Exploratory Analysis of Strain-Derived Features for Machine-Learning Estimation of State of Health in 18650 Lithium-Ion Cells

Malichi Flemming*, George Anthony†, Ryan L. Limbaugh‡, and Austin R.J. Downey§

Accurate monitoring of lithium-ion battery state of health (SOH) is essential for ensuring safety and long-term performance. This study explores the use of strain sensing as a non-invasive diagnostic tool for SOH estimation in 18650 lithium-ion cells. Strain data were collected using external strain gauges during standard charge-discharge cycling, revealing mechanical deformation patterns that evolve with cell aging. To extract meaningful indicators from the strain response, we introduce a decomposition method that separates cyclic strain signals into three components: offset, slope, and residual. The offset captures irreversible deformation and structural degradation, the slope reflects reversible strain correlated with the state of charge, and the residual component captures nonlinear behaviors associated with lithium intercalation dynamics. Time-domain features were extracted from the residual strain signals to quantify changes in mechanical response. Features used include mean strain, maximum strain, root mean square strain, strain skewness, strain kurtosis, and maximum temperature. These features were found to correlate with capacity fade and serve as effective proxies for degradation. A machine learning framework was developed to predict battery SOH using a random forest model trained on the extracted strain-based features. This approach enables interpretable and scalable predictions by identifying the most informative features and mitigating overfitting. Results demonstrate the feasibility of strain-informed feature analysis coupled with data-driven modeling for real-time SOH monitoring in lithium-ion batteries. This study provides foundational evidence for integrating mechanical strain sensing into battery management systems to enhance reliability and safety in energy storage technologies.

Nomenclature

SOH	=	State of Health
SEI	=	Solid Electrolyte Interphase
CCCV	=	Constant Current Constant Voltage
SOC	=	State of Charge
RMS	=	Root Mean Square
RMSE	=	Root Mean Square Error
MAE	=	Mean Absolute Error

I. Introduction

Lithium-ion batteries play a crucial role in modern energy storage, powering electric vehicles, smart grids, and consumer electronics. As their prevalence grows, ensuring their safety and longevity becomes increasingly important. One key aspect of battery health management is monitoring the state of health (SOH), particularly under abnormal conditions like overcharging, which can lead to catastrophic failures.

*Graduate Research Assistant, Department of Mechanical Engineering, University of South Carolina, Columbia, SC 29208, USA.
†Graduate Research Assistant, Department of Mechanical Engineering, University of South Carolina, Columbia, SC 29208, USA.
‡Undergraduate Research Assistant, Department of Mechanical Engineering, University of South Carolina, Columbia, SC 29208, USA.
§Associate Professor, Department of Mechanical Engineering and Department of Chemical Engineering, University of South Carolina, Columbia, SC 29208, USA. Contact Author: austindowney@sc.edu

Traditional SOH monitoring methods typically rely on electrical parameters, including voltage, current, and impedance. However, these approaches have limitations in identifying the mechanical changes caused by internal electrochemical reactions. Mechanical strain measurement has emerged as a valuable diagnostic tool for detecting such abnormalities, as it provides insights into the physical transformations occurring within the cell. Overcharging or misuse of lithium-ion cells can result in significant internal pressure variations, which can be detected through strain measurements. Kirchev et al. [1] found that similar 18650 cells exhibited swelling almost immediately after an overcharge event, while Anthony et al. [2] demonstrated that strain data could be used to identify the high current that leads to the failure of the current interrupt device (CID) in 18650 cells.

Lithium-ion intercalation between the anode and cathode is known to induce expansion, particularly in the graphite anode of 18650 cells. As lithium ions migrate from the cathode to the anode during charging, they embed between graphene layers, causing swelling perpendicular to the graphene planes as seen in figure 1 [3]. While this expansion is reversible, other forms of cell expansion are not; such as those caused by solid electrolyte interphase (SEI) formation. The SEI layer, which forms early in a battery's life, captures lithium ions while preventing further electrolyte decomposition. Over time, SEI growth diminishes. Weng et al. [4] modeled both reversible and irreversible expansion processes, highlighting their impact on battery performance.

Once a reliable method for monitoring SOH is established, the logical next step is the implementation of machine learning models that can predict battery SOH. SOH profiles formed from collected charge-discharge cycling data can be employed to build and train these models.

The contributions of this paper are twofold. First, this study explores the feasibility and accuracy of using strain measurements to evaluate the state of health (SOH) of an 18650 Samsung 30Q NMC cell, presenting a method for decomposing strain measurements into specific features, which are subsequently analyzed to identify correlations between mechanical strain, SOH, and cell capacity. Second, this paper presents a machine learning model trained from the preliminary SOH profile to estimate SOH.

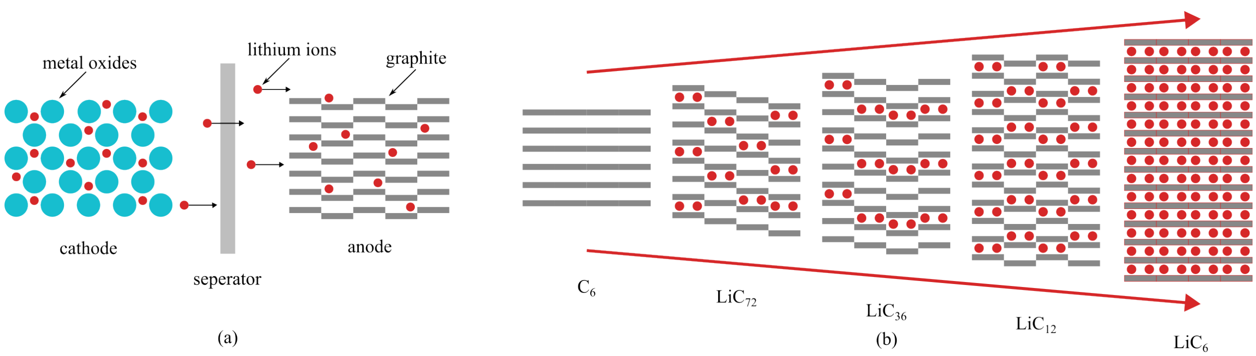

Fig. 1 Discharge ion flow, illustrating: (a) lithium-ion migration from cathode to anode and (b) lithium-ion intercalation into graphene planes

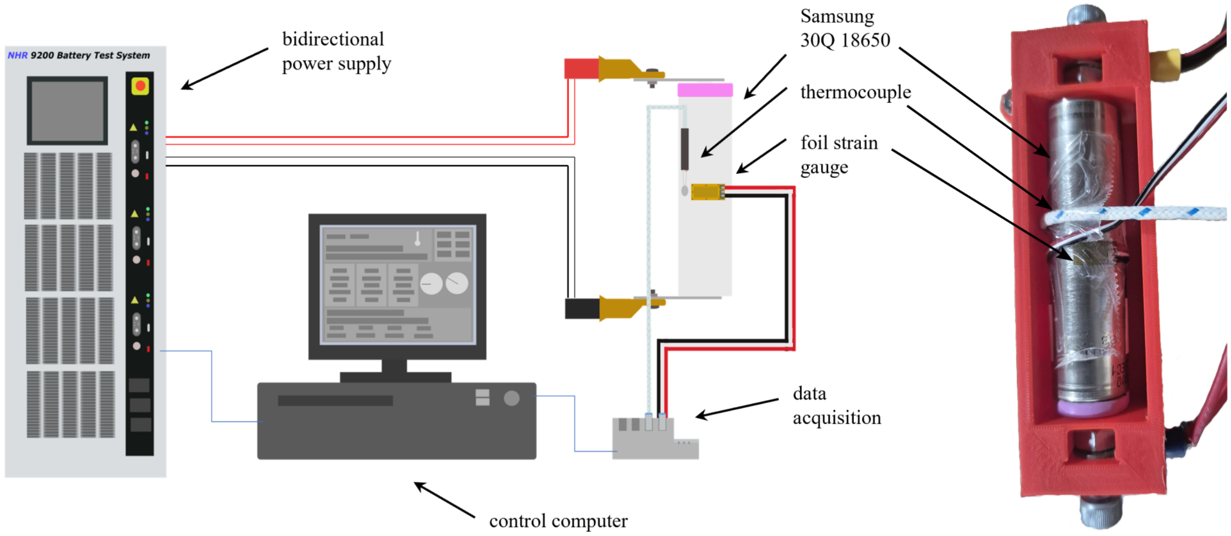

Fig. 2 Conceptual testing setup configuration.

II. Methodology

This section outlines the experimental procedures and data processing techniques employed in this study, including strain data collection, decomposition, and subsequent feature extraction utilized for machine learning modeling.

A. Data Collection

A single 18650 lithium-ion cell (Samsung 30Q), rated at 3000 mAh, was utilized for experimental cycling tests. The cell underwent repeated cycling characterized by a constant current-constant voltage (CCCV) charge at 1C, followed by a constant current (CC) discharge at 1C. Each charge-discharge cycle was separated by a one-hour rest period to ensure thermal equilibrium. Throughout testing, external strain gauges were surface-mounted to measure hoop strain during discharge phases, as illustrated in figure 2. Concurrently, current, voltage, and temperature were recorded continuously until the cell reached an 80% capacity threshold, marking its effective end-of-life. A single CC discharge test is shown in figure 3.

Capacity at each cycle was quantified using coulomb counting, integrating the discharge current over time. The calculated nominal state of health (SOH) is expressed as a percentage of the cell's measured capacity relative to its original rated capacity.

B. Strain Decomposition

To accurately interpret mechanical strain responses, each discharge cycle's strain data was decomposed into three distinct components: offset, slope, and residual, as shown in figure 4. The offset represents the initial strain at the start of discharge and is subtracted from all subsequent data points. Next, a linear slope connecting the initial and final strain values is computed and removed from the offset-corrected strain data. The remaining nonlinear strain response, termed the residual strain, encapsulates the complex mechanical behavior associated with lithium intercalation and electrode structural changes. Figure 5 illustrates the temporal evolution of the offset (part a), slope (part b), and residual strain (part c) for several discharge cycles, highlighting the distinct mechanical behaviors captured by each strain component.

C. Feature Extraction

From the residual strain data, meaningful quantitative features were extracted and normalized using both time-domain (root mean square, skewness, kurtosis, crest factor, and impulse factor) and frequency-domain (RMS frequency, frequency center) analyses. These features, shown in Figure 6, were selected to capture various aspects of the strain response indicative of mechanical and electrochemical degradation processes occurring within the cell. To further evaluate the relevance of each feature, a Random Forest feature importance analysis was conducted. The resulting importance scores

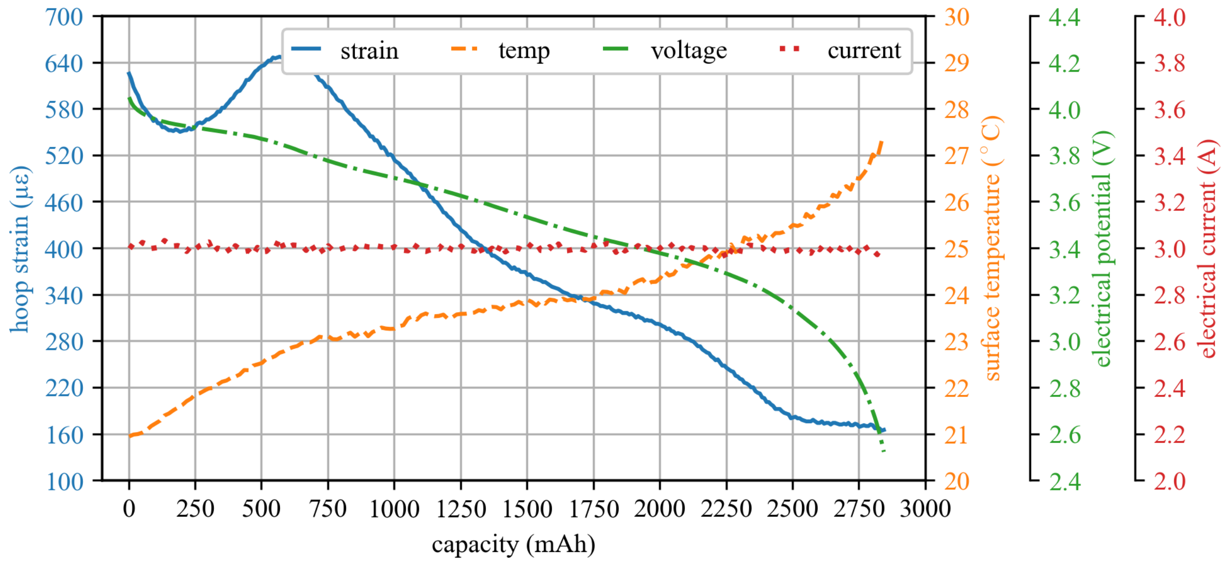

Fig. 3 Discharge profile of single cell with strain, temperature, voltage, and current measurements.

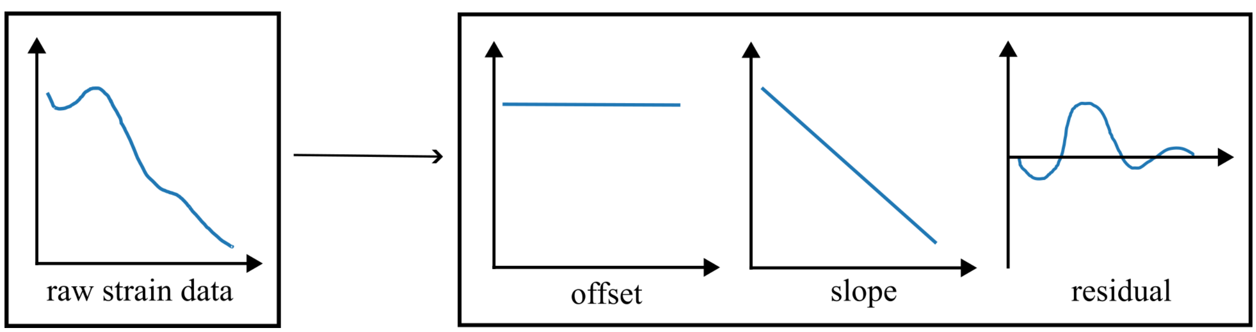

Fig. 4 Strain decomposition methodology.

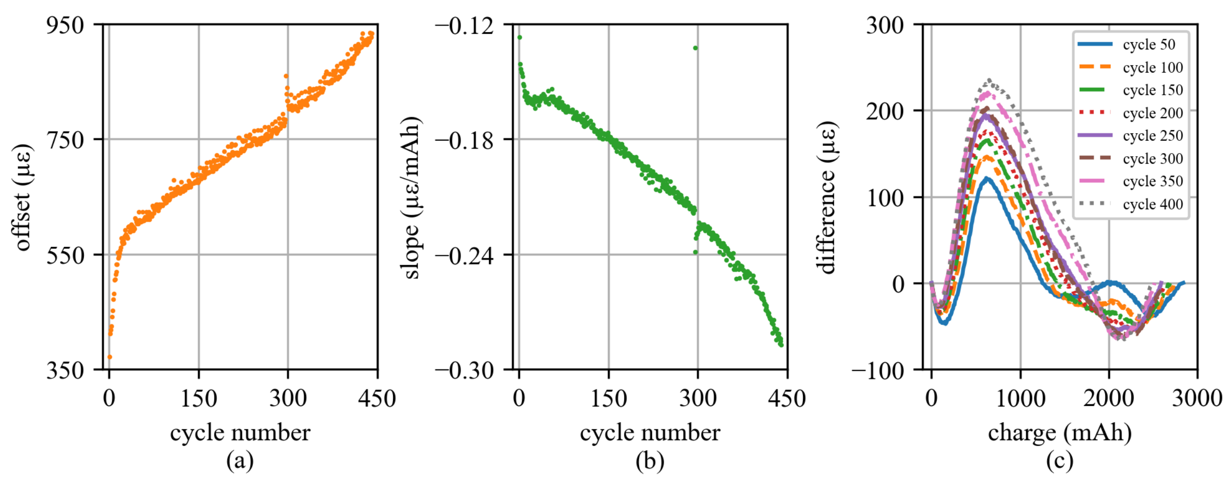

Fig. 5 Hoop strain cycling temporal evolution, depicting: (a) offset component for each cycle, (b) slope component for each cycle, and (c) residual strain profiles over multiple discharge cycles.

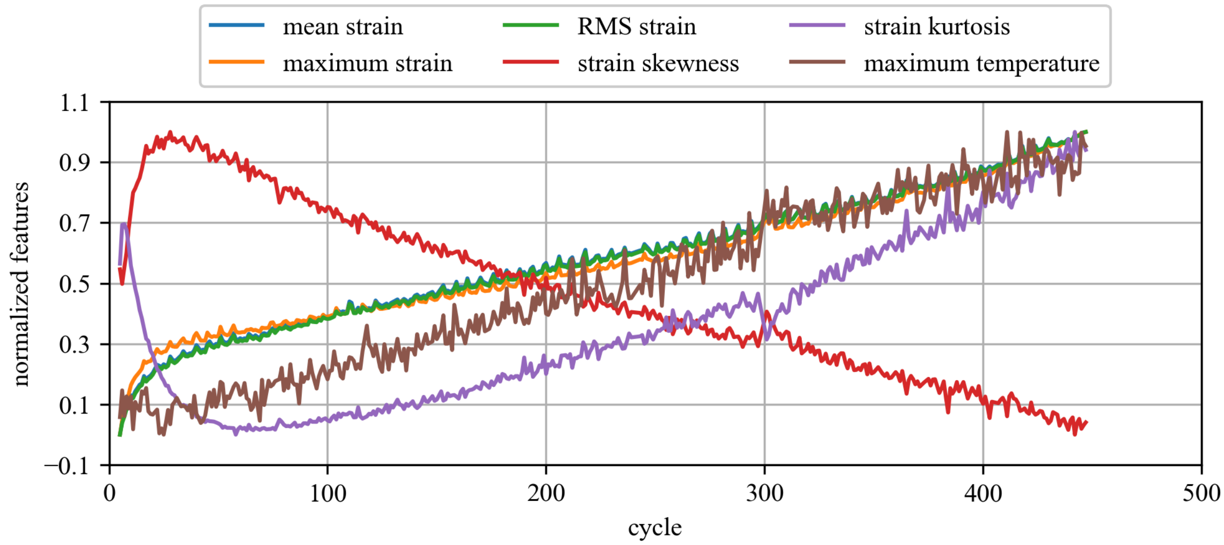

Fig. 6 Normalized features extracted from cell charging data.

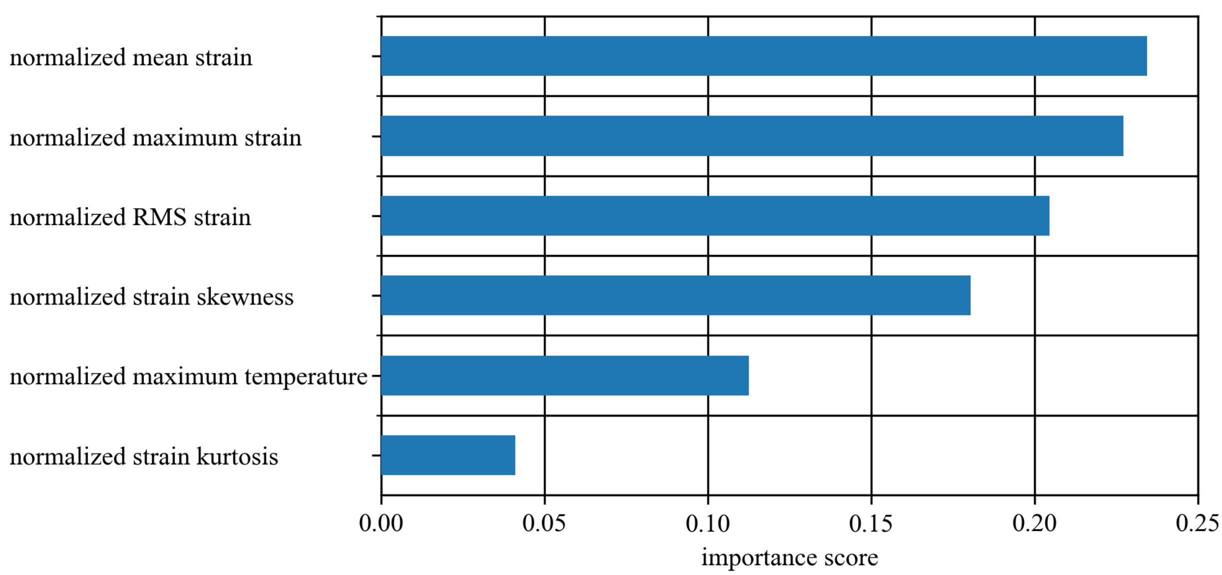

Fig. 7 Impurity-based feature importance score for random forest regression model. Higher scores indicate a greater contribution of the corresponding feature to SOH prediction.

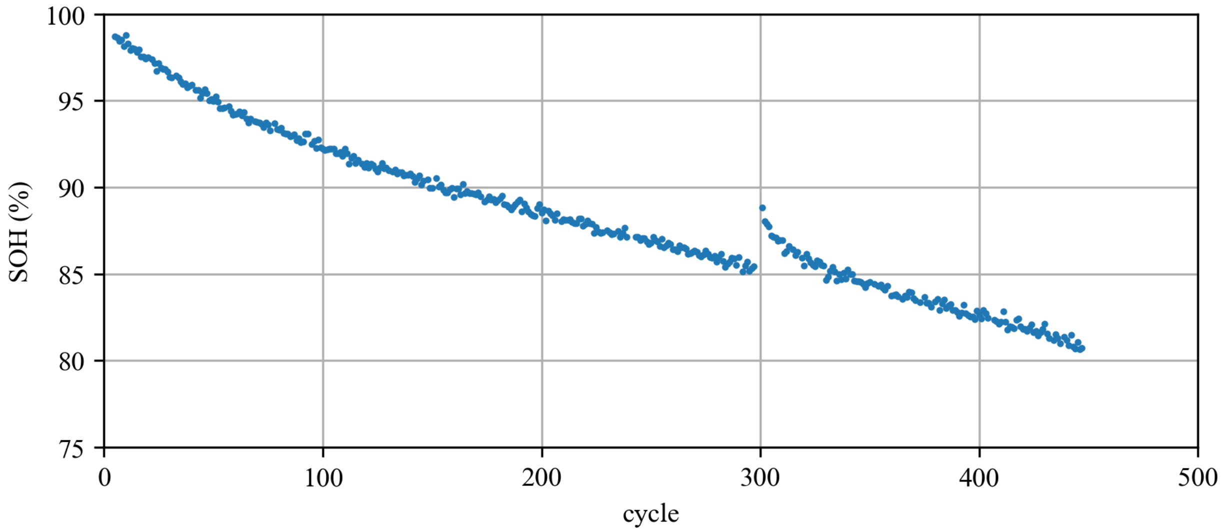

Fig. 8 Cycling SOH degradation results.

quantified the relative contribution of each feature to the SOH prediction task, thereby providing interpretability and insight into the physical mechanisms most strongly associated with battery degradation. Figure 7 presents the ranked importance scores, highlighting the dominant role of strain-derived parameters in the predictive model.

D. Machine Learning Modeling

Extracted strain-derived features were utilized to train a Random Forest regression model for predicting battery state of health (SOH). The Random Forest approach was selected due to its robustness against overfitting, its ability to capture complex nonlinear relationships, and its interpretability through feature importance metrics.

SOH was calculated using the conventional capacity-fade definition, expressed as the ratio of the measured discharge capacity at a given cycle to the nominal rated capacity of the cell and plotted in figure 8. Mathematically, this is given by

$$SOH(\%) = \frac{Q_{\text{measured}}}{Q_{\text{rated}}} \times 100, \tag{1}$$

where Q_{measured} is the discharge capacity obtained during cycling and Q_{rated} is the initial rated capacity of the battery. This formulation provides a normalized measure of degradation that decreases monotonically with cycling, making it suitable for regression modeling.

The Random Forest model was implemented with 100 decision trees, each constrained to a maximum depth of 10. To regulate tree growth and reduce overfitting, a minimum of five samples was required to initiate a split, and each terminal leaf contained at least one sample. At each node, the number of features considered was restricted to the square root of the available predictors, introducing randomness and decorrelation among trees. A fixed random seed was applied to ensure reproducibility. The model was trained on engineered features derived from strain and temperature measurements, supplemented with cycle-aware variables to capture degradation progression. Performance was evaluated using standard regression metrics, including the coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE).

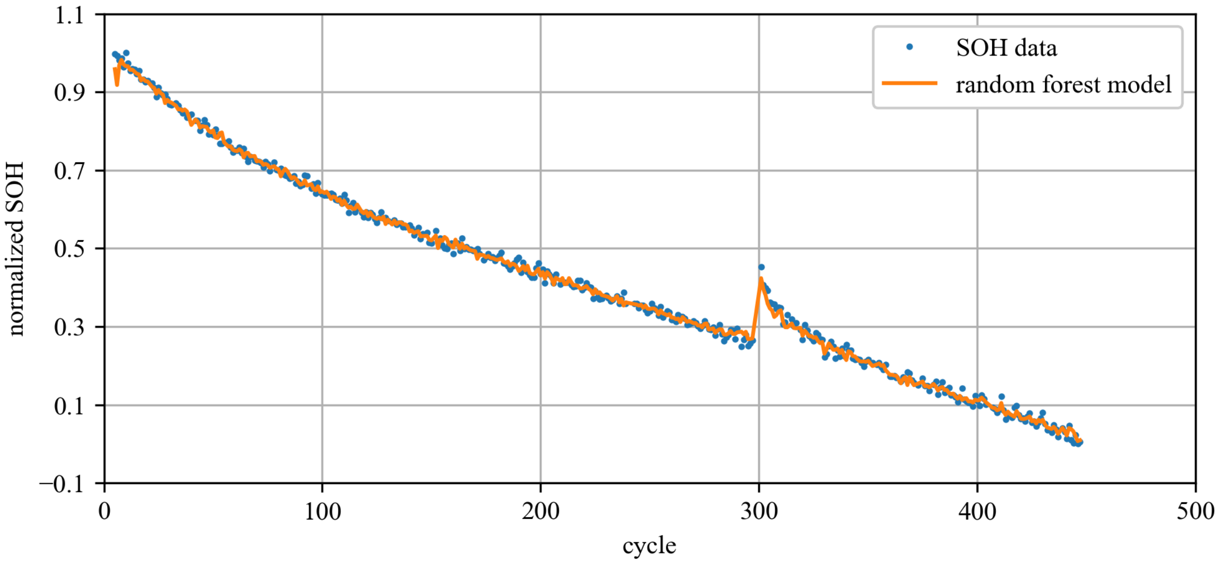

Fig. 9 Random Forest regression results for battery SOH prediction. Blue markers represent experimental SOH data, while the orange line shows the Random Forest model predictions.

III. Results

The Random Forest regression model demonstrated predictive accuracy for battery SOH estimation. The evaluation yielded an R^2 score of 0.9949, indicating that the model explained more than 99% of the variance in SOH. The mean absolute error (MAE) was 0.0151, reflecting minimal average deviation between predicted and observed values. The mean squared error (MSE) was 0.00035, and the corresponding root mean squared error (RMSE) was 0.0187, confirming that prediction errors were consistently small in magnitude. These results highlight the model's ability to capture nonlinear dependencies between strain, temperature, and cycle progression features, producing predictions that closely align with experimental SOH trajectories.

Table 2 Performance metrics of the Random Forest regression model for SOH prediction.

Metric	Value
R^2 Score	0.9949
Mean Absolute Error (MAE)	0.0151
Mean Squared Error (MSE)	0.00035
Root Mean Squared Error (RMSE)	0.0187

Figure 9 illustrates the comparison between experimental SOH data and Random Forest predictions. Blue markers represent the measured SOH values, while the orange line corresponds to the Random Forest model output.

IV. Conclusion

The Random Forest regression model proved to be a potentiality effective tool for data-driven estimation of battery state of health. By combining multiple decision trees with controlled complexity and feature randomness, the model achieved alignment with experimental data. The results demonstrate that ensemble learning methods can capture the complex, nonlinear relationships inherent in battery degradation without requiring explicit mechanistic formulations.

This approach provides a robust framework for predictive diagnostics in lithium-ion systems, offering both accuracy and interpretability through feature importance analysis. Future work will extend this methodology by incorporating additional cycle-aware features, exploring monotonic constraints, or validating performance across diverse operating conditions to further enhance generalization and practical applicability.

Acknowledgments

This material is partially supported by the the National Science Foundation (NSF) grant number CPS - 2237696. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Kirchev, A., Guillet, N., Lonardoni, L., and Dumenil, S., "Li-Ion Cell Safety Monitoring Using Mechanical Parameters, Part 3: Battery Behaviour during Abusive Overcharge," *Batteries*, Vol. 9, No. 7, 2023. doi:10.3390/batteries9070338, URL `https://www.mdpi.com/2313-0105/9/7/338`.

[2] Anthony, G., Madden, C., Ogunniyi, E., Downey, A. R. J., Limbaugh, R., Peskar, J., Bao, J., and Huang, X., "Exploratory investigation of early detection for high-C discharge-induced failure in 18650 lithium-ion batteries," *Health Monitoring of Structural and Biological Systems XVIII*, Vol. 12951, edited by Z. Su, K. J. Peters, F. Ricci, and P. Rizzo, International Society for Optics and Photonics, SPIE, 2024, p. 129510Q. doi:10.1117/12.3011097, URL `https://doi.org/10.1117/12.3011097`.

[3] Hendricks, C., Sood, B., and Pecht, M., "Lithium-Ion Battery Strain Gauge Monitoring and Depth of Discharge Estimation," *Journal of Electrochemical Energy Conversion and Storage*, Vol. 20, No. 1, 2022. doi:10.1115/1.4054340.

[4] Weng, A., Movahedi, H., Wong, C., Siegel, J. B., and Stefanopoulou, A., "Current Imbalance in Dissimilar Parallel-Connected Batteries and the Fate of Degradation Convergence," *Journal of Dynamic Systems, Measurement, and Control*, Vol. 146, No. 1, 2024. doi:10.1115/1.4064028.

[5] Song, J., Li, Y., Ren, T., Wang, J., and Yang, Z., "Two Strain Modes and Transition Point of 18650 Lithium-Ion Battery at Different Charging Rates," 2023. doi:10.2139/ssrn.4526882.

[6] Riedmiller, M., "Advanced supervised learning in multi-layer perceptrons - From backpropagation to adaptive learning algorithms," *Computer Standards and Interfaces*, Vol. 16, No. 3, 1994, pp. 265–278. doi:10.1016/0920-5489(94)90017-5.

[7] Ren, W., Zheng, T., Piao, C., Benson, D. E., Wang, X., Li, H., and Lu, S., "Characterization of commercial 18,650 Li-ion batteries using strain gauges," *Journal of Materials Science*, Vol. 57, No. 28, 2022, pp. 13560–13569. doi:10.1007/s10853-022-07490-4.

[8] Rahimi-Eichi, H., Ojha, U., Baronti, F., and Chow, M.-Y., "Battery Management System: An Overview of Its Application in the Smart Grid and Electric Vehicles," *IEEE Industrial Electronics Magazine*, Vol. 7, No. 2, 2013, pp. 4–16. doi:10.1109/mie.2013.2250351.

[9] ARTS-Lab, "Dataset Cycling with Strain Monitoring for Samsung 30Q Cell," GitHub, n.d. URL `https://github.com/ARTS-Laboratory/dataset-cycling-with-strain-monitoring-for-samsung-30Q-cell`.