Supplementary Information

for

High C-rate Li-NMC/Graphite Pouch Cell End-of-Life Prediction via Cycle-Dependent Variations and Machine Learning

Jung-goo Choi[a, b], Jethro Daniel Pascasio[a, b], Jaeyoung Lee[a, b, c]*

[a] *School of Earth Sciences and Environmental Engineering, Gwangju Institute of Science and Technology (GIST), 123 Cheomdangwagi-Ro, Buk-gu, Gwangju 61005, Republic of Korea*

[b] *International Future Research Center of Chemical Energy Storage and Conversion Processes (ifRC-CHESS), GIST, 123 Cheomdangwagi-Ro, Buk-gu, Gwangju 61005, Republic of Korea*

[c] *Ertl Center for Electrochemistry and Catalysis, GIST, 123 Cheomdangwagi-Ro, Buk-gu, Gwangju 61005, Republic of Korea*

jaeyoung@gist.ac.kr (J Lee)*

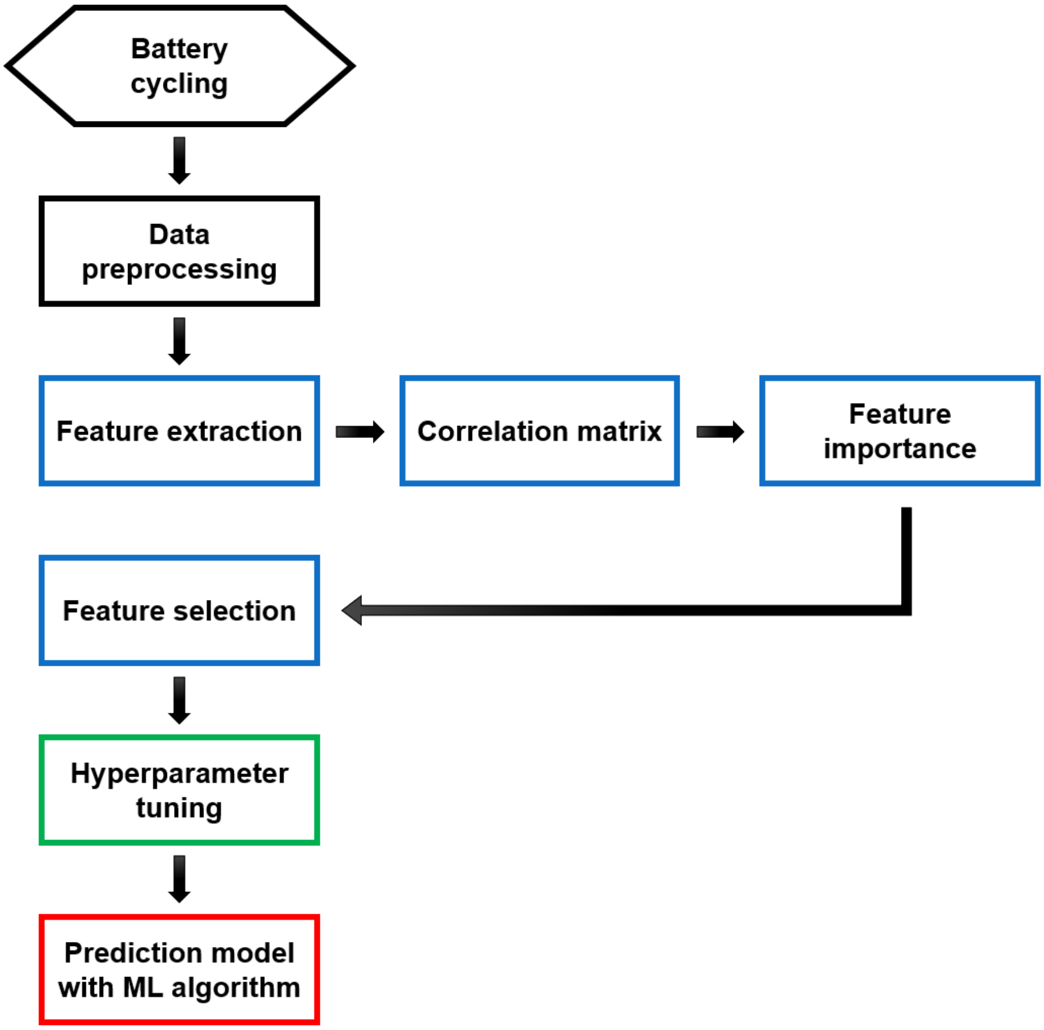

Figure S1. Flowchart illustrating the process of ML-based battery degradation modeling, including feature engineering, selection, and model optimization.

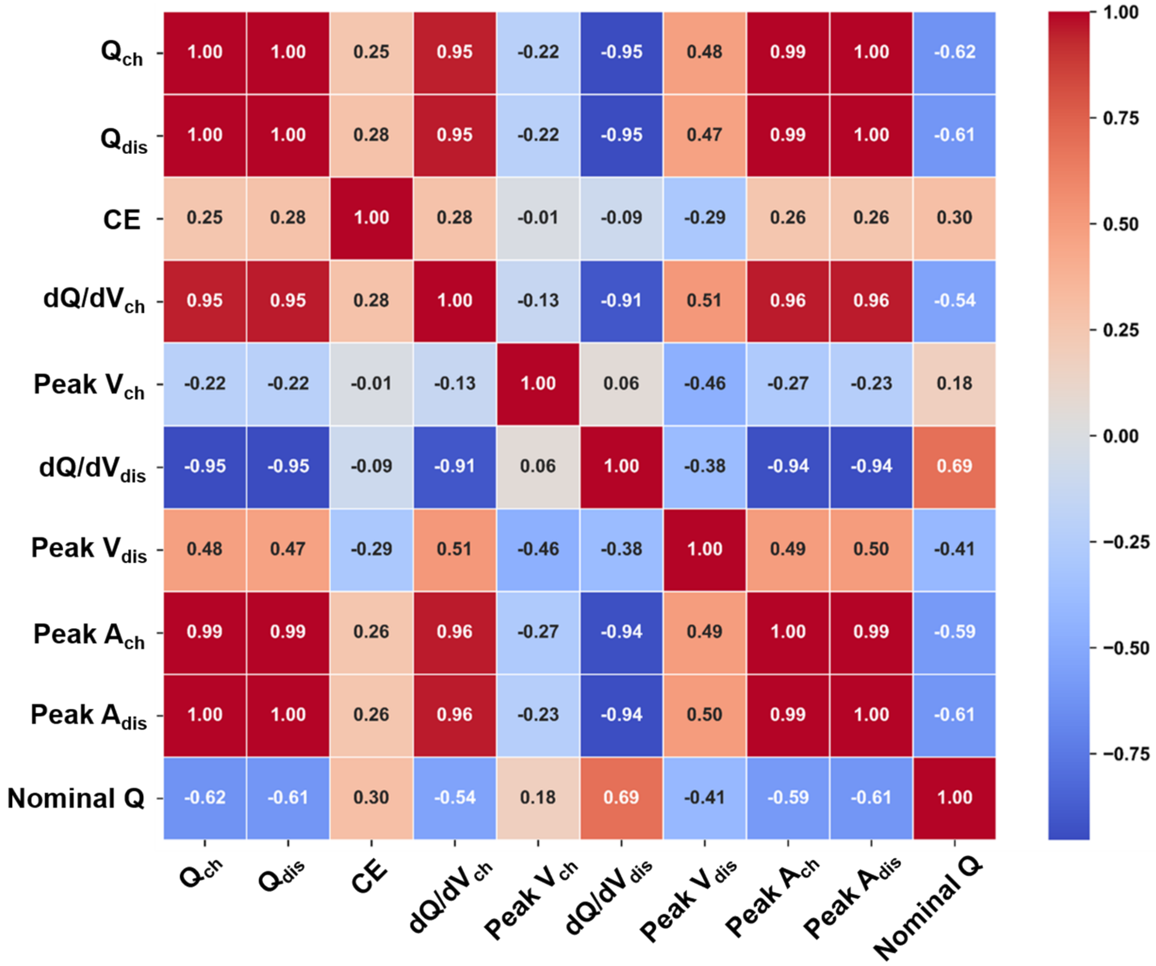

Figure S2. Correlation matrix of the initially selected features for end-of-life prediction. The heatmap visualizes pairwise Pearson correlation coefficients between features such as charge capacity (Q_{ch}), discharge capacity (Q_{dis}), coulombic efficiency (CE), and other cycle-dependent metrics. Here, Q denotes capacity, with the subscripts ch and dis indicating charge and discharge, respectively. Peak V and Peak A refer to the voltage and area under the curve at the dQ/dV peak. Darker red and blue colors indicate stronger positive and negative correlations, respectively. Strong correlations are observed between features from the charge and discharge processes.

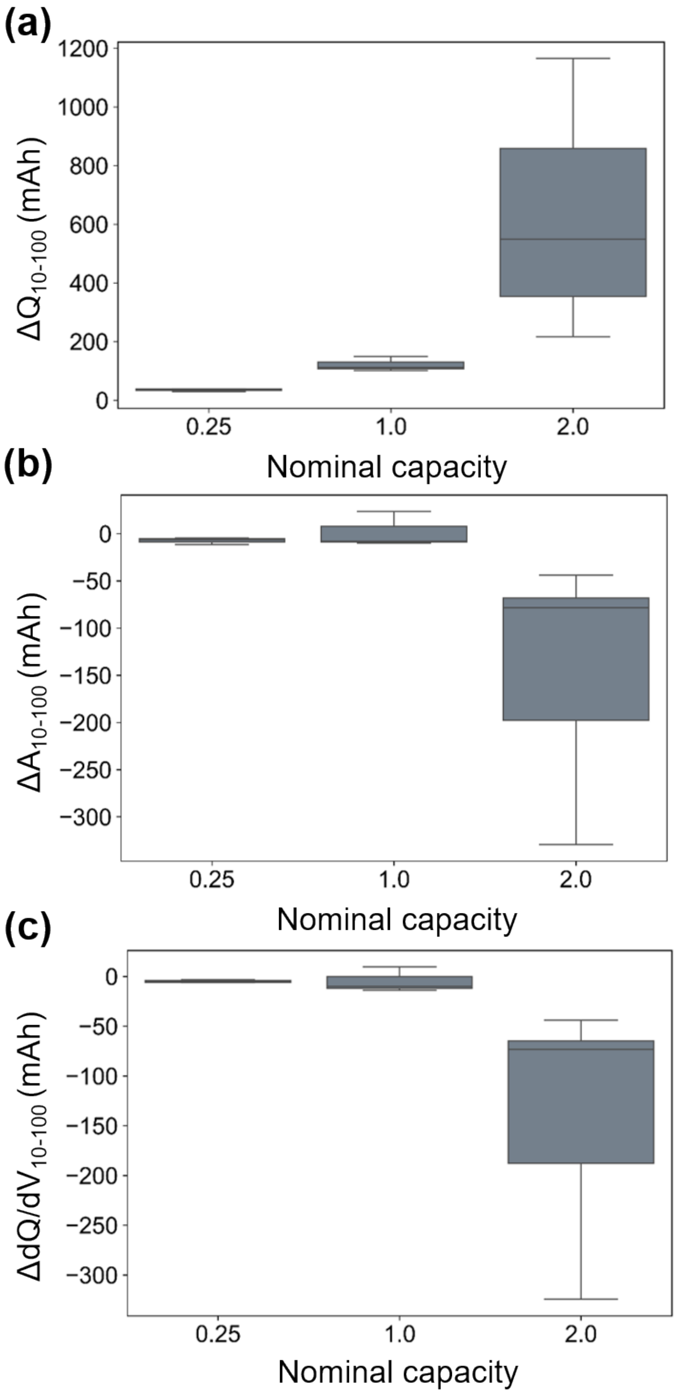

Figure S3. Boxplots of early-cycle degradation indicators grouped by rated cell capacity (0.25 Ah, 1.0 Ah, 2.0 Ah). (a) Difference in discharge capacity between cycle 100 and 10. (b) Difference in dQ/dV minimum value between cycle 100 and 10. (c) Difference in dQ/dV negative peak area between cycle 100 and 10. All three features indicate significantly larger degradation in 2.0 Ah cells, suggesting stronger structural or kinetic changes during early-stage cycling.

Model	Parameter	Optimal Value
Gradient Boosting	learning_rate	0.1
	max_depth	3
	n_estimators	100
Random Forest	n_estimators	100
	max_depth	None
CatBoost	learning_rate	0.03
	iterations	700
	depth	8

Table S1. Optimized hyperparameters for the machine learning models used in this study. Parameters were tuned using GridSearchCV with 5-fold cross-validation for each model: Gradient Boosting, Random Forest, and CatBoost.