# Machine Learning Enabled Prediction of Lithium-Ion Battery Degradation

Ari Everett, Computer Science
Mentor: Dr. Nicholas Rolston, Assistant Professor
School of Electrical, Computer & Energy Engineering

## Introduction

Batteries are a pillar of modern energy, but forecasting their degradation remains a major challenge. **Electrochemical Impedance Spectroscopy (EIS)** and **Machine Learning (ML)** offer complementary ways to address this problem where EIS provides a non-invasive, info-rich reading, while ML builds predictive models that link these signals to capacity and health. This research applies **Gaussian Process Regression (GPR)** and **eXtreme Gradient Boosting (XGB)** to map EIS to capacity, identify degradation patterns, and improve data-driven forecasting of battery **state of health (SoH)**. GPR offers interpretability and uncertainty estimation. XGB is explored for faster, scalable ensemble learning for higher accuracy and transferability.

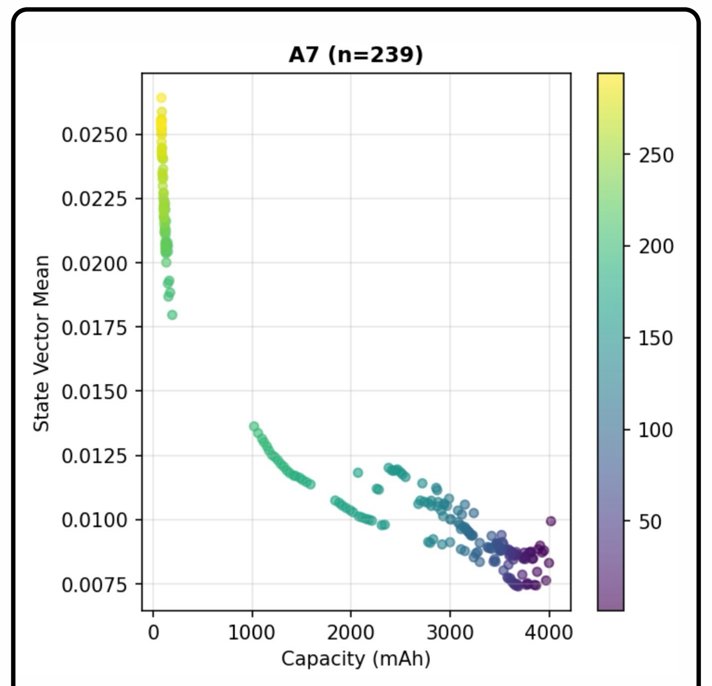


**Fig. 1 Battery Degradation Curve:** Captured data shows capacity consistently declines and impedance increases with cycle number, showing typical battery aging over time. This trend provides the basis for this research and motive using EIS and machine learning to predict health from impedance data

## Methods

**Collect:** EIS data at multiple charge rates and temps; clean, normalize, and extract impedance features from Nyquist spectra.

**Construct:** state vectors combining real & imaginary impedance across frequencies.

**Bin:** data by capacity ranges to balance degradation levels and reduce temporal bias.

**Split: Leave-One-Subject-Out (LOSO)** validation to test model generalization.

**Train:** GPR and XGB to learn EIS to capacity relationships and predict battery degradation.

**Analyze:** feature importance w/ **Automatic. Relevance Determination (ARD)** to see impedance ranges that impact predictions.

**Compare:** model accuracy on test data using regression metrics (**$R^2$, RMSE, MAE, MSE**).

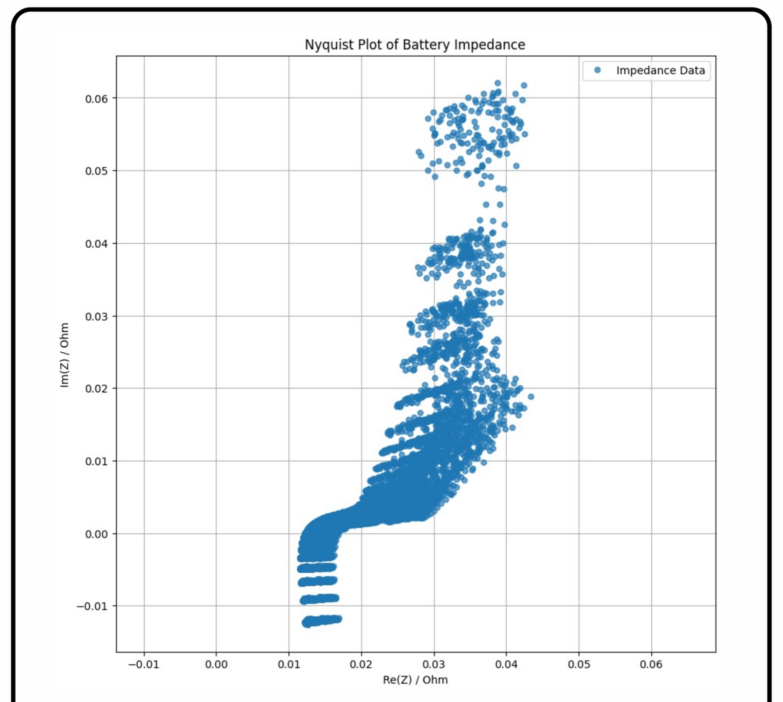


**Fig. 4 Data Integrity:** Nyquist plot from collected EIS data showing a typical impedance curve, with high frequencies near the lower tail and mid to low frequencies forming the semicircular region. The systematic shift in the high-frequency tail suggests behavioral changes that may have led to the elevated feature importance seen in the ARD analysis

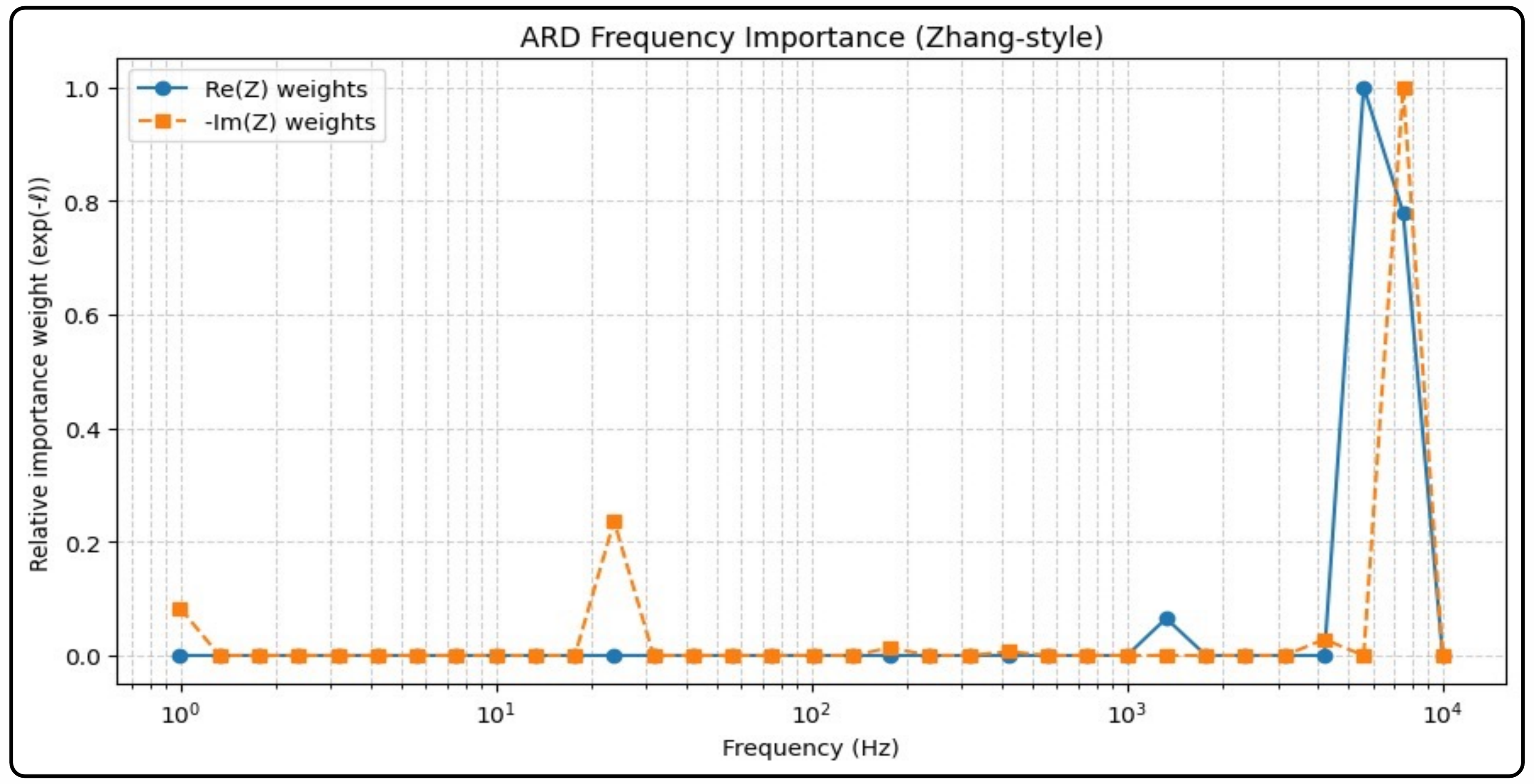


**Fig. 5 Feature importance:** Frequency Weighting

**Experiment:** ARD analysis identified which impedance frequencies most influenced model performance

**Findings:** Unlike prior studies emphasizing 1–100 Hz, higher importance appeared in the **$10^3$–$10^4$ Hz** range, suggesting a shift toward high-frequency sensitivity

**Analysis:** Results indicate possible effects from data balance or preprocessing, warranting further study of frequency weighting and model interpretation

## Results

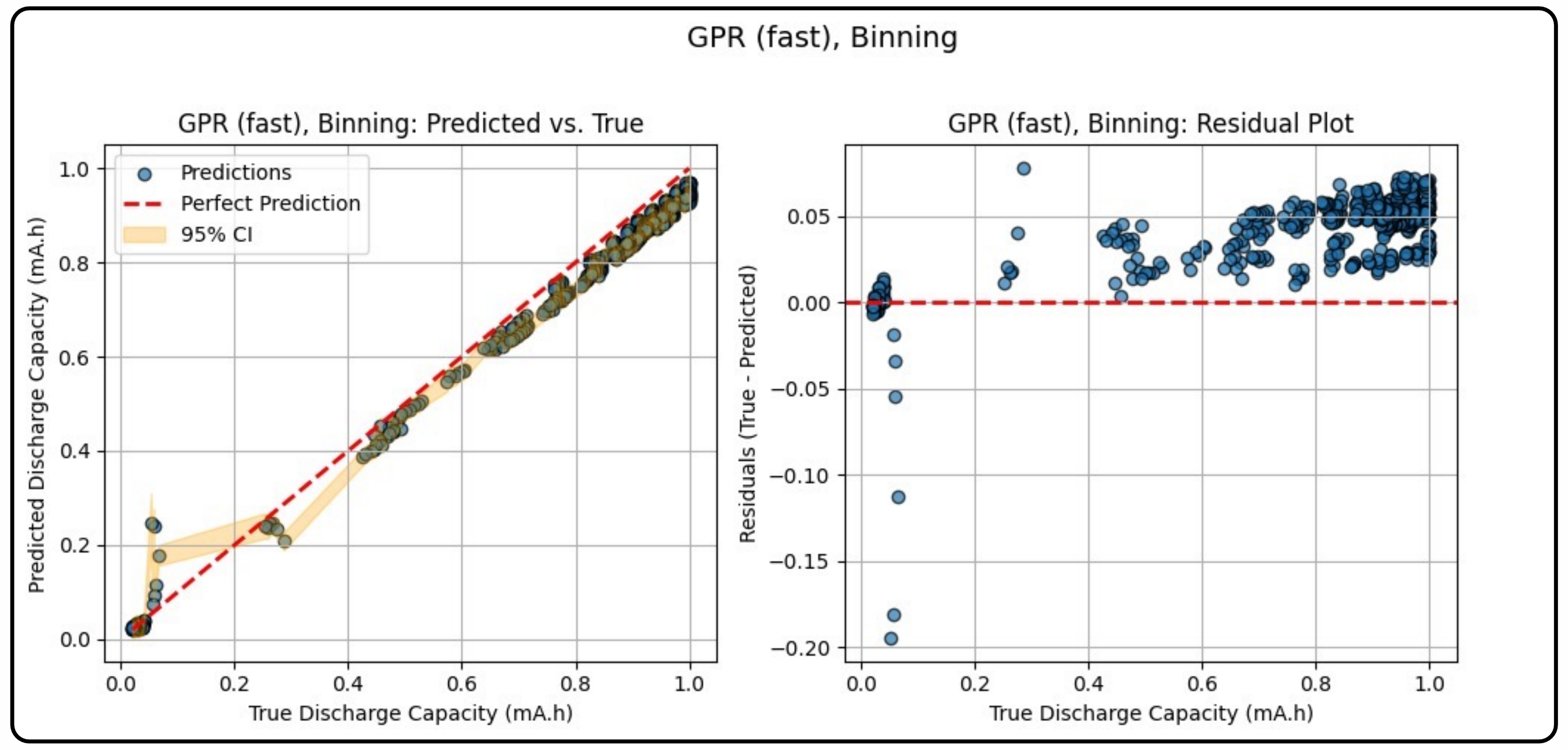


**Fig. 2 GPR performance:** binned GPR model achieved **$R^2$ = 0.87, RMSE = 0.06**, and **MAE = 0.04**, showing a trend with variability between predicated and true capacities

**Predictive behavior:** GPR captured general degradation trends with consistent confidence intervals but showed moderate variability across cells

**Residual analysis:** Errors remain near zero but increased a bit at lower capacities, may have limits in kernel generalization

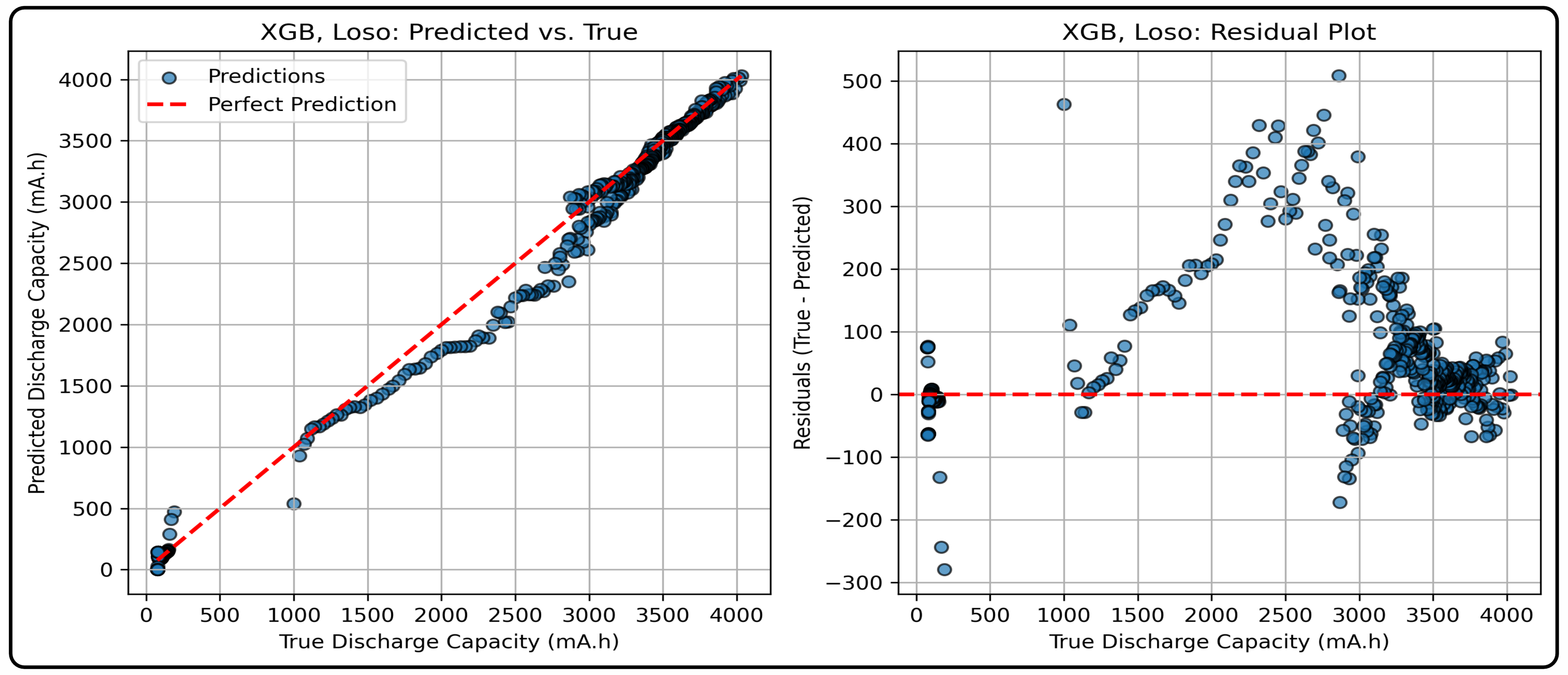


**Fig. 3 XGB performance:** binned + LOSO XGB model achieved **$R^2$ = 0.992**, **RMSE = 116.8**, **MAE = 71.3**, showing strong overall correlation but large absolute errors in predicted capacity

**Generalization:** LOSO indicates transferability across cells, but accuracy weakens for certain capacity ranges/cells

**Residual behavior:** large, not evenly distributed, suggesting bias and pointing to improvements in preprocessing, feature scaling, and frequency selection

## Conclusions

**GPR** achieved moderate accuracy w/ **$R^2$ = 0.87, RMSE = 0.06** and captured general degradation trends, while **XGB** performed stronger at **$R^2$ = 0.99**, but less precise at **RMSE = 116.7** and generalized better under **LOSO**. **XGB** showed larger residual errors, revealing the need for improved preprocessing and feature scaling.

**Capacity binning** improved data balance but may have reduced the visibility of long-term degradation trend, while **ARD analysis** indicated unexpected high-frequency weighting at $10^3$–$10^4$ Hz, highlighting areas for refinement in data processing, physical interpretation, and more experimenting.

## Future Work

**Improve data preparation and grouping** to better track how batteries age over time.

**Adjust models to make predictions more stable** and consistent across different cells.

**Test lower frequency ranges** to confirm the unusual high-frequency behavior and improve data pipeline.

## References & Acknowledgements

1. Zhang, X. et al. *Nat. Commun.* 2020, **11**, 15235
2. Margoschis, S. *ASU Thesis*, 2024
3. Rolston, N. et al. *NEI STC Tech. Vol.*, 2023

Special thanks to **Dr. Nicholas Rolston** for mentorship, **Nithin Jakrebet** for guidance, **Patricio Solana** and the Rolston Labs team for data collection, & **my friends and family** for support

FURI

Ira A. Fulton Schools of Engineering
Arizona State University