****

**Lithium-Ion Battery Degradation Analysis for Electric Vehicles Using Real-Driving Data**

**Submitted By**

**Ahmet Hakan Yıldız 2000003057**

**Yusuf Abdullatif 2100006330**

Department of Computer Engineering

İstanbul Kültür University

<Year><Term>

# ABSTRACT

The rise of electric vehicles (EVs) as a sustainable alternative to fossil fuel-powered transportation has highlighted the critical role of lithium-ion (Li-ion) batteries, valued for their high energy density and reliability. However, real-world usage leads to battery degradation, characterized by capacity loss and internal resistance growth, which poses significant challenges to EV efficiency, range, and longevity. This study investigates Li-ion battery aging patterns using real-driving discharge profiles, including Urban Dynamometer Driving Schedule (UDDS) data and periodic diagnostic tests such as capacity tests, Hybrid Pulse Power Characterization (HPPC), and Electrochemical Impedance Spectroscopy. By developing machine learning and physics-informed models, this research aims to predict key battery health metrics, including State of Health (SOH) and Remaining Useful Life (RUL). The findings have potential implications for improving EV battery management systems, optimizing charging strategies, and advancing the design of more durable and sustainable battery technologies.

TABLE OF CONTENTS

[ABSTRACT I](#_Toc5951944)

[ACKNOWLEDGEMENTS II](#_Toc5951946)

[TABLE OF CONTENTS III](#_Toc5951947)

[LIST OF TABLES IV](#_Toc5951948)

[LIST OF FIGURES V](#_Toc5951949)

[SYMBOLS & ABBREVIATIONS VI](#_Toc5951950)

[1. INTRODUCTION 1](#_Toc5951951)

[1.1. Problem Statement 1](#_Toc5951952)

[1.2. Project Purpose 1](#_Toc5951953)

[1.3. Project Scope 1](#_Toc5951954)

[1.4. Objectives and Success Criteria of the Project 1](#_Toc5951955)

[1.5. Report Outline 1](#_Toc5951956)

[2. RELATED WORK 2](#_Toc5951957)

[2.1. Existing Systems 2](#_Toc5951958)

[2.2. Overall Problems of Existing Systems 2](#_Toc5951959)

[2.3. Comparison Between Existing and Proposed Method 2](#_Toc5951960)

[3. METHODOLOGY 3](#_Toc5951961)

[3.1. Requirement Analysis 3](#_Toc5951962)

[3.2. Design 3](#_Toc5951963)

[3.3. Implementation 3](#_Toc5951964)

[3.4. Testing 3](#_Toc5951965)

[3.1. Overview of the Dataset/Model 3](#_Toc5951966)

[3.2. Tools and Technology 3](#_Toc5951967)

[3.3. Proposed Approach 3](#_Toc5951968)

[3.1. Design Overview 4](#_Toc5951969)

[3.2. System Architecture 4](#_Toc5951970)

[3.2.1. Module A 4](#_Toc5951971)

[3.2.2. Module B (and more, if necessary) 4](#_Toc5951972)

[3.3. System Software 4](#_Toc5951973)

[4. EXPERIMENTAL RESULTS 5](#_Toc5951974)

[5. DISCUSSION 6](#_Toc5951975)

[6. CONCLUSIONS 7](#_Toc5951976)

[REFERENCES 8](#_Toc5951977)

[APPENDIX 9](#_Toc5951978)

# LIST OF TABLES

[Table 2.1: Comparison of methods 2](#_Toc5872571)

# LIST OF FIGURES

[Figure 4.1: Comparison with the current best algorithm and our algorithm 4](#_Toc5872579)

# SYMBOLS & ABBREVIATIONS

ACM: Association for Computing Machinery

APA: American Psychological Association

IEEE: Institute of Electrical and Electronics Engineers

# INTRODUCTION

## Problem Statement

Lithium-ion batteries in electric vehicles (EVs) combine high energy density with efficiency and reliability, making them critical components of sustainable transportation. However, their degradation—manifested as capacity loss and increased internal resistance—remains a significant challenge, negatively impacting vehicle range, performance, and consumer satisfaction. Understanding battery degradation patterns under real-world traffic conditions is essential to develop solutions that enhance battery life and performance.

## Project Purpose

This project aims to investigate lithium-ion battery degradation in EVs using real-world driving data. Leveraging the dataset provided by Gabriele Pozzato and colleagues, the study seeks to uncover the mechanisms of battery aging and develop predictive maintenance models. The findings of this research have the potential to improve battery management systems and extend the operational lifespan of EV batteries.

## Project Scope

The scope of this study includes data preprocessing, visualization of key relationships, and the development of machine learning models to predict battery degradation trends. The study will evaluate parameters such as temperature, test duration, voltage, current, and other factors related to the internal structure of the batteries to identify significant patterns and correlations.

## Objectives and Success Criteria of the Project

The objectives of this study are:

* Identify the impact of external factors on charge/discharge cycles.
* Analyze and explore correlations between temperature, internal resistance, and energy loss.
* Develop machine learning models to predict battery degradation based on current usage patterns.

The success of this project will be evaluated based on the accuracy of the battery degradation predictions and their applicability to real-world battery management systems.

## Report Outline

Our report starts with the literature review(Section II), continues by the methodology of analyzing dataset (Section III). In section IV we present the results of our study, showing key trends and our ml predictive model. In section V we discuss our findings, and in Section VI we finally conclude our project and include some recommendations for future research in this field.

### 2. RELATED WORK

This section provides an overview of previous research and systems that address similar problems related to lithium-ion battery degradation analysis and real-driving data in electric vehicles (EVs). It discusses existing methods, identifies challenges, and compares them with the proposed approach in this study.

**2.1. Existing Systems**

Several studies have explored various methods for analyzing lithium-ion battery aging and degradation, particularly in the context of electric vehicles. One of the notable works, *Pozzato et al.* (2022), proposed a dataset based on real-driving data profiles to monitor battery degradation. This dataset has been instrumental in understanding battery performance and degradation under real-world driving conditions, offering a more practical approach compared to laboratory tests. Similarly, *Xia et al.* (2021) employed machine learning techniques to predict battery health, demonstrating a strong correlation between driving patterns and battery wear.

Other approaches, such as *Wang et al.* (2020), used physics-based models to simulate battery aging, while *Jiang et al.* (2019) focused on incorporating temperature effects into degradation models. Although these studies have made significant contributions, they typically rely on controlled conditions or simulations, which may not fully capture the complexities of real-world driving scenarios.

**2.2. Overall Problems of Existing Systems**

Despite the advancements made by these systems, several challenges remain, especially when attempting to use real-driving data for accurate aging predictions. A common problem is the limited availability of real-driving datasets, which often leads to models that are less generalizable. For example, the models proposed by *Wang et al.* and *Jiang et al.* struggle with accounting for the variations in driving behavior, road conditions, and environmental factors that significantly influence battery aging. Furthermore, the complexity of real-world driving profiles requires models that can effectively handle large amounts of dynamic data and account for the non-linear nature of battery degradation.

**2.3. Comparison Between Existing and Proposed Method**

The proposed method in this study seeks to address these challenges by utilizing a dataset based on real-driving discharge profiles, which includes diverse driving conditions and battery usage patterns. The following table compares the key features of the proposed method with existing approaches.

**Table 2.1: Comparison of Methods**

| **Method A (Pozzato et al., 2022)** | **Method B (Xia et al., 2021)** | **Method C (Wang et al., 2020)** | **Method D (Jiang et al., 2019)** | **Our Method** |
| --- | --- | --- | --- | --- |
| Real-driving data based | Machine learning for SOH prediction | Physics-based aging model | Temperature effects modeling | Real-driving dataset & ML-based aging prediction |
| Performance: High correlation with driving profiles | Performance: Moderate accuracy | Performance: High in lab conditions | Performance: Moderate with environmental adjustments | Performance: High in real-driving scenarios |
| Accuracy: High for driving patterns | Accuracy: High for battery state estimation | Accuracy: Limited by controlled testing | Accuracy: Limited by environmental assumptions | Accuracy: High for real-world applications |
| Scalability: Limited dataset availability | Scalability: Scalable with more data | Scalability: Low for large datasets | Scalability: Moderate with sensor data | Scalability: High, real-time data processing |
| Limitation: Limited to specific datasets | Limitation: Dependent on driving patterns | Limitation: Unable to replicate real-world driving conditions | Limitation: Temperature factors may not cover all conditions | Limitation: Requires continuous data collection |

Note: Ensure that all figures, tables, and other elements in the document are referenced correctly. If adapting data from another source, provide full citations as per academic standards (e.g., "Reprinted from [Pozzato et al., 2022]").

# [METHODOLOGY](#_Toc470871184)

This section tells how you conducted your project. It should be detailed enough to guide someone who wants to reproduce your study.

Consult your supervisor to choose only one of the sub-section groups to implement your report!

- For Data/Model-driven Research Projects;

## 3.1. Overview of the Dataset/Model

The dataset used in this project was provided by Gabriele Pozzato and colleagues and consists of real-driving discharge profiles, including the Urban Dynamometer Driving Schedule (UDDS), periodic diagnostic tests like capacity tests, Hybrid Pulse Power Characterization (HPPC), and Electrochemical Impedance Spectroscopy. The dataset spans a 23-month period and includes data from 10 INR21700 M50T lithium-ion cells, simulating real-world EV driving and charging conditions.

Key features of the dataset include:

* **Voltage and Current**: Recorded to analyze energy efficiency and degradation.
* **Test Time**: Tracks the time elapsed during various tests.
* **Cycle Count**: Monitors the number of charge-discharge cycles.

Preprocessing steps involved handling missing data, removing outliers, and normalizing features for consistency across the dataset.

## 3.2. Tools and Technology

The following tools and technologies were utilized for this project:

* **Programming Language**: Python, due to its extensive libraries for data analysis and machine learning.
* **Libraries and Frameworks**:
  + **Pandas**: For data cleaning and manipulation.
  + **Matplotlib/Seaborn**: For visualizing relationships and trends in the data.
  + **Scikit-learn**: For building and evaluating machine learning models.
  + **TensorFlow/PyTorch** (if applicable): For advanced deep learning models.
  + **NumPy**: For numerical computations.

The analysis was conducted on Google Collab. Version control was managed using Git, and Python IDEs like Jupyter Notebook were used for code development.

## 3.3. Proposed Approach

1. **Data Preprocessing**:

* Missing data was imputed where necessary, and outliers were removed based on statistical thresholds.
* Features were normalized to ensure uniform scaling.
* New features were engineered, such as depth of discharge and cycle life estimations, to enrich the dataset.

1. **Exploratory Data Analysis (EDA)**:

* Visualizations such as line plots, scatter plots, and heatmaps were used to uncover trends in SOH, RUL, temperature, and internal resistance.
* Correlation matrices were generated to identify relationships between critical parameters, such as temperature and capacity loss.

1. **Model Development**:

* Machine learning models, including Linear Regression, Random Forest, and Gradient Boosting, were implemented to predict SOH and RUL.
* Hybrid approaches incorporating physics-based principles were explored for greater accuracy.
* Model performance was evaluated using metrics like Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R²).

1. **Validation**:

* The dataset was split into training and testing subsets using an 80:20 ratio, and cross-validation was applied to avoid overfitting.
* Models were tested on unseen data to ensure robustness and generalization.

## 3.4. Code Implementation

For the implementation, we structured the code into several steps to ensure clarity and reproducibility. Below is the breakdown of our process:

1. **Data Loading and Preprocessing**  
   We began by importing the necessary libraries, such as pandas, numpy, matplotlib, and seaborn. After loading the dataset, we carefully inspected it for missing values, duplicates, and outliers. Missing data was handled using interpolation, and any data points identified as anomalies were removed. We applied normalization to numerical features like voltage, temperature, and cycle count to ensure consistency across the dataset. Additionally, we engineered new features, such as Depth of Discharge (DoD) and resistance trends, to better capture the behavior of the batteries over time.
2. **Exploratory Data Analysis (EDA)**  
   To understand the dataset, we will create visualizations, including scatter plots, line graphs, and heatmaps. These visualizations will help us identify patterns and relationships between features such as temperature, internal resistance, SOH, and RUL. We will also generate a correlation matrix to pinpoint which variables have the strongest relationships with the target metrics (SOH and RUL).
3. **Feature Selection and Transformation**  
   Using the results from EDA, we will select the most relevant features for modeling, such as temperature, current, voltage, and cycle count.
4. **Model Training and Evaluation**  
   We will split the dataset into training and testing subsets, typically using an 80:20 split. Multiple machine learning models will be trained, including Random Forest and Gradient Boosting, to predict SOH and RUL. To evaluate the models, we will use metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R²). We will also plot the predicted values against the actual values to visualize the accuracy of the models.
5. **Validation and Visualization**  
   We will perform k-fold cross-validation to ensure the robustness of the models. This helped minimize overfitting and provided a more reliable measure of performance. For visual representation, we will plot residuals, feature importance, and other key metrics to better understand the behavior of the models.

## IV. EXPERIMENTAL RESULTS

1. **A. Data Preprocessing and Analysis**

The dataset was initially analyzed for consistency and completeness. Histograms, capacity test trends, EIS (Electrochemical Impedance Spectroscopy) results, and HPPC (Hybrid Pulse Power Characterization) tests were evaluated across the 10 lithium-ion cells (G1, V4, V5, W3, W4, W5, W7, W8, W9, W10). Cells with incomplete or insufficient data (e.g., missing cycle ranges, absent EIS tests) were excluded from downstream analysis.

**Key Observations:**

* Certain cells displayed irregularities in cycle counts and test frequencies.
* Histogram, capacity trends, and test-specific graphs were generated for each cell to determine usability. *(Figures to be added here)*

1. **B. Prediction of Discharge Capacity**

Machine Learning models were applied to predict the discharge capacity using an 80% training and 20% testing split. The models included:

* **Linear Regression**
* **XGBoost**
* **LightGBM Regressor**
* **Gradient Boosting with GridSearchCV**

**Model Performance Metrics:**

* **Linear Regression:** R²: 98.0%
* **XGBoost:** R²: 87.35%
* **LightGBM:** R²: 87.81%
* **Gradient Boosting:** R²: 87.35%

Due to time constraints, further optimization using GridSearchCV was halted. Overall, Linear Regression achieved the highest accuracy despite its simplistic approach.

*(Include individual model performance graphs and a combined comparison graph here.)*

1. **C. State of Health (SOH) Estimation**

Initial SOH calculations using step\_index resulted in inaccurate estimations. Feature engineering was performed by incorporating Charge\_Discharge\_Counts and Cumulative\_Charge\_Discharge from the original dataset.

**Correlation Analysis:**

* A Pearson Correlation analysis revealed a strong negative correlation (-0.7686) between cycle count and capacity retention. *(Correlation heatmap to be added here.)*

**Model Performance for SOH Prediction:**

| **Model** | **MAE** | **MSE** | **R²** |
| --- | --- | --- | --- |
| Linear Regression | 0.0308 | 0.0009 | 0.9687 |
| Decision Tree | 0.0040 | 0.0000 | 0.9984 |
| Random Forest | 0.0130 | 0.0004 | 0.9876 |
| Support Vector Regressor | 0.1027 | 0.0113 | 0.6339 |

Decision Tree achieved the best overall performance, with minimal error and high predictive accuracy.

*(Graphs for each model and a combined comparison graph should be included.)*

1. **D. Cross-Cell Model Evaluation (W8, W9, W10)**

The SOH prediction model trained on Cell W8 was tested on Cells W9 and W10.

**Performance on W9:**

* **Linear Regression:** R²: 0.8689
* **Decision Tree:** R²: 0.9938
* **Random Forest:** R²: 0.9757
* **Support Vector Regressor:** R²: 0.6620

**Performance on W10:**

* **Linear Regression:** R²: 0.8609
* **Decision Tree:** R²: 0.9815
* **Random Forest:** R²: 0.9648
* **Support Vector Regressor:** R²: 0.6401

Decision Tree consistently outperformed other models in both W9 and W10, demonstrating robustness across cells.

*(Include W9 and W10 performance comparison graphs.)*

1. **E. Statistical Analysis**

To validate the significance of the results, the following statistical tests were conducted:

* **T-Test:** T-statistic = 10211.72, P-value = 0.0
* **ANOVA Test:** F-statistic = 10316961.75, P-value = 0.0
* Both tests indicate statistically significant differences across the predicted and actual SOH values.

Figure 4.1: Comparison with the current best algorithm and our algorithm

1. **V. DISCUSSION**

This study aimed to predict the discharge capacity and State of Health (SOH) of lithium-ion batteries using real-driving discharge profile data. Accurate SOH estimation is critical for ensuring battery longevity, optimizing charging strategies, and improving the reliability of electric vehicles. The experimental results demonstrated that advanced machine-learning models could effectively predict SOH across multiple cells, highlighting both model strengths and limitations.

1. **Key Findings and Significance**

* **Model Accuracy:** Among the applied models, the Decision Tree Regressor consistently outperformed other algorithms, achieving an R² score of **0.9938** for W9 and **0.9815** for W10. This indicates that Decision Trees are highly effective at capturing the non-linear relationships in battery degradation data.
* **Cross-Cell Generalization:** The model trained on Cell W8 successfully generalized its predictions to Cells W9 and W10, confirming that patterns of battery degradation share similarities across cells from the same batch.
* **Correlation Insight:** The Pearson correlation coefficient of **-0.7686** revealed a strong negative correlation between cycle count and capacity retention, aligning with existing literature on battery degradation behavior.
* **Statistical Validation:** The T-test (**T-statistic: 10211.72, P-value: 0.0**) and ANOVA test (**F-statistic: 10316961.75, P-value: 0.0**) confirmed the statistical significance of the predictive models' results.

1. **Potential Sources of Error and Anomalies**

Despite high performance, some limitations were observed:

* **Data Variability:** Differences in cell manufacturing or operating conditions could introduce inconsistencies in model predictions across datasets.
* **Feature Engineering Limitations:** While key features such as Cumulative\_Charge\_Discharge and Charge\_Discharge\_Counts improved accuracy, additional derived features might further enhance predictions.
* **Model Overfitting:** Certain models, like Decision Trees, are prone to overfitting on smaller datasets, which could affect scalability.

1. **Impact and Applications**

The findings of this study contribute to the ongoing effort to enhance Battery Management Systems (BMS) in electric vehicles. Accurate SOH estimation models enable proactive battery maintenance, reducing unexpected failures and optimizing overall performance. Furthermore, these models can be integrated into cloud-based battery analytics platforms, offering real-time insights for EV fleet operators.

1. **Future Directions**

* **Integration with Real-Time Systems:** Implement predictive models directly into real-time BMS frameworks.
* **Dataset Expansion:** Train and validate the models on larger, more diverse datasets across different battery chemistries and use cases.
* **Hybrid Models:** Explore ensemble approaches combining the strengths of Decision Trees, Random Forest, and Gradient Boosting.

1. **VI. CONCLUSIONS**

This study focused on developing predictive models for lithium-ion battery degradation using real-driving discharge profile data. By leveraging advanced machine-learning algorithms, the research successfully addressed the challenge of accurate SOH prediction.

The **Decision Tree Regressor** emerged as the most effective model, demonstrating high accuracy across training and validation datasets, as well as during cross-cell evaluation on W9 and W10. Statistical tests further validated the reliability of these predictions, and correlation analysis provided key insights into the relationship between cycle count and capacity retention.

The findings hold significant implications for improving **Battery Management Systems (BMS)** in electric vehicles, enabling more efficient battery monitoring, predictive maintenance, and enhanced operational efficiency.

In the future, integrating these models into real-time systems and expanding the datasets to include varied operational profiles will pave the way for even more robust predictive frameworks. This research serves as a foundation for future advancements in electric vehicle battery analytics and contributes to the broader goal of sustainable energy solutions.