

# Long Short-Term Memory Network with Transfer Learning for Lithium-ion Battery Capacity Fade and Cycle Life Prediction

Yixiu Wang<sup>a,\*</sup>, Jiangong Zhu<sup>b,\*</sup>, Liang Cao<sup>a</sup>, Bhushan Gopaluni<sup>a,\*\*</sup>, Yankai Cao<sup>a,\*\*</sup>

<sup>a</sup>*Department of Chemical and Biological Engineering, University of British Columbia, BC V6T 1Z3, Canada*

<sup>b</sup>*School of Automotive Studies, Tongji University, 201804, Shanghai, China*

---

## Abstract

Machine Learning (ML) is a promising technique for battery health estimation and prediction. However, with more and more types of batteries entering the market, building an ML model from scratch for each new battery requires collecting a large amount of data, which is very expensive and time-consuming. This paper proposes a transfer learning approach to reduce the amount of data that needs to be recollected for a new battery. The key idea is to train an ML model for a new battery of interest (i.e., target battery) with a limited amount of data by transferring the knowledge contained in a well-studied battery (i.e., source battery) with sufficient data. We illustrate this approach using two types of batteries, i.e., the battery with  $\text{Li}_{0.86}\text{Ni}_{0.86}\text{Co}_{0.11}\text{Al}_{0.03}\text{O}_2$ -based positive electrode (NCA battery, source bat-

---

\*These authors contribute equally

\*\*Corresponding authors

*Email addresses:* `bhushan.gopaluni@ubc.ca` (Bhushan Gopaluni), `yankai.cao@ubc.ca` (Yankai Cao)

tery) and the battery with  $\text{Li}_{0.84}\text{Ni}_{0.83}\text{Co}_{0.11}\text{Mn}_{0.07}\text{O}_2$ -based positive electrode (NCM battery, target battery), which have similar degradation patterns but dramatically different cycle life. Specifically, we first pre-train a long short-term memory (LSTM) network, using cycling data of 20 NCA cells at 25 °C and at 45 °C, to predict the following capacity fade based on the previous capacity sequence. Then, to make the model applicable to NCM cells, we employ the transfer learning method to retrain the model, using cycling data of only 2 NCM cells at 25 °C, and propose a two-stage approach to further improve the model performance. The proposed two-stage model can predict the cycle life of NCM cells at 45 °C using the capacities of the first 13 cycles and obtain a cycle life root-mean-squared-error (RMSE) of 25.23 cycles and a capacity trajectory RMSE of 17.80 mAh (0.51 %).

*Keywords:* Lithium-ion battery; Long short-term memory network; Transfer learning; Capacity fade; Cycle life

---

## 1. Introduction

Lithium-ion batteries are widely used in portable electronic devices, micro-grids, electric vehicles and other applications because of their high energy, and power density [1]. However, batteries of any chemistry will gradually lose their initial capacity and power during usage, which leads to degraded performance and potentially higher operating costs [2]. Therefore, it is critical to predict the capacity fade and cycle life accurately. In general, lithium-ion battery aging models are divided into three main categories: physical models

[3], semi-empirical models [4], and data-driven models [5].

The physical modelling of battery aging generally starts with the cell's internal chemical reactions and couples the battery's aging law into the model through certain relevant parameters. Classical theoretical models of cells, such as porous electrode theory and molecular dynamics [6, 7, 8], are often used in this group of models. Ramadass et al. [9] developed the first-principle formulation for battery capacity fade and quantitatively discussed the influence of some parameters. Shrihari et al. [10] proposed a simple one-dimensional model coupling diffusion and dynamics to study the capacity decay of lithium-ion batteries. Purewal et al. [11] obtained the growth law of the solid electrolyte interphase (SEI) film on the negative electrode surface from the expansion of graphite particle cracks and then gave a formulation for calculating the capacity fade per cycle. However, physical models require many chemical kinetic parameters of the cell, some of which are difficult to obtain in practice.

A semi-empirical model is typically an empirical mathematical formulation obtained by fitting accelerated aging experimental data from the laboratory [12]. John et al. [13] proposed a semi-empirical model for capacity fade, and they pointed out that the dependence of capacity fade on the rate is exponential. At the same time, that temperature is a quadratic polynomial, and for a charge, throughput is linear. Long et al. [14] and Saurabh et al. [15] proposed aging models considering the State of Charge (SOC) ranges for  $\text{LiFePO}_4$  cells and  $\text{LiCoO}_2$  cells, respectively. However, one drawback of this

group of models is that they are not very precise, as simple mathematical formulations are difficult to capture the complex dynamics of the battery accurately.

In recent years, data-driven models have become popular in the research field of batteries [16, 17, 18]. Instead of directly reflecting the aging mechanism inside the cell, they are purely based on statistical analysis to model the nonlinear capacity fade law by mining the intrinsic correlation between inputs and outputs. For the task of cycle life prediction, many previous studies have used different machine learning approaches such as Support Vector Machine [19, 20, 21], Gaussian Process Regression [22, 23, 24], Neural Networks [25, 26, 27], to predict future values of capacity and cycle life based on previous capacity trajectory. These data-driven methods that use battery capacity as input require first obtaining the capacity, which can be derived from the cyclic charging/discharging data in the laboratory or estimated from the partial charging profile [28, 29] for practical applications. Recent literatures have proved that battery capacity can be estimated accurately and efficiently for operating data of real-world electric vehicles with incremental capacity analysis [30], Extreme Gradient Boosting (XGBoost) [31], and hybrid deep neural network [32], which provide a data basis for cycle life modeling and demonstrates the possibility of online implementation. However, one disadvantage of these models is that they require nearly half of the capacity trajectory to obtain an accurate cycle life prediction. Recently, Kristen et al. [33] and Yu et al. [34] proposed to use the features of the early-cycle stage to

predict the cycle life directly, and the results exhibited high accuracy. One disadvantage of these data-driven strategies reported in the literature is that they only give a prediction of cycle life but cannot predict the capacity trajectory for each cycle.

Another challenge with data-driven models is that they are specific to a particular battery type. With more and more different types of batteries entering the market, building an aging model of each type of battery will require a significant amount of time and cost to run extensive laboratory experiments. Transfer learning is a promising approach for predicting different battery types' capacity and cycle life. It can transfer the information learned from a well-studied source cell to the target cell. Xiaopeng et al. [35] proposed an input-output slope and bias correction method to capture the aging of the target cell from the source cell, and Sheng et al. [36] proposed a deep convolutional neural network combined with transfer learning for capacity prediction. The limitation of this work is that they only transfer information between different individual cells of the same battery type. However, to the best of our knowledge, approaches that migrate information between different types of batteries for cycle life prediction have not been reported in the literature.

In this paper, we generated a cycling dataset of 32 cells containing two types of batteries, i.e., the battery with  $\text{Li}_{0.86}\text{Ni}_{0.86}\text{Co}_{0.11}\text{Al}_{0.03}\text{O}_2$ -based positive electrode (NCA battery) and the battery with  $\text{Li}_{0.84}\text{Ni}_{0.83}\text{Co}_{0.11}\text{Mn}_{0.07}\text{O}_2$ -based positive electrode (NCM battery), at two temperatures (25 °C and 45

°C). The data of 20 NCA cells are first used to pre-train a long short-term memory (LSTM) network. The LSTM model predicts the capacity fade in the next cycle based on the previous 13 cycles. By recursively feeding the network, the model can predict the long-term capacity trajectory and cycle life. Then, to make the model applicable to NCM cells, we employ the transfer learning method to retrain the model, using cycling data of only 2 NCM cells at 25 °C. The model is tested on 10 NCM cells at 45 °C and achieves good performance. The principal contributions of this study can be summarized as follows:

- 1) The LSTM network is trained with data from all cycles of NCA cells, rather than only early-stage cycling data, so it can capture the capacity trend of the battery at any life point and accurately predict the long-term capacity trajectory with a small amount of initial capacity data.
- 2) Transfer learning is applied to transfer information between different types of batteries for the long-term capacity trajectory and cycle life prediction, which can significantly reduce the amount of data needed to build a data-driven model for new types of batteries.

The paper is organized as follows. Section 2 describes the experimental setup and data generation. Section 3 introduces the preprocessing of the data and the proposed methodology. Section 4 reports the prediction results of the model and discussion. Section 5 summarizes the key ideas of the paper.

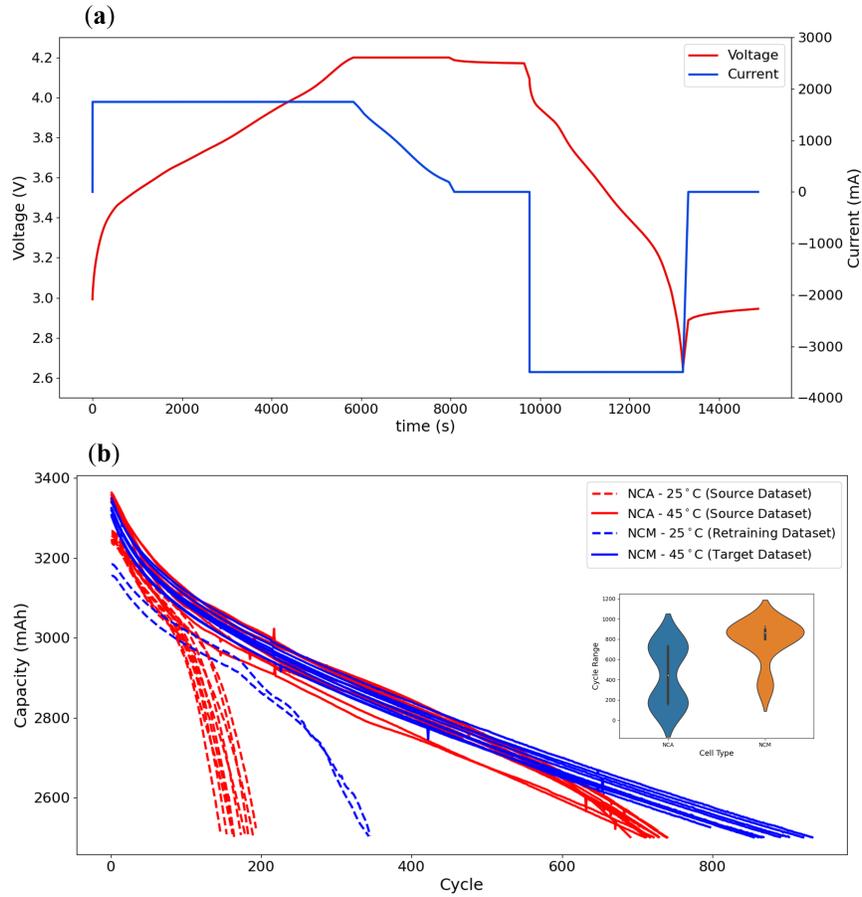
## 2. Data acquisition

Two types of commercial 18650-type batteries with 3500 mAh nominal capacity are cycled at two different temperatures to collect experimental data to model and verify capacity fade and cycle life prediction. The battery specifications are listed in Table 1. 20 NCA cells (10 cells at 25 °C and 10 cells at 45 °C) and 12 NCM cells (2 cells at 25 °C and 10 cells at 45 °C) are used in our study. All charging rates were set to be 0.5 C, and all discharging rates were set to be 1 C, where the current rate is determined on the basis of the nominal capacity of the batteries, i.e., 1 C is equal to 3500 mAh. Specifically, the cell cycling process involves constant current charging to 4.2 V at a rate of 0.5 C, followed by a constant voltage charging step at 4.2 V until the current corresponds to 0.05 C. A constant current is employed at 3500 mA for the discharging process until the voltage drops to 2.65 V for the NCA cells and 2.5 V for the NCM cells, respectively. More detailed descriptions of the experiments and datasets can be found in our previous study [29].

Table 1: Specifications of NCA and NCM batteries

Name	NCA battery	NCM battery
Battery type		18650
Anode material		Graphite/Si
Cathode material	$\text{Li}_{0.86}\text{Ni}_{0.86}\text{Co}_{0.11}\text{Al}_{0.03}\text{O}_2$ [37]	$\text{Li}_{0.84}\text{Ni}_{0.83}\text{Co}_{0.11}\text{Mn}_{0.07}\text{O}_2$ [37]
Electrolyte	Solution of lithium hexafluorophosphate ( $\text{LiPF}_6$ )	
Nominal voltage		3.6 V
Cut-off voltage	2.65 V ~ 4.2 V	2.5 V ~ 4.2 V
Nominal capacity		3500 mAh
Battery mass		45.0 g

Voltage and current are the basic data recorded in the experiment, and Fig. 1 (a) shows the voltage and current profile in one complete cycle for an NCA cell, which includes the charging and discharging process. Capacity is obtained by integrating the current over time during the discharge process at specific ambient temperatures, so that we can get the capacity at each cycle. The datasets of NCA cells and NCM cells are collected, respectively. The NCA cell dataset is used for the training of the source model, while the NCM cell dataset is used for the retraining and final testing of the target model. Figure 1 (b) shows the change of capacity with cycles for all NCA and NCM cells, and the cycle number ranges from 146 to 933 until the capacity decays to 2500 mAh. The cycling data of NCA cells are plotted as red lines, where the dotted line is for cells cycling at 25 °C and the solid line is for cells at 45 °C. Fatigue down to 2500 mAh is found to be around 175 cycles at 25 °C cycling temperature and around 725 cycles at 45 °C cycling temperature, exhibiting temperature is an important factor affecting the rate of battery degradation. The cycling data of NCM cells are shown as blue lines, and it is evident that the capacity of 2500 mAh is reached after 350 cycles and 900 cycles at 25 °C and 45 °C respectively. Different peak positions of the violin plot in the figure also indicate that NCA cells and NCM cells exhibit different cycling characteristics due to the difference in the cathode material.



**Fig. 1. Battery cycling dataset:** (a) The plot of the voltage and current profile in one complete cycle for a NCA cell. (b) Capacity versus cycle number of all cells.

### 3. Methodology

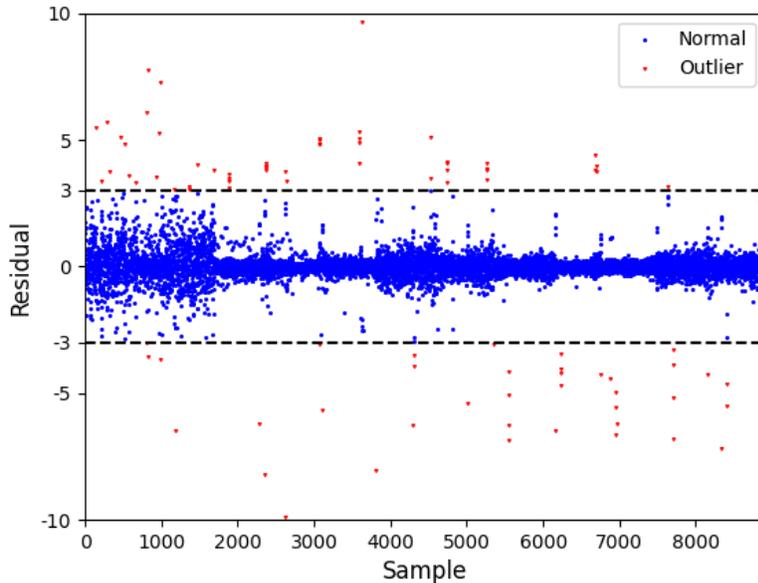
#### 3.1. Data preprocessing

The outliers in the dataset can reduce the accuracy of the model. In this paper, the outliers are detected by the residual analysis. For each cell in the dataset, the sequence of capacity data is smoothed with the moving average

method, and the standard residual is calculated by

$$d_i = \frac{Q_i - \tilde{Q}_i}{\sqrt{MSE}} \quad (1)$$

where  $Q_i$  is the experimental capacity,  $\tilde{Q}_i$  is the smoothed capacity and  $MSE$  is the mean-squared-error between the original and smoothed data. Fig. 2 shows the standard residual of all samples for NCA cells, and the samples with standard residual outside the 99% confidence interval are recognized as outliers and will be replaced by linear interpolation. For some cells, capacity increases slightly in the first few cycles. These cycles with increasing capacity are removed to facilitate the model training, and we only consider cycles after capacity fade is observed. In other words, the first cycle is the one where the capacity rises to the highest value. All capacity data after outlier removal are smoothed with the Savitzky-Golay method [38] to reduce noise. The capacity data in the following discussion are all smoothed data.

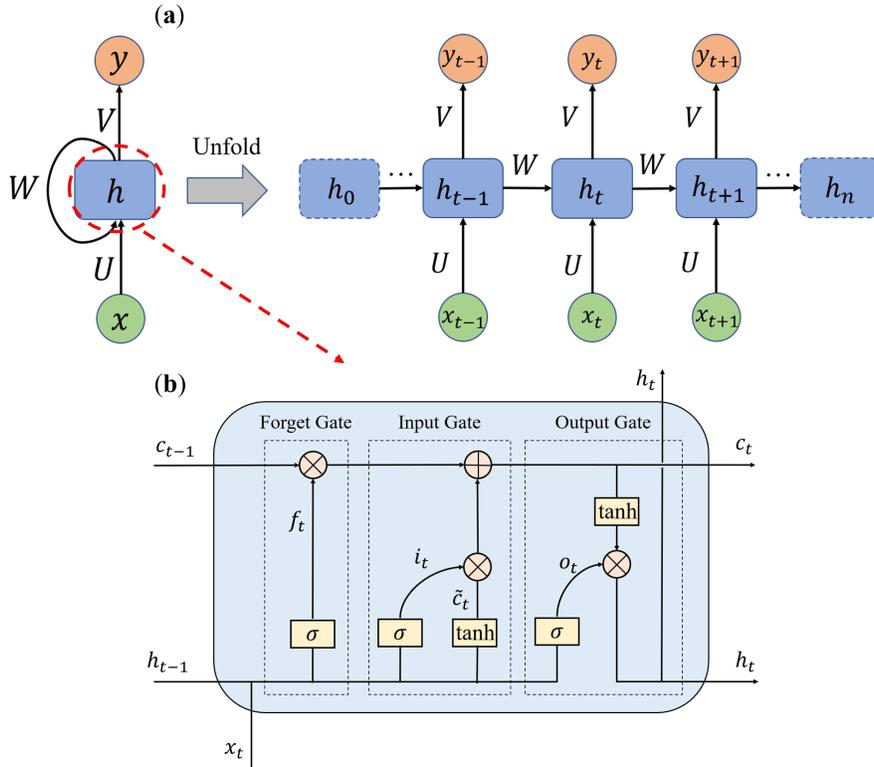


**Fig. 2.** Residual analysis for battery capacity with 99% confidence interval for NCA cells.

### 3.2. Long short-term memory network

The goal of our model is to predict future battery capacity trends based on previous capacity trajectories. Here we choose the recurrent neural network (RNN) [39] for battery capacity prediction because of its nonlinear mapping power and strong prediction performance based on the historical sequence information. It is a variant of artificial neural networks used to process continuous sequential data. It demonstrates high accuracy in many fields of application, such as machine translation [40] and multilingual language processing [41]. It can read one input at a time and store some information through the hidden state that gets passed from one time-step to the next. A basic structure of a typical RNN is shown in Fig. 3 (a), where  $x$  is the input,

$y$  is the output, and  $h$  is the hidden state updated over time.



**Fig. 3. Illustration of the LSTM network:** (a) A basic structure of a RNN. (b) The structure of an LSTM cell.

One challenge in training an RNN model with gradient-based methods is that the gradients that are propagated can tend to zero (vanish) or tend to infinity (explode) as the number of layers and time-steps increase. These phenomena make the optimization algorithm fail to find the optimal parameters of the networks. To overcome this challenge, the LSTM network uses a more carefully designed cell structure to avoid the gradient vanishing and

exploding problems [42], as shown in Fig. 3 (b). In addition to the usual hidden state  $h$ , LSTM cells have an extra memory state  $c$ . An LSTM cell consists of a forget gate and an input gate that allows the information to be removed from or written to the memory state  $c$ , and an output gate that decides which values of the memory state to output. The sigmoid function in these gates can decide how much of each component should be let through by outputting numbers between 0 and 1, so that the gates can selectively pass information.

In this paper, the LSTM network is trained to predict the capacity fade for the next cycle based on the capacities of previous  $m$  cycles and cycling temperature. Note that the capacity fade in this paper means the difference in capacity between two consecutive cycles, not a loss of capacity relative to the nominal capacity or initial capacity. As shown in Fig. 4, in order to generate samples, a window of length  $m + 1$  ( $m$  inputs and 1 output) is used to scan capacity data for each cell to obtain input-output pairs. More generally, consider a capacity sequence of a cell at temperature  $T$  is  $Q = [Q_1, \dots, Q_t, \dots, Q_N]$ , where  $Q_t$  denotes the capacity of  $t$ -th cycle and  $N$  is the sequence length. For any continuous  $m + 1$  capacity data in the sequence  $[Q_i, Q_{i+1}, \dots, Q_{i+m-1}, Q_{i+m}]$  ( $1 \leq i \leq N - m$ ), we can have an input-output pair to train the LSTM network. Concretely, the input to the network is

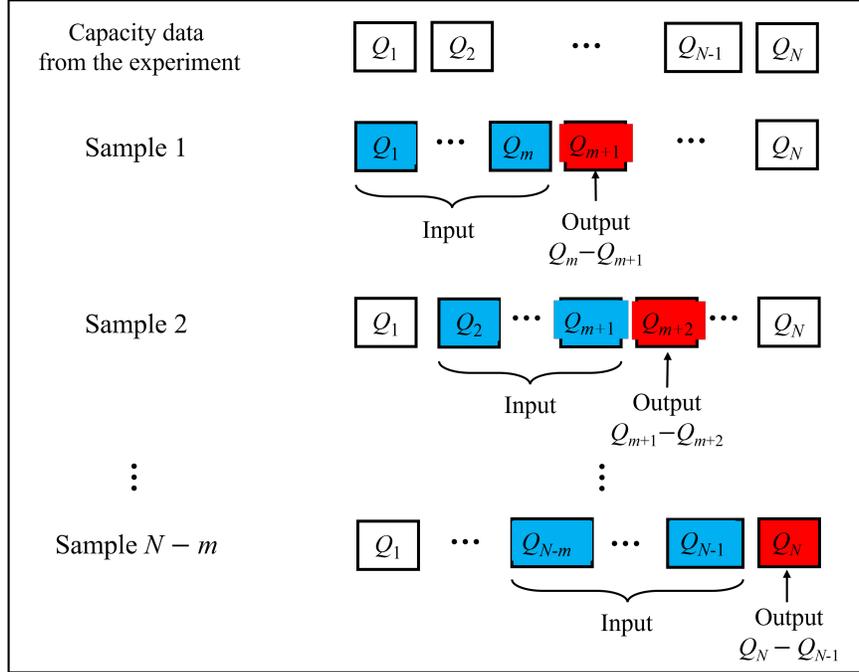
$$x_i = [x_i^{(1)}; \dots; x_i^{(j)}; \dots; x_i^{(m)}] \quad (2)$$

where  $x_i^{(j)} = (Q_{i+j-1}, T)$ , and the output is

$$y_i = Q_{i+m-1} - Q_{i+m} \quad (3)$$

Moreover, to reduce overfitting and improve the robustness of the model, a GaussianNoise layer is added before the LSTM cell. A fully connected network containing four hidden layers and one output layer is also added after the LSTM cell to improve the model’s predictive performance. Each hidden layer uses a *sigmoid* activation function to introduce nonlinearity, whereas the output layer uses a *softplus* activation function to ensure that the output is greater than 0. We choose to predict the capacity change rather than the capacity in this paper. One reason is that the capacity change of each cycle is small relative to the capacity, and directly predicting the capacity value may cause the model to output the last value of the input capacity sequence directly. Another reason is that predicting the capacity change ensures the capacity decreases over cycles for long-term capacity trajectory prediction, which is consistent with the prior knowledge on the battery that the battery capacity shows a downward trend due to the change of internal electrochemical reaction with the degradation of the battery [43]. In addition, because the battery has a faster degradation rate at 25 °C, there are significantly more samples at 45 °C than at 25 °C. To balance the sample inconsistency, the samples are weighted by the total number of samples at two temperatures, so samples at 25 °C are given a higher weight. The overall loss function for

the LSTM network is the mean squared error loss.



**Fig. 4.** Diagram of the sample generation.

### 3.3. Cycle life prediction algorithm

The LSTM network proposed in the previous subsection can predict the capacity change of the next cycle; in this subsection, we illustrate how to use the model to make predictions for a longer forecasting horizon, i.e., the entire capacity trajectory before reaching the end of life and the corresponding cycle life. We chose to feed the predicted capacities back into the LSTM model recursively. The details for the recursive method are illustrated as follows.

The trained LSTM network is essentially a nonlinear mapping between in-

puts  $x$  and outputs  $y$ , and the notation  $x \rightarrow f(x)$  is adopted to describe this relationship. Given an input  $x_1 = [(Q_1, T); (Q_2, T); \dots; (Q_m, T)]$  (capacities of the first  $m$  cycles and cycling temperature) for one cell in the test set, the model can get the one-step capacity fade prediction  $\hat{y}_1 = f(x_1)$ . The superscript  $\wedge$  here means that it is the predicted value obtained by the model. We compute the predicted capacity of the next cycle  $\hat{Q}_{m+1}$  using  $\hat{Q}_{m+1} = Q_m - \hat{y}_1$  and it can be used to update the input  $x_2 = [(Q_2, T); \dots; (Q_m, T); (\hat{Q}_{m+1}, T)]$ . For simplicity of the expression, we use  $\hat{y}_1$  directly to denote the new input  $x_2$ , and then the two-step prediction is  $\hat{y}_2 = f(\hat{y}_1) = f(f(x_1))$ . After repeated iterations, the  $k$ -step prediction is

$$\hat{y}_k = f(\hat{y}_{k-1}) = f(f(\hat{y}_{k-2})) = \dots = \underbrace{f(\dots f(\hat{y}_1))}_{k-1} = \underbrace{f(\dots f(x_1))}_k \quad (4)$$

Based on the above recursive idea, the cycle life prediction algorithm is shown below. Here we set the battery cell to reach 75% of the nominal capacity as the end of life, that is, the capacity reached 2625 mAh.

---

**Algorithm 1** Cycle Life Prediction

---

**Input:**

The first  $m$  capacity data of a cell and the cycling temperature  
The capacity value of the cell reaching the end of life  $Q_{end}$

**Output:**

The predicted cycle life of the cell

- 1: Initial the cycle number  $l = m$ ;
  - 2: Predict the capacity fade for the next cycle  $\hat{y}_{l-m+1}$  with previous  $m$  capacity values and temperature;
  - 3: Compute the capacity for the next cycle  $\hat{Q}_{l+1} = \hat{Q}_l - \hat{y}_{l-m+1}$ ;
  - 4: set  $l = l + 1$ ;
  - 5: **if**  $\hat{Q}_l > Q_{end}$  **then**
  - 6:   Go back to step 2
  - 7: **else**
  - 8:   **return**  $l$  as the cycle life of the cell
  - 9: **end if**
- 

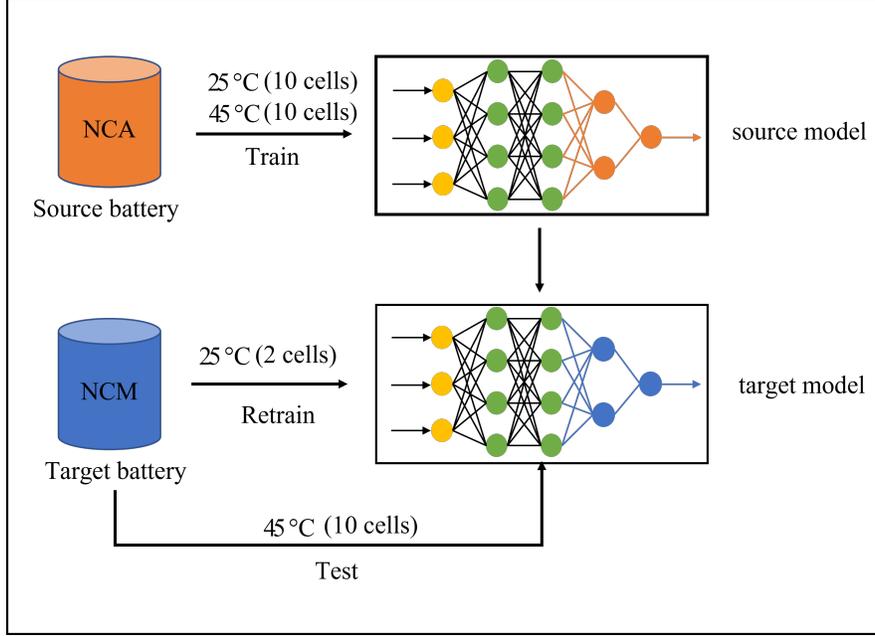
### 3.4. Transfer learning

Generally, the cycling characteristics of different batteries are not the same. However, traditional machine learning techniques assume that train data and test data have identical statistical distribution, which is not guaranteed between different batteries. In this case, to predict the cycling performance of a new type of battery, many experiments on the new battery are required, which may take months to years. Fortunately, different batteries share the same degradation pattern, i.e., lithium loss due to the solid electrolyte interphase (SEI) film formation. Therefore, it is crucial to transfer some relevant important information from one type of battery to another. Transfer learning provides a practical framework for this issue.

Transfer learning is a machine learning method that focuses on applying

knowledge gained while solving one task to a related task [44, 45]. This paper proposes a transfer learning method to transfer the model information trained on NCA cells (source battery) to NCM cells (target battery). The core idea of transfer learning is to improve the prediction accuracy for the target battery using the knowledge in the source battery. Specifically, we first pre-train a long short-term memory (LSTM) network, using cycling data of 20 NCA cells to predict the capacity fade and cycle life. Then, to make the model applicable to NCM cells, we employ the transfer learning method to retrain the model, using cycling data of only 2 NCM cells at 25 °C. In the implementation of transfer learning, fine-tuning of the last few layers is a common strategy for neural networks [46, 47, 48], as the last few layers are usually task-specific layers that need to be fine-tuned based on new data and tasks. In this study, we kept the variables of the early and middle layers of the LSTM network unchanged, and only retrained the last two layers of the network to learn the difference between the source and target batteries. Fig. 5 shows a flowchart of TL, and here are the details of our method:

- 1) Train an LSTM network to predict the capacity fade based on experimental data of NCA cells (source dataset). Data for both 25 °C and 45 °C are used here. This model is denoted as the source model.
- 2) Freeze other layers and retrain the last two layers of model\_source based on experimental data of 2 NCM cells (target dataset) at 25 °C to get a transfer learning model. The performance of this model is assessed on 10 NCM cells at 45 °C. This model is denoted as the target model.



**Fig. 5.** Flowchart of transfer learning.

## 4. Results and discussion

### 4.1. Metrics

To evaluate the predictive performance of the LSTM networks, Root-Mean-Squared-Error (RMSE) is used in the paper, and it is given by

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (5)$$

In our paper, 3 different RMSEs are proposed to describe the accuracy of capacity fade and cycle life prediction:

1) RMSE of Capacity Fade Prediction (RCF) is used to evaluate the performance of the LSTM network in predicting the capacity fade of the next cycle based on the previous capacities. Here,  $\hat{Y}_i$  is the predicted value given by the LSTM network,  $Y_i$  is the corresponding target value in the test set, and  $n$  is the total number of test samples.

2) RMSE of Cycle Life Prediction (RCL) is used to evaluate the performance of our cycle life prediction algorithm. Here,  $\hat{Y}_i$  is the predicted cycle life obtained by the cycle life prediction algorithm,  $Y_i$  is the observed cycle life of the corresponding cell, and  $n$  is the total number of test cells.

3) RMSE of Capacity Trajectory Prediction (RCT) is used to evaluate the performance of the cycle life prediction algorithm on long-term capacity trajectory prediction. Here,  $\hat{Y}_i$  is the predicted capacity of each cycle obtained by recursively feeding the LSTM network predictions,  $Y_i$  is the corresponding observed capacity, and  $n$  is the total number of predicted capacities.

Moreover, considering the difference in the cycle life of cells under different operation conditions, Percentage Error of Cycle Life (PECL) is also proposed to describe the accuracy of the cycle life prediction algorithm. It is defined as

$$\text{PECL} = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i} \times 100 \quad (6)$$

where  $\hat{Y}_i$  is the predicted cycle life of each cell,  $Y_i$  is the observed cycle life of the corresponding cell, and  $n$  is the total number of test cells.

#### 4.2. Performance of source model on NCA cells

We first train an LSTM network based on experimental data of NCA cells (source dataset). The LSTM model predicts the capacity fade for the next cycle based on the capacities of previous  $m$  cycles and cycling temperature.

##### 4.2.1. Influence of the number of input capacities

To investigate the influence of the different number of input capacities  $m$  on the prediction performance, we compare six models with different  $m$ . The source dataset containing 20 NCA cells is randomly divided into a training set with 16 cells (8 cells at 25 °C and eight cells at 45 °C) and a validation set with four cells (2 cells at 25 °C and two cells at 45 °C). Six models are trained on the training set, and the prediction results on the validation set are shown in Table 2.

Table 2: Comparison of source model performance on NCA cells with different  $m$

$m$	RCF (mAh)	RCL (cycles)	RCT	PECL
5	0.3471	23.96	43.94 (1.26%)	6.43%
8	0.2557	25.54	50.72 (1.45%)	6.63%
10	0.1812	20.90	35.98 (1.03%)	4.87%
13	0.1424	<b>14.12</b>	<b>33.69 (0.96%)</b>	<b>4.67%</b>
15	0.1298	25.37	45.74 (1.31%)	7.84%
20	<b>0.1167</b>	27.49	37.58 (1.07%)	6.93%

As shown in Table 2, it is clear that RCF decreases as  $m$  increases, as more capacities can provide more information. The best capacity fade prediction error at  $m$  equals 20 is 0.1167 mAh, only about 1/3 of that at  $m$  equals 5. However, unlike the capacity fade prediction results, we also notice that

a larger  $m$  does not guarantee a higher accuracy of cycle life prediction. RCL, RCT and PECL at  $m$  equal 15 and 20 are significantly higher than those at  $m$  equal 10 and 13. One possible explanation for this is that more inputs make the model more prone to overfitting. When we perform cycle life prediction, the input to the LSTM network is the predicted value rather than the actual value, and the error is more likely to accumulate when the model is overfitted. Considering the prediction accuracy of capacity fade and cycle life comprehensively,  $m$  is chosen to be 13 in the following discussion.

#### *4.2.2. Comparison of different machine learning models*

The effectiveness of the proposed LSTM network is benchmarked with other three commonly used machine learning models, including the elastic net [49], support vector regression (SVR) [50] and Extreme Gradient Boosting (XGBoost) [51], covering a range of linear to nonlinear scenarios. The elastic net is an extension of the linear model that combines the  $L_1$  and  $L_2$  penalties of the lasso and ridge methods. SVR is a special case of support vector machine in regression, which was designed to find the optimal decision boundary. The advantage of SVR over the elastic net is that it can introduce nonlinearity through different kernel functions, i.e., radial basis function (rbf) kernel, to improve the performance of the model. XGBoost is an efficient and scalable implementation of a gradient boosting framework designed for speed and performance. It gives a prediction result in the form of an ensemble of decision trees and can handle nonlinear relationships.

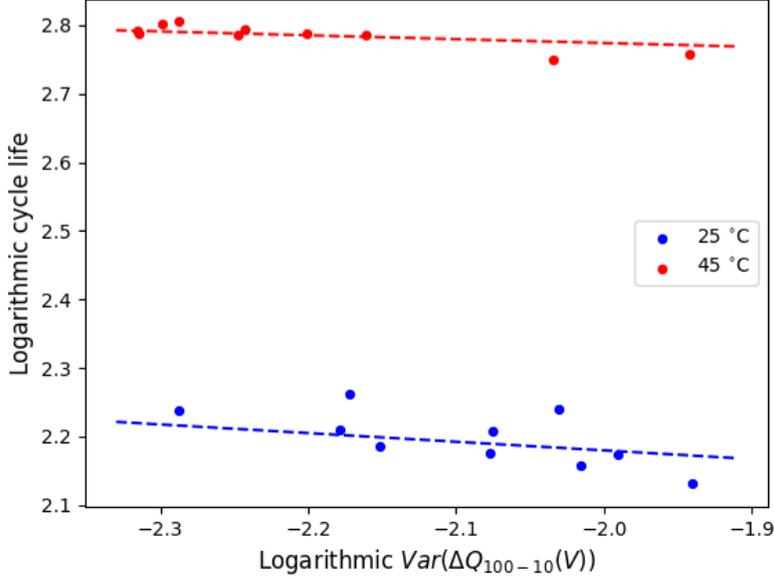
Table 3: Performance comparison of different source models on NCA cells

Model	RCF (mAh)	RCL (cycles)	RCT	PECL
The elastic net	1.6216	169.62	88.45 (2.53%)	22.73%
SVR	0.3168	23.15	45.28 (1.29%)	6.12%
XGBoost	0.3189	22.62	43.52 (1.24%)	6.17%
LSTM network	<b>0.1424</b>	<b>14.12</b>	<b>33.69 (0.96%)</b>	<b>4.67%</b>

Same as in Section 4.2, 20 NCA cells are randomly divided into training and validation sets at a 4:1 ratio. All models are trained on the same training set, and the prediction results on the validation set are reported. Moreover, to ensure a fair comparison between different models, we compared the performance of the model with different hyperparameters. There are two hyperparameters in SVR, i.e., *epsilon* and the regularization parameter *C*. As shown in Table A1 and Table A2 in Appendix, it is observed that RCF decreases as *C* increases, but a larger *C* does not guarantee a higher accuracy of cycle life prediction. After comprehensively considering the prediction accuracy of capacity fade and cycle life, the performance of SVR with *epsilon* of 0.2 and *C* of 0.1 was reported in the paper. For XGBoost, we have tested the performance with a different number of estimators *n\_estimators* from 1 to 10. The results shown in in Table A3 in Appendix suggested that *n\_estimators* of 5 have good accuracy for both capacity fade and cycle life prediction. Table 3 exhibits the performance of different models. The elastic net shows the worst performance because it cannot capture the complex nonlinear dynamics of

the battery. SVR and XGBoost demonstrate better prediction accuracy, but they are still worse than the LSTM network, demonstrating the LSTM network’s superiority for modelling on time series data.

We also compared the performance with the method proposed in Kristen et al. [33], which established an early-prediction model to predict the cycle life using data from the first 100 cycles. To be more specific, they proposed a feature  $Var(\Delta Q_{100-10}(V))$ , which is the variance of the change in discharge voltage curves between cycles 10 and 100, and found a strong linear correlation between the logarithm of  $Var(\Delta Q_{100-10}(V))$  and the logarithm of the cycle life. An elastic net regression combined L1 and L2 regularizer were then proposed to predict the cycle life based on  $Var(\Delta Q_{100-10}(V))$  with reasonable accuracy. Fig. 6 shows the relationship between the logarithmic  $Var(\Delta Q_{100-10}(V))$  and the logarithmic cycle life of our dataset from 20 NCA cells. For all cells, the correlation coefficient is  $-0.51$ , which does not demonstrate a strong correlation. If we consider cells cycled at different temperatures separately, the correlation coefficient is  $-0.69$  for cells at  $25\text{ }^{\circ}\text{C}$  and  $-0.88$  for cells at  $45\text{ }^{\circ}\text{C}$ . Further, we train a model separately for cells cycled at different temperatures, RCL and RECL on four validation cells are 15.35 cycles and 4.75%, respectively. The results are relatively accurate but are still slightly worse than our model. Another advantage of our proposed method is that it can also predict the long-term capacity trajectory, which cannot be predicted using methods in Kristen et al. Moreover, our model only needs the capacities of the first 13 cycles as inputs instead of 100 cycles.



**Fig. 6.** Logarithmic cycle life vs logarithmic  $Var(\Delta Q_{100-10}(V))$

#### 4.2.3. Computational efficiency analysis

To evaluate the feasibility of the practical implementation of the proposed model, Table 4 compares the computational cost of the four source models in terms of training time, average testing time of capacity fade prediction per cycle, and average testing time of cycle life prediction per cell. We use Python to implement the proposed models and our implementation runs on a computer with Intel(R) Core(TM) i7-8700K CPU running at 3.70 GHz. As shown in the table, it can be noticed that the proposed LSTM network consumes more than 500 times the training time compared to the other models because of the complex structure. However, once we have obtained the model parameters, the LSTM network only requires a similar amount of time as the

other models to perform the capacity fade prediction and cycle life prediction. It should be noted that the LSTM network can be deployed in onboard battery management systems after offline training. In this case, the training time is not a concern, and we confirm that the proposed method can be used in practical applications.

Table 4: Computational time of different models on NCA cells

Model	Training (s)	Testing (s)	
		Capacity fade	Cycle life
The elastic net	$4.987 \times 10^{-3}$	$1.70 \times 10^{-7}$	0.0147
SVR	0.5027	$1.15 \times 10^{-4}$	0.0798
XGBoost	0.0628	$1.70 \times 10^{-6}$	0.6188
LSTM network	243.6612	$7.13 \times 10^{-5}$	0.4548

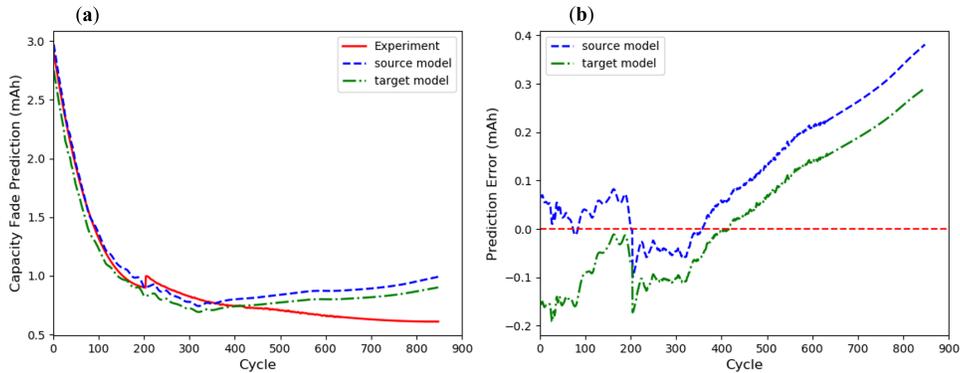
#### 4.3. Effects of transfer learning on predicting NCM cells

After training the source model on NCA cells, to make the model applicable to NCM cells, we employ the transfer learning method to obtain the target model, using cycling data of only 2 NCM cells.

##### 4.3.1. Performance of target model on capacity fade prediction of NCM cells

To compare the performance of the source model and target model on capacity fade prediction, Fig. 7 (a) shows the predicted capacity fade of one random test cell, while 'Experiment' is the observed capacity fade. To visualize the results, the prediction error is also shown in Fig. 7 (b), where the red line in the figure is the 'zero' line, and the closer the data point is to this line, the smaller the error. As shown in Fig. 7 (b), compared to the source model,

the prediction error is reduced for most of the cycles after transfer learning. But, we also noticed that the source model exhibited higher accuracy in the capacity fade prediction for the first few cycles, which may be because there are so few samples (only 660 samples) to retrain the model that the target model cannot capture the aging characteristics of this interval.



**Fig. 7. Capacity fade prediction results for one test cell:** (a) Capacity fade versus cycle number. (b) Prediction error versus cycle number.

To illustrate the average results across all cells in the test set, Table 5 shows the comparison of RCF for all test samples. Here, all of the test data are derived from the cycling date of NCM cells at 45 °C and test samples are independent since our LSTM network directly maps the sequence of the previous capacities and the capacity fade for the next cycle. From the results in the table, it is clear that transfer learning can improve the accuracy of the capacity fade prediction and the RCF can be reduced from 0.1892 mAh to 0.1536 mAh after transfer learning.

Table 5: Performance comparison of different model on NCM cells at 45 °C

Model	RCF (mAh)	RCL (cycles)	RCT	PECL
source model	0.1892	67.39	18.54 (0.53%)	9.35%
target model	<b>0.1536</b>	32.68	35.35 (1.01%)	4.47%
TL0 model	1.7104	288.82	60.52 (1.73%)	43.00%
TS model-3050 mAh	-	<b>25.23</b>	<b>17.80 (0.51%)</b>	<b>2.74%</b>

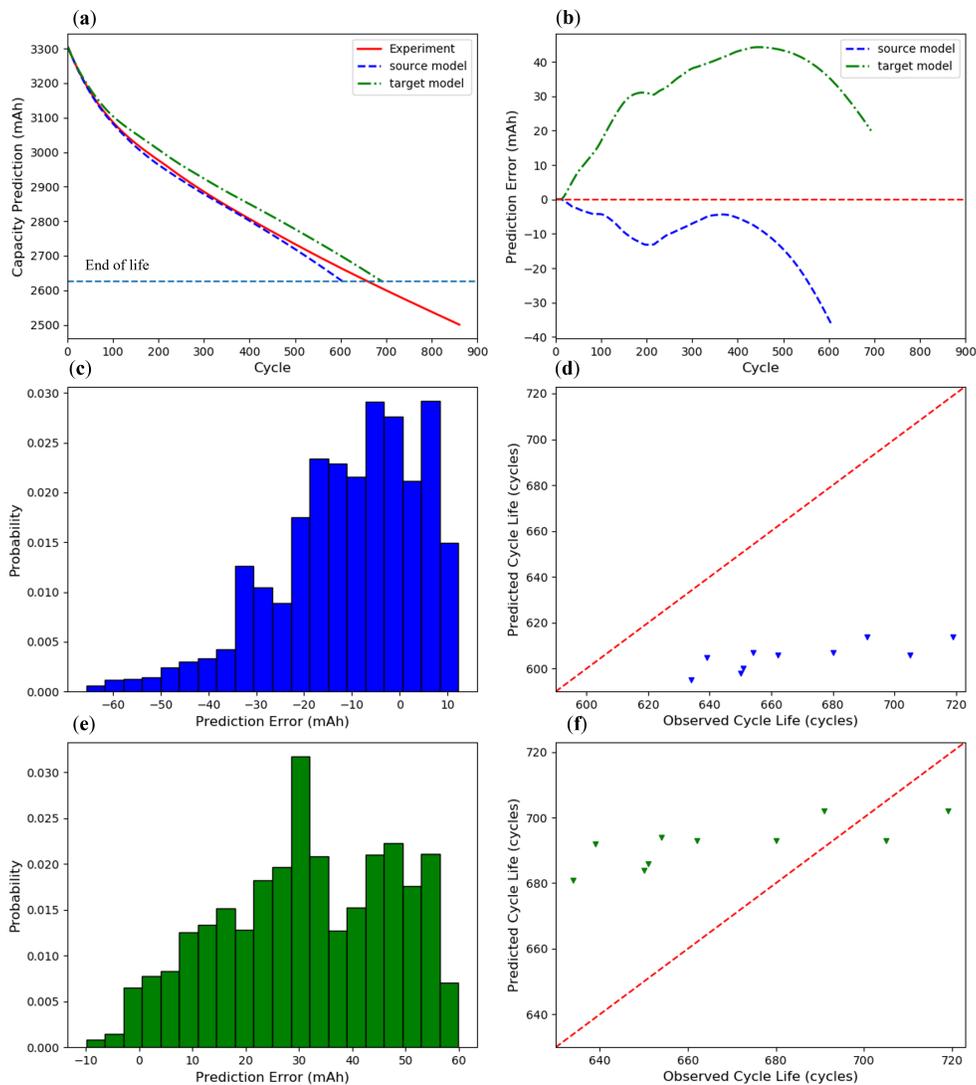
To further validate the effectiveness of transfer learning, we also show the performance of the TL0 model, which is trained totally from scratch based on the experimental data of 2 NCM cells cycled at 25 °C only. The RCF and RCL on 10 test cells at 45 °C are 1.7104 mAh and 288.82 cycles, respectively. The poor prediction results show that the influence of temperature cannot be ignored, and we cannot accurately predict the cycling characteristics of the battery cycled at 45 °C using only cycling data at 25 °C without transfer learning. We will only compare the performance of the model\_source and model\_target in the following discussion.

#### 4.3.2. Performance of target model on cycle life prediction of NCM cells

We also evaluate the effects of transfer learning on the cycle life prediction, which is critical for battery usage and management. Here the cycle life prediction algorithm mentioned in Section 3.3 is implemented. As shown in Table 5, the source model has an RCL of 67.39 cycles and an RCT of 18.54 mAh (0.53%), while the target model achieves an RCL of 32.68 cycles and an RCT of 35.35 mAh (1.01%). Specifically, Fig. 8 (a) shows the prediction

of capacity trajectory over cycles for one random NCM cell in the test set, where "Experiment" means the observed capacity curve and the number of cycles when the cell capacity reaches the end of life (2625 mAh) is defined as cycle life. Fig. 8 (b) further demonstrates the prediction error with different cycles. Note that the cumulative error due to the fact that the inputs to the model are predicted values rather than experimental values is much larger than the prediction error in Fig. 7 (b). As shown in Fig. 8 (a) and (b), the source model demonstrates higher accuracy in predicting the long-term capacity trajectory, especially for the first 400 cycles, but it fails in predicting cycle life. When we updated the LSTM network with some of the experimental data from NCM cells, it could be noticed that the accuracy of the cycle life prediction started to improve. Still, the long-term capacity trajectory prediction results are not satisfactory because it does not have good prediction accuracy in the early-cycle stage, and error accumulates as the number of cycles increases. Moreover, for all test cells, Fig. 8 (c) shows the capacity trajectory prediction error distribution for the source model. Most of the prediction errors are concentrated in the range of -20 mAh to 10 mAh, which is excellent relative to the nominal capacity of 3500 mAh. Fig. 8 (d) illustrates the corresponding cycle life predictions, and all of the predicted cycle life are lower than the observed cycle life, indicating the difference in the cycling characteristics of the two batteries. For the target model, Fig. 8 (e) shows the distribution of the capacity trajectory prediction error, and most of the prediction errors are concentrated in the range of 20 mAh to

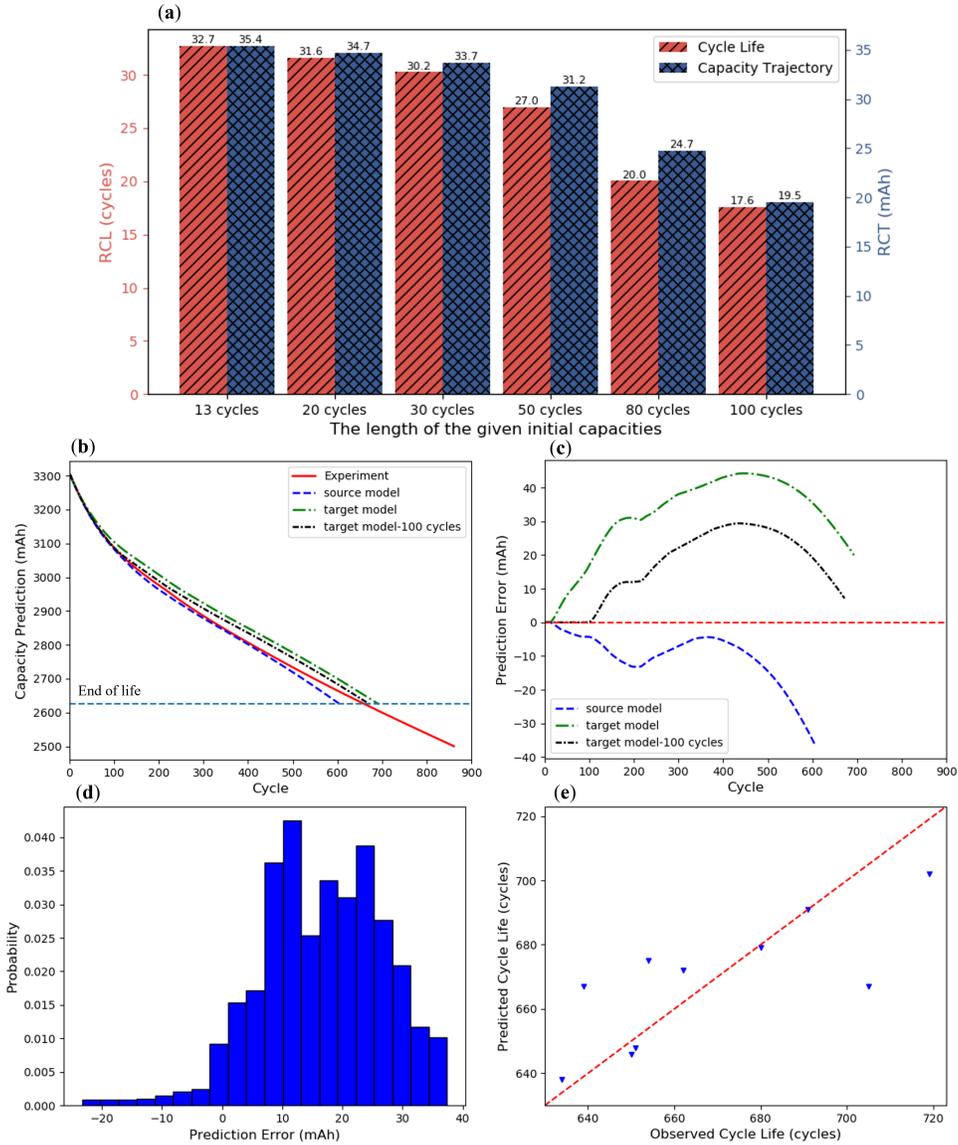
55 mAh, which are significantly worse than source model. But, for cycle life prediction, Fig. 8 (f) shows that the target model has higher accuracy compared to the source model, which is consistent with the results in Table 5.



**Fig. 8. Cycle life prediction results of NCM cells:** (a) Capacity trajectory prediction and (b) prediction error for one test cell. (c) Distribution of capacity trajectory prediction error and (d) cycle life prediction for source model. (e) Distribution of capacity trajectory prediction error and (f) cycle life prediction for target model.

We further analyze the effect of the length of the given initial capacities on the prediction accuracy, and the results are shown in Fig. 9. Fig. 9 (a)

shows RCL and RCT change with the length of the given initial capacities. Since RCL and PECL have the same trend with the different lengths of the given initial capacities, we choose to show only RCL. Here, if the length of the given initial capacities is greater than 13, we only choose the last 13 capacities as the initial input to the model. For example, when the given initial capacities length is 50 cycles, we choose the 38th to 50th cycle capacities as the model input. As shown in the figure, it is clear that the more the initial capacities are given, the higher the model prediction accuracy. Fig. 9 (b) and (c) compare capacity trajectory prediction results of source model, target model and target model with the first 100 initial capacities for one random test cell, and we can notice that the target model with the first 100 initial capacities has a significantly higher accuracy in capacity fade and cycle life prediction relative to the other two models. Fig. 9 (d) demonstrates the capacity trajectory prediction error distribution for the target model with the first 100 initial capacities. Most of the prediction errors are in the range of 5 mAh to 30 mAh, which are better than the result in Fig. 8 (e). Fig. 9 (e) shows the corresponding cycle life prediction result, which exhibits a higher level of accuracy compared to Fig. 8 (d) and (f). Concretely, for the target model, RCT, RCT and PECL with the first 100 initial capacities are only 17.61 cycles, 19.55 mAh (0.56%) and 1.87%, respectively, which are almost half of these with 13 initial capacities.



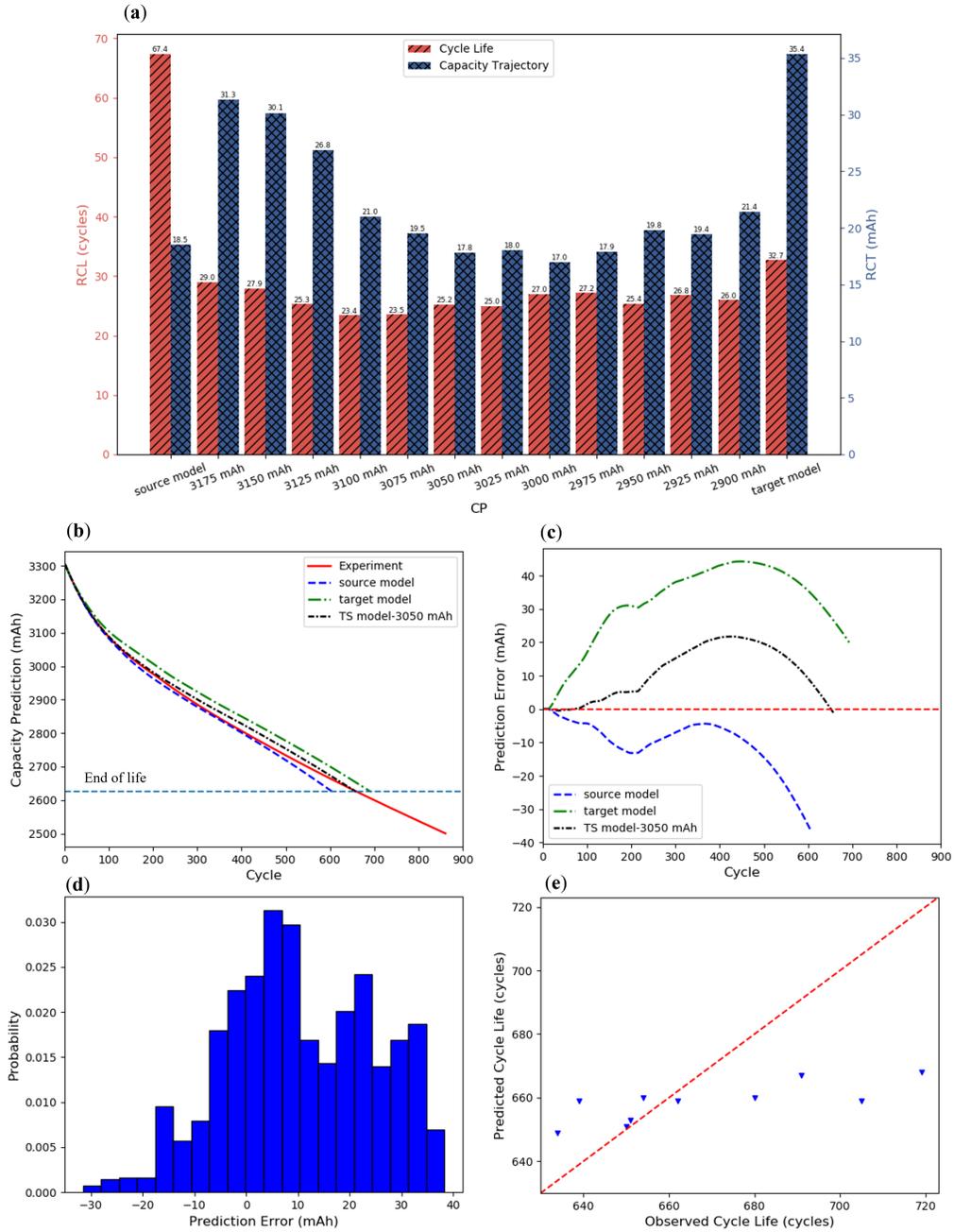
**Fig. 9. Results for the effect of the length of the given initial capacities:** (a) Prediction accuracy versus the length of the given initial capacities. (b) Capacity trajectory prediction and (c) prediction error for one test cell. (d) Distribution of capacity trajectory prediction error and (e) cycle life prediction for target model with the first 100 initial capacities.

#### 4.4. Two-stage model

A two-stage model is proposed to improve the prediction accuracy further to address the long-term capacity trajectory prediction for the target model. The motivation is that the target model loses its prediction accuracy in the early-cycle stage, so we propose to train two target models, one for the early stage and the other for the late stage. Specifically, we define a change point (CP). Cycling data of 2 NCM cells at 25 °C with capacities greater than CP is used to retrain the source model and obtain the target model 1. Similarly, target model 2 is obtained by retraining the source model using the cycling data of 2 NCM cells with capacities less than CP. To make predictions on test cells at 45 °C, target model 1 is used when the capacity of the cell is above CP, and we start to use target model 2 when the capacity drops to CP. We denote our two-stage model as the TS model.

CP balances the data used to retrain the two target models, and different values of CP have a significant impact on the accuracy of our two-stage model. Fig. 10 (a) shows the RCL and RCT change with CP for model\_TS. Same as in Fig. 9 (a), we choose to show only RCL here as RCL and PECL have the same trend as CP changes. Note that when CP equals 3175 mAh, no data are used to retrain target model 1, and all data are used to retrain target model 2, so target model 1 is the source model, and target model 2 is the target model. As shown in Fig. 10 (a), both RCL and RCT first demonstrate a downward trend, followed by slight fluctuations when the CP decreases from 3175 mAh to 2900 mAh, but they cannot be optimal at one CP value.

After comprehensively considering the prediction accuracy of cycle life and capacity trajectory, 3050 mAh is selected as the best CP value for TS model in our paper. Fig. 10 (b) shows the comparison of capacity fade prediction of 3 models for one random test cell, and we can notice that the cycle life prediction error of the TS model is further improved relative to the target model. To visualize the results of capacity fade prediction, the prediction error is also shown in Fig. 10 (c). As shown in Fig. 10 (c), compared to the source model and target model, the prediction error of the TS model for most of the cycles is significantly reduced. Fig. 10 (d) shows the distribution of capacity trajectory prediction error for all cells, and most prediction errors for the TS model are in the range of -5 mAh to 30 mAh, which are better than the results of the target model. Fig. 10 (e) demonstrates the cycle life prediction result of the TS model, and it is clear that there is a significant improvement in accuracy compared to the source model and target model, which confirms the effectiveness of the proposed TS model. Specifically, as shown in the Table 5, the TS model with 3050 mAh as CP can achieve an RCL of 25.23 cycles and a PECL of 2.74% with a guaranteed high RCT accuracy (17.80 mAh, 0.51%), which are all better than source model and target model.



**Fig. 10. Performance of the proposed two-stage model:** (a) RCL and RCT change with CP for model.TS. (b) Capacity trajectory prediction and (c) prediction error for one test cell. (d) Distribution of capacity trajectory prediction error and (e) cycle life prediction for TS model.

#### 4.5. Comparison of different transfer learning strategies

To further verify the effectiveness of the proposed transfer learning strategy of retraining the last two layers of the LSTM network, we compare the performance of retraining only the last layer and retraining the last three layers, denoted as LSTM-TL1 and LSTM-TL3, respectively. The results show that LSTM-TL1 has an RCF of 0.1418 mAh, which is lower than the proposed target model from retraining the last two layers of the LSTM network, but an RCL of 42.40 cycles, an RCT of 36.61 mAh (1.05%) and a PECL of 5.62% are worse than target model. With regards to LSTM-TL3, it achieves an RCL of 24.64 cycles, an RCT of 31.12 mAh (0.89%) and a PECL of 3.29%, which are better than target model, but an RCF of 0.1607 mAh is inferior to target model. Additionally, the proposed TS model-3050 mAh outperforms LSTM-TL3 in terms of RCL, RCT and PECL.

We also apply the two-stage model to LSTM-TL1 and LSTM-TL3 to further demonstrate the superiority of the proposed method. Table 6 and Table 7 compare the performance of the two-stage model on LSTM-TL1 and LSTM-TL2 with different CPs, respectively. As shown in Table 6, for the two-stage model on LSTM-TL1, RCL, RCT and PECL first show a downward trend, followed by a slight upward trend and fluctuations when CP changes from 3175 mAh to 2900 mAh. Considering both the cycle life and capacity trajectory predictions, the optimal CP is determined to be 3000 mAh, and the model achieves an RCL of 23.75 cycles, which is slightly better than TS model-3050 mAh, but an RCT of 18.00 mAh (0.51%) and a PECL of 2.84%

are worse than the proposed TS model. As shown in Table 7, the two-stage model on LSTM-TL3 has the best capacity trajectory prediction at CP of 3000 mAh, with an RCT of 15.37 mAh (0.44%), but with remarkably high RCL of 33.57 cycles and PECL of 3.68%, which are inferior to the proposed TS model-3050 mAh. By comprehensively comparing the performance of the proposed transfer learning strategy with LSTM-TL1 and LSTM-TL2, it can be concluded that our TS model-3050 mAh provides a better balance between cycle life prediction and capacity trajectory prediction.

Table 6: Performance of the two-stage model on LSTM-TL1 with different CPs

CP	RCL (cycles)	RCT	PECL
source model	67.39	18.54 (0.53%)	9.35%
3175 mAh	39.74	34.33 (0.98%)	5.33%
3150 mAh	38.57	33.47 (0.96%)	5.20%
3125 mAh	34.38	30.53 (0.87%)	4.67%
3100 mAh	26.76	24.71 (0.71%)	3.58%
3075 mAh	25.07	22.72 (0.65%)	3.33%
3050 mAh	23.29	19.90 (0.57%)	3.07%
3025 mAh	23.15	19.81 (0.57%)	3.05%
3000 mAh	<b>23.75</b>	<b>18.00 (0.51%)</b>	<b>2.84%</b>
2975 mAh	24.24	18.19 (0.52%)	2.73%
2950 mAh	23.40	19.77 (0.56%)	2.84%
2925 mAh	24.80	18.90 (0.54%)	2.74%
2900 mAh	24.36	20.60 (0.59%)	2.71%
LSTM-TL1	42.40	36.61 (1.05%)	5.62%

Table 7: Performance of the two-stage model on LSTM-TL3 with different CPs

CP	RCL (cycles)	RCT	PECL
source model	67.39	18.54 (0.53%)	9.35%
3175 mAh	23.48	27.43 (0.78%)	3.15%
3150 mAh	23.29	26.29 (0.75%)	3.07%
3125 mAh	23.41	23.29 (0.67%)	2.89%
3100 mAh	27.24	18.21 (0.52%)	2.87%
3075 mAh	28.50	17.03 (0.49%)	3.01%
3050 mAh	31.24	15.89 (0.45%)	3.40%
3025 mAh	31.27	15.97 (0.46%)	3.42%
3000 mAh	<b>33.57</b>	<b>15.37 (0.44%)</b>	<b>3.68%</b>
2975 mAh	33.22	16.19 (0.46%)	3.65%
2950 mAh	30.50	17.89 (0.51%)	3.32%
2925 mAh	31.59	17.91 (0.51%)	3.45%
2900 mAh	29.93	19.70 (0.56%)	3.23%
LSTM-TL3	24.64	31.12 (0.89%)	3.29%

## 5. Conclusions

In this paper, an LSTM network combined with transfer learning is proposed to predict lithium-ion batteries' capacity fade and cycle life. The number of input capacities  $m$  is 13 to provide both high capacities fade and cycle life prediction accuracy. This LSTM network was initially trained on cycling data from 20 NCA cells, and after transfer learning with cycling data from 2 NCM cells at 25 °C, the RMSE of capacity fade prediction for NCM cells at 45 °C can be reduced from 0.1892 mAh to 0.1536 mAh. Feeding the LSTM neural network predictions can yield long-term capacity trajectory and cycle life prediction. Our results show that the model after transfer learning can reduce the cycle life RMSE from 67.39 cycles to 32.68 cycles with only the

first 13 initial capacities. The model prediction accuracy can be improved if we give more initial capacities, and the proposed model\_target can realize a cycle life RMSE of 17.61 cycles and a capacity trajectory RMSE of 19.55 mAh (0.56%) with the first 100 initial capacities. A two-stage model is further proposed to improve prediction accuracy, and CP is defined to balance the data used to retrain the two target models. By selecting an appropriate CP, the two-stage model achieves a cycle life RMSE of 25.23 cycles based on ensuring accurate capacity trajectory prediction (17.80 mAh, 0.51%). The success of the proposed method confirms the potential of transfer learning in modeling battery aging for different types of lithium-ion batteries. More data under different temperatures and dynamic loading profiles will be collected in the future to help the model learn different degradation patterns and improve the generalizability of the model.

## Appendix

Table A1: RCF (mAh) comparison for different  $\epsilon$  and  $C$  in SVR

$\epsilon \backslash C$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	0.2990	0.2194	0.1756	0.1526	0.1383	0.1304	0.1238	0.1192	0.1153	0.1105
0.2	<b>0.3168</b>	0.2450	0.2224	0.1984	0.1863	0.1805	0.1779	0.1717	0.1663	0.1602
0.3	0.3360	0.2794	0.2621	0.2469	0.2473	0.2507	0.2456	0.2420	0.2263	0.2200
0.4	0.3639	0.3326	0.3211	0.3147	0.3103	0.3153	0.3041	0.2874	0.2822	0.2803
0.5	0.4062	0.3852	0.3791	0.3715	0.3619	0.3577	0.3360	0.3180	0.3059	0.3008
0.6	0.4735	0.4426	0.4464	0.4597	0.4463	0.4186	0.4036	0.3926	0.3856	0.3794
0.7	0.5304	0.5087	0.5166	0.5320	0.5410	0.5132	0.4909	0.4824	0.4775	0.4672
0.8	0.5912	0.5753	0.5853	0.5987	0.6136	0.6150	0.6095	0.6091	0.6020	0.6029
0.9	0.6475	0.6425	0.6608	0.6752	0.6839	0.6849	0.6805	0.6736	0.6764	0.6818
1.0	0.7115	0.7233	0.7379	0.7598	0.7651	0.7565	0.7546	0.7518	0.7526	0.7543

Table A2: RCL (cycles) comparison for different  $\epsilon$  and  $C$  in SVR

$\epsilon \backslash C$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	26.98	26.47	41.96	72.31	85.17	111.72	128.55	152.66	157.22	169.12
0.2	<b>23.15</b>	43.64	99.70	100.39	102.13	106.63	115.65	108.03	99.14	72.24
0.3	32.33	82.30	119.09	130.23	138.20	143.48	144.54	143.45	131.63	119.82
0.4	56.26	102.46	123.66	134.47	138.96	144.24	142.20	139.41	142.96	146.06
0.5	73.41	110.48	123.26	132.77	135.86	140.05	136.56	128.20	119.23	116.45
0.6	91.18	112.55	132.38	142.79	145.58	144.88	142.79	142.79	143.93	144.94
0.7	104.04	125.81	145.33	160.71	165.61	167.33	168.74	171.97	176.18	178.33
0.8	127.90	142.48	163.51	173.32	178.93	183.19	188.81	193.06	195.21	195.91
0.9	141.03	161.32	177.49	185.25	190.16	195.12	198.67	200.08	202.89	204.29
1.0	157.07	178.85	190.82	198.59	202.80	207.06	209.59	210.58	211.28	211.99

Table A3: Performance comparison of XGBoost with different  $n_{estimators}$ 

$n_{estimators}$	RCL (mAh)	RCL (cycles)	RCT	PECL
1	0.5819	67.87	76.98 (2.20%)	15.39%
2	0.3659	28.55	46.02 (1.31%)	5.29%
3	0.3437	23.22	43.69 (1.25%)	5.54%
4	0.3292	23.22	43.26 (1.24%)	5.98%
5	<b>0.3189</b>	<b>22.62</b>	<b>43.52 (1.24%)</b>	<b>6.17%</b>
6	0.3101	22.68	44.25 (1.26%)	6.32%
7	0.3039	24.92	45.32 (1.29%)	6.77%
8	0.2975	24.11	43.99 (1.26%)	6.28%
9	0.2911	25.19	45.09 (1.29%)	6.98%
10	0.2865	24.19	44.59 (1.27%)	7.01%

### 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.

## Acknowledgements

Yankai Cao acknowledges funding from the discovery program of the Natural Science and Engineering Research Council of Canada under grant RGPIN-2019-05499. Bhushan Gopaluni would like to acknowledge the financial support from NSERC Discovery grant. Jiangong Zhu would like to thank the foundation of the National Natural Science Foundation of China (NSFC, Grant No. 52107230). We gratefully acknowledge the computing resources provided by SciNet ([www.scinethpc.ca](http://www.scinethpc.ca)) and Compute Canada ([www.computecanada.ca](http://www.computecanada.ca)).

## References

- [1] 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, *Applied energy* 232 (2018) 197–210.
- [2] 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.
- [3] M. Safari, M. Morcrette, A. Teyssot, C. Delacourt, Multimodal physics-based aging model for life prediction of li-ion batteries, *Journal of The Electrochemical Society* 156 (3) (2008) A145.
- [4] E. Sarasketa-Zabala, E. Martinez-Laserna, M. Berecibar, I. Gandiaga,

- L. M. Rodriguez-Martinez, I. Villarreal, Realistic lifetime prediction approach for li-ion batteries, *Applied energy* 162 (2016) 839–852.
- [5] L. Wu, X. Fu, Y. Guan, Review of the remaining useful life prognostics of vehicle lithium-ion batteries using data-driven methodologies, *Applied Sciences* 6 (6) (2016) 166.
- [6] M. Doyle, T. F. Fuller, J. Newman, Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell, *Journal of the Electrochemical society* 140 (6) (1993) 1526.
- [7] A. Jokar, B. Rajabloo, M. Désilets, M. Lacroix, Review of simplified pseudo-two-dimensional models of lithium-ion batteries, *Journal of Power Sources* 327 (2016) 44–55.
- [8] R. B. Smith, M. Z. Bazant, Multiphase porous electrode theory, *Journal of The Electrochemical Society* 164 (11) (2017) E3291.
- [9] P. Ramadass, B. Haran, P. M. Gomadam, R. White, B. N. Popov, Development of first principles capacity fade model for Li-ion cells, *Journal of the Electrochemical Society* 151 (2) (2004) A196.
- [10] S. Sankarasubramanian, B. Krishnamurthy, A capacity fade model for lithium-ion batteries including diffusion and kinetics, *Electrochimica Acta* 70 (2012) 248–254.
- [11] J. Purewal, J. Wang, J. Graetz, S. Soukiazian, H. Tataria, M. W. Verbrugge, Degradation of lithium ion batteries employing graphite nega-

- tives and nickel-cobalt-manganese oxide + spinel manganese oxide positives: Part 2, chemical–mechanical degradation model, *Journal of power sources* 272 (2014) 1154–1161.
- [12] J. Wang, P. Liu, J. Hicks-Garner, E. Sherman, S. Soukiazian, M. Verbrugge, H. Tataria, J. Musser, P. Finamore, Cycle-life model for graphite-LiFePO<sub>4</sub> cells, *Journal of power sources* 196 (8) (2011) 3942–3948.
- [13] J. Wang, J. Purewal, P. Liu, J. Hicks-Garner, S. Soukiazian, E. Sherman, A. Sorenson, L. Vu, H. Tataria, M. W. Verbrugge, Degradation of lithium ion batteries employing graphite negatives and nickel-cobalt-manganese oxide + spinel manganese oxide positives: Part 1, aging mechanisms and life estimation, *Journal of Power Sources* 269 (2014) 937–948.
- [14] L. Lam, P. Bauer, Practical capacity fading model for Li-ion battery cells in electric vehicles, *IEEE transactions on power electronics* 28 (12) (2012) 5910–5918.
- [15] S. Saxena, C. Hendricks, M. Pecht, Cycle life testing and modeling of graphite/LiCoO<sub>2</sub> cells under different state of charge ranges, *Journal of Power Sources* 327 (2016) 394–400.
- [16] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, K. Dietmayer, Health diagnosis and remaining useful life prognostics of lithium-ion bat-

- teries using data-driven methods, *Journal of power sources* 239 (2013) 680–688.
- [17] J. Wu, C. Zhang, Z. Chen, An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks, *Applied energy* 173 (2016) 134–140.
- [18] K. Liu, Y. Shang, Q. Ouyang, W. D. Widanage, A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery, *IEEE Transactions on Industrial Electronics* 68 (4) (2020) 3170–3180.
- [19] D. Wang, Q. Miao, M. Pecht, Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model, *Journal of Power Sources* 239 (2013) 253–264.
- [20] M. A. Patil, P. Tagade, K. S. Hariharan, S. M. Kolake, T. Song, T. Yeo, S. Doo, A novel multistage Support Vector Machine based approach for lithium-ion battery remaining useful life estimation, *Applied energy* 159 (2015) 285–297.
- [21] D. Gao, M. Huang, Prediction of remaining useful life of lithium-ion battery based on multi-kernel support vector machine with particle swarm optimization, *Journal of Power Electronics* 17 (5) (2017) 1288–1297.
- [22] D. Liu, J. Pang, J. Zhou, Y. Peng, M. Pecht, Prognostics for state of health estimation of lithium-ion batteries based on combination gaussian

- process functional regression, *Microelectronics Reliability* 53 (6) (2013) 832–839.
- [23] Y.-J. He, J.-N. Shen, J.-F. Shen, Z.-F. Ma, State of health estimation of lithium-ion batteries: A multiscale gaussian process regression modeling approach, *AIChE Journal* 61 (5) (2015) 1589–1600.
- [24] R. R. Richardson, M. A. Osborne, D. A. Howey, Gaussian process regression for forecasting battery state of health, *Journal of Power Sources* 357 (2017) 209–219.
- [25] Y. Zhang, R. Xiong, H. He, M. G. Pecht, Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries, *IEEE Transactions on Vehicular Technology* 67 (7) (2018) 5695–5705.
- [26] X. Li, L. Zhang, Z. Wang, P. Dong, Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and elman neural networks, *Journal of Energy Storage* 21 (2019) 510–518.
- [27] L. Mao, J. Xu, J. Chen, J. Zhao, Y. Wu, F. Yao, A LSTM-STW and GS-LM fusion method for lithium-ion battery RUL prediction based on EEMD, *Energies* 13 (9) (2020) 2380.
- [28] R. R. Richardson, C. R. Birkl, M. A. Osborne, D. A. Howey, Gaussian

- process regression for in situ capacity estimation of lithium-ion batteries, *IEEE Transactions on Industrial Informatics* 15 (1) (2018) 127–138.
- [29] J. Zhu, Y. Wang, Y. Huang, R. Bhushan Gopaluni, Y. Cao, M. Heere, M. J. Mühlbauer, L. Mereacre, H. Dai, X. Liu, et al., Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation, *Nature Communications* 13 (1) (2022) 1–10.
- [30] C. She, L. Zhang, Z. Wang, F. Sun, P. Liu, C. Song, Battery state of health estimation based on incremental capacity analysis method: Synthesizing from cell-level test to real-world application, *IEEE Journal of Emerging and Selected Topics in Power Electronics* (2021).
- [31] Z. Wang, C. Song, L. Zhang, Y. Zhao, P. Liu, D. G. Dorrell, A data-driven method for battery charging capacity abnormality diagnosis in electric vehicle applications, *IEEE Transactions on Transportation Electrification* 8 (1) (2021) 990–999.
- [32] Q. Wang, Z. Wang, L. Zhang, P. Liu, L. Zhou, A battery capacity estimation framework combining hybrid deep neural network and regional capacity calculation based on real-world operating data, *IEEE Transactions on Industrial Electronics* (2022).
- [33] 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 pre-

- diction of battery cycle life before capacity degradation, *Nature Energy* 4 (5) (2019) 383–391.
- [34] Y. Zhang, Z. Peng, Y. Guan, L. Wu, Prognostics of battery cycle life in the early-cycle stage based on hybrid model, *Energy* 221 (2021) 119901.
- [35] X. Tang, K. Liu, X. Wang, F. Gao, J. Macro, W. D. Widanage, Model migration neural network for predicting battery aging trajectories, *IEEE Transactions on Transportation Electrification* 6 (2) (2020) 363–374.
- [36] S. Shen, M. Sadoughi, M. Li, Z. Wang, C. Hu, Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries, *Applied Energy* 260 (2020) 114296.
- [37] D. R. Sørensen, M. Heere, J. Zhu, M. S. Darma, S. M. Zimnik, M. J. Mühlbauer, L. Mereacre, V. Baran, A. Senyshyn, M. Knapp, et al., Fatigue in high-energy commercial li batteries while cycling at standard conditions: an in situ neutron powder diffraction study, *ACS Applied Energy Materials* 3 (7) (2020) 6611–6622.
- [38] W. H. Press, S. A. Teukolsky, Savitzky-Golay smoothing filters, *Computers in Physics* 4 (6) (1990) 669–672.
- [39] Z. C. Lipton, J. Berkowitz, C. Elkan, A critical review of recurrent neural networks for sequence learning, arXiv preprint arXiv:1506.00019 (2015).
- [40] M. Vathsala, G. Holi, Rnn based machine translation and translitera-

- tion for twitter data, *International Journal of Speech Technology* 23 (3) (2020) 499–504.
- [41] D. Gillick, C. Brunk, O. Vinyals, A. Subramanya, Multilingual language processing from bytes, *arXiv preprint arXiv:1512.00103* (2015).
- [42] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural computation* 9 (8) (1997) 1735–1780.
- [43] D. Yang, X. Zhang, R. Pan, Y. Wang, Z. Chen, A novel gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve, *Journal of Power Sources* 384 (2018) 387–395.
- [44] S. J. Pan, Q. Yang, A survey on transfer learning, *IEEE Transactions on knowledge and data engineering* 22 (10) (2009) 1345–1359.
- [45] S. Ruder, M. E. Peters, S. Swayamdipta, T. Wolf, Transfer learning in natural language processing, in: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials, 2019*, pp. 15–18.
- [46] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, *Advances in neural information processing systems* 27 (2014).
- [47] M. Hon, N. M. Khan, Towards alzheimer’s disease classification through transfer learning, in: *2017 IEEE International conference on bioinformatics and biomedicine (BIBM), IEEE, 2017*, pp. 1166–1169.

- [48] N. Wang, G. Zhao, Y. Kang, W. Wang, A. Chen, B. Duan, C. Zhang, Core temperature estimation method for lithium-ion battery based on long short-term memory model with transfer learning, *IEEE Journal of Emerging and Selected Topics in Power Electronics* (2021).
- [49] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *Journal of the royal statistical society: series B (statistical methodology)* 67 (2) (2005) 301–320.
- [50] M. Awad, R. Khanna, Support vector regression, in: *Efficient learning machines*, Springer, 2015, pp. 67–80.
- [51] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.