Research papers

Robust State of Health estimation of lithium-ion batteries using convolutional neural network and random forest☆

Niankai Yang [a], Ziyou Song [a,*], Heath Hofmann [b], Jing Sun [a]

[a] *Department of Naval Architecture and Marine Engineering, University of Michigan, Ann Arbor, MI 48109, USA*
[b] *Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA*

ARTICLE INFO

ABSTRACT

The State of Health (SOH) of lithium-ion batteries is directly related to their safety and efficiency, yet effective assessment of SOH remains challenging for real-world applications. In this paper, the estimation of SOH (i.e., capacity fading) under partial discharge with different initial and final State of Charge (SOC) levels is investigated. The challenge lies in the fact that partial discharge causes the truncation of the data available for SOH estimation, thereby leading to the loss or distortion of common SOH indicators. To address this challenge, we utilize the convolutional neural network (CNN) to extract indicators for both SOH and changes in SOH (ΔSOH) between two successive charge/discharge cycles. The random forest algorithm is then adopted to produce the final SOH estimate by exploiting the indicators from the CNNs. Performance evaluation is conducted using the partial discharge data with different SOC ranges created from a fast-discharging dataset. The proposed approach is compared with (i) a differential-analysis-based approach and (ii) two CNN-based approaches using only SOH and ΔSOH indicators, respectively. Through comparison, the proposed approach demonstrates improved estimation accuracy and robustness. Sensitivity analysis of the CNN and random forest models further validates that the proposed approach makes better use of the available partial discharge data.

1. Introduction

Lithium-ion (Li-ion) batteries have been well established as an effective energy storage technology for various applications due to their low self-discharge rate, high energy density, and falling cost [1, 2]. To maintain safe and reliable operation, an accurate and robust battery State of Health (SOH) estimation is of critical importance. Generally, the battery SOH is characterized by either its capacity or its internal resistance [3]. Compared to resistance, capacity is a more direct indicator for SOH, as it represents the energy storage capability of a battery [4]. Therefore, the estimation of capacity fading is the focus of most SOH monitoring works. To assess the SOH of a battery, one can completely discharge a fully charged battery and compute its capacity using Coulomb counting [5]. However, fully discharging or charging the battery for SOH estimation may not be feasible in some applications (e.g., electric vehicle), and can accelerate battery degradation [6]. Therefore, for general applications, SOH estimation is performed using battery partial charge or discharge operating data [7].

To estimate SOH using partial charge or discharge data, a mapping from the battery operating data (e.g., current, voltage, and temperature over time) to SOH needs to be constructed. However, given its high dimensionality [8], the direct construction of this mapping requires a large amount of data. Without a comprehensive set of data, this mapping can be established with a two-step procedure. First, SOH indicators (e.g., parameters in physics-based models or signatures in experimental data) are identified from charge or discharge data based on expert knowledge. Then, a low-dimensional mapping is developed between these indicators and SOH [9].

Equivalent circuit models (ECMs) were used with different state estimation techniques such as least squares [10], Kalman filters [11], and particle filters [12] to obtain open circuit voltage (OCV) versus State of Charge (SOC) curves. Then, SOH was inferred based on the relationship between OCV, SOC, and SOH. ECM-based approaches provide a computationally-efficient SOH estimation under different battery operating conditions (e.g., partial or pulse charge/discharge [13]). However, given the complex electrochemical nature of batteries, the oversimplified ECMs can have significant unmodeled dynamics, leading to degraded estimation accuracy. In addition, ECM-based approaches have fundamental difficulties extracting SOH indicators from the flat region of OCV–SOC curves, i.e., the mid-SOC range [14].

To better capture the electrochemical mechanisms in batteries, electrochemical models were adopted for extracting SOH indicators [15, 16]. Various physics-informed model parameters, such as solid electrolyte interphase growth and solid state diffusion coefficient, can be fitted from the battery operating data using the electrochemical models. SOH is then tracked based on its correlation with these model parameters [17,18]. The intensive computation of electrochemical-model-based approaches, however, makes them infeasible for most real-time applications [19].

For a trade-off between computational complexity and model accuracy, differential analysis (DA) was applied to identify SOH indicators. For example, incremental capacity analysis (ICA), which differentiates the battery charge capacity versus the terminal voltage, was utilized in [20,21]. In [22], SOH indicators were extracted by computing the voltage change per unit charge capacity change (i.e., differential voltage analysis). The ratio of the temperature and voltage differentials (i.e., differential thermal voltammetry) was leveraged in [23] for SOH indicator extraction. Data-driven DA-based approaches can extract SOH indicators better representing the true battery condition. However, due to the use of differential operations, data smoothing is required before applying DA, which could lead to distortion or loss of information. Moreover, SOH indicators from DA can be mostly located near the high or low SOC range, thereby limiting its effectiveness under partial charge or discharge.

Considering the aforementioned challenges in extracting SOH indicators and the increasing data availability, deep-learning-based approaches offer a promising alternative to the above methods. By leveraging the strong learning ability of neural networks (NNs), deep-learning-based approaches can directly approximate the mapping from charge or discharge data to SOH (i.e., end-to-end SOH estimation) without relying on expert knowledge. For instance, a vanilla recurrent NN (RNN) was adopted to estimate SOH from the battery charge and discharge data in [24]. A variant of vanilla RNN, a long short-term memory network, was utilized in [25] to capture the long-term dependency of SOH on the past and present battery operating conditions. In [26], a convolutional NN (CNN) was applied to handle the SOH estimation under partial charge. However, one drawback for the deep-learning-based approaches that perform end-to-end SOH estimation is the lack of interpretability, as there is no explicit SOH indicator extraction procedure. In addition, it is intrinsically difficult to incorporate expert knowledge into the estimation procedure for these approaches.

This paper proposes a deep-learning-based SOH estimation algorithm for a single battery cell under partial discharge. The proposed approach adopts a two-step procedure for SOH estimation to facilitate better interpretability and the incorporation of expert knowledge. In the first step, motivated by the work in [8,26], CNNs are utilized to handle SOH indicator extraction under partial discharge. Two CNNs are established to extract the indicators related to the SOH and the change of SOH between two successive discharge cycles (ΔSOH). In the second step, considering the potential correlation among indicators extracted from two CNNs, random forest [27], which uses an ensemble of decision trees to capture an input–output relationship, is adopted to produce the final SOH estimate based on the extracted indicators. To validate the effectiveness of the proposed approach, it is compared with the following approaches: (i) a DA-based approach using ICA for SOH indicator extraction, and (ii) two CNN-based approaches using SOH and ΔSOH indicators, respectively, for SOH estimation. Based on a fast-discharging dataset provided in [28], it has been demonstrated that the proposed approach can improve both estimation accuracy and robustness.

The contribution of this paper is as follows.

- It presents a two-step procedure for SOH estimation that uses CNN for indicator extraction and random forest for indicator fusion. This framework facilitates the exploitation of both the learning ability of NNs and physics-based knowledge for SOH estimation.

- Sensitivity analysis of CNN and random forest models is performed to interpret the data-driven models and demonstrate the effectiveness of the proposed algorithm.
- A novel perspective for SOH estimation based on ΔSOH is proposed, which can enhance the SOH detectability under partial charge/discharge by leveraging the indicators complementary to the SOH-related indicators.

The remainder of this paper is organized as follows: Section 2 introduces the procedure for partial discharge data collection and formulates the SOH estimation problem under partial discharge. The limitation of the DA-based approach under partial discharge is shown in Section 3. In Section 4, two CNNs extracting indicators related to SOH and ΔSOH are first developed and examined. Based on the sensitivity analysis of two CNNs, a random forest model is then developed to fuse the outputs from the CNNs for enhanced SOH estimation performance. In Section 5, the proposed approach is verified under partial discharge with different initial and final SOC levels. Conclusions and future work are provided in Section 6.

2. Background

In this section, the data to be used for developing and validating the proposed approach is first presented. Then, the SOH estimation problem under partial discharge is formulated. In this study, an open-source dataset provided in [28], which contains aging data for 124 commercial batteries, is used to demonstrate the concept and validate the approach. The batteries under investigation are lithium–iron-phosphate/graphite cells manufactured by A123 Systems (model APR18650M1A). The nominal capacity of the cells is 1.1 Ah, and the lower and upper cut-off voltages are 2.0 V and 3.6 V, respectively. The battery is cycled in a temperature-controlled environmental chamber of $30\,°C$ by charging and discharging the cell repeatedly. A varied charging rate and a constant discharging rate are adopted for cycling. Since the focus of this work is to address the SOH estimation under incomplete charge and discharge (instead of different C rates), the discharge data will be used for the subsequent algorithm development and evaluation.

The cells are discharged from a fully charged state with a constant current–constant voltage (CC–CV) policy. In particular, the cell is first discharged at 4 C until its terminal voltage reaches the low cutoff voltage (i.e., 2.0 V). Then, a constant voltage discharge is performed until the cell current reaches the cutoff current (i.e., 1/50 C). The aging data consists of the discharge capacity, voltage, and cell capacity of each discharge cycle, where the cell capacity is obtained as the maximum discharge capacity over the entire cycle. Based on the cell capacity, the SOH for each discharge cycle is defined as the ratio between the cell capacity and the nominal capacity. The full voltage-discharge capacity curves (i.e., full discharge curves) from one sample cell under different SOH levels are given in Fig. 1.

Using the full discharge curves, a partial discharge dataset is created by truncating the full discharge curves with different initial and final depth of discharge (DoD) values. The initial DoD (DoD_i) is determined randomly according to a Gaussian distribution, and the final DoD (DoD_f) is obtained as

$$DoD_f = DoD_i + \frac{Q^{max}}{C_{cell}}, \qquad (1)$$

where Q^{max} is a uniformly sampled maximum incremental discharge capacity, and C_{cell} is the cell capacity. Note that, $Q^{max} = 0$ implies that $DoD_f = DoD_i$, namely, no discharge. Then, the voltage (V) and incremental discharge capacity (Q) sequences from the initial to final DoD values are considered as the partial discharge curve for each cycle (see Fig. 2). The SOH estimation problem under partial discharge is formulated as estimating the SOH for the present discharge cycle based on the present and past partial discharge curves.
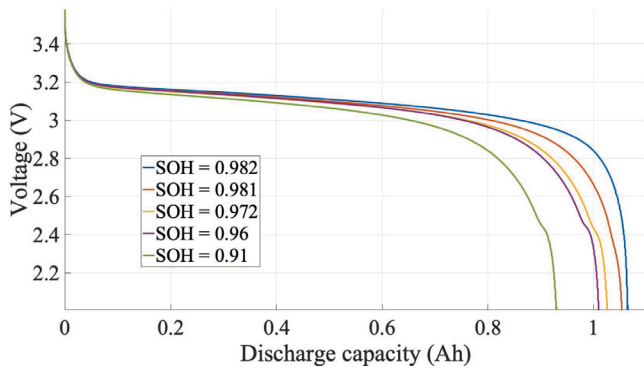
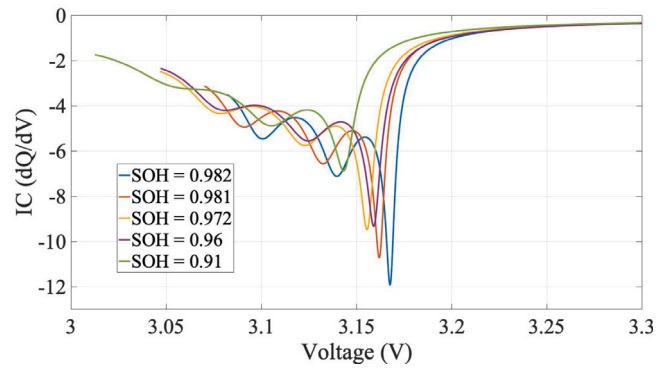Fig. 1. Full discharge curves under different SOH levels.

Fig. 2. Illustration of data truncation for creating partial discharge curve.

Fig. 3. IC curves under different SOH levels for the cell in Fig. 1 under partial discharge with the low DoD region.
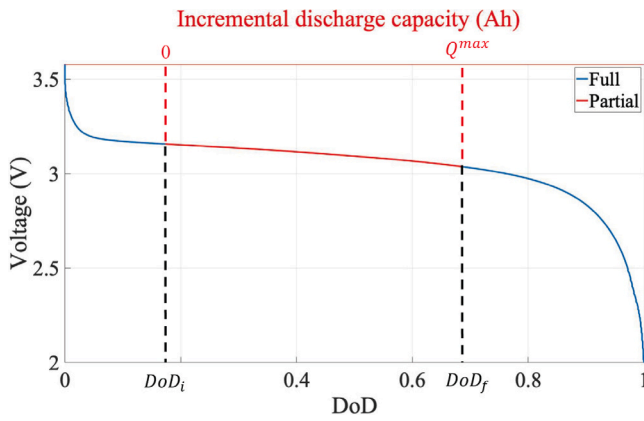

Fig. 4. The lowest IC minimum vs. SOH from five different cells under partial discharge with the low DoD region.

3. DA-based approaches for SOH estimation under partial discharge

In this section, a DA-based approach is applied to estimate SOH under partial discharge to illustrate the limitation of DA-based approaches and the characteristics of SOH indicators. The DA-based approach that uses ICA for indicator extraction is adopted given its demonstrated effectiveness for SOH estimation under partial charge/discharge [29–32]. Typical ICA takes the following steps to extract SOH indicators from the discharge curves. First, data pre-processing is performed to smooth out the measurement noise. In this study, the support vector regression (SVR) method proposed in [21,30] is utilized, which fits the discharge curves (i.e., V–Q curves) for smoothing. The choice of SVR is based on the finding in [21,30] that SVR can effectively remove noise and has minimal information loss under partial charge/discharge. With the smoothed discharge curves, the IC values are then computed as the gradient of V with respect to Q (i.e., $\Delta Q/\Delta V$). Finally, by examining the IC-voltage curves, salient features can be extracted as the SOH indicators.

Considering that the IC features are mostly in the high SOC (i.e., low DoD) region [21], a partial discharge dataset with $DoD_i = 0$ and $Q^{max} \sim \mathcal{U}(0.65, 0.75)$ is created, where $\mathcal{U}(\underline{Q}^{max}, \bar{Q}^{max})$ denotes a uniform distribution with $\underline{Q}^{max}$ and $\bar{Q}^{max}$ being the lower and upper bounds, respectively. Note that the distribution for sampling Q^{max} is chosen such that ICA can effectively extract SOH indicators from the partial discharge curves. By applying SVR and differentiation, the IC curves for the same cell in Fig. 1 are shown in Fig. 3. From Fig. 3, we have the following observations:

- There are multiple IC minima over the available voltage range, which correlate with SOH.
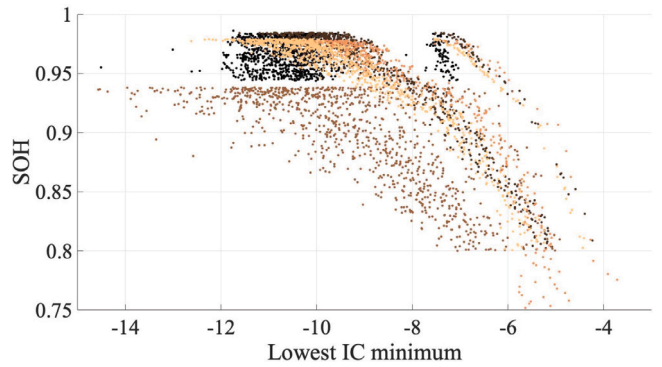
- The location of these IC minima shifts irregularly along the voltage axis under different SOH levels.
- Only the lowest IC minimum remains identifiable under different SOH levels.

The above observations indicate that, under partial discharge, the SOH estimation performance of the DA-based approach that uses ICA heavily relies on the existence and the quality of the lowest IC minimum.

To examine the quality of the lowest IC minimum for different cells, we plot the lowest IC minimum of five different cells and its corresponding SOH levels in Fig. 4, where the dots with the same color indicate the data from the same cell. It can be seen from Fig. 4 that:

- The lowest IC minimum of the cell may not be consistently correlated with its SOH.
- The relationship between the lowest IC minimum and SOH can vary for different cells.

Since the correlation between the lowest IC minimum and SOH is corrupted under partial discharge and sensitive to the cell-to-cell variation, the DA-based approach can suffer from performance degradation when estimating the SOH.

To further illustrate the robustness issue with the DA-based approach under partial discharge, another partial discharge dataset with $DoD_i \sim \mathcal{N}(0.2, 1/900)$ and $Q^{max} \sim \mathcal{U}(0.45, 0.55)$ is created, where $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with μ and σ being the mean and standard deviation, respectively. The distributions for the sampled initial and final DoD values are shown in Fig. 5. This partial discharge setup represents a normal battery operating condition in a typical electric vehicle application. The IC curves from this dataset for the same cell in Fig. 3 is plotted in Fig. 6. In Fig. 6, we see that:
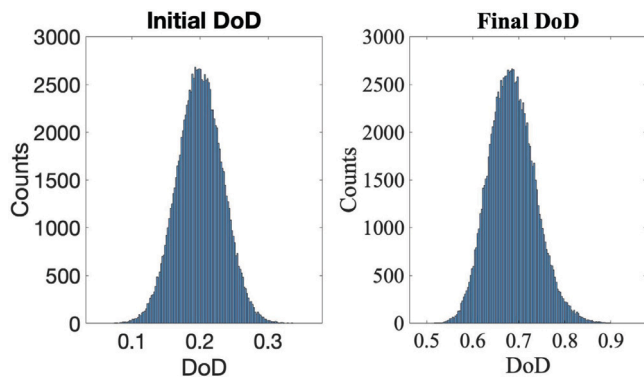
Fig. 5. Initial DoD and final DoD distributions for the partial discharge dataset without the low DoD region.
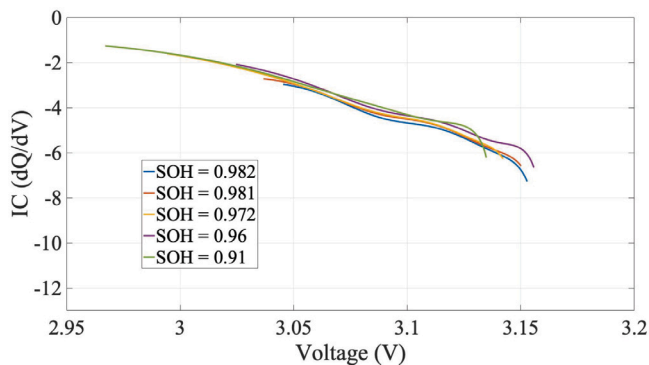

Fig. 6. IC curves under different SOH levels for the cell in Fig. 1 under partial discharge without the low DoD region.

- Due to the data truncation caused by partial discharge, the lowest IC minimum between 3.15 V and 3.2 V disappears.
- The IC minima located between 3.05 V and 3.15 V in Fig. 3 also disappear due to the data smoothing.

4. CNNs and random forest for SOH estimation under partial discharge

In this section, CNNs and random forest are used to address SOH estimation under partial discharge. First, we develop a CNN that directly estimates SOH from the discharge data and evaluate its performance under partial discharge without the low DoD region. In view of the low consistency between successive SOH estimates from the CNN for direct SOH estimation, a CNN for incremental SOH estimation is then proposed. Since both CNNs have pros and cons, a random forest model is then designed to fuse two CNNs for enhanced estimation performance.

4.1. Direct SOH estimation using CNN

In Section 3, it has been shown that DA-based approaches become ineffective when the expert knowledge (e.g., IC minima) is not applicable. In the absence of applicable expert knowledge, NN can be a promising solution, since it can automatically extract features from the data that correlate with the desired output. For SOH estimation, three types of NNs have been investigated, i.e., multilayer perception (MLP) [33], RNN [24,34], and CNN [26]. In order to select the best NN for our problem, we recall the observations in Section 3 that the IC minimum is a local property and may shift when the SOC range of the partial discharge curves differs. Consequently, CNN is chosen to extract locally-connected features whose locations in the input data
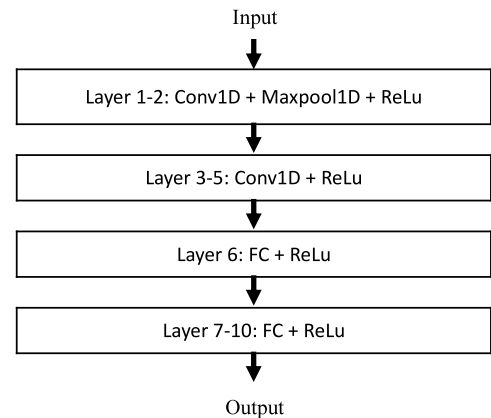

Fig. 7. Architecture of the proposed CNN.

Table 1
Hyperparameters of the proposed CNNs.

Hyperparameter	Value
Convolution filter size	3
Convolution filter stride	1
Number of convolution filters	50
Pooling size	3
Pooling stride	3
Layer 6 FC dimension	550
Layer 7-10 FC dimension	200

may shift [35]. The advantage of CNN over MLP and RNN lies in its convolution and pooling operations. The former enables the extraction of locally-connected features such as the regional minima in IC curves, and the latter can identify shifted features by taking the extreme value within a local region as its output.

We use a ten-layer CNN model with both convolutional and fully-connected (FC) layers (see Fig. 7). "Conv1D" and "Maxpool1D" denote the one-dimensional convolution and max-pooling operation. The hyperparameters of the CNN are listed in Table 1. The architecture and hyperparameters of the proposed CNN are designed based on the results presented in [26] and through trial and error. The output of the CNN is the SOH of the present discharge cycle. The CNN has two input channels. The first channel is the sequence of incremental discharge capacity of the present cycle, i.e., a vector of evenly spaced points between zero and the upper bound of the uniform distribution for Q^{max} ($\bar{Q}^{max}$). The second channel is the sequence of cell terminal voltage of the present cycle, i.e., a vector of terminal voltage values corresponding to the points in the incremental discharge capacity sequence. Since the discharge ends at Q_{max}, zero padding is performed in the terminal voltage sequence from Q_{max} to $\bar{Q}^{max}$. To train the CNN, we use 60% of the data for training, 20% for validation, and 20% for testing. Five-folds cross-validation is performed to remove the effect of the dataset partition on the estimation accuracy. For performance evaluation, the following definition of mean absolute error (MAE) between the true SOH (SOH_t) and the estimated SOH (SOH_e) is adopted:

$$MAE = \frac{|SOH_t - SOH_e|}{SOH_t} \times 100\%. \tag{2}$$

The CNN with the above setup (SOH-CNN) is applied to the same dataset used for DA-based approach, namely, with the initial and final DoD values located around 0.2 and 0.7 (see Fig. 5). In this study, the development of CNNs is performed in PyTorch [36] with the Adam optimizer. The average MAE from SOH-CNN on the testing data (i.e., 27 cells with cycle life ranging from approximately 500 to 2000) is 1.28%, and its representative SOH estimation performance is presented in Fig. 8. Fig. 8 shows that satisfactory SOH estimation performance can be achieved by SOH-CNN without explicitly incorporating the salient
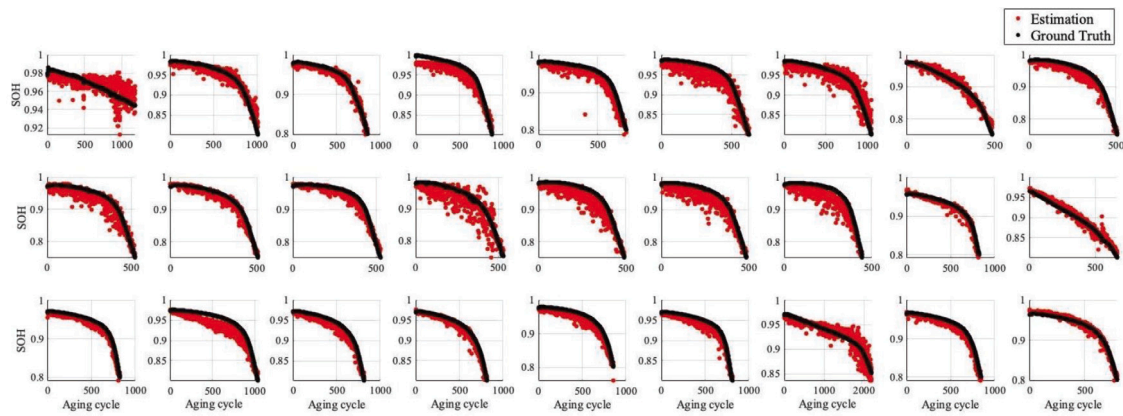
Fig. 8. Representative SOH estimation performance of SOH-CNN under partial discharge without the low DoD region (Each subplot contains the SOH trajectory of a different cell).

SOH indicator used in the DA-based approach (i.e., the lowest IC minimum). This is because the indicators that are numerically correlated with SOH, but may not be physically interpretable, can be extracted by SOH-CNN. Moreover, since no additional data smoothing is required, the potential information loss caused by data processing can be avoided. Despite its success in handling SOH estimation under partial discharge, SOH-CNN does not yield consistent SOH estimates in successive cycles, as shown in Fig. 8. This is due to the following two reasons. First, the information provided with the partial discharge curve may not be enough, thereby leading to degraded estimation accuracy. Second, only the partial discharge data of present cycle is used for SOH estimation. Consequently, the correlation between present and past SOH values cannot be captured.

4.2. Incremental SOH estimation using CNN

To enhance consistency amongst successive SOH estimates, we need an NN formulation that can better capture the accumulative and evolutionary nature of SOH based on the discharge data. Although RNNs can capture the dependency of SOH on battery operating history, they are ineffective in extracting locally-connected and shifted features from discharge curves, which can be important for characterizing battery degradation [28,37]. Therefore, a CNN that estimates the difference of SOH between the present and past SOH values (ΔSOH) is proposed in an attempt to capture the change in SOH over time. The number of layers and architecture of each layer of the CNN are the same as those of the SOH-CNN. The output of the CNN is ΔSOH. Since the change of SOH depends on the degradation history, the CNN is designed to use both the past and present partial discharge curves. Namely, the CNN additionally takes in the sequences of incremental discharge capacity and cell terminal voltage of the past cycle compared to the SOH-CNN. Note that we do not constrain the ΔSOH from the CNN to be strictly negative for a monotonically decreasing SOH trajectory, as ΔSOH can be positive due to various factors [28]. Based on the estimated ΔSOH from the CNN, the SOH estimate for the present cycle (SOH_e^t) can be computed as the sum of the SOH estimate for the past cycle (SOH_e^{t-1}) and the estimated ΔSOH for the present cycle (ΔSOH_e^t), i.e.,

$$SOH_e^t = SOH_e^{t-1} + \Delta SOH_e^t. \tag{3}$$

For the same dataset used in Section 4.1, the average MAE from using the CNN that outputs ΔSOH, i.e., ΔSOH-CNN, on the testing data is 1.57%, and representative SOH estimation results are shown in Fig. 9. Compared to the SOH estimation results in Fig. 8, ΔSOH-CNN improves the consistency among the successive SOH estimates. This is attributed to two reasons. First, as both the present and past discharge curves are used as inputs to ΔSOH-CNN, richer information is available for estimating SOH. More importantly, by formulating the output of the CNN as ΔSOH, ΔSOH-CNN could better learn the relationship between

successive SOHs by extracting ΔSOH-related features, leading to an improved consistency. On the downside, ΔSOH-CNN has a worse MAE than that of SOH-CNN, as the estimation errors will be accumulated over time like any integrator-based estimators. Furthermore, since the ΔSOH is typically much smaller than SOH, the ground truth ΔSOH used for CNN training is more likely to be corrupted by the noise in experimental data, which could affect the training and deteriorate the estimation accuracy. Finally, it should be pointed out that since ΔSOH depends on the degradation mechanism of a cell, the performance of ΔSOH-CNN can be more sensitive to the cell-to-cell variation, as is the case for the 1st and 18th cells in Fig. 9.

4.3. CNN and random forest based SOH estimator

Both SOH-CNN and ΔSOH-CNN can handle SOH estimation under partial discharge and outperform the DA-based approach. The SOH-CNN has a smaller MAE but a worse consistency in terms of successive SOH estimates compared to ΔSOH-CNN. Motivated by their complementary characteristics, we investigate if these two CNN models can supplement each other for SOH estimation. To this end, we analyze the CNN model sensitivity by computing the partial derivative of the output with respect to the inputs. In particular, considering that SOH is closely related to OCV, the sensitivity of the CNN output with respect to the input voltage curves are studied, as plotted in Fig. 10. It shows that SOH-CNN has larger partial derivative values in the first half of the present voltage curve under partial discharge (see Fig. 10(a)), while ΔSOH-CNN shows larger partial derivative values in the second half of the present and past voltage curves (see Fig. 10(b) and Fig. 10(c)). This observation motivates the use of an ensemble algorithm to combine these two CNN models for better performance.

In order to combine two CNN models for estimating SOH, we treat the SOH estimates from two CNNs as two SOH indicators. A regression model can then be developed to estimate SOH. Considering that the SOH estimates from CNNs are both produced based on the discharge curves, there can be strong correlation between these two SOH estimates. Therefore, the random forest algorithm is chosen to mitigate the potential effect caused by multicollinearity [38]. By combining the CNNs and random forest models, a random forest-CNN (RF-CNN) SOH estimator is proposed as shown in Fig. 11. In RF-CNN, two CNNs first take the partial discharge curves from the past and present cycles to produce two estimates of SOH. Then, the random forest model fuses two SOH estimates from the CNNs to provide the final estimate of SOH.

To develop the RF-CNN, the SOH-CNN trained in Section 4.1 and the ΔSOH-CNN trained in Section 4.2 can be directly used. In addition, a random forest model is developed using the statistics and machine learning toolbox in MATLAB based on the SOH estimates from the CNNs and the true SOH values. The random forest model contains 25 regression trees, and each tree uses interaction tests [39] to select
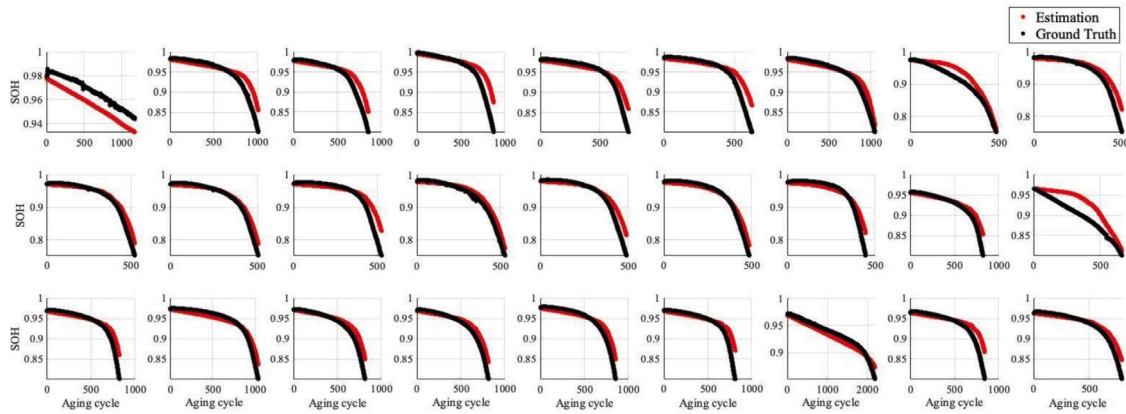
Fig. 9. Representative SOH estimation performance of ΔSOH-CNN under partial discharge without the low DoD region (Each subplot contains the SOH trajectory of a different cell).
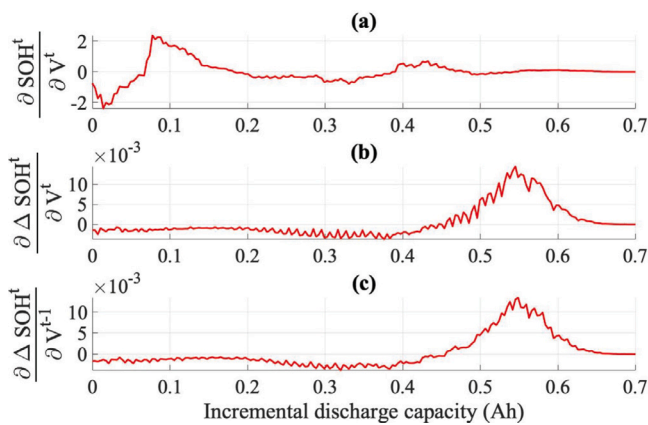

Fig. 10. Sensitivity of the output with respect to the input voltage curves in CNNs (V_t and V_{t-1} are the present and past voltage curves, (a) is for SOH-CNN, and (b) and (c) are for ΔSOH-CNN).
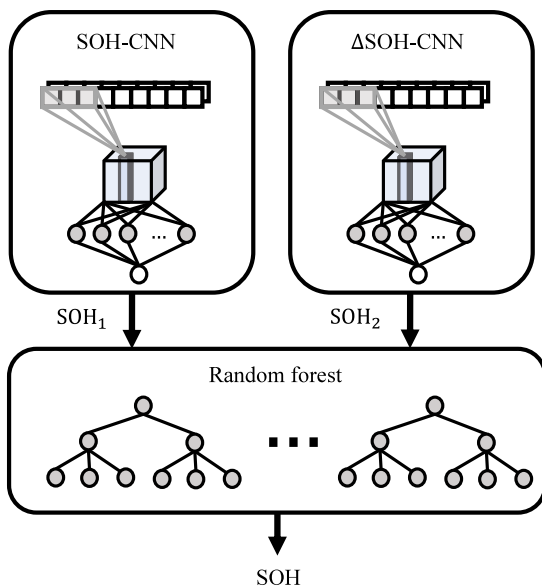

Fig. 11. Overall structure of the RF-CNN (SOH$_1$ and SOH$_2$ are the estimated SOH from SOH-CNN and ΔSOH-CNN).

the best split predictor at each tree node. The average MAE from RF-CNN is 0.85% for the same case study in Sections 4.1 and 4.2, and its

representative SOH estimation performance is shown in Fig. 12. Since the CNNs in the RF-CNN are the same as those developed in Sections 4.1 and 4.2, it can be seen that RF-CNN achieves a better consistency and a 34% improvement in accuracy compared to SOH-CNN (see Fig. 8). Compared to ΔSOH-CNN (see Fig. 9), RF-CNN has a worse consistency but suffers less from the cumulative error, leading to a 46% reduction in MAE. To further illustrate the reason why RF-CNN can enhance the estimation performance, we study the importance of the outputs from two CNNs on the SOH estimate from the random forest model. In particular, the ratios of the importance of the SOH-CNN output to that of the ΔSOH-CNN output over five folds in cross-validation are reported, which are 1.41, 1.38, 1.87, 1.53 and 1.00. From the reported ratios, it can be seen that both CNN models contribute substantially to the final SOH estimate, thereby resulting in better estimation accuracy than that from each individual CNN.

5. Performance evaluation

In this section, to validate the effectiveness of the proposed RF-CNN, we first compare it with the DA-based approach discussed in Section 3. Then, an extensive comparison between RF-CNN, SOH-CNN, and ΔSOH-CNN is performed under partial discharge with different DoD ranges. Finally, several aspects concerning the practical implementation of the proposed approach are discussed.

5.1. Comparison with DA-based approaches

To perform comparison with the DA-based approach, the partial discharge data with the low DoD region presented in Section 3, i.e., $DoD_i = 0$, is used. For a fair comparison, the random forest algorithm is adopted to establish the mapping from the lowest IC minimum to SOH. By applying the DA-based approach that uses ICA and random forest (RF–ICA), its average MAE is 2.31%, while the MAE from RF-CNN is 0.72%. Therefore, it is verified that RF-CNN can achieve better MAE under partial discharge thanks to the richness and robustness of the SOH indicators extracted by CNN. Furthermore, as is shown in Sections 3 and 4.3, RF-CNN has better robustness compared to the RF–ICA under partial discharge without the low DoD region.

5.2. Comparison with SOH-CNN and ΔSOH-CNN

The RF-CNN is further compared with SOH-CNN and ΔSOH-CNN under partial discharge with different DoD ranges to demonstrate the robustness of the RF-CNN. The following four partial discharge conditions are considered:

(i) $DoD_i \sim \mathcal{N}(0.1, 1/900)$ and $Q^{max} \sim \mathcal{U}(0.67, 0.77)$;
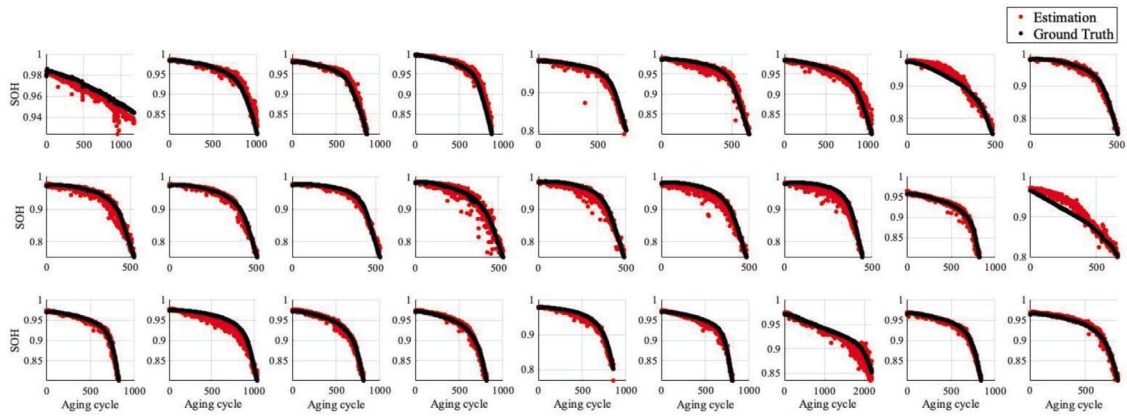(ii) $DoD_i \sim \mathcal{N}(0.2, 1/900)$ and $Q^{max} \sim \mathcal{U}(0.45, 0.55)$;

Fig. 12. Representative SOH estimation performance of the RF-CNN under partial discharge without the low DoD region (Each plot contains the SOH trajectory of a different cell).

Table 2
SOH estimation MAE of different CNN-based estimators under partial discharge with different DoD ranges.

Method	DoD 0.1–0.8	DoD 0.2–0.7	DoD 0.3–0.6	DoD 0.4–0.5
SOH-CNN	0.99%	1.28%	1.61%	2.92%
ΔSOH-CNN	1.04%	1.57%	1.54%	2.30%
RF-CNN	0.62%	0.85%	1.01%	1.51%

(iii) $DoD_i \sim \mathcal{N}(0.3, 1/900)$ and $Q^{max} \sim \mathcal{U}(0.25, 0.35)$;
(iv) $DoD_i \sim \mathcal{N}(0.4, 1/900)$ and $Q^{max} \sim \mathcal{U}(0.05, 0.15)$.

The resulted final DoD distribution for each condition will approximately be a Gaussian distribution as that in Fig. 5 with its mean located near 0.8, 0.7, 0.6, and 0.5. On each partial discharge datasets, a SOH-CNN, a ΔSOH-CNN, and a RF-CNN are trained and tested. The average MAE from different CNN-based estimators is summarized in Table 2. As can be seen from Table 2, RF-CNN can outperform SOH-CNN and ΔSOH-CNN on all four conditions. In addition, although the MAE increases as the data range decreases for all three approaches, RF-CNN maintains better robustness (i.e., smaller MAE increase) compared to the other two CNN-based estimators due to richer SOH indicators.

5.3. Discussion

Extensive discussion about the proposed approach is provided concerning the following four aspects.

(1) Computational complexity: On a computer with a 2.9 GHz Intel Core i5 processor and 16 GB RAM, the computation time for SOH-CNN or ΔSOH-CNN is about 0.001 s, and the computation time for random forest is about 0.05 s. In the future, cloud computing technology may be used to implement the proposed approach on platforms with limited on-board computing resources, e.g., electric vehicles and smartphones [40, 41].

(2) Partial charge/discharge condition: Since the charge data in the adopted dataset is collected under different multi-step charging policies, the constant-current discharge data is used for developing a solution to SOH estimation under partial charge/discharge. In reality, it is easier to achieve a constant-current charge than a constant-current discharge as the battery discharge rate depends on the driver demand [42]. Considering that both random forest and CNN are model-free methods, the proposed approach should be applicable to charge data. For example, it has been shown in [43] that the random forest model can estimate SOH based on SOH indicators under constant-current charge. CNNs are proven to be effective for SOH indicator extraction on charge data in [44]. Meanwhile, when studying the partial discharge problem, we have investigated the sensitivity of the proposed modeling approach to the SOC range (e.g., see Table 2). The

results show that with limited discharge range of 10%, the proposed algorithm can still achieve MAE of less than 2%. Unlike SOC, SOH estimation does not need to run continuously. Consequently, SOH can be updated with the proposed approach when a period of constant-current discharge is expected (e.g., the vehicle is moving at a relatively constant speed).

(3) Battery chemistry: The battery chemistry can impact SOH estimation as battery dynamics (e.g., internal resistance and OCV–SOC curve) can be significantly different [45]. In particular, for batteries with a flat OCV–SOC curve, the detectability of SOH is inherently weak [13]. Consequently, the SOH estimation accuracy of the proposed approach may be impaired. However, the lithium–iron-phosphate battery considered in this study has a relatively flat OCV–SOC curve when compared to other chemistries (e.g., NCM/NMC cells) [11]. Therefore, for batteries with other chemistries, the proposed approach may extract richer SOH indicators due to its automatic feature extraction ability and achieve better estimation accuracy.

(4) Battery temperature: The development and evaluation of the proposed approach are conducted on data collected in a temperature-controlled chamber of 30 °C. In a real battery application, variations in the ambient temperature will inevitably exist [46], even with the best thermal management system. When temperature goes outside the pre-defined window, the SOH estimation algorithm with the model trained under constant temperature should be switched off. Meanwhile, temperature variations can be addressed within the proposed framework if additional degradation data with different temperatures is available. The SOH-CNN and the ΔSOH-CNN of RF-CNN can be retrained with additional input channels taking in the temperature sequences. Suppose there is insufficient data for retraining deep learning models. In that case, a correction mechanism can be developed alternatively which takes the estimate from the RF-CNN and corrects it based on the ambient temperature. Experiments are being planned in the future to collect battery degradation data under different temperature conditions to explore the applicability of the proposed method.

6. Conclusions and future work

In this paper, we consider the SOH estimation problem for a single battery cell under partial discharge and propose RF-CNN as a solution. Two CNNs are used simultaneously to extract the indicators correlating to SOH and change of SOH between two consecutive discharge cycles (ΔSOH) from partial discharge curves. Based on the outputs from the CNNs, a random forest model is then designed to integrate two CNNs to produce the final SOH estimate. Evaluation of the proposed approach is performed based on the partial discharge data with different DoD ranges created from a fast-discharging dataset. By comparing RF-CNN with RF–ICA, SOH-CNN, and ΔSOH-CNN, enhanced accuracy and robustness are verified for the proposed approach. The sensitivity analysis

of the CNN and random forest models further validates that richer indicators can be extracted by RF-CNN for SOH estimation.

In this work, the proposed approach is evaluated on a fast discharging dataset with an identical 4 C discharge rate under 30 °C ambient temperature. Our future work will focus on evaluating the effectiveness of the proposed approach under different battery chemistries, charge/discharge profiles, and ambient temperature conditions using experimental data. Validation of the proposed approach on parallel-connected battery cells will also be our focus in the future.

CRediT authorship contribution statement

Niankai Yang: Simulation, Formal analysis, Writing. **Ziyou Song:** Supervision, Formal analysis, Editing. **Heath Hofmann:** Supervision, Formal analysis, Writing – review & editing. **Jing Sun:** Supervision, Formal analysis, Writing – review & editing.

Declaration of competing interest

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

References

[1] B. Nykvist, M. Nilsson, Rapidly falling costs of battery packs for electric vehicles, Nature Clim. Change 5 (4) (2015) 329–332.

[2] Z. Song, J. Hou, H.F. Hofmann, X. Lin, J. Sun, Parameter identification and maximum power estimation of battery/supercapacitor hybrid energy storage system based on Cramer–Rao bound analysis, IEEE Trans. Power Electron. 34 (5) (2018) 4831–4843.

[3] X. Han, L. Lu, Y. Zheng, X. Feng, Z. Li, J. Li, M. Ouyang, A review on the key issues of the lithium ion battery degradation among the whole life cycle, ETransportation 1 (2019) 100005.

[4] Y. Li, K. Liu, A.M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, J. Van Mierlo, H.E. Hoster, Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review, Renew. Sustain. Energy Rev. 113 (2019) 109254.

[5] K.S. Ng, C.-S. Moo, Y.-P. Chen, Y.-C. Hsieh, Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries, Appl. Energy 86 (9) (2009) 1506–1511.

[6] F.C. Correa, J.J. Eckert, L.C. Silva, F.M. Santiciolli, E.S. Costa, F.G. Dedini, Study of different electric vehicle propulsion system configurations, in: 2015 IEEE Vehicle Power and Propulsion Conference, VPPC, pp. 1–6.

[7] Z. Chen, M. Sun, X. Shu, R. Xiao, J. Shen, Online state of health estimation for lithium-ion batteries based on support vector machine, Appl. Sci. 8 (6) (2018) 925.

[8] S. Shen, M. Sadoughi, X. Chen, M. Hong, C. Hu, Online estimation of lithium-ion battery capacity using deep convolutional neural networks, in: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 51753, American Society of Mechanical Engineers, 2018, pp. V02AT03A058.

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

[10] Z. Wei, K.J. Tseng, N. Wai, T.M. Lim, M. Skyllas-Kazacos, Adaptive estimation of state of charge and capacity with online identified battery model for vanadium redox flow battery, J. Power Sources 332 (2016) 389–398.

[11] S. Tong, M.P. Klein, J.W. Park, On-line optimization of battery open circuit voltage for improved state-of-charge and state-of-health estimation, J. Power Sources 293 (2015) 416–428.

[12] J. Bi, T. Zhang, H. Yu, Y. Kang, State-of-health estimation of lithium-ion battery packs in electric vehicles based on genetic resampling particle filter, Appl. Energy 182 (2016) 558–568.

[13] Y. Zou, X. Hu, H. Ma, S.E. Li, Combined state of charge and state of health estimation over lithium-ion battery cell cycle lifespan for electric vehicles, J. Power Sources 273 (2015) 793–803.

[14] Z. Song, X. Wu, X. Li, J. Sun, H.F. Hofmann, J. Hou, Current profile optimization for combined state of charge and state of health estimation of Lithium ion battery based on Cramer–Rao bound analysis, IEEE Trans. Power Electron. 34 (7) (2018) 7067–7078.

[15] Y. Merla, B. Wu, V. Yufit, R.F. Martinez-Botas, G.J. Offer, An easy-to-parameterise physics-informed battery model and its application towards lithium-ion battery cell design, diagnosis, and degradation, J. Power Sources 384 (2018) 66–79.

[16] K. Uddin, S. Perera, W.D. Widanage, L. Somerville, J. Marco, Characterising lithium-ion battery degradation through the identification and tracking of electrochemical battery model parameters, Batteries 2 (2) (2016) 13.

[17] S. Dey, B. Ayalew, P. Pisu, Combined estimation of state-of-charge and state-of-health of li-ion battery cells using SMO on electrochemical model, in: 2014 13th International Workshop on Variable Structure Systems, VSS, IEEE, 2014, pp. 1–6.

[18] J. Li, K. Adewuyi, N. Lotfi, R.G. Landers, J. Park, A single particle model with chemical/mechanical degradation physics for lithium ion battery state of health (SOH) estimation, Appl. Energy 212 (2018) 1178–1190.

[19] T. Kim, W. Qiao, A hybrid battery model capable of capturing dynamic circuit characteristics and nonlinear capacity effects, IEEE Trans. Energy Convers. 26 (4) (2011) 1172–1180.

[20] M. Dubarry, B.Y. Liaw, Identify capacity fading mechanism in a commercial LiFePO4 cell, J. Power Sources 194 (1) (2009) 541–549.

[21] C. Weng, Y. Cui, J. Sun, H. Peng, On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression, J. Power Sources 235 (2013) 36–44.

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

[23] Y. Merla, B. Wu, V. Yufit, N.P. Brandon, R.F. Martinez-Botas, G.J. Offer, Novel application of differential thermal voltammetry as an in-depth state-of-health diagnosis method for lithium-ion batteries, J. Power Sources 307 (2016) 308–319.

[24] H. Chaoui, C.C. Ibe-Ekeocha, State of charge and state of health estimation for lithium batteries using recurrent neural networks, IEEE Trans. Veh. Technol. 66 (10) (2017) 8773–8783.

[25] P. Li, Z. Zhang, Q. Xiong, B. Ding, J. Hou, D. Luo, Y. Rong, S. Li, State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network, J. Power Sources 459 (2020) 228069.

[26] S. Shen, M. Sadoughi, X. Chen, M. Hong, C. Hu, A deep learning method for online capacity estimation of lithium-ion batteries, J. Energy Storage 25 (2019) 100817.

[27] A. Liaw, M. Wiener, et al., Classification and regression by randomforest, R News 2 (3) (2002) 18–22.

[28] K.A. Severson, P.M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M.H. Chen, M. Aykol, P.K. Herring, D. Fraggedakis, et al., Data-driven prediction of battery cycle life before capacity degradation, Nat. Energy 4 (5) (2019) 383–391.

[29] M. Dubarry, V. Svoboda, R. Hwu, B.Y. Liaw, Incremental capacity analysis and close-to-equilibrium OCV measurements to quantify capacity fade in commercial rechargeable lithium batteries, Electr. Solid State Lett. 9 (10) (2006) A454.

[30] C. Weng, X. Feng, J. Sun, H. Peng, State-of-health monitoring of lithium-ion battery modules and packs via incremental capacity peak tracking, Appl. Energy 180 (2016) 360–368.

[31] Y. Li, M. Abdel-Monem, R. Gopalakrishnan, M. Berecibar, E. Nanini-Maury, N. Omar, P. van den Bossche, J. Van 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 373 (2018) 40–53.

[32] X. Li, Z. Wang, L. Zhang, C. Zou, D.D. Dorrell, State-of-health estimation for li-ion batteries by combing the incremental capacity analysis method with grey relational analysis, J. Power Sources 410 (2019) 106–114.

[33] D. Andre, A. Nuhic, T. Soczka-Guth, D.U. Sauer, Comparative study of a structured neural network and an extended Kalman filter for state of health determination of lithium-ion batteries in hybrid electric vehicles, Eng. Appl. Artif. Intell. 26 (3) (2013) 951–961.

[34] G.-W. You, S. Park, D. Oh, Diagnosis of electric vehicle batteries using recurrent neural networks, IEEE Trans. Ind. Electron. 64 (6) (2017) 4885–4893.

[35] W. Zhang, K. Itoh, J. Tanida, Y. Ichioka, Parallel distributed processing model with local space-invariant interconnections and its optical architecture, Appl. Opt. 29 (32) (1990) 4790–4797.

[36] P. Adam, G. Sam, C. Soumith, C. Gregory, Y. Edward, D. Zachary, L. Zeming, D. Alban, A. Luca, L. Adam, Automatic differentiation in PyTorch, in: Proceedings of Neural Information Processing Systems, 2017.

[37] J. Hong, D. Lee, E.-R. Jeong, Y. Yi, Towards the swift prediction of the remaining useful life of lithium-ion batteries with end-to-end deep learning, Appl. Energy 278 (2020) 115646.

[38] D.E. Farrar, R.R. Glauber, Multicollinearity in regression analysis: the problem revisited, Rev. Econ. Stat. (1967) 92–107.

[39] W.-Y. Loh, Regression tress with unbiased variable selection and interaction detection, Statist. Sinica (2002) 361–386.

[40] J.G. Ibanez, S. Zeadally, J. Contreras-Castillo, Integration challenges of intelligent transportation systems with connected vehicle, cloud computing, and internet of things technologies, IEEE Wirel. Commun. 6 (22) (2015) 122–128.

[41] H. Zhang, Q. Zhang, X. Du, Toward vehicle-assisted cloud computing for smartphones, IEEE Trans. Veh. Technol. 64 (12) (2015) 5610–5618.

[42] J. Tian, R. Xiong, W. Shen, State-of-health estimation based on differential temperature for lithium ion batteries, IEEE Trans. Power Electron. 35 (10) (2020) 10363–10373.

[43] Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J.C.-W. Chan, P. Van den Bossche, J. Van Mierlo, N. Omar, Random forest regression for online capacity estimation of lithium-ion batteries, Appl. Energy 232 (2018) 197–210.

[44] Y. Fan, F. Xiao, C. Li, G. Yang, X. Tang, A novel deep learning framework for state of health estimation of lithium-ion battery, J. Energy Storage 32 (2020) 101741.

[45] D.I. Stroe, M. Swierczynski, A.I. Stan, V. Knap, R. Teodorescu, S.J. Andreasen, Diagnosis of lithium-ion batteries state-of-health based on electrochemical impedance spectroscopy technique, in: IEEE Energy Conversion Congress and Exposition, ECCE, 2014, pp. 4576–4582.

[46] G.-w. You, S. Park, D. Oh, Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach, Appl. Energy 176 (2016) 92–103.