

CHAPTER III

EVALUATING THE ROLE OF OCV IN LSTM MODEL

3.1 Introduction

Estimating the degradation of a battery, often referred to as the SOH, is a crucial step for monitoring its performance over time. As a battery is used, its ability to store charge gradually decreases, which directly affects its overall performance and lifespan. Accurate SOH estimation is therefore essential for effective energy management in electrical devices and electric vehicles. Currently, SOH estimation often relies on machine learning techniques applied to data collected during the charging and discharging processes. The accuracy of these models strongly depends on the quality of the input features. Features that effectively capture the electrical behavior of the battery can help the model learn the degradation patterns more accurately. This chapter focuses on comparing the performance of models using features derived from the OCV against models using standard features alone. The goal is to investigate whether including OCV-related features can improve the accuracy of SOH estimation. The study uses data from the Oxford Battery Dataset, a widely recognized dataset containing detailed charge and discharge records of LIBs. Features extracted include the AUC_v and the time taken for each charge–discharge cycle, along with features related to OCV. These features are then used to evaluate the performance of the model. The results show that incorporating OCV-related features improves the accuracy of SOH estimation compared to using only basic features. This highlights that OCV is an important parameter reflecting battery degradation and can be effectively used to enhance model performance for SOH estimation. Figure 3.1 shows the proposed estimation approach framework.

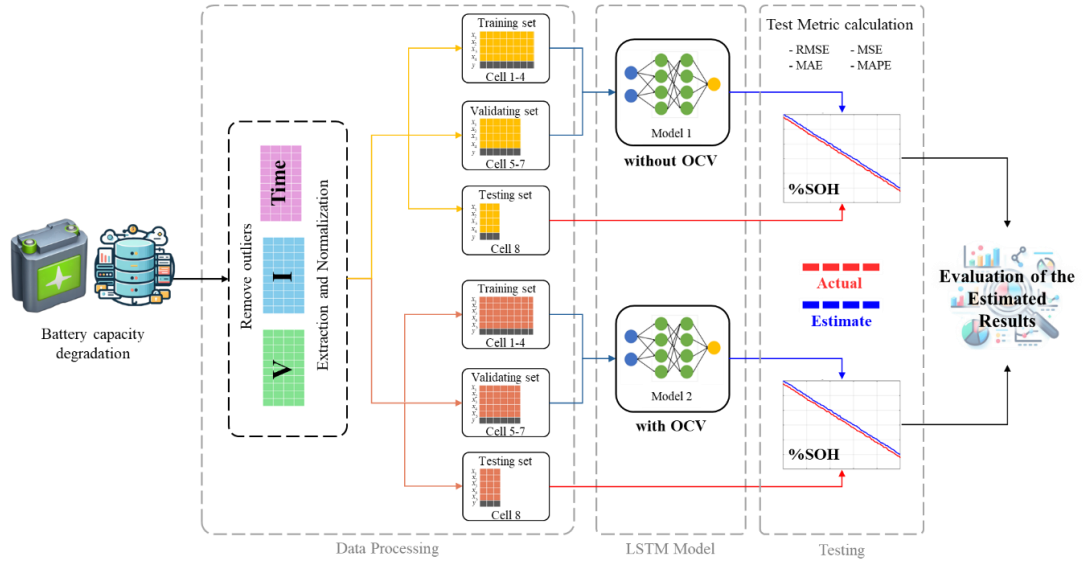


Figure 3.1 Framework of comparative analysis between LSTM models with and without OCV feature inclusion

3.2 Battery Datasets Utilized

As discussed in Section 3.2, various open-access battery datasets have been proposed for SOH estimation. In this chapter, the focus is placed solely on the Oxford Battery Degradation Dataset.

3.2.1 Oxford Battery Degradation Dataset

The Oxford Battery Degradation Dataset 1 contains ageing data from eight small Kokam SLPB533459H4 Li-ion pouch cells, each with a nominal capacity of 740 mAh. The test subjects consist of 8 × Kokam CO LTD, SLPB533459H4 cells. All tests were conducted in a thermal chamber maintained at 40 °C. The cells underwent a constant-current–constant-voltage (CC–CV) charging profile, followed by a drive cycle discharging profile derived from the urban Artemis driving pattern. Characterisation measurements, including 1-C charge/discharge and pseudo-OCV tests, were recorded every 100 drive cycles. In this study, the 1-C charge/discharge test data were selected for analysis and visualization. Each 1-C test was conducted using a constant current of 740 mA. The sampling rate was 1 second per measurement point. Each measurement

includes the following parameters: time (t , in seconds), voltage (v , in Volts), charge (q , in mAh), and temperature (T , in degrees Celsius). However, there are data recording mistakes or electrical noise segments in the dataset. Therefore, before doing feature extraction for additional analysis, data cleaning and preprocessing are necessary. From the recorded data during the charging and discharging processes, the voltage and current profiles for selected cycles are plotted. Figures 3.2 and 3.3 show the voltage and current profiles during the charging process, respectively, while Figs. 3.4 and 3.5 present the corresponding profiles during discharging. These plots illustrate the typical behavior of the battery under the applied test conditions and highlight the fundamental characteristics of the dataset. However, to understand how these profiles evolve over multiple cycles in relation to the battery's SOH, further analysis is provided in the Feature Extraction section.

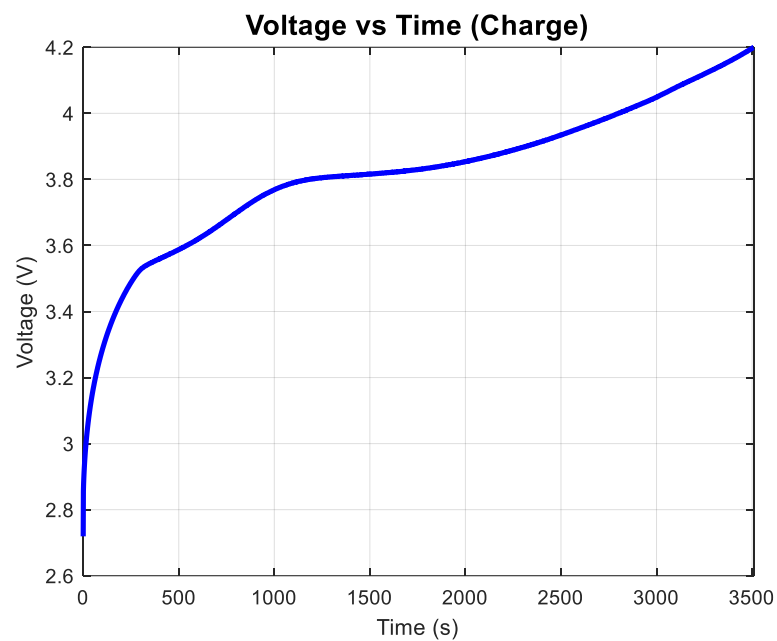


Figure 3.2 Voltage profiles of Cell 1 during charging cycle from the Oxford Battery Dataset

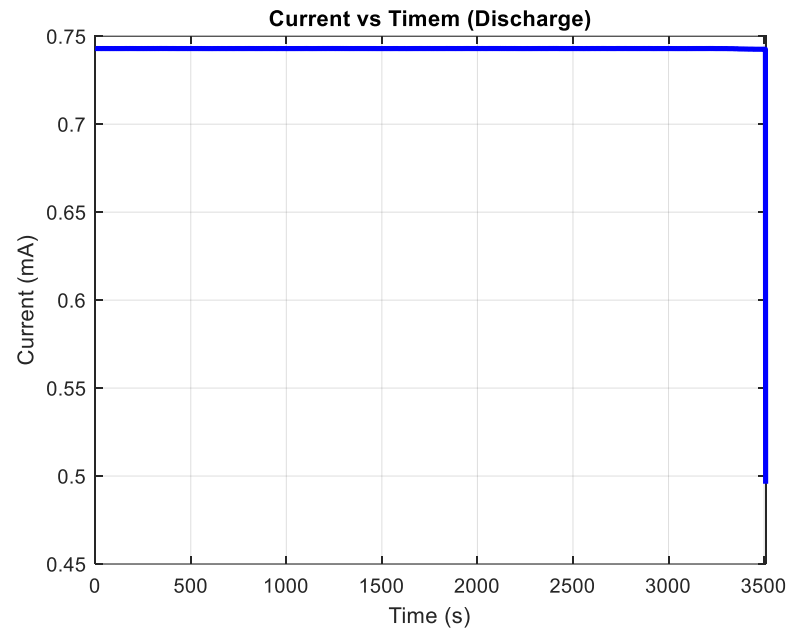


Figure 3.3 Current profiles of Cell 1 during charging cycle from the Oxford Battery

Dataset

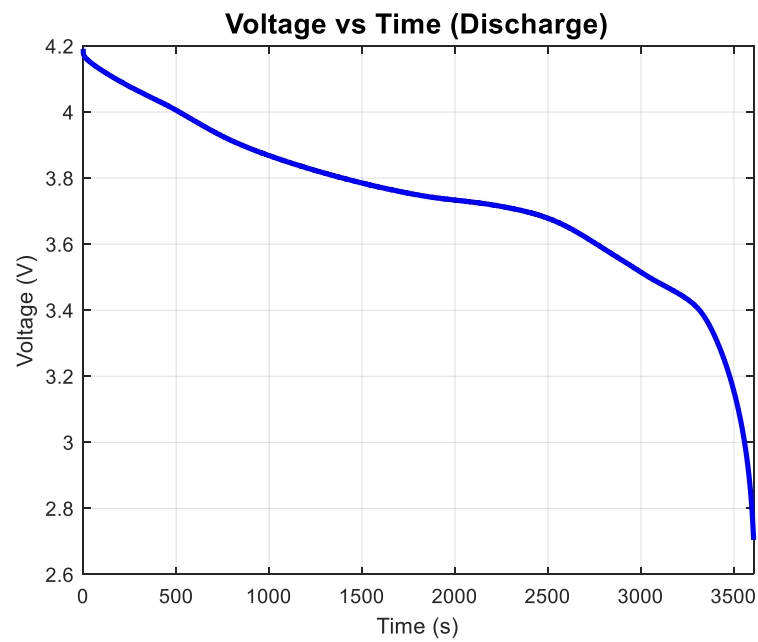


Figure 3.4 Voltage profiles of Cell 1 during discharging cycle from the Oxford

Battery Dataset

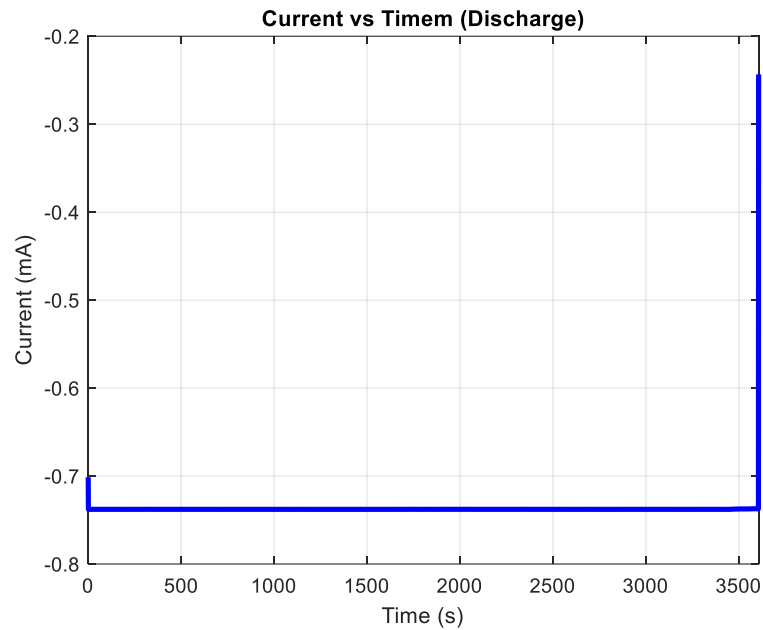


Figure 3.5 Current profiles of Cell 1 during discharging cycle from the Oxford Battery Dataset

During the charging and discharging processes, voltage cut-off limits are defined to ensure safety and stability. In the charging cycle, the voltage increases from 2.7 V to 4.2 V, whereas in the discharging cycle, it decreases from 4.2 V to 2.7 V.

To better understand the degradation behavior of the battery over time, the SOH of each cell was plotted as a function of the number of charge-discharge cycles. Figure 3.6 helps highlight the gradual decline in capacity and provides insight into the ageing characteristics under repeated cycling.

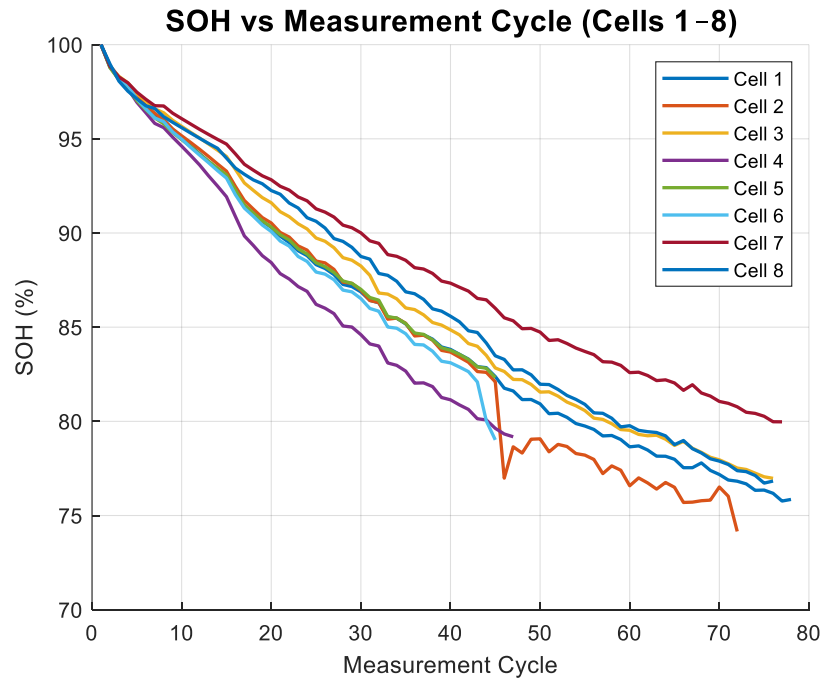


Figure 3.6 SOH degradation curves of all 8 cells from the Oxford Battery

Degradation Dataset

3.2.2 OCV Characteristics as a Health-Aware Feature

The OCV is a fundamental parameter that reflects the internal electrochemical state of a battery cell. OCV has a direct relationship with the SOC and can be used as an indicator of the electrode material's integrity and the amount of active lithium remaining within the cell. As the battery undergoes degradation over long-term usage, the shape of the OCV–SOC curve gradually changes, making it a valuable feature for estimating the SOH effectively. Physically, variations in OCV arise from changes in the chemical potential between the positive and negative electrodes. These changes occur as a result of internal structural degradation, such as the loss of active lithium ions, deterioration of electrode materials, or the thickening of the SEI layer. All of these mechanisms contribute to deviations in the open-circuit voltage at specific SOC levels from their nominal values. Consequently, the shape of the OCV curve is often used for analysis or feature extraction to represent the degradation state of a battery cell. For example, in this study, the voltage curve area during the charging and discharging phases is selected as one of the features to assist in the estimation of

the battery SOH, along with other parameters such as the potential difference at certain SOC points or the slope of the OCV curve. These characteristics provide insight into the underlying chemical changes occurring throughout the battery's operational cycles.

In the Oxford Battery Degradation Dataset used in this work, the voltage data were collected in the form of Pseudo-OCV cycles, obtained under a constant current of 40 mA. Although these measurements are not true OCV readings taken under equilibrium conditions, they provide a reliable approximation of OCV behavior and closely reflect the electrochemical dynamics of the cells. Therefore, the Pseudo-OCV data can be effectively used as additional features to enhance the performance of SOH estimation models.

3.2.3 Data Preprocessing

Before feature extraction, the raw dataset was carefully examined and cleaned. Some charging and discharging cycles exhibited abnormal behavior that clearly deviated from the expected trend, such as sudden voltage or current fluctuations. These anomalies were likely caused by measurement errors or hardware issues during data acquisition. Therefore, such cycles were removed from the dataset to ensure that the data used for model training accurately represented the real behavior of the battery.

After the cleaning process, the dataset was divided into three subsets: training, validation, and testing. To ensure a reliable evaluation of the model's generalization capability on unseen data, this division was performed based on battery cell identity, guaranteeing that no data from the same cell appeared across different subsets.

- 1) Training set (Cells 1–4): This subset, consisting of data from Cells 1–4, was used to train the model parameters.
- 2) Validation set (Cells 5–7): This subset included data from Cells 5–7, which served to fine-tune model hyperparameters and prevent overfitting during the training process.

3) Testing set (Cell 8): Finally, Cell 8 was reserved exclusively for testing to assess the model's generalization performance on data from a completely unseen battery.

3.3 Feature Extraction Methodology

To effectively estimate the SOH of LIBs, it is essential to extract relevant features that represent the battery's degradation behavior over time. These features are typically derived from raw data collected during charging and discharging cycles, including voltage and current. Feature extraction helps reduce data dimensionality and highlights important patterns or characteristics that are valuable for machine learning model training.

3.3.1 Area under curve

The AUC is a quantitative measure used to represent the accumulated value of a variable over a specific range. In this study, AUC is calculated from the voltage and current profiles recorded during both charging and discharging cycles. It is used as a feature that reflects the energy behavior and performance characteristics of the battery over time. The AUC is computed using the trapezoidal integration method as shown in the following equation (3.1).

$$AUC = \int_{t_0}^{t_n} x(t) = \sum_{i=1}^{n-1} \frac{(x_i + x_{i+1})}{2} \cdot (t_{i+1} - t_i) \quad (3.1)$$

where x_t represents either the voltage or current at time t , and t_0 to t_n are the time intervals during the cycle. This integration provides insight into the total energy processed during each cycle, and variations in AUC values can be correlated with battery ageing and SOH degradation. When plotting voltage versus time and current versus time for every cycle of the battery cell, it can be observed that the curves tend to become smaller as the number of cycles increases. This indicates a gradual decline in both voltage and current profiles, reflecting the degradation of battery performance over time. Figures 3.7 and 3.8 illustrate the voltage profiles of Cell 1 during all charge and discharge cycles, respectively. Similarly, Figs. 3.9 and 3.10

present the corresponding current profiles for Cell 1. The profiles of the remaining cells are provided in the Appendix A for completeness.

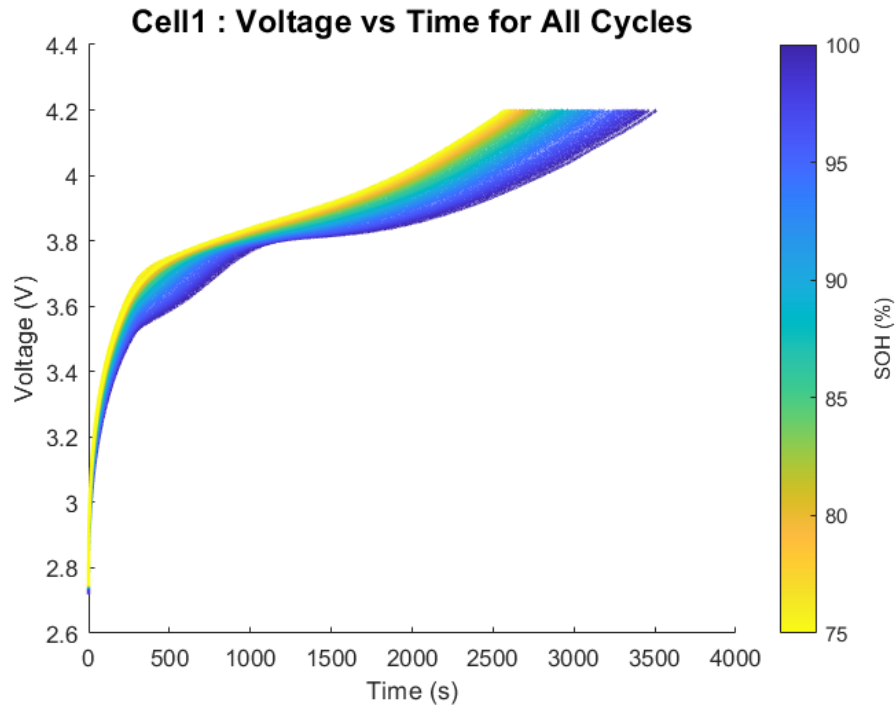


Figure 3.7 Voltage profiles Cell 1 of Oxford are used to charge all the cycles.

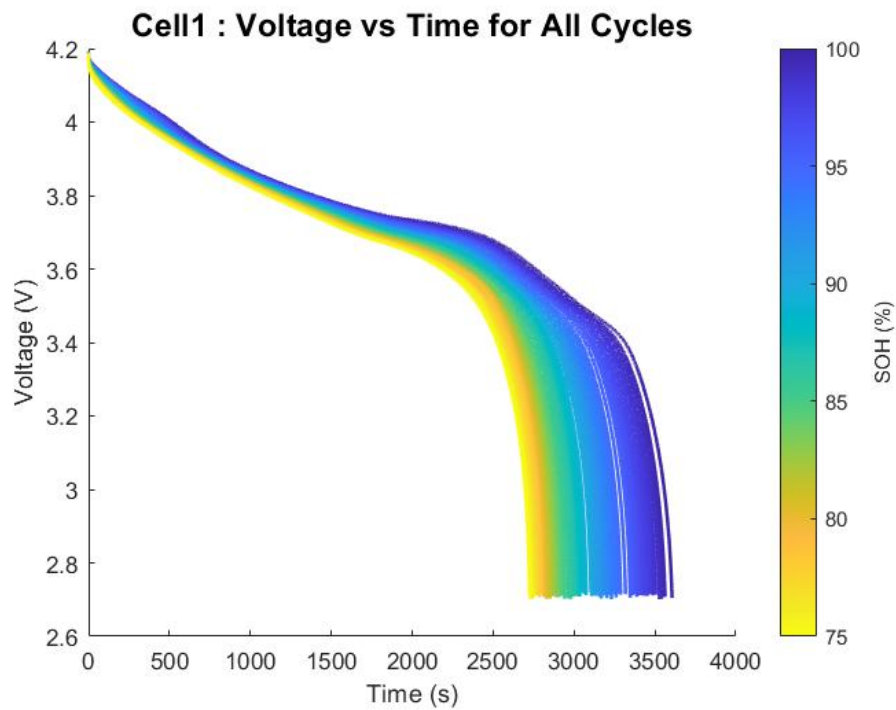


Figure 3.8 Voltage profiles Cell 1 of Oxford are used to discharge all the cycles.

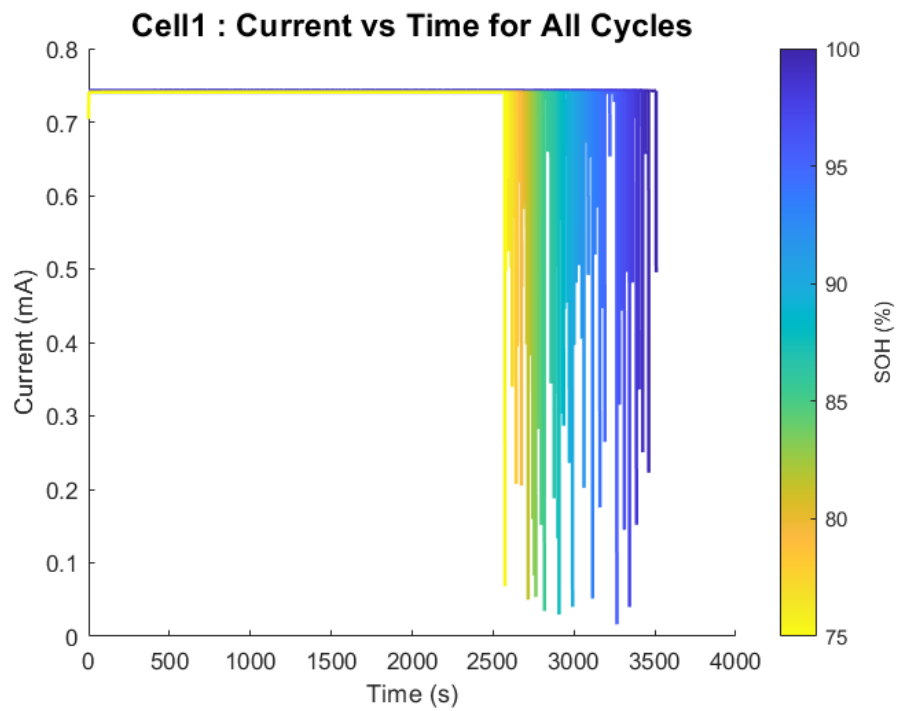


Figure 3.9 Current profiles Cell1 of Oxford are used to charge all the cycles.

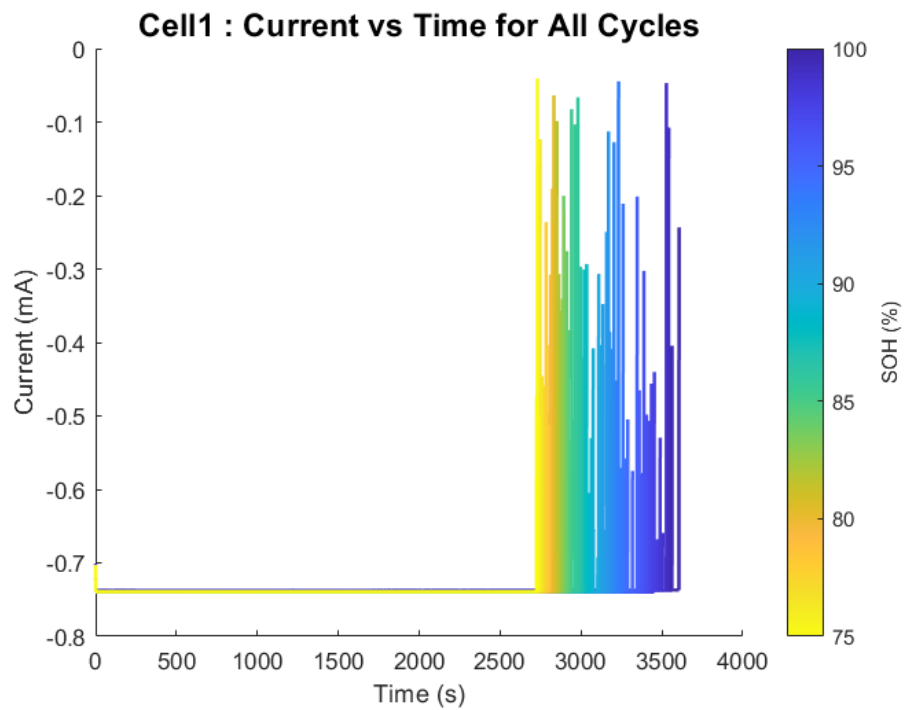


Figure 3.10 Current profiles Cell1 of Oxford are used to discharge all the cycles.

Consequently, when calculating the AUC using voltage and current data, it is found that the AUC values tend to decrease as the number of operating cycles increases. Before visualization, Min-Max Normalization was applied to the data. The normalized AUC graphs of voltage and current during the charging process for the Oxford dataset are illustrated in Fig. 3.11 and Fig. 3.12, respectively. Similarly, Fig. 3.13 and Fig. 3.14 show the normalized AUC graphs of voltage and current during the discharging process.

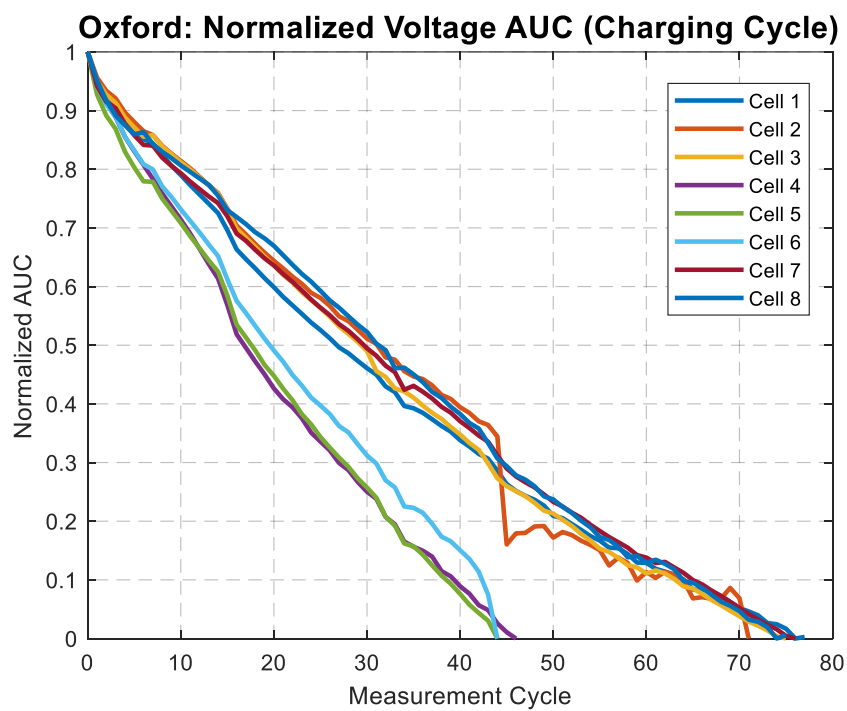


Figure 3.11 Normalized AUC of voltage during the charging cycle from the Oxford dataset

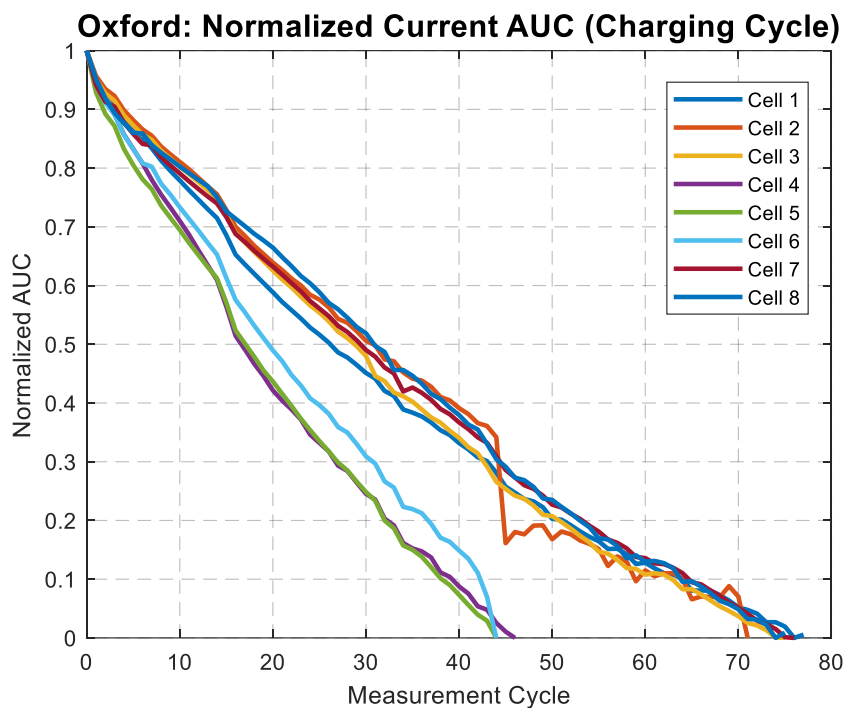


Figure 3.12 Normalized AUC of current during the charging cycle from the Oxford dataset

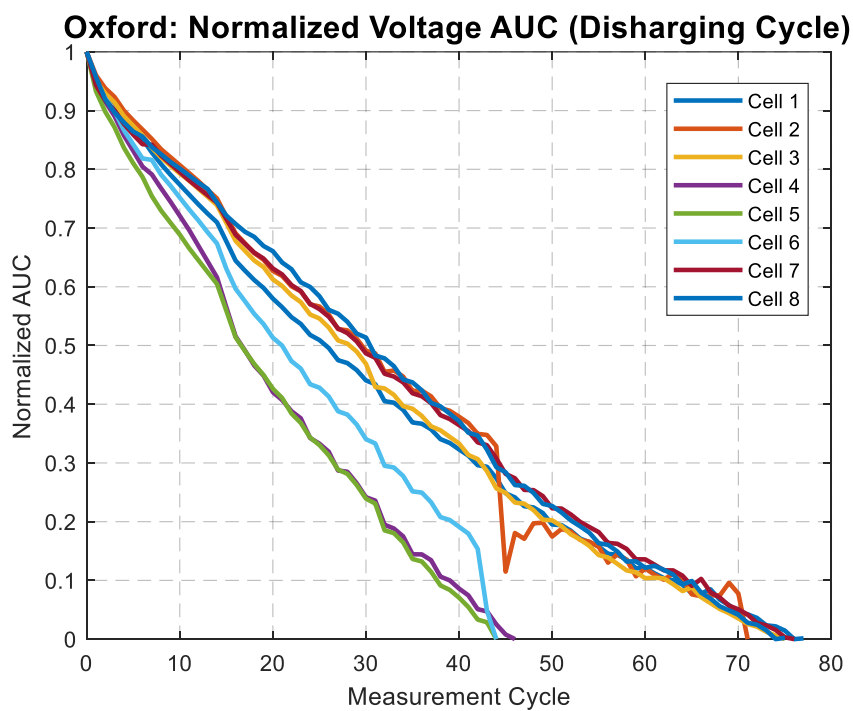


Figure 3.13 Normalized AUC of voltage during the discharging cycle from the Oxford dataset.

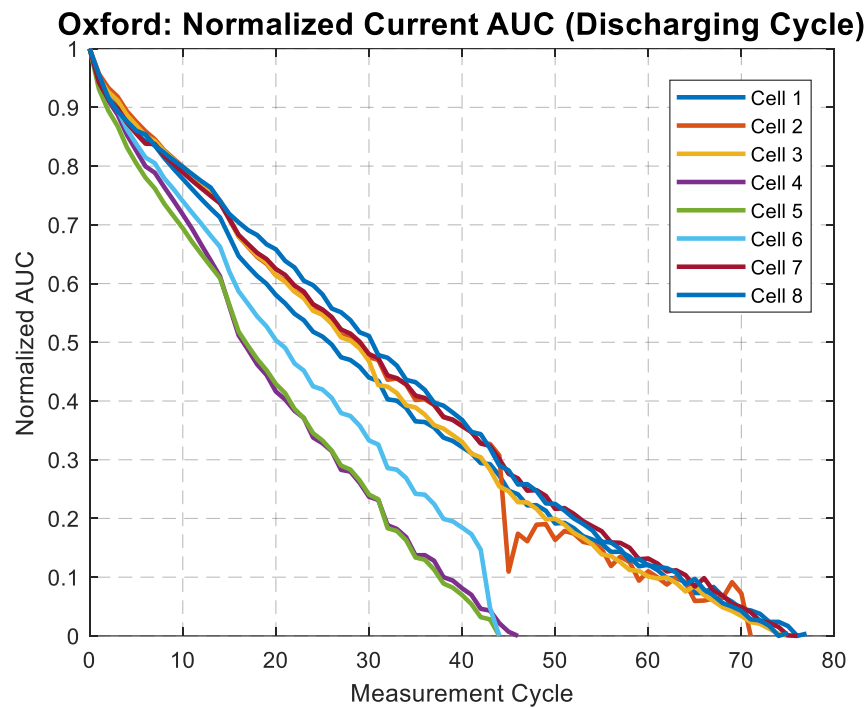


Figure 3.14 Normalized AUC of current during the discharging cycle from the Oxford dataset

The overall trends of the normalized AUC values derived from both voltage and current data show a similar decreasing pattern to the battery's SOH, indicating a strong correlation between AUC behavior and battery degradation.

3.3.2 The charging and discharging durations

The charging and discharging durations are used as features for further analysis. The charging time refers to the period from the start of the charging process until the battery reaches its maximum voltage threshold, while the discharging time corresponds to the period during which the battery releases energy until its voltage drops to a predefined minimum. These time durations, recorded across operating cycles, tend to vary as the battery ages and can therefore serve as useful indicators for assessing the battery's SOH. However, it was observed that the NASA dataset contains anomalies towards the end of each cycle, particularly after the battery has been fully charged or completely discharged. The data recording continued beyond the actual end point, resulting in an overestimation of the charging and discharging

durations. Therefore, it was necessary to remove these erroneous segments prior to feature extraction in order to obtain accurate time duration values.

A correlation matrix was then constructed to analyze the relationships between the extracted features and the battery's SOH. The coefficient of determination (R^2) was used to quantify the linear correlation between each feature and the SOH value. The results showed that all three selected features voltage curve area (AUC_V), charging and discharging time, and AUC of voltage during the OCV measurement cycle exhibited strong correlations with the SOH, with R^2 values close to 1. This indicates that these features effectively capture the underlying physical behavior related to battery degradation. The results of this analysis are presented in Fig. 3.15.

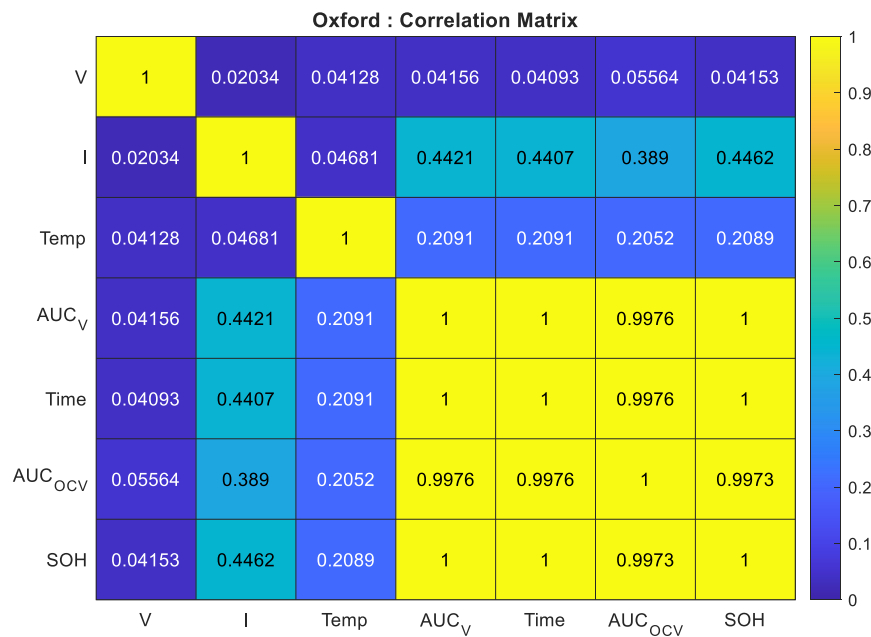


Figure 3.15 The correlation matrix

3.3.3 Hyperparameter Settings

Hyperparameter configuration plays a crucial role in determining the performance of a neural network model. Properly tuned hyperparameters enable the model to learn efficiently and help prevent overfitting, which occurs when the model performs well on training data but fails to generalize to unseen data. In this study, the same hyperparameter settings were applied to both models the one without OCV

information and the one with OCV information to ensure a fair comparison and to isolate the effect of incorporating the OCV feature on the model's SOH estimation accuracy. The complete configuration of hyperparameters used in the LSTM models in Table 3.1.

Table 3.1 Hyperparameter Settings for Training

Parameter	Value
Number of Hidden Units	10
Max Epochs	300
Mini-Batch Size	1024
Initial Learn Rate	0.1
Learn Rate Drop Factor	0.5
Learn Rate Drop Period	40 epochs
Optimizer	Adam

3.4 Results of the Proposed SOH Estimation Method

After completing data preprocessing and feature extraction, the prepared dataset was divided into training, validation, and testing subsets as previously detailed. The training set was used to train the model parameters. The validation set served to fine-tune model hyperparameters and prevent overfitting. Finally, the testing set was reserved exclusively for testing to assess the model's generalization performance on unseen data. This separation ensures that the evaluation results reflect the true predictive capability of the proposed SOH estimation model rather than memorization of previously observed patterns. The estimation results for Cell 8 are presented to evaluate the performance of the proposed SOH models. In particular, the predictions from the model that includes the OCV feature are compared with those from the model without OCV. This comparison highlights the impact of incorporating OCV on the accuracy of SOH estimation and provides insight into how additional electrochemical information can improve predictive performance. The results are illustrated in Fig. 3.16, where the predicted SOH values are plotted alongside the true measurements for clear visual comparison.

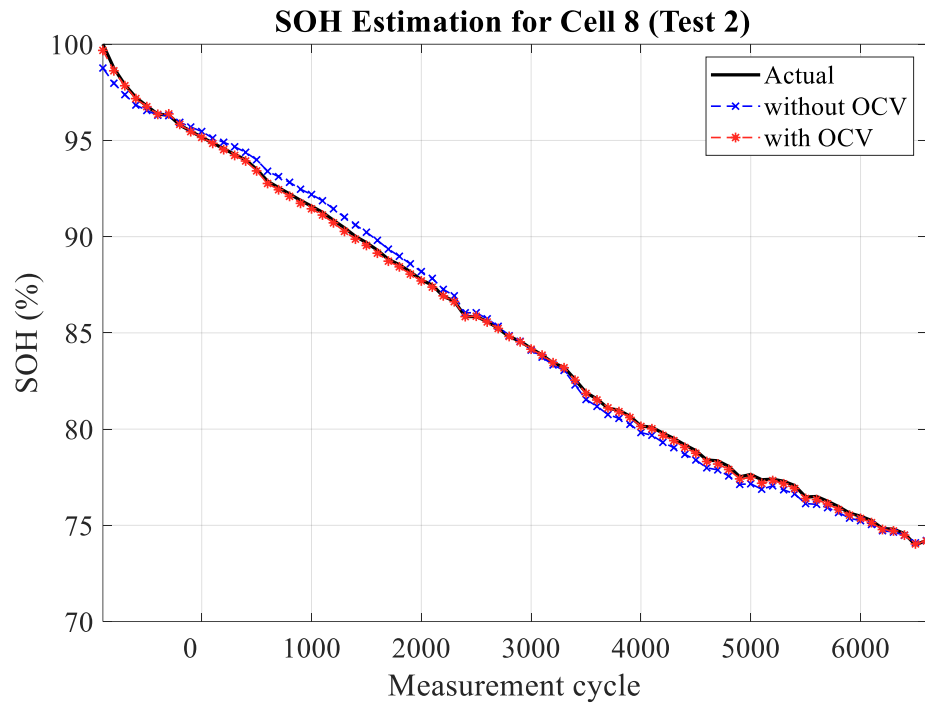


Figure 3.16 SOH estimation of Cell 8 from the Oxford dataset

As shown in the figures, the model without OCV already provides reasonably accurate SOH predictions. However, the model with OCV demonstrates improved performance, indicating that including the OCV feature contributes to enhancing the estimation accuracy. To ensure the reliability of the results, the experiments were repeated three times, and consistent outcomes were obtained. The quantitative results of the SOH estimation, evaluated using the four metrics, are summarized in Table 3.2.

Table 3.2 Quantitative Evaluation of SOH Estimation for Cell 8 Using Different Models

Test	Model	Metrics			
		<i>RMSE</i>	<i>MSE</i>	<i>MAE</i>	<i>MAPE</i>
1	without OCV	0.4518	0.2041	0.404	0.4790
	with OCV	0.1071	0.0115	0.0677	0.0778
2	without OCV	0.4130	0.1706	0.3610	0.4208
	with OCV	0.1040	0.0108	0.0859	0.1022
3	without OCV	0.4331	0.1881	0.3873	0.4537
	with OCV	0.1128	0.0127	0.0868	0.1060

The estimation results show that adding the OCV parameter to the model improves the estimation accuracy significantly. The model without OCV achieved an RMSE of 0.4130, MSE of 0.1706, MAE of 0.3610, and MAPE of 0.4208. In contrast, the model with OCV achieved much better performance with an RMSE of 0.1040, MSE of 0.0108, MAE of 0.0859, and MAPE of 0.1022. Although including the OCV parameter enhances accuracy, it requires additional measurement equipment and may not be practical for real-world applications.

3.5 Summary

This study investigates the estimation of Li-ion battery SOH by comparing LSTM models with and without the inclusion of the OCV parameter. The results indicate that the model incorporating OCV achieves higher accuracy in SOH estimation, as OCV provides additional information about the battery state, enabling the model to predict SOH more precisely. The advantage of using OCV lies in its direct correlation with the SOC and its lower noise compared to voltage measurements during charging or discharging, allowing the model to capture battery changes more accurately. However, including OCV increases the complexity of data processing and also extends the training time of the model, due to the higher dimensionality and detail of the input data. In addition, adding the OCV variable may raise the risk of overfitting, potentially affecting the model's generalization capability. To mitigate overfitting, increasing the diversity of the training data is beneficial. This can be achieved by incorporating a wider

range of operating conditions and battery states, ensuring that the model generalizes well across different scenarios.

In summary, incorporating OCV improves the accuracy of SOH estimation but requires additional complexity and longer training time. The full findings of this foundational study have been published, and the paper is included in Appendix E. This concept serves as a basis for the current research, which utilizes a larger dataset and additional feature extraction along with resampling techniques to reduce data and enhance the performance of deep learning models for SOH estimation.