Special Collection

Supervised Learning of Synthetic Big Data for Li-Ion Battery Degradation Diagnosis

Karthik S. Mayilvahanan,[a] Kenneth J. Takeuchi,[b, c, d, e] Esther S. Takeuchi,[b, c, d, e] Amy C. Marschilok,[b, c, d, e] and Alan C. West*[a, f]

Models to understand and predict degradation can play a key role in improving the utility of Li-ion batteries. While classical mechanistic models can describe the complex physics of degradation, more recently, data-driven machine learning models have been increasingly utilized for state estimation and lifetime prediction. In this study, the physical grounding of mechanistic models is combined with the power of machine learning via the analysis of published synthetic low rate charge curves generated by a mechanistic model for different thermodynamic degradation modes. The analysis is applied to LFP, NMC, and NCA cells. A step-by-step procedure for developing interpretable machine learning models, including data set splitting, featurization, and model fitting for regression and classification tasks is outlined. Random forest regressors trained on features from incremental capacity analysis of the low rate charge curves can estimate degradation modes with a root mean squared error of 5%. Further discussion is provided on what can be learned about feature importances, and these leanings are cross-checked with expert defined features.

1. Introduction

Li-ion batteries can undergo significant degradation during use and storage, which result in capacity fade and power fade. Models to understand, diagnose, and predict degradation are necessary in management of commercialized batteries and development of new batteries. The existing models for diagnosis and prognosis of the state of Li-ion batteries include empirical, mechanistic, and physics based models. Physics based models describe physical phenomena inside the battery using coupled partial differential equations.[1,2] These models allow for description of a wide variety of degradation processes in Li-ion batteries, such as solid electrolyte interphase (SEI) formation[3–5] and particle cracking.[6] Mechanistic models do not explore the root causes of degradation, but model the evolution of the system using the net effect of the degradation processes, such as loss of active material, loss of lithium inventory, and resistance increase.[2,7] While these models encode physically relevant information, their applicability may be limited by more cumbersome evaluation and the large number of parameters that must be identified.

Empirical models include algebraic equations for capacity prediction[8] as well as more flexible data-driven machine learning (ML) models.[9] These models have been reported extensively in the literature for both state of health (SOH) estimation[10–14] and remaining useful lifetime (RUL) prediction.[15–20] While they are easy to evaluate, the lack of incorporated physics, and relatively small data sets make them limited in scope, often to the specific data set, chemistry, and cycling conditions of the data on which they were trained.

Recently, there has been interest in the community in combining the physical grounding and interpretability of mechanistic models with the adaptability and ease of evaluation of empirical ML models.[21,22] Specifically, supplementing or replacing experimental training data for ML models with synthetic training data generated by a physics based or mechanistic model has been proposed as a solution to the data limitation problem.[23]

In this study, supervised learning models are trained on low rate charge curves generated by a mechanistic model published by Dubarry et al.[24–26] C/25 charge curves are provided for three different cells (graphite-LFP, graphite-NCA, and graphite-NMC811), and full-cell voltage vs. capacity curves are provided for varying degrees of loss of lithium inventory (LLI)

[a] K. S. Mayilvahanan, Prof. A. C. West
Department of Chemical Engineering
Columbia University
New York, New York 10027, USA
E-mail: acw7@columbia.edu
[b] Prof. K. J. Takeuchi, Prof. E. S. Takeuchi, Prof. A. C. Marschilok
Department of Chemistry
State University of New York at Stony Brook
Stony Brook, New York 11794, United States
[c] Prof. K. J. Takeuchi, Prof. E. S. Takeuchi, Prof. A. C. Marschilok
Institute for Electrochemically Stored Energy
State University of New York at Stony Brook
Stony Brook, New York 11794, United States
[d] Prof. K. J. Takeuchi, Prof. E. S. Takeuchi, Prof. A. C. Marschilok
Interdisciplinary Science Department
Brookhaven National Laboratory
Upton, New York 11973, United States
[e] Prof. K. J. Takeuchi, Prof. E. S. Takeuchi, Prof. A. C. Marschilok
Department of Materials Science and Chemical Engineering
State University of New York at Stony Brook
Stony Brook, New York 11794, United States
[f] Prof. A. C. West
Department of Earth and Environmental Engineering
Columbia University
New York, New York 10027, US

Supporting information for this article is available on the WWW under https://doi.org/10.1002/batt.202100166

Special Collection An invited contribution to a Special Collection on Artificial Intelligence in Electrochemical Energy Storage

loss of active material (LAM) in either electrode with extremely high resolution, creating a big-data training set for the development of ML algorithms. The authors of this data set have published an analysis alongside a model that uses predefined predictive features from the voltage curves and a lookup table to predict the values of the degradation modes with very high accuracy (1%).[27] Others have trained ML models on similar, smaller data sets to regress values of the degradation modes as well.[28]

The goal of this study is to provide a detailed implementation and comparison of supervised machine learning models to diagnose battery degradation. We take some common transformations of the low rate charge curves that are known to contain information relevant for diagnosis, including incremental capacity and differential voltage curves, and use these curves to train machine learning models to diagnose battery degradation. We first train models to regress values of the common degradation modes and compare model accuracies to the state of the art. Results indicate that random forest regressors trained on features from incremental capacity curves perform the best, with root mean squared errors of about 5%. Next, we train classifiers to identify the limiting electrode during charge and discharge, which can be identified to 97% accuracy. In deliberately choosing interpretable models in this study, we aim to make two main contributions. First, we quantify interpretable models' performance, i.e. the ability to learn the relationship between the voltage curves and the degradation modes. Knowing these values are important when benchmarking more complex deep learning algorithms, as a simpler model should be chosen if it gives comparable performance to a more complex model. Second, we explore the ability of interpretable models to understand which features are indicators of degradation, since models that are able to perform well and provide same physical understanding at the same time are of great value to the community. We find that with large synthetic training sets, these simple, interpretable machine learning approaches can achieve high accuracy while simultaneously offering interpretability after fitting.

2. Degradation Modes and Limiting Electrodes

A helpful framework provided by Birkl et al.[29] categorizes degradation into causes, mechanisms, modes, and effects. Causes include factors such as temperature, voltage cutoffs, current, and mechanical stress that can affect the extent of degradation. Mechanisms include physical phenomena such as SEI growth, particle cracking, Li plating, etc. that are influenced in a highly coupled way by the causes. Mechanisms manifest in degradation modes, which are broad categories including loss of lithium inventory (LLI), loss of active material (LAM), and ohmic resistance increase (ORI). The degradation modes are also coupled to various combinations of degradation mechanisms. The final category, effects, provides the observable results of the degradation, namely capacity fade and power fade.

Studies understanding the relationships between causes and effects of degradation are being conducted on commercial cells.[30–32] Such data sets are well positioned for data driven studies because the training data can be generated under different cycling conditions to observe the resulting power and capacity fade, but generating large enough data sets that explore all the causes and combinations of causes remains challenging.

Mapping causes to mechanisms or mechanisms to effects is more challenging because there are several possible mechanisms for a given cell with specific active materials, electrolyte, binders, conductors, etc. under different cycling conditions. Physics-based models for different mechanisms and their relation to the causes have been proposed and used to some extent[2,33,34] but the resulting parameter space is large unless some cell-specific expert knowledge is applied for reduction.

Degradation modes represent the collective effects of the degradation mechanisms on the capacity and voltage. Since the degradation modes can fall into a few broad categories, it is a more achievable task to identify them.[7] In this study, we examine slow rate diagnostic cycles. Because of the current rate, thermodynamic degradation modes (LLI, LAM) are probed, and ORI from slower charge transfer kinetics or slower transport will not appear. Future work could include analysis of voltage curves at different current rates, which would allow for ORI diagnosis as well.

The data sets from Dubarry et al. analyzed here consider three degradation modes: loss of lithium inventory (LLI), loss of active material in the positive electrode (LAM_{PE}), and loss of active material in the negative electrode (LAM_{NE}). LAM_{PE} and LAM_{NE} can occur in the lithiated or delithiated state. In this case, LAM_{PE} and LAM_{NE} are assumed to occur in the delithiated state. Any loss of lithiated active material is counted in both LAM and LLI.

The effects of these three modes on the low rate charge curves are well documented.[7,29] Figure 1, adapted from Ref. [29] shows examples of the effects of these modes on the low rate charge curves. In Figure 1a, the case for no degradation is shown. The *x*-axis is normalized to the capacity of the positive electrode. Two characteristic parameters of the half-cell and full cell low rate charge curves are defined by Dubarry et al.:[7] the loading ratio LR and the offset OFS. LR is simply the ratio of capacities of the anode to the cathode, i.e. $LR = Q_{NE}/Q_{PE}$. OFS represents the offset, or slippage, between the cathode and the anode. When a cell is assembled, the fully lithiated cathode is aligned with the fully delithiated anode at 0% SOC_{PE}. Typically, after some formation cycles where an SEI is formed and there is loss of lithiated anode, the anode curve (red, Figure 1a) shrinks to the left such that an offset OFS between the two electrodes is formed, reducing the amount of lithium that can be cycled in the full cell.

As demonstrated in Figure 1b, LLI causes an increase in the OFS between the two electrodes. This can be simply explained by considering the fact that if the anode and cathode both do not lose sites that can intercalate lithium, the width of the PE and NE curves cannot change. Therefore, the decrease in capacity due to lost lithium inventory must result from a shift

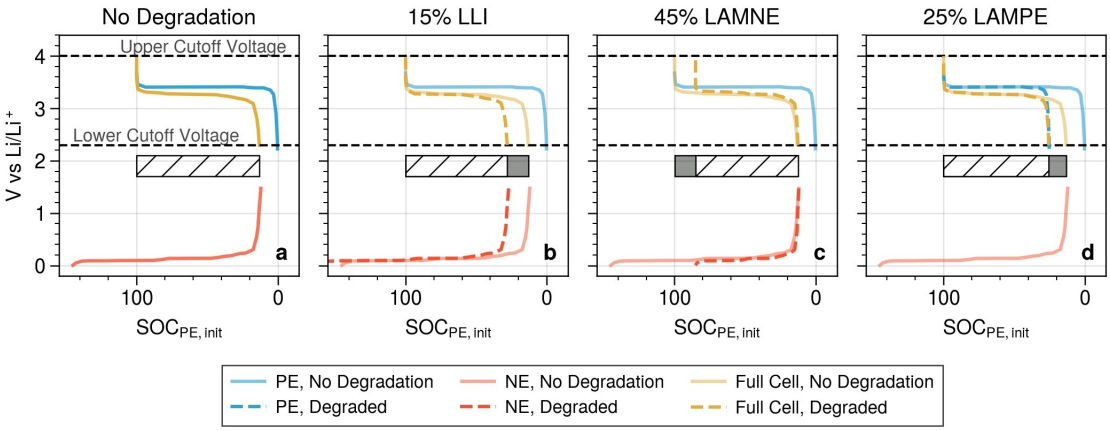

Figure 1. Exemplary low rate charge curves for an LFP cell. The half-cell voltages and resulting full cell voltage for the case of no degradation (a), 15% **LLI**, (b), 45% **LAM_{NE}**, (c), 25% **LAM_{PE}**, (d). The hashed bar in the center of each plot represents the cyclable lithium (usable capacity). The grey regions in the bars in (b–d) show the amount of capacity lost relative to the no degradation case in (a). Figure is adapted from Ref. [29] with permission. Copyright (2019) Elsevier.

in the curves relative to each other, resulting in a lower full cell capacity. LAM_{NE} (Figure 1c) results in a reduction in the width of the NE half-cell curve, whereas LAM_{PE} (Figure 1d) results in a reduction in the width of the PE half-cell curve.

Based on Figure 1a, when the cell is discharged, the sharp drop in cell voltage to reach the lower cutoff voltage at the bottom of discharge can be attributed to the sharp rise in the anode potential. Conversely, during charge, the sharp increase in cell voltage to reach the upper cutoff voltage at the top of charge can be attributed to the sharp rise in cathode potential. Thus, even from the low rate charge curves, we can determine which electrode limits the capacity on both charge and discharge. When there is sufficient LAM_{PE}, the limiting electrode for discharge switches from the anode to the cathode, as seen in Figure 1d. On charge, when there is sufficient LAM_{NE}, because the anode reversible potential is flat at high states of lithiation, the upper cutoff voltage will not be reached when the available active material in the anode is fully lithiated (Figure 1c). Instead, lithium plating will occur. While this may not immediately lower the capacity under ordinary cycling conditions, after some cycles the adverse effects of lithium plating may accelerate capacity fade.[35,36]

The onset of these phenomena can be described by the following, adapted from Ref. [7] and Ref [37].

$$OFS = OFS_{init} + LLI - LAM_{PE} \tag{1}$$

$$LAM_{NE,PT} = 1 - \frac{1 - OFS_{init} - LLI}{LR_{init}} \tag{2}$$

where LR_{init} and OFS_{init} are the initial loading ratio and offset of the electrodes after formation, before degradation, as in Figure 1a. These values are scaled to the initial SOC_{PE}, such that Figures 1b-d can be compared against Figure 1a to verify the applicability of Equations (1) and (2). $LAM_{NE,PT}$ refers to a reversible plating threshold. The limiting electrodes on charge and discharge can then be determined based on the values of OFS and $LAM_{NE,PT}$ [Eqs. (3), (4)].

$$\text{Discharge}: \begin{cases} \text{PE limiting if} & OFS \leq 0 \\ \text{NE limiting} & \text{otherwise} \end{cases} \tag{3}$$

$$\text{Charge}: \begin{cases} \text{NE limiting if} & LAM_{NE} \geq LAM_{NE,PT} \\ \text{PE limiting} & \text{otherwise} \end{cases} \tag{4}$$

Changes in the limiting electrode that occur over the course of cycling, especially the onset of lithium plating, have been linked to "knee-points" behavior in capacity retention,[35,36] so identification of these shifts could serve as valuable information during diagnosis.

In this study, we compare supervised machine learning models and voltage curve transformations for the following tasks: (1) regression of the values of LLI, LAM_{PE}, and LAM_{NE} and (2) classification of the limiting electrode on charge and discharge. The details of the supervised learning approach, including the selection of the training and testing sets, processing of the low rate charge curves, and choice of models are outlined in the following section.

3. Supervised Learning Approach

Figure 2 provides an overview of the workflow for the trained models in this study. For a given input low rate charge curve, a series of processing steps is taken to extract features. These features are passed to models for prediction of two different kinds of targets. The first target is the simultaneous quantitative prediction of the values of LAM_{PE}, LAM_{NE}, and LLI that correspond to the low rate charge curve input. The second target is the label for which electrode will limit capacity when the cell is subsequently charged and discharged at higher rates. The correct labels for each low rate charge curve are identified using Equations (1)–(4) with the values of OFS_{init} and LR_{init} provided in Ref. [27] The following subsections outline how subsets of the published data by Dubarry et al. were selected

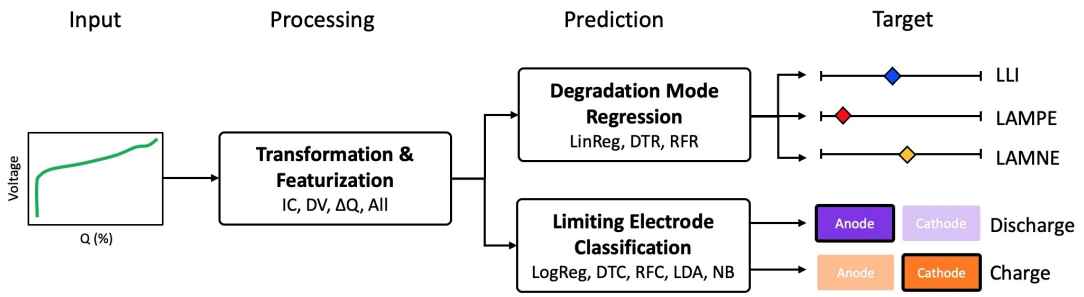

Figure 2. An overview of the two prediction tasks for trained models explored in this study. Low rate charge curves are taken as inputs, features are extracted, and predictions are made on the values of the three degradation modes and the limiting electrodes on charge and discharge.

to train these models and subsequently test them, the featurization steps, and the selection of candidate models for the regression and classification tasks.

3.1. Train and Test Set Selection

The data sets provided by Dubarry et al. include low rate charge curves generated with combinations of LLI, LAM_{NE}, and LAM_{PE} at a resolution of 0.85% between 0 and 85% for LFP, NMC, and NCA.[24–26] In total, the number of voltage curves provided exceeded 500,000 for each cell. To reduce times for training and testing, a more sparse subset of the provided data was chosen. The training set was chosen with a resolution of 5% between 0 and 80% for each degradation mode, resulting in a training set size of $N_{train} = 4{,}096$. This corresponds to 4,096 low rate charge curves and the corresponding 4,096 vectors of [LAM_{PE},LAM_{NE},LLI]. The testing set was chosen to test the interpolative ability of the models, and therefore was chosen to have a resolution of 2.5% between 0 and 80% for each degradation mode, resulting in a testing set size of $N_{test} = 32{,}768$. As exact values of the three degradation modes may not be represented in these sets, the SciPy Spatial KDTree function was used to look up the nearest neighbors for the chosen grid.[38] These samples were used to represent the training and testing sets.

Notably, a validation set was not separately chosen here. These data sets are highly structured, and the training set was designed to span the whole parameter space of interest such that the trained models would be used only for interpolation and not extrapolation. As a result, we allowed our models to learn the dataset without any regularization that may require some hyperparameter tuning. This is discussed further in the model development section.

3.2. Featurizing Low Rate Charge Curves

At the low rate charge condition, differential analysis of the voltage curve is known to contain signatures of degradation. Incremental capacity (IC) analysis takes the derivative of the capacity with respect to voltage, emphasizing the voltages at which phase equilibria exist.[39–44] Differential voltage (DV) analysis is the inverse of IC, taking the derivative of voltage with respect to capacity, featuring the transitions between phase equilibria.[41,45–47] Peak intensities, areas, locations can all be use as features to track degradation and for state estimation.[7,9,27,48]

Experimentally, IC/DV analyses may be complicated due to the presence of noise in the data.[29] In addition, some peaks may disappear over the course of aging.[28] An alternative approach that still emphasizes changes in the profile due to degradation was employed by Severson and Attia, calculating the difference in capacity ΔQ as a function of voltage for a degraded voltage profile and the voltage profile before degradation.[16,20]

All three approaches were employed to feature low rate charge curves, and are shown in Figure 3. Figure 3a shows the no degradation curve and a degraded low rate charge curve. In Figure 3b, the ΔQ curve is extracted by subtracting the two curves in Figure 3a at each voltage. Figure 3c and 3d show the IC and DV curves, respectively.

To translate these curves into a set of features that can be used for machine learning models, 20 evenly spaced points are taken from each curve. For the IC and ΔQ curves, the 20 points are chosen at evenly spaced voltages between the lower and upper cutoff voltages. For the DV curve, the 20 points are chosen at 5% SOC increments, where the SOC scale goes from 0 to 100 (the maximum capacity delivered before any degradation). One challenge with the DV curve is that when capacity is lost due to degradation, the value of the DV curve at high SOC is undefined. These features are set to 0. Some information is clearly lost in choosing these 20 points, especially in the IC curve in Figure 3c.

One key advantage to the approach described here is the automation of discovery of important features from the IC, DV, and ΔQ curves. Expert identification of features from these curves have shown highly accurate results in terms of regressing values of the three degradation modes and for estimating state of health.[27,48] However, extracting these features requires some subjectivity, such as choice of a voltage window to detect a specific peak from which an intensity is extracted. In the present study, if the voltage window is too small, significant degradation may cause the peak to move out of the chosen window. If the window is too large, it may capture multiple peaks. While some have automated peak

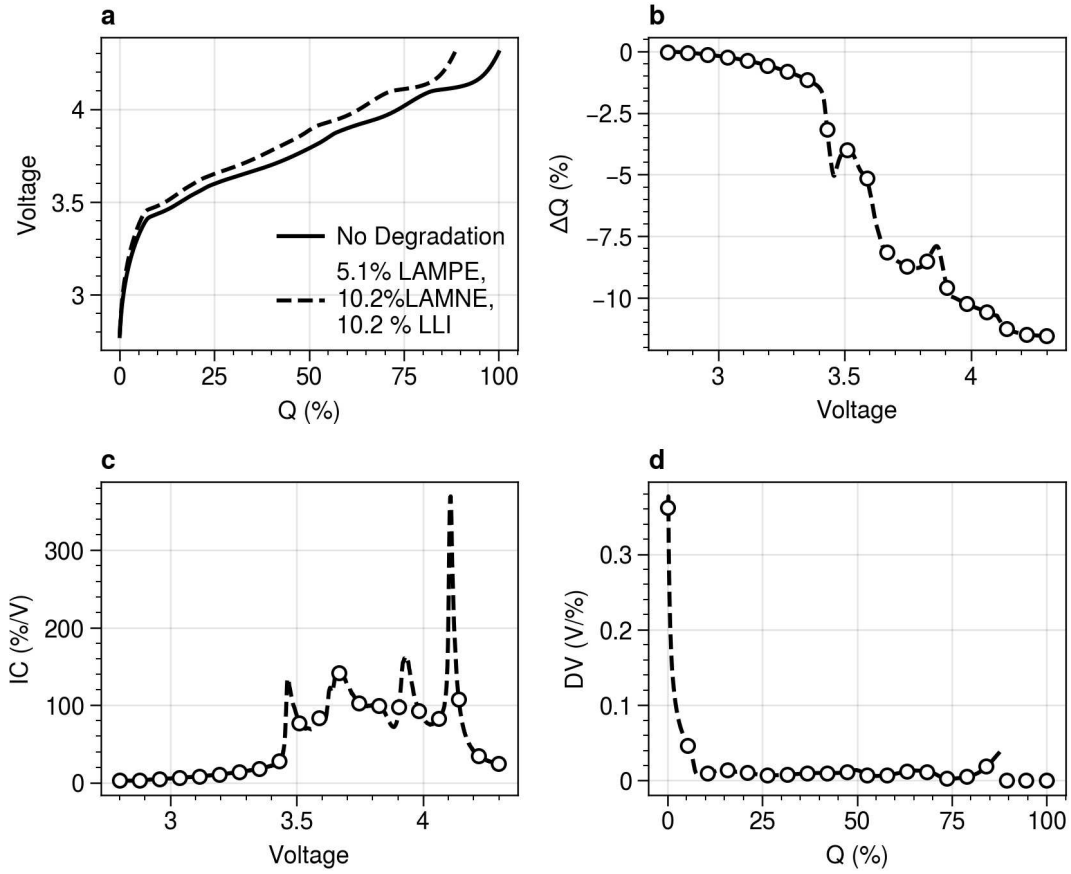

Figure 3. Low rate charge curves before and after degradation (a). The processing steps calculating $\Delta \boldsymbol{Q}$, IC, and DV are shown in b, c, and d, respectively. In c and d, the transformations shown are for the low rate charge curve after degradation. The open circles in b–d correspond to the 20 evenly spaced points chosen as features from each curve.

identification from these differential analyses,[44] allowing the most important features to be determined from the values of these three curves by fitting to the training data offers a simple alternative that removes the need for predefined featurization. An example of this is shown in Figure S1, where one peak in the IC curve after degradation disappears entirely, and significant shifts of the other peaks are observed.

Each of these three processing steps was explored individually to determine which gave the most accurate results. In addition to three different types of processing steps, a fourth was included that combines the 20 features from each curve to yield 60 features. This was done by joining the three 20×1 vectors of features into a new 60×1 feature vector. The feature normalization step was done before this concatenation. This was included since a sensitivity analysis on the Dubarry data sets revealed that both IC and DV peaks were correlated with degradation.[27] Discussion on potential overfitting is provided in the results section.

To ensure all features were on the same order of magnitude, the features were rescaled based on the distribution of the feature across all the training samples, a standard practice in feature processing for machine learning. In this case, the MinMaxScaler from the preprocessing module in scikit-learn was used.[49]

3.3. Interpretable Models

In the choice of a data-driven model for a supervised learning task, there is often a tradeoff between model performance and model interpretability.[50] Many machine learning models, such as neural networks, are able to achieve excellent accuracies in regression and classification, but lack the ability to explain the relationship between the input features and the target being predicted. To this end, we specifically consider machine learning models that are inherently interpretable, and look to establish the baseline performance of these models relative to the state of the art. As pointed out by Attia et al., who applied interpretability-focused statistical learning approaches to remaining useful lifetime prediction, these simpler models can serve as a baseline to more powerful but less interpretable techniques like deep learning.[20]

In order to regress values of the three degradation modes, we employ linear regression and two tree-based models – decision tree and random forest regressors. In linear regression, a weight is assigned to each feature such that the linear sum of the products of each feature and corresponding weight predicts the target. The weight for each of the features is readily interpreted to provide information about the degree of correlation of its feature with the target.

Tree-based methods often outperform linear models when features are highly correlated and the relationship between features and the target is nonlinear.[50] Decision trees take the approach of segmenting the feature space into regions based on threshold values, an approach that is analogous to human decision making and is therefore useful for interpretation. Feature importances can be assigned during fitting based on the improvement in predictive performance for a segmentation of the features. Random forests take subsets of the available features and subsets of the training samples, fit individual decision trees to each subset, and average the results. This process, known as bagging, reduces the high variance of an individual decision tree.[51]

Classification of the limiting electrode is a multilabel task, with two labels that have two classes each (discharge and charge, positive or negative electrode limiting). The same tree-based models that are used for regression are capable of multilabel classification as well, and are used. In addition, several other interpretable classification models are used, including logistic regression, linear discriminant analysis, and Naive Bayes.[50] The reader is referred to Ref. [49] and Ref. [51] for more detailed descriptions of these models.

We do not include regularization in these models. Regularization, or shrinkage, involves including a penalty during the fitting of a model that excludes or reduces the effect of redundant features. Regularization helps ensure that models do not overfit to the training data, especially when there are many features of the model. Regularization can also help with interpretability by identifying the most important features. However, because of the highly structured nature of the synthetic data set and the interpolation-only objective, we did not include any regularization in our models. For linear models such as linear regression, logistic regression, and LDA, L1/L2-norm penalties were not applied. For the tree-based methods, minimum depth or tree pruning were not considered.[51]

All models in this study were implemented via scikit-learn in python.[49] A full table of the models used in this study is provided in Table 1. For the random forests, the default settings in the scikit-learn implementation (e.g., 100 trees) were used. For the interpretation of model weights in linear regression and logistic regression, standard errors for the feature weights are not returned by scikit-learn. These standard errors are required for determination of feature importances, and are calculated using the linear and logistic regression implementations in statsmodels python module.[52]

Table 1. Summary of regression and classification models used. All models were implemented using scikit-learn.[49]

Algorithm	Abbreviation	Notes
Linear Regression	LinReg	Multioutput Regression
Decision Tree Regressor	DTR	Regression
Random forest Regressor	RFR	Regression
Decision Tree Classifier	DTC	Multilabel Classification
Random forest Classifier	RFC	Multilabel Classification
Logistic Regression	LogReg	Multilabel Classification via Binary Relevance
Linear Discriminant Analysis	LDA	Multilabel Classification via Binary Relevance
Naive Bayes	NB	Multilabel Classification via Binary Relevance

4. Results

4.1. Degradation Mode Regression

The results of the regression using the different models and processing conditions are shown for each degradation mode and material in Figure 4. The results are provided as root mean square errors (RMSE), in units of %. RMSE is chosen here as a more conservative metric since it penalizes larger errors more.

The model-processing combination that yields the lowest RMSE in each heatmap in Figure 4 is outlined in red. Across the board, linear regression models appear to underperform relative to the tree-based models, likely because the latter are better suited to handle the nonlinear relationship between the features and the degradation modes.[50,51] For most cases, random forests trained on IC features perform the best. These results indicate that random forests improve on the generalizability of the decision trees. This can be seen in Figure S2, which compare testing and training errors. The random forest results consistently have slightly higher training error but a lower test error. Amongst the linear models, the inclusion of DV features leads to overfitting in LFP, but for the other materials and featurization steps, the poor performance can be attributed to the use of a linear model itself. Regularization would likely not improve performance.

Other than the case described above, models trained on DV features tend to not perform as well. This is because the valuable information in DV curves include features like the distance between peaks,[27,45,53] which cannot be captured in our simple featurization methodology. As discussed in the example of incremental capacity analysis, knowledge of which peaks to choose to calculate peak differences requires some knowledge of which peaks are informative a priori, as many local maxima can be found (see Figure 1 in Ref. [27]). Our approach forgoes this process entirely by taking 20 evenly spaced features from each DV curve, and is able to achieve accuracies on the order of 10% RMSE. A more nuanced featurization approach that takes into account distance between DV peaks will achieve better accuracies at the cost of more expert knowledge. Also of note is the performance of models trained on ΔQ features, which are often second best to the IC features. While some accuracy is forfeited, the avoidance of having to calculate derivatives and deal with disappearing peaks may render this approach useful in practice.

Compared to the results from the models and features proposed by Dubarry et al.,[27,48] which are able to diagnose the degradation modes to within 1% using a lookup table, the best interpretable models used here can only diagnose degradation modes to within 2 to 7%, depending on the degradation mode and material. While the expert-defined features predetermined to be correlated with the degradation modes can be expected to be more predictive, a direct comparison cannot be made because a much larger training set was used to build the

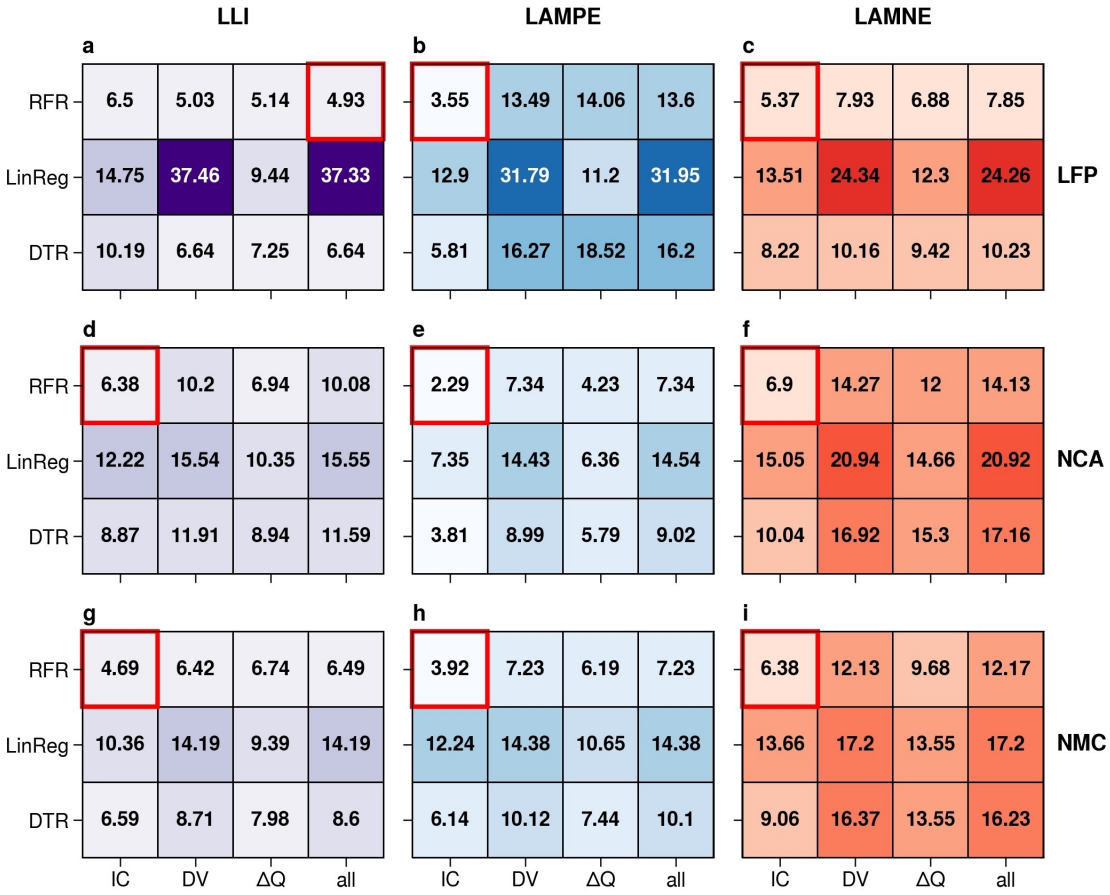

Figure 4. RMSE results, in units of %, for each degradation mode (columns) and each active material (rows). Within each heatmap, regression results are compared for each model (y-axis) and processing condition (x-axis). The model-processing combination that yields the lowest RMSE in each heatmap is outlined in red.

lookup table used for diagnosis in Ref. [27]. The accuracies achieved here without having to calculate peak areas of convoluted peaks or specify voltage or capacity ranges for identifying peaks are notable.

4.2. Limiting Electrode Classification

The results of the evaluation of the various classification models and the processing conditions for the detection of the limiting electrode on (dis)charge in the test set are shown in Figure 5. The *x*-axis shows the accuracy for predicting the limiting electrode on charge, and the *y*-axis shows the accuracy for predicting the limiting electrode for discharge.

Some key takeaways emerge from Figure 5. It appears that most of the models and processing conditions can achieve accuracies as high as 90%. Because the training set spanned such a wide space of low rate charge curves and extrapolation was not a concern on the test set, it appears that with enough training data, several models will perform well. The results for all the models are provided in Figure S3.

As with regression of the degradation modes, the best performing model-processing combination across all materials is the random forest trained on the IC features. The results for this model for each cell is shown in Table 2. In addition to the classification models considered, classification based on the regression estimates were also performed, as a baseline to test the need for classification at all. If LAM_{PE}, LAM_{NE}, and LLI can be estimated with high accuracy, Equations (1)–(4) can be used to classify the limiting electrodes. The random forest classifier trained on the IC features slightly outperforms the regression models for the limiting electrode classification task.

4.3. Interpretation of Trained Models

In this section, we look to interpret the trained models to gain some understanding of the learnings of these models.

First, we look at the weights from linear regression, as they are the easiest to understand. Since the linear regression

Table 2. Classification results for the Random Forest Classifier trained on incremental capacity features.

	Discharge	Charge
LFP	99.1%	98.6%
NCA	98.4%	97.6%
NMC	99.6%	98.3%

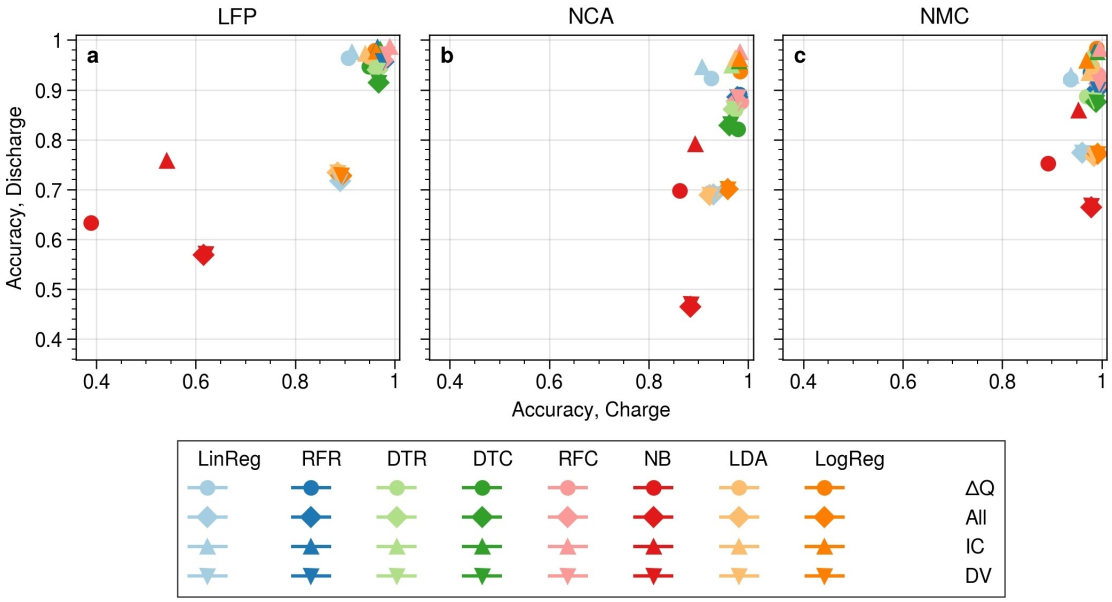

Figure 5. Limiting electrode classification accuracies for charge and discharge for LFP (a), NCA (b), and NMC (c).

models do not perform very well compared to the tree-based methods, the identified features should be taken with caution. To illustrate an example of feature importances from linear regression, we look at the linear model used to regress LAM_{PE} in NCA using IC features, which yielded an RMSE of 7.35 %.

Figure 6a shows the weights of each feature in the linear model, which indicates the correlation of the IC value at each of the 20 voltages to the value of LAM_{PE} in NCA. Each estimated weight also has an associated standard error, given by the error bars. The feature importance of a linear model can be calculated by dividing the magnitude of the weight by its standard error. Therefore, weights that are known more precisely are given more importance while weights that are not known precisely are rendered less important. The results of this transformation are shown in Figure 6b.

Also on Figure 6 are the features from the IC curve used by Dubarry et al., delineated in Ref. [27]. These features include peak locations, peak intensities, valley (local minima) locations, valley intensities, and peak areas in predefined voltage windows. Specifically, the authors find that three features are predictive of LAM_{PE} in NCA: (1) the intensity and voltage location of the IC peak between 3.0 and 3.6 V (Peak V + Peak I, pink), (2) the area underneath the IC peak between 4.02 and 4.05 V (area, green), and the area underneath the IC peak between 4.15 and 4.225 V (area, blue). These voltage windows and quantities are shown in the background of Figure 6.

The direct comparison between the feature importances identified here and by Dubarry et al. can only be made qualitatively, since these features are merely intensity values at discrete voltages. Some notable alignments are observed with the peak intensity features above 4 V. The two corresponding area features in Dubarry et al.'s approach are the two most correlated to LAM_{PE}. Instead of the peak intensity and location

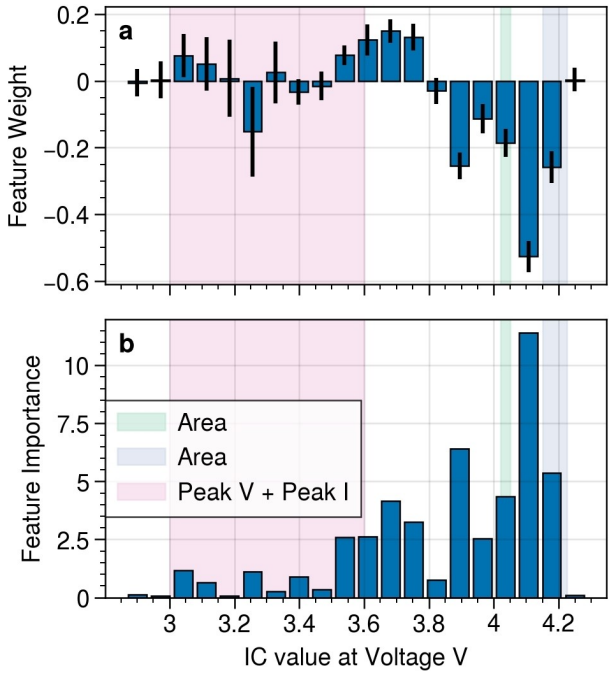

Figure 6. Coefficients and feature importances for a linear regression model trained on IC features to predict **LAM_{PE}** in NCA, which yielded an RMSE of 7.35 % on the test set. The shaded regions indicate the specified voltage window over which features were calculated from the IC curve in Ref. [27]. The legend indicates what feature was calculated in that voltage window, e.g. area under the IC curve (Area) in the voltage window, intensity of the IC peak in the voltage window (Peak I), or the voltage at which an IC peak is observed in the voltage window (Peak V).

between 3 and 3.6 V, our analysis identifies peak intensities between 3.5 and 3.8 V as predictive features for LAM_{PE}.

Next, we examine feature importances from the random forest regressors trained on IC features, which performed the best at regressing the three degradation modes. Feature

importance is determined by the total improvement in training error due to segmentation of each feature averaged over each tree in the random forest.[51]

The feature importance for the best trained model for each cell are shown in Figure 7. As with the linear regression results, the voltage windows and IC curve features in those voltage windows as defined by Dubarry et al. are shaded in the background of Figures 7a-c and labeled to the side of each subplot. Unlike linear regression, random forest regression allows for simultaneous prediction of all three degradation modes, so we cannot correlate a feature with a degradation mode. Rather, we can see which features are most important in determining the specific combination of [LLI,LAM_{PE},LAM_{NE}]. In some cases, the features determined to be important in this analysis are consistent with the expert analysis. For example, the IC features at the higher voltages are most important in our analysis, and the peak area at these voltages were chosen as features by Dubarry et al. However, at voltages where the published analyses identified IC valley intensities and locations as important features (i.e. between 3.6 and 3.98 V for NMC), our analyses did not identify important features.

Finally, we look at an example of feature importance from the classification task. We take the example of detection of the limiting electrode on charge. This task is related to the detection of reversible plating as described in Ref. [27]. The results of the logistic regression model trained on IC features to predict the limiting electrode on charge in NMC are shown in Figure 8 as an example. The logistic regression model achieved an accuracy of 94% for this classification task (Figure S3f).

As with the linear regression example in Figure 6, we have the feature weights in panel a and the importances, calculated by dividing the magnitude of the weight by its standard error, in panel b. The shading corresponds to the voltage window in which the IC peak area is expected to correlate with reversible lithium plating in Ref. [27]. The logistic regression model also identifies the IC intensity features in this voltage window as important, along with some other features in the lower voltages. Figure 6 also offers an example of an important caution when interpreting weights in linear models. A larger weight does not necessarily mean a larger feature importance. In Figure 6a we see a large negative weight followed by two smaller positive weights at higher voltages. However, the standard error on the positive weights are much smaller than that of the negative weight, so their feature importance is much more.

5. Outlook

The results presented here, while promising, can likely be improved through method optimization. A larger training set could be used, and preliminary analysis shows that the featurization of IC, DV, and ΔQ curves into more than 20 features can provide some marginal improvements as well. However, results seem to indicate that the random forests trained on IC show the most promise for both regression of the

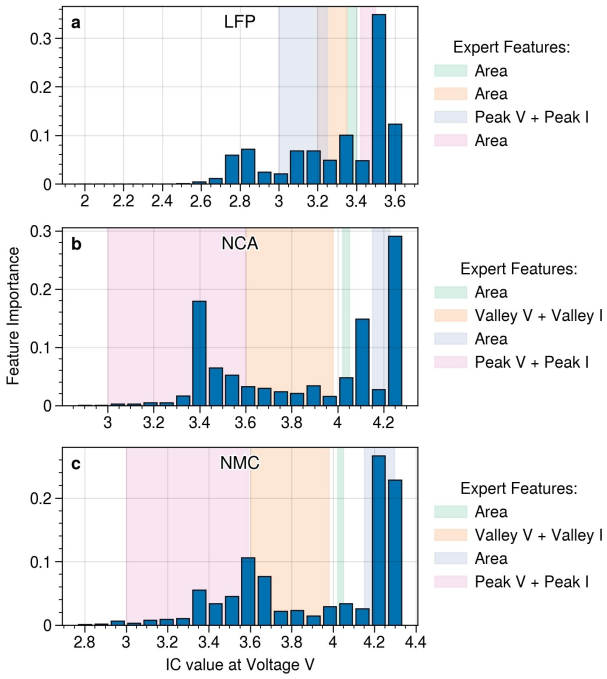

Figure 7. Feature importances for random forest regressors trained on IC curves for LFP (a), NCA (b), and NMC (c).The shaded regions indicate the specified voltage window over which features were calculated from the IC curve in Ref. [27]. The legend indicates what feature was calculated in that voltage window, e.g. area under the IC curve (Area) in the voltage window, intensity of the IC peak or minimum in the voltage window (Peak I, Valley I), or the voltage at which an IC peak or minimum is observed in the voltage window (Peak V, Valley V).

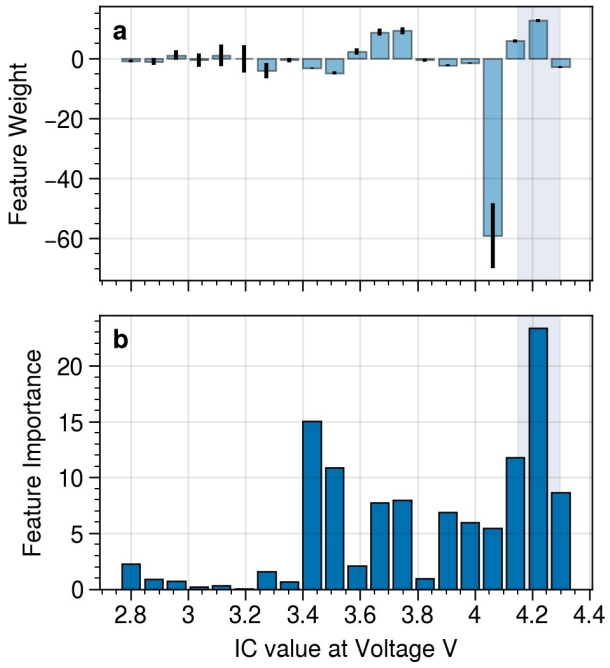

Figure 8. Weights and feature importances for a logistic regression model trained on IC features to predict the limiting electrode on charge in NMC. The shading corresponds to the voltage window in which the IC peak area is expected to correlate with reversible lithium plating in Ref. [27].

degradation modes and the limiting electrode classification problem.

The utility of the methodology presented in this study will depend on knowledge of the cell specifications, the variability in cell-to-cell behavior, and experience with certain active materials. For example, if the offset and loading ratio are known initially, fitting an OCV model that is parameterized by the three degradation modes may be sufficient.[29] If the expected voltage or capacity windows for IC or DV peaks are well known, then more nuanced features can be defined and selected to give better predictive performance than is shown here.[27,48] The value of the approach presented here is that without any knowledge of the behavior of the system other than the post-formation loading ratio and offset, decent performance can be achieved alongside some interpretability of the features that are most predictive of changes in the cell.

One limitation of this approach and other similar near OCV estimation approaches is that the results are highly dependent on the initial loading ratio and offset value being known. To make these approaches more robust from a cell-to-cell variance perspective (all cells will of course not lose the same amount of material and lithium inventory during formation, for example) some of this sample to sample variance could be incorporated into the training data by including a distribution of OFS_{init} and LR_{init} values from which to choose for a specific simulated low rate charge curve. This would then better replicate cell to cell variance, and with some regularization, important features that stand out over this variation could be identified.

Finally, only thermodynamic degradation modes are explored in the data set used for model development here, but increase of resistances in cells can also be explored using this type of analysis.[7] Expanding the design space for synthetic voltage curves to include different experimental conditions that go beyond the C/25 charge will be informative in diagnosing resistance increase in addition to loss of active material and loss of lithium inventory.

6. Conclusions

The above study provides the detailed implementation of a machine learning analysis of a synthetic big data set containing low rate charge curves generated from different combinations of thermodynamic degradation modes. Synthetic voltage curves allow for a faster exploration of the entire range of outcomes due to all the different combinations of degradation modes, a space that may be too large to explore experimentally. Training machine learning models to learn the relationships between the outputs from a physics based or mechanistic model, which mirror experimental observations, and the physical parameters used to generate that synthetic data can enable more physically relevant state estimators that provide a better understanding of the evolution of the system.

Here, we report performance of interpretable machine learning models for regression of degradation modes and classification of the limiting electrode on charge and discharge. Random forests are shown to outperform linear regression for estimation of the degradation modes. Based on the comparison of different featurization of the low rate charge curves, we find intensities of incremental capacity analysis to be informative. The performance of the models reported here can serve as benchmarks for more complex machine learning models such as neural networks. Finally, the understanding gained from the trained models is analyzed, and compared to expert-identified features, some of which are recovered in our machine learning analysis.

Acknowledgements

This research was supported by the Center for Mesoscale Transport Properties, an Energy Frontier Research Center supported by the DOE-BES, under Award #DE-SC0012673.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The code to reproduce the results of this study is available in a github repository online and can be obtained from the corresponding author upon reasonable request.

Keywords: Li-ion battery degradation · degradation modes · machine learning · classification · supervised learning

[1] J. Newman, K. E. Thomas-Alyea, *Electrochemical Systems*, 3rd ed., John Wiley & Sons, Inc., Hoboken, NJ, **2004**.
[2] J. M. Reniers, G. Mulder, D. A. Howey, *J. Electrochem. Soc.* **2019**, *166*, A3189–A3200.
[3] M. Safari, M. Morcrette, A. Teyssot, C. Delacourt, *J. Electrochem. Soc.* **2009**, *156*, A145.
[4] Y. Dai, L. Cai, R. E. White, *J. Electrochem. Soc.* **2013**, **160**, A182–A190.
[5] R. Fu, S.-Y. Choe, V. Agubra, J. Fergus, *J. Power Sources* **2015**, *278*, 506–521.
[6] R. Narayanrao, M. M. Joglekar, S. Inguva, *J. Electrochem. Soc.* **2013**, *160*, A125–A137.
[7] M. Dubarry, C. Truchot, B. Y. Liaw, *J. Power Sources* **2012**, *219*, 204–216.
[8] P. Gasper, K. Gering, E. Dufek, K. Smith, *J. Electrochem. Soc.* **2021**, *168*, 020502.
[9] Y. Li, K. Liu, A. M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, J. Van Mierlo, H. E. Hoster, *Renewable Sustainable Energy Rev.* **2019**, *113*, 109254.
[10] Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J. C.-W. Chan, P. van den Bossche, J. Van Mierlo, N. Omar, *Appl. Energy* **2018**, *232*, 197–210.
[11] R. R. Richardson, C. R. Birkl, M. A. Osborne, D. A. Howey, *IEEE Trans. Industr. Inform.* **2019**, *15*, 127–138.
[12] Y. Deng, H. Ying, J. E. H. Zhu, K. Wei, J. Chen, F. Zhang, G. Liao, *Energy* **2019**, *176*, 91–102.
[13] X. Hu, Y. Che, X. Lin, S. Onori, *IEEE Trans. Transp. Electrification* **2020**, 1–1.
[14] A. Ran, Z. Zhou, S. Chen, P. Nie, K. Qian, Z. Li, B. Li, H. Sun, F. Kang, X. Zhang, G. Wei, *Adv. Theory Simul.* **2020**, *3*, 2000109.
[15] R. R. Richardson, M. A. Osborne, D. A. Howey, *J. Power Sources* **2017**, *357*, 209–219.

[16] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, R. D. Braatz, *Nat. Energy* **2019**, *4*, 383–391.
[17] P. Fermín-Cueto, E. McTurk, M. Allerhand, E. Medina-Lopez, M. F. Anjos, J. Sylvester, G. dos Reis, *Energy and AI* **2020**, *1*, 100006.
[18] V. Sulzer, P. Mohtat, S. Lee, J. B. Siegel, A. G. Stefanopoulou, *arXiv:2010.07460 [cs, eess, stat]* (2020) http://arxiv.org/abs/2010.07460.
[19] J. Hong, D. Lee, E.-R. Jeong, Y. Yi, *Appl. Energy* **2020**, *278*, 115646.
[20] P. M. Attia, K. A. Severson, J. D. Witmer, *arXiv:2101.01885 [cond-mat, stat]* (2021) http://arxiv.org/abs/2101.01885.
[21] M. Aykol, C. B. Gopal, A. Anapolsky, P. K. Herring, B. van Vlijmen, M. D. Berliner, M. Z. Bazant, R. D. Braatz, W. C. Chueh, B. D. Storey, *J. Electrochem. Soc.* **2021**, *168*, 030525.
[22] A. Mistry, A. A. Franco, S. J. Cooper, S. A. Roberts, V. Viswanathan, *ACS Energy Lett.* **2021**, 1422–1431.
[23] M. Dubarry, D. Beck, *J. Power Sources* **2020**, *479*, 228806.
[24] M. Dubarry. Graphite//LFP Synthetic V vs. Q Dataset. 2021, https://data.mendeley.com/datasets/bs2j56p.
[25] M. Dubarry. Graphite//NCA Synthetic V vs. Q & Duty Cycle Datasets. 2021, https://data.mendeley.com/datasets/2h8cpsz.
[26] M. Dubarry. Graphite//NMC Synthetic V vs. Q & Duty Cycle Datasets. 2021, https://data.mendeley.com/datasets/pb5xpv8.
[27] M. Dubarry, D. Beck, *Energies* **2021**, *23*.
[28] S. Lee, Y. Kim, in *2020 American Control Conference (ACC)*, **2020**, p. 1137–1142.
[29] C. R. Birkl, M. R. Roberts, E. McTurk, P. G. Bruce, D. A. Howey, *J. Power Sources* **2017**, *341*, 373–386.
[30] G. Ning, B. Haran, B. N. Popov, *J. Power Sources* **2003**, *117*, 160–169.
[31] T. Waldmann, A. Iturrondobeitia, M. Kasper, N. Ghanbari, F. Aguesse, E. Bekaert, L. Daniel, S. Genies, I. J. Gordon, M. W. Löble, E. De Vito, M. Wohlfahrt-Mehrens, *J. Electrochem. Soc.* **2016**, *163*, A2149–A2164.
[32] Y. Preger, H. M. Barkholtz, A. Fresquez, D. L. Campbell, B. W. Juba, J. Romàn-Kustas, S. R. Ferreira, B. Chalamala, *J. Electrochem. Soc.* **2020**, *167*, 120532.
[33] V. Ramadesigan, K. Chen, N. A. Burns, V. Boovaragavan, R. D. Braatz, V. R. Subramanian, *J. Electrochem. Soc.* **2011**, *158*, A1048.
[34] K. S. Mayilvahanan, J. Kuang, A. H. McCarthy, L. Wang, K. J. Takeuchi, A. M. Marschilok, E. S. Takeuchi, A. C. West, *J. Electrochem. Soc.* **2021**, *168*, 050525.
[35] Y. Preger. Degradation of Commercial Li–Ion Cells Beyond 80% Capacity. Sandia National Lab, 2019, https://www.osti.gov/servlets/purl/1646309 .
[36] X.-G. Yang, Y. Leng, G. Zhang, S. Ge, C.-Y. Wang, *J. Power Sources* **2017**, *360*, 28–40.
[37] G. Baure, M. Dubarry, *Batteries* **2019**, *5*, 42.
[38] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, *Nat. Methods* **2020**, *17*, 261–272.
[39] M. Dubarry, B. Y. Liaw, *J. Power Sources* **2009**, *194*, 541–549.
[40] A. J. Smith, J. C. Burns, J. R. Dahn, *Electrochem. Solid-State Lett.* **2011**, *14*, A39.
[41] E. Sarasketa-Zabala, F. Aguesse, I. Villarreal, L. M. Rodriguez-Martinez, C. M. López, P. Kubiak, *J. Phys. Chem. C* **2015**, *119*, 896–906.
[42] M. Berecibar, M. Dubarry, N. Omar, I. Villarreal, J. Van Mierlo, *WEVJ* **2016**, *8*, 350–361.
[43] D. Anseán, V. M. García, M. González, C. Blanco-Viejo, J. C. Viera, Y. F. Pulido, L. Sánchez, *IEEE Trans. Ind. Appl.* **2019**, *55*, 2992–3002.
[44] B. Pan, D. Dong, J. Wang, J. Nie, S. Liu, Y. Cao, Y. Jiang, *Electrochim. Acta* **2020**, *362*, 137101.
[45] I. Bloom, A. N. Jansen, D. P. Abraham, J. Knuth, S. A. Jones, V. S. Battaglia, G. L. Henriksen, *J. Power Sources* **2005**, *139*, 295–303.
[46] M. Safari, C. Delacourt, *J. Electrochem. Soc.* **2011**, *158*, A1123.
[47] P. Keil, S. F. Schuster, J. Wilhelm, J. Travi, A. Hauser, R. C. Karl, A. Jossen, *J. Electrochem. Soc.* **2016**, *163*, A1872–A1880.
[48] M. Dubarry, M. Berecibar, A. Devie, D. Anseán, N. Omar, I. Villarreal, *J. Power Sources* **2017**, *360*, 59–69.
[49] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
[50] C. Molnar, *Interpretable machine learning: A guide for making black box models explainable*, **2019**.
[51] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, Springer-Verlag, New York, **2013**.
[52] S. Seabold, J. Perktold, in *9th python in science conference*, 2010.
[53] W. M. Dose, C. Xu, C. P. Grey, M. F. L. De Volder, *Cell Reports Physical Science* **2020**, *1*, 100253.

Manuscript received: July 15, 2021
Revised manuscript received: August 11, 2021
Accepted manuscript online: August 24, 2021
Version of record online: October 4, 2021