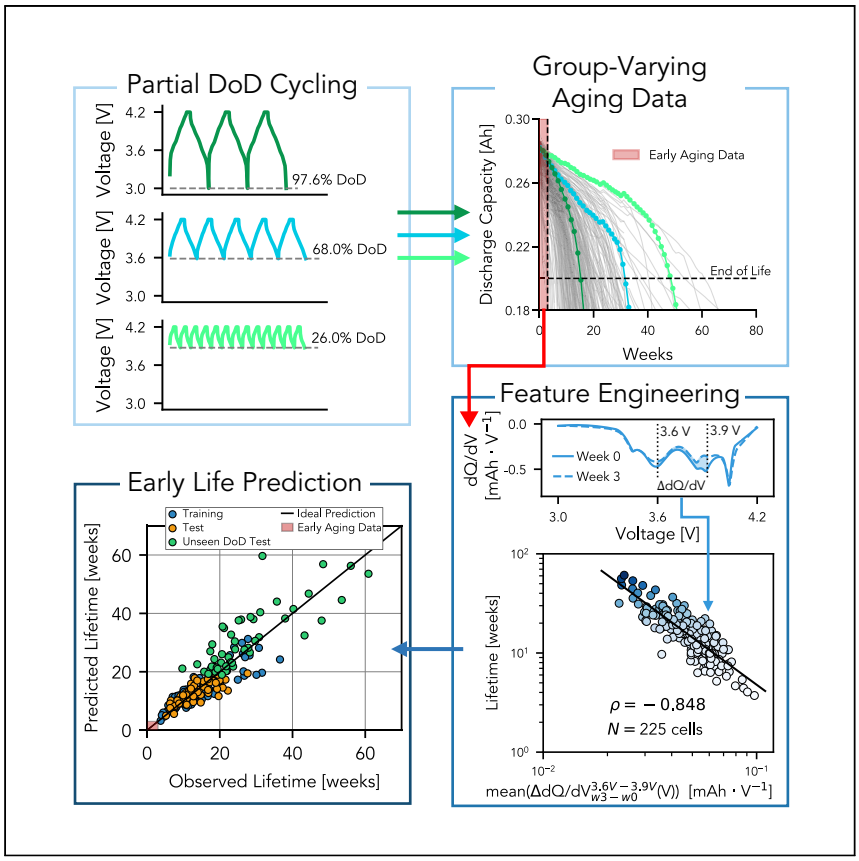
Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

Predicting battery lifetime under varying usage conditions from early aging data



Li and Zhou et al. demonstrate a method for predicting the lifetime of cells under widely varying cycling conditions using early-life measurements. This method utilizes degradation-informed features from early-life data and captures the hierarchical structure of battery aging data, showing potential for extension to different chemistries.

Tingkai Li, Zihao Zhou, Adam Thelen, David A. Howey, Chao Hu

chao.hu@uconn.edu

Highlights

A new early prediction problem, considering multiple stress factors, is defined

Predictive features with degradation-mode information are introduced

The hierarchical structure of battery aging data is captured in lifetime prediction

An aging dataset of 225 cells subjected to diverse cycling conditions is shared

Li et al., Cell Reports Physical Science 5, 101891
April 17, 2024
https://doi.org/10.1016/j.xcrp.2024.101891

Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

Predicting battery lifetime under varying usage conditions from early aging data

Tingkai Li,[1,4] Zihao Zhou,[2,4] Adam Thelen,[3] David A. Howey,[2] and Chao Hu[1,5,*]

SUMMARY

Accurate battery lifetime prediction is important for maintenance, warranties, and cell design. However, manufacturing variability and usage-dependent degradation make life prediction challenging. Here, we investigate new features derived from capacity-voltage data in early life to predict the lifetime of cells cycled under varying charge rates, discharge rates, and depths of discharge. The early-life features capture a cell's state of health and the change rate of component-level degradation modes. Using a newly generated dataset from 225 nickel-manganese-cobalt/graphite lithium-ion cells aged under a wide range of conditions, we demonstrate a lifetime prediction of in-distribution cells with 15.1% mean absolute percentage error (MAPE). A hierarchical Bayesian model shows improved performance on extrapolation, achieving 21.8% MAPE for out-of-distribution cells. Our approach highlights the importance of using domain knowledge of battery degradation to inform feature engineering and model construction. Further, a new publicly available battery lifelong aging dataset is provided.

INTRODUCTION

Understanding the long-term degradation of lithium-ion (Li-ion) batteries is crucial for their optimal manufacturing, design, and control.[1,2] However, repeatedly assessing cell performance via aging experiments is a time- and cost-intensive task.[3] Manufacturers and researchers need quick and accurate methods to screen long-term performance and to quantify the impact of new designs and control changes without having to cycle cells to the end of life (EOL) each time a new question arises. Models using data from early life could significantly shorten the time needed to make accurate predictions of long-term degradation,[4] and this could lead to rapid screening of new battery performance and optimization of charging protocols.[5–7]

The idea that lifetime can be predicted using measurements from the early stages of battery aging experiments has its roots in research from over a decade ago by J. Dahn and researchers at Dalhousie University, who were investigating the impact of new electrolyte additives and electrode designs on battery performance. In late 2009, they published a paper describing how high-precision measurements of Coulombic efficiency during the first few cycles could be used to predict cell lifetime and rank it qualitatively against other cells.[8] Coulombic efficiency is an important performance metric, and it is calculated as the discharge-to-charge capacity ratio, where an ideal value of unity indicates perfect cyclic efficiency. Measuring cell Coulombic efficiency with an error of $< 0.01\%$ can indicate cell-to-cell differences caused by different rates of undesirable side reactions that lead to capacity fade. Using purpose-built high-precision equipment, the Dalhousie team published a paper in 2011 that compared long-term cycling data (> 750 cycles) with predicted lifetimes

[1]School of Mechanical, Aerospace, and Manufacturing Engineering, University of Connecticut, Storrs, CT 06269, USA

[2]Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK

[3]Department of Mechanical Engineering, Iowa State University, Ames, IA 50011, USA

[4]These authors contributed equally

[5]Lead contact

*Correspondence: chao.hu@uconn.edu

https://doi.org/10.1016/j.xcrp.2024.101891

extrapolated from short-term (<500 h) high-precision Coulombic efficiency measurements.[9]

Since this work, many new studies have been published on "early-life prediction." In 2013, the Dalhousie University group published another paper demonstrating the lifetime ranking of 160 Li-ion cells with various electrolyte additives, using high-precision Coulombic efficiency measurements from the first 50 cycles of data.[10] The Coulombic efficiency measurements strongly correlated with the cells' lifetimes. However, many researchers and industry professionals do not have access to high-precision machines for testing. Furthermore, it would be even more useful to predict lifetimes using early-life measurements made during faster cycling experiments and under a broader range of operating conditions, enabling the technology to be deployed in more research areas and even for cells operating in the field.

Research by Baumhöfer et al.[11] and Harris et al.[12] investigated alternative approaches not requiring the use of a high-precision cycler. Baumhöfer et al. developed a lifetime prediction model on 48 cells cycled under identical conditions.[11] Hundreds of early-life features extracted from impedance spectra, pulse characterization tests at different states of charge, and standard capacity tests were reduced to a set of 24 features and used for prediction. The model using 24 features was accurate within 16 cycles; however, further analysis showed that model accuracy was highly dependent on the number of features used, with more features generally being better, suggesting the model may possibly be overfitting the small dataset ($N = 48$). Harris et al. examined the failure statistics of 24 cells cycled under identical conditions and established a weak correlation between the cells' capacity at cycle 80 and the capacity at cycle 500.[12] These works suggest simpler and more easily obtainable early-life features might be found to correlate with eventual lifetime.

Severson et al.[5] in 2019 demonstrated an early-life prediction model using features extracted from the discharge capacity vs. voltage ($Q(V)$) curves during regular cycling. The feature extraction method was unique, quantifying the cells' degradation rates by tracking the early-life variation of their $Q(V)$ curves between cycles 10 and 100, referred to as $\Delta Q_{100-10}(V)$. The approach was also used in follow-up work by Attia et al.[6] to accelerate an experimental campaign to optimize the constant current portion of a fast charging protocol. The researchers in these papers generated a large battery aging dataset from 169 lithium-iron-phosphate/graphite (LFP) cells cycled under various fast charging protocols. This was made publicly available, and many other researchers have investigated methods of further improving predictive performance and feature extraction techniques using these data.[13–22] Notably, Paulson et al.[22] demonstrated accurate early-life prediction on six different metal oxide cathode chemistries. Fermin-Cueto et al.[20] investigated predicting the knee point (when capacity begins to decrease rapidly) in a cell's capacity degradation curve using early-life features. Similarly, Li et al.[21] demonstrated a prediction model capable of projecting the entire capacity degradation trajectory from early-life features.

Despite this growing body of research, many fundamental questions about battery life modeling remain unanswered. One fundamental issue is that, in order to train machine learning models to predict lifetime from early-life cycles, data from the *entire* lifetime are required. Therefore, these approaches are best suited to applications such as screening cells after manufacturing, or relative comparisons, rather than quantitatively absolute predictions. A second issue is a lack of publicly available battery lifetime data that covers a wide range of conditions. The dataset published

by Severson et al. and Attia et al.[5,6] was specifically generated to study high-rate fast charging protocols for LFP cells, leaving the discharge rate and depth of discharge fixed. Even though the dataset is relatively large compared to existing publicly available datasets (N = 169 cells), the limited range of operating conditions, in this case, induced a single dominant degradation mode (loss of active material at the anode or negative electrode, "LAM_{NE}"), causing all of the capacity degradation trajectories to have very similar shapes, and perhaps making lifetime prediction easier.[23] While the relationships between cell operating conditions and the corresponding degradation modes are well understood,[1,3,24,25] it remains unclear how the $\Delta Q(V)$ feature transfers to cells of different chemistries and to situations where multiple interacting degradation modes are present. This is especially the case for cells that experience milder degradation resulting in less obvious changes in the $Q(V)$ curve. Furthermore, all cells in the dataset from Severson et al. and Attia et al.[5,6] were cycled under a fixed depth of discharge, making it easy to extract features from any cycle along the cell's degradation trajectory. However, in practice, the depth of discharge of cells depends on applications and user preferences, and cells may not be subjected to full depth-of-discharge cycles in many cases. So, there is a need to explore alternative methods of collecting early-life feature data and validating results using periodic reference performance tests or other means.

In this work, we investigate new early-life features derived from capacity-voltage data that can be used to predict the lifetimes of cells cycled under a wide range of charge rates, discharge rates, and depths of discharge. To study this, we generated a new battery aging dataset from 225 nickel-manganese-cobalt/graphite (NMC/Gr) cells, cycled in groups of four per condition, under a much wider range of operating conditions than existing publicly available datasets.[26] The cells in our dataset exhibit larger variations in their capacity degradation trajectories than previous open datasets, driven by the interactions and accumulations of various component-level degradation mechanisms.[1,23] To predict the lifetimes of cells experiencing different degradation pathways accurately, we introduce new early-life features extracted from the differential voltage (dV/dQ vs. Q) and incremental capacity (dQ/dV vs. V) data gathered during regular weekly reference performance tests (RPTs). The RPTs, two complete cycles at full depth of discharge, enable consistent feature extraction and lifetime prediction for cells that normally cycle at fractional depths of discharge, some as low as 4.0%. Using as little as the first 5% of the aging data, we achieve a mean absolute percentage error (MAPE) of 22% on the lifetime. Including up to 15% of the entire cell lifetime data, we achieve an average prediction error of 15.1% MAPE and a root-mean-square error (RMSE) of 2.8 weeks on in-distribution test sets when testing the new features in traditional machine learning models built with regularized linear regression. Given that our dataset has a hierarchical structure (i.e., the "group" level and the "cell" level) in nature, we also explore the possibility of applying hierarchical Bayesian linear modeling to predict lifetime, which achieves better extrapolation performance on out-of-distribution samples, viz., 7.3 weeks RMSE and 21.8% MAPE lifetime prediction error.

The major contributions of this work are 4-fold:

(1) proposing a new idea to categorize early-life features into two hierarchical levels, the condition (upper) and the cell (lower) level, in order to capture an inherent hierarchical structure in the battery aging data and enable greater generalization especially to out-of-distribution data;
(2) creating a hierarchical Bayesian model (HBM) to address the hierarchical nature of the aging data and quantify the uncertainty in lifetime predictions;

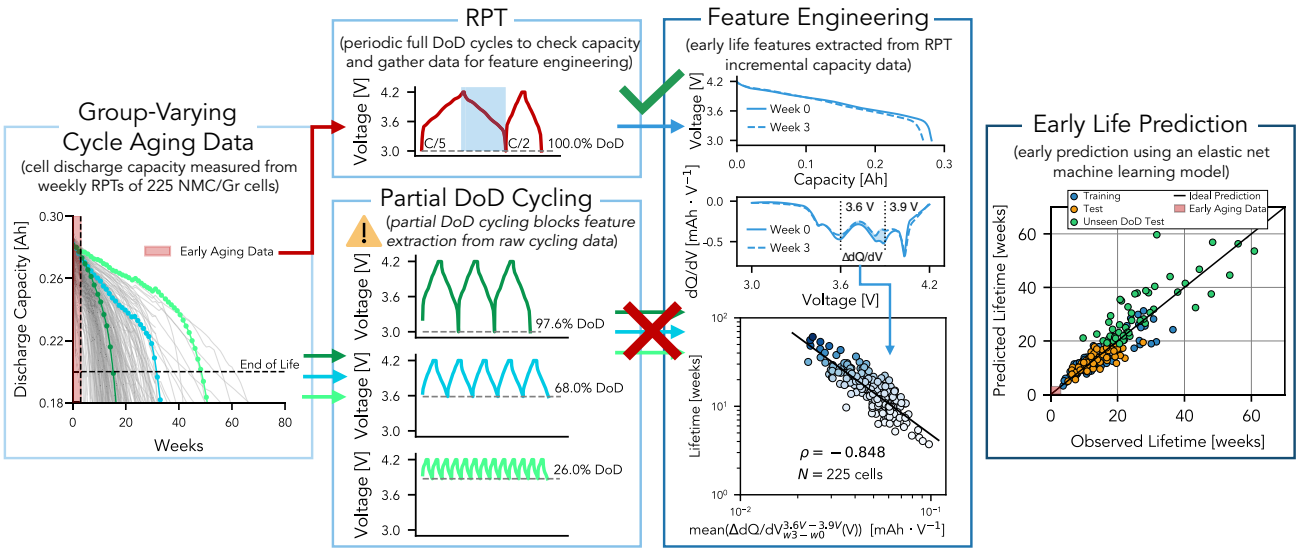


Figure 1. High-level overview of our approach

Unlike existing approaches for early prediction, we extract features from periodic reference performance tests instead of regular cycling data. In this example, we extract a feature from a partial voltage window of incremental capacity that is highly correlated with lifetime. From this and other features, we build a machine learning model to predict the lifetimes of new unseen cells.

(3) demonstrating a new method for extracting predictive features from incremental capacity curves, incorporating optimization of voltage window to improve correlations with lifetime;

(4) generating and publicly sharing a large battery aging dataset consisting of 225 NMC cells cycled under a wide range of operating conditions, enabling researchers without access to battery testing equipment to study lifetime modeling and other related topics.

RESULTS AND DISCUSSION

Approach for predicting lifetime from early-life data

Prediction of lifetime from early data is more challenging when there are multiple varying stress factors, because this leads to diverging capacity trajectories. Our approach, outlined in Figure 1, differs from the prior art[5,8–10] in several ways.

First, to apply early prediction to cells cycled under different depths of discharge, we extract features from periodic RPTs instead of regular cycling data. This means that the discharge voltage curves obtained from periodic RPTs are complete and consistent for every cell, making feature extraction more consistent. Second, we develop new features based on partial voltage windows of $Q(V)$ curves and their derivatives (differential voltage and incremental capacity data). Using a new feature extraction method (see details in "extracting features from incremental capacity data"), we find features that better correlate with cell lifetime for our dataset than existing features reported in the literature.[5,15,19] Additionally, we explore using cycling protocol information (C_{chg}, C_{dchg}, DoD) as features to predict lifetime, establishing a link between the two. All extracted features are reduced to a highly predictive subset using a feature selection method (see "feature selection"). Then, the selected features are used as input to a machine learning model to predict cell lifetime. In what follows, we outline our approach to feature engineering for early-life prediction and discuss the challenges of applying existing feature engineering methodologies proven on

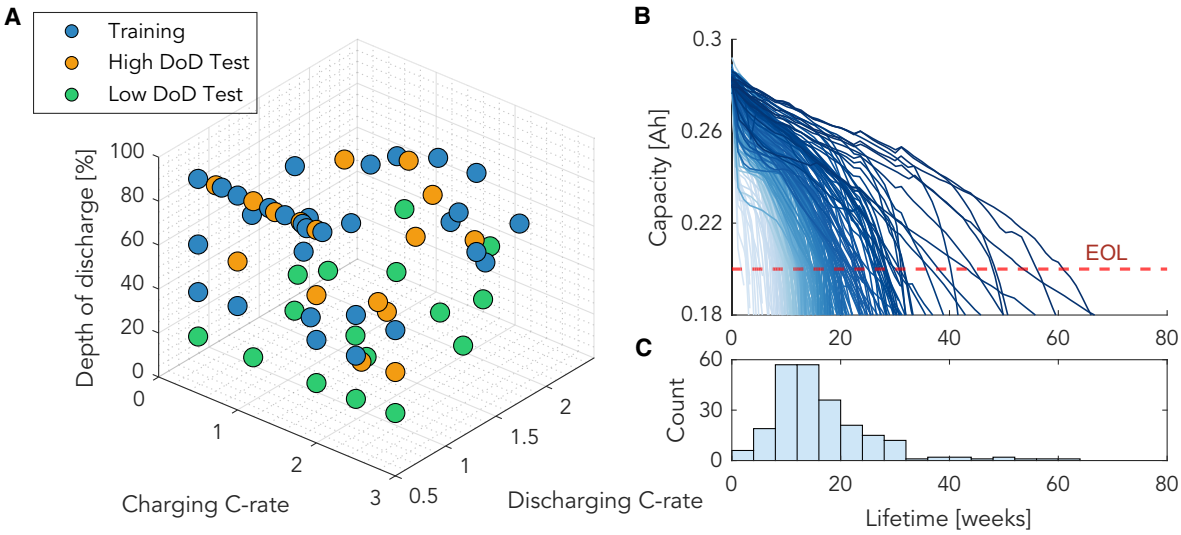


Figure 2. Overview of battery aging test conditions and capacity data

(A) 3D scatterplot showing train-test split and cycling conditions used—each point represents conditions for a group of four cells, and marker color indicates a data subset used to generate prediction results in "predicting lifetime using machine learning models."

(B) Discharge capacity fade curves for all 225 NMC/graphite cells plotted past 80% their rated capacity (250 mAh); color of each curve is scaled by cell lifetime.

(C) Histogram of the cell lifetimes at end of life (EOL) using 80% of rated capacity as threshold.

LFP/Gr to our NMC/Gr cells that are cycled under a wider range of operating conditions. Last, we introduce HBMs for early-life prediction.

Lithium-ion battery dataset under varying usage conditions

Publicly available datasets, such as those from NASA's Prognostics Center of Excellence,[27,28] the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland, College Park,[29,30] and Sandia National Laboratories,[31] contain cells of different chemistries cycled under a range of charge rates, discharge rates, and temperatures. These datasets are frequently used in research studies for battery modeling since aging commercial-grade Li-ion cells is slow and expensive. Also, these datasets report measurements including capacity, internal resistance (NASA and CALCE), voltage, current, and temperature, enabling researchers to study different aspects of battery modeling. However, the relatively small size of these datasets (roughly 30 cells per group) makes investigating machine learning-based approaches to early-life prediction challenging. On the other hand, in recent years, datasets such as those from the Toyota Research Institute[5,6] and Argonne National Laboratory[22] contain many more cells (>150 cells). However, they focus on a limited range of operating conditions—fast charging and symmetric C/2 cycling, respectively—making it difficult to build machine learning models that generalize across cycling conditions.

In light of this, we designed our battery aging dataset to study more cells under a broader range of operating conditions than current publicly available datasets.[26] Our dataset comprises 225 cells cycled in groups of four to capture some of the intrinsic cell-to-cell aging variability.[32] A unique feature of our dataset is the many capacity degradation trajectories that reflect different accumulated degradation modes induced by the various operating conditions. These trajectories, shown in Figure 2, exhibit different one-, two-, and three-stage degradation trends driven by the interaction and accumulation of hidden, threshold, and snowballing

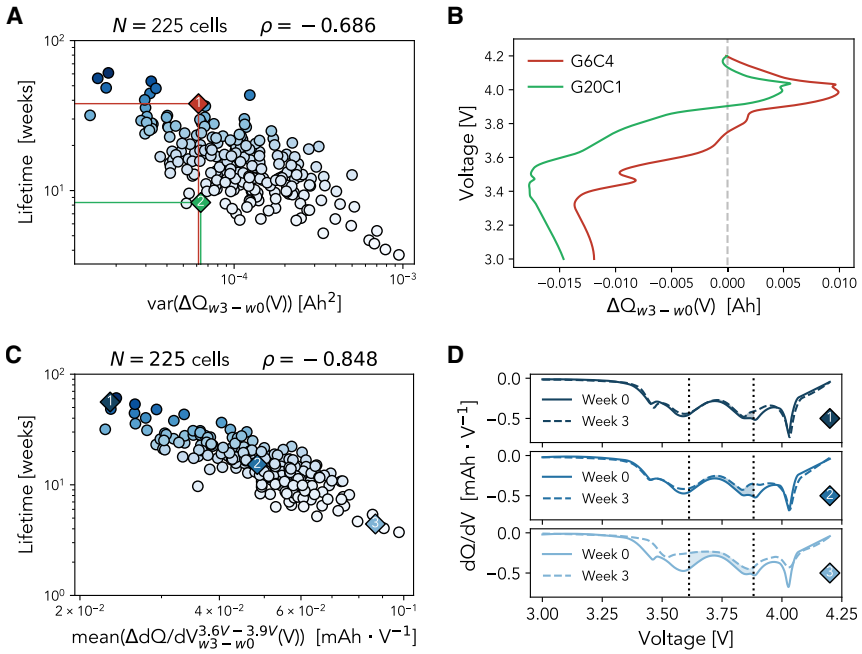


Figure 3. Feature extraction from early-life data

Well-known early-life features do not explain the variance in our dataset, and a newly extracted feature from incremental capacity curves correlates better with lifetime.

(A) Cell lifetime for 225 NMC cells plotted as a function of var($\Delta Q_{w3-w0}(V)$); Pearson correlation coefficient −0.686. The two cells highlighted have similar values of var($\Delta Q_{w3-w0}(V)$) but very different lifetimes.

(B) Difference between discharge capacity curves as a function of voltage between week 3 and 0 for the two cells highlighted in (A).

(C) Cell lifetime plotted as a function of optimized feature mean($\Delta dQ/dV_{w3-w0}^{3.60V-3.90V}(V)$), Pearson correlation coefficient − 0.848.

(D) Incremental capacity curves from weeks 3 and 0 for three representative cells (G1C2, G16C1, and G53C2, respectively); the shaded areas indicate the change of curves between the voltage bound (3.60 − 3.90 V) after 3 weeks of cycling.

degradation modes.[23] These varying trends produce cell lifetimes from 1.5 to 60.9 weeks. Experimental details and testing procedures used to generate the dataset can be found in "battery aging test design" and Note S1 and Note S2 of the supplemental information.

Extracting predictive features from early usage data

Initially, we extracted features previously reported to correlate strongly with cell lifetime.[5,15,19] We adopt the notation $\Delta Q_{w3-w0}(V)$ to describe the features, where the subscripts w3 and w0 correspond to data obtained from the RPTs from weeks 3 and 0, respectively. Preliminary testing of these well-established early-life features reveals that they do not fully explain the variance in our dataset. This is illustrated in Figure 3A, where we extract the var($\Delta Q(V)$) feature reported by Severson et al.[5] using discharge data from RPTs var($\Delta Q_{w3-w0}(V)$) and plot it against lifetime, revealing a large unexplained variance in the predicted lifetimes.

To understand why this occurs, consider two cells (G6C4 and G20C1) that have similar feature values but vastly different lifetimes. In this case, even though the $\Delta Q(V)$ curves have the same variance, they do not have the same shape and location (Figure 3B). It can be seen that the group twenty cell (G20C1) experienced more

significant capacity loss during this time, evident by the endpoint of $\Delta Q(V)$ at 3.0 V. Other noticeable changes exist in the $dV/dQ(Q)$ curves that differ between the cells (shown in Note S7 of the supplemental information), indicating additional but more subtle degradation modes are present. However, these differences in the evolution of the $Q(V)$ curve during early life are not captured by the $\mathrm{var}(\Delta Q_{w3-w0}(V))$ feature, causing the unexplained variance in the dataset. A further analysis compares the $\Delta Q_{w3-w0}(V)$ curves from our dataset with the $\Delta Q_{100-10}(V)$ curves from the dataset of Severson et al.,[5] shown in Figure S8 of the supplemental information. From this analysis, we find that the $\mathrm{var}(\Delta Q_{w3-w0}(V))$ feature only captures the capacity fade within the first 3 weeks in our dataset, which can be attributed to the differences in the trend of the $Q(V)$ curve evolution. Both the differences in cell chemistries between these two datasets and different discharging profiles from which the variance feature is extracted could contribute to the trend differences. The former impacts the overall discrepancy of voltage responses due to different phase change mechanisms involved during discharging. The latter, especially for the difference in C-rates, would cause different levels of heat generation during discharging. The discharge curve for which Severson et al.[5] introduced their feature is at 4 C, and the effect of self-heating on the voltage response cannot be neglected under this high C-rate, especially toward the end of discharge. In contrast, we extracted the equivalent $\mathrm{var}(\Delta Q_{w3-w0}(V))$ feature from measurements at C/5, where self-heating is much less of a concern.

While we only showed an example in Figure 3 for this particular feature, $\mathrm{var}(\Delta Q_{w3-w0}(V))$, the unexplained variance in the data persists using most other early-life features we tested. Typically, it is not a requirement that all model input features exhibit a strong correlation with cell lifetime, but finding a few features that do correlate well is generally advantageous because it can improve model fit and accuracy. Considering this, we explored extracting features from differential voltage and incremental capacity curves using partial voltage intervals in order to capture the diverse cell-specific degradation trends observed in our dataset more accurately.

Differential voltage (dV/dQ vs. Q) and incremental capacity (dQ/dV vs. V) curves have been widely adopted in battery aging diagnostics because certain features (e.g., peaks and valleys) on these curves are closely associated with phase transitions of electrodes and allow us to investigate electrode-specific aging modes. For example, simply plotting dQ/dV vs. V curves over successive cycles within a long-term aging test and observing changes in positions, amplitudes, and widths of certain peaks over cycling can help detect the underlying degradation modes (e.g., loss of active materials on the negative and positive electrodes and loss of lithium inventory associated with capacity fade) that drive these changes. This diagnostic property of differential voltage and incremental capacity curves could even allow reasonably accurate quantification of degradation modes.[33] Also, several battery lifetime prediction studies include features describing the locations and magnitudes of peaks from differential voltage and incremental capacity curves.[15,19,22] Here, we present a method to find an optimal incremental capacity feature, which is backed up by our feature selection process in "feature selection" and Note S11 and Note S12 of the supplemental information.

Rather than only using locations and magnitudes of peaks from incremental capacity ($dQ/dV(V)$) curves, we examine how the incremental capacity curve evolves over selected voltage intervals. Specifically, segments of the incremental capacity curve within a certain voltage interval are extracted from two RPT tests (the week 0 RPT and

the week 3 RPT, respectively). Then, two major summary statistics—the mean and the variance—of the difference between these two extracted segments are used as features to reflect the evolution of the incremental capacity curve within this voltage interval during the 3 weeks. Rather than manually identifying the voltage interval, a grid search method is employed to find an optimal voltage interval with the highest correlation between feature and lifetime (see "extracting features from incremental capacity data" for more details). The voltage interval search result using the mean summary statistics is shown in Figure S9 of the supplemental information.

We find the voltage interval that produces the highest linear correlation with cell lifetime is a mid-range where the upper and lower voltage limits are centered around prominent peaks in the incremental capacity curves at 3.60 V and 3.90 V. Figure 3C shows that the change in incremental capacity in this range is inversely proportional to lifetime. The new feature captures the change in incremental capacity intensity, calculated as the mean change in mAh/V over the middle voltage interval, $\text{mean}(\Delta dQ/dV_{w3-w0}^{3.60V-3.90V}(V)) = \text{mean}(dQ/dV_{w3}^{3.60V-3.90V}(V) - dQ/dV_{w0}^{3.60V-3.90V}(V))$; see Figure 3D. This new feature shows a much stronger correlation with cell lifetime and better explains the variance in our dataset compared with the traditional feature $\text{var}(\Delta Q_{w3-w0}(V))$.

The new feature derived from the incremental capacity curve likely captures the rate of active material loss during early life. This idea is supported by degradation diagnostics literature, which shows that changes in the intensity of the incremental capacity (mAh/V) curve at constant voltage correspond to a loss of active material.[1,33–35] Also, a close relationship between this feature and average cycling stress, with a Pearson correlation coefficient of 0.89, can be found in Figures S11 and S14 of the supplemental information, where an increase in our proposed feature is accompanied by an increase in the Stress_{avg} feature introduced in "extracting features from cycling conditions." Based on the assumption that higher stresses induced during cycling are likely to cause more loss of active materials, the found relationship helps back up our hypothesis that this incremental capacity feature captures the loss of active materials in general. Additional analysis to understand this feature regarding degradation information is included in Note S8 of the supplemental information. Additionally, we use the upper and lower voltage limits imposed during cycling to create two more intervals, 3.00 – 3.60 V and 3.90 – 4.20 V. We then extract two features from each voltage interval using the mean and variance summary statistics. In total, we extracted six features from $\Delta dQ/dV(V)$, two from each of the three voltage intervals using the mean and variance summary statistics.

Lifetime modeling work on NMC/Gr cells by Smith et al.[36] showed that the capacity fade rate due to cycling tracked nearly linearly with the square-root-of-cycling throughput, calculated as $(C_{chg}\text{DoD})^{0.5}$, where C_{chg} is charging C-rate, and DoD is depth of discharge for the experiments. This metric is described as tracking the concentration gradient of lithium ions in the cathode active material and is a proxy for diffusion-induced stress.[36–38] We further investigate this feature as a model input for early-life prediction ("extracting features from cycling conditions") and as a condition-level grouping variable for our hierarchical Bayesian modeling approach ("hierarchical Bayesian models for early prediction").

In addition to the incremental capacity and stress-related features, the processes of extracting features from other data sources (e.g., differential voltage, constant-voltage

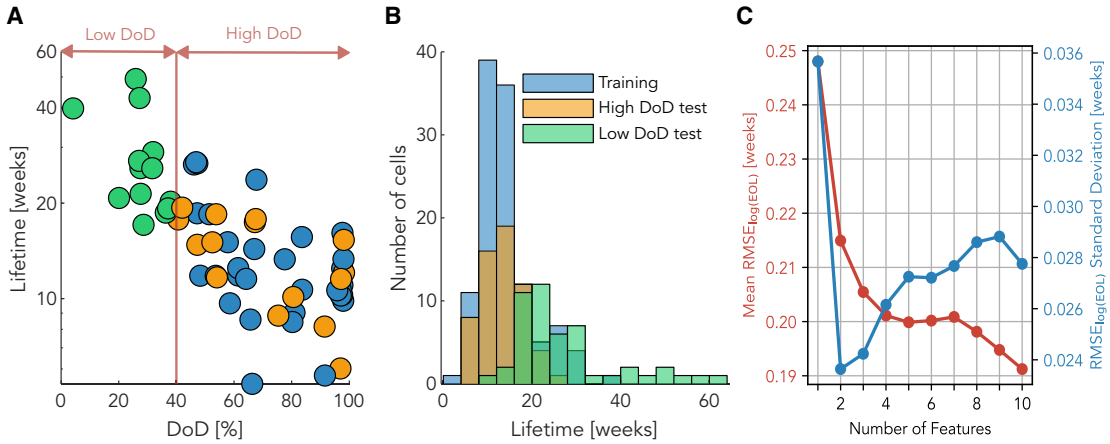


Figure 4. Overview of data partition and feature selection

(A) Scatterplot of mean group lifetime vs. DoD; marker color indicates train/test subset.

(B) Histogram showing each subset's distribution of cell lifetimes.

(C) Mean and standard deviation of $RMSE_{log(EOL)}$ for 5-fold repeated cross-validation on the ten candidate models.

charging segment, and capacity fade) are reported in Note S9 and Note S10 of the supplemental information. By extracting features from a total of 6 sources, we obtained a set of 29 candidate features, outlined in Table S2 of the supplemental information. All the early-life features in this work were extracted from C/5 cycles in RPTs.

Partitioning data for machine learning and feature selection

Dataset partitioning was done at the group rather than the cell level, for three reasons. First, practical battery aging tests for product validation typically cycle multiple cells under the same conditions to capture the aging variability due to manufacturing. Second, it is desirable to build an early prediction model to predict the lifetimes of cells cycled under previously untested conditions. Finally, although building an early prediction model with cells tested under rapidly accelerated aging conditions is useful in minimizing the time and costs of collecting aging data, one cannot preemptively know the lifetime (before tests), so grouping must be done using an alternative indicator of cell lifetime. Since the depth of discharge is the dominant cycling stress factor impacting the battery lifetimes in our aging dataset (Figure 4A), this was used to determine the dataset partitioning.

We first separate our dataset into a high-DoD region and a low-DoD region, with a boundary at 40% depth of discharge (Figure 4A). In the high-DoD region, we further divide the data into a training set and an in-distribution high-DoD test set. The high-DoD test set is used to evaluate the model's prediction accuracy for cells with conditions similar to the ones the model was trained on. Last, we assign all data in the low-DoD region ($<40\%$) to a second test set used to test the model's ability to extrapolate to unseen test conditions. The dataset split is also visualized in Figure 2A, where each axis is one of the three cycle aging stress factors (C_{chg}, C_{dchg}, DoD), and the marker color indicates the data subset that the group belongs to. The training set contains cells with lifetimes ranging from 3.7 to 36.6 weeks, and the high-DoD test set has cells with lifetimes between 5.2 and 31.6 weeks. On the other hand, the low-DoD test set is more diverse, with lifetimes ranging from 9.7 to 60.9 weeks. Histograms of cell lifetimes for each data subset are visualized in Figure 4B.

After extracting a total of 29 features, down-selecting a smaller feature subset is required before training machine learning models for a couple of reasons. First,

Table 1. Stepwise forward search results

Step number	Selected feature	Description
1	$\log(\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.6V-3.9V}(V))$	best incremental capacity feature from "extracting features from incremental capacity data" (shown in Figure 3C)
2	$\log(\lvert\Delta \mathrm{CV\ Time}_{w3-w0}\rvert)$	change in CV hold time (see Note S10 of the supplemental information)
3	DoD	depth of discharge
4	ΔQ_{w3-w0}^{1}	change in DVA-based capacity $Q^{DVA,1}$ (see Note S9 of the supplemental information)
5	$C_{chg}{}^{0.5}\mathrm{DoD}^{0.5}$	charge-induced stress (see "extracting features from cycling conditions")
6	C_{chg}	charging C-rate
7	$\log(\mathrm{var}(\Delta dQ/dV_{w3-w0}^{3.0V-3.6V}(V))$	variance of low-voltage incremental capacity feature (see "extracting features from incremental capacity data")
8	ΔQ_{w3-w0}^{3}	change in DVA-based capacity $Q^{DVA,3}$ (see Note S9 of the supplemental information)
9	$\log(\lvert\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.0V-3.6V}(V)\rvert)$	mean of low-voltage incremental capacity feature (see "extracting features from incremental capacity data")
10	$\log(\lvert\Delta \mathrm{CV\ Time}_{w0}\rvert)$	CV hold time of the initial RPT (see Note S10 of the supplemental information)

some features are strongly correlated with one another, known as multicollinearity. A model trained with collinear features can be sensitive to minor changes in the feature values and may extrapolate poorly.[39] Second, while the new dataset is large (225 cells) relative to existing publicly available aging datasets, it is still considered small for training a machine learning model. Small datasets require special care to avoid overfitting and to improve generalization performance on unseen test data. This is especially the case when the number of training points is not significantly larger than the number of features ($N_{data} \gg N_{features}$). Therefore, it is crucial to select a subset of highly predictive features before training.[40,41] We performed stepwise forward feature selection on the training dataset following the method described in "feature selection." To avoid poor performance on the test sets due to overfitting, we performed a 5-fold cross-validation study five times, using up to 10 features. With the cross-validated errors, including the mean and standard deviation for each candidate model, one can evaluate the candidate feature subsets from two aspects. As the mean of $RMSE_{EOL}$ decreases, the feature subset fits this training set better in general. As the standard deviation increases, the feature subset yields larger run-to-run variation for different folds, which indicates unstable performance on prediction. So, when selecting features, one must consider both aspects simultaneously to ensure good performance, especially for extrapolating to the low-DoD test set. The trends of the mean and standard deviation of cross-validation $RMSE_{log(EOL)}$ of this trial are reported in Figure 4C, and the selected feature in each step is listed in Table 1. The model with two features, namely $\log(\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.6V-3.9V}(V))$ and $\log(\lvert\Delta \mathrm{CV\ Time}_{w3-w0}\rvert)$, has the lowest run-to-run variance and relatively low mean error $RMSE_{EOL}$. Adding a third feature to the set, DoD, produces a model with lower mean $RMSE_{EOL}$ but increases the run-to-run variation. For a more comprehensive evaluation, we compare the results of models trained using both two and three features.

Predicting lifetime using machine learning models

There are two types of machine learning models used in this study: one is elastic net regression, and the other one is the HBM. Except for the dummy model and HBM, the rest of the models compared in this section are trained with elastic net models

Table 2. Prediction errors for selected models tested using the high- and low-DoD test datasets

		MAPE (%)			RMSE (weeks)		
Model	$N_{feature}$	Training	High DoD	Low DoD	Training	High DoD	Low DoD
Dummy model	0	35.0	31.5	47.5	6.5	4.8	18.5
Cycling conditions	3	24.8	19.0	23.7	4.0	3.3	9.8
Discharge model[5]	5[a]	23.9	28.0	24.8	4.6	4.7	11.5
Degradation-informed	2	17.3	16.0	24.4	3.2	3.0	7.8
Degradation-informed	3	16.5	15.1	33.0	3.1	2.8	9.7
HBM	2[b]	18.6	16.9	21.8	3.3	3.1	7.3
HBM	3[b]	17.4	15.8	24.1	3.1	2.9	7.5

[a]The discharge model[5] contains six features, with one of them being the difference between the maximum capacity and capacity at cycle two, ΔQ_{max-2}. However, this feature cannot be calculated for our dataset due to the partial depth of discharge cycling and the continuously decreasing capacity-fade curves for all cells and has thus been omitted.

[b]The number of features listed refers to the number of cell-level input features. For both HBMs, a single cycling condition-level feature is used for grouping cells, and, as indicated in the table, either two or three cell-level features are used for regression.

on different feature sets. To predict the lifetime, we initially establish a pair of baseline models. The first baseline model is a dummy model that does not use any input features or have any trainable parameters, and instead, it predicts the mean cell lifetime of the training set for all cells. This is a good way to determine if a more complex model is truly learning new information from the input data or instead only appears to be learning because of similar training/test dataset distributions that lead to similar error metrics. When tested on the two test datasets, the dummy model achieves $MAPE_{EOL}$ of 31.5% and 47.5% on the high-DoD and low-DoD test sets, respectively. The error metrics for all models tested are shown in Table 2.

The second baseline model is built using only the cycling condition parameters as input features. This model predicts lifetimes without using cell-specific aging measurements. This model achieves an $MAPE_{EOL}$ of 19.0% and 23.7% on the high-DoD and low-DoD test sets, respectively. The substantial decrease in prediction error over the dummy model shows that the usage parameters convey a significant amount of information that can be used to predict lifetime accurately. This result is expected, as a great deal of battery lifetime modeling work[36,42,43] has already explored the strong connection between usage and degradation. However, only using condition-level cycling features does not account for intrinsic cell-to-cell variability. Hence, the next set of models we tested included cell-level features extracted from the early aging data.

The first cell-level feature model is the "discharge model" described in Severson et al.[5] and "extracting predictive features from early usage data." This model, and all other models built on cell-level inputs, use features extracted from the RPTs of weeks 0 and 3, which is just under 18% of the average lifetime. The original "discharge model" has six features, with five of them from statistics of $\Delta Q(V)$ curves and one feature ΔQ_{max-2} capturing the capacity recovery at initial cycles. Due to the monotonic decreasing capacity trajectories observed in our dataset, an analogous feature for ΔQ_{max-2} cannot be extracted and thus is omitted. The main feature included is $var(\Delta Q_{w3-w0}(V))$; however, we found that this did not completely describe the variance in our dataset. When tested on the high- and low-DoD test datasets, the discharge model achieved 28.0% and 24.8% $MAPE_{EOL}$, respectively. The performance on the two test datasets is slightly worse than the cycling condition

model, yet still better than the dummy model, indicating that the features used in the discharge model do carry useful information but are not optimal for our dataset (see Table 2).

The remaining models we compare are the degradation-informed model and HBM. We refer to our elastic net models as *degradation-informed* in Table 2 because of the newly developed degradation-based features used as model inputs. Both the degradation-informed model and HBM use the same sets of input features, and for thoroughness, we compare models built using two and three features each. Compared to the cycling condition baseline, the two-feature elastic net model shows decreased $MAPE_{EOL}$ on the high-DoD test of 16.0% and a slight increase in error on the low-DoD test set to 24.4%. However, the $RMSE_{EOL}$ of the low-DoD test set drops considerably from 9.8 to 7.8 weeks. For the HBMs, we observe small increases in the training and the high-DoD test errors with a noticeable improvement in the low-DoD test errors over the degradation-informed models using the same set of features.

For both the degradation-informed and hierarchical models, we observe that including the third feature decreases model prediction error on the training and high-DoD test datasets but increases error for the low-DoD test dataset. When the third feature is added, both models overfit the training dataset and exhibit poor extrapolation capability to the low-DoD test dataset where the cells have longer lifetimes. Regardless, the HBM trained with three features still performs better when predicting the low-DoD test set compared with its elastic net counterpart. Generally, by comparing the evaluation metrics of the two models (degradation-informed model and HBM), we find that the HBM has better generalizability to the low-DoD test set but at the cost of slightly higher training and high-DoD test errors.

The large improvement in performance observed for models using cell-level (as opposed to only using cycling condition features) features prompts us to further investigate why the feature $\log(\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.6V-3.9V}(V))$ explains cell-to-cell variability better than other features. Firstly, it is more accurate to use measured health metrics from individual cells in operation to predict their lifetime. This reveals the intrinsic cell-to-cell variability that could cause different aging behaviors under identical cycling conditions. Secondly, this optimized feature, which likely captures how much loss of active material happens during early life, has a balanced representation of the variability within the group and among the entire dataset.

In summary, we find that the best feature $\log(\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.6V-3.9V}(V))$ explains the cell-to-cell variability well for a majority of cells. The remaining variance in the feature-lifetime correlation may be contributed jointly by measurement inaccuracy and unexplained manufacturing variability. Hence, our analysis of the results suggests that a predictive early-life feature should capture the variability introduced by the difference in cycling conditions and information about intrinsic cell-to-cell variation that causes different performances under identical loads. Also, our feature engineering methodology ("extracting features from incremental capacity data") can be extended to find good features for other cell chemistries. Additional analyses on benchmarking different feature subsets and machine learning models are included in Note S16 of the supplemental information.

Analysis of HBM results

The probabilistic nature of HBMs enables us to extract a deeper understanding by considering both the mean and the uncertainty of lifetime predictions. Assuming

individual cluster fitting parameters and noise variance, $\boldsymbol{\theta}_j$ and σ_j respectively, are independent, the posterior predictive distribution can be written as

$$p\left(y_j^*\middle|Y_j\right) = \iint p(\sigma_j|Y_j)p(\boldsymbol{\theta}_j|Y_j)p\left(y_j^*\middle|\boldsymbol{\theta}_j,\sigma_j\right)d\boldsymbol{\theta}_j d\sigma_j, \quad \text{(Equation 1)}$$

where y_j^* is the predicted lifetime of a new cell belonging to the j th group, and Y_j represents all lifetime observations associated with the j th group. For a point-wise prediction, one can estimate the mean value of $p(y_j^*|Y_j)$. Table 2 lists the performance of the HBM built using two different feature sets. The first uses two cell-level features, $\log(\left|\text{mean}(\Delta dQ/dV_{\text{w3}-\text{w0}}^{3.6V-3.9V}(V)\right|)$ and $\log(\Delta\text{CV Time}_{\text{w3}-\text{w0}})$, and achieves 3.1 weeks RMSE and 16.9% MAPE for the high-DoD test set, which is almost the same as the performance of the degradation-informed model using the same feature set, while for the low-DoD test set, the HBM achieves 7.3 weeks RMSE and 21.8% MAPE, which outperforms the degradation-informed model by 7% and 10% for RMSE and MAPE, respectively.

Similar to the degradation-informed model, we observe that the HBM model overfits the training dataset when the third feature (DoD) is added. This is evident by the increased performance on the training and high-DoD test set but worse performance on the low-DoD test set. Specifically, under the high-DoD test set, RMSE improved from 3.1 to 2.9 weeks, and MAPE improved from 16.9% to 15.8%. However, for the low-DoD test set, RMSE increased from 7.3 to 7.5 weeks, and MAPE increased from 21.8% to 24.1%. Notably, the HBM shows more resistance to overfitting than the degradation-informed model, whose performance decreased substantially more than the HBM when the third feature was included in the feature set. Similar results can also be found in Tables S3 and S4 of the supplemental information, where the HBM consistently shows better robustness on unseen test samples (i.e., the low-DoD test set) compared to elastic net models on different feature sets or different machine learning algorithms on the same feature set.

Figure 5B shows the uncertainty (2 standard deviations) of $p(y_j^*\left|Y_j\right.)$ for posterior lifetime predictions of each cluster. The uncertainty levels for clusters 0 and 1 are around $\pm$ 4.5 weeks (at 2 SD), whereas for clusters 2 and 3, the uncertainty levels are around $\pm$ 9.5 and 10.5 weeks, respectively, which reflects the model's uncertainty when predicting cells from unseen cycling conditions. According to Table 3, there are only 12 cells from cluster 3 in the training set, while there are 23 cells from cluster 3 in the low-DoD test set. Due to the lack of data, the uncertainty for all regression parameters $(\boldsymbol{\theta_3},\sigma_3)$ for cluster 3 is much larger than that of clusters 0 and 1. On the other hand, as the prediction uncertainty becomes large for long-life cells, uncertainty itself can be used as an indicator to denote whether one should include more early-life data for feature calculation. For example, when running HBM in a forward mode (using the trained model to give predictions), for test samples in cluster 3, large prediction uncertainty is observed (>10 weeks). One may consider including the 4th or 5th week of training data to retrain the model so that the prediction uncertainty on cluster 3 test samples can be reduced. Since the used 3 weeks of training data only take up 7% of the average lifetime for cluster 3 samples, using 1–2 more weeks train data still only covers the very early stage of these long-life cells.

Further analysis of uncertainty for model parameters can be found in Note S15 of the supplemental information. This uncertainty on both lifetime predictions and model parameters can be more beneficial to real-world applications compared to only a point-wise prediction. For example, instead of knowing the exact EOL lifetime,

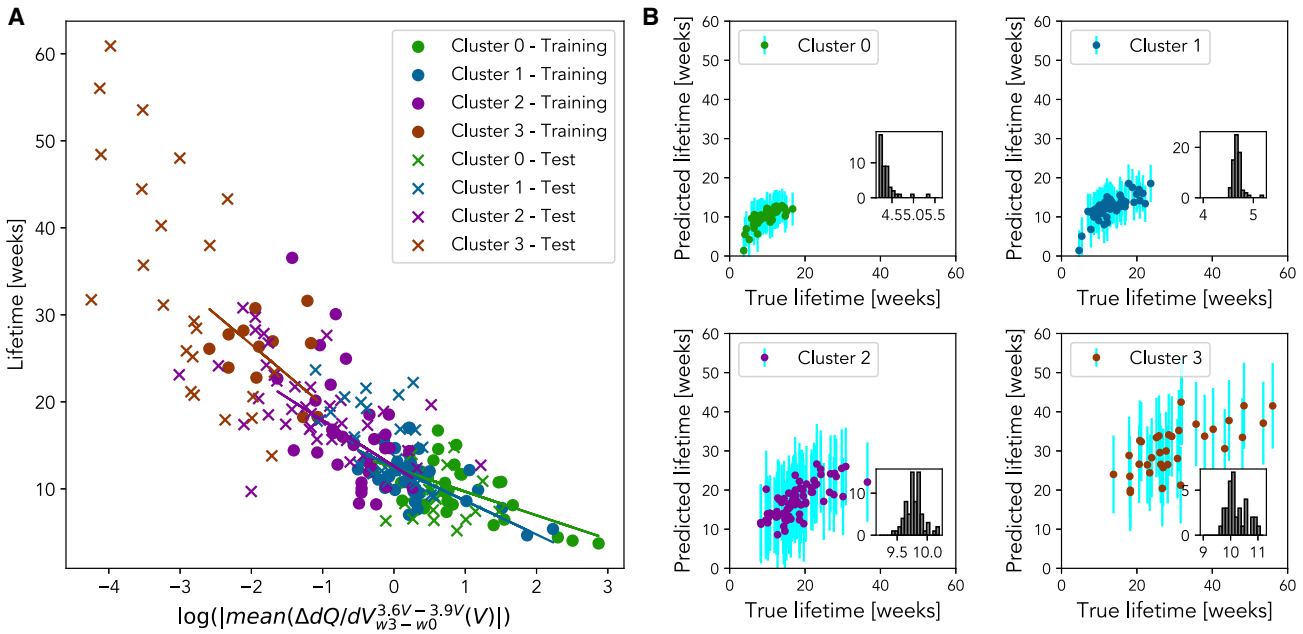


Figure 5. Overview of HBM prediction results

(A) Relationship between $\log(|\text{mean}(\Delta dQ/dV_{w3-w0}^{3.6V-3.9V}(V)|)$ and true lifetime across different clusters and train-test split ("Test" denotes samples from both high- and low-DoD sets). Fits, corresponding to mean parameter values, are plotted for each cluster.

(B) Predictions for each cluster with 2 standard deviations as the corresponding error bar for each sample. The embedded histograms show a summary of error bars.

customers care more about a warranty for the worst-case lifetime, which can be satisfied by using the standard deviation of prediction distributions.

In this study, we have developed two data-driven models to tackle the problem of battery early-life prediction on a large and unique aging dataset, which consists of 225 NMC cells cycled under a wide range of charge and discharge C-rates (0.5 C–3 C) and DoDs (4%–100%). Our feature engineering process identifies a new predictive feature, $\text{mean}(\Delta dQ/dV_{w3-w0}^{3.60V-3.90V}(V))$, derived from incremental capacity curves and closely related to the degradation induced by loss of active materials. Also, our analysis shows that the widely used $\Delta Q(V)$ features in the existing early prediction literature may not explain cell-to-cell lifetime variability within our dataset.

In terms of results, two distinct machine learning models are trained to predict the lifetime. Our degradation-informed model, trained using elastic net regression, yields 3.0 and 7.8 weeks RMSE and 15.1% and 33.0% MAPE on the high- and low-DoD test sets, respectively. The HBM produces 3.1 and 7.3 weeks RMSE and 16.9% and 21.8%

Table 3. Summary of train-test split for each cluster

		n samples		
Cluster ID	Stress_{avg}	Training	High-DoD test	Low-DoD test
0	1.12	30	18	0
1	0.95	41	24	4
2	0.76	33	18	22
3	0.51	12	0	23
Total	0.86	116	60	49

MAPE for the high- and low-DoD test sets, respectively. While the HBM shows performance improvement for point-wise predictions on the low-DoD test set, it also gives uncertainty information for its predictions, which can be used in applications like the cell lifetime warranty. And we found that the uncertainty grows across groups with the decrease of cycling stress factor $Stress_{avg}$, which indicates the lack of observability for cell-to-cell differences from early-life features, and thus more cycling time range may need to be included for cells under mild cycling conditions.

A limitation of this work is that the models are demonstrated on battery aging data collected in a well-controlled laboratory setting under constant cycling conditions over the life of the cells. However, depending on the applications, battery data from real-world applications may be more variable and noisy, posing a challenge to feature extraction and lifetime prediction. To investigate this further, we will expand the dataset by aging cells using electric grid duty cycle protocols (e.g., protocols simulating peak shaving and frequency regulation). This duty cycle dataset is currently being collected in the Reliability Engineering and Informatics Laboratory at the University of Connecticut. It will add a valuable addition to the ISU-ILCC dataset, particularly in evaluating lifetime models on cells with dynamic cycling profiles.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Requests for any additional information about this work should be directed to the lead contact, Chao Hu (chao.hu@uconn.edu).

Materials availability

No new material was generated in this study.

Data and code availability

- The battery aging dataset collected and used for this work is available for download from the open-access data repository of Iowa State University (Iowa State University DataShare) at https://doi.org/10.25380/iastate.22582234.[44] Please refer to the dataset as the *ISU-ILCC NMC/Gr battery aging dataset*. A sample code for preprocessing the data is included.
- The code for feature extraction and early prediction modeling is available at https://doi.org/10.5281/zenodo.10648587.

Cell and tester specifications

The Li-ion cells used in this study were commercial 502030 size Li-polymer cells with NMC as the positive electrode and graphite as the negative electrode, manufactured by Honghaosheng Electronics in Shenzhen, China. The rated capacity is 250 mAh (giving 1 C as 250 mA), and the operating voltage ranges from 3.0 to 4.2 V. All cells were tested on two 64-channel Neware BTS4000 battery testers, in thermal chambers set at 30°C.

Battery aging test design

The aging experiments were designed around three main stress factors that impact battery lifetime: charge rate (C_{chg}), discharge rate (C_{dchg}), and depth of discharge (DoD). To make the scope of our aging campaign manageable and work within the limitations of our equipment, we decided not to introduce temperature as an additional variable stress factor. Therefore, as also mentioned in "cell and tester specifications," all cells were placed in thermal chambers set at 30°C over the entire span of the aging experiments. The design of experiments for our aging campaign

involved deliberately subjecting various groups of cells to different stress levels between RPTs, achieved by randomly sampling the three stress factors within a wide design space. This treatment contributed to a wide distribution of the EOL in this dataset. However, this aging dataset is limited in reflecting realistic usage profiles, particularly dynamic charging/discharging profiles where the charge and discharge rates vary rapidly over time.

To track the full discharge capacity of cells with partial depths of discharge cycling, we periodically ran RPTs that measured cell capacity and gathered complete $Q(V)$ data for feature engineering. Each RPT consisted of two cycles performed at slow rates (C/2 and C/5) to capture cell voltage response while minimizing the impact of the cell kinetics. Before beginning the aging tests, an initial RPT was conducted to determine the beginning-of-life health. Aging tests consisted of 1 week of cycling followed by an RPT, and they were repeated until cell capacity decreased below 200 mAh (80% of the rated capacity). In real field operations, the necessity of incorporating RPTs to obtain consistent measurements (e.g., capacity-voltage data) from cells varies. Taking electric vehicles (EVs) as an example, the onboard charge controller could regulate the charging process, ensuring all EVs of the same model operate under certain constant-current conditions until sufficient data are collected. Then, data-driven diagnostics or prognostics can be achieved solely based on these data.[45] However, in grid storage applications, both the charging and discharging profiles can be random and noisy in operation, especially the profiles for frequency regulation.[46] Extracting clear predictive features can be practically infeasible in these applications. Therefore, periodic RPT cycles can help collect consistent measurements to estimate the state of health and predict lifetime. Since the grid may have multiple power-generation sources (e.g., power plants, wind farms, and solar farms) and large battery storage systems, the downtime of battery storage systems during diagnostic cycles, ideally executed in alternating batches, would have less impact on the overall operation of the grid or storage system. Furthermore, the frequency of RPTs can be less often, i.e., monthly or quarterly, to provide flexibility and avoid disruption in real applications—the weekly frequency in our aging campaign is to capture the rapid degradation due to the nature of the accelerated aging test.

As previously mentioned, four cells were cycled at each test condition. We refer to a specific cell using its group number and cell identifier, e.g., G7C3, where the numbers following each letter indicate the group and cell, respectively. Initially, we aimed to study two stress factors: DoD and C_{chg}. Conditions were selected using a grid search, with the discharge rate fixed at 0.5 C for all cells. Later, we expanded the dataset to study the third stress factor, C_{dchg}. Additional conditions were then selected using random sampling. The charge/discharge rates and depths of discharge were sampled evenly from the ranges 0.5 C–3 C and 25%–100%, respectively.

The cycling conditions for all cell groups can be found in Table S1 of the supplemental information. However, the depth of discharge design values do not exactly match the measured depths of discharge from the cycling experiments. When we programmed the cycling protocols, we determined the cutoff voltages using a reference discharge capacity vs. voltage curve from a cell cycled at C/2. Unfortunately, the voltage hysteresis that the cells experience under C/2 discharge causes the cells to reach the cutoff voltage quicker than expected, thus causing the difference between the measured and designed depth of discharge. In this paper, we present and discuss the depth of discharge using the actual

measured values since they more accurately represent the test conditions the cells experienced.

Extracting features from incremental capacity data

Extracting features from incremental capacity curves is a natural extension to using the $Q(V)$ discharge curve since it is defined over the same fixed voltage range for every cell. After fitting a spline and downsampling each cell's $Q(V)$ curve to 1,000 points, we calculated incremental capacity ($dQ/dV(V)$) as a finite difference approximation (difference quotient) of the first derivative of $Q(V)$ based on measurements of the Q and V time series.[5] It is well documented that incremental capacity analysis is an effective method for cell degradation diagnostics.[1,33,47] Measuring changes to the incremental capacity curve over the lifetime enables the diagnosis of different degradation modes, specifically loss of lithium inventory, and loss of active material in each electrode. Hence, we calculate core summary statistics of $\Delta dQ/dV(V)$ over a partial voltage interval so as to focus the feature extraction on specific areas that may correspond to specific degradation modes. This approach is inspired by work in Greenbank and Howey,[13] where the authors showed a strong correlation between the time a cell spends in a specific voltage interval and its capacity loss, although here the incremental capacity curve is a result of degradation rather than a cause. Instead of manually specifying the voltage interval to calculate the summary statistics, we exhaustively searched the entire $3.0 - 4.2$ V range in increments of 0.01 V, with a minimum window size of 0.02 V searching for the maximum Pearson correlation coefficient.

Literature in this field reports a wide range of possible features for lifetime prediction, either derived from capacity, voltage, and temperature measurements during cycling or other measurements such as impedance spectra from electrochemical impedance spectroscopy.[5,11,13,19,22] Most of these features are hand-crafted based on direct mathematical manipulation and are not always optimized for maximum correlation with lifetime. As a distinction from other feature extraction processes, the method we used to extract the $\Delta dQ/dV(V)$ feature optimizes the correlation coefficient by an exhaustive grid search. So, even though the optimal voltage interval identified for our aging dataset, i.e., $3.60 - 3.90$ V, may not directly apply to other aging datasets, the general methodology of optimizing a voltage interval when identifying early-life features allows researchers the possibility of extracting features of higher predictive power from their own datasets.

Extracting features from cycling conditions

As briefly mentioned in "extracting predictive features from early usage data," we consider a set of stress-related features for early-life prediction, which is $\text{Stress}_{\text{chg}} = {C_{\text{chg}}}^{0.5}\text{DoD}^{0.5}$. This feature captures the square-root-of-cycling charge throughput and is a proxy for diffusion-induced stress in the electrode active materials.[36–38] In addition to the charge-based feature, we also calculate a discharge feature, $\text{Stress}_{\text{dchg}} = {C_{\text{dchg}}}^{0.5}\text{DoD}^{0.5}$. Further, to capture the effects of different charge and discharge rates in a single feature, we calculate an average stress feature as $\text{Stress}_{\text{avg}} = (\text{Stress}_{\text{chg}} + \text{Stress}_{\text{dchg}})/2$ and also calculate a multiplicative stress feature as $\text{Stress}_{\text{mult}} = \text{Stress}_{\text{chg}} \cdot \text{Stress}_{\text{dchg}}$. For all features, we use the measured DoD from the first week of cycling in the calculation. A unique characteristic of these features is that they require no cell-specific measurements, assuming the calculation of DoD is accurate and accounts for voltage hysteresis. For this reason, these features are excellent candidates as condition-level grouping variables in our hierarchical Bayesian modeling approach to early prediction (see "hierarchical Bayesian models for early prediction").

Feature selection

To minimize collinearity and the risk of overfitting, we perform stepwise forward selection using a linear model and repeated cross-validation with $RMSE_{EOL}$ as the evaluation metric. Starting with a null model, one feature is added to the model for each step until the number of selected features reaches a preset threshold ($N_{feature} = 10$). During each step, all features are tested in the model, and the feature that reduces the mean of the cross-validation $RMSE_{EOL}$ the most is selected and added to the model for the next step. Simultaneously, we evaluated the selected model at each step using the standard deviation of the cross-validation $RMSE_{EOL}$. We then select the features to use corresponding to the set with a balance between low mean and small standard deviation of cross-validated $RMSE_{EOL}$. In practice, we tend toward selecting fewer features so that the resulting model will be less complex and extrapolate better.

Elastic net regression for lifetime prediction

To predict cell lifetime, we formulate a regression problem with the extracted early-life features $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m]$ as inputs and the measured cell lifetimes $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]^T$ in logarithmic scale as outputs, where m is the number of early-life features, and n is the number of cells. Each element of $\boldsymbol{X}$ is a column vector containing the specific features selected through the technique introduced in "feature selection." We assume that the lifetime is a linear function of the early-life features, giving

$$\widehat{\boldsymbol{y}} = f(\boldsymbol{X}) = \boldsymbol{\beta}_0 + \boldsymbol{X}\boldsymbol{\beta}_1, \qquad \text{(Equation 2)}$$

where $\boldsymbol{\beta}_0$ is an $n \times 1$ column vector of the intercept, and $\boldsymbol{\beta}_1$ is a vector of coefficients, one for each feature, $\boldsymbol{\beta}_1 = [\beta_1, \beta_2, \ldots, \beta_m]^T$.

To find the coefficients of this equation, we formulate an optimization problem with elastic net regularization, which is a combination of L_1 and L_2 penalization. The objective function is

$$\widehat{\boldsymbol{\beta}} = \underset{\beta_0, \boldsymbol{\beta}_1}{\operatorname{argmin}} \left(\|\boldsymbol{y} - \boldsymbol{\beta}_0 - \boldsymbol{X}\boldsymbol{\beta}_1)\|_2^2 + \lambda \left(\frac{1-\alpha}{2} \|\boldsymbol{\beta}\|_2^2 + \alpha \|\boldsymbol{\beta}\|_1 \right) \right), \qquad \text{(Equation 3)}$$

where α and λ are hyperparameters that control the balance between the L_1 and L_2 penalties and the magnitude of regularization, respectively. To select optimal values of α and λ, we perform repeated cross-validation using randomized dataset splits.

Hierarchical Bayesian models for early prediction

As a comparison and contrast to the method in the previous section, we also consider HBMs for lifetime prediction. These have a layered structure that can model changes in the feature-target relationship throughout the dataset. HBMs have been applied to model naturally structured data in various research fields, from ecology to sociology, psychology, and computer vision.[48,49] Several studies in the battery field explored the concept of HBM to solve different problems, such as estimating the state of health against Ah throughput[50] and identifying parameters of an equivalent circuit model.[51] However, their applications were not early-life prediction, and the model structures adopted in their studies differed from those used in this work.

Clustering for hierarchical modeling

For our problem of early-life prediction, features can be viewed as coming from two levels: the "cycling condition" level and the "individual cell" level. Condition-level features relate to user-defined test protocols rather than measured data. For our dataset, the charge/discharge C-rates and depth of discharge (C_{chg}, C_{dchg}, DoD), and any mathematical combination of these are all condition-level features. In contrast,

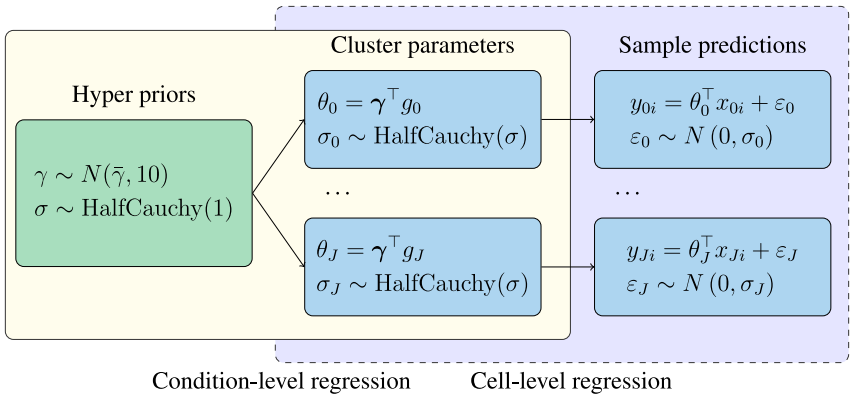


Figure 6. Overview of HBM structure

Model parameters can be classified as either individual-level ($\boldsymbol{\theta}_j, \boldsymbol{\sigma}_j$) or conditional-level ($\boldsymbol{\gamma}, \boldsymbol{\sigma}$); j represents cycling condition group index, i represents individual cell index, y_{ji} represents lifetime of i th cell in j th cycling group. The two-level structure allows the individual cell-level feature-label ($x_{ji} - y_{ji}$) relationship to vary with cycling condition based on cycling condition-level features ($\boldsymbol{g}_j$). The hyper-priors are assumed to be $\gamma \sim N(\bar{\gamma} = 0, 10)$ and $\sigma \sim \text{HalfCauchy}(\beta = 1)$.

features that require specific cell measurements during cycling are considered cell-level features. Features such as mean($\Delta dQ/dV_{\text{w3}-\text{w0}}^{3.60\text{V}-3.90\text{V}}(V)$) and var($\Delta Q_{\text{w3}-\text{w0}}(V)$) are examples of cell-level features that are unique to each cell.

To validate the hypothesis that conditional-level features have a strong impact on the relationship between cell-level features and lifetime, we calculate the condition-level feature $\text{Stress}_{\text{avg}} = (\text{C}_{\text{chg}}{}^{0.5}\text{DoD}^{0.5} + \text{C}_{\text{dchg}}{}^{0.5}\text{DoD}^{0.5})/2$ described in "extracting features from cycling conditions." This represents the average diffusion-induced stress that a cell experiences.[36] To take advantage of an HBM's ability to model the change in feature-target relationship across different levels, we investigate clustering cell data based on cycling conditions, quantified by average stress ($\text{Stress}_{\text{avg}}$). In general, we expect cells with similar average stress levels to share the same feature-lifetime relationship, enabling the HBM to better fit the dataset. We adopt a constrained K-means clustering algorithm,[52] which is an improved version of the traditional K-means algorithm that imposes minimum and maximum cluster size limits. An optimal cluster number $K = 4$ is used in later analysis, where details can be found in Note S15 of the supplemental information.

Bayesian hierarchical linear model

Similar to the HBM used in our preliminary study,[53] the model in this study also has a bi-level structure, as shown in Figure 6. The first level considers the cycling condition parameters. As mentioned previously, cells are first divided into four clusters (indexed from 0) based on their average stress $\text{Stress}_{\text{avg}}$, calculated using the cycling condition parameters.

At this level, we aim to find the mapping (parameterized by $\boldsymbol{\gamma}, \sigma$) between condition-level features ($\boldsymbol{g}_j$) and the cell-level regression parameters ($\boldsymbol{\theta}_j, \sigma_j$). Notice that, different from Zhou and Howey,[53] the noise terms σ_j are assumed to come from the same hyper-prior distribution HalfCauchy(σ):

$$\begin{aligned} \boldsymbol{\theta}_j &= \boldsymbol{\gamma}^\top \boldsymbol{g}_j \\ \sigma_j &\sim \text{HalfCauchy}(\sigma) \end{aligned} \qquad \text{(Equation 4)}$$

After the coefficients ($\boldsymbol{\theta}_j, \sigma_j$) are decided for each cluster, the individual cell-level regression is built as the second level of the HBM. The cell-level regression uses

individual health features (x_{ji}) and coefficients (θ_j, σ_j) to give lifetime predictions (y_{ji}) for individual cells.

$$y_{ji} \sim N\left(\theta_j{}^{\top} x_{ji}, \sigma_j^2\right) \quad \text{(Equation 5)}$$

The overall training objective is to infer posterior distributions for both the condition-level model and the individual cell-level models, $P(\theta_j|Y_j)$ and $P(\gamma|\{Y\})$ respectively, where Y_j represents lifetimes from only the j th group but $\{Y\}$ represents data from all lifetimes. More details about the training procedure are included in Note S14 of the supplemental information.

Model evaluation metrics

We use two standard error metrics to evaluate the lifetime prediction accuracy of our approaches, namely, mean absolute percentage error ($MAPE_{EOL}$) and root mean squared error ($RMSE_{EOL}$), both calculated using the measured and predicted values of cell lifetime on a linear scale. The metrics are

$$MAPE_{EOL} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad \text{(Equation 6)}$$

$$RMSE_{EOL} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \quad \text{(Equation 7)}$$

where y are the measured cell lifetimes, $\hat{y}$ are the predicted cell lifetimes, and n is the number of cells. These two error metrics are commonly used to measure the prediction accuracy in regression problems. MAPE measures the overall prediction error on a percentage scale, while RMSE penalizes predictions with larger errors, making this error metric more sensitive to outliers.[54] Lower values are more desirable for both error metrics, indicating the model has better prediction accuracy.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrp.2024.101891.

ACKNOWLEDGMENTS

We acknowledge the hard work of Jinqiang Liu from Iowa State University and Chad Tischer and Reuben D. Schooley from Iowa Lakes Community College for executing and maintaining the battery aging tests. We also want to acknowledge Murtaza Zohair for assembling the half-cells used in this study.

The work at Iowa State University and the University of Connecticut was partly supported by the Iowa Economic Development Authority under the Iowa Energy Center grant no. 20-IEC-018 and partly by the US National Science Foundation under grant no. ECCS-2015710. The China Scholarship Council and the Department of Engineering Science supported the work at the University of Oxford. Any opinions, findings, or conclusions in this paper are those of the authors and do not necessarily reflect the sponsors' views.

AUTHOR CONTRIBUTIONS

Conceptualization, T.L., A.T., Z.Z., C.H., and D.H.; data collection, data management, and raw data processing, T.L.; investigation, methodology, visualization,

software, formal analysis, and writing – original draft, T.L., A.T., and Z.Z.; writing – review and editing, T.L., A.T., Z.Z., C.H., and D.H.

DECLARATION OF INTERESTS

D.H. is a co-founder of Brill Power Ltd.

Received: August 25, 2023
Revised: November 22, 2023
Accepted: February 29, 2024
Published: March 22, 2024

REFERENCES

1. Birkl, C.R., Roberts, M.R., McTurk, E., Bruce, P.G., and Howey, D.A. (2017). Degradation diagnostics for lithium ion cells. J. Power Sources *341*, 373–386. https://doi.org/10.1016/j.jpowsour.2016.12.011.

2. Sulzer, V., Mohtat, P., Aitio, A., Lee, S., Yeh, Y.T., Steinbacher, F., Khan, M.U., Lee, J.W., Siegel, J.B., Stefanopoulou, A.G., and Howey, D.A. (2021). The challenge and opportunity of battery lifetime prediction from field data. Joule *5*, 1934–1955. https://doi.org/10.1016/j.joule.2021.06.005.

3. Thelen, A., Lui, Y.H., Shen, S., Laflamme, S., Hu, S., Ye, H., and Hu, C. (2022). Integrating physics-based modeling and machine learning for degradation diagnostics of lithium-ion batteries. Energy Storage Mater. *50*, 668–695. https://doi.org/10.1016/j.ensm.2022.05.047.

4. Kunz, M.R., Dufek, E.J., Yi, Z., Gering, K.L., Shirk, M.G., Smith, K., Chen, B.-R., Wang, Q., Gasper, P., Bewley, R.L., and Tanim, T.R. (2021). Early battery performance prediction for mixed use charging profiles using hierarchal machine learning. Batter. Supercaps *4*, 1186–1196. https://doi.org/10.1002/batt.202100079.

5. 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.

6. 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.

7. Dave, A., Mitchell, J., Kandasamy, K., Wang, H., Burke, S., Paria, B., Póczos, B., Whitacre, J., and Viswanathan, V. (2020). Autonomous discovery of battery electrolytes with robotic experimentation and machine learning. Cell Reports Physical Science *1*, 100264. https://doi.org/10.1016/j.xcrp.2020.100264.

8. Smith, A.J., Burns, J.C., Trussler, S., and Dahn, J.R. (2010). Precision measurements of the coulombic efficiency of lithium-ion batteries and of electrode materials for lithium-ion batteries. J. Electrochem. Soc. *157*, A196. https://doi.org/10.1149/1.3268129.

9. Burns, J.K., Nashed, J.Y., Blohm, G., Eberman, K., Scott, E., Gardner, J., and Dahn, J. (2011). Evaluation of effects of additives in wound Li-ion cells through high precision coulometry. J. Vis. *11*, 3. https://doi.org/10.1149/1.3531997.

10. Burns, J.C., Kassam, A., Sinha, N.N., Downie, L.E., Solnickova, L., Way, B., and Dahn, J.R. (2013). Predicting and extending the lifetime of Li-ion batteries. J. Electrochem. Soc. *160*, A1451–A1456. https://doi.org/10.1149/2.060309jes.

11. Baumhöfer, T., Brühl, M., Rothgang, S., and Sauer, D.U. (2014). Production caused variation in capacity aging trend and correlation to initial cell performance. J. Power Sources *247*, 332–338. https://doi.org/10.1016/j.jpowsour.2013. 08.108.

12. Harris, S.J., Harris, D.J., and Li, C. (2017). Failure statistics for commercial lithium ion batteries: A study of 24 pouch cells. J. Power Sources *342*, 589–597. https://doi.org/10.1016/j.jpowsour.2016.12.083.

13. Greenbank, S., and Howey, D. (2022). Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life. IEEE Trans. Ind. Inf. *18*, 2965–2973. https://doi.org/10.1109/TII. 2021.3106593.

14. Zhang, Y., Peng, Z., Guan, Y., and Wu, L. (2021). Prognostics of battery cycle life in the early-cycle stage based on hybrid model. Energy *221*, 119901. https://doi.org/10.1016/j.energy.2021.119901.

15. Yang, F., Wang, D., Xu, F., Huang, Z., and Tsui, K.-L. (2020). Lifespan prediction of lithium-ion batteries based on various extracted features and gradient boosting regression tree model. J. Power Sources *476*, 228654. https://doi.org/10.1016/j.jpowsour.2020.228654.

16. 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*, 2971–2992. https://doi.org/10.1016/j.joule.2021.09.015.

17. Saxena, S., Ward, L., Kubal, J., Lu, W., Babinec, S., and Paulson, N. (2022). A convolutional neural network model for battery capacity fade curve prediction using early life data. J. Power Sources *542*, 231736. https://doi.org/10.1016/j.jpowsour.2022.231736.

18. Herring, P., Balaji Gopal, C., Aykol, M., Montoya, J.H., Anapolsky, A., Attia, P.M., Gent, W., Hummelshøj, J.S., Hung, L., Kwon, H.-K., et al. (2020). BEEP: A python library for battery evaluation and early prediction. SoftwareX *11*, 100506. https://doi.org/10.1016/j.softx.2020.100506.

19. Fei, Z., Yang, F., Tsui, K.-L., Li, L., and Zhang, Z. (2021). Early prediction of battery lifetime via a machine learning based framework. Energy *225*, 120205. https://doi.org/10.1016/j.energy.2021.120205.

20. Fermín-Cueto, P., McTurk, E., Allerhand, M., Medina-Lopez, E., Anjos, M.F., Sylvester, J., and Dos Reis, G. (2020). Identification and machine learning prediction of knee-point and knee-onset in capacity degradation curves of lithium-ion cells. Energy and AI *1*, 100006. https://doi.org/10.1016/j.egyai.2020.100006.

21. Li, W., Sengupta, N., Dechent, P., Howey, D., Annaswamy, A., and Sauer, D.U. (2021). One-shot battery degradation trajectory prediction with deep learning. J. Power Sources *506*, 230024. https://doi.org/10.1016/j.jpowsour.2021. 230024.

22. Paulson, N.H., Kubal, J., Ward, L., Saxena, S., Lu, W., and Babinec, S.J. (2022). Feature engineering for machine learning enabled early prediction of battery lifetime. J. Power Sources *527*, 231127. https://doi.org/10.1016/j.jpowsour. 2022.231127.

23. Attia, P.M., Bills, A., Brosa Planella, F., Dechent, P., Dos Reis, G., Dubarry, M., Gasper, P., Gilchrist, R., Greenbank, S., Howey, D., et al. (2022). "Knees" in lithium-ion battery aging trajectories. J. Electrochem. Soc. *169*, 060517. https://doi.org/10.1149/1945-7111/ac6d13.

24. Waldmann, T., Hogg, B.-I., and Wohlfahrt-Mehrens, M. (2018). Li plating as unwanted side reaction in commercial Li-ion cells–A review. J. Power Sources *384*, 107–124. https://doi.org/10.1016/j.jpowsour.2018.02.063.

25. 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.

26. Dos Reis, G., Strange, C., Yadav, M., and Li, S. (2021). Lithium-ion battery data and where to find it. Energy and AI *5*, 100081. https://doi.org/10.1016/j.egyai. 2021.100081.

27. Bole, B., Kulkarni, C.S., and Daigle, M. (2014). Adaptation of an electrochemistry-based li-ion battery model to account for deterioration

observed under randomized use. In Annual Conference of the PHM Society. https://doi.org/10.36001/phmconf.2014.v6i1.2490.

28. Saha, B., Goebel, K., Poll, S., and Christophersen, J. (2009). Prognostics methods for battery health monitoring using a Bayesian framework. IEEE Trans. Instrum. Meas. *58*, 291–296. https://doi.org/10.1109/TIM.2008.2005965.

29. He, W., Williard, N., Osterman, M., and Pecht, M. (2011). Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method. J. Power Sources *196*, 10314–10321. https://doi.org/10.1016/j.jpowsour.2011.08.040.

30. Xing, Y., Ma, E.W., Tsui, K.-L., and Pecht, M. (2013). An ensemble model for predicting the remaining useful performance of lithium-ion batteries. Microelectron. Reliab. *53*, 811–820. https://doi.org/10.1016/j.microrel.2012.12.003.

31. 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.

32. Dechent, P., Greenbank, S., Hildenbrand, F., Jbabdi, S., Sauer, D.U., and Howey, D.A. (2021). Estimation of Li-Ion Degradation Test Sample Sizes Required to Understand Cell-to-Cell Variability. Batter. Supercaps *4*, 1821–1829. https://doi.org/10.1002/batt.202100148.

33. Pastor-Fernández, C., Uddin, K., Chouchelamane, G.H., Widanage, W.D., and Marco, J. (2017). A comparison between electrochemical impedance spectroscopy and incremental capacity-differential voltage as Li-ion diagnostic techniques to identify and quantify the effects of degradation modes within battery management systems. J. Power Sources *360*, 301–318. https://doi.org/10.1016/j. jpowsour.2017.03.042.

34. Berecibar, M., Dubarry, M., Omar, N., Villarreal, I., and Van Mierlo, J. (2016a). Degradation mechanism detection for NMC batteries based on Incremental Capacity curves. World Electric Vehicle Journal *8*, 350–361. https://doi.org/10.3390/wevj8020350.

35. Berecibar, M., Dubarry, M., Villarreal, I., Omar, N., and Van Mierlo, J. (2016b). "Degradation mechanisms detection for HP and HE NMC cells based on incremental capacity curves". In 2016 IEEE Vehicle Power and Propulsion Conference (VPPC) (IEEE), pp. 1–5. https://doi.org/10.1109/VPPC.2016.7791648.

36. Smith, K., Gasper, P., Colclasure, A.M., Shimonishi, Y., and Yoshida, S. (2021). Lithium-Ion Battery Life Model with Electrode Cracking and Early-Life Break-in Processes. J. Electrochem. Soc. *168*, 100530. https://doi.org/10.1149/1945-7111/ac2ebd.

37. 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.

38. Smith, K., and Wang, C.-Y. (2006). Solid-state diffusion limitations on pulse operation of a lithium ion cell for hybrid electric vehicles. J. Power Sources *161*, 628–639. https://doi.org/10.1016/j.jpowsour.2006.03.050.

39. Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography *36*, 27–46. https://doi.org/10.1111/j.1600-0587.2012. 07348.x.

40. Cai, J., Luo, J., Wang, S., and Yang, S. (2018). Feature selection in machine learning: A new perspective. Neurocomputing *300*, 70–79. https://doi.org/10.1016/j.neucom.2017.11.077.

41. Hastie, T.J., Tibshirani, R.J., and Friedman, J.H. (2013). *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. Eng, 2nd. (Springer).

42. Gasper, P., Collath, N., Hesse, H.C., Jossen, A., and Smith, K. (2022). Machine-Learning Assisted Identification of Accurate Battery Lifetime Models with Uncertainty. J. Electrochem. Soc. *169*, 080518. https://doi.org/10.1149/1945-7111/ac86a8.

43. Jiang, B., Gent, W.E., Mohr, F., Das, S., Berliner, M.D., Forsuelo, M., Zhao, H., Attia, P.M., Grover, A., Herring, P.K., et al. (2021). Bayesian learning for rapid prediction of lithium-ion battery-cycling protocols. Joule *5*, 3187–3203. https://doi.org/10.1016/j.joule.2021.10.010.

44. Li, T., Thelen, A., Liu, J., Tischer, C., and Hu, C. (2023) (ISU-ILCC Battery Aging Dataset). https://doi.org/10.25380/iastate.22582234.

45. Richardson, R.R., Birkl, C.R., Osborne, M.A., and Howey, D.A. (2019). Gaussian Process Regression for In Situ Capacity Estimation of Lithium-Ion Batteries. IEEE Trans. Ind. Inf. *15*, 127–138. https://doi.org/10.1109/TII.2018.2794997.

46. Kim, N., Shamim, N., Crawford, A., Viswanathan, V.V., Sivakumar, B.M., Huang, Q., Reed, D., Sprenkle, V., and Choi, D. (2022). "Comparison of Li-ion battery chemistries under grid duty cycles". en. J. Power Sources *546*, 231949. https://doi.org/10.1016/j.jpowsour.2022.231949.

47. Dubarry, M., and Anseán, D. (2022). Best practices for incremental capacity analysis. Front. Energy Res. *10*. https://doi.org/10.3389/fenrg.2022.1023555.

48. Lake, B.M., Salakhutdinov, R., and Tenenbaum, J.B. (2015). Human-level concept learning through probabilistic program induction. Science *350*, 1332–1338. https://doi.org/10.1126/science.aab3050.

49. Pedersen, E.J., Miller, D.L., Simpson, G.L., and Ross, N. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ *7*, e6876. https://doi.org/10.7717/peerj.6876.

50. Jafari, M., Brown, L.E., and Gauchia, L. (2019). Hierarchical Bayesian model for probabilistic analysis of electric vehicle battery degradation. IEEE Trans. Transp. Electrific. *5*, 1254–1267. https://doi.org/10.1109/TTE.2019.2956350.

51. Mishra, M., Martinsson, J., Rantatalo, M., and Goebel, K. (2018). Bayesian hierarchical model-based prognostics for lithium-ion batteries. Reliab. Eng. Syst. Saf. *172*, 25–35. https://doi.org/10.1016/j.ress.2017. 11.020.

52. Bhattacharya, A., Jaiswal, R., and Kumar, A. (2018). Faster algorithms for the constrained k-means problem. Theor. Comput. Syst. *62*, 93–115. https://doi.org/10.1007/s00224-017-9820-7.

53. Zhou, Z., and Howey, D.A. (2023). Bayesian hierarchical modelling for battery lifetime early prediction. IFAC-PapersOnLine *56*, 6117–6123. https://doi.org/10.1016/j.ifacol.2023.10.708.

54. Naser, M.Z., and Alavi, A.H. (Dec. 2023). Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences. Archit. Struct. Constr. *3*, 499–517. https://doi.org/10.1007/s44150-021-00015-8.