

PAPER

## Data-driven estimation of battery state-of-health with formation features

To cite this article: Weilin He *et al* 2024 *J. Micromech. Microeng.* **34** 075004

View the [article online](#) for updates and enhancements.

### You may also like

- [Improved Particle Swarm Optimization-Extreme Learning Machine Modeling Strategies for the Accurate Lithium-ion Battery State of Health Estimation and High-adaptability Remaining Useful Life Prediction](#)  
Chu-yan Zhang, Shun-li Wang, Chun-mei Yu *et al.*
- [State of Health Prediction of Lithium-Ion Batteries Using Combined Machine Learning Model Based on Nonlinear Constraint Optimization](#)  
Yawen Liang, Shunli Wang, Yongcun Fan *et al.*
- [State-of-Health Prediction for Lithium-Ion Batteries based on Empirical Mode Decomposition and Bidirectional Gated Recurrent Unit Neural Network Optimized by Slime Mould Algorithm](#)  
Jing Sun and Xiaodong Zhang



**PRIME<sup>TM</sup>**  
**PACIFIC RIM MEETING**  
ON ELECTROCHEMICAL  
AND SOLID STATE SCIENCE



**HONOLULU, HI**  
October 6-11, 2024

*Joint International Meeting of*  
The Electrochemical Society of Japan (ECSJ)  
The Korean Electrochemical Society (KECS)  
The Electrochemical Society (ECS)

Early Registration Deadline:  
**September 3, 2024**

**MAKE YOUR PLANS  
NOW!**

# Data-driven estimation of battery state-of-health with formation features

Weilin He<sup>1</sup>, Dingquan Li<sup>2</sup>, Zhongxian Sun<sup>1</sup>, Chenyang Wang<sup>3</sup>, Shihai Tang<sup>3</sup>, Jing Chen<sup>3</sup>, Xin Geng<sup>3</sup>, Hailong Wang<sup>3</sup>, Zhimeng Liu<sup>3</sup>, Linyu Hu<sup>4</sup>, Dongchen Yang<sup>5</sup>, Haiyan Tu<sup>1</sup>, Yuanjing Lin<sup>4</sup>  and Xin He<sup>1,3,\*</sup> 

<sup>1</sup> College of Electrical Engineering, Sichuan University, Chengdu 610065, People's Republic of China

<sup>2</sup> Peng Cheng Laboratory, Shenzhen 518055, People's Republic of China

<sup>3</sup> School of Chemical Engineering, Sichuan University, Chengdu 610065, People's Republic of China

<sup>4</sup> School of Microelectronics, Southern University of Science and Technology, Shenzhen 518055, People's Republic of China

<sup>5</sup> Faculty of Engineering, Architecture and Information Technology (EAIT), University of Queensland, St Lucia, QLD 4072, Australia

E-mail: [xinhe@scu.edu.cn](mailto:xinhe@scu.edu.cn)

Received 13 April 2024, revised 16 May 2024

Accepted for publication 30 May 2024

Published 17 June 2024



## Abstract

Accurately estimating the state-of-health (SOH) of a battery is crucial for ensuring battery safe and efficient operation. The lifetime of lithium-ion batteries (LIBs) starts from their manufacture, and the performance of LIBs in the service period is highly related to the formation conditions in the factory. Here, we develop a deep transfer ensemble learning framework with two constructive layers to estimate battery SOH. The primary approach involves a combination of base models, a convolutional neural network to combine electrical features with spatial relationships of thermal and mechanical features from formation to subsequent cycles, and long short-term memory to extract temporal dependencies during cycling. Gaussian process regression (GPR) then handles SOH prediction based on this integrated model. The validation results demonstrate highly accurate capacity estimation, with a lowest root-mean-square error (RMSE) of 1.662% and a mean RMSE of 2.512%.

Characterization on retired cells reveals the correlation between embedded formation features and their impact on the structural, morphological, and valence states evolution of electrode material, enabling reliable prediction with the corresponding interplay mechanism. Our work highlights the value of deep learning with comprehensive analysis through the relevant features, and provides guidance for optimizing battery management.

Supplementary material for this article is available [online](#)

Keywords: state-of-health estimation, ensemble model, formation cycle, thermal feature, mechanical feature

## 1. Introduction

Lithium-ion batteries (LIBs) are widely utilized in consumer electronics, energy storage systems, and electric vehicles due

to their high energy densities, low costs, and long lifetimes [1, 2]. However, LIBs still face practical issues, such as electrolyte decomposition [3, 4], loss of active materials [5], formation of dead Li, and growth of dendritic Li [6, 7] that may continue to occur during the cycling process. These lead to capacity fading over time and may result in severe safety concerns [8]. The accumulation of irreversible reactions is primarily responsible

\* Author to whom any correspondence should be addressed.

for the structural degradation, but they are collected in the form of electrical performance in commercialized battery management system (BMS) [9]. Therefore, the term of battery state-of-health (SOH) that defined with the ratio between the present capacity and the initial capacity [10–12], is challenging to achieve highly accurate estimation merely based on the limited electrical feedbacks. In addition, the aging process highly depends on the usage activities and operating environmental condition [13, 14]. For example, LIBs charged at high current rates or low temperatures are easier to induce Li-plating [15–17], which leads to short circuit of battery. On the contrary, the high temperature accelerates electrolyte decomposition, which increases the internal resistance [18, 19] and raises the risk of thermal abuse. As a result, capacity fading is hard to predict and it typically follows a non-linear trend. The occurrence of electrical, mechanical and thermal failures present tremendous barriers in achieving precise SOH estimation.

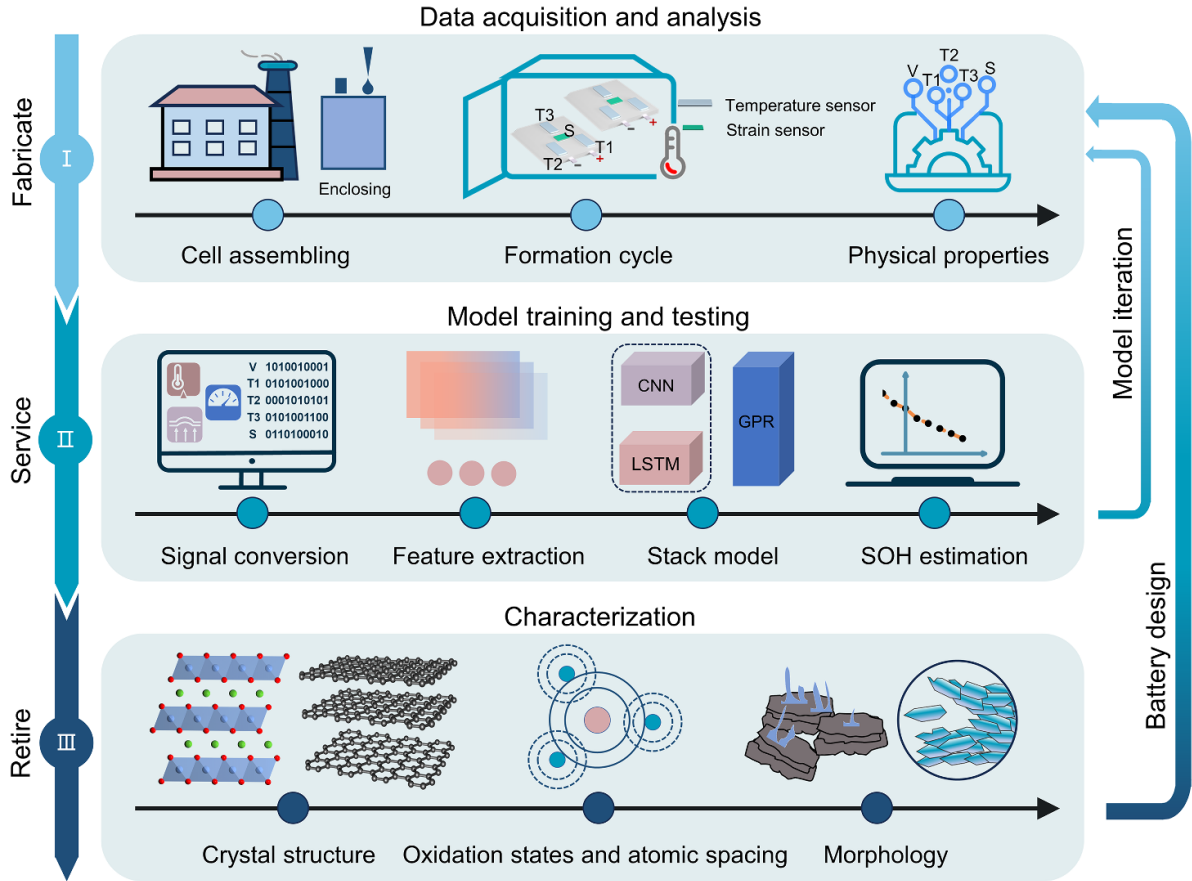
When it comes to handling datasets with complex patterns, machine learning (ML) emerges as a powerful approach [20–22]. The model is capable of adapting intricate relationships between inputs and outputs, allowing the study to continue without being confined by pre-existing knowledge in the system [22–24]. Each model possesses unique advantages, necessitating a thoughtful consideration of suitability for certain scenarios. Given the diverse nature of distinct features, it is evident that no single model is universally applicable to all cases [25]. For instance, the convolutional neural networks (CNN) model, utilizing the increment capacity (IC) features extracted from the specific cells in the NASA dataset, achieves superior accuracy in capacity estimation compared to the long short-term memory (LSTM), support vector machine (SVM), and random forest models [26]. However, when applying the IC features to other cells in the dataset, the LSTM model demonstrates higher accuracy than SVM and artificial neural network models [27]. Generally, an integration of different models could enhance the accuracy and robustness in predicting capacity across various applications.

Most of current battery SOH estimation is proceeded on the basis of electrical features as inputs, which are collected by BMS during their service period [28]. The practical usage of LIBs involves various operating protocols and complex aging conditions in most applications, with only a partial section of stored energy being utilized according to the user activity. Compared to the dataset collected from real scenarios, the data generated during formation cycles, which comes from the battery manufacturer, is known to contain complete cycles with full voltage range. Thus, the dataset collected from the formation cycles by the manufacturers can be utilized to fill up the deficiency of training cases. More importantly, the solid electrolyte interphase (SEI) layer, which works as an effectively prohibition of further electrolyte decomposition and enables Li-ion diffusion in-between the electrode and electrolyte during the electrochemical process [29–31], is formed during the formation process. The SEI layer plays a critical role on the Li-ions diffusion dynamics and maintains interfacial stability in subsequent cycles [32]. The structure, morphology, and chemical composition of SEI layer is determined by various factors, leading to distinct interfacial

properties [33]. It has been reported that SEI layer formed at 25 °C has a stable LiF-rich inorganic film with medium thickness and low energy barriers facilitates ionic diffusion, resulting good electrochemical performance. This result reveals a strong correlation between the formation temperature and interfacial chemistry [34]. Besides the factor of temperature, the duration affects the interfacial chemistry as well. The intermediate formation protocol with a duration of 26–30 h facilitates capacity retention with minimal impedance growth. In contrast, shorter formation time leads to poorer electrochemical performance, and an extended formation time has no further contribution for performance improvement [35]. Above all, the formation process is essential for creating a stable SEI layer and minimizing irreversible reactions over the lifetime of LIBs.

Considering the significance of the formation process, collecting a comprehensive dataset, which includes electrical, thermal, and mechanical features to enrich the data dimensions, is believed to facilitate the improvement on the accuracy of SOH estimation. However, the functional mechanism and the weight of each feature on SOH estimation are still unclear, which may mislead the model construction or result a reduced accuracy on certain circumstance. The end-of-life characterizations on retired electrodes have been proven to be a feasible approach for enabling a clear interpretation of the capacity fading mechanism. Since the cycling performance is strongly related to the formation process, they are believed to have certain interplay with the relevant formation features. Thus, analyzing the linear correlation coefficient may help identify battery failures along with their corresponding features, providing guidance on model construction to improve estimation accuracy.

In this work, we develop a deep transfer ensemble learning framework with two constructive layers to estimate battery SOH. The primary layer involves a combination of CNN model and LSTM model as integrated base models for ML approach. The CNN model is capturing electrical feature and spatial relationships of thermal and mechanical features from formation to subsequent cycles, while LSTM model is functionalized to extract temporal dependencies during cycling. Then the Gaussian process regression (GPR) in second layer handles SOH prediction with this integrated base models. As depicted in figure 1, we start to collect the real-time physical parameters such as surface temperatures, surface strain, and electrochemical data of current, voltage, and capacity and transform as digital datasets. From the above data, features with spatial relationships and temporal dependencies are extracted for the input of the integrated base models to accurately estimate capacity. Since the physical and electrochemical parameters of cells are essentially determined by the chemical composition, microstructure, and morphology of the electrode materials, the correlation of the chemical characterization results to the formation and cycling information is further analyzed at the end-of-life. It can bridge the gap between chemical analysis and data modeling towards the precise prediction of the SOH of cells, benefitting the advancement of the design of high-performance batteries and high-accuracy BMS.



**Figure 1.** Flowchart of the deep transfer ensemble framework for battery SOH estimation. I. Data acquisition and analysis for the formation process by the manufacture. II. Model training and testing based on the electrical, thermal and mechanical features. III. Chemical and structural characterization after disassembling to reveal the correlation coefficient.

## 2. Results and discussion

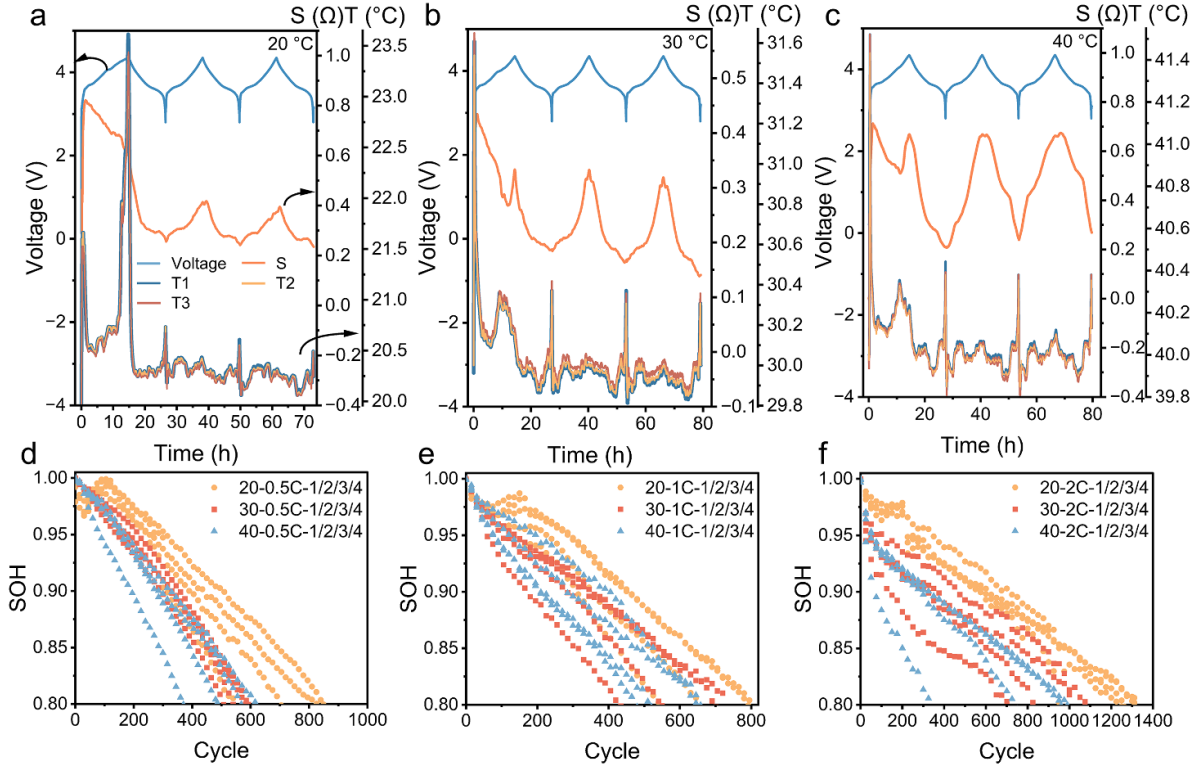
### 2.1. Data generation

From a ML perspective, all the relevant features associated with the electrochemical process have the potential to significantly impact the accuracy of SOH estimation. To fully leverage the potential value of these features to the best advantage, formation and cycling data of 36 lab-assembled pouch cells (NCM523/graphite, nominal capacity 1.1 Ah) were collected for model training and testing. The cells are denoted as  $T-C-N$ , where  $T$  represents the environmental temperature during formation,  $C$  stands for the current rate during cycling, and  $N$  denotes the sequence number of the cells operated under identical conditions. As illustrated in supplementary table 1, 40 pouch cells were divided into five groups based on the formation temperature (0 °C, 20 °C, 30 °C, 40 °C and 60 °C). However, the formation cycles proceeded at 0 °C arise incomplete decomposition/reaction of the electrolyte, resulting in the generation of organic-rich SEI layers that increase the resistance for Li-ions kinetics [36]. As a consequence, the pouch cells experienced severe swelling, accompanied by a remarkably low coulombic efficiency of 17.613% in the first cycle (supplementary figure 1). Formation

cycles at 60 °C display a low initial capacity (0.840 Ah), along with a sustained capacity increasing trend (supplementary figure 2). This possibly attribute to the formation of thick SEI layer at high temperature [37]. Therefore, only 36 cells were utilized in further step. To be more realistic for the practical application, the formation temperature is confined in the optimized temperature range from 20 °C to 40 °C, which is capable of offering capacities close to the nominal level (1.1 Ah). Each group is further classified into three sub-groups according to the cycling current rates (0.5 C, 1 C and 2 C). Within each sub-group, there are 4 parallel cells operated under the same condition, each marked with a sequence number. The cells numbered 1, 2, and 3 from each sub-group are utilized for training the model, while cell numbered 4 is reserved for testing the performance of the trained model.

To systematically track the thermal and mechanical responses during the formation process, spatially distributed temperature sensors and strain sensor are employed to record the temperature and strain variation of the cells (supplementary figure 3). Figures 2(a)–(c) display the charge-and-discharge profiles associating with real-time temperature and strain variations of three representative cells throughout three formation cycles with the formation temperature at 20 °C, 30 °C, and 40 °C. Obviously, the thermal and mechanical





**Figure 2.** The performance of cells at formation and subsequent cycles. The charge and discharge profiles, temperatures ( $T1$ ,  $T2$ ,  $T3$ ) curves, and strain ( $S$ ) curve of representative cells when formation process are proceeded at (a) 20 °C, (b) 30 °C, and (c) 40 °C. Real SOH as the function of strain of cycle numbers at the cycling current rate of (d) 0.5C, (e) 1C, and (f) 2C under distinct formation temperatures.

index curve performs inconsistent patterns as the first cycle charge-and-discharge profile, which is actually caused by the decomposition of electrolyte and initial formation of SEI layer. Though these processes are not directly reflected on the electrochemical profile, the thermal and mechanical curve has provided supplementary information to reveal the corresponding mechanisms. The gas generated by electrolyte decomposition [38, 39], induces a dramatic increase in strain at the initiation stage. Subsequently, the gas is expelled into the gas bag due to the applied external pressure, gradually decreasing the surface strain even with continues Li-ions insertion into graphite anode. The inner pressure will eventually reach a state close to equilibrium at which point the process is dominated by the Li-ions intercalation in the form of a rapid growth of strain [40, 41].

Meanwhile, temperatures experience large-scale fluctuations due to the inevitable heat generation is during the electrochemical process. The heat generation equation can be expressed as follows:

$$Q = I(V - U) + IT \frac{\partial U}{\partial T} - \sum_i \Delta H_i^{\text{avg}} r_i - \int \sum_j (\bar{H}_j - \bar{H}_j^{\text{avg}}) \frac{\partial c_j}{\partial t} dv \quad (1)$$

where  $I$  stands for the operating current,  $U$  represents the open voltage,  $V$  represents the operating voltage,  $T$  means ambient

temperature, and  $\frac{\partial U}{\partial T}$  means voltage temperature coefficient [42].

The term  $I(V-U)$  refers to irreversible heat generated from resistive dissipation. This over-potential determines the internal heat generation and is performed as the multiplication of operating current  $I$  and internal resistance  $R$  (including Joule resistance, polarizable resistance, and contact resistance), thus this part can be conducted as a linear increment with the square of current. The term  $IT \frac{\partial U}{\partial T}$  is assigned to the reversible entropic heat, and it remains the endothermic nature of the charge chemical reaction and exothermic during discharge, which increases linearly with the current. The part  $\sum_i \Delta H_i^{\text{avg}} r_i$  refers to heat production or consumption by any chemical reactions [42]. The last term  $\int \sum_j (\bar{H}_j - \bar{H}_j^{\text{avg}}) \frac{\partial c_j}{\partial t} dv$  represents the heat of mixing [43], which can be neglected under regular conditions [42]. The effect of heat generation on temperature is more complicated even at the current rate of 0.1 C during formation cycles, especially for the first cycle, with increased internal resistance due to electrolyte decomposition and SEI layer formation. Though cells deliver same level of capacity, they perform distinct thermal and mechanical patterns that during formation cycle. These features provide more instant and richer information on internal reactions compared to the electrical features, and the summarized dataset of those signals can be fruitfully exploited by our integrated model.

After the first formation cycle, the electrode surface is passivated by the SEI layer to prevent further electrolyte decomposition. The strain response is consistent with

charge-discharge profile in subsequent cycles, correlating with the highly reversible lithiation and delithiation processes [44–46]. The differences in the pattern and intensity of the strain profiles attribute to the interfacial solvation structure and ionic diffusion dynamics [38, 44]. Though temperature sensors demonstrate similar trend under different formation temperatures, the heat generation is far more complicated, which influenced by the state of charge, operating temperature, thickness of SEI layer and other relevant factors [47–49]. The gradient of current flow through the current collector and distribution of heterogeneous SEI layer mainly responsible for the slightly difference of spatially distributed temperature sensors, which is consistent with the observation from infrared thermography images (supplementary figure 4), could be an indicator of internal thermal feature.

The normalized SOH shown in figures 2(d)–(f) reflects the capacity fading index of the batteries with slow time-varying dynamics. In order to evaluate the cycling performance, all the cells are cycled at 30 °C and stopped as the SOH drops to 80% of the initial level. After formation at 20 °C, the cells demonstrate a better capacity retention compare to the cases at 30 °C and 40 °C. Despite formation at 20 °C causes lower capacity at varied cycling current rates due to the existence of LiF-rich SEI layer (supplementary figure 5), the integrity and stability of interface is well maintained to support a longer cycling life. This result provides clear evidence on the fact that the formation process has a great impact on the performance of battery in subsequent cycles. The information generated during the formation process is critical vital for the SOH estimation.

## 2.2. Model building

The dataset generated during the formation and cycling processes is not suitable for direct input to the base models. Handling such large datasets can indeed be computationally intensive and may pose challenges in terms of storage, processing power, and memory requirements. In addition, the raw data may lack clearly valid information; it may be noisy, ambiguous, or may not conform to the quality standards required for effective input into the base models. Valid and reliable information is crucial for training models effectively, and if the dataset contains inaccuracies or uncertainties, it can negatively impact the performance of the models. Therefore, it is necessary to transform it into an appropriate format for effective utilization, which mainly involves four steps. Firstly, five representative feature curves are extracted from the formation process, encompassing voltage (figure 3(a)), temperature of  $T_1$ ,  $T_2$ , and  $T_3$  (figure 3(b) depicts the average temperature of  $T_1$ ,  $T_2$ , and  $T_3$ ), and strain (figure 3(c)). Then, these curves are fitted over time and linearly interpolated to a fixed length of 100 to ensure the consistency of the model, generating a formation information sample with a shape of  $1 \times 3 \times 100 \times 5$  (i.e. 1 sample, 3 cycles, 100 sampling points and 5 feature profiles). Figure 3(d) illustrates the charging profiles with the cycling current rate of 1C. These curves demonstrate consistently systematic changes, suggesting a robust correlation between voltage and cycle number. Therefore, the data collected within the incomplete voltage ranges during the cycling process is

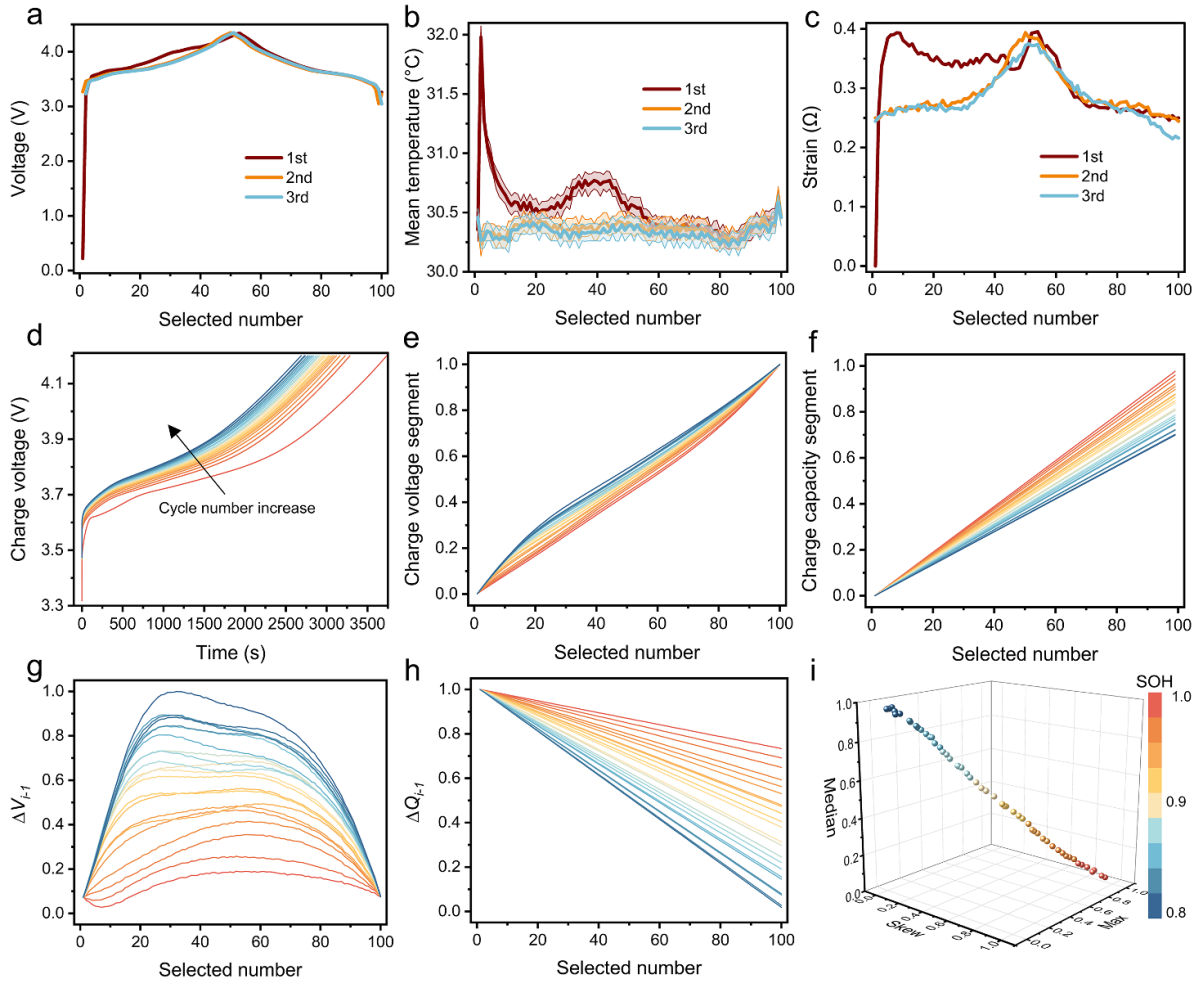
subsequently transformed into an appropriate segment format. Four electrical features are extracted spanning 3.70 V–3.85 V of the incremental charge profiles, including the partial charge voltage curves (figure 3(e)), partial charge capacity profiles (figure 3(f)), difference of charge voltage profiles between each cycle and the first cycle ( $\Delta V_{i-1}$ ) (figure 3(g)), and difference of charge capacity profiles between each cycle and the first cycle ( $\Delta Q_{i-1}$ ) in (figure 3(h)), are linearly interpolated to a fixed length of 100, resulting in a cycling information sample with a shape of  $1 \times 100 \times 4$  (i.e. 1 sample, 100 sampling points and 4 feature profiles).

To construct the model framework for this work, CNN and LSTM are employed as the base models in the deep transfer ensemble learning approach. The CNN capture source from both formation and cycling data as input, where the cycling information sample are replicated 3 times to generate a new sample with the shape of  $1 \times 3 \times 100 \times 4$ , with the aim of being horizontally concatenated with the formation information sample ( $1 \times 3 \times 100 \times 5$ ) to compose the CNN input ( $1 \times 3 \times 100 \times 9$ ). The target-labeled of the CNN is the discharge capacity of present cycle. The LSTM input is extracted from cycling information due to its capability of temporal processing, where three statistical features associated with capacity, including the median and skewness (Skew) from the charge voltage curves, and the maximum value (Max) from the charge capacity curves are extracted (figure 3(i)). The sequence length of the LSTM input is 10, and thus each LSTM input sample has a shape of  $1 \times 10 \times 3$  (i.e. 1 sample, 10 cycles and 3 features). The target-labeled of LSTM is the discharge capacity of present cycle.

The outputs of the above base models are handled by GPR. Hence the inputs of the proposed model include the inputs from CNN and LSTM. Afterwards, the deep transfer ensemble learning model is developed by training the CNN and LSTM models once, and the GPR model twice. In specific, foundational training and testing of the proposed model are performed using the datasets of 2C at first, and the trained models (CNN, LSTM and GPR) are saved in this process. Then the base models are employed to estimate the datasets of 0.5C and 1C, and these results are utilized for secondary training of GPR, which is also saved after this process. This training approach reduces the computational resources required for model training, and facilitates the accurate estimations of battery SOH at various formation temperatures and current rates.

## 2.3. Battery SOH estimation

The effectiveness of ensemble learning and transfer learning is firstly verified by the minimum average RMSE shown in supplementary table 2 and supplementary table 3. Our GPR model exhibits the best estimation accuracy among the ensemble models, including the least square regression, kernel ridge regression, GPR, and the single CNN or LSTM model. Thus, GPR is integrated to the base models and then transfer learning is utilized to enhance the estimation accuracy. After training with the extracted relevant features from available dataset, then the proposed model is utilized to estimate the capacity of cells with sequence number 4. The algorithm performance on



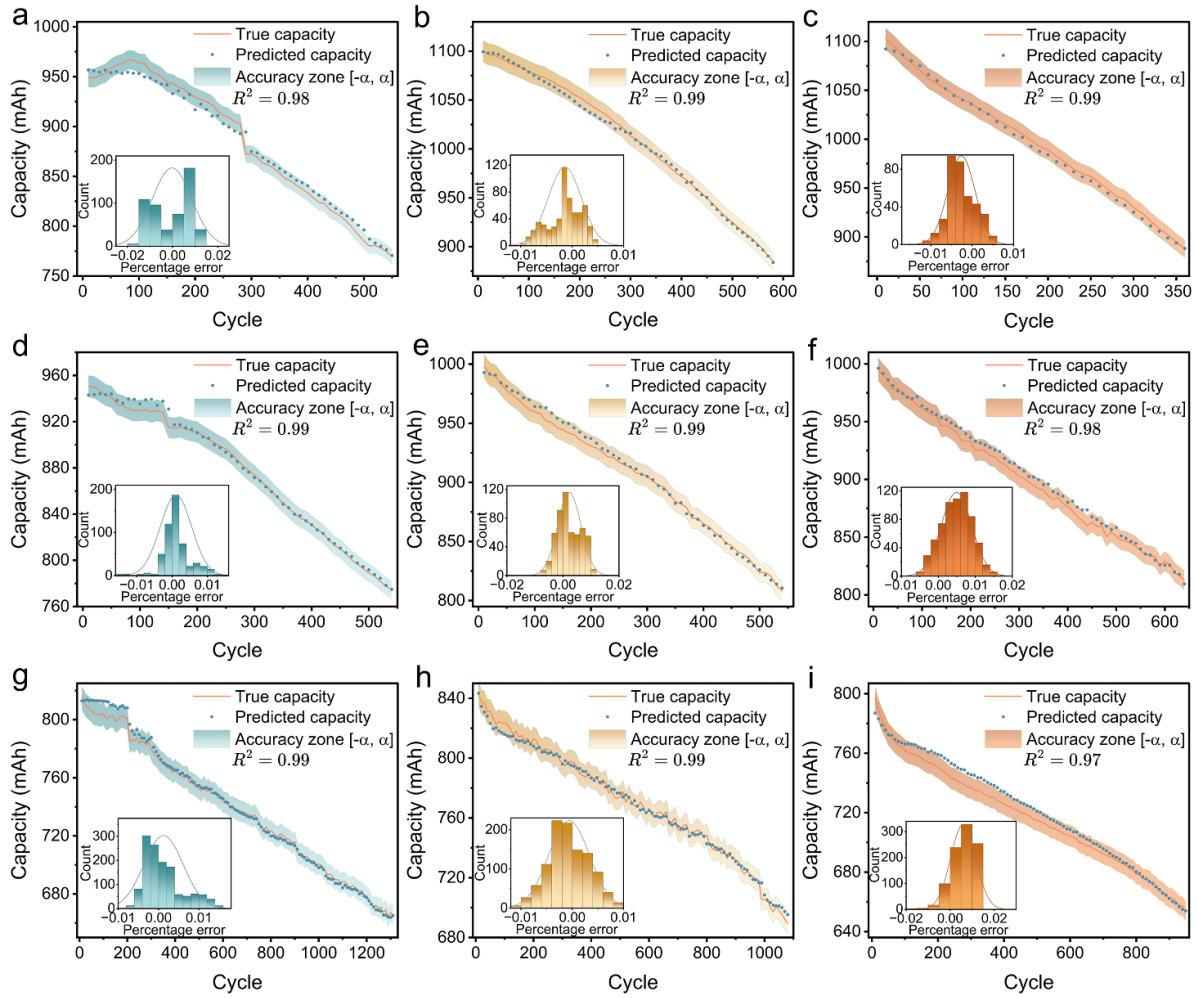
**Figure 3.** Extracted features from the formation and cycling process. Cell 30-1C-1 is employed as a representative sample. During the formation cycles, the extracted features include: (a) the voltage profile, (b) the average temperature curve of three points, and (c) the strain curve. The shaded regions in (b) denote the margin of error. (d) The charge profiles during cycling. Four corresponding features spanning 3.70 V–3.85 V: (e) partial charge voltage profiles, (f) partial charge capacity profiles, (g) difference of charge voltage profiles between each cycle and the first cycle ( $\Delta V_{i-1}$ ) and (h) difference of charge capacity curves between each cycle and the first cycle ( $\Delta Q_{i-1}$ ). (i) Extracted statistical features from (e) partial charge voltage profiles and (f) partial charge capacity profile, encompassing median, skewness (Skew), and maxima (Max) in relation to the SOH of cell.

the variation of estimated and actual capacity as a function of cycle numbers is shown in figure 4 (the scatter plot of estimated and true values in shown supplementary figure 6), and the average RMSE of each cell is summarized in supplementary table 2. Here, most of the predicted capacities closely follow the trajectories of the delivered true capacities, which display the relative errors within the interval of  $[-1\%, 1\%]$ .

The prediction errors are relatively large in the early stages, but the convergence rate gradually increases, and the error bars more effectively capture the variability associated with battery aging. Thus, it is more reliable for predicting the SOH after hundreds of cycles, especially when the cells are close to failure. The model validation achieves an RMSE of 2.512% based on the estimation results from all the cells with sequence number 4, and the optimal estimation result displays an RMSE of 1.662%, demonstrating the effectiveness of the proposed model and indicating that the capacity for each cycle of the new cell is estimated precisely. It is worth to mention that the

accuracy of estimation is highly dependent on the formation and operating conditions according to our results.

The estimation capacity for the dataset at the formation temperature of 40 °C displays a decreased RMSE with increased current densities. The RMSE at 0.5C is only 1.887% (figure 4(c)), whereas the 1C and 2C dataset yields an RMSE of 2.843% (figure 4(f)) and 3.716% (figure 4(i)), respectively. Under identical cycling conditions with varied formation temperatures, the similar phenomenon is also observed. The estimation results corresponding to the formation temperature of 20 °C and the cycling current rate of 1C registers an RMSE of 2.197% (figure 4(d)), while the one at formation temperature of 30 °C and 40 °C registers an RMSE of 2.207% (figure 4(e)) 2.843% (figure 4(f)), respectively. However, we should also notice that there is a variation on the accuracy of individual cells, which may relate to the cell fabrication and the occurrence of independent random event. In the real scenarios, power failures occur (as depicted in figures 4(a), (d),



**Figure 4.** The validation of testing cells. The estimated capacity and true capacity for (a) 20–0.5C, (b) 30–0.5C, (c) 40–0.5C, (d) 20–1C, (e) 30–1C, (f) 40–1C, (g) 20–2C, (h) 30–2C and (i) 40–2C. The graph embedded in (a-i) represents the percentage error histogram corresponding to the estimation result. The shaded regions in (a-i) denote the  $\pm 1\%$  error range of the true capacity.

(g)), which disrupt the consistency of original capacity fading trend. Subsequently, the system enters an almost unpredictable initial new trend with huge capacity gap, posing more complex but practical cases for the model training.

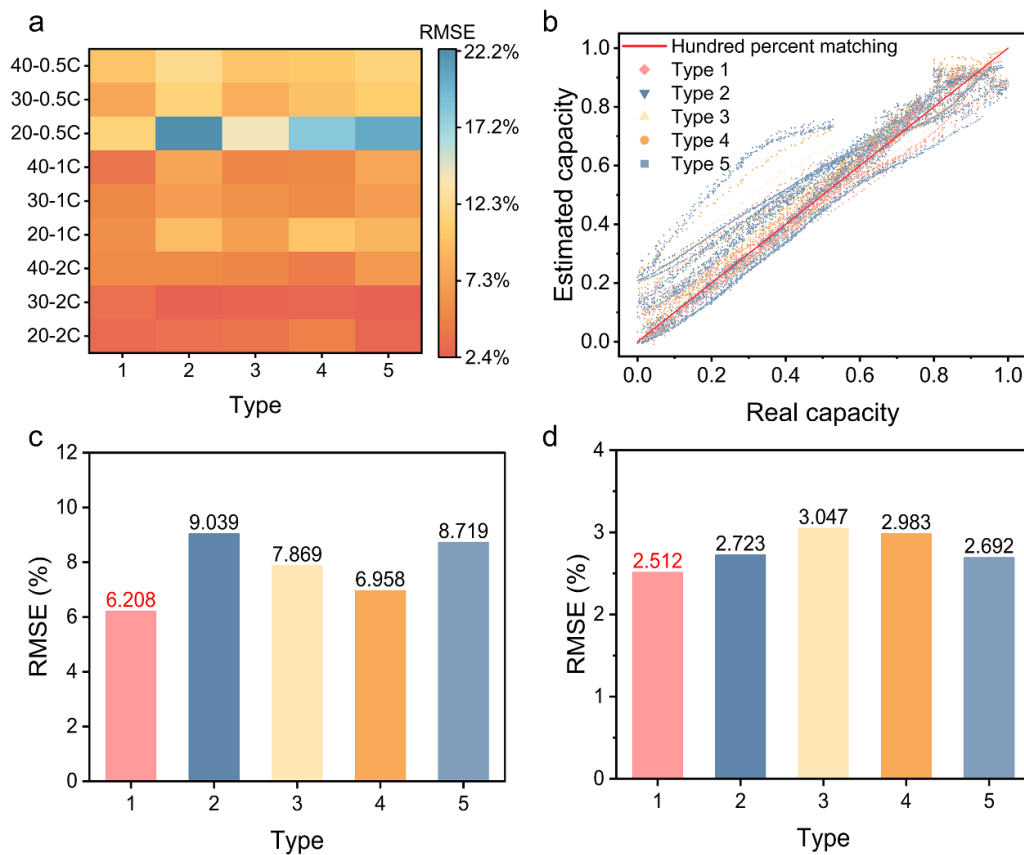
#### 2.4. Impact on the dimension of relevant features

The potential impact of formation information inputs on model performance is further evaluated by comparing their corresponding accuracy of SOH estimation. Five types of dataset inputs are investigated, which are categorized as follows: (1) temperature, strain, and voltage; (2) none; (3) strain and voltage; (4) temperature and voltage; (5) voltage. Figure 5(a) illustrates the RMSE derived from estimation results of 9 cells that employ these 5 types of inputs. Despite the fullest formation information used for type1, not all cells yield the lowest RMSE (figure 5(a)). This partly attributes to the inherent uncertainty of ML as a black-box approach [50, 51]. However, individual differences cannot stand for the statistical result from all the sample cells. Figure 5(b) visually presents the normalized actual capacity versus predicted capacity results of the

9 cells based on these 5 types of input. It is apparent that type1 generally yield estimations closest to true values, indicating an enhancement in model accuracy attributed to the inclusion of all formation information. The average RMSE obtained from validation for the formation datasets with sequence number 4 with CNN model is shown in figure 5(c) and supplementary table 4, we should notice that adding a single thermal feature or mechanical features on the basis of voltage input can increase the model accuracy, as evidenced by the lower RMSE values of 7.869% and 6.985% for types 3 and 4, respectively. More importantly, Type1 displays the minimum average RMSE of 6.208%, proves the effect on accuracy improvement with increased dimension of relevant features from formation process.

However, the estimation accuracy is reduced with combined features type 3 and type 4 in figure 5(d) based on our proposed model, which possibly attributes to the mismatch between the CNN model and LSTM model. In addition, an overemphasis on one relevant feature to the detriment of the main electrical feature, could lead to model distortion. The concurrent utilization of three features (type 1) can effectively





**Figure 5.** The accuracy of estimation with different feature types. (a) The RMSE value across sample cells with 5 types of inputs (type: 1) temperature, strain, and voltage; (2) none; (3) strain and voltage; (4) temperature and voltage; (5) voltage). (b) The normalized real capacity versus the normalized estimated capacity. The average RMSE of (c) CNN model and (d) proposed integrated model in estimations with 5 types of inputs.

reduce the RMSE, implying that the integrated model could better reflect the completeness and realism of the circumstances within the confined dimensions.

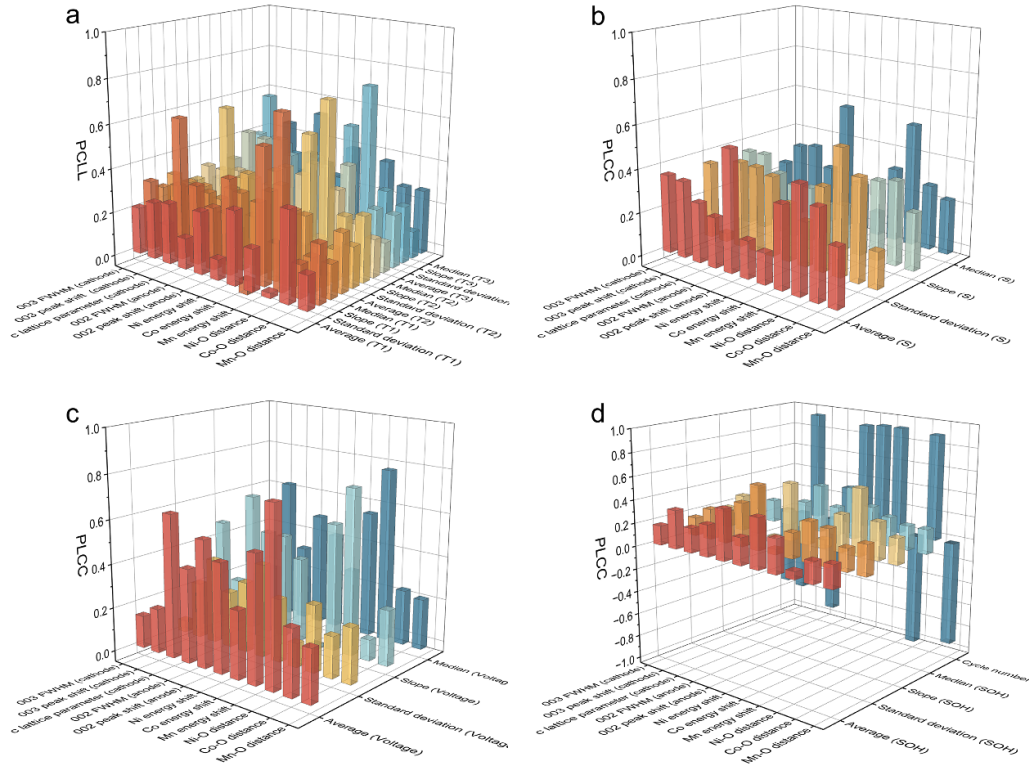
### 2.5. Correlation analysis with end-of-life characterization

Aging of batteries proceed in the form of accumulated degradation, which initiates from the formation process and continues influence the structural, morphological, and oxidation states evolution. To quantify the correlation between formation features and their corresponding impact on battery degradation, retired cells are disassembled for a series of characterizations on electrode materials, as shown in supplementary figures 9–42. The x-ray diffraction (XRD) spectra and Rietveld refinement of NCM cathode at both pristine and end-of-life states (shown in supplementary figures 9–12) reveal the mechanical integrity of particles through the full width at half maximum (FWHM) with peak shifting to higher angle. Since the initial cycle, the nonuniform local stress undermines the mechanical stability and release along the interparticle boundaries by nucleation and propagation of microcracks, particularly evident for the width of (003) peak (summarized in supplementary figure 13). The peak shifts (summarized in supplementary figure 14) attributable to Li-ions insertion/extraction reversibility and the extent of structural inhomogeneity [52].

In addition, the layer to spinel phase transitions and trapped Li-ions in the host structure causes the abrupt anisotropic expansion of lattice structures, especially for the  $c$  lattice parameter (shown in supplementary figure 15), which becomes larger as the current rates increase. The scanning electron microscopic (SEM) images shown in supplementary figure 37 provide clear-evidence for formation of intergranular microcracks at high current densities, which offer paths for electrolyte to penetrate through the particle and trigger internal phase transitions [53].

The degradation also happens at anode side, as shown in supplementary figures 16–21, the wider of FWHM and right shift of (002) peak after cycling reveals destruction in the form of lower graphite crystallinity [54]. Moreover, the SEI layer formed on graphite surface during formation cycles continues influence the Li-insertion/extraction behaviors. As shown in supplementary figures 39 and 41, the graphite particles are covered with inhomogeneous SEI layer with random distributed Li-dendrites. Besides the morphology changes in the regular area on electrode surface, more Li-plating could be observed in supplementary figures 40 and 42, leading to a capacity loss during cycling.

As the redox of transition metals (TMs) start from the initial electrochemical process, the reaction activity undergoes reduction and transformation with structural evolution. The



**Figure 6.** Pearson linear correlation coefficient (PLCC) analysis. The correlation coefficient between statistical characterization and (a) the thermal feature  $T$  (tables S5–7), (b) the mechanical feature  $S$  (table S8), (c) the electrical feature  $V$  (table S9), and (d) cycling information (table S10). The statistical characterization refers to structure (003) peak FWHM, (003) peak shift,  $c$  lattice parameter, (002) FWHM and (002) peak shift), valence states ( $k$ -edge energy shift of Ni, Co and Mn) and the atomic coordination (Ni–O interatomic distance, Co–O interatomic distance, and Mn–O interatomic distance).

valence states of TMs and local atomic coordination at pristine state and after cycling is measured by x-ray absorption spectroscopy (XAS). The  $k$ -edge x-ray absorption near edge spectroscopy (XANES) of Ni, Co and Mn, and their corresponding energy shifts are shown in supplementary figures 22–24. The extended x-ray absorption fine structure (EXAFS) analysis is performed to directly investigate the local environment of TMs after cycling, and the Fourier transform magnitude of the TM–O and TM–M interatomic distances (supplementary figures 25–36) serve as indicators of the stability and reversibility of local coordination environment [55].

To reveal the relationship between the datasets of each specific feature from formation cycles and the statistical characterization at the end-of-life, we employ Pearson linear correlation coefficient (PLCC)  $\rho$  as an indicator to quantify the correlation, as shown in figure 6. It is essentially a normalized quantification on the strength and direction of each pair of datasets, represented by a covariance value between each pair of datasets. Given a pair of variables, the formula for  $\rho$  is:

$$\rho_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

where  $n$  is the number of samples.  $\bar{x}$  and  $\bar{y}$  represent the average of the two variables. Initially, we conduct statistical analysis on the dataset comprising relevant features ( $T1$ ,  $T2$ ,  $T3$ ,  $S$ , and  $V$ )

collected during formation cycles. From this analysis, the average, median, mode, and standard deviation values are extracted to calculate  $\rho$  after the mapping of four-parameter logistic function (supplementary note 2).

As shown in figure 6(a) and supplementary tables 5–7, the PLCC between the fitted statistical thermal features (including  $T1$ ,  $T2$ , and  $T3$  in supplementary figures 43–54) and the characterization results demonstrate that the temperature deviation is closely related to the  $c$  lattice parameter, energy shift of Mn and Ni–O distance, suggesting that uneven distribution of thermal effect has great impact on the phase transformation [56]. Figure 6(b) and supplementary table 8 displays the PLCC of the fitted mechanical feature of  $S$  (supplementary figures 55–58). The mechanical feature  $S$  stands for the internal strain variation, whereas the high correlation coefficient value of average and median  $S$  with (002) peak shift and Ni–O distance, indicate the destruction of graphite crystallinity by irreversible reactions and accumulated stress with cation disordering in NCM cathode [57], respectively. The PLCC between the fitted electrical features of Voltage (supplementary figures 59–62) and the characterized results is presented in figure 6(c) and supplementary table 9, which demonstrate a higher correlation than the thermal and mechanical features, suggesting the weight electrical feature on SOH estimation would be more crucial. Each individual cells perform relative similar electrochemical behavior without significant fluctuating on the voltage profile, thus the voltage deviation has lower

correlation. In contrast, the average, slope and median voltage is closely related to several characterization results, indicating that both the absolute value and the variation trend during the whole electrochemical process benefits the SOH estimation. Figure 6(d) and supplementary table 10 shows the PLCC between the fitted statistical features of SOH and cycle number (supplementary figures 63–67) and the characterized results, whereas cycle number is the dominate factor for capacity fading, but realistically proceeded in the form of structural degradation. Therefore, the electrical features utilized by the LSTM model has advantages on SOH estimation in cycling process.

### 3. Conclusion

In this paper, we demonstrate that our deep transfer ensemble learning model accurately estimate battery SOH leveraging the knowledge spanning the entire lifetime of the cells. This approach provides more comprehensive understanding on the internal relationship of relevant features based on cells cycled under various conditions, and thus enables BMS more reliable suggestions for the users to dynamically adjust the usage plans to avoid safety accidents. By systematically analyzing the correlation between the characterization results of retired cells and the formation information, the fundamental principle of utilizing the formation information to enhance the estimation accuracy of models is revealed. The addition of insufficient or non-effective information may generate counterproductive effects. The future work should focus on the improvement for the integration of each model, and fully functioning the relevant features to enhance the accuracy and robust.

Accordingly, battery manufacturers can monitor the battery SOH in a more efficient way by making full use of the formation information. Above all, we highlight the pivotal role of formation process in the battery performance, and provide a valuable guidance to the effective estimation of battery SOH by coupling the formation information with ML, which is conducive to the advancement of the next-generation battery modeling and the deployment of intelligent BMS in real application.

## 4. Experimental procedures

### 4.1. Pouch cell preparation

The NCM523/graphite pouch cell, free of electrolyte, were purchased from Dongguan Detai Energy Co., Ltd The electrolyte was procured from Duoduo Chemical Co., Ltd, contained 1.0 M LiPF<sub>6</sub> in EC and DEC with a weight ratio of 1:1. The pouch cell was dried at 45 °C in dynamic vacuum overnight to remove residual moisture. Thereafter the cell was sealed completely by heat lamination at 90 kPa vacuum after the injection of 3 ml of electrolyte. Before the attachment of temperature and strain sensors to the surface, cells were allowed to stand for 12 h to guarantee the completion of wetting. Galvanostatic cycling tests were conducted following the

application of an initial pressure of 40–50 N to the cells by the mold in a temperature-controlled chamber using NEWARE Battery Test System (CT-ZWJ-4'S-T-1U, Shenzhen Data generation, China).

### 4.2. Sensor attachment

The strain sensors (BF350-3AA/1.5AA) employs Kang Copper wire with a resistance of 350 ohms were sourced from Loga Technology Co., Ltd the temperature sensors (DS18B20) were purchased from the North China Sensor Instrument Factory Store. Three temperature sensors were affixed respectively to the positive electrode, negative electrode, and bottom of the battery using thermally conductive silicone. Positioned in the center of the battery, the strain sensor was attached with glue. The specific sensor placements are illustrated in supplementary figure 3.

### 4.3. Feature engineering

In this paper, we extracted nine feature curves along with three statistical features. Among them, five curves were extracted from the full voltage range of the formation data, including  $T1$ ,  $T2$ ,  $T3$ , strain, and voltage. Additionally, four curves were obtained from partial cycling charge data (charge voltage range from 3.70 V to 3.85 V), including partial charge voltage curve ( $V$ ), partial charge capacity curve ( $Q$ ),  $\Delta V_{(i-1)}$ , and  $\Delta Q_{(i-1)}$ . Subsequently, these curves were fitted over time and linearly interpolated to a fixed length of 100. The three statistical features consist of median, skewness (Skew), and maximum value (Max). The calculations for  $\Delta V_{(i-1)}$ ,  $\Delta Q_{(i-1)}$ , and Skew are as follows:

$$\Delta V_{(i-1)} = V_i - V_1 \quad (3)$$

$$\Delta Q_{(i-1)} = Q_i - Q_1 \quad (4)$$

$$\text{Skew} = \frac{1}{n} \sum_{i=1}^n \left[ \left( \frac{v_i - \mu}{\sigma} \right)^3 \right] \quad (5)$$

where  $i$  is the number of cycles. The  $\mu$  and  $\sigma$  are the average and standard deviation of the  $V$ , respectively.

### 4.4. Model development

This section introduces the proposed deep transfer learning model. In essence, the GPR handles discharge capacity prediction with the base models. In the training phase, two base models are first trained based on the extracted features (As described in Model building) and the true discharge capacity are learned. Second, the CNN model and LSTM are used to perform  $k$ -fold cross validation on the training set. The folds number  $k$  chosen in this paper is 9. Each base model outputs a prediction set of the discharge capacity. These prediction sets are then stacked horizontally as feature sets. The true discharge capacity labels of the training set are combined with the feature sets to construct a new dataset. Finally, GPR model is selected as a second-level learner for training on the constructed dataset.

With the remaining model conditions unchanged, we configured CNNs with varying numbers of layers and subsequently compared the SOH estimation accuracy to validate the appropriateness of the CNN layer depths. It is evident that the average prediction error is minimized when the CNN consists of 3 layers (supplementary table 11). As the number of CNN layers increases, the model training time escalates, yet the prediction accuracy becomes hard to be improved. Consequently, a 3-layer CNN architecture was adopted. There are three convolutional layers, three max-pooling layers, and one fully connected layer (supplementary figure 7(a)).

LSTM is a recurrent neural network endowed with special memory capabilities, of which the structure is showed in supplementary figure 7(b). For an LSTM neuron at any given time step ( $t$ ), the inputs include the current input ( $x$ ), the previous hidden layer output ( $h_{t-1}$ ), and the previous memory cell state output ( $C_{t-1}$ ). The outputs comprise the current hidden layer state ( $h_t$ ) and the current cell state ( $C_t$ ). Control over the neuron's cell state in LSTM is facilitated through a structure called 'gates', which include the input gate, forget gate, and output gate. Specifically, the forget gate and input gate determine the content of the memory cell ( $C$ ). The forget gate regulates the proportion of the previous time step's memory cell state ( $C_t$ ) retained in the current time step's memory cell state ( $C_t$ ), and the input gate governs the proportion of the current time step's input ( $x$ ) preserved in the current time step's memory cell state ( $C_t$ ). In this study, the LSTM model comprises one hidden layer (32 neurons) and one fully connected layer (32 neurons).

The ReLU and Sigmoid activation function are used to introduce nonlinear factors to enhance the ability of modeling in complex patterns in the base models.

The GPR model is a nonparametric model that achieves state prediction through prior knowledge within a Bayesian framework. For our implementation, we make use of the standard radial basis kernel.

#### 4.5. Transfer learning

In pretraining, nine cells from the 2C dataset were selected for  $K$ -fold cross-validation, resulting in nine folds, where one cell was utilized for testing datasets in each fold, and the remaining cells were used for training datasets, (supplementary table 1). Consequently, nine pretrained CNN models and nine LSTM models were trained using the 2 C dataset. Then the GPR model is trained using the predictions from both CNN and LSTM. After the pretraining steps, transfer learning was employed to swiftly and effortlessly fine-tune the model based on the knowledge from the source dataset, aiding in addressing the target task. Therefore, in transfer learning, only partial data from 0.5 C and 1 C were used to train the GPR model to adapt to the target dataset. The performance of the proposed model was compared with DCNN-ETL method [58] and the deep learning framework proposed by Ma *et al* [22], as shown in supplementary figure 8. The method proposed in this paper can make more accurate capacity estimations.

#### 4.6. Characterization

The XANES and EXAFS spectroscopy measurements were carried out using a Rapid XAFS 2 M laboratory (XAS, Anhui Absorption Spectroscopy Analysis Instrument Co., Ltd, China.) in the transmission mode, and Si (110), Si (533) and Si (551) spherically bent crystal analyzers with a radius of curvature of 500 mm was used to acquire the Mn, Co and Ni XAS spectra, respectively.

Powder XRD data was collected using DX-2800 with Cu  $K\alpha$  at  $\lambda = 1.54051 \text{ \AA}$  under 40 kV and 40 mA. Data were collected at a step width of  $0.02^\circ$  and  $0.03^\circ$  for both positive and negative materials, ranging from  $10^\circ$  to  $90^\circ$ . All XRD data were refined using GSAS/EXPGUI software.

The morphologies of cathodes and anodes were observed by SEM (ZEISS GeminiSEM 300, Germany) instrument under an acceleration voltage of 15 kV.

#### Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

#### Acknowledgments

The authors acknowledge the financial support of the program for National Natural Science Foundation of China (No. 52203272), Research Foundation for the Postdoctoral Program of Sichuan University (No. 2022SCU12068), and Sichuan Science and Technology Program (2023NSFSC1131).

#### Author contributions

Conceptualization, X H and W H; data curation, X H, W H and D L; formal analysis, X H, W H, D L, Z S, C W, S T, J C and H W; funding acquisition, X H; investigation, W H, C Y, S T and H W; methodology, X H, W H, and D L; project administration, X H, Z L, L H and H T; resources Y L; software, W H, D L and Z S; supervision, X H; validation W H, D L and Z S; visualization, W H, X G and H W writing—original draft, W H, C W, S T and X G; and writing—review & editing, X H, W H, H W, L H and X H.

#### Conflict of interest

There are no conflicts to declare.

#### ORCID iDs

Yuanjing Lin  <https://orcid.org/0000-0002-8568-1786>  
Xin He  <https://orcid.org/0000-0002-0272-9079>



## References

- [1] Viswanathan V, Epstein A H, Chiang Y M, Takeuchi E, Bradley M, Langford J and Winter M 2022 The challenges and opportunities of battery-powered flight *Nature* **601** 519–25
- [2] Costa C M, Barbosa J C, Gonçalves R, Castro H, Campo F J D and Lanceros-Méndez S 2021 Recycling and environmental issues of lithium-ion batteries: advances, challenges and opportunities *Energy Storage Mater.* **37** 433–65
- [3] Zeng J and Liu S 2023 Research on aging mechanism and state of health prediction in lithium batteries *J. Energy Storage* **72** 108274
- [4] Xiong R, Pan Y, Shen W, Li H and Sun F 2020 Lithium-ion battery aging mechanisms and diagnosis method for automotive applications: recent advances and perspectives *Renew. Sustain. Energy Rev.* **131** 110048
- [5] Li Y, Liu K, Foley A M, Zülke A, Berecibar M, Nanini-Maury E, Van Mierlo J and Hoster H E 2019 Data-driven health estimation and lifetime prediction of lithium-ion batteries: a review *Renew. Sustain. Energy Rev.* **113** 109254
- [6] Barré A, Deguilhem B, Grolleau S, Gérard M, Suard F and Riu D 2013 A review on lithium-ion battery ageing mechanisms and estimations for automotive applications *J. Power Sources* **241** 680–9
- [7] Xiao Y, Wen J, Yao L, Zheng J, Fang Z and Shen Y 2023 A comprehensive review of the lithium-ion battery state of health prognosis methods combining aging mechanism analysis *J. Energy Storage* **65** 107347
- [8] Jia Y, Li J, Yuan C, Gao X, Yao W, Lee M and Xu J 2021 Data-driven safety risk prediction of lithium-ion battery *Adv. Energy Mater.* **11** 2003868
- [9] Lin C, Xu J and Mei X 2023 Improving state-of-health estimation for lithium-ion batteries via unlabeled charging data *Energy Storage Mater.* **54** 85–97
- [10] Roman D, Saxena S, Robu V, Pecht M and Flynn D 2021 Machine learning pipeline for battery state-of-health estimation *Nat. Mach. Intell.* **3** 447–56
- [11] Basia A, Simeu-Abazi Z, Gascard E and Zwolinski P 2021 Review on state of health estimation methodologies for lithium-ion batteries in the context of circular economy *CIRP J. Manuf. Sci. Technol.* **32** 517–28
- [12] Hossain Lipu M S, Hannan M A, Karim T F, Hussain A, Saad M H M, Ayob A, Miah M S and Indra Mahlia T M 2021 Intelligent algorithms and control strategies for battery management system in electric vehicles: progress, challenges and future outlook *J. Clean. Prod.* **292** 126044
- [13] Wang H, Liu F, Yu R, Xiao Z, Zhu Z, Zhou L and Wu J 2022 Co-gradient Li-rich cathode relieving the capacity decay in lithium-ion batteries *Nano Energy* **100** 107439
- [14] Palacı M R 2018 Understanding ageing in Li-ion batteries: a chemical issue *Chem. Soc. Rev.* **47** 4924–33
- [15] Yang Y, Xu L, Yang S J, Yan C and Huang J Q 2022 Electrolyte inhomogeneity induced lithium plating in fast charging lithium-ion batteries *J. Energy Chem.* **73** 394–9
- [16] Angeles Cabañero M, Altmann J, Gold L, Boaretto N, Müller J, Hein S, Zausch J, Kallo J and Latz A 2019 Investigation of the temperature dependence of lithium plating onset conditions in commercial Li-ion batteries *Energy* **171** 1217–28
- [17] Qin Y, Zuo P, Chen X, Yuan W, Huang R, Yang X, Du J, Lu L, Han X and Ouyang M 2022 An ultra-fast charging strategy for lithium-ion battery at low temperature without lithium plating *J. Energy Chem.* **72** 442–52
- [18] Gao T, Bai J, Ouyang D, Wang Z, Bai W, Mao N and Zhu Y 2023 Effect of aging temperature on thermal stability of lithium-ion batteries: part a—high-temperature aging *Renew. Energy* **203** 592–600
- [19] Zhang G, Wei X, Chen S, Wei G, Zhu J, Wang X, Han G and Dai H 2023 Research on the impact of high-temperature aging on the thermal safety of lithium-ion batteries *J. Energy Chem.* **87** 378–89
- [20] Liu K, Shang Y, Ouyang Q and Widanage W D 2021 A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery *IEEE Trans. Ind. Electron.* **68** 3170–80
- [21] Li W, Sengupta N, Dechent P, Howey D, Annaswamy A and Sauer D U 2021 Online capacity estimation of lithium-ion batteries with deep long short-term memory networks *J. Power Sources* **482** 228863
- [22] Ma G, Xu S, Jiang B, Cheng C, Yang X, Shen Y, Yang T, Huang Y, Ding H and Yuan Y 2022 Real-time personalized health status prediction of lithium-ion batteries using deep transfer learning *Energy Environ. Sci.* **15** 4083–94
- [23] Hu X, Xu L, Lin X and Pecht M 2020 Battery lifetime prognostics *Joule* **4** 310–46
- [24] Zhu J et al 2022 Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation *Nat. Commun.* **13** 2261
- [25] Lin C, Xu J, Hou J, Liang Y and Mei X 2023 Ensemble method with heterogeneous models for battery state-of-health estimation *IEEE Trans. Ind. Inform.* **19** 10160–9
- [26] Yao J and Han T 2023 Data-driven lithium-ion batteries capacity estimation based on deep transfer learning using partial segment of charging/discharging data *Energy* **271** 127033
- [27] Zhang Z, Min H, Guo H, Yu Y, Sun W, Jiang J and Zhao H 2023 State of health estimation method for lithium-ion batteries using incremental capacity and long short-term memory network *J. Energy Storage* **64** 107063
- [28] Zhang Y, Tang Q, Zhang Y, Wang J, Stimming U and Lee A A 2020 Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning *Nat. Commun.* **11** 1706
- [29] Wang L et al 2019 Identifying the components of the solid–electrolyte interphase in Li-ion batteries *Nat. Chem.* **11** 789–96
- [30] Wood D L, Li J and An S J 2019 Formation challenges of lithium-ion battery manufacturing *Joule* **3** 2884–8
- [31] Adenusi H, Chass G A, Passerini S, Tian K V and Chen G 2023 Lithium batteries and the solid electrolyte interphase (SEI)—progress and outlook *Adv. Energy Mater.* **13** 2203307
- [32] Zhang S, Li Y, Bannenberg L J, Liu M, Ganapathy S and Wagemaker M 2024 The lasting impact of formation cycling on the Li-ion kinetics between SEI and the Li-metal anode and its correlation with efficiency *Sci. Adv.* **10** eadj8889
- [33] Weng A, Mohtat P, Attia P M, Sulzer V, Lee S, Less G and Stefanopoulou A 2021 Predicting the impact of formation protocols on battery lifetime immediately after manufacturing *Joule* **5** 2971–92
- [34] Mo Y, Liu G, Chen J, Zhu X, Peng Y, Wang Y, Wang C, Dong X and Xia Y 2023 Unraveling the temperature-responsive solvation structure and interfacial chemistry for graphite anodes *Energy Environ. Sci.* **17** 227–37
- [35] Mao C, An S J, Meyer H M, Li J, Wood M, Ruther R E and Wood D L 2018 Balancing formation time and electrochemical performance of high energy lithium-ion batteries *J. Power Sources* **402** 107–15
- [36] Weng S et al 2023 Temperature-dependent interphase formation and Li<sup>+</sup> transport in lithium metal batteries *Nat. Commun.* **14** 4474
- [37] Wang J, Huang W, Pei A, Li Y, Shi F, Yu X and Cui Y 2019 Improving cyclability of Li metal batteries at elevated

- temperatures and its origin revealed by cryo-electron microscopy *Nat. Energy* **4** 664–70
- [38] Li Y *et al* 2022 Operando decoding of surface strain in anode-free lithium metal batteries via optical fiber sensor *Adv. Sci.* **9** 2203247
- [39] Deng Z, Huang Z, Shen Y, Huang Y, Ding H, Luscombe A, Johnson M, Harlow J E, Gauthier R and Dahn J R 2020 Ultrasonic scanning to observe wetting and “unwetting” in Li-ion pouch cells *Joule* **4** 2017–29
- [40] Yusuf A, Sai Avvaru V, De la Vega J, Zhang M, Garcia Molleja J and Wang D Y 2023 Unveiling the structure, chemistry, and formation mechanism of an in-situ phosphazene flame retardant-derived interphase layer in LiFePO<sub>4</sub> cathode *Chem. Eng. J.* **455** 140678
- [41] Xu J 2022 Critical review on cathode–electrolyte interphase toward high-voltage cathodes for Li-ion batteries *Nano-Micro. Lett.* **14** 166
- [42] Wang S, Wu T, Xie H, Li C, Zhang J, Jiang L and Wang Q 2022 Effects of current and ambient temperature on thermal response of lithium ion battery *Batteries* **8** 203
- [43] Forgez C, Vinh Do D, Friedrich G, Morcrette M and Delacourt C 2010 Thermal modeling of a cylindrical LiFePO<sub>4</sub>/graphite lithium-ion battery *J. Power Sources* **195** 2961–8
- [44] Louli A J, Ellis L D and Dahn J R 2019 Operando pressure measurements reveal solid electrolyte interphase growth to rank Li-ion cell performance *Joule* **3** 745–61
- [45] Zhang W, Schröder D, Arlt T, Manke I, Koerver R, Pinedo R, Weber D A, Sann J, Zeier W G and Janek J 2017 (Electro)chemical expansion during cycling: monitoring the pressure changes in operating solid-state lithium batteries *J. Mater. Chem. A* **5** 9929–36
- [46] Han S Y, Lee C, Lewis J A, Yeh D, Liu Y, Lee H W and McDowell M T 2021 Stress evolution during cycling of alloy-anode solid-state batteries *Joule* **5** 2450–65
- [47] Liu S, Zhang H and Xu X 2021 A study on the transient heat generation rate of lithium-ion battery based on full matrix orthogonal experimental design with mixed levels *J. Energy Storage* **36** 102446
- [48] Lyu P, Huo Y, Qu Z and Rao Z 2020 Investigation on the thermal behavior of Ni-rich NMC lithium ion battery for energy storage *Appl. Therm. Eng.* **166** 114749
- [49] Lin C, Xu S and Liu J 2018 Measurement of heat generation in a 40 Ah LiFePO<sub>4</sub> prismatic battery using accelerating rate calorimetry *Int. J. Hydrog. Energy* **43** 8375–84
- [50] Rudin C 2019 Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead *Nat. Mach. Intell.* **1** 206–15
- [51] Loyola-Gonzalez O 2019 Black-box vs. white-box: understanding their advantages and weaknesses from a practical point of view *IEEE Access* **7** 154096–113
- [52] Ryu H H, Namkoong B, Kim J H, Belharouak I, Yoon C S and Sun Y K 2021 Capacity fading mechanisms in Ni-rich single-crystal NCM cathodes *ACS Energy Lett.* **6** 2726–34
- [53] Sun H H, Kim U H, Park J H, Park S W, Seo D H, Heller A, Mullins C B, Yoon C S and Sun Y K 2021 Transition metal-doped Ni-rich layered cathode materials for durable Li-ion batteries *Nat. Commun.* **12** 6552
- [54] He H, Huang C, Luo C W, Liu J J and Chao Z S 2013 Dynamic study of Li intercalation into graphite by in situ high energy synchrotron XRD *Electrochim. Acta* **92** 148–52
- [55] Ou X *et al* 2022 Enabling high energy lithium metal batteries via single-crystal Ni-rich cathode material co-doping strategy *Nat. Commun.* **13** 2319
- [56] Jiang M, Danilov D L, Eichel R A and Notten P H L 2021 A review of degradation mechanisms and recent achievements for Ni-rich cathode-based Li-ion batteries *Adv. Energy Mater.* **11** 48
- [57] Su Y, Zhang Q, Chen L, Bao L, Lu Y, Chen S and Wu F 2022 Stress accumulation in Ni-rich layered oxide cathodes: origin, impact, and resolution *J. Energy Chem.* **65** 236–53
- [58] Shen S, Sadoughi M, Li M, Wang Z and Hu C 2020 Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries *Appl. Energy* **260** 114296