Analysis of Second-Life Battery Feedstock

Technical Assessment for Summer Internship

Aastik Wagadre

November 27, 2025

1 Problem Statement

In the domain of second-life battery adoption, a critical challenge is the lack of standardization in feedstock originating from multiple Original Equipment Manufacturers (OEMs). Variations in cell chemistry, form factor, and State of Health (SOH) complicate the reuse of these cells.

The objective of this report is to analyze electrical test data from three different OEMs to:

1. Identify the physical meaning of the unknown "Feature 10".
2. Perform a comparative analysis of cell degradation across OEMs.
3. Develop predictive models for State of Health (SOH) estimation.

2 Dataset Overview

The dataset comprises electrical test data organized into three folders, representing three distinct OEMs. Each folder contains '.txt' files representing individual cells. The test protocols include 5-step and 7-step procedures capturing charge-discharge behaviors.

3 Objective 1: Identification of Feature 10

The dataset included an unidentified column labeled "Feature 10". To determine its physical significance, a correlation analysis was performed against known electrical parameters (Voltage, Current, Discharge Capacity, and Power).

3.1 Analysis

A Pearson correlation matrix was generated using the combined dataset.

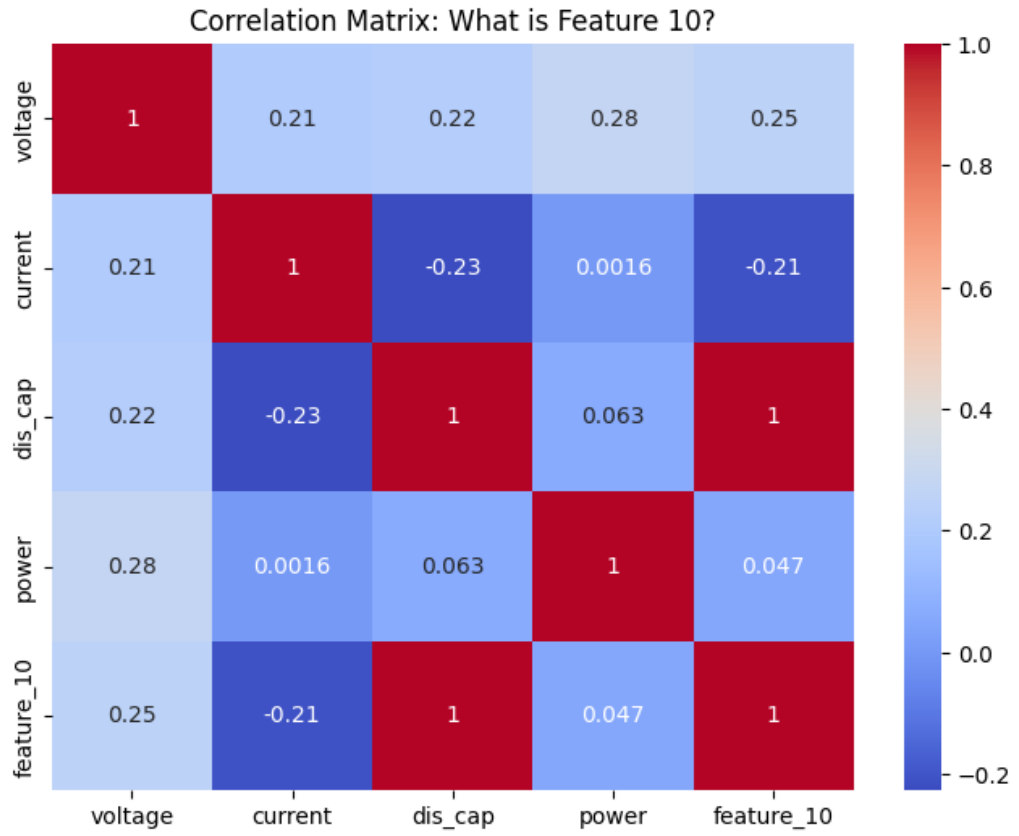

Figure 1: Correlation Matrix showing dependencies between Feature 10 and other parameters.

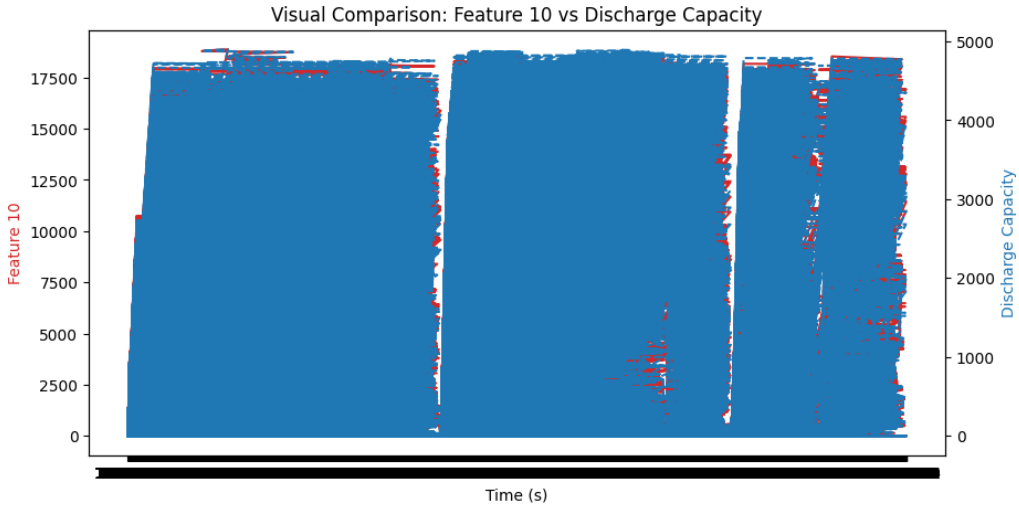

Figure 2: Visual Comparison: Feature 10 v/s Discharge Capacity

3.2 Conclusion

As observed in Figure 1, **Feature 10 exhibits a correlation coefficient of 1.0 with Discharge Capacity (‘dis_cap‘)**.

Justification: The perfect linear correlation indicates that Feature 10 is not a distinct derivative metric but rather a direct record of the cumulative discharge capacity (Ah) for the cell. Visual inspection of the time-series data confirms that Feature 10 accumulates identically to the discharge capacity column.

4 Objective 2: Comparative Analysis

An inter-OEM comparison was conducted to analyze the variation in battery health across different manufacturers. The maximum discharge capacity was extracted for each cell to serve as a proxy for the State of Health (SOH).

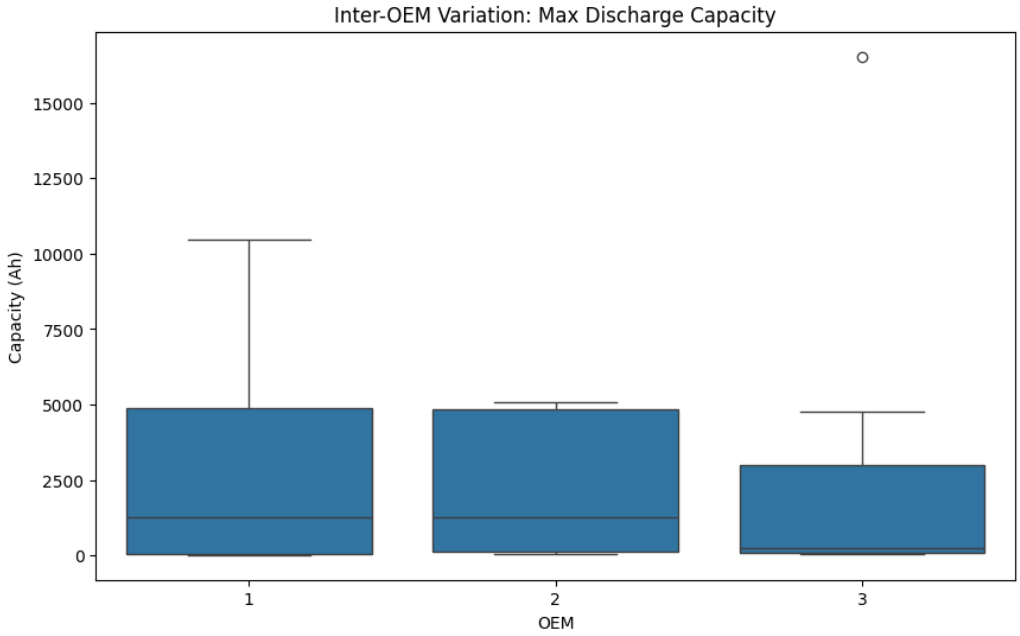

Figure 3: Inter-OEM Variation in Max Discharge Capacity.

4.1 Observations

- **OEM 1 and OEM 2:** These manufacturers show a wide inter-quartile range (taller boxes), indicating significant inconsistency in cell quality. The feedstock contains both healthy cells and highly degraded cells.
- **OEM 3:** This manufacturer demonstrates a significantly tighter distribution (shorter box). While the absolute capacity might differ, the cells are highly consistent, suggesting better manufacturing tolerance or a more uniform previous usage history.

5 Objective 3: SOH Estimation Model

A Random Forest Regressor was trained to estimate the State of Health (SOH) based on extracted features (Max Voltage, Total Time, Average Current). The model was evaluated using 5-fold cross-validation.

5.1 Model Performance Metrics

The following table summarizes the error metrics for each OEM model:

Metric	OEM 1	OEM 2	OEM 3
RMSE (Root Mean Square Error)	0.3575	0.3842	**0.2137**
MAE (Mean Absolute Error)	0.3575	0.3842	0.1822
MAPE (Mean Abs. % Error)	41.67%	**11.24%**	26.57%

Table 1: Cross-Validation Results for SOH Estimation.

5.2 Findings

- **Best Fit:** The model for **OEM 3** achieved the lowest RMSE (0.2137), likely due to the lower variance in the input data (as seen in the box plots).
- **Error Analysis:** OEM 1 showed the highest relative error (MAPE $\approx 41\%$), suggesting that simple voltage/time features may be insufficient for this specific chemistry, or that the degradation patterns are more complex (non-linear) compared to OEM 2 and 3.

6 Conclusion

The analysis successfully identified Feature 10 as Discharge Capacity. The comparative study revealed that OEM 3 offers the most consistent feedstock, which is preferable for second-life grading. Finally, the Random Forest model established a baseline for SOH estimation, with OEM 3 yielding the most reliable predictions.