Prognostics of Lithium Ion Batteries
DS2500 Team Project Final Report
Victor Li (li.victor@northeastern.edu)
Nataly Chan (chan.natal@northeastern.edu)
Section 2: 01:35 pm - 03:15 pm at SH 335
December 9, 2025

Introduction:

Goal:

Our goal with this project was to develop prognostic models to predict the capacity and remaining useful life (RUL) of lithium ion batteries at a certain point in time given a variety of data that would help to make this prediction.

Problem Statement and Background:

We hope to analyze and model the degradation behavior of lithium-ion batteries using real NASA battery aging data. Through data analysis, visualization, and machine learning, the goal is to uncover patterns in capacity loss, temperature effects, and to predict the remaining useful life (RUL) of batteries.

Why This Matters:

Battery health prognostics and degradation analysis is critical for electric vehicles, renewable energy storage, and portable electronics, where failure prediction directly affects performance and safety. Without reliable degradation prediction, manufacturers face higher testing expenses, users experience reduced device reliability, and aging batteries contribute to greater environmental waste. By visualizing degradation trends and developing accurate, interpretable RUL and capacity predictions, this project can help reduce maintenance costs, improve energy efficiency, and enable smarter battery management systems that anticipate aging before failures occur.

Dataset:

Overview

We sourced our data from this Kaggle dataset which is a clone and cleaned version of this official NASA dataset. The dataset's experiments took place from April 2, 2008 to August 29, 2010 and therefore would not cover any recent developments in Li-ion battery technology and manufacturing. The experiments were performed on a total of 34 batteries and the data is divided into two layers, the top level metadata and the secondary level data. The metadata is a CSV file with a total of 7565 experiments logged within it, each of those experiments having a corresponding CSV with the measured data for that experiment. There are three types of experiments: charge experiments had around 1500 rows, discharge experiments had around 600 rows, and impedance experiments had around 50 rows.

Data Collection Methodology

This data set was collected from a custom built battery prognostics testbed at the NASA Ames Prognostics Center of Excellence (PCoE). Li-ion batteries were run through 3 different test types (charge, discharge and Electrochemical Impedance Spectroscopy) at different temperatures. Discharges were carried out until the battery voltage fell to predetermined voltage thresholds. Some of these thresholds were lower than that recommended by the OEM (2.7 V) in order to induce deep discharge aging effects. Repeated charge and discharge cycles result in accelerated aging of the batteries. The 34 batteries were divided into subsets of 3 to 4 batteries that were subjected to a variety of testing conditions including varying temperatures, discharge types and discharge currents. All batteries are meant to start with a capacity of ~2Ah and the experiments would continue until an end-of-life criterion were met, usually 70-80% of the original capacity, being 1.4-1.6 Ah. The testbed was comprised of: commercially available Li-ion 18650 sized rechargeable batteries, programmable 4-channel DC electronic load, programmable 4-channel DC power supply, voltmeter, ammeter, thermocouple sensor suite, custom EIS equipment, and an environmental chamber to impose various operational conditions.

Key Variables

For our purposes, we primarily examined the metadata alongside discharge and impedance tests.

Metadata Key Variables:

Variable	Description	Type	Example Value
start_time	Time when experiment began	string	[2010. 7. 21. 15. 0. 35.093]
ambient_temperature	Temperature of the environment	integer	24
type	Type of test	string	discharge
battery_id	ID of battery	string	B0005
filename	Name of corresponding csv file	string	05122.csv
Capacity	Measured capacity of battery, only for discharge tests, null otherwise	string	1.856487
Re	Estimated electrolyte resistance, only for impedance tests, null otherwise	string	0.044669
Rct	Estimated charge transfer resistance, only for impedance tests, null otherwise	string	0.069456

Discharge Key Values:

Variable	Description	Type	Example Value
Voltage_measured	Battery terminal voltage	float	4.246711253516259
Current_measured	Battery output current	float	-0.9967310600613466
Temperature_measured	Temperature of the battery	float	6.212696115916043
Current_load	Current measured at load	float	1.0
Time	Seconds since experiment began	float	36.406

Impedance Key Values:

Variable	Description	Type	Example Value
Battery_impedance	Battery impedance computed from raw data	string (complex number)	(0.19021741554080737+0.07913959666077047j)
Rectified_Impedance	Calibrated and smoothed battery impedance	string (complex number)	(0.17493022756754967-0.02331644173631698j)

Ethical Consideration

There are few ethical considerations to be had with this data because it is publicly available and collected under extremely controlled laboratory conditions. Additionally, the data does not involve any living subjects such as humans or animals, so there is no external concern for the wellbeing of subjects (in this case, batteries). The only potential source of bias is in representation bias; the dataset represents specific batteries and conditions, so the results may not necessarily generalize to all batteries and conditions.

Methods:

Data Preprocessing and Preparation

Instead of making a model read through hundreds of rows for each individual test CSV file, Victor wrote a variety of preprocessing functions that would allow for data points to be drawn from the discharge and impedance files. For example, instead of reading all the measured voltage values for a discharge cycle, the max, min, mean, voltage drop, and mean dV/dt would be found from the test. A cycle consists of one charge and one discharge test, so each discharge test can be counted to count the cycle number. The impedance tests would be merged with the discharges, but one issue we faced is that there was no consistent pattern in which the researchers would perform the impedance test, at times with multiple discharge tests passed before a single impedance test had occurred. To account for this, the data we extracted from the impedance tests would be paired with the discharge test closest to it by time. The start_time was converted from a string to datetime value for all rows in metadata because it was inconsistently formatted originally. The capacity, Re, and Rct values were stored as strings, which had to be converted to float values. However, for a select few tests, the Re and Rct were stored as complex numbers, for which only the real part was taken. Also, there were 25 discharge cycles where the capacity was null and because there were so few of them, the rows were dropped. No other null value issues were encountered.

Exploratory Data Analysis

We performed exploratory data analysis to investigate any general trends between different attributes of the batteries. Our initial correlation heatmap and r-value calculations relieved that there were moderate to high correlations between capacity (Capacity) and average temperature (mean_temperature), voltage drop, and discharge time. There was also correlation revealed between RUL and capacity, average temperature, and the rectified impedance phase mean. The correlation between capacity and RUL is intuitive because the remaining useful life is dependent on when the capacity goes below the EOL threshold. Overall though, there were no extremely strong correlations relating to RUL or capacity, suggesting that they exhibit complex behavior patterns and are not easily determinable by simply looking at a few features.

Modeling and Evaluation

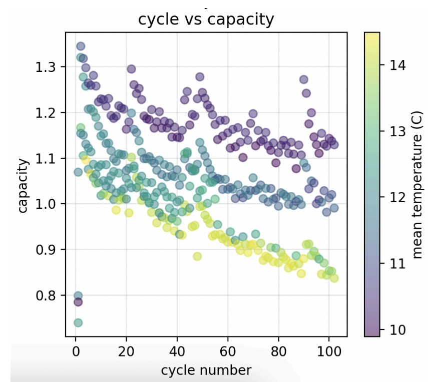
We trained a total of four models, KNN Regression, Linear Regression, Random Forest Regression, and XGBoost Regression. The models were evaluated using Mean Absolute Error, Mean Squared Error, and R^2 metrics. These are all common metrics used to evaluate regression models. MAE was included because, unlike MSE and R^2, it is less sensitive to outliers because it averages the absolute differences between predictions and actual values making it easier to interpret in real-world units. We decided to use Random Forest and XGBoost because they handle nonlinear relationships, noisy features, and complex interactions much better than linear models or KNN, making them well-suited for real-world battery degradation patterns even though they were not covered in class.

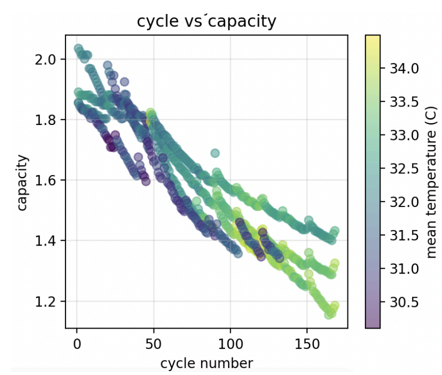
When splitting the training, testing, and validation sets, we had to split them by battery ID instead of by individual cycles because cycles from the same battery are highly correlated, and mixing them would leak information, causing the model to artificially perform well without truly generalizing to new, unseen batteries. We used the validation sets to test out what features and hyperparameters would be ideal to train the models on and then used those features on the test sets. These features were primarily decided through the correlation matrix and trial and error. At first, we had only derived very basic features from the discharge csv files such as mean_voltage, mean_current, mean_temperature, etc. After we had applied some of the physics-derived features and the impedance-derived features, all of the models' accuracies had gone up in the validation set, most noticeably with the XGBoost R^2 score increasing by a difference of 0.0339 and the Linear Regression's increasing by 0.0860.

Results and Conclusions:

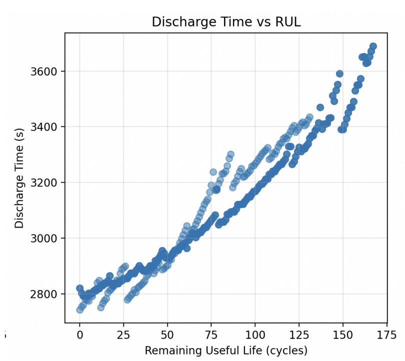
Key Findings

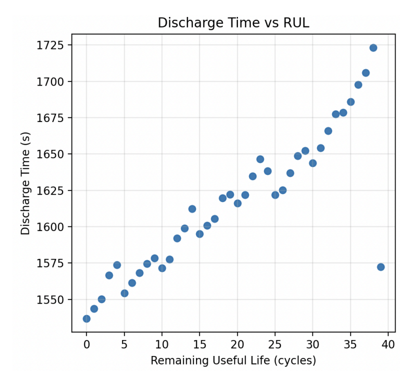
Finding 1: A higher average internal temperature causes capacity to degrade faster. The graphs show the capacity of the batteries in groups H and I over the courses of their cycles, color coded by temperature. They demonstrate that the capacity is clearly lower and degrades faster for the cycles with a higher average temperature.





Finding 2: The duration of a discharge is highly correlated with the remaining useful life of a battery, with a shorter discharge time signifying that the battery has less remaining useful life. This is intuitive because as a battery's capacity decreases, it'll contain less charge and thus take less time to discharge. The graphs below show clearly that battery groups B and I have a lower RUL with a lower discharge time.





Model Performance

For each test battery, a corresponding visualization was generated. To avoid having 18 images, only the most notable ones were selected. Below are the measured results of our models:

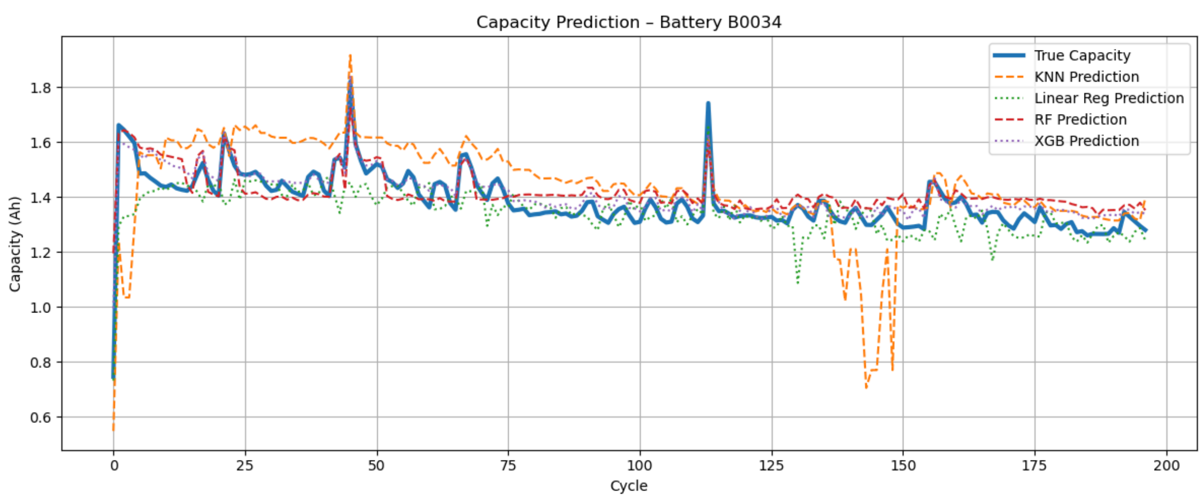
Predicting Capacity:

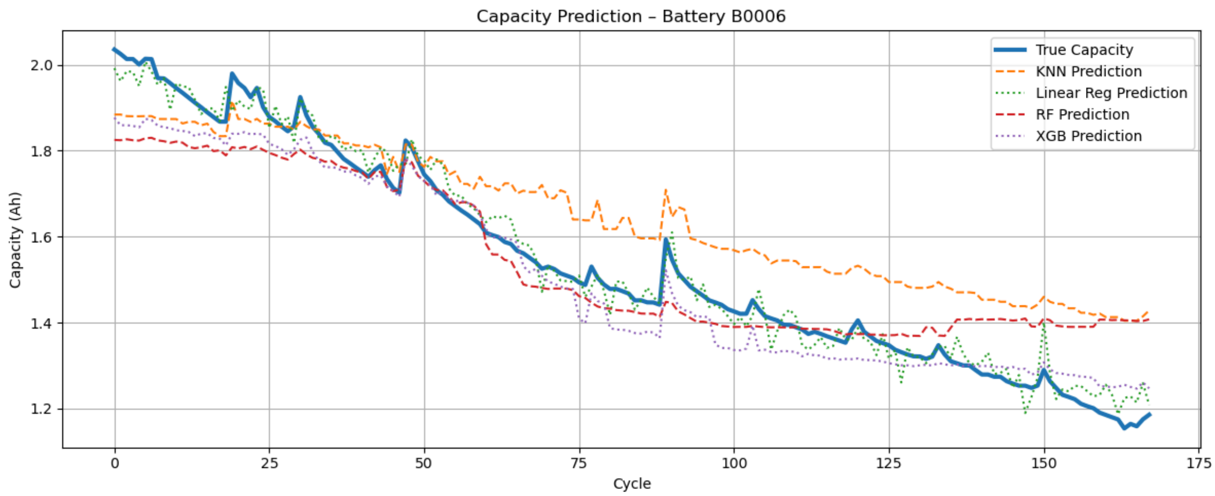
Model	MAE	MSE	R^2
Random Forest	0.08353	0.01569	0.89118
XGBoost	0.06504	0.09198	0.94134
KNN	0.11226	0.15705	0.82895

Linear	0.12019	0.16626	0.80832

For predicting capacity, XGboost clearly demonstrated the best performance, having the lowest error and highest R^2 scores. Random forest also performed rather well, ranking second overall. XGBoost and random forest also both had rectified_impedance_phase_mean and r_internal as extremely important features. These features describe how charge transfer resistance and internal ohmic resistance increase as the battery ages respectively, with the model suggesting that they are very closely tied to a battery's capacity fade. KNN and Linear performed similarly and noticeably worse, indicating that capacity does not follow a linear trend and is likely influenced by interactions between many features. Linear regression's most important features were voltage_drop, Battery_impedance_mag_mean, and mean_temperature, suggesting that these features change more linearly and gradually with aging. The results suggest that more flexible, non-linear models are better suited for capacity prediction than simpler ones.

Below are a few visualizations, in each the performance patterns stated above can easily be seen:



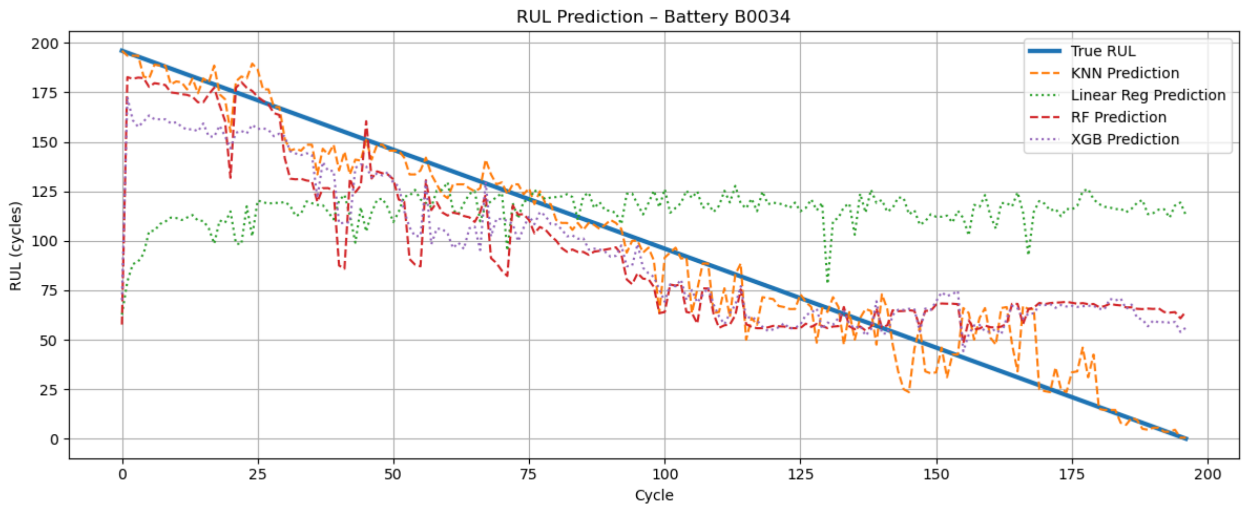


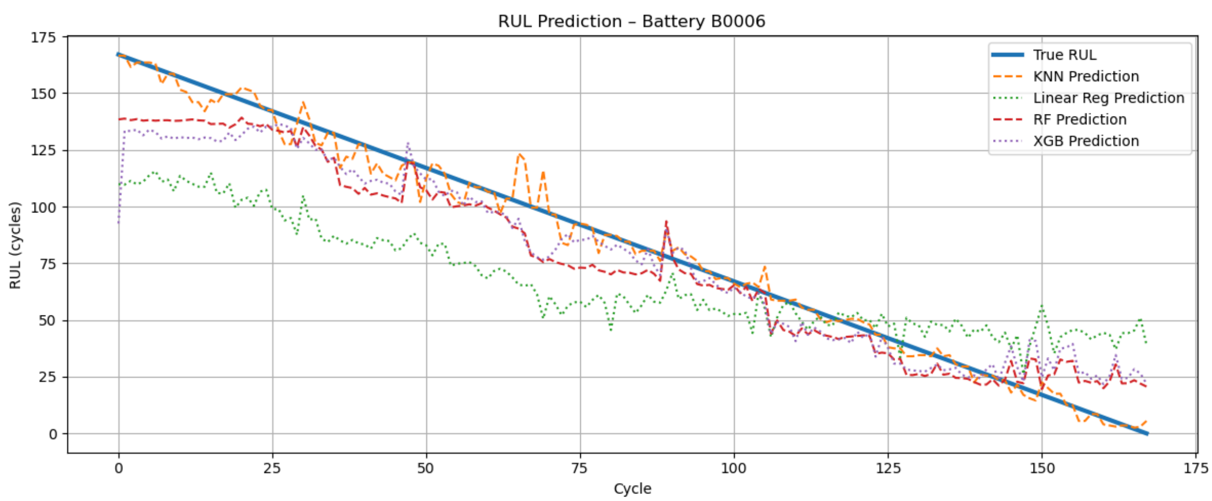
Predicting RUL:

Model	MAE	MSE	R^2
Random Forest	15.72795	443.90902	0.81934
XGBoost	15.36225	419.93619	0.82910
KNN	4.72061	78.39342	0.96810
Linear	33.95341	41.96058	0.28346

For predicting RUL, KNN performed far better than the rest, with the lowest errors and highest R^2 scores. This likely means that local similarity between cycles is a strong indicator for RUL. Random forest and XG boost performed similarly and decently, showing that these models are not as good as capturing local relationships. Random forest and XGboost both had discharge_time and mean_current as very important features, XGboost also having ambient_temperature as rather important. This is intuitive because RUL is directly related to how aggressively the battery was treated with higher discharge currents and longer discharge durations generally accelerating degradation. XGboost's assigning of importance to ambient_temperature may also indicate that environmental conditions such as temperature make a significant impact on battery degradation. The linear model performed extremely poorly, indicating that RUL is extremely non-linear.

Below are a few visualizations, in each the performance patterns stated above can easily be seen:





Conclusions

Returning to our original question in finding key factors to predict capacity and RUL: our analysis demonstrates that the average temperature is a strong indicator for capacity and discharge time for RUL. Furthermore, in our machine learning models, the phase of the rectified impedance and the internal resistance of the battery were the most important features in determining capacity while discharge time and current were the most important features for determining the RUL. We were successfully able to create models to predict RUL and capacity with great accuracy, XGboost being the best for capacity and KNN the best for RUL.

Future Work & Limitations

Limitations of Current Analysis

The dataset features a relatively small number of batteries, 34 to be exact, and among those batteries, only a few modes of operation and environmental factors. Additionally, because test profiles are relatively controlled and laboratory-based, real-world stressors such as highly variable loads, extreme temperatures, and irregular charge patterns are not fully represented. These factors combined greatly restricts the generalizability of the models. Originally, during the planning phase, we had wanted to make a purely physics based model that did not rely on machine learning to predict RUL and Capacity. This proved to be more difficult than anticipated and was eventually determined to be unfeasible due to a lack of necessary data and time. Another limitation is that the models had to be trained on simplified versions of the test csv files (the extracted features) and it would be ideal to have been able to give the models far more data to work with to improve their predictions.

An assumption made was that all batteries' last available cycle was that in which they reached their end-of-life (EOL) criterion as stated in the README files. Originally, we tried to manually test when the battery went below the capacity threshold but this proved to be unreliable as some batteries would dip below it for a few cycles before returning to a general trend and others would get very close and never reach it. Another assumption made is that, because the open circuit voltage was never measured by the researchers, when calculating the internal resistance, the open circuit voltage had to be approximated by adding a small offset (0.04V) to the initial voltage. This approximation may introduce inaccuracies in the derived internal resistance and therefore the model accuracy.

Future Research Directions

Future work could explore more physics based modeling approaches that combine machine learning with electrochemical models, allowing predictions to better reflect real battery behavior. We could also try to use time-series methods such as recurrent neural networks, temporal convolutional networks, or sequence-to-sequence models to better capture long-term nonlinear degradation patterns. It would also be good to try to extract more features or just train models on the entirety of the discharge and impedance csv files, although doing so in an efficient manner would prove difficult. Alternatively, instead of hardcoding the features we wish to extract, it may even be possible to represent each cycle using transformer or autoencoder frameworks.

References

https://www.kaggle.com/datasets/patrickfleith/nasa-battery-dataset/data
https://data.nasa.gov/dataset/li-ion-battery-aging-datasets
https://www.mdpi.com/1996-1073/12/9/1803