KEVIN-CYUBAHIRO

REAL-TIME BATTERY HEALTH
MONITORING FOR
ELECTRIC VEHICLE USING MACHINE

MASTER THESIS

2023

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING

REAL-TIME BATTERY HEALTH MONITORING FOR ELECTRIC VEHICLE USING MACHINE LEARNING

M.Sc. THESIS

Kevin Cyubahiro

Nicosia

May, 2023

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF ELECTRICAL AND
ELECTRONIC ENGINEERING

REAL-TIME BATTERY HEALTH MONITORING FOR ELECTRIC VEHICLE USING MACHINE LEARNING

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF
APPLIED SCIENCES OF
NEAR EAST UNIVERSITY

BY

KEVIN CYUBAHIRO

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONICS ENGINEERING

SUPERVISOR

PROF.DR. BÜLENT BILGEHAN

NICOSIA

MAY, 2023

Approval

We certify that we have read the thesis submitted by Kevin Cyubahiro titled **"Real-Time Battery Health Monitoring for Electric Vehicle using Machine Learning**" and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Educational Sciences.

Examining Committee	Name-Surname	Signature
Head of the Committee:	Assoc. Prof. Dr. Kamil Dimililer	
Committee Member*:	Assoc.Prof. Dr. Sertan Serte	
Supervisor:	Prof. Dr. Bülent Bilgehan	

Approved by the Head of the Department

…../…../20…

Prof.Dr. Bülent Bilgehan.

Title, Name-Surname

Head of Department

Approved by the Institute of Graduate Studies

NEAR EAST UNIVERSITY INSTITUTE OF GRADUATE STUDIES 1988

…../…../20…

Prof. Dr. Kemal Hüsnü Can Başer

Head of the Institute

Declaration

I hereby declare that all information, documents, analysis and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of Institute of Graduate Studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Kevin Cyubahiro
25/04/2023

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Bülent Bilgehan, for his invaluable guidance, support, and encouragement throughout my thesis project. His expertise, patience, and insightful feedback have been crucial to the completion of this work.

I also wish to thank my colleagues for their unwavering support and encouragement during the ups and downs of this journey. Their constructive feedback, discussions, and encouragement have been instrumental in shaping my ideas and refining my work.

Finally, I would like to extend my sincere appreciation to Near East University for providing me with the necessary resources, tools, and space to carry out my research. The university's commitment to academic excellence and research has been a significant factor in my academic growth and success."

Kevin Cyubahiro

Abstract

Real-Time Battery Health Monitoring for
Electric Vehicle using Machine Learning

Cyubahiro, Kevin

Prof.Dr. Bülent Bilgehan

M.SC Department of Electrical and Electronics Engineering

May,2023

Electric vehicles (EVs) are gaining a big market share in the automotive sector. Most of the automotive companies are emerging in producing a big number of them. But there are still challenges on how to control the battery managements system of the EVs by ensuring Efficiency. This reasons have caused the EVs not to be accepted on a high rate compare to the combustion engine cars. By giving precise information about the Charge state and health state of the lithium-ion batteries. The safety, dependability, and energy efficiency of EVs can all be improved through battery health state classification.

This study focuses on differentiating the performance of three machine learning algorithms which are: Support Vector Machine (SVM), Bagged Trees (BT) and Artificial Neural Network (ANN). A sizable dataset of battery voltage, current, and temperature measurements from EV usage was gathered and pre-processed to train and test each algorithm. The accuracy, efficiency, and computational complexity of each approach were evaluated to determine the most suitable algorithm for battery health monitoring. The suggested methods can be used to low-power edge devices and is capable of battery health monitoring with excellent accuracy and minimal computational complexity. This thesis offers knowledge for future study and advances the development of battery health monitoring systems for electric vehicles.

Key Words*:* Support Vector Machine; Electric Vehicles; State of Charge; Machine Learning; Artificial Neural Network.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANN: Artificial Neural Network

AUC: Area Under a Curve

BT: Bagged Trees

CNN: Convolutional Neural Network

EV: Electric Vehicle

FNR: False Negative Rate

FPR: False Positive Rate

I: Current

KF: Kalman Filter

LSTM: Long Short Term Memory

MAE: Mean Absolute Error

RBF: Radial Basis Function

ReLU: Rectified Linear Activation Function

RMSE: Root Mean Square Error

RNN: Recurrent Neural Network

ROC: Receiver Operating Characteristics

RUL: Remaining Useful Life

SOC: State of Charge

SOH: State of Health

SVM: Support Vector Machine

T: Temperature

TNR: True Negative Rate

TPR: True Positive Rate

V: Voltage

CHAPTER I

Introduction

1.1 Background Study

As the electric vehicles continue to rise the need on how to manage the battery systems in the effective way must be introduced. The battery mounted on the electric vehicle plays a major important function, managing well its functions will make the vehicle perform well and bring the safety to the whole car. The State Health of the battery defines some major parameter for the EV such as: the overall cost, performance and the vehicles range.(Rimpas et al., 2022)

In the traditional ways, to monitor the health of the battery the rule based systems that applies some predefined algorithms is implemented. Considering these methods, it is vivid that they are not effective to classify the battery health status. For instance, the battery may work in abnormal way and the system will not be able to predict the situation. (Q. Zhang et al., 2016)

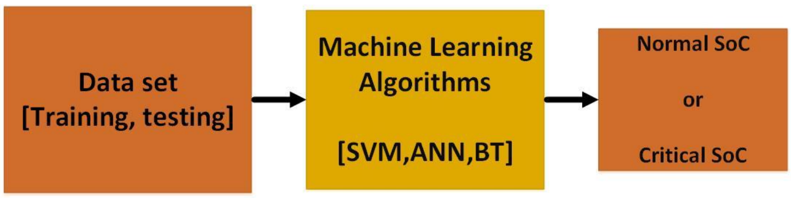

Figure 1 : *Machine Learning Classification*

With machine learning techniques the accuracy for the results are reliable and effective. The advancements of the machine learning have improved the classification for the battery systems. In this project I have proposed the use of three machine learning for the battery health monitoring in electric vehicles which are: Artificial Neural Networks (ANNs), Support Vector Machine (SVM) and Bagged Trees (BT). The Artificial Neural network has been implemented in the related applications due to its ability of being able to study the historical data and make the classifications.(Toughzaoui et al., 2022). Support Vector Machine is a type of machine learning algorithm which learn in a supervision mode. SVM is widely used in regression and classification applications.BT are known as ensemble method that put together the multiple weak learners to produce a classification model.(Fan et al., 2020)

In the recent years, machine learning techniques has been presented as the solution which can replace the rule based methods. With the equipped knowledge of learning from the historical data it is possible to make the classification based on the knowledge given. With this in consideration it is possible to give more accurate result compare to the rule based systems. The machine learning models are able to monitor the degradation of the battery (State of Health) and remaining useful of life battery (State of Charge). (Fan et al., 2020)

Machine Learning algorithms are classified in two categories which are the following: Supervised and Unsupervised methods. With the supervised method the data is labeled while the unsupervised algorithm does not use the labeled data instead they use clustering algorithms to group similar observations.(Roman et al., 2021) The implementation of increasing the range for electric vehicles and reduce several cost that can go along with it is a good step in automotive industry. DE-carbonization is a major challenge the world is facing now; many automotive industries are changing to electric vehicles. The production of electric vehicles has reduced a number of carbon emission. The lithium-ion batteries used in electric cars can face some major difficulties because of the environmental factors. Those factors include: aging degradation, fading capacity and life end. Machine Learning has showed great potential in analyzing the battery state of health and battery state of charge. The accuracy provided the machine learning algorithms are reliable and precise. The fact that, trained model can also predict the operation of the battery in the future is an added advantage of applying machine learning model in the monitoring the battery.(Rauf et al., 2022)

Machine learning capabilities has advanced the estimation of the battery health status. Yongzhi Zhang and Mingyuan Zhao have developed a moving window which can extract the cells data and then predict the aging of the battery. The accuracy of their method is created using machine learning model. as the machine learning has continued to advance its importance has been revealed in the automotive industry.

There are a lot of other ways of controlling and Classifying the battery life. For instance: Battery Management Systems are used in most of the electric vehicles nowadays and the manage all the functions of the high voltage placed in the EV. Battery Management Systems can get all real time information from the battery cells. As the researches, will continue advance in this field Machine Learning algorithms will soon be integrated in most of the electric vehicles. Some big companies like Tesla are using artificial intelligence in their systems.(Y. Zhang & Zhao, 2023).

The objective of this Project is to apply the machine learning techniques to classify the State of Charge and State of Health of the battery in real time. The classification will be done using the data collected from the battery sensors. The Classification

will help us to detect any potential issue that may arise before it may become critical. This Classification will alert the vehicles owner about the battery status. The results that will be produced by this study will provide the effectiveness of the machine learning model in monitoring the battery health status. The outcome of the study will also contribute to the existing methods being applied in Electric vehicles industries. The Machine Learning techniques has been improved and their performance currently are assured to give the real classification. The use of this method in the battery monitoring systems ensures a good performance of the EVs. A big number of situations has occurred on in the past which shows how the EVs has not been able to withstand the malfunction of the battery systems. Battery health monitoring in electric vehicles will enhance the safety of the Battery systems by keeping the car and the owner safe.

My research will implement supervised machine learning. The method will be used specifically on the following machine learning model: Artificial Neural Network (ANNs), SVM (Support Vector Machine) networks, and Bagged Trees (BT). These algorithms have shown to be effective in various classification tasks and are well suited for the task of the battery health monitoring. Before applying any machine learning model to the battery health monitoring systems it is necessary to apply the pre-processing techniques. these techniques used in the pre-processing includes: feature extraction, normalization and outlier detection.

The normalization process means to transform the data and turn them to zero mean and unit variance. The process ensures that all the features are consistent and thereby preventing some features to dominate others. Feature Extraction is the process of extracting certain features in the data as mean of using the data for the Classification purpose. The outlier detection is the process of removing the outliers from the data. These outliers can have the negative impact on the machine learning models. The data will be processed by using the techniques elaborated above. The data will be processed to ensure the quality, suitability and efficient for the machine learning. The which will be processed will be used for training and testing.

1.2 Thesis Problems

1. By using the SVM (Support Vector Machine), Bagged Trees (BT) and Artificial Neural Network (ANN) in real time Classification for the SOH and SOC. Which model will bring the best results in performance?

2. By applying the Real time sensor data including: Temperature, current and voltage in machine learning models, how does the battery health monitoring accuracy can be improved? With this in consideration how will this affect the system in estimation or Classifying the SOH and SOC for lithium-ion batteries?

1.3 Aims of the Research

The main aim of the research is to compare the performance of three machine learning algorithms (SVM, BT, and ANN) for battery health monitoring of lithium-ion batteries used in electric vehicles. The study aims to evaluate the accuracy, efficiency, and computational complexity of each approach using a dataset of battery voltage, current, and temperature measurements from EV usage. The goal is to determine the most suitable algorithm for battery health monitoring, which can improve the safety, dependability, and energy efficiency of electric vehicles by providing precise information about the charge state and health state of the batteries. The research aims to advance the development of battery health monitoring systems for electric vehicles and provide knowledge for future studies in this area.

1.4 Significance of the Research

Battery health monitoring for electric vehicles has some great significance. One is the improved safety of the vehicle, the model system will be able to predict any failure which might occur in the battery and alert the driver beforehand. This would increase the safety of the car, environment and their owner. Secondly, the battery life span of the car is another great major factor to consider. By monitoring the health status of the battery it can help in maintaining well the battery thereby increasing battery lifespan overtime. In case, the user is aware of the usage this will help in maintaining the battery.

Thirdly, battery health monitoring can offer suggestions on how to enhance energy consumption and lower waste by examining data on battery use and performance. Drivers can adapt their driving practices to enhance the economy of their car with the aid of battery health monitoring.

Finally, electric vehicles are still seen as sustainable and they are seen as good alternative for the cars powered by the gasoline. Thought electric cars are sustainable the disposal of the lithium ions batteries can cause some danger to the environment too.Battery health monitoring research will also discover how these batteries can be more sustainable.

1.5 Limitations of the Research

The availability of data is one of this research's main difficulties. Large volumes of data are needed to train machine learning systems that can properly predict battery health. However, due to privacy concerns or proprietary information, access to this material may be restricted. The real-time operation of machine learning algorithms can be quite computationally intensive and include complicated algorithms. Electric vehicles, which have constrained processing capabilities yet demand quick reactions to assure safety and efficiency, may be limited by this. Although machine learning algorithms may be able to estimate battery health precisely, there is always a chance of error or inaccuracy.

Particularly in safety-critical applications like electric automobiles, this can be an issue. Battery health monitoring systems can be expensive to implement, especially for smaller manufacturers or those with constrained funding. This may restrict the use of the technology and impede advancement in the field.

1.6 Thesis Structure

Chapter 1: Introduction

Chapter 2: Literature review, covering battery technology, battery health monitoring techniques, and machine learning algorithms used for predictive maintenance in other researches.

Chapter 3: Methodology, detailing the data collection, preprocessing, and machine learning techniques used in the research.

Chapter 4: Results and discussion, presenting the performance evaluation of the machine learning models and the comparison with traditional battery health monitoring techniques.

Chapter 5: Conclusion and future work, summarizing the key findings of the research and providing recommendations for future work.

CHAPTER II

Literature Review

2.1 Overview

Battery Health Monitoring for electric vehicle using Machine learning will bring an additional contribution to the research made previously in technologies for detecting the age of battery and state of charge (SOC) for the lithium ion cells. While previous studies have developed similar systems for lithium-ion batteries in general, my research will focus on the unique requirements of electric vehicle batteries, such as the need for quick and accurate monitoring to ensure safety and optimize driving range. (Q. Zhang et al., 2016)

The following are various previous researches done on the topic of battery health monitoring. Various methods have been implemented by various researches and ongoing researches are keep adding to the existing ones and my research also will contribute to the current ones. The method that will be used are based on machine learning and three machine learning algorithms will be applied. Those algorithms are: Support Vector Machine, Neural Network model and Bagged trees which uses ensemble learning. (Driscoll et al., 2022)

2.2 Kalman Filter Based SVM

This article outlines a model for Lead acid regulated valve batteries it then demonstrates the use of techniques for state estimating the battery state of Charge and battery state of health in real time mode. The approach employs a Kalman filter (KF) using a machine learning algorithm named as support vector machine.

This method considers factors related to the battery's capacity using input vector with multi-dimensional, and then the time series components is trained using SVM to predict any new data which might be introduced on the model later on. The KF predictor is applied to the SVM innovation to perform step-by-step prognosis. The method proposed by this author uses a real-time data and proves to be effective in estimating the battery state of health and battery state of charge for the Lead Acid Regulated cells with an accuracy equal to 4% for determining the SOH and its deterioration trend.(Chang & Xiaoluo, 2011)

2.3 Model Classification using Machine Learning Models

Machine learning field is a branch for the artificial intelligence. From machine learning algorithms it is possible to build different algorithms which can learn from data given and improve the learning process. Machine learning models are able to take decisions based on the experience it has and knowledge supplied to it. Machine learning algorithms does not need to be reprogrammed again to work. When it comes in Classifying the parameters for any energy storing device machine learning models have more efficient ways for Classifying.

The authors proposed two different methods on how to find the battery health for lithium ion cells. Those methods include: model bases on physics and models based on equivalent circuit. But these methods presented that the current has some limitation. Therefore, the authors then have suggested the machine learning models which can give better responses for battery SOC and SOH Classification. this paper also provides some overview on how real time Classification for battery can be optimized.(Chang & Xiaoluo, 2011)

2.4 Model Classification using Artificial Neural Network (ANN).

The author in this study describes the application of Neural Networks to predict the battery health status for the lithium cell. Battery packs are critical and expensive components of electric vehicles, so comprehensive monitoring and control is essential. By using the artificial neural network algorithms is trained on experimental information's from a commercial (3.6V) battery cells using SOC/OCV relationship derived from the R-RC model. The model get various parameters as input which include (current, voltage, temperature) and then deliver the State of Charge as output.

Then the model which have been trained using neural network is tested with drive recorded drive cycles to accurately estimate battery SOC. This paper emphasizes that a well-trained single-layer neural network can capture nonlinear properties of batteries and estimate unmeasurable parameters such as battery SOC levels. The use of ANNs in her battery health status Classification is promising and offers a potential solution for comprehensive battery monitoring and control in electric vehicles.(Ismail et al., 2017)

2.5 Model Classification using CNN with Parameters Optimized

The author continues to describes a principle for accurately Classifying the battery health status (RUL) used for the lithium, which is crucial for ensuring the safety for their applications. The authors then give details about aging mechanism of lithium cells which brings a challenge to current technologies. Further on, they propose a convolutional neural network model for the Classification of the battery health status of the battery cells. The CNN is optimized using an orthogonal method, which is known to reduce the training time of the model.(Li & Yang, 2020)

The method which is proposed is confirmed using the RUL Classification and large dataset is reported to exceed 90.9 %. The RME and MAE are reported to be limited to 35.1 and 13.7, respectively. The authors state that their proposed method is suitable for Classification Lithium Remaining Useful time for batteries used in energy storage devices and EV. (Li & Yang, 2020)

2.6 Model Classification using Convolutional Neural Network Approach

(Fan et al., 2020) The intelligent and reliable battery state of health estimation is very critical when considering the use and safety for the Lion batteries. The application for this is being employed different areas such: smartphone, electric vehicle and smart grid. The author for this paper indicate that a brand-new structure built on a convolutional neural network that uses data on

voltage, current, and temperature gathered during battery charging to directly estimate SOH. The proposed algorithm has been trained using cells data numbered to twenty-eight which were aged up to two different temperatures with a random profile usage. The effects of different CNN configurations with layers numbered from one to six and neurons between 32-256 neurons are explored, the accuracy was improved by augmenting the training data with noise and errors. Significantly, partial charges, which are typical in many real-world applications, were taken into account when validating the suggested Convolution Neural Network model. While using a very limited charge window of 85 percent to 97 percent for indicating the battery state of charge, the suggested algorithm is Convolution Neural Network model SoH estimate methodology still managed a respectable MAE of 1.6 percent over the course of the battery life.

Even though it was demonstrated that the suggested Convolutional Neural Network algorithm can estimate State of Health effectively for data with partial charge and 2 temperature values, more research can examine other ranges of temperatures, various power charging with a constant and charge currents. Overall, the Convolution Neural Network-based framework offers a reliable and efficient approach for calculating SOH straight from charging data, which is crucial for guaranteeing the secure and dependable operation of Li-ion cells in a variety of applications.(Fan et al., 2020)

2.7 Model Classification using Forest Approach for Lithium-ion Battery

(Mawonou et al., 2021) Although this author proposes this system there can be challenges to apply them to the lithium-ions batteries which are used in electric vehicles. It is recommended to use machine learning models as they are the ones which can be implemented on batteries used in electric vehicles. One of the key concerns for users of electrified vehicles is maintaining a power supply with consistent range for the lifetime of an electric vehicle. Lithium cells has an important feature toward achieving this objective, but assessing the health of these batteries is essential for ensuring their performance and longevity. The State of

Health for lithium cells assessment is an assured method for achieving this. The author of this paper presents the indicators for aging which are numbered to two. The indicators delivered from the collected data stored during driving and charging events. These indicators make assessment for the State of Health with the error which have low estimation, in addition to this a forest algorithm is introduced to predict how the battery perform in its aging process which then produces the battery state of health estimation error up to 1.27%. Collecting data and ranking analysis could not be enough to bring the best in Classifying the age of the batteries. Real time monitoring system is advised here to make the systems works at its best. (Mawonou et al., 2021)

2.8 Model Classification using Novel Informed Deep Learning

The battery cells used in electric vehicles imposed some safety issues in operation, and accurate estimation of its condition and battery health status is very important in energy storage management of the batteries. In recent years, deep learning-based Classification frameworks have become established for this task due to their high accuracy. This paper combines empirical models and deep neural networks to propose a new approach to predict the battery health status for lithium cells by applying impedance-related properties.(Kim et al., 2022) The proposed framework includes three phases including feature gathering, model training, and probabilistic Classification. The model is protected with Monte Carlo dropouts to ensure robustness and reliability.

The proposed approach significantly outperforms basic deep neural networks in terms of accuracy and robustness. Hierarchical relevance propagation is used to track feature importance and provide scientific justification for output. The outcome of this research carry some valuable insight into the use of deep neural networks for monitoring battery health and for power and energy decision-making and remediation planning used in the electric vehicles cells battery pack. It shows the importance of an explainable, uncertainty-based pipeline.(Kim et al., 2022)

2.9 Model Classification using Recurrent Neural Network (RNN)

Zhang, Yongzhi and Zhang, Yongzhi proposed a model for Classifying the battery health status by using the algorthim of Recurrent Neural Network. The RUL Classification has shown an effective performance in assessing the battery usefulness and reducing the risks which might arise, as it can determine the advent of failure. However, the current remaining useful life Classification techniques learning are inefficient when it the dependencies are very long in determining the battery degradation.

(Toughzaoui et al., 2022)The authors continue by suggesting that employing deep learning approaches to predict the health status of the battery will tackle the problem in an efficient way. To know which long phase dependencies that involves in the lithium cells degradation, two models plays a major role. Those models are RNN (Recurrent Neural Network) and long short term Memory. In order to advance the RNN (Recurrent Neural Network) and LSTM (Long Short Term Memory) a specific method is used namely mean square back propagation.

The created models RNN (Recurrent Neural Network) and LSTM (Long Short Term Memory) can be able to generate an explicitly capacity-oriented Remaining Useful life predictor and this get the information about the degraded capacities for cells. The filter model is introduced to for the SVM and RNN which are then contrasted with learning ability of the RNN and LSTM. A probabilistic battery health status forecast is produced using Monte Carlo simulation. To verify, compare, and build models. In case the training data is offered online, the new method can estimate the battery health status independent of any data, and it can do so earlier than using conventional techniques. In general, the article implies that by identifying long-term connections among the deteriorated capacities, deep learning approaches, in particular RNN (Recurrent Neural Network) and LSTM (Long Short Term Memory), can be useful in Classifying the battery health status of the battery (RUL). The suggested approach may enhance battery dependability assessment and lessen battery danger.(Toughzaoui et al., 2022)

2.10 Model Classification using Hybrid K-means CNN-Long Short Term Memory

Toughzaoui and Yassine discusses the use of lithium batteries in numerous areas and the need for real-time monitoring of Remaining Useful Life. There are two methods which needs to employed to detect the Remaining Useful life which includes the methods based on data provided and another one is model-based methods. Artificial neural networks have shown good performance in Classification of State of Health and RUL in several studies. Artificial Neural Network has some good performing models which includes: Long Short Term Memory and Recurrent neural network. Moreover, convolutional neural networks have shown to function well while processing time series data. Some studies propose combining different models for Classifying the Battery State of Health, such as CNN LSTM hybrid algorithm. This author also discusses the use of the NASA open source dataset to perform the evaluation and do the training of the model using three metrics: RMSE (Root Mean Square Error), and MAE (Mean Absolute Error). The author concludes with a discussion on the model development in Classifying remaining capacitance values until the battery end is achieved.(Toughzaoui et al., 2022)

Classification for health status of battery uses a multiple machine learning algorithms which includes the following: Bagged trees (BT), Artificial Neural Network (ANN) and Support Vector Machine(SVM). Comparative analysis of the performance of each model for battery health monitoring will be provided. This will help identify robustness and flaw for those various machine learning algorithms and determine which one is most suitable for this particular application. My research paper will also build on previous work in feature engineering and selection, as well as model calibration, to optimize the operation of the algorithms for battery health monitoring.

This will help to ensure that the models are accurate and reliable in estimating the battery health status. By making suggestions on how to create robust data-driven models for estimating battery SOH and highlighting the need of reliability boundaries around the assumption, the research will contribute to the broader field of health monitoring and diagnostics for lithium-ion batteries. This might provide guidance for the creation of comparable systems for other crucial components that need real-time SOH estimation.

CHAPTER III
Methodology

3.1 Research Design

In this study, machine learning algorithms are used to check the battery health for electric vehicles using machine learning. Specifically, three popular machine learning models will be implemented: Support Vector Machine (SVM), Bagged Trees (BT), and Artificial Neural Network (ANN), to predict the battery's health status. Data for battery pack be collected from various data centres which provides electric vehicles battery historical data. Then, the most important features that affect battery health will be selected and be trained on each model using the pre-processed data. The working condition of each algorithm model will be evaluated by using various metrics such as precision, recall, accuracy and F1 score and select the best-performing. Finally, the algorithms will be validated by testing it under different conditions and continuously monitor and improve it by incorporating new data and optimizing the training process**.**

3.2 Data Collection

To get the data for my experiment I researched through various online research data centres and found Kaggle platform

. Data scientists and machine learning experts now frequently use Kaggle as a platform for learning, working together, and competing in real-world data problems. More Collaborators along with Scholarly keep adding the data for research purposes. The datasets relevant to my research question was picked and downloaded from Kaggle platform. The datasets contained information on battery health parameters such as voltage, temperature, and state of charge for electric vehicles.

To be able to use the data they had to be cleaned and processed to ensure that the algorithms are getting the exact inputs. This involved removing missing values, normalizing the data, and creating new features. By using Support Vector Machine (SVM), Boosted trees (BT), and Convolutional Neural Network (CNN) it was possible to validate the results of my analysis by comparing them with existing literature and conducting sensitivity analyses. (*BATTERY AND HEATING DATA IN REAL DRIVING CYCLES | Kaggle*, n.d.)

In general, the data used was obtained through an online platform named Kaggle. Relevant datasets were selected to be cleaned, pre-processed, and analysed to answer my research questions.

3.3 Machine Learning Model Development

To be able to make classification there are steps which are normally used in machine learning development. The first step taken is to define the problem which includes identifying the relevant data sources and specifying the types of outputs to be generated.

Once the problem is defined, data were then collected and prepared. To produce the best results, the data had to be cleaned and transform them into the desired input. Then data must also be separated into validation set, testing set and training set.

After preparing the data, machine learning algorithms that will be used to build the models were selected which are: Artificial Neural network, Support Vector Machine and Bagged Trees(BT). Then algorithms were evaluated based on their performance on the training and validation sets, as well as their suitability.

Furthermore, a new prepared data set was used to optimize the performance and working conditions of the model algorithm. After training the models, they were evaluated their performance on the testing data. This also ensures all new data will be Classified with high accuracy.

When the trained model has resulted in good results they can be used to make classification on the unseen data. This involves integrating the models into a larger system and ensuring that they are capable of handling new data in real-time.

In summary, first the problem is to define, prepare and then collect the data, select the models and train them, evaluate their performance, deploy them to make classification and Classification on new data. (*Machine Learning Development | Machine Learning Consulting*, n.d.)

3.4. Data Pre-processing

The data pre-processing stage includes machine preparing the data for machine learning. The main importance of doing pre-processing is to improve the accuracy, precision and quality for the data. Pre-processing also helps make it easier for the interpretation and analysing the data.

The steps include in Data-pre-processing are follow.

- Handling Missing data: In any dataset there might be some missing data which can give wrong results in training the model. There are some techniques used to handle missing such as data imputation and removal. Data imputations uses some techniques of calculating the mean, mode, finding the average to predict any missing value based on the given features. While removal involves in deleting the columns and rows that contains the missing values.
- Removing Duplicates: in various data set we can have duplicates rows and columns and this can cause inaccuracy for any machine learning model. It is better to remove the duplicates in data. There are some methods used in removing duplicated which includes comparing the rows and columns and then remove duplicates. Another method used is clustering which involves in grouping similar rows and columns and remove the duplicates.
- Outliers: Outliers are included in the dataset and this data have a significant difference with other data. This can create distortion and make the learning machine algorithms give bad results. The way of removing the outliers is to use the median or mean value of other features.
- Feature scaling: Feature scaling includes normalizing the data into a common scale. For instance, the normalization range can be from 0 to 1. Some algorithms such SVM and KNN can perform bad if features have different ranges of values. The techniques involve in feature scaling includes: Normalization standalirzation, min-max scaling.
- Feature Engineering: Feature engineering is the process of creating the new featured from the one which already exist. The feature engineering can be achieved by applying mathematical formulas such as logarithms or square root. For instance, two number can be multiplied to create a new feature.

- Evaluation Metrics: the evaluation metrics are used to analyse and compare the performances of the machine learning algorithms. There are some evaluation metrics which are commonly used such as: accuracy, precision, F1 score, recall and ROC.

3.5 Implemented Machine Learning algorithms

3.5.1 Artificial Neural Network (ANN)

Neural Networks are also named as Artificial Neural Network. These kind models are mostly used in deep learning application. Neural Network algorithm mimics the brain of a human and they use supervised machine learning. By the using the hidden layer it is possible to learn the features of the data and classify them to their class.

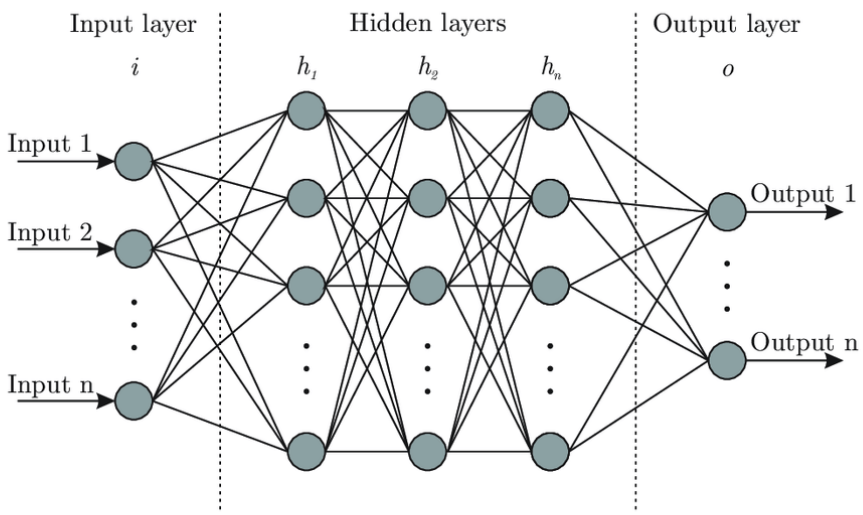

Figure 2 *: Artificial Neural Network*

The methodology which must be applied in training the Neural network model involves various steps and they are listed below:

1. Data Preparation: The input data which might be training set or testing set are processed and in this steps missing values, feature engineering, feature scaling is managed.
2. Architecture Design: By establishing the required number of layers, the number of neurons for each layer, and the activation function for each layer, the architectural of NN is formed.

3. Initialization: The weights and biases have to be initialized in a random way. These will break the symmetry for the NN network and then proceed with training.
4. Forward Propagation: The training data will be fed at the input of the neural network algorithm model and this data will pass through each layer to be processed and then an activation function will be used to give the Classified outcome.
5. Loss Calculation: During training or testing there is some difference between the actual output and Classified output. To find their difference a loss function must be applied. This loss function is also named a cost function. There are some popular loss functions used in this process which includes: categorical cross-entropy, cross-entropy and Mean Square Error. This loss function will be used depending on the nature of question. The problem might be (regression, classification, multi class classification)
6. Backward Propagation: The weights and biases of the neurons must be updated to minimize the loss of the model. backpropagation must be applied in to adjust the parameters of the NN model.
7. Hyper parameter Tuning: Neural network models have various parameters to be tuned which includes: number of epochs, learning rate and droop out methods.
8. Model Evaluation: The trained neural network is used on the testing set, validation set to examine its performance. The effectiveness of the models is evaluated using several evaluation metrics. Those metrics includes: ROC, recall, F1-score, precision and accuracy.
9. Classification: A new dataset can be provided to the neural network model's learnt biases and weights after training. The data will be supplied in a forward direction.

Neural Networks can be very useful in capturing the complexity of the patterns. Neural network requires the tuning hyper parameters for achieving the optimal performance. Numerous applications, including speech recognition, natural language processing, image recognition, and others, use neural networks algorithms. (*Types of Neural Network Algorithms in Machine Learning with Examples*, n.d.)

3.5.2 Bagged Trees (BT)

Bagged trees algorithm is included in the ensemble learning methods for machine learning. Bagging means the bootstrap Aggregation in other words this mean the ensemble methods. The ensemble methods mean in other words layering the data, algorithms and models. The bagged trees are delivered from the decision tree. It was observed that the decision tree like to give high variance. Bagged trees is created by combining the decision trees and then the average will be made to find the output. The bagged trees can be composed with 10 trees, 30 trees, 60 trees and more. (*Bagged Trees: A Machine Learning Algorithm Every Data Scientist Needs | by Rob W | Towards Data Science*, n.d.)

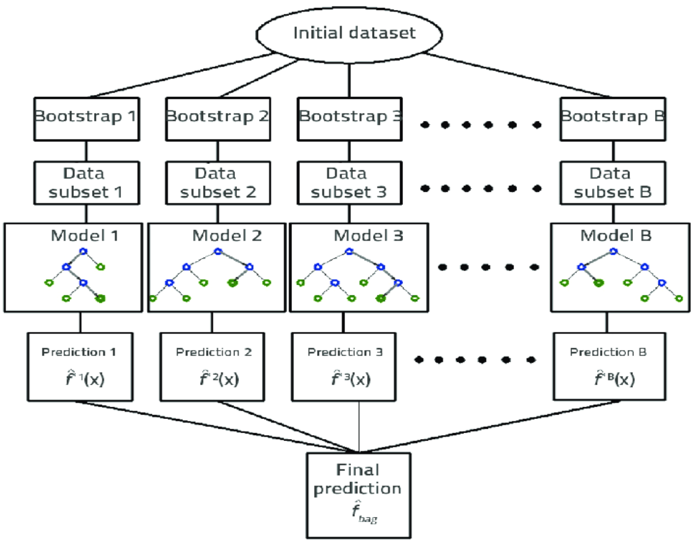

Figure 3 : *Bagged Trees*

The bagged trees are methodology follow the steps below:

1. Data Preparation: The input data which might be training set or testing set are processed and in this steps missing values, feature engineering, feature scaling is managed.

2. Bootstrap Sampling: Through bootstrapping, random subsets of the training data are produced. Each subset, referred to as a "bootstrap sample," is the same size as the initial training data, but some samples have been repeated and others have been left out.

3. Base Learner Training: Each independent bootstrap sample is used to train a fundamental learning algorithm, such as a decision tree. Multiple base learners are produced as a result, each with their own set of learnt rules and classification.

4. Ensemble Creation: The final classification is created by combining the classifications from each base learner. Voting for classification problems or averaging for regression problems can be used to determine the ensemble Classification.

5. Evaluation and Classification: A different validation or test dataset is used to evaluate the bagged trees model's performance. When the model is deemed adequate, classification based on fresh, unforeseen data can be made using it.

When compared to a single decision tree, bagged trees are notable for their capacity to reduce overfitting by averaging the classification of several base learners, which can enhance accuracy and generalization performance. Bagged trees can accommodate a high number of input features and are robust to noisy data. They might not be as easy to comprehend as a single decision tree, and for optimum performance, the ensemble size and base learner parameters might need to be adjusted.(Chandran et al., 2021)

3.5.3 Support Vector Machine (SVM)

Support Vector Machines (SVM) is a supervised learning algorithms and it is generally used for regression and classification tasks.

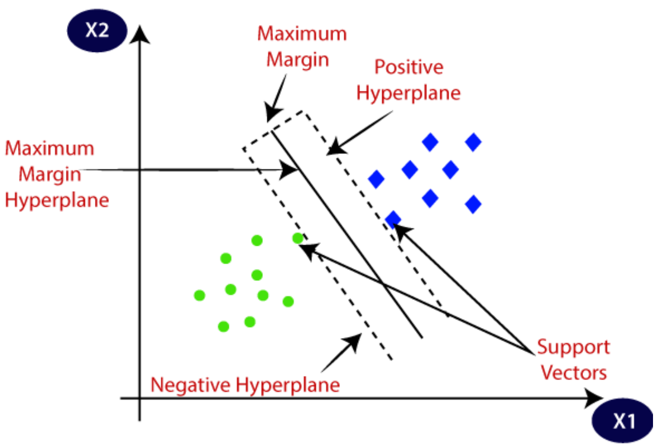

***Figure 4** : Support Vector Machine*

The methodology for the SVM includes the following steps:

1. Data Preparation: Pre-processing is performed on the initial training data, which may include addressing missing values, feature scaling, and feature engineering if necessary.

2. Feature Representation: Each sample is represented by a vector of values according to its features in the feature vector representation of the data.

3. Labelling: Each sample in the training data is marked with the goal value or class to which it belongs.

4. Training: SVM determines a decision boundary, also known as a hyperplane, as the one that best separates the samples into several classes. The SVM approach aims to maximize the margin, which is the separation between the hyperplane and the closest samples from every group.

5. Kernel Transformation (Optional): If the data cannot be divided linearly, SVM can transform the feature vectors into a higher-dimensional space with the use of a kernel function, enabling the division of the samples along a linear hyperplane. The radial basis function (RBF), sigmoid, linear, and polynomial kernel functions are often used.

6. Model Evaluation: The effectiveness of the trained SVM model is evaluated using a different validation or test dataset. Common evaluation criteria include accuracy, precision, recall, the F1-score, and the area under the receiver operating characteristic (ROC) curve.

7. Hyperparameter Tuning: To maximize the model's performance, SVM contains a number of hyperparameters that need to be adjusted, including the kind of kernel, kernel parameters, regularization parameter (C), and class weights.

8. Classification: By applying the learnt decision boundary to the feature vectors of the samples, the SVM model can be trained and improved to make classification on new, unforeseen data.

SVM is renowned for its capacity to handle both linearly and non-linearly separable data as well as for its use of kernel functions to detect intricate patterns in the data. SVM may be computationally expensive, though, especially when working with huge datasets, and careful hyper parameter optimization is necessary for it to perform at its best. In comparison to simpler algorithms like decision trees, the interpretability of SVM models can be constrained because they often produce a binary decision border without explicit feature importance metrics.(*Support Vector Machine (SVM) Technique - Creative Biolabs*, n.d.)

In General Machine Learning Development includes the following steps:

1. Data preparation
2. Data Pre-Processing
3. Model selection
4. Model training
5. Model tuning
6. Model evaluation
7. Model deployment.

CHAPTER IV

Findings and Discussion

4.1 Results and Findings

As a result of this research, Neural Networks (NN), Support Vector Machines (SVM), and Bagged Trees (BT) were clearly demonstrated. All these models have performed well in Classifying the State of Charge (SOC) for an electric vehicle battery. Among the models used for this study the Neural Network has achieved a higher accuracy of 99.1 % followed by Bagged trees at 98.4 % and the SVM model at 92.5 %.

The confusion matrices have demonstrated that the algorithm can classify and predict well the SOC of the battery as either Normal or Critical. For all the three models the Neural network and Bagged trees has a good number of True positive rate compared to the SVM. Neural Network model has performed well both in training and testing compared to other models. The negative outcome is Classified as ‘Critical SOC’ where as the positive outcome is Classified as the ‘Normal SOC’. When the model gives the output of Normal SOC it means that SOC of the battery is in the good conditions. When the model gives the output of Critical SOS it means the battery health has some problems which might lead to battery degradation. Which also means that, current may be getting very low; voltage might be rising above expectation or the temperature might higher.

The RoC curves were also created for each algorithm to analyze and visualize their performance. The ROC curves for all the three algorithm used show a high area under a curve (AUC), this indicates that the models have a good discriminatory power. ROC curves has a low discriminatory power compared to the curves produced the other two models.

Finally, the scatter plots were also created to analyze and visualize the relationship between the Classified SOC outcome and the predictors which are: Current (I), Voltage (V) and Temperature (T). The scatter plots Cleary indicates that the predictors are correlated with Classified SOC outcome, with a separation between the two outcomes which are “Normal SOC” and “Critical SOC”.

All in all, the outcome of the study indicates the Neural Network algorithm with medium layers applied is the most effective in making the Classification of SOC for the electric vehicles batteries. Though the Neural Network performs well than others each algorithm presented in this study has a good accuracy on the data supplied to it.

4.2 Data Analysis

The dataset used consisted of 1500 observations with 3 features. The predictors are Voltage, Current and Temperature. The response of the model must be either ‘Normal SOC’ or Critical ‘SOC’. The data was pre-processed by scaling and splitting the data into testing and training sets using 70% (1050) samples for training and 30 % (450) samples of testing. For each model the data was trained using training set and tested using testing data.

4.3 Model Performance

The performance of the models utilized in this investigation is clearly displayed in tables 1 and 2. Table 1 displays the effectiveness of neural network (NN), support vector machine (SVM), and bagged tree (BT) models on training data, whereas table 2 displays the effectiveness of the same models on testing data. The Neural Network have achieved the highest accuracy.

Table 1 : Models Performance on Training data (Accuracy)

Neural Network	99.1%
Bagged Trees	98.4%
Support Vector Machine	92.5%

Table 2 : Models Performance on the testing data (Accuracy)

Neural Network	77.5%
Bagged Trees	71.6%
Support Vector Machine	63.6%

4.4 Neural Network Model Classification Parameters

Table 3 : Neural Network Model Performance Parameters

Accuracy	99.1%
Classification Speed	32000 obs/sec
Training time	8.5477 sec
Neural Network type	Medium Neural Network
Number of fully connected layers	3
First layer size	25
Second layer	20
Third layer size	20
Activation function	ReLU
AUC	0.99

4.4.1 Confusion Matrix for Neural Network Model

The neural network achieved an impressive accuracy of 99.1%, indicating that it is making correct classification for the majority of samples. The True Positive Rate (TPR) for critical SOC was 99.5%, and for normal SOC it was 97.2%, indicating that the model is correctly identifying the majority of samples belonging to these classes. The False Negative Rate for normal SOC was 2.8%, and for critical SOC it was 0.5%, indicating that the model is correctly identifying most of the positive samples.

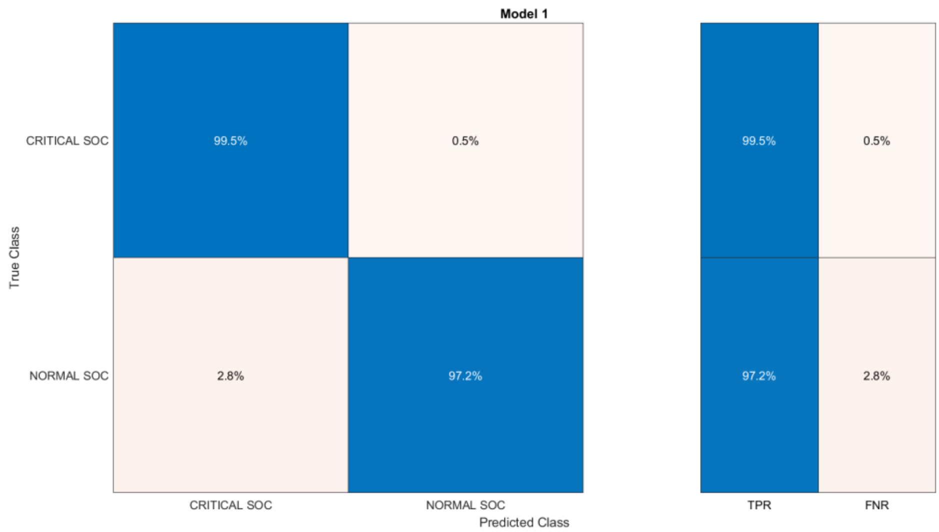

Figure 5 *: TPR and FNR for Neural Network Model*

The neural network achieved an impressive accuracy of 99.1%, indicating that it is making correct classification for the majority of samples. The True Positive Rate (TPR) for critical SOC was 99.5%, and for normal SOC it was 97.2%, indicating that the model is correctly identifying the majority of samples belonging to these classes. The False Negative Rate for normal SOC was 2.8%, and for critical SOC it was 0.5%, indicating that the model is correctly identifying most of the positive samples.

Overall, the neural network's performance is excellent in terms of accuracy, TPR, and false negative rate, suggesting that it is accurately classifying the majority of samples for both normal and critical SOC classes, with low false negative rates.

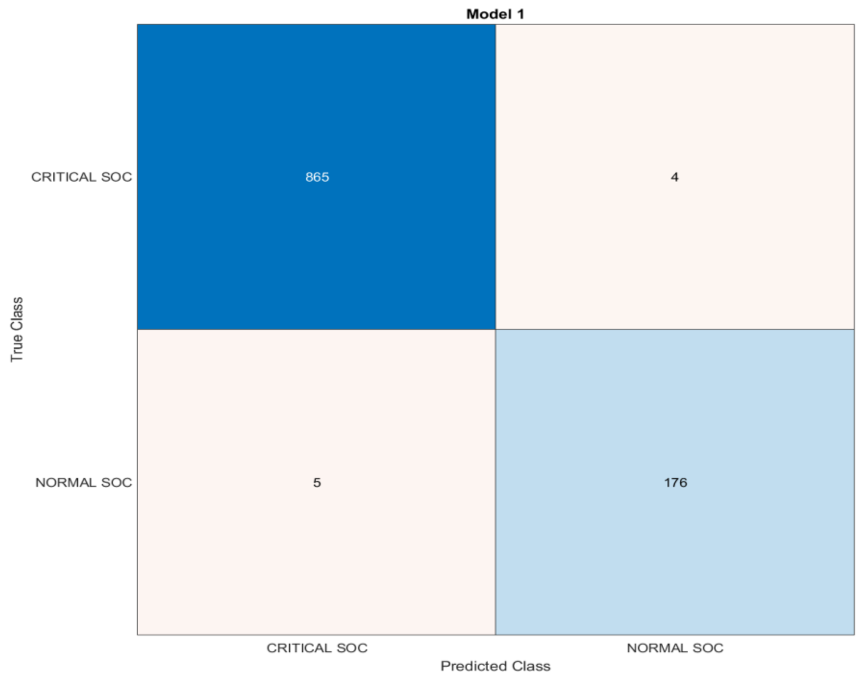

Figure 6 *: : Confusion Matrix for Neural Network Model*

Finally, electric vehicles are still seen as sustainable and they are seen as good alternative for the cars powered by the gasoline. Thought electric cars are sustainable the disposal of the lithium ions batteries can cause some danger to the environment too. Battery health monitoring research will also discover how these batteries can be more sustainable.

The dataset consisted of 1500 samples, out of which 1050 were used for training the neural. The model classified 865 samples as critical SOC and 176 samples as normal SOC.There were 5 samples from the normal SOC class that were misclassified as critical SOC, and 4 samples from the critical SOC class that were misclassified as normal SOC. Despite these misclassifications, the overall performance of the neural network was still excellent, with an accuracy of 99.1% and high TPR values for both critical SOC (99.5%) and normal SOC (97.2%).

It's important to further investigate and address the misclassifications to improve the model's performance, such as through fine-tuning, feature engineering, or adjusting the decision threshold.

4.4.2 Scatter Plot for Neural Network Model

On the test dataset, which included 450 samples (1500 - 1050) that were not used for training, a scatter plot was made to show how well the neural network performed. Each data point in the scatter plot represented a sample, with the true labels (critical SOC or normal SOC) displayed on the y-axis and the Classified labels (critical SOC or normal SOC) on the x-axis.

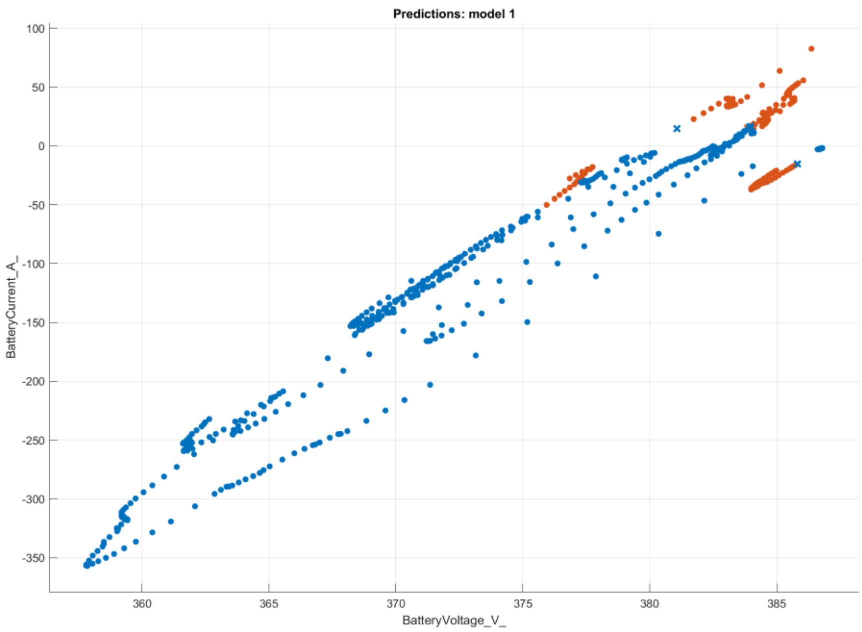

Figure 7 *: Scatter Plot for Neural Network Model*

The scatter plot revealed that the majority of samples were clustered around the diagonal line, indicating that the neural network's classifications were in agreement with the true labels for most of the However, there were a few misclassified samples that were located off the diagonal line, indicating instances where the Classified label did not match the true label.

Specifically, there were 5 samples from the normal SOC class that were misclassified as critical SOC, and 4 samples from the critical SOC class that were misclassified as normal SOC, as indicated by the confusion matrix. samples.

4.4.3 ROC Graph for Neural Network Model

ROC (Receiver Operating Characteristic) curve was generated to evaluate the performance of the neural network, using the Classified labels and true labels from the test dataset.

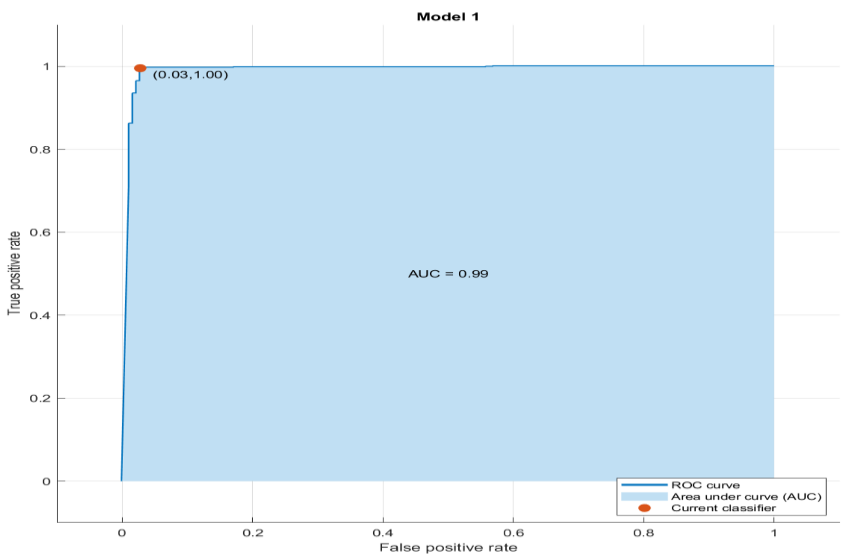

Figure 8 *: ROC Graph for Neural Network Model*

At various classification thresholds, the ROC curve showed the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR), with TPR on the y-axis and FPR on the x-axis.

The ROC curve demonstrated excellent performance, with an AUC of 0.99, which is close to the perfect number of one. A higher AUC suggests that the model can distinguish between positive and negative samples more accurately.The point on the ROC curve where it meets the top left corner (0.03, 1.00) represents a perfect classification scenario where the model achieves a TPR of 1.0 (100%) while maintaining a low FPR of 0.03 (3%).

This implies that the neural network has a high True Positive Rate and a low False Positive Ratio for successfully categorizing the critical SOC and normal SOC data, this indicates indicating excellent discriminatory power.

4.5 Support Vector Machine Model Classification Parameters.

Table 4 : Support Vector Model Classification Parameters

Accuracy	92.5%
Classification Speed	28000 obs/sec
Training time	1.7001 sec
SVM model type	Linear SVM
Kernel function	Linear
Kernel scale	Automatic
Box constraint level	1
Multi Class method	One-vs-one
AUC	0.76

4.5.1 Confusion Matrix for Support Vector Machine

The SVM model achieved an accuracy of 92.5% on the test dataset, which consisted of 450 samples (1500 - 1050) that were not used for training.The confusion matrix revealed that out of 869 critical SOC samples, the SVM correctly classified 869 samples as critical SOC, resulting in a True Positive Rate (TPR) of 100% for critical SOC.

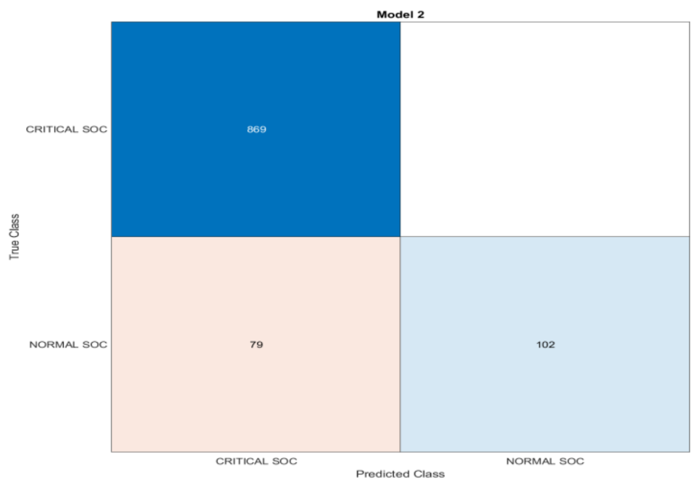

Figure 9 : *Confusion Matrix for Support Vector Machine*

Similarly, out of 102 normal SOC samples, the SVM correctly classified 102 samples as normal SOC, resulting in a True Negative Rate (TNR) of 100% for normal SOC.

However, there were 79 samples from the normal SOC class that were misclassified as critical SOC, resulting in a False Positive Rate (FPR) of 79% for normal SOC. Importantly, there were no misclassifications of critical SOC samples as normal SOC, resulting in a False Negative Rate (FNR) of 0% for critical SOC.

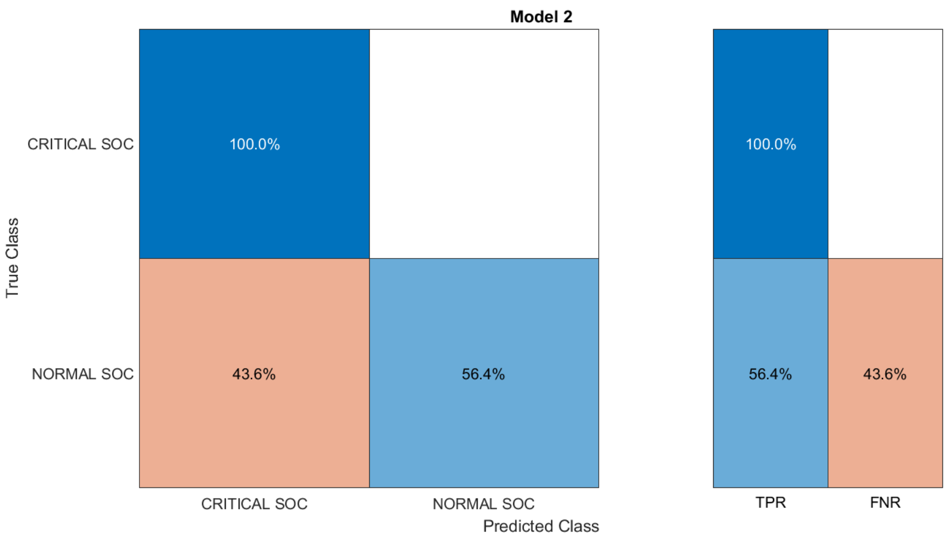

Figure 10 : *TPR and FNR for Support Vector Machine*

The SVM model achieved a perfect TPR of 100% for critical SOC samples, indicating that it correctly classified all critical SOC samples as critical SOC. This indicates a high sensitivity of the model in detecting critical SOC samples.

However, the SVM model had a lower TPR of 56.4% for normal SOC samples, indicating that it correctly classified only 56.4% of normal SOC samples as normal SOC. This indicates a lower sensitivity of the model in detecting normal SOC samples compared to critical SOC samples**.**

The FNR, which the percentage of samples from a particular class that were misclassified as the opposite class, was 43.6% for normal SOC samples and 0% for critical SOC samples by the SVM model. This means that the SVM model had a high rate of false

4.5.2 Scatter plot for Support Vector Machine

A scatter plot was created to visualize the performance of the SVM on the test dataset, with the Classified labels (critical SOC or normal SOC) on the x-axis and the true labels on the y-axis.

The scatter plot revealed that most of the samples were clustered around the diagonal line, indicating that the SVM's classifications were in agreement with the true labels for most of the samples.

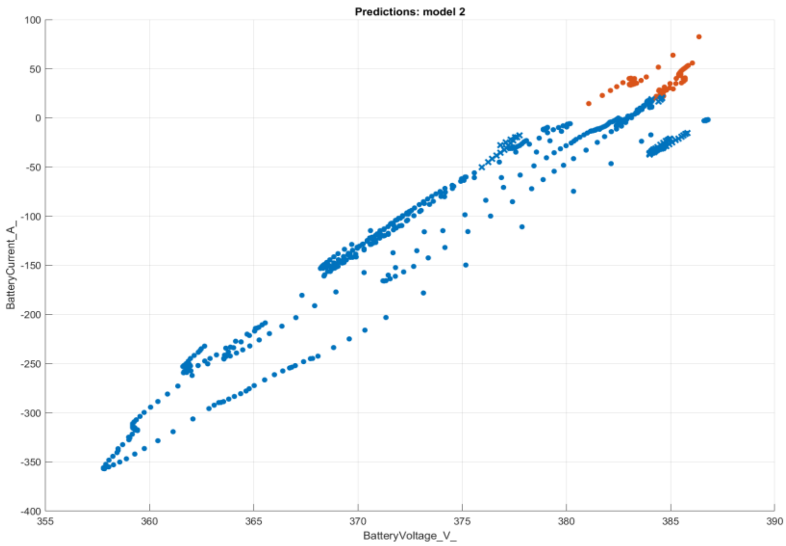

Figure 11 : *Scatter Plot for Support Vector Machine*

However, there were some misclassified samples that were located off the diagonal line, indicating instances where the Classified label did not match the true label. Specifically, there were 79 samples from the normal SOC class that were misclassified as critical SOC, as indicated by the confusion matrix.

4.5.3 ROC plot for Support Vector Machine

Using the test dataset's true labels and Classified labels, the ROC curve was created to assess the SVM's performance. The ROC curve, with TPR on the y-axis and FPR on the x-axis, showed the trade-off between the TPR and FPR at various categorization levels.

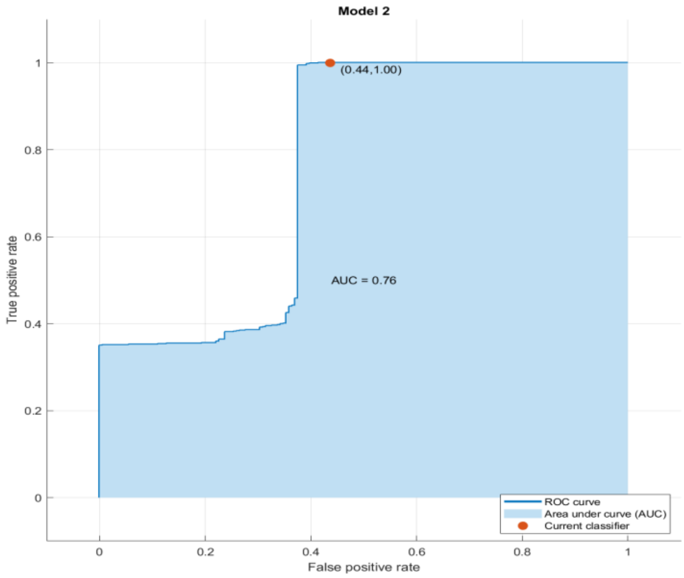

Figure 12 : *ROC curve for Support Vector Machine*

The ROC curve demonstrated moderate performance, with an AUC of 0.76, which indicates fair discriminatory ability of the model to correctly classify positive and negative samples. The point on the ROC curve where it meets the coordinates (0.44, 1.00) represents the classification threshold where the SVM achieves a TPR of 1.0 (100%) while maintaining an FPR of 0.44 (44%).This suggests that the SVM is capable of accurately classifying critical SOC samples with a high TPR, but with a relatively higher FPR for normal SOC samples.

4.6 Bagged Trees Model Classification Parameters

Table 5 : Bagged Trees Model Classification Parameters.

Accuracy	98.4%
Classification speed	3900 obs/sec
Training time	4.3311 sec
Model type	Bagged trees
Ensemble method	Bag
Learner type	Decision tree
Maximum number of splits	1050
Numbers of learners	30
AUC	1

4.6.1 Confusion Matrix for Bagged Trees.

The Bagged Trees model achieved an accuracy of 98.4% on the test dataset, which consisted of 450 samples (1500 - 1050) that were not used for training. The confusion matrix revealed that out of 862 critical SOC samples, the Bagged Trees model correctly classified 862 samples as critical SOC, resulting in a True Positive Rate (TPR) of 100% for critical SOC.

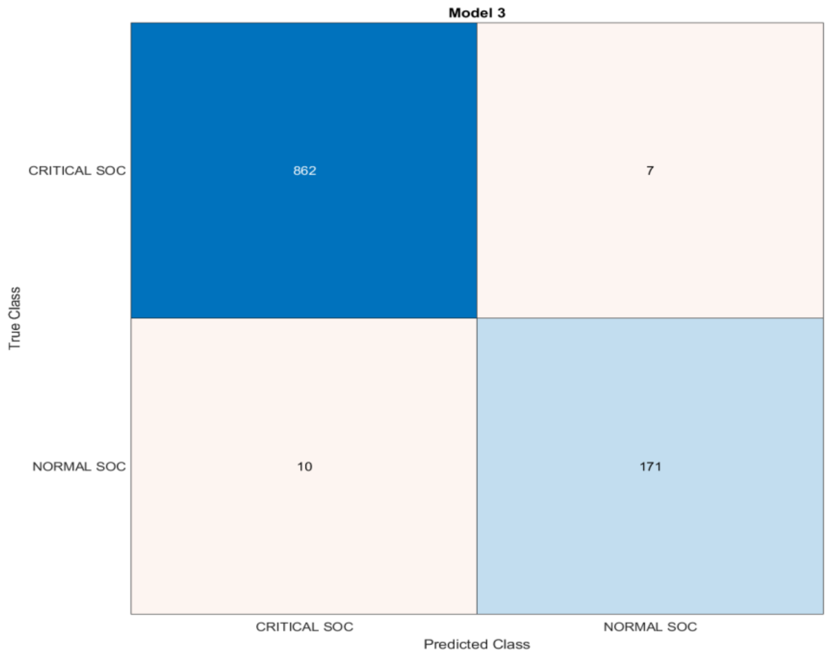

Figure 13 *: Confusion Matrix for Bagged Trees*

Similarly, out of 171 normal SOC samples, the Bagged Trees model correctly classified 171 samples as normal SOC, resulting in a True Negative Rate (TNR) of 100% for normal SOC. However, there were 10 samples from the normal SOC class that were misclassified as critical SOC, resulting in a False Positive Rate (FPR) of 5.8% for normal SOC. Additionally, there were 7 samples from the critical SOC class that were misclassified as normal SOC, resulting in a False Negative Rate (FNR) of 0.8% for critical SOC.

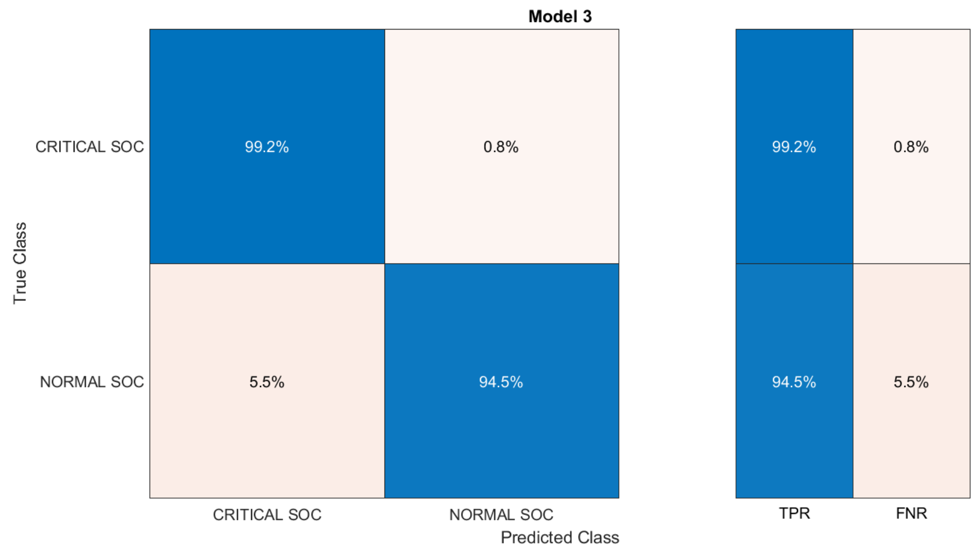

Figure 14 *: TPR and FNR for Bagged Trees*

The Bagged Trees model achieved a TPR of 99.2% for critical SOC samples, indicating that it correctly classified 99.2% of critical SOC samples as critical SOC. This indicates a high sensitivity of the model in detecting critical SOC samples. The Bagged Trees model also achieved a TPR of 94.5% for normal SOC samples, indicating that it correctly classified 94.5% of normal SOC samples as normal SOC. This indicates a slightly lower sensitivity of the model in detecting normal SOC samples compared to critical SOC samples.

The FNR, which represents the percentage of samples from a particular class that were misclassified as the opposite class, was 0.8% for normal SOC samples and 5.5% for critical SOC samples by the Bagged Trees model. This means that the Bagged Trees model had a low rate of false negatives for normal SOC samples but a slightly higher rate for critical SOC samples.

4.6.2 Scatter Plot for Bagged Trees.

A scatter plot was created to visualize the performance of the Bagged Trees model on the test dataset, with the Classified labels (critical SOC or normal SOC) on the x-axis and the true labels on the y-axis.The scatter plot revealed that most of the samples were clustered around the diagonal line, indicating that the Bagged Trees model's classifications were in agreement with the true labels for most of the samples.

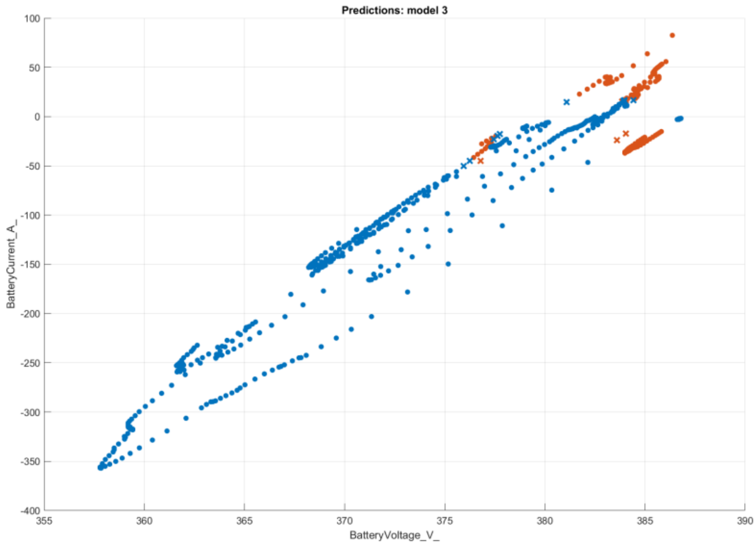

Figure 15 *: Scatter Plot for Bagged Trees*

However, there were some misclassified samples that were located off the diagonal line, indicating instances where the Classified label did not match the true label. Specifically, there were 10 samples from the normal SOC class that were misclassified as critical SOC, and 7 samples from the critical SOC class that were misclassified as normal SOC, as indicated by the confusion matrix.

4.6.3 Scatter Plot for Bagged Trees.

The ROC curve was generated to evaluate the performance of the Bagged Trees model, using the Classified labels and true labels from the test dataset.

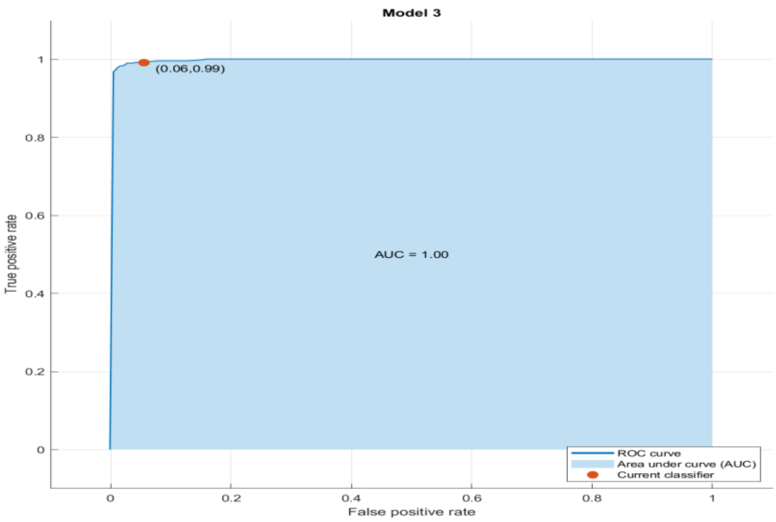

Figure 16 *:ROC curve for Bagged Trees*

Using the true labels and Classified labels from the test dataset, a ROC curve was created to assess how well the Bagged Trees model performed. The ROC curve, with TPR on the y-axis and FPR on the x-axis, showed the trade-off between the TPR and FPR at various categorization levels.

The ROC curve demonstrated excellent performance, with an AUC of 1.0, which indicates perfect discriminatory ability of the model to correctly classify positive and negative samples.

The point on the ROC curve where it meets the coordinates (0.06, 0.99) represents the classification threshold where the Bagged Trees model achieves a TPR of 0.99 (99%) while maintaining a low FPR of 0.06 (6%).This suggests that the Bagged Trees model is capable of accurately classifying both critical SOC and normal SOC samples with high TPR and low FPR, as indicated by the AUC

4.7 Comparison Table for SVM, NN and BT

Models	Accuracy	AUC	Classification Speed	Training Time	Activation Function
Neural Network	99.1%	0.99	32000 obs/sec	8.5477 sec	ReLU
Bagged Trees	98.4%	1	3900 obs/sec	4.3311 sec	Linear
SVM	92.5%	0.76	28000 obs/sec	1.7001 sec	No activation function

CHAPTER VI

Conclusion and Recommendations

5. Conclusion

In this study, Artificial Neural Network (ANN) , Bagged trees and Support Vector Machines (SVM) are evaluated and compared by analysing their performance to a set of data. The model which performed well among the three models is Neural Network model which used three hidden layer and ReLU activation function. Neural Network model has shown its good performance compare to Bagged trees and Support Vector Machines (SVM). These models were used to make the Classification for the electric vehicle battery State of Charge (SOC). The SOC was classified as either 'Normal SOC' or Critical SOC'. The accuracy of the models in training was 99.1%, Bagged trees 98.4% and SVM 92.5%. Whereas the accuracy of the models in testing was 77.5%, Bagged Trees 71.6% and SVM 63.6%. The predictors used in this research were Voltage (V), Temperature (T), Current (A). The predictors have provided best results when they were given as inputs to models.

Furthermore, the results of the models have shown that the Neural Network and the Bagged trees has performed well in classifying the battery health condition as 'Normal SOC' or 'Critical SOC'. These two models have shown the high accuracy with Higher True Positive Rate (TPR) and lower False Negative (FNR). the SVM model accuracy had some limitations in detecting which also resulted in high False Negative Rate (FNR). The observation of the results also shows that AUC values and ROC curves has performed well in classifying the battery health condition as 'Normal SOC' or 'Critical SOC'. However, the trade-off between specify, sensitivity and false negatives can depend on specific requirements and priorities of the model application.

5.1 Recommendations

According to the three models results, it is evident that using machine learning to classify the State of Charge (SOC) of electric vehicles can be very efficient and act as solution to prevent the damages that can occur in the batteries for electric vehicles. The fact that, the classification is done in real time is added advantage as this will keep alerting the driver on any possible fault in the driver. Using Machine learning in Classification for SOC in electric vehicles could increase the safety for both drive, environment and the legacy of the car manufacturer. The use of machine learning in this application could help in the optimization of the battery and extend its lifespan. Though three machines learning algorithms were implemented in this research, I would suggest that the future research could explore more machine learning algorithms to see how also they perform on the battery dataset. I also propose if the predictors can be changed also to observe if there is any change in Classifying the battery health conditions.

CHAPTER VII

References

1. Rob W. Bagged Trees: A Machine Learning Algorithm Every Data Scientist Needs. Towards Data Science. Retrieved April 28, 2023, from https://towardsdatascience.com/bagged-trees-a-machine-learning-algorithm-every-data-scientist-needs-d8417ec2e0d9
2. Kaggle. BATTERY AND HEATING DATA IN REAL DRIVING CYCLES. Retrieved April 28, 2023, from https://www.kaggle.com/datasets/atechnohazard/battery-and-heating-data-in-real-driving-cycles?select=TripA02.csv
3. Chandran, V., Patil, C. K., Karthick, A., Ganeshaperumal, D., Rahim, R., & Ghosh, A. (2021). State of Charge Estimation of Lithium-Ion Battery for Electric Vehicles Using Machine Learning Algorithms. World Electric Vehicle Journal 2021, Vol. 12, Page 38, 12(1), 38. https://doi.org/10.3390/WEVJ12010038
4. Chang, L., & Xiaoluo, J. (2011). Kalman filter based on SVM innovation update for Classifying state-of health of VRLA batteries. Communications in Computer and Information Science, 225 CCIS(PART 2), 455–463. https://doi.org/10.1007/978-3-642-23220-6_58/COVER
5. Driscoll, L., de la Torre, S., & Gomez-Ruiz, J. A. (2022). Feature-based lithium-ion battery state of health estimation with artificial neural networks. Journal of Energy Storage, 50, 104584. https://doi.org/10.1016/J.EST.2022.104584
6. Fan, Y., Xiao, F., Li, C., Yang, G., & Tang, X. (2020). A novel deep learning framework for state of health estimation of lithium-ion battery. Journal of Energy Storage, 32, 101741. https://doi.org/10.1016/J.EST.2020.101741
7. Ismail, M., Dlyma, R., Elrakaybi, A., Ahmed, R., & Habibi, S. (2017). Battery state of charge estimation using an Artificial Neural Network. 2017 IEEE Transportation and Electrification Conference and Expo, ITEC 2017, 342–349. https://doi.org/10.1109/ITEC.2017.7993295
8. Kim, S. W., Oh, K. Y., & Lee, S. (2022). Novel informed deep learning-based prognostics framework for on-board health monitoring of lithium-ion batteries. Applied Energy, 315, 119011. https://doi.org/10.1016/J.APENERGY.2022.119011
9. Li, D., & Yang, L. (2020). Remaining useful life Classification of lithium battery using convolutional neural network with optimized parameters. Proceedings - 2020

5th Asia Conference on Power and Electrical Engineering, ACPEE 2020, 840–844. https://doi.org/10.1109/ACPEE48638.2020.9136289

10. LeewayHertz. Machine Learning Development | Machine Learning Consulting. Retrieved April 28, 2023, from https://www.leewayhertz.com/machine-learning-services/
11. 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. https://doi.org/10.1016/J.JPOWSOUR.2020.229154
12. Rauf, H., Khalid, M., & Arshad, N. (2022). Machine learning in state of health and remaining useful life estimation: Theoretical and technological development in battery degradation modelling. Renewable and Sustainable Energy Reviews, 156, 111903. https://doi.org/10.1016/J.RSER.2021.111903
13. Rimpas, D., Kaminaris, S. D., Aldarraji, I., Piromalis, D., Vokas, G., Papageorgas, P. G., & Tsaramirsis, G. (2022). Energy management and storage systems on electric vehicles: A comprehensive review. Materials Today: Proceedings, 61, 813–819. https://doi.org/10.1016/J.MATPR.2021.08.352
14. Roman, D., Saxena, S., Robu, V., Pecht, M., & Flynn, D. (2021). Machine learning pipeline for battery state-of-health estimation. Nature Machine Intelligence, 3(5), 447–456. https://doi.org/10.1038/S42256-021-00312-3
15. Support Vector Machine (SVM) Technique - Creative Biolabs. (n.d.). Retrieved April 28, 2023, from https://www.creative-biolabs.com/drug-discovery/therapeutics/support-vector-machine-svm-technique.html
16. Toughzaoui, Y., Chaoui, H., Louahlia, H., Petrone, R., & Gualous, H. (2022). State of Health Estimation and Remaining Useful Life Classification Using Hybrid Kmeans CNN-Lstm Network. ECS Meeting Abstracts, MA2022-02(28), 1080–1080. https://doi.org/10.1149/ma2022-02281080mtgabs
17. Types of Neural Network Algorithms in Machine Learning with Examples. (n.d.). Retrieved April 28, 2023, from https://omdena.com/blog/types-of-neural-network-algorithms-in-machine-learning/
18. Zhang, Q., Deng, W., Zhang, S., & Wu, J. (2016). A Rule Based Energy Management System of Experimental Battery/Supercapacitor Hybrid Energy Storage System for Electric Vehicles. Journal of Control Science and Engineering, 2016. https://doi.org/10.1155/2016/6828269

19. Zenati, A., Desprez, P., Razik, H., & Rael, S. (2012). A methodology to assess the state of health of lithium-ion batteries based on the battery's parameters and a fuzzy logic system. *2012 IEEE International Electric Vehicle Conference, IEVC 2012*. https://doi.org/10.1109/IEVC.2012.6183268
20. Tang, X., Liu, K., Wang, X., Gao, F., MacRo, J., & Widanage, W. D. (2020). Model Migration Neural Network for Predicting Battery Aging Trajectories. *IEEE Transactions on Transportation Electrification*, *6*(2), 363–374. https://doi.org/10.1109/TTE.2020.2979547
21. Hannan, M. A., Hoque, M. M., Hussain, A., Yusof, Y., & Ker, P. J. (2018). State-of-the-Art and Energy Management System of Lithium-Ion Batteries in Electric Vehicle Applications: Issues and Recommendations. *IEEE Access*, *6*, 19362–19378. https://doi.org/10.1109/ACCESS.2018.2817655
22. Su, C., & Chen, H. J. (2017). A review on prognostics approaches for remaining useful life of lithium-ion battery. *IOP Conference Series: Earth and Environmental Science*, *93*(1). https://doi.org/10.1088/1755-1315/93/1/012040
23. Li, L., Wang, P., Chao, K. H., Zhou, Y., & Xie, Y. (2016). Remaining useful life prediction for lithium-ion batteries based on Gaussian processes mixture. *PLoS ONE*, *11*(9). https://doi.org/10.1371/JOURNAL.PONE.0163004
24. Park, K., Choi, Y., Choi, W. J., Ryu, H. Y., & Kim, H. (2020). LSTM-Based Battery Remaining Useful Life Prediction with Multi-Channel Charging Profiles. *IEEE Access*, *8*, 20786–20798. https://doi.org/10.1109/ACCESS.2020.2968939
25. Severson, K. A., Attia, P. M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M. H., Aykol, M., Herring, P. K., Fraggedakis, D., Bazant, M. Z., Harris, S. J., Chueh, W. C., & Braatz, R. D. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, *4*(5), 383–391. https://doi.org/10.1038/S41560-019-0356-8
26. Yun, Z., Qin, W., Shi, W., & Ping, P. (2020). State-of-health prediction for lithium-ion batteries based on a novel hybrid approach. *Energies*, *13*(18). https://doi.org/10.3390/EN13184858

27. Jia, J., Liang, J., Shi, Y., Wen, J., Pang, X., & Zeng, J. (2020). SOH and RUL prediction of lithium-ion batteries based on Gaussian process regression with indirect health indicators. *Energies*, *13*(2). https://doi.org/10.3390/EN13020375
28. Liu, D., Luo, Y., Liu, J., Peng, Y., Guo, L., & Pecht, M. (2014). Lithium-ion battery remaining useful life estimation based on fusion nonlinear degradation AR model and RPF algorithm. *Neural Computing and Applications*, *25*(3–4), 557–572. https://doi.org/10.1007/S00521-013-1520-X
29. Chang, C., Wang, Q., Jiang, J., & Wu, T. (2021). Lithium-ion battery state of health estimation using the incremental capacity and wavelet neural networks with genetic algorithm. *Journal of Energy Storage*, *38*, 102570. https://doi.org/10.1016/J.EST.2021.102570
30. Wang, Y., Zhang, C., & Chen, Z. (2016). An adaptive remaining energy prediction approach for lithium-ion batteries in electric vehicles. *Journal of Power Sources*, *305*, 80–88. https://doi.org/10.1016/J.JPOWSOUR.2015.11.087
31. Wu, J., Zhang, C., & Chen, Z. (2016). An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks. *Applied Energy*, *173*, 134–140. https://doi.org/10.1016/J.APENERGY.2016.04.057
32. Dai, H., Zhao, G., Lin, M., Wu, J., & Zheng, G. (2019). A novel estimation method for the state of health of lithium-ion battery using prior knowledge-based neural network and markov chain. *IEEE Transactions on Industrial Electronics*, *66*(10), 7706–7716. https://doi.org/10.1109/TIE.2018.2880703
33. Zhang, S., Zhai, B., Guo, X., Wang, K., Peng, N., & Zhang, X. (2019). Synchronous estimation of state of health and remaining useful lifetime for lithium-ion battery using the incremental capacity and artificial neural networks. *Journal of Energy Storage*, *26*, 100951. https://doi.org/10.1016/J.EST.2019.100951

Appendices

Appendix A

1. Neural Network Model Parameters

Model 1: Trained

Training Results

Accuracy (Validation): 99.1%

Total cost (Validation): Not applicable

Classification speed: ~32000 obs/sec

Training time: 8.5477 sec

Model Type

Preset: Medium Neural Network

Number of fully connected layers: 3

First layer size: 25

Second layer size: 20

Third layer size: 20

Activation: ReLU

Iteration limit: 1000

Regularization strength (Lambda): 0

Standardize data: Yes

Optimizer Options

Hyperparameter options disabled

Feature Selection

All features used in the model, before PCA

PCA

PCA disabled

Misclassification Costs

Not supported

2. Support Vector Machine Parameters

Model 2: Trained

Training Results

Accuracy (Validation): 92.5%

Total cost (Validation): 79

Classification speed: ~28000 obs/sec

Training time: 1.7001 sec

Model Type

Preset: Linear SVM

Kernel function: Linear

Kernel scale: Automatic

Box constraint level: 1

Multiclass method: One-vs-One

Standardize data: true

Optimizer Options

Hyperparameter options disabled

Feature Selection

Cost matrix: default

3. Bagged Trees Model Parameters

Model 3: Trained

Training Results

Accuracy (Validation): 98.4%

Total cost (Validation): 17

Classification speed: ~3900 obs/sec

Training time: 4.3311 sec

Model Type

Preset: Bagged Trees

Ensemble method: Bag

Learner type: Decision tree

Maximum number of splits: 1049

Number of learners: 30

Optimizer Options

Hyperparameter options disabled

Feature Selection

NEAR EAST UNIVERSITY

SCIENTIFIC RESEARCH ETHICS COMMITTEE

11.05.2023

Dear Kevin Cyubahiro

Your project **"Real-Time Battery Health Monitoring for Electric Vehicle usingMachine Learning"** has been evaluated. Since only secondary data will be used the project does not need to go through the ethics committee. You can start your research on the condition that you will use only secondary data.

Prof. Dr. Aşkın KİRAZ

The Coordinator of the Scientific Research Ethics Committee

Appendix X
Turnitin Similarity Report

kevin thesis

ORIGINALITY REPORT

8%	7%	4%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	1%
2	Sanbao Zhang, Shanshan Wang, Ruibin Xue, Jian Zhu, Aimon Tanvir, Danran Li, Bin Zhou. " Impact Assessment of COVID - 19 Lockdown on Vertical Distributions of NO and HCHO From MAX - DOAS Observations and Machine Learning Models ", Journal of Geophysical Research: Atmospheres, 2022 Publication	<1%
3	www.mdpi.com Internet Source	<1%
4	medium.com Internet Source	<1%
5	www.ijraset.com Internet Source	<1%
6	link.springer.com Internet Source	<1%
7	mdpi-res.com Internet Source	<1%

8	Submitted to Heriot-Watt University Student Paper	<1%
9	Submitted to National University of Ireland, Galway Student Paper	<1%
10	Submitted to Edith Cowan University Student Paper	<1%
11	Submitted to Universidad Internacional de la Rioja Student Paper	<1%
12	dehesa.unex.es Internet Source	<1%
13	www.science.gov Internet Source	<1%
14	Amir Muhammed Sa\'ad. "NEMO: a neural motion estimator for mooring line failure detection of offshore platforms.", Universidade de Sao Paulo, Agencia USP de Gestao da Informacao Academica (AGUIA), 2023 Publication	<1%
15	www.medrxiv.org Internet Source	<1%
16	www.wseas.us Internet Source	<1%

17	keytodatascience.com Internet Source	<1 %
18	mhealth.jmir.org Internet Source	<1 %
19	towardsdatascience.com Internet Source	<1 %
20	worldwidescience.org Internet Source	<1 %
21	site.ieee.org Internet Source	<1 %
22	www.frontiersin.org Internet Source	<1 %
23	dc.library.okstate.edu Internet Source	<1 %
24	A. Varnek, D. Fourches, V. Solov'ev, O. Klimchuk, A. Ouadi, I. Billard. "Successful “In Silico” Design of New Efficient Uranyl Binders", Solvent Extraction and Ion Exchange, 2007 Publication	<1 %
25	hdl.handle.net Internet Source	<1 %
26	pdfcoffee.com Internet Source	<1 %
27	www.ncbi.nlm.nih.gov Internet Source	<1 %

28	Lecture Notes in Computer Science, 2014. Publication	<1 %
29	mafiadoc.com Internet Source	<1 %
30	patents.google.com Internet Source	<1 %
31	ruor.uottawa.ca Internet Source	<1 %
32	www.essays.se Internet Source	<1 %
33	www.slideshare.net Internet Source	<1 %
34	Muhammad Attique khan, Tallha Akram, Muhammad Sharif, Tanzila Saba. "Fruits diseases classification: exploiting a hierarchical framework for deep features fusion and selection", Multimedia Tools and Applications, 2020 Publication	<1 %
35	etd.aau.edu.et Internet Source	<1 %
36	idr.l1.nitk.ac.in Internet Source	<1 %
37	ijettjournal.org Internet Source	<1 %

38	Somayeh Bakhtiari Ramezani, Logan Cummins, Brad Killen, Richard Carley et al. "Scalability, Explainability and Performance of Data-Driven Algorithms in Predicting the Remaining Useful Life: A Comprehensive Review", Institute of Electrical and Electronics Engineers (IEEE), 2023 Publication	<1 %
39	bassantgz30.medium.com Internet Source	<1 %
40	escholarship.org Internet Source	<1 %
41	theses.hal.science Internet Source	<1 %
42	www.sciencegate.app Internet Source	<1 %
43	Sam C. Bose. "GPS Spoofing Detection by Neural Network Machine Learning", IEEE Aerospace and Electronic Systems Magazine, 2022 Publication	<1 %
44	Sepp Hochreiter, Jürgen Schmidhuber. "Long Short-Term Memory", Neural Computation, 1997 Publication	<1 %

45	M.S. Hossain Lipu, Shaheer Ansari, Md Sazal Miah, Kamrul Hasan et al. "A review of controllers and optimizations based scheduling operation for battery energy storage system towards decarbonization in microgrid: Challenges and future directions", Journal of Cleaner Production, 2022 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		