Comprehensive Report on

Machine Learning-Based State of Health Estimation for Lithium-Ion Batteries Using NASA Dataset

Prepared For

Dr. G C Nayak (Project Supervisor)

Dr. Pritam Kumar Jana (Project Co-Supervisor)

Prepared By -

Sidartha Sankara Pati: 2020B2A70687P

Abstract

The accurate estimation of the State of Health (SoH) is critical for ensuring the safe, efficient, and sustainable operation of lithium-ion batteries, which are integral to applications ranging from electric vehicles to renewable energy systems. This project focuses on developing and evaluating machine learning models for predicting battery SoH using publicly available datasets, particularly the NASA lithium-ion battery dataset. Both univariate and multivariate approaches were explored, leveraging features such as capacity and other operational parameters to improve prediction accuracy.

The study implemented convolutional neural networks (CNNs) to extract meaningful patterns from the dataset, utilizing batteries B05, B18, B33, and B34 for training and B46, B47, and B48 for testing. The models were evaluated based on performance metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), with results demonstrating the effectiveness of CNNs in capturing complex degradation trends. The findings highlight the potential of machine learning in advancing battery health management systems and provide insights into the applicability of data-driven approaches for real-world scenarios. This work not only contributes to the field of battery health monitoring but also paves the way for further advancements in predictive maintenance and sustainability.

Acknowledgement

I would like to express my heartfelt gratitude to all those who have guided and supported me throughout the course of this project. First and foremost, I extend my sincere thanks to my supervisor, Prof. G. C. Nayak, for his invaluable guidance, encouragement, and insightful feedback, which have been instrumental in shaping this work. His expertise and unwavering support provided me with the confidence to navigate the complexities of this research.

I am equally grateful to my co-supervisor, Prof. Pritam Kumar Jana, for his constant mentorship, constructive suggestions, and keen insights, which greatly enriched the quality of my work. His attention to detail and hands-on approach made this learning journey both inspiring and rewarding.

I would also like to acknowledge the faculty members of the Department of Computer Science and Chemistry at BITS Pilani for their continuous encouragement and for equipping me with the foundational knowledge necessary for this interdisciplinary research.

Special thanks go to my family, whose unwavering belief in me has been my greatest source of motivation. Lastly, I extend my gratitude to my peers and friends, whose support and camaraderie made this endeavor more enjoyable and fulfilling.

This project represents not only an academic milestone but also a reflection of the collective support I have received, for which I am truly grateful.

Table Of Contents

List of Figures

1. Introduction

1.1 What is battery health management and why is it important?

Battery health management (BHM) involves monitoring and maintaining the optimal performance and longevity of batteries, especially lithium-ion batteries, by assessing factors like State of Health (SoH), capacity, and internal resistance. Effective BHM practices use algorithms and predictive models to estimate remaining capacity and expected lifespan based on real-time data and usage patterns.

Importance of Battery Health Management:

1. Safety: Deteriorating batteries can lead to overheating, swelling, or even combustion. BHM helps to detect such issues early, reducing safety risks.
2. Cost Efficiency: By predicting the battery's remaining useful life, BHM reduces the frequency of replacements and the overall cost of ownership.
3. Optimal Performance: BHM ensures batteries operate at peak performance, maintaining their efficiency in powering devices or vehicles over time.
4. Sustainability: Effective BHM reduces the number of discarded batteries, helping to lower environmental impact and support sustainability goals.

Battery health management enhances safety, performance, and longevity while contributing to environmental conservation—a critical factor in today's growing reliance on battery-powered technology.

1.2. What is State Of Health (SoH)?

State of Health (SoH) is a measure used to indicate the overall condition of a battery relative to its ideal or original state. It considers factors like capacity fade, internal resistance, and charge acceptance, which affect a battery's ability to store and deliver energy effectively. Typically expressed as a percentage, SoH reflects how much the battery's capacity and performance have degraded over time; for example, an SoH of 80% means the battery can hold 80% of its original capacity when new. Monitoring SoH is crucial in battery health management, as it helps predict lifespan, ensure safety, and maintain optimal performance in devices or systems relying on battery power.

1.3 What are some methods to compute SoH?

The State of Health (SoH) of a battery can be calculated using various methods, each suited to different applications and data availability. Below are some commonly used approaches along with their mathematical formulations:

1. Capacity Fade Measurement - This method evaluates the degradation of a battery by comparing its current capacity to its nominal (initial) capacity:

$$SoH = \frac{Current\ Capacity}{Initial\ Capacity} \times 100\%$$

   As the battery ages, its ability to store charge decreases, making this approach one of the most direct ways to measure SoH.

2. Internal Resistance Analysis - As a battery degrades, its internal resistance increases. SoH can be estimated using the following formula:

$$SoH = \left(1 - \frac{Current\ Internal\ Resistance - Initial\ Internal\ Resistance}{Initial\ Internal\ Resistance}\right) \times 100\%$$

   This method relies on measuring the deviation of internal resistance from its initial value, where larger deviations indicate greater degradation.

3. Voltage Curve Analysis - Voltage curve analysis involves assessing the battery's voltage response during charging and discharging cycles.

$$SoH = f(Voltage\ Response)$$

   This approach compares the measured voltage profile against reference profiles at different SoH levels, enabling an inference of the battery's condition.

4. Coulomb Counting (Charge Throughput) - This approach measures the total charge the battery can deliver over time and compares it to its initial discharge capacity:

$$SoH = \frac{Discharge\ Capacity\ over\ Time}{Initial\ Discharge\ Capacity} \times 100\%$$

   It is a practical method for systems where charge and discharge data are regularly recorded.

5. Data-Driven and Machine Learning Models - Advanced data-driven techniques utilize machine learning algorithms to analyze historical and real-time data, including capacity, voltage, temperature, and current. While these models do not follow a universal formula,

they use predictive algorithms to estimate SoH based on patterns and correlations in the data. These models require extensive datasets but offer high accuracy and adaptability.

1.4 Applications of SoH

State of Health (SoH) estimation has numerous practical applications across industries, primarily focusing on the efficient, safe, and sustainable utilization of batteries. Some of its key applications include:

1. Electric Vehicles (EVs) - SoH estimation is critical for monitoring battery degradation in EVs, helping predict remaining driving range and scheduling maintenance or replacement.It ensures the safe operation of batteries, preventing failures that could lead to safety hazards.
2. Renewable Energy Systems - In solar and wind energy storage systems, SoH helps assess the reliability and capacity of battery banks. It aids in scheduling timely replacements to ensure consistent energy availability.
3. Consumer Electronics - Devices like smartphones, laptops, and wearables rely on SoH metrics to inform users about battery life and suggest replacements. Accurate SoH estimation enables manufacturers to optimize power management systems, extending device longevity.
4. Aerospace and Defence - SoH monitoring is vital for critical systems like satellites, drones, and military equipment, where battery failure could have severe consequences. It ensures operational reliability under extreme conditions.
5. Logistics and Supply Chain - In battery-powered forklifts, automated guided vehicles (AGVs), and drones, SoH monitoring ensures uninterrupted operations by identifying batteries at risk of failure.
6. Smart Grids and Energy Storage - SoH estimation allows efficient energy distribution by monitoring the health of storage systems. It supports predictive maintenance, improving the overall stability of smart grids.

2. CNN

2.1 What is a Convolutional Neural Network?

An artificial neural network type called a convolutional neural network (CNN) is made specifically for processing and evaluating visual data. In tasks like object detection, image categorization, and image recognition, it has shown to be incredibly effective. The following are a CNN's essential parts:

a. Convolutional Layers: In order to identify patterns and features in various spatial locations, these layers apply convolution operations to the input data using filters or kernels.

b. Pooling Layers: In order to reduce the amount and complexity of the input data, pooling layers downsample the spatial dimensions of the data. CNNs frequently use two popular techniques: max pooling and average pooling.

c. Activation Functions: By adding non-linearity to the network, activation functions—like Rectified Linear Units, or ReLUs—allow the network to recognize intricate patterns and correlations in the data.

d. Fully Connected Layers: Based on the features retrieved from preceding layers, these levels link every neuron in the network to every other neuron in the layers above and below. This helps the network generate final predictions.

e. Flattening: The output of the convolutional and pooling layers is converted into a one-dimensional vector prior to the fully connected layers.

CNNs are particularly effective in handling spatial hierarchies and translational invariance, making them well-suited for tasks involving visual data. They have been crucial in the advancement of computer vision. They have applications in many areas, such as autonomous cars, medical image analysis, and image and video identification.

2.2 Relevance of CNN in SoH Prediction

Convolutional Neural Networks (CNNs) are highly relevant and effective for predicting the State of Health (SoH) of batteries due to their ability to extract meaningful patterns from complex data. Some reasons why CNNs are particularly suited for this task:

1. Feature Extraction from Sequential Data - Battery data, such as voltage, current, and temperature over time, can be represented as time-series or structured matrices. CNNs excel at automatically extracting features from such data without the need for extensive manual engineering.They identify subtle, non-linear relationships between variables, which are crucial for accurate SoH prediction.
2. Robustness to Noise - CNNs are resilient to noise in the input data, a common challenge in battery systems due to measurement errors or environmental variations. This robustness ensures more reliable predictions even with imperfect data.
3. Scalability and Generalization - Once trained, CNNs can generalize well across different battery types and operating conditions, provided the training dataset is diverse enough. This makes them adaptable to various battery chemistries and use cases.
4. Integration with Multi-Model Data - CNNs can process multi-channel inputs, such as combining time-series voltage and temperature data or other sensor readings, enabling comprehensive analysis for SoH prediction.
5. High Predictive Accuracy - CNNs can capture local and global features in data, allowing them to model complex degradation processes effectively. They often outperform traditional methods and simpler machine learning models in SoH estimation accuracy.
6. Applications in Real-Time Monitoring - CNNs can be implemented in embedded systems for real-time SoH prediction, making them suitable for practical applications like electric vehicles and renewable energy storage systems.

CNNs offer a powerful tool for SoH prediction by leveraging their ability to process complex data patterns, adapt to diverse scenarios, and deliver high accuracy, making them a valuable asset in battery health management systems.

3. Literature Review

3.1 Battery State-of-Health Estimation Using Machine Learning and Preprocessing with Relative State-of-Charge - Sungwoo Jo , Sunkyu Jung and Taemoon Roh

The introduction sets the stage for the importance of lithium-ion batteries in modern technology, particularly in electric vehicles and renewable energy systems. The authors emphasize the need for accurate state-of-health (SOH) estimation to ensure the efficiency and safety of these batteries. They discuss the challenges associated with traditional model-based methods, which often require complex electrochemical knowledge and can be inefficient in terms of training and accuracy. The authors provide a brief overview of existing methods for SOH estimation, highlighting the shift towards data-driven approaches, particularly machine learning. They note that while machine learning can bypass the need for detailed electrochemical modeling, the effectiveness of these models heavily relies on the quality of the input data and preprocessing techniques. The authors introduce a new preprocessing method that focuses on the relative state of charge (SOC) rather than the typical SOC, which is often calculated based on design capacity. The relative SOC is defined as a simpler indicator that correlates more directly with the battery's energy level. This transformation from time-domain data to SOC-domain data is crucial because battery characteristics are more closely related to their energy levels than to time intervals. The preprocessing method involves transforming time-domain data into SOC-domain data, which allows for a more accurate representation of the battery's state. This transformation is essential for improving the performance of machine learning models, as it aligns the input data more closely with the underlying physical characteristics of the battery. The paper discusses various machine learning models used for SOH estimation, including FNN amd LSTM. The authors highlight the effectiveness of simpler FNN models when combined with the proposed preprocessing method. While LSTMs are powerful for sequential data, the authors argue that simpler models can achieve comparable results with the right preprocessing. The authors describe the experimental setup used to validate their proposed method. They detail the datasets used for training and testing, including the number of cycles and the specific batteries involved. The results section presents a comprehensive comparison of SOH estimation accuracy between the two approaches (time-based vs. SOC-based). The authors provide quantitative metrics, such as mean absolute error (MAE) and root mean square error (RMSE), to demonstrate the improvements achieved through the proposed preprocessing method. The authors acknowledge certain limitations in their study, including the need for further validation across different battery chemistries and operating conditions. The paper includes visualizations, such as graphs and charts, to illustrate the performance differences between the models. These visual aids help to convey the effectiveness of the SOC-based preprocessing method in enhancing estimation accuracy. They suggest that future work could explore the integration of additional features and more complex models to further improve SOH estimation.

3.2 A Machine Learning-Based Robust State of Health (SOH) Prediction Model for Electric Vehicle Batteries - Khalid Akbar, Yuan Zou, Qasim Awais, Mirza Jabbar Aziz Baig and Mohsin Jamil

The paper outlines the significance of battery health monitoring in electric vehicles (EVs) due to the increasing reliance on battery systems for energy storage. The state of health (SOH) is a critical parameter that indicates the battery's condition and performance over time. Traditional methods for assessing SOH often involve costly and time-consuming experimental processes. This study proposes a data-driven ML approach to create a robust SOH prediction model that can serve as an alternative to these conventional methods. The methodology section describes the development of the ML model, which utilizes experimental battery data. The authors employed a systematic approach that included data collection, model development, and validation. High-quality battery testing data was gathered, which included various operational parameters. The ML model was constructed using advanced algorithms capable of learning from the data to predict SOH accurately. The model's predictions were validated against experimental results to ensure reliability and accuracy. The report emphasizes the use of a high-end computer system equipped with an NVIDIA GeForce GPU to facilitate the ML processes, ensuring efficient data processing and model training. The results section presents a thorough analysis of the model's performance. The model achieved an impressive accuracy rate of 99.98% with a mean squared error (MSE) of 0.03, indicating a high level of precision in predicting battery capacity and SOH. The study includes visualizations (Figures 3-8) that compare the predicted SOH and capacity against experimental data for multiple batteries (batteries 25, 26, and 27). The results demonstrate a strong correlation between the predicted and actual values, validating the model's effectiveness. The report discusses the error estimation methods used to assess the model's performance, highlighting the low average error of 0.02. The authors suggest avenues for future research, including the exploration of additional data sources and the integration of the model into real-time monitoring systems for enhanced battery management.

3.3 State of Health Estimation for Li-Ion Batteries Using Machine Learning Algorithms - Yunus Koc

The study begins by emphasizing the importance of SOH estimation for Li-Ion batteries, as their performance diminishes with usage. The paper outlines the need for accurate SOH estimation methods to optimize battery usage and prolong their lifespan. The author highlights the role of machine learning (ML) techniques in analyzing battery data to predict SOH effectively. The research utilizes a dataset from the Hawaii Natural Energy Institute (HNEI), which includes over 1000 charge/discharge cycles of 18650 Li-Ion batteries. The dataset provides various parameters, including charge and discharge rates, voltage, and capacity, under controlled conditions. The study focuses on two charging/discharging methods: Constant Current

(CC) and Constant Voltage (CV). The features for SOH estimation are categorized into long-term and short-term features. Long-term features require complete charge/discharge cycles, while short-term features can be derived from smaller portions of the charge/discharge curves. The paper discusses the significance of these features in the context of machine learning models. The study employs several regression methods to estimate SOH, including Random Forest Regressor, Decision Tree Regressor, Ridge Regressor, Bayesian Ridge Regressor etc. Each model is trained and tested using both long-term and short-term features, with a training/testing split of 75% and 25%, respectively. The performance of these models is evaluated based on their accuracy in predicting SOH. The results indicate that the Bayesian Ridge Regression model achieves the highest accuracy (98.2%) when using long-term features. In contrast, when short-term features are employed, the Random Forest Regression model performs best, demonstrating its adaptability to smaller datasets. The paper includes detailed performance metrics, showcasing the accuracy of each model under different feature sets. The study also calculates mutual information scores to assess the importance of various features in predicting SOH. Long-term features generally yield higher scores, indicating their relevance in accurate SOH estimation. However, some short-term features also show promising results, suggesting their potential for time-critical applications. The paper emphasizes the need for further research to define more efficient short-term features that can enhance accuracy without compromising speed.

4. Implementation Details

4.1 Dataset Overview

The NASA battery dataset, widely used for research on lithium-ion batteries, served as the foundation for this project. It contains data collected from multiple lithium-ion cells subjected to various charge-discharge cycles under controlled conditions. Each battery in the dataset was tested until it reached its end-of-life criteria, characterized by a significant capacity drop below 70-80% of its initial capacity. The dataset includes detailed measurements such as cycle number, capacity, voltage, current, SoH, and temperature, providing a comprehensive view of battery behavior over time. Individual batteries in the dataset exhibit varying degradation patterns due to differences in their operating conditions, such as charge-discharge rates and temperatures. These variations allow for the development and testing of predictive models that account for diverse real-world scenarios. By leveraging this dataset, the project aimed to create robust SoH prediction models capable of generalizing across different battery chemistries and usage conditions. Specific batteries from the NASA dataset—namely B05, B18, B33, B34, B46, B47, and B48—were selected for training and testing the SoH prediction models. These batteries were chosen to ensure diversity in the training data, as they exhibit varying degradation rates, charge-discharge characteristics, and operational conditions.

To provide a clearer understanding of the dataset, scatter plots illustrating the relationship between cycle count and State of Health (SoH) for battery B05 are presented below. These visualizations highlight the gradual decline in SoH as the cycle count increases, reflecting the battery's degradation over time. A similar decline is observed in all other batteries as well.

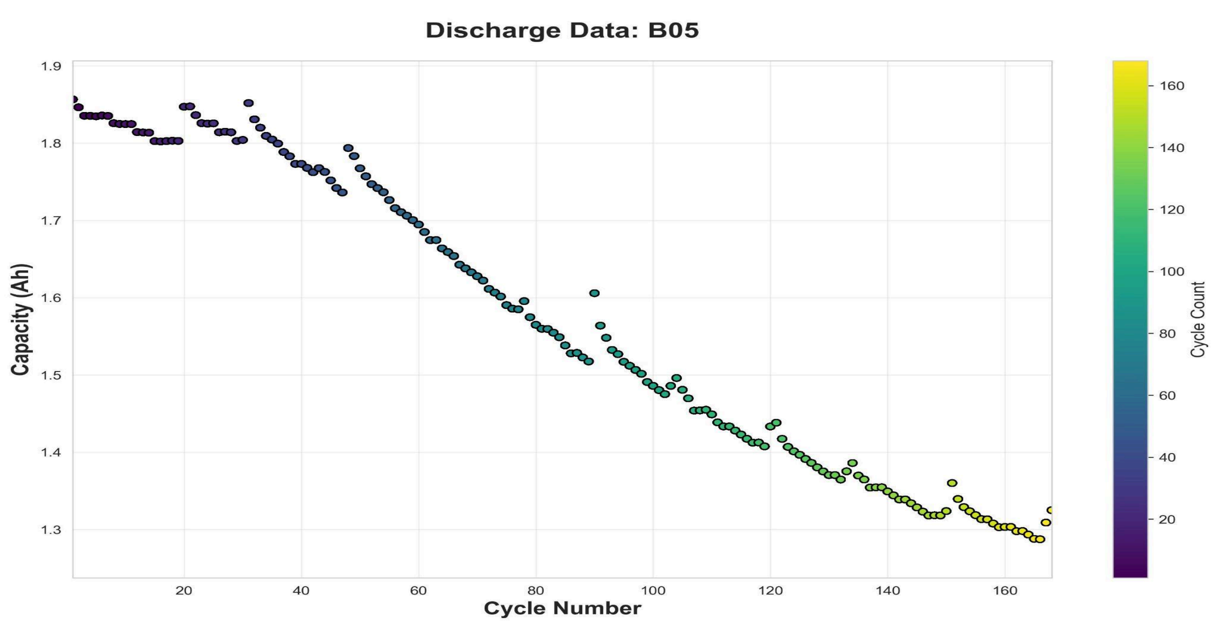

Fig 1. Capacity vs Cycle Number for batter B05

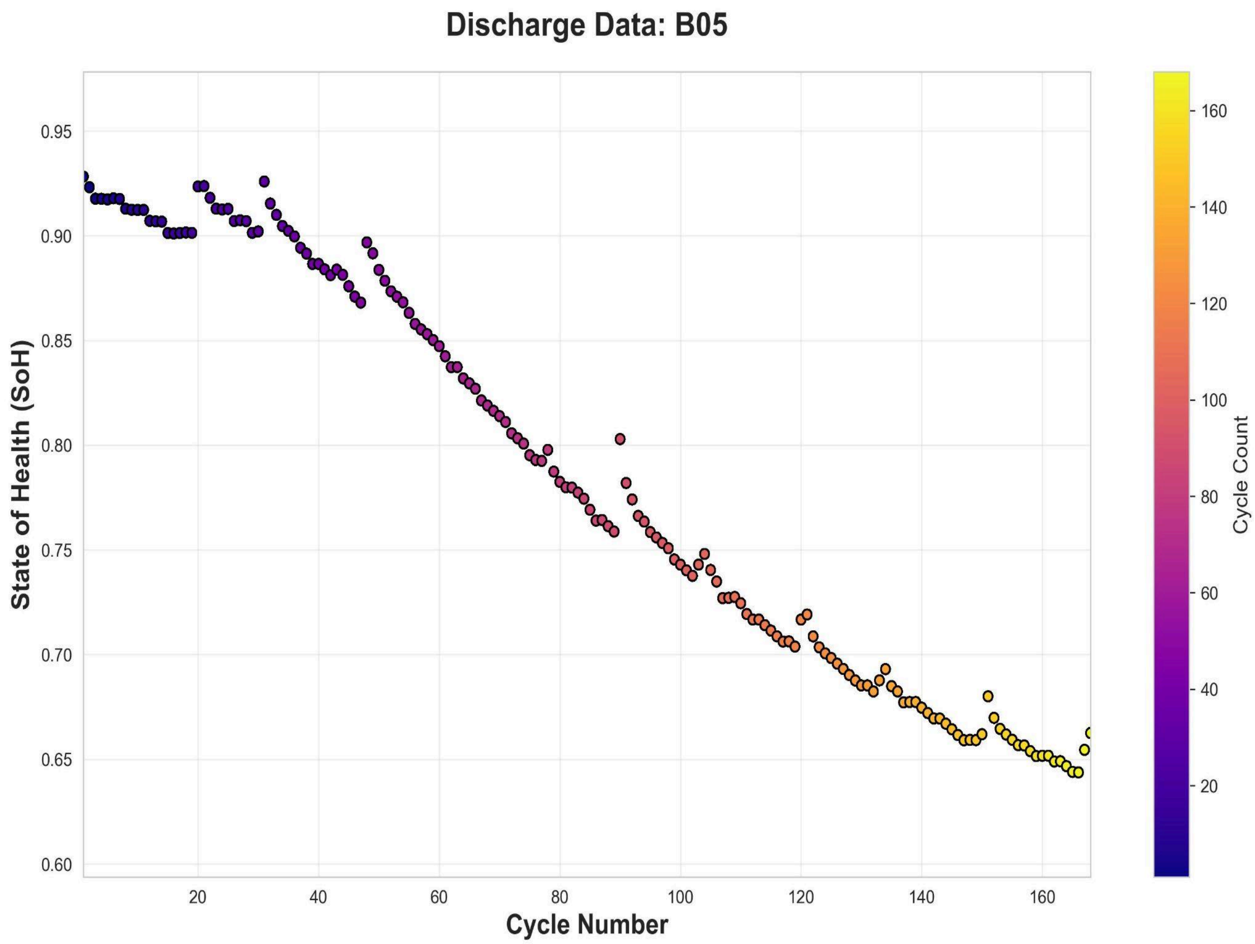

Fig 2. SoH vs Cycle Number for battery B05

4.2 Data Set Cleaning and Preprocessing

The data cleaning and preprocessing phase was critical to ensure the quality and reliability of the input data for accurate State of Health (SoH) prediction. The dataset, sourced from NASA's Prognostics Center of Excellence, consisted of multiple CSV files, each corresponding to a specific battery cell (e.g., B05, B07, B18, etc.). Each file included key attributes such as cycle, capacity, and SoH recorded across multiple charge-discharge cycles. The first step was to systematically load each file, extract the relevant columns, and organize them into individual DataFrames for each battery. This organization provided clarity in handling the data and allowed for battery-specific analysis. To identify any potential issues, an initial exploratory data analysis (EDA) was conducted, which highlighted patterns, trends, and anomalies in the data. Missing values, if present, were addressed either by imputation using interpolation methods or by removing the affected rows where appropriate.

The next step was to structure the data for model training. For the multivariate approach, where both cycle and capacity were used as input features, a combined data frame was prepared, with SoH designated as the target variable.

$$Multivariate\ approach{:}\quad SoH\ =\ f(cycle\ count,\ capacity)$$

As the data was measured on different scales, normalization was applied using the MinMaxScaler to transform all features to a range between 0 and 1. This is particularly useful when features have different scales, as it ensures that no feature dominates the learning process due to its magnitude, in this case, the cycle count had a magnitude in multiples of capacity. This ensured that cycle count did not dominate the learning process. Mathematically, the Min-Max Scaler transforms a feature x using the formula:

$$xscaled\ =\ \frac{x - xmin}{xmax - xmin}$$

For the univariate approach, the data was restructured to use only previous SoH values as the input feature, enabling the evaluation of its predictive power in isolation.

$$Univariate\ approach{:}\quad SoH\ =\ f(SoH\ values\ at\ previous\ times)$$

To account for temporal dependencies inherent in sequential data, a look-back window of 5 cycles was introduced for the CNN models. This meant that the model could consider data from the previous five cycles to predict the SoH for the next cycle. The input data was reshaped into a 3D array of the form (samples, time steps, features), which is required for such deep learning models.

Conversely, for the no-lookback technique, the data was flattened into 2D arrays of the form (samples, features) to evaluate the performance of simpler machine learning algorithms such as XGBoost and MLPs.It was observed that disregarding temporal dependencies resulted in significantly poorer performance compared to the approach that incorporated a lookback mechanism. Consequently, the lookback approach was adopted for further analysis and model development.

In both approaches, the dataset was split into training and testing subsets, with 80% of the data reserved for testing to ensure robust model evaluation.

4.3 Model Development for multivariate approach

The multivariate approach for predicting the State of Health (SoH) incorporates a combination of features—capacity, cycle count, and SoH itself—to better understand and predict the degradation patterns in batteries. By leveraging multiple input features, this approach captures complex interdependencies that influence battery health, aiming for improved prediction accuracy. Each of these features plays a critical role: capacity reflects the battery's charge-holding ability, which declines over time; cycle count represents the cumulative number of charge-discharge cycles, offering insight into the battery's usage; and SoH, the target variable, quantifies the battery's overall health. Before training, these features are normalized using Min-Max scaling to bring them into a consistent range, thus facilitating faster convergence and better model performance.

The model architecture is based on a Convolutional Neural Network (CNN), structured to handle multivariate time-series data with a look-back mechanism. The input layer is designed to accept data shaped as (samples, look_back, features), where "look_back" refers to the number of previous time steps considered for prediction. For instance, with a look-back of 3, the input dimensions are (3,3) —three time steps and three features. The first layer of the model is a 1D convolutional layer with 64 filters and a kernel size of 2, which extracts patterns and correlations across the input sequence. ReLU activation is applied to introduce non-linearity, enabling the model to learn intricate data relationships. This is followed by a max-pooling layer with a pooling size of 2, which reduces the spatial dimensions of the feature maps while retaining critical information, thereby making the model computationally efficient. The flattened feature maps are then passed through a dense layer comprising 50 neurons with ReLU activation to learn complex feature representations. Finally, a single-neuron output layer predicts the SoH value.

The model employs the Mean Absolute Error (MAE) as the loss function, optimizing it using the Adam optimizer, which dynamically adjusts learning rates for faster and more stable convergence.This CNN configuration strikes a balance between complexity and performance, enabling accurate predictions while being computationally manageable.

4.4 Model Development for univariate approach

The univariate approach focuses on modeling a single input feature, making it particularly suitable for datasets where a primary variable, such as capacity, exhibits a strong correlation with the target variable, SoH. For this approach, a Convolutional Neural Network (CNN) was implemented to capture patterns and dependencies in the data. CNNs are powerful tools for extracting local features from sequences, even for univariate datasets.

The architecture was designed to process the data in a time-series format, ensuring that the temporal structure of the input data was preserved. The input data was reshaped into the format of (samples, look-back, features). The model starts with a 1D convolutional layer with 64 filters and a kernel size of 1, enabling it to capture feature-level patterns in the data effectively. The ReLU activation function was employed in this layer to introduce non-linearity and enhance the model's ability to learn complex relationships.

The convolutional layer is followed by a max-pooling layer with a pool size of 1, which reduces the dimensionality while retaining the most relevant information from the features. Afterward, the data is flattened to prepare it for fully connected layers. The first dense layer consists of 50 neurons with ReLU activation, helping to learn higher-level representations of the data. The final output layer is a single neuron designed to predict the SoH value directly. The model was compiled with the Adam optimizer for efficient gradient-based optimization and a mean absolute error (MAE) loss function, which is well-suited for regression problems.

4.5 Model- Training Details

The training process for the CNN model in both approaches was conducted with meticulous attention to data preparation and parameter tuning. The dataset was divided into training and testing sets using an 80-20 split. This split ensured that the model was trained on a substantial portion of the data while preserving enough unseen samples for robust evaluation. The training data was used to optimize the model's parameters, while the test data served to assess the generalization performance.

Key training parameters included a batch size of 20, allowing the model to process small chunks of data efficiently and enabling stable gradient updates. The model was trained for 100 epochs, a sufficient number to allow convergence while avoiding overfitting. The shuffled approach was disabled to maintain the temporal order of the data, preserving any inherent sequential patterns crucial for accurate SoH prediction.

4.6 Evaluation Metrics

In this study, three evaluation metrics were utilized to assess the performance of the model: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Median Absolute Error. Each metric provides unique insights into the model's predictive accuracy, helping evaluate its performance comprehensively.

Root Mean Squared Error (RMSE) is calculated as the square root of the average squared differences between the predicted and actual values. It is particularly sensitive to large errors, making it suitable for identifying outliers and quantifying the magnitude of prediction errors in high-impact cases. RMSE is expressed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{y_i}\right)^2}$$

Here y_i is the actual value, $\widehat{y_i}$ is the predicted value, and N is the total number of data points.

Mean Absolute Error (MAE) measures the average magnitude of errors without considering their direction, making it easy to interpret. Unlike RMSE, it is less sensitive to outliers, providing a balanced view of the overall prediction error. It is given by:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \widehat{y_i}\right|$$

Here y_i is the actual value, $\widehat{y_i}$ is the predicted value, and N is the total number of data points.

Median Absolute Error calculates the median of absolute differences between predicted and actual values. By focusing on the median, it is robust against outliers, offering a reliable measure of central tendency for errors.

5. Results

5.1 Multivariate Approach

5.1.1 Predicted vs Actual Graphs

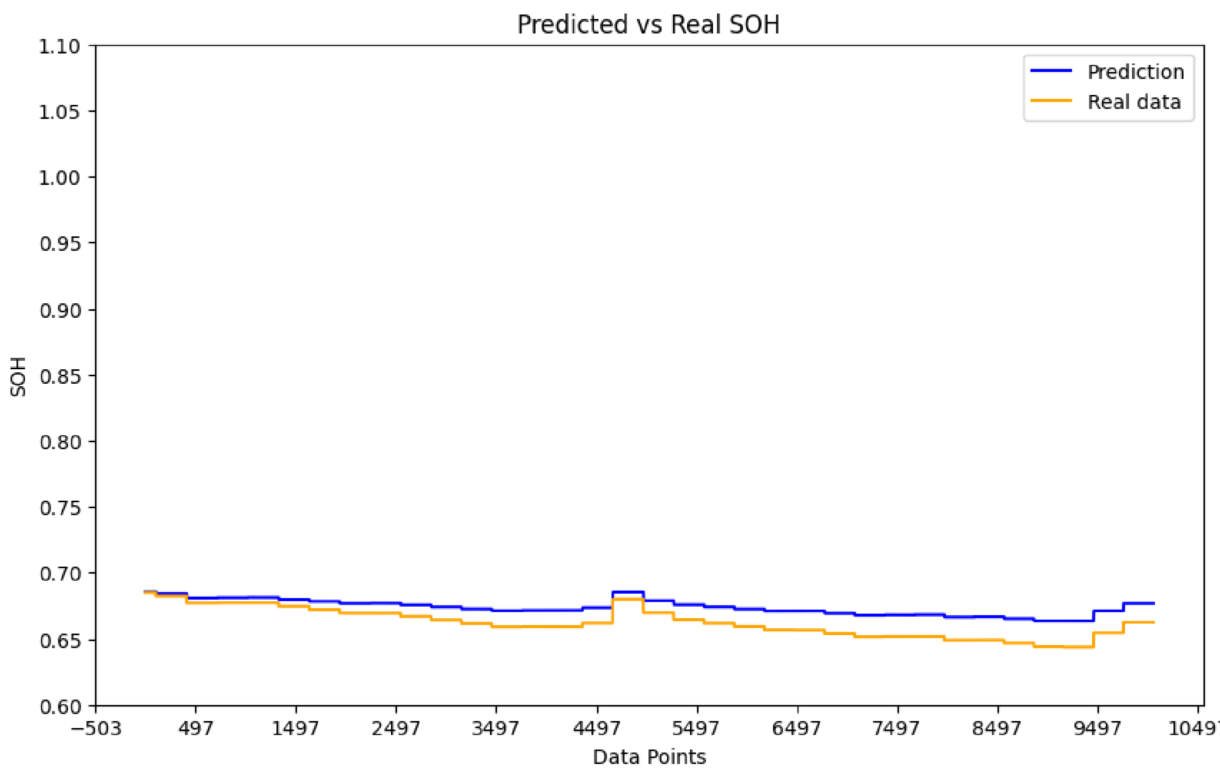

Fig 3. SoH Predicted vs Actual graph for B05

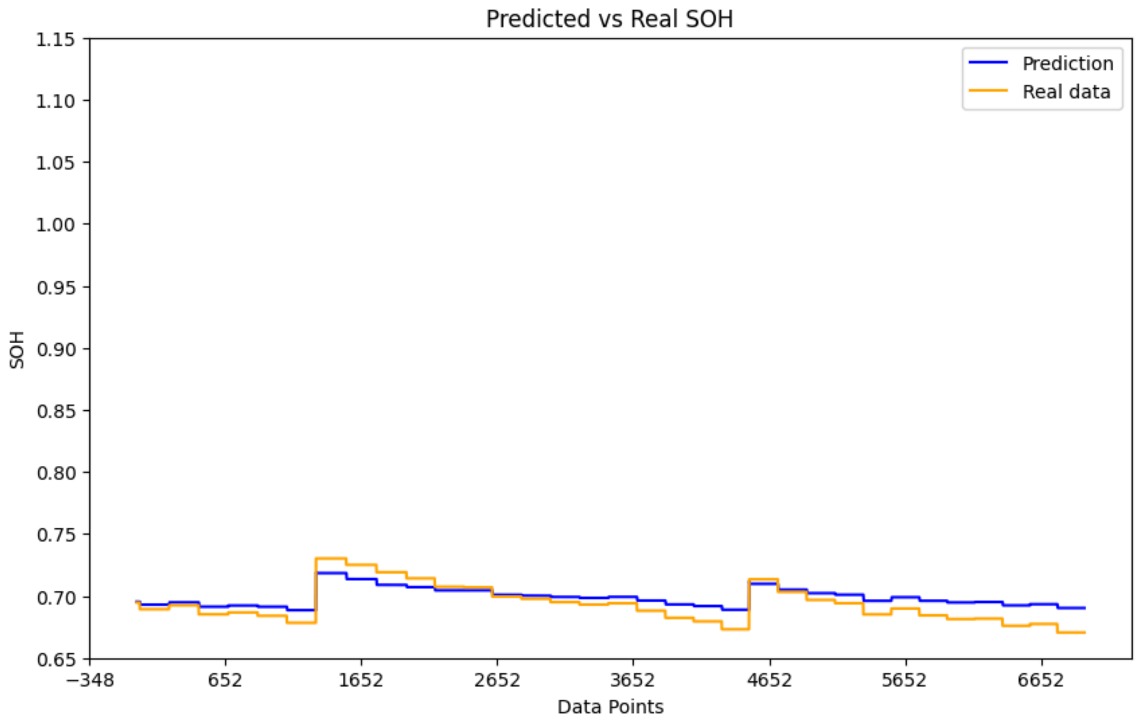

Fig 4. SoH Predicted vs Actual for B18

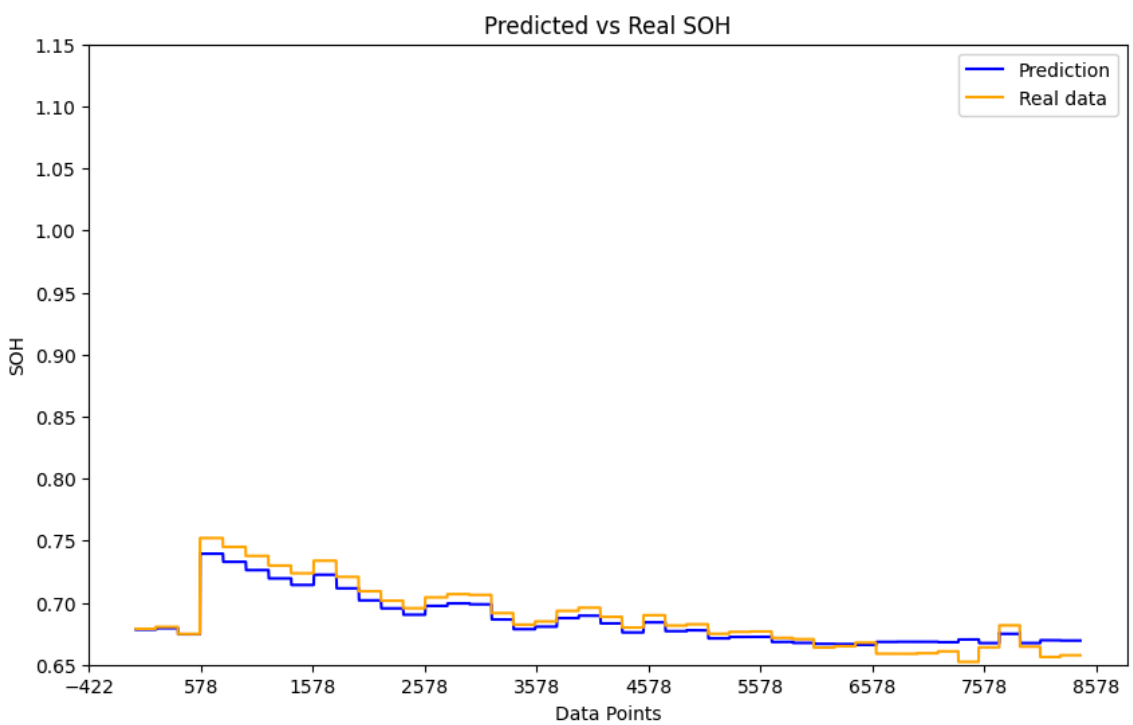

Fig 5. SoH Predicted vs Actual for B33

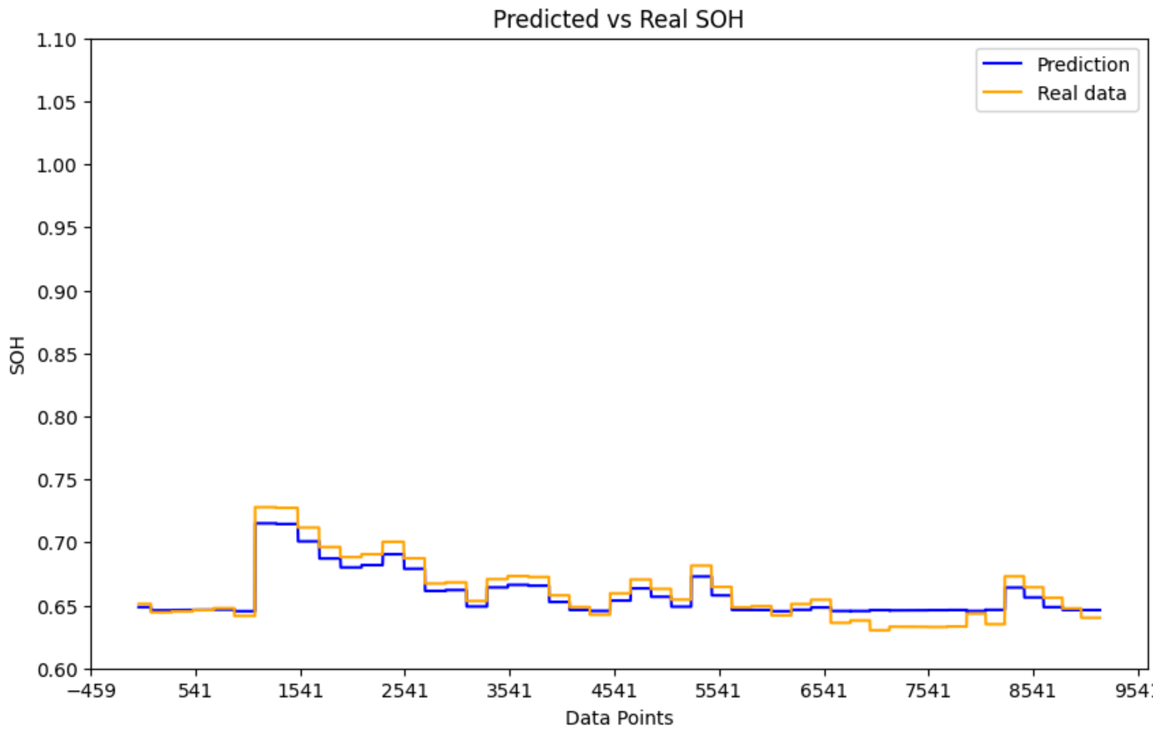

Fig 6. SoH Predicted vs Actual for B34

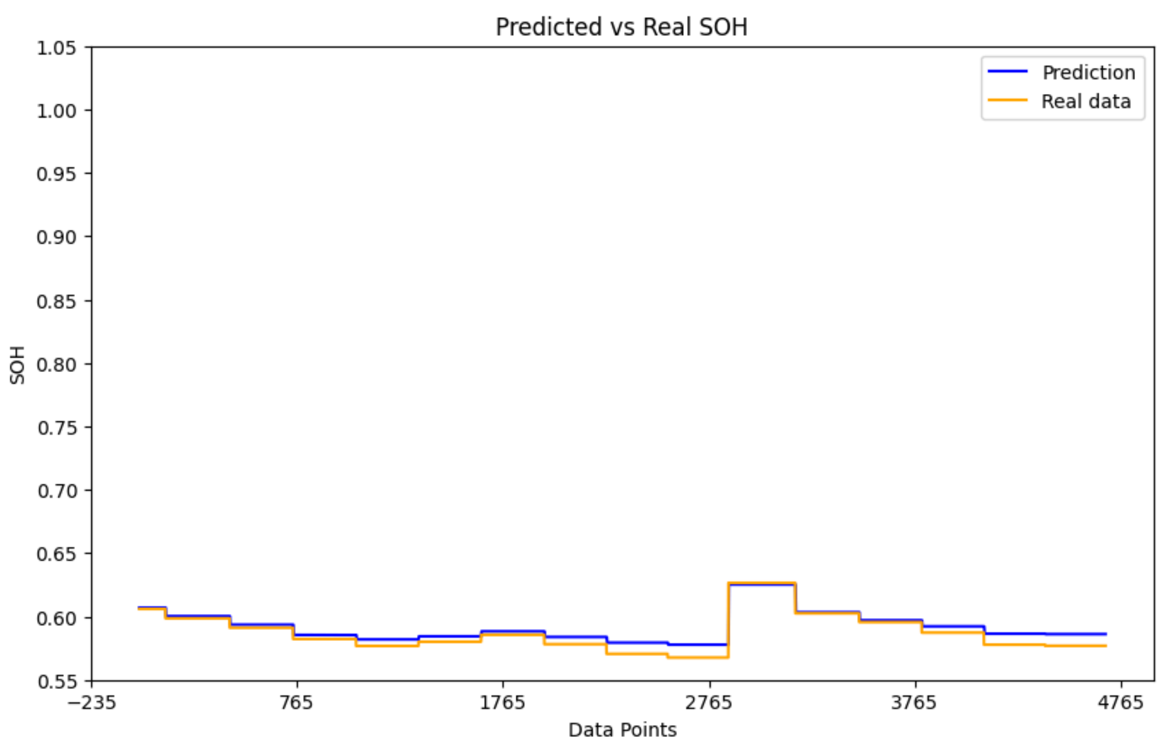

Fig 7. SoH Predicted vs Actual for B46

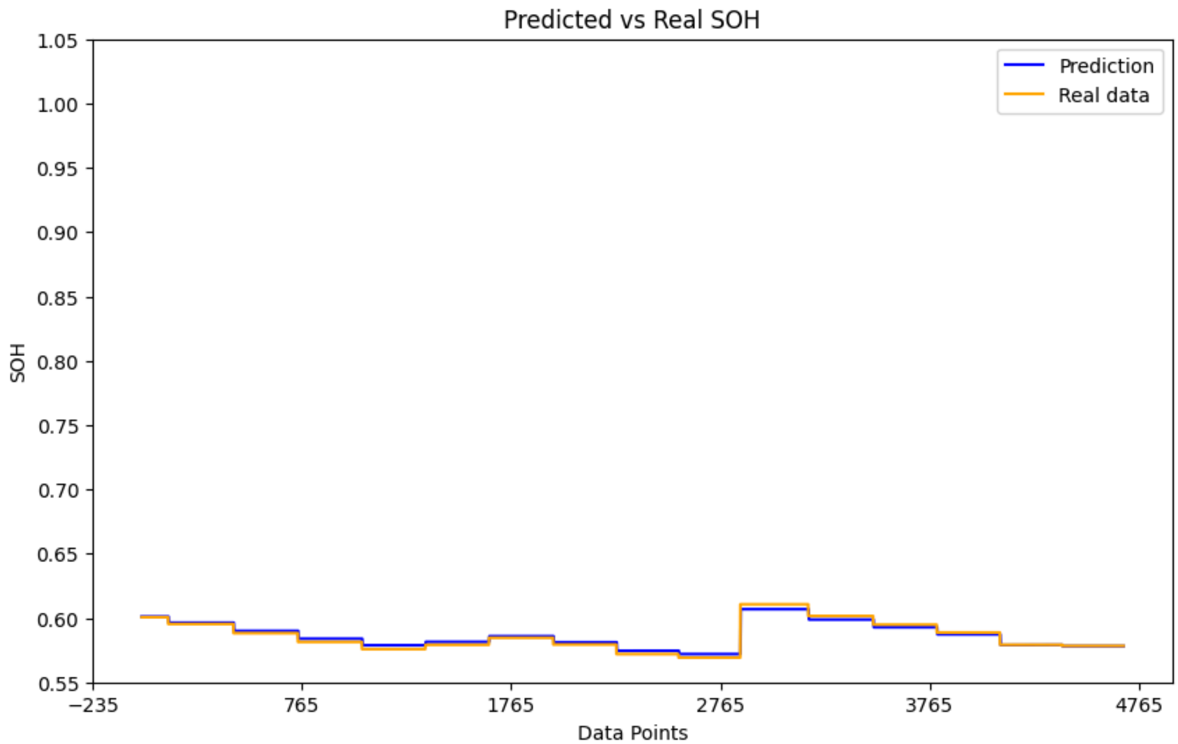

Fig 8. SoH Predicted vs Actual for B47

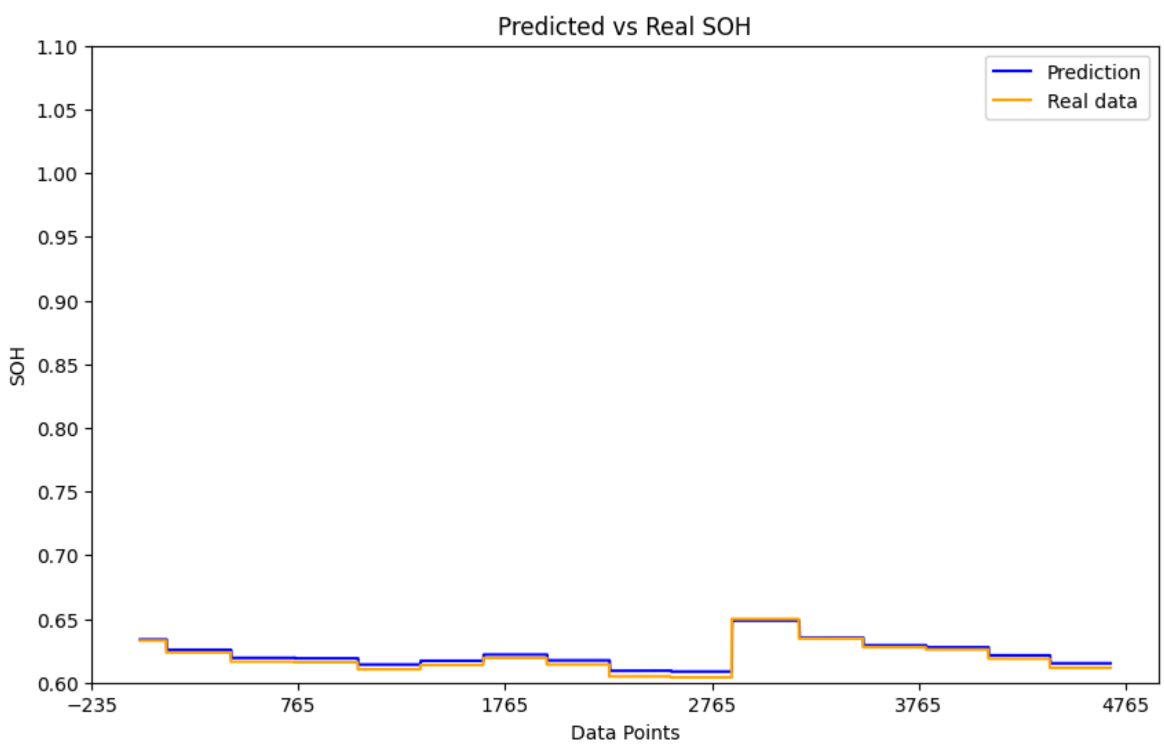

Fig 9. SoH Predicted vs Actual for B48

5.1.2 Performance Metrics Evaluation

S.No	Battery No	RMSE	MAE	Median Absolute Error
1.	B05	0.013	0.012	0.012
2.	B18	0.010	0.008	0.007
3.	B33	0.008	0.007	0.006
4.	B34	0.008	0.007	0.007
5.	B46	0.006	0.005	0.004
6.	B47	0.002	0.002	0.002
7.	B48	0.003	0.003	0.003

Table: Results obtained after running the model for multivariate approach on dataset

5.2 Univariate Approach

5.2.1 Predicted vs Actual Graphs

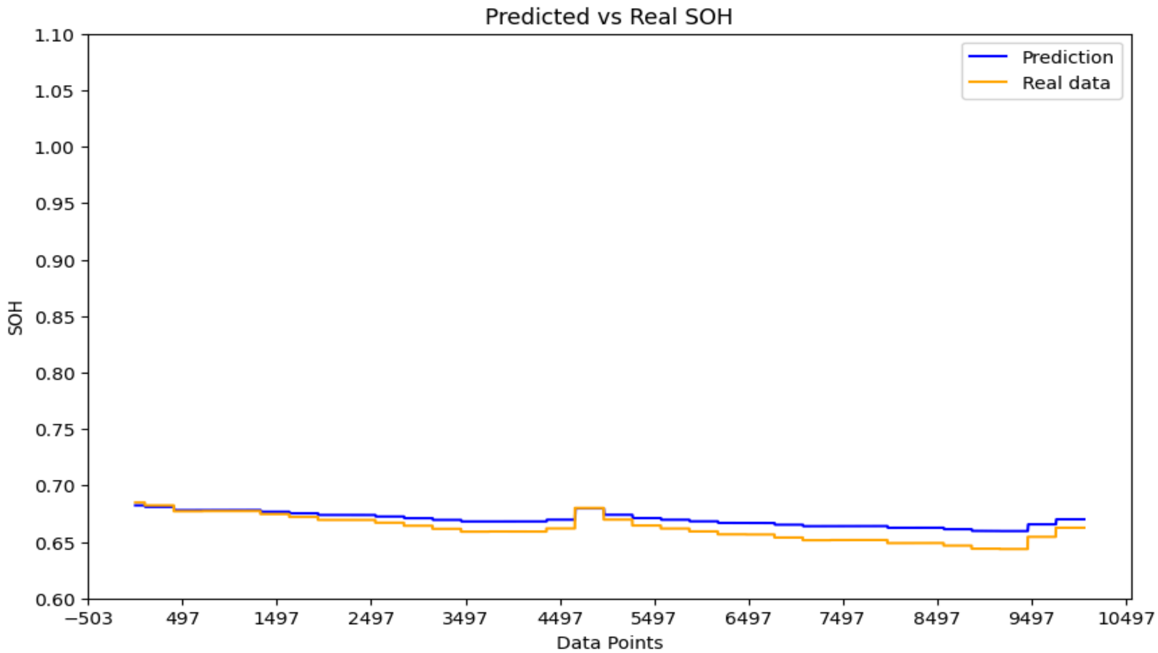

Fig 10. SoH Predicted vs Actual for B05

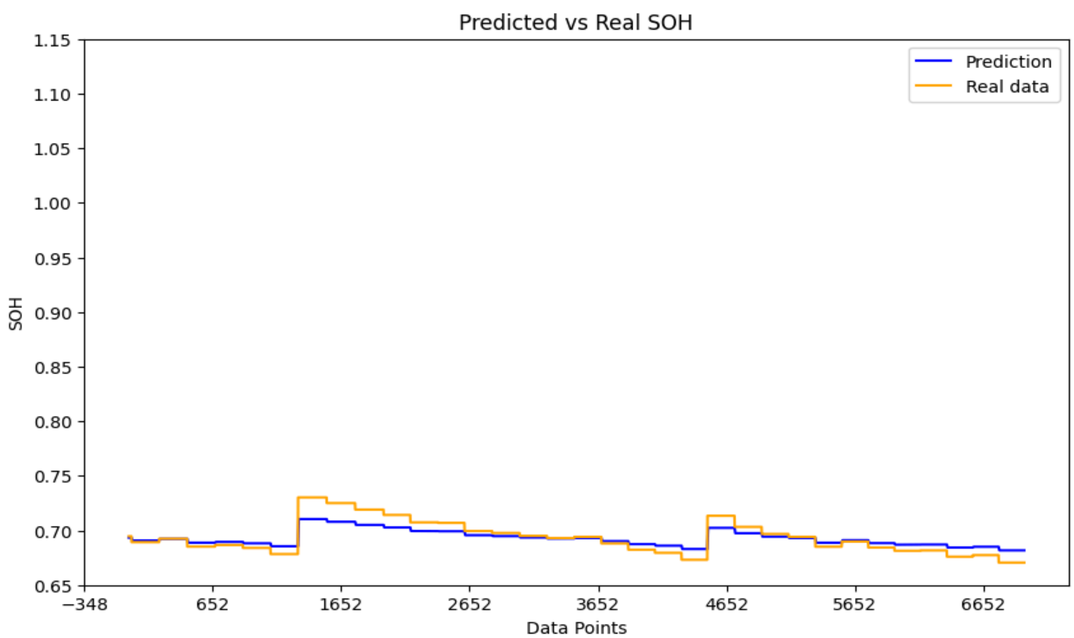

Fig 11. SoH Predicted vs Actual for B18

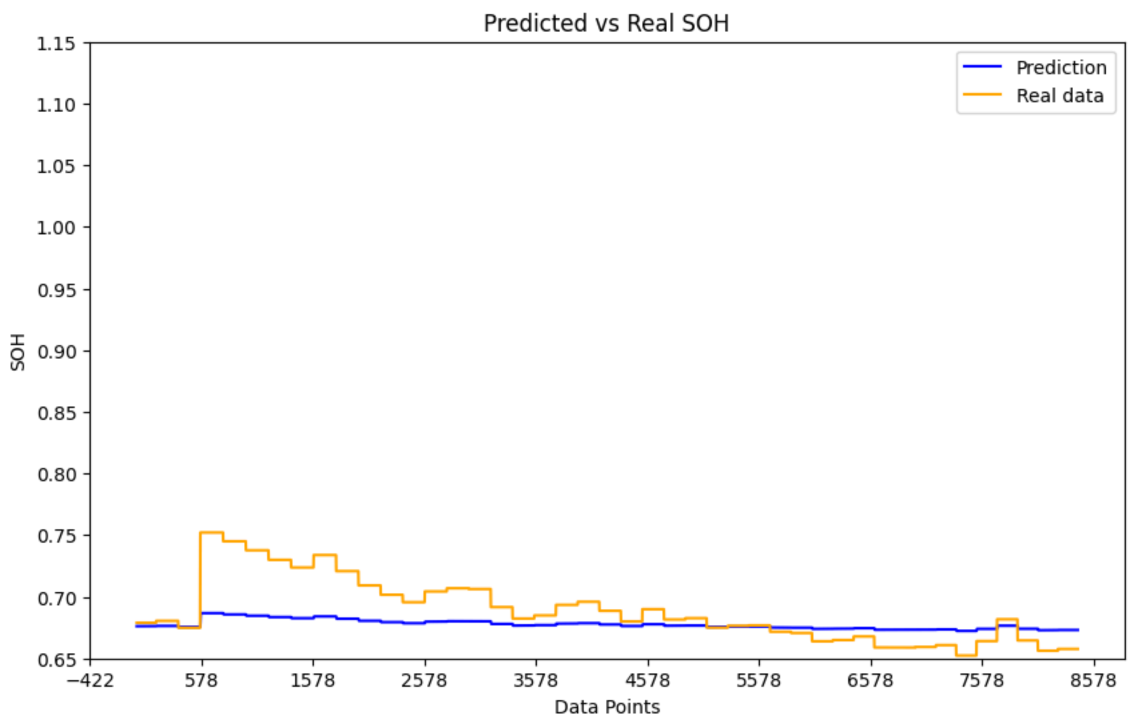

Fig 12. SoH Predicted vs Actual for B33

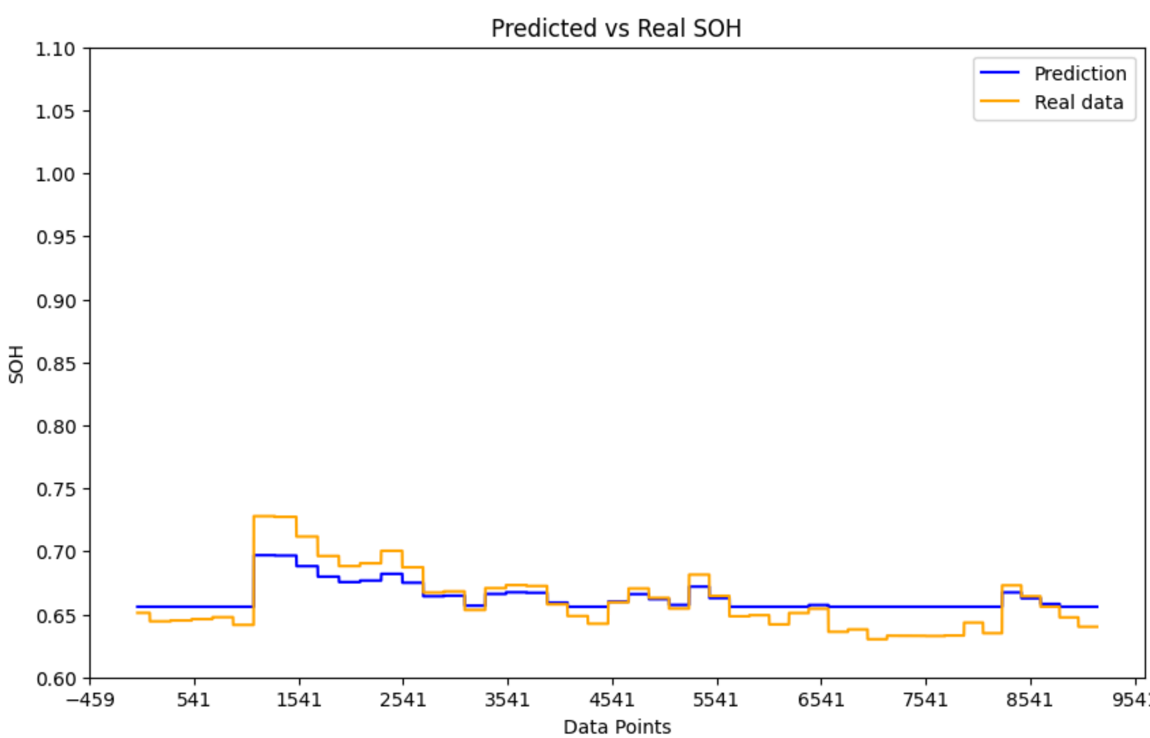

Fig 13. SoH Predicted vs Actual for B34

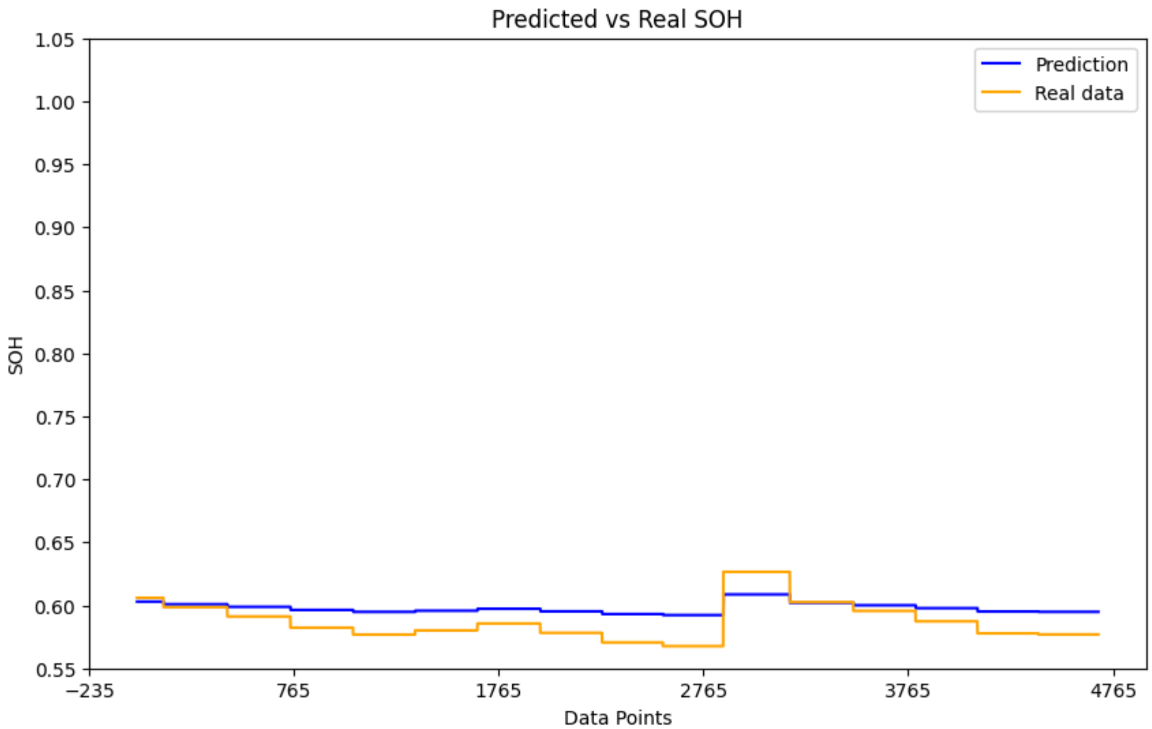

Fig 14. SoH Predicted vs Actual for B46

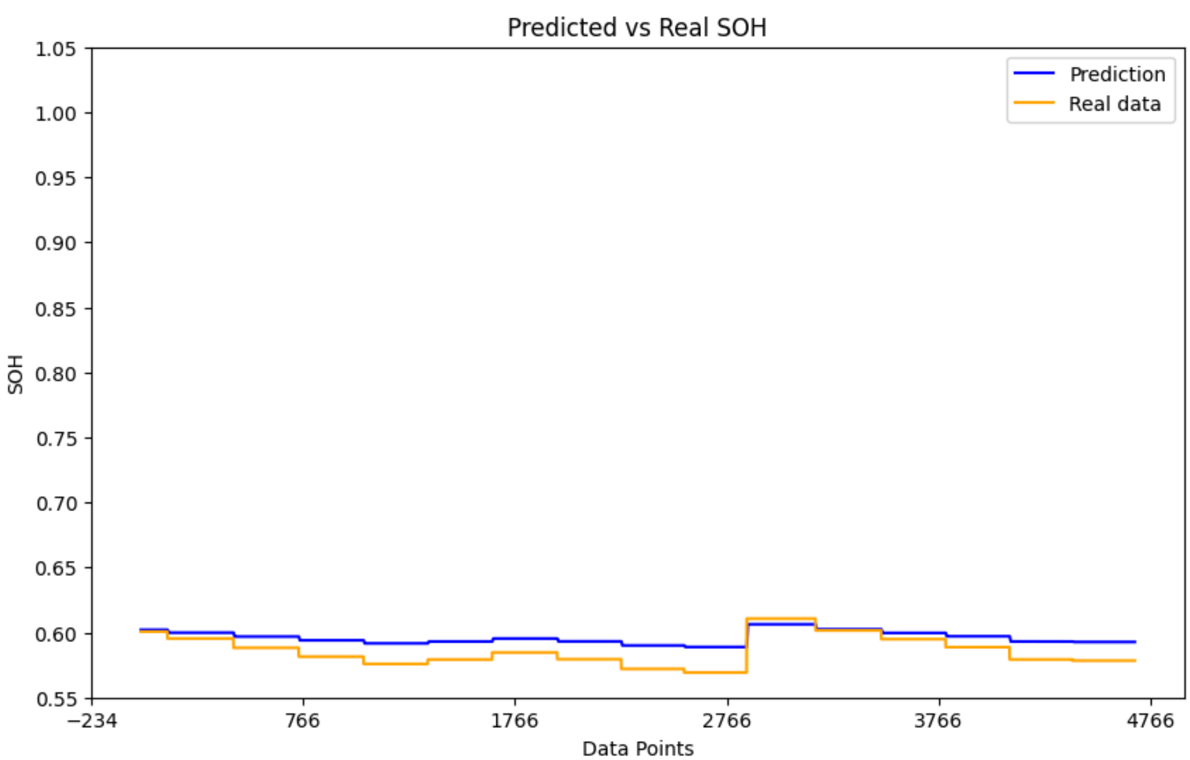

Fig 15. SoH Predicted vs Actual for B47

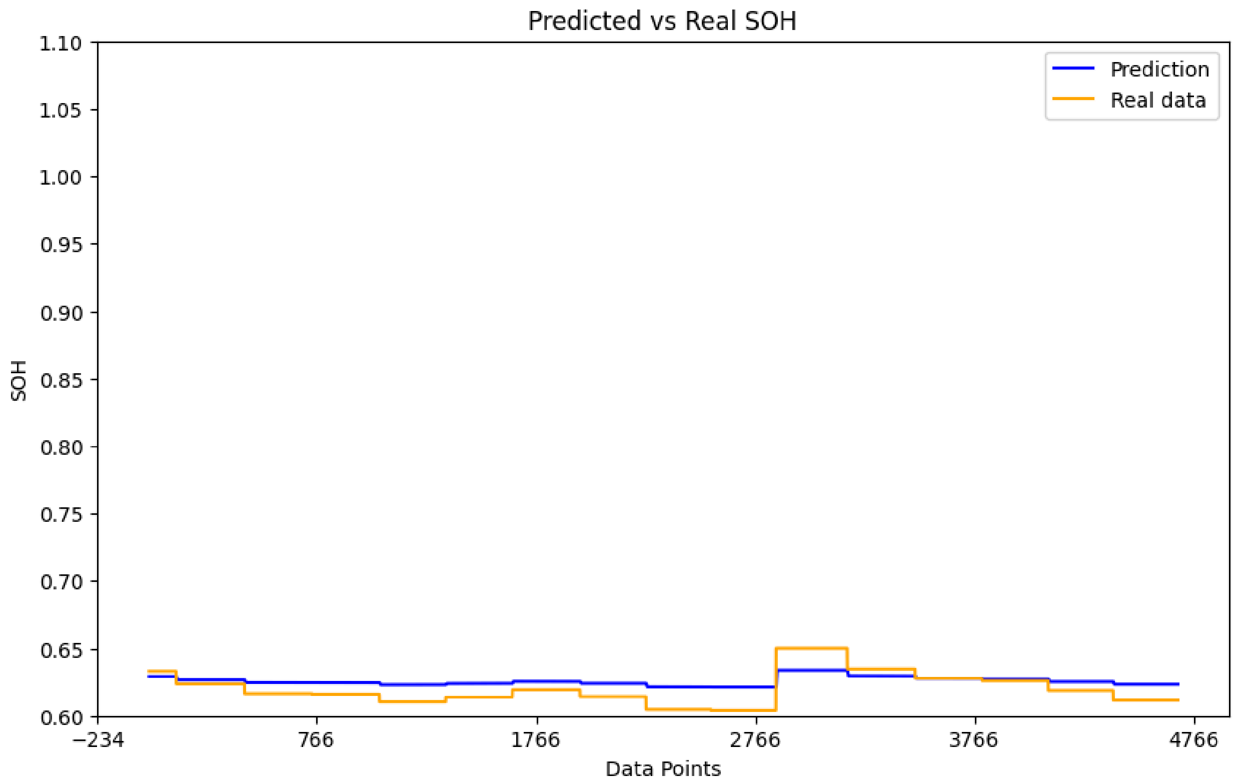

Fig 16. SoH Predicted vs Actual for B48

` **5.2.2 Performance Metrics Evaluation (Univariate approach)**

S.No.	**Battery No.**	**RMSE**	**MAE**	**Median Absolute Error**
1.	B05	0.009	0.008	0.008
2.	B18	0.008	0.006	0.005
3.	B33	0.025	0.018	0.014
4.	B34	0.014	0.011	0.010
5.	B46	0.015	0.013	0.016
6.	B47	0.012	0.011	0.013
7.	B48	0.010	0.009	0.009

Table: Results obtained after running the model for univariate approach on dataset

5.3 Observations and Analysis

From the performance metrics recorded for both the multivariate and univariate approaches, it is evident that there are notable differences in the predictive accuracy of the two methods across various batteries. The metrics considered—Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Median Absolute Error—serve as robust indicators of the models' precision.

For the **multivariate approach**, the RMSE values generally range between 0.002 and 0.013, MAE values between 0.002 and 0.012, and Median Absolute Error values between 0.002 and 0.012. The smaller errors for batteries like B47 and B48 indicate superior predictions, highlighting the multivariate model's ability to effectively utilize additional features such as cycles and other covariates. On the other hand, batteries like B05 and B18 show slightly higher error values, which may indicate variations in battery characteristics that were harder to capture even with multivariate inputs.

In the **univariate approach**, the RMSE values span a broader range, from 0.008 to 0.025, with batteries like B33 experiencing higher errors. MAE values also show a similar trend, ranging between 0.006 and 0.018, and the Median Absolute Error reflects comparable variability. The higher error values in the univariate approach compared to the multivariate approach suggest that relying solely on the historical state of health (SoH) values without additional contextual features is less effective in predicting future states. Batteries like B05 and B18 still maintain relatively low errors in the univariate setup, indicating some robustness in prediction for those cases.

When comparing both approaches, it is clear that the multivariate model consistently outperforms the univariate model across all batteries. For instance, the RMSE for B33 in the multivariate setup is 0.008, while it rises to 0.025 in the univariate case—a substantial difference. Similarly, the MAE and Median Absolute Error values for all batteries are consistently lower in the multivariate approach. This demonstrates the advantage of incorporating additional input features in capturing the underlying patterns in battery degradation more accurately.

The results underline the importance of leveraging multiple covariates in predictive modeling, particularly for applications where temporal dependencies and feature interactions significantly impact the outcome. While the univariate approach provides a simpler setup and is computationally less intensive, its limitations in accuracy make it less suitable for applications requiring precise predictions, such as state-of-health estimation for lithium-ion batteries. The multivariate approach demonstrates better overall performance and is the recommended method for reliable SoH prediction in this context.

6. Conclusion

This study explored and compared the performance of univariate and multivariate approaches for predicting the state of health (SoH) of lithium-ion batteries using convolutional neural networks (CNN). The multivariate approach, which leveraged multiple input features such as cycle count, capacity, and historical SoH values, consistently outperformed the univariate model, which relied solely on past SoH values. Metrics such as RMSE, MAE, and Median Absolute Error demonstrated that the inclusion of additional features significantly enhanced predictive accuracy.

The findings of this report hold practical relevance for real-world applications, particularly in electric vehicles, renewable energy storage systems, and portable electronics, where accurate SoH estimation is crucial for safety, performance optimization, and cost management. By enabling more reliable battery health predictions, the multivariate approach can contribute to extending battery life and reducing waste in sustainable energy ecosystems.

Future work could involve exploring more advanced architectures or transformers to better capture temporal dependencies. Additionally, incorporating more diverse datasets, including those with environmental factors like temperature, could improve the generalizability of the model. Finally, developing lightweight, real-time implementations of the proposed approach would further enhance its utility in industrial and consumer applications.

References

1. *Akbar, K., Zou, Y., Awais, Q., Baig, M. J. A., & Jamil, M. (2022). A Machine Learning-Based Robust State of Health (SOH) prediction model for electric vehicle batteries. Electronics, 11(8), 1216. https://doi.org/10.3390/electronics11081216*
2. *Aloisio, D., Campobello, G., Leonardi, S. G., Sergi, F., Brunaccini, G., Ferraro, M., Antonucci, V., Segreto, A., Donato, N., Institute of Advanced Energy Technologies "Nicola Giordano", National Research Council of Italy, & University of Messina, Department of Engineering. (2020). A machine learning approach for evaluation of battery state of health. In 24th IMEKO TC4 International Symposium, 22nd International Workshop on ADC and DAC Modelling and Testing [Conference-proceeding]. https://www.imeko.org/publications/tc4-2020/IMEKO-TC4-2020-25.pdf*
3. *Jo, S., Jung, S., & Roh, T. (2021). Battery State-of-Health Estimation Using Machine Learning and Preprocessing with Relative State-of-Charge. Energies, 14(21), 7206. https://doi.org/10.3390/en14217206*
4. *State of Health Estimation using Machine Learning for Li-ion battery on Electric Vehicles. (2021, October 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9699273*
5. *Jha, A., Annamalai, K. R., Varshini, C. R. A., Tiwari, A., Deepa, K., Sailaja, V., & Department of Electrical and Electronics Engineering, Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham, India. (2021). Survey on Estimation Methods for EV Battery Health using ML Techniques. In Department of Electrical and Electronics Engineering, Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham,*

India (pp. 1–7) [Journal-article]. https://www.riverpublishers.com/pdf/ebook/chapter/RP_P9788770229630C24.pdf

6. *Zhang, M., Yang, D., Du, J., Sun, H., Li, L., Wang, L., & Wang, K. (2023). A Review of SOH Prediction of Li-Ion Batteries Based on Data-Driven Algorithms. Energies, 16(7), 3167. https://doi.org/10.3390/en16073167*
7. Lin, J., Yan, G., & Wang, C. (2021). Li-ion battery state of health Prediction based on Long Short-Term Memory Recurrent Neural Network. *Journal of Physics Conference Series*, *2010*(1), 012133. https://doi.org/10.1088/1742-6596/2010/1/012133

Appendix

Link to the code implementation and the dataset required to run the code:

> https://drive.google.com/drive/folders/1d3ZFznVpREzTgo1vgLCugasyB_EHgUvG?usp=drive_link