Proceedings of CONF-APMM 2026 Symposium: Applied Mathematics and Quantitative Modeling
DOI: 10.54254/2753-8818/2026.PU38480

Estimating Lithium-Ion Battery Capacity from Charging-Curve Health Indicators: How Far Does It Generalise to an Unseen Cell?

Tianxun Wu

School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK
sbjq7836@leeds.ac.uk

Abstract. Accurate battery capacity estimation is important for evaluating the health of lithium-ion batteries. However, a complete discharge test is difficult to perform when a battery is already installed in a vehicle. This study investigates whether battery capacity can be estimated from health indicators extracted from a single charging curve and whether the prediction can generalise to an unseen cell. The NASA ageing data from four 18650 lithium-ion battery cells were used. After charge-discharge pairing and data cleaning, 628 charging samples were retained. Ten health indicators were extracted and divided into time, voltage and temperature groups. Linear Regression, Random Forest and Gradient Boosting were evaluated using a pooled random split and leave-one-cell-out validation. Under the pooled random split, Random Forest with all features achieved the lowest mean absolute error (MAE) of 0.0119 Ah. However, the tree models showed clear performance reductions under leave-one-cell-out validation. Linear Regression with time and voltage features achieved the best cross-cell result, with an MAE of 0.0329 Ah and an R^2 of 0.9446. Adding voltage features reduced the cross-cell MAE by 38.6% compared with using time features alone, while temperature features did not provide a further improvement. B0006 was the most difficult cell to predict. The results show that charging-curve indicators can estimate battery capacity effectively, but random splitting may overestimate the generalisation ability of complex models. Testing on unseen cells is therefore necessary before practical application.

Keywords: lithium-ion battery, state of health, health indicator, charging curve, cross-cell generalisation

1. Introduction

1.1. Background and research gap

Lithium-ion batteries are widely used in electric vehicles, portable devices and energy storage systems because they have high energy density and long cycle life. However, the amount of energy that a battery can store gradually decreases during use. This ageing process affects the driving range, safety and reliability of an electric vehicle. Therefore, estimating the current battery capacity is an important function of a battery management system [1].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.PU38480

Battery capacity is also commonly used to calculate the state of health [1, 2]. A direct capacity measurement normally requires the battery to be fully discharged under controlled conditions. This method is suitable for a laboratory experiment, but it is not practical for a battery that is already installed in a vehicle. During normal operation, the battery management system can more easily measure voltage, current, time and temperature. For this reason, information from the charging process provides a possible way to estimate battery capacity without completing a separate discharge test [1, 2].

Battery ageing changes the shape of the charging curve. For example, the duration of the constant-current and constant-voltage (CV) stages may change as capacity decreases. The voltage may also rise at a different rate, while the temperature response may show changes during charging. These measurable changes can be converted into health indicators and used as inputs for machine learning models [3-5].

Previous studies have used charging curves, health indicators and data-driven models to estimate battery capacity or state of health [6-8]. Many studies mainly focus on improving prediction accuracy. However, training and test samples are often randomly selected from the same group of batteries. As a result, different cycles from one battery may appear in both sets. A model can therefore learn the ageing pattern of a particular cell instead of a general relationship that works for other cells. Good performance under random splitting may not fully represent the situation of applying the model to a new battery. This creates an important research gap between model accuracy within known cells and generalisation to an unseen cell [9, 10].

1.2. Research questions and methods

This study investigates whether lithium-ion battery capacity can be estimated from health indicators extracted from a single charging curve. It uses the NASA Battery Data Set [11], including four 18650 lithium-ion cells: B0005, B0006, B0007 and B0018. Each retained charging record is paired with the capacity measured in the following discharge cycle.

After pairing and cleaning, the dataset contains 628 samples. Ten health indicators are extracted and divided into three groups. The first group contains three time features. The second group adds three voltage features, while the third group also includes four temperature features. Linear Regression, Random Forest and Gradient Boosting are used to compare a simple linear model with two nonlinear tree-based models.

This study asks three main questions. First, how accurately can capacity be estimated using information from one charging curve? Second, how much improvement is obtained by adding voltage and temperature information to the time features? Third, how much does model performance change when the test battery is completely unseen during training? To answer the third question, pooled random splitting is compared with leave-one-cell-out validation.

1.3. Objectives and contributions

The first contribution of this work is a clear process for converting the original NASA time-series records into one feature row for each charging cycle. The cleaning rules, charge-discharge pairing method and feature definitions are reported so that the analysis can be repeated.

The second contribution is the comparison of three feature groups based on their sensor requirements. This makes it possible to examine whether the improvement from an additional signal is large enough to justify its use.

DOI: 10.54254/2753-8818/2026.PU38480

The third contribution is a direct comparison between random splitting and leave-one-cell-out validation. The cell-wise results are also examined to identify which battery is most difficult to predict.

2. Methodology

2.1. Data source and experimental conditions

This study used the BatteryAgingARC-FY08Q4 subset of the NASA Battery Data Set [11]. It contains ageing test data from four 18650 lithium-ion cells: B0005, B0006, B0007 and B0018. The cells were tested at an ambient temperature of 24°C. During charging, a constant current of 1.5 A was used until the voltage reached 4.2 V. The process then changed to constant-voltage charging. Discharging was carried out using a constant current of 2 A.

The original MATLAB files contained 644 charge operations, 636 discharge operations and 887 impedance measurements. In this study, the value to be predicted was the capacity measured during the discharge after each charge. The unit of capacity was ampere-hour (Ah).

2.2. Charge-discharge pairing and data cleaning

The MATLAB files were loaded using scipy.io.loadmat with simplify_cells=True. All experimental operations were processed in their original time order. Each valid charge was matched with its following discharge, and the measured Capacity value was used as the prediction target. Impedance measurements between a charge and a discharge were ignored.

Of the 644 charge records, eight could not be matched one-to-one because another charge appeared before the following discharge. Three final charge records contained only five samples and had no later capacity measurement. This produced 633 charge–discharge pairs.

Several cleaning steps were then applied. Non-finite values were checked first. Short negative-current signals at the start of charging were treated as switching noise and removed when they occurred within the first 30 seconds and the measured current was below −0.1 A. A record was considered an incomplete top-up charge if its starting voltage was above 3.9 V and it contained fewer than 50 constant-current samples. Five records met this rule and were removed.

The final dataset contained 628 samples. B0005, B0006 and B0007 each provided 166 samples, while B0018 provided 130 samples.

2.3. Charging health indicators

The constant-current (CC) stage was defined as the period when the measured current was higher than 1.4 A. Ten health indicators were extracted from every cleaned charge. These indicators described charging time, voltage behaviour and temperature behaviour [3, 4].

Table 1. Definition of charging health indicators

Group	Feature	Definition	Unit
Time	CC duration	Duration of the constant-current stage	s
Time	CV duration	Time from the end of the CC stage to the end of charging	s
Time	CC fraction	CC duration divided by the total CC and CV duration	-
Voltage	3.8-4.1 V rise time	Time taken for voltage to rise from 3.8 V to 4.1 V	s

DOI: 10.54254/2753-8818/2026.PU38480

Table 1. (continued)

Voltage	Starting voltage	First retained voltage value in the charge	V
Voltage	CC voltage slope	Voltage slope fitted during the CC stage	V s^{-1}
Temperature	Temperature rise	Maximum temperature minus minimum temperature	°C
Temperature	Peak temperature	Highest measured temperature	°C
Temperature	Time to peak	Time taken to reach the highest temperature	s
Temperature	Mean temperature	Mean temperature during the charge	°C

The voltage-rise time was calculated from the first time the voltage reached 3.8 V to the first time it reached 4.1 V. The CC voltage slope was calculated by fitting a straight line to the voltage values in the CC stage.

2.4. Feature sets and model inputs

The indicators were divided into three feature sets. The first set, Time, contained the three time features. The second set, Time + Voltage, contained six time and voltage features. The third set, Time + Voltage + Temperature, contained all ten features. This design was used to examine whether adding another type of sensor signal improved prediction.

Battery identity and charge number were not used as model inputs. They were only used for grouping and plotting. Charge number directly shows how long a battery has been tested, so including it could make prediction unrealistically easy. It may also be unknown when the model is applied to a used battery with an uncertain history.

2.5. Regression models and evaluation

Three regression models were compared. Linear Regression was used as a simple baseline model. A StandardScaler was fitted before Linear Regression so that features with different units were placed on similar scales. Random Forest was selected because it can learn non-linear relationships [12]. It used 300 trees and random_state=42. Gradient Boosting was used as a second non-linear model because it learns by correcting the errors of earlier trees [13]. It also used random_state=42. Test data were not used to tune the models.

Two evaluation methods were applied. First, the 628 samples were randomly divided into 80% training data and 20% testing data using random_state=42. The split was stratified by battery, so all four cells were included in both parts. This method is called the pooled random split.

Second, leave-one-cell-out (LOCO) evaluation was used. In each round, three cells were used for training and the remaining cell was used only for testing. The process was repeated four times, so every cell was used once as an unseen test cell.

Model performance was measured using mean absolute error (MAE) and the coefficient of determination (R^2):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Q_i - \hat{Q}_i\right| \tag{1}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Q_i - \hat{Q}_i\right)^2}{\sum_{i=1}^{n}\left(Q_i - \bar{Q}\right)^2} \tag{2}$$

DOI: 10.54254/2753-8818/2026.PU38480

Here, Q_i is the measured capacity, $\hat{Q}_i$ is the predicted capacity, $\bar{Q}$ is the mean measured capacity, and n is the number of samples.

A lower MAE and a higher R^2 represent better performance. Finally, leave-one-feature-out ablation was applied to the best LOCO model. Each feature was removed in turn, the model was trained again using the same LOCO procedure, and the change in MAE was recorded.

3. Results

3.1. Descriptive analysis

Figure 1 shows the capacity changes of the four battery cells during ageing. The capacity of every cell decreased as the number of charge cycles increased, although several short increases were also observed. This means that the degradation process was not completely smooth. B0006 had the highest initial capacity, at around 2.04 Ah, but its final capacity was the lowest. In comparison, B0007 kept a higher capacity near the end of the experiment.

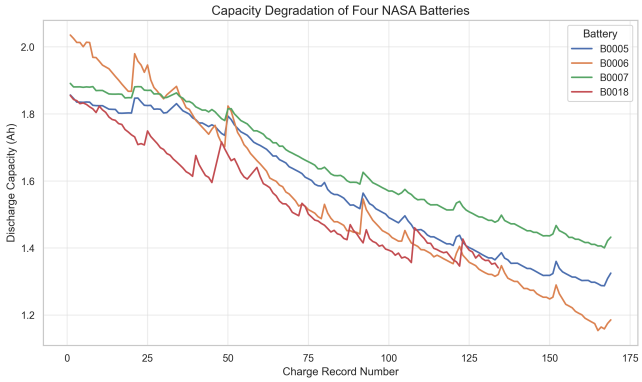


Figure 1. Capacity degradation of the four NASA battery cells

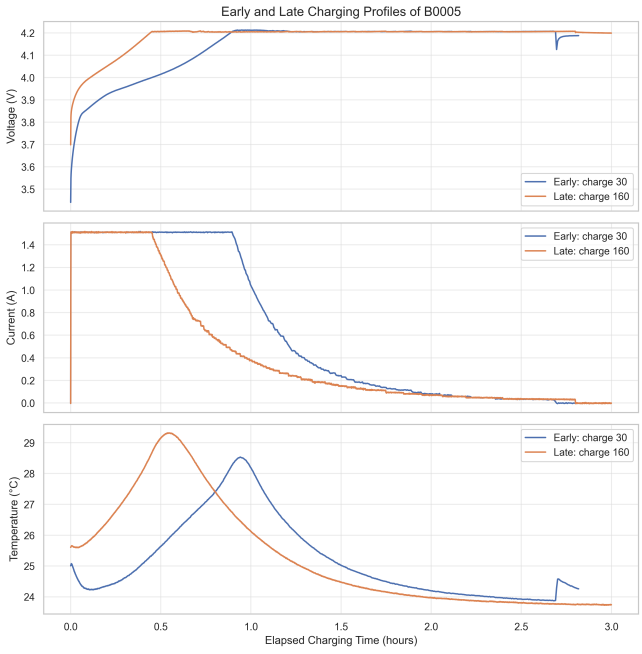


Figure 2. Early and late charging profiles of B0005

DOI: 10.54254/2753-8818/2026.PU38480

Figure 2 compares an early and a late charging cycle of B0005. In the later cycle, the battery reached the voltage limit earlier. The constant-current stage became shorter, while the constant-voltage stage formed a larger part of the charging process. Changes could also be observed in the voltage and temperature curves. These differences show that battery ageing left visible information in a single charging curve.

The correlation heatmap in Figure 3 shows that CC duration, CC fraction and the 3.8–4.1 V rise time had strong positive relationships with capacity. CV duration had a negative relationship with capacity. Most temperature features showed weaker relationships. Some time and voltage features were also highly correlated with each other.

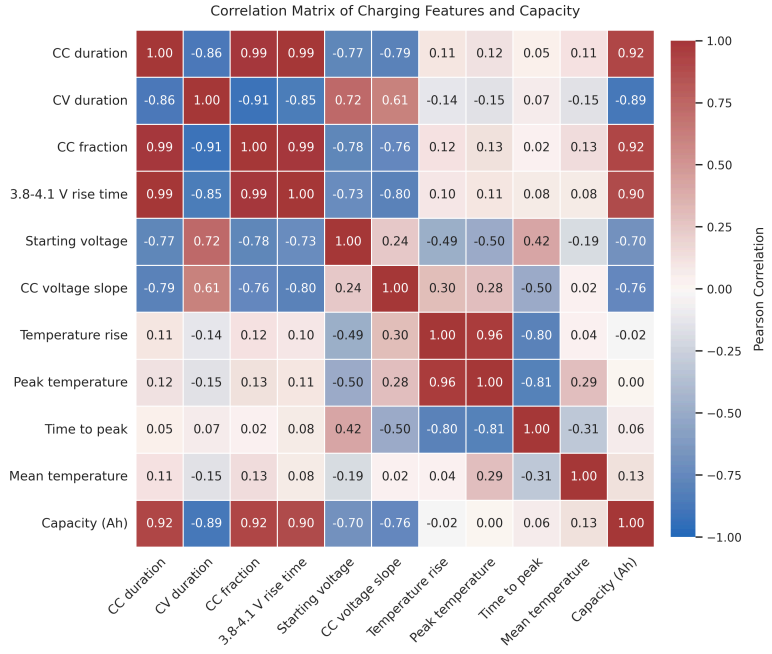


Figure 3. Correlation matrix of charging features and capacity

3.2. Comparison of models and feature sets

Table 2 presents the results from the nine model and feature combinations under the two validation methods. Under the pooled random split, all models achieved relatively strong results. The MAE values ranged from 0.0119 Ah to 0.0460 Ah, while the R^2 values ranged from 0.9026 to 0.9924.

Table 2. Model performance under two validation methods

Feature set	Model	Random MAE (Ah)	Random R^2	LOCO MAE (Ah)	LOCO R^2
Time	Linear Regression	0.0460	0.9026	0.0536	0.8673
Time	Random Forest	0.0269	0.9664	0.0546	0.8034
Time	Gradient Boosting	0.0257	0.9648	0.0542	0.8030
Time + Voltage	Linear Regression	0.0318	0.9492	0.0329	0.9446
Time + Voltage	Random Forest	0.0130	0.9891	0.0466	0.8885
Time + Voltage	Gradient Boosting	0.0165	0.9876	0.0515	0.8425
All features	Linear Regression	0.0290	0.9550	0.0375	0.9246
All features	Random Forest	0.0119	0.9888	0.0488	0.8894

DOI: 10.54254/2753-8818/2026.PU38480

Table 2. (continued)

All features	Gradient Boosting	0.0126	0.9924	0.0506	0.8795

Using only the three time features, Gradient Boosting achieved the lowest MAE of 0.0257 Ah. Adding the voltage features improved all three models. When all ten features were used, Random Forest produced the lowest MAE of 0.0119 Ah, while Gradient Boosting achieved the highest R^2 of 0.9924. Therefore, the temperature features provided a small improvement under random splitting.

The results changed under leave-one-cell-out validation. The best result was produced by Linear Regression with the time and voltage features. It achieved an MAE of 0.0329 Ah and an R^2 of 0.9446. Compared with the time-only Linear Regression model, adding voltage reduced the MAE from 0.0536 Ah to 0.0329 Ah, which was an improvement of about 38.6%. However, adding temperature increased the MAE to 0.0375 Ah.

The tree models showed a larger difference between the two validation methods. For example, the MAE of the full-feature Random Forest increased from 0.0119 Ah to 0.0488 Ah. The full-feature Gradient Boosting model increased from 0.0126 Ah to 0.0506 Ah. In contrast, the MAE of the time-and-voltage Linear Regression model only increased from 0.0318 Ah to 0.0329 Ah. Figure 4 presents this difference more clearly.

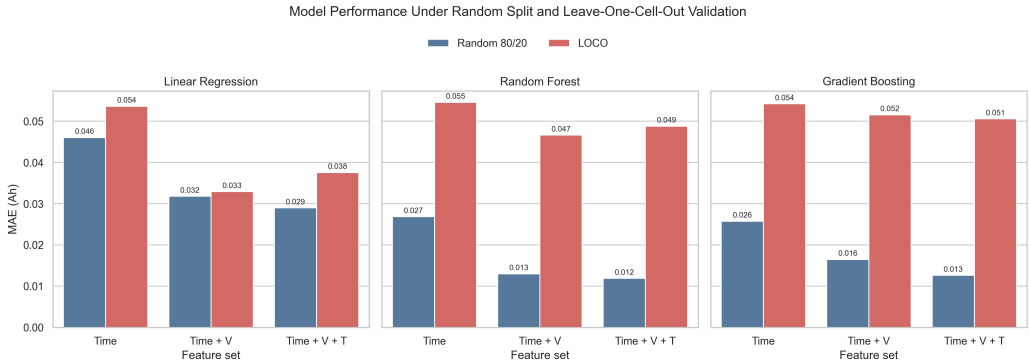


Figure 4. MAE under pooled random split and leave-one-cell-out validation

3.3. Cell-wise generalisation

Table 3 gives the cell-wise results of the best leave-one-cell-out model. B0005 and B0007 had the smallest errors, with MAE values of 0.0258 Ah and 0.0255 Ah. The results for B0018 were also strong, with an MAE of 0.0290 Ah.

Table 3. Cell-wise performance of the best cross-cell model

Held-out cell	Test samples	MAE (Ah)	R^2	Mean error (Ah)	Main direction
B0005	166	0.0258	0.9645	+0.0165	Slight overestimation
B0006	166	0.0505	0.9294	-0.0357	Underestimation
B0007	166	0.0255	0.9473	-0.0074	Slight underestimation
B0018	130	0.0290	0.9371	+0.0180	Slight overestimation

Note: Mean error = predicted capacity − measured capacity.

B0006 was the most difficult cell to predict. Its MAE was 0.0505 Ah and its R^2 was 0.9294. The predictions for this cell were, on average, 0.0357 Ah lower than the measured values. In comparison,

DOI: 10.54254/2753-8818/2026.PU38480

the model slightly overestimated B0005 and B0018. The four scatter plots in Figure 5 show that most points remained close to the 45-degree reference line, but B0006 had a wider error range.

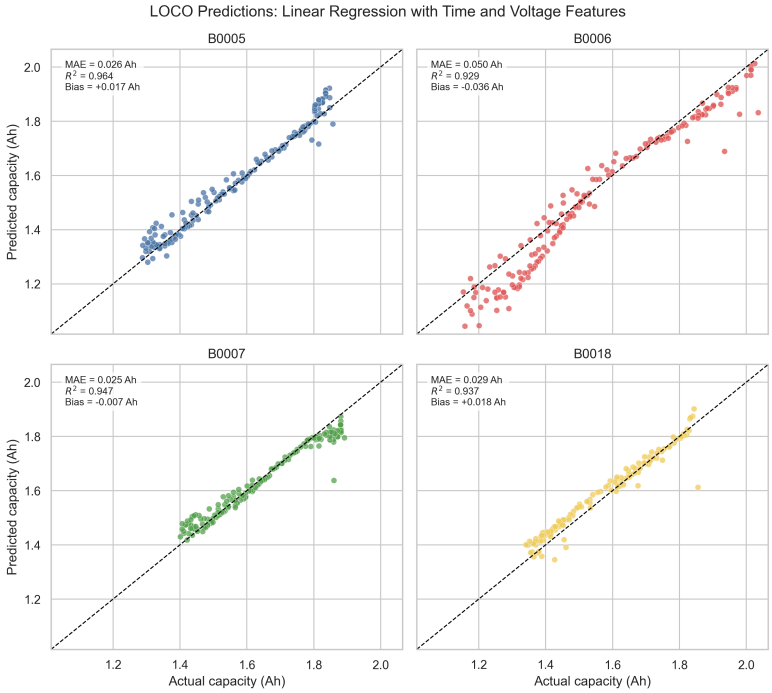


Figure 5. LOCO predictions for the four held-out cells using Linear Regression with time and voltage features

3.4. Feature contribution

Leave-one-feature-out ablation was used to examine feature contribution for the best leave-one-cell-out model. CC duration was the most useful feature. Removing it increased the MAE from 0.0329 Ah to 0.0658 Ah and reduced R^2 from 0.9446 to 0.8134.

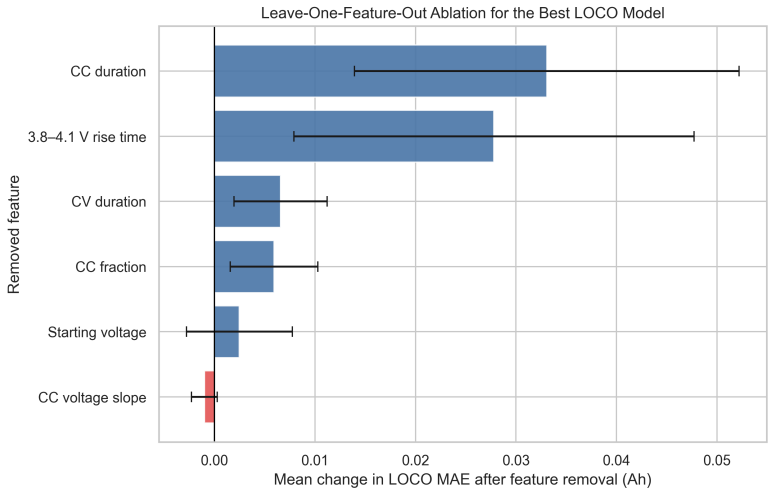


Figure 6. Leave-one-feature-out ablation results for the best LOCO model

The 3.8–4.1 V rise time was also important. After this feature was removed, the MAE increased to 0.0600 Ah. CV duration and CC fraction provided smaller but still useful contributions. Starting

DOI: 10.54254/2753-8818/2026.PU38480

voltage had only a limited effect. Removing the CC voltage slope slightly improved the result, showing that this feature did not provide useful information for cross-cell prediction. The feature contribution results are presented in Figure 6.

4. Discussion

4.1. Interpretation of the results

The main finding of this study is that the validation method changed the ranking of the models. Random Forest and Gradient Boosting performed very well under the pooled random split, but their performance decreased clearly when a completely unseen cell was used for testing. In the random split, charging cycles from the same battery could appear in both the training and test sets. As the cycles from one battery followed a similar ageing pattern, the tree models may have learned relationships that were specific to individual cells. These relationships were less useful when the models were tested on another battery [9, 10].

In comparison, Linear Regression with time and voltage features showed the best cross-cell performance. Its MAE only changed from 0.0318 Ah under random splitting to 0.0329 Ah under leave-one-cell-out validation. Although this model was simpler, it may have captured a more general relationship between the charging curve and battery capacity. It was also less likely to depend on small patterns from individual cells. This result suggests that a more complex model does not always provide better performance when the final purpose is to predict an unseen battery.

The feature comparison also gives useful information about sensor selection. Adding voltage features to the time-only Linear Regression model reduced the cross-cell MAE by about 38.6%. This shows that the shape of the voltage curve provided useful information beyond charging time. However, adding temperature features did not further improve the best cross-cell result. One possible reason is that all experiments were completed at the same ambient temperature of 24°C. Therefore, the temperature signals may not have contained enough independent information. Some temperature features may also have introduced noise or cell-specific differences.

B0006 was the most difficult cell to predict, and its capacity was mainly underestimated. Its initial capacity and degradation path were different from the other three cells. This difference was also visible in the capacity degradation curve. The result shows that variation between batteries can affect prediction even when the cells have a similar type and are tested under the same conditions.

4.2. Limitations and future work

This study has several limitations. First, only four battery cells were included, so the leave-one-cell-out evaluation contained only four test cases. All cells were also similar 18650 lithium-ion batteries. Second, the experiments used a fixed ambient temperature and a single constant-current and constant-voltage charging method. The results may therefore be different under fast charging, low temperature or other operating conditions. In addition, the data were collected in a controlled laboratory. Real vehicle data may contain incomplete charging periods, sensor noise and more changes in operating conditions. The cycles from each battery also formed a continuous ageing sequence and were not completely independent observations.

Future work should include more batteries, different temperatures, charging rates and battery chemistries. The method should also be tested with partial charging curves because a complete charge may not always be available in real use [14, 15]. External datasets could be used to examine whether the same features work under different experimental conditions [9, 16]. Finally, future

DOI: 10.54254/2753-8818/2026.PU38480

models could include prediction uncertainty and use a small amount of data from a new battery for calibration or transfer learning [9, 10, 16].

5. Conclusion

This study investigated whether lithium-ion battery capacity could be estimated from health indicators extracted from a single charging curve. The original NASA ageing records from four battery cells were processed into 628 valid samples. Ten time, voltage and temperature features were tested with Linear Regression, Random Forest and Gradient Boosting under two validation methods.

Under the pooled random split, Random Forest with all ten features achieved the lowest MAE of 0.0119 Ah. However, the performance of the tree models decreased clearly when an unseen battery was used for testing. Under leave-one-cell-out validation, Linear Regression with time and voltage features produced the best result, with an MAE of 0.0329 Ah and an R^2 of 0.9446. Its performance was also similar under both validation methods, showing better stability across different cells.

Adding voltage features reduced the cross-cell MAE by 38.6% compared with using time features alone. In contrast, adding temperature features did not provide a further improvement. The results also showed that B0006 was the most difficult cell to predict because its capacity and degradation path were different from the other cells.

Overall, charging-curve health indicators can provide useful estimates of battery capacity. However, random splitting may give an overly positive result for complex models. Testing on unseen cells is necessary before a model is considered suitable for a practical battery management system. In practice, the method could use voltage, current and time recorded during charging to estimate capacity without a separate full-discharge test. For a new cell, model selection should mainly consider the MAE under leave-one-cell-out validation and the consistency of results across cells. Further testing under real operating conditions is still needed before the method is used in a battery management system.

References

[1] Yang, B., Qian, Y., Li, Q., Chen, Q., Wu, J., Luo, E., Xie, R., Zheng, R., Yan, Y., Su, S., & Wang, J. (2024). Critical summary and perspectives on state-of-health of lithium-ion battery. *Renewable and Sustainable Energy Reviews, 190*, 114077. https://doi.org/10.1016/j.rser.2023.114077

[2] Zhou, W., Lu, Q., & Zheng, Y. (2022). Review on the selection of health indicator for lithium ion batteries. *Machines, 10*(7), 512. https://doi.org/10.3390/machines10070512

[3] Guo, Y., Huang, K., & Hu, X. (2021). A state-of-health estimation method of lithium-ion batteries based on multi-feature extracted from constant current charging curve. *Journal of Energy Storage, 36*, 102372. https://doi.org/10.1016/j.est.2021.102372

[4] Lin, C., Xu, J., Shi, M., & Mei, X. (2022). Constant current charging time based fast state-of-health estimation for lithium-ion batteries. *Energy, 247*, 123556. https://doi.org/10.1016/j.energy.2022.123556

[5] Ruan, H., Wei, Z., Shang, W., Wang, X., & He, H. (2023). Artificial intelligence-based health diagnostic of lithium-ion battery leveraging transient stage of constant current and constant voltage charging. *Applied Energy, 336*, 120751. https://doi.org/10.1016/j.apenergy.2023.120751

[6] Li, Y., Stroe, D.-I., Cheng, Y., Sheng, H., Sui, X., & Teodorescu, R. (2021). On the feature selection for battery state of health estimation based on charging-discharging profiles. *Journal of Energy Storage, 33*, 102122. https://doi.org/10.1016/j.est.2020.102122

[7] Driscoll, L., de la Torre, S., & Gómez-Ruiz, J. A. (2022). Feature-based lithium-ion battery state of health estimation with artificial neural networks. *Journal of Energy Storage, 50*, 104584. https://doi.org/10.1016/j.est.2022.104584

[8] Marri, I., Petkovski, E., Cristaldi, L., & Faifer, M. (2023). Comparing machine learning strategies for SoH estimation of lithium-ion batteries using a feature-based approach. *Energies, 16*(11), 4423. https://doi.org/10.3390/en16114423

DOI: 10.54254/2753-8818/2026.PU38480

[9] Che, Y., Zheng, Y., Onori, S., Hu, X., & Teodorescu, R. (2023). Increasing generalization capability of battery health estimation using continual learning. *Cell Reports Physical Science, 4*(12), 101743. https://doi.org/10.1016/j.xcrp.2023.101743
[10] Sheng, H., Zhou, Y., Bai, L., & Shi, L. (2022). Transfer state of health estimation based on cross-manifold embedding. *Journal of Energy Storage, 47*, 103555. https://doi.org/10.1016/j.est.2021.103555
[11] Saha, B., & Goebel, K. (2007). Battery data set. *NASA Prognostics Data Repository*, NASA Ames Research Center. https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/
[12] Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32. https://doi.org/10.1023/A:1010933404324
[13] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189-1232. https://doi.org/10.1214/aos/1013203451
[14] Xue, Q., Li, J., Chen, Z., Zhang, Y., Liu, Y., & Shen, J. (2023). Online capacity estimation of lithium-ion batteries based on deep convolutional time memory network and partial charging profiles. *IEEE Transactions on Vehicular Technology, 72*(1), 444-457. https://doi.org/10.1109/TVT.2022.3205439
[15] Zhang, Z., Gu, X., Zhu, Y., Wang, T., Gong, Y., & Shang, Y. (2025). Data-driven available capacity estimation of lithium-ion batteries based on fragmented charge capacity. *Communications Engineering, 4*, 32. https://doi.org/10.1038/s44172-025-00372-y
[16] Wang, Q., Ye, M., Cai, X., Sauer, D. U., & Li, W. (2023). Transferable data-driven capacity estimation for lithium-ion batteries with deep learning: A case study from laboratory to field applications. *Applied Energy, 350*, 121747. https://doi.org/10.1016/j.apenergy.2023.121747