UNIVERSITY OF GHANA, LEGON

MODELLING THE STATE-OF-HEALTH OF LITHIUM BATTERIES IN ELECTRIC AND HYBRID ELECTRIC VEHICLES USING RANDOM FOREST REGRESSION

BY:

SHERIFF ALIDU (10710834)

IN PARTICULAR FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF BACHELOR OF ARTS IN MATHEMATICS AND STATISTICS, THIS THESIS IS SUBMITTED TO THE DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE, LEGON.

SEPTEMBER, 2022

DECLARATION

I, the undersigned, declare that this project work is the result of my original study and that no part of it has been submitted for another university's degree program. All borrowed items, however have been properly acknowledged.

NAME OF STUDENT

SHERIFF ALIDU (10710834)

…….. ……

Signature **Date**

SUPERVISOR'S DECLARATION

I declare that the preparation of this project work is in accordance with the laid down guidelines of the University of Ghana, Legon.

NAME OF SUPERVISOR

DR. EDWARD ACHEAMPONG

…… ……

Signature **Date**

ACKNOWLEDGEMENT

I am grateful to my Creator Allah for the gift of life to work through this with my supervisor, Dr Edward Acheampong whose unbelievable kindness and motivation stuck with me during the difficult times of this journey.

DEDICATION

I dedicate this work to my supervisor, Dr Edward Acheampong and lecturers at the Department of Mathematics and Department of Statistics & Actuarial Science for bequeathing knowledge to me over my four years at the University of Ghana.

ABSTRACT

Using data from the National Aeronautics and Space Association (NASA) after over 137 discharging cycles, this research work predicted the state-of-health of lithium ion batteries. It explained the two key terms in the battery management system (state-of-health (SOH) and state-of-charge (SOC)). For the purpose of this research at the Department of Statistics, this study specialised in data-based approaches for predicting the SOH even though past work has been done on other model based and experimental approaches. In estimating the SOH, the random forest regression was used and capacitor, a variable in the data acted as the indicator of SOH. For the dataset, the random forest regression was at its best when the number of trees grown in the forest was 1000 and the maximum number of terminal nodes possible was 100. This algorithm produced a minimal error rate of 0.02679 and R^2 value of 0.9837 which showed higher prediction accuracy, when compared with the linear regression model.

KEYWORDS : State-of-health, state-of-charge, random forest regression, machine learning, lithium battery, battery health**.**

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1.1 Background of Study

Day-in day-out, the world's reliance on batteries continue to increase. The plenty usefulness of batteries has resulted in the production of different types of batteries to carry out specific tasks. Since 1970, lithium batteries have evolved to be one of the most used and popular battery types. Known for its longevity, lithium batteries are used in machines which lasts 15 or more years. Take a cylindrical 21700 cell, which can be used in a torch could still be used in a large grid-connected energy storage installation. This versatility, the zero maintenance cost, the high energy densities it stores and the life expectancy are among several reasons why they are expensive but are the most sought after battery type in the world at the moment.

But how did we get here?

Lithium was discovered by Johan August Arfwedsonnt and Jöns Jacob Berzelius following a holistic analysis of the petalite ore ($LiAlSi_4O_{10}$) in 1817. Four years later William Thomas Brande and Sir Humphrey Davy separated the element through electrolysis of the lithium chloride (LiCl) . It did not end there. American chemist, William Sidney Harris in his studies named Electrochemical Studies in Cyclic Esters at University of California in 1958, studied the solubility of lithium in esters where he found out that the element lithium has high solubility properties and the solution could be used for batteries, which resulted in the studies and discoveries on the stability of lithium batteries.

Only about 0.0007% of the earth's crust contains lithium, which can be found in the magma deep down of rocks. The chemical compound, as established earlier, is very reactive. They react with water (but not as rigorously as sodium) and moisture (even in air).Furthermore,

lithium has 2.0×10 milligrams per kilogram in the earth crust and 1.8×10^{-1} milligrams per litre in the ocean. It is very light and can be combined with other well-known metals like aluminium, copper and manganese to produce devices.

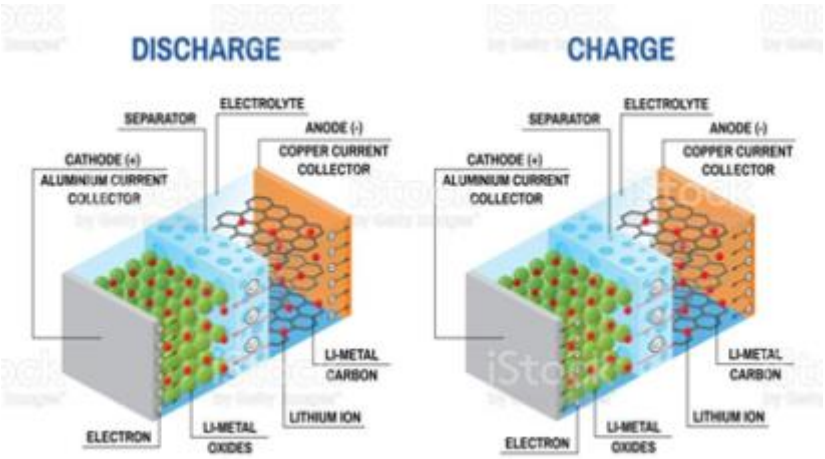

Figure 1. 1: A diagram describing the state-of-health of lithium-ion batteries at charge and discharge rates.

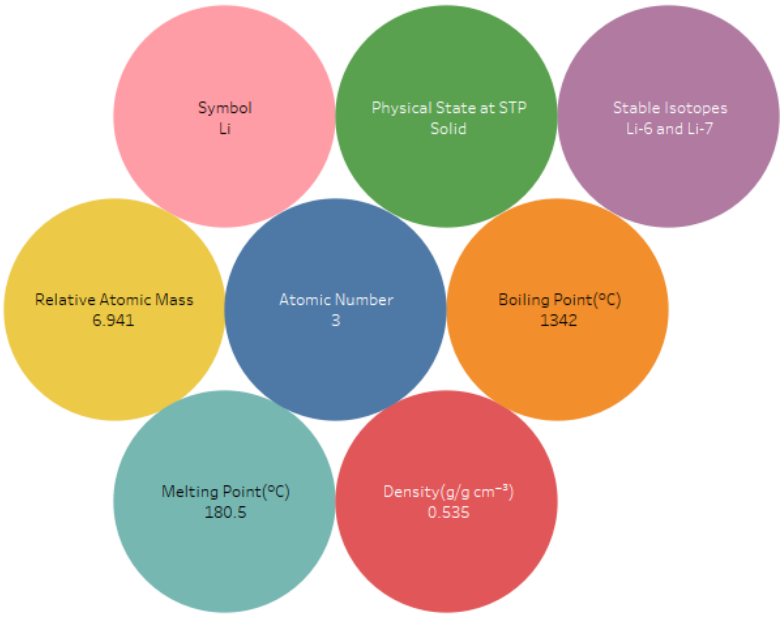

Figure 1. 2: A diagram of some properties of lithium-ion.

STP means standard temperature and pressure, normally at room temperature

1.2 Economic Importance of Lithium-Ion Battery

These properties of lithium have made it possible for innovation of lithium batteries, which has been a key stakeholder in the growth of aviation, electronic technology and health. As the world looks set to move on from fossil fuels to emissions free electrification, the emergence of lithium batteries could not be more crucial at this time. Several countries have set targets to be zero emission of carbon dioxide by 2025. Asia is the biggest manufacturer of lithium batteries in the world. About 79 per cent of the world's total production of lithium batteries are produced in China, with 5 per cent also produced in South Korea and Japan. The growing sales of electronic vehicles and consumer electronics in North America has alerted the United States to re-strategize its production and this has seen them produce about 6.5 per cent of the world's lithium batteries. This figure is expected to grow in the next few years. The European Commission and the Batteries European Association have launched a public-private partnership to encourage research into battery technology and production in countries like Germany, France, Spain and the rest of Europe. In Africa, although South Africa has joined countries in the gulf regions to accelerate efficient use of the lithium batteries, high producing lithium ore countries in the world like Ghana, Namibia, Democratic Republic of Congo, Zimbabwe and Mali have not prioritised the production of lithium batteries yet.

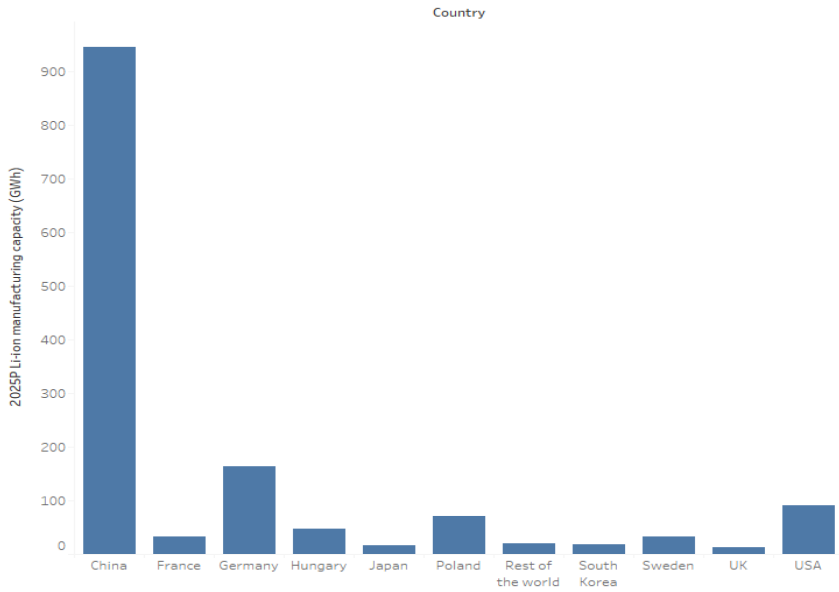

Figure 1. 3: A histogram of 2025 projections of lithium-ion manufacturing capacity in GWh. GWh means Gigawatt hours

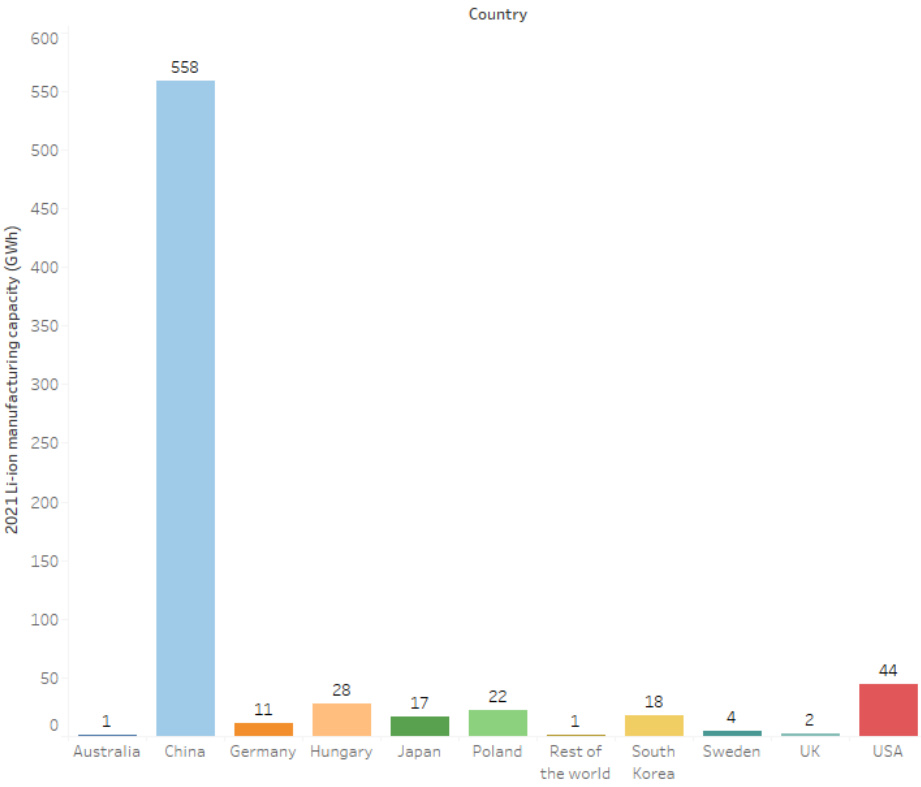

Figure 1. 4: A histogram of manufacturing capacity (in GWh) of countries in 2021.

Figure 1. 5: A global distribution of lithium-ion battery manufacturing capacity in 2021. Numbers attached to country names are in GWh.

As per data sighted by Statista, a renowned website in publishing world economic data, in 2020, companies made over $40.6billion in revenue in the production of lithium batteries, a figure that sounds staggering given the consequences of COVID-19 that year. That figure increased to about $46.3billion in 2021 and looks set to reach $53.22billion by the end of 2022. By 2026, the total revenue forecast would have reached $90billion, with Asia earning about 60-70 per cent of that sum, and Africa, unfortunately earning only just below 1 percent.

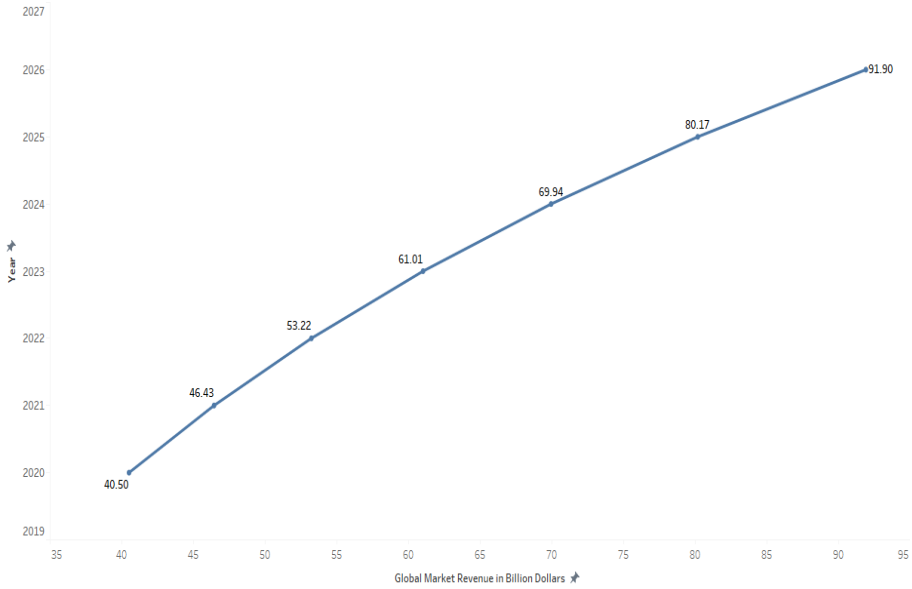

Figure 1. 6: A curve of the (projected) revenue (in $bn) companies are expected to make from the sale of lithium batteries.

1.3 Lithium in Ghana

With the recent rise of lithium exploitation in Ghana, the West African country can promote sustainable economic development through the establishment of lithium-ion battery factories. In line with the Government of Ghana's "Ghana Beyond Aid", this policy would have provided more jobs to the Ghanaian youth, substantiated the exchange rate and promoted infrastructural development. Of course, it needs mouth-watering long-term investments, but the potential returns of establishing a lithium plant in Ghana makes sense all round.

The availability of lithium, which will act as a raw material to any prospective lithium battery plant has been discovered in the Volta, Ashanti and Central regions of the country, as per the Ghana Minerals Commission, makes it very convenient for the proposed project.

Figure 1.7: A picture of lithium found in Ghana. (Asamoah, 2022)

1.4 State-of-Health Modelling

The rise in the production of lithium-ion batteries over the years has called for critical analysis into the battery management systems. Since the performance of batteries fade with its usage, manufacturers are on the look-out for relevant models that can prolong the battery lifespan while ensuring reliability. That is why modelling the state-of-health (SOH) and state-of-charge (SOC) of these batteries have evolved over the years with various models. In this work, the researcher focused on SOH and explained it as a good indicator of degradation of batteries. The SOH shows the contemporary ability of a battery to save and provide energy with respect to that at the start of its life.

In modelling the state-of-health of lithium-ion batteries, advanced research has been done already and several methods have been proposed. Here, these methods were grouped into three. The first is the experimental approach which often requires a lot of time and money. The battery's internal impedance measurement, energy level and internal resistance measurement are some ways used in this experimental approach. These experimental approaches have high accuracy rate when done in stable environments and the electrochemical impedance spectroscopy (EIS) is used to measure impedance and internal resistance.

The model-based method refers to another set of models for estimating the SOH of lithium batteries. It involves building mathematical models to address the existing problems of battery failure in the degradation process. A prominent feature of this method is time flexibility. The Kalman based filters is one of the model based methods which is made up of a number of equations that are regularly evaluated. Kalman filter (KF), extended Kalman filter (EKF), unscented Kalman filter (UKF) and dual EKF are some the approaches used but need high computational requirements. The electrochemical model is another model-based tool which relies on the electrochemical reaction within the battery and non-linear differential equations. Other models like the observers, least square and equivalent circuit methods are very good ones but tedious computationally.

When we are interested in predicting the future behaviour of batteries, the data driven approaches provide a very good option. These approaches involve the linear regression, random forests, vector machines, Gaussian processes, neural networks and other statistical based models. It relies on data to make predictions and during the process, the data is normally grouped into the test and train data, before the model used is validated. Usually, researchers use two or three of these data driven models and test its predictive power in order of importance using the R^2, mean absolute error, maximum absolute error and the mean squared error.

The remainder of this work is arranged as follows: chapter II involves a literature review of past works regarding SOH modelling, chapter III is a methodology of the process to be used, chapter IV involves thorough analysis of results and lastly, chapter V is a conclusion of the work done.

CHAPTER TWO

LITERATURE REVIEW

2.1 Background

The relevance of lithium-ion batteries has over the decade created awareness for incredible research to be done in predicting the state-of-charge (SOC) and state-of-health (SOH) of lithium-ion batteries. SOC and SOH are the most relevant measures to study in the battery management system. To start with, there has not been a widely accepted unique definition for state-of-health and state-of-charge but there have been closely related ones. Plett (2015) explains SOC of a battery as the representation of the ratio of its current available capacity to its total capacity. The battery's capacity in its contemporary condition as in relation to the capacity in its fully charged state is what SOC essentially refers to (equivalent of a fuel gauge) (Conduit et al., 2020). If we consider a battery pack, we can say the SOC is the amount of power still in the pack to deliver battery power. Simple SOC indications depend on an invariant cell performance model to produce their output and fail to consider how a pack is altering with time. (Tulsyan et al., 2016).Mathematically, we represent the SOC as

$$\frac{C_{curr}}{C_{full}} \times 100\%, \qquad (2.1)$$

where C_{curr} is the capacity of battery at present-day state and C_{full} is the battery capacity at fully charged state.

The SOH of a battery is another key performance indicator in the battery management system. According to Zhou et al. (2018), SOH has four definitions:

1. In order to determine the battery SOH based on the battery's remaining power, its definition can be provided by

$$SOH = \frac{\text{Qaged}}{\text{Qnew}} \times 100\%, \qquad (2.2)$$

where Q_{aged} is the present largest power of battery and Q_{new} is the beginning power.

2. From the perspective of starting power to define SOH, the expression is given by

$$SOH = \frac{\text{CCAocmp} - \text{CCAmin}}{\text{CCAnew} - \text{CCAmin}} \times 100\%, \qquad (2.3)$$

where CCA_{ocmp} is the real-time starting power, CCA_{min} is the needed minimum starting power and CCA_{new} is the predicted starting power released by battery when the capacity of the battery in its charged state with respect to its fully charged state is 100%.

3. The SOH, using impedance is represented by

$$SOH = \frac{R_i}{R_o} \times 100\%, \qquad (2.4)$$

where R_o is the starting impedance and R_i is the ith measurement of impedance that varies with charging and discharging cycles.

4. From the battery capacity point of view, SOH is expressed as

$$SOH = \frac{\text{Ci}}{\text{Co}} \times 100\% \qquad (2.5)$$

, where C_i is the ith capacitance measurement in time and C_O is the initial value.

As the battery ages, its performance begins to degrade with time. This degradation is often related to the chemistry of the battery. Initially, the battery electrical efficiency is decreased when a solid electrolyte interface (SEI) layer forms. Due to this, the battery's high-frequency resistance rises, decreasing its ability to provide the greatest amount of power (Troltzsch, 2006). The continuous loss of battery will grant the vehicle ineffective in the long run. The formulas seen in the definition above inform us that impedance and capacity provide a way for researchers to estimate the SOH of batteries. In this text, the methods of estimating the SOH

of a battery are grouped into three: the experimental method, the model-based approaches and the data-driven method (machine learning approach).

2.1 The Experimental Approach

Often, the approaches under this require a lot of time and highly costly. Due to this, a lot of researchers often opt for other less expensive approaches. Below are some of the experimental methods.

2.1.1 The Battery's Internal Impedance Measurement

Impedance is the combination of internal resistance and reactance. Electrical resistance, also known as ohmic resistance, and ionic resistance make up a battery's internal resistance. Ionic resistance is a measurement of the resistance to current flow caused by internal factors such as electrode surface area and electrolyte conductivity, whereas electrical resistance is a measure of the resistance to current flow in an electrical circuit. When used as a trending technique, internal ohmic values (AC resistance) can serve to show the SOH of a battery being tested. Although internal resistance varies between different manufacturers and types, the relationship between impedance and capacity is clear. The term "battery capacity" (AH) refers to the amount of current that can be drawn from a battery while it can still power a load up to the point at which the voltage of each cell drops below a specific threshold. A battery's nominal capacity, for instance, is 60 AH if it can inject 6 amps for 10 hours without its voltage dropping below its critical margin. (Abdi et al., 2017). Because lower internal resistance generally indicates higher capacity, an increase in internal resistance over time can indicate declining battery capacity or degradation. Measuring the trend over time will reveal which cells are the most vulnerable or the least strong. Normally, we use the Electrochemical Impedance Spectroscopy (EIS) to measure the battery's impedance The EIS uses sinusoidal AC current to measure the output voltage. As an advantage, the EIS gives accurate measurements of the

ageing occurrence in batteries. In Cuadras & Ovejas (2018), the authors used EIS to find that the two main occurrences of battery ageing. They found that one of them happened due to the positive electrode transfer charge and the other, induced by lithium ion transport across the solid electrolyte interphase (SEI) layer. Also Vetter et al. (2005) gave instances where EIS was a specialized test technique for non-destructive continuous monitoring of the effects of cycling and storage under field working circumstances on the life-time of Li-ion batteries. This way of measuring impedance provide accurate measurements, except that it needs environmental stability and takes time.

2.1.2 The Battery's Internal Resistance Measurement

As another indicator for estimating the battery SOH, the internal resistance is heavily affected by ageing and degradation. In calculating internal resistance, one could use the EIS where internal resistance is calculated from impedance. Another way of measuring internal resistance is using the calorimeter in a joule's law approach. Since heat is developed by internal resistance, the internal resistance can be shown by heat loss during the cell operation. In using the calorimeter to measure chemical reactions, the chemical reactions that occur may add to the heat or absorb it. (Conway, 1984).The reversible heat effect is wiped out when a battery is cycled symmetrically (equal quantities of charge and discharge) around a particular SOC, and only Joule's equation influences heat released by the battery. If this is the case, joule's law may be used to measure internal resistance (Schweiger et al., 2010). Furthermore, another way of measuring internal resistance is using the current pulse. Similar to the two approaches discussed already, the current pulse mostly requires stable environment and takes longer so it is mostly done in laboratories. It uses the Ohm's law to calculate the drop in voltage and hence estimates the internal resistance using

$$R = \frac{OCV - V}{I}$$

, where OCV is the open circuit voltage, V is the voltage and I is the current applied.

2.1.3 The Battery's Energy Level

As battery ages, its capacity falls. This approach allows us to estimate SOH by measuring the decrease in capacity with time. Li el al. (2019) studied the capacity as the voltage decreased for each of the cycles, which resulted in them having a maximum test error of 4%. Also, at 25 °C and 40 °C, Xiong et al. (2019) tested 800 cycles to estimate the SOH of two different cells where the capacity estimates had an error rate less than 1.5%. It was also established that in repeated trials, 95% of all the cases will contain the true mean which is between 0- 20 cycles. This way of estimating SOH is only limited to laboratories where the battery is tested until its end of life.

In addition to the three above, some researchers have discovered the Incremental Capacity Analysis (ICA) and Differential Voltage Analysis (DVA) methods but both demand time. For example, Stroe & Schaltz (2020) used ICA to tie the capacity fading of an electric vehicle lithium-ion battery to distinct peaks and valleys and their related voltage levels based on findings from an 11-month calendar ageing test at 35 °C and 50 percent state-of-charge. Ultrasonic inspection has shown to be useful in detecting interior flaws in batteries because of its age, using a wave generating test bench (D'Orazio et al., 2008). Destructive methods done in the laboratory use X-ray diffraction to analyze changes in structure within the battery.(Kostecki & McLarnon , 2003)

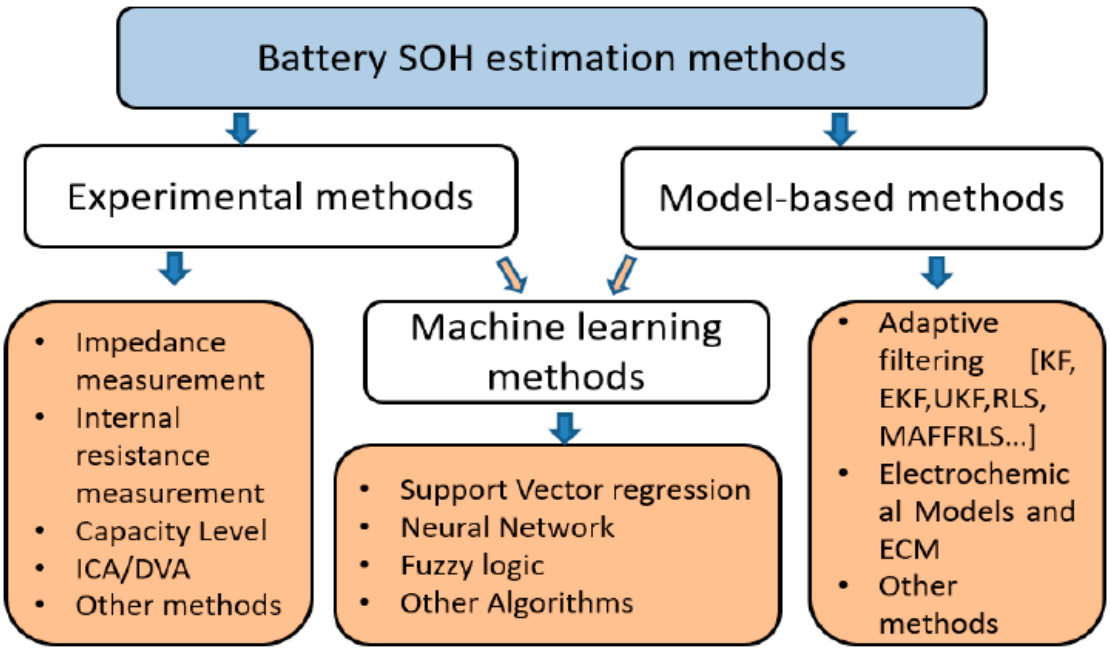

Figure 2. 1: The various methods in SOH estimation. (Noura et al., 2020)

2.2 The Model Based Methods

Here, we use knowledge acquired beforehand to build mathematical models that describe the failure of the battery in its degradation process. The auto industry has huge interests in these methods due to its time flexibility. We discuss some of the main methods of model based below.

2.2.1 The Kalman Based Filters

The Kalman filter is made up of a series of equations that are periodically evaluated as the system runs. These methods involve estimating the battery states and parameters through adaptive filtering. Kalman filter (KF), extended Kalman filter (EKF), unscented Kalman filter (UKF) and dual EKF are some the approaches used (Noura et al., 2020). In Zhu et al. (2019), the authors used fractional order adaptive approach to estimate battery's SOC. Also, Cho & Kim (2011) covered the use of an extended Kalman filter (EKF) in conjunction with a per-unit (p.u.) system to identify acceptable battery model variables for the prediction of a lithium-ion

deteriorated battery's SOC and SOH. The dual extended Kalman filter was used, in (Omariba et al., 2020) and it showed that the dual EKF produces more accurate and resilient state estimate results. Again, Omariba et al. (2020) identified the battery parameters and then described the influence of vibration and temperature dynamics in battery state estimation. Moreover, in (Chen et al., 2014), the authors estimated the battery's internal resistance for SOH prediction based on the battery state and estimation for the UKF process. Significantly, the Kalman filter methods automatically gives a changing estimation error bounds on these estimates. The authors undertook this method to give energertic performance from the battery pack, without caution of bringing destruction by over-charge or over-discharge (Plett, 2004). Generally, the Kalman based methods need high performances controllers and improved versions like EKF and UKF are computationally tedious.

2.2.2 The Electrochemical models

This model is mostly based on the electrochemical reaction process that occurs within a lithium-ion battery. Even though they are reliant on non-linear differential equations, these models can be simplified and joined with adaptive filtering. The electrochemical parameters are identified for SOH estimation. In (Prasad & Rahn, 2013), the electrochemical model was streamlined before internal resistance and time of diffusion, the two SOH indicators were identified applying online recursive parameter identification. Also, Peng & Pedram (2006) estimates the remaining capacity of SOC by taking temperature and cycling age effects into account. The same technique can be used to calculate the highest capacity of a battery for SOH estimation. Generally the EM has a good theoretical foundation and can identify the precise internal electrochemical reaction mechanism and reaction intensity in the ageing phase of the battery. It properly characterizes the movement law of lithium ions as well as the changing tendency of agile chemicals in positive and negative electrodes at various SOH locations. (Yao

et al. 2021a). The main disadvantage of the EM models are its complex equations, leading to sometime poor prediction accuracy.

2.2.3 The Equivalent Circuit Model.

This method simply means formulating a mathematical model to describe the capacity fading in the lithium-ion battery over a period of time. The ECM is built on the basis of electrotechnics; the battery as a black box, as per the input-output relating to the building of electrical components. (Yao et al. 2021b). The most common ECMs are found in Figure 2.2.

The Rint model defines the OCV by exploring an ideal voltage source. The resistance and OCV are functions of SOC, SOH and temperature. The Rint model is normally convenient to use as the battery characteristics are linearly transformed. Although it is easy to use, its error margin is large.

In the Thevenin model, a parallel RC connection is done in series based on the Rint model. As shown in (b) below, it mainly consists of the OCV, internal resistances and the equivalent capacitances. U_o, the terminal voltage and ohmic and polarization resistance represent the two internal resistance labels. Also, during charging and discharging mechanism, the equivalent capacitance is used to explain the transient response. (He et al., 2011)

Also, the PNVG model is built on a Thevenin model by connecting a series of capacitance to explain the varying OCV in the time of the load current. It is shown in (c) below. Liu et al. (2018a) merged PNGV model and the KF to find model parameters under SOH and SOC at charge and discharge rates, measured the OCV and SOC before comparing them with the battery's internal resistance method. Also in (Liu et al., 2018), it was discovered that PNGV could better describe the characteristics of lithium-ion batteries than the battery's internal resistance measurement method (an experimental approach) in the charging and discharging mechanisms.

Designed by the popular SAFT company, the RC model has a very good industry application. Consisting of two capacitors and three resistors, the RC model also has OCV and load voltage. Its bulk capacitor allows SOC to be estimated through the voltage across the bulk capacitor is determined. Terminal resistor, capacitor resistor and end resistor are the three resistors involved in the RC model, as shown in (d) below. For battery with non-linear features, the multistage RC model yields high prediction accuracy. In (He et al., 2012) the authors found out that a second-order RC model is better than a first-order RC model due to complex calculations surrounding the latter. The Thevenin model can also be improved by the multistage RC model through the addition of several RC series modules but the addition of SOH and SOC parameters make recognition of model parameters. Thus, in spite of obtaining a high accuracy level and robust models, it is difficult to apply this in real time (Yao et al., 2021c). There is also the hysteresis RC battery model, which decreases estimated errors as a result of the hysteresis effect. (Chen et al., 2022)

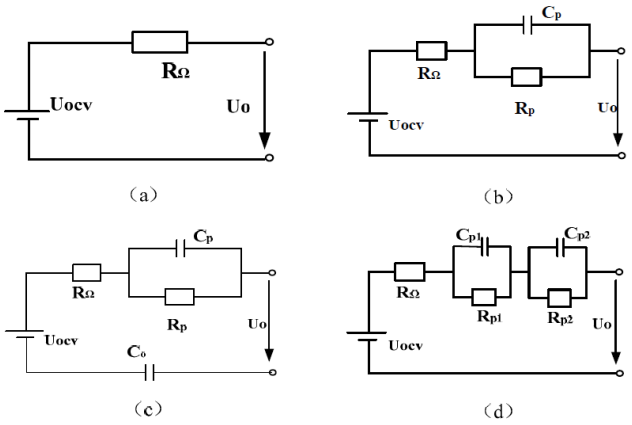

Figure 2. 2: The various types of ECM models (a) Rint Model ; (b) Thevenin model ; (c) partnership for a new generation of vehicles (PNGV); (d) multi-stage RC model.

(Yao et al., 2021)

2.2.4 Least Square method

This is another adaptive filtering method. In recent times, particular attention has been given to the recursive least squares (RLS). This technique estimates clearly relevant parameters like the OCV for SOC and internal resistance for SOH. In (Rijanto et al., 2017), a new RLS method was used. In the process, a particle swarm algorithm was used to determine system parameters before an urban dynamometer driving schedule (UDDS) evaluated that multiple adaptive forgetting factors (MAFFs) is a better algorithm than multiple fixed forgetting factors (MFFFs) for RLS based on SOC and the internal resistance's mean square error (MSE).

2.2. 5 Observers

The observer is another adaptive filtering method. Chen et al.(2022) saw the accuracy of the battery SOC estimation improved after the introduction of a hysteresis RC battery model which reduced error estimate also. When two observer methods were compared, terminal sliding mode observer (TSMO) resulted in a higher estimation accuracy and quicker response speed than with the Luenberger observer-based method due to hysteresis. The authors of (Ning et al., 2018) used adaptive sliding method for SOH by following the derivative of OCV online estimate. Using UDDS driving cycle, the authors discovered accuracy of online parameter estimates, less than 2% of absolute error SOC and a high chance of convergence of the capacity of lithium-ion battery to its actual error rate with small capacity. To add more, in (Kim, 2010), the resistance degradation and capacity decline was estimated using a dual-sliding mode observer. This observer has two ways of estimating model parameters: the fast-paced time-varying observer and the slow-paced time-varying observer. While the former estimates SOC and terminal voltage, and polarization effects, the latter predicts the SOH in terms of the capacity decline and resistance degradation. The Lyapunov equation was used to verify convergence of the dual –sliding mode observer and an error rate less than 3% was recorded when the fast-paced time-varying observer was used to estimate SOC. The dual-sliding mode

observer is also easy to use and thus have robust ways of reducing error rates. Like the ECM models, the observers are robust against variability in temperature and high level of accuracy is always sought for.

Other model-based methods include the use of optimization to recognise parameters that are SOH indicators. Chen et al. (2013) used a resistance-capacitance circuit model to generate data to study the superb correlation between diffusion capacitance and SOH of a lithium-ion battery. Then, an optimization algorithm, genetic algorithm (GA), is used to estimate the diffusion capacitance of a battery ECM. The use of optimization requires high computational time. (Chen et al., 2013)

2.3 The Data Driven Methods (Machine Learning Methods)

Often times, we are interested in predicting the future of batteries. For instance, how long an EV can go, the best behaviour of the EV battery, SOC over one charge/discharge or SOH after several charge/discharge cycles associated with the lithium-ion batteries. In order to address these questions, machine learning models use current battery states to predict future occurrences, even without domain knowledge. Although there is no perfect model, it is important to (a) select the best machine learning models from the lot and (b) validate the model. In order to validate the model, we train about 70-80% of the data and the error rate of the remaining data (20-30%), unseen to the model, known as test error, provides a benchmark for our prediction accuracy. In choosing the best models, we choose the ones with low test error rate. The principle of cross-validation, where datasets are divided into any arbitrary number of sets (typically 5 0r 10) can be used to validate the model too. In the validation set approach, data is randomly divided into training and test set, fit on the training set and the fitted model is used to predict the error rate, bias and standard deviation on the parameters. The leave-one-out

cross-validation (LOOCV), k-fold cross-validation and the validation set approach are cross-validation approaches with only few distinct features separating them.

2.3.1 Battery Parameters

For successful design, implementation and prediction of battery features, it is important to understand the parameters of the battery. Usually, we choose to ignore some parameters depending on what we want and for purposes of simplification. In machine learning, battery inputs/parameters are grouped into categorical, continuous and integer variables. Categorical variables take on qualitative parameters like lithium-ion and lead acid and cannot be grouped into a list. As for integer variables, they are basically the number of cycles during the charge/discharge processes while continuous variables can have any value such as but not limited to voltage, load, current measured etc.

Also, short timescale over single charge/discharge cycle to study SOC and long timescale over several charge/discharge cycles are the two classes of outputs. One charge/discharge cycle predicts the battery's evolution. Some of the things predicted include state-of-charge, current and the concentration and level of degradation in battery. By following closely, any point in the battery can be addressed and estimated but errors might occur when there are too many cycles. As for the long timescale over several charges, the battery's progression from the identical place cycle after cycle throughout a lot cycles is predicted. The technique can be easily implemented throughout finitely many cycles covering the battery's whole lifespan, but it is not possible to be done in the process of a specific cycle but can only begin from and transform to a certain determined end throughout the process of the cycle, such as when completely charged. (Conduit et al., 2020)

In predicting SOH, remaining user life (RUL), SOC and error rates, Conduit et al. (2020) used sample data to test several machine learning techniques: Neural network, support-vector

machine (SVM), Gaussian/Bayesian, Regression, Random forest. The authors used voltage, current, temperature, capacity, power and geometry as variables get their needed results.

Table 2 | Summary of recent work on machine learning for battery state predictions

Ref	Method						Feature set inputs								Output			
	Neural network	Support-vector machine	Gaussian/ Bayesian	Regression	Random forest/ tree	Kalman filter	Voltage	Current	Temperature	Cycle number	Capacity	Power	Geometry	P2D model parameters	SOC	SOH	RUL	Error (%)
3				x			x	x		x	x						x	9.1
4		x					x		x	x	x					x		6.4
44			x				x	x		x							x	8.7
45	x							x					x			x		7
46	x						x	x	x			x			x			0.1
47	x						x	x	x						x			2.4
48	x						x	x	x						x			1.3
49	x	x		x	x		x										x	3.3
50	x				x		x	x		x					x			13
51						x	x	x	x						x			3
52	x	x	x	x			x	x	x								x	6.7
53	x						x	x		x							x	0.6
54			x				x	x	x						x			0.8
55		x					x	x	x						x			12.2
56	x						x	x	x						x			3.8
57	x						x	x			x				x			10
58	x						x	x							x			1.7
59	x						x	x	x						x			3
60	x						x			x						x		5
61		x	x				x	x			x				x			2.1
62	x	x		x			x	x		x	x					x		1.6
63			x				x	x		x	x						x	3.2
64	x						x			x	x						x	1
65		x					x	x	x	x		x			x			0.2
66				x			x	x		x							x	0.2
67	x					x	x	x							x			1.5
68	x						x	x								x		5
69	x				x									x	x			5

Figure 2. 3: A previous work on battery health prediction using data based methods. (Ng et al., 2020)

Below, we discuss some of the machine learning models that have been used in predicting battery states in the past.

2.3.2 Linear Regression

This is the simplest possible data-driven method for calculating the SOH of lithium-ion batteries. It takes the form $y = mx + c$, where y is the output, x is the vector of input parameters, m is the vector of fitting parameters corresponding to the gradients, and c is the shifting fitting parameter. A function in the form of mean square error is commonly utilized to select the fitting parameters. The model is very fast and easy to use and doesn't only work for

a straight line. Many batteries show non-linear behaviours and as such, these behaviours are captured by advancing the method to be non-linear (comprising of quadratic, cubic and order polynomials) using Taylor's expansion. (Severson et al., 2019) used a linear regression of nine battery parameters to predict the remaining user life of lithium iron phosphate and graphite batteries after 100 charging and discharging processes.

2.3.3 Vector Machines

Considered as one of the best black box tools, the support vector machine (SVM) was developed by computer scientists in the 1990s and its applications continue to gain popularity. The SVM is a non-linear method that provides accurate predictions and estimates in the case of few data samples. Originally, the SVM is characterised by very small number of unknown scattered parameters (Yao et al., 2021). However, Klass et al. (2012) thinks SVM is a robust method that can handle large amount of real life battery data. Nuhic et al. (2013) predicted the probability density function, SOH and other battery states using the SVM algorithm after training the payload data. Also, SVM models were trained to encapsulate the features of battery charging curves at various SOHs in Feng et al. (2019). Two commercial lithium-ion batteries (Li(NiCoMn)1/3O2 cathode and graphite anode) were trained, validated and tested using SVM and with an error rate of SOH less than 2% in 80% of the cases and less than 3% in 95% of the cases. SVM has been combined with principal component analysis (PCA) to make full use of the data, prevent noise bad points and overcome data untunefulness. The method made use of several performance degradation parameters that show the battery states. (Pan et al., 2020) proposed an intelligent predictor for SOH of battery based on sample entropy characteristic of discharge voltage. The predictor as set-up imbibing SVM and relevance vector machine (RVM). The study also found out that RVM was a better predictor of accurate SOH results than the SVM.

Based on sparse Bayesian theory, the RVM is a rejuvenated machine learning algorithm. Unlike SVM, its reliance on kernel function and penalty factor selection is minimal. The RVM also adjusts parameters and controls overfitting and underfitting, making computations of parameters very convenient. In (Dong et al., 2013), a predictor for battery capacity was set-up. This predictor was made up of a conditional three-parameter capacity fading model and RVM. The RVM was used to generate relevance vectors that was useful to finding representative training vectors comprising the relevance vectors' cycles and predictive values at the relevance vectors' cycles. In order to fit predictive values at cycles of relevance vectors, a restrictive three-parameter capacity fading approach was created. The remaining useful life of lithium-ion batteries was estimated by extrapolating the restrictive three-parameter capacity fading approach to a failure point. Three case studies were carried out to affirm the newly created method. The results demonstrated that the developed method could forecast the future health of lithium-ion batteries. For the online application of the RVM model, a prediction method of battery SOH by Zhang et al. (2016) based on an optimized RVM model was proposed. It decreases the quantity of data in online training samples when compared to traditional RVM. It increased the efficiency of parameter identification, introducing a fresh concept for embedded computation of battery SOH prediction.

2.3.4 Gaussian Processes (GP)

GP is a stochastic approach that provides a probability distribution of possible outcomes. It is given by $f(x) = N\big(\mu(x), \sigma^2(x)\big)$. $f(x)$ is normally distributed having mean $\mu(x)$ and variance $\sigma^2(x)$.

Practically, the $\mu(x)$ is used as a point estimate for the function output, with the $\sigma^2(x)$ translating into uncertain bounds regarding this estimate. Whereas $\sigma^2(x)$ corresponds to the model's uncertainty (and thus varies depending on the earlier and the local data complexity), it is useful information because it allows us to quantify the uncertainty associated

with the prediction. (Girard & Murray-Smith, 2005). Despite the absence of a cost function, the covariance contains a statistical prior distribution, which is commonly assumed to be normally distributed in the range of the expected point from the training set. As a result, the whole training set must be saved as the model's foundation. Once the input variables are provided, it measures the fundamental fitting functions' probability distribution, which are typically Gaussian distributions, and the training set at run-time. (Sahinoglu et al., 2017) estimated the SOC of Li-ion batteries using Gaussian process regression. As inputs, the model takes battery parameters such as temperature, voltage and current. The Gaussian processes outperformed SVM and neural network predictions for SOC within 0.8 percent. Li & Wang (2019) used four datasets from NASA repository to demonstrate the predicting ability and efficacy of battery fading using Gaussian regression. Robust and reliable results of SOH were obtained using the proposed model. The GP method generally is very certain in its predictions near to known training set, but uncertain when the predictions farther away the training set is made or when the set has too much noise. This implies the GP approach is cost prohibitive.

2.3.5 Random Forests

A random forest is a collection of broad sweeping classification trees, each trained with data chosen at random. The split of each layer is frequently employed to optimize the variance reduction of the leftover training data. A new request is sent down the trees, yielding an ensemble of predictions that are averaged to give the estimated value plus an uncertainty. A successful application of random forests was done by Nikolakopoulos et al. (2017) when they predicted the RUL of a lithium-ion battery. The authors' goal was to extend the flying time window by concentrating on batteries in aircraft without crew. They discovered that a random forest with only voltage variation as input presented a pretty standard estimation error in the remaining user life of 3.3% , surpassing linear models, SVMs, and ANN. Another example is

indicated in Mawonou et al. (2021), where a random forest approach was used for SOH estimation within a prediction error of 1.27%. Variables such as distance, speed, temperature, charging power, and others are included in the data. Another approach in random forest regression, purely dependent on signals like time, voltage and current measured was used to predict the battery states of lithium nickel manganese cobalt with various ageing trends in (Li et al., 2018) . The proposed method needed very low computations and recorded a root MSE less than 1.3%. As for Chen et al.(2018), they discovered that their novel random forest approach could produce more accurate results regarding the health of a lithium-ion battery than least squares support vector machine (LSSVM) and SVM after using the partial charging voltage dataset in constant-current (CC) configuration and current value in constant voltage (CV) configuration. This method, however, didn't factor in the effects of temperature changes to their method.

2.3.6 Neural Networks

Probably the often applied machine learning approach in predicting battery states, neural networks became very popular in in the 1980s. They were the talk of town and the inspiration for the famous Neural Information Processing Systems meetings (NeurIPS, formerly known as NIPS). Then a platform was created where statisticians, mathematicians and machine learners met to improve algorithms and substantiated the methods. It didn't take long before boosting, SVMs and random forests took over from neural network because they were more convenient and less complex.

All this while, an important group of neural network admirers were working hard on new computer architecture and datasets. Deep learning revived around 2010 with new architectures, extra bells and whistles, and a string of success stories on some specialized issues such as image and video categorization, text and speech modelling. Many experts feel that the primary

availability of ever-larger training datasets is the cause behind these accomplishments, enabled by the widespread use of digitisation in research and industry. (James et al., 2021)

In estimating battery states, long short term memory (LSTM), the single layer, multi-layer, recurrent, artificial and convolutional neural networks have been proposed. In (Zhu et al., 2021), the article propounded a time convolutional memory model for SOH estimation. A combination of convolutional neural networks (CNN), LSTM, exponential data and smoothing was used to build the new model. Zahid et al. (2018) demonstrated a generic neural network model that could be applied to the three battery groups. This model took voltage, current, power losses, and power into account to forecast SOC to within 0.1 percent. This was a significant strategy because it allowed knowledge from one battery system to guide the behaviour of less well-studied systems, as well as provided suggestions for future prospective battery families. The time convolutional data was compared with the back propagation(BP) model, Grey model-based neural networks, CNN, LSTM using data from the National Aeronautics and Space Association(NASA), and results confirming the new proposed model was better than the ones before it, after recording a prediction error of 3.79% and average error of 0.143%. Also, Yang et al. (2017) proposed a BP model for calculating the SOH. The model had a low cost of calculation as it only took 21milliseconds for training the neural network, was easy to understand and didn't need a high memory.

The Artificial Neural Network (ANN) is made up of numerous neurons that follow particular principles and has a basic structure as well as a good learning capacity. ANN can attempt any non-linear function by adding more hidden layers and neurons. In Simeone et al. (2020), the SOH classification was modelled by ANN and cells were classified into safe, latent risk, and unsafe. In this regard, the simulation-derived deformation characteristics were fed into an ANN pattern classifier that had been empirically trained to predict the cells' SOH inside the module. (Fleischer et al., 2013) also combined a system of both fuzzy logic model and ANN into an

adaptive neuro-fuzzy inference system (ANFIS) to predict state-of-available power (SOAP), where SOC and SOH were input parameters. The article (Xia & Qahouq, 2021) learnt that due to the non-linear relationship between capacity decline and minimum impedence, ANN was deployed to study the relationship between them and estimator to predict the capacity of battery which in turn calculated the SOH. In Naha et al. (2020), battery parameters (temperature, current and voltage) were recorded after 400 cycles and trained to predict the SOH. The method didn't necessarily need full charging data, required less training data and less than 1% prediction error was achieved. The special method do not require prior history of recent SOH, SOC

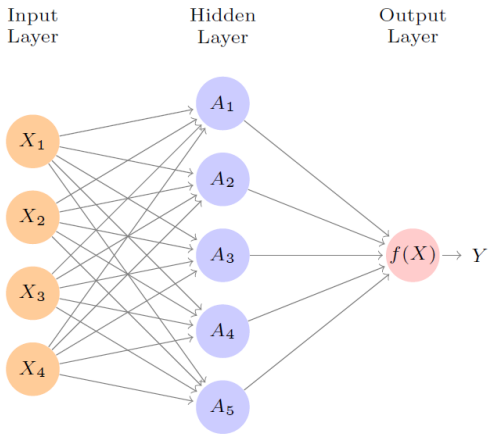

Figure 2. 4: A single hidden layer neural network. The hidden layer calculates non-linear transformations of linear combinations of $\boldsymbol{X_k}$, where k = 1,2,…,n. The $\boldsymbol{f(X)}$ is a linear model that takes $\boldsymbol{A_k}$ as inputs. (James et al., 2020)

vs. OCV relationship and complete SOC figure. To add more, Sarmah et al. (2020) built an ANN model to show the effect of stress on battery capacity, with capacity decline represent SOH estimation in real time. The model had an MSE of 0.000309 and R- squared of 0.849687, implying confidence in the accuracy of our prediction. The ANN was tested or validated against a simple linear regression, which also produced a result not very far from what had been

produced earlier, after MSE being 0.000438 and R-squared 0.819347. A collection of five ANNs have been used to tackle complications in regression and classification in (Bonfitto, 2020), where a combination estimation of SOC and SOH in EVs were proposed after there was a belief that SOC relied on the recognition of ageing of battery and SOH relied on the characteristics of SOC in closed-loop recursion. The model obtained an R-squared of 0.97.

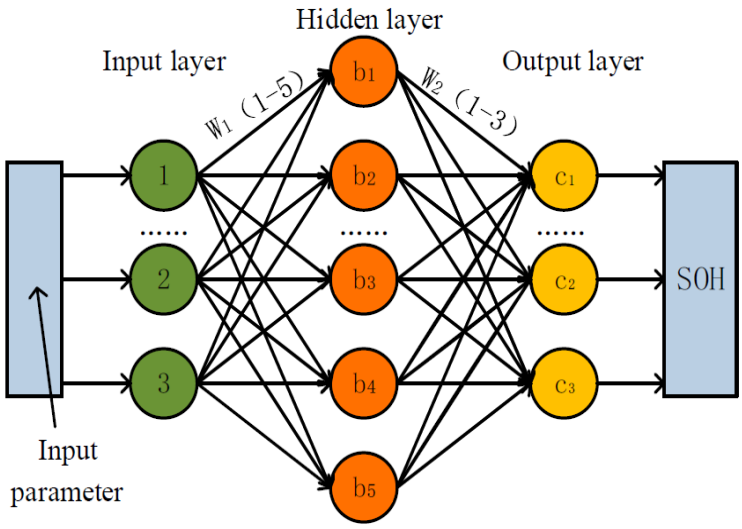

Figure 2.5: A typical neural network. b1 to b5 show the framework of the hidden layer, c1-c3 show the output layer framework. W1 (1-5) is the input weight and W2 (1-3) is the output weight. (Yao et al., 2021)

The CNN is a common deep learning algorithm model that layers the single layer CNN many times. The output of the previous layer is utilized as the input of the final layer as the multi-layer hidden layers are processed layer by layer. This basic model may be utilized to perform the task of target learning while greatly reducing the calculation results' accuracy faults. (Li & Yang, 2020). Researchers are frequently enthusiastic about estimating the growth of the SOC over a charge or discharge cycle or the evolution of the SOH over many cycles in batteries. A convolutional neural network is useful for various challenges involving the passage of time. This is a specialized fitting function that is useful for systems that exhibit temporal invariance, capturing, for example, the behaviour of a battery regardless of the time of day it was used

(Conduit et al. 2020). Fan et al. (2020) suggested a hybrid technique based on gate-recursive element convolutional neural network (GRU-CNN) to evaluate and investigate the charging voltage curves of lithium batteries. SOH is estimated online using measurable data such as voltage, current, and temperature. It is a unique approach for estimating SOH and predicting life. The article achieved a prediction error of 4.3 %, which shows GRU-CNN is a very effective way of estimating SOH. The downside of CNN is, it requires a lot of time and large training data sets in order to perform effectively.

RNN use feedback to complete a cycle by converting output parameters to input and sending data back to the network. As a result, the network model can recall previous data and use it for prediction. When there is a significant appropriate gap in information, however, the reverse propagation of the RNN network results in gradient explosion. Researchers tweaked the authentic neurons and constructed a more complicated archetypal structural model to increase the model's performance. To address this issue, a long short-term memory (LSTM) network was developed. As a time series analysis approach, the LSTM algorithm is a version of RNN. It is employed in the estimate and prediction of SOH in lithium-ion batteries due to its exceptional performance. Zhang et al. (2020) used LSTM to predict RUL and degree of health degradation. The novel LSTM RNN captured the underlying long-term relationships among the fading capacities and built an explicitly capacity-oriented RUL predictor, whose lengthy development output was better when compared to the support vector machine model, the particle filter model, and the basic RNN model.

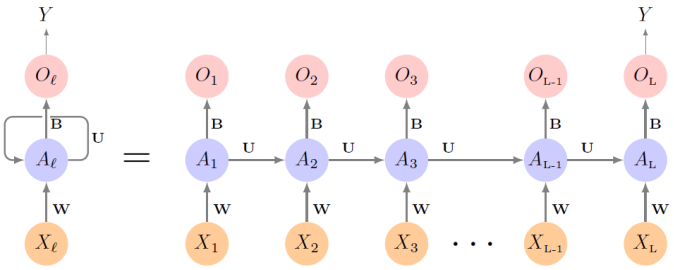

Figure 2. 5: A simple RNN (James et al., 2019)

In (Tan & Zhao, 2019), the researchers suggested LSTM completely combined with fully connected (FC) layers. The LSTM would learn the reduce errors in the prediction model by learning the long-term dependencies of battery ageing and FC layers would serve as “firewall” when layers are transferred. Their method achieved optimal results when contrasted to other data-based methods.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Background of research

In this work, we designed an algorithm for estimating the SOH of a lithium-ion battery with the help of data from the National Aeronautics and Space Administration. The algorithm is a random forest regression. In estimating the SOH, the cycle and capacitor were two important factors used. The capacitor acted as the SOH indicator and cycles, the independent variable. The datasets used were divided into test and training sets, to avoid overfitting. About 25 per cent of the data was test, with the remaining being training data so that the model's level of accuracy would be detected through calculating the test errors.

3.2 Data Description

The data consists of eight key variables after a discharge process of a lithium-ion battery. The variables are cycles measured after each discharge process, time measured in seconds, voltage measure calculated in volts, current measure measured in ampere, temperature measured in degree Celsius, voltage load measured in watts, current load measured in ampere and capacitance measured in farad.

The data was recorded after 167 discharging cycles and loaded into the trained model for state-of-health calculation after the discharge. For each cycle, at least 190 different sets of data of the other variables were measured.

3.3 Random Forests

Decision trees are a type of non-parametric supervised learning approach that may be used for classification and regression. The objective is to build a model that forecasts the value of the

target variable using basic decision rules derived from data attributes. An extension of decision trees is random forest regression which provides more accurate predictions. Random forests are a tree predictor combination in which each tree is reliant on the values of a randomly selected vector sampled separately and with the same distribution for all trees in the forest. The random forest creates a large number of decision trees, where its final result is the

Table 3. 1: The first few observations of the data used.

	Cycle	Time	Voltage Measure	Current measure	Temperature	Voltage Load	Current load	Capacitor
1	1	0.000	4.191492	-4.901589e-03	24.33003	0.000	-0.0006	1.856487
2	1	16.781	4.190749	-1.478006e-03	24.32599	4.206	-0.0006	1.856487
3	1	35.703	3.974871	-2.012528e+00	24.38909	3.062	-1.9982	1.856487
4	1	53.781	3.951717	-2.013979e+00	24.54475	3.030	-1.9982	1.856487
5	1	71.922	3.934352	-2.011144e+00	24.73139	3.011	-1.9982	1.856487
6	1	90.094	3.920058	-2.013007e+00	24.90982	2.991	-1.9982	1.856487
7	1	108.281	3.907904	-2.014400e+00	25.10588	2.977	-1.9982	1.856487
8	1	126.453	3.897036	-2.011603e+00	25.31702	2.967	-1.9982	1.856487
9	1	144.641	3.887477	-2.018015e+00	25.50942	2.959	-1.9982	1.856487
10	1	162.844	3.878959	-2.013135e+00	25.70360	2.951	-1.9982	1.856487
11	1	181.016	3.871016	-2.013162e+00	25.88721	2.943	-1.9982	1.856487
12	1	199.219	3.863480	-2.013030e+00	26.07781	2.935	-1.9982	1.856487
13	1	217.391	3.856661	-2.013710e+00	26.25431	2.930	-1.9982	1.856487
14	1	235.594	3.850014	-2.014150e+00	26.43671	2.925	-1.9982	1.856487
15	1	253.750	3.843842	-2.013464e+00	26.61128	2.925	-1.9982	1.856487
16	1	271.984	3.837688	-2.012646e+00	26.78135	2.919	-1.9982	1.856487
17	1	290.141	3.831938	-2.013409e+00	26.95746	2.913	-1.9982	1.856487
18	1	308.359	3.826414	-2.013145e+00	27.12980	2.907	-1.9982	1.856487
19	1	326.500	3.821096	-2.012041e+00	27.30582	2.902	-1.9982	1.856487
20	1	344.750	3.815897	-2.014463e+00	27.46482	2.897	-1.9982	1.856487
21	1	362.906	3.810697	-2.013140e+00	27.61796	2.892	-1.9982	1.856487
22	1	381.047	3.805837	-2.009226e+00	27.77548	2.887	-1.9982	1.856487
23	1	399.187	3.801122	-2.014666e+00	27.93133	2.882	-1.9982	1.856487
24	1	417.281	3.796293	-2.013218e+00	28.07897	2.877	-1.9982	1.856487
25	1	435.484	3.791449	-2.011514e+00	28.23350	2.872	-1.9982	1.856487
26	1	453.672	3.786625	-2.012202e+00	28.37878	2.867	-1.9982	1.856487

Showing 1 to 27 of 49,976 entries, 8 total columns

average output of all the decision trees. Decision nodes and leaf nodes make up each decision tree. The decision nodes use a test function to evaluate each fed-in sample and route it to one of several branches based on the sample's characteristics.

Suppose X is the input vector containing m variables represented as $X = \{X_1, X_2, X_1, \dots, X_n\}$ and Y the output scalar, then the training data containing n number of observations can be given in the form $Vn = \{(X1, Y1), (X2, Y2), (X3, Y3), \dots, (Xn, Yn)\}$, where $X \in \mathbb{R}^m$, $Y \in \mathbb{R}$.

During training, the algorithm splits the input data at each node so that the variables of the split functions can be maximized to fit the set. During the first step, the decision tree must determine the best split across all variables. This splitting process begins at the root and proceeds through each node, with each node applying its own split function to the new input X. This is done iteratively until a terminal node (tree leaves) is reached. It is common practice to stop the tree when a certain number of levels are attained or when a node has fewer than a certain number of observations. Finally, a function $\hat{h}(X, Vn)$ for prediction is built over V_n.

The random forest uses L-tree structured base classifier $h(X, \Theta_k)$ where $k = 1,2,3,4 \dots, \Theta$ and Θ_k belongs to a group of independent and identically distributed (IID) random vectors that combine several decision trees. Usually, bagging or bootstrap aggregation is used to carry out this process. Bagging is a popular ensemble learning method for reducing variance in a noisy dataset. Proposed by Leo Breiman, bagging selects a random sample of data from a training set with replacement, which means that individual data points can be chosen multiple times. Following the generation of several data samples, these weak models are trained independently, and based on the task— regression or classification, for example—the average or majority of those predictions yield a more accurate estimate. The process of collecting individual random samples is known as bootstrapping and each sample has a probability of 1/n to be selected. Bagging algorithm chooses several bootstrap samples ($Vn^{\Theta 1}, \dots, Vn^{\Theta q}$) and employs the

previous tree decision algorithm on these samples to generate a set of q prediction trees $\hat{h}(X, Vn^{\Theta 1}), \ldots, \hat{h}(X, Vn^{\Theta q})$. The ensemble generates q outputs, one for each tree $\hat{Y}_1 = \hat{h}(X, Vn^{\Theta 1}), \hat{Y}_2 = \hat{h}(X, Vn^{\Theta 2}), \ldots, \hat{Y}q = \hat{h}(X, Vn^{\Theta q})$. By finding the average of all outputs, aggregation is done and Y is obtained by $\hat{Y} = \frac{1}{q} \sum_{k=0}^{q} \hat{h}(X, Vn^{\Theta k})$.

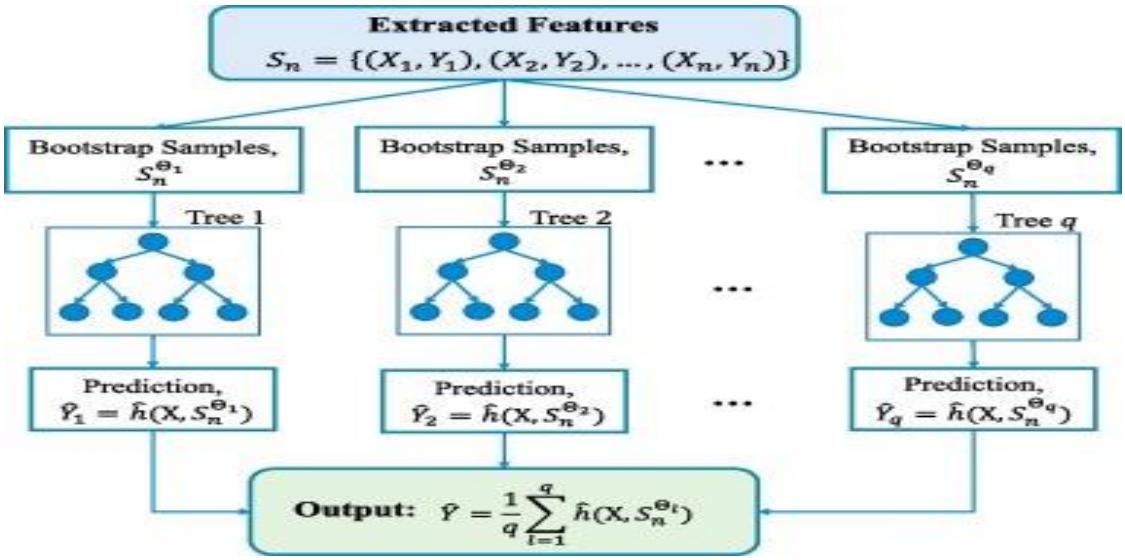

Figure 3. 1: An illustration of the construction of random forest regression. Let $\boldsymbol{Vn^{\Theta q}} = \boldsymbol{Sn^{\Theta q}}$ (Li et al., 2020)

As an advantage, bagging is immune to noise because it generates non-correlated trees from variety of training samples. Although a weak predictor may be dependent on noise, the mean of several de-correlated decision trees can significantly reduce noise sensitivity. (Lahouar & Slama, 2017). A prominent characteristic of RF is that the trees within grow without being pruned, making them algorithmically light. Furthermore, preventing the correlation of several trees, and ramping up the distinctiveness of the trees by growing them from various training data subsets formed in random forest is a phenomenon advantage of using random forest regression. (Rodriguez-Galiano et al., 2015). Certain data may be used multiple times during training, whereas others may never be employed. As a result of the use of bagging, greater RF

stability has been achieved, making RF regression more robust when dealing with minor tweaks in input data. (Breiman, 2001).

In RF, the number of trees (n_{trees}) and random features (m_{try}) for every split in the forest to construct are the two parameters to tune. In broad sense, more and more trees that are grown in the forest, the more accurate and reliable the prediction can be. An improved number of trees, on the other hand, can result in an increased computational requirements. The generalization error converges as the number of trees grows, implying that after a particular point, the estimation accuracy cannot be enhanced. The quantity of trees must be very large to necessitate for this convergence, with the default value of 500 commonly used for prediction. The amount of random features for every split is a delicate parameter that affects model resilience and describes the strength of every tree in the forest in addition to the correlation between any two of the forest's trees. Expanding the number of random features can raise the strength of each tree, but it also increases the correlation between trees simultaneously. (Peters et al., 2007)

As for estimating the error rate, two methods have been proposed. The first one involves predicting the out-of-bag data at each step of bootstrap employing the tree grown from the sample of bootstrap. Since each bagged tree uses approximately two-thirds of the observations on average. Out-of-bag (OOB) observations are the remaining one-third of the observations that were not used to fit a given bagged tree. Because the response for each observation is predicted using only the trees that weren't fit to use that observation, the resulting OOB error is a valid estimate of the test error for the bagged model. The other method is to add up the OOB predictions. On average, each data point would be OOB 36% of the time, so combine them and determine the error rate and refer to it as the OOB error rate estimate.

The random forests algorithm (for both classification and regression) is as follows:

- *Draw ntree bootstrap samples from the original data.*
- *For each of the bootstrap samples, grow an unpruned classification or regression tree, with the following modification: at each node, rather than choosing the best split among all predictors, randomly sample m_{try} of the predictors and choose the best split from among those variables. (Bagging can be thought of as the special case of random forests obtained when $m_{try} = p$, the number of predictors.)*
- *Predict new data by aggregating the predictions of the ntree trees (i.e., majority votes for classification, average for regression).*

Figure 3. 2: An illustration of the processes of RF. (Liaw & Wiener, 2002)

.

3.4 Feature Importance

Given that there are eight variables in the dataset recorded during the discharging process, it is particularly important that features are grouped in order of importance in order to know which ones are relevant for predicting the capacitor. In doing that, node impurities in RF shall be used to check the importance of every feature and the feature with much relevance to prediction of capacitor would be used in the analysis

3.5 Performance Evaluation Tools

Below are some metrics used in calculating the performance and standard of predictions of our random forests regression.

a. Root Mean Square Error/Deviation: The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a commonly used measure of the differences between predicted values and observed values. RMSD is never negative, and a value of 0 (never practicable) indicates a perfect fit to the data. A lower RMSD is generally preferable to a higher one. It is remarkably comparable to mean absolute error, except that it

marginalizes larger absolute values by giving them more weight than mean absolute error. $RMSD$ is represented as

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}}, \qquad (3.1)$$

where i is any variable, N the number of non-missing data points, x_i is the actual observations in time series and $\hat{x}_i$ is the estimated time series.

b. Goodness-of-fit (R^2) : When R^2 is 1, it implies 100% of our variability in Y is explained by the model but that is not the often times. Using the proportion of overall variation explained by the model, R^2 gives an indication of how well observed outcomes are reproduced by the model. (Draper & Smith, 2005). The R^2 is given by

$$R^2 = \frac{1 - RSS}{TSS}, \qquad (3.2)$$

where RSS is the sum of squares residuals and TSS is the total sum of squares.

3.6 Test Model

The linear regression would be used to test the validity of the model by dividing the data into test and train of 75% and 25% respectively. Then the adjusted R^2 and $RMSE$ of the model would be compared with those obtained for the random forest regression

CHAPTER FOUR

ANALYSIS AND RESULTS

4.1 Correlation

The SOH of battery degrades as number of cycles increase and the correlation between the cycles and the estimated SOH has been found as illustrated below. Clearly, there is a near-perfect relationship between the SOH and the number of cycles. From the plot, as the number of cycles increase, the SOH decreases substantially. In fact, the Pearson, the Kendall and Spearman's methods of calculating correlation coefficients gave results of -0.987, -0.953 and -0.994 respectively, showing a strong relationship between the number of cycles and SOH. A further test for significance indicated that the correlation between the two variables were significant at 5 per cent significance level, the Kendall and Spearman's test resulted in very small values(less than 0.5) and the 95 per cent confidence interval for the Pearson test was (-0.9873, -0.9868) with a very small p-value, implying the significance of the test.

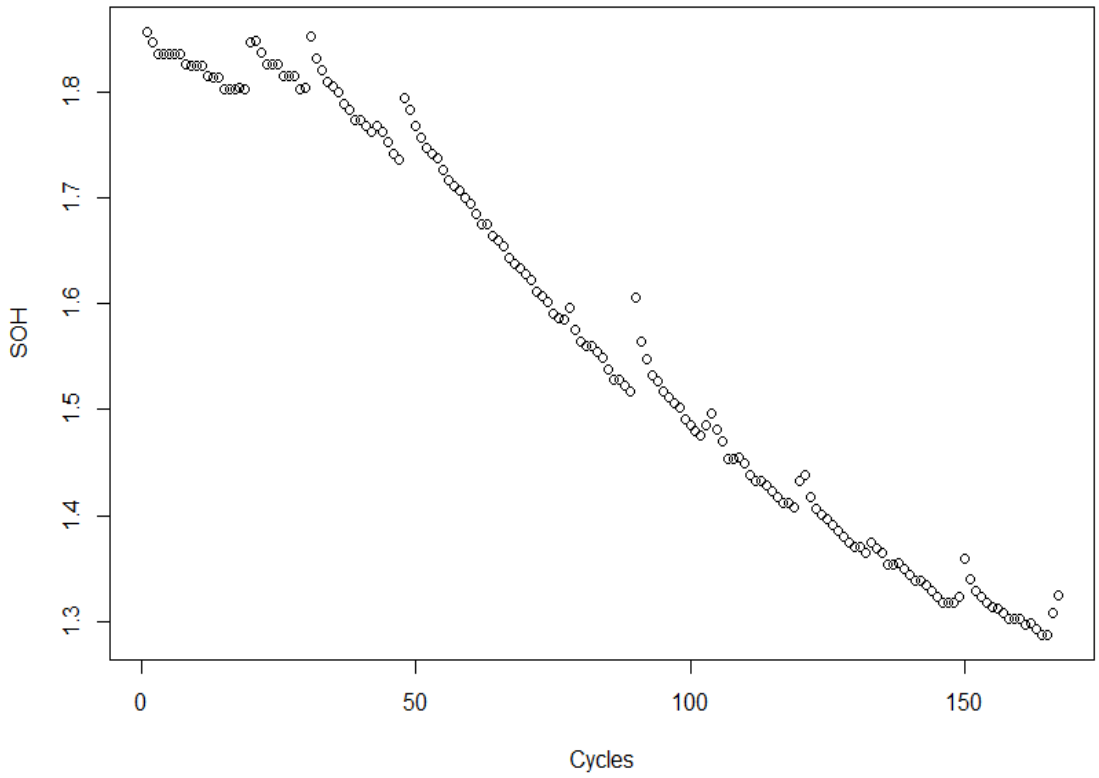

Figure 4. 1: A scatter plot showing the inverse relationship between the number of cycles and capacitor (the estimated SOH)

4.2 Random Forest Regression

As in the previous chapter, the random forest does not require scaling of data. As such, the number of features and target should be defined. However, numerous features can also be created by adding the attributes available but it won't be done for now. Here, the researcher constructed a random forest regression for predicting the SOH. As said earlier, 75 per cent of the data used for model fitting was trained and the remaining 25 per cent would be the test data. For a start, a pre-trained model was used to predict the values for the test data. The pre-trained model was set at 8 n_{trees}, the number of trees to grow ensuring that every variable gets predicted a few times. Also, the maximum number of terminal nodes the forest could have was eight and

couldn’t be set at a larger number. Then, the predicted values were compared with the actual data to check the accuracy of the model. It could be seen in the diagrams below that the prediction of the capacitor, the estimated indicator for SOH was quite different from the real value. Using the ggplot package in R, both the real and predicted values of the capacitor was plotted.

Table 4. 1: A dataframe showing the real and predicted values of the capacitor.

	Cycle	Time	Voltage.Measure	Current.measure	Temperature	Voltage.Load	Current.load	capacitor	prediction
1	1	271.984	3.837688	-2.012646e+00	26.78135	2.919	-1.9982	1.856487	1.760647
2	1	326.500	3.821096	-2.012041e+00	27.30582	2.902	-1.9982	1.856487	1.760647
3	1	435.484	3.791449	-2.011514e+00	28.23350	2.872	-1.9982	1.856487	1.760647
4	1	453.672	3.786625	-2.012202e+00	28.37878	2.867	-1.9982	1.856487	1.760647
5	1	490.203	3.777175	-2.012131e+00	28.66590	2.858	-1.9982	1.856487	1.760647
6	1	599.656	3.749629	-2.012847e+00	29.40275	2.831	-1.9982	1.856487	1.760647
7	1	654.422	3.736700	-2.013385e+00	29.68143	2.819	-1.9982	1.856487	1.760647
8	1	672.594	3.732483	-2.012629e+00	29.76670	2.815	-1.9982	1.856487	1.760647
9	1	690.812	3.728353	-2.015505e+00	29.83908	2.810	-1.9982	1.856487	1.760647
10	1	837.016	3.696758	-2.012316e+00	30.28051	2.779	-1.9982	1.856487	1.760647
11	1	873.578	3.689177	-2.012080e+00	30.38628	2.772	-1.9982	1.856487	1.760647
12	1	1056.922	3.652327	-2.010716e+00	30.85971	2.735	-1.9982	1.856487	1.760647
13	1	1112.000	3.642076	-2.013710e+00	31.00631	2.725	-1.9982	1.856487	1.760647
14	1	1130.328	3.638464	-2.012817e+00	31.06152	2.722	-1.9982	1.856487	1.760647
15	1	1240.797	3.618461	-2.014194e+00	31.37758	2.702	-1.9982	1.856487	1.760647
16	1	1277.547	3.611882	-2.012210e+00	31.49875	2.696	-1.9982	1.856487	1.760647
17	1	1443.609	3.583442	-2.012771e+00	31.96442	2.667	-1.9982	1.856487	1.760647
18	1	1462.047	3.580485	-2.012823e+00	32.01316	2.664	-1.9984	1.856487	1.761305
19	1	1499.031	3.574369	-2.011585e+00	32.11931	2.658	-1.9982	1.856487	1.760647
20	1	1536.109	3.568721	-2.012228e+00	32.21985	2.652	-1.9984	1.856487	1.761305
21	1	1591.859	3.559784	-2.010843e+00	32.37084	2.643	-1.9982	1.856487	1.760647
22	1	1628.953	3.553974	-2.013477e+00	32.44709	2.638	-1.9984	1.856487	1.761305
23	1	1684.656	3.545806	-2.013022e+00	32.59900	2.629	-1.9982	1.856487	1.760647
24	1	1815.047	3.527548	-2.012969e+00	32.90916	2.610	-1.9982	1.856487	1.760647
25	1	1964.891	3.508805	-2.012129e+00	33.23716	2.591	-1.9982	1.856487	1.760647
26	1	2417.797	3.460766	-2.013968e+00	34.34762	2.542	-1.9982	1.856487	1.778129
27	1	2455.969	3.456376	-2.010760e+00	34.45520	2.538	-1.9984	1.856487	1.778786
28	1	2474.984	3.453821	-2.012178e+00	34.53097	2.536	-1.9984	1.856487	1.778159
29	1	2647.234	3.430897	-2.012669e+00	35.09557	2.513	-1.9984	1.856487	1.778159
30	1	2685.812	3.424844	-2.011097e+00	35.24659	2.507	-1.9982	1.856487	1.777502
31	1	2839.937	3.396131	-2.009720e+00	35.82367	2.479	-1.9982	1.856487	1.766042
32	1	2878.578	3.387048	-2.012472e+00	35.99557	2.470	-1.9982	1.856487	1.766042
33	1	2994.969	3.348677	-2.013447e+00	36.48121	2.434	-1.9982	1.856487	1.754387
34	1	3111.828	3.273716	-2.012311e+00	37.06922	2.364	-1.9982	1.856487	1.754387
35	1	3150.672	3.233103	-2.011235e+00	37.30958	2.327	-1.9982	1.856487	1.754387
36	1	3248.625	3.065382	-2.014808e+00	37.98252	2.173	-1.9982	1.856487	1.777137
37	1	3268.328	3.013061	-2.011095e+00	38.13325	2.125	-1.9982	1.856487	1.777137
38	1	3307.625	2.866018	-2.010609e+00	38.47645	1.998	-1.9982	1.856487	1.792343

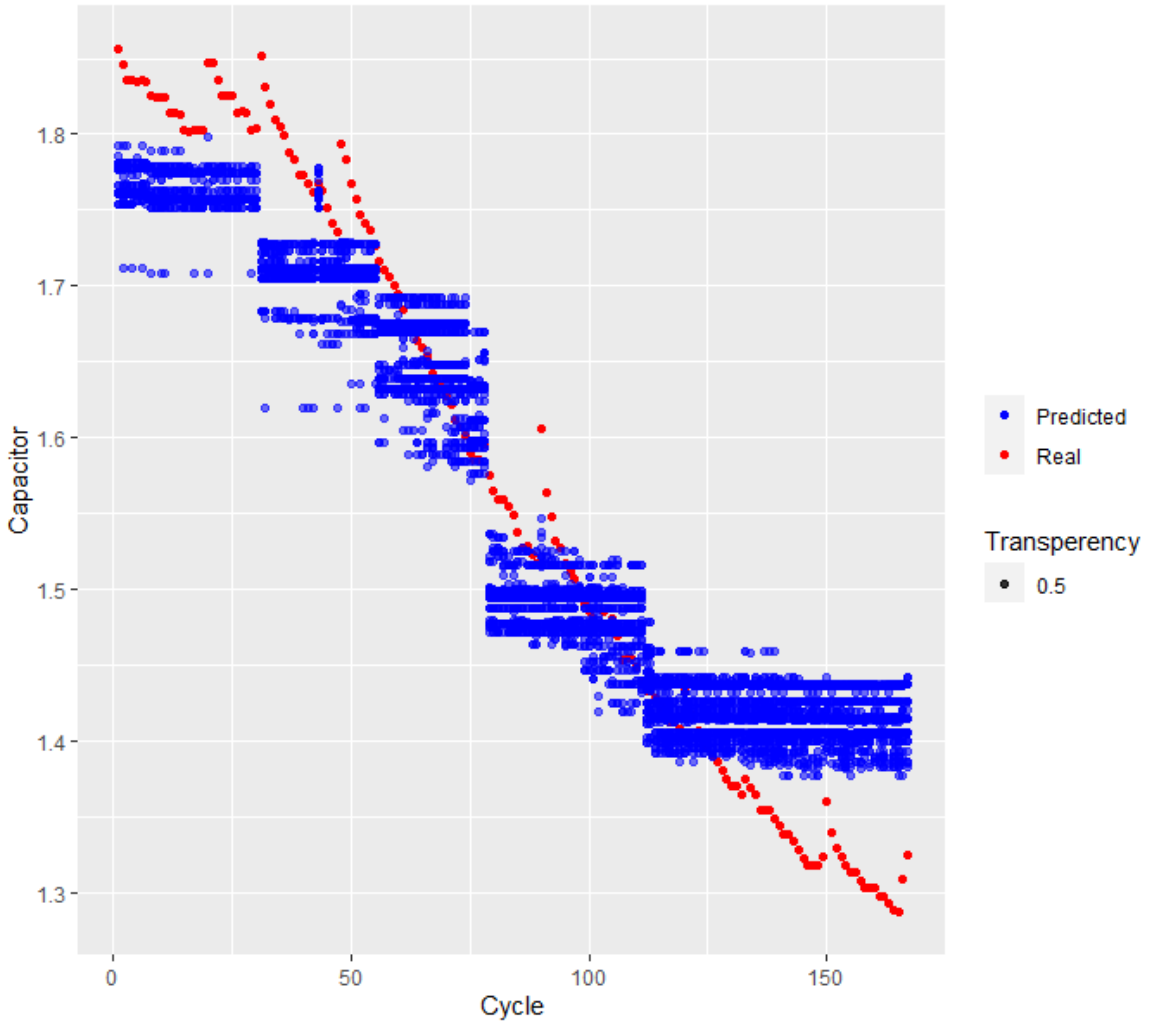

Figure 4. 2: The plotted values of the real and predicted values of the capacitor.

To estimate the model more precisely, the R^2 of the test data was computed and 0.927 was recorded. Also, the RMSE recorded was 0.065.

Although the RMSE and R^2 have good values, we was still a possible improvement of the power of prediction of the model by tuning the hyper-parameters of the model. It could be done manually but in order to save time, the R software was used. To tune the parameters: the number of trees in the forest (n_{trees}) and the maximum number of terminal nodes possible for trees to have in the forest(maxnodes), a custom random forest model was built to obtain the best set of parameters and compare the results for the different combinations of the parameters.

The model was at its best at maxnodes 100 and n_{trees} 1000 and that these values would be used in the final model to predict the SOH of the lithium-ion battery after the discharge.

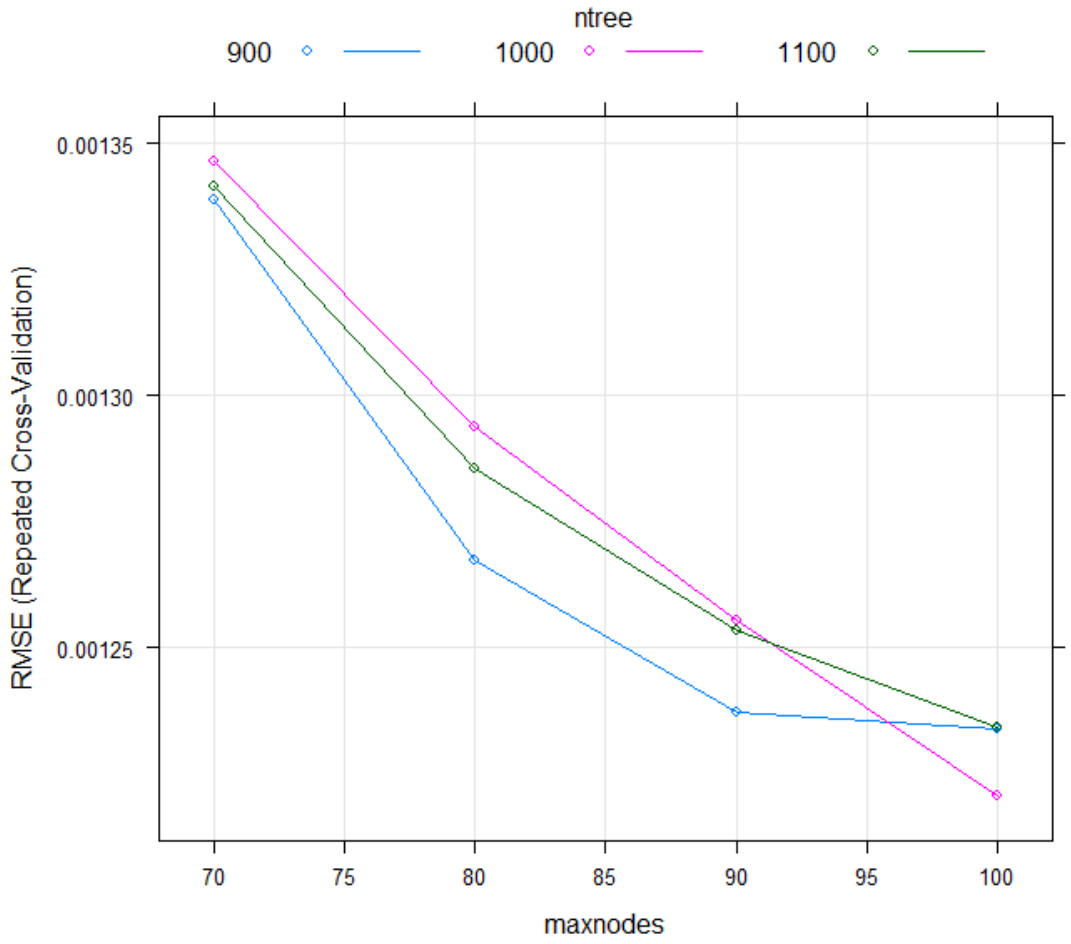

Figure 4. 3: A Cross-Validation plot of the impact of the tuned parameters on RMSE.

4.3 Feature Importance

For this model, we used all the features available. In predicting the capacitor of the battery, it was clear only cycles had significant importance in the prediction. On the vertical axis, an incremental reduction of node impurities from dividing on the variable is seen and it's calculated by the sum of square residuals. In general, random forests rarely gives casual interpretation.

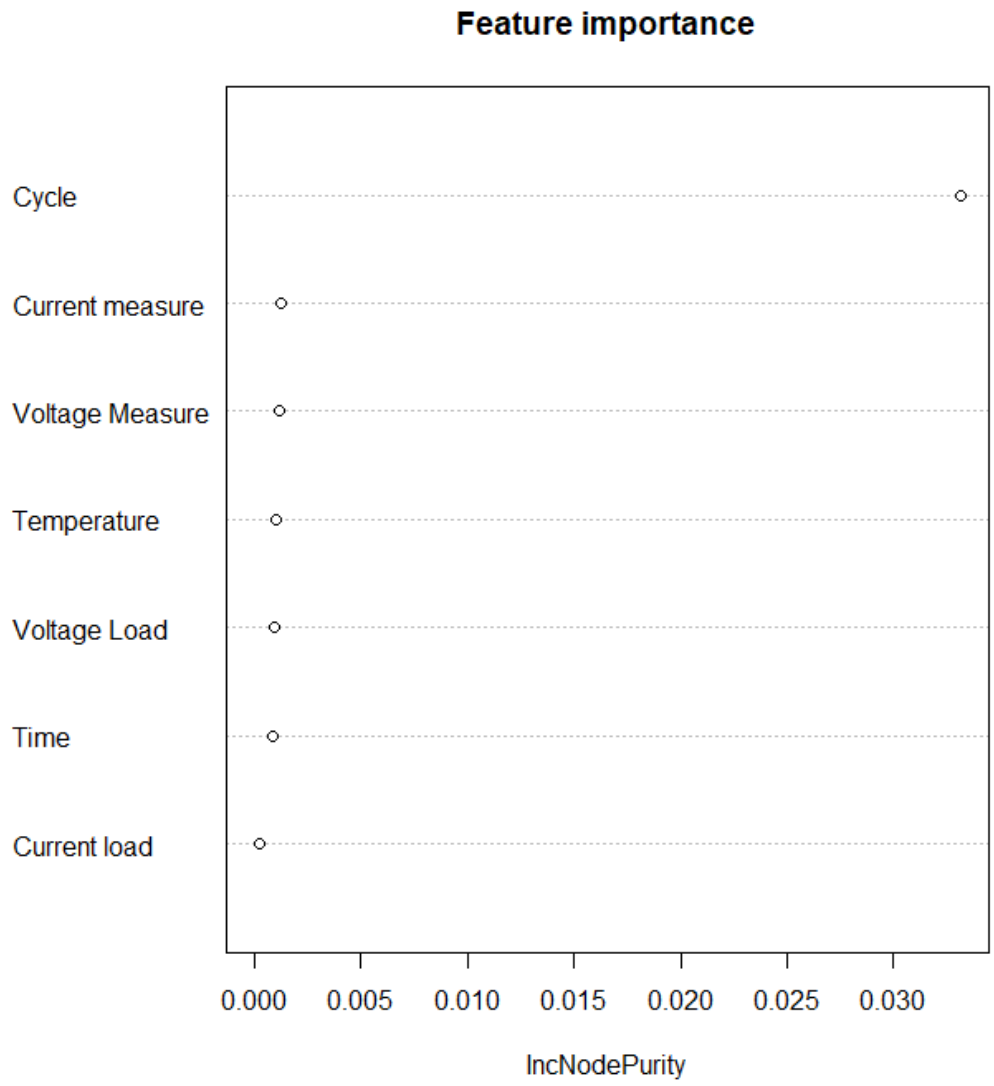

Figure 4. 4: The cycle is the major part of predicting the capacitor of the battery.

After realising the model shall attain its highest predictive power at maxnodes and n_{trees} at 100 and 1000 respectively, a final model is built to predict the capacitor of the lithium battery. The results of the final model is shown below. It can be seen that in the final model to estimate the capacitor, the predicted values are closer to the real values than the pre-trained model.

The following images show that.

Table 4. 2: Using the final model to predict the capacitor of the battery.

	Cycle	Voltage.Measure	Current.measure	Temperature	Voltage.Load	Current.load	capacitor	prediction
1	1	3.837688	-2.012646e+00	26.78135	2.919	-1.9982	1.856487	1.830243
2	1	3.821096	-2.012041e+00	27.30582	2.902	-1.9982	1.856487	1.830219
3	1	3.791449	-2.011514e+00	28.23350	2.872	-1.9982	1.856487	1.829964
4	1	3.786625	-2.012202e+00	28.37878	2.867	-1.9982	1.856487	1.830020
5	1	3.777175	-2.012131e+00	28.66590	2.858	-1.9982	1.856487	1.829869
6	1	3.749629	-2.012847e+00	29.40275	2.831	-1.9982	1.856487	1.830014
7	1	3.736700	-2.013385e+00	29.68143	2.819	-1.9982	1.856487	1.829919
8	1	3.732483	-2.012629e+00	29.76670	2.815	-1.9982	1.856487	1.829718
9	1	3.728353	-2.015505e+00	29.83908	2.810	-1.9982	1.856487	1.829721
10	1	3.696758	-2.012316e+00	30.28051	2.779	-1.9982	1.856487	1.830788
11	1	3.689177	-2.012080e+00	30.38628	2.772	-1.9982	1.856487	1.831030
12	1	3.652327	-2.010716e+00	30.85971	2.735	-1.9982	1.856487	1.831451
13	1	3.642076	-2.013710e+00	31.00631	2.725	-1.9982	1.856487	1.831534
14	1	3.638464	-2.012817e+00	31.06152	2.722	-1.9982	1.856487	1.831398
15	1	3.618461	-2.014194e+00	31.37758	2.702	-1.9982	1.856487	1.831961
16	1	3.611882	-2.012210e+00	31.49875	2.696	-1.9982	1.856487	1.832085
17	1	3.583442	-2.012771e+00	31.96442	2.667	-1.9982	1.856487	1.832325
18	1	3.580485	-2.012823e+00	32.01316	2.664	-1.9984	1.856487	1.834978
19	1	3.574369	-2.011585e+00	32.11931	2.658	-1.9982	1.856487	1.832457
20	1	3.568721	-2.012228e+00	32.21985	2.652	-1.9984	1.856487	1.835153
21	1	3.559784	-2.010843e+00	32.37084	2.643	-1.9982	1.856487	1.832863
22	1	3.553974	-2.013477e+00	32.44709	2.638	-1.9984	1.856487	1.835671
23	1	3.545806	-2.013022e+00	32.59900	2.629	-1.9982	1.856487	1.833244
24	1	3.527548	-2.012969e+00	32.90916	2.610	-1.9982	1.856487	1.833461
25	1	3.508805	-2.012129e+00	33.23716	2.591	-1.9982	1.856487	1.834143
26	1	3.460766	-2.013968e+00	34.34762	2.542	-1.9982	1.856487	1.834863
27	1	3.456376	-2.010760e+00	34.45520	2.538	-1.9984	1.856487	1.838120
28	1	3.453821	-2.012178e+00	34.53097	2.536	-1.9984	1.856487	1.838427
29	1	3.430897	-2.012669e+00	35.09557	2.513	-1.9984	1.856487	1.838555
30	1	3.424844	-2.011097e+00	35.24659	2.507	-1.9982	1.856487	1.834982
31	1	3.396131	-2.009720e+00	35.82367	2.479	-1.9982	1.856487	1.835997
32	1	3.387048	-2.012472e+00	35.99557	2.470	-1.9982	1.856487	1.834984
33	1	3.348677	-2.013447e+00	36.48121	2.434	-1.9982	1.856487	1.834462
34	1	3.273716	-2.012311e+00	37.06922	2.364	-1.9982	1.856487	1.833711
35	1	3.233103	-2.011235e+00	37.30958	2.327	-1.9982	1.856487	1.833931
36	1	3.065382	-2.014808e+00	37.98252	2.173	-1.9982	1.856487	1.834923
37	1	3.013061	-2.011095e+00	38.13325	2.125	-1.9982	1.856487	1.835205
38	1	2.866018	-2.010609e+00	38.47645	1.998	-1.9982	1.856487	1.836533
39	1	2.757252	-2.014112e+00	38.66860	1.906	-1.9984	1.856487	1.837804
40	1	3.172906	-1.014384e-03	37.76735	0.001	-0.0006	1.856487	1.840316
41	1	3.228753	-3.457734e-03	36.47916	0.000	-0.0006	1.856487	1.840500
42	1	3.238095	2.314058e-04	36.18014	0.000	-0.0006	1.856487	1.840053
43	1	3.245844	7.285853e-04	35.87476	0.000	-0.0006	1.856487	1.840340

To estimate the accuracy of the final model, the RMSE and R^2 are calculated. The RMSE here left a value of 0.02679 and R^2 , value of 0.9837. These values showed more accuracy and preciseness than the pre-trained model.

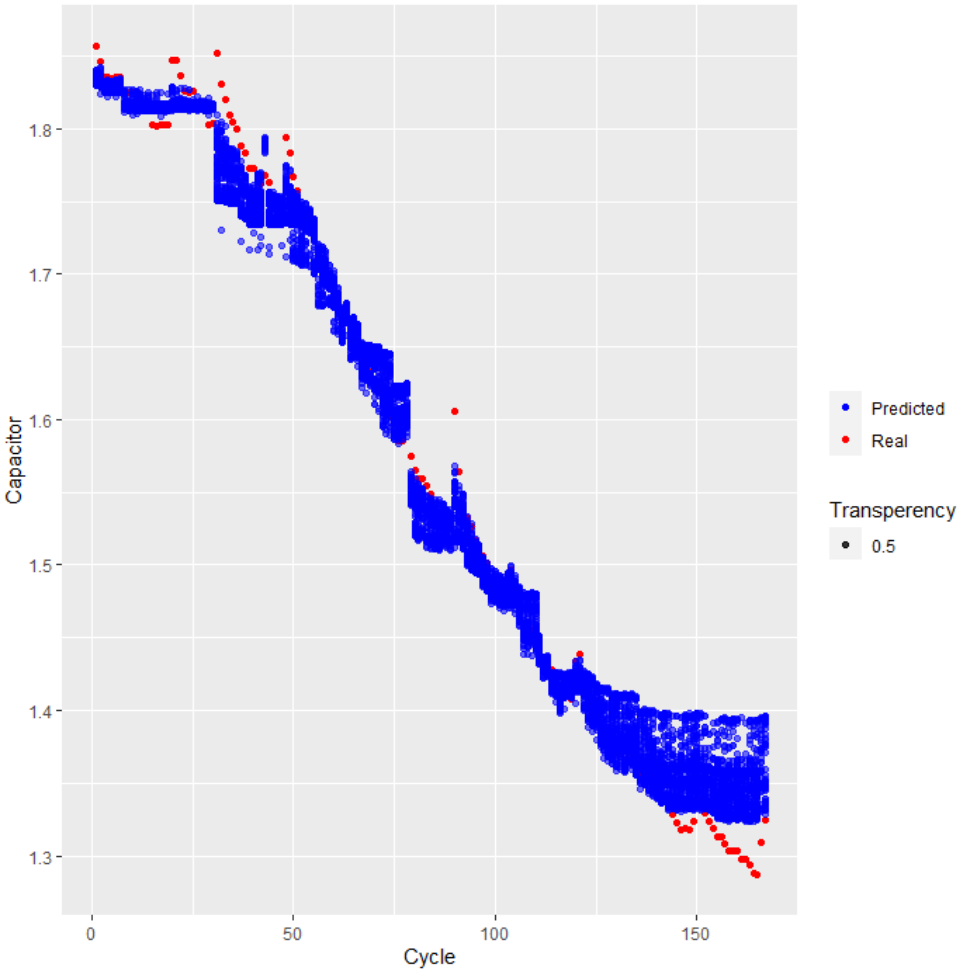

Figure 4. 5: A confirmation of how close the predicted values is to the real ones, with both mostly colliding**.**

4.4 Linear Regression

The linear regression model is tested against the random forest regression to evaluate which model fits best for the prediction of capacitor. Using the lm function in R, the data was split into test and train, with 75% of the data responsible for training and 25% used to validate the model adequacy by calculating its error.

The linear regression model resulted in an RMSE of 0.0274 and R^2 recorded at 0.9778.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Summary

The study used the random forest regression model to predict capacitor, believed to be an estimate of state-of-health of the lithium-ion battery. During discharging of the battery, the research discovered the key factor(s) influencing the capacitor. As a result, the finding will assist manufacturers of electric vehicle, hybrid-electric vehicles and any stakeholders relevant to the field of lithium-ion battery on how best their productions equipped with lithium-ion battery can prolong its state until the end-of-life.

5.2 Summary of Findings

This part summarizes the data analysis concerning the study topics.

Key Findings:

i. Of the seven variables obtained to predict the capacitor of the battery, only the number of cycles was relevant.

ii. Using Pearson, Kendall and Spearman's methods of calculating correlation coefficients, a result of -0.987, -0.953 and -0.994 were obtained , depicting that as the number of cycles increase, the strength of the capacitor reduces

iii. At 95% confidence interval, the relationship between number of cycles and capacitor was significant.

iv. In predicting the capacitor of the battery, the random forest regression algorithm is at its best if the maximum number of terminal nodes in the forest is 100 and the number of trees to be grown is 1000.

v. The random forest regression provides more accurate predictions for SOH of lithium batteries than the linear regression model.

5.3 Conclusions

Upon the several ways of estimating the state-of-health of lithium batteries, the research specialised on random forest regression, a tree-based model. Typically, the battery SOH is 100 per cent at the time of production but decreases with time and can even deteriorate before its end of life. As a result, manufacturers are on the look -out for models that are able to estimate the SOH with high accuracy and list factors that affect them. That is exactly what this study sought to achieve. The results of our model suggests that the random forest regression model is a robust way of accurately predicting the capacitor, as a SOH indicator of the battery, even though the linear regression also produced fairly good results. This study is the most recent achievement of online battery state-of-health modelling.

5.4 Recommendations

Based on the findings of this study, this empirical research would like to present the following recommendations.

- In order to check the accuracy of the model, analysis on more rigorous data-based models like the Neural Network, Support Vector Machines and Gaussian Processes should be done and compared with the Random Forest Regression in future.
- Big industries concerned with the use of lithium batteries should make battery data accessible to researchers so more research can be done.
- In terms of feature importance, more variables if added to the study could mean the factors affecting the capacitor of the lithium battery might be more than just the number of cycles given environmental factors hugely contribute to state-of-health.
- In terms of lithium batteries, more research work should be done in Africa in a bid to create factories that produce the batteries with the discovery of lithium ore in several African lands.

References

Chen,Z., Sun,M., Shu,X., Shen,J. & Xiao, R. (2018). On-board state of health estimation for lithium-ion batteries based on random forest. *IEEE International Conference on Industrial Technology (ICIT)*, 1754-1759. doi:10.1109/ICIT.2018.8352448

Feng,X., Weng,C., He, X., Han,X., Lu,L., Ren, D. & Ouyang, M. (2019). Online State-of-Health Estimation for Li-Ion Battery Using Partial Charging Segment Based on Support Vector Machine. *IEEE Transactions on Vehicular Technology*, 8583 - 8592.

Kim,J. & Cho,B.H. (2011). State-of-Charge Estimation and State-of-Health Prediction of a Li-Ion Degraded Battery Based on an EKF Combined With a Per-Unit System. *IEEE Transactions on Vehicular Technology, 60*(9), 4249 - 4260.

Lahouar,A. & Slama, J.B.H. (2017). Hour-ahead wind power forecast based on random forests. *Renewable Energy*, 529-541.

Li,X. Wang,Z. , Zhang,L., Zou,C. & Dorrell, D.D. (2019). State-of-health estimation for Li-ion batteries by combing the incremental capacity analysis method with grey relational analysis. *Journal of Power Sources*, 113.

Mawonou,K.S.R., Eddahech,A., Dumur,D.,Beauvois,D. & Godoy,E. . (2021). State-of-health estimators coupled to a random forest approach for lithium-ion battery aging factor ranking. *Journal of Power Sources, 484*, 229154. doi:https://doi.org/10.1016/j.jpowsour.2020.229154

Ng,M.-F., Zhao,J., Yan,Q., Conduit, G.J. & Seh, Z.W. (2020). Predicting the state of charge and health of batteries using data-driven machine learning. *Nature Machine Intelligence*, 161–170.

Ozcan,S.G., Milutin,P., Zafer,S., Yebin,W., Orlik P.V. & Toshihiro,L. (2017). Battery State of Charge Estimation Based on Regular/Recurrent Gaussian Process Regression. *IEEE Transactions on Industrial Electronics*, 4311–4321.

Peters,J., De Baets,B., Verhoest,N.E., Samson,R. , Degroeve,S. , De Becker,P. & Huybrechts,W. (2007). Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling*, 304–318.

Rijanto, E., Rozaqi, L. , Nugroho,A. & Kanarachos,S. (2017). RLS with Optimum Multiple Adaptive Forgetting Factors for SoC and SoH Estimation of Li-Ion Battery. *In Proceedings of the 2017 5th International Conference on Instrumentation, Control, and Automation (ICA)*, (pp. 73-77). Yogyakarta, Indonesia.

Tan,Y. & Zhao,G. . (2019). Transfer Learning With Long Short-Term Memory Network for State-of-Health Prediction of Lithium-Ion Batteries. *IEEE Transactions on Industrial Electronics*, 8723 - 8731.

Zhou,D., Yin,Y., Fu,P. , Song, X., Lu, W., Yuan,L., & Fu, Z. (2018). Prognostics for State of Health of Lithium-Ion Batteries Based on Gaussian Process Regression. *Mathematical Problems in Engineering*.

Abdi,H., Nezhad,R.R. & Salehimaheh, M. (2017). Fuel Cells. *Distributed Generation Systems*.

Bartlett, A. (2015). *Electrochemical Model-Based State of Charge and State of Health of Lithium-Ion Batteries.*

Bonfitto, A. (2020). A Method for the Combined Estimation of Battery State of Charge and State of Health Based on Artificial Neural Networks. *Energies, 13*, 2548.

Breiman, L. (2001). Random Forest. *Machine Learning* , 5-32.

Chen, N., Hu, X., Gui, W. & Zou, J. (2014). Estimation of li-ion battery state of charging and state of healthy based on unsented Kalman filtering. *In Proceedings of the 26th Chinese Control and Decision Conference (2014 CCDC)*, (pp. 4725–4729). Changsha, China.

Chen, Z., Mi, C.C., Fu, Y., Xu, J. & Gong, X. (2013). Online battery state of health estimation based on Genetic Algorithm for electric and hybrid vehicle applications. *Journal of Power Sources, 240*, 184-192.

Chen,M., Han,F., Shi,L., Feng,Y. Xue, C., Gao, W. & Xu, J. (2022). Sliding Mode Observer for State-of-Charge Estimation Using Hysteresis-Based Li-Ion Battery Model. *Energies*, 2658.

D'Orazioa,T., Leoa,M., Distantea,A.,Guaragnellab,C., Pianesec, V.&Cavaccinic, G. (2008). Automatic ultrasonic inspection for internal defect detection in composite materials. *NDT & E International, 41*(2), 145-154.

Donepudi, V.S. & Conway, B.E. (1984). Electrochemical calorimetry of the zinc and bromine electrodes in zinc-bromine and zink-air batteries. *Electrochem. Soc*, 1477–1485.

Dong, W., Qiang, M. & Pecht, M. (2013). Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model. *Journal of Power Sources*, 253–264.

Draper, N. R. & Smith, H. (2005). *Applied Regression Analysis.* Wiley-Interscience.

Fan, Y., Xiao, F., Li, F., Yang, G. & Tang, X. (2020). A novel deep learning framework for state of health estimation of lithium-ion battery. *Journal of Energy Storage*, 101741.

Fleischer, C., Waag, W., Bai, Z., & Sauer, D. U. (2013). Adaptive On-line State-of-available-power Prediction of Lithium-ion Batteries. *Journal of Power Electronics*, 516-527.

Girard, A. & Murray-Smith, R. (2005). Gaussian Processes: Prediction at a Noisy Input and Application to Iterative Multiple-Step Ahead Forecasting of Time-Series. *In Proceedings of Hamilton Summer School on Switching and Learning in Feedback Systems* (pp. 158–184). Maynooth, Ireland: Springer.

Gregory, L. (2004). Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs. Part 2. Modeling and identification. *Journal of Power Sources*, 262-276.

He, H., Xiong, R., Guo, H. & Li, S. (2012). Comparison study on the battery models used for the energy management of batteries in electric vehicles. *Energy Conversation Management*, 113-121.

He,H., Xiong,R. & Fan,J. (2011). Evaluation of Lithium-Ion Battery Equivalent Circuit Models for State of Charge Estimation by an Experimental Approach. *Energies*, 582-598.

James, G. ,. (n.d.). *An Introduction to Statistical Learning.*

James, G., Witten,D., Hastie,T. & Tibshirani, R. (2021). *An Introduction to Statistical Learning.* Springer Texts in Statistics.

Kim, I. (2010). A Technique for Estimating the State of Health of Lithium Batteries Through a Dual-Sliding-Mode Observer. *IEEE Transactions on Power Electronics, 25*(4), 1013 - 1022.

Klass,V. , Behm,M. & Lindbergh,G. (2012). Evaluating Real-Life Performance of Lithium-Ion Battery Packs in Electric Vehicles. *Journal of The Electrochemical Society*, A1856.

Asamoah, S.K. (2022, May 31). *A picture of lithium found in Ghana*. Retrieved from GhanaWeb: https://www.ghanaweb.com/GhanaHomePage/features/Who-owns-Ghana-s-Lithium-deposits-1549805

Kostecki, R. & McLarnon, F. (2003). Microprobe study of the effect of Li intercalation on the structure of graphite. *Journal of Power Sources*, 550-554.

Lei Yao, S. X. (n.d.). Schematic diagram of equivalent circuit mode. *A Review of Lithium-Ion Battery State of Health Estimation and Prediction Methods.* Henan Engineering Research Center of New Energy Vehicle Lightweight Design and Manufacturing, Zhengzhou University of Light Industry, Zhengzhou, China.

Li, D. & Yang, L. (2020). Remaining useful life prediction of lithium battery using convolutional neural network with optimized parameters. *In Proceedings of the 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE),.* Chengdu, China.

Li, Y., Zou,C., Berecibar,M., Nanini-Maury,E., Chan, J.C.-W., Bossche,P.V.D., Van Mierlo,J. & Omar,N. (2018). Random forest regression for online capacity estimation of lithium-ion batteries. *Applied Energy, 232*, 197-210. doi:https://doi.org/10.1016/j.apenergy.2018.09.182

Liaw, A. & Wiener, M. (2002). Classification and Regression by Random Forest. 23.

Liu, X., Li, W. & Zhou, A. (2018). PNGV Equivalent Circuit Model and SOC Estimation Algorithm for Lithium Battery Pack Adopted in AGV Vehicle. *IEEE Access*, 23639–23647.

Mansouri, S. S., Karvelis, P., Georgoulas, G. & Nikolakopoulos, G. (2017). Remaining useful battery life prediction for UAVs based on machine learning. *IFAC-PapersOnLine*, 4727–4732.

Naha,A. ,Han, S. , Agarwal, S., Guha,A. , Khandelwal,A. , Tagade, P., Hariharan, K.S., Kolake, S.M., Yoon,J. & Oh, B. (2020). An Incremental Voltage Difference Based Technique for Online State of Health Estimation of Li-ion Batteries. *Scientific Reports*, 2045-2322. doi:https://doi.org/10.1038/s41598-020-66424-9

Ng,M.-F., Zhao,J., Yan,Q., Conduit, G.J. & Seh, Z.W. (2020). Predicting the state of charge and health of batteries using multiple regression. *Nature Machine Intelligence*, 161–170.

Ning, B.,Cao, B., Wang, B. & Zou, Z. (2018). Adaptive sliding mode observers for lithium-ion battery state estimation based on parameters identified online. *Energy*, 732-742.

Noura, N.,Boulon, L. & Jemeï, S. (2020). A Review of Battery State of Health Estimation Methods: Hybrid Electric Vehicle Challenges. *World Electric Vehicle Journal*, 5.

Nuhic, A., Terzimehic, T., Soczka-Guth, T., Buchholz, M. & Dietmayer, K. (2013). Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods. *Journal of Power Sources*, 680–688.

Ovejas,V.J. & Cuadras,A. (2018). Impedance Characterization of an LCO-NMC/Graphite Cell: Ohmic Conduction, SEI Transport and Charge-Transfer Phenomenon. *Batteries, 4*, 43.

Peng, R. & Pedram, M. (2006). An analytical model for predicting the remaining battery capacity of lithium-ion btteries. *IEEE Transactions on Very Large Scale Integration(VLSI) Systems*, 441-451.

Prasad, G.K. & Rahn, C.D. (2013). Model based identification of aging parameters in lithium ion batteries. *Journal of Power Sources*, 79-85.

Rodriguez-Galiano,V., Sanchez-Castillo,M., Chica-Olmo,M. & Chica-Rivas,M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 804-818.

Sarmah, S.B. , Kalita, P., Das, B., Garg, A. & Sarma, M. (2020). Numerical and experimental investigation of state of health of Li-ion battery. *International Journal of Green Energy*, 510-520.

Schweiger, H.-G., Obeidi, O., Komesker, O., Raschke, A., Schiemann, M., Zehner, C., Gehnen, M., Keller, M., & Birke, P. (2010). Comparison of Several Methods for Determining the Internal Resistance of Lithium Ion Cells. *Sensors*, 5604–5625. Retrieved from https://doi.org/10.3390/s100605604

Severson, K.A. , Attia, P.M., Jin,N., Perkins, N., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., Bazant, M.Z., Harris S.J., Chueh, W.C., Braatz R.D. & Jiang, B. . (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 383-391.

Song, Z., Gao, J., Pan, L., & Xi J. (2020). Lithium-ion battery health status prediction based on principal component analysis and improved support vector machines. *International Journal of Automotive Technology*, 21-27.

Stroe, D.-L. & Schaltz, E. (2020). Lithium-Ion Battery State-of-Health Estimation Using the Incremental Capacity Analysis Technique. *IEEE Transactions on Industry Applications*, 678-685.

Troltzsch, U., Kanoun, O. & Trankler, H. (2006). Characterizing aging effects of lithium ion batteries by impedence spectrocopy. *Electrochimica Acta, 5(8,9)*, 1664-1672.

Tulsyan,A.,Tsai,Y., Gopaluni R.B. & Braatz R.D. (2016). State-of-charge estimation in lithium-ion batteries: A particle filter. *Journal of Sources of Power(331)*, 208-223.

Vetter, J., Novák, P., Wagner, M.R., Veit, C., Möller, J., Besenhard, J.O. , Winter, M., Wohlfahrt-Mehrens, M., Vogler, C. & Hammouche, A. (2005). Ageing mechanisms in lithium-ion batteries. *Journal of Power Sources*(0378-7753), 269-281. Retrieved from https://doi.org/10.1016/j.jpowsour.2005.01.006.

Xia, Z. & Qahouq, J.A. (2021). State-of-charge Balancing of Lithium-ion Batteries with State-of-health Awareness Capability. *IEEE Transactions on Industry and Applications*, 673-684. doi:https://ieeexplore.ieee.org/document/9217974

Xiaoyu Li, Z. W. (2019). State of health estimation for lithium-ion battery by combining incremental capacity analysis with Gaussian process regression. *arXiv : Signal Processing*.

Xiong, R., Zhang, Y.,Wang, J., He, H., Peng, S. & Pecht, M. (2019). Lithium-Ion Battery Health Prognosis Based on a Real Battery Management System Used in Electric Vehicles. *IEEE Transactions on Vehicular Technology*, 4110 - 4121.

Yang, D., Wang, Y., Pan, R., Chen, R., & Chen, Z. (2017). A Neural Network Based State-of-Health Estimation of Lithium-ion Battery in Electric Vehicles. *Energy Procedia, 105*, 2059-2064. doi:https://doi.org/10.1016/j.egypro.2017.03.583

Yao, L.,Xu, S, Tang A., Zhou, F., Hou, J., Xiao, Y. & Fu, Z. (2021). A Review of Lithium-Ion Battery State of Health Estimation and Prediction methods. *World Electric Vehicle Journal*, 113. doi:https://doi.org/10.3390/wevj12030113

Zahid, T., Xu, K, Li, W., Li, C. & Li, H. (2018). State of charge estimation for electric vehicle power battery using advanced machine learning algorithm under diversified drive cycles. *Energy*, 871-882.

Zhan, J., Liu, X., Chen, C., & Simeone, A. (2020). An Intelligent Deformation-Based Approach to the State of Health Estimation of Collided Lithium-Ion Batteries for Facilitating Battery Module Safety Evaluation. *Energy Technology*. doi:https://doi.org/10.1002/ente.202000624

Zhang, Y., Xiong, R., He, H., & Pecht, M. (2018). Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction. *IEEE Transactions on Vehicular Technology*, 5695 - 5705. doi:10.1109/TVT.2018.2805189

Zhang, Z., Huang, M., Chen, Y. & Zhu, S. (2016). Prediction of Lithium-ion Battery's Remaining Useful Life Based on Relevance Vector Machine. *SAE International Journal of Alternative Powertrains*, 30-40.

Zhu ,C., Zheng,B., He,Z.,Gao,M., Sun,C. & Bao,Z. (2021). State of Health Estimation of Lithium-Ion Battery Using Time Convolution Memory Neural Network. *Mobile Information Systems*, 1-16.

Zhu, Q., Xu, M., Liu, W. & Zheng, M. (2019). A state of charge estimation method for lithium-ion batteries based on fractional order adaptive extended kalman filter. *Energy*, 187.