.
- [21] M. Dubarry, C. Truchot, and B. Y. Liaw, "Cell degradation in commercial LiFePO₄ cells with high-power and high-energy designs," *J. Power Sources*, vol. 258, pp. 408–419, Jul. 2014.
- [22] Y. Jiang, J. Jiang, C. Zhang, W. Zhang, Y. Gao, and Q. Guo, "Recognition of battery aging variations for LiFePO₄ batteries in 2nd use applications combining incremental capacity analysis and statistical approaches," *J. Power Sources*, vol. 360, pp. 180–188, Aug. 2017.
- [23] D. Ansean, V. M. Garcia, M. Gonzalez, C. Blanco-Viejo, J. C. Viera, Y. F. Pulido, and L. Sanchez, "Lithium-ion battery degradation indicators via incremental capacity analysis," *IEEE Trans. Ind. Appl.*, vol. 55, no. 3, pp. 2992–3002, May 2019.
- [24] J. Lu and F. Gao, "Model migration with inclusive similarity for development of a new process model," *Ind. Eng. Chem. Res.*, vol. 47, no. 23, pp. 9508–9516, Dec. 2008.
- [25] X. Tang, C. Zou, K. Yao, J. Lu, Y. Xia, and F. Gao, "Aging trajectory prediction for lithium-ion batteries via model migration and bayesian monte carlo method," *Appl. Energy*, vol. 254, p. 113591, Nov. 2019.
- [26] X. Tang, K. Liu, X. Wang, F. Gao, J. Macro, and W. D. Widanage, "Model migration neural network for predicting battery aging trajectories," *IEEE Trans. Transport. Electrification*, vol. 6, no. 2, pp. 363–374, Jun. 2020.
- [27] C. Birkl, "Diagnosis and prognosis of degradation in lithium-ion batteries," Ph.D. dissertation, University of Oxford, 2017.
- [28] J. K. Ko, K. M. Wiaderek, N. Pereira, T. L. Kinnibrugh, J. R. Kim, P. J. Chupas, K. W. Chapman, and G. G. Amatucci, "Transport, phase reactions, and hysteresis of iron fluoride and oxyfluoride conversion electrode materials for lithium batteries," *ACS Appl. Mat. Interfaces*, vol. 6, no. 14, pp. 10 858–10 869, 2014.
- [29] R. Malik, F. Zhou, and G. Ceder, "Kinetics of non-equilibrium lithium incorporation in LiFePO₄," *Nature Mat.*, vol. 10, no. 8, pp. 587–590, 2011.
- [30] A. Fly and R. Chen, "Rate dependency of incremental capacity analysis (dQ/dV) as a diagnostic tool for lithium-ion batteries," *J. Energy Storage*, vol. 29, p. 101329, Jun. 2020.
- [31] I. Bloom, A. N. Jansen, D. P. Abraham, J. Knuth, S. A. Jones, V. S. Battaglia, and G. L. Henriksen, "Differential voltage analyses of high-power, lithium-ion cells," *J. Power Sources*, vol. 139, no. 1-2, pp. 295–303, Jan. 2005.
- [32] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Comput.*, vol. 1, no. 2, pp. 281–294, Jun. 1989.
- [33] M. Kamal and D. Yu, "Model-based fault detection for proton exchange membrane fuel cell systems," *Int. J. Eng. Sci. Technol.*, vol. 3, no. 9, pp. 1–15, 2011.
- [34] Y. Li, C. Zou, M. Bercibar, E. Nanini-Maury, J. C.-W. Chan, P. van den Bossche, J. Van Mierlo, and N. Omar, "Random forest regression for online capacity estimation of lithium-ion batteries," *Appl. Energy*, vol. 232, pp. 197–210, 2018.
- [35] Z. Ding, Z. Chen, T. Ma, C.-T. Lu, W. Ma, and L. Shaw, "Predicting the hydrogen release ability of LiBH₄-based mixtures by ensemble machine learning," *Energy Storage Mater.*, vol. 27, pp. 466–477, May 2020.
- [36] T. K. Ho, "Random decision forests," in *Proc. Int. Conf. Document Analysis Recog.* IEEE Comput. Soc. Press, 14–16 Aug. 1995.
- [37] M. J. Er, S. Wu, J. Lu, and H. L. Toh, "Face recognition with radial basis function (RBF) neural networks," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 697–710, May 2002.
- [38] J.-L. Yuan and T. Fine, "Neural-network design for small training sets of high dimension," *IEEE Trans. Neural Netw.*, vol. 9, no. 2, pp. 266–280, Mar. 1998.
- [39] M. J. L. Orr, "Regularization in the selection of radial basis function centers," *Neural Computat.*, vol. 7, no. 3, pp. 606–623, May 1995.

- [40] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *ArXiv Preprint*, 2015.
- [41] X. Zhang, Y. Zou, and W. Shi, "Dilated convolution neural network with LeakyReLU for environmental sound classification," in *2017 22nd Int. Conf. Digital Signal Process. (DSP)*. London, UK: IEEE, Aug. 2017, pp. 1–5.
- [42] S. Zhang, B. Zhai, X. Guo, K. Wang, N. Peng, and X. Zhang, "Synchronous estimation of state of health and remaining useful lifetime for lithium-ion battery using the incremental capacity and artificial neural networks," *J. Energy Storage*, vol. 26, p. 100951, Dec. 2019.



Chengqi She received the B.S. degree in vehicle engineering from the Hunan University, Changsha, Hunan, China, in 2016. He is currently working toward the Ph.D. degree in mechanical engineering with the Beijing Institute of Technology, Beijing, China. His research interests include modeling and state of health estimation of lithium-ion batteries based on real-world big data.



Yang Li (Member, IEEE) received the B.E. degree in electrical engineering from Wuhan University, Wuhan, China, in 2007, and the M.Sc. and Ph.D. degrees in power engineering from Nanyang Technological University (NTU), Singapore, in 2008 and 2015, respectively. From 2015 to 2018, he was a Research Fellow with the Energy Research Institute, NTU and the School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, QLD, Australia. Since 2019, he has been with the School of Automation, Wuhan University of Technology, Wuhan. He is currently a Researcher with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden. His research interests include modeling and control of energy storage systems in power systems and transport sector.

Dr. Li was a recipient of the EU Marie Skłodowska-Curie Action Individual Fellowship in 2020.



Changfu Zou (Member, IEEE) received the Ph.D. degree in automation and control engineering from the Department of Mechanical Engineering, University of Melbourne, VIC, Australia, in 2017. He was a Postdoctoral Researcher with the Chalmers University of Technology, Gothenburg, Sweden, where he is currently an Assistant Professor with the Automatic Control Group, Department of Electrical Engineering.

His current research interests include modeling and control of energy storage systems for automotive and power grid applications.



Torsten Wik (Member, IEEE) received the M.Sc. degree in chemical engineering (major in applied mathematics), the Licentiate of Engineering degree in control engineering, the Ph.D. degree in environmental sciences (major in automatic control), and the Docent degree in electrical engineering from the Chalmers University of Technology, Gothenburg, Sweden, in 1994, 1996, 1999, and 2004, respectively. From 2005 to 2007, he was a Senior Researcher with Volvo Technology, Gothenburg, researching on control system design for combustion engine test cells, and combined reformer and fuel cells. He is a Professor and the Head of the Automatic Control Group, Department of Electrical Engineering, Chalmers University of Technology. His current research interests include optimal control, model reduction, and systems with model uncertainties, with applications to energy storage, environmental, and biological systems.



Zhenpo Wang received the Ph.D. degree in automotive engineering from the Beijing Institute of Technology, Beijing, China, in 2005. He is currently a Professor with the School of Mechanical Engineering, Beijing Institute of Technology, and the Director of National Engineering Laboratory for Electric Vehicles. He has authored or coauthored four monographs and translated books as well as more than 80 technical papers. He also holds more than 60 patents. His current research interests include pure electric vehicle integration, packaging

and energy management of battery systems, and charging station design. Prof. Wang is the recipient of numerous awards including the Second National Prize for Progress in Science and Technology and the First Prize for Progress in Science and Technology from the Ministry of Education, China, and the Second Prize for Progress in Science and Technology from Beijing Municipal, China.



Fengchun Sun received the M.S.E. and Ph.D. degrees both in vehicle engineering from the Beijing Institute of Technology, Beijing, China, in 1984 and 1989, respectively. He studied at the Technical University of Berlin, Berlin, Germany as a joint Ph.D. degree student, from 1987 to 1989.

He is currently an Academician of the Chinese Academy of Engineering, Beijing. He has been conferred the title of Cheung Kong Scholar by the Ministry of Education, China. He has published more than 200 papers and is the holder of 19 patents.

His research interests include electric vehicles, electric drive systems, electric vehicle demonstration, and infrastructure.

Dr. Sun is the recipient of the second prize from the Chinese National Science and Technology Progress Awards, in 2008, the second prize from the Chinese National Technological Innovation Awards, in 2004 and 2009, and the award for industrial innovation from the Ho Leung Ho Lee Foundation, in 2007.