

Battery Second-Life Analysis and SOH/SOP Estimation

Lohum Internship Assignment

Kamakshi Garg

November 27, 2025

Agenda

- 1 Dataset Overview
- 2 Feature-10 Identification
- 3 OEM Comparison
- 4 SOH/SOP Modeling
- 5 Conclusion

Dataset Summary

- Cell test logs from **3 OEMs**, identified by filename prefixes:
 - OEM1: 27_*.txt
 - OEM2: 21_*.txt
 - OEM3: 10_*.txt and 11_*.txt
- Each file contains time-series measurements for up to **256 cells**
- Common 13-column structure across all OEMs
- Key variables: voltage, current, capacity, power, and an unknown Feature 10

Column Mapping

Column	Meaning
V_mV	Voltage [mV]
I_mA	Current [mA]
Q_mAh	Capacity [mAh]
P_W	Power [W] = V × I
E_mWh	Feature 10 (unknown)
t_min	Time [min]
step	Test step index

What is Feature 10?

Feature 10 increases:

- only during discharge
- proportionally to power \times time
- remains constant when power = 0

We compute:

$$E_i^{calc} = \frac{P_i + P_{i+1}}{2} \cdot \frac{\Delta t_i}{60} \times 1000$$

Result: Feature-10 exactly matches computed energy increments.

→ **Feature 10 = cumulative discharge energy (mWh)**

Discharge Step Identification

- OEM1: discharge step = **3**
- OEM2: discharge step = **3**
- OEM3: discharge step = **5**

All OEMs follow the same general structure with different step indices.

OEM Capacity Comparison

Conclusions from Q_max (capacity):

- **OEM3** — Highest capacity, tight distribution, few degraded cells
- **OEM1** — Medium quality, moderate degradation
- **OEM2** — Worst performance, many near-zero or failed cells

OEM3 is the strongest candidate for second-life battery packs.

Targets

SOH Definition:

$$SOH = \frac{Q_{\max}}{Q_{ref}}$$

SOP Definition:

$$SOP = \frac{E_{\max}}{E_{ref}}$$

where Q_{ref}, E_{ref} are the 95th percentiles.

Targets are normalized [0, 1].

Input Features

Simple discharge summary features:

- Discharge duration
- Initial voltage
- Mean voltage
- Mean current
- Mean power

→ **No full time-series needed**

Model and Validation

- Model: **Random Forest Regressor**
- Validation: **5-fold cross-validation**
- Dataset used: **OEM3 cells**

Results

Target	RMSE	MAE	MAPE (%)
SOH	0.00547	0.00403	0.42
SOP	0.00871	0.00641	0.66

SOH error = 0.4% SOP error = 0.7%

These errors are far below typical industry thresholds (5%).

Conclusion

- Feature-10 is **cumulative discharge energy (mWh)**
- OEM3 has the healthiest and most consistent cells
- Simple discharge features are enough for accurate prediction
- Random Forest achieves:
 - SOH MAPE = **0.42%**
 - SOP MAPE = **0.66%**
- Approach is scalable for second-life battery screening

Thank You!