Article

Battery health evaluation using a short random segment of constant current charging

Zhongwei Deng, Xiaosong Hu, Yi Xie, Le Xu, Penghua Li, Xianke Lin, Xiaolei Bian

dengzhongw@cqu.edu.cn (Z.D.)
xiaosonghu@ieee.org (X.H.)

Highlights

A short random charging segment enables battery health evaluation

Two features with high correlations to battery capacity are extracted

Three typical machine learning methods are compared for battery health estimation

The method is verified by four types of batteries cycled under different conditions

Article

Battery health evaluation using a short random segment of constant current charging

Zhongwei Deng,[1,5,*] Xiaosong Hu,[1,*] Yi Xie,[1] Le Xu,[1] Penghua Li,[2] Xianke Lin,[3] and Xiaolei Bian[4]

SUMMARY

Accurately evaluating the health status of lithium-ion batteries (LIBs) is significant to enhance the safety, efficiency, and economy of LIBs deployment. However, the complex degradation processes inside the battery make it a thorny challenge. Data-driven methods are widely used to resolve the problem without exploring the complex aging mechanisms; however, random and incomplete charging-discharging processes in actual applications make the existing methods fail to work. Here, we develop three data-driven methods to estimate battery state of health (SOH) using a short random charging segment (RCS). Four types of commercial LIBs (75 cells), cycled under different temperatures and discharging rates, are employed to validate the methods. Trained on a nominal cycling condition, our models can achieve high-precision SOH estimation under other different conditions. We prove that an RCS with a 10mV voltage window can obtain an average error of less than 5%, and the error plunges as the voltage window increases.

INTRODUCTION

Lithium-ion batteries (LIBs) have made our daily lives more convenient and colorful by powering our smartphones, computers, electric vehicles, and so forth. Their advantages in energy density, power density, and long lifetime have been accelerating their penetration in various energy storage applications (Larcher and Tarascon, 2015). However, LIBs inevitably age during use and non-use time, mainly owing to the loss of lithium-ion inventory (LLI) and loss of active materials (LAM) inside the batteries (Barré et al., 2013). The direct effect of aging on battery performance is the decrease in capacity and the increase in internal resistance (Han et al., 2019). To improve the safety, efficiency, and economy of using LIBs, it is indispensable to conduct battery safety monitoring (Deng et al., 2018), residual value assessment, and timely maintenance (Hu et al., 2020), all of which heavily depend on the battery health evaluation. However, the health degradation of LIBs affected by temperature, current rate, mechanical stress, and historical operational conditions presents highly nonlinear dynamics (Attia et al., 2020; Severson et al., 2019).

From a physical point of view, the most direct method for battery health evaluation is to quantify the microscopic degradation processes of the battery, such as solid electrolyte interphase (SEI) growth (Wang et al., 2019), particle cracking (Yan et al., 2017), and lithium plating (Xiao, 2019). However, these degradation processes are coupled with each other, and their measurement is currently destructive to the battery (Reniers et al., 2019) and non-destructive techniques need to be explored. Moreover, these microscopic inspections usually require high-cost *ex situ* techniques and cannot be conducted in the field. To avoid these intractable problems, researchers developed various semi-empirical models to quantify the degradation caused by different stress factors (Schimpe et al., 2018), such as temperature, current rate, and depth of discharge. A large number of experiments are required to establish an accurate semi-empirical model, and its adaptivity to other operational conditions is still questionable. Battery electrochemical models (Doyle et al., 1993) and equivalent circuit models (Plett, 2004a) are widely used to simulate battery behaviors. As the battery health-related parameters, such as capacity and internal resistance, can be derived from the battery model, parameter identification (Rahimi-Eichi et al., 2014) and state estimation methods (Plett, 2004b) are used to obtain these parameters. However, owing to the model uncertainty and limited measurements for feedback (only voltage and temperature), it is very difficult to obtain reliable estimation results with clear physical meaning.

Recently, with the development of machine learning techniques and the availability of a large amount of high-quality battery data, various data-driven methods (Li et al., 2019; Ng et al., 2020; Shu et al., 2021) have been

[1]College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China

[2]College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

[3]Department of Mechanical Engineering, Ontario Tech University, ON L1G 0C5, Canada

[4]Department of Chemical Engineering, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden

[5]Lead contact

*Correspondence:
dengzhongw@cqu.edu.cn
(Z.D.),
xiaosonghu@ieee.org (X.H.)

https://doi.org/10.1016/j.isci.2022.104260
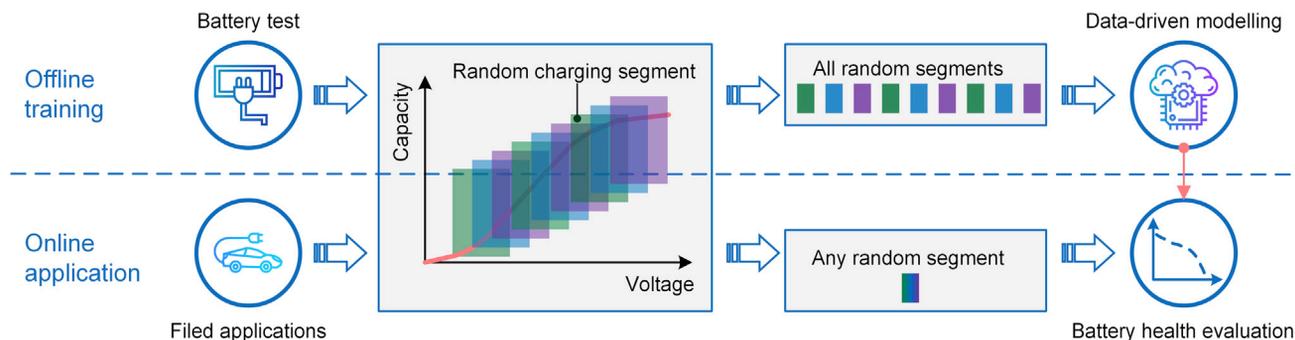
Figure 1. Schematic of the proposed method

proposed for battery health prognostics. According to the input types, these methods can be roughly divided into two categories: feature-based methods and sequence-based methods. Feature-based methods use the extracted features as the inputs and then utilize lightweight machine learning algorithms to model the latent functions between the features and the output target, such as multiple linear regression (MLR) (Deng et al., 2021), support vector machine (Deng et al., 2016), relevance vector machine (Li et al., 2014), and Gaussian process regression (GPR) (Richardson et al., 2019; Roman et al., 2021). Various features can be extracted from the voltage, current, temperature curves during the charging/discharging process, and electrochemical impedance spectrum (Zhang et al., 2020). For example, incremental capacity (IC) and differential voltage (DV) analysis (Han et al., 2014) are two useful methods to extract features to evaluate battery health, and typical features include the peak values of the IC curves (Jiang et al., 2020; Tang et al., 2021), the valley values of the DV curves (Li et al., 2018), and the curve area within a given voltage range. In contrast, sequence-based methods directly use time-series data as the input and employ deep learning methods to achieve automatically feature extraction and nonlinear modeling, e.g., deep neural network (Roman et al., 2021; Tian et al., 2021), long short-term memory network (Deng et al., 2022b; Li et al., 2020), deep convolutional neural network (DCNN) (Shen et al., 2020), and their variants. These techniques usually use time-series data of battery current, temperature, voltage, and accumulated charge under complete or partial charging/discharging conditions as the input. Many studies have shown that both feature-based and sequence-based methods can achieve outstanding performance under specific conditions.

In many applications, the battery discharge process is a bit dynamic, while the charging process is relatively stable and usually pre-defined, such as in electric vehicles and smartphones. Therefore, many researchers developed health evaluation models based on the charging data (Jiang et al., 2020; Li et al., 2018; Shen et al., 2020; Tian et al., 2021). In these studies, a specific voltage range and a fixed start/endpoint are required to ensure that the curves of different cycles have the same reference points. However, in practical applications, the charging behaviors of users are random (Zhao et al., 2021), which means the charging start and endpoints are not fixed, and a complete charging process is very difficult to capture (Deng et al., 2022a). Furthermore, for series battery packs, the inconsistency between battery cells causes them to have different charging voltage curves and a narrow voltage overlap window (Tian et al., 2020), especially for aged cells, which hinders the health evaluation of each cell. In short, it is still a significant challenge to conduct battery health evaluation based on a random and short charging segment in real applications.

In this work, we develop data-driven methods to accurately estimate battery state of health (SOH) using a random charging segment (RCS) extracted from the constant current process. The proposed methods are validated with four types of commercial batteries (75 cells in total) cycling under different temperatures and discharging rates. As schematically shown in Figure 1, we first divide the constant current (CC) charging curve into dozens of segments and extract a capacity increment sequence in each segment. We show that the capacity increment sequence evolves in a certain pattern as the battery ages, and its average value and SD have high correlations with the battery SOH. Then, we analyze the correlations under different numbers of segments and find that even a short segment is highly correlated with battery SOH. Finally, two types of machine learning algorithms (features-based and sequence-based) are used to model the SOH estimators. In the training process of the data-driven models, all RCSs are used as input, but only an RCS is required as input for the online application. We prove that the developed methods can achieve high accuracy using even an RCS with an extremely small voltage window (i.e., 10mV).
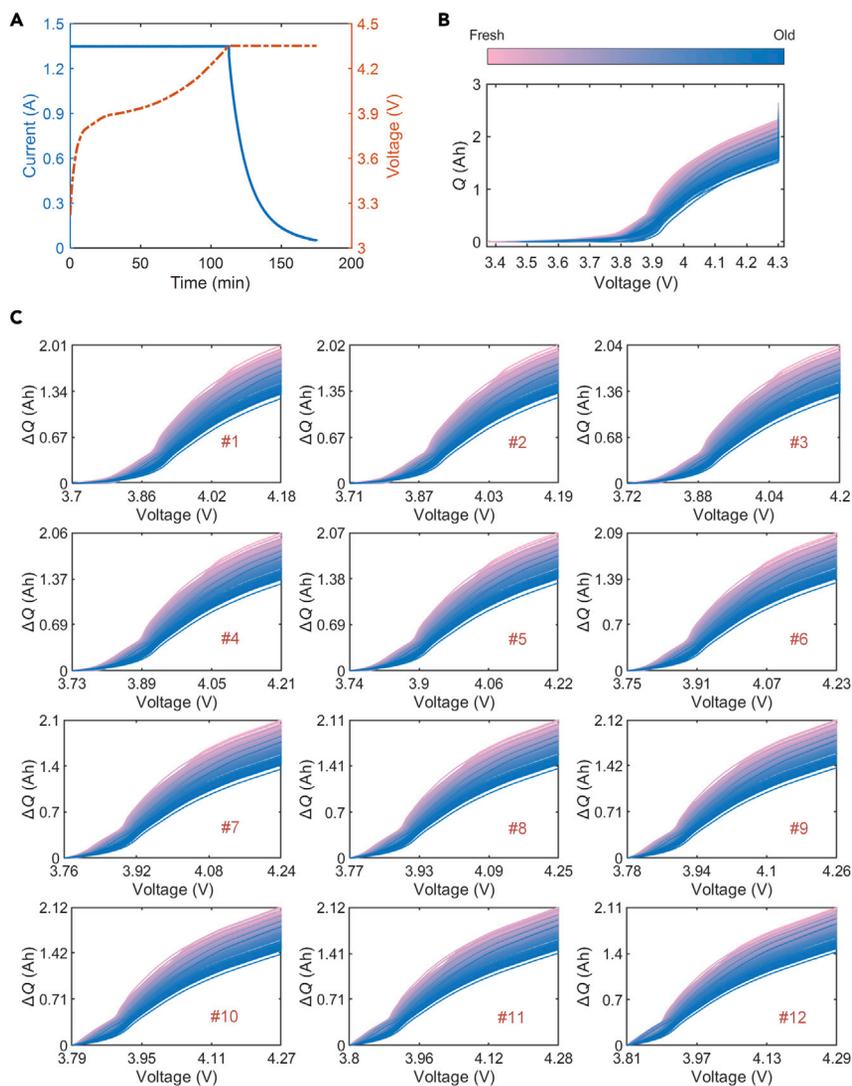
Figure 2. Charge capacity evolution as battery ages

An NMC-LCO cell is taken as an example.

(A) CC-CV charging policy.

(B) Battery charge capacity curves as a function of voltage at different aging levels.

(C) Capacity increment sequences in different voltage segments. The charge capacity sequence corresponding to the voltage range of 3.70–4.29V is divided into 12 segments, and each segment is denoted by a symbol #x ($1 \leq x \leq 12$).

RESULTS

Charge capacity evolution

Taking the data of a $LiNi_xMn_yCo_{1-x-y}O_2$ blended with $LiCoO_2$ (NMC-LCO) cell as an example, we investigate the evolution of battery charge capacity with the number of cycles. Figure 2A and Figure 2B show the constant current-constant voltage (CC-CV) charging profiles and the corresponding charge capacity (Q) as a function of voltage, where the color denotes the number of cycles. It shows that the capacity curves gradually shift downward as the battery ages. Table 3 presents parameters setting used to extract the Q from different batteries. Furthermore, we divide the voltage range of 3.70–4.29V into 12 segments according to (Equation 2) and extract the capacity increment sequences (ΔQ_{seg}) from the segments (Figure 2C). After this segmentation, each segment has a 0.48V voltage window. Some patterns in the evolution of charge capacity as the battery ages can also be observed in these partial charging segments.
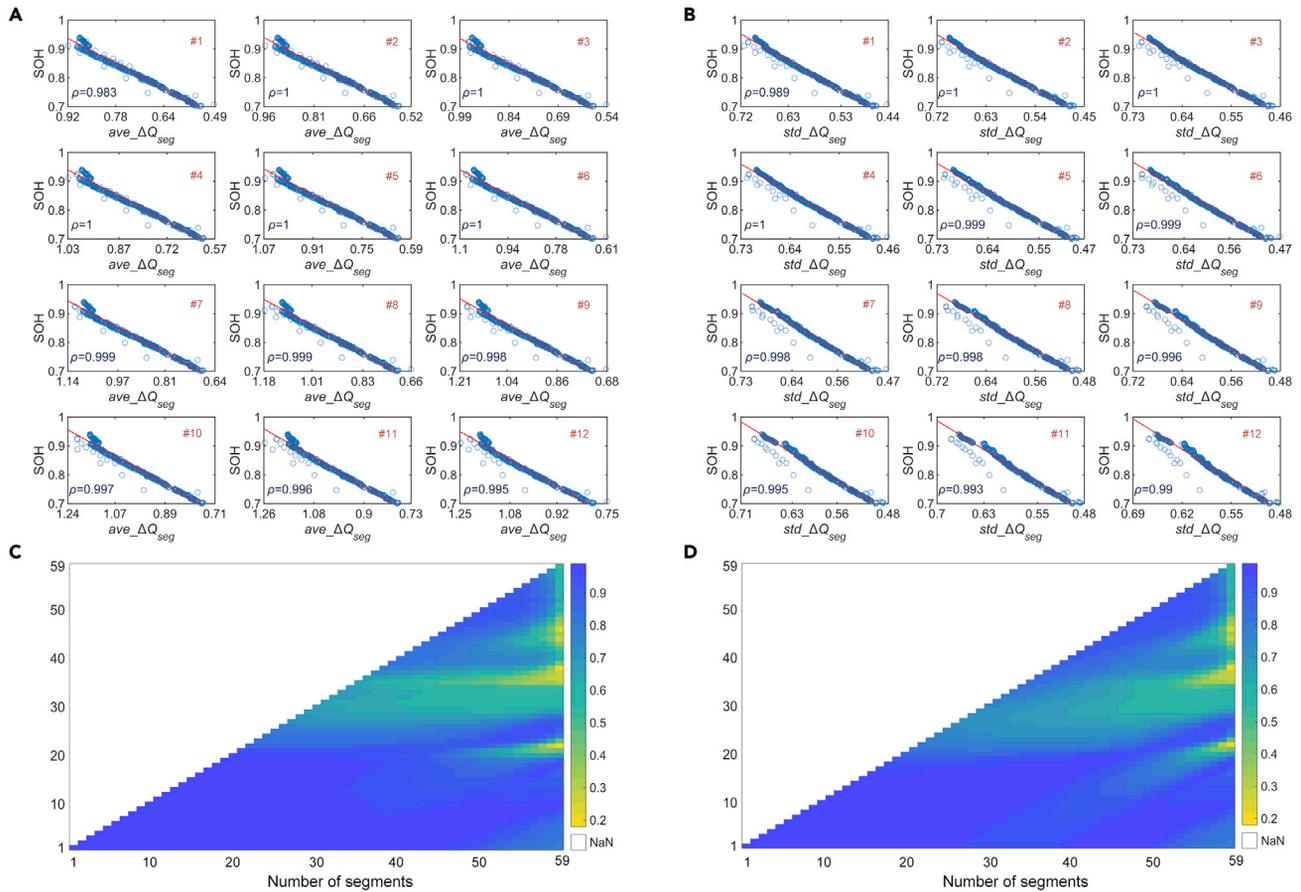
Figure 3. Correlation analysis of extracted features for NMC-LCO battery

The charge capacity sequence corresponding to the voltage range of 3.70–4.29V is divided into m segments (ΔQ_{seg}). The ρ between the features and battery SOH are analyzed.

(A) Correlation of $ave\_\Delta Q_{seg}$ for each segment when m is equal to 12.

(B) Correlation of $std\_\Delta Q_{seg}$ for each segment when m is equal to 12.

(C) Correlation of $ave\_\Delta Q_{seg}$ for each segment as m varies from 1 to 59.

(D) Correlation of $std\_\Delta Q_{seg}$ for each segment as m varies from 1 to 59.

Correlation analysis

To evaluate the usefulness of the partial capacity segments for the battery health evaluation, we calculate two statistical characteristics of $\Delta \boldsymbol{Q}_{seg}$ and analyze the correlations (ρ) between them and battery SOH. We choose the average value of $\Delta \boldsymbol{Q}_{seg}$ ($ave\_\Delta \boldsymbol{Q}_{seg}$) and the SD of $\Delta \boldsymbol{Q}_{seg}$ ($std\_\Delta \boldsymbol{Q}_{seg}$) as features, and their correlations with the SOH of NMC-LCO battery are shown in Figure 3. The correlations when the voltage range is divided into 12 segments are illustrated in Figures 3A and 3B. It can be observed that the two features have a ρ value close to one for almost all segments. To analyze the effect of the number of segments on the correlations, we further calculate the correlations of different segments under various segmentation operations, and the results are shown in Figures 3C and 3D. According to (Equation 2), we know that a smaller voltage window can be obtained with a larger number of segments. From the two heatmaps, it is clear that a high correlation (>0.8) can be maintained for all segments until the number of segments exceeds 30, and the first 20 segments always have high correlations no matter how many segments are defined. We also analyze the correlations of features for the other three types of batteries (see Figures S1–S3).

For each segmenting operation, we calculate the average absolute correlation (AAC) of each feature for different segments. Figure 4 compares the variation of the AACs with the number of segments for different batteries. All the batteries are cycling in a 25°C chamber with the same charging policy (0.5C CC-CV). For the discharging policy, the NMC-LCO battery is discharged at 1.5 CC, while the other three types are discharged at 0.5 CC. Owing to the difference in electrochemistry, the correlations of the four batteries present different patterns of
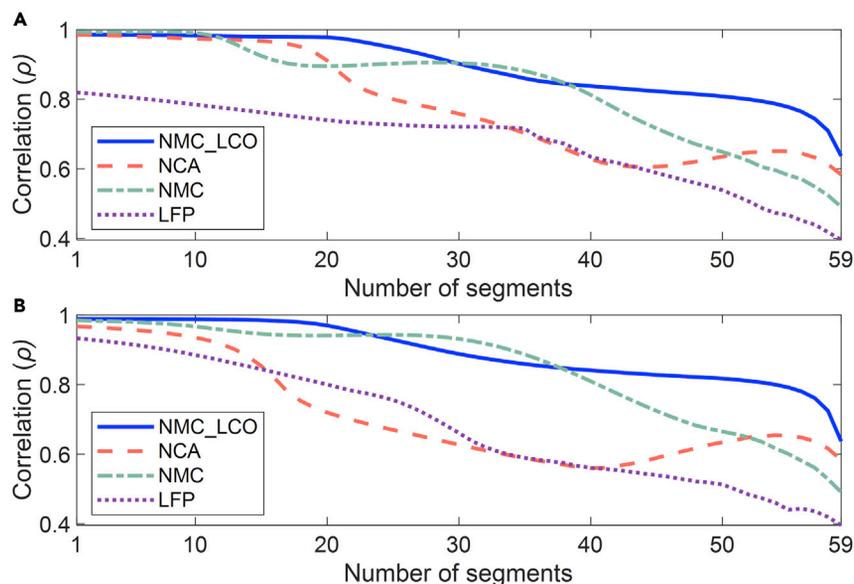
Figure 4. Comparison of feature correlations for different batteries

The capacity sequence is divided into m segments, and m varies from 1 to 59. For each segmentation operation, the AACs of features for different segments are calculated.

(A) AACs of $ave\_\triangle Q_{seg}$.

(B) AACs of $std\_\triangle Q_{seg}$.

change. In contrast, for the same battery chemistry, the correlation variation of its two features is highly consistent. The correlation generally decreases as the number of segments increases, except for a partial recovery in NCA battery. For all types of batteries, high correlations of the two features with the battery SOH appear when the number of segments is less than 20. Even when the number of segments is 59, which corresponds to a voltage window of 10mV, a correlation over 0.5 is obtained for NMC-LCO, NCA, and NMC batteries, and around 0.4 for LFP battery. Compared with other battery types, it is more difficult for LFP battery to extract features highly related to battery SOH based on the $\triangle Q_{seg}$, which increases the difficulty of its SOH estimation.

Battery state of health estimation

We use two types of data-driven methods to model the battery SOH estimator. The first one is the feature-based method, which uses selected features as input. By comparison, the second one, i.e., the sequence-based method, can directly use the raw data sequence as input.

In this study, one simple algorithm, (MLR) and two state-of-the-art machine learning algorithms (sparse GPR and DCNN) corresponding to the above-mentioned two types of methods are employed to construct battery SOH estimators. The MLR is a typical method to model the linear relationship between input features and output target. In general, the better the linear correlation between the features and the output, the higher the accuracy of MLR estimation. The sparse GPR (SGPR) can efficiently capture the nonlinear relationships between the inputs features and output and provide a probabilistic prediction of the target. For the DCNN, it can infinitely approximate the nonlinear characteristics of the process owing to its deep learning mechanism. In this SOH estimation problem, the MLR and SGPR take the two features ($ave\_\triangle Q_{seg}$ and $std\_\triangle Q_{seg}$) and the mean value of the corresponding voltage sequence as inputs, while the DCNN uses the capacity increment sequence ($\triangle Q_{seg}$) and the corresponding voltage sequence as inputs directly. We convert the high dimensional sequences into images as input to the DCNN, which can automatically extract features from the front layers of the network.

As illustrated in Figure 1, to achieve battery SOH estimation based on any RCS, the training samples need to cover all segments. This increases the sample size by dozens of times (usually up to 10,000), resulting in a regular GPR being unable to complete the training on a regular computer (computational complexity is $O(n^3)$, n is the sample size). In this regard, we use a sparse GPR and it can significantly reduce the computational burden by introducing inducing points (Candela and Rasmussen, 2005).
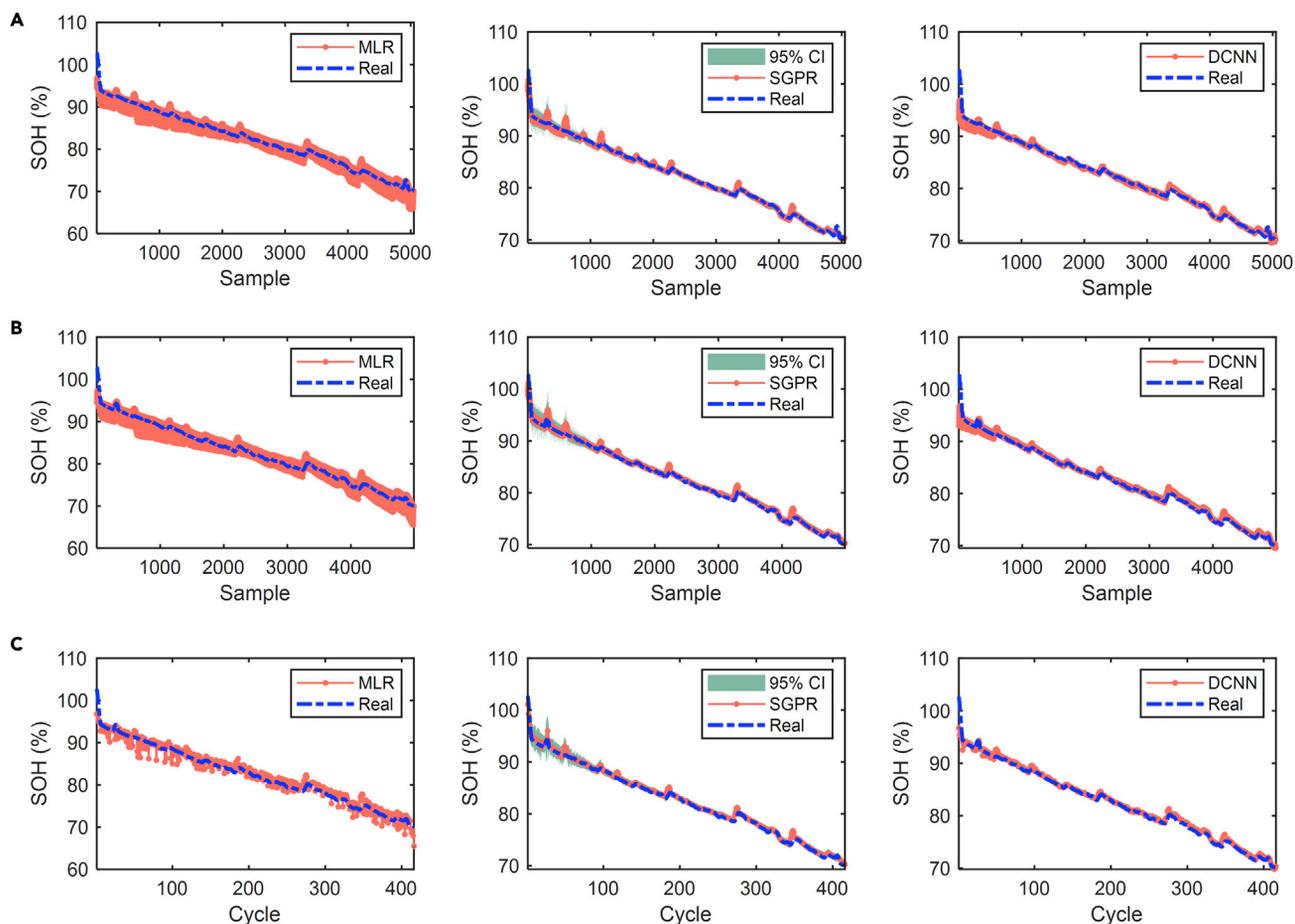
Figure 5. SOH estimation results of NMC-LCO batteries

The capacity sequence is divided into 12 segments. MLR, SGPR, and DCNN methods are used to estimate battery SOH. The Dash line represents actual values and the dashed-dotted line represents the estimated values.

(A) Training results.

(B) Test results using all segments for each cycle.

(C) Test results using a random segment for each cycle.

The SOH estimation results for NMC-LCO batteries are presented in Figure 5, including the training process, the test process using all segments, and the test process using a random segment for each cycle. The results of the other three types of batteries are shown in Figures S4–S6. The statistical errors of SOH estimation for the four types of batteries are summarized in Table 1. All the mentioned methods are trained on a cell under 25°C, 0.5C CC-CV charging and 0.5C discharging conditions (except for 1.5C for NMC-LCO), and are tested on another cell with the same battery type and cycling condition. We define the above cycling condition as a nominal cycling condition. In the modeling processes, we also divide the voltage range into 12 segments, which means that 12 samples are generated per cycle. Therefore, the x-axis labels in the training (Figure 5A) and test results (Figure 5B) are denoted by "sample" rather than "cycle." To mimic the random charging behaviors of users, we perform uniformly distributed sampling and select one from 12 segments for each cycle as the final estimated value of the cycle (Figure 5C). It is worth noting that for the SGPR and DCNN methods, the model obtained after each training is different owing to the random setting of initial parameters. To obtain reliable results, we run 20 times of training and test processes, and take their average values as the final results.

The above results show that all the methods can obtain high-precision SOH estimation for different battery chemistries, with a mean absolute error (MAE) lower than 2% and a root-mean-square error (RMSE) lower than 2.5% in the test process. The DCNN achieves the best performance, and its MAEs and RMSEs are lower than 1 and 1.2%, respectively. Owing to the inability to model nonlinear characteristics, the

Table 1. Statistical errors of SOH estimation for different batteries

	Errors	NMC-LCO			NCA			NMC			LFP		
		MLR	SGPR	DCNN	MLR	SGPR	DCNN	MLR	SGPR	DCNN	MLR	SGPR	DCNN
Training	MAE (%)	0.97	0.29	0.32	1.80	0.31	0.32	0.89	0.80	0.31	1.40	0.20	0.23
	RMSE (%)	1.28	0.52	0.66	2.29	0.63	0.52	1.35	1.40	0.46	1.82	0.70	0.41
Test	MAE (%)	1.02	0.30	0.37	1.88	1.97	0.95	0.84	0.83	0.40	1.30	0.34	0.27
	RMSE (%)	1.30	0.50	0.68	2.48	2.07	1.18	1.38	1.51	0.54	1.74	0.81	0.49
Test (random)	MAE (%)	1.05	0.29	0.36	1.93	1.96	0.95	0.91	0.91	0.39	1.31	0.35	0.27
	RMSE (%)	1.33	0.47	0.63	2.55	2.07	1.18	1.46	1.60	0.53	1.72	0.83	0.47

estimation accuracy of the MLR is always lower than the SGPR and DCNN. Besides, we can find that the estimation errors based on random segments are almost the same as that of using all test segments. This is because of the use of uniformly distributed sampling. According to the above results, we demonstrate that a high-accuracy data-driven estimator can be built as long as the data changes in some pattern with the output, whether it is a feature-based method or a sequence-based method.

We further analyze the effect of the number of segments on the accuracy of SOH estimation. Note that with a larger number of segments, we can get a smaller voltage window for each segment. For instance, when the number is 59, the voltage window is only 10 mV. Besides, for battery health evaluation, it is of practical importance to construct a model based on a specific cell but can maintain its accuracy for other cells. Therefore, we further test the accuracy of the models on other cells of the same chemistry. The cycling conditions of the four types of batteries are explained in Table 2. NMC-LCO batteries have the same cycling conditions, while NCA, NMC, and LFP have different cycling temperatures and discharging rates. The relationship between the capacity and the cycle number is shown in Figure S7 for all the batteries. This figure indicates that the rate of capacity degradation is significantly affected by temperature and discharging rates. In addition, we can see that the capacity of parts of batteries decays too fast under certain cycling conditions and reach a capacity range (area under black dashed-dotted line in Figure S7), to which the batteries under the nominal cycling condition have not decayed. To examine the accuracy of the models under severe capacity degradation, we still use the models established under nominal cycling conditions to estimate the SOH in this area.

The MAEs of SOH estimation for the four types of batteries are shown in Figure 6, in which the errors are plotted as a function of the number of segments and cycling conditions (except for NMC-LCO, which uses cell numbers). Owing to the requirement of convolutional operation in the DCNN, the length of the input sequence cannot be too small. We set the lowest limit of the length equal to five for the DCNN, thus there are at most 55 segments for each cycle.

For NMC-LCO cells, the estimation accuracy of different cells is almost the same, and the maximum MAEs of the three models are lower than 6%. Even using an extremely small voltage window (10 mV), an acceptable estimation result can still be obtained. And when a voltage window of 100 mV is used, a MAE of SOH lower than 3% can be obtained by the SGPR and the DCNN models. However, the above outstanding performance is owing to the same cycling condition between the trained and test cells.

For NCA, NMC, and LFP batteries, when several cells are cycled under the same condition, the average values of MAEs are used to indicate the results of this cycling condition. We can find that the MAE increases

Table 2. Battery information and experimental setting

Battery types	Contributors	Manufacturers	Nominal capacity (Ah)	Temperature (°C)	Charge & discharge policies	Battery Numbers
NMC-LCO	HNEI	LG Chem	2.8	25	0.5C CC-CV & 1.5C CC	14
NCA	SNL	Panasonic	3.2	15/25/35	0.5C CC-CV & 0.5C/1C/2C CC	18
NMC	SNL	LG Chem	3	15/25/35	0.5C CC-CV & 0.5C/1C/2C/3C CC	22
LFP	SNL	A123	1.1	15/25/35	0.5C CC-CV & 0.5C/1C/2C/3C CC	21
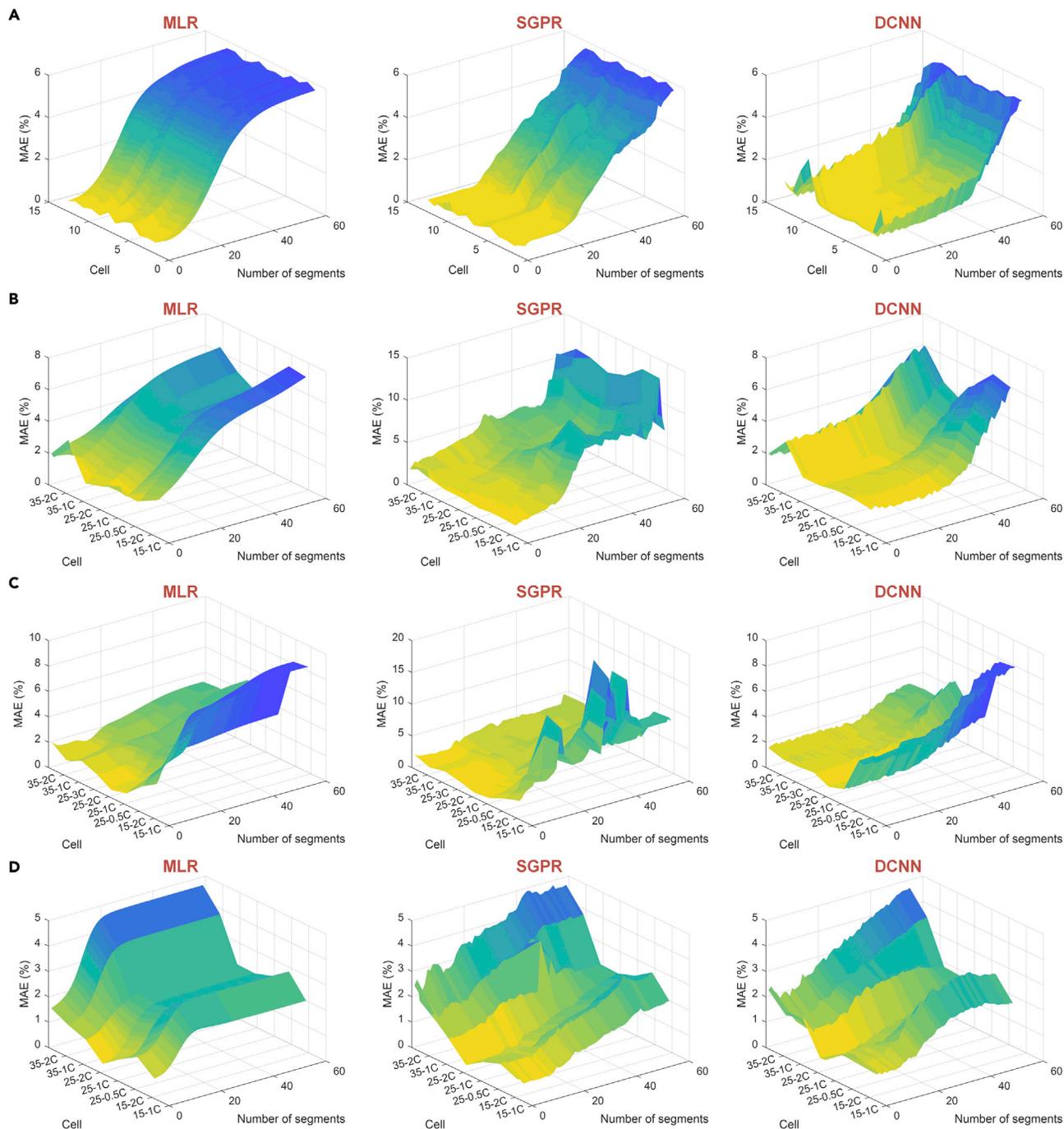
Figure 6. The MAEs of SOH estimation for four types of batteries

Using the data of one cell to train MLR, SGPR and DCNN models, and the remaining cells are used to test the models. The variation of errors with the number of segments is also given. The symbol "T1-C1" in tick labels denotes the cell is cycled in a chamber with T1 temperature and C1 discharging rate.

(A) NMC-LCO cells.

(B) NCA cells.

(C) NMC cells.

(D) LFP cells.

as the number of segments increases or the cycling conditions deviate from the nominal cycling condition. The influence of temperature on the estimation errors is different for the three types of batteries. For NMC batteries, a lower temperature increases the error; for LFP batteries, a higher temperature increases the
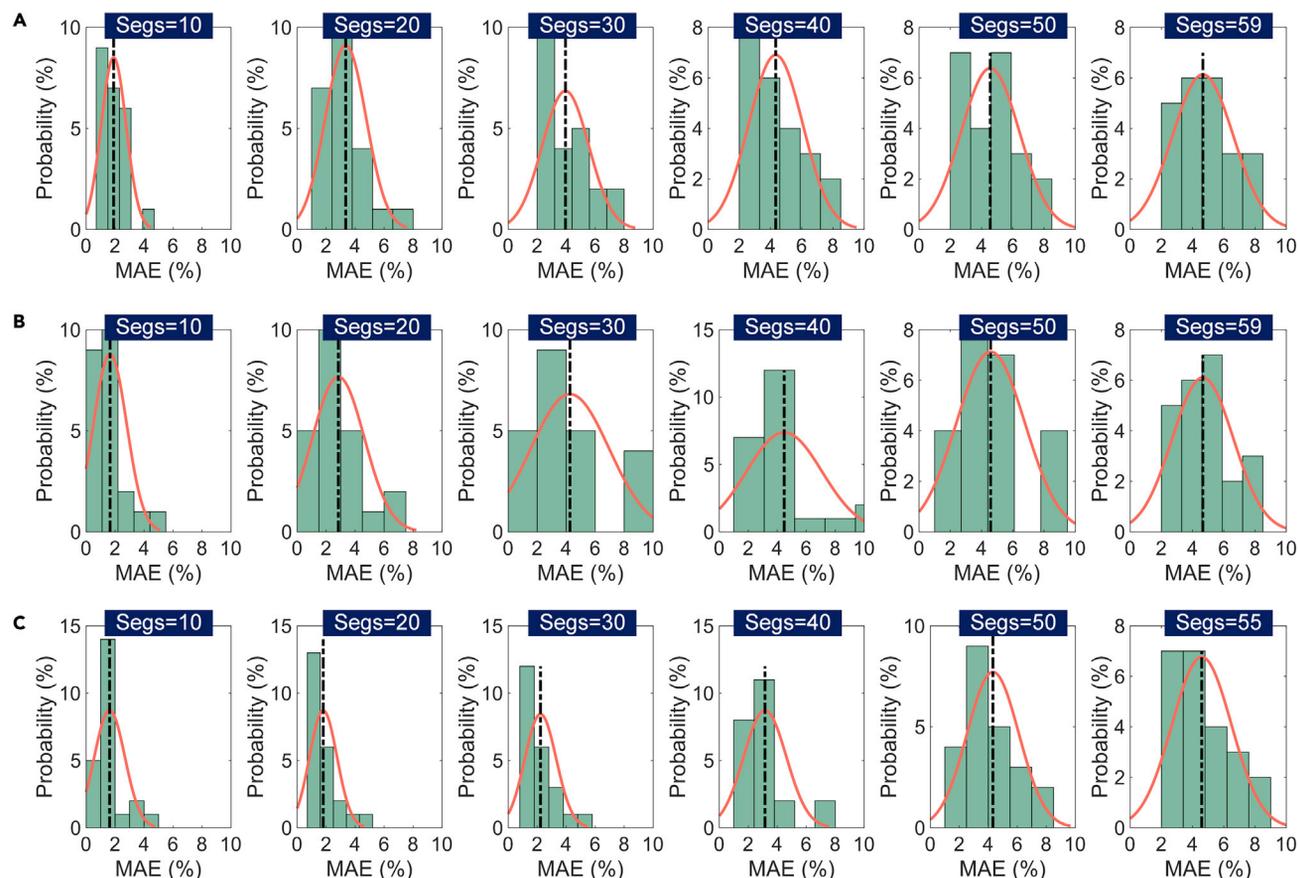
Figure 7. The distributions of MAEs at different numbers of segments
The vertical dotted line represents the position of the mean value, and the symbol "segs" denotes the number of segments.
(A) MLR method.
(B) SGPR method.
(C) DCNN method.

error; while for NCA batteries, both a higher temperature and a lower temperature increases the error. In contrast, the influence of discharging rate on the estimation errors is insignificant. In addition to the metric of MAE, the RMSEs of SOH estimation for four types of batteries are also shown in Figure S8, and the same influence law from temperature can be observed.

To better evaluate the robustness of the models constructed using different numbers of segments and under different cycling conditions, the distributions of MAEs at different segments are shown in Figure 7 (the distributions of RMSEs are shown in Figure S9). The distribution is calculated based on the results under the same number of segments for the four types of batteries under different cycling conditions. It can be observed that when the number of segments is less than 10, a mean MAE of less than 2% can be obtained for each method. Even if the number of segments is up to 55 or 59 (corresponds to a 10mV voltage window), a mean MAE less than 5% and a mean RMSE around 6% can be guaranteed. Note that a larger number of segments means a smaller voltage window for SOH estimation. This proves that an RCS with a 10mV voltage window can achieve an acceptable SOH estimation. Meanwhile, we can observe that a sequence with a big voltage window can capture more battery degradation information and has better generalization for different working conditions. However, it is difficult to obtain a big voltage window in the charging process in many applications. Therefore, we have to sacrifice some accuracy to ensure the availability of the models.

DISCUSSION

In this article, we develop data-driven models to achieve accurate battery health evaluation using an RCS extracted from the constant current charging process. We prove that capacity increment sequences in the

Table 3. Parameter settings for capacity increment sequence extraction

Battery types	V_{start} (V)	V_{end} (V)	$\triangle V$ (V)	n
NMC-LCO	3.7	4.29	0.01	60
NCA/NMC	3.6	4.19	0.01	60
LFP	3.0	3.59	0.01	60

CC charging process are informative for battery health evaluation. Two features extracted from these sequences are closely related to battery SOH. Two feature-based methods (i.e., the MLR and SGPR) and one sequence-based method (i.e., the DCNN) are used to construct data-driven models for SOH estimation. The proposed models are trained using the data of a cell under the nominal cycling condition and subsequently tested on other cells with the same or different conditions (difference in temperature and discharging current rates). The developed methods are validated using four types of batteries (75 cells in total), with the error lower than 2% when the voltage window is up to 500mV. Moreover, an average estimation error lower than 5% can be obtained even when the voltage window is less than 10mV. Our work substantiates that it is promising to use an RCS with a narrow voltage window to achieve accurate health evaluation for LIBs. As only the short random charging data is needed, the proposed methods can be applied to a variety of scenarios.

Limitations of the study

In this article, the proposed methodology is verified under cycling conditions with constant ambient temperatures, constant discharging currents, and full charge and discharge. It is valuable to conduct more verifications under other cycling conditions, such as changing ambient temperature, dynamic discharge, and shallow charge and discharge. These will be investigated in our future work.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- METHOD DETAILS
  - Dataset description
  - Capacity increment sequence extraction
  - Multiple linear regression
  - Gaussian process regression
  - Deep convolutional neural network
  - Evaluation criteria

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104260.

AUTHOR CONTRIBUTIONS

Z. Deng and L. Xu analyzed the datasets and conceived the study. Z. Deng, X. Hu, and Y. Xie developed the models. Z. Deng, X. Lin, and X. Bian interpreted the results. All authors edited and reviewed the article. X. Hu and X. Lin supervised the work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Attia, P.M., Grover, A., Jin, N., Severson, K.A., Markov, T.M., Liao, Y.-H., Chen, M.H., Cheong, B., Perkins, N., Yang, Z., et al. (2020). Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature *578*, 397–402. https://doi.org/10.1038/s41586-020-1994-5.

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 Sourc. *241*, 680–689. https://doi.org/10.1016/j.jpowsour.2013.05.040.

Candela, J.Q., and Rasmussen, C.E. (2005). A unifying view of sparse approximate Gaussian process regression. J. Mach. Learn. Res. 6, 1939–1959.

Deng, J., Bae, C., Marcicki, J., Masias, A., and Miller, T. (2018). Safety modelling and testing of lithium-ion batteries in electrified vehicles. Nat. Energy *3*, 261–266. https://doi.org/10.1038/s41560-018-0122-3.

Deng, Z., Hu, X., Li, P., Lin, X., and Bian, X. (2022a). Data-driven battery state of health estimation based on random partial charging data. IEEE Trans. Power Electron. *37*, 5021–5031. https://doi.org/10.1109/TPEL.2021.3134701.

Deng, Z., Hu, X., Lin, X., Xu, L., Che, Y., and Hu, L. (2021). General discharge voltage information enabled health evaluation for lithium-ion batteries. IEEE/ASME Trans. Mechatron. *26*, 1295–1306. https://doi.org/10.1109/TMECH.2020.3040010.

Deng, Z., Lin, X., Cai, J., and Hu, X. (2022b). Battery health estimation with degradation pattern recognition and transfer learning. J. Power Sourc. *525*, 231027. https://doi.org/10.1016/j.jpowsour.2022.231027.

Deng, Z., Yang, L., Cai, Y., Deng, H., and Sun, L. (2016). Online available capacity prediction and state of charge estimation based on advanced data-driven algorithms for lithium iron phosphate battery. Energy *112*, 469–480. https://doi.org/10.1016/j.energy.2016.06.130.

Devie, A., Baure, G., and Dubarry, M. (2018). Intrinsic variability in the degradation of a batch of commercial 18650 lithium-ion cells. Energies *11*, 1031. https://doi.org/10.3390/en11051031.

Doyle, M., Fuller, T.F., and Newman, J. (1993). Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. J. Electrochem. Soc. *140*, 1526–1533. https://doi.org/10.1149/1.2221597.

Gao, T., and Lu, W. (2021). Machine learning toward advanced energy storage devices and systems. iScience *24*, 101936. https://doi.org/10.1016/j.isci.2020.101936.

Han, X., Lu, L., Zheng, Y., Feng, X., Li, Z., Li, J., and Ouyang, M. (2019). A review on the key issues of the lithium ion battery degradation among the whole life cycle. eTransportation *1*, 100005. https://doi.org/10.1016/j.etran.2019.100005.

Han, X., Ouyang, M., Lu, L., Li, J., Zheng, Y., and Li, Z. (2014). A comparative study of commercial lithium ion battery cycle life in electrical vehicle: aging mechanism identification. J. Power Sourc. *251*, 38–54. https://doi.org/10.1016/j.jpowsour.2013.11.029.

Hu, X., Xu, L., Lin, X., and Pecht, M. (2020). Battery lifetime prognostics. Joule *4*, 310–346. https://doi.org/10.1016/j.joule.2019.11.018.

Jiang, B., Dai, H., and Wei, X. (2020). Incremental capacity analysis based adaptive capacity estimation for lithium-ion battery considering charging condition. Appl. Energy *269*, 115074. https://doi.org/10.1016/j.apenergy.2020.115074.

Larcher, D., and Tarascon, J.M. (2015). Towards greener and more sustainable batteries for electrical energy storage. Nat. Chem. *7*, 19–29. https://doi.org/10.1038/nchem.2085.

Li, H., Pan, D., and Chen, C.L.P. (2014). Intelligent prognostics for battery health monitoring using the mean entropy and relevance vector machine. IEEE Trans. Syst. Man Cybern. Syst. *44*, 851–862. https://doi.org/10.1109/TSMC.2013.2296276.

Li, P., Zhang, Z., Xiong, Q., Ding, B., Hou, J., Luo, D., Rong, Y., and Li, S. (2020). State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network. J. Power Sourc. *459*, 228069. https://doi.org/10.1016/j.jpowsour.2020.228069.

Li, Y., Abdel-Monem, M., Gopalakrishnan, R., Berecibar, M., Nanini-Maury, E., Omar, N., van den Bossche, P., and Van Mierlo, J. (2018). A quick on-line state of health estimation method for Li-ion battery with incremental capacity curves processed by Gaussian filter. J. Power Sourc. *373*, 40–53. https://doi.org/10.1016/j.jpowsour.2017.10.092.

Li, Y., Liu, K., Foley, A.M., Zülke, A., Berecibar, M., Nanini-Maury, E., Van Mierlo, J., Hoster, H.E.J.R., and reviews, s.e. (2019). Data-driven health estimation and lifetime prediction of lithium-ion batteries. Renew. Sustain. Energy Rev. *113*,

109254. https://doi.org/10.1016/j.rser.2019.109254.

Ng, M.-F., Zhao, J., Yan, Q., Conduit, G.J., and Seh, Z.W. (2020). Predicting the state of charge and health of batteries using data-driven machine learning. Nat. Mach. Intell. *2*, 161–170. https://doi.org/10.1038/s42256-020-0156-7.

Plett, G.L. (2004a). Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs. J. Power Sourc. *134*, 262–276. https://doi.org/10.1016/j.jpowsour.2004.02.032.

Plett, G.L. (2004b). Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs. J. Power Sourc. *134*, 277–292. https://doi.org/10.1016/j.jpowsour.2004.02.033.

Preger, Y., Barkholtz, H.M., Fresquez, A., Campbell, D.L., Juba, B.W., Romàn-Kustas, J., Ferreira, S.R., and Chalamala, B. (2020). Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. J. Electrochem. Soc. *167*, 120532. https://doi.org/10.1149/1945-7111/abae37.

Rahimi-Eichi, H., Baronti, F., and Chow, M.Y. (2014). Online adaptive parameter identification and state-of-charge coestimation for lithium-polymer battery cells. IEEE Trans. Ind. Electron. *61*, 2053–2061. https://doi.org/10.1109/TIE.2013.2263774.

Reniers, J.M., Mulder, G., and Howey, D.A. (2019). Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries. J. Electrochem. Soc. *166*, A3189–A3200. https://doi.org/10.1149/2.0281914jes.

Richardson, R.R., Osborne, M.A., and Howey, D.A. (2019). Battery health prediction under generalized conditions using a Gaussian process transition model. J. Energy Storage *23*, 320–328. https://doi.org/10.1016/j.est.2019.03.022.

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*, 447–456. https://doi.org/10.1038/s42256-021-00312-3.

Schimpe, M., von Kuepach, M.E., Naumann, M., Hesse, H.C., Smith, K., and Jossen, A. (2018). Comprehensive modeling of temperature-dependent degradation mechanisms in lithium iron phosphate batteries. J. Electrochem. Soc. *165*, A181–A193. https://doi.org/10.1149/2.1181714jes.

Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M.,

Herring, P.K., Fraggedakis, D., et al. (2019). Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy *4*, 383–391. https://doi.org/10.1038/s41560-019-0356-8.

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. https://doi.org/10.1016/j.apenergy.2019.114296.

Shu, X., Shen, S., Shen, J., Zhang, Y., Li, G., Chen, Z., and Liu, Y. (2021). State of health prediction of lithium-ion batteries based on machine learning: advances and perspectives. iScience *24*, 103265. https://doi.org/10.1016/j.isci.2021.103265.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 1–9. https://doi.org/10.1109/CVPR.2015.7298594.

Tang, X., Wang, Y., Liu, Q., and Gao, F. (2021). Reconstruction of the incremental capacity trajectories from current-varying profiles for lithium-ion batteries. iScience *24*, 103103. https://doi.org/10.1016/j.isci.2021.103103.

Tian, J., Wang, Y., Liu, C., and Chen, Z. (2020). Consistency evaluation and cluster analysis for lithium-ion battery pack in electric vehicles. Energy *194*, 116944. https://doi.org/10.1016/j.energy.2020.116944.

Tian, J., Xiong, R., Shen, W., Lu, J., and Yang, X.-G. (2021). Deep neural network battery charging curve prediction using 30 points collected in 10 min. Joule *5*, 1521–1534. https://doi.org/10.1016/j.joule.2021.05.012.

Wang, L., Menakath, A., Han, F., Wang, Y., Zavalij, P.Y., Gaskell, K.J., Borodin, O., Iuga, D., Brown, S.P., Wang, C., et al. (2019). Identifying the components of the solid–electrolyte interphase in Li-ion batteries. Nat. Chem. *11*, 789–796. https://doi.org/10.1038/s41557-019-0304-z.

Williams, C.K., and Rasmussen, C.E. (2006). Gaussian Processes for Machine Learning (MIT press).

Xiao, J. (2019). How lithium dendrites form in liquid batteries. Science *366*, 426–427. https://doi.org/10.1126/science.aay8672.

Yan, P., Zheng, J., Gu, M., Xiao, J., Zhang, J.-G., and Wang, C.-M. (2017). Intragranular cracking as a critical barrier for high-voltage usage of layer-structured cathode for lithium-ion batteries. Nat. Commun. *8*, 14101. https://doi.org/10.1038/ncomms14101.

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. https://doi.org/10.1038/s41467-020-15235-7.

Zhao, Y., Wang, Z., Shen, Z.-J.M., and Sun, F. (2021). Assessment of battery utilization and energy consumption in the large-scale development of urban electric vehicles. Proc. Natl. Acad. Sci. U S A *118*, e2017318118. https://doi.org/10.1073/pnas.2017318118.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Battery Archive	The battery datasets are from Hawaii Natural Energy Institute (HNEI) and Sandia National Laboratories (SNL)	https://www.batteryarchive.org/index.html
Software and algorithms		
MATLAB R2018a	MathWorks	https://www.mathworks.com
GPML Matlab Code version 4.2	The code is written by Carl Edward Rasmussen and Hannes Nickisch	http://www.gaussianprocess.org/gpml/code/matlab/doc/

RESOURCE AVAILABILITY

Lead contact
Further requests for information should be directed and will be handled by the corresponding author and lead contact, Zhongwei Deng (dengzhongw@cqu.edu.cn).

Material availability
This study did not generate new materials.

Data and code availability
The raw dataset used in this study is available at https://www.batteryarchive.org.

The code for data processing, features extraction and data-driven modeling is completely available at https://github.com/TengMichael/battery-health-evaluation.

METHOD DETAILS

Dataset description
The cell cycling dataset of four different types of batteries is used in this work, and all of them come from an open-source battery data website. The information of batteries and experimental settings are listed in Table 2. All the batteries use graphite material as their negative electrodes, but material of their electrolytes are not disclosed. The battery cells are divided into different types based on the positive materials, namely, $LiNi_xCo_yAl_{1-x-y}O_2$ (NCA), $LiNi_xMn_yCo_{1-x-y}O_2$ (NMC), $LiFePO_4$ (LFP), and a blend of NMC and $LiCoO_2$ (NMC-LCO). The dataset of NMC-LCO is contributed by Hawaii Natural Energy Institute (HNEI) (Devie et al., 2018), and the other three are contributed by Sandia National Laboratories (SNL) (Preger et al., 2020). All the battery cells are cycled under the same constant current-constant voltage (CC-CV) charging profile (0.5C CC-CV), but different CC discharging profiles with different current rates (0.5C, 1.5C, 1C, 2C, and 3C). Although the dataset of SNL including cells cycled under different depths of discharge (DODs), only the data of 100% DOD is used for this study. The benchmark SOH of the cell is calculated by,

$$SOH_i = \frac{C_i}{C_n} \qquad \text{(Equation 1)}$$

where C_n is the nominal capacity of the battery, and C_i is the actual capacity at ith cycle, which is equal to the maximum discharge capacity under CC discharging.

Capacity increment sequence extraction
In a battery charging process, battery charge capacity can be calculated by integrating battery current over time. For constant current (CC) charging, the battery voltage is usually monotonically increasing (or filtered to maintain monotonicity). Given a constant voltage range $[V_{start}, V_{end}]$ and a fixed voltage interval $\triangle V$, a charge capacity sequence ($Q = [Q_1, Q_1, ..., Q_n]$) can be extracted by interpolating the charge capacity curve

with respect to the voltage sequence, where $n = (V_{start} - V_{end})/\Delta V + 1$ (Deng et al., 2021). Considering the difference in voltage platforms for the four types of batteries, different parameters setting is used to extract the Q from different batteries and are listed in Table 3. The V_{start} is set to a voltage point corresponding to about 5% SOC of the battery because 5% SOC is usually reserved to avoid battery over-discharge in practical applications. The V_{end} is set to a value slightly away from the charging cut-off voltage to avoid the influence of data fluctuations when switching to the constant voltage (CV) charging stage.

To achieve accurate battery SOH estimation based on partial charging data, the charge capacity sequence can be further divided into dozens of segments (Q_{seg}). Given a fixed length (h) of the segment and a stride size (s), m segments can be extracted,

$$m = floor\left(\frac{n-h}{s}\right) + 1 \qquad \text{(Equation 2)}$$

where the function $floor(.)$ gets the largest integer no more than the input value. In this study, s is set to 1, corresponding to one ΔV voltage interval. As listed in Table 3, a charge capacity sequence, $Q = [Q_1, Q_1, ..., Q_n]$, can be extracted with $n = 60$ for all batteries. Setting $h = 49$, which corresponds to a 0.48V voltage window, then $m = 12$, which means Q can be divided into 12 Q_{seg}, as shown in Figure 1 for NMC-LCO battery. Due to the random charging behaviors of users, the charging start point is not fixed, thus it is impossible to calculate the exact (or absolute) charge capacity values in practical applications. To overcome this problem, the capacity sequence is replaced by the capacity increment sequence ($\Delta Q_{seg} = Q_{seg} - Q_{seg,1}$) in each segment.

For each ΔQ_{seg}, its average value ($ave\_\Delta Q_{seg}$) and standard deviation ($std\_\Delta Q_{seg}$) can be extracted as features. The Pearson correlation coefficient (ρ) is used to provide the strength of the linear correlation between the features and battery SOH. The correlation analysis results in four different types of batteries are illustrated in Figures 2 and S1–S3.

Multiple linear regression

To model the linear relationship between the input features and output target, an MLR is a commonly used method. Its expression is,

$$\widehat{y}_i = \sum_{j=1}^{n} \widehat{w}_j x_j + \widehat{w}_0 \qquad \text{(Equation 3)}$$

where $\widehat{y}_i$ is the predicted SOH, x_j is the input feature, n is the number of features, and $\widehat{w}_j$ is the weight. The objective function of this regression problem is often defined to minimize the mean square error of the output. When there are many features, a regularization technique can be introduced to prevent the model from overfitting during the training process (Severson et al., 2019).

Gaussian process regression

To better capture the nonlinear relationship between the input features and the battery SOH, the GPR technique is employed. GPR is a machine-learning framework with non-parametric modeling and uncertainty evaluation (Williams and Rasmussen, 2006). For a typical regression problem, observations usually contain Gaussian white noises and can be modelled as,

$$y_i = f\left(x_i\right) + N\left(0, \sigma_n^2\right) \qquad \text{(Equation 4)}$$

where x_i is the ith input features, σ_n^2 is the noise covariance, and $f(x) = [f(x_1), f(x_2), ..., f(x_n)]$ is a Gaussian process. $f(x)$ can be described as $f(x) \sim N(0, K)$, where $K_{ij} = k(x_i, x_j)$ is the covariance kernel function, which is a measure of distance between points x_i and x_j. A squared exponential kernel function is the most widely used, and is expressed as,

$$k\left(x_i, x_j\right) = \sigma_f^2 exp\left(\frac{-\|x_i - x_j\|^2}{2l^2}\right) \qquad \text{(Equation 5)}$$

where σ_f and l are hyperparameters, which determine the amplitude of the kernel function and the importance of each input feature, respectively. Given training samples, the hyperparameters $[\sigma_f, l, \sigma_n]$ of GPR can be optimized by maximizing the marginal likelihood. Due to the matrix inversion in solving the maximum likelihood estimation problem, the computational complexity is $O(n^3)$ for a regular GPR. When the size of

the training space is very large, the training of the regular GPR is intractable. To overcome this problem, a sparse GPR with a specific number of inducing points is employed (Candela and Rasmussen, 2005). Only m training samples are selected from the original training set, thus the computational complexity can be reduced to $O(m^2 n)$. In this paper, the GPR-based SOH estimation is realized by using the Gaussian processes for machine learning (GPML) toolbox (Williams and Rasmussen, 2006).

Deep convolutional neural network

The DCNN has been successfully used in image recognition, and the elements, lines, and shapes of a picture can be captured by different layers (Szegedy et al., 2015). Due to its ability to nonlinear modeling and automatic feature extraction (Gao and Lu, 2021), we use it to estimate battery SOH directly based on the $\triangle Q_{seg}$. In addition to the $\triangle Q_{seg}$, the corresponding voltage sequence is also used as the input of the DCNN. Unlike color images that have an input size equal to $n \times n \times 3$, the above features can only form input with a size equal to $n \times 2 \times 1$, thus 1D CNN is used to construct the network. In this study, the DCNN-based SOH estimation model mainly consists of two 1D convolutional layers, one maximum pooling layer, and one fully-connected layer. The structure of the developed DCNN is shown in Table S1. Since the input size is changing (n varies from 6 to 60 in this case), to ensure that the size of the input to the maximum pooling layer is constant, the size of the first two convolutional layers is also set to be variable. The stride size is default to one, and no padding is used. In each convolutional layer, a batch normalization technique is applied to improve the performance and stability, and a rectified linear unit (ReLU) activation function is subsequently used to learn the nonlinear relationships.

Evaluation criteria

Two statistical characteristics of the SOH estimation errors are chosen to evaluate the model performance. The mean absolute error (MAE), and root mean square error (RMSE) are respectively defined as,

$$MAE = \frac{\sum\limits_{i=1}^{n} |y_i - \widehat{y}_i|}{n} \qquad \text{(Equation 6)}$$

$$RMSE = \sqrt{\frac{\sum\limits_{i=1}^{n} (y_i - \widehat{y}_i)^2}{n}} \qquad \text{(Equation 7)}$$

where y_i is the observed battery SOH, $\widehat{y}_i$ is the estimated SOH, and n is the total number of samples.