RESAMPLING-BASED DATA SIMPLIFICATION FOR LITHIUM-ION BATTERY STATE OF HEALTH ESTIMATION USING DEEP LEARNING

SARAN TECHANOK

มหาวิทยาลัยเทคโนโลยีสุรนารี

A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Engineering in Electrical Engineering
Suranaree University of Technology
Academic Year 2025

การประมาณสถานะสุขภาพของแบตเตอรี่ลิเธียมไอออนด้วยการเรียนรู้เชิงลึกบนข้อมูลที่ลดความซับซ้อนด้วยเทคนิคการสุ่มตัวอย่างใหม่

นายศรัณย์ เตชะนอก

มหาวิทยาลัยเทคโนโลยีสุรนารี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2568

RESAMPLING-BASED DATA SIMPLIFICATION FOR LITHIUM-ION BATTERY STATE OF HEALTH ESTIMATION USING DEEP LEARNING

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for a Master's Degree.

Thesis Examining Committee

Somchat Jiriwibhakorn
.....
(Assoc. Prof. Dr. Somchat Jiriwibhakorn)
Chairperson

.....
(Assoc. Prof. Dr. Keerati Chayakulkheeree)
Member (Thesis Advisor)

.....
(Asst. Prof. Dr. Nitikorn Junhuathon)
Member (Thesis Co-Advisor)

.....
(Prof. Dr. Thanatchai Kulworawanichpong)
Member

Tosaphol Ratniyomchai
.....
(Asst. Prof. Dr. Tosaphol Ratniyomchai)
Member

Yupaporn Ruksakulpiwat
.....
(Assoc. Prof. Dr. Yupaporn Ruksakulpiwat)
Acting Vice Rector for Academic Affairs and Quality Assurance

.....
(Assoc. Prof. Dr. Pornsiri Jongkol)
Dean of Institute of Engineering

มหาวิทยาลัยเทคโนโลยีสุรนารี

ศรัณย์ เตชะนอก : การประมาณสถานะสุขภาพของแบตเตอรี่ลิเธียมไอออนด้วยการเรียนรู้เชิงลึกบนข้อมูลที่ลดความซับซ้อนด้วยเทคนิคการสุ่มตัวอย่างใหม่ (RESAMPLING-BASED DATA SIMPLIFICATION FOR LITHIUM-ION BATTERY STATE OF HEALTH ESTIMATION USING DEEP LEARNING)
อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร. กีรติ ชยะกุลคีรี, 146 หน้า.

คำสำคัญ : แบตเตอรี่ลิเธียมไอออน, การประมาณค่าสุขภาพแบตเตอรี่, เทคนิคการสุ่มตัวอย่างใหม่, การเรียนรู้เชิงลึก

วิทยานิพนธ์นี้นำเสนอแนวทางการประมาณค่าสุขภาพแบตเตอรี่ (State of Health - SOH) ของแบตเตอรี่ลิเธียมไอออน เพื่อแก้ปัญหาข้อมูลขนาดใหญ่และความซับซ้อนของรูปแบบการเสื่อมสภาพ งานวิจัยนี้เน้นการประยุกต์ใช้ เทคนิคการสุ่มตัวอย่างใหม่ (Resampling Technique) เพื่อลดขนาดข้อมูล ร่วมกับคุณลักษณะที่สกัดได้ (Extracted Features) อาทิ พื้นที่ใต้กราฟ (AUC) ระยะเวลาการชาร์จ/คายประจุ และคุณลักษณะจากการวิเคราะห์ความจุส่วนเพิ่ม (ICA) เพื่อประมาณค่าด้วยเทคนิคการเรียนรู้เชิงลึก (Deep Learning) โดยทดสอบกับชุดข้อมูลมาตรฐาน Oxford และ NASA ที่มีพฤติกรรมทางกายภาพแตกต่างกัน

จากผลการศึกษาแสดงให้เห็นว่า การประยุกต์ใช้เทคนิคการสุ่มตัวอย่างใหม่สามารถลดภาระทางคอมพิวเตอร์และระยะเวลาในการฝึกฝนโมเดลลงได้อย่างมีนัยสำคัญเมื่อเทียบกับการใช้ข้อมูลดิบ โดยยังคงรักษาข้อมูลสารสนเทศที่สำคัญซึ่งจำเป็นต่อความแม่นยำในการประมาณค่าไว้ได้ การเลือกช่วงเวลาในการสุ่มตัวอย่างที่เหมาะสมกับลักษณะของข้อมูลช่วยให้สามารถจัดการกับสัญญาณรบกวนหรือรักษารายละเอียดของพฤติกรรมทางเคมีไฟฟ้าที่ซับซ้อนได้อย่างมีประสิทธิภาพ นอกจากนี้ ผลการเปรียบเทียบเชิงประจักษ์ยืนยันว่าโมเดลการเรียนรู้เชิงลึกมีความทนทานและความแม่นยำในการทำนายค่า SOH ที่เหนือกว่าวิธีการทางสถิติแบบดั้งเดิม สามารถปรับตัวเข้ากับรูปแบบการเสื่อมสภาพที่หลากหลาย และเป็นแนวทางที่เหมาะสมสำหรับการประยุกต์ใช้ในระบบบริหารจัดการแบตเตอรี่ที่มีข้อจำกัดด้านทรัพยากร

สาขาวิชาวิศวกรรมไฟฟ้า
ปีการศึกษา 2568

ลายมือชื่อนักศึกษา .....
ลายมือชื่ออาจารย์ที่ปรึกษา.....
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

SARAN TECHANOK : RESAMPLING-BASED DATA SIMPLIFICATION FOR LITHIUM-ION BATTERY STATE OF HEALTH ESTIMATION USING DEEP LEARNING
THESIS ADVISOR : ASSOC. PROF. KEERATI CHAYAKULKHEEREE, Ph.D., 146 PP.

Keyword : Lithium-ion Battery / State of Health (SOH) Estimation / Resampling Technique / Deep Learning

This thesis presents an approach for State of Health (SOH) estimation of lithium-ion batteries to address challenges regarding large-scale data volume and complex degradation patterns. The research focuses on applying a resampling technique to reduce data complexity, utilizing extracted features such as the Area Under the Curve (AUC), Charging/Discharging Duration, and Incremental Capacity Analysis (ICA) for deep learning estimation. The methodology was evaluated using Oxford and NASA datasets, which exhibit distinct physical characteristics and degradation behaviors. The experimental results demonstrate that applying the resampling technique significantly reduces the size and complexity of the measurement data, leading to a substantial reduction in model training time compared to using raw data, while preserving critical information required for accurate estimation. Selecting a resampling interval appropriate for the specific data characteristics effectively filters noise or retains fine-grained details of complex electrochemical behaviors. Furthermore, empirical comparisons confirm that deep learning models utilizing these extracted features exhibit superior robustness and accuracy in SOH prediction compared to traditional statistical methods. The findings suggest that the proposed approach is capable of adapting to diverse degradation patterns and serves as a viable solution for implementation in battery management systems with limited computational resources

School of Electrical Engineering
Academic Year 2025

Student's Signature .....
Advisor's Signature .....
Co-Advisor's Signature .....

ACKNOWLEDGEMENT

The authors are particularly grateful to the Kittibandit Scholarship at Suranaree University of Technology, for which I am sincerely grateful.

I appreciate my advisors, Assoc. Prof. Dr. Keerati Chayakulkheeree and Asst. Prof. Dr. Nitikorn Junhuathon, whose guidance, insightful feedback, and constant encouragement have been instrumental in the completion of this work. Beyond academic advice, I am deeply grateful for the valuable life skills and practical wisdom I have gained from him in my daily life. His mentorship has not only shaped my research abilities but also inspired me to develop discipline, perseverance, and a positive attitude in facing challenges both in my studies and in everyday living. I also extend my sincere thanks to the committee members: Assoc. Prof. Dr. Somchat Jiriwibhakorn, Prof. Dr. Thanatchai Kulworawanichpong, and Asst. Prof. Dr. Tosaphol Ratniyomchai, for their valuable advice, suggestions, and support throughout the thesis evaluation process

Special appreciation is given to my parents for their enduring support, both emotionally and financially. Their belief in me has been a cornerstone of my academic journey. I am deeply grateful for everything they have done for me.

Furthermore, I would like to acknowledge all my lecturers at the School of Electrical Engineering for providing me with foundational knowledge and academic mentorship during my studies. I am also grateful to my friends, seniors, and fellow members of the Energy and Power Engineering Research Group (EPERG), whose collaboration, encouragement, and assistance greatly contributed to this work.

Lastly, I would like to thank the faculty members and administrative staff at Suranaree University of Technology for their support and kind assistance throughout my academic life.

Saran Techanok

TABLE OF CONTENTS

Page

ABSTRACT (THAI) I
ABSTRACT (ENGLISH) II
ACKNOWLEDGEMENT III
TABLE OF CONTENTS IV
LIST OF TABLES VII
LIST OF FIGURES VIII
LIST OF ABBREVIATIONS XIII
CHAPTER
I INTRODUCTION 1
1.1 General Introduction 1
1.2 Problem Statement 2
1.3 Research Objectives 3
1.4 Scope and limitations 3
1.4.1 Scope 3
1.4.2 limitations 4
1.5 Conception 4
1.6 Research Benefits 5
1.7 Structure of thesis 5
II LITERATURE REVIEW 7
2.1 Introduction 7
2.2 Literature Overview 7
2.3 Model Development and Hybridization Approaches 17
2.4 Feature Extraction and Input Variable Selection 18
2.5 Battery Dataset 20
2.6 Battery Degradation 27

มหาวิทยาลัยเทคโนโลยีสุรนารี

TABLE OF CONTENTS (Continued)

Page

2.7 Definition and Calculation of SOH 29
2.8 Model Definition 29
2.8.1 LSTM Model Architecture 29
2.8.2 Bi-LSTM Model Architecture 31
2.8.3 AM-LSTM Model Architecture 33
2.8.4 GRU Model Architecture 34
2.8.5 GPR Model Architecture 36
2.9 Evaluation Metrics 37
2.9.1 Root Mean Squared Error 37
2.9.2 Mean Squared Error 38
2.9.3 Mean Absolute Error 38
2.9.4 Mean Absolute Percentage Error 38
III EVALUATING THE ROLE OF OCV IN LSTM MODEL 40
3.1 Introduction 40
3.2 Battery Datasets Utilized 41
3.2.1 Oxford Battery Degradation Dataset 41
3.2.2 OCV Characteristics as a Health-Aware Feature 45
3.2.3 Data Preprocessing 46
3.3 Feature Extraction Methodology 47
3.3.1 Area under curve 47
3.3.2 The charging and discharging durations 52
3.3.3 Hyperparameter Settings 53
3.4 Results of the Proposed SOH Estimation Method 54
3.5 Summary 56
IV INVESTIGATING THE IMPACT OF DATA RESAMPLING ON MODEL ACCURACYAND COMPLEXITY 58

มหาวิทยาลัยเทคโนโลยีสุรนารี

TABLE OF CONTENTS (Continued)

Page

4.1 Introduction 58
4.2 Battery Datasets Utilized 59
4.2.1 NASA Prognostics Center of Excellence Dataset 59
4.2.2 Dataset Partitioning Strategy 62
4.3 Data Resampling 62
4.4 Feature Extraction Methodology 64
4.4.1 Area under curve 64
4.4.2 The charging and discharging durations 69
4.4.3 Incremental Capacity Analysis 70
4.5 Feature Correlation Analysis 73
4.6 Hyperparameter Settings 75
4.7 Experimental Results and Discussion 76
4.7.1 Hyperparameter Tuning Analysis 76
4.7.2 Comparative SOH Estimation Results 78
4.8 Summary 91
V CONCLUSION 93
5.1 Conclusion 93
5.2 Suggestions 94
REFERENCES 96
APPENDIX A AUC Feature Extraction for Oxford Battery Data 102
APPENDIX B AUC Feature Extraction for NASA Battery Data 116
APPENDIX C ICA Feature Extraction for Oxford Battery Data 122
APPENDIX D ICA Feature Extraction for NASA Battery Data 126
APPENDIX E List of publications 128
BIOGRAPHY 146

มหาวิทยาลัยเทคโนโลยีสุรนารี

LIST OF TABLES

Table **Page**

2.1 The research considering battery dataset 8
2.2 The research considering feature extraction 10
2.3 The research considering model development 13
2.4 Presents a comparison of previous research 22
3.1 Hyperparameter Settings for Training 54
3.2 Quantitative Evaluation of SOH Estimation for Cell 8 Using Different Models 56
4.1 Dataset Partitioning for Oxford and NASA Datasets 62
4.2 Hyperparameter Settings for Training 76
4.3 RMSE results from the Grid Search analysis on the Oxford dataset 77
4.4 Comparison of SOH Estimation Results on Oxford Dataset 83
4.5 Execution Time of Each Model for SOH Estimation on Oxford Dataset 84
4.6 Comparison of SOH Estimation Results on NASA Dataset 90
4.7 Execution Time of Each Model for SOH Estimation on NASA Dataset 91

มหาวิทยาลัยเทคโนโลยีสุรนารี

LIST OF FIGURES

Figure **Page**

2.1 The comparison of battery datasets 25
2.2 The comparison of feature engineering techniques 25
2.3 The comparison of model usage rates 26
2.4 An example of the LSTM cell 31
2.5 The Bi-LSTM network architecture 33
2.6 The ATT-LSTM network architecture 34
2.7 The GRU network architecture 36
3.1 Framework of comparative analysis between LSTM models with and without OCV feature inclusion 41
3.2 Voltage profiles of Cell 1 during charging cycle from the Oxford Battery Dataset 42
3.3 Current profiles of Cell 1 during charging cycle from the Oxford Battery Dataset 43
3.4 Voltage profiles of Cell 1 during discharging cycle from the Oxford Battery Dataset 43
3.5 Current profiles of Cell 1 during discharging cycle from the Oxford Battery Dataset 44
3.6 SOH degradation curves of all 8 cells from the Oxford Battery Degradation Dataset 45
3.7 Voltage profiles Cell 1 of Oxford are used to charge all the cycles 48
3.8 Voltage profiles Cell 1 of Oxford are used to discharge all the cycles 48
3.9 Current profiles Cell1 of Oxford are used to charge all the cycles 49
3.10 Current profiles Cell1 of Oxford are used to discharge all the cycles 49
3.11 Normalized AUC of voltage during the charging cycle from the Oxford dataset 50

LIST OF FIGURES (Continued)

Figure **Page**

3.12 Normalized AUC of current during the charging cycle from the Oxford dataset 51
3.13 Normalized AUC of voltage during the discharging cycle from the Oxford dataset 51
3.14 Normalized AUC of current during the discharging cycle from the Oxford Dataset 52
3.15 The correlation matrix 53
3.16 SOH estimation of Cell 8 from the Oxford dataset 55
4.1 Experimental workflow for SOH estimation using data resampling 58
4.2 Voltage profiles of B0005 during charging cycle from the NASA Battery Dataset 60
4.3 Current profiles of B0005 during charging cycle from the NASA Battery Dataset 60
4.4 Voltage profiles of B0005 during discharging cycle from the NASA Battery Dataset 61
4.5 Current profiles of B0005 during discharging cycle from the NASA Battery Dataset 61
4.6 SOH degradation curves of all 8 cells from the Oxford Battery Degradation Dataset 62
4.7 Voltage profiles B0005 of NASA are used to charge all the cycles 65
4.8 Voltage profiles B0005 of NASA are used to discharge all the cycles 65
4.9 Current profiles B0005 of NASA are used to charge all the cycles 66
4.10 Current profiles B0005 of NASA are used to discharge all the cycles 66
4.11 Normalized AUC of voltage during the charging cycle from the NASA dataset 67
4.12 Normalized AUC of voltage during the discharging cycle from the NASA dataset 67

LIST OF FIGURES (Continued)

Figure **Page**

4.13 Normalized AUC of current during the charging cycle from the NASA dataset 68
4.14 Normalized AUC of current during the discharging cycle from the NASA Dataset 68
4.15 Charging and discharging time duration across cycles from the Oxford Dataset 69
4.16 Charging and discharging time duration across cycles from the NASA dataset 70
4.17 ICA curves for Oxford Cell 1 72
4.18 ICA curves for NASA B0005 72
4.19 The correlation matrix from the Oxford dataset..... 73
4.20 The correlation matrix from the NASA dataset 74
4.21 Comparison of RMSE values across different model architecture configurations 77
4.22 Baseline comparison of SOH estimation results using raw data (Non-resampling) for all models on the Oxford dataset 79
4.23 Comparison of SOH estimation results using the Best Resampling at 100s for all models on the Oxford dataset..... 80
4.24 Analysis of the impact of resampling on the AM-LSTM model (Oxford dataset) comparing Non-resampling vs. the optimal interval (100s)..... 81
4.25 RMSE comparison of all models on the Oxford dataset..... 82
4.26 Baseline comparison of SOH estimation results using raw data (Non-resampling) for all models on the NASA dataset 85
4.27 Comparison of SOH estimation results using the Best Resampling at 20s for all models on the NASA dataset 86

LIST OF FIGURES (Continued)

Figure **Page**

4.28 Analysis of the impact of resampling on the AM-LSTM model (NASA dataset) comparing Non-resampling vs. 20s interval 87
4.29 RMSE comparison of all models on the NASA dataset 70
A.1 Voltage profiles Cell 2 of Oxford are used to charge all the cycles 102
A.2 Voltage profiles Cell 3 of Oxford are used to charge all the cycles 102
A.3 Voltage profiles Cell 4 of Oxford are used to charge all the cycles 103
A.4 Voltage profiles Cell 5 of Oxford are used to charge all the cycles 103
A.5 Voltage profiles Cell 6 of Oxford are used to charge all the cycles 104
A.6 Voltage profiles Cell 7 of Oxford are used to charge all the cycles 104
A.7 Voltage profiles Cell 8 of Oxford are used to charge all the cycles 105
A.8 Voltage profiles Cell 2 of Oxford are used to discharge all the cycles 105
A.9 Voltage profiles Cell 3 of Oxford are used to discharge all the cycles 106
A.10 Voltage profiles Cell 4 of Oxford are used to discharge all the cycles 106
A.11 Voltage profiles Cell 5 of Oxford are used to discharge all the cycles 107
A.12 Voltage profiles Cell 6 of Oxford are used to discharge all the cycles 107
A.13 Voltage profiles Cell 7 of Oxford are used to discharge all the cycles 108
A.14 Voltage profiles Cell 8 of Oxford are used to discharge all the cycles 108
A.15 Current profiles Cell 2 of Oxford are used to charge all the cycles 109
A.16 Current profiles Cell 3 of Oxford are used to charge all the cycles 109
A.17 Current profiles Cell 4 of Oxford are used to charge all the cycles 110
A.18 Current profiles Cell 5 of Oxford are used to charge all the cycles 110
A.19 Current profiles Cell 6 of Oxford are used to charge all the cycles 111
A.20 Current profiles Cell 7 of Oxford are used to charge all the cycles 111
A.21 Current profiles Cell 8 of Oxford are used to charge all the cycles 112
A.22 Current profiles Cell 2 of Oxford are used to discharge all the cycles 112
A.23 Current profiles Cell 3 of Oxford are used to discharge all the cycles 113
A.24 Current profiles Cell 4 of Oxford are used to discharge all the cycles 113

LIST OF FIGURES (Continued)

Figure **Page**

A.25 Current profiles Cell 5 of Oxford are used to discharge all the cycles 114
A.26 Current profiles Cell 6 of Oxford are used to discharge all the cycles 114
A.27 Current profiles Cell 7 of Oxford are used to discharge all the cycles 115
A.28 Current profiles Cell 8 of Oxford are used to discharge all the cycles 115
B.1 Voltage profiles B0006 of NASA are used to charge all the cycles 116
B.2 Voltage profiles B0007 of NASA are used to charge all the cycles 116
B.3 Voltage profiles B0018 of NASA are used to charge all the cycles 117
B.4 Voltage profiles B0006 of NASA are used to discharge all the cycles 117
B.5 Voltage profiles B0007 of NASA are used to discharge all the cycles 118
B.6 Voltage profiles B0018 of NASA are used to discharge all the cycles 118
B.7 Current profiles B0006 of NASA are used to charge all the cycles 119
B.8 Current profiles B0007 of NASA are used to charge all the cycles 119
B.9 Current profiles B0018 of NASA are used to charge all the cycles 120
B.10 Current profiles B0006 of NASA are used to discharge all the cycles 120
B.11 Current profiles B007 of NASA are used to discharge all the cycles 121
B.12 Current profiles B0018 of NASA are used to discharge all the cycles 121
C.1 ICA curves for Oxford Cell 2 122
C.2 ICA curves for Oxford Cell 3 122
C.3 ICA curves for Oxford Cell 4 123
C.4 ICA curves for Oxford Cell 5 123
C.5 ICA curves for Oxford Cell 6 124
C.6 ICA curves for Oxford Cell 7 124
C.7 ICA curves for Oxford Cell 8 125
D.1 ICA curves for NASA B0006 126
D.2 ICA curves for NASA B0007 126
D.3 ICA curves for NASA B0018 127

มหาวิทยาลัยเทคโนโลยีสุรนารี

LIST OF ABBREVIATIONS

AM-LSTM	=	Attention mechanism-long short-term memory
AUC	=	Area Under the Curve
Bi-LSTM		Bidirectional long short-term memory
BMS	=	Battery management system
EVs	=	Electric vehicles
GPR	=	Gaussian process regression
GRU	=	Gated recurrent unit
ICA	=	Incremental capacity analysis
LIBs	=	Lithium-ion batteries
Li-ion	=	Lithium-ion
LSTM	=	Long short-term memory
MAE	=	Mean absolute error
MAPE	=	Mean absolute percentage error
MSE	=	Mean squared error
OCV	=	Open circuit voltage
RMSE	=	Root mean squared error
RNN	=	Recurrent neural network
RUL	=	Remaining useful life
SEI	=	Solid electrolyte interphase
SOC	=	State of charge
SOH	=	State of Health

มหาวิทยาลัยเทคโนโลยีสุรนารี

CHAPTER I
INTRODUCTION

1.1 General Introduction

In today's rapidly advancing technological era, Li-ion batteries have emerged as the heart of global energy storage systems. It is undeniable that these batteries serve as the foundation for numerous innovations driving our daily lives, ranging from small portable devices like smartphones and tablets to large-scale industries such as EVs and smart grid systems. The preference for Li-ion batteries over other types stems from their superior characteristics, including high energy density, lightweight properties, and long operational lifespan, which perfectly meet the high-efficiency energy demands of the modern world.

However, despite their high performance, Li-ion batteries remain electrochemical devices with inherent natural limitations. As they undergo repeated usage and charging cycles over extended periods, degradation is inevitable. Internal chemical reactions gradually shift, leading to a decrease in energy storage capacity and an inability to deliver power as efficiently as when they were new. If this degradation is left unmonitored, it can lead to various risks, such as unexpected device shutdowns or, in severe cases, thermal runaway, which poses a safety hazard to users. For this reason, the BMS was developed to act as a controller and safety guardian. One of the most critical functions of the BMS is to estimate the SOH. Knowing the SOH is analogous to a health check-up, informing us of how much remaining life the battery has or what percentage of its efficiency remains. This information is vital for decision-making regarding maintenance planning, replacing batteries before failure, and ensuring that systems operate with maximum safety and cost-effectiveness.

1.2 Problem Statement

Currently, global warming and climate change have become urgent agendas prioritized by governments worldwide. Nations are collaborating through global initiatives such as the Paris Agreement and the United Nations' Sustainable Development Goals (SDGs) to reduce greenhouse gas emissions. These efforts are the primary drivers shifting the world from fossil fuels to clean energy, leading to an exponential rise in the demand for renewable energy and EVs. Consequently, the role of LIBs has become increasingly significant as the key enabler of this transition.

As battery usage creates a surge in demand, a consequential challenge arises regarding the management of aging batteries. Generally, when an EV battery's SOH falls below 80%, it is considered unsuitable for vehicle propulsion. However, these batteries still retain sufficient energy capacity to be utilized for other purposes, such as stationary energy storage for homes or buildings (Second-life applications). The problem lies in determining which batteries are still of good quality for reuse. Accurate SOH estimation is, therefore, essential to ensure that battery resources are utilized to their maximum economic potential while minimizing electronic waste.

Nevertheless, a major technical barrier existing today is "data complexity and volume." In real-world applications, modern battery systems continuously monitor and record status parameters such as voltage, current, and temperature at very high frequencies, generating massive amounts of Big Data. Processing all this raw data instantaneously to estimate SOH requires immense computational resources, which often exceeds the capabilities of the standard embedded chips found in vehicles or general devices. This leads to latency and excessive power consumption. This thesis aims to directly address this problem by introducing a resampling technique to reduce data size and complexity while preserving critical details. This approach allows for the effective application of Deep Learning technologies to estimate SOH quickly, accurately, and practically within resource-constrained systems.

1.3 Research Objectives

Building upon the challenges and limitations discussed previously, particularly regarding data complexity and computational constraints in battery management systems, this research aims to develop a robust and efficient framework for SOH estimation. The primary goal is to bridge the gap between high-accuracy theoretical models and practical implementation by integrating effective data management strategies with advanced deep learning techniques.

The main objectives of this thesis are as follows,

1) To apply a resampling technique to reduce the size and complexity of battery measurement data while preserving critical information necessary for accurate SOH estimation

2) To apply deep learning techniques for estimating the SOH of LIBs using multiple open-source datasets

1.4 Scope and limitations

1.4.1 Scope

The scope of this research is defined by the specific methodologies, datasets, and parameters selected for investigation, as detailed below:

1) Algorithms and Models: The study employs Deep Learning techniques, specifically LSTM, Bi-LSTM, GRU, and AM-LSTM. Additionally, GPR is utilized as a statistical baseline method (non-Deep Learning) to compare estimation performance.

2) Resampling Technique: A resampling approach is applied to reduce data size and complexity. The investigation covers specific resampling intervals for each dataset: 13 distinct intervals for the Oxford dataset and 8 distinct intervals for the NASA dataset.

3) Data Inputs: The study utilizes two open-source datasets (Oxford and NASA). The input features for all models are strictly derived from Voltage and Current measurements obtained during the charging and discharging cycles.

1.4.2 Limitations

The primary limitation of this research stems from its exclusive reliance on pre-existing open-source datasets rather than primary data generated through controlled laboratory experiments. Consequently, the researcher cannot control or modify the experimental conditions, such as ambient temperature, aging protocols, or discharge rates, as these factors were fixed during the original data collection. This implies that the findings and the developed models are inherently constrained by the specific physical characteristics and operational environments of the batteries in the Oxford and NASA datasets. Therefore, the models may not be directly generalizable to other battery chemistries or usage scenarios that differ significantly from the conditions recorded in these datasets without further retraining or adaptation.

1.5 Conception

The conception of this thesis is established upon a data-driven framework designed to optimize the SOH estimation process for LIBs. The core concept integrates Deep Learning techniques with a strategic Resampling approach. The fundamental idea is that by systematically reducing the sampling rate of raw voltage and current data, the computational complexity and data volume can be minimized without losing the critical electrochemical features such as ICA and AUC required for accurate health assessment. To realize this concept, the study utilizes reliable open-source datasets (Oxford and NASA) to simulate real-world degradation scenarios across different battery chemistries. The technical implementation and verification are conducted within the MATLAB environment, which serves as the primary platform for data preprocessing, model training, and performance evaluation. The overall conception aims to demonstrate that a simplified data structure, when processed by advanced neural networks, can achieve high estimation accuracy comparable to models using raw, high-resolution data To realize this concept, the study utilizes reliable open-source datasets (Oxford and NASA) to simulate real-world degradation scenarios across different battery chemistries. The technical implementation and verification are conducted within the MATLAB environment, which serves as the primary platform for data preprocessing, model training, and performance evaluation. The overall conception aims to

demonstrate that a simplified data structure, when processed by advanced neural networks, can achieve high estimation accuracy comparable to models using raw, high-resolution data.

1.6 Research Benefits

Accurate estimation of battery SOH can significantly enhance the efficiency, reliability, and safety of battery management systems, contributing to the optimization of system performance and lifecycle management. By applying a resampling technique to simplify input data while preserving essential features, the proposed method is expected to reduce model complexity, improve computational efficiency, and facilitate the deployment of SOH estimation models in practical applications.

1.7 Structure of thesis

This thesis is organized into five chapters, outlining the progression from theoretical background to experimental validation and conclusion. The structure is defined as follows:

Chapter I: Introduction presents the general background, problem statement, research objectives, scope, limitations, and the conceptual framework of the study.

Chapter II: Literature Review and Theoretical Background provides a review of related literature and fundamental theories regarding LIBs and their degradation mechanisms. It also covers the basic principles of the deep learning algorithms used in this study and defines the evaluation metrics for performance assessment.

Chapter III: Impact of OCV Features on SOH Estimation investigates the effect of integrating OCV features into the estimation process. This chapter focuses on utilizing AUC and charging duration as input features for a LSTM model.

Chapter IV: SOH Estimation using Resampling Technique and Deep Learning details the core methodology of this thesis. It presents the application of the resampling technique to reduce data complexity and evaluates the performance of multiple deep learning models (LSTM, Bi-LSTM, GRU, AM-LSTM) and GPR using Oxford and NASA datasets.

Chapter V: summarizes the research findings, discusses the key contributions.

CHAPTER II
LITERATURE REVIEW

2.1 Introduction

LIBs tend to degrade over time, and their RUL decreases as the number of charge-discharge cycles increases. This has driven considerable research attention toward the estimation of battery SOH, aiming to ensure optimal performance and safe operation of battery systems. Many studies have proposed various methods for estimating SOH, including approaches based on deep learning, which have shown promising results in capturing complex patterns from time-series battery data. In addition, hybrid models that combine different machine learning techniques have also been developed to enhance estimation accuracy. Researchers have also focused on extracting useful features from battery measurement data collected during charge and discharge cycles. Among these, the ICA has been widely used as a feature extraction technique that helps to reveal internal battery degradation mechanisms and improve model performance.

2.2 Literature Overview

In recent years, the challenge of effectively estimating the SOH of LIBs has received growing attention. Numerous research approaches have been proposed to address this issue, differing in their use of experimental setups, model-based techniques including adaptive algorithms and data-driven methods, such as machine learning and deep learning. The variety of datasets used in different studies further contributes to the diversity in approaches and outcomes. This section provides a summary and overview of related research in a table format to highlight the current landscape and the remaining research gaps. The literature review has been divided

into three tables: Table 2.1 Battery Dataset. Table 2.2 Feature Extraction and Table 2.3 Model Development.

Table 2.1 The research considering battery dataset

Reference	Method	Objective	Description
(Xiong et al., 2023)	- GPR - LSTM	- NASA - Oxford - Real-world	- Using random charging data and reconstructing missing health features with GPR-LSTM
(Liang Zhang et al., 2023)	- Ensemble	- NASA - Oxford	- Uses an ensemble model based on improved LSTM neural network to estimate SOH
(Wu et al., 2023)	- GPR - LSTM - BP	- NASA - CALCE - Oxford - UL-PUR	- HFs from multiple datasets
(Lijun Zhang et al., 2023)	- LSTM	- NASA - CALCE	- Using the LSTM model to predict the available capacity and RUL of LIBs.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.1 The research considering battery dataset (Continued)

Reference	Method	Objective	Description
(Dineva, 2024)	- ML and AI	- NASA - CALCE - Oxford	- A review of Li-ion battery SOH estimation techniques in EVs, focusing on advancements in ML and AI.
(Olabi et al., 2024)	- ANN - FNN - CNN	- FUDS - DST - Oxford - NASA	- Reviews the use of ANN in estimating Li-ion battery states, including SOC, SOH, RUL, thermal state, and other parameters.

From table 2.1 The research considering Battery Dataset, it is observed that most studies prefer to use open-source battery datasets. These datasets typically provide comprehensive recordings from the very beginning of battery usage, allowing for continuous monitoring of battery behavior and degradation over time.

Table 2.2 The research considering feature extraction

Reference	Method	Objective	Description
(B. Zou et al., 2023)	- Bi-LSTM model with AM	- NASA - Oxford	- A DL model is constructed using Bi-LSTM with AM to focus on important parts of the features.
(Sun et al., 2022)	- Bi-LSTM	- NASA	- Extraction of IC curve peaks and corresponding
(Bian et al., 2022)	- ANN	- NASA - CALCE	- Uses a fusion-based method combining OCV model and ICA with ANN to estimate SOH
(Z. Zou et al., 2023)	- BP - PCR	- Information gathered from past battery cells that the researchers	- Uses ICA and PCA combined with Multi-chaotic Sparrow Search Algorithm to estimate SOH
(Stephens et al., 2024)	- GPR	- NASA	- ICA-DTA outperforms using ICA and DTA individually

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.2 The research considering feature extraction (Continued)

Reference	Method	Objective	Description
(Wei et al., 2025)	- MSHHO-DELM	- Oxford	- Uses the MSHHO-DELM framework improved from Harris Hawks Optimization and Deep Extreme Learning Machine to estimate SOH
(Rao et al., 2024)	- LSTM - DNN - GRU	- NASA	- Battery data is processed using Fourier Resampling to produce uniform and denoised signals, enhancing the accuracy of SOH prediction. - Incremental capacity-based input datasets have been considered as input to GRU

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.2 The research considering feature extraction (Continued)

Reference	Method	Objective	Description
(Techanok et al., 2025)	- LSTM	- Oxford	- Compare the results of evaluating the role of OCV in estimating the SOH of batteries. - Use the variable AUC to effectively highlight differences between battery cycles.

The table 2.2 The research considering Feature Extraction summarizes studies focusing on the extraction of informative features from battery usage data for use as model inputs in SOH estimation. These studies highlight the critical role of proper feature selection in improving model accuracy. Various techniques have been applied, such as ICA, and in some cases, resampling of raw signals to enhance temporal resolution.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.3 The research considering model development

Reference	Method	Objective	Description
(Fan et al., 2020)	- GRU-CNN	- NASA - Oxford	- Using a hybrid GRU-CNN model, which can learn shared information and time dependencies of the charging curve with deep learning technology.
(Aliberti et al., 2022)	- LSTM - GRU - 1D-CNN - CNN-LSTM	- NASA	- Various neural network models - 1D-CNN best predicts SOH
(Audin et al., 2021)	- Autoencoders and LSTM	- SNL - MIT - NASA	- Using Autoencoders and LSTM Neural Networks applied to usage data (voltage, current, temperature)
(Nguyen Van & Quang, 2023)	- LSTM - FNN	- NASA	- LSTM has higher accuracy than FNN in estimating SOH and internal resistances

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.3 The research considering model development (Continued)

Reference	Method	Objective	Description
(Oyewole et al., 2022)	- CDTL	- Information gathered from past battery cells that the researchers	- Construction of a CDTL model consisting of two LSTM networks: source LSTM and target LSTM
(Wang et al., 2022)	- GPR	- Oxford	- Uses Pearson correlation coefficient to select health features and GPR to estimate the SOH
(Li et al., 2023)	- GPR	- NASA -Oxford	- Estimation of SOH and RUL of batteries using a hybrid Temporal Convolutional Network-Long Short-Term Memory (TCN-LSTM) model.
(Gong et al., 2021)	- CNN - ULSAM - GRU	- NASA - Oxford	- Uses a hybrid CNN-ULSAM-GRU model to estimate SOC of LIBs

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.3 The research considering model development (Continued)

Reference	Method	Objective	Description
(Hossain Lipu et al., 2022)	- DL	- Data from various studies related to the estimation of SOC, SOH, and RUL of EV batteries	- Reviews the use of DL in estimating SOC, SOH, and RUL of EV batteries, including methods, implementations, strengths, weaknesses, accuracy, and future recommendations
(Jarraya et al., 2025)	- SOH-KLSTM	- NASA	- Integrating KAN with LSTM to estimate SOH of LIBs networks: source LSTM and target LSTM
(Li et al., 2025)	- MC-CNN-TimesNet.	- NASA - CALCE -. MIT	- Estimation of Li-ion battery SOH and RUL using the MC-CNN-TimesNet model.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.3 The research considering model development (Continued)

Reference	Method	Objective	Description
(Ma et al., 2024)	- Bagged-GPR	- LMBs dataset	- Uses a data-driven health estimation framework and Bagged-GPR to estimate SOH of lithium metal batteries, validated with data from test cells.
(Safavi et al., 2024)	- CNN - LSTM	- NASA	- 2D voltage data and use the sliding window technique to create a new dataset, automatically extracting health features. - The sampling rate was increased to have one voltage measurement every 20 s

มหาวิทยาลัยเทคโนโลยีสุรนารี

Based on table 2.3 research considering Model Development, it is evident that many studies have focused on the development of hybrid models by combining the strengths of various approaches. the strengths of various approaches. These hybrid strategies have shown significant improvements in SOH estimation accuracy. The literature reviewed in this table highlights a clear trend in recent research toward multi-method integration, aiming to enhance model performance in response to the complex characteristics of battery data.

2.3 Model Development and Hybridization Approaches

Estimating a battery's SOH is essential for precisely forecasting its performance and lifespan, allowing for preventive maintenance, and boosting the dependability of ESS. To increase the accuracy of SOH estimate models, a lot of research has been done. Hybrid modeling techniques have drawn a lot of interest among these initiatives. To improve prediction performance, these strategies integrate the advantages of several methodologies, such as combining several learning models or combining machine learning with model-based techniques. In order to improve robustness and capture the intricate, nonlinear behavior seen in battery degrading processes, hybrid models are intended to address the shortcomings of individual models.

The three main components of the hybrid neural network model suggested by (Qingrui Gong et al. 2021) are a GRU module, an ULSAM module, and a convolutional module. The GRU module processes the feature sequences to estimate the SOC over time after the convolutional and ULSAM modules collaborate to extract features from sequences of sampled voltage, current, and temperature data. Both temporal relationships and deep feature representations in the data are successfully captured by this approach. (Yaxiang Fan et al. 2020) proposed a hybrid modeling approach for battery SOH estimation by combining GRU and CNN. The GRU is utilized to capture temporal dependencies, while the CNN extracts key features from the charging curves under CC-CV mode. This hybrid structure enables the model to effectively learn relevant patterns associated with battery health. (Aliberti et al. 2022) developed and compared various neural network architectures for estimating the SOH of LIBs using a

data-driven approach at the single-cell level. The models evaluated include LSTM, GRU, 1D-CNN, and CNN-LSTM, combining sequential learning and feature extraction techniques to identify the most suitable structure for battery condition analysis. (Li et al. 2023) proposed a novel hybrid model combining TCN and LSTM for SOH and RUL estimations of batteries. This architecture leverages both temporal sequence modeling and deep feature extraction capabilities. Additionally, Bayesian optimization is applied to fine-tune the hyperparameters of each layer in the model. (Chaoran Li et al. 2023) proposed a model centered on creating a hybrid model for estimating batteries' SOH and RUL. TCN and LSTM are combined in its suggested TCN-LSTM architecture to take advantage of sequential learning and temporal feature extraction. We use Bayesian optimization to optimize the model's hyperparameters.

The objective of these projects is to create sophisticated models that use real-world usage data, such as voltage, current, and temperature, to forecast the SOH of LIBs. LSTM networks, which are well-known for their ability to effectively simulate time-series data, are at the heart of both strategies. In order to improve the input representation prior to SOH prediction, (Paul AUDIN et al. 2021) combine LSTM with an autoencoder, which serves as a feature extractor. In order to better represent intricate, nonlinear degradation processes, (Imen Jarraya et al. 2025). suggest a more advanced hybrid model that combines LSTM with KAN. Even if the model architectures differ, both approaches aim to improve prediction accuracy and robustness under realistic and variable operating conditions.

2.4 Feature Extraction and Input Variable Selection

Input feature extraction and selection play a vital role in the accuracy of battery SOH estimation models. Typically, input variables are derived from recorded voltage, current, and temperature data during battery charge and discharge cycles. One widely used technique for feature extraction is ICA, which reveals subtle degradation patterns in the battery. This section provides a summary of prior research focusing on the methods and strategies for selecting and extracting input features, highlighting the diversity of approaches proposed in the literature.

(Daniel Stephens et al. 2024) provide a unique method to enhance the efficiency of Li-ion battery SOH estimate by combining ICA with DTA. When combined, ICA and DTA's complimentary qualities can help overcome the drawbacks of each technique alone, as the work demonstrates. SOH estimation's stability and accuracy are improved by this methodical integration. The results emphasize how crucial ICA is as a tool for obtaining features linked to deterioration, especially when paired with other analytical techniques. (Zhanxin Zou et al. 2023) work uses ICA, a crucial method for gleaning health-related characteristics from LIBs constant current charging phase. PCA is used to reduce the dimensionality of the four HFs that were extracted via ICA. This results in a single principal component that is then utilized for linear regression. To optimize the regression model, a learning mechanism-based Multi-Chaotic Sparrow Search Algorithm is used. This method demonstrates how ICA-derived features may successfully enable correct modeling even in low-resource contexts, highlighting the relevance of ICA in providing lightweight and computationally efficient SOH estimate.

Both studies share a common approach in leveraging ICA to extract features that better reflect battery degradation (Hanlei Sun et al. 2020). utilize ICA to extract peak points and corresponding voltages as additional inputs to a Bi-LSTM model. (Bosong Zou et al. 2023). combine ICA with other features like equal voltage interval discharge time and thermal analysis before feeding them into a Bi-LSTM model with an attention mechanism. Although the implementation differs, both studies agree on the crucial role of ICA in enhancing the accuracy and robustness of SOH estimation. (K. Dhananjay Rao et al. 2024) investigates SOH estimation utilizing a variety of machine learning algorithms, including LSTM, DNN, and GRU, with a focus on optimal hyperparameter optimization. The integration of ICA-based input datasets into the GRU model is a significant study feature. By reflecting realistic internal degradation parameters like battery chemistry and operating conditions, these ICA-derived inputs help the GRU model more closely resemble real-world aging behaviorBoth studies share a common approach in leveraging ICA to extract features that better reflect battery degradation (Hanlei Sun et al. 2020). utilize ICA to extract peak points and corresponding voltages as additional inputs to a Bi-LSTM model. (Bosong Zou et al. 2023). combine ICA with other features like equal voltage interval discharge time and

thermal analysis before feeding them into a Bi-LSTM model with an attention mechanism. Although the implementation differs, both studies agree on the crucial role of ICA in enhancing the accuracy and robustness of SOH estimation. (K. Dhananjay Rao et al. 2024) investigates SOH estimation utilizing a variety of machine learning algorithms, including LSTM, DNN, and GRU, with a focus on optimal hyperparameter optimization. The integration of ICA-based input datasets into the GRU model is a significant study feature. By reflecting realistic internal degradation parameters like battery chemistry and operating conditions, these ICA-derived inputs help the GRU model more closely resemble real-world aging behavior.

(S. Techanok, N. Junhuathon, and K. Chayakulkheeree 2025) investigates the impact of incorporating OCV as an input feature for SOH estimation using an LSTM model trained on battery degradation data from Oxford University. The model uses inputs such as charge/discharge current, voltage, time, and temperature, and the inclusion of OCV is shown to significantly improve estimation accuracy. Additionally, the AUC is used to highlight the differences between battery cycles, providing insights into voltage behavior changes due to degradation. This research underscores the importance of OCV and AUC as critical variables in enhancing SOH prediction performance.

2.5 Battery Dataset

To estimate the SOH of a battery, it is essential to use data collected during the battery's charge and discharge cycles, including voltage, current, and temperature measurements. Many reviewed studies use a variety of datasets some are publicly available open-source datasets, while others are obtained through laboratory experiments conducted by the researchers themselves. These datasets differ in terms of quality, structure, and experimental conditions, which can significantly affect model training and estimation performance. This section provides an overview of the datasets used in related research.

(Jiang Wu et al. 2023) utilized datasets from multiple sources including NASA, CALCE, Oxford, and UL-PUR for the selection of Li-ion battery HFs prior to SOH estimation. The work by (Adrienn Dineva 2024) presents a review of SOH estimation

techniques for LIBs in EVs, pointing out that a number of papers use NASA, CALCE, and Oxford databases as important assessment references. (Vahid Safavi et al. 2024) utilizes datasets from NASA, Oxford, and real-world sources to develop a GPR-LSTM model aimed at improving the efficiency of SOH estimation across various data sources. Many studies that review SOH estimation techniques for LIBs commonly utilize open-source datasets from NASA and Oxford.

From the literature review, it is evident that most research has focused on developing hybrid models in order to minimize the estimation error of battery SOH. While this approach enhances model accuracy, improving the practical usability of such models such as reducing training time, simplifying input data, and lowering computational complexity is equally important and often underexplored.

Therefore, this study proposes the use of a resampling technique to adjust data resolution to a more manageable level, along with the selection of the ICA feature, which is widely recognized in the literature for its effectiveness in reflecting battery degradation behavior. The datasets used in this research are the Oxford Battery Dataset and the NASA Battery Dataset

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.4 Presents a comparison of previous research

Reference	Dataset		ICA	Resampling	Method
	Oxford	NASA			
Yaxiang Fan et al. (2020)	✓	✓		✓	GRU-CNN
Alessandro Aliberti et al. (2022)		✓			LSTM, GRU, 1D-CNN, CNN-LSTM
Paul AUDIN et al. (2021)		✓			Autoencoders and LSTM
Chi Nguyen Van et al. (2023)		✓			LSTM, FNN
Isaiah Oyewole et al. (2022)					CDTL
Zuolu Wang et al. (2022)	✓				GPR
Chaoran Li et al. (2023)	✓	✓			TCN-LSTM
Qingrui Gong et al. (2021)	✓	✓		✓	CNN-ULSAM-GRU
M.S. Hossain Lipu et al. (2022)	✓	✓			Deep Learning
Imen Jarraya et al. (2025)	✓				SOH-KLSTM

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.4 Presents a comparison of previous research (Continued)

Reference	Dataset		ICA	Resampling	Method
	Oxford	NASA			
Yanming Li et al. (2025)		✓			MC-CNN-TimesNet.
Wenting Ma et al. (2025)			✓		Bagged-GPR
Vahid Safavi et al. (2024)		✓		✓	CNN, LSTM
Daniel Stephens et al. (2024)		✓	✓		GPR
Zhiyuan Wei et al. (2025)	✓		✓		MSHHO-DELM
K. Dhananjay Rao et al. (2024)		✓	✓		LSTM, DNN, GRU
Bosong Zou et al. (2023)	✓	✓	✓		Bi-LSTM with AM
Hanlei Sun et al. (2020)		✓	✓		Bi-LSTM
Xiaolei Bian et al. (2022)		✓	✓		ANN
Zhanxin Zou et al. (2023)			✓		BP, PCR
Xin Xiong et al. (2023)	✓	✓	✓	✓	GPR, LSTM

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.4 Presents a comparison of previous research (Continued)

Reference	Dataset		ICA	Resampling	Method
	Oxford	NASA			
Liang Zhang et al. (2023)	✓	✓			Ensemble
Jiang Wu et al. (2023)	✓	✓			GPR, LSTM, BP
Lijun Zhang (2023)		✓			LSTM
Adrienn Dineva (2024)	✓	✓			ML and AI
A.G. Olabi et al. (2024)	✓	✓			ANN, FNN, CNN
S. Techanok, N. Junhuathon and K. Chayakulkheeree (2025)	✓				LSTM
Proposed Framework	✓	✓	✓	✓	LSTM, Bi-LSTM, AM-LSTM, GRU, GPR

Based on Table 2.4, which presents a comparative review of previous research related to battery SOH estimation including aspects such as model selection, data preparation techniques, and dataset usage a set of summary charts has been created to clearly illustrate the prevailing research trends. Figure 2.1 shows the proportion of studies using datasets such as Oxford and NASA. Figure 2.2 summarizes the application of feature extraction techniques, notably ICA and resampling, while Fig. 2.3 compares the frequency of use for various model types, including LSTM, Bi-LSTM, AM-LSTM, GRU,

and GPR. Together, these figures provide a comprehensive overview of common practices in SOH estimation.

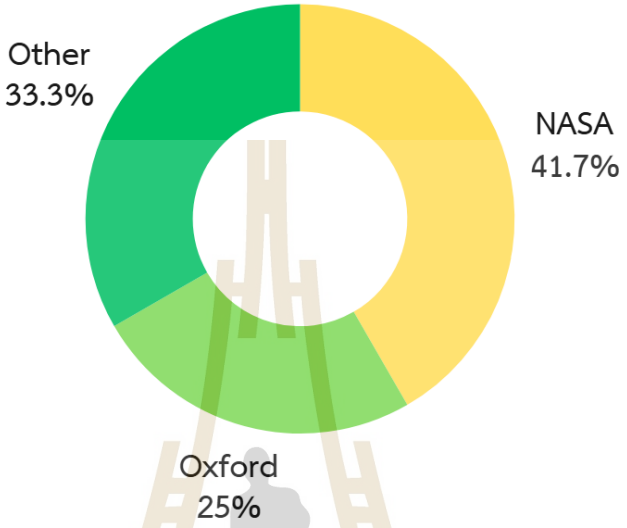


Figure 2.1 The comparison of battery datasets

Figure 2.1 shows the distribution of battery datasets used in the reviewed research. The NASA dataset is the most frequently used, appearing in 41.7% of the studies, followed by the Oxford dataset at 25%. Other datasets account for 33.3%.

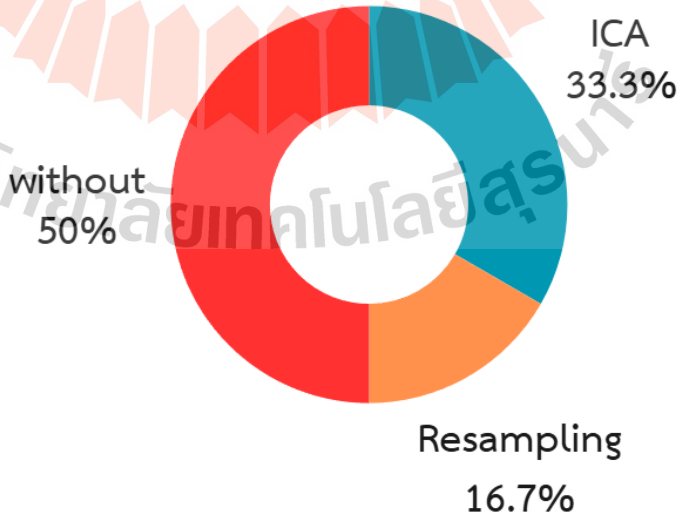


Figure 2.2 The comparison of feature engineering techniques

Figure 2.2 illustrates the use of data preprocessing techniques in the reviewed studies. Among all studies, 33.3% applied ICA for feature extraction, while only 16% employed resampling techniques. Notably, 50% of the studies did not apply either ICA or resampling, indicating that such preprocessing methods are not yet widely adopted compared to the focus on model development.

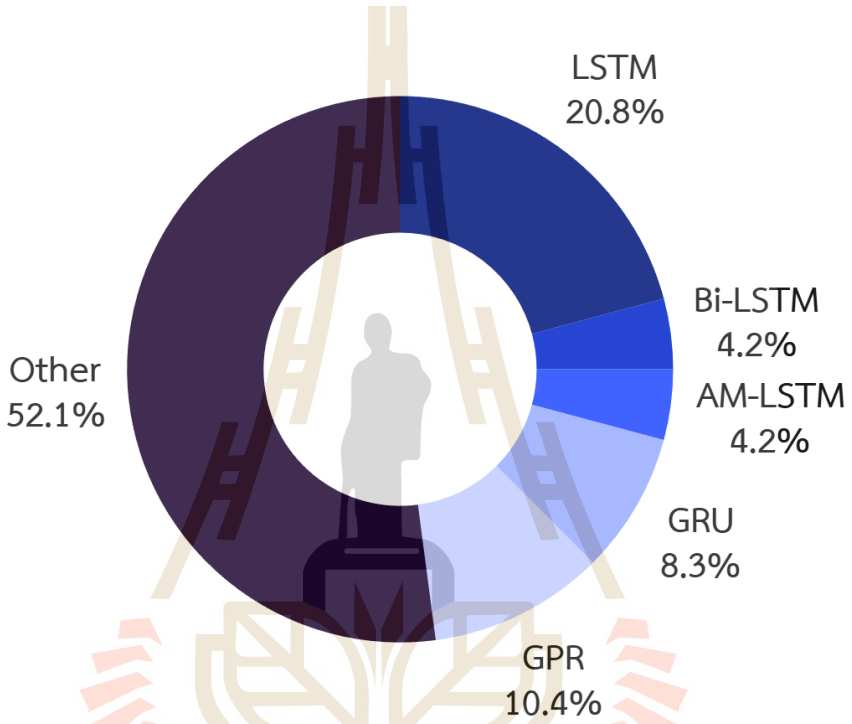


Figure 2.3 The comparison of model usage rates

Figure 2.3 illustrates the distribution of model types used for SOH estimation. Among them, LSTM is the most frequently adopted, accounting for 20.8% of the studies, followed by Bi-LSTM and AM-LSTM at 4.2%, GRU at 8.3% and GPR at 10.4%. Other models not included in the main categories make up 52.1% of the total.

2.6 Battery Degradation

The degradation of LIBs is an inevitable process that occurs progressively with repeated charge–discharge cycles. It arises from both electrochemical and structural changes within the cell, which gradually deteriorate its performance over time. During normal operation, lithium ions shuttle between the cathode and anode through the electrolyte. Over many cycles, several irreversible phenomena occur, including:

1) Formation of the SEI: This layer forms on the anode surface as a result of side reactions between the electrolyte and electrode materials. Although the SEI layer initially helps stabilize the interface, it continues to grow with cycling, increasing internal resistance and hindering Li-ion transport.
2) Loss of Active Lithium Ions: Some lithium ions become trapped within the SEI layer or participate in unwanted side reactions, reducing the number of active ions available for charge transfer. This directly contributes to the gradual loss of battery capacity.
3) Electrode Structural Deformation: The repetitive expansion and contraction of electrode materials during lithiation and delithiation can lead to particle cracking or detachment, decreasing the effective reaction surface area.

Lithium Plating: Under conditions such as fast charging or low temperature, metallic lithium may deposit on the anode surface instead of intercalating into the carbon structure. This not only accelerates capacity fading but also poses safety risks due to potential short circuits.

The degradation of LIBs, as described earlier, directly affects their performance, particularly in terms of capacity and internal resistance. These two parameters have long been the main indicators used to evaluate the SOH of a battery before modern estimation techniques were introduced. In the early stages of battery research, SOH assessment relied heavily on experimental and electrochemical analysis methods. Researchers often measured the loss of capacity after each full charge–discharge cycle or observed the increase in internal resistance using techniques such as Electrochemical Impedance Spectroscopy (EIS) or DC resistance measurement. Some studies also analyzed open-circuit voltage (OCV) curves to identify chemical or

structural changes inside the cell. Although these traditional approaches provided accurate insights into degradation behavior, they required time-consuming experiments and specialized laboratory equipment. Moreover, they were not suitable for real-time monitoring, as the battery had to be taken out of operation for each test. (Berecibar et al. (2016))

To overcome these limitations, researchers later developed model-based estimation methods, such as the Equivalent Circuit Model (ECM) and Electrochemical Model (EM), which aimed to represent battery behavior through mathematical and physical equations. However, these models often required complex parameter identification and detailed information about internal cell characteristics, which are difficult to obtain in practice.

With the advancement of computing power and data availability, the focus gradually shifted toward data-driven approaches, which estimate SOH by analyzing measurable parameters such as voltage, current, temperature, and time during battery operation. Instead of relying on physical modeling, these methods allow the system to learn degradation patterns directly from operational data, leading to more flexible and efficient SOH estimation.

มหาวิทยาลัยเทคโนโลยีสุรนารี

2.7 Definition and Calculation of SOH

SOH is a measure of the overall condition of a battery compared to its ideal or new state. It reflects the battery's ability to store and deliver energy effectively. SOH is expressed as a percentage, where 100% represents a battery in perfect condition, and lower percentages indicate varying degrees of degradation. SOH is influenced by various factors such as the number of charge-discharge cycles, temperature, and usage patterns. It helps in assessing the battery's performance and predicting its RUL. SOH can be calculated using the following formula in Equation (2.1).

$$\%SOH = \frac{C_{remain}}{C_{initial}} \tag{2.1}$$

Where: C_{remain} is the maximum charge, the battery can hold at its current state. $C_{initial}$ is the rated capacity of the battery when it was new. As the battery ages, C_{remain} decreases with increased usage.

2.8 Model Definition

Deep learning models come in various forms, and choosing the appropriate one depends on the nature of the data and the objectives of the analysis. In the context of battery charge/discharge processes, it is essential to employ models that can learn sequential patterns and temporal dependencies. This ability allows the model to detect subtle variations that reflect battery degradation more accurately. Therefore, this section introduces the selected model used for assessing battery performance based on the extracted features.

2.8.1 LSTM Model Architecture

The LSTM neural network is a type of RNN developed to address the vanishing gradient problem, which commonly occurs in traditional RNNs when dealing with long or complex time sequences. LSTM is designed to efficiently remember and selectively retain important information across time sequences.

The structure illustrated in the figure represents the internal architecture of a LSTM cell, which is composed of three main gates: the Forget Gate,

Input Gate, and Output Gate. These gates collectively regulate the flow of information through time, enabling the network to retain long-term dependencies while discarding irrelevant information effectively. The overall architecture of the LSTM model used in this study is shown in Fig. 2.4.

1) Forget Gate (f_t) This gate determines which information from the previous cell state (C_{t-1}) should be retained or discarded. A sigmoid activation function (σ) is applied to the combination of the current input (x_t) and the previous hidden state (h_{t-1}), producing a value between 0 and 1. A value closer to 0 indicates that the information in C_{t-1} is largely forgotten, while a value near 1 means it is mostly preserved.

2) Input Gate (i_t, g_t) The input gate controls how much new information from the current input $x_t$and the previous hidden state h_{t-1}should be added to the cell state. The sigmoid function (σ) computes the input weight (i_t), while the tanh activation generates a candidate input (g_t). These two components are then multiplied together and integrated into the cell state, allowing selective updating of memory with relevant information.

3) Output Gate (o_t) The output gate decides which part of the cell state will be passed to the hidden state h_t. A sigmoid activation generates the output weight (O_t), while the updated cell state C_t passes through a tanh activation to produce the final hidden state $h_t = o_t \times \tanh(C_t)$, which serves as the LSTM's output at time step t.

4) Cell State Update (C_t) The updated cell state $C_t$is obtained by combining two components:

- The retained portion of the previous state ($f_t \times C_{t-1}$)
- The newly added information from the input gate ($i_t \times g_t$)

This mechanism enables the LSTM to preserve essential long-term information across multiple time steps.

5) Fully Connected Layer and SOH Estimation The resulting hidden state h_t is then fed into a fully connected layer, which captures both linear and nonlinear relationships among features before producing the final output representing the estimated SOH of the battery.

These mechanisms enable LSTM to effectively learn temporal relationships without losing important information from previous sequences, while also discarding irrelevant data. This makes it highly suitable for processing time-series data, such as voltage, current, and temperature measurements of batteries, which vary continuously over charge and discharge cycles. For further details, see (Nguyen Van & Quang, 2023)

Therefore, LSTM is an appropriate model for estimating the SOH of batteries, as it can capture gradual degradation patterns occurring across multiple charge–discharge cycles.

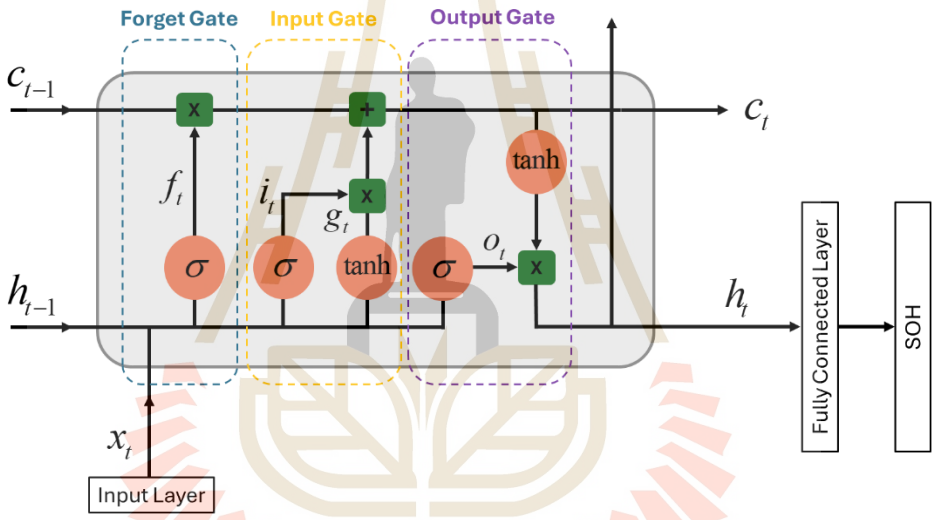


Figure 2.4 An example of the LSTM cell

2.8.2 Bi-LSTM Model Architecture

The Bi-LSTM network is an advanced extension of the conventional LSTM architecture designed to enhance the model's ability to learn temporal dependencies from both past and future contexts. In a Bi-LSTM layer, two parallel LSTM units operate simultaneously: one processes the input sequence in the forward direction (from past to future), while the other processes it in the reverse direction (from future to past). The outputs from both directions are then concatenated to form a unified representation that captures bidirectional temporal information within the sequence.

The architecture of the Bi-LSTM model, as illustrated in the Fig. 2.5, demonstrates the bidirectional processing of sequential data using two parallel LSTM layers: the Forward Layer and the Backward Layer. This structure allows the model to capture temporal dependencies in both directions, enhancing its ability to understand complex time-series relationships.

In the Forward Layer, data are processed sequentially from left to right, where each input at time step x_t is propagated to the next time step $t + 1$, allowing the model to learn temporal patterns in the normal direction. Meanwhile, the Backward Layer processes the same sequence in the opposite direction from right to left enabling the model to retain contextual information from both past and future time steps simultaneously.

At each time step, the outputs from both directions are concatenated and passed through an activation layer to form a comprehensive feature representation that integrates information from the entire sequence. This bidirectional mechanism enables the Bi-LSTM to better capture dependencies that a unidirectional LSTM might miss, leading to smoother and more context-aware predictions.

For the SOH estimation task, the combined output from the forward and backward layers is passed through a fully connected layer to produce the final estimation. This structure enhances the model's capability to learn the complex, nonlinear degradation behavior of batteries over time, resulting in improved accuracy and robustness in prediction performance. For further details, see (B. Zou et al., 2023)

Although incorporating the backward direction increases the model's structural complexity and requires more training time, the benefit lies in its enhanced ability to capture temporal patterns with greater smoothness and precision than the standard LSTM. This makes it particularly suitable for estimation tasks that rely heavily on the continuity and consistency of time-series data.

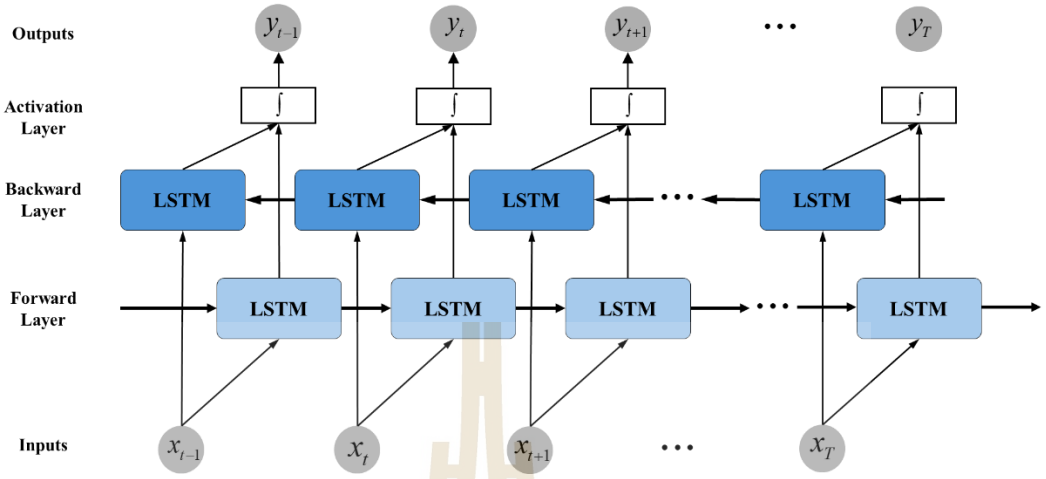


Figure 2.5 The Bi-LSTM network architecture

2.8.3 AM-LSTM Model Architecture

The AM-LSTM network is an enhanced architecture designed to improve the model's ability to focus on the most relevant parts of a temporal sequence. The attention mechanism acts as a dynamic weighting process that evaluates the importance of each time step with respect to the prediction target. Instead of treating all input information equally, the model assigns higher weights to time steps that contribute more significantly to the output while reducing the influence of less relevant data.

The structure of the Attention-based LSTM model, as illustrated in the Fig. 2.6, By integrating an Attention Layer, the model can "focus" on the most relevant time steps that significantly influence the prediction, instead of treating all input sequences with equal importance as in a standard LSTM.

Within the LSTM Layer, the model processes the sequential inputs $x_t, x_{t+1}, \ldots, x_T$ to generate a series of hidden states $(h_t, h_{t+1}, \ldots, h_T)$, each capturing contextual information at a specific time step. These hidden states are then fed into the Attention Mechanism, which computes the importance weights (attention scores) corresponding to each temporal step. In the Attention Layer, the tanh activation function is used to transform each hidden state into a relevance vector that reflects its contribution to the output. Then, the Softmax function normalizes these values into a probability distribution, ensuring that the attention weights sum to one. This allows the model to determine the relative importance of each time step. The final

context vector (Z) is obtained by computing the weighted sum of all hidden states according to their respective attention weights.

Subsequently, the output from the Attention Layer is passed to a Fully Connected Layer, which performs the final regression to estimate the SOH. The incorporation of the Attention mechanism enables the model to effectively capture critical patterns in specific time intervals that reflect battery degradation behavior, thereby improving the accuracy and robustness of SOH estimation compared to the conventional LSTM model. For further details, see (B. Zou et al., 2023)

In the context of battery SOH estimation, the attention mechanism allows the model to concentrate on critical moments such as abrupt voltage or current variations that indicate degradation behaviors. This selective focus leads to more accurate and robust SOH predictions, as the model can better capture the underlying dynamics of battery aging over multiple charge-discharge cycles.

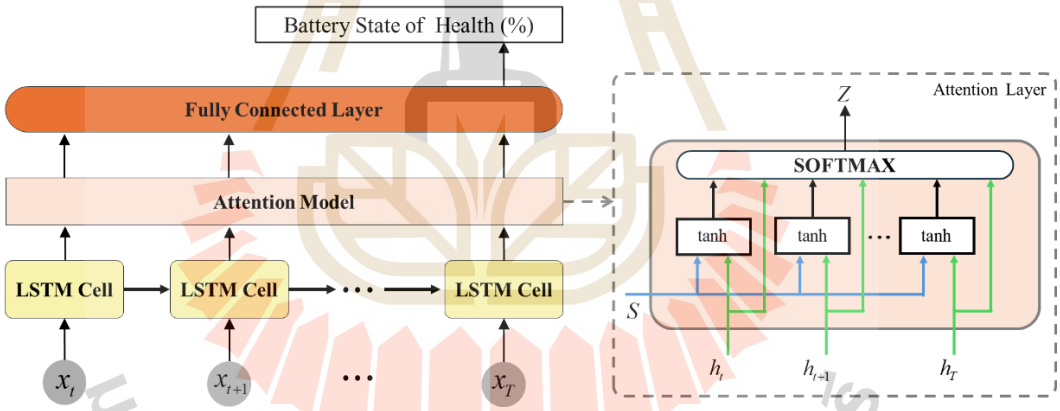


Figure 2.6 The ATT-LSTM network architecture

2.8.4 GRU Model Architecture

The GRU is an advanced RNN architecture designed to address the vanishing gradient problem commonly encountered in traditional RNNs. Similar in concept to the LSTM network, GRU introduces a simplified yet efficient structure that reduces computational complexity and the number of trainable parameters, while still maintaining the ability to capture long-term dependencies within sequential data. Structurally, GRU employs two main gating mechanisms: the update gate and the reset gate.

1) Reset Gate (r_t), This gate acts like a filter for past information. It takes the current input (x_t) and the previous hidden state (h_{t-1}) to decide how much of the old memory (h_{t-1}) is irrelevant to the current input. Essentially, it determines what to "forget" or "reset" from the past before the cell calculates its new candidate information ($\tilde{h}_t$).

2) Update Gate (z_t), This gate is key for blending the data. The update gate (z_t) learns (from $x_t$and h_{t-1}) to decide the "proportion" of new information to add. Its value (between 0 and 1) determines how much of the old memory (h_{t-1}) to keep (the $1 - z_t$ part) and how much of the new candidate information ($\tilde{h}_t$) to accept (the z_t part)

3) New Hidden State (h_t), This is the final output for this time step. It's the "merged" result of the old and new information, based on the proportions set by the update gate (z_t) in the equation $h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$. This h_t becomes the new "updated memory," which is then passed on to the next time step ($t - 1$) to continue analyzing the battery's SOH trend.

The architecture of the GRU is presented in Fig. 2.7. Unlike LSTM, GRU eliminates the separate memory cell state and relies solely on the hidden state to store and update information. For further details, see (Fan et al., 2020). This design simplification allows for faster training and reduced computational cost, while still achieving comparable accuracy in learning long-term temporal relationships. Due to its efficiency and compact architecture, GRU often provides faster convergence when handling large or repetitive datasets, such as charging–discharging sequences in battery systems.

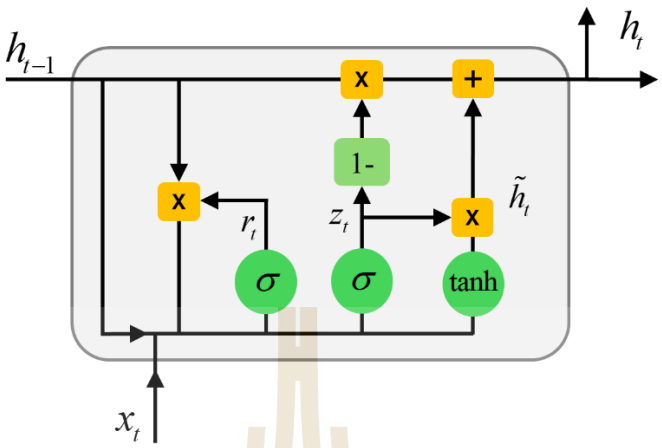


Figure 2.7 The GRU network architecture

2.8.5 GPR Model Architecture

GPR is a statistical technique based on the concept of the Gaussian distribution, which is used to model the relationship between input and output variables without requiring an explicit predefined functional form. GPR can be viewed as an extension of linear regression to a more flexible, non-parametric approach by defining a Gaussian process over possible functions rather than focusing solely on specific model parameters. The core principle of GPR lies in the covariance function, also known as the kernel function, which determines how similar or related different data points are to one another. This kernel-based formulation allows GPR to capture complex and nonlinear relationships in the data effectively. By computing the correlations between all input samples, GPR constructs a probabilistic model that provides smooth and consistent predictions even when the data distribution is irregular or noisy. One of the main advantages of GPR is its ability to provide both predicted values and uncertainty estimates simultaneously. This characteristic is particularly beneficial in battery SOH estimation, where understanding the model's confidence level in each prediction is as important as the prediction itself. The uncertainty quantification enhances the interpretability and reliability of the estimation results. For further details, see (Wang et al., 2022)

However, GPR has a major limitation in terms of computational cost when dealing with large datasets, since it must compute correlations between all pairs

of data points. As the dataset size increases, the amount of required computation grows significantly, resulting in longer training times and higher memory consumption. Therefore, GPR is generally more suitable for problems with smaller datasets or for use as a benchmark model to compare against deep learning-based approaches in terms of accuracy and interpretability.

Based on the reviewed literature, various models have been employed for estimating the SOH of batteries, ranging from statistical to deep learning approaches. Each model possesses distinct principles and structural designs to handle the complexity and time-dependent characteristics of battery data. Deep learning models such as LSTM, Bi-LSTM, GRU, and Attention-based LSTM have demonstrated strong capabilities in capturing temporal dependencies and degradation behaviors effectively, while the statistical model GPR offers advantages in interpretability and in providing uncertainty estimates for predictions. Comparing these models highlights their respective strengths and limitations, providing a clearer understanding of different methodological perspectives in battery health estimation.

2.9 Evaluation Metrics

To evaluate the performance of the models used for estimating the battery SOH, which is expressed as a percentage (%SOH), several evaluation metrics are employed to assess both accuracy and consistency. In this study, four commonly used metrics are selected: MSE, RMSE, MAE, and MAPE. Each of these metrics provides a different perspective on the model's estimation capability.

2.9.1 Root Mean Squared Error

RMSE, defined in Equation (2.2), is the square root of MSE and provides an interpretable measure of error in the same unit as the output variable (%SOH). It offers an intuitive understanding of the model's estimation precision and is sensitive to large errors. RMSE is especially useful in this study for evaluating how well the model can maintain consistent estimation accuracy across different degradation stages.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} SOH(i) - SOH_e(i))^2} \quad (2.2)$$

Where n is a number of cycles SOH, is actual capacity, SOH_e is estimated capacity.

2.9.2 Mean Squared Error

MSE measures the average of the squared differences between the estimated and actual SOH values, as shown in Equation (2.3). This metric emphasizes larger deviations by squaring the error terms, making it particularly useful for identifying models that minimize large prediction discrepancies. A lower MSE value indicates that the model's estimations are closer to the actual SOH across the dataset.

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (SOH(i) - SOH_e(i)) \quad (2.3)$$

2.9.3 Mean Absolute Error

MAE represents the average magnitude of the absolute differences between estimated and actual SOH values, as expressed in Equation (2.4). Unlike MSE or RMSE, MAE does not square the differences, making it less sensitive to outliers. This metric highlights the model's general estimation capability and provides an interpretable measure of average error that directly corresponds to the deviation percentage of SOH.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |SOH(i) - SOH_e(i)| \quad (2.4)$$

2.9.4 Mean Absolute Percentage Error

MAPE, formulated in Equation (2.5), expresses the average estimation error as a percentage of the actual SOH values. This allows for straightforward comparison between datasets with different scales. Since the SOH output in this study is represented in percentage form (%SOH), MAPE provides an intuitive metric for understanding how far, on average, the estimated battery health deviates from its true value.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{SOH(i) - SOH_e(i)}{SOH(i)}\right| \times 100\% \quad (2.5)$$

Using all four metrics provides a comprehensive evaluation of model accuracy, covering aspects such as absolute error, squared error, and relative error in percentage terms.

มหาวิทยาลัยเทคโนโลยีสุรนารี

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.

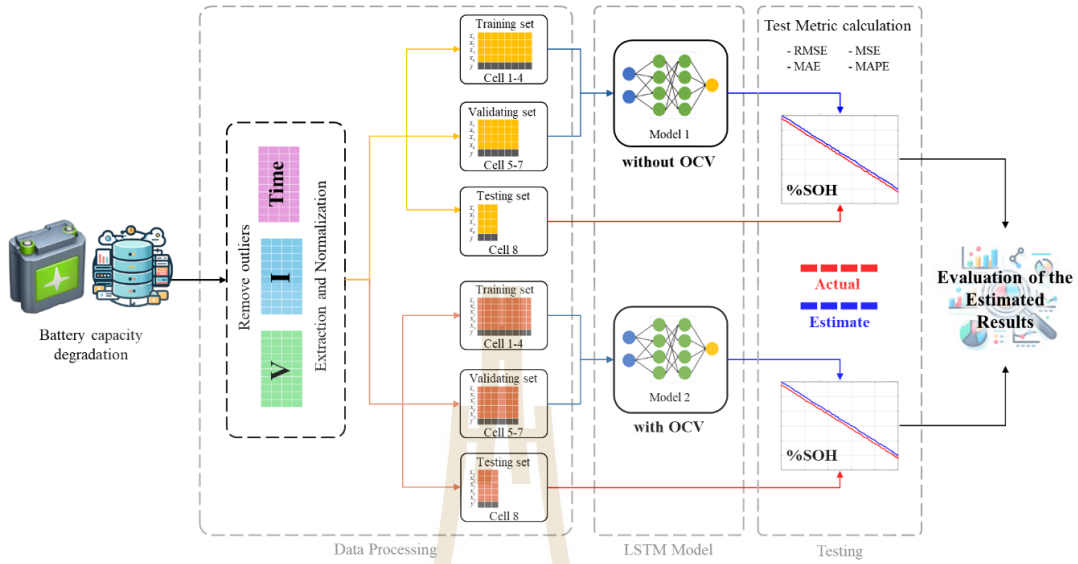


Figure 3.1 Framework of comparative analysis between LSTM models with and without OCV feature inclusion

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.

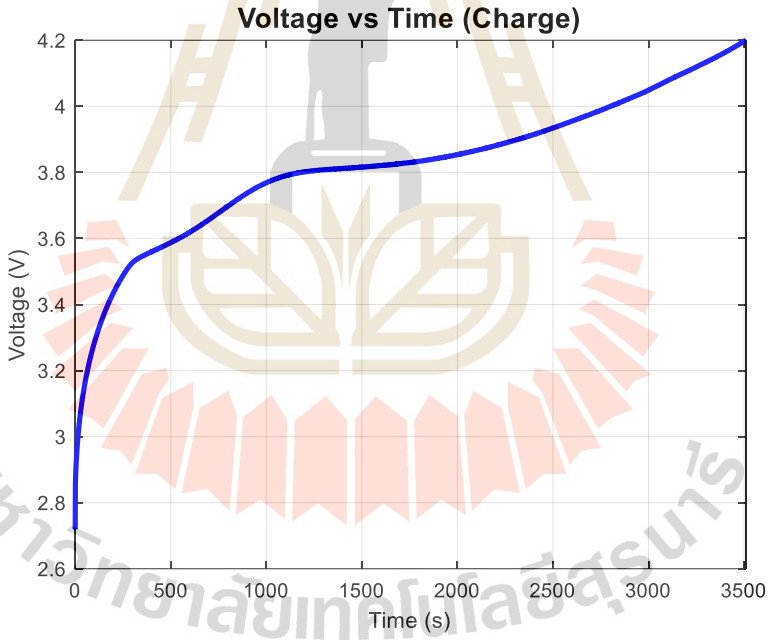


Figure 3.2 Voltage profiles of Cell 1 during charging cycle from the Oxford Battery Dataset

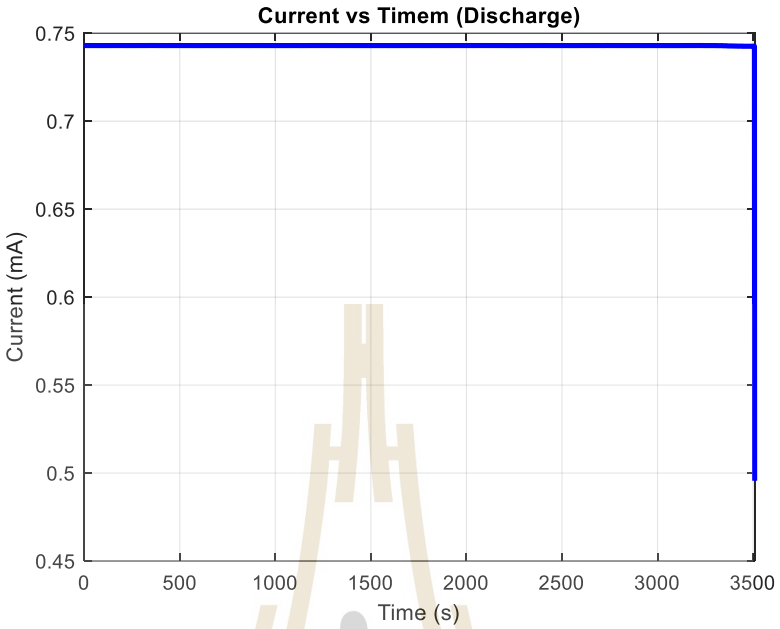


Figure 3.3 Current profiles of Cell 1 during charging cycle from the Oxford Battery Dataset

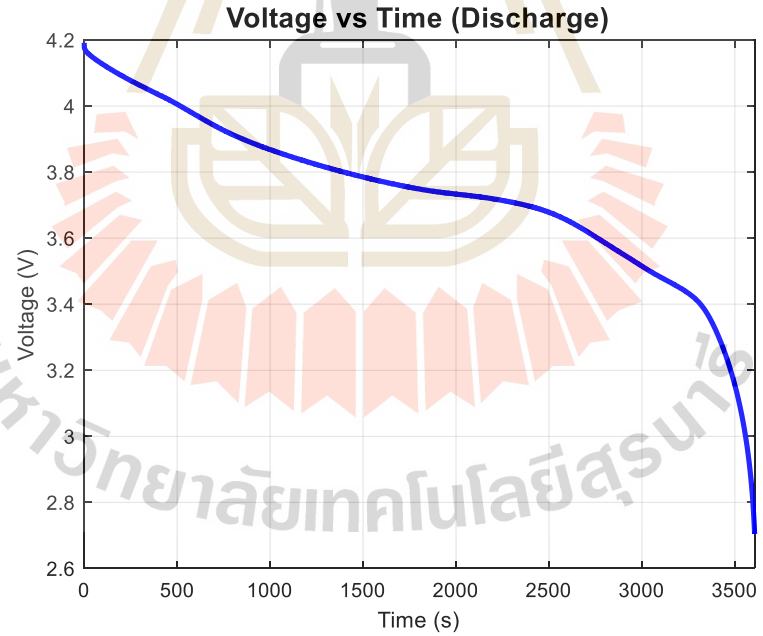


Figure 3.4 Voltage profiles of Cell 1 during discharging cycle from the Oxford Battery Dataset

มหาวิทยาลัยเทคโนโลยีสุรนารี

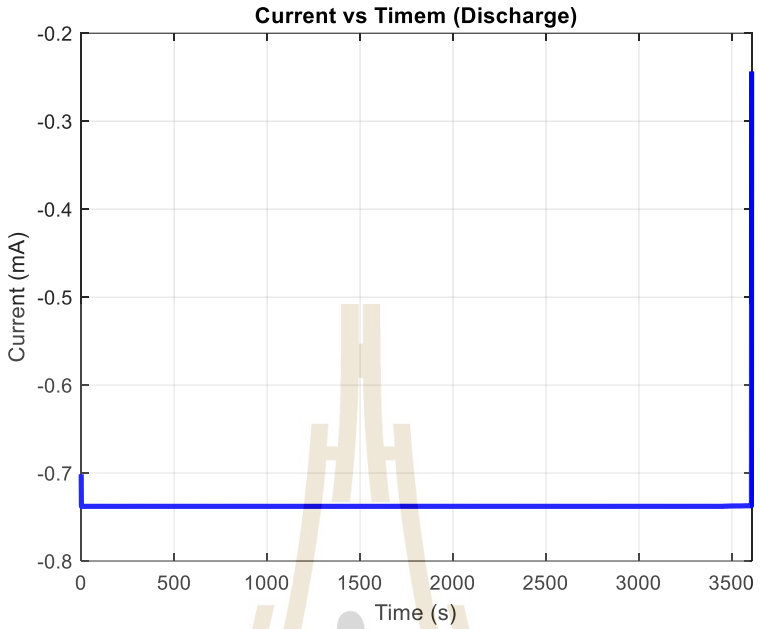


Figure 3.5 Current profiles of Cell 1 during discharging cycle from the Oxford Battery Dataset

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.

มหาวิทยาลัยเทคโนโลยีสุรนารี

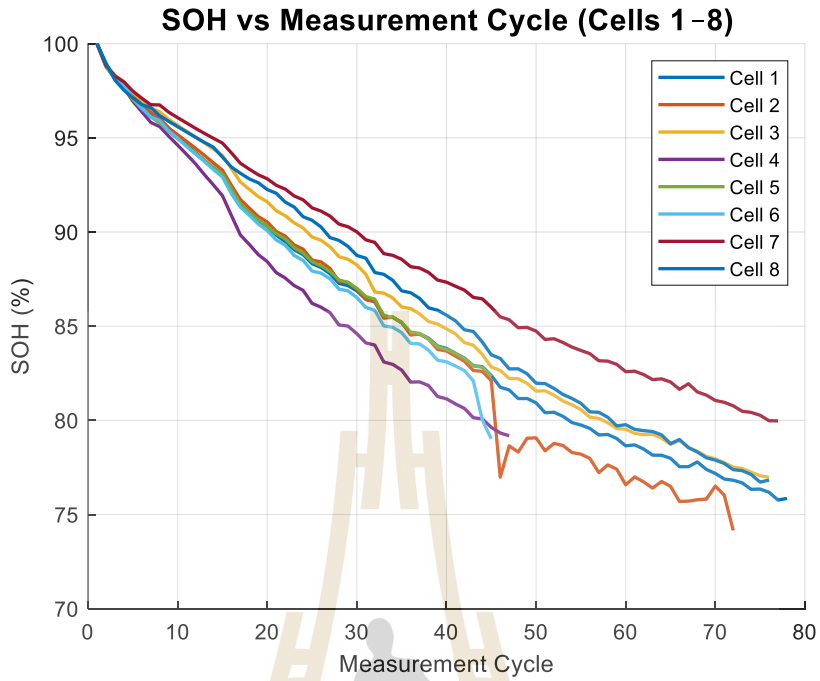


Figure 3.6 SOH degradation curves of all 8 cells from the Oxford Battery Degradation Dataset

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.

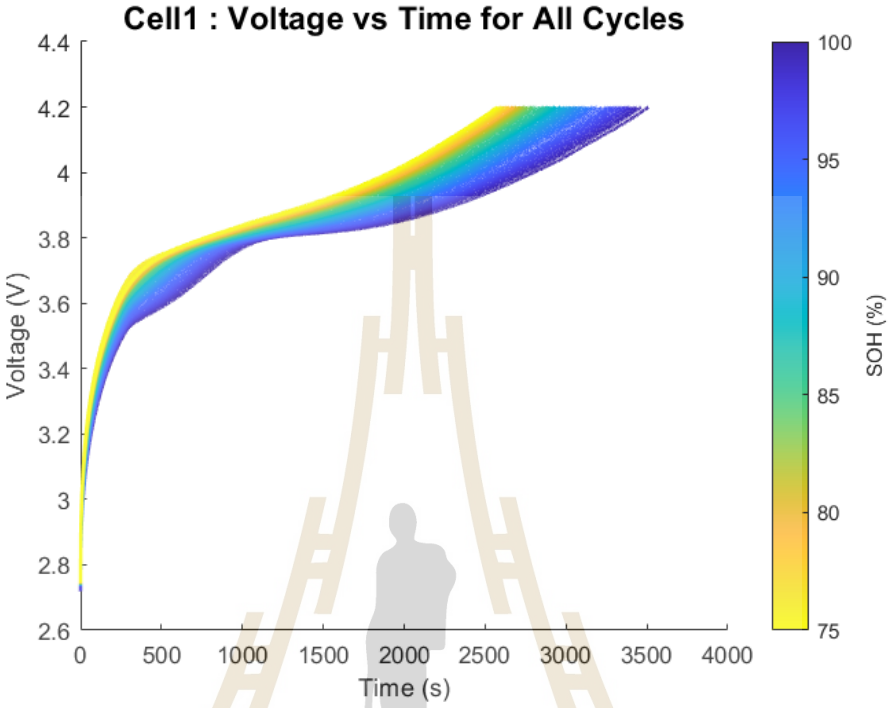


Figure 3.7 Voltage profiles Cell 1 of Oxford are used to charge all the cycles.

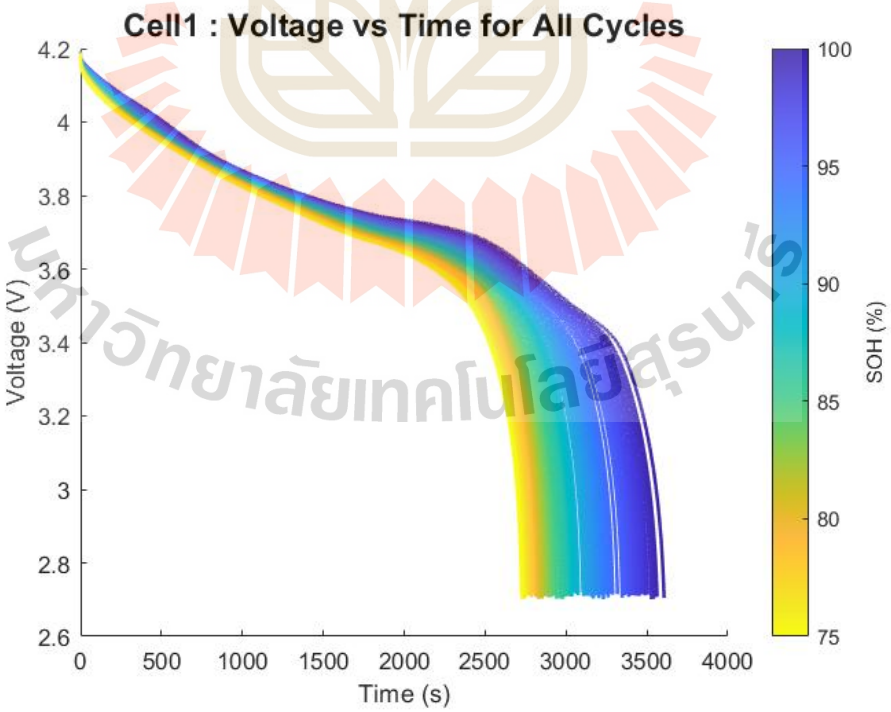


Figure 3.8 Voltage profiles Cell 1 of Oxford are used to discharge all the cycles.

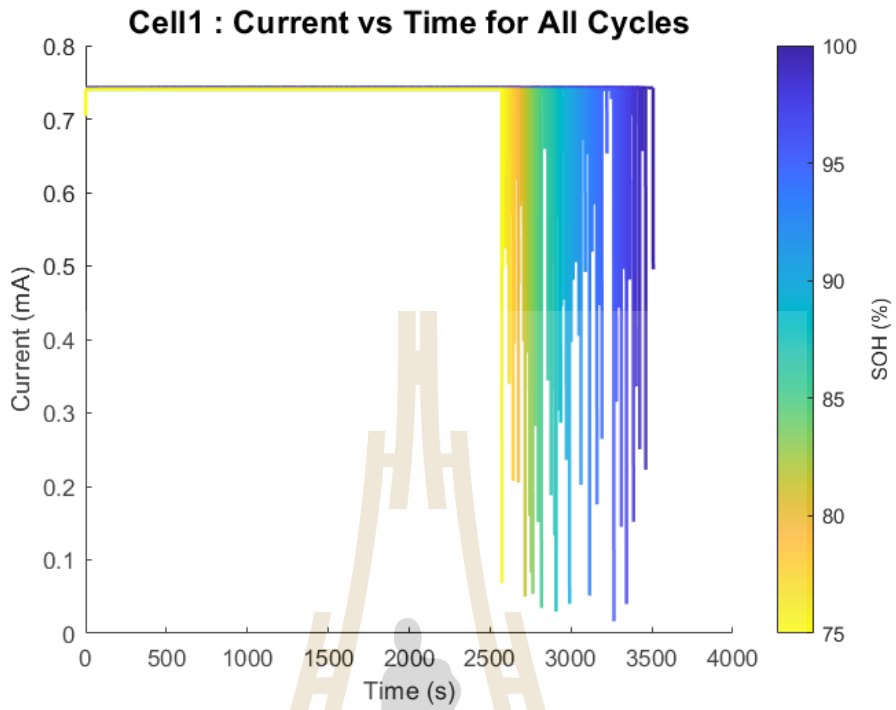


Figure 3.9 Current profiles Cell1 of Oxford are used to charge all the cycles.

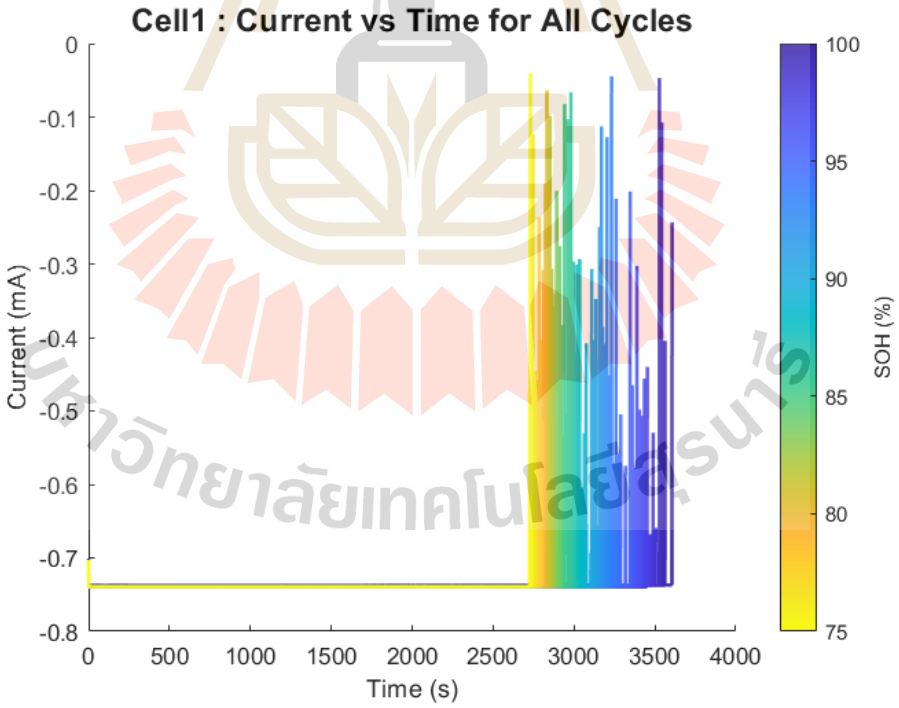


Figure 3.10 Current profiles Cell1 of Oxford are used to discharge all the cycles.

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.

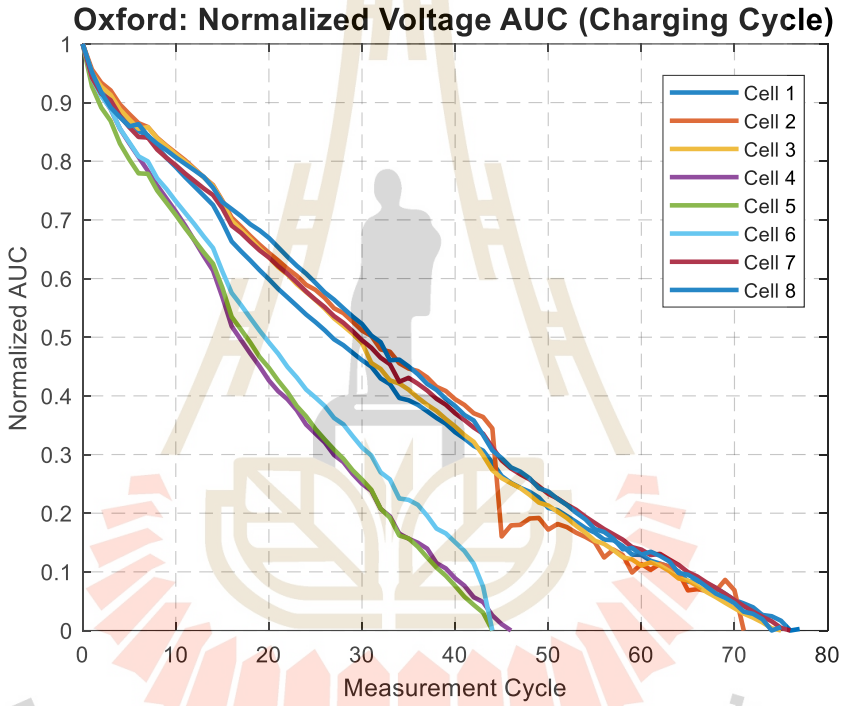


Figure 3.11 Normalized AUC of voltage during the charging cycle from the Oxford dataset

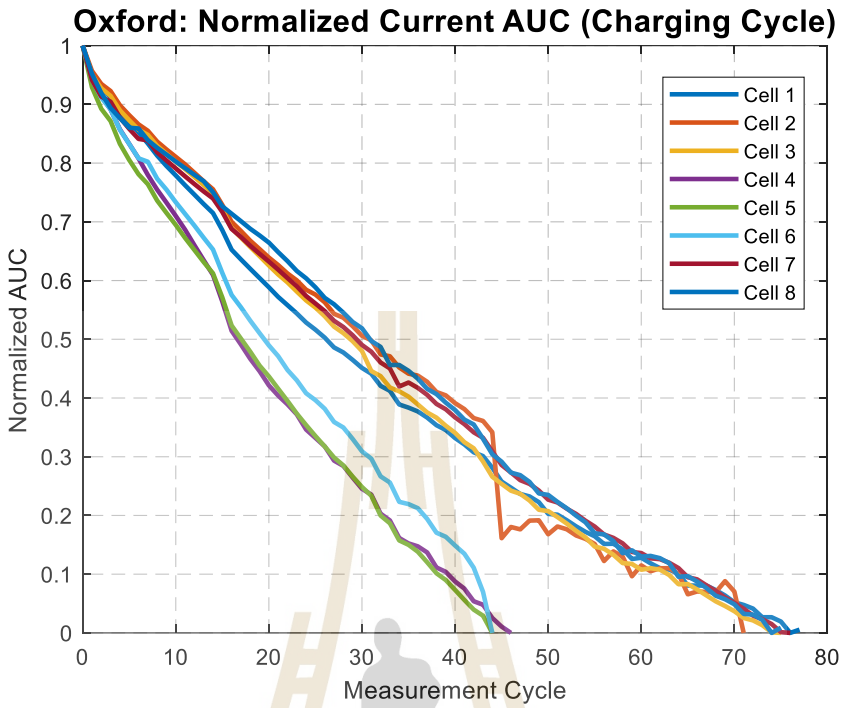


Figure 3.12 Normalized AUC of current during the charging cycle from the Oxford dataset

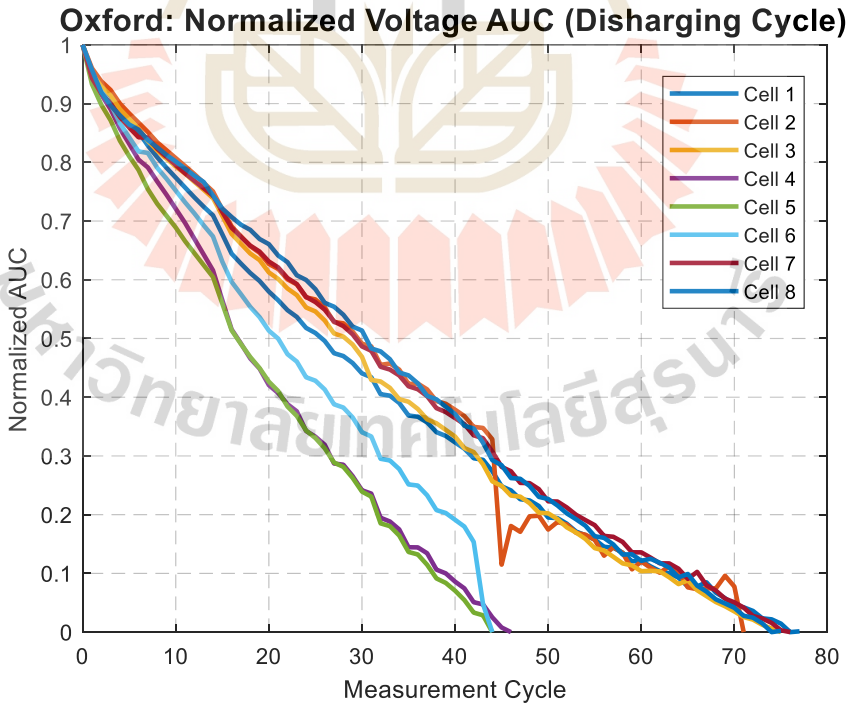


Figure 3.13 Normalized AUC of voltage during the discharging cycle from the Oxford dataset.

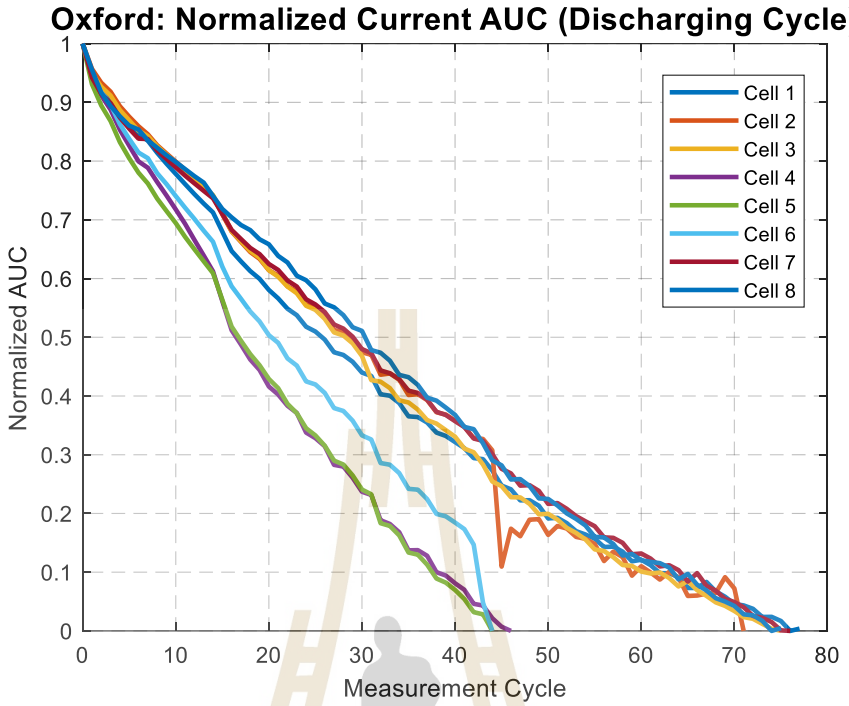


Figure 3.14 Normalized AUC of current during the discharging cycle from the Oxford dataset

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.

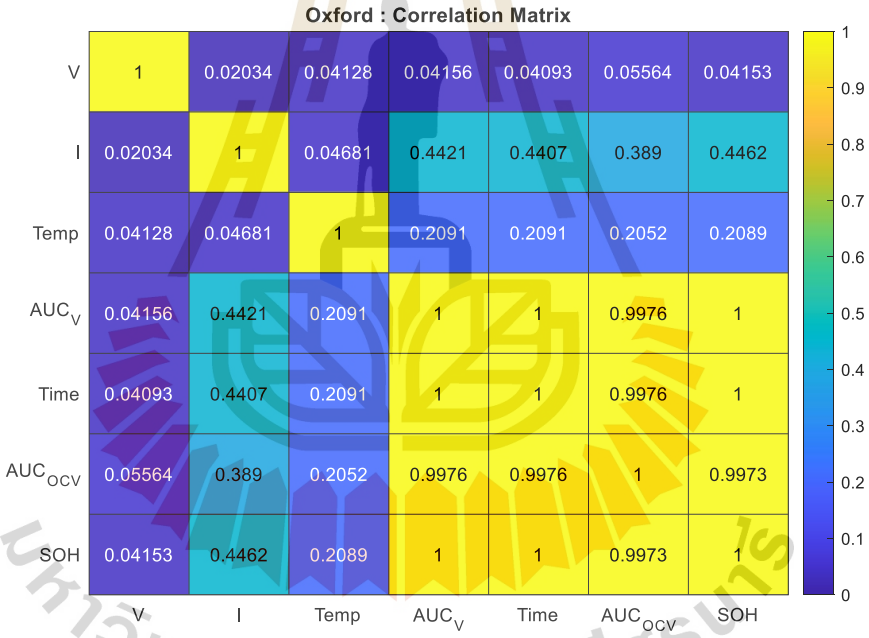


Figure 3.15 The correlation matrix

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.

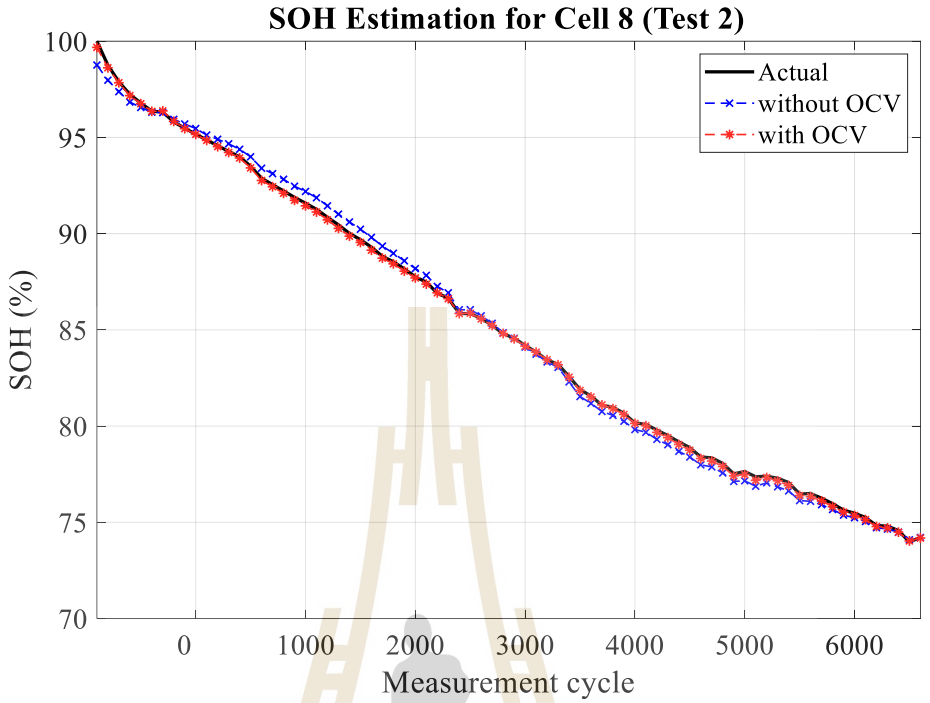


Figure 3.16 SOH estimation of Cell 8 from the Oxford dataset

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.

มหาวิทยาลัยเทคโนโลยีสุรนารี

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

Test	Model	Metrics			
		RMSE	*MSE*	*MAE*	*MAPE*
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.

มหาวิทยาลัยเทคโนโลยีสุรนารี

CHAPTER IV

INVESTIGATING THE IMPACT OF DATA RESAMPLING ON MODEL ACCURACY AND COMPLEXITY

4.1 Introduction

This chapter details the core experimental methodology and results of this study. The focus is on the application and evaluation of data resampling techniques for SOH estimation. The overall experimental framework for this chapter is illustrated in Fig. 4.1. The methodology will be applied to two distinct battery datasets: the Oxford Battery Degradation Dataset and the NASA Prognostics dataset. The performance of this approach will be systematically benchmarked across five different predictive models: LSTM, Bidirectional-LSTM (Bi-LSTM), AM-LSTM, GRU, and GPR.

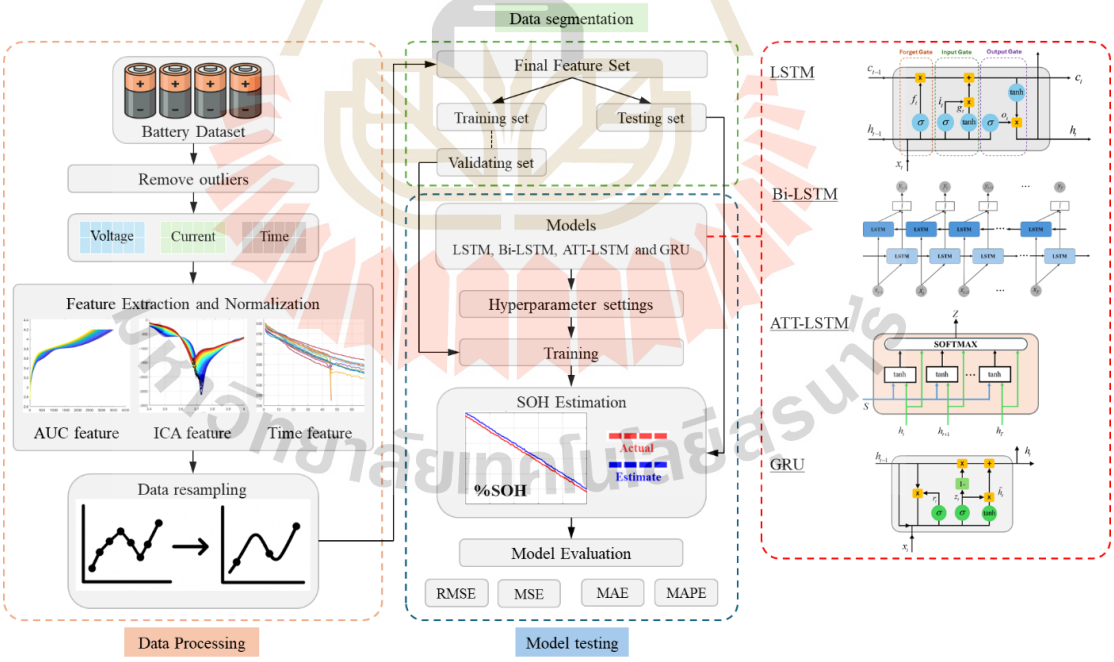


Figure 4.1 Experimental workflow for SOH estimation using data resampling.

4.2 Battery Datasets Utilized

In addition to the Oxford Battery Degradation Dataset, which was used and detailed in Chapter III, this chapter also incorporates the NASA Prognostics dataset for analysis and evaluation.

4.2.1 NASA Prognostics Center of Excellence Dataset

The NASA Prognostics Center of Excellence (PCoE) dataset comprises ageing data collected from four LIBs: B0005, B0006, B0007, and B0018. These cells were subjected to three operational modes: charging, discharging, and impedance measurements, all conducted under room temperature conditions. During charging, each battery was charged using a constant current (CC) mode at 1.5 A until the terminal voltage reached 4.2 V, after which charging continued in a constant voltage (CV) mode until the charging current decreased to 20 mA. The characteristic charging voltage profile is illustrated in Fig. 4.2, and the corresponding current profile is shown in Fig. 4.3. Discharge testing was performed using a constant current of 2 A, and the discharge process was terminated once the cell voltage dropped to predetermined cut-off levels, which were 2.7 V for cell B0005, 2.5 V for cells B0006 and B0018, and 2.2 V for cell B0007. Typical discharge voltage and current profiles are presented in Fig. 4.4 and Fig. 4.5, respectively.

Each battery's data was stored in a separate .mat file, and the recorded parameters included the measured voltage, current, and temperature, along with the corresponding charging current, charging voltage, and the timestamp in seconds. In this study, only the measured voltage and measured current were selected for both the charging and discharging cycles to ensure a consistent basis for analysis. It should be noted that this dataset contains certain limitations. Some cells exhibit inconsistencies in data logging, and a few segments show the presence of noise. For example, in some charging cycles, even after the cell voltage had reached 4.2 V and the current dropped to zero, the voltage occasionally decreased slightly, a phenomenon that is considered non-ideal and treated as noise. Similarly, during discharging, some records continued even after the battery had reached its designated cut-off voltage. To ensure data

integrity, all such anomalous segments were removed during a dedicated data-cleaning phase prior to feature extraction.

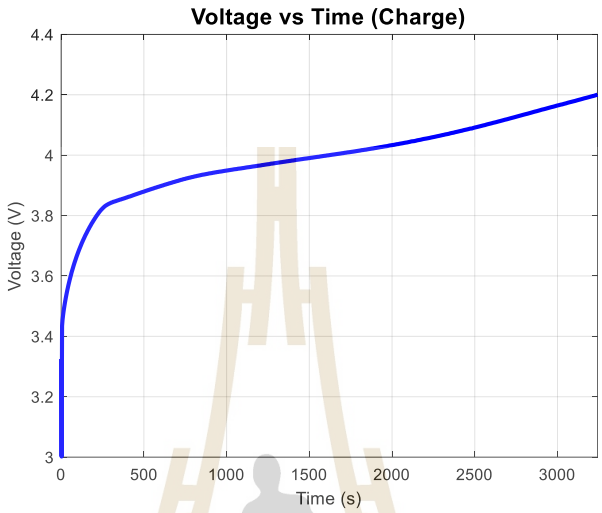


Figure 4.2 Voltage profiles of B0005 during charging cycle from the NASA Battery Dataset

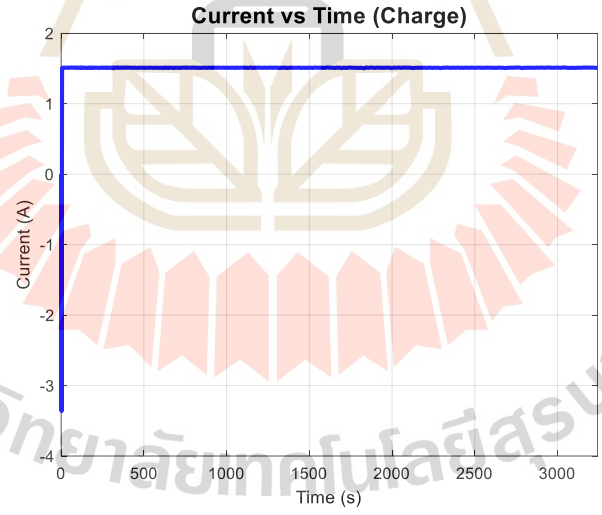


Figure 4.3 Current profiles of B0005 during charging cycle from the NASA Battery Dataset

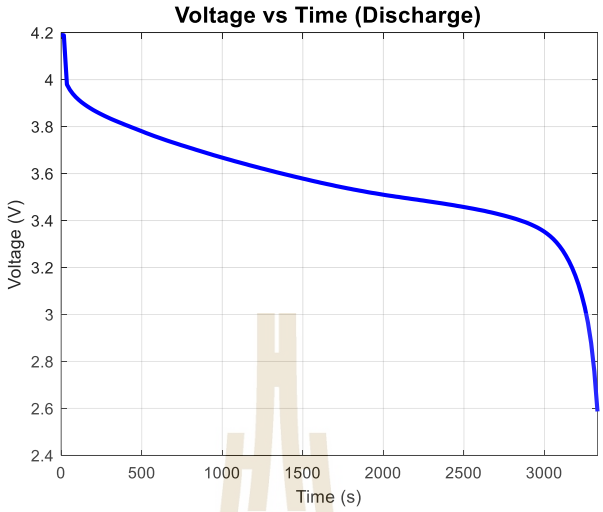


Figure 4.4 Voltage profiles of B0005 during discharging cycle from the NASA Battery Dataset

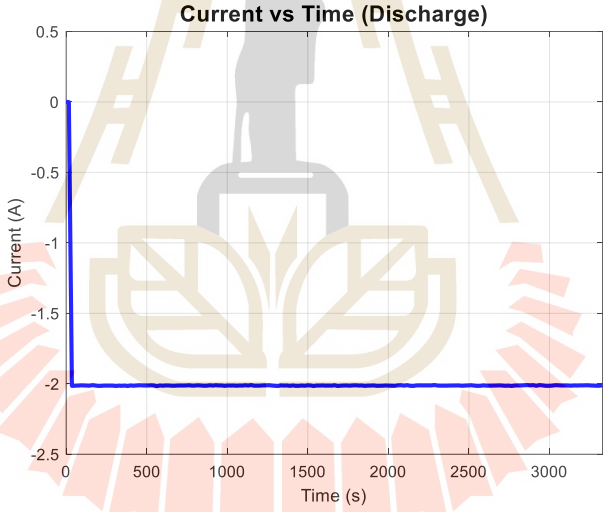


Figure 4.5 Current profiles of B0005 during discharging cycle from the NASA Battery Dataset

The degradation behavior under different discharge cut-off voltages and operating profiles, the SOH trends of the four LIBs in the NASA dataset were plotted.

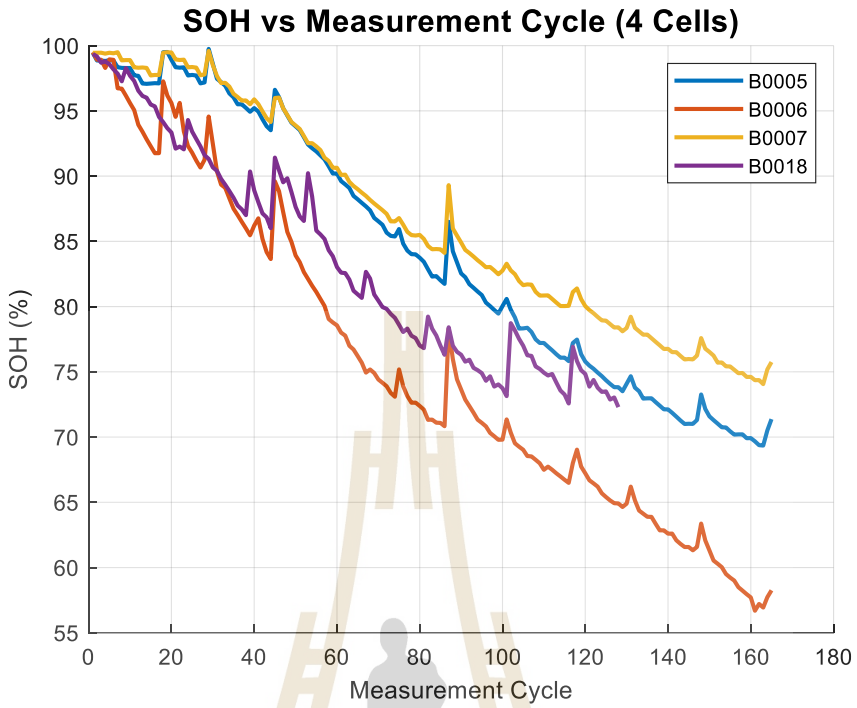


Figure 4.6 SOH degradation curves of all 8 cells from the Oxford Battery Degradation Dataset

4.2.2 Dataset Partitioning Strategy

Following the data preparation described in Chapter 3, the datasets were partitioned based on battery cell identity. This ensures that the testing set consists of a completely "unseen" battery. The specific allocation for both Oxford and NASA datasets is summarized in Table 4.1.

Table 4.1 Hyperparameter Settings for Training

Dataset	Training Set	Validation Set	Testing Set
Oxford	Cells 1, 2, 3, 4	Cells 5, 6, 7	Cell 8
NASA	B0006, B0007	B0005	B0018

4.3 Data Resampling

Data resampling is a process used to alter the temporal resolution of time-series data. This can involve increasing the data frequency (upsampling), often through interpolation, or decreasing it (downsampling), which involves selecting data at sparser

intervals rather than using every available data point. The primary goal of resampling is to adapt the dataset to improve the efficiency, speed, or analytical suitability of the model being applied. This study focuses exclusively on downsampling. The objective is to simplify the input data fed into the deep learning models, thereby reducing the computational load and training time, while critically aiming to preserve the accuracy of the battery's SOH estimation. To implement this, a fixed-step downsampling strategy was employed. Recognizing that the two datasets possess different characteristics and data logging frequencies, a distinct range of resampling intervals was applied to each. These specific intervals were determined based on preliminary exploratory experiments to establish a range that captures the spectrum from minor data reduction to highly sparse data. The selected intervals were:

1) For the Oxford Dataset: Intervals of 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 300, and 400 seconds were tested.

2) For the NASA Dataset: Intervals of 10, 15, 20, 25, 30, 40, and 60 seconds were tested.

For both datasets, the performance at these downsampled intervals was compared against the baseline (labeled 'None'), which utilizes the original, high-resolution (non-resampled) data. This systematic approach allows for a thorough evaluation of the threshold at which reduced data still maintains acceptable accuracy and the point at which information loss begins to significantly degrade model performance. Ultimately, this experiment serves as a key step in identifying the optimal trade-off between data volume and model performance.

The determination of resampling intervals in this research was based on criteria tailored to the specific temporal characteristics of each dataset, ensuring that the resampled data retained essential features for accurate estimation:

For the Oxford dataset, the interval selection was based on the Relative Temporal Resolution framework. Given the average charging duration (T_{avg}) of approximately 3,600 seconds, a baseline interval of 20 seconds was selected. This value represents approximately 0.5% of the total duration, serving as a high-fidelity baseline. Subsequent intervals were increased in arithmetic progression to investigate the impact of resolution degradation.

Conversely, for the NASA dataset, the interval selection was guided by the Signal-to-Noise Ratio Enhancement principle. Due to the inherent sampling irregularity of approximately 2.5 to 3 seconds in the raw data, a baseline interval of 10 seconds was established.

This value corresponds to a smoothing factor (k) of approximately 3 to 4 times the raw sampling rate, chosen based on Variance Reduction principles. Statistically, aggregating data within this range significantly mitigates random noise and provides a necessary safety margin against sampling jitter. Consequently, the resampling process effectively acts as a low-pass filter, standardizing time-series stability before model training.

4.4 Feature Extraction Methodology

Following the data cleaning and preprocessing for both the Oxford and NASA datasets, the feature extraction process was performed. This involved calculating key degradation-aware indicators such as the AUC, Charging and Discharging Time Duration, and features derived from ICA.

4.4.1 Area under curve

The detailed methodology for calculating the AUC as a health-aware feature was previously established in Section 3.3.1.

For the Oxford dataset, the identical set of AUC features presented in that section is utilized in this analysis. For the NASA dataset, the AUC features were calculated anew. To illustrate the raw data characteristics from which these features were extracted, representative cycle profiles for cell B0005 are shown in Figs. 4.7–4.10: the charge voltage profile (Fig. 4.7), charge current profile (Fig. 4.8), discharge voltage profile (Fig. 4.9), and discharge current profile (Fig. 4.10). The corresponding AUC data for the remaining NASA cells are presented in the Appendix B.

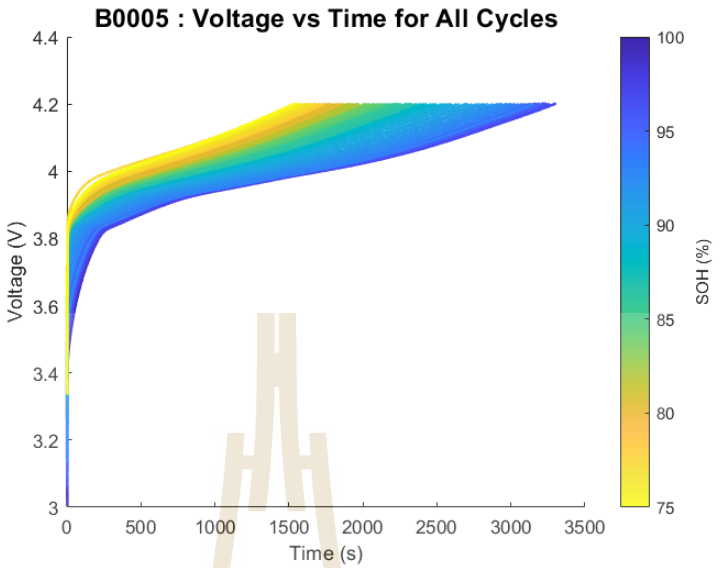


Figure 4.7 Voltage profiles B0005 of NASA are used to charge all the cycles

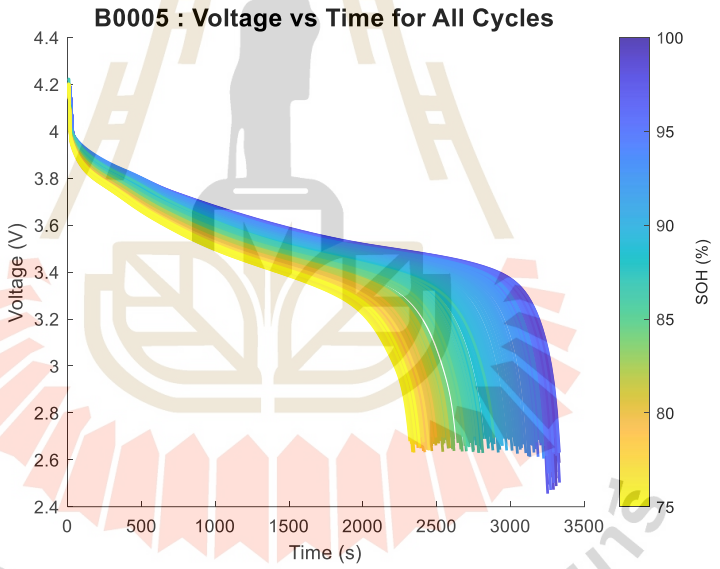


Figure 4.8 Voltage profiles B0005 of NASA are used to discharge all the cycles

มหาวิทยาลัยเทคโนโลยีสุรนารี

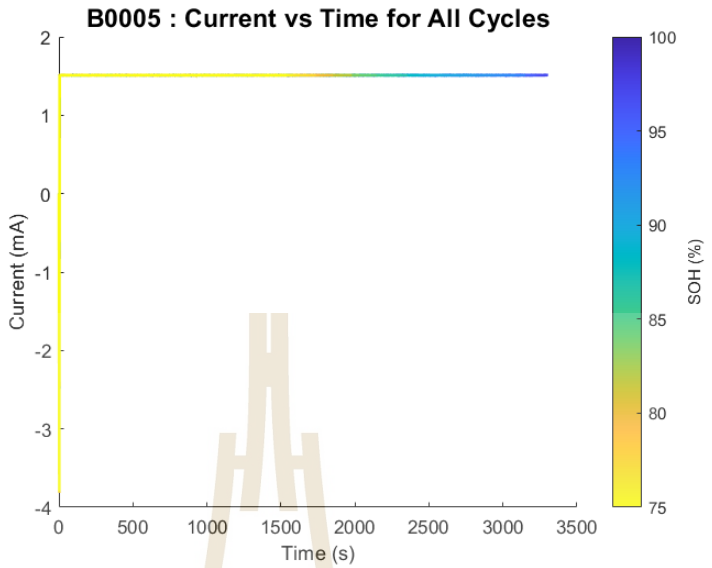


Figure 4.9 Current profiles B0005 of NASA are used to charge all the cycles

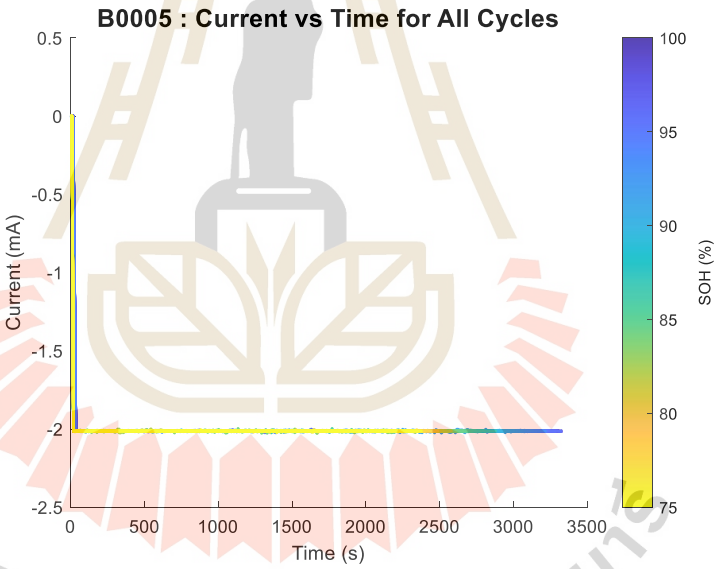


Figure 4.10 Current profiles B0005 of NASA are used to discharge all the cycles

After applying the AUC calculation to all cycles for all NASA cells, the resulting degradation features were normalized using Min-Max Normalization for visualization. The complete normalized AUC degradation trends for all NASA cells (B0005, B0006, B0007, and B0018) are presented in Figs. 4.11–4.14, showing the charge voltage AUC (Fig. 4.11), discharge voltage AUC (Fig. 4.12), charge current AUC (Fig. 4.13), and discharge current AUC (Fig. 4.14), respectively.

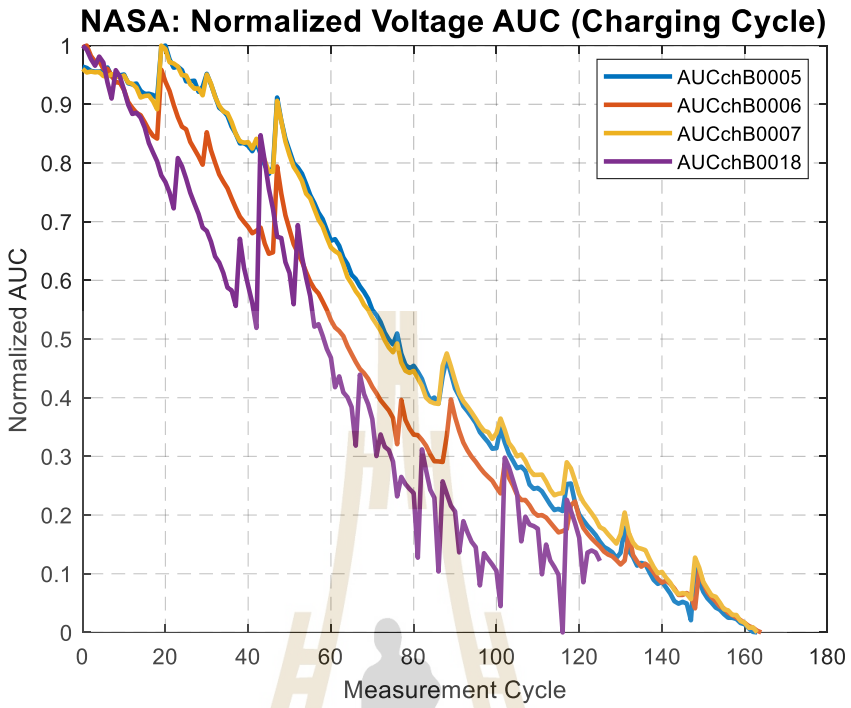


Figure 4.11 Normalized AUC of voltage during the charging cycle from the NASA dataset

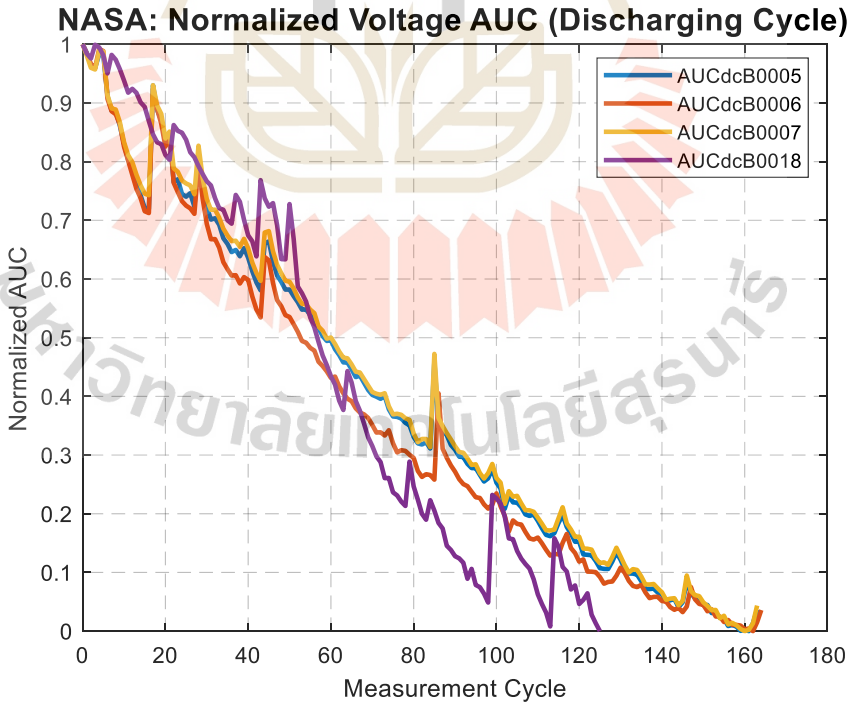


Figure 4.12 Normalized AUC of voltage during the discharging cycle from the NASA dataset

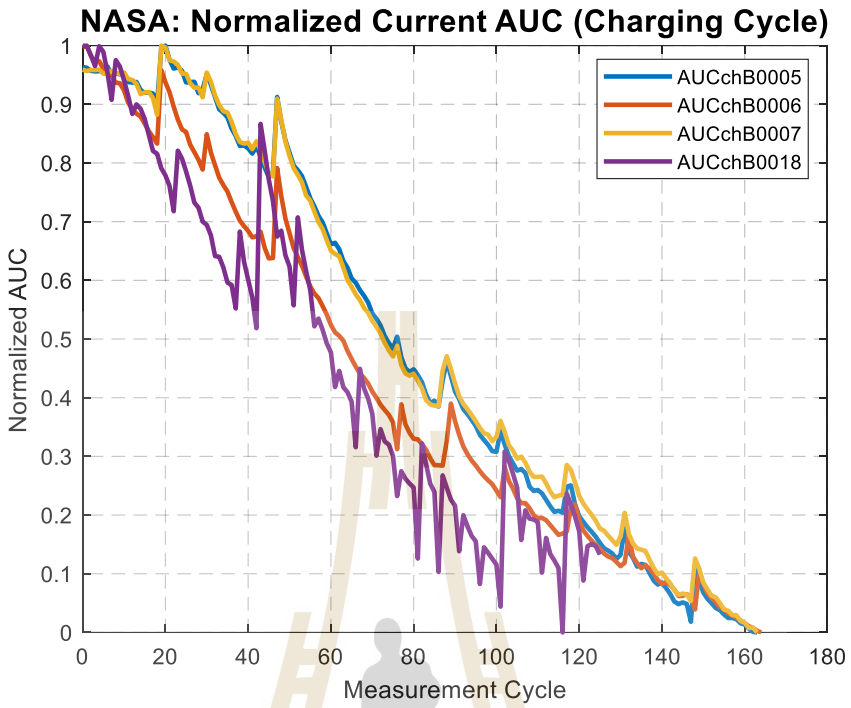


Figure 4.13 Normalized AUC of current during the charging cycle from the NASA dataset

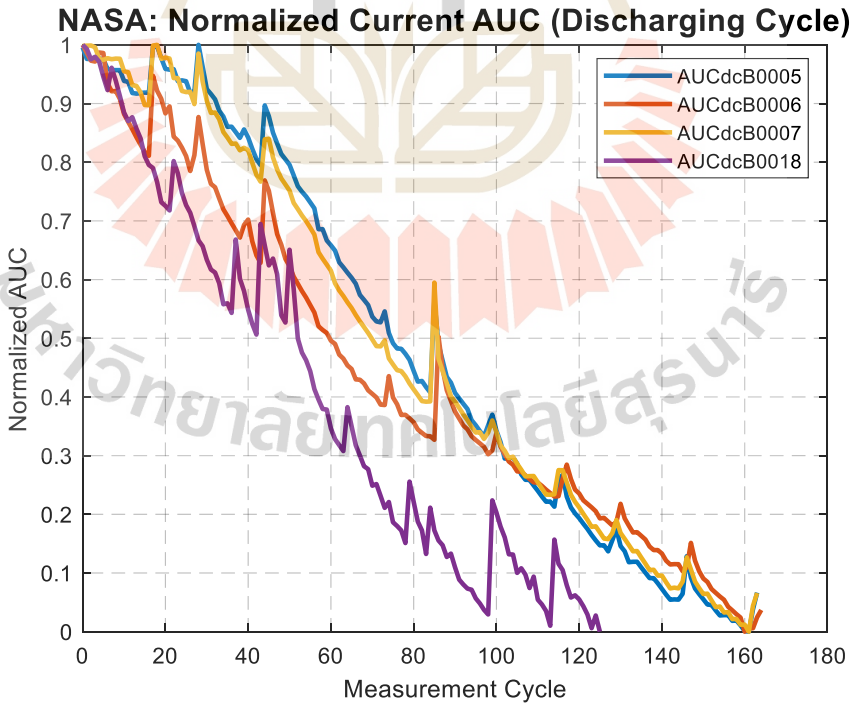


Figure 4.14 Normalized AUC of current during the discharging cycle from the NASA Dataset

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.

4.4.2 The charging and discharging durations

For both the Oxford and NASA datasets, the charging and discharging processes were recorded with corresponding timestamps. Therefore, the final recorded time value for each cycle can be extracted and used as a feature representing the charging and discharging duration.

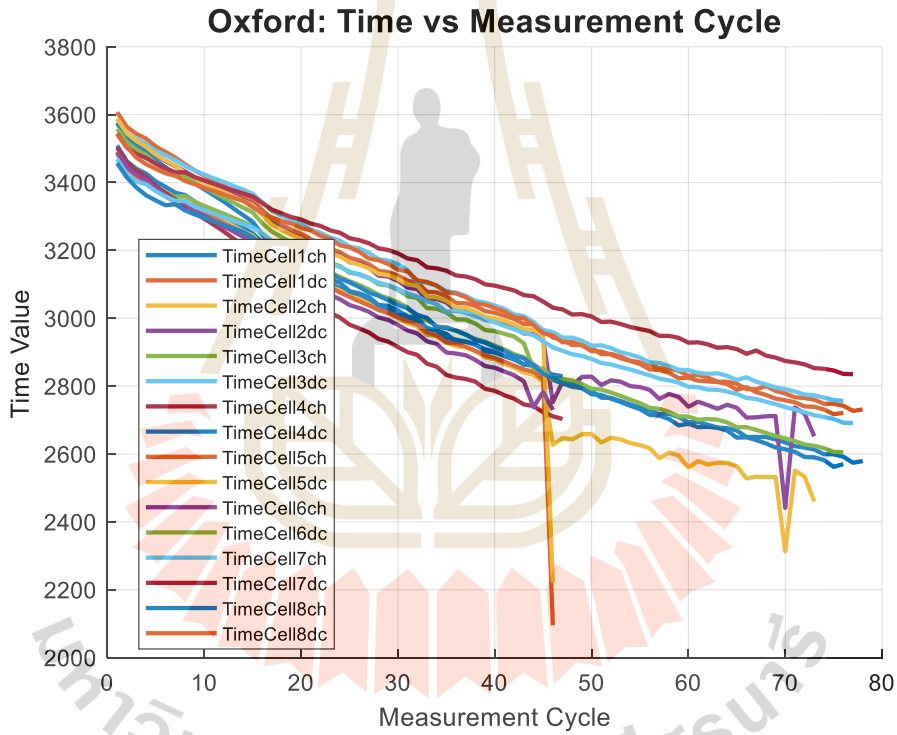


Figure 4.15 Charging and discharging time duration across cycles from the Oxford Dataset

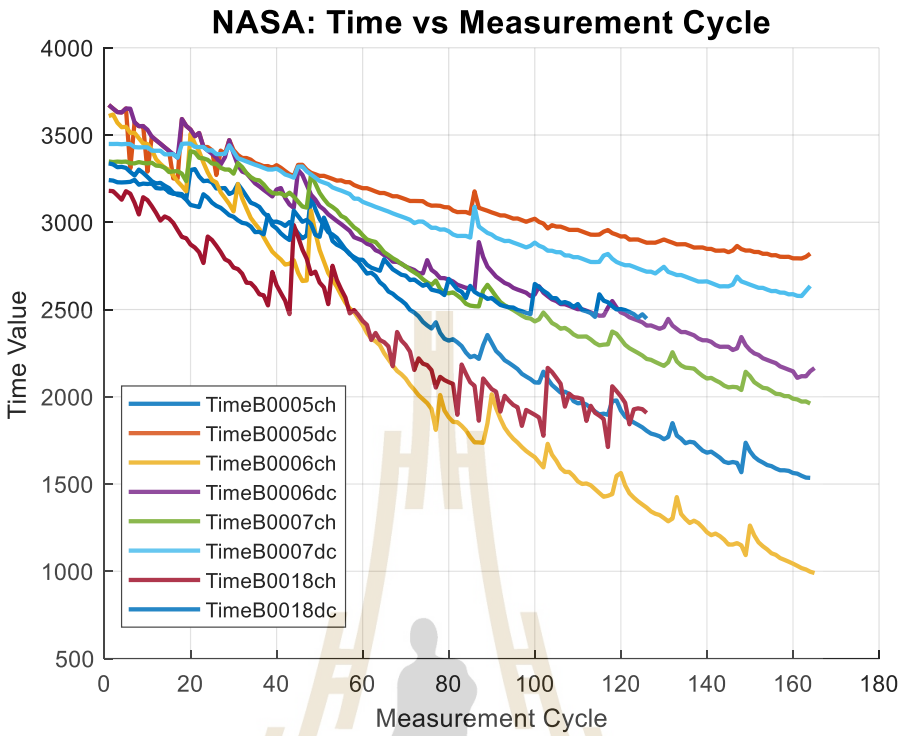


Figure 4.16 Charging and discharging time duration across cycles from the NASA dataset

As shown in Fig. 4.15, which presents the charging and discharging duration from the Oxford dataset, and Fig. 4.16 from the NASA dataset, the overall trend of time duration tends to decrease with increasing cycle count, which corresponds to the decline in the battery's SOH.

4.4.3 Incremental Capacity Analysis

ICA is an in-depth analytical technique used to investigate the behavior and diagnose the health of batteries during charging or discharging processes. It relies on calculating the derivative of electrical capacity (Q) with respect to voltage (V), known as shown in the equation (4.1)

$$\frac{dQ}{dV} \tag{4.1}$$

The hallmark of the ICA technique is its ability to transform the relatively flat voltage plateaus, which normally occur in a voltage-capacity (V-Q) graph, into distinct peaks or valleys on the dQ/dV plot. These peaks correspond directly to

the phase transitions of the active materials within the electrodes. The characteristics of these peaks, such as their height, area, width, and voltage position, are highly sensitive to the electrochemical changes within the battery cell. As the battery degrades, major degradation mechanisms such as Loss of Lithium Inventory (LLI), Loss of Active Material (LAM), and Increased Internal Resistance will cause these peak characteristics to change for example, the peaks may shift, shrink (in height or area), or broaden. Consequently, tracking the changes in key features on the dQ/dV curve such as the minimum point or peaks can serve as an effective indicator for evaluating the degradation of the battery's ability to store or deliver charge. This reflects internal structural changes, such as the loss of active material surface area or the aforementioned lithium loss. For this reason, ICA has been widely adopted in battery aging studies and is considered an effective indicator for evaluating the battery's SOH.

To improve the accuracy of the analysis, the raw dQ/dV data must first undergo a noise filtering process. In this study, a Moving Average technique was applied to smooth the dQ/dV curves. For the Oxford dataset, a window size of 400 was used, while for the NASA dataset, which has different characteristics, a smaller window size of 25 was chosen. These values were selected to match the characteristics and resolution of each dataset, enabling clearer and more reliable identification of battery degradation trends throughout its life cycle. Examples of the calculation results for Cell 1 of the Oxford dataset are shown in Fig. 4.17, and for B0005 of the NASA dataset are shown in Fig. 4.18. The ICA profiles for the remaining batteries from the Oxford and NASA datasets are provided in Appendix C and Appendix D, respectively.

มหาวิทยาลัยเทคโนโลยีสุรนารี

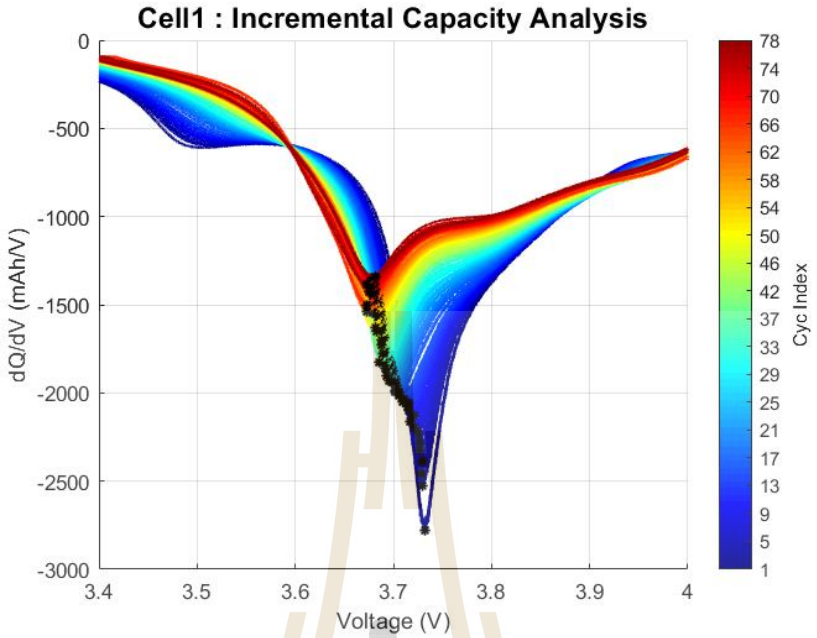


Figure 4.17 ICA curves for Oxford Cell 1

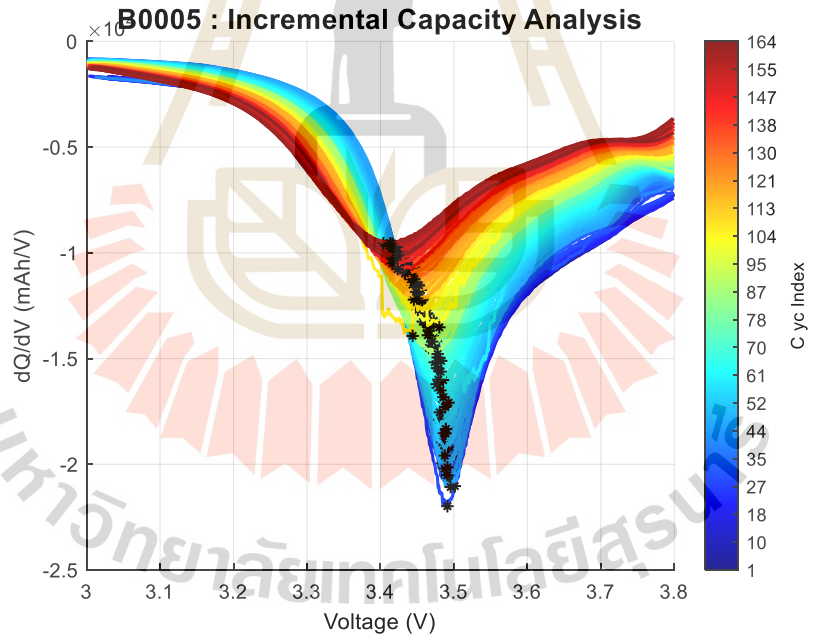


Figure 4.18 ICA curves for NASA B0005

As illustrated in the figures, the ICA curves reveal noticeable changes as the number of cycles increases. In this study, the minimum ICA value (the value at a key dQ/dV valley) from each cycle is selected as a feature for extraction. This feature is chosen because it is directly related to the electrochemical changes within the battery cell and can effectively reflect the battery's degradation condition.

All extracted features, including AUC, time-based features, and ICA features, are normalized using Min-Max Normalization, scaling the values into a 0 to 1 range. This process allows features with different units and scales to be fairly compared and processed on the same footing. It also helps the learning algorithm to converge faster and prevents the model from being biased toward features with larger numerical ranges. As a result, the normalization process enhances the model's stability and prediction accuracy over time.

4.5 Feature Correlation Analysis

After all relevant features (AUC, time-based features, and ICA) were extracted, a correlation analysis was performed to investigate the relationships between these features themselves (inter-feature correlation) and, more importantly, their correlation with the battery's SOH. For this analysis, a Correlation Matrix was constructed, following the same approach mentioned in Chapter 3, to evaluate the linear relationships between variables. The results for the Oxford dataset are presented in Fig. 4.19, and the results for the NASA dataset are shown in Fig. 4.20.

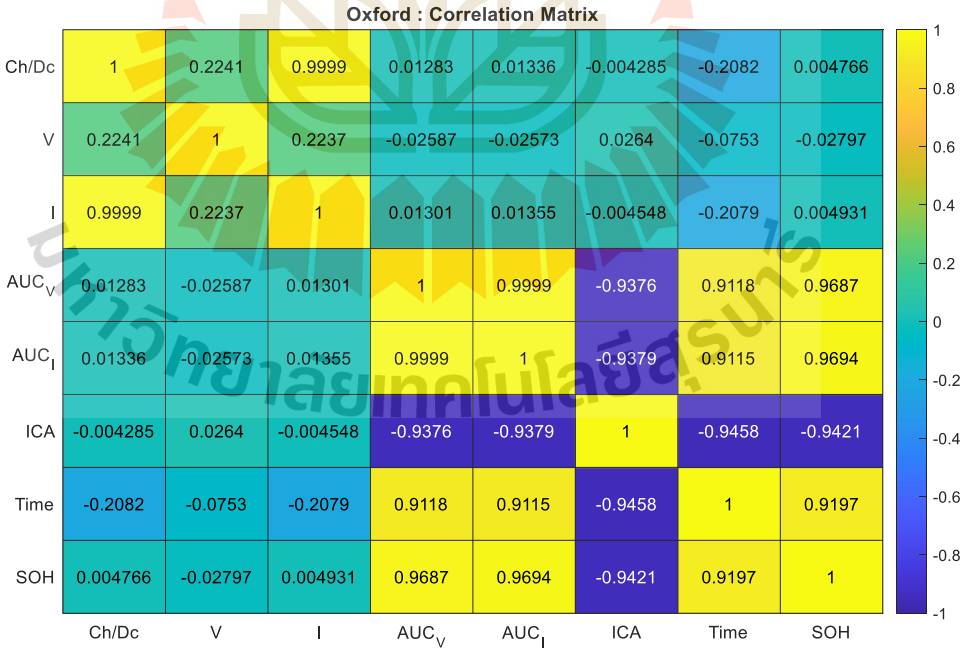


Figure 4.19 The correlation matrix from the Oxford dataset

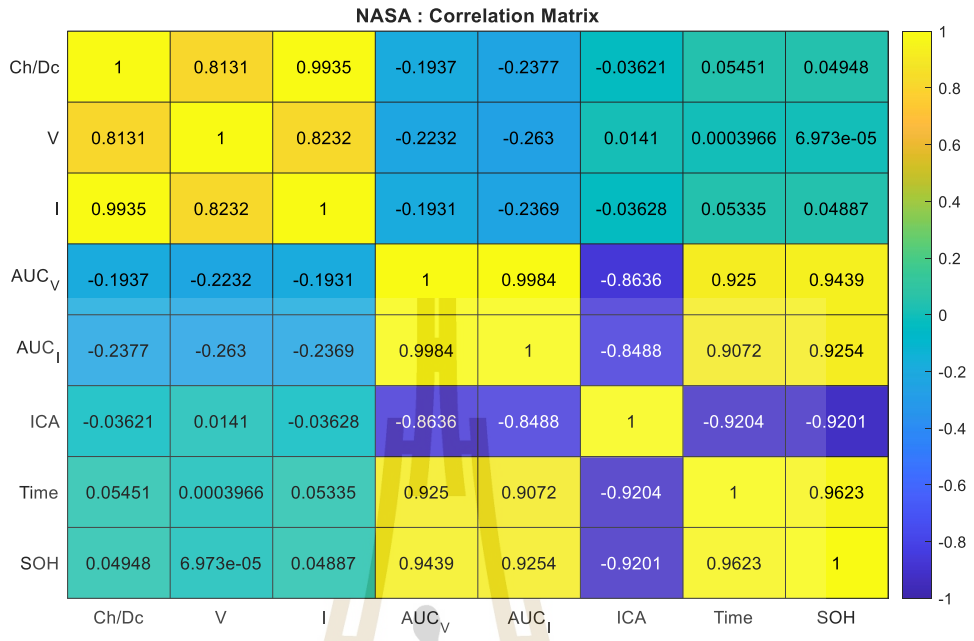


Figure 4.20 The correlation matrix from the NASA dataset

An examination of both correlation matrices reveals that all extracted features show a very high degree of correlation with the battery's SOH. The features AUC_v, AUC_i, and Time all exhibit an extremely strong positive correlation. In the Oxford dataset (Fig. 4.19), their coefficients are 0.9687, 0.9694, and 0.9197, respectively. The NASA dataset (Fig. 4.20) shows clearly consistent results, with values of 0.9439, 0.9254, and 0.9623, respectively. This strong positive correlation confirms that as the battery degrades (SOH decreases), the values of these features (such as the AUC or the time duration) also decrease, which aligns with the expected physical behavior of a degrading battery.

Meanwhile, the ICA feature, which was extracted from the minimum value of the dQ/dV curve, shows a very strong negative correlation (a coefficient close to -1) with SOH, posting a value of -0.9421 in the Oxford dataset and -0.9201 in the NASA dataset. This strong negative coefficient is highly significant; it does not imply it is a poor feature but rather indicates a clear inverse relationship. In other words, as the battery's SOH decreases (the battery degrades), the extracted ICA feature value (the value at the minimum point/valley) tends to increase. The fact that the coefficient is nearly -1 confirms that the change in this dQ/dV valley is also a highly reliable

indicator of degradation. In summary, this analysis confirms that all selected features (AUC_v, AUC_i, ICA, and Time) are excellent predictors of SOH, supporting their suitability for use in the subsequent modeling phase.

4.6 Hyperparameter Settings

The performance and accuracy of Deep Learning models depend heavily on proper hyperparameter settings. In this study, to ensure a fair and standardized comparison between models, and specifically to clearly evaluate the effectiveness of the Resampling technique without the influence of differing model settings, most basic parameters were kept identical for the LSTM, Bi-LSTM, and GRU models.

Key training parameters, including the number of hidden units, maximum epochs, mini-batch size, as well as the learning rate schedule and optimizer details, are summarized in Table 4.2.

However, the Attention-LSTM model required slightly different settings. The number of Hidden Units was set to 256, with an additional Attention Dimension (attnDim) of 64. This adjustment was necessary to satisfy the structural constraints of the Multi-Head Attention mechanism, which require the input dimension (Hidden Units) to be exactly divisible by the number of attention heads. Using 200 units, as in the other models, caused calculation errors due to this mismatch. In contrast, 256 (a power of two) ensures full compatibility and efficient operation within the attention layer's architecture.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 4.2 Hyperparameter Settings for Training

Parameter	Value
Number of Hidden Units	200
Fully Connected Layer	500
Max Epochs	100
Mini-Batch Size	1024
Initial Learn Rate	0.01
Learn Rate Drop Factor	0.5
Learn Rate Drop Period	40 epochs
Optimizer	Adam

The parameters listed in Table 4.2 were selected to balance accuracy with computational efficiency. Certain key values were adopted from the optimization methodology in (Jafari & Byun, 2025) as a robust baseline, while others were empirically fined-tuned. A more in-depth analysis regarding hyperparameter tuning and its impact on model performance will be presented and discussed in the following section.

4.7 Experimental Results and Discussion

This section presents the experimental results, organized into two main parts. First, the Hyperparameter Tuning Analysis details the process of identifying optimal model settings. Second, the Comparative SOH Estimation Results evaluates and compares the accuracy of the proposed models using these finalized parameters.

4.7.1 Hyperparameter Tuning Analysis

In this study, we performed hyperparameter tuning primarily on the Oxford dataset (specifically the version resampled at 100 cycles). Our goal was to find the settings that allow the model to learn degradation patterns most effectively.

Once the best parameters were identified, we applied these exact same settings to the NASA dataset (resampled at 20 cycles) without any further adjustments. We did this to test the model's robustness and ability to generalize. This approach

proves that the proposed model architecture works well even on datasets with different characteristics.

To determine the optimal network architecture, we used a Grid Search strategy. We simultaneously varied two key parameters: the number of hidden units (numHiddenUnits) and the size of the fully connected layer (hiddenLayerSize), using the RMSE as our main guide. The numerical results are listed in Table 4.2, and the trends are visualized in the bar chart in Fig. 4.21.

Table 4.2 RMSE results from the Grid Search analysis on the Oxford dataset

Hidden Layer Size	Num of Hidden Units		
(Nodes)	100	200	300
200	0.4224	0.3357	0.4736
300	0.3975	0.3215	0.4127
400	0.358	0.3286	0.3436
500	0.3353	0.3045	0.3239
600	0.3887	0.4376	0.3684
700	0.3872	0.3378	0.4783

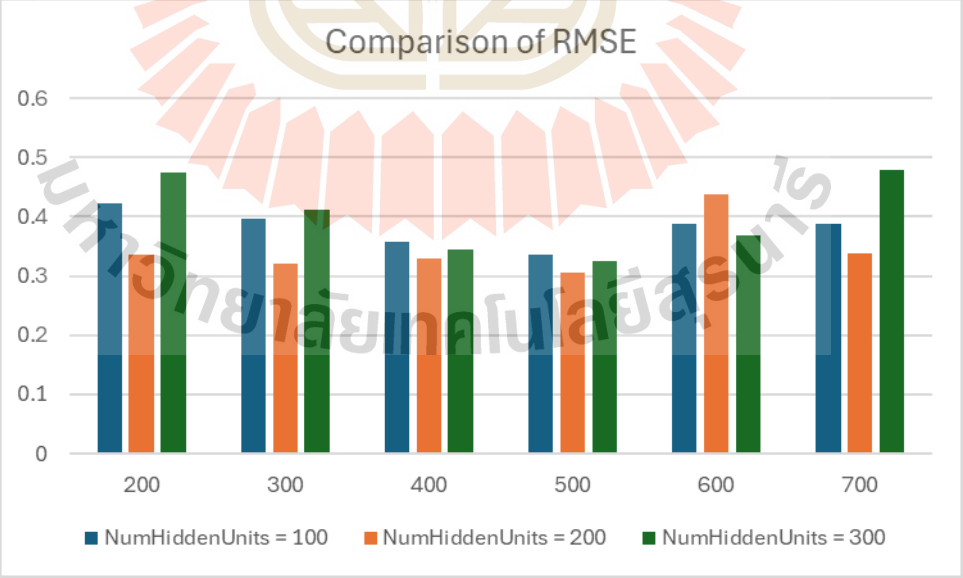


Figure 4.21 Comparison of RMSE values across different model architecture configurations

Impact of Hidden Layer Size: As clearly shown by the orange bars (where numHiddenUnits is fixed at 200), the RMSE tends to decrease as the hiddenLayerSize increases from 200 up to 500. The error drops from 0.3357 down to 0.3045 at size 500, which is the lowest point (minimum error) in this experiment. However, when we increased the size further to 600, the RMSE spiked significantly to 0.4376. This suggests the model became too complex (over-parameterized), leading to overfitting, where it started memorizing noise instead of learning the actual patterns.

Impact of Hidden Units: Similarly, if we compare the height of the bars at a hiddenLayerSize of 500, we see that using 100 units (blue bar) or 300 units (green bar) results in higher errors compared to 200 units (orange bar). Their values are 0.3353 and 0.3239, respectively. This indicates that 200 units offer the best balance between learning capability and model complexity.

Based on this grid search analysis, we confirmed that the combination of (numHiddenUnits = 200, hiddenLayerSize = 500) is the optimal choice for this dataset. Therefore, this setting was selected as the standard for the comparative experiments in the next section.

4.7.2 Comparative SOH Estimation Results

The performance evaluation of the four deep learning models LSTM, Bi-LSTM, GRU, AM-LSTM and the statistical model (GPR) begins with an analysis of the Oxford dataset, which is characterized by its high-resolution data. The quantitative results presented in Table 4.3 and Table 4.4 highlight significant differences in both estimation accuracy and computational time between using raw data and applying the proposed resampling technique.

When considering the baseline case Non-resampling, most models struggled to predict SOH accurately. This is illustrated in Fig. 4.21 Baseline Comparison, which compares the prediction results of all models against the actual values using raw data. The prediction curves for the standard models, particularly LSTM and GRU, exhibited considerable fluctuations throughout the battery's life cycle. This visual observation aligns with the high RMSE values of 0.922 and 1.548, respectively. The

volatility seen in Fig. 4.22 reflects the presence of significant noise in the raw data, preventing the models from effectively capturing the global degradation trend, resulting in jagged curves that fail to track the actual SOH closely.

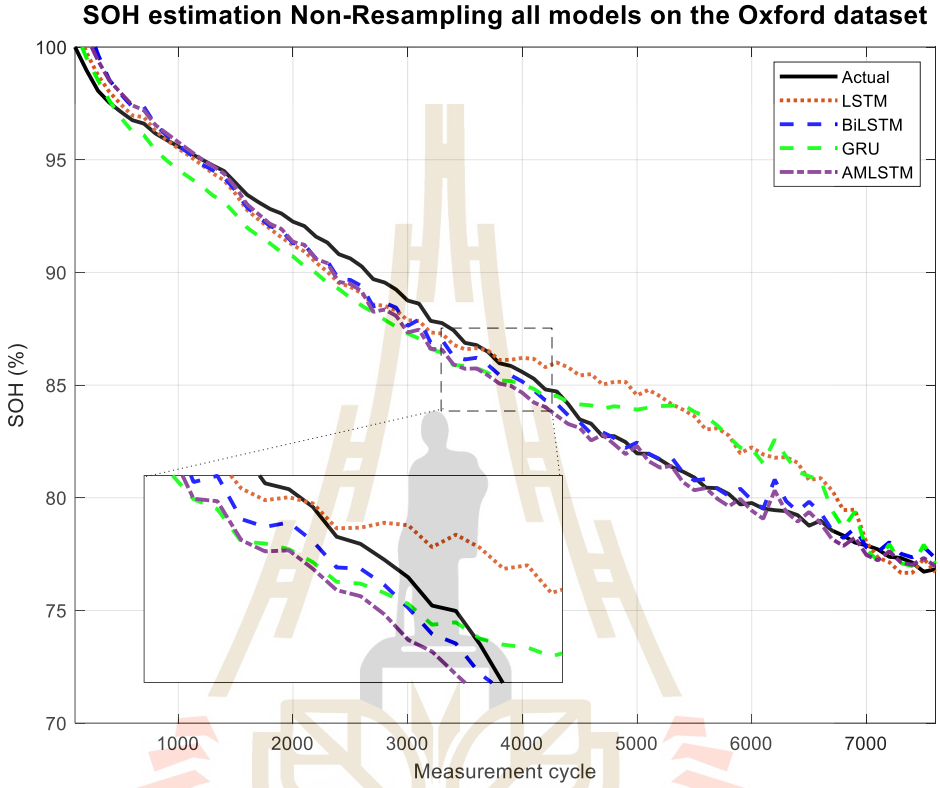


Figure 4.22 Baseline comparison of SOH estimation results using raw data (Non-resampling) for all models on the Oxford dataset

However, when the data underwent resampling at the optimal interval of 100 seconds, the results improved significantly, as shown in Fig. 4.23 Proposed Method Comparison. The prediction curves for all models became much smoother and less volatile. Comparing the models, the proposed AM-LSTM (purple line) demonstrated the best ability to track the actual value curve (black line), especially during the final stages of the battery life where degradation is most severe. This is supported by the data in Table 4.3, where AM-LSTM achieved the lowest RMSE of 0.292, compared to LSTM (0.305) and Bi-LSTM (0.311).

It is also important to address the specific status of the GPR model in Table 4.3. Readers will notice that the entry for GPR under the raw data ('None')

category is marked with a hyphen (-). This symbol indicates that the model could not be successfully executed in this scenario due to its inherent computational limitations; GPR faces significant challenges with large-scale datasets as its computational complexity increases drastically with the number of data points. Consequently, it was unable to process the massive volume of the raw Oxford data directly. However, once the data was reduced via the resampling technique at the 100-second interval, the GPR model became operational. As shown in Fig. 4.23, although its resulting RMSE of 1.541 was higher than that of the deep learning models, the GPR model offers a distinct advantage by providing a 95% confidence interval. This feature offers valuable insight into the uncertainty of the predictions, even if the model lacks the capacity to capture the complex, non-linear relationships of the battery data as effectively as the neural networks.

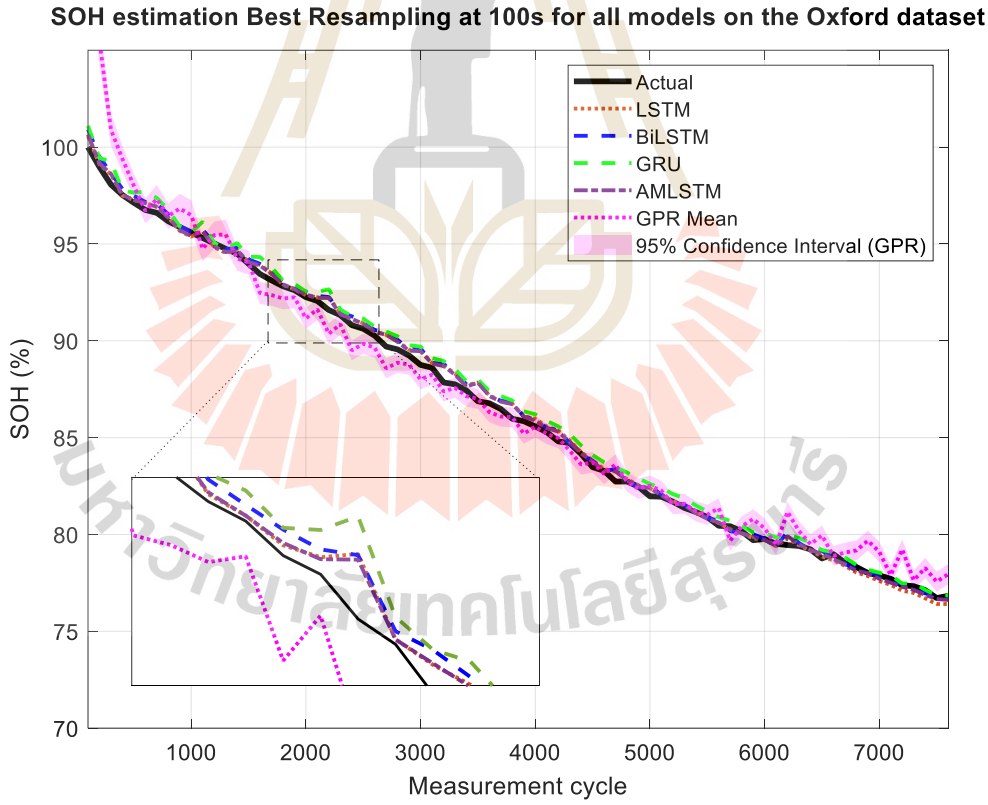


Figure 4.23 Comparison of SOH estimation results using the Best Resampling at 100s For all models on the Oxford dataset.

To confirm the necessity and effectiveness of the proposed approach, the specific impact of the resampling technique on the AM-LSTM model is analyzed in Fig. 4.24 Impact of Resampling Analysis. This graph clearly contrasts the prediction using raw data (yellow line) versus the prediction using data resampled at 100 seconds (blue line). As seen in the figure, the yellow line is widely scattered with high error, whereas the blue line is tight, concise, and converges accurately to the actual SOH values. This visual evidence confirms that selecting an appropriate resampling interval effectively removes redundant noisy data, allowing the model to learn the underlying degradation structure more clearly.

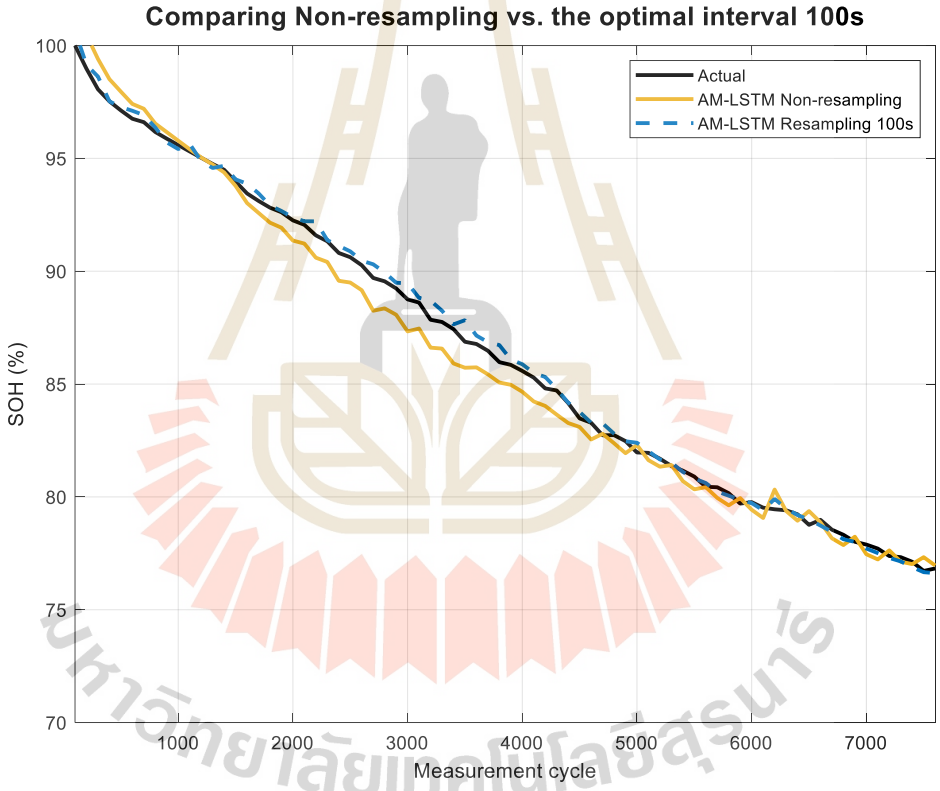


Figure 4.24 Analysis of the impact of resampling on the AM-LSTM model (Oxford dataset) comparing Non-resampling vs. the optimal interval (100s).

However, a key observation from the experiments is that a trade-off exists: expanding the resampling interval too much can negatively affect accuracy. As indicated in Table 4.3, when the resampling interval extends beyond the optimal point (e.g., beyond 120 seconds), the RMSE of the model starts to rise again. This

phenomenon is caused by excessive information loss; subtle degradation patterns are removed along with the noise. Consequently, the prediction curve becomes too coarse and fails to capture the detailed behavior of the battery. Therefore, the 100-second mark represents the "sweet spot" for this study, balancing maximum noise reduction with the preservation of essential data required for model learning. Fig. 4.25 illustrates the RMSE comparison of all models for the Oxford dataset.

Beyond improvements in accuracy, computational efficiency is another major achievement of this study. Data from Table 4.4 reveals that training the models with raw data requires a substantial amount of time. For instance, the AM-LSTM model took approximately 1.59×10^{4} seconds (about 4.4 hours). However, by applying resampling at 100 seconds, the processing time dropped significantly to just 169.1 seconds (under 3 minutes). This dramatic reduction in training time is not merely due to data reduction, but because the model can learn faster when the input is compact and noise-free. This allows the optimization process to converge more easily, making the proposed method highly suitable for real-world applications where both accuracy and speed are critical.

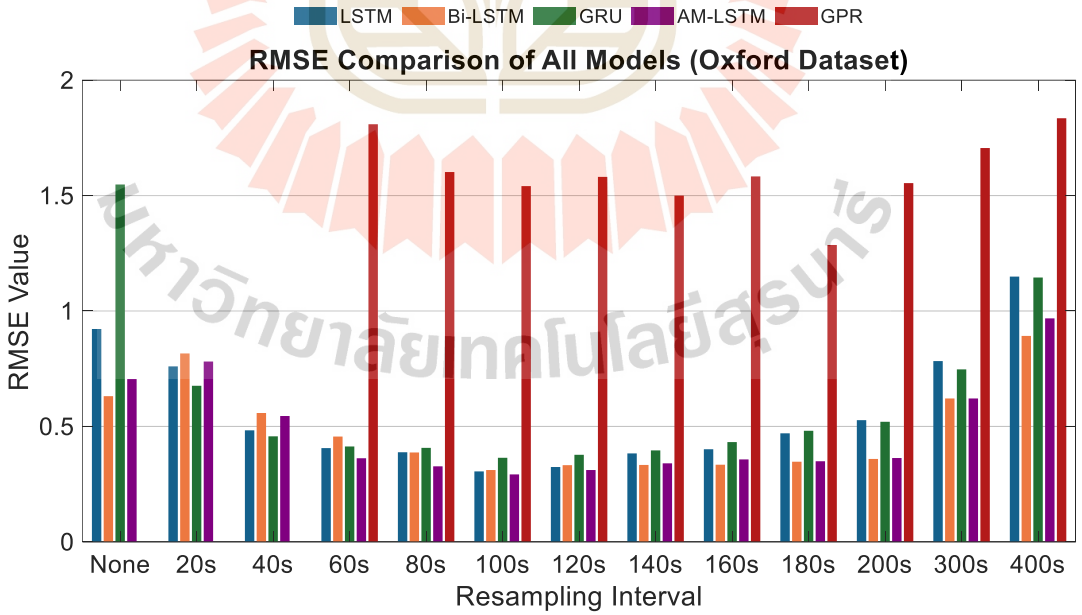


Figure 4.25 RMSE comparison of all models on the Oxford dataset

Table 4.3 Comparison of SOH Estimation Results on Oxford Dataset

Dataset	Resampling Interval	LSTM				Bi-LSTM				GRU				AM-LSTM				GPR			
		RMSE	MSE	MAE	MAPE	RMSE	MSE	MAE	MAPE	RMSE	MSE	MAE	MAPE	RMSE	MSE	MAE	MAPE	RMSE	MSE	MAE	MAPE
Oxford	None	0.922	0.850	0.722	0.810	0.631	0.398	0.523	0.596	1.548	2.397	1.331	1.564	0.705	0.497	0.585	0.669	-	-	-	-
	20 s	0.760	0.577	0.609	0.684	0.816	0.666	0.667	0.747	0.676	0.456	0.400	0.445	0.781	0.610	0.592	0.658	-	-	-	-
	40 s	0.483	0.233	0.364	0.420	0.558	0.312	0.427	0.478	0.457	0.209	0.362	0.433	0.545	0.297	0.452	0.534	-	-	-	-
	60 s	0.406	0.165	0.332	0.385	0.456	0.208	0.382	0.450	0.413	0.170	0.328	0.373	0.362	0.131	0.286	0.339	1.809	3.272	0.869	0.972
	80 s	0.388	0.150	0.340	0.299	0.387	0.150	0.311	0.367	0.407	0.166	0.317	0.363	0.327	0.107	0.256	0.294	1.602	2.566	0.867	0.986
	100 s	0.305	0.093	0.242	0.279	0.311	0.097	0.249	0.290	0.364	0.132	0.294	0.336	0.292	0.085	0.232	0.271	1.541	2.375	0.828	0.929
	120 s	0.324	0.105	0.257	0.296	0.332	0.110	0.254	0.289	0.377	0.142	0.308	0.350	0.311	0.097	0.239	0.271	1.581	2.500	0.907	1.029
	140 s	0.383	0.147	0.314	0.360	0.333	0.111	0.263	0.299	0.396	0.157	0.325	0.369	0.340	0.115	0.270	0.305	1.500	2.250	0.892	1.019
	160 s	0.401	0.161	0.325	0.368	0.334	0.112	0.258	0.292	0.432	0.186	0.349	0.396	0.357	0.128	0.278	0.313	1.583	2.505	1.048	1.203
	180 s	0.470	0.221	0.389	0.440	0.347	0.120	0.272	0.308	0.481	0.232	0.394	0.447	0.349	0.122	0.278	0.316	1.286	1.654	0.817	0.929
	200 s	0.527	0.278	0.446	0.503	0.359	0.129	0.287	0.326	0.520	0.270	0.448	0.508	0.363	0.132	0.279	0.313	1.554	2.416	0.950	1.083
	300 s	0.783	0.613	0.694	0.795	0.621	0.385	0.527	0.591	0.747	0.557	0.669	0.760	0.621	0.385	0.527	0.593	1.706	2.909	0.985	1.124
	400 s	1.149	1.321	1.006	1.156	0.892	0.795	0.790	0.893	1.145	1.311	0.999	1.152	0.968	0.937	0.849	0.971	1.835	3.366	1.040	1.200

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 4.4 Execution Time of Each Model for SOH Estimation on Oxford Dataset

Dataset	Resampling Interval	LSTM (s)	Bi-LSTM (s)	GRU (s)	AM-LSTM (s)	GPR (s)
Oxford	None	14,056	17,071	13,403	15,891	-
	20 s	447.4362	556.6291	403.2034	848.2249	-
	40 s	217.598	261.4257	205.7885	382.1823	-
	60 s	160.9488	184.5931	139.2921	264.0181	1187.7686
	80 s	105.9322	139.2008	103.1165	228.6973	282.9889
	100 s	84.0947	115.3445	85.6528	169.1163	148.4209
	120 s	70.8816	97.7156	72.7832	131.5215	100.683
	140 s	58.675	82.5298	59.7725	106.8197	73.4618
	160 s	51.8306	66.6126	51.5024	93.0716	56.3858
	180 s	48.366	64.9802	51.0482	96.7402	48.5944
	200 s	44.8542	57.1521	45.14	84.3684	43.021
	300 s	30.0948	40.1414	30.8055	57.0947	27.4732
	400 s	23.5741	28.96	23.4261	41.0468	26.0582

Following the analysis of the Oxford dataset, this study extended the evaluation to the NASA dataset to assess the model's generalization capability and robustness under different data conditions. The quantitative results regarding estimation errors and computational time are summarized in Table 4.5 and Table 4.6, respectively.

Upon examining the results, the behavior of the NASA dataset differs interestingly from the Oxford one. In the baseline case using raw data (None), most models performed reasonably well, as shown in Fig. 4.26 Baseline Comparison. Notably, the AM-LSTM model (purple line) achieved the lowest error with an RMSE of 0.2351, representing the best performance in this experiment. This indicates that the Attention-LSTM architecture is highly effective at handling the raw NASA data directly without requiring complex preprocessing.

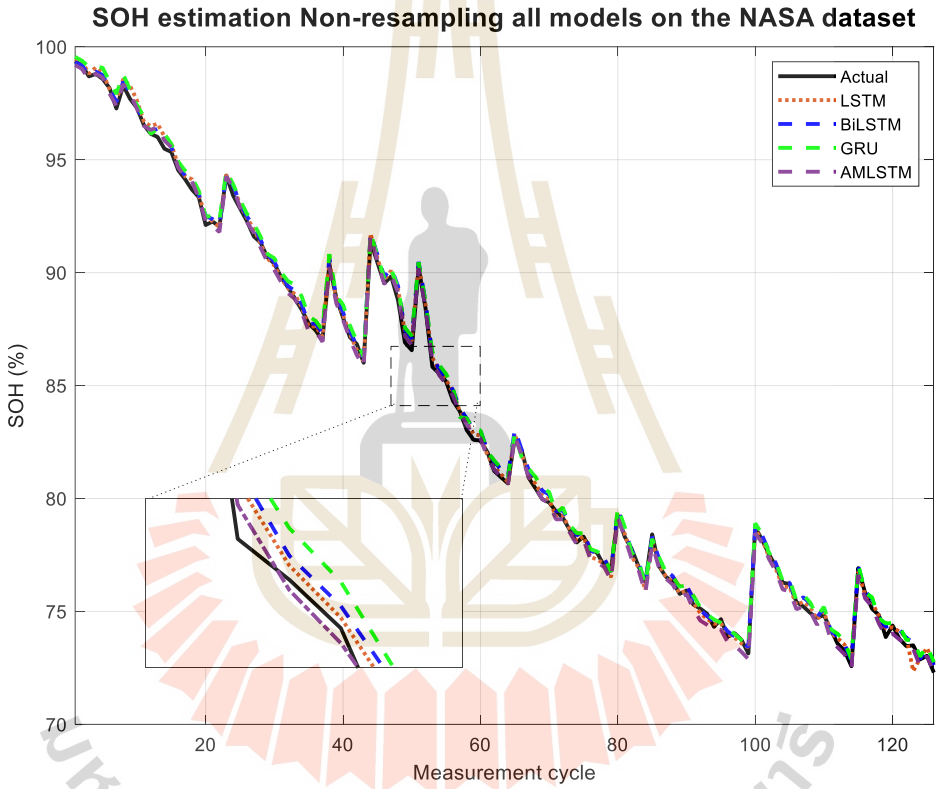


Figure 4.26 Baseline comparison of SOH estimation results using raw data (Non-resampling) for all models on the NASA dataset

However, a key insight emerges when applying the resampling technique at a 20-second interval (the optimal setting for this dataset). The results demonstrate the robustness of the model. Although the RMSE of the AM-LSTM increased slightly to 0.311, it remained highly accurate compared to other baseline models. This is visible in Fig. 4.27 Robustness Check, where the prediction curve

continues to track the actual degradation trend satisfactorily. This proves the model is flexible enough to handle data reduction without a critical loss of performance.

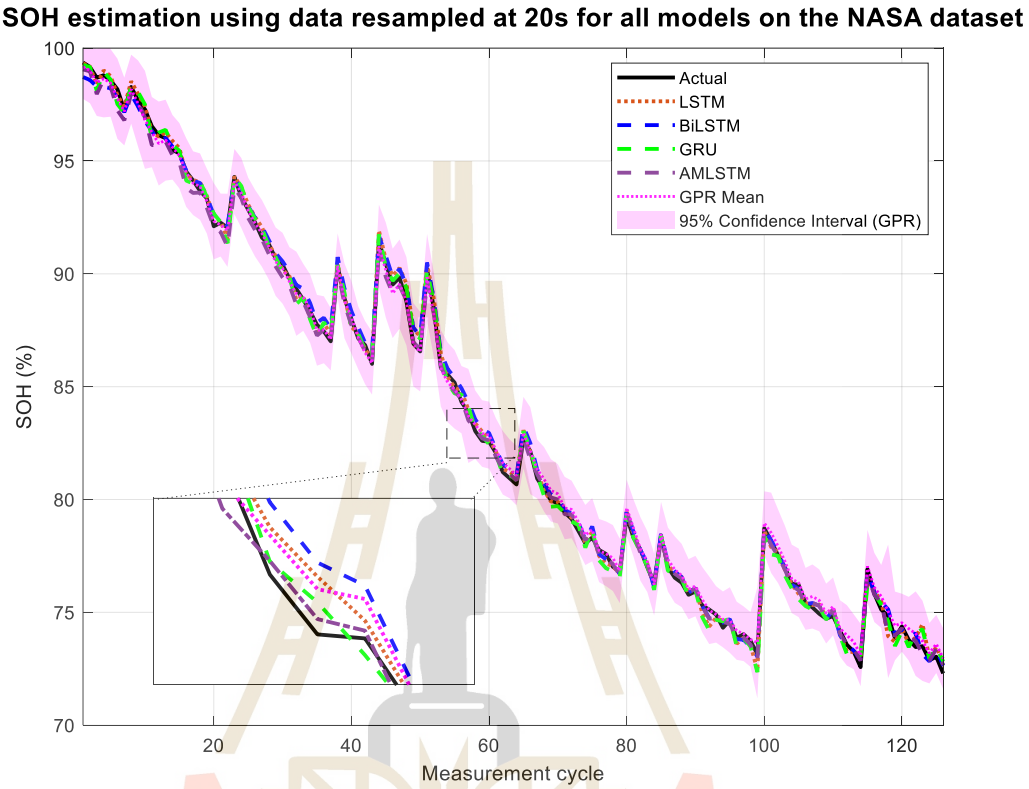


Figure 4.27 Comparison of SOH estimation results using the Best Resampling at 20s for all models on the NASA dataset

To further illustrate this robustness and the choice of the optimal interval, Fig. 4.28 Impact of Resampling Analysis compares the AM-LSTM model using raw data (yellow line) versus the 20-second resampling (blue line). It can be observed that the two lines almost overlap and follow the same trajectory smoothly, confirming that the 20-second interval is a "Safe Zone" for data reduction. Conversely, if the resampling interval is extended too far (e.g., 40 or 60 seconds), the error spikes rapidly RMSE exceeding 1.45, as seen in Table 4.5 due to the loss of fine-grained details which are much smaller in the NASA dataset compared to Oxford.

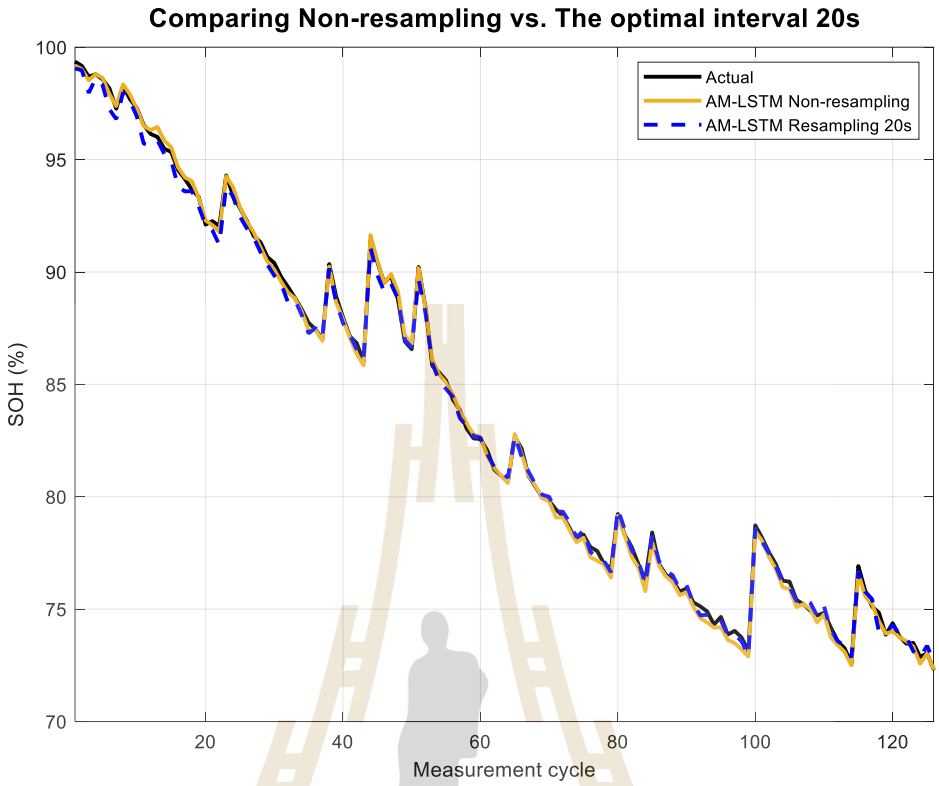


Figure 4.28 Analysis of the impact of resampling on the AM-LSTM model (NASA dataset) comparing Non-resampling vs. 20s interval

Regarding the GPR model, the same limitation persists: it could not process the large raw dataset (indicated by the hyphen symbol '-' in the table). However, it became operational once the data size was reduced. While its accuracy is not as high as the deep learning models, it remains useful for providing confidence intervals.

The deciding factor for using resampling with the NASA dataset is computational efficiency. Data from Table 4.6 clearly shows that training the AM-LSTM model with raw data took approximately 2,759 seconds. However, using resampling at 20 seconds reduced this time to just 121.8 seconds, a decrease of over 95%. Therefore, although resampling involves a slight trade-off in accuracy compared to raw data, the massive gain in processing speed makes it a worthwhile option for real-world applications with limited computational resources.

Comparing the SOH prediction results between the two datasets reveals a distinct relationship between the physical characteristics of the data and the optimal data processing strategy. For the Oxford, the degradation pattern exhibits a relatively smooth, monotonic decay but is superimposed with high-frequency noise throughout its lifecycle. This characteristic explains why the 100-second resampling interval yielded the best results; in this context, resampling acts effectively as a denoising filter. By averaging out the fluctuations, the model can more clearly discern the underlying global degradation trend, leading to improved accuracy.

In contrast, the NASA dataset displays a fundamentally different behavior. The SOH curve does not follow a simple linear decline but is marked by significant Capacity Regeneration Phenomena or capacity recovery. These are observed as sudden upward spikes in SOH during certain cycles, typically caused by the battery's relaxation effect during rest periods, allowing active ions to redistribute.

This unique characteristic of the NASA dataset is the key reason why using raw data (None) or a very short resampling interval (20s) resulted in the highest accuracy. These upward "bounces" are not noise, but significant features reflecting the battery's actual health status. If a wider resampling interval were applied here, the data reduction process would smooth out these critical peaks, causing the model to lose vital information required to predict the regeneration phases accurately.

However, the ultimate validation of the proposed AM-LSTM model lies in its adaptability. The experimental results demonstrate that the model is robust enough to handle these opposing data extremes. It successfully generated a smooth, trend-following curve for the noisy Oxford data, while simultaneously being responsive enough to track the sharp fluctuations and regeneration peaks of the NASA data without significant delay. This confirms that the proposed architecture coupled with the attention mechanism is capable of capturing complex degradation patterns regardless of the specific data characteristics. Fig. 4.29 presents the RMSE comparison of all models for the NASA dataset.

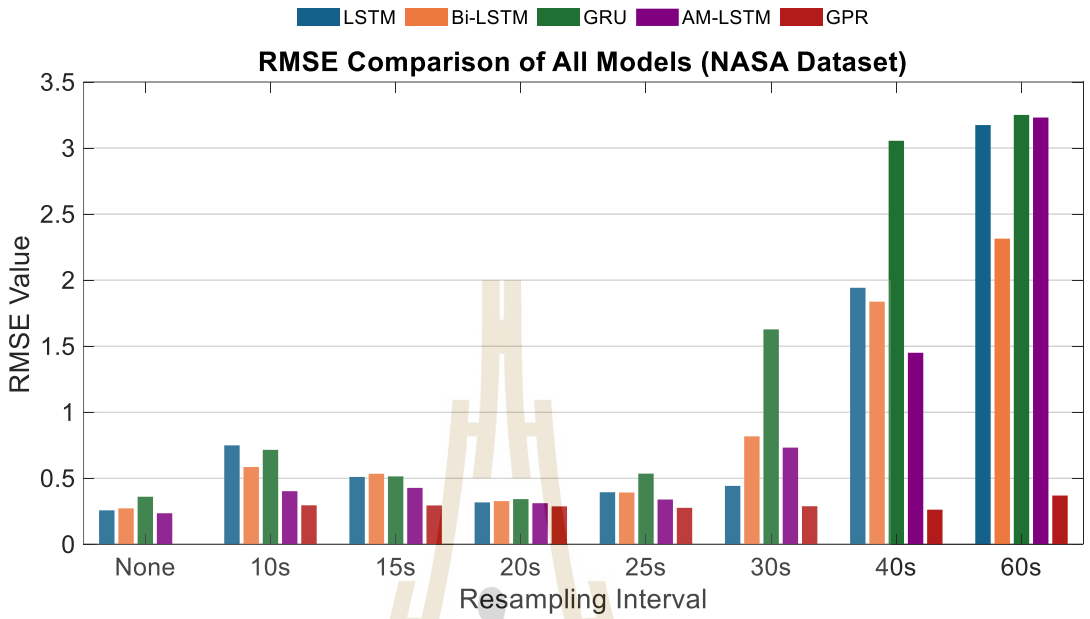


Figure 4.29 RMSE comparison of all models on the NASA dataset

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 4.5 Comparison of SOH Estimation Results on NASA Dataset

Dataset	Resampling Interval	LSTM				Bi-LSTM				GRU				AM-LSTM				GPR			
		RMSE	MSE	MAE	MAPE	RMSE	MSE	MAE	MAPE	RMSE	MSE	MAE	MAPE	RMSE	MSE	MAE	MAPE	RMSE	MSE	MAE	MAPE
NASA	None	0.257	0.066	0.201	0.242	0.272	0.074	0.228	0.271	0.360	0.130	0.317	0.377	0.235	0.055	0.187	0.229	-	-	-	-
	10 s	0.749	0.562	0.636	0.737	0.585	0.342	0.498	0.617	0.715	0.512	0.648	0.795	0.402	0.161	0.336	0.407	0.295	0.087	0.240	0.298
	15 s	0.510	0.260	0.437	0.525	0.534	0.286	0.420	0.485	0.514	0.264	0.415	0.485	0.427	0.182	0.344	0.413	0.294	0.086	0.244	0.301
	20 s	0.317	0.100	0.252	0.306	0.327	0.107	0.273	0.325	0.342	0.117	0.276	0.336	0.311	0.097	0.247	0.290	0.287	0.082	0.237	0.293
	25 s	0.394	0.156	0.313	0.380	0.392	0.153	0.310	0.369	0.535	0.286	0.451	0.554	0.339	0.115	0.269	0.325	0.276	0.076	0.223	0.278
	30 s	0.442	0.195	0.350	0.424	0.817	0.667	0.643	0.807	1.627	2.645	1.568	1.905	0.732	0.535	0.582	0.725	0.288	0.083	0.238	0.294
	40 s	1.942	3.769	1.890	2.277	1.837	3.373	1.748	2.087	3.055	9.329	2.752	3.344	1.450	2.102	1.344	1.601	0.262	0.069	0.212	0.262
	60 s	3.174	10.073	2.762	3.358	2.314	5.356	2.259	2.719	3.251	10.570	2.820	3.437	3.231	10.439	2.813	3.419	0.369	0.137	0.236	0.295

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 4.6 Execution Time of Each Model for SOH Estimation on NASA Dataset

Dataset	Resampling Interval	LSTM (s)	Bi-LSTM (s)	GRU (s)	AM-LSTM (s)	GPR (s)
NASA	None	1,749.56	2,175.9	1,602.4	2,759.3	-
	10 s	134.4283	231.0435	134.9282	254.1212	1,249.0148
	15 s	108.6076	122.1542	91.2231	179.9841	224.0689
	20 s	68.7088	102.988	66.7452	119.223	98.7963
	25 s	62.6125	75.981	57.9775	107.0138	47.8137
	30 s	45.5109	58.8247	44.2024	83.9645	33.577
	40 s	34.8527	43.4884	33.3974	62.2606	17.8264
	60 s	28.7726	34.8007	26.8101	44.6983	11.0508

3) Summary

This chapter presented an analysis of the proposed methodology for SOH estimation, focusing on the interaction between dataset characteristics and processing strategies. The experimental results highlighted fundamental differences between the Oxford and NASA datasets: Oxford is characterized by high-density data with significant noise, while NASA exhibits complex electrochemical behaviors, specifically the capacity regeneration phenomena.

Empirical evidence confirmed that the AM-LSTM model achieved superior performance compared to baseline models. For the Oxford dataset, the model attained its best accuracy with an RMSE of 0.2920 when paired with a resampling interval of 100 seconds, which effectively filtered out high-frequency noise. Conversely, for the NASA dataset, the model achieved its optimal performance with an RMSE of

0.2351 by utilizing raw data or short-interval resampling, thereby preserving the fine-grained details required to track capacity recovery.

Beyond accuracy, the most significant advantage of applying the resampling technique proved to be computational efficiency. Specifically, for the Oxford dataset, the results demonstrated a drastic reduction in model training time by over 98%, accompanied by a significant 58.58% reduction in estimation error compared to the raw data baseline. In the case of the NASA dataset, although resampling led to a 32.34% increase in RMSE, this is considered an acceptable trade-off given the massive reduction in training time of over 95%. These findings confirm that the proposed approach successfully strikes a balance between high accuracy and computational feasibility, making it highly suitable for implementation in real-world battery management systems with limited resources.

มหาวิทยาลัยเทคโนโลยีสุรนารี

CHAPTER V
CONCLUSION

5.1 Conclusion

This thesis presents a comprehensive study on SOH estimation for LIBs by integrating effective data management strategies with advanced deep learning techniques. The experimental results lead to the following conclusions:

First, the application of the resampling technique proved effective in reducing the size and complexity of battery measurement data. This reduction directly correlated with a significant decrease in computational load and model training time compared to using raw data. Crucially, this process preserved the essential information required for accurate estimation. By selecting a resampling interval aligned with the physical characteristics of the specific dataset, the method successfully filtered out noise in high-fluctuation scenarios while retaining the fine-grained details of complex electrochemical behaviors.

Second, the evaluation across multiple open-source datasets (Oxford and NASA) demonstrated the robustness of the proposed approach under distinct operating conditions. The findings indicate that deep learning models, particularly the developed architecture, possess the capability to learn and adapt to diverse degradation patterns, ranging from monotonic decay to highly fluctuating trends. When compared to traditional statistical methods like GPR, deep learning techniques exhibited superior accuracy and stability in handling large-scale, complex datasets. These outcomes suggest that the proposed approach serves as a viable and efficient solution for practical implementation in battery management systems constrained by limited computational resources.

5.2 Suggestions

Based on the findings and limitations of this study, several recommendations are proposed for future research to further enhance the reliability and applicability of SOH estimation models:

1) Validation with Real-world Driving Profiles: While this study utilized standard laboratory datasets, future work should verify the proposed method using data collected from real-world electric vehicle operations. Real-world conditions often involve highly dynamic current loads, irregular charging habits, and varying ambient temperatures, which could test the robustness of the resampling and estimation algorithms more rigorously.

2) Integration of Additional Health Indicators: To improve estimation accuracy, particularly for different battery chemistries, future research could incorporate additional input features beyond voltage and current. Integrating temperature data or integrating mechanical stress sensors (e.g., expansion/strain measurements) could provide a more holistic view of the battery's health status.

3) Spatial Representation in Large-Scale Systems: For practical deployment in large-scale energy storage systems, environmental factors such as temperature distribution often vary significantly depending on the physical location of the battery cells. Relying on training data from a single localized cluster may fail to capture these environmental heterogeneities. Therefore, future implementations should adopt a spatially stratified sampling strategy, selecting representative cells from diverse physical locations across the entire battery pack to ensure the model generalizes well to all environmental conditions within the system.

4) Model Retraining and Lifecycle Adaptability: Since battery degradation is a non-linear and time-varying process, a static model trained solely on historical data may eventually experience performance degradation due to concept drift where the underlying data distribution shifts as the battery ages. To maintain estimation accuracy throughout the entire lifespan, it is recommended to implement a periodic retraining strategy. Future implementations should consider triggering model updates based on usage intensity or performance degradation metrics for instance, recalibrating the model every 50–100 cycles or whenever the prediction error exceeds a user-defined

threshold (e.g., RMSE > 3%). Establishing these optimal triggers is a critical area for further investigation.

Therefore, while this research confirms the technical efficacy of the proposed framework, its transition to practical application requires addressing the dual challenges of interpretability and adaptability. As deep learning models inherently lack transparency, ensuring explainable outcomes is crucial for meeting rigorous safety standards. Simultaneously, the rapid evolution of energy storage demands a system capable of adapting to emerging battery chemistries and novel degradation mechanisms. Addressing these factors is essential to guarantee that the estimation framework remains both robust and relevant in a continuously advancing technological landscape.

มหาวิทยาลัยเทคโนโลยีสุรนารี

REFERENCES

Aliberti, A., Boni, F., Perol, A., Zampolli, M., Jaboeuf, R. J. P., Tosco, P., Macii, E., & Patti, E. (2022, 27 June-1 July 2022). Comparative Analysis of Neural Networks Techniques for Lithium-ion Battery SOH Estimation. 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC),

Audin, P., Jorge, I., Mesbahi, T., Samet, A., Beuvron, F. D. B. D., & Boné, R. (2021, 13-16 Dec. 2021). Auto-encoder LSTM for Li-ion SOH prediction: a comparative study on various benchmark datasets. 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA),

Bian, X., Wei, Z., Li, W., Pou, J., Sauer, D. U., & Liu, L. (2022). State-of-Health Estimation of Lithium-Ion Batteries by Fusing an Open Circuit Voltage Model and Incremental Capacity Analysis. *IEEE Transactions on Power Electronics,* 37(2), 2226-2236. https://doi.org/10.1109/TPEL.2021.3104723

Dineva, A. (2024). Evaluation of Advances in Battery Health Prediction for Electric Vehicles from Traditional Linear Filters to Latest Machine Learning Approaches. *Batteries,* 10(10).

Fan, Y., Xiao, F., Li, C., Yang, G., & Tang, X. (2020, 2020/12/01/). A novel deep learning framework for state of health estimation of lithium-ion battery. *Journal of Energy Storage,* 32, 101741. https://doi.org/https://doi.org/10.1016/j.est.2020.101741

มหาวิทยาลัยเทคโนโลยีสุรนารี

Gong, Q., Wang, P., Cheng, Z., & Zhang, J. a. (2021, 2021/11/18). A Method for Estimating State of Charge of Lithium-Ion Batteries Based on Deep Learning. *Journal of The Electrochemical Society,* 168(11), 110532. https://doi.org/10.1149/1945-7111/ac3719

Hossain Lipu, M. S., Ansari, S., Miah, M. S., Meraj, S. T., Hasan, K., Shihavuddin, A. S. M., Hannan, M. A., Muttaqi, K. M., & Hussain, A. (2022, 2022/11/25/). Deep learning enabled state of charge, state of health and remaining useful life estimation for smart battery management system: Methods, implementations, issues and prospects. *Journal of Energy Storage,* 55, 105752. https://doi.org/https://doi.org/10.1016/j.est.2022.105752

Jarraya, I., Ben Atitallah, S., Alahmed, F., Abdelkader, M., Driss, M., Abdelhadi, F., & Koubaa, A. (2025, 2025/06/30/). SOH-KLSTM: A hybrid Kolmogorov-Arnold Network and LSTM model for enhanced Lithium-ion battery Health Monitoring. *Journal of Energy Storage,* 122, 116541. https://doi.org/https://doi.org/10.1016/j.est.2025.116541

Li, C., Han, X., Zhang, Q., Li, M., Rao, Z., Liao, W., Liu, X., Liu, X., & Li, G. (2023, 2023/12/25/). State-of-health and remaining-useful-life estimations of lithium-ion battery based on temporal convolutional network-long short-term memory. *Journal of Energy Storage,* 74, 109498. https://doi.org/https://doi.org/10.1016/j.est.2023.109498

Li, Y., Qin, X., Chai, M., Wu, H., Zhang, F., Jiang, F., & Wen, C. (2025, 2025/09/01/). SOH evaluation and RUL estimation of lithium-ion batteries based on MC-CNN-TimesNet model. *Reliability Engineering & System Safety,* 261, 111125. https://doi.org/https://doi.org/10.1016/j.ress.2025.111125

มหาวิทยาลัยเทคโนโลยีสุรนารี

Ma, W., Ha, S., Pozzato, G., Saatchi, A., Aliahmad, N., Littau, K., & Onori, S. (2024, 2025/01/06). Engineering Testing Protocols for Machine Learning-Based SoH Estimation in Lithium Metal Batteries. *Journal of The Electrochemical Society,* 171(12), 120531. https://doi.org/10.1149/1945-7111/ad9cc8

Nguyen Van, C., & Quang, D. T. (2023, 2023/06/01/). Estimation of SoH and internal resistances of Lithium ion battery based on LSTM network. *International Journal of Electrochemical Science,* 18(6), 100166. https://doi.org/https://doi.org/10.1016/j.ijoes.2023.100166

Olabi, A. G., Abdelghafar, A. A., Soudan, B., Alami, A. H., Semeraro, C., Al Radi, M., Al-Murisi, M., & Abdelkareem, M. A. (2024, 2024/02/01/). Artificial neural network driven prognosis and estimation of Lithium-Ion battery states: Current insights and future perspectives. *Ain Shams Engineering Journal,* 15(2), 102429. https://doi.org/https://doi.org/10.1016/j.asej.2023.102429

Oyewole, I., Chehade, A., & Kim, Y. (2022, 2022/04/15/). A controllable deep transfer learning network with multiple domain adaptation for battery state-of-charge estimation. *Applied Energy,* 312, 118726. https://doi.org/https://doi.org/10.1016/j.apenergy.2022.118726

Rao, K. D., Anand, N. V., Pandraju, T. K. S., Alsaif, F., & Ustun, T. S. (2024). Optimally Tuned Gated Recurrent Unit Neural Network-Based State of Health Estimation Scheme for Lithium Ion Batteries. *IEEE Access,* 12, 58597-58607. https://doi.org/10.1109/ACCESS.2024.3392432

Safavi, V., Bazmohammadi, N., Vasquez, J. C., & Guerrero, J. M. (2024). Battery State-of-Health Estimation: A Step towards Battery Digital Twins. *Electronics,* 13(3).

Stephens, D., Zhou, Q., Hofmann, H., & Sun, J. (2024, 21-23 Aug. 2024). Integrated Incremental Capacity Analysis and Differential Thermal Analysis for Improved Robustness in Li-Ion Battery State of Health Estimation. 2024 IEEE Conference on Control Technology and Applications (CCTA),

Sun, H., Sun, J., Zhao, K., Wang, L., & Wang, K. (2022, 2022/01/01). Data-Driven ICA-Bi-LSTM-Combined Lithium Battery SOH Estimation. *Mathematical Problems in Engineering,* 2022(1), 9645892. https://doi.org/https://doi.org/10.1155/2022/9645892

Techanok, S., Junhuathon, N., & Chayakulkheeree, K. (2025, 5-7 March 2025). A Comparative Study on Feature Extraction for Battery State of Health Estimation: Evaluating the Role of Open Circuit Voltage in LSTM Model. 2025 13th International Electrical Engineering Congress (iEECON),

Wang, Z., Feng, G., Sun, X., Zhen, D., Gu, F., & Ball, A. D. (2022, 2022/03/01). Feature Extraction from Charging Profiles for State of Health Estimation of Lithium-ion Battery. *Journal of Physics: Conference Series,* 2184(1), 012024. https://doi.org/10.1088/1742-6596/2184/1/012024

Wei, Z., Li, Y., Sun, X., Liu, W., Liu, C., & Lu, H. (2025, 2025/06/01/). SOH estimation of lithium-ion batteries based on multi-feature extraction and improved DLEM. *Journal of Energy Storage,* 120, 116460. https://doi.org/https://doi.org/10.1016/j.est.2025.116460

Wu, J., Liu, Z., Zhang, Y., Lei, D., Zhang, B., & Cao, W. (2023, 2023/12/20/). Data-driven state of health estimation for lithium-ion battery based on voltage variation curves. *Journal of Energy Storage,* 73, 109191. https://doi.org/https://doi.org/10.1016/j.est.2023.109191

Xiong, X., Wang, Y., Li, K., & Chen, Z. (2023, 2023/11/25/). State of health estimation for lithium-ion batteries using Gaussian process regression-based data reconstruction method during random charging process. *Journal of Energy Storage,* 72, 108390. https://doi.org/10.1016/j.est.2023.108390

Zhang, L., Ji, T., Yu, S., & Liu, G. (2023). Accurate Prediction Approach of SOH for Lithium-Ion Batteries Based on LSTM Method. *Batteries,* 9(3).

Zhang, L., Zhang, J., Gao, T., Lyu, L., Wang, L., Shi, W., Jiang, L., & Cai, G. (2023, 2023/12/25/). Improved LSTM based state of health estimation using random segments of the charging curves for lithium-ion batteries. *Journal of Energy Storage,* 74, 109370. https://doi.org/https://doi.org/10.1016/j.est.2023.109370

Zou, B., Xiong, M., Wang, H., Ding, W., Jiang, P., Hua, W., Zhang, Y., Zhang, L., Wang, W., & Tan, R. (2023). A Deep Learning Approach for State-of-Health Estimation of Lithium-Ion Batteries Based on a Multi-Feature and Attention Mechanism Collaboration. *Batteries,* 9(6).

Zou, Z., Li, Y., Liu, X., Zhang, Z., Duan, B., & Zhang, C. (2023, 27-29 Oct. 2023). Fast Estimation of State of Health for Retired Lithium-ion Batteries Based on Principal Component Regression. 2023 7th CAA International Conference on Vehicular Control and Intelligence (CVCI),

Jafari, S., & Byun, Y.-C. (2025). Interpretable AI for explaining and predicting battery state of health using PSO-enhanced deep learning models. *Energy Reports, 14*, 1779-1798. https://doi.org/10.1016/j.egyr.2025.07.027

APPENDIX

มหาวิทยาลัยเทคโนโลยีสุรนารี

APPENDIX A

AUC Feature Extraction for Oxford Battery Data

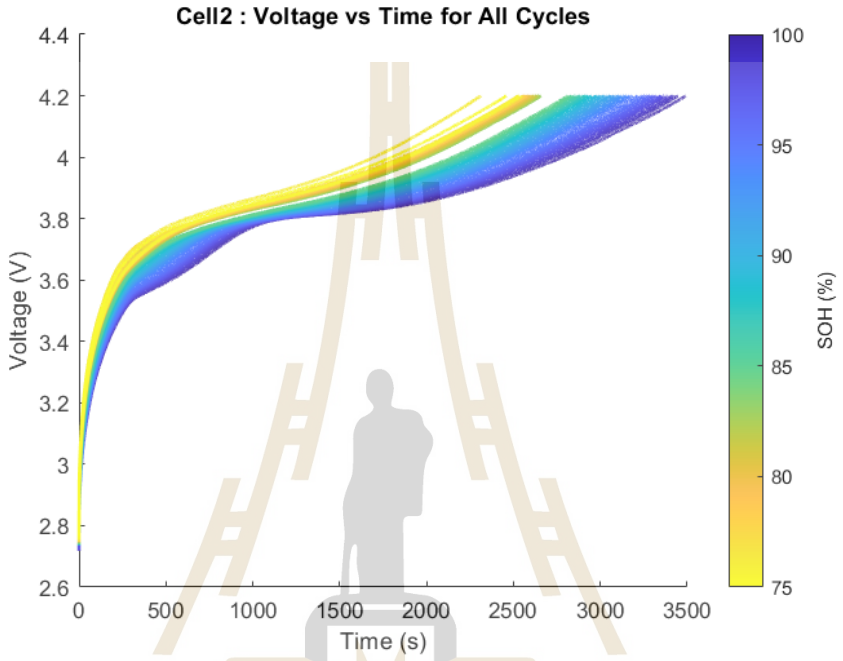


Figure A.1 Voltage profiles Cell 2 of Oxford are used to charge all the cycles.

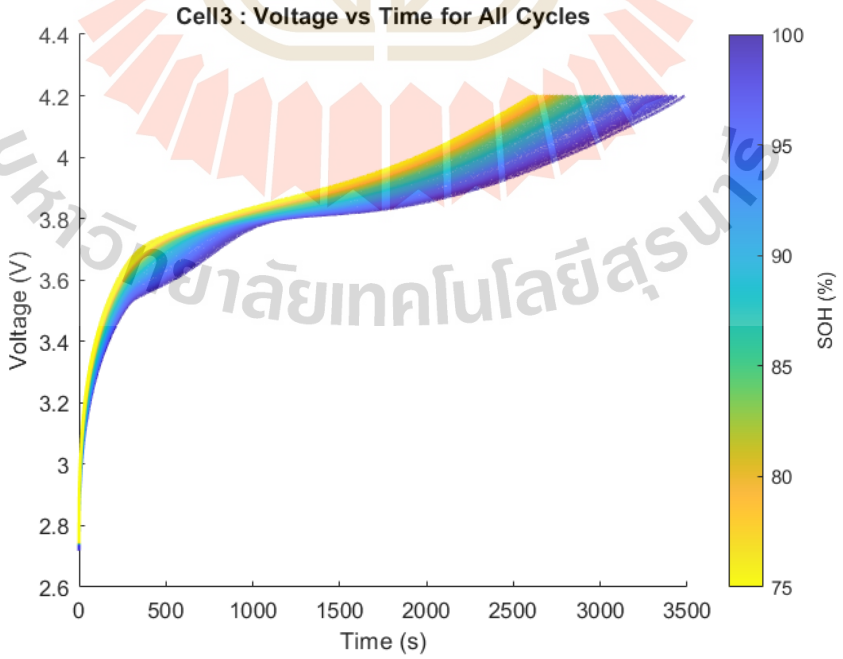


Figure A.2 Voltage profiles Cell 3 of Oxford are used to charge all the cycles.

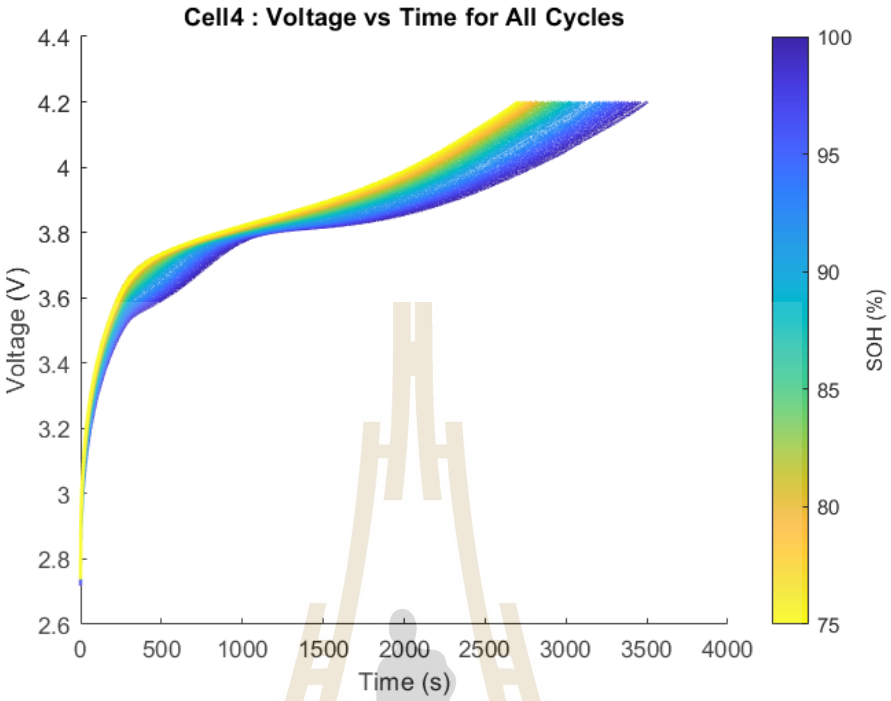


Figure A.3 Voltage profiles Cell 4 of Oxford are used to charge all the cycles.

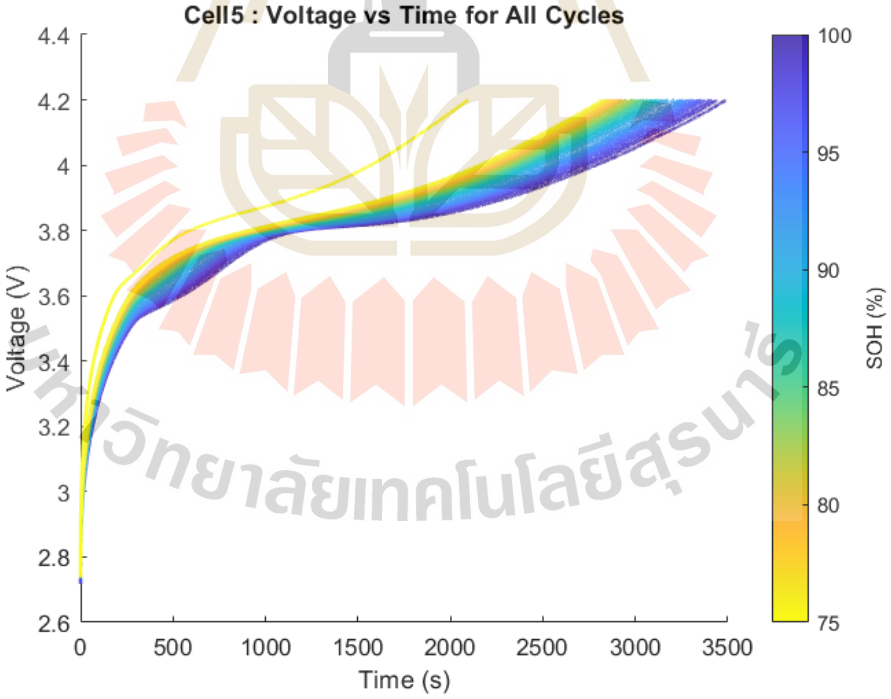


Figure A.4 Voltage profiles Cell 5 of Oxford are used to charge all the cycles.

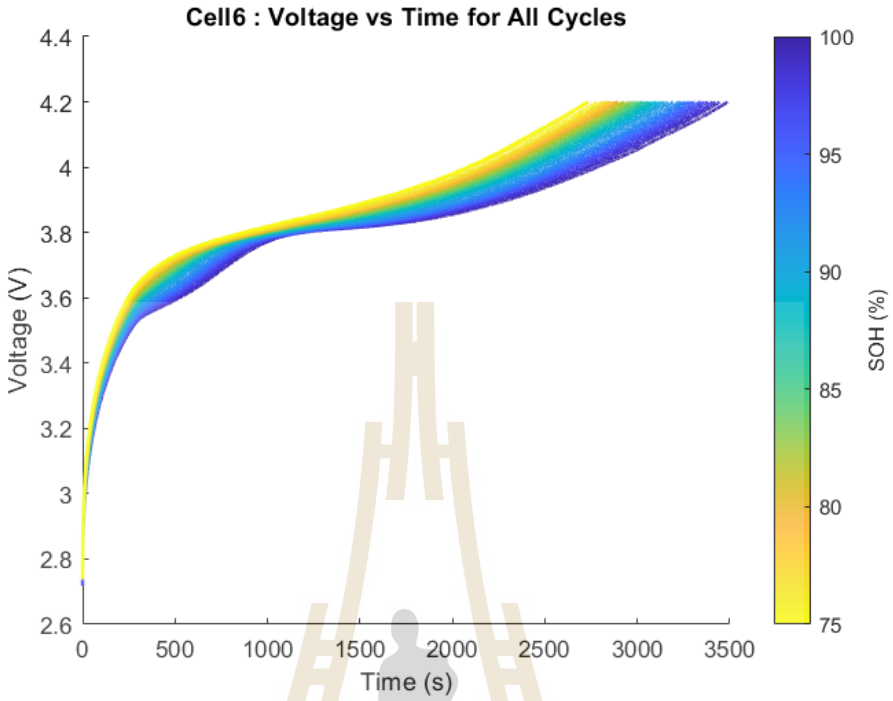


Figure A.5 Voltage profiles Cell 6 of Oxford are used to charge all the cycles.

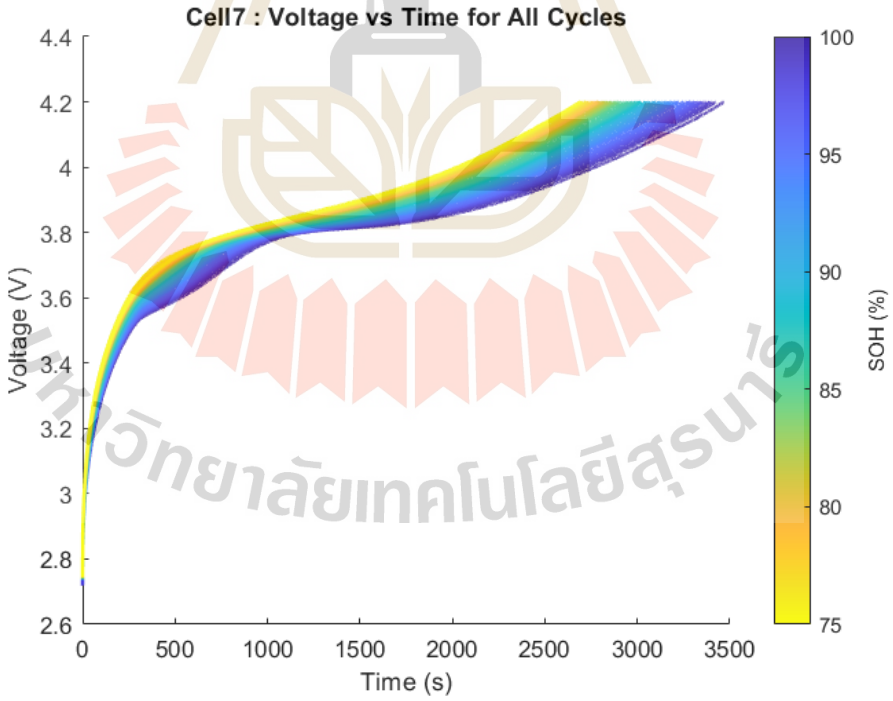


Figure A.6 Voltage profiles Cell 7 of Oxford are used to charge all the cycles.

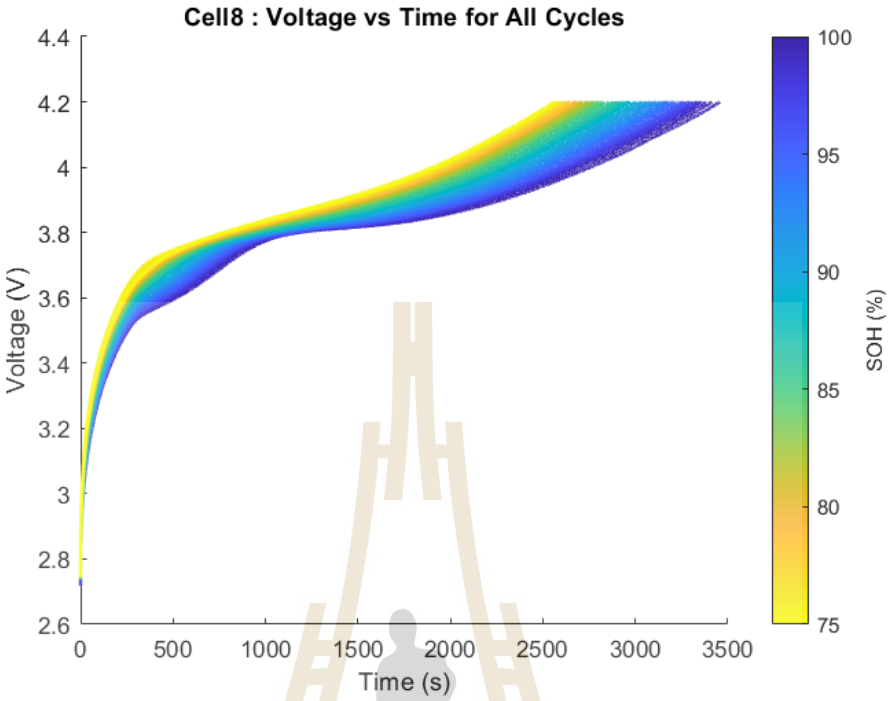


Figure A.7 Voltage profiles Cell 8 of Oxford are used to charge all the cycles.

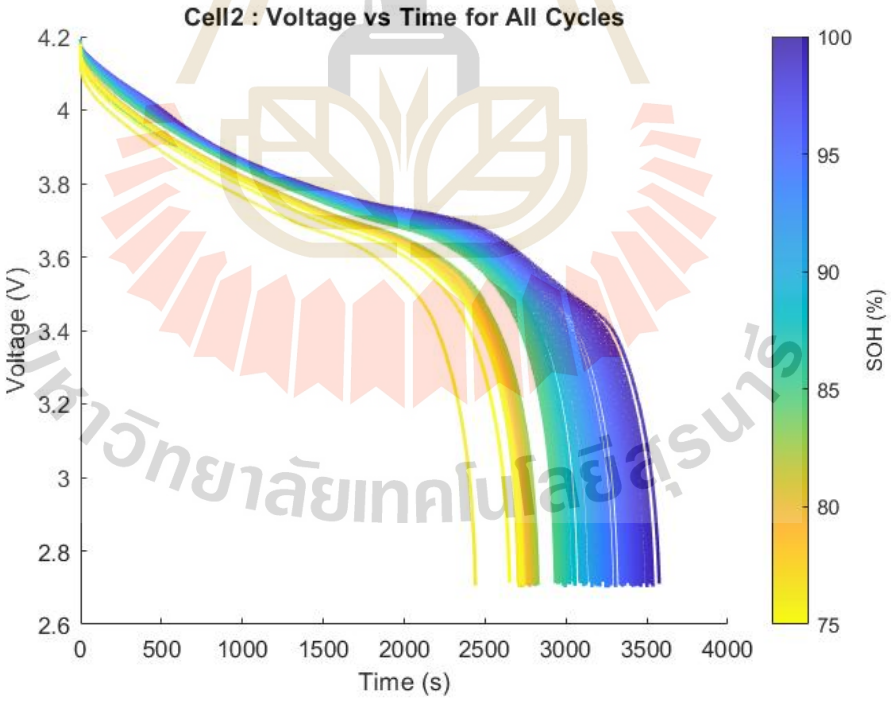


Figure A.8 Voltage profiles Cell 2 of Oxford are used to discharge all the cycles.

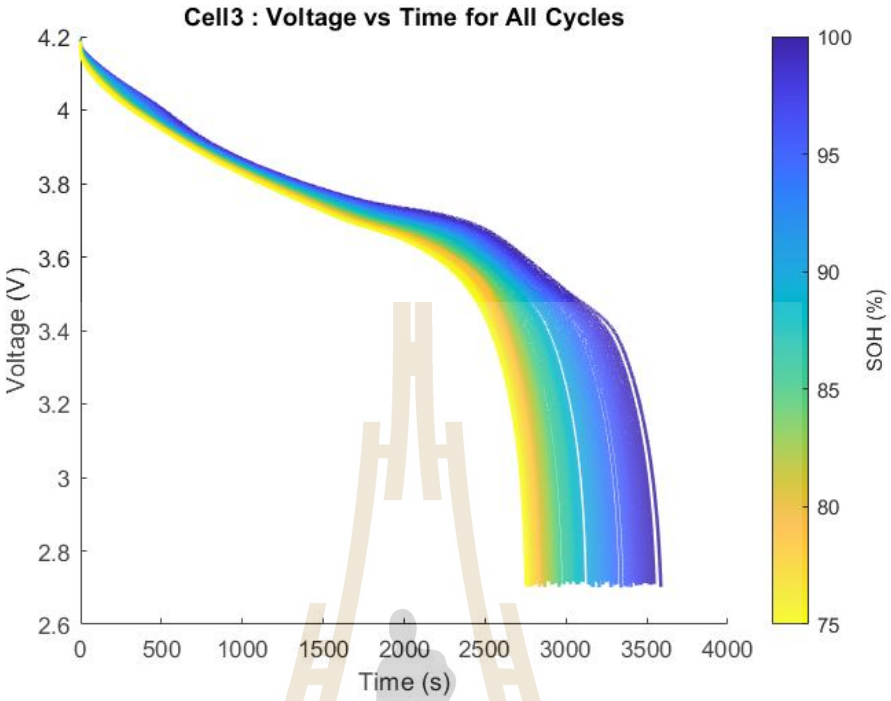


Figure A.9 Voltage profiles Cell 3 of Oxford are used to discharge all the cycles.

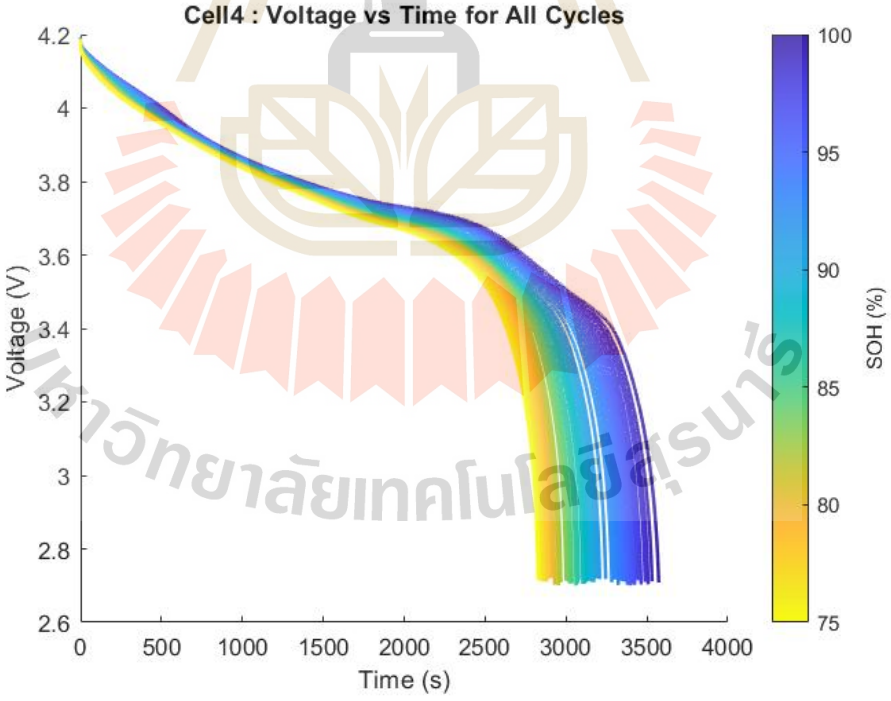


Figure A.10 Voltage profiles Cell 4 of Oxford are used to discharge all the cycles.

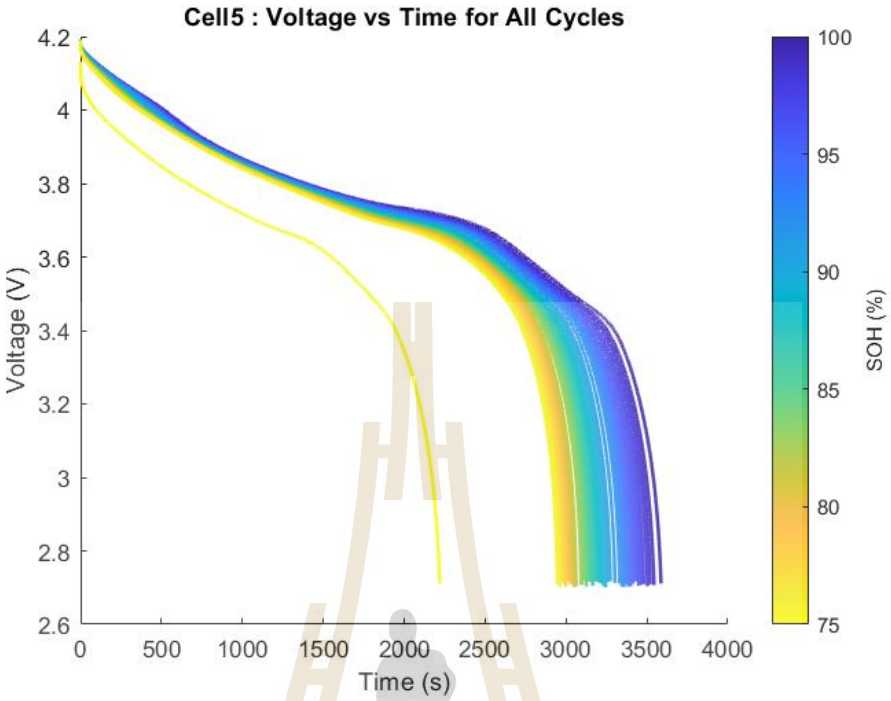


Figure A.11 Voltage profiles Cell 5 of Oxford are used to discharge all the cycles.

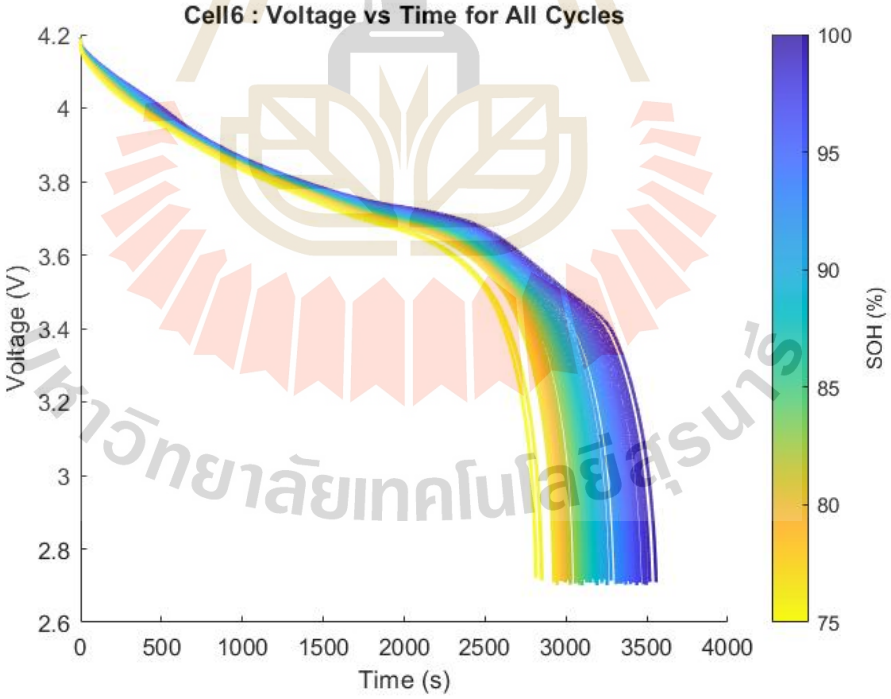


Figure A.12 Voltage profiles Cell 6 of Oxford are used to discharge all the cycles.

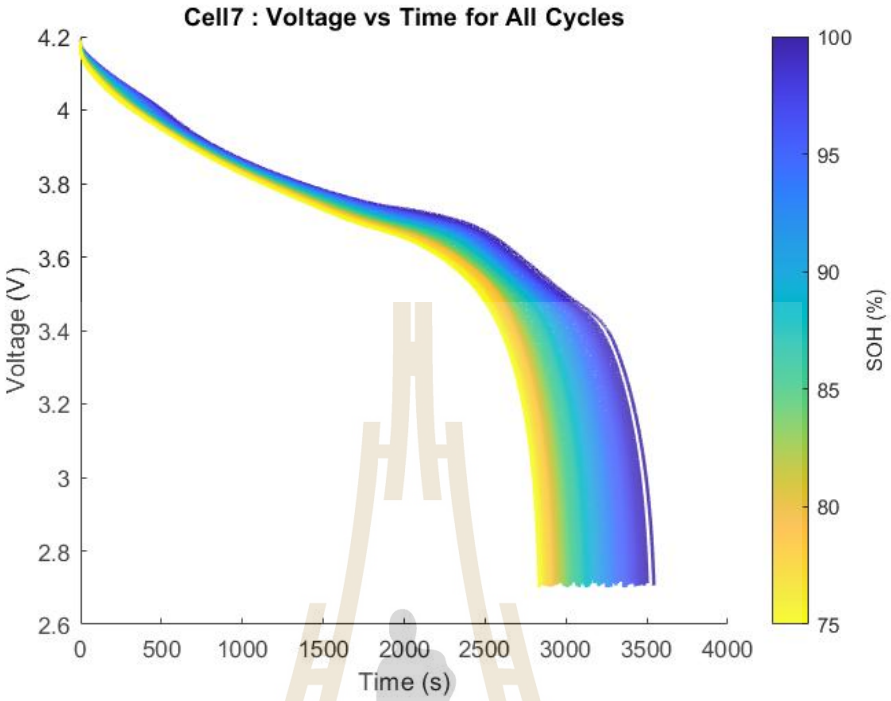


Figure A.13 Voltage profiles Cell 7 of Oxford are used to discharge all the cycles.

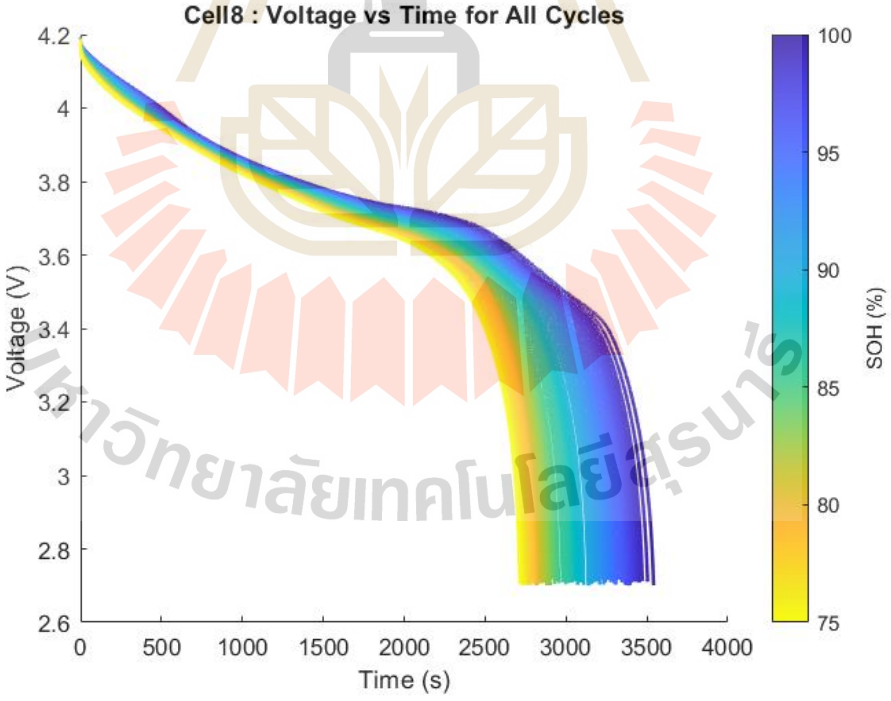


Figure A.14 Voltage profiles Cell 8 of Oxford are used to discharge all the cycles.

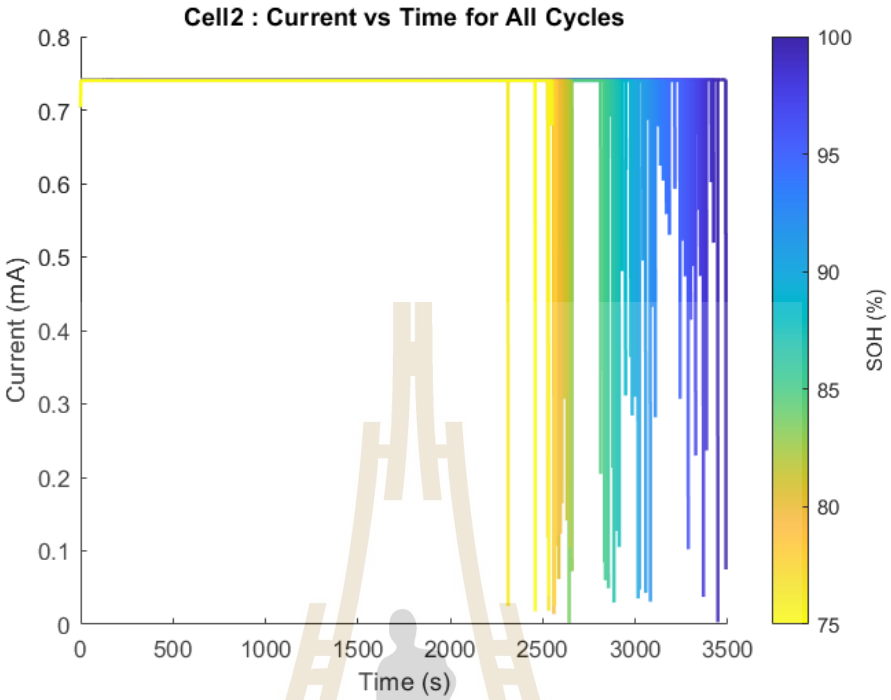


Figure A.15 Current profiles Cell 2 of Oxford are used to charge all the cycles.

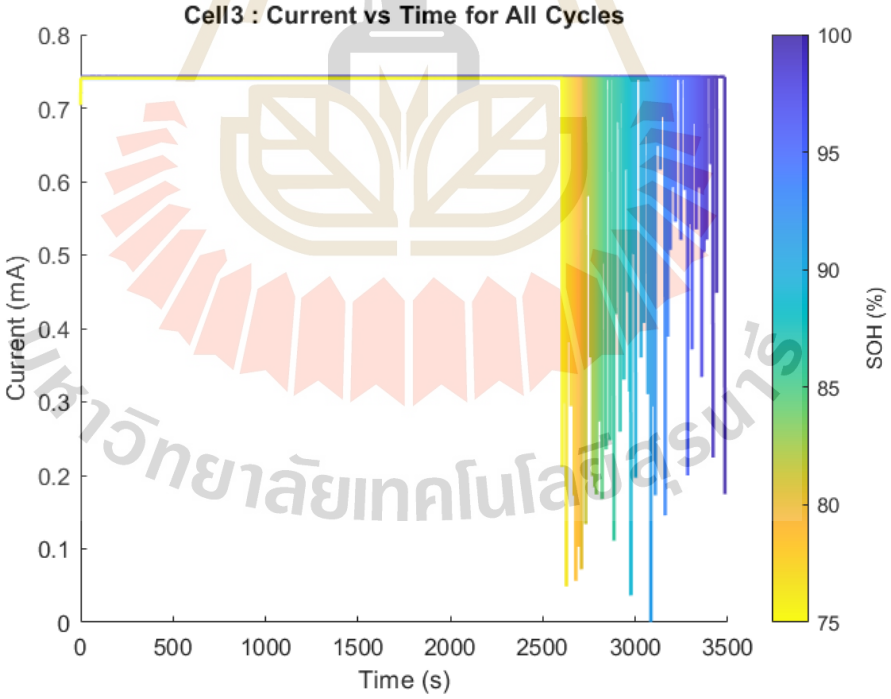


Figure A.16 Current profiles Cell 3 of Oxford are used to charge all the cycles.

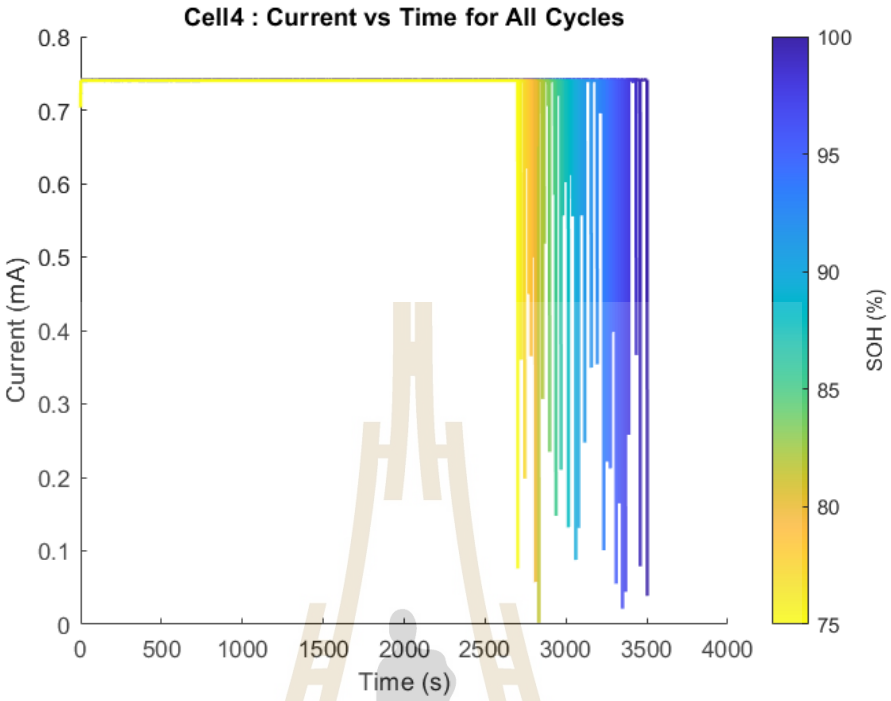


Figure A.17 Current profiles Cell 4 of Oxford are used to charge all the cycles.

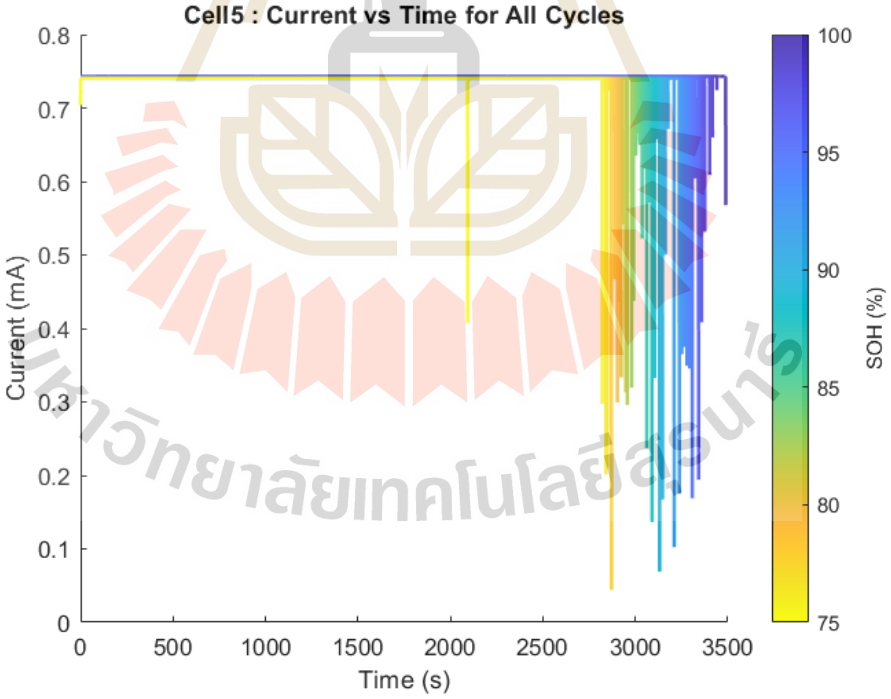


Figure A.18 Current profiles Cell 5 of Oxford are used to charge all the cycles.

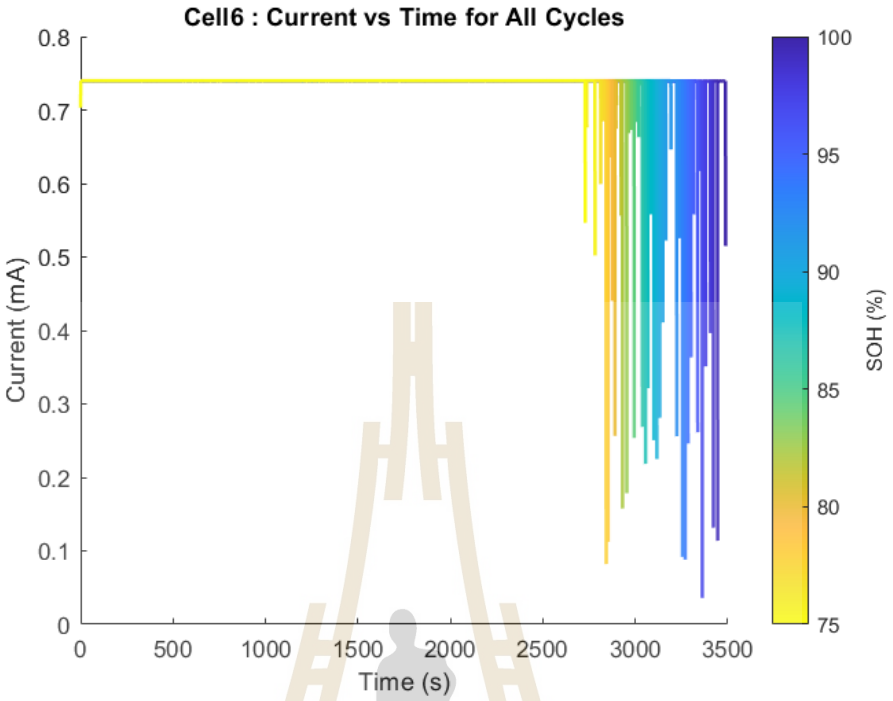


Figure A.19 Current profiles Cell 6 of Oxford are used to charge all the cycles.

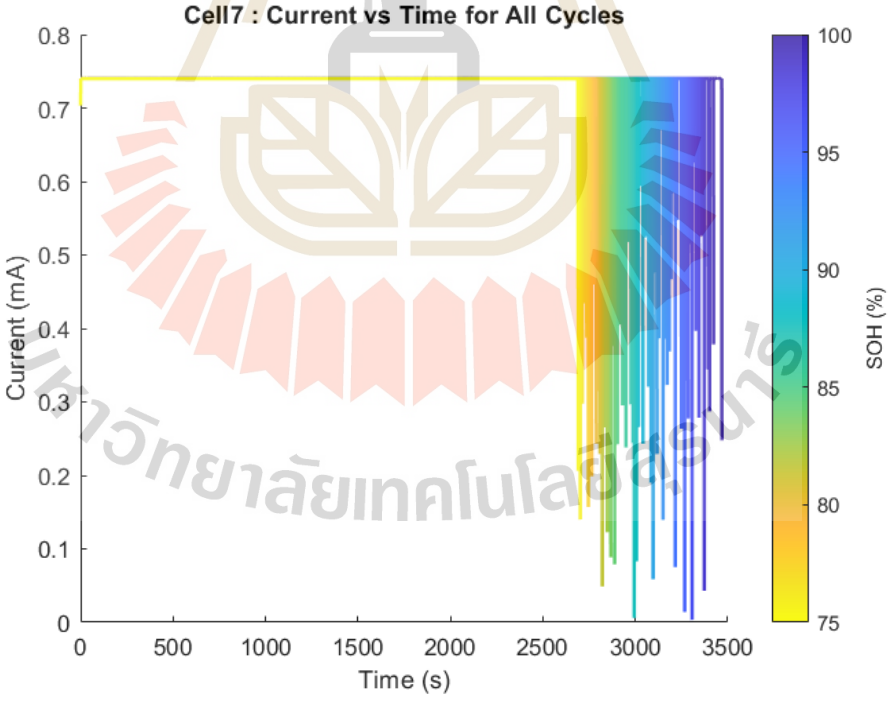


Figure A.20 Current profiles Cell 7 of Oxford are used to charge all the cycles.

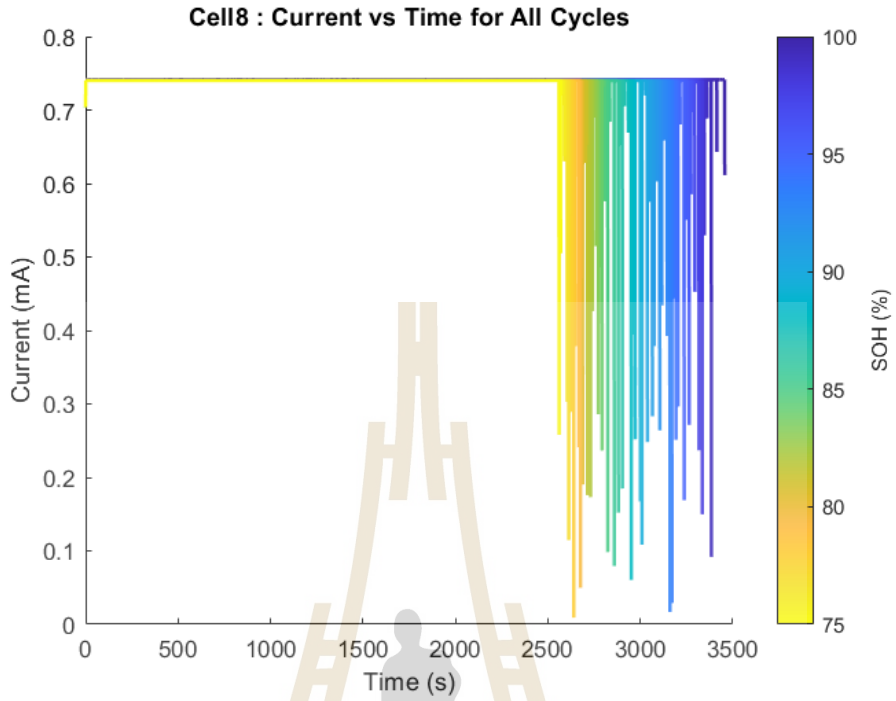


Figure A.21 Current profiles Cell 8 of Oxford are used to charge all the cycles.

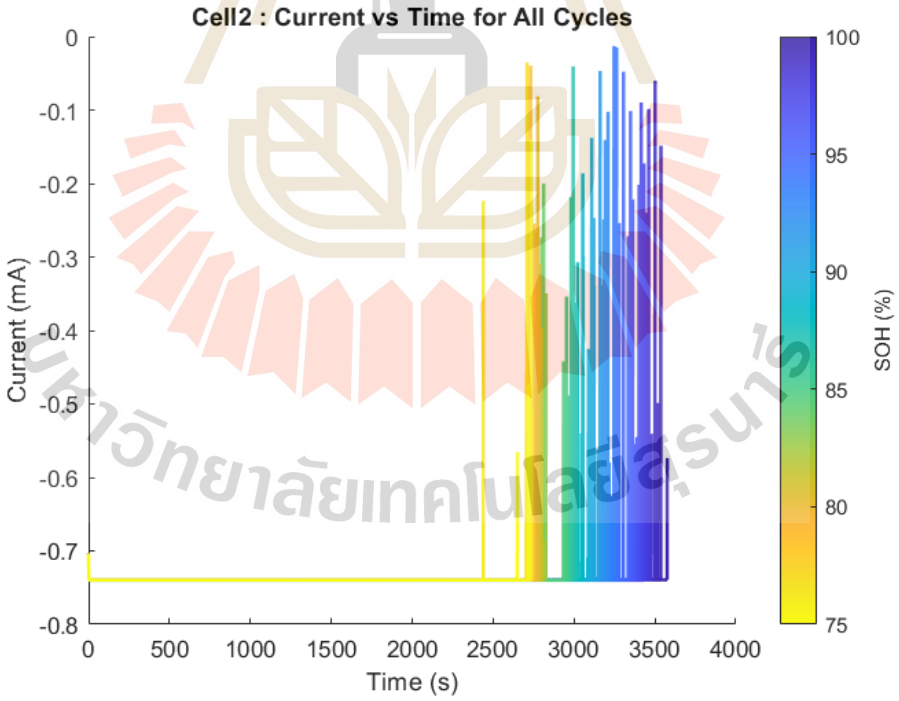


Figure A.22 Current profiles Cell 2 of Oxford are used to discharge all the cycles.

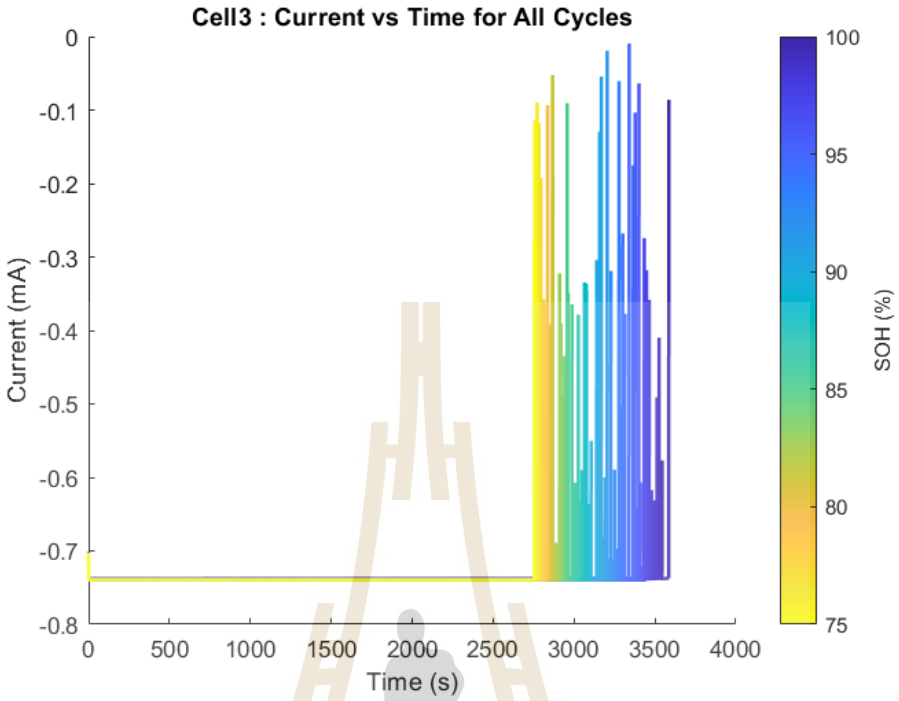


Figure A.23 Current profiles Cell 3 of Oxford are used to charge all the cycles.

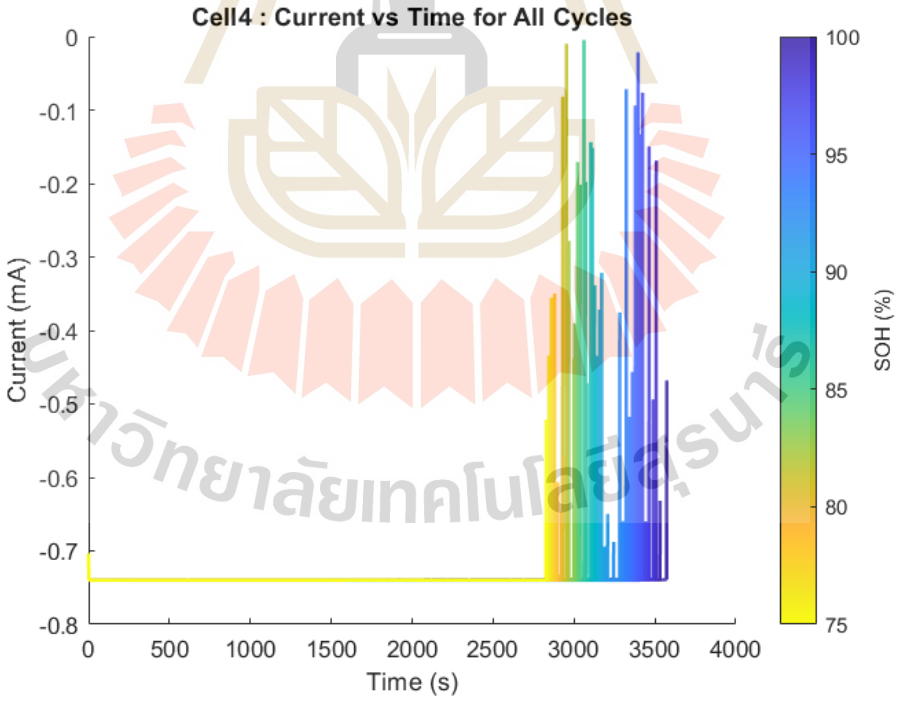


Figure A.24 Current profiles Cell 4 of Oxford are used to discharge all the cycles.

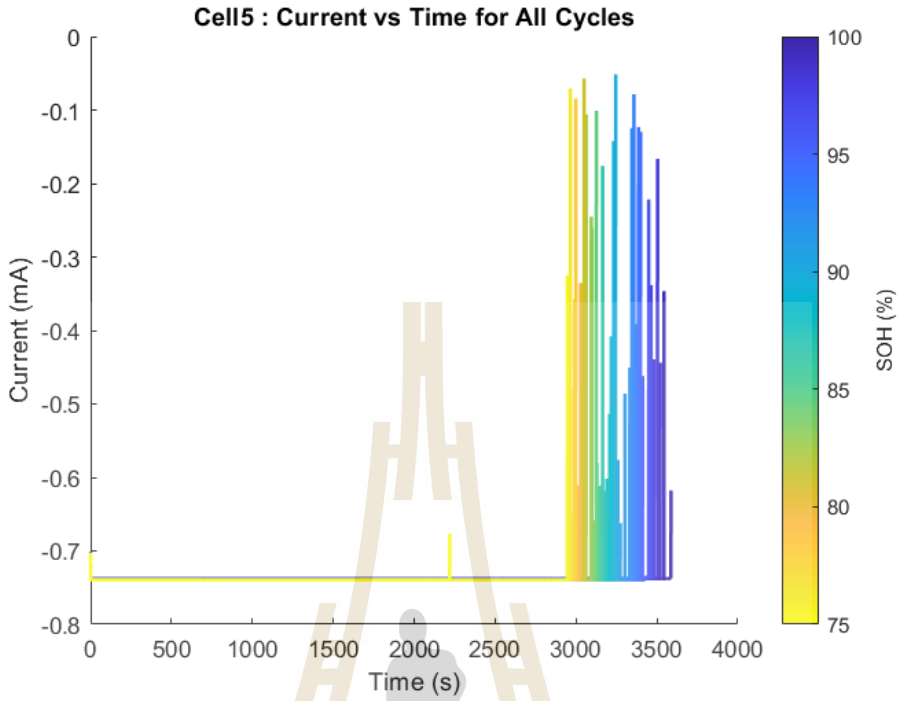


Figure A.25 Current profiles Cell 5 of Oxford are used to charge all the cycles.

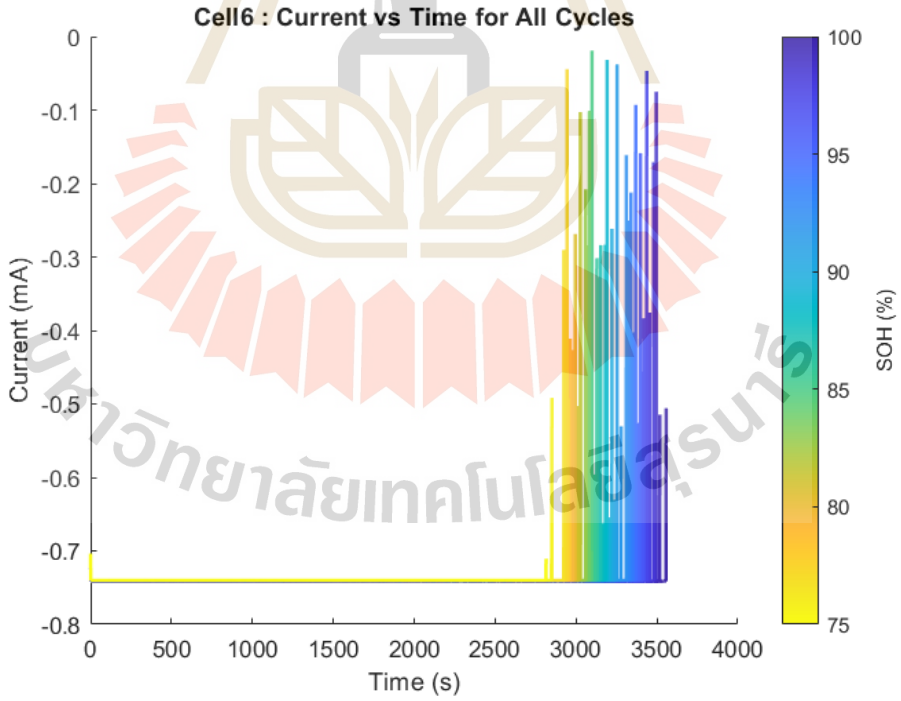


Figure A.26 Current profiles Cell 6 of Oxford are used to discharge all the cycles.

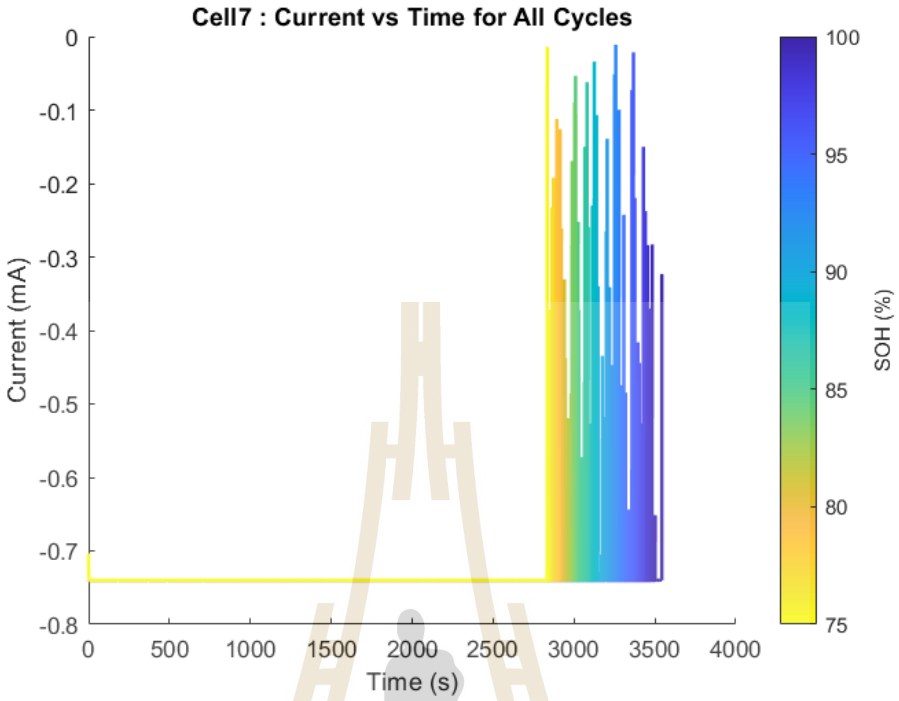


Figure A.27 Current profiles Cell 7 of Oxford are used to charge all the cycles.

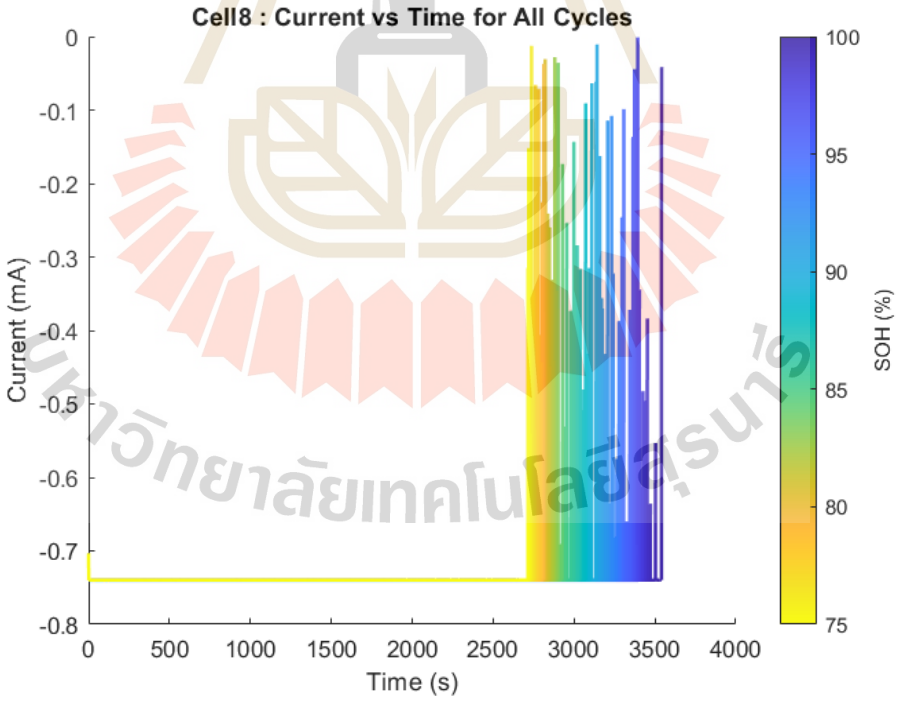


Figure A.28 Current profiles Cell 8 of Oxford are used to discharge all the cycles.

APPENDIX B

AUC Feature Extraction for NASA Battery Data optimization

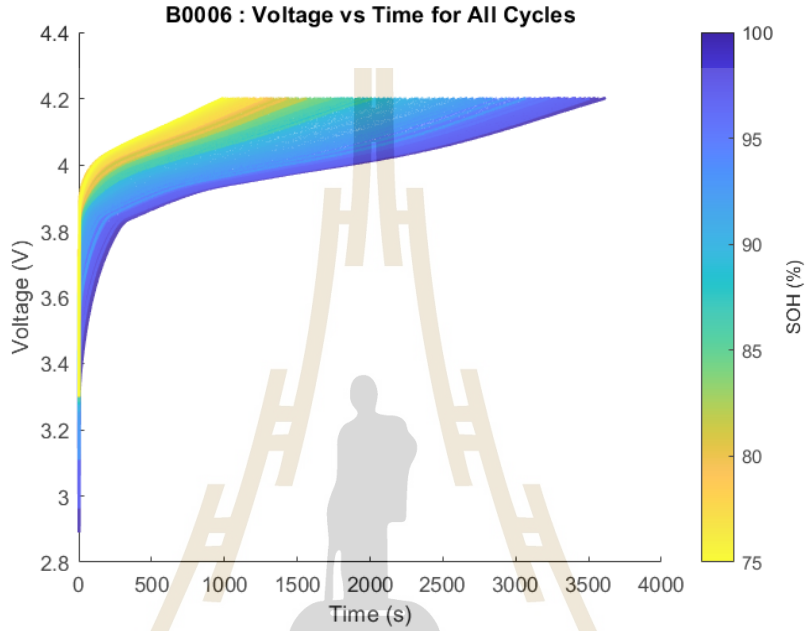


Figure B.1 Voltage profiles B0006 of NASA are used to charge all the cycles.

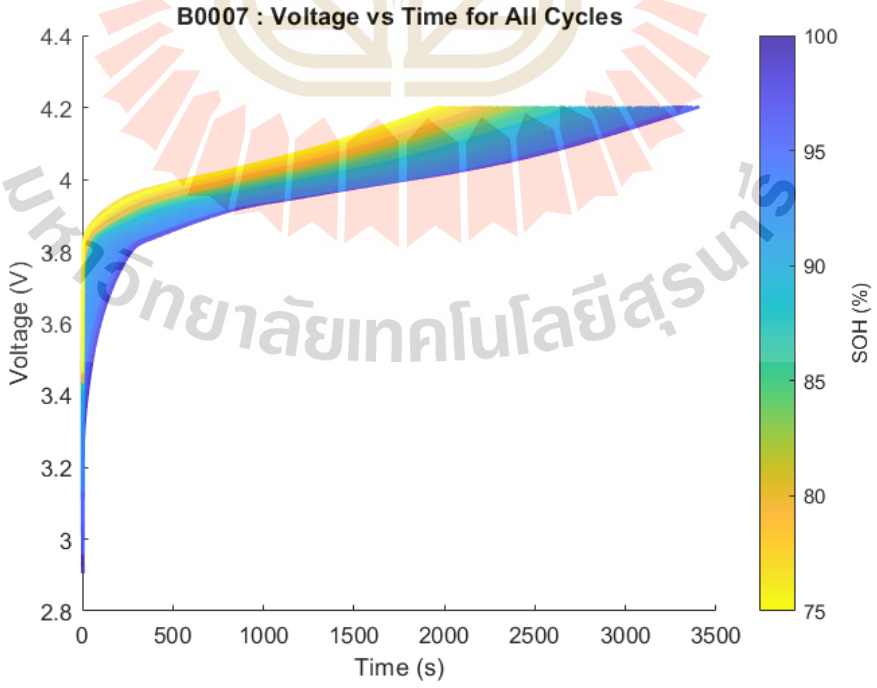


Figure B.2 Voltage profiles B0007 of NASA are used to charge all the cycles.

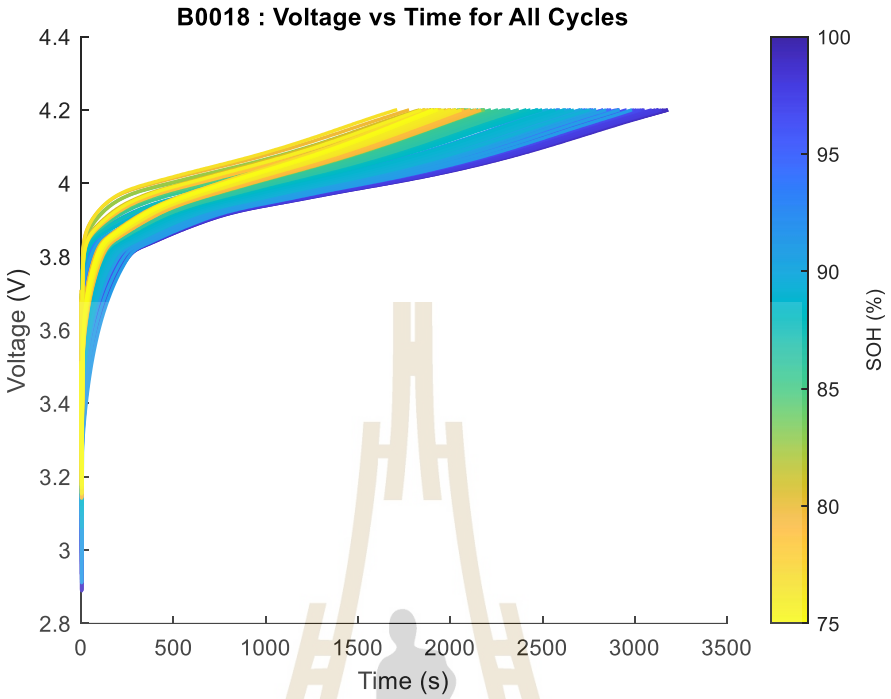


Figure B.3 Voltage profiles B0018 of NASA are used to charge all the cycles.

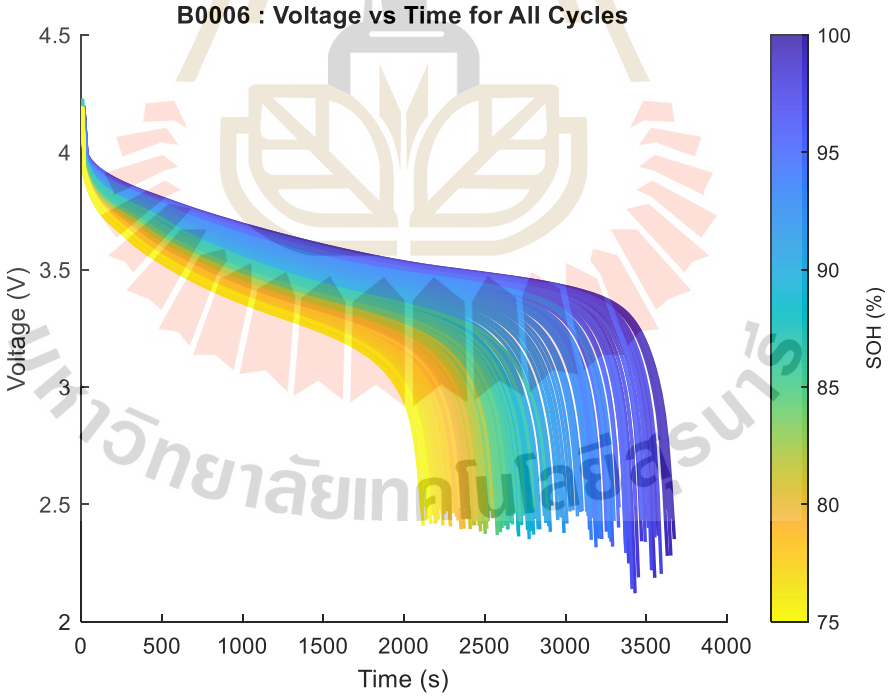


Figure B.4 Voltage profiles B0006 of NASA are used to discharge all the cycles.

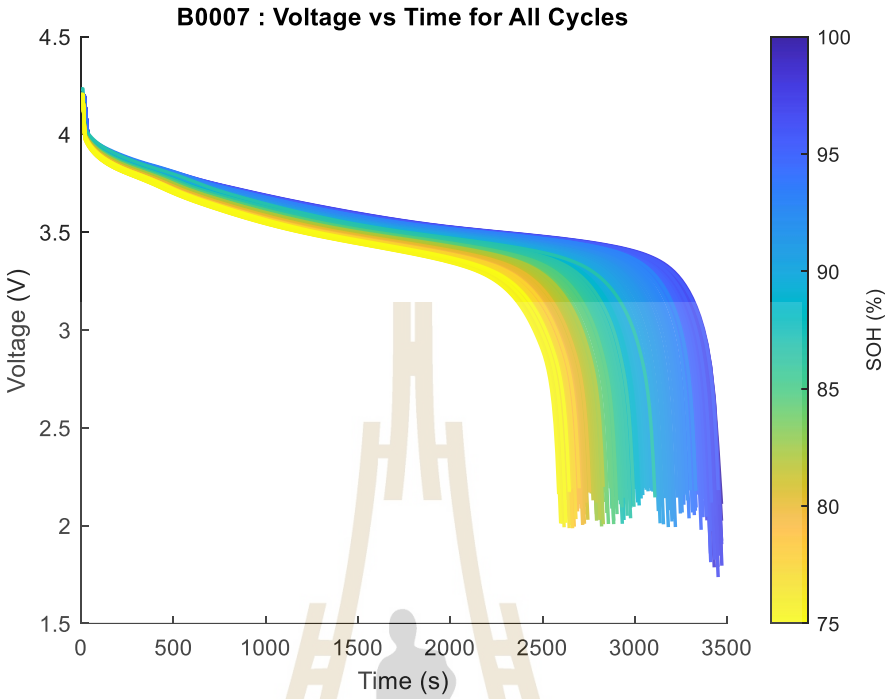


Figure B.5 Voltage profiles B0007 of NASA are used to discharge all the cycles.

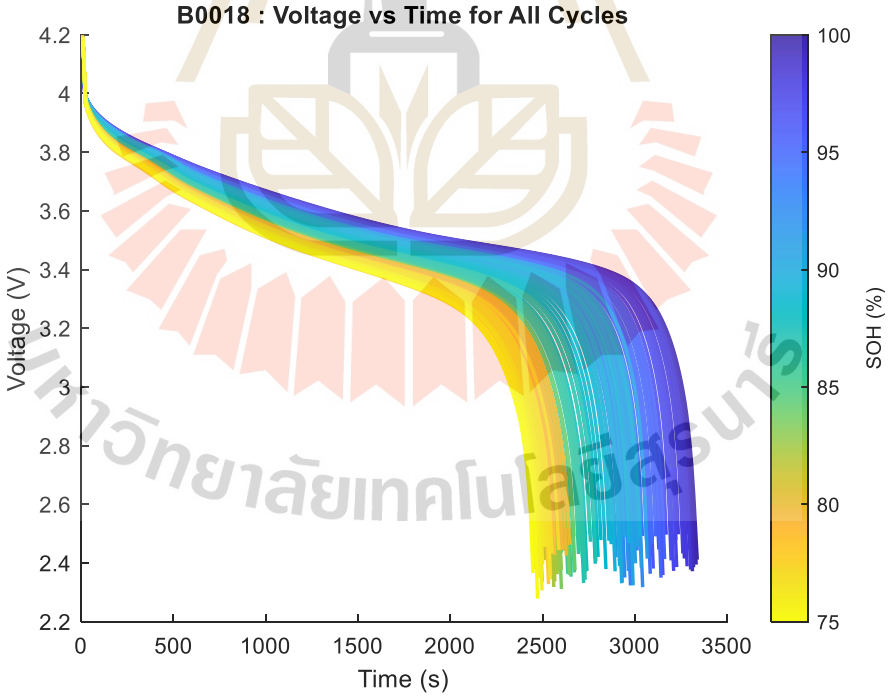


Figure B.6 Voltage profiles B0018 of NASA are used to discharge all the cycles.

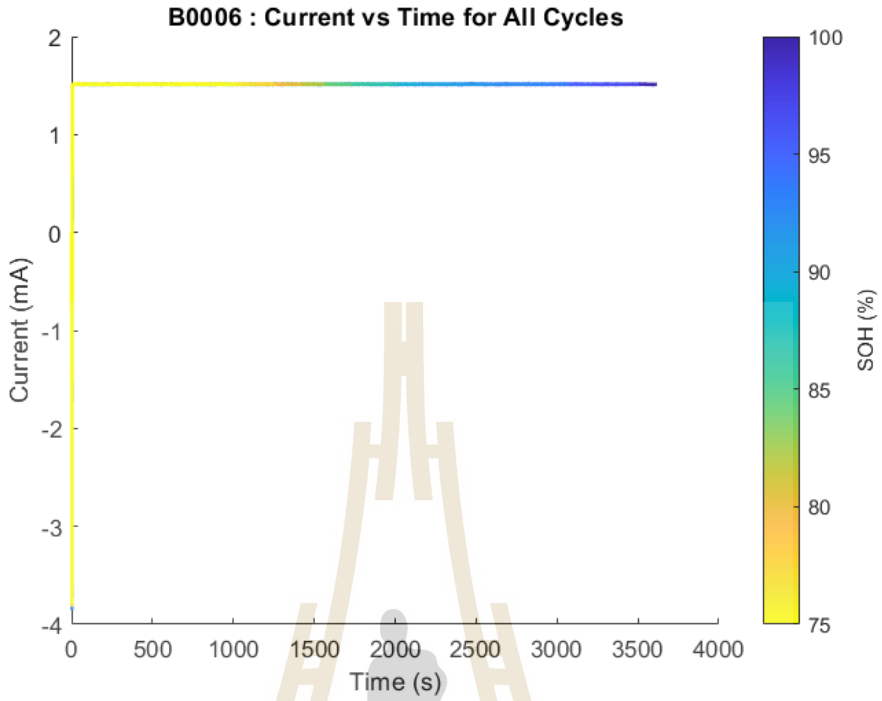


Figure B.7 Current profiles B0006 of NASA are used to charge all the cycles.

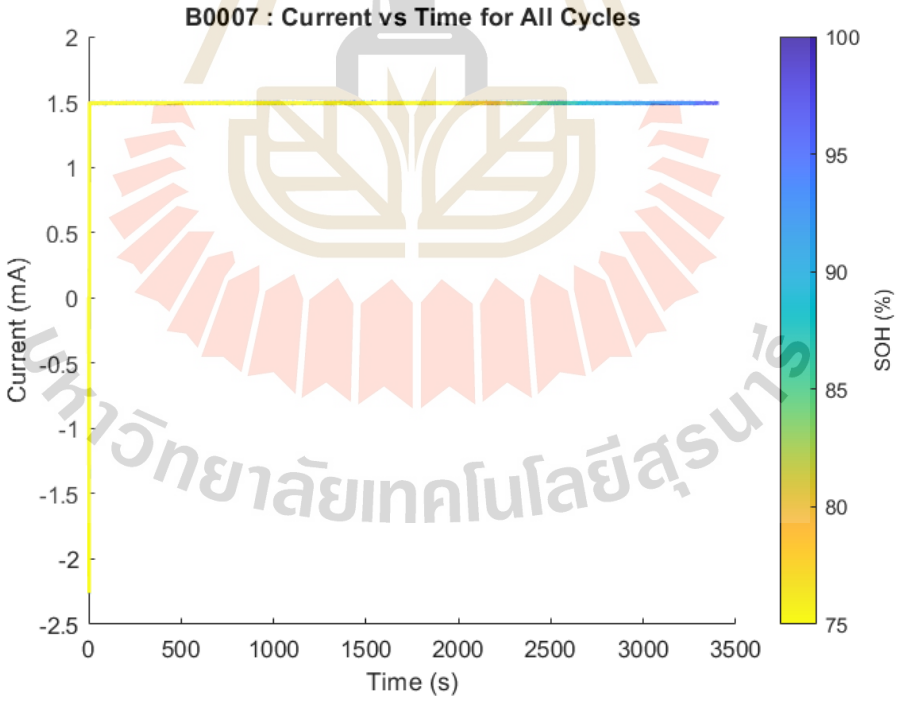


Figure B.8 Current profiles B0007 of NASA are used to charge all the cycles.

มหาวิทยาลัยเทคโนโลยีสุรนารี

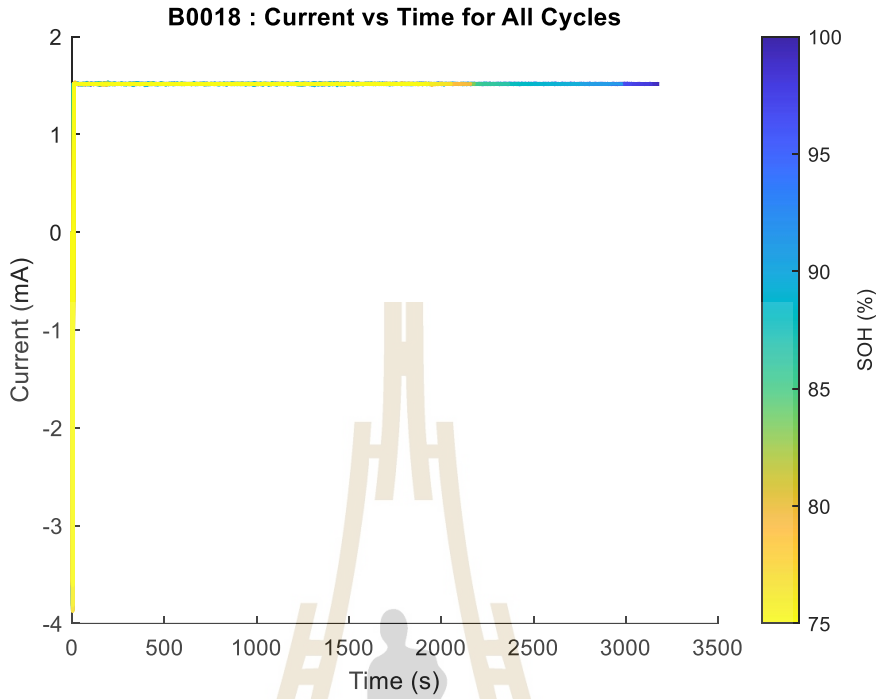


Figure B.9 Current profiles B0018 of NASA are used to charge all the cycles.

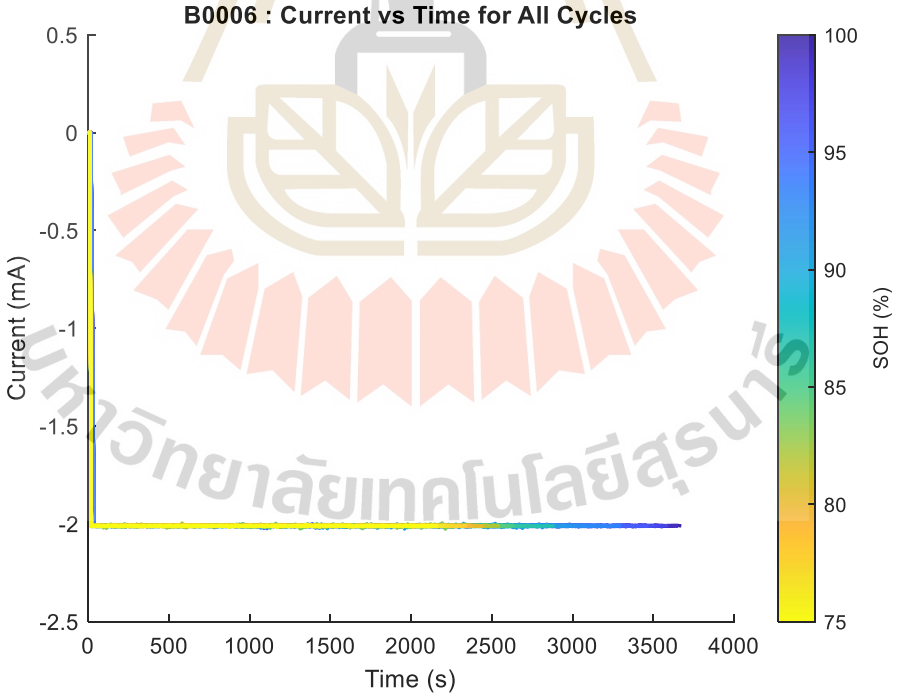


Figure B.10 Current profiles B0006 of NASA are used to discharge all the cycles.

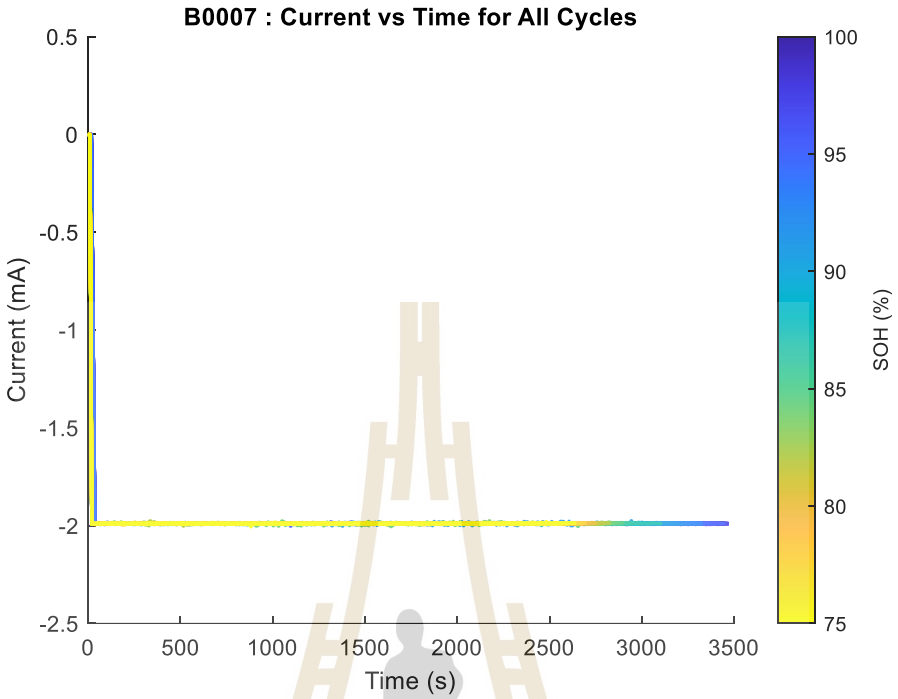


Figure B.11 Current profiles B007 of NASA are used to discharge all the cycles.

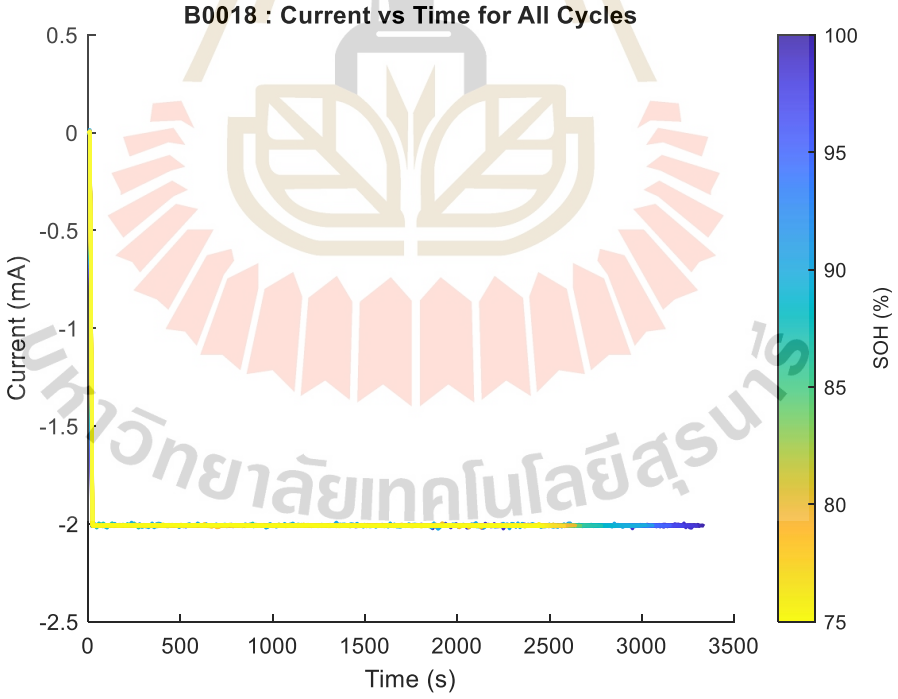


Figure B.12 Current profiles B0018 of NASA are used to discharge all the cycles.

APPENDIX C

ICA Feature Extraction for Oxford Battery Data

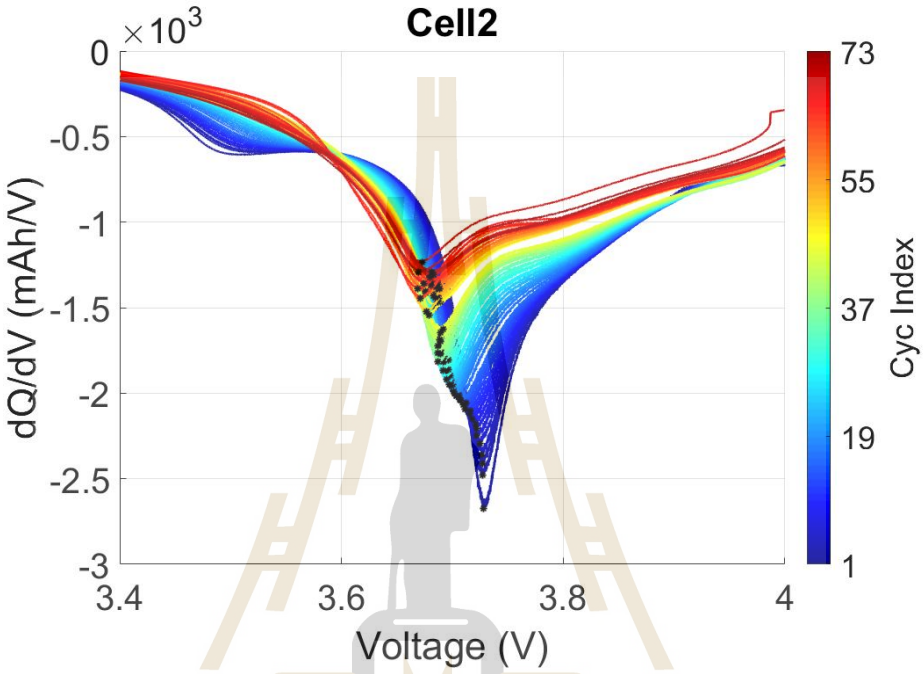


Figure C.1 ICA curves for Oxford Cell 2.

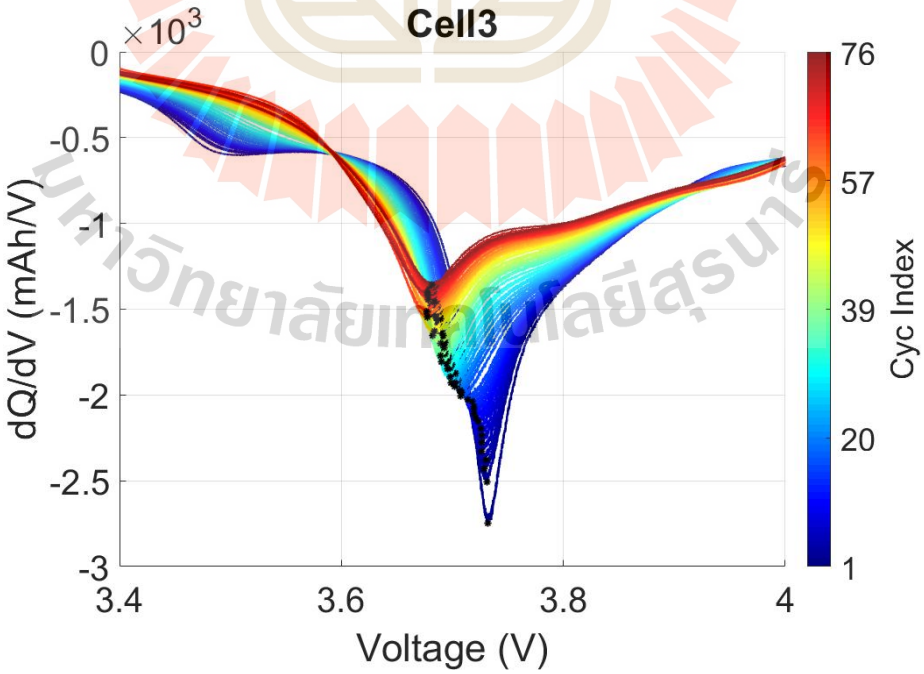


Figure C.2 ICA curves for Oxford Cell 3.

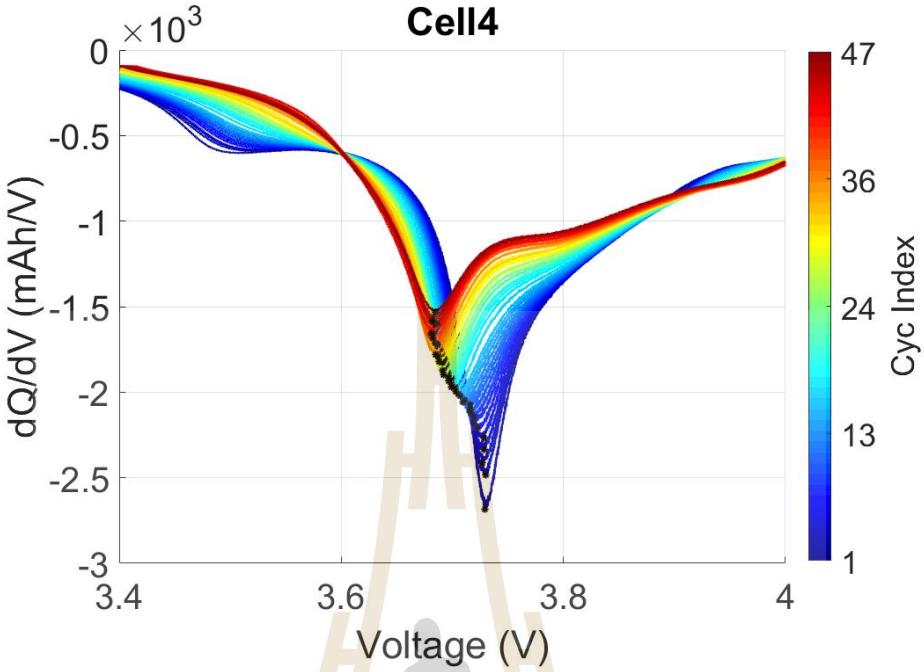


Figure C.3 ICA curves for Oxford Cell 4.

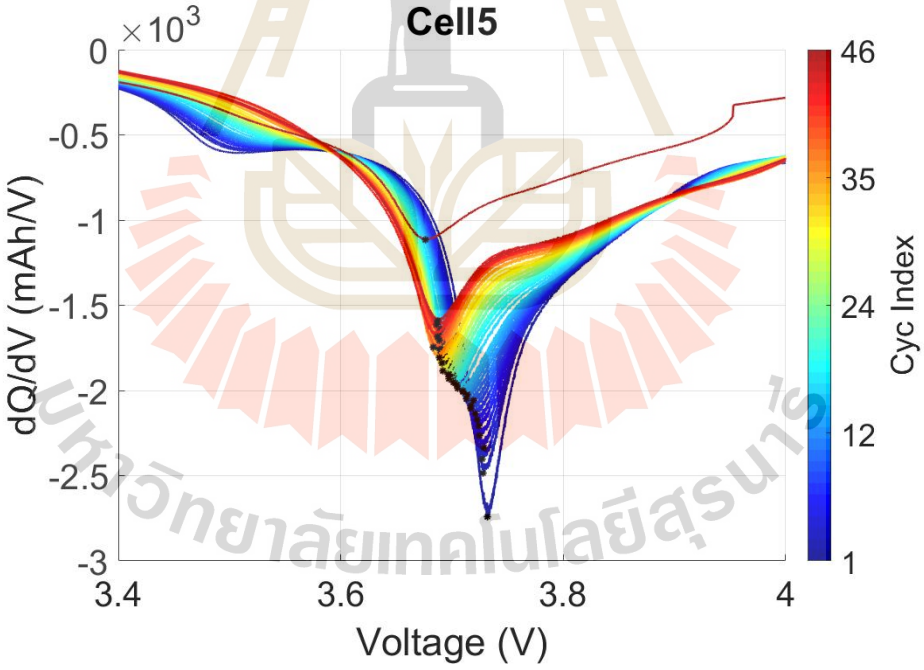


Figure C.4 ICA curves for Oxford Cell 5.

มหาวิทยาลัยเทคโนโลยีสุรนารี

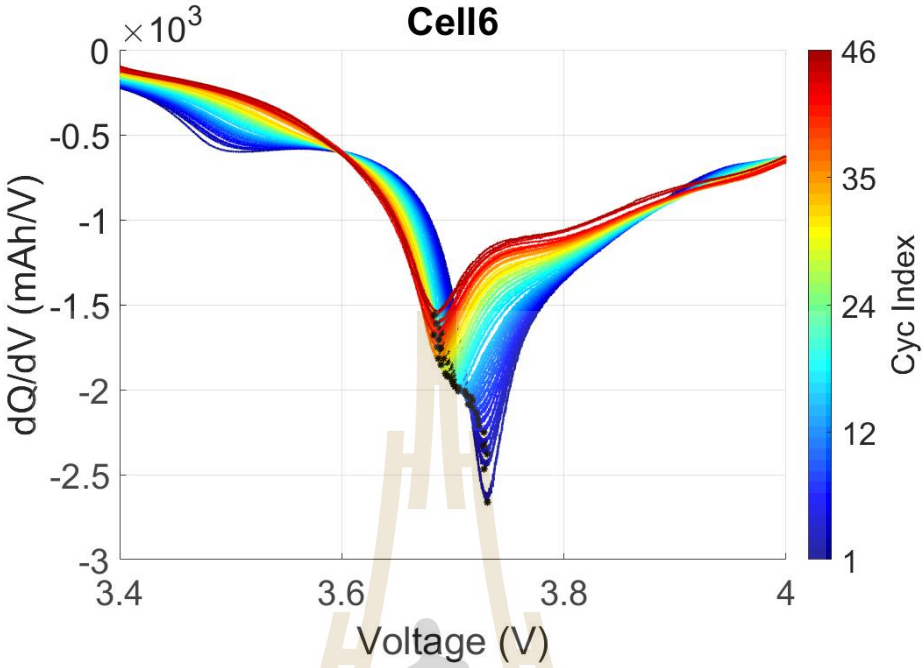


Figure C.5 ICA curves for Oxford Cell 6.

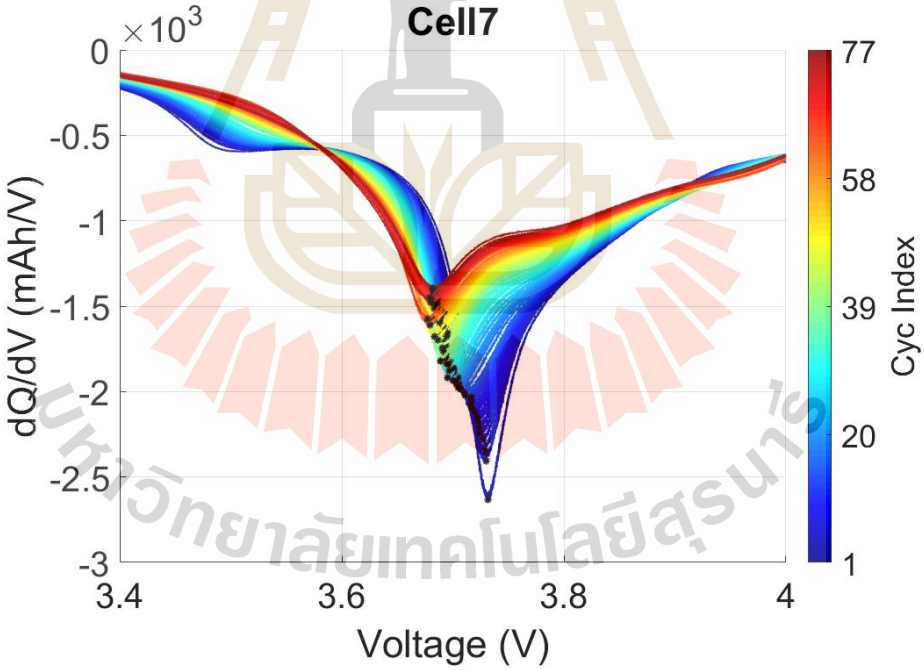


Figure C.6 ICA curves for Oxford Cell 7.

มหาวิทยาลัยเทคโนโลยีสุรนารี

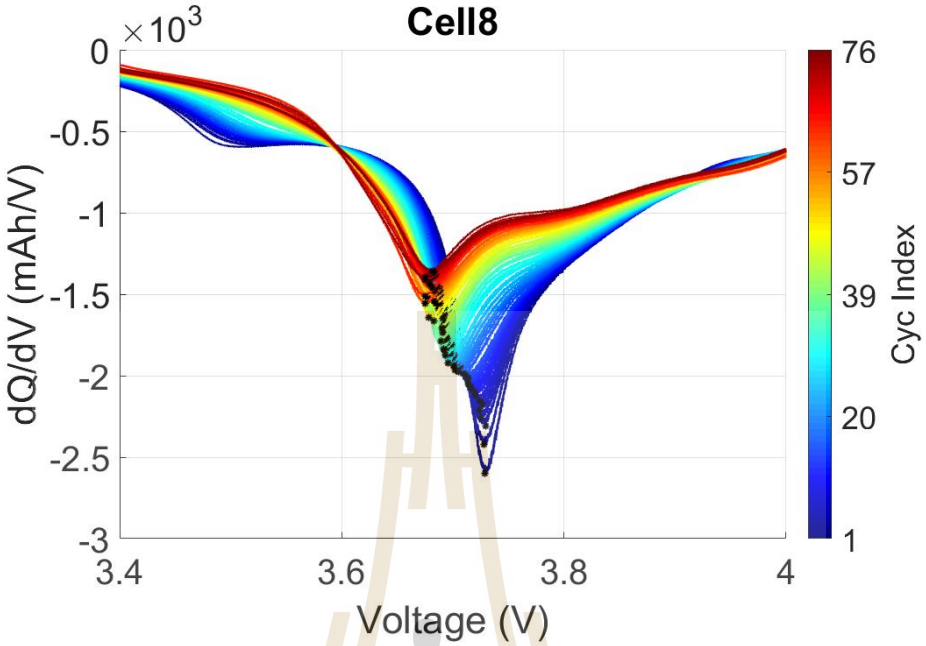


Figure C.7 ICA curves for Oxford Cell 8.

มหาวิทยาลัยเทคโนโลยีสุรนารี

APPENDIX D

ICA Feature Extraction for NASA Battery Data

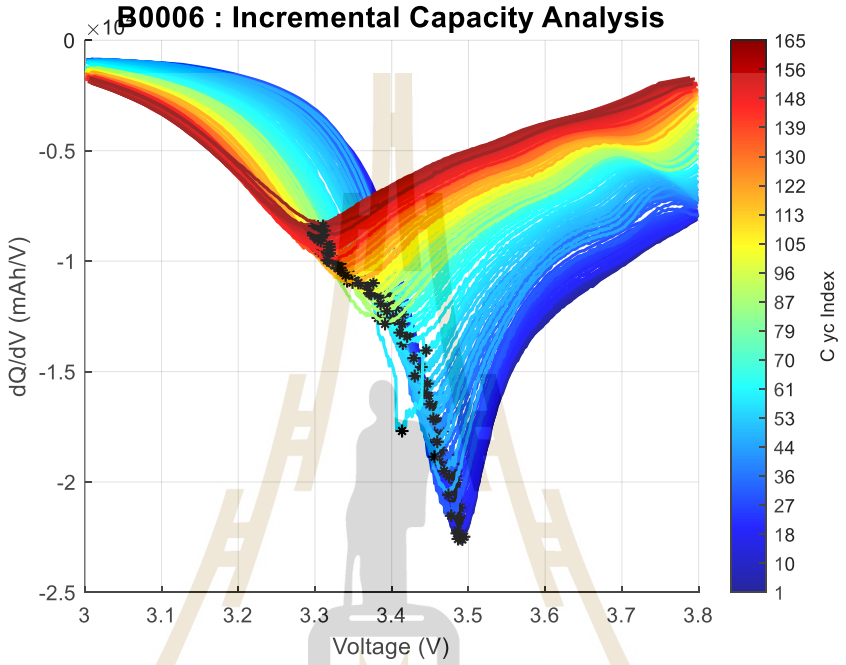


Figure D.1 ICA curves for NASA B0006.

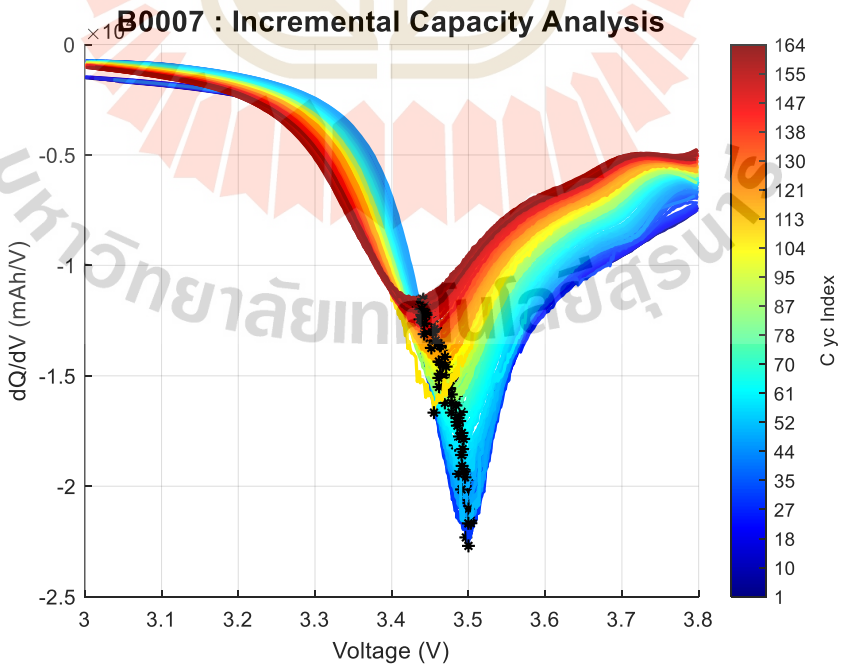


Figure D.2 ICA curves for NASA B0007.

มหาวิทยาลัยเทคโนโลยีสุรนารี

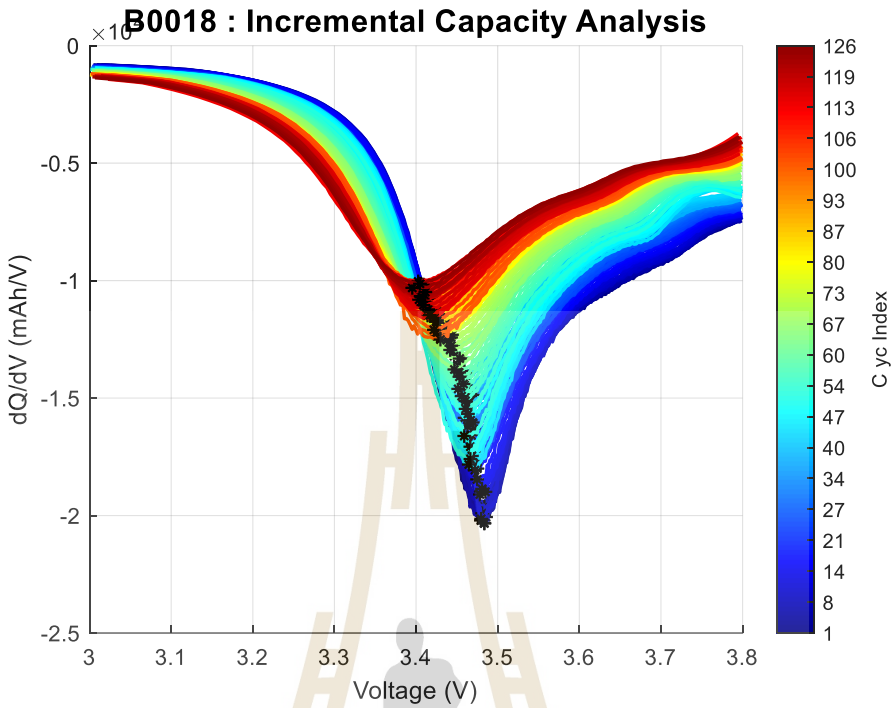


Figure D.5 ICA curves for NASA B0018.

มหาวิทยาลัยเทคโนโลยีสุรนารี

APPENDIX E

List of publications

S. Techanok, N. Junhuathon and K. Chayakulkheeree, (2025). A Comparative Study on Feature Extraction for Battery State of Health Estimation: Evaluating the Role of Open Circuit Voltage in LSTM Model, 2025 13th International Electrical Engineering Congress (iEECON)

S. Techanok, N. Junhuathon and K. Chayakulkheeree, (2025). Lithium-Ion Battery State of Health Estimation Using Resampling-Based Data Simplification Deep Learning Techniques, Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology Association (ECTI).

มหาวิทยาลัยเทคโนโลยีสุรนารี

A Comparative Study on Feature Extraction for Battery State of Health Estimation: Evaluating the Role of Open Circuit Voltage in LSTM Model

Saran Techanok
School of electrical engineering
Institute of Engineering
Suranaree University of Technology
Nakhon Ratchasima, Thailand
saran.techn@gmail.com

Nitikorn Junhuathon
Department of electrical engineering
Faculty of engineering
Rajamangala University of Technology Thanyaburi
Pathum Thani, Thailand
nitikorn_j@rmutt.ac.th

Keerati Chayakulkheeree*
School of electrical engineering
Institute of Engineering
Suranaree University of Technology
Nakhon Ratchasima, Thailand
keerati.ch@sut.ac.th

2025 13th International Electrical Engineering Congress (iEECON) | 979-8-3315-4395-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/iEECON64081.2025.10987628

***Abstract*— This paper investigates the addition of the open circuit voltage (OCV) parameter for state of health (SOH) estimation using a long short-term memory (LSTM) model. The battery degradation data of Oxford University is used in the investigation. The variable used for training the models are charging and discharging currents and voltage, times and temperatures. The comparison shows that the model incorporating OCV data performs better in estimation accuracy than the model without OCV.**

Keywords*— *State of Health, Lithium-ion batteries, Open Circuit Voltage

I. Introduction

Governments worldwide have actively campaigned in recent years to reduce greenhouse gas emissions. To address this issue, there is an emphasis on transitioning to clean and renewable energy. Additionally, the growth of renewable energy has caused the rise in development of energy storage systems (ESSs). As result Lithium-ion batteries (LIBs) are critical components in modern electric technologies. Accurately estimating the SOH of batteries is critical for ensuring system reliability, safety, and performance. LIBs have gained significant popularity in ESSs and electric vehicle (EV) systems . The key challenge with LIBs is their performance degradation over time and repeated usage cycles. The decline in efficiency has resulted in the battery no longer meeting the required energy demands as initially designed Once the battery's condition falls below its original operational threshold, it is recycled or repurposed for second-life applications. To maximize the benefits and value of battery usage, it's crucial to be aware of the battery's lifespan. Accurate tools for estimating battery health are essential to ensure safety and reliability, even as batteries deteriorate. One of the key indicators providing information of a battery cell is the SOH. This reflects its current ability to store and deliver energy compared to its capacity at the beginning of its lifespan.

There are several methods available for estimating the SOH of a battery. Currently, the most popular approach is utilizing machine learning and deep learning techniques. These methods enable accurate estimations of a battery's condition by analyzing historical usage data, such as voltage, current, and temperature. Accurately estimating SOH is critical for enhancing the safety of ESS. Various methodologies have been proposed to enhance models and algorithms for estimating battery SOH. Numerous research has been conducted to evaluate and compare the accuracy of different models in estimating the SOH to achieve the lowest possible error and ensure suitability for practical usage. These studies encompass comparisons between models developed using deep learning techniques [1] and algorithms from various other methods to identify the most accurate and efficient options for estimating battery SOH [2]. Including a review of techniques ranging from traditional methods to the latest machine learning approaches [3], Research has utilized LSTM to estimate battery SOH due to It has strength in handling time series data, Comparisons with recurrent neural networks (RNN) show that LSTM achieves higher accuracy and adaptability, making it more suitable for applications requiring precise, stable Estimations [4], Additional techniques are being employed to extract features from battery data. These enhancements aim to increase the accuracy and flexibility of predicting the battery's SOH [5]. The LSTM model offers high accuracy in forecasting when working with time-sequenced data [6]. The work in [7] utilizes OCV data for incremental capacity analysis (ICA) to assess the battery's SOH. Furthermore, it employs a Bi-LSTM neural network to improve accuracy The findings show that when compared to other models utilized for comparison, the ICA-Bi-LSTM model has greater accuracy. In [8] the OCV is used as an indicator of the SOH for batteries. It is analyzed alongside other parameters, such as voltage and current, to accurately detect modifications to the battery's internal structure. The work in [9] propose a new method for evaluating SOH that integrates the OCV and ICA models. Develop a novel OCV model to extract the OCV curve data from the terminal voltage measured during constant current charging. Finally, in [10] compares the C1 charging graph with the OCV graph to analyze the battery's deterioration characteristics under different charging conditions. After that, the researcher chooses traits with a high correlation with battery degradation to develop an LSTM model, aimed at improving SOH forecasting accuracy.

However, most research on battery SOH estimation focuses primarily on improving model training performance without comparing parameter data prior to model input. Studying the benefits of incorporating OCV data into the model can offer useful data for enhancing model accuracy. Therefore, the objective of this paper is to perform feature extraction on the existing data, add OCV as an addition parameter, and compare the results between models using standard charge-discharge cycle data and models with added OCV data to assess the impact on SOH estimation accuracy, using identical LSTM model settings for fair comparison.

The article is arranged into four sections. Section II represents the datasets and feature selection. The definition of the model in section III. Section IV simulation result. Finally, Section V is the conclusion.

*Corresponding author

979-8-3315-4395-2/25/$31.00 ©2025 IEEE

Authorized licensed use limited to: Suranaree University of Technology provided by UniNet. Downloaded on January 09,2026 at 16:03:34 UTC from IEEE Xplore. Restrictions apply.

II. Datasets and Feature Extraction

A. Battery Model

Among the available open-source datasets, this study utilizes a dataset provided by Oxford University [11]. The testing procedure encompasses both charging and discharging at a rate of 1C, alongside measurements taken at the Pseudo-OCV, which approximates the OCV obtained from voltage measurements under controlled conditions. The battery parameters for the tested system appear in Table I.

TABLE I. Specification of Battery Datasets

Typcs	I (mA)	Cut-off voltage (V)	
		Charging	*Discharging*
1-C cycles	740	2.7/4.2	4.2/2.7
Pseudo-OCV cycles	40	2.7/4.2	4.2/2.7

From Fig. 1, The voltage profile from the discharge cycles of cells 1 and 7 in the dataset demonstrates deterioration mechanisms, evident from the decrease in discharge time. This aligns with the expected trend as the quantity of usage cycles rises. To assess the remaining useful life of the battery for maintenance or replacement purposes, it is necessary to establish an appropriate SOH. The SOH is defined by the remaining capacity (C_{remain}) compared to the initial capacity ($C_{initial}$) [12], as shown in equation 1

$$\%SOH = \frac{C_{remain}}{C_{initial}} \times 100 \quad (1)$$

B. Feature Extraction

In the experimental lithium-ion battery dataset, anomalies such as missing or unrecorded charging or discharging cycles were detected and subsequently removed to ensure data quality and accuracy prior to feeding the model. Feature extraction involved separating charging and discharging data to enhance cycle-specific analysis. This separation also enhances model precision and aids in better understanding the battery's behavior under varying conditions. Furthermore, data from the area under the voltage curve (Voltage Curve Area) was extracted to emphasize differences between cycles more effectively. This feature is crucial for predicting the battery's SOH. Thus, feature extraction is a crucial step in deep learning, as it reduces the amount of data to be processed without losing relevant information, allowing the model to focus on critical data while minimizing noise and redundancy, thereby enhancing estimation accuracy. Meanwhile, normalization using min-max scaling was applied to the charging and discharging time, ensuring that data is on a comparable scale, which contributes to model stability and high performance. A correlation matrix was generated to analyze the relation between variables. This analysis provides insights into how strongly each variable is associated with others. Fig. 2 displays the correlation matrix for the input features, including voltage, current, temperature, the area under the voltage curve, and charging/discharging time, as well as the recently included OCV data. The results demonstrate that the OCV data exhibits a strong positive correlation with the SOH, indicating it enhances the model's capacity for estimation. Additionally, the high correlations

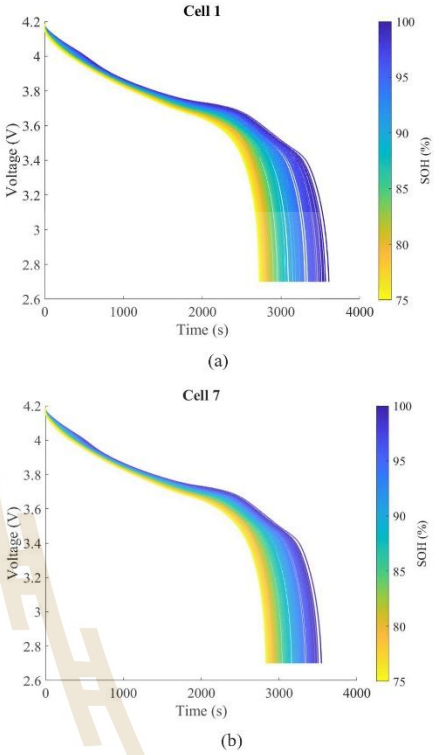


Fig. 1. (a) Voltage profiles of Cell 1 are used to discharge all the cycles. (b) Voltage profiles of Cell 7 are used to discharge all the cycles.

among some features reinforce their importance in capturing the underlying dynamics of the battery system

After feature extraction and normalization, the input data is fed into the LSTM model for training. In this study, two distinct models are employed. The first model uses input data that includes voltage, current, temperature, and the data processed through feature extraction and normalization as previously described, along with other processed data. The second model incorporates additional data from the area under the OCV curve. This additional data has been featuring extraction and processed using the moving mean technique, alongside the same information as in the first model. The second model enables a deeper analysis of subtle variations in charging cycle data. SOH is calculated based on the maximum capacity after a full charge in each cycle. The implementation of the LSTM model enables precise SOH estimation by effectively capturing and analyzing the temporal dynamics present in the dataset. The process of designing estimator will be shown in Fig. 3. Where V is voltage, I is current, T is temperature. The OCV data is highly beneficial for evaluating the SOC, as the OCV directly correlates with the remaining energy within the battery. Analyzing OCV curves can provide insights into battery lifespan by monitoring voltage changes over time and usage. Additionally, OCV data can assist in

Authorized licensed use limited to: Suranaree University of Technology provided by UniNet. Downloaded on January 09,2026 at 16:03:34 UTC from IEEE Xplore. Restrictions apply.

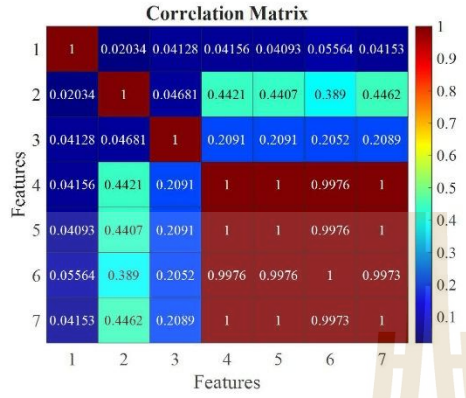


Fig. 2. The correlation matrix.

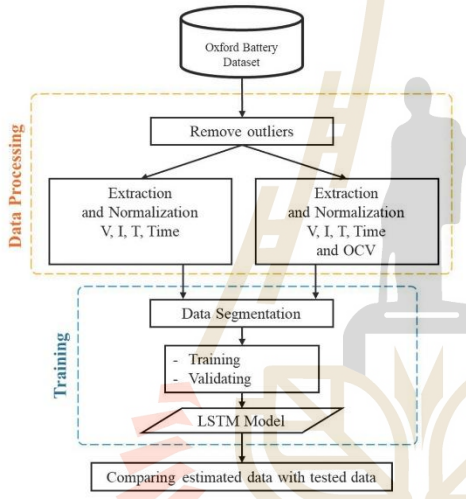


Fig. 3. The process to design the estimator.

identifying abnormalities in battery capacity, such as unexpected degradation, allowing for timely corrective actions. Incorporating OCV data into battery modeling enhances the accuracy of predictions regarding battery behavior under various conditions. OCV information is essential for battery management systems (BMS), aiding in SOC estimation, remaining energy prediction, capacity assessment, charging cut-off, and cell balancing. Figure 4. shows the proposed estimation approach framework.

III. Model Definition

The LSTM network is a type of RNN that is very useful for processing sequential data since it is very good at remembering information across long sequences. Because LSTM can manage long-term dependencies, they are frequently utilized in applications including time-series data, natural language processing, and battery health estimates. The LSTM consists of three types of gates and a memory cell [13], as shown in Fig. 5.

- The forget gate functions to eliminate irrelevant information from the memory cell.
- The input gate regulates the storage of new data in memory by determining which inputs are important.
- The output gate decides which information should be used as output in the current time step.

These gates interact to make it possible for the LSTM to update and remove unnecessary data while preserving important long-term information. The memory cell fixes the vanishing gradient issue that RNNs frequently encounter by ensuring that important information is carried over numerous time steps. Improper adjustments or storing excessive irrelevant data can make LSTMs prone to overfitting. Furthermore, LSTMs may need more training time and computational resources than simpler models, even though they are excellent at capturing long-term dependencies. The configuration of both models was designed to be identical for the training process, as shown in Table II. These settings were carefully tuned to suit the dataset and the LSTM architecture employed for SOH estimation.

TABLE II. 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

The configuration of parameters such as learn rate schedule and learn rate prop factor improves the model's capacity to adjust itself in later training stages. In addition to the settings outlined in Table II, shuffling is performed to rearrange the data every epoch, ensuring randomness and reducing the risk of overfitting. Furthermore, the data is separated into a validation set, which is essential for assessing the model's performance during training on unseen data. The validation data is assessed at regular intervals defined by the Validation Frequency parameter, making certain that the training process is periodically evaluated. All things considered, the aforementioned parameter combinations significantly lower the chance of overfitting, increasing the model's overall efficacy.

The training error for both models was used to calculate the particle velocity and served as the objective metrics for evaluating them. The first comparative index considered was RMSE, following equation 2.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(SOH(i) - SOH_e(i))^2} \quad (2)$$

The second comparative index used is the Mean Square Error (MSE) can be stated as,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(SOH(i) - SOH_e(i)) \quad (3)$$

Authorized licensed use limited to: Suranaree University of Technology provided by UniNet. Downloaded on January 09,2026 at 16:03:34 UTC from IEEE Xplore. Restrictions apply.

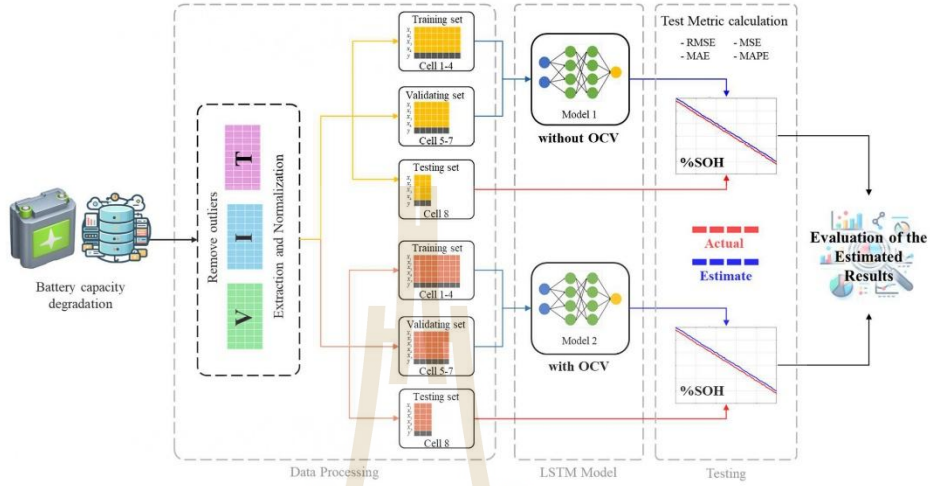


Fig. 4. Framework of the proposed approach for battery SOH estimation.

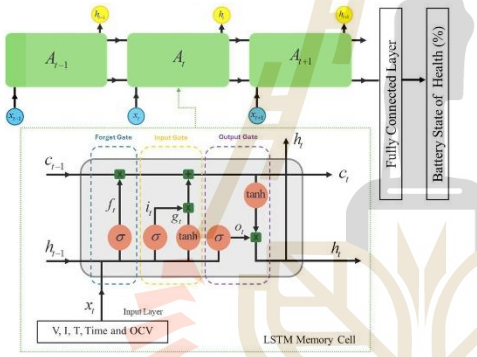


Fig. 5. An example of the LSTM cell for the SOH estimation.

Where x_t is the input to the neuron, h_t means the information of the hidden layer, and c_t is the state of the neuron cell, at the moment t [5].

Mean Absolute Error (MAE) can be shown as,

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|SOH(i) - SOH_e(i)\right| \quad (4)$$

Equation 5 can be used to express the Mean Average Percentage Error (MAPE).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{SOH(i) - SOH_e(i)}{SOH(i)}\right| \times 100\% \quad (5)$$

Where n is a number of cycles, SOH is actual capacity, SOH_e is estimated capacity. Figure 3 shows the suggested estimating approach.

IV. Simulation Result

The Oxford battery degradation dataset 1 consists of cell1, cell2, cell3, and cell4 as training data, cell5, cell6, and cell7 as validation data, and cell8 as test data, ensures data variety for validating the model. Using an LSTM model with identical parameter settings for both models, Three distinct tests were carried out to guarantee the consistency and dependability of the SOH estimation results. Every test was created to confirm the model's functionality in the same circumstances. Fig. 6 illustrates the results of the first SOH estimation test, comparing the outcomes of the two models. Fig. 7 presents the second test results, while Fig. 8 illustrates the third.

The simulation results for the Cell8 battery model highlight the performance differences between the models without and with OCV data across all three test iterations. For Test 1, the model without OCV data records RMSE, MSE, MAE, and MAPE values of 0.4518, 0.2041, 0.404, and 0.4790, respectively. In contrast, the model with OCV data achieves values of 0.1071, 0.0115, 0.0677, and 0.0778, respectively. In Test 2, the model without OCV data shows values of 0.4130, 0.1706, 0.3610, and 0.4208, respectively, while the model with OCV data attains values of 0.1040, 0.0108, 0.0859, and 0.1022, respectively. For Test 3, the model without OCV data produces values of 0.4331, 0.1881, 0.3873, and 0.4537, respectively, whereas the model with OCV data achieves values of 0.1128, 0.0127, 0.0868, and 0.1060, respectively. The average performance metrics across the three tests for the model without OCV data are 0.4326 for RMSE, 0.1876 for MSE, 0.3841 for MAE, and 0.4512 for MAPE. Similarly, for the model with OCV data, the average values are 0.1080 for RMSE, 0.0117 for MSE, 0.0801 for

Authorized licensed use limited to: Suranaree University of Technology provided by UniNet. Downloaded on January 09,2026 at 16:03:34 UTC from IEEE Xplore. Restrictions apply.

MAE, and 0.0953 for MAPE, the average results from the three test runs demonstrate that the four objective metrics indicate the model with OCV achieves more accurate estimations compared to the model without OCV. Table III displays a summary of the estimated outcomes.

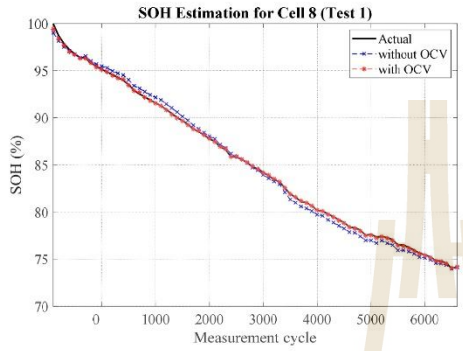


Fig. 6. SOH estimation of Cell 8 (Test 1).

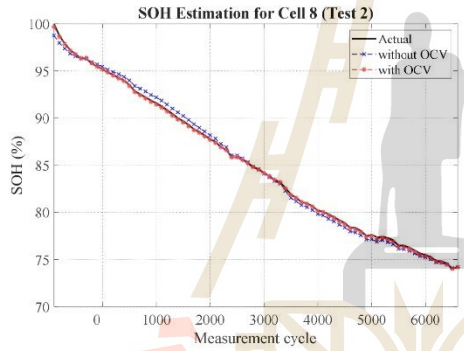


Fig. 7. SOH estimation of Cell 8 (Test 2).

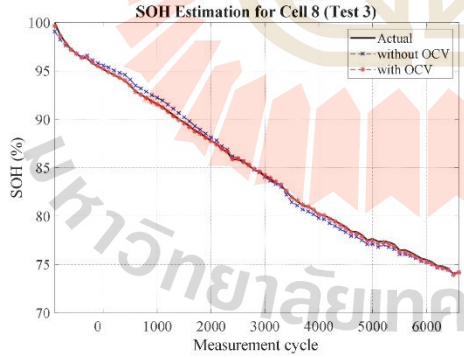


Fig. 8. SOH estimation of Cell 8 (Test 3).

TABLE III. SIMULATION RESULTS

Test	Model	Metrics			
		RMSE	*MSE*	*MAE*	*MAPE*
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

V. Conclusion

This research introduces a method for estimating the SOH of lithium-ion batteries by comparing the inclusion of the OCV parameter in SOH estimation using an LSTM model. The comparison between the model without OCV and the model with OCV shows that the model with OCV performs better. Including OCV helps improve the accuracy of SOH estimation by providing additional information about the battery's state, which enhances the model's ability to predict SOH more precisely. The advantage of OCV is that it has a direct relationship with the SOC and has less noise compared to voltage during charging or discharging, thus capturing battery changes more accurately. However, including OCV in the model adds complexity to the data processing. Collecting accurate and complete OCV data requires more expensive equipment and takes more time to process. Additionally, adding the OCV variable may increase the risk of overfitting, which could affect the model's predictive capability. To manage the risk of overfitting, increasing the diversity of the data used for training the model can be beneficial. This involves incorporating a wider range of operating conditions and battery states to ensure the model generalizes well to different scenarios. In real-world applications, using OCV in the system helps the BMS to more accurately track long-term changes in the battery, reducing the likelihood of premature degradation or damage. This is particularly important for BMS used in EVs and ESS. Therefore, it is crucial to balance the improved accuracy with the increased complexity and cost.

Acknowledgment

This work is fully supported by Suranaree University of Technology.

References

[1] A. Aliberti, F. Boni, A. Perol, M. Zampolli, R. J. P. Jabocuf, P. Tosco, E. Macii, and E. Patti, "Comparative analysis of neural networks techniques for lithium-ion battery SOH estimation," *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 1355-1361, 2022, doi: https://doi.org/10.1109/COMPSAC54236.2022.00214.

[2] L. Qian, L. Xuan, and J. Chen, "Battery SOH estimation based on decision tree and improved support vector machine regression algorithm," *Frontiers in Energy Research*, vol. 11, 2023. doi: https://doi.org/10.3389/fenrg.2023.1218580.

[3] A. Dineva, "Evaluation of Advances in Battery Health Prediction for Electric Vehicles from Traditional Linear Filters to Latest Machine Learning Approaches," *Batteries*, vol. 10, no. 10, p. 356, 2024. doi: https://doi.org/10.3390/batteries10100356.

[4] B. Zou, M. Xiong, H. Wang, W. Ding, P. Jiang, W. Hua, Y. Zhang, L. Zhang, W. Wang, and R. Tan, "A Deep Learning Approach for State-of-Health Estimation of Lithium-Ion Batteries Based on a Multi-Feature and Attention Mechanism Collaboration," *Batteries*, vol. 9, no. 6, p. 329, 2023. doi: https://doi.org/10.3390/batteries9060329.

[5] A. Al-Dulaimi, M. Al-Ani, and M. Al-Maamari, "Auto-encoder LSTM for Li-ion SOH prediction: a comparative study on various benchmark

Authorized licensed use limited to: Suranaree University of Technology provided by UniNet. Downloaded on January 09,2026 at 16:03:34 UTC from IEEE Xplore. Restrictions apply.

datasets," in *Proceedings of the IEEE International Conference on Big Data*, 2021, pp. 123-130. doi: https://doi.org/10.1109/ICMLA52953.2021.00246.

[6] C. Nguyen Van, and D. Ta Quang, "Estimation of SoH and internal resistances of Lithium ion battery based on LSTM network," *International Journal of Electrochemical Science*, vol. 4, p. 100166, 2023. doi: https://doi.org/10.1016/j.ijoes.2023.100166.

[7] H. Sun, J. Sun, K. Zhao, L. Wang, and K. Wang, "Data-Driven ICA-Bi-LSTM-Combined Lithium Battery SOH Estimation," *Mathematical Problems in Engineering*, vol. 2022, p. 9645892, 2022. doi: https://doi.org/10.1155/2022/9645892.

[8] K. Das, and R. Kumar, "Electric vehicle battery capacity degradation and health estimation using machine-learning techniques: a review," *Clean Energy*, vol. 7, no. 6, pp. 1268-1281, 2023. doi: https://doi.org/10.1093/ce/zkad054.

[9] X. Bian, Z. Wei, W. Li, J. Pou, D. U. Sauer and L. Liu, "State-of-Health Estimation of Lithium-Ion Batteries by Fusing an Open Circuit Voltage Model and Incremental Capacity Analysis," in *IEEE Transactions on Power Electronics*, vol. 37, no. 2, pp. 2226-2236, Feb. 2022, doi: https://doi.org/10.1109/TPEL.2021.3104723.

[10] L. Zhang, L. Jiang, J. Zhang, and G. Cai, "Improved LSTM based state of health estimation using random segments of the charging curves for lithium-ion batteries," *Journal of Energy Storage*, vol. 45, pp. 109370, 2023. doi: https://doi.org/10.1016/j.est.2023.109370.

[11] C. R. Birkl, "Oxford Battery Degradation Dataset 1," *University of Oxford*, 2017. [Online]. Available: Oxford Battery Degradation Dataset 1 - ORA - Oxford University Research Archive. [Accessed: Nov. 6, 2024].

[12] N. Junhuathon, G. Sakunphaisal and K. Chayakulkheeree, "Li-ion Battery Aging Estimation Using Particle Swarm Optimization Based Feedforward Neural Network," *2020 International Conference on Power, Energy and Innovations (ICPEI)*, Chiangmai, Thailand, 2020, pp. 73-76, doi: https://doi.org/10.1109/ICPEI49860.2020.9431432.

[13] P. Elefteriadis, M. Gangi, S. Leva, A. V. Rey, E. Groppo, and L. Grande, "Comparative study of machine learning techniques for the state of health estimation of Li-Ion batteries," *Electric Power Systems Research*, vol. 164, pp. 110889, 2024. doi: https://doi.org/10.1016/j.epsr.2024.110889

Biography



Saran Techanok received his B.Eng in EE (First Class Honors) from Suranaree University of Technology (SUT), Thailand, in 2024. He is now enrolled in M.Eng in EE at School of EE, Institute of Engineering, SUT. His research area is on power system and AI applications in power systems.



Nitikorn Junhuathon He received his B. Eng., M. Eng., and Ph.D. In EE from SUT, in 2017, 2019 and 2023, respectively. He is a lecturer at Department of EE, Faculty of Engineering, Raja Mangala University of Technology Thanyaburi (RMUTT), Thailand. His research interests are machine learning and optimization technique for power forecasting, electric vehicle application, and related fields.



Keerati Chayakulkheeree received his B. Eng. in EE from King Mongkut's Institute of Technology Ladkrabang (KMITL) in 1995, and M. Eng. and D. Eng. in Electric Power System Management from the Asian Institute of Technology (AIT), Thailand, in 1999 and 2004, respectively. He is an Associate Professor at the School of EE, Institute of Engineering, SUT. His research area is power system optimization, AI applications in power systems, smart grid technology, power system protection, electricity market restructuring, renewable energy integration, and carbon emissions in the power sector.

Authorized licensed use limited to: Suranaree University of Technology provided by UniNet. Downloaded on January 09,2026 at 16:03:34 UTC from IEEE Xplore. Restrictions apply.

Lithium-Ion Battery State of Health Estimation Using Resampling-Based Data Simplification Deep Learning Techniques

Saran Techanok[1], Nitikorn Junhuathon[2,*] and Keerati Chayakulkheeree[1,*]

[1] School of Electrical Engineering, Institute of Engineering, Suranaree University of Technology
[2] Department of Electrical Engineering, Faculty of engineering, Rajamangala University of Technology Thanyaburi,
Email: saran.techn@gmail.com, nitikorn_j@rmutt.ac.th, keerati.ch@sut.ac.th
* Corresponding Author

ABSTRACT

Lithium-ion batteries (Li-ion) are widely used in various applications due to their high efficiency and reliability. However, as these batteries are continuously used, their performance gradually degrades over time, making accurate estimation of the State of Health (SOH) of increasing significance. One of the key challenges in SOH estimation lies in the nature of the measurement data collected during charge and discharge processes, which is typically time-series data with large volume and complexity. This results in increased computational load and reduced efficiency of the estimation models. To address this issue, this research proposes a data simplification method using a resampling technique aimed at identifying the optimal sampling level that maintains estimation accuracy while reducing computational cost. Four deep learning (DL) models are employed in this work: Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), Attention-based LSTM (ATT-LSTM), and Gated Recurrent Unit (GRU). These models are trained and evaluated using public battery datasets that contain complete charge-discharge cycles. The proposed methods were tested with the Oxford Battery Dataset, and the experimental results demonstrate that the proposed approach achieves higher estimation accuracy while significantly reducing computation time.

Keywords: Lithium-ion battery, State of Health, Deep learning, Resampling

1. INTRODUCTION

Due to the severe growth of climate change caused by increasing greenhouse gas emissions, especially from the energy and transportation sectors, many countries around the world have begun taking action to curb these emissions. One major step is the Paris Agreement, which seeks to reduce the impact of climate change and encourages nations to set long-term carbon-reduction targets, thereby supporting a more environmentally sustainable future. Other significant frameworks, such as COP26 and the European Green Deal, likewise emphasize the transition to clean energy and a reduced reliance on fossil fuels. To meet these goals, renewable energy sources like solar and wind power have expanded rapidly, together with the development of energy storage systems (ESS) that store excess electricity for use during peak-demand periods. At the same time, electric vehicles (EVs) are becoming more popular as alternatives to traditional fuel-powered vehicles. Currently, lithium-ion batteries (Li-ion) are among the most popular battery types due to their high energy density, lightweight construction, and extended lifespan. Nevertheless, as they age, Li-ion batteries gradually degrade through chemical and physical changes, leading to diminished performance that can affect system safety and reliability. Consequently, monitoring and estimating a battery's state of health (SOH) is essential for predicting its remaining useful life, preventing unexpected failures, and enabling predictive maintenance strategies.

In recent years, the challenge of estimating the SOH of batteries has grown significantly, driven by the increasing adoption of renewable energy systems, EVs, and energy storage technologies. Batteries that are reliable and accurately monitored throughout their lifetime are crucial for these systems. As a result, SOH estimation has become an active area of research, especially approaches that learn directly from real-world battery data. Today, physics-based models for SOH estimation are becoming less reliable due to limitations in accuracy and flexibility. This has led to a growing shift toward using machine learning (ML) and artificial intelligence (AI) techniques to build data-driven models instead[1]. One basic model in the family of deep learning (DL) is the Artificial Neural Network (ANN), which has been applied in the estimation of various battery states such as the State of Charge (SOC), SOH, and remaining useful life (RUL) [2]. In [3], a multi-layer Long Short-Term Memory (LSTM) network combined with an attention mechanism was proposed, and then a backpropagation neural network was applied to integrate SOH estimations from different parts of the charging curve. Gaussian Process Regression (GPR) was used to reconstruct missing data, and an LSTM model was trained to estimate SOH from a section of the voltage sequence without relying on complex feature extraction [4]. A hybrid Gated Recurrent Unit (GRU)-CNN model was tested on NASA and Oxford datasets and showed a maximum estimation error of no more than 4.3% [5]. A separate evaluation comparing LSTM, GRU, 1D-CNN, and CNN-LSTM models found that 1D-CNN achieved the highest accuracy with the lowest variance [6].

Autoencoders were combined with LSTM to estimate SOH from voltage, current, and temperature data [7]. Controllable Deep Transfer Learning (CDTL) was used to estimate SOC in both short- and long-term periods during early-stage degradation, using two LSTM modules and a Multiple Domain Adaptation technique to transfer knowledge from batteries with extensive histories to new batteries with limited data [8]. The SOH-KLSTM model, which fuses LSTM with the Kolmogorov-Arnold Network (KAN), improved SOH estimation by capturing both time-based patterns and nonlinear degradation behavior [9]. The MC-CNN-TimesNet model was developed to estimate the SOH and RUL of Li-ion [10]. Apart from improving model architectures, feature extraction is also a key focus in many studies. Research [11] selected battery health features based on the area under constant current charging and discharging voltage curves across four types of Li-ion, showing a strong correlation between these extracted features and SOH. In [12], SOH was estimated using multiple features such as constant voltage discharge time, incremental capacity (IC), and differential thermal voltage (DTV) with a Bidirectional Long Short-Term Memory (Bi-LSTM) model incorporating an attention mechanism and achieved a minimum test RMSE of 0.162%. Moreover, feature extraction techniques that combine Differential Thermal Analysis (DTA) and Incremental Capacity Analysis (ICA) have been used to increase estimation accuracy [13]. The analysis of IC curve peaks combined with other variables has led to better SOH estimation results in [14] and [15]. A fusion-based method that integrates an OCV model and ICA with an ANN was presented in [16]. In [17], Fourier resampling was applied to process the signals, resulting in more uniform and denoised inputs. These processed signals were then used as model inputs, with IC features specifically selected to improve the accuracy of SOH estimation. The study in [18] introduced the MSHHO-DELM framework, which combines IC analysis with feature selection using Pearson correlation and Random Forest. It applies an improved Harris Hawks Optimization algorithm with multi-strategy mechanisms to avoid local optima during training. This method produced a lowest test RMSE of 0.32 on the Oxford dataset. In [19], nine data-driven approaches for Li-ion SOH estimation are compared. The method includes ICA and feature extraction, then uses a variety of machine learning algorithms to predict SOH. Two publicly available datasets are used to assess the models' processing time and estimation error. Results show that ML methods offer faster processing than experimental techniques, making them suitable for accelerating industrial battery testing.

Although research in battery SOH estimation has explored a wide range of model types and hybrid model developments aimed at reducing estimation error, another important perspective that can also improve performance lies in enhancing the quality of input data before training. This includes reducing training time, simplifying the input structure, and minimizing computational complexity, which are crucial factors for enabling real-world deployment of SOH models. The overall framework of the proposed method utilized to address these challenges is illustrated in Fig. 1. Despite their importance, these aspects often remain underexplored in the existing literature. The related studies are summarized in Table 1. The table compares whether each study applied resampling techniques. For those that did, it considers whether the sampling intervals were defined with sufficient granularity to capture broader patterns and trends.

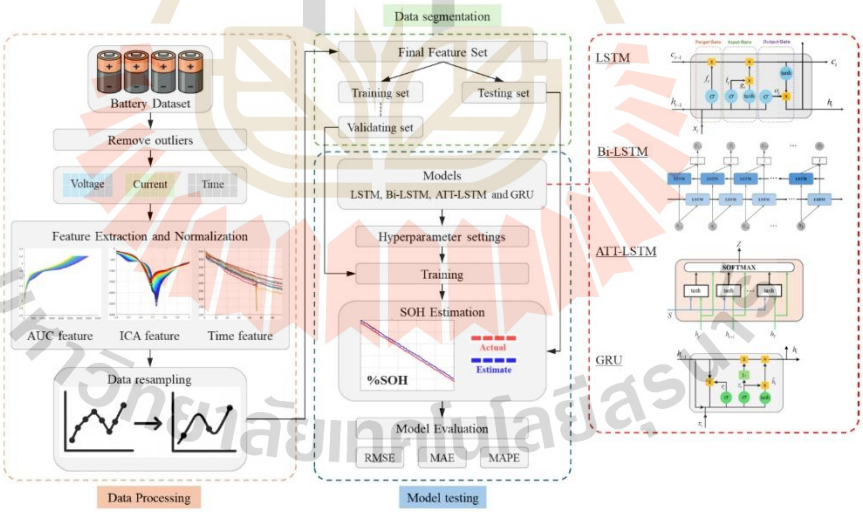


Fig.1: *Overall Framework of the Proposed Battery SOH Estimation*

Table 1: *Review of Related Literature*

Ref.	Year	Resampling	Intervals Investigation	Algorithm	Input features and output
[14]	2020	-	-	Bi-LSTM	Input = [ICA], Output = SOH
[5]	2020	✓	-	GRU-CNN	Input = [Charging curves], Output = SOH
[7]	2021	-	-	Autoencoders- LSTM	Input = [V, I, T], Output = RUL
[20]	2021	✓	3 intervals	CNN, ULSAM, GRU	Input = [V, I, T], Output = SOC
[16]	2022	-	-	ANN	Input = [OCV, ICA], Output = SOH
[6]	2022	-	-	LSTM, GRU, 1D-CNN, CNN-LSTM	Input = [V, I, T], Output = SOH
[8]	2022	-	-	CDTL	Input = [V, I, T, V_{avg}], Output = SOH
[21]	2022	-	-	GPR	Input = [voltage curves], Output = SOH
[22]	2022	-	-	DL	Input/Output = Various (as reviewed)
[12]	2023	-	-	Bi-LSTM-AM	Input = [ICA, DTV], Output = SOH
[15]	2023	-	-	BP, PCR	Input = [ICA, PCA], Output = SOH
[23]	2023	-	-	LSTM, FNN	Input = [V, I, T], Output = SOH, RUL
[24]	2023	-	-	TCN-LSTM	Input = [V, I, T, t], Output = SOH, RUL
[11]	2023	-	-	GPR, LSTM, BP	Input = [voltage curve], Output = SOH
[25]	2023	-	-	LSTM	Input = [Capacity], Output = RUL
[4]	2023	✓	-	GPR, LSTM	Input = [ICA], Output = SOH
[3]	2023	-	-	Ensemble	Input = [Charging curves], Output = SOH
[13]	2024	-	-	GPR	Input = [ICA, DTA], Output = SOH
[17]	2024	✓	-	LSTM, DNN, GRU	Input = [ICA], Output = SOH
[26]	2024	✓	-	CNN-LSTM	Input = [V, T, capacity], Output = SOH
[1]	2024	-	-	ML and AI	Input/Output = Various (as reviewed)
[2]	2024	-	-	ANN, FNN, CNN	Input/Output = Various (as reviewed)
[19]	2024	-	-	9 data-driven techniques	Input = [ICA], Output = SOH
[18]	2025	-	-	MSHHO-DELM	Input = [ICA, V], Output = SOH
[27]	2025	-	-	LSTM	Input = [AUC_v, V, I, T], Output = SOH
[9]	2025	-	-	SOH-KLSTM	Input = [V, I, T, capacity], Output = SOH
[10]	2025	-	-	MC-CNN-TimesNet	Input = [Charging curves], Output = RUL
[28]	2025	-	-	Bagged-GPR	Input = [R, V], Output = SOH
Proposed Models		✓	9 intervals	LSTM, Bi-LSTM, ATT-LSTM, GRU	Input = [AUC, ICA], Output = SOH

It also compares the input and output variables used in each study. This highlights a research gap in previous works regarding the comprehensive selection and analysis of resampling intervals. In summary, the main contributions of this analysis are:

- Applying a resampling technique to reduce the resolution of time-series data, making it more manageable for model training and computation.
- Utilizing input features based on ICA and area under the curve (AUC) extraction to enhance SOH estimation.
- Comparing four DL models: LSTM, Bi-LSTM, Attention-based LSTM (ATT-LSTM), and GRU to identify the most effective model for estimating SOH.

This article is organized into several sections. Section 2 presents the battery datasets used in this study. Section 3 explains the feature extraction process based on a data-driven approach. Section 4 describes the DL models selected for SOH estimation, including the techniques used for data handling. Section 5 presents the comparative performance of all four models across the different resampling intervals. Section 6 concludes the study.

2. BATTERY DATASETS

The Oxford Battery Degradation Dataset 1, containing aging data from eight 740 mAh Kokam SLPB533459H4 Li-ion pouch cells, was employed in this study. All tests were conducted in a thermal chamber maintained at a constant temperature of 40 °C. The cells underwent a constant-current–constant-voltage (CC–CV) charging process, followed by discharging, with a 1C charge/discharge test conducted every 100 drive cycles. The resulting data were used for analysis and visualization.

Each test applied a constant current of 740 mA, with measurements recorded at a sampling rate of one second. The recorded parameters include time (t, seconds), voltage (v, volts), charge (q, mAh), and temperature (T, °C). Figure 2 shows the voltage and current profiles during the charging and discharging cycles. In this study, the data were trimmed to use only the voltage range between 2.7 V and 4.2 V: charging corresponds to the voltage rising from 2.7 V to 4.2 V, while discharging corresponds to the voltage decreasing from 4.2 V down to 2.7 V. These graphs provide an overview of the general battery behavior under the given test conditions. In Fig 3, the SOH of each cell was plotted as a function of the number of charge-discharge cycles. This visualization helps highlight the gradual decline in capacity and gives an overview of the aging characteristics under repeated cycling. The SOH is defined by the remaining capacity (C_{remain}) compared to the initial capacity ($C_{initial}$) [27], as shown in Eq. (1):

$$\%SOH = \frac{C_{remain}}{C_{initial}} \quad (1)$$

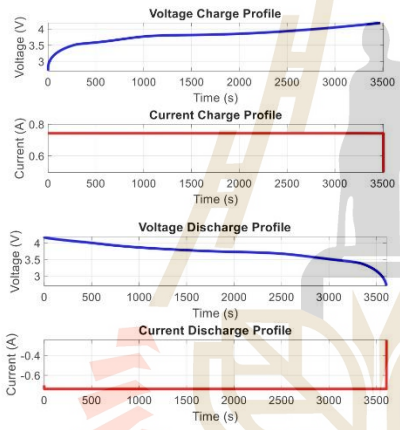


Fig.2: *Voltage and Current Profiles During Charging and Discharging cycles at a 1C rate.*

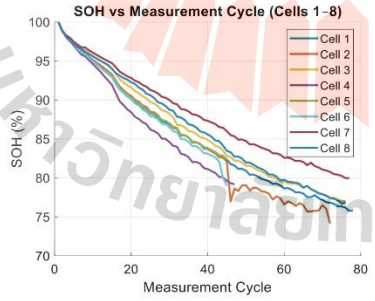


Fig.3: *SOH Degradation Curves of all 8 cells*

3. FEATURE EXTRACTION

Extracting features that represent the battery's deterioration behavior over time is crucial for accurate SOH estimation. These characteristics are usually obtained from unprocessed data, including voltage and current readings, that are captured during charge-discharge cycles. Feature extraction highlights the most pertinent attributes for ML model training while assisting in reducing the dimensionality of the input. However, the raw data may contain recording errors or segments affected by electrical noise in some cells. Therefore, data cleaning and preprocessing are necessary before conducting analysis or extracting features.

3.1 Area under the curve

AUC is a metric used to measure the total magnitude of a variable over a period of time. In this study, the AUC is used to analyze the voltage and current patterns recorded during the battery's charging and discharging cycles. It is utilized as a feature to indicate how the battery uses energy and how this behavior changes over time. Because the AUC aggregates the overall signal instead of focusing on single points, it can give a clearer picture of the battery's behavior, even if there are small fluctuations or noise in the data. This makes it effective for monitoring gradual changes that occur as the battery degrades. The AUC is calculated using the trapezoidal integration method, as shown in the following Eq. (2):

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

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. Figure 4 shows the voltage profiles during the discharging process for all cycles, plotted against time. This allows for a direct comparison of how the voltage curve changes over time as the battery ages. The AUC value represents the total voltage over each cycle, starting from the first cycle through to the later ones. This information can be used to analyze the relationship between battery degradation and the SOH.

From the voltage profiles in Fig. 4, another feature can be extracted: the charging and discharging time duration (Time). This feature represents the accumulated time from the start of the cycle until the voltage (or current) reaches the specified cut-off endpoint. Degradation mechanisms cause the battery voltage to reach the cut-off threshold significantly faster, resulting in a continuous decrease in the total duration. This trend correlates strongly with the reduction in battery capacity.

Consequently, we utilized the time at which the voltage reaches the cut-off endpoint as one of the input variables, as it effectively reflects the battery's health status.

3.2 Incremental Capacity Analysis

ICA is a non-destructive analytical technique used to examine the electrochemical behavior of a battery during

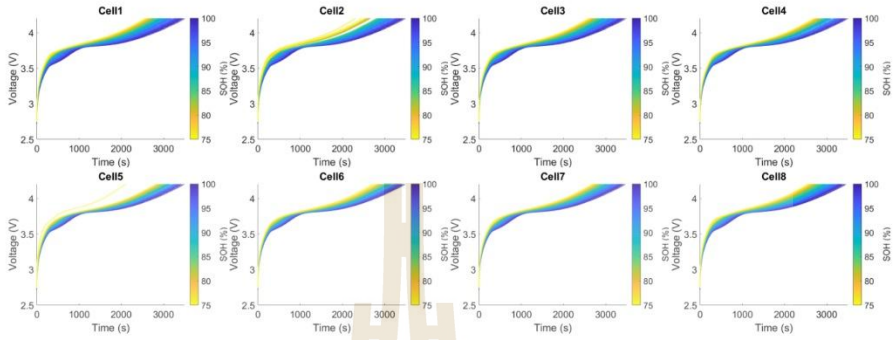


Fig.4: *Voltage Profiles are Used to Charge all the cycles*

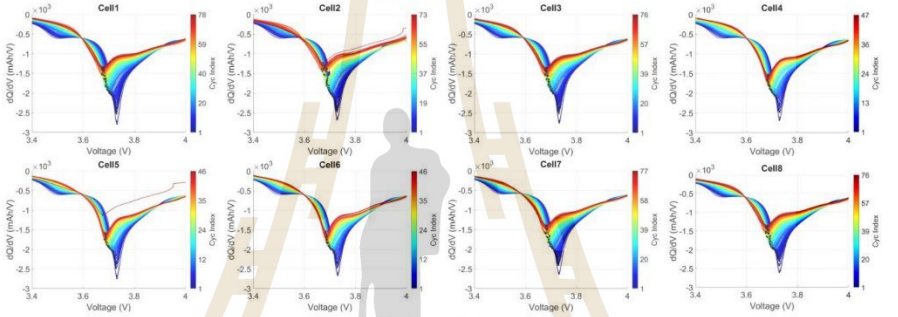


Fig.5: *ICA Curve all the cycles 8 cells*

charging and discharging cycles. This method is based on the derivative of battery capacity (Q) with respect to voltage (V), expressed by Eq. (3):

$$ICA = dQ/dV \quad (3)$$

The resulting dQ/dV curve which translates raw voltage data into a representation of electrochemical activity, reveals distinct features that evolve as the battery ages. Specifically, the minimum point of the dQ/dV curve for each cycle serves as a key indicator of degradation. This feature is highly sensitive to internal changes, such as the loss of active material and lithium inventory. For this reason, ICA is widely regarded as an effective approach for generating reliable features to estimate the battery's SOH. However, a practical challenge arises from the fact that the raw dQ/dV curves often exhibit significant noise, which can interfere with accurate feature extraction. Therefore, to improve data quality, a Moving Average (MA) filter was applied for data smoothing. Figure 5 presents the resulting ICA curves after this filtering process. In this study, a filter window size of 400 was selected. This value was empirically chosen to effectively reduce noise while preserving the essential shape of the curve. This smoothing process is a crucial pre-processing step that ensures that the subsequent identification of key degradation features is both more reliable and robust.

Following the feature extraction process, all extracted features AUC, Charging and Discharging Time Duration, and ICA are normalized using the Min-Max Normalization method to rescale all feature values to a uniform range between 0 and 1. Preventing features with wider numerical ranges from controlling the model's learning process is the main goal of this normalization. Every feature will contribute more equally to the final estimation thanks to this standardization.

Furthermore, to quantitatively validate the relevance of the selected features, a correlation analysis was performed. The results are presented as a correlation matrix in Fig. 6.

This matrix illustrates the strength and direction of the linear relationship between each feature and the battery's SOH. A high correlation coefficient, whether positive or negative, indicates a strong relationship, confirming that the extracted features are indeed sensitive and reliable indicators of the battery's health degradation.

3.3 Data Resampling

Data resampling is a process of adjusting the structure or resolution of time-series data, with the goal of making

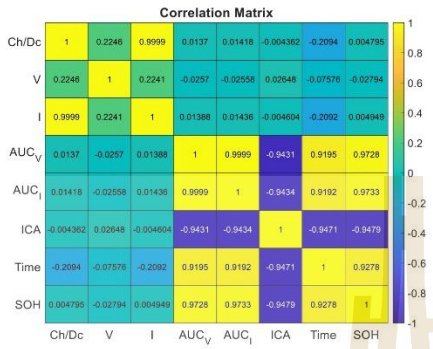


Fig.6: *The Correlation Matrix of all Parameters*

the data more suitable for analysis or model training. In this study, the focus is placed on reducing the frequency of input data. The main objective is to simplify the data in order to reduce training time and computational cost, while still maintaining an acceptable level of accuracy in the estimation of battery SOH.

For the experiments, thirteen resampling intervals were considered: no resampling (None), and down-sampling by skipping data every 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 300, and 400 seconds. This design allows an evaluation of how reducing the amount of input data affects the model's performance, and whether prediction accuracy can still be preserved effectively. Ultimately, this experiment represents an important step in identifying the optimal balance between data volume and model performance.

3.4 Min-Max Normalization

Min-Max Normalization is a method of scaling data into a predefined range from 0 to 1, which helps address the issue of varying magnitudes across different features. This process is important for ensuring that input data remain proportionate and balanced, making them more suitable for model training. The general formula is defined in Eq. (4):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4)$$

Where x represents the original data value x_{min} and are x_{max} the minimum and maximum values of the data, respectively, and x' is the normalized value. In this study, Min-Max Normalization was applied to the selected features, including ICA, AUC, and Time. By scaling all features into the same range, the model can better capture the underlying patterns of the data and improve the effectiveness of SOH estimation.

4. DEEP LEARNING MODELS

As previously discussed, various DL models are employed for battery SOH estimation. In this study, the resampled data are input into LSTM, Bi-LSTM, ATT-LSTM, and GRU models to compare their estimation performance. The following subsections provide a brief description of each model.

4.1 Long Short-Term Memory

LSTM is a type of recurrent neural network (RNN) developed to address the problem of learning long-term dependencies in time-series data. Three main gates complement the memory cell that serves as the foundation of LSTM. The input gate regulates how new data is added to the memory cell, the output gate chooses what data is output at each time step, and the forget gate chooses which data to discard. The memory cell is capable of storing important information for extended periods of time, as shown in Fig. 7. These mechanisms enable the model to retain important information, learn patterns in sequential data, and adapt to complex temporal dependencies effectively, additional information about LSTM can be found in [26].

4.2 Bidirectional Long Short-Term Memory

Bi-LSTM builds upon the conventional LSTM by simultaneously processing input sequences in both forward and backward directions, enabling the network to draw on contextual information from earlier and later time steps at once. This dual-directional design strengthens the model's ability to learn intricate temporal relationships. Although it maintains the memory cells and gate operations used in LSTM, its distinctive feature is merging the outputs from both directions to form a richer representation. Further information about Bi-LSTM is provided in [12], and the architecture of the model is illustrated in Fig. 8.

4.3 Attention-based LSTM

An attention mechanism is included into the memory cell of the ATT-LSTM, an enhanced variant of the conventional LSTM. As a result, rather than considering every time step identically, the model can concentrate on the most crucial portions of the input sequence. The attention mechanism creates a context vector by combining the hidden states of each time step with different weights according to their importance, as illustrated in Fig. 9. While keeping the memory cells and gate structure of a regular LSTM, ATT-LSTM emphasizes key time steps, helping the model capture complex relationships in the sequence and retain the most relevant information effectively. A detailed explanation of ATT-LSTM is provided in [12].

4.4 Gated Recurrent Unit

The GRU is a simplified version of recurrent neural networks aimed at capturing long-term dependencies in sequential data. In contrast to LSTM, it features a more straightforward structure with just two primary gates: the update gate, which controls the merging of new inputs with existing information, and the reset gate, which decides how much of the prior context should be abandoned. Unlike the LSTM, the GRU does not use a separate memory cell but

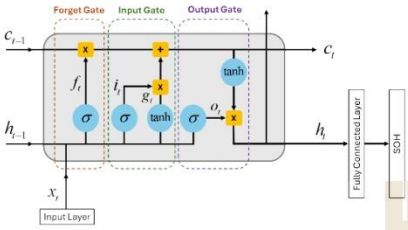


Fig.7: The LSTM Network Architecture

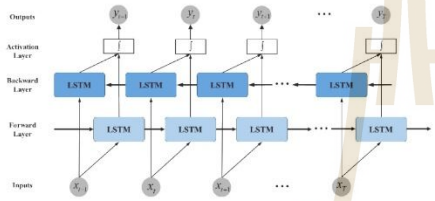


Fig.8: The Bi-LSTM Network Architecture

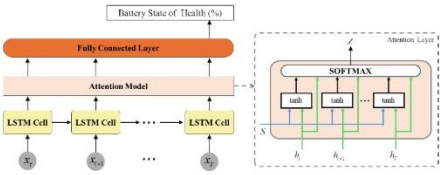


Fig.9: The ATT-LSTM Network Architecture

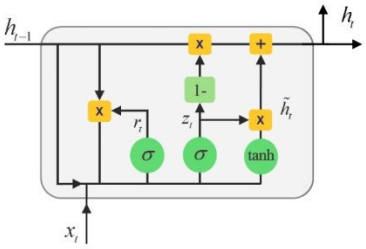


Fig.10: The GRU Network Architecture

instead incorporates the memory mechanism directly into its hidden state, resulting in fewer parameters, faster training, and an effective ability to model sequential patterns. For more detailed information on the GRU, readers are referred to [17]. The architecture of the GRU is presented in Fig. 10.

In this study, four models LSTM, Bi-LSTM, ATT-LSTM, and GRU were used for comparison. To ensure fairness and reliability of the results, the same hyperparameter settings were applied to all models. The selected values were chosen to balance accuracy and computational time. Using identical hyperparameters also allows the impact of resampling to be evaluated more clearly without being influenced by differences in model configurations. This approach ensures that the learning behavior and performance of each model can be assessed under the same conditions. The detailed hyperparameter settings are presented in Table 2.

Table 2: *Hyperparameter Settings*

Parameter	Value
Optimizer	adam
Max Epochs	100
Mini-Batch Size	1024
Initial Learning Rate	0.01
Learning Rate Schedule	piecewise
Learning Rate Drop Factor	0.5
Learning Rate Drop Period	40
Shuffle	every-epoch
Validation Frequency	500

Table 2 presents the hyperparameter settings that are applied consistently across all models. The only difference lies in the Output Mode. The ATT-LSTM model is configured to use the sequence mode, since the Attention Layer requires the complete sequence of outputs. In contrast, the LSTM, Bi-LSTM and GRU models use the last mode, which considers only the final output of the sequence for prediction.

4.5 Evaluation Metrics

To evaluate the performance of battery SOH estimation, three common metrics were selected: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The first, RMSE, is frequently used to display, in terms of root mean square, the average difference between estimated and actual values. which is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(SOH(i) - SOH_e(i))^2} \quad (5)$$

Where n is a number of cycles, SOH is actual capacity, SOH_e is estimated capacity. Next, MAE focuses on the overall error using absolute differences, which helps reduce the effect of extreme mistakes defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|SOH(i) - SOH_e(i)| \quad (6)$$

Finally, MAPE makes it simple to compare accuracy across several scales by expressing the error as a percentage. MAPE is defined as:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{SOH(i) - SOH_e(i)}{SOH(i)}\right| \times 100\% \quad (7)$$

5. COMPARATIVE ANALYSIS

In this section, a comparative analysis is presented by evaluating the effect of different resampling intervals. Thirteen cases are considered: None, 20, 40, 60, 80, 100,

120, 140, 160, 180, 200, 300, and 400 seconds, as described earlier. The objective is to identify the most suitable interval for accurate SOH estimation of Li-ion batteries. In addition, the performance of four DL models, LSTM, Bi-LSTM, ATT-LSTM, and GRU, is compared to determine which model can extract features most effectively from the data derived from AUC, ICA, and Time. In this study, to ensure that the data sets are non-overlapping and to reduce bias in the evaluation, the Oxford battery data was divided into three main sets. Cells 1–6 were used for training the models, allowing them to learn the degradation patterns of the batteries. Cell 7 was assigned as the validation set to fine-tune model parameters and prevent overfitting. Cell 8 was set aside for testing, offering an objective evaluation of the model's capacity to calculate SOH from unknown data. By using this data partitioning technique, the outcomes will more precisely represent how well the model generalizes to new batteries.

The experimental results show that applying resampling initially leads to a steady reduction in error, indicating that selecting an appropriate resampling interval helps the models learn and capture the data relationships more effectively. This improves the accuracy of SOH estimation. However, when the resampling interval becomes too large, specifically greater than 100-second, the error starts to increase again. This trend is observed starting from 120-second, where the error gradually rises as the gap widens up to 400-second. This suggests that as the data becomes too coarse, the models lose important details, making it more difficult to capture meaningful patterns. This trend is observed across all four models. The best performance is achieved at the 100-second interval, where the MAPE for LSTM is 0.279, while the error values for Bi-LSTM, ATT-LSTM which is the best among all models, and GRU are 0.290, 0.271, and 0.336 respectively. These results are illustrated in Fig. 11, which shows the SOH estimation of all four models at their best-performing interval. On the other hand, the worst results are observed at the 400-second interval for LSTM, Bi-LSTM, and ATT-LSTM, with error values of 1.156, 0.893, and 0.971 respectively, while the worst case for GRU occurs when no resampling is applied, with an error value of 1.564. These results are shown in Fig. 12, which presents the SOH estimation of the four models under the least favorable conditions. In addition, bar charts are provided to compare the error values at the best interval, which are shown in Fig. 13, and at the worst interval, which are shown in Fig. 14 across all models, highlighting the performance differences more clearly. A complete summary of these results is provided in Table 3. In addition to evaluating SOH estimation using the three previously mentioned

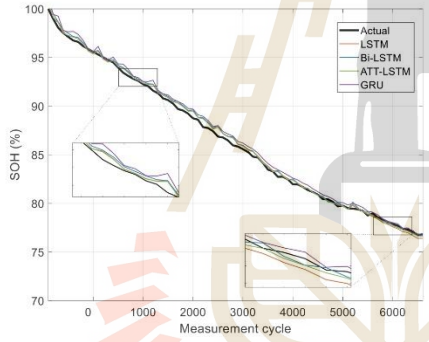


Fig.11: SOH Estimation Results at the Best Interval

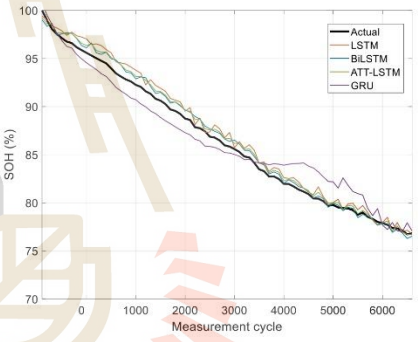


Fig.12: SOH Estimation Results at the Worst Interval

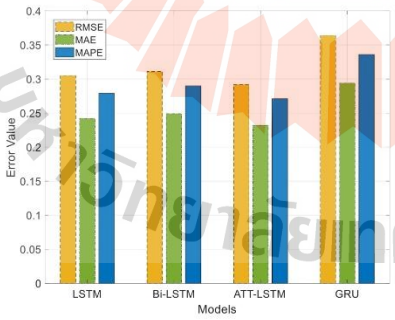


Fig.13: Error Comparison at the Best Interval

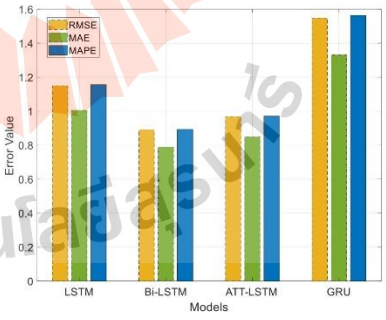


Fig.14: Error Comparison at the Worst Interval

Table 3: *SOH Estimation Results*

Resampling Interval	LSTM			Bi-LSTM			ATT-LSTM			GRU		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
None	0.922	0.722	0.810	0.631	0.523	0.596	0.705	0.585	0.669	1.548	1.331	1.564
20 s	0.760	0.609	0.684	0.816	0.667	0.747	0.781	0.592	0.658	0.676	0.400	0.445
40 s	0.483	0.364	0.420	0.558	0.427	0.478	0.545	0.452	0.534	0.457	0.362	0.433
60 s	0.406	0.332	0.385	0.456	0.382	0.450	0.362	0.286	0.339	0.413	0.328	0.373
80 s	0.388	0.340	0.299	0.387	0.311	0.367	0.327	0.256	0.294	0.407	0.317	0.363
100 s	0.305	0.242	0.279	0.311	0.249	0.290	0.292	0.232	0.271	0.364	0.294	0.336
120 s	0.324	0.257	0.296	0.332	0.254	0.289	0.311	0.239	0.271	0.377	0.308	0.350
140 s	0.383	0.314	0.360	0.333	0.263	0.299	0.340	0.270	0.305	0.396	0.325	0.369
160 s	0.401	0.325	0.368	0.334	0.258	0.292	0.357	0.278	0.313	0.432	0.349	0.396
180 s	0.470	0.389	0.440	0.347	0.272	0.308	0.349	0.278	0.316	0.481	0.394	0.447
200 s	0.527	0.446	0.503	0.359	0.287	0.326	0.363	0.279	0.313	0.520	0.448	0.508
300 s	0.783	0.694	0.795	0.621	0.527	0.591	0.621	0.527	0.593	0.747	0.669	0.760
400 s	1.149	1.006	1.156	0.892	0.790	0.893	0.968	0.849	0.971	1.145	0.999	1.152

metrics, the models were also assessed based on Processing Time, which refers to the time each model requires to train and generate the SOH estimation results. Analyzing processing time provides insight into not only the accuracy of each model but also which models are faster or slower in computation. The experimental results presented in Table 4 compare the processing times of each model across different resampling intervals, showing that processing time generally decreases as the resampling interval increases. When comparing the models at their best-performing resampling interval of 100 seconds, the LSTM model demonstrated the fastest performance, requiring only 84.09 seconds. It was followed closely by the GRU at 85.65 seconds, while the Bi-LSTM and ATT-LSTM models required longer processing times of 115.34 and 169.12 seconds, respectively. This indicates that while all models benefit from resampling, the standard LSTM architecture offers the greatest computational speed among the evaluated models.

Table 4: *Execution Time of Each Model*

Resampling Interval	Processing Time (seconds)			
	LSTM	Bi-LSTM	ATT-LSTM	GRU
None	14,056	17,071	15,891	13,403
20 s	447.44	556.63	848.22	403.20
40 s	217.60	261.42	382.18	205.79
60 s	160.95	184.59	264.02	139.29
80 s	105.93	139.20	228.70	103.12
100 s	84.09	115.34	169.12	85.65
120 s	70.88	97.72	131.52	72.78
140 s	58.68	82.53	106.82	59.77
160 s	51.83	66.61	93.07	51.50
180 s	48.37	64.98	96.74	51.05
200 s	44.85	57.15	84.37	45.14
300 s	30.09	40.14	57.09	30.80
400 s	23.57	28.96	41.05	23.43

Regarding overall performance, the ATT-LSTM model achieved the highest accuracy with a MAPE of 0.271, yet it consumed the most processing time (169.12 seconds). In contrast, the LSTM model required only 84.09 seconds, approximately half the duration of the ATT-LSTM, while achieving a very similar accuracy with a MAPE of 0.279. When comparing the LSTM model's performance between the raw data (None) and the 100-second resampling interval, the estimation error decreased by 65.56%, while the processing time was significantly reduced by 99.40%. Therefore, considering the trade-off between high estimation accuracy and computational efficiency, the LSTM is identified as the most suitable model for real-world applications where resource constraints are a concern.

6. CONCLUSIONS

This research focused on the estimation of the SOH of Li-ion by investigating the impact of resampling on the performance of DL models. Several models, including LSTM, Bi-LSTM, GRU, and ATT-LSTM, were evaluated under different resampling intervals to analyze both estimation accuracy and processing time. The results showed that selecting an appropriate resampling interval can reduce data complexity and suppress noise, which improves the learning efficiency of the models. This leads to lower estimation errors and significantly shorter processing times. However, when the resampling interval becomes too large, the models may lose important details in the data, making it harder to capture the relationships needed for accurate SOH estimation.

In practice, resampling offers potential for real-world applications such as battery health monitoring in ESS and EVs. By reducing the amount of data while maintaining accuracy, it allows faster processing and more efficient use of computational resources. This makes it possible to apply these models in systems that require reliable and near real-time SOH estimation.

7. ACKNOWLEDGEMENT

This work is fully supported by Suranaree University of Technology.

8. REFERENCES

[1] A. Dineva, "Evaluation of Advances in Battery Health Prediction for Electric Vehicles from

Traditional Linear Filters to Latest Machine Learning Approaches," *Batteries*, vol. 10, no. 10, p. 356, Oct. 2024.

[2] A. G. Olabi, A. A. Abdelghafar, B. Soudan, A. H. Alami, C. Semeraro, M. Al Radi, M. Al-Murisi, and M. A. Abdelkareem, "Artificial Neural Network Driven Prognosis and Estimation of Lithium-Ion Battery States: Current Insights and Future Perspectives," *Ain Shams Engineering Journal*, vol. 15, no. 2, p. 102429, Feb. 2024.

[3] L. Zhang, J. Zhang, T. Gao, L. Lyu, L. Wang, W. Shi, L. Jiang, and G. Cai, "Improved LSTM Based State of Health Estimation Using Random Segments of the Charging Curves for Lithium-Ion Batteries," *Journal of Energy Storage*, vol. 74, p. 109370, Dec. 2023.

[4] X. Xiong, Y. Wang, K. Li, and Z. Chen, "State of Health Estimation for Lithium-Ion Batteries Using Gaussian Process Regression-Based Data Reconstruction Method During Random Charging Process," *Journal of Energy Storage*, vol. 72, p. 108390, Nov. 2023.

[5] Y. Fan, F. Xiao, C. Li, G. Yang, and X. Tang, "A Novel Deep Learning Framework for State of Health Estimation of Lithium-Ion Battery," *Journal of Energy Storage*, vol. 32, p. 101741, Dec. 2020.

[6] A. Aliberti, F. Boni, A. Perol, M. Zampolli, R. J. P. Jaboeuf, P. Tosco, E. Macii, and E. Patti, "Comparative Analysis of Neural Networks Techniques for Lithium-Ion Battery SOH Estimation," in *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, Los Alamitos, CA, USA, 2022, pp. 1355-1361.

[7] P. Audin, I. Jorge, T. Mesbahi, A. Samet, F. D. B. D. Beuvron, and R. Boné, "Auto-Encoder LSTM for Li-Ion SOH Prediction: A Comparative Study on Various Benchmark Datasets," in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Pasadena, CA, USA, 2021, pp. 1529-1536.

[8] I. Oyewole, A. Chehade, and Y. Kim, "A Controllable Deep Transfer Learning Network with Multiple Domain Adaptation for Battery State-of-Charge Estimation," *Applied Energy*, vol. 312, p. 118726, May 2022.

[9] I. Jarraya, S. Ben Atitallah, F. Alahmed, M. Abdelkader, M. Driss, F. Abdelhadi, and A. Koubaa, "SOH-KLSTM: A Hybrid Kolmogorov-Arnold Network and LSTM Model for Enhanced Lithium-Ion Battery Health Monitoring," *Journal of Energy Storage*, vol. 122, p. 116541, Jun. 2025.

[10] Y. Li, X. Qin, M. Chai, H. Wu, F. Zhang, F. Jiang, and C. Wen, "SOH Evaluation and RUL Estimation of Lithium-Ion Batteries Based on MC-CNN-TimesNet Model," *Reliability Engineering & System Safety*, vol. 261, p. 111125, Sep. 2025.

[11] J. Wu, Z. Liu, Y. Zhang, D. Lei, B. Zhang, and W. Cao, "Data-Driven State of Health Estimation for Lithium-Ion Battery Based on Voltage Variation Curves," *Journal of Energy Storage*, vol. 73, p. 109191, Dec. 2023.

[12] B. Zou, M. Xiong, H. Wang, W. Ding, P. Jiang, W. Hua, Y. Zhang, L. Zhang, W. Wang, and R. Tan, "A Deep Learning Approach for State-of-Health Estimation of Lithium-Ion Batteries Based on a Multi-Feature and Attention Mechanism Collaboration," *Batteries*, vol. 9, no. 6, p. 329, Jun. 2023.

[13] D. Stephens, Q. Zhou, H. Hofmann, and J. Sun, "Integrated Incremental Capacity Analysis and Differential Thermal Analysis for Improved Robustness in Li-Ion Battery State of Health Estimation," in *2024 IEEE Conference on Control Technology and Applications (CCTA)*, Newcastle upon Tyne, UK, 2024, pp. 729-734.

[14] H. Sun, J. Sun, K. Zhao, L. Wang, and K. Wang, "Data-Driven ICA-Bi-LSTM-Combined Lithium Battery SOH Estimation," *Mathematical Problems in Engineering*, vol. 2022, p. 9645892, Mar. 2022.

[15] Z. Zou, Y. Li, X. Liu, Z. Zhang, B. Duan, and C. Zhang, "Fast Estimation of State of Health for Retired Lithium-Ion Batteries Based on Principal Component Regression," in *2023 7th CAA International Conference on Vehicular Control and Intelligence (CVCI)*, Changsha, China, 2023, pp. 1-6.

[16] X. Bian, Z. Wei, W. Li, J. Pou, D. U. Sauer, and L. Liu, "State-of-Health Estimation of Lithium-Ion Batteries by Fusing an Open Circuit Voltage Model and Incremental Capacity Analysis," *IEEE Transactions on Power Electronics*, vol. 37, no. 2, pp. 2226-2236, Feb. 2022.

[17] K. D. Rao, N. V. Anand, T. K. S. Pandraju, F. Alsaif, and T. S. Ustun, "Optimally Tuned Gated Recurrent Unit Neural Network-Based State of Health Estimation Scheme for Lithium Ion Batteries," *IEEE Access*, vol. 12, pp. 58597-58607, 2024.

[18] Z. Wei, Y. Li, X. Sun, W. Liu, C. Liu, and H. Lu, "SOH Estimation of Lithium-Ion Batteries Based on Multi-Feature Extraction and Improved DLEM," *Journal of Energy Storage*, vol. 120, p. 116460, May 2025.

[19] P. Eleftheriadis, M. Gangi, S. Leva, A. V. Rey, E. Groppo, and L. Grande, "Comparative Study of Machine Learning Techniques for the State of Health Estimation of Li-Ion Batteries," *Electric Power Systems Research*, vol. 235, p. 110889, Oct. 2024.

[20] Q. Gong, P. Wang, Z. Cheng, and J. Zhang, "A Method for Estimating State of Charge of Lithium-Ion Batteries Based on Deep Learning," *Journal of The Electrochemical Society*, vol. 168, no. 11, p. 110532, Nov. 2021.

[21] Z. Wang, G. Feng, X. Sun, D. Zhen, F. Gu, and A. D. Ball, "Feature Extraction from Charging Profiles for State of Health Estimation of Lithium-Ion Battery," *Journal of Physics: Conference Series*, vol. 2184, no. 1, p. 012024, Jan. 2022.

[22] M. S. Hossain Lipu, S. Ansari, M. S. Miah, S. T. Meraj, K. Hasan, A. S. M. Shihavuddin, M. A. Hannan, K. M. Muttaqi, and A. Hussain, "Deep Learning Enabled State of Charge, State of Health and Remaining Useful Life Estimation for Smart Battery

Management System: Methods, Implementations, Issues and Prospects," *Journal of Energy Storage*, vol. 55, p. 105752, Nov. 2022.

[23] C. Nguyen Van and D. T. Quang, "Estimation of SOH and Internal Resistances of Lithium Ion Battery Based on LSTM Network," *International Journal of Electrochemical Science*, vol. 18, no. 6, p. 100166, Jun. 2023.

[24] C. Li, X. Han, Q. Zhang, M. Li, Z. Rao, W. Liao, X. Liu, X. Liu, and G. Li, "State-of-Health and Remaining-Useful-Life Estimations of Lithium-Ion Battery Based on Temporal Convolutional Network-Long Short-Term Memory," *Journal of Energy Storage*, vol. 74, p. 109498, Dec. 2023.

[25] L. Zhang, T. Ji, S. Yu, and G. Liu, "Accurate Prediction Approach of SOH for Lithium-Ion Batteries Based on LSTM Method," *Energy Reports*, vol. 9, pp. 3206-3214, Nov. 2023.

[26] V. Safavi, N. Bazmohammadi, J. C. Vasquez, and J. M. Guerrero, "Battery State-of-Health Estimation: A Step Towards Battery Digital Twins," *Electronics*, vol. 13, no. 2, p. 384, Jan. 2024.

[27] S. Techanok, N. Junhuathon, and K. Chayakulkheeree, "A Comparative Study on Feature Extraction for Battery State of Health Estimation: Evaluating the Role of Open Circuit Voltage in LSTM Model," in *2025 13th International Electrical Engineering Congress (iEECON)*, Phuket, Thailand, 2025, pp. 1-6.

[28] W. Ma, S. Ha, G. Pozzato, A. Saatchi, N. Aliahmad, K. Littau, and S. Onori, "Engineering Testing Protocols for Machine Learning-Based SOH Estimation in Lithium Metal Batteries," *Journal of The Electrochemical Society*, vol. 171, no. 12, p. 120531, Dec. 2024.

Biography



Saran Techanok received his B.Eng in EE (First Class Honors) from Suranaree University of Technology (SUT), Thailand, in 2024. He is now enrolled in M.Eng in EE at School of EE, Institute of Engineering, SUT. His research area is on power system and AI applications in power systems.



Nitikorn Junhuathon He received his B. Eng., M. Eng., and Ph.D. In EE from SUT, in 2017, 2019 and 2023, respectively. He is a lecturer at Department of EE, Faculty of Engineering, Raja Mangala University of Technology Thanyaburi (RMUTT), Thailand. His research interests are machine learning and optimization technique for power forecasting, electric vehicle application, and related fields.



Keerati Chayakulkheeree received his B. Eng. in EE from King Mongkut's Institute of Technology Ladkrabang (KMITL) in 1995, and M. Eng. and D. Eng. in Electric Power System Management from the Asian Institute of Technology (AIT), Thailand, in 1999 and 2004, respectively. He is an Associate Professor at the School of EE, Institute of Engineering, SUT. His research area is power system optimization, AI applications in power systems, smart grid technology, power system protection, electricity market restructuring, renewable energy integration, and carbon emissions in the power sector.

มหาวิทยาลัยเทคโนโลยีสุรนารี

BIOGRAPHY

Mr. Saran Techanok was born on August 1, 2002, in Chonburi, Thailand. He received his Bachelor of Engineering degree in Electrical Engineering (First Class Honors) from Suranaree University of Technology (SUT) in 2023. In the same year, he began his graduate studies in the Master of Engineering Program in Electrical Engineering at the Graduate School of Engineering, Suranaree University of Technology. His academic interests include power systems, electric vehicle (EV) integration, and the application of Artificial Intelligence (AI) in power systems.

During his graduate studies, he received the Kittibandit Scholarship and served as a teaching assistant for courses related to power systems and load flow analysis. This role enhanced his skills in academic coordination, communication, and instructional support. He also presented his research at the International Electrical Engineering Congress (iEECON 2025), gaining valuable experience in professional academic communication.

He is passionate about contributing to the advancement of smart grid technology and sustainable energy infrastructure in Thailand. He aspires to become an expert in AI applications for power systems or work directly in the field of EV batteries to support national development goals.

มหาวิทยาลัยเทคโนโลยีสุรนารี