2025

A Model for estimating the state of health of retired lithium-ion EV batteries based on machine learning.

Rugami, Vanessa
School of Computing and Engineering Sciences
Strathmore University

A Model for Estimating the State of Health of Retired Lithium-ion EV Batteries Based on Machine Learning

Vanessa Rugami
098646

A dissertation submitted in partial fulfillment of the requirements of the Degree of Master of Science in Sustainable Energy Transition at Strathmore University

School of Computing and Engineering Sciences
Strathmore University
Nairobi, Kenya

June, 2025

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the dissertation contains no material previously published or written by another person except where due reference is made in the dissertation itself.

Student's Name: Vanessa Rugami

Sign: _____ Date: ____30/03/2025_____

Approval

The dissertation of Vanessa Rugami was reviewed and approved (for examination) by the following:

Sign: _____ Date: 4/4/2025

Eng. Dr. Julius Butime, School of Computing & Engineering Sciences,
Strathmore University

Dr. Julius Butime,
Dean, School of Computing & Engineering Sciences,
Strathmore University

Prof. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

The electric vehicle market is growing rapidly and with it comes subsequent growth in the number of lithium-ion batteries that reach end of life in electric vehicle applications. Instead of being discarded in landfills, these batteries can be used in other applications such as energy storage since they still retain about 70% to 80% of their original capacity. This is known as battery repurposing, and it helps to manage battery waste.

To repurpose batteries, their state of health must be tested to determine if they are adequately safe and reliable to use in second life applications. Current testing methods are time-consuming. Long testing times inhibit the scalability of repurposing operations to match the rapidly increasing number of electric vehicles, hence retired electric vehicle batteries.

In this study, a machine learning model was developed to determine the SOH of used batteries. The model was based on quantum particle swarm optimization-support vector regression (QPSO-SVR) and used partial discharge data from differential capacity curves to estimate SOH. It was trained on data obtained from cycling used battery cells. The model achieved best MAE of 0.6139, RMSE of 0.7875, and R^2 of 0.8481.

Table of Contents

Declaration ii
Abstract iii
List of Figures vi
List of Tables vii
List of Abbreviations viii
Definition of Terms ix
Acknowledgement x
Dedication xi
Chapter One: Introduction 1
1.1 Background of the Study 1
1.2 Problem Statement 2
1.3 Research Objectives 3
1.3.1 General Objective 3
1.3.2 Specific Objectives 3
1.4 Research Questions 3
1.5 Justification 3
1.6 Scope 4
1.7 Limitations 4
Chapter Two: Literature Review 5
2.1 Introduction 5
2.2 Theoretical Review 5
2.2.1 Lithium-ion Battery 5
2.2.2 Lithium-ion Battery Cell Dimensions 7
2.2.3 Battery Testing 10
2.2.4 SOH Determination Methods 11
2.3 Empirical Review 12
2.4 Study Gaps 14
Chapter Three: Methodology 15
3.1 Introduction 15
3.2 Research Design 15
3.3 Data Collection 15
3.3.1 Battery Cycling 15
3.3.2 Feature Extraction 17
3.4 Model Development 22

3.4.1 QPSO-SVR Model 22
3.4.2 Model Training 26
3.4.3 Model Validation 26
3.5 Method Flow Diagram 27
Chapter Four: Results 28
4.1 Introduction 28
4.2 Battery Cycling 28
4.2 Feature Extraction 33
4.3 Model Results 38
Chapter Five: Discussion 40
5.1 Introduction 40
5.2. Battery Cycling 40
5.3 Feature Extraction 40
5.4 Model Evaluation 41
Chapter Six: Conclusion 42
6.1 Introduction 42
6.2 Conclusion 42
6.3 Future Research 42
References 43
Appendices 50
Appendix A: Similarity Report 50
Appendix B: Ethical Review Approval 52
Appendix C: NACOSTI License 53
Appendix D: Codes 54

List of Figures

Figure 2. 1: Lithium-ion Cell (UL Research Institutes, 2021) 6
Figure 2. 2: Coin Cells (Idelah, 2020) 8
Figure 2. 3: Prismatic Cell (Idelah, 2020) 8
Figure 2. 4: Pouch Cell (Idelah, 2020) 9
Figure 2. 5: Cylindrical Cells (Idelah, 2020) 9

Figure 3. 1: Neware CT-4008T-5V6A-S1 Cell Tester 16
Figure 3. 2: QPSO-SVR Flow Chart 25
Figure 3. 3: Method Flow Diagram 27

Figure 4. 1: Sample Cycling Record 28
Figure 4. 2: Capacity Changes Down Cycles 31
Figure 4. 3: Energy Changes Down Cycles 32
Figure 4. 4: SOH Changes Down Cycles 32
Figure 4. 5: DCIR Changes Down Cycles 33
Figure 4. 6: Differential Capacity Curves for Cell 1 34
Figure 4. 7: Smoothed Differential Capacity Curves for Cell 1 34
Figure 4. 8: Portion of Curve Extracted from Cell 1, Cycle 1 35
Figure 4. 9: Feature Correlation Heatmap 37
Figure 4. 10: QPSO-SVR and SVR Prediction Comparison 39

List of Tables

Table 4. 1: Battery Parameters Extracted 28
Table 4. 2: Initial and Final Values of Battery Parameters 31
Table 4. 3: ANOVA Results 36
Table 4. 4: Correlation Coefficients of Selected Features 37
Table 4. 5: QPSO-SVR and SVR Comparison 38
Table 4. 6: Model Performance with Different Training Sets 39

List of Abbreviations

ANOVA	Analysis of Variance
BMS	Battery Management System
CC	Constant Current
CV	Constant Voltage
CC-CV	Constant Current Constant Voltage
DCIR	Direct Current Internal Resistance
EV	Electric Vehicle
GHG	Green House Gas
LFP	Lithium Iron Phosphate
LMO	Lithium Manganese Oxide
MAE	Mean Absolute Error
NCA	Nickel Cobalt Aluminum Oxide
NDCs	Nationally Determined Contributions
NMC	Nickel Manganese Cobalt Oxide
QPSO	Quantum Particle Swarm Optimisation
QPSO-SVR	Quantum Particle Swarm Optimisation – Support Vector Regression
RMSE	Root Mean Square Error
SDG	Sustainable Development Goal
SOH	State of Health
SVM	Support Vector Machines
SVR	Support Vector Regression

Definition of Terms

Battery Capacity	The maximum amount of energy a battery can store (Wei, 2023).
State of Health	The difference between the capacity of a battery when it was new and its current capacity (Dunn, 2022).
End of Life	The point below which the capacity of a battery is no longer useful for its desired application (Martin, 2016).
Cut-off Voltage	The voltage at which discharging is terminated to prevent over discharge (Large Power, 2022).
C-rate	The rate at which a battery is charged or discharged relative to its rated capacity. At a discharging c-rate of 0.5C or C/2, the discharge current applied discharges the battery in 2 hours. At 1C, the battery will be discharged in 1 hour. The same applies for charging c-rate. Therefore, a 100Ah battery discharging at 1C rate, will provide 100A for 1 hour. At 0.5C, it will provide 50A for 2 hours (Bollini, 2022).

Acknowledgement

The accomplishment of this work would not have been possible without the support of my supervisor, Dr. Eng. Julius Butime. Thank you for your guidance and for encouraging me to take up this study. I would also like to acknowledge Dr. Eng. Fenwicks Musonye and my classmate, Thomas Bundi, for sharing their knowledge with me. I have learnt many things from you. To all I say, thank you and God bless you.

Dedication

I dedicate this work to God for His grace and guidance, and for being my source of strength throughout this journey. To my father for supporting me in my education and encouraging me to reach greater heights. To my mother for her constant prayers and advice. To my sister whose encouragement and belief in me pushed me to persevere even in the most challenging moments. This work is a testament to your love, support, and faith in me.

Chapter One: Introduction

1.1 Background of the Study

The electric vehicle (EV) market is expanding as the world transitions to clean energy. As the demand for electric vehicles rises, so does the number of retired EV batteries. Xu et al. (2022) reports that the amount of retired EV batteries is expected to increase to more than one million tons by 2025. This poses an environmental challenge due to increased battery waste (Chung, 2021). Such batteries, however, are still suitable for other applications such as energy storage and fast charging (Balaraman, 2023) since they still retain about 70% to 80% of their original capacity (Dunn, 2022).

Giving used EV batteries a new application as opposed to discarding them is known as battery repurposing. Repurposing prolongs the life of a battery and reduces the need to mine for metals like nickel, lithium, and cadmium which are used to manufacture batteries (Gaia, 2024). This contributes to environmental conservation as the mining processes of these metals are very energy intensive and environmentally destructive. Moreover, repurposing reduces battery waste which is hazardous due to the toxic chemicals contained in batteries that can seep into the ground and contaminate soil and water sources (Gregory, 2024).

EV batteries usually come in packs with modules that contain individual battery cells (Automotive Cells Company, 2022). Most of the time, whole battery packs or modules cannot be used straight out of EVs due to the likelihood of some cells being degraded. Degraded cells within a battery pack can significantly reduce the efficiency and safety of the battery, compromising its performance in the systems it is employed. Hence, it is necessary for end-of-life batteries to be disassembled and individual cells to be evaluated to ensure that they are suitable and safe for use (Mohan et al., 2023).

Battery capacity decreases over time due to degradation of the electrolyte and electrodes (Etengoff, 2023). State of Health (SOH) is a paramount indicator of degradation. It refers to the capacity of a battery in present-time compared to when it was new. SOH is typically tested by subjecting battery cells to multiple charge/discharge cycles. However, this process is

lengthy (Dunn, 2022) and could take several hours depending on the charge/discharge rates used (Battery University, 2021).

As the number of EVs increases, so does the number of batteries that reach their end of life (EOL) and are at risk of being dumped in landfills. Long SOH testing times for retired EV batteries make it difficult to scale repurposing operations to keep up with the growing number of used EV batteries which is expected to rise significantly over the next few years (Ahmeid et al., 2022). Therefore, it is necessary to develop reliable techniques for determining the SOH of retired EV batteries. In this study, a machine learning model was developed for SOH determination.

1.2 Problem Statement

The growing adoption of EVs has led to an increase in the number of retired EV batteries, which can cause environmental degradation if not properly managed. These batteries can be repurposed since they still retain up to about 70% to 80% of their original capacity. While repurposing offers a sustainable solution, not all cells in retired battery packs are suitable for reuse due to degradation. It is therefore necessary for battery cells to be tested before repurposing to determine their SOH hence suitability in second life applications.

Current SOH determination methods are time-consuming hence not scalable to the expected significant rise in end-of-life EV batteries. This inhibits repurposing efforts, and hence battery waste management. Moreover, while research has been undertaken on battery SOH monitoring, most works focus on SOH estimation of EV batteries that are still in use. Others require batteries to undergo full cycling tests which are lengthy. This work aims to fill these gaps by developing a model for estimating the SOH of retired EV batteries using partial discharge information and QPSO-SVR.

1.3 Research Objectives

1.3.1 General Objective

To develop a method of determining SOH using machine learning

1.3.2 Specific Objectives

1. To analyse the characteristics of used batteries in the market
2. To develop a model that uses partial discharge information for determining the SOH of used batteries
3. To train the model for prediction of SOH of used batteries
4. To validate the model

1.4 Research Questions

1. What are the characteristics of used batteries in the market?
2. How can partial discharge information be used to develop a model that can determine the SOH of used batteries?
3. How can the model be trained to effectively determine SOH of used batteries?
4. How effective is the developed model?

1.5 Justification

The development of a model to determine the SOH of retired EV batteries in Kenya aligns with both global and national sustainability goals. For example, Kenya's updated Nationally Determined Contributions (NDCs) include a commitment to reduce the country's greenhouse gas (GHG) emissions by 32% relative to its business-as-usual scenario by 2030 (Kenya, 2022). In line with this, the country aims to increase its share of renewables on the national grid and adopt a low carbon transportation system. Adoption of EVs contributes to the decarbonisation of the transport sector, and efficient SOH determination facilitates repurposing of used EV batteries which can be used for energy storage in renewable energy systems. Additionally,

repurposing extends the useful life of batteries thus reducing battery waste and the demand for new batteries, which helps to lower the carbon footprint associated with battery disposal and production (Farmer & Watkins, 2023).

The model would also aid in advancing Sustainable Development Goal (SDG) 7 which emphasizes universal access to affordable, reliable, sustainable, and modern energy. Part of the targets of this goal is to increase the share of renewable energy in the global energy mix (Trinh & Chung, 2023). Accurate SOH estimation allows for repurposing of retired EV batteries in energy storage applications for renewable energy sources, which facilitates grid integration. Furthermore, using retired batteries instead of new ones reduces energy storage costs since new batteries are more expensive. This enhances energy affordability (Akram & Abdul-Kader, 2024).

1.6 Scope

This work focused on developing a model for determining the SOH of retired lithium-ion EV batteries in Kenya. The SOH was determined for individual cells and not entire battery modules or packs. Partial discharge information from differential capacity curves was used to determine battery capacity hence SOH. The model was based on QPSO-SVR.

1.7 Limitations

The model does not consider the effects of ambient conditions on the SOH. In real sense however, conditions like relative humidity and ambient temperature might have effects on the SOH of the batteries.

Chapter Two: Literature Review

2.1 Introduction

This chapter discusses the theoretical and empirical concepts related to the topic under study. The theoretical concepts are discussed in the theoretical review section which is divided into four subsections. The subsections dwell on the basic structure and operation of lithium-ion batteries, lithium-ion battery cell dimensions, battery testing, and SOH determination methods. The empirical concepts are presented in the empirical review section. Here, past studies have been critically reviewed.

2.2 Theoretical Review

2.2.1 Lithium-ion Battery

A lithium-ion battery is a rechargeable battery in which lithium ions move between electrodes during charging and discharging. Its major parts are the anode, cathode, separator, positive current collector, negative current collector, and electrolyte. The anode stores lithium ions and releases them to the cathode during discharging, while the cathode stores lithium ions and releases them to the anode during charging. The separator prevents electrons from flowing within the battery and allows lithium ions to move from the anode to the cathode and vice versa. The positive current collector receives electrons from the negative current collector during discharging, while the negative current collector receives electrons from the positive current collector during charging. The function of the electrolyte is to transport lithium ions between the anode and cathode (Korthauer, 2018). The set up for the lithium-ion battery is presented in Figure 2.1.

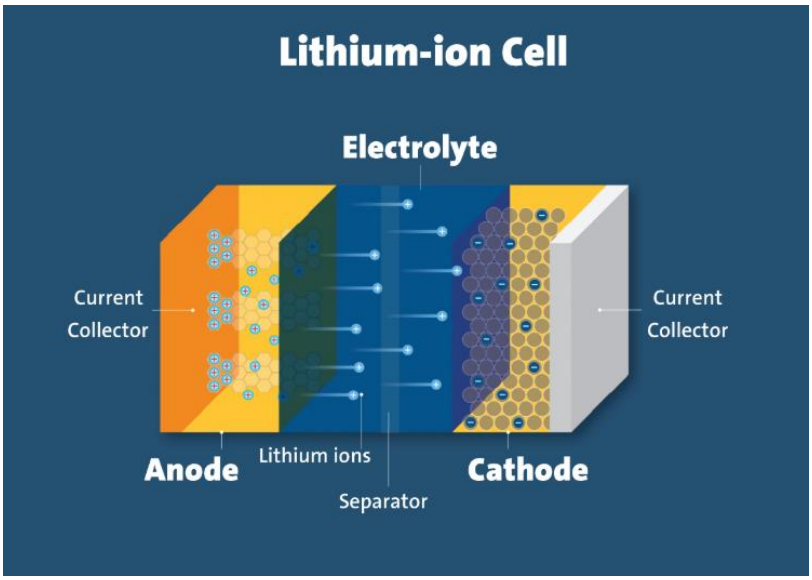


Figure 2. 1: Lithium-ion Cell (UL Research Institutes, 2021)

During charging, lithium ions move from the cathode to the anode through the separator. At the same time, electrons move from the positive current collector to the negative current collector in the external circuit. During discharging, lithium ions move from the anode to the cathode through the separator while electrons move from the negative current collector to the positive current collector in the external circuit (UL Research Institutes, 2021).

Compared to other batteries, for example, lead acid batteries, lithium-ion batteries typically have a longer cycle life and higher energy and power density. Higher energy density enables lithium-ion batteries to store more energy per unit weight, making them lighter than other batteries of the same capacity. Higher power density, on the other hand, enables them to deliver energy faster. A longer cycle life means they can undergo more charge and discharge cycles before degrading, giving them a longer operational lifespan. These characteristics make them particularly suitable for certain applications such as in EVs which require adequate power density for acceleration and high energy density for increased range (Dragonfly Energy, 2024). Even so, lithium-ion battery characteristics vary across different chemistries.

The chemistry of a lithium-ion battery is determined by the material of its cathode. Usually, the anodes of lithium-ion batteries are made of graphite. However, cathode materials vary. For example, the cathode of Lithium Iron Phosphate (LFP) batteries is made of lithium-ion

phosphate, the cathode of Lithium Manganese Oxide (LMO) batteries is made of lithium manganese oxide, the cathode of Lithium Nickel Manganese Cobalt Oxide (NMC) batteries is made of nickel, manganese, and cobalt oxides, and the cathode of Lithium Nickel Cobalt Aluminum Oxide (NCA) batteries is made of lithium nickel cobalt aluminum oxide. Of these, NMC and NCA have higher specific energy, while LMO and LFP have higher specific power and thermal stability (Battery University, 2023). Despite their advantages, lithium-ion batteries suffer several drawbacks.

One drawback is that they lose capacity over time due to repeated charging and discharging which reduces their long-term efficiency. They are also susceptible to thermal runaway which can lead to fires or explosions, if the battery is damaged or improperly handled (Khan et al., 2024). For example, Quraishi et al. (2023) reported a fire at an e-bike shop which was caused by malfunctioning of a lithium-ion battery. Additionally, manufacturing and disposal processes of lithium-ion batteries contribute to environmental degradation. Mining of raw materials, such as lithium and nickel, for battery production causes soil erosion and water contamination. Improper disposal of batteries can lead to toxic chemical leakage. For example, batteries disposed of in landfills can release toxic substances such as heavy metals which seep into soil and water sources, posing a risk to human health (Kang et al., 2013). These limitations highlight the need for research into enhancing the safety, performance, and environmental sustainability of lithium-ion batteries.

2.2.2 Lithium-ion Battery Cell Dimensions

Lithium-ion batteries come in different shapes. These include prismatic, coin, cylindrical, and pouch. These further come in various sizes. The naming convention of lithium-ion batteries is such that their dimensions can be identified from their names. For example, prismatic cells are named according to their length, thickness, and width. A 103450 prismatic cell, for instance, has a thickness of 10 mm, width of 34 mm, and length of 50 mm (Idelah, 2020).

Coin cells get their name from their shape which resembles that of a coin. They are small and named based on their thickness and diameter. For instance, a CR2032 cell is 20mm in diameter and 32 mm thick. Cylindrical cells, on the other hand, are named according to their diameter and length, with common sizes being 18650, 26650, and 21700. The last digit, which is a zero, indicates that the battery is cylindrical while the first four numbers describe the dimensions of

the battery. 18650 cells have a diameter of 18 mm and 65 mm length, 26650 cells have a diameter of 26 mm and 65 mm length, and 21700 cells have a diameter of 21 mm and 70 mm length (Idelah, 2020). Figures 2.2 to 2.5 show the various cell shapes.

Figure 2. 2: Coin Cells (Idelah, 2020)

Figure 2. 3: Prismatic Cell (Idelah, 2020)

Figure 2. 4: Pouch Cell (Idelah, 2020)

Figure 2. 5: Cylindrical Cells (Idelah, 2020)

Each shape is suitable for particular applications. For example, coin cells are typically used where compactness and low power is required, such as in watches. Prismatic and polymer cells are used in handheld devices that have a thin profile, such as tablets. Cylindrical cells are used in a wide range of applications including medical and industrial applications. Due to their symmetrical form, they can be packed efficiently, hence they are commonly used in EVs (Idelah, 2020).

2.2.3 Battery Testing

Battery testing entails evaluating the performance of a battery to ensure its reliability and safety. Testing also helps to identify failure, degradation, and other issues that can affect battery performance. For example, capacity fade testing measures the ability of a battery to store charge over time which helps to estimate battery lifespan (Mulpuri et al., 2023). This typically involves subjecting a battery to multiple charge and discharge cycles as in the charge/discharge test. The charge/discharge test is a common technique used to evaluate battery performance. It is used to determine how efficiently a battery can store and deliver energy over multiple cycles. This test involves repeatedly charging and discharging a battery under controlled conditions (Kumar, 2023).

Battery charging can be carried out in various modes, typically involving constant current (CC) and constant voltage (CV) processes. For example, in CC charging mode, a fixed current is applied to the battery as voltage gradually increases. This proceeds until the maximum voltage is reached. Provision of a steady charging current ensures the battery gains charge efficiently. However, there is risk of the battery overcharging, overheating, or getting damaged if it stays in this mode for too long. CV charging mode usually follows CC charging to prevent battery overcharge, overheating, and damage. Once the maximum voltage is reached, charging transitions to CV mode. Here, voltage is held constant as the battery approaches full charge, while current decreases (Ufine Battery, 2024). Constant current constant voltage (CC-CV) charging is a combination of CC and CV modes.

Similar to charging, discharging can occur in CC, CV, and CC-CV modes. In CC discharge, current remains constant as voltage decreases ensuring a consistent current output. This makes it suitable for applications that require a stable supply of current such as power supplies. However, there is risk of over discharge if discharging proceeds for long. In CV discharging, on the other hand, voltage is held constant as current decreases. This mode is suitable where a stable voltage output is required. However, holding voltage constant can cause over drainage of the battery which can lead to cell damage or capacity reduction. CC-CV discharging is a combination of CC and CV modes. It starts with the CC phase then transitions to CV once a predefined voltage limit is reached. This mode helps to prevent over discharge and maintain battery performance (Ufine Battery, 2024).

2.2.4 SOH Determination Methods

SOH represents the current ability of a battery to store and deliver energy compared to when it was new. It is determined by the ratio of current battery capacity to initial battery capacity (Li et al., 2019) and is usually expressed as a percentage. Accurate SOH estimation is essential for ensuring optimum battery performance, safety in operation, and preventing failures due to battery malfunctions (Shu et al., 2021). Several methods have been developed for SOH estimation. These can be broadly classified into model-based methods and data-driven methods.

Model-based methods involve the development of battery models that simulate cell behaviour. For example, electrochemical models use partial differential equations to simulate internal physical and chemical battery processes such as mass and charge transfer kinetics (Li et al., 2019). These models tend to provide accurate SOH predictions. However, they have high computational complexity which limits their use, especially in real-time applications (Xu et al., 2022). Equivalent circuit models, on the other hand, represent batteries as electrical networks of resistors, capacitors, and voltage sources. These models assess battery behaviour through changes in parameters such as internal resistance and capacity. Despite having lower computational complexity, equivalent circuit models tend to be less accurate than electrochemical models. This is because they do not capture internal electrochemical battery processes (Tomasov et al., 2019).

Data-driven methods are gaining attention since they do not rely on complex physical models. They use battery data and machine learning techniques to detect patterns in battery performance (Li et al., 2019). These methods can process large amounts of battery cycling data to identify relationships between parameters such as voltage and battery SOH. Two common machine learning approaches are supervised learning and unsupervised learning, which differ primarily in the data used for training and the purpose of the models.

Unsupervised learning is a machine learning approach that uses unlabelled data for training models. The purpose of unsupervised learning is to detect patterns and relationships in data (Singh, 2024). Algorithms used in this approach include K-Means for clustering, autoencoders for anomaly detection, and principal component analysis for dimensionality reduction (Noema, 2022). Supervised learning, however, uses labelled data for model training. This involves using

input-output pairs to train a model to make accurate predictions on unseen data by learning relationships between inputs and their corresponding outputs (Singh, 2024). This approach uses algorithms such as Support Vector Machines (SVM), decision trees, and random forest. For example, Qian et al. (2023) estimated battery SOH using a model based on decision tree and Support Vector Regression (SVR) algorithms. SVR is a type of SVM.

SVR aims to fit a function such that most training data points lie within a margin of ϵ from predicted values. This forms a hyperplane that resembles a tube around the line of best fit. This tube is called the ϵ-sensitive tube. Errors of data points that fall outside this tube are computed, while errors of points that lie within the tube are disregarded (Salunkhe, 2021). This reduces sensitivity to minor deviations.

Additionally, SVR uses kernel functions to map input data points into higher dimensional spaces for modelling complex non-linear relationships. Different kernel functions exist and are selected based on the characteristics of the data and complexity of the relationship to be modelled. These include the linear kernel, polynomial kernel, Radial Basis Function (RBF), and sigmoid function. The linear kernel is simple and suitable for modelling linear relationships, the polynomial kernel is used for data with polynomial features, RBF is used for complex non-linear relationships, and the sigmoid function is mainly used for data whose distribution resembles a neural network (Abhishek, 2024). Optimisation algorithms such as Quantum Particle Swarm Optimisation (QPSO) can be used to enhance accuracy (Wang et al., 2018).

2.3 Empirical Review

Various studies have been carried out on SOH estimation of EV batteries. For example, Petkovski et al. (2023) developed an SVR model for SOH estimation using features derived from full and partial discharge capacity curves and battery temperature data. The model showed good performance, however, the study was geared towards determining the SOH of batteries still in use rather than retired batteries. Moreover, the study used publicly available cycling data which may not fully capture the degradation patterns of used batteries.

Lin et al. (2023) investigated a method for SOH estimation without the need for full cycle charging and discharging using piecewise linear regression. The method depicted robustness

against outliers and truncation errors. However, the method was tailored for in-use batteries. It also used Battery Management System (BMS) data for SOH estimation. A BMS collects data about EV batteries while they are in use but once they are removed from vehicles, this data is not readily available (Dunn, 2022).

There have been some studies carried out specifically for SOH determination of retired EV batteries. For example, Salek et al. (2024) developed time-series models to estimate SOH of retired batteries already employed in second life applications using data from their first life applications. While the models depicted high accuracy, they relied on first-life data of EV batteries which are not always available.

Some researchers have investigated using partial charge/discharge information for SOH estimation. This is an attempt to reduce the time taken to carry out charge/discharge tests by not having to go through full cycles. For example, Ahmeid et al. (2022) investigated the use of incremental capacity, equivalent circuit model, and manipulated coulomb counting to establish the capacity of retired EV battery packs from a partial discharge profile. Although the method was accurate in capacity estimation and portrayed good repeatability, it was still time consuming as it required batteries to be fully charged and discharged.

Zhang et al. (2021) also investigated a method to determine SOH of used batteries. The study employed impedance spectroscopy, charge/discharge, incremental capacity analysis, and average Fréchet distance. Tests were conducted on a battery module as opposed to individual cells. Although the study was able to identify significant features for SOH extraction, this required full curves to be obtained.

2.4 Study Gaps

The following were the gaps identified from the analysed literature:

1. Some studies focused on SOH estimation of EV batteries that were still in use
2. Some studies relied on first-life data or BMS data which are usually not accessible once batteries are removed from vehicles
3. Some studies used publicly available cycling datasets which may not fully capture the degradation patterns of used batteries
4. Some studies relied on full charge/discharge cycles for SOH estimation

This study aims to fill these gaps by developing a machine learning model that uses partial discharge data trained on actual cycling data from retired batteries for SOH estimation.

Chapter Three: Methodology

3.1 Introduction

This chapter describes the method that was used to achieve the study objectives and answer the research questions. The chapter has been organised in sections. Section one briefly discusses the research design used in this study. It is followed by the data collection method, including battery cycling and feature extraction. Subsequent sections discuss model development and validation.

3.2 Research Design

This study employed an experimental research design. The design entails establishing a relationship between variables, usually by manipulating an independent variable and observing the consequent effect on a dependent variable (Romanchuk, 2023). For example, in this study, the dependent variable was SOH. The independent variables were features of interest extracted from differential capacity curves.

3.3 Data Collection

3.3.1 Battery Cycling

Eight used 18650 lithium-ion EV battery cells, with a rated capacity of 2.2 Ah, were cycled. The batteries were first charged in CC-CV mode until they reached 4.2 V. They were then discharged in CC-CV mode until they reached a lower cut-off voltage of 2.75 V. The cells were cycled ten times with five-minute rest periods between charging and discharging. The C-rate used was 0.5C.

The testing system comprised of a Neware CT-4008T-5V6A-S1 cell tester and a desktop computer where the cycling software was installed. The Neware CT-4008T-5V6A-S1 cell tester consisted of nine modules, each with eight charging ports where cells were loaded. Once the cells were loaded, the charging and discharging protocols described above were set on the software. The cycling process was then initiated and cycling data was automatically recorded through the software as cycling continued. Figure 3.1 shows a Neware CT-4008T-5V6A-S1 cell tester.

Figure 3. 1: Neware CT-4008T-5V6A-S1 Cell Tester

After the cycling data was obtained, records of capacity, energy, and Direct Current Internal Resistance (DCIR) were extracted and used for analysis of the cycled cells. SOH was obtained from the capacity using equation 1.

$$SOH = \left(\frac{Q_{current}}{Q_{rated}}\right) \times 100\% \qquad (1)$$

Where:

$Q_{current}$ = the current cycle capacity

Q_{rated} = the rated capacity

3.3.2 Feature Extraction

The cycling data was used to plot differential capacity curves from which features of interest were extracted. These curves were obtained by differentiating the discharge capacity with respect to voltage, and this was done using equation 2.

$$\frac{dQ}{dV} = \frac{Q_{i+1} - Q_i}{V_{i+1} - V_i} \qquad (2)$$

Where:

$\frac{dQ}{dV}$ = the differential capacity

V_i = the voltage at a point i

Q_i = the capacity corresponding to V_i

V_{i+1} = the voltage measurement immediately after V_i

Q_{i+1} = the capacity corresponding to V_{i+1}

The differential capacity curves were then smoothed using a Savitsky-Golay filter. This filter uses a moving window and polynomials of specific degrees to fit data. It is a good method for reducing noise while preserving features of interest (Konstantinovsky, 2024). This is implemented using equation 3.

$$y_i^{smoothed} = \sum_{j=-\frac{m-1}{2}}^{\frac{m-1}{2}} c_j y_{i+j} \qquad (3)$$

Where:

$y_i^{smoothed}$ = smoothed value at index *i*

y_{i+j} = original data value at the *jth* neighbouring point

c_j = coefficients of the polynomial fitting

Features of interest were then obtained from the initial segment of discharge, that is, from 3.6 V to 4.2 V. The features of interest extracted were included peak characteristics, curve characteristics, and statistical characteristics of the extracted differential capacity curve. A peak was identified at a point *i* if equation 4 held and a valley was identified if equation 5 held.

$$f(V_i) > f(V_{i-1}) \; and \; f(V_i) > f(V_{i+1}) \qquad (4)$$

$$f(V_i) < f(V_{i-1}) \; and \; f(V_i) < f(V_{i+1}) \qquad (5)$$

Where:

$f(V_i)$ = dQ/dV value at V_i

V_i = voltage at i

V_{i-1} *and* V_{i+1} = neighbouring voltage points

Peak and valley widths were computed using equation 6.

$$w = V_{right} - V_{left} \qquad (6)$$

Where:

w= width of peak or valley

V_{right}, V_{left} = voltages where peak height is half its maximum value

The area under the portion of the extracted curve, also called the dQ/dV integral was computed using Simpson's rule. This is represented by equation 7.

$$\int_{V_{min}}^{V_{max}} f(V)dV \approx \frac{h}{3} \sum_{i=0,2,\ldots}^{n-2} \left(f(V_i) + 4f(V_{i+1}) + f(V_{i+2})\right) \qquad (7)$$

Where:

$f(V)$ = dQ/dV values

V_{min} = minimum voltage

V_{max} = maximum voltage

h = step size between voltage points

n = number of voltage points

The slope at 3.7 V represents the differential capacity at 3.7 V. It was determined using equation 8.

$$\left.\frac{dQ}{dV}\right|_{V=3.7\,V} \approx \frac{dQ/dV[i+1] - dQ/dV[i-1]}{V[i+1] - V[i-1]} \quad (8)$$

Where:

i = index of voltage closest to 3.7 V

$dQ/dV[i+1], dQ/dV[i-1]$ = neighbouring dQ/dV values

$V[i+1], V[i-1]$ = neighbouring voltage values

The voltage at the maximum slope was computed using equation 9.

$$V_{\max slope} = V_i \; where \max\left(\frac{df}{dV}\right) = max \frac{dQ/dV[i+1] - dQ/dV[i]}{V[i+1] - V[i]} \quad (9)$$

Where:

$V_{\max slope}$ = maximum slope voltage

i = maximum gradient point

The mean and standard deviation were computed using equation 10 and 11.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} dQ/dV_i \qquad (10)$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(dQ/dV - \mu)^2} \qquad (11)$$

Where:

μ = mean dQ/dV

σ = standard deviation of dQ/dV

n = number of points

After the features were extracted, statistical analysis using Analysis of Variance (ANOVA) was then used to select statistically significant features. ANOVA was computed using the F-statistic which was in turn used for comparison with the error term of 0.05. The F-statistic is determined using equation 12.

$$F = \frac{MSB}{MSW} \qquad (12)$$

Where:
MSB is the Mean Square Between Groups

MSW is the Mean Square Within Groups

MSB and MSW are determined by equations 13 and 14 respectively.

$$MSB = \frac{SSB}{df_B} \qquad (13)$$

$$MSW = \frac{SSW}{df_W} \qquad (14)$$

Where:

SSB is the between group sum of squares

df_B is the between group degrees of freedom

SSW is the within group sum of squares

df_W is the within group degrees of freedom

df_B is determined using equation 15 while df_W is determined using equation 16.

$$df_B = k - 1 \qquad (15)$$

$$df_W = N - k \qquad (16)$$

Where:

N is the total number of samples

k is the total number of groups

SSB is computed using equation 17 and SSW is computed using equation 18.

$$SSB = \sum_{i=1}^{k} n(\bar{x}_i - \bar{x})^2 \qquad (17)$$

$$SSW = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i\right)^2 \qquad (18)$$

Where:

$\bar{x}_i$ = the mean of group *i*

$\bar{x}$ = the overall mean across all groups

x_{ij} = the *j*th observation in the *i*th group

n_i = the number of observations in group *i*

Pearson correlation analysis was then performed on identified significant features to quantify their correlation to SOH. This was done using equation 19.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \qquad (19)$$

Where:

r = correlation coefficient

xi = x-variable values

yi = y-variable values

$\bar{x}$=mean of x-variable values

$\bar{y}$=mean of y-variable values

3.4 Model Development

3.4.1 QPSO-SVR Model

The machine learning model developed in this study was implemented using a QPSO-SVR technique presented by Wang et al. (2018). In this method, QPSO is used to solve the optimization problem of SVR. SVR aims to find a function *f(x)* that approximates the relationship between an input *x* and an output *y*, while minimising error within a margin ϵ. This is computed using equation 20.

$$f(x) = w^T \phi(x) + b \qquad (20)$$

Where:

w^T = the weight vector

$\phi(x)$ = the kernel function

b = the bias term

The aim is to minimise the objective function:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$

subject to the constraints:

$$y_i - (w^T\phi(x_i) + b) \leq \epsilon + \xi_i$$

$$(w^T\phi(x_i) + b) - y_i \leq \epsilon + \xi_i^*$$

$$\xi_i\xi_i^* \geq 0$$

Where:

ϵ = margin of tolerance

C = regularisation parameter

ξ_i and ξ_i^* = slack variables allowing deviations beyond ϵ

In QPSO, the position of a particle *i* is updated using equation 21.

$$x_i^{(t+i)} = p_i^{mbest} \pm \beta \cdot \left|p_i^{mbest} - x_i^{(t)}\right| \cdot ln\left(\frac{1}{u}\right) \qquad (21)$$

Where:

$x_i^{(t)}$ = current position of particle *i* at iteration *t*

β = contraction-expansion coefficient

u = random number uniformly distributed in (0,1)

p_i^{mbest} = mean best position of the swarm

p_i^{mbest} is determined using equation 22.

$$p_i{}^{mbest} = \frac{1}{N}\sum_{i=1}^{N} p_i \qquad (22)$$

Where:

p_i = the best position of particle i

N = the total number of particles in the swarm

The QPSO-SVR method by Wang et al. (2018) begins by defining a search space, which includes the parameters to be optimized and other parameters such as the particle size population, *M*, the number of iterations, *MT*, and the fitness function. The parameters to be optimized are the penalty parameter, *C*, the error term, *ε*, and the kernel bandwidth, *σ*. The fitness function is determined using equation 23.

$$F_{fitness} = \frac{1}{N}\sum_{j=1}^{N}(y_i - \hat{y}_i)^2 \qquad (23)$$

Where:

y_i = real value of *j*th sample

$\hat{y}_i$ = training output value of *j*th sample

N = number of training samples

Particles are then initialized randomly. If the position of a particle is outside the search space, its fitness function value is set to infinity. If the position of a particle is within the search space, it is used to train the SVR model and its fitness value is computed using equation 23. The position of the particles is then updated according to the fitness value. The position of the particles is then checked again to see if they lie within or outside the search space, and the process is repeated until the number of iterations set is reached. Figure 3.2 presents a flowchart of the process.

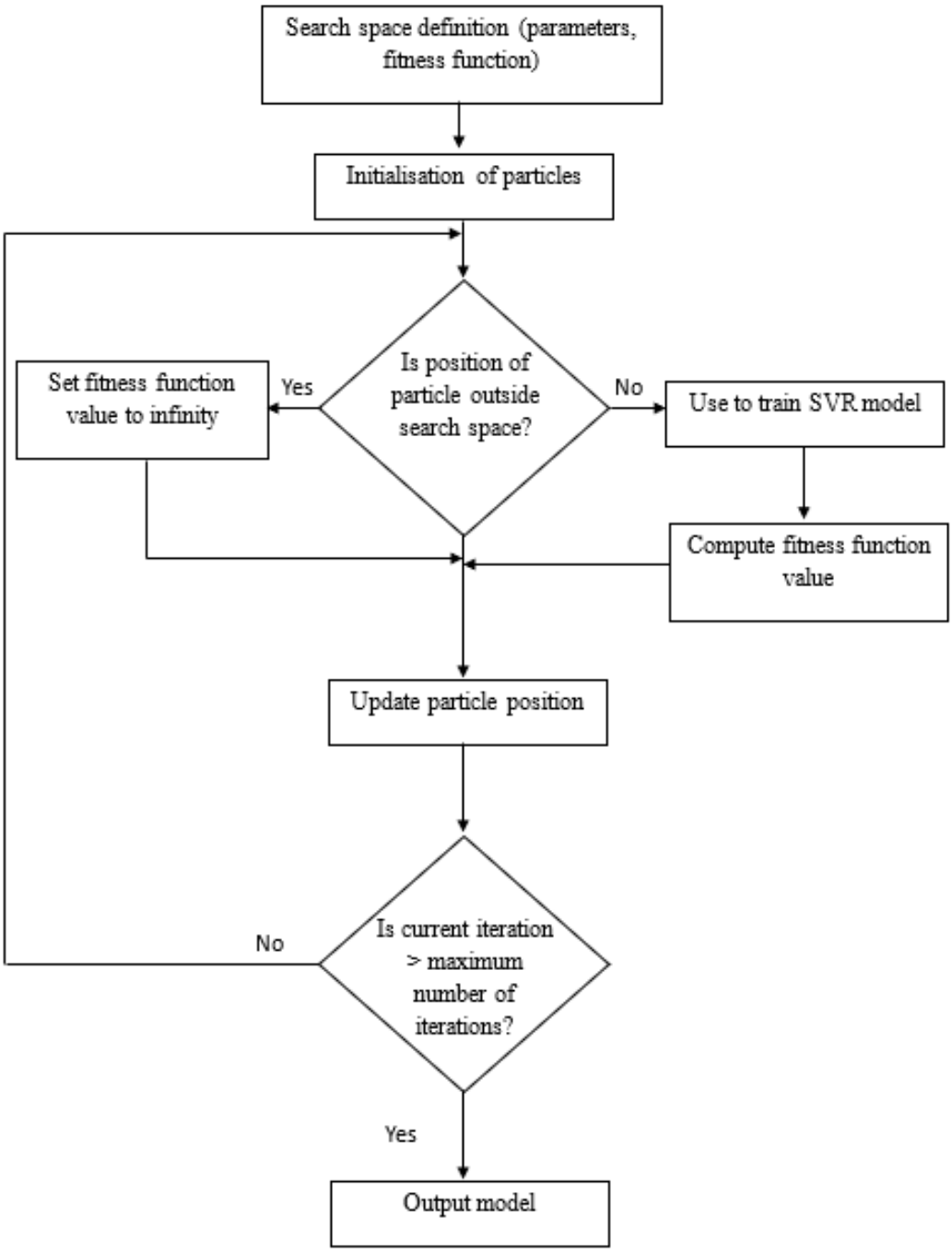


Figure 3. 2: QPSO-SVR Flow Chart

3.4.2 Model Training

The features of interest obtained from the feature extraction process were used as inputs for the model. Once these features were obtained, they were paired with SOH labels. The labeled data was then split into a training set and a testing set for evaluating its performance.

3.4.3 Model Validation

The model was validated using evaluation metrics. These were Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2). RMSE, MAE, and R^2 are computed using equations 24, 25, and 26 respectively.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (24)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (25)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (26)$$

Where:

y_i = actual value for the *i*th data point

$\hat{y}_i$ = predicted value for the *i*th data point

n = total number of data points

$\bar{y}$ = mean of the actual values

$\bar{y}$ is determined using equation 27.

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (27)$$

Feature extraction and model development and validation was implemented using python on Google Colab.

3.5 Method Flow Diagram

Discharge data, which included voltage and capacity data, was used to construct a partial differential capacity curve from which features of interest were extracted. The portion of the curve used for feature extraction was that ranging from 3.6 V to 4.2 V. After features of interest were obtained, the results were divided into a training data set and a testing data set. The training data set was used to train the model, which was based on SVR and used QPSO for optimization. The testing data set was then used for validating the performance of the model. Figure 3.3 shows the method flow diagram.

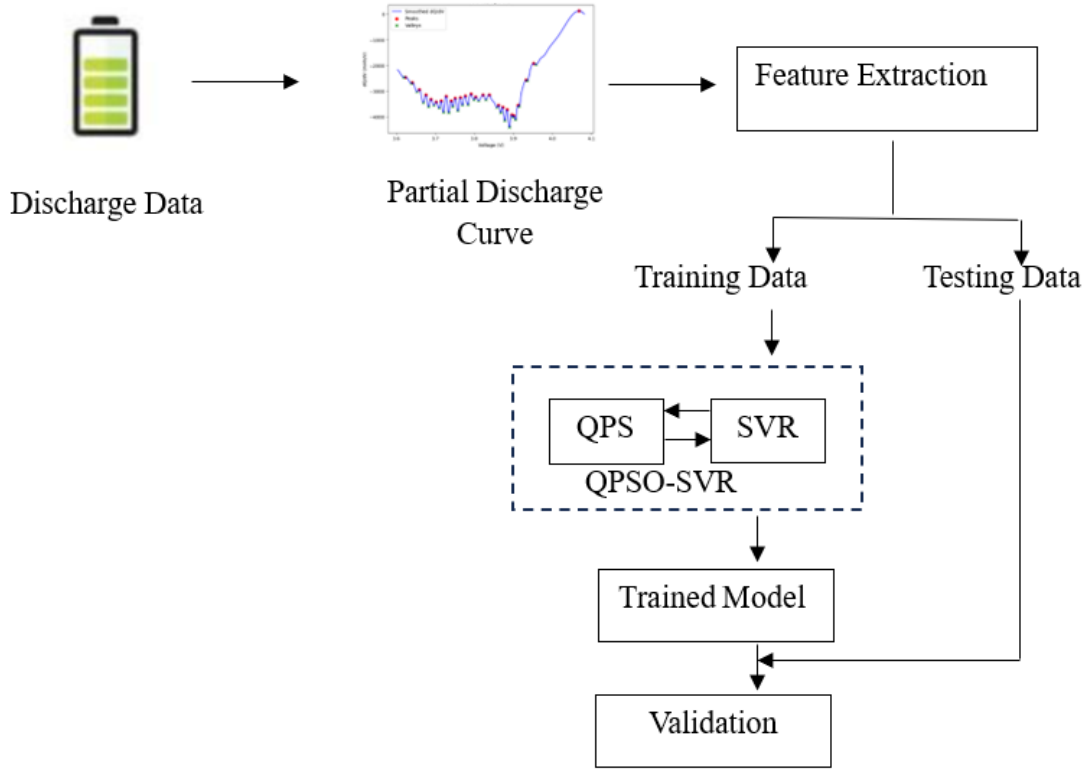


Figure 3. 3: Method Flow Diagram

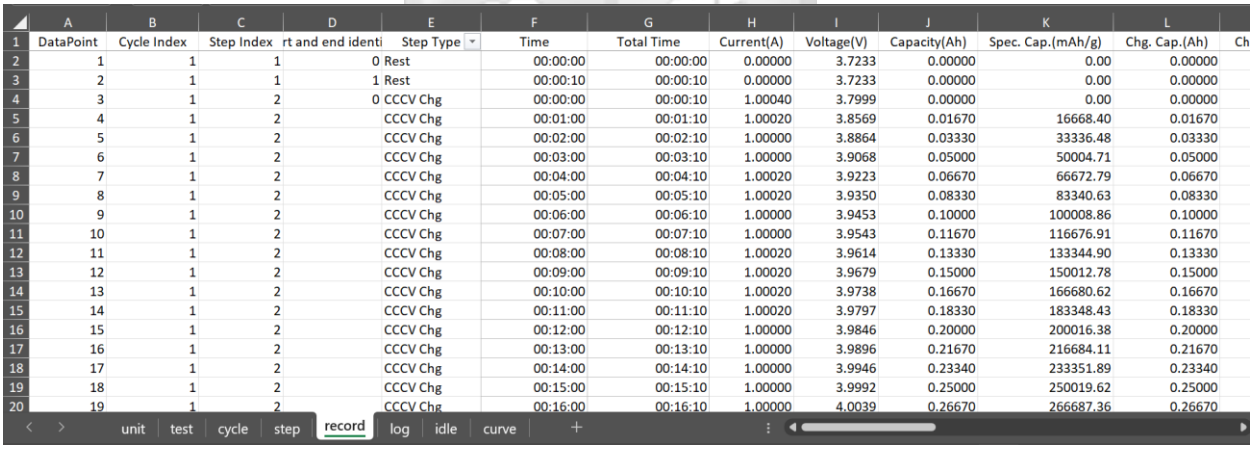
Chapter Four: Results

4.1 Introduction

This chapter presents the results obtained from the methods used to achieve the study objectives. The results from battery cycling are presented in the first section, followed by the results from feature extraction in the second section. The third section discusses results from model development.

4.2 Battery Cycling

During cycling, data was recorded automatically every ten seconds as the process proceeded. This data was recorded on Excel worksheets and included current, voltage, time, DCIR, energy, and capacity values. The data was obtained for eight cells each cycled ten times. A sample record is shown in Figure 4.1.



	A	B	C	D	E	F	G	H	I	J	K	L	
1	DataPoint	Cycle Index	Step Index	rt and end identi	Step Type	Time	Total Time	Current(A)	Voltage(V)	Capacity(Ah)	Spec. Cap.(mAh/g)	Chg. Cap.(Ah)	Ch
2	1	1	1	0	Rest	00:00:00	00:00:00	0.00000	3.7233	0.00000	0.00	0.00000	
3	2	1	1	1	Rest	00:00:10	00:00:10	0.00000	3.7233	0.00000	0.00	0.00000	
4	3	1	2	0	CCCV Chg	00:00:00	00:00:10	1.00040	3.7999	0.00000	0.00	0.00000	
5	4	1	2		CCCV Chg	00:01:00	00:01:10	1.00020	3.8569	0.01670	16668.40	0.01670	
6	5	1	2		CCCV Chg	00:02:00	00:02:10	1.00000	3.8864	0.03330	33336.48	0.03330	
7	6	1	2		CCCV Chg	00:03:00	00:03:10	1.00000	3.9068	0.05000	50004.71	0.05000	
8	7	1	2		CCCV Chg	00:04:00	00:04:10	1.00020	3.9223	0.06670	66672.79	0.06670	
9	8	1	2		CCCV Chg	00:05:00	00:05:10	1.00020	3.9350	0.08330	83340.63	0.08330	
10	9	1	2		CCCV Chg	00:06:00	00:06:10	1.00000	3.9453	0.10000	100008.86	0.10000	
11	10	1	2		CCCV Chg	00:07:00	00:07:10	1.00000	3.9543	0.11670	116676.91	0.11670	
12	11	1	2		CCCV Chg	00:08:00	00:08:10	1.00020	3.9614	0.13330	133344.90	0.13330	
13	12	1	2		CCCV Chg	00:09:00	00:09:10	1.00020	3.9679	0.15000	150012.78	0.15000	
14	13	1	2		CCCV Chg	00:10:00	00:10:10	1.00020	3.9738	0.16670	166680.62	0.16670	
15	14	1	2		CCCV Chg	00:11:00	00:11:10	1.00020	3.9797	0.18330	183348.43	0.18330	
16	15	1	2		CCCV Chg	00:12:00	00:12:10	1.00000	3.9846	0.20000	200016.38	0.20000	
17	16	1	2		CCCV Chg	00:13:00	00:13:10	1.00000	3.9896	0.21670	216684.11	0.21670	
18	17	1	2		CCCV Chg	00:14:00	00:14:10	1.00000	3.9946	0.23340	233351.89	0.23340	
19	18	1	2		CCCV Chg	00:15:00	00:15:10	1.00000	3.9992	0.25000	250019.62	0.25000	
20	19	1	2		CCCV Chg	00:16:00	00:16:10	1.00000	4.0039	0.26670	266687.36	0.26670	

unit | test | cycle | step | record | log | idle | curve

Figure 4. 1: Sample Cycling Record

Values of capacity, energy, and DCIR were extracted for analysis. The SOH of the batteries was computed and included in the analysis. These are shown in Table 4.1.

Table 4. 1: Battery Parameters Extracted

Cell	Cycle	Capacity (Ah)	Energy (Wh)	DCIR (mΩ)	SOH (%)
1	1	1.7182	6.3389	68.77	78.1000
	2	1.7178	6.2665	86.48	78.0818
	3	1.6905	6.2437	67.18	76.8409
	4	1.6948	6.2688	65.09	77.0364
	5	1.6651	6.1211	71.96	75.6864

	6	1.648	6.0478	73.47	74.9091
	7	1.6313	5.9769	76.19	74.1500
	8	1.6226	5.949	76.19	73.7545
	9	1.6381	6.0468	71.87	74.4591
	10	1.6504	6.1107	66.28	75.0182
2	1	0	0.0001	538.92	0.0000
	2	0	0.0001	608.11	0.0000
	3	0	0.0001	608.11	0.0000
	4	0	0.0001	608.11	0.0000
	5	0	0.0001	151.26	0.0000
	6	0	0.0001	588.24	0.0000
	7	0.0001	0.0002	106.38	0.0045
	8	0.0001	0.0002	256.72	0.0045
	9	0	0.0002	148.19	0.0000
	10	0.0001	0.0002	372.67	0.0045
3	1	1.653	6.1205	64.8	75.1364
	2	1.6145	5.9245	64.8	73.3864
	3	1.6221	5.9967	65.71	73.7318
	4	1.6306	6.0394	63.79	74.1182
	5	1.6152	5.9516	66.31	73.4182
	6	1.5898	5.8454	69.7	72.2636
	7	1.5747	5.7747	71.91	71.5773
	8	1.5672	5.7447	72.3	71.2364
	9	1.5724	5.787	70.01	71.4727
	10	1.5945	5.8993	64.22	72.4773
4	1	1.6656	6.1391	66.29	75.7091
	2	1.611	5.868	66.99	73.2273
	3	1.6309	6.015	66.69	74.1318
	4	1.634	6.0394	64.8	74.2727
	5	1.6112	5.9218	68.89	73.2364
	6	1.5876	5.8257	71	72.1636
	7	1.5659	5.7312	73.5	71.1773
	8	1.5568	5.696	74.39	70.7636
	9	1.5626	5.7464	71.89	71.0273
	10	1.5815	5.8481	66	71.8864
5	1	1.5824	5.8545	65.96	71.9273
	2	1.5134	5.5388	64.76	68.7909
	3	1.5303	5.6387	67.55	69.5591
	4	1.5291	5.6482	63.47	69.5045
	5	1.5066	5.5337	66.06	68.4818
	6	1.4661	5.3573	71.56	66.6409
	7	1.4404	5.2414	74.07	65.4727
	8	1.4241	5.1773	74.07	64.7318
	9	1.4139	5.1471	74.69	64.2682
	10	1.4327	5.2538	69.36	65.1227

6	1	1.7325	6.4229	64.78	78.7500
	2	1.6892	6.1847	63.19	76.7818
	3	1.7044	6.3129	65.68	77.4727
	4	1.7124	6.3505	62.57	77.8364
	5	1.6862	6.2159	67.59	76.6455
	6	1.6707	6.1475	68.8	75.9409
	7	1.6552	6.0781	71.3	75.2364
	8	1.6484	6.0547	71.01	74.9273
	9	1.6645	6.1522	67.89	75.6591
	10	1.6804	6.231	63.49	76.3818
7	1	1.6619	6.1517	64.8	75.5409
	2	1.6144	5.9107	64.19	73.3818
	3	1.6265	6.0136	65.1	73.9318
	4	1.6292	6.0355	62.9	74.0545
	5	1.6087	5.9269	66.13	73.1227
	6	1.5839	5.8246	70.11	71.9955
	7	1.5655	5.7413	71.93	71.1591
	8	1.5574	5.7091	72.81	70.7909
	9	1.5621	5.7511	70.13	71.0045
	10	1.5822	5.8576	65.1	71.9182
8	1	1.6883	6.2183	67.9	76.7409
	2	1.6258	5.8903	87.2	73.9000
	3	1.6554	6.1072	66.69	75.2455
	4	1.6596	6.1328	62.87	75.4364
	5	1.63	5.9858	69.49	74.0909
	6	1.6105	5.9062	71.59	73.2045
	7	1.5889	5.8111	74.71	72.2227
	8	1.5781	5.7702	75.3	71.7318
	9	1.5884	5.8444	71	72.2000
	10	1.6042	5.934	65.08	72.9182

From the results, it was observed that Cell 2 failed. This was depicted in its capacity, energy, and SOH values being very low, and its DCIR values being very high compared to the other cells. These low capacity and energy values show significant degradation in the cell, for example, the energy values are almost zero. This means the cell has lost its ability to store energy.

The initial capacity, energy, DCIR, and SOH were different for each cell. Initial values here refers to the values obtained from the first cycle. At the end of 10 cycles, the capacity, energy, and SOH had decreased while the DCIR had increased. This was observed in all cells except Cell 2 which had failed. Table 4.2 shows the comparison between initial and final values, where final values refers to the values obtained from the tenth cycle.

Table 4. 2: Initial and Final Values of Battery Parameters

Cell	Initial Capacity (Ah)	Final Capacity (Ah)	Initial Energy (Wh)	Final Energy (Wh)	Initial DCIR (mΩ)	Final DCIR (mΩ)	Initial SOH (%)	Final SOH (%)
1	1.7182	1.6504	6.3389	6.1107	68.77	66.28	78.1000	75.0182
2	0	0.0001	0.0001	0.0002	538.92	372.67	0.0000	0.0045
3	1.653	1.5945	6.1205	5.8993	64.8	64.22	75.1364	72.4773
4	1.6656	1.5815	6.1391	5.8481	66.29	66	75.7091	71.8864
5	1.5824	1.4327	5.8545	5.2538	65.96	69.36	71.9273	65.1227
6	1.7325	1.6804	6.4229	6.231	64.78	63.49	78.7500	76.3818
7	1.6619	1.5822	6.1517	5.8576	64.8	65.1	75.5409	71.9182
8	1.6883	1.6042	6.2183	5.934	67.9	65.08	76.7409	72.9182

Capacity, energy, and SOH decreased down cycles while DCIR increased. Figure 4.2, 4.3, 4.4, and 4.5 show the changes in capacity, energy, SOH, and DCIR respectively.

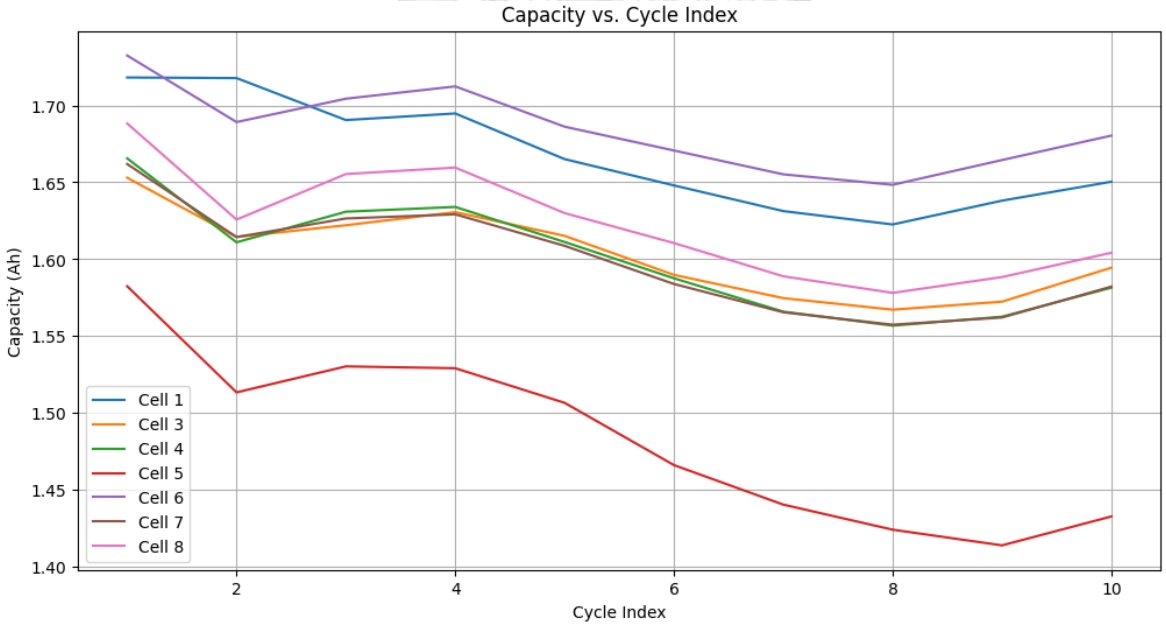


Figure 4. 2: Capacity Changes Down Cycles

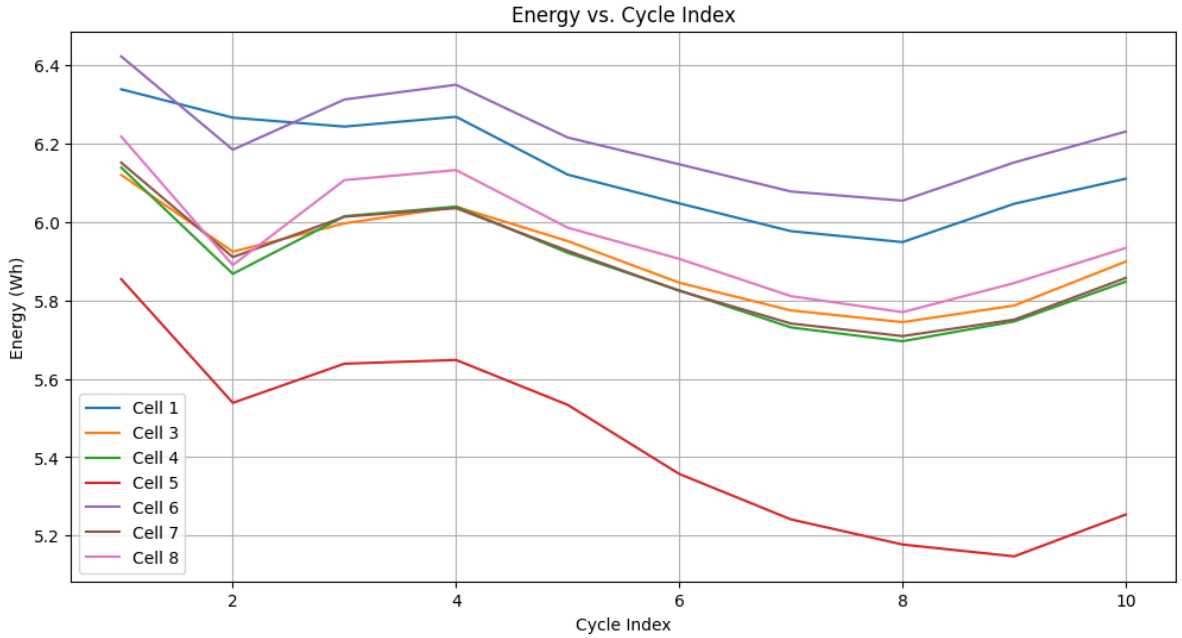


Figure 4. 3: Energy Changes Down Cycles

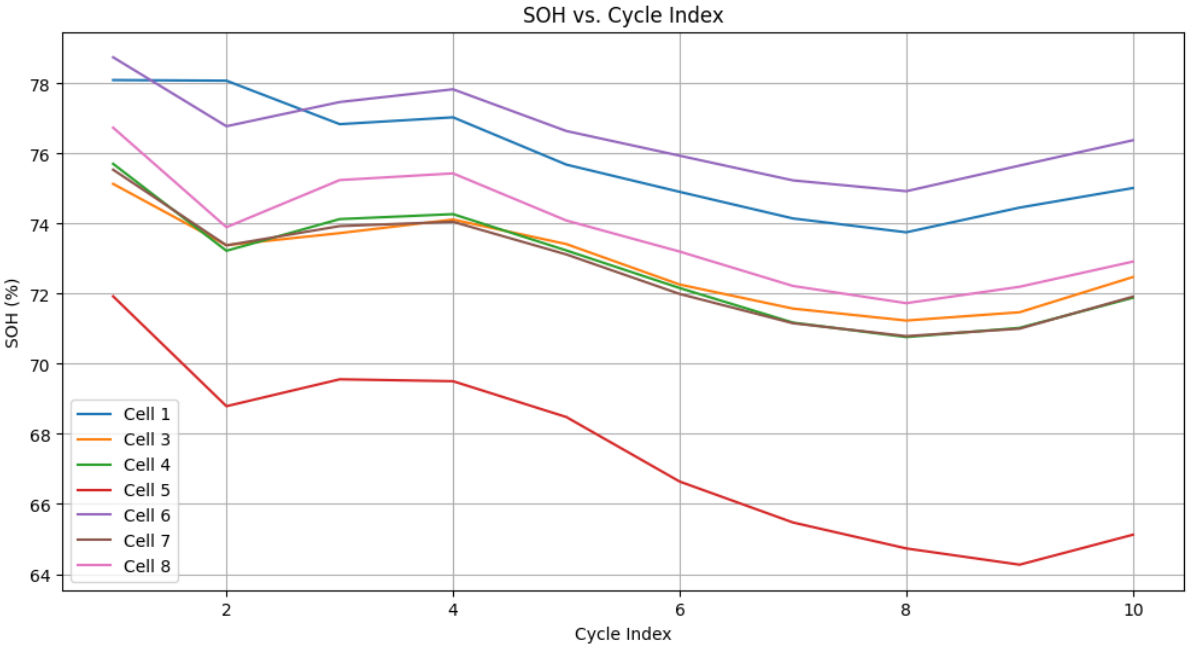


Figure 4. 4: SOH Changes Down Cycles

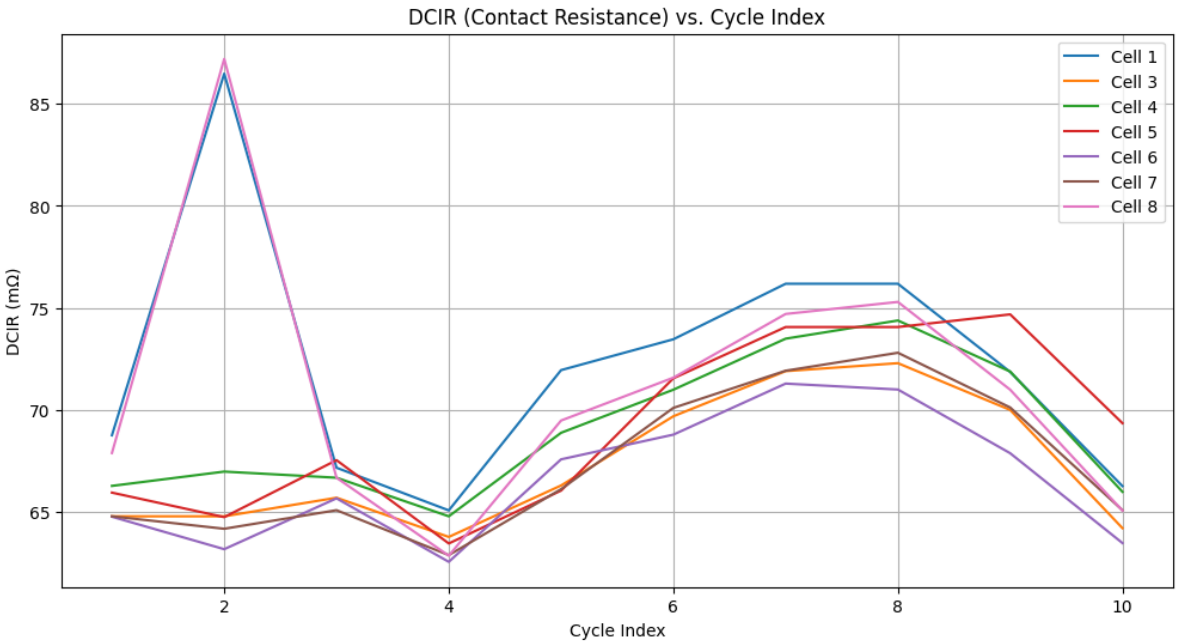


Figure 4. 5: DCIR Changes Down Cycles

Figures 4.2 to 4.5 show a direct relationship between capacity, energy, and SOH, and an indirect relationship between these three characteristics and DCIR.

4.2 Feature Extraction

Differential capacity curves were obtained for each cycle of every cell. These curves were constructed using discharge data. The portion of the curve selected for feature extraction was the portion ranging from 3.6 V to 4.2 V. Within this range, electrochemical reactions in a cell are well captured and peak shifts show stronger correlation with battery aging. Additionally, outside this range, differential capacity signals tend to be noisier and are more influenced by side reactions which reduces their reliability for SOH estimation (Lai et al., 2020).

Differential capacity curves were plotted for all the cells except Cell 2 which failed. Differential capacity values could not be obtained for Cell 2 since the cycling results produced capacity values of zero. Hence, no curves were generated for this cell. Figure 4.6 shows the differential capacity curves for Cell 1.

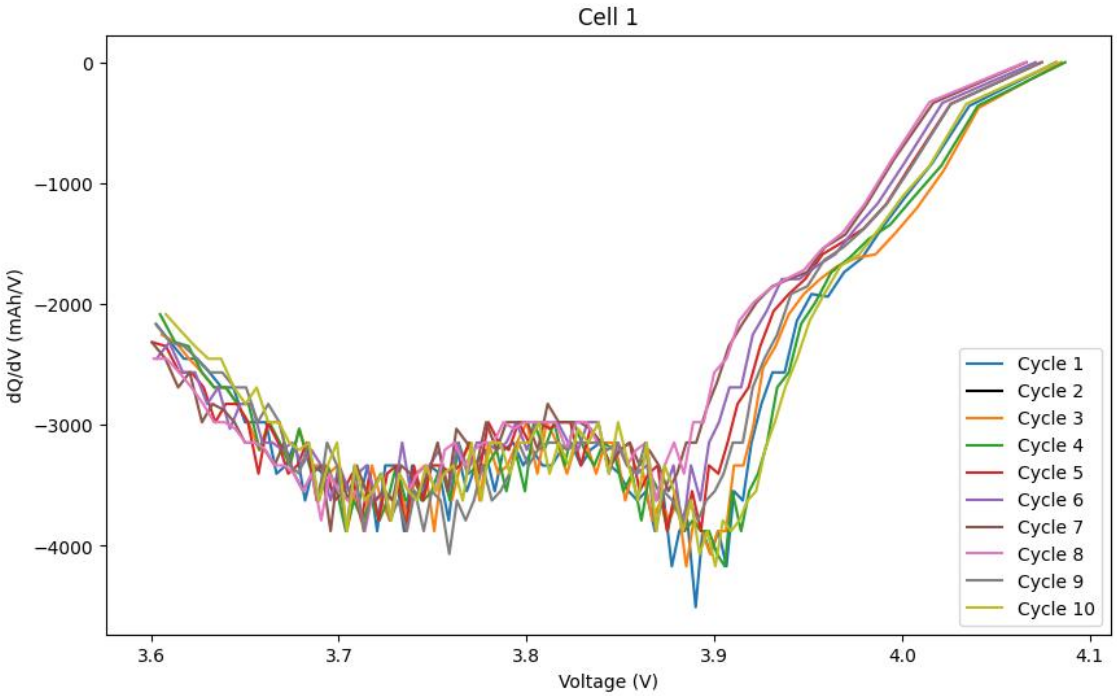


Figure 4. 6: Differential Capacity Curves for Cell 1

The curves were then smoothed before being used for feature extraction. This was done using a Savitsky-Golay filter. Figure 4.7 shows the smoothed curves for Cell 1.

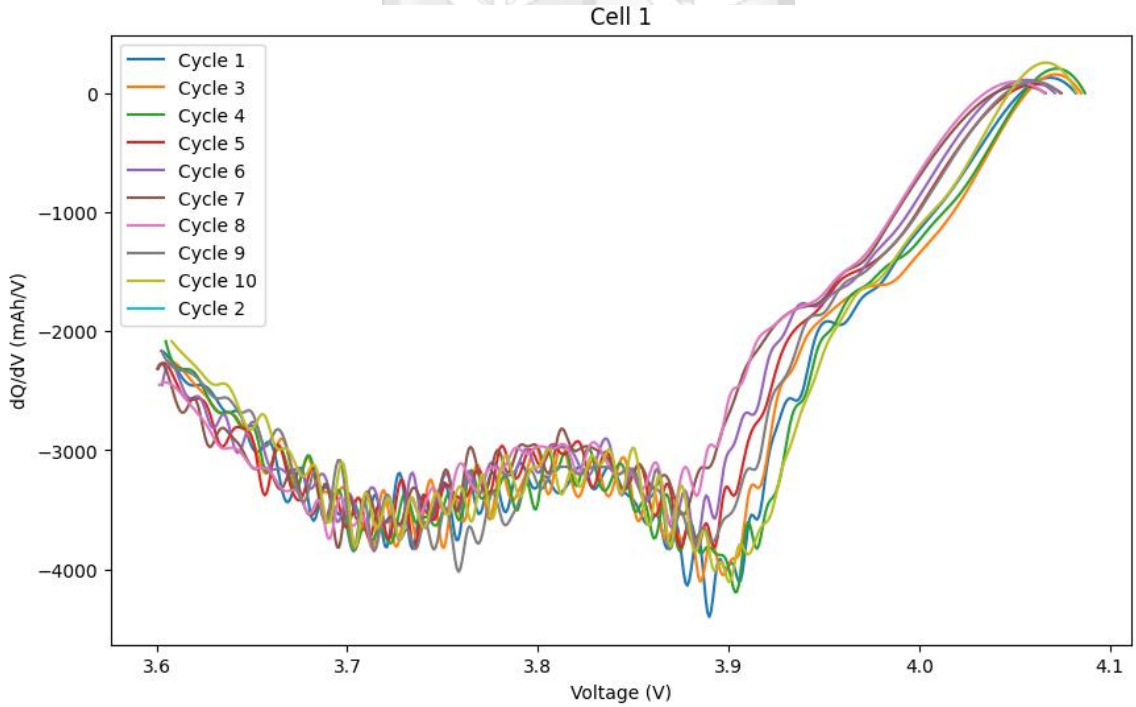


Figure 4. 7: Smoothed Differential Capacity Curves for Cell 1

The differential capacity curves were observed to have peaks and valleys at different positions. From Figure 4.7, it can be seen that the position of these peaks and valleys varies down cycles. There are differences in the heights and widths of the peaks and valleys as well. Peaks in differential capacity curves are as a result of electrochemical processes occurring within the cell (Lai et al., 2020).

Peak characteristics, curve characteristics, and statistical characteristics of the differential capacity curves were considered as features for developing the model. Figure 4.8.shows the portion of the curve obtained for the first cycle of Cell 1, from which features were derived.

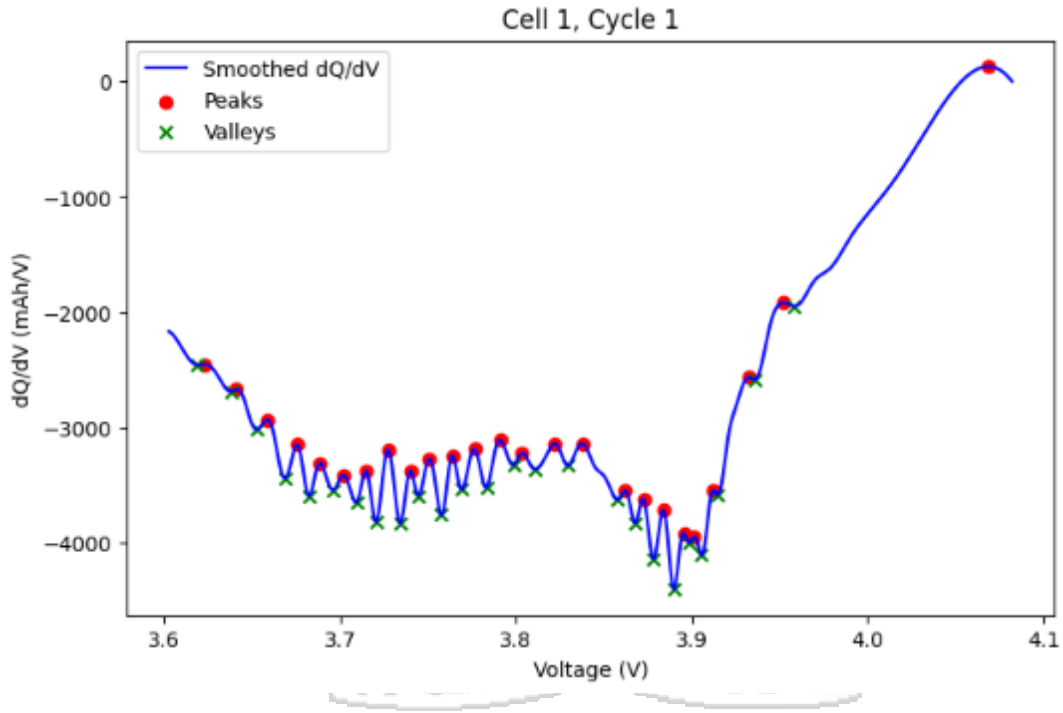


Figure 4. 8: Portion of Curve Extracted from Cell 1, Cycle 1

Peak characteristics included the number of peaks, the voltage at which the first peak occurred, the height of the first peak, the width of the first peak, the number of valleys, the voltage at which the first valley occurred, the depth of the first valley, and width of the first valley. The curve characteristics considered were the area under the curve, the voltage at which the maximum slope occurred, and the slope at 3.7 V.

Statistical characteristics taken under consideration were the mean and standard deviation of the dQ/dV distribution. These features were extracted then Analysis of Variance (ANOVA)

was performed to determine their statistical significance. The results of the ANOVA test are shown in Table 4.3.

Table 4. 3: ANOVA Results

Feature	F-statistic	p-value
Number of peaks	3.23484163	0.007798568
Voltage at first peak	1.51220362	0.188746991
Height of first peak	1.22588443	0.305123899
Width of first peak	1.25377479	0.291617862
Number of valleys	3.48165138	0.004906076
Voltage at first valley	2.04178772	0.073027727
Depth of first valley	0.92935042	0.48028855
Width of first valley	1.50166388	0.19221664
Area under the curve	8.69457365	6.72E-07
Slope at 3.7 V	10.2754569	6.55E-08
Maximum slope voltage	1.01585923	0.423254286
Mean	13.8207607	5.95E-10
Standard deviation	4.97096252	0.000318877

Only features with p-values of less than 0.05 were selected, while the rest were excluded. Six features were identified to have met this criterion. These were the number of peaks, the number of valleys, the area under the curve – represented by the dQ/dV integral, the slope at 3.7 V, the mean, and the standard deviation. The p-values obtained for these features depicted statistically significant variations. This could be attributed to different degradation patterns across the cells.

Pearson correlation was then performed on the six significant features to determine their correlation with SOH. Figure 4.9 shows the results of correlation, which are summarized in Table 4.4.

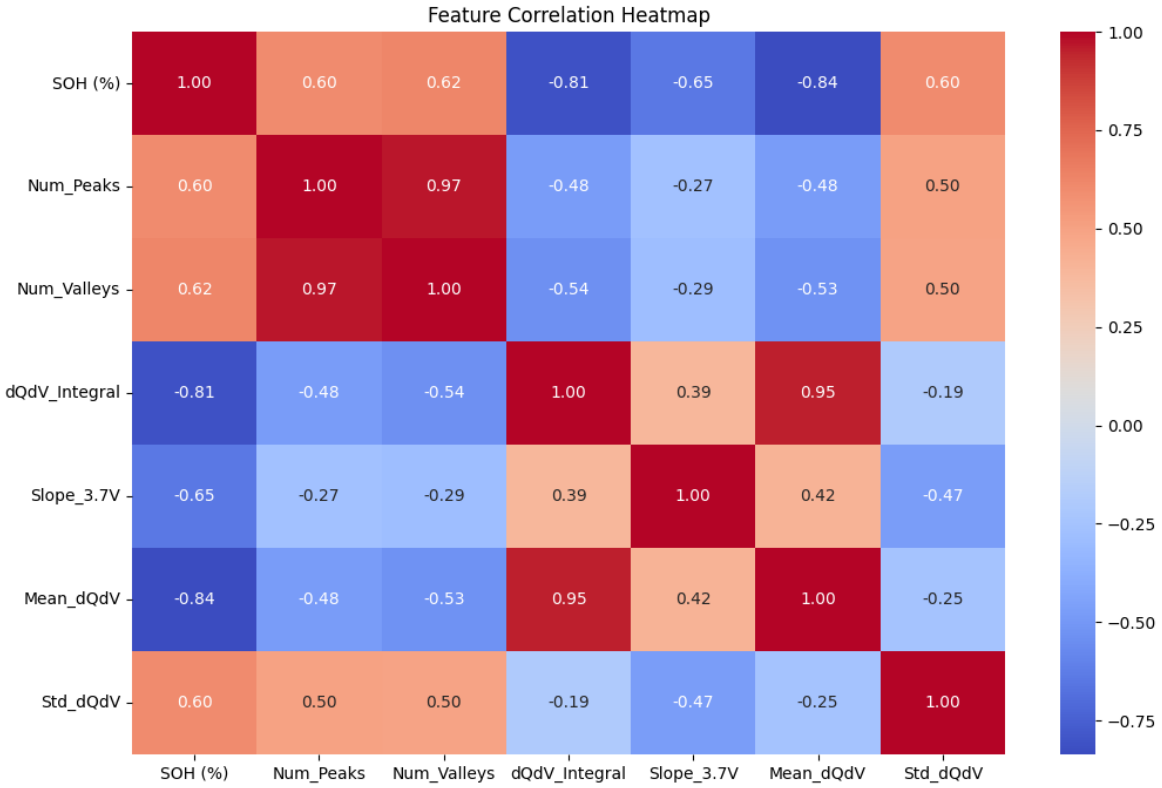


Figure 4. 9: Feature Correlation Heatmap

Table 4. 4: Correlation Coefficients of Selected Features

Feature	**Feature Label**	**Correlation Coefficient (*r*)**
Mean	Mean_dQdV	0.835694308
Area under the curve	dQdV_Integral	0.809697643
Slope at 3.7 V	Slope_3.7V	0.648764587
Number of Valleys	Num_Valleys	0.620474193
Number of Peaks	Num_Peaks	0.602833361
Standard Deviation	Std_dQdV	0.602529676

All the features had correlation coefficients higher than 0.5 which means that these features were substantially correlated to SOH. The coefficients were positive which suggests a direct relationship between the features and SOH.

4.3 Model Results

The input to the machine learning model were the selected features of interest and the output was SOH. The model was implemented using python. After the features of interest were extracted, the data was split into a training set and a testing set. The training set was used for training the model while the testing set was used for validation. The model was first implemented using SVR then optimized with QPSO. The results of the SVR and QPSO-SVR model are depicted in Table 4.5.

Table 4. 5: QPSO-SVR and SVR Comparison

	MAE	**RMSE**	**R^2**
QPSO-SVR	0.9286	1.1577	0.7951
SVR	1.1865	1.6247	0.5965

Lower MAE and RMSE values are obtained using QPSO-SVR compared to SVR. Additionally, a higher R^2 is obtained for QPSO-SVR. Lower MAE and RMSE values correspond to higher model accuracy while higher R^2 indicates a better fit (Chugh, 2020). Therefore, the QPSO-SVR model is more accurate and fits the data better. This is also shown in Figure 4.10.

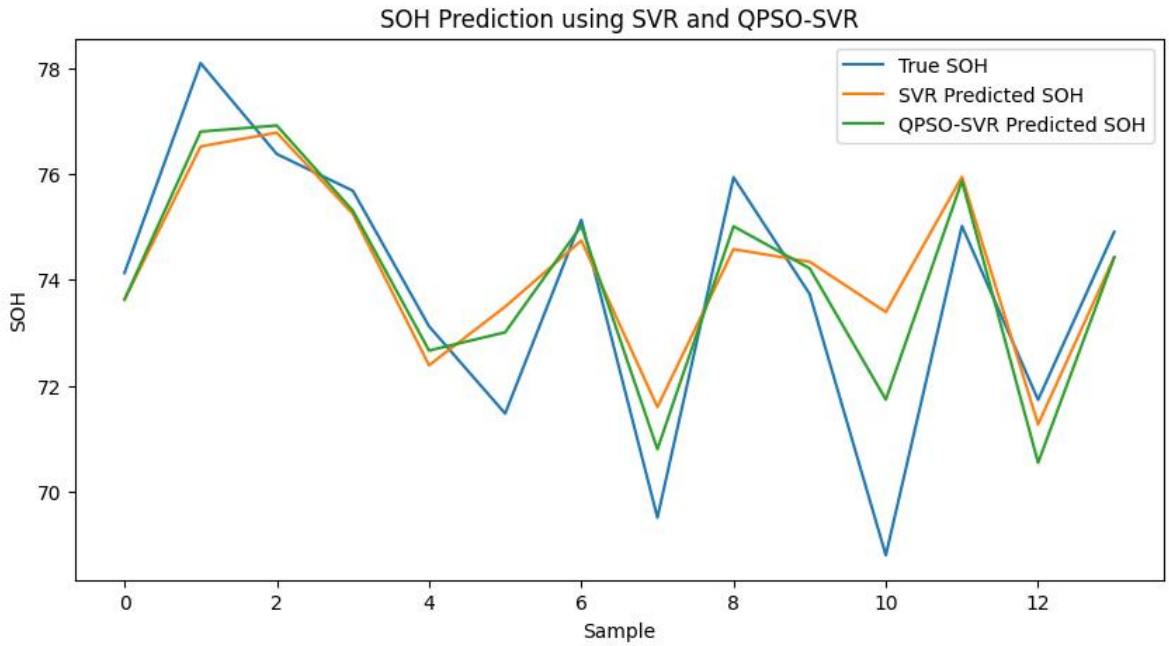


Figure 4. 10: QPSO-SVR and SVR Prediction Comparison

The model was also analysed for different sets of training and testing data. Table 4.6 compares the performance metrics for both the SVR and QPSO-SVR models trained on varying percentages of training data.

Table 4. 6: Model Performance with Different Training Sets

	QPSO-SVR			**SVR**		
Training Data Portion	**90%**	**80%**	**70%**	**90%**	**80%**	**70%**
MAE	0.6139	0.9286	0.9944	0.8180	1.1865	1.2555
RMSE	0.7875	1.1577	1.2824	1.1258	1.6246	1.7114
R^2	0.8481	0.7951	0.7870	0.6897	0.5965	0.6207

Improved performance is observed in both the QPSO-SVR and SVR models with increase in the size of the training set. MAE and RMSE decrease while R^2 increases as the size of the training set increases.

Chapter Five: Discussion

5.1 Introduction

This chapter discusses the results of battery cycling, feature extraction, and model evaluation.

5.2. Battery Cycling

The results obtained from cycling the cells show expected degradation trends as batteries are cycled. The batteries, being used cells, have an initial SOH lower than 100%. All the cells also have initial SOH below 80%, which shows that they are not suitable for use in EVs, which require SOH of 80% and above for optimum performance. However, the SOH of all the cells, except cell 2, is suitable for second life applications such as secondary storage. SOH levels of between 70% to 80% are sufficient for these applications (Dunn, 2022). Cell 2 cannot hold charge hence cannot be repurposed.

Capacity, energy, and SOH decreased down cycles, while DCIR increased. As a battery ages, it degrades. Degradation could be attributed to factors such as loss of electrolyte or electrode material. This decreases the ability of the battery to store charge, hence decrease in capacity. This also means that the amount of energy the battery can store decreases, hence decreasing trends in energy down cycles. SOH is obtained from capacity hence when capacity decreases so does SOH. Degradation causes increase in internal resistance of a battery. This can be attributed to growth of solid electrolyte interface layer on the battery anode. This layer has resistive properties and increases as a battery is cycled leading to increase in DCIR (Mulpuri et al., 2023).

5.3 Feature Extraction

Changes in differential capacity curves can be used to analyse degradation in cells. These changes include shifts in the position of peaks and valleys and changes in peak heights and widths. Decrease in height and shifting of peaks is usually associated with loss of lithium inventory (Biologic, 2024). This study sought to use partial discharge data for battery SOH estimation from differential capacity curves. The targeted region for feature extraction was the portion of the curve ranging from 3.6 V to 4.2 V. In the curves generated, several peaks and valleys were identified in this region from which features of interest were extracted.

When tested using correlation analysis, the selected features yielded correlation coefficients above 0.5 supporting their correlation to SOH. The Pearson correlation coefficient usually takes on values ranging from -1 to +1. Where a negative sign indicates a negative relationship and a positive sign indicates a positive relationship between variables. Although there is no fixed threshold for evaluation, generally, values between 0 and 0.3 indicate a weak relationship, values between 0.3 and 0.7 indicate a moderate relationship, and values between 0.7 and 1 indicate a strong relationship (Ratner, 2009).

5.4 Model Evaluation

The QPSO-SVR model employed in this study was first compared to a standard SVR model to confirm its robustness. From Table 4.5, it is evident that QPSO-SVR is superior to SVR in terms of performance as it produced lower RMSE and MAE values signifying greater accuracy, and a higher R^2 value signifying that it fits the data better.

The model was first trained on 80% training data. The training data was varied to see if this would have an effect on its performance. Therefore, the model was trained again using a 70% training set and 90% training set. The model showed highest accuracy and best fit when the 90% training set was used. The performance was poorest when trained on 70% training data, the smallest set. This suggests training the model on more data would increase its accuracy and fit.

Chapter Six: Conclusion

6.1 Introduction

This chapter summarises the findings of this study and provides recommendations for future research.

6.2 Conclusion

The aim of this study was to develop a method for estimating battery SOH. SOH determination is crucial in battery repurposing to ensure safety and optimum performance of batteries in second-life applications. However, the time-consuming nature of current battery testing methods inhibits the scalability of battery repurposing. This study therefore set out to develop a machine learning model to determine SOH of retired batteries using partial discharge information.

Features of interest were extracted from the portion of differential capacity curves ranging from 3.6 V to 4.2 V, which is the initial discharge segment. Using these features, a machine learning model was developed based on QPSO-SVR. The model was compared to an SVR model and it was found to perform better. Different datasets were tested on the model and the model performed best when trained on the largest data set. Hence, increase in training data could improve model performance. The results obtained demonstrate that this method can be used to estimate the SOH of retired lithium-ion batteries, although, improvements could be made to increase its accuracy.

6.3 Future Research

Ambient conditions were not included in the development of this model. Batteries operate under different ambient conditions that affect their degradation patterns and consequently SOH. These include varying temperature and humidity. Future studies could therefore incorporate different cycling conditions in their data to capture the varying degradation patterns of used cells, hence improving the ability of the model to accurately determine SOH.

References

Abhishek, J. (2024, September 11). *SVM kernels and its type.* Medium. https://medium.com/@abhishekjainindore24/svm-kernels-and-its-type-dfc3d5f2dcd8

Ahmeid, M., Muhammad, M., Lambert, S., Attidekou, P. S., & Milojevic, Z. (2022). A rapid capacity evaluation of retired electric vehicle battery modules using partial discharge test. *Journal of Energy Storage*, *50*, 104562. https://doi.org/10.1016/j.est.2022.104562

Akram, M. N., & Abdul-Kader, W. (2024). Repurposing Second-Life EV Batteries to Advance Sustainable Development: A Comprehensive review. *Batteries*, *10*(12), 452. https://doi.org/10.3390/batteries10120452

Automotive Cells Company. (2022, June 30). *Battery Cell, Module or Pack. What's the difference?* https://www.acc-emotion.com/stories/battery-cell-module-or-pack-whats-difference-infographics

Balaraman, K. (2023, July 11). EV batteries can be repurposed as grid storage to reduce battery supply chain impacts: report. *Utility Dive*. https://www.utilitydive.com/news/evbatteries-repurpose-recycle-grid-storagemicrogrid-nrdc/686200/

Battery University. (2021, October 25). *BU-409: Charging lithium-ion. https://batteryuniversity.com/article/bu-409-charging-lithium-ion*

Battery University. (2023, December 8). *BU-204: Types of lithium ion. https://batteryuniversity.com/article/bu-205-types-of-lithium-ion*

Biologic. (2024, November 18). *Investigating battery aging using Differential Capacity Analysis (DCA).* https://www.biologic.net/topics/investigating-battery-ageing-using-differential-capacity-analysis-dca/

Bollini, R. (2022, January 20). *Understanding charge-discharge curves of li-ion cells.* EVreporter.com. https://evreporter.com/understanding-charge-discharge-curves-of-li-ion-cells/

Brorein, E. (2024, February 6). *Characteristics of DC Power Source Priority Modes*. https://www.keysight.com/blogs/en/tech/bench/2024/02/06/characteristics-of-dc-power-source-priority-modes

Brorein, E. (2022, August 30). *Lithium-Ion cell charging and discharging during life cycle testing versus formation*. https://www.keysight.com/blogs/en/tech/bench/2022/08/30/lithium-ion-cell-charging-and-discharging-during-life-cycle-testing-versus-formation

Chugh, A. (2020, Dec 8). *MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Square – Which Metric is Better?* Medium. https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e

Chung, H. C. (2021). Charge and discharge profiles of repurposed LiFePO4 batteries based on the UL 1974 standard. *Scientific Data*, *8*(165). https://doi.org/10.1038/s4159702100954-3

Dunn, J. (2022, October 7). *Battery stateof health - What is it? Why is it important?* The Equation. https://blog.ucsusa.org/jessica-dunn/battery-state-of-health-what-is-itwhyis-it-important/

Dragonfly Energy. (2024, January 16). *The effect energy density has on the power of your battery.* https://dragonflyenergy.com/the-effect-energy-density-has-on-the-power-of-your-battery/

Etengoff, A. (2023, October 30). How does EV battery aging affect range and performance? EV Engineering & Infrastructure. https://www.evengineeringonline.com/how-doesevbattery-aging-affect-range-and-performance/

Farmer, A., & Watkins, E. (2023). *Managing waste batteries from electric vehicles: The case of the European Union and Japan* (Report). Institute for European Environmental Policy.

Gaia. (2024). *Electric Vehicle Battery Repurposing and Second Life* [Info sheet]. https://www.no-burn.org/wp-content/uploads/2024/06/04-Battery-Infosheet-ElectricVehicle-Battery-Repurposing-and-Second-Life.pdf

Gregory, D. (2024, June 24). Addressing the Environmental and Health Risks in Battery Manufacturing. Environment Energy Leader. https://www.environmentenergyleader.com/2024/06/addressing-theenvironmentaland-health-risks-in-batterymanufacturing/#:~:text=The%20manufacturing%20process%20generates%20hazardo us,poses%20a%20significant%20environmental%20threat.

Idelah, I. (2020, June 26). *Li-ion Cell Types.* Inventus Power. https://news.inventuspower.com/blog/li-ion-cell-types

Kang, D. H. P., Chen, M., & Ogunseitan, O. A. (2013). Potential environmental and human health impacts of rechargeable lithium batteries in electronic waste. *Environmental Science & Technology*, *47*(10), 5495–5503. https://doi.org/10.1021/es400614y

Kenya. (2022). *Kenya's Updated Nationally Determined Contribution (NDC)*. https://unfccc.int/documents/497612

Khan, M. I., Gilani, R., Hafeez, J., Ayoub, R., Zahoor, I., & Saira, G. (2024). Advantages and disadvantages of lithium-ion batteries. In *Elsevier eBooks* (pp. 47–64). https://doi.org/10.1016/b978-0-443-13338-1.00017-4

Konstantinovsky, T. (2024, November 20). *Introduction to the Savitzky-Golay Filter: A Comprehensive Guide (Using*

Python). Medium. https://medium.com/pythoneers/introduction-to-the-savitzky-golay-filter-a-comprehensive-guide-using-python-b2dd07a8e2ce

Korthauer, R. (2018). Lithium-Ion batteries: Basics and applications. In *Springer eBooks*. https://doi.org/10.1007/978-3-662-53071-9

Kumar, V. (2023, May 8). *Comprehending battery charge and discharge testing.* Infinitalab. https://infinitalab.com/blogs/analytical-tests/common-uses-of-batterychargedischarge-testing/

Lai, J., Chao, D., Wu, A., & Wang, C. (2020). Combining machine learning algorithms and an incremental capacity analysis on 18650 cell under different cycling temperature and SOC range. *E3S Web of Conferences*, *182*, 03007. https://doi.org/10.1051/e3sconf/202018203007

Large Power. (2022, July 15). *Lithium Battery Discharge Cutoff Voltage.* Linkedin. https://www.linkedin.com/pulse/lithium-battery-discharge-cutoff-voltage-

Li, Y., Liu, K., Foley, A. M., Zülke, A., Berecibar, M., Nanini-Maury, E., Van Mierlo, J., & Hoster, H. E. (2019). Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review. *Renewable and Sustainable Energy Reviews*, *113*, 109254. https://doi.org/10.1016/j.rser.2019.109254

Lin, K., Huang, C., & Sou, K. (2023). Lithium-Ion battery state of health estimation using simple regression model based on incremental capacity analysis features. *Energies*, *16*(20), 7066. https://doi.org/10.3390/en16207066

Martin, J. (2016, Aug 3). *What is battery 'end of life?'* Solar Choice. https://www.solarchoice.net.au/blog/battery-end-of-life-explained/

Mohan, A. K., Papafilippou, E., Paulik, T., Petrie, R. D., Sargood, A., Spiesshofer, N., Wang, J., Ward-Berry, J. N., & Watt, M. J. (2023, August 25). *LiFETIME: A low-cost,*

opensource solution for cell health testing for second life battery applications. MRes Team Challenge, Centre for Doctoral Training in Sensor Technologies for a Healthy and Sustainable Future, Department of Chemical Engineering & Biotechnology, University of Cambridge.

Mulpuri, S. K., Sah, B., & Kumar, P. (2023). Unraveling capacity fading in lithium-ion batteries using advanced cyclic tests: A real-world approach. *iScience*, *26*(10), 107770. https://doi.org/10.1016/j.isci.2023.107770

Noema, Y. (2022, January 6). *Top 8 most important unsupervised machine learning algorithms with python code references.* Medium. https://medium.com/imagescv/top-8-most-important-unsupervised-machine-learning-algorithms-with-python-code-references-1222393b5077

Petkovski, E., Marri, I., Cristaldi, L., & Faifer, M. (2023). State of Health estimation procedure for Lithium-Ion batteries using partial discharge data and support vector regression. *Energies*, *17*(1), 206. https://doi.org/10.3390/en17010206

Qian, L., Xuan, L., & Chen, J. (2023). Battery SOH estimation based on decision tree and improved support vector machine regression algorithm. Frontiers in Energy Research, 11. https://doi.org/10.3389/fenrg.2023.1218580

Quraishi, A., Zalani, A., Beard, R., & Mercedes, D. (2023, August 15). Lithium-ion battery fires from electric cars, bikes and scooters are on the rise. Are firefighters ready? CBS News. https://www.cbsnews.com/news/lithium-ion-battery-fires-electric-cars-bikes-scooters-firefighters/

Ratner, B. (2009). The correlation coefficient: Its values range between +1/−1, or do they? *Journal of Targeting Measurement and Analysis for Marketing*, *17*(2), 139–142. https://doi.org/10.1057/jt.2009.5

Romanchuk, J. (2023, January 28). *The four types of research design – everything you need to know.* Hubspot. https://blog.hubspot.com/marketing/types-of-research-design

Salek, F., Resalati, S., Azizi, A., Babaie, M., Henshall, P., & Morrey, D. (2024). State of Health prediction of electric vehicles' retired batteries based on First-Life historical degradation data using predictive Time-Series algorithms. Mathematics, 12(7), 1051. https://doi.org/10.3390/math12071051

Salunkhe, V. (2021, July 22). *Support vector regression.* Medium. https://medium.com/@viveksalunkhe80/support-vector-regression-652bbdae7808

Shu, X., Shen, S., Shen, J., Zhang, Y., Li, G., Chen, Z., & Liu, Y. (2021). State of health prediction of lithium-ion batteries based on machine learning: Advances and perspectives. *iScience*, *24*(11), 103265. https://doi.org/10.1016/j.isci.2021.103265

Singh, A. (2024, September 20). Types of machine learning in artificial intelligence. Digital Regenesys. https://www.digitalregenesys.com/blog/types-of-machine-learning-in-artificial-intelligence

Tomasov, M., Kajanova, M., Bracinik, P., & Motyka, D. (2019). Overview of Battery Models for Sustainable Power and Transport Applications. *Transportation Research Procedia*, *40*, 548–555. https://doi.org/10.1016/j.trpro.2019.07.079

Trinh, V., & Chung, C. (2023). Renewable energy for SDG-7 and sustainable electrical production, integration, industrial application, and globalization: Review. *Cleaner Engineering and Technology*, *15*, 100657. https://doi.org/10.1016/j.clet.2023.100657

Ufine Battery. (2024, April 22). *How to read lithium battery discharge curve and charging curve?* https://www.ufinebattery.com/blog/how-to-read-lithium-batterydischargecurve-andchargingcurve/#:~:text=Specifically%2C%20its%20discharge%20curve%20shows,lit hium%2 0battery%20will%20gradually%20decrease.

UL Research Institutes. (2021, September 14). *What are lithium-ion batteries?* https://ul.org/research-updates/what-are-lithium-ion-batteries/

Wang, Y., Li, K., Peng, P., & Chen, Z. (2022). Health diagnosis for Lithium-Ion battery by combining partial incremental capacity and deep belief network during insufficient discharge profile. *IEEE Transactions on Industrial Electronics*, *70*(11), 11242–11250. https://doi.org/10.1109/tie.2022.3224201

Wang, Z., Zeng, S., Guo, J., & Qin, T. (2018). Remaining capacity estimation of lithium-ion batteries based on the constant voltage charging profile. *PLoS ONE*, *13*(7), e0200169. https://doi.org/10.1371/journal.pone.0200169

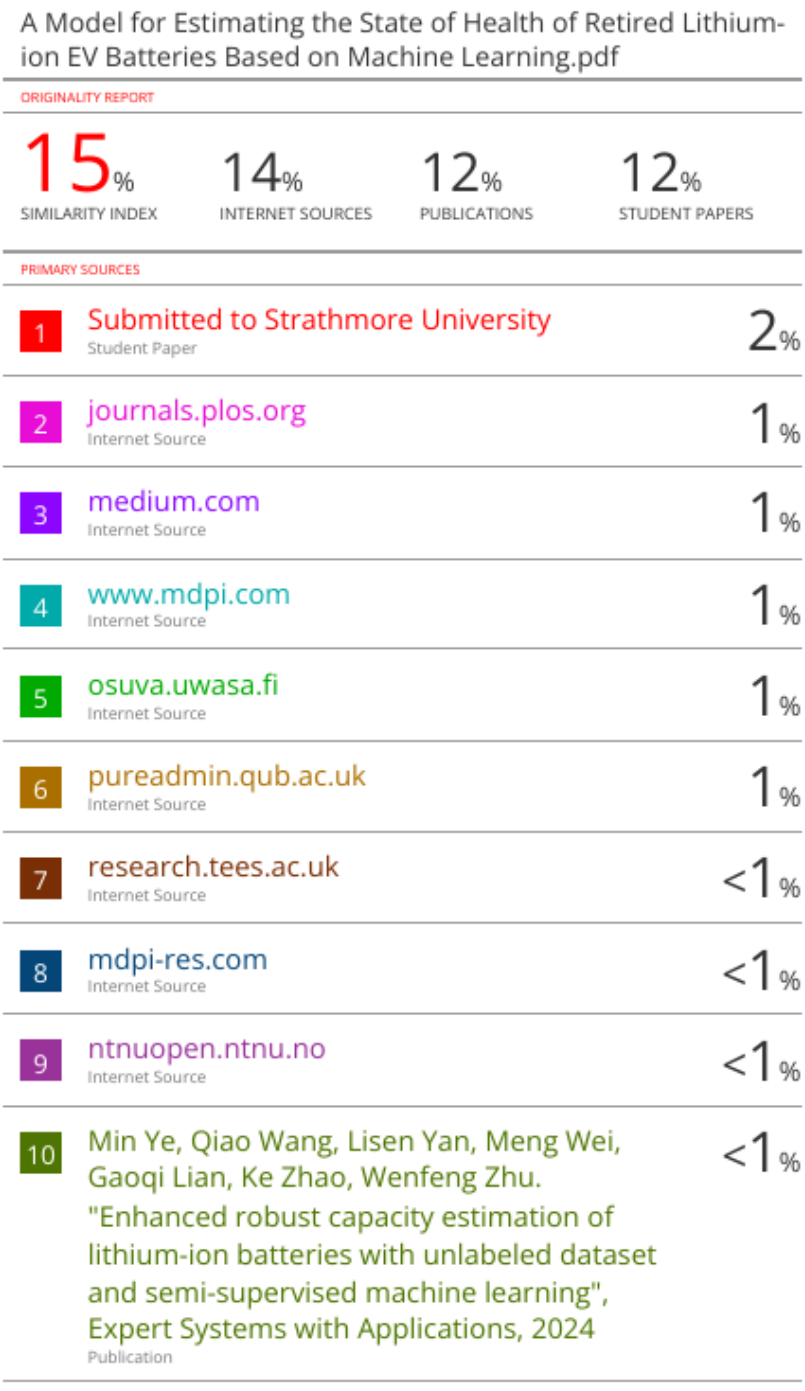
Wei, Y. (2023). Prediction of state of health of Lithium-Ion battery using Health Index Informed Attention model. *Sensors*, *23*(5), 2587. https://doi.org/10.3390/s23052587

Xu, Z., Li, H., Yazdi, M., Ouyang, K., & Peng, W. (2022). Aging Characteristics and State-of-Health Estimation of Retired Batteries: An Electrochemical Impedance Spectroscopy Perspective. *Electronics*, *11*(23), 3863. https://doi.org/10.3390/electronics11233863

Zhang, Q., Li, X., Du, Z., & Liao, Q. (2021). Aging performance characterization and state-of-health assessment of retired lithium-ion battery modules. *Journal of Energy Storage*, *40*, 102743. https://doi.org/10.1016/j.est.2021.102743

Appendices

Appendix A: Similarity Report



A Model for Estimating the State of Health of Retired Lithium-ion EV Batteries Based on Machine Learning.pdf

ORIGINALITY REPORT

15%	14%	12%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Strathmore University Student Paper	2%
2	journals.plos.org Internet Source	1%
3	medium.com Internet Source	1%
4	www.mdpi.com Internet Source	1%
5	osuva.uwasa.fi Internet Source	1%
6	pureadmin.qub.ac.uk Internet Source	1%
7	research.tees.ac.uk Internet Source	<1%
8	mdpi-res.com Internet Source	<1%
9	ntnuopen.ntnu.no Internet Source	<1%
10	Min Ye, Qiao Wang, Lisen Yan, Meng Wei, Gaoqi Lian, Ke Zhao, Wenfeng Zhu. "Enhanced robust capacity estimation of lithium-ion batteries with unlabeled dataset and semi-supervised machine learning", Expert Systems with Applications, 2024 Publication	<1%

11	Submitted to University of Ulster Student Paper	<1%
12	Submitted to Southern New Hampshire University - Continuing Education Student Paper	<1%
13	Submitted to University of California, Los Angeles Student Paper	<1%
14	Submitted to Curtin University of Technology Student Paper	<1%
15	Submitted to University of St Mark and St John Student Paper	<1%
16	Submitted to University of Technology, Sydney Student Paper	<1%
17	d21zja6o12zyp0.cloudfront.net Internet Source	<1%
18	openaccess.nhh.no Internet Source	<1%
19	research.chalmers.se Internet Source	<1%
20	Submitted to Chester College of Higher Education Student Paper	<1%
21	research-portal.uws.ac.uk Internet Source	<1%
22	pub.h-brs.de Internet Source	<1%
23	Submitted to Excelsior University Student Paper	<1%

Appendix B: Ethical Review Approval

Strathmore UNIVERSITY

14th November 2024

Ms Rugami Vanessa,
vanessa.rugami@strathmore.edu

Dear Ms Rugami,

RE: A Model for Estimating the State of Health of Retired Lithium-ion EV Batteries using Quantum Particle Swarm Optimisation - Support Vector Regression (QPSO-SVR)

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2452/24.** The approval period is from **14th November 2024 to 13th November 2025.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

Appendix C: NACOSTI License

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **121442**

Date of Issue: **06/March/2025**

RESEARCH LICENSE

This is to Certify that Miss.. Vanessa Mukami Rugami of Strathmore University, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Nairobi on the topic: A Model for Estimating the State of Health of Retired Lithium-ion EV Batteries Based on Quantum Particle Swarm Optimisation - Support Vector Regression (QPSO-SVR) for the period ending : 06/March/2026.

License No: **NACOSTI/P/25/416158**

121442

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

See overleaf for conditions

Appendix D: Codes

Plotting Differential Capacity Curves

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Upload file to Google Colab
from google.colab import files
uploaded = files.upload()  # Prompt to upload file
file_path = list(uploaded.keys())[0]  # Get the name of the uploaded file
df = pd.read_excel(file_path)

# Define unique battery IDs
battery_ids = df['Battery_ID'].unique()

# Set up subplots
num_batteries = len(battery_ids)
cols = 3  # Number of columns in subplot grid
rows = -(-num_batteries // cols)  # Ceiling division for rows
fig, axes = plt.subplots(rows, cols, figsize=(15, 10), sharex=True, sharey=True)
axes = axes.flatten()  # Flatten in case of uneven grid

# Plot each battery's dQ/dV curves
for i, battery in enumerate(battery_ids):
    ax = axes[i]
    battery_data = df[df['Battery_ID'] == battery]

    # Loop over 10 cycles
    for cycle in range(1, 11):
        cycle_data = battery_data[battery_data['Cycle Index'] == cycle]
        sns.lineplot(
            x=cycle_data['Voltage(V)'],
            y=cycle_data['dQ/dV(mAh/V)'],
            label=f'Cycle {cycle}',
            ax=ax,
            alpha=0.7
        )

    ax.set_title(f'Cell {battery}')
    ax.set_xlabel('Voltage (V)')
    ax.set_ylabel('dQ/dV (mAh/V)')
    ax.legend(loc='upper right', fontsize=8)
    ax.grid(True)

# Remove empty subplots if batteries < grid size
```

```
for j in range(i + 1, len(axes)):
    fig.delaxes(axes[j])

plt.tight_layout()
plt.show()
```

Feature Extraction

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.signal import find_peaks, peak_widths
from scipy.integrate import simpson
from scipy.interpolate import interp1d
from scipy.signal import savgol_filter
from scipy.stats import skew, kurtosis, entropy

# Load data (assuming a single CSV file with all batteries)
from google.colab import files
uploaded = files.upload()  # Prompt to upload file
file_path = list(uploaded.keys())[0]  # Get the name of the uploaded file
df = pd.read_excel(file_path)

# Filter discharge data only
df_discharge = df[df["Current(A)"] < 0]

# Initialize feature storage
features_list = []

# Iterate through each battery and cycle
for battery in df_discharge["Cell_ID"].unique():
    df_battery = df_discharge[df_discharge["Cell_ID"] == battery]

    for cycle in df_battery["Cycle Index"].unique():
        cycle_data = df_battery[df_battery["Cycle Index"] == cycle]

        # Extract voltage and dQ/dV values
        V = cycle_data["Voltage(V)"].values
        dQdV = cycle_data["dQ/dV(mAh/V)"].values

        # Filter data within the voltage range 4.2V to 3.6V
        mask = (V <= 4.2) & (V >= 3.6)
        V_filtered = V[mask]
        dQdV_filtered = dQdV[mask]
```

```
        if len(V_filtered) < 10:
            continue

        # Remove duplicate voltage values by averaging dQ/dV
        df_clean = pd.DataFrame({"Voltage": V_filtered, "dQ/dV": dQdV_filtered})
        df_unique = df_clean.groupby("Voltage", as_index=False).mean()
        V_clean = df_unique["Voltage"].values
        dQdV_clean = df_unique["dQ/dV"].values

        # Interpolation for smooth curve
        V_new = np.linspace(V_clean.min(), V_clean.max(), num=500)
        interp_func = interp1d(V_clean, dQdV_clean, kind="cubic", fill_value="extrapolate")
        dQdV_interp = interp_func(V_new)

        # Apply Savitzky-Golay filter for smoothing
        dQdV_smooth = savgol_filter(dQdV_interp, window_length=11, polyorder=3)

        # Find peaks (local maxima) and valleys (local minima)
        peaks, _ = find_peaks(dQdV_smooth)
        valleys, _ = find_peaks(-dQdV_smooth)

        # Compute peak and valley properties
        peak_voltages = V_new[peaks] if len(peaks) > 0 else [np.nan]
        peak_heights = dQdV_smooth[peaks] if len(peaks) > 0 else [np.nan]
        valley_voltages = V_new[valleys] if len(valleys) > 0 else [np.nan]
        valley_depths = dQdV_smooth[valleys] if len(valleys) > 0 else [np.nan]

        # Compute peak and valley widths
        peak_widths_vals = peak_widths(dQdV_smooth, peaks)[0] if len(peaks) > 0 else [np.nan]
        valley_widths_vals = peak_widths(-dQdV_smooth, valleys)[0] if len(valleys) > 0 else
[np.nan]

        # Compute area under dQ/dV curve
        dq_dv_integral = simpson(y=dQdV_smooth, x=V_new)

        # Compute slope at 3.7V
        target_voltage = 3.7
        if V_new.min() <= target_voltage <= V_new.max():
            slope_index = np.argmin(np.abs(V_new - target_voltage))
            slope_at_target = dQdV_smooth[slope_index]
        else:
            slope_at_target = np.nan

        # Compute voltage at maximum slope
        max_slope_index = np.argmax(np.abs(np.gradient(dQdV_smooth, V_new)))
```

```
max_slope_voltage = V_new[max_slope_index]

# Compute statistical and shape features
mean_dQdV = np.mean(dQdV_smooth)
std_dQdV = np.std(dQdV_smooth)
skewness_dQdV = skew(dQdV_smooth, nan_policy='omit')
kurtosis_dQdV = kurtosis(dQdV_smooth, nan_policy='omit')
entropy_dQdV = entropy(np.abs(dQdV_smooth) + 1e-10)
smoothness_dQdV = np.std(np.gradient(dQdV_smooth, V_new))

# Store extracted features
features = {
    "Cell": battery,
    "Cycle": cycle,
    "Num_Peaks": len(peaks),
    "Peak_Voltage_1": peak_voltages[0],
    "Peak_Height_1": peak_heights[0],
    "Peak_Width_1": peak_widths_vals[0],
    "Num_Valleys": len(valleys),
    "Valley_Voltage_1": valley_voltages[0],
    "Valley_Depth_1": valley_depths[0],
    "Valley_Width_1": valley_widths_vals[0],
    "dQdV_Integral": dq_dv_integral,
    "Slope_3.7V": slope_at_target,
    "Max_Slope_Voltage": max_slope_voltage,
    "Mean_dQdV": mean_dQdV,
    "Std_dQdV": std_dQdV,
    #"Skewness_dQdV": skewness_dQdV,
    #"Kurtosis_dQdV": kurtosis_dQdV,
   # "Entropy_dQdV": entropy_dQdV,
    #"Smoothness_dQdV": smoothness_dQdV
}

features_list.append(features)

# Plot dQ/dV curve with marked peaks and valleys
plt.figure(figsize=(8, 5))
plt.plot(V_new, dQdV_smooth, label="Smoothed dQ/dV", color="blue")
plt.scatter(peak_voltages, peak_heights, color="red", label="Peaks", marker="o")
plt.scatter(valley_voltages, valley_depths, color="green", label="Valleys", marker="x")
plt.axvline(x=3.7, color="purple", linestyle="--", label="Slope at 3.7V")
plt.xlabel("Voltage (V)")
plt.ylabel("dQ/dV (mAh/V)")
plt.title(f"Battery {battery}, Cycle {cycle}")
plt.legend()
```

```
        plt.show()

# Convert to DataFrame
features_df = pd.DataFrame(features_list)

# Save to CSV for machine learning use
features_df.to_csv("dQ_dV_features.csv", index=False)

print(features_df.head()
```

ANOVA Test

```
import pandas as pd
import numpy as np
from scipy.stats import f_oneway
from google.colab import files

# Load extracted features data
uploaded = files.upload()
file_path = list(uploaded.keys())[0]
features_df = pd.read_csv(file_path)

# List of numerical features to test
num_features = [col for col in features_df.columns if col not in ["Battery", "Cycle"]]

# One-way ANOVA
anova_battery_results = {}
for feature in num_features:
        groups = [features_df[features_df["Battery"] == b][feature] for b in
features_df["Battery"].unique()]
   f_stat, p_value = f_oneway(*groups)
   anova_battery_results[feature] = {"F-statistic": f_stat, "p-value": p_value}

# Convert results to DataFrame and save
anova_battery_df = pd.DataFrame.from_dict(anova_battery_results, orient="index")
anova_battery_df.to_csv("anova_battery.csv")

# Display results
print("One-way ANOVA Results")
display(anova_battery_df)
```

Correlation Analysis

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import files

# Upload and load data
uploaded = files.upload()
file_path = list(uploaded.keys())[0]
df = pd.read_csv(file_path)

# Perform Pearson correlation analysis
correlation_matrix = df.corr()

# Plot heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Feature Correlation Heatmap")
plt.show()

# Save the heatmap as an image file
heatmap_path = "/content/correlation_heatmap.png"
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Feature Correlation Heatmap")
plt.savefig(heatmap_path)

# Identify most correlated features with SOH
correlation_with_SOH = correlation_matrix["SOH (%)"].drop("SOH (%)")
correlated_features = correlation_with_SOH.abs().sort_values(ascending=False)

# Save the correlated features as a CSV file
correlated_features_path = "/content/top_correlated_features.csv"
correlated_features.to_csv(correlated_features_path, header=True)

# Provide download links
files.download(heatmap_path)  # Download the heatmap image
files.download(correlated_features_path)  # Download the CSV file with top correlated features

print("Top correlated features with SOH:")
print(correlated_features)
```

Model

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.svm import SVR
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from pyswarm import pso  # For QPSO optimization
from google.colab import files
from math import sqrt

# Upload file
uploaded = files.upload()
filename = list(uploaded.keys())[0]  # Get the uploaded file name

# Load dataset
df = pd.read_csv(filename)
print("Dataset Loaded:")
print(df.head())

# Assume first column is SOH, rest are features
y = df.iloc[:, 0].values
X = df.iloc[:, 1:].values

# Standardize features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Split data
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.3, random_state=42)

# Train baseline SVR model
svr = SVR(kernel='rbf')
svr.fit(X_train, y_train)
y_pred = svr.predict(X_test)

# Evaluate SVR
svr_metrics = {
    "MAE": mean_absolute_error(y_test, y_pred),
    "MSE": mean_squared_error(y_test, y_pred),
    "RMSE": sqrt(mean_squared_error(y_test, y_pred)),
    "R2": r2_score(y_test, y_pred)
}
```

```
print("SVR Performance:")
print(svr_metrics)

# Save SVR results
svr_results_df = pd.DataFrame([svr_metrics])
svr_results_file = "svr_results.csv"
svr_results_df.to_csv(svr_results_file, index=False)
print(f"SVR results saved to {svr_results_file}")

# Define QPSO-SVR optimization
def svr_loss(params):
    C, epsilon, gamma = params
    svr = SVR(kernel='rbf', C=C, epsilon=epsilon, gamma=gamma)
    scores = cross_val_score(svr, X_train, y_train, scoring='neg_mean_squared_error', cv=5)
    return -np.mean(scores)  # Minimize MSE

# Define bounds
lb = [0.1, 0.001, 0.0001]  # Lower bounds for C, epsilon, gamma
ub = [100, 1, 10]  # Upper bounds

# Run QPSO optimization
best_params, _ = pso(svr_loss, lb, ub, swarmsize=20, maxiter=50)
C_opt, epsilon_opt, gamma_opt = best_params
print("Optimized Parameters:", best_params)

# Train optimized QPSO-SVR model
qpso_svr = SVR(kernel='rbf', C=C_opt, epsilon=epsilon_opt, gamma=gamma_opt)
qpso_svr.fit(X_train, y_train)
y_pred_qpso = qpso_svr.predict(X_test)

# Evaluate QPSO-SVR
qpso_svr_metrics = {
    "MAE": mean_absolute_error(y_test, y_pred_qpso),
    "MSE": mean_squared_error(y_test, y_pred_qpso),
    "RMSE": sqrt(mean_squared_error(y_test, y_pred_qpso)),
    "R2": r2_score(y_test, y_pred_qpso)
}

print("QPSO-SVR Performance:")
print(qpso_svr_metrics)

# Save QPSO-SVR results
qpso_svr_results_df = pd.DataFrame([qpso_svr_metrics])
qpso_svr_results_file = "qpso_svr_results.csv"
qpso_svr_results_df.to_csv(qpso_svr_results_file, index=False)
```

```
print(f"QPSO-SVR results saved to {qpso_svr_results_file}")

# Plot results
plt.figure(figsize=(10,5))
plt.plot(y_test, label='True SOH')
plt.plot(y_pred, label='SVR Predicted SOH')
plt.plot(y_pred_qpso, label='QPSO-SVR Predicted SOH')
plt.legend()
plt.xlabel("Sample")
plt.ylabel("SOH")
plt.title("SOH Prediction using SVR and QPSO-SVR")
plt.show()
```