Design and Development of Advanced Machine Learning Algorithms for Lithium-ion Battery State-of-Charge Estimation

by

Manjot Singh Sidhu

A thesis submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

Master of Applied Science in Electrical and Computer Engineering

Faculty of Engineering and Applied Science

Department of Electrical, Computer and Software Engineering

Ontario Tech University

Oshawa, Ontario, Canada

November 2019

THESIS EXAMINATION INFORMATION

Submitted by: **Manjot Sidhu**

Master of Applied Science in **Electrical and Computer Engineering**

Thesis title: Design and Development of Advanced Machine Learning Algorithms for Lithium-ion Battery State-of-Charge Estimation.

An oral defense of this thesis took place on November 19, 2019 in front of the following examining committee:

Examining Committee:

Chair of Examining Committee	Dr. Ying Wang
Research Supervisor	Dr. Sheldon Williamson
Examining Committee Member	Dr. Min Dong
Thesis Examiner	Dr. Masoud Makrehchi

The above committee determined that the thesis is acceptable in form and content and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate during an oral examination. A signed copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

Abstract

Design and Development of Advanced Machine Learning Algorithms for Lithium-ion Battery State-of-Charge Estimation

by MANJOT SIDHU

Batteries have been becoming more and more popular because of their long life and lightweight. Accurate estimation of the SOC help in making plans in an application to conserve and further enhance battery life. State of Charge (SOC) estimation is a difficult task made more challenging by changes in battery characteristics over time and their nonlinear behavior. In recent years, intelligent schemes for the estimation of the SOC have been proposed because of the absence of the formula for calculating SOC which is hard to deduce because of the effect of external factors like temperature. As the traditional methods only considered certain aspects which with the aging and degradation of the battery results in errors. To tackle this problem several methods were proposed which made use of now evolving artificial intelligence technologies. This paper presents a new SOC estimation algorithm based on kNearest neighbor and random forest regression and a comparison study is done using four algorithms Support Vector Regression, Neural Network Regression, Random Forest Regression and kNearest Neighbor. Their performance is evaluated using data from two drive cycles.

Acknowledgements

I would like to express my gratitude and thankfulness to my supervisor, Dr. Sheldon S. Williamson for his guidance and support throughout the course of this work. His encouragement and thoughtful insights on variety of topics made this thesis possible.

To all my colleagues in the STEER group at Ontario Tech University for their invaluable suggestions and knowledge for this work.

To my family for all their love and support all throughout my life. I dedicate this thesis to my family for their never ending love and support and for being my pillar of strength.

Dedicated to My Family.

Statement of Contributions

Part of the work described in Chapter 3 has been published as:

1. M. S. Sidhu, D. Ronanki, and S. Williamson, “Hybrid state of charge estimation approach for lithium-ion batteries using k-nearest neighbour and gaussian filter-based error cancellation,” in *2019 IEEE 28th International Symposium on Industrial Electronics (ISIE)*, IEEE, 2019, pp. 1506–1511.

2. M. S. Sidhu, D. Ronanki, and S. Williamson, "State of Charge Estimation of Lithium-ion Batteries using Hybrid Machine Learning Technique" in 45^{th} *Annual Conference of the IEEE Industrial Electronics Society, 2019*. (Accepted)

I performed the model creations, validations and performance evaluations. I also did majority of writing of the manuscripts.

Contents

List of Figures

List of Tables

List of Abbreviations

SOC	**S**tate **o**f **C**harge
ML	**M**achine **L**earning
SVR	**S**upport **V**ector **R**egression
SVM	**S**upport **V**ector **M**achine
k**NN**	k**N**earest **N**eighbor
NN	**N**eural **N**etwork
BMS	**B**attery **M**anagement **S**ystem
EV	**E**lectric **V**ehicle
LNMC	**L**ithium **N**ickel **M**anganese **C**obalt oxide
SOH	**S**tate **O**f **H**ealth
OCV	**O**pen **C**ircuit **V**oltage
KF	**K**alman **F**ilter
EKF	**E**xtended **K**alman **F**ilter
UKF	**U**nscented **K**alman **F**ilter
SPKF	**S**igma **P**oint **K**alman **F**ilter
CDKF	**C**entral **D**ifference **K**alman **F**ilter
AEKF	**A**daptive **E**xtended **K**alman **F**ilter
ALS	**A**uto-covariance **L**east **S**quare
GA	**G**enetic **A**lgorithm
ANFIS	**A**daptive **N**euro **F**uzzy **I**nference **S**ystem
LCA	**L**inear **C**orrelation **A**nalysis
NCA	**N**onparametric **C**orrelation**A**nalysis
PSO	**P**article **S**warm **O**ptimization
RBF	**R**adial **B**asis **F**unction
BPNN	**B**ack **P**ropagation **N**eural **N**etwork
BSA	**B**acktracking **S**earch **A**lgorithm

USABC	**U**nited **S**tates of **A**merica **B**attery **C**onsortium
UDDS	**U**rban **D**ynamometer **D**riving **S**chedule
DST	**D**ynamic **S**tress **T**est
FUDS	**F**ederal **U**rban **D**riving **S**chedule
US06	**U**nited **S**tates highway schedule

Chapter 1

Introduction

1.1 Introduction

As the world moves away from petroleum-based fuels due to their impact on the environment, alternative forms of energy are becoming more mainstream. This paradigm shift towards energy security, lower emission, and higher fuel efficiency led to the electrification of the transport sector through a wide spread of electric vehicle (EV) technology. These systems utilize a battery as the main energy source, which plays a significant role in the performance and lifetime of the EVs. To meet the required performance, longer-lasting battery life and their performance have become an active area of research [1].

Electric vehicles are gaining popularity due to environmental hazards caused by internal combustion engines. While the conventional energy sources are efficient in energy conversion but the amount of carbon emitted to the environment is very high [2]. As the world moves towards clean energy, EV's have taken the center stage because of their zero carbon emissions which have put the focus on efficient energy storage systems. Their growth is mainly dependent on battery management

systems which help increase the driving range of the vehicle as well as their reliability and safety [3]. Their popularity is mainly hindered by range anxiety about how long the battery will last, time is taken to charge the battery which has put the focus on rechargeable batteries as the clean source for electric energy storage.

1.2 Battery Management Systems (BMS)

To manage battery packs in EVs, BMS is employed. BMS helps protect battery by monitoring and safeguarding battery. Main duties of BMS inlude:

- Help battery fulfill requirements of a vehicle for its safe operation.
- Enhance the life of battery packs of vehicles.
- Monitor the cell temperature, series cell voltages, pack current and estimate SOC.
- Ensure the working of battery in a specific range of voltage and temperature to stop it from being damaged and establish a battery's safe operation.

1.3 Lithium-ion Batteries

Lithium-ion batteries are rechargeable batteries and are commonly used in consumer electronics like phones, watches, EV's, etc. In a Lithium-ion battery, during the discharge cycle, Lithium-ions move from anode to cathode and in opposite direction during the charge cycle. Lithium-ion batteries come in different chemistries such as Lithium iron phosphate ($LiFePO_4$), Lithium manganese oxide (LMO) and

Lithium nickel manganese cobalt oxide ($LNMC$). LNMC batteries are the most common chemistry used in EV's.

Their main advantage of Li-ion batteries over the other chemistries is their energy density, which can go over 160kWh/kg making it possible for the long range drives needed in the EV application. Furthermore, its calendar life is impressive of more than 1000 full cycles, where it is possible to implement them in a fashion that makes the EV battery pack last at least 8 to 10 years. In addition, it is possible to have a good operating temperature range, parameter that is very crucial for the battery performance and safety.

On the other hand, the disadvantages are also present in the usage of LNMC over the other chemistries. For example, the external electronics circuits are needed for the composition of the battery pack with series connections, in order to balance their voltage and for its safe and reliable operation. Also, they are more expensive due to their high quality and high energy density.

1.4 State of Charge (SOC)

One of the main functions of BMS is SOC estimation. Accurate SOC estimation helps in protecting the battery and increasing its life by helping prevent over-discharge and helps in implementing energy saving strategies for the battery. SOC is the ratio of remaining capacity of the battery to the full charge capacity of the battery.

SOC of the battery is given by:

$$SOC = \frac{q_{remaining}}{q_{rated}} \times 100\% \tag{1.1}$$

where,

$q_{remaining}$ represents remaining capacity of the battery.

q_{rated} represents rated capacity of the battery.

1.5 SOC Estimation Methods

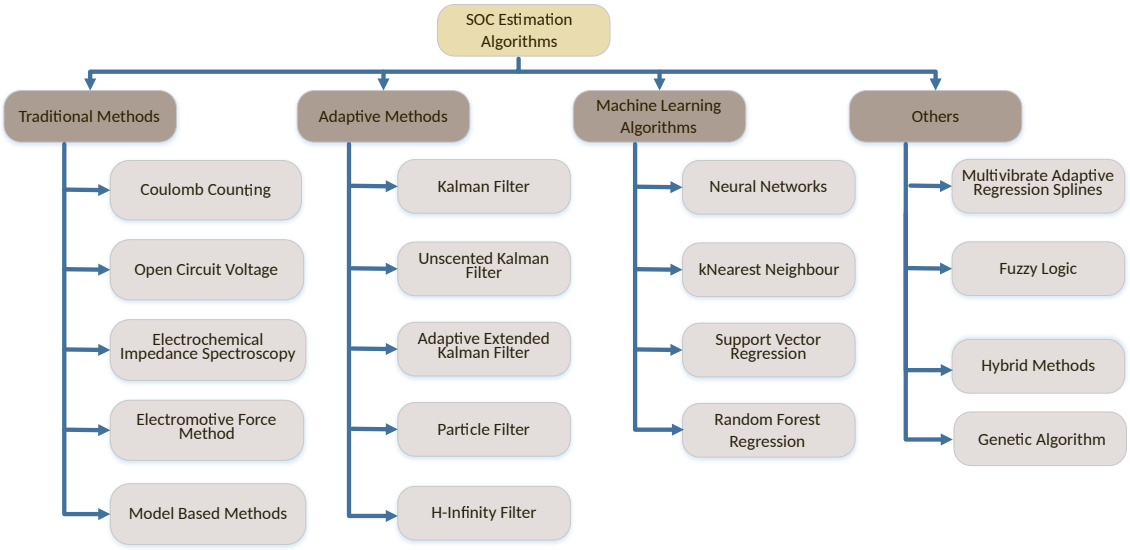


FIGURE 1.1: Classification of popular SOC estimation algorithms.

The estimation of accurate SOC of a battery is indispensable to energy management strategies developed for efficient use of energy left. Researchers have proposed numerous SOC estimation algorithms over the years. Figure 1.1 shows existing SOC estimation techniques proposed in literature. These techniques are classified into four main categories. Traditional methods include SOC estimation

techniques based on voltage and current like Coulomb counting, open circuit voltage method, etc. Adaptive methods include SOC estimation techniques based statistical methods like Kalman filter and its extensions. ML-based methods include techniques that learn from data. And rest of estimation techniques are grouped into last category.

1.5.1 Coulomb Counting Approach

Coulomb counting is one of the popular techniques employed for the estimation of SOC, due to its low computational requirement and ease of implementation. This method calculates the charge being transferred to and from the battery, mathematically expressed as:

$$SOC(t) = SOC(t_0) - \frac{1}{C_N} \int_{t_0}^{t} \eta . I(t) . dt \tag{1.2}$$

where,

$SOC(t_0)$	is SOC at initial time,
η	is coulombic efficiency,
$I(t)$	represents current which is positive at discharge and negative at charge and
C_n	is the rated capacity of the battery.

In [4], a technique to estimate SOC and State of Health (SOH) for lithium-ion batteries focusing on correction of operating efficiency and evaluation of SOH. This Coulomb counting approach was improved by the use of the least squares method [5].

Drawbacks:

- Coulomb counter being an open loop SOC estimator accumulates the errors in the current detector which means the longer the estimator is operated larger the cumulative error is. The estimator produces faster and incorrect results for worsening error in the current detector [4], [6].

- It does not take into account the age of the battery. Although the aging characteristics of the battery can be saved in the software but that is not an effective solution. The estimations could go incorrect as the battery ages if the battery does not follow the expected course for aging [7].

- Coulomb counting estimation algorithm has an average error of $\pm 15\%$. Also, the coulomb counter estimates the starting SOC with the help of battery pack voltage. There is no way to correct or detect the error in the starting SOC [7].

- Because the estimation is based on the readings of the current sensor, measurement drift can have some impact [5].

Also, this does not take into account temperature, voltage which influences the capacity of the battery [8] which leads to restricting its usage on a larger scale [3].

1.5.2 Open Circuit Voltage based Estimation Method

Open circuit voltage (OCV) based method is simple and has very high precision. Figure 1.2 shows relationship between OCV and SOC. As the SOC increases, OCV also increases. However, this relationship between OCV and SOC does not hold for all kinds of batteries [9]. This method has been modified [10] and also combined with other methods [11, 12, 13].

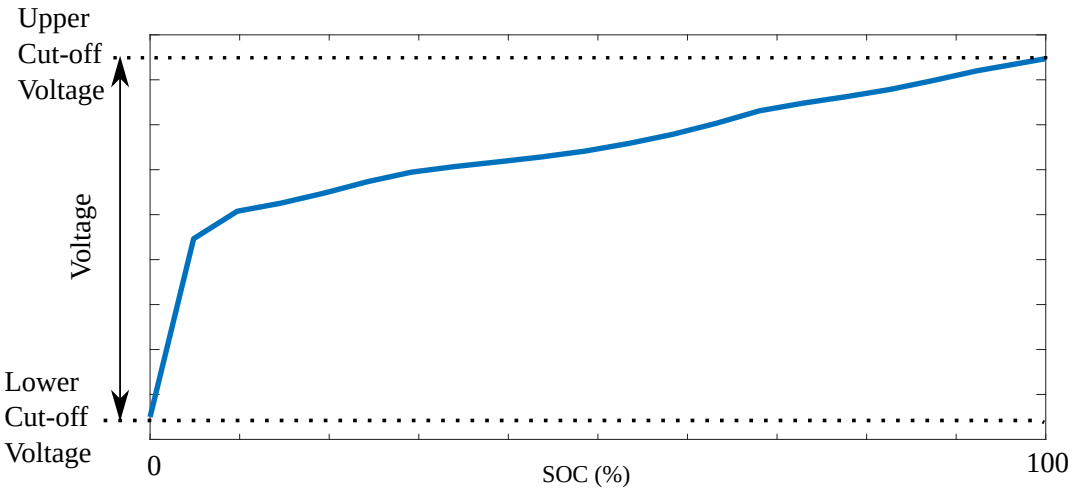


FIGURE 1.2: Relationship of OCV and SOC of a Li-ion battery.

Drawbacks:

- This method is not very popular because the relation between SOC and OCV only holds if the battery has been disconnected from the load and in state of rest for a couple of hours which makes its use for batteries in real-time very difficult [14], [15].
- The SOC at the same OCV can have two different values because of the amount of current being received as the upper limit voltage is received faster if more current is supplied to the battery [16].

1.5.3 Kalman Filter

Kalman filter is an intelligent tool that uses measurements taken over time to produce the output of the unknown variable. The equations for Kalman Filter are defined in [17]. The Kalman filter algorithm involves five major steps:

- Prediction of state space variables

- Prediction of covariance matrix
- Update of the Kalman gain
- Estimation of the state variable and correction of the prediction and
- Calculation of the estimation error

The uncertainty of the initial SOC is reduced by the Kalman Filter due to the equations being repeatedly evaluated during the operation of the system. have studied the covariance noise parameters of Kalman Filters on lithium-ion battery of two different geometric shapes (cylindrical and prismatic) have been studied by [18]. After analyzing the parameters calculated by the filter it showed the dependence of these values on the type of battery. In [19], researchers use KF to estimate SOC which was tested using Matlab-Simulink and dSPACE. In [20], researchers propose a model combining coulomb counting and OCV with KF which showed improved accuracy.

While Kalman filter can only be used for linear systems but for a nonlinear system, linear time-varying system can be used to approximate the nonlinear system known as extended Kalman filter (EKF) [16]. In EKF, taylor series expansion used for linearization increases the error. Unscented Kalman Filter (UKF) proposed in [21] gives a lesser error when compared with EKF. Two of the most commonly used Sigma Point Kalman filters (SPKF) are UKF and central difference Kalman filter (CDKF) [22]. The accuracy of UKF was further improved by proposing adaptively updating the process and measurement noise covariance by Adaptive Extended Kalman Filter (AEKF) [23]. EKF, AEKF, and UKF were compared by [23] to conclude the accuracy of AUKF was better. A new technique based on UKF and

Autocovariance least square (ALS) is proposed by [24] to estimate SOC which has high relative convergence speed and immunity to wrong initialization.

1.5.4 Genetic Algorithm

Genetic Algorithm (GA) is used to find the optimal model parameters of the non-linear complex system. In SOC estimation applications, parameters of the battery are to be optimized to give us SOC as the result. The Genetic algorithm follows the iterative process of selection, crossover, and mutation to find the optimal solution [16].

In [25], a method to calculate the capacity of a $LiFePO_4$ battery pack using voltage capacity rate curve which is tested by 4 cells connected in series and it reports an error of less than 1%. In [26], GA is used for the fitting process in the Neural network which has one hidden layer and gives SOC as output. Several modifications of the Genetic Algorithm can be found in the literature [27], [28].

1.5.5 Fuzzy Logic

Fuzzy logic is another algorithm that is used to simplify noisy and vague input data [29]. The implementation of the Fuzzy logic is divided into 4 parts which include fuzzification in which measured values are converted into fuzzy sets and classified into membership functions, fuzzy rule base, inference engine which transforms fuzzy rules into linguistic outputs and defuzzification to translate the linguistic outputs into analog outputs [29]. A fuzzy rule-based system to compute the SOC through recursive filtering is proposed in [30]. In [31], the noise of the samples is

reduced by least square SVM and by applying fuzzy-based interference and correlation measurement. They compared this method with ANN and SVR to show a better accuracy of the Fuzzy-least square Support Vector Machine (Fuzzy-LSSVM). In [32], researchers make use of an adaptive neuro fuzzy inference system (ANFIS) which combines the fuzzy system and adaptive systems. It first uses the circuit model to calculate OCV which is further used as input to get the SOC as output for a NiMH battery.

1.5.6 Machine Learning-based Methods

Machine Learning (ML) based SOC estimation methods make use of SOC estimation techniques like support vector regression (SVR) and neural networks (NN). These methods estimate SOC by mapping inputs and outputs using ML methods. To create a model to estimate SOC, first, it needs to be trained using data from a battery. It can take any number of inputs like current, voltage, temperature, etc. Data used to train the model (training data) needs to be accurate to reduce unintended errors in ready to be deployed model. These methods require high computation power and higher processing tools. ML-based SOC estimation algorithms are reviewed in chapter 2.

1.6 Summary

In this chapter, Battery management systems are briefly discussed. SOC estimation is discussed with a brief introduction to commonly used traditional methods like coulomb counting, open circuit voltage, etc. In this thesis, a novel SOC estimation

method, based on kNearest Neighbor (kNN), is proposed. Another SOC estimation method based on random forest regression (RFR) is proposed. The proposed method is compared with the existing ML-based estimation technique to bring out the superiority of the method. To compare estimation algorithms, a novel generalized estimation algorithm is proposed. Four SOC estimation algorithms based on SVR, NN, RFR, and kNN are compared based on their accuracy using the same data sets.

1.7 Motivation and Scope of Thesis

Traditional SOC estimation methods ignore effect of ambient temperature and aging of battery which can result in unintended increase in errors during estimation process. This thesis presents SOC estimation techniques based on ML methods which are kNearest Neighbor (kNN) and random forest regression (RFR). The superiority of these techniques is established by comparing them with other methods present in literature like Neural network (NN) and support vector regression (SVR). The experimental results show that kNN outperforms all other SOC estimation techniques. The studies are performed at three different temperatures which are cold temperature (0°C), room temperature (25°C) and hot temperature (45°C). This shows working of proposed methods in all temperature conditions. Their suitability for EVs is shown by testing them on two different drive cycles namely federal urban driving schedule (FUDS) and united states highway schedule (US06).

1.8 Thesis Outline

In this thesis, SOC estimation techniques based on ML methods are proposed and compared. SOC estimation methods based on kNN and RFR are proposed and compared with SOC estimation techniques based on NN and SVR.

- Chapter 2 presents a literature review of ML-based SOC estimation techniques.
- Chapter 3 shows implementation and results from kNN and RFR based SOC estimation algorithms.
- Chapter 4 compares techniques presented in Chapter 3 with existing SOC estimation techniques based on SVR and NN.
- Chapter 5 summarizes the thesis and discusses future work.

Chapter 2

Literature Review

2.1 Introduction

SOC estimation, the most important task in battery management systems, plays a vital role to prevent the battery from over charging and to know the remaining capacity of battery. SOC is one of the important parameters of the energy management system which shows the ratio of remaining capacity of the battery and the nominal capacity [33]. While measuring SOC directly is not possible but it can be estimated by measuring the voltage, current, temperature [34]. Precise estimation of SOC is very important as it helps in avoiding batteries from overcharge which can damage the battery [35]. Machine learning is very helpful in establishing relationship between any kind of input and output. Support vector Regression which is motivated by the results from the statistical learning theory, maps the input data into hyperplane of high dimensionality to estimate output while neural networks build a neural architecture consisting of hidden layers and neurons.

Most of the SOC estimation algorithms focus on getting the accurate result but they do not take into account thermal influence and impact of degradation [36]

which may result in change in the battery behaviour. Many SOC estimation algorithms have been proposed over the years which are reviewed in [37], [38].

2.2 Machine Learning

Machine learning is the ability of a program to learn and improve on specific task with the help of data without actually being programmed to do the task. There are two steps before a ML model can be implemented in real life. For a ML model to learn from data, it needs a dataset of previously identified solutions, called training dataset. Training dataset is used to map inputs and solutions of the problems. This helps create a model which is then used to predict new values from diverse and never seen before set of inputs. After learning the task provided, the same model can then be used to perform the task with the data it has not encountered before. Machine learning techniques are divided into two major categories:

- Supervised learning
- Unsupervised learning

Algorithms in both categories are suitable for different kind of problems. For SOC estimation problem, we will be using supervised learning algorithm called regression. In the coming sections, we will discuss both of these categories briefly.

2.2.1 Supervised learning

In supervised learning, algorithms are trained on labeled data. In this learning, input data and corresponding output are clearly labeled. Supervised learning is

less complex than unsupervised learning. Algorithms used for supervised learning problems are classification, regression etc. Advantages of supervised learning include learning from previous experience and various real world problems can be solved by using supervised learning algorithms. Disadvantages of supervised learning algorithms are training data constraints like selecting lot of good examples as training data and training process is time consuming.

2.2.2 Unsupervised learning

When data we have is not labeled and unstructured, we make use use of unsupervised learning techniques like clustering. These algorithms are used to find patterns in unknown data. These methods are not as accurate as unsupervised learning algorithms in which we labeled data.

2.3 SOC Estimation using ML

Machine learning methods include models based on Neural Network and Support Vector Machines. The main steps included in these methods, as shown in Figure 2.1, are training and testing before the model can be deployed for real time applications. The training process of Neural Networks and Support Vector Regression is discussed briefly in the subsequent sections. In these methods, there are many algorithms used during the training process. Machine learning methods are compared in table 2.1.

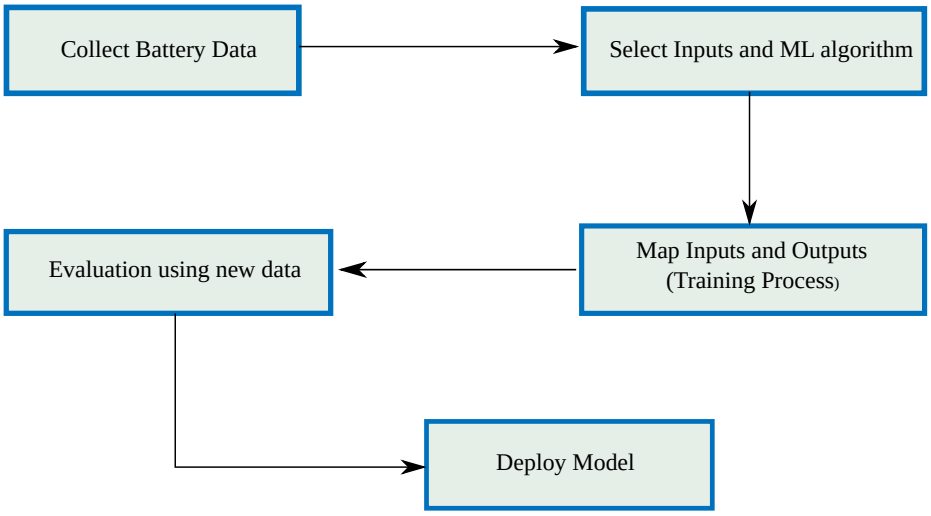


FIGURE 2.1: Working of a Machine Learning Algorithm for SOC Estimation

2.3.1 Neural Network Based Methods

Artificial Neural network is a powerful and intelligent algorithm to map nonlinear inputs to a target output. A neural network consists of three main parts: input layer, output layer and hidden layer(s). Inputs go from input layer to output layer from which outputs are produced, which is then compared to the target output which may contain an error. To reduce the error, weights and biases are added in the hidden layer. During initialization, weights and biases are set to random variable with a range and we can also define maximum number of iterations. During each iteration, outputs of neurons are calculated in hidden layer and in output layer. Information about the errors in output layer and hidden layer is calculated using which the weights and bias in the hidden layer are updated to reflect the errors. This process is repeated until we reach the maximum number of iterations or the value of error becomes negligible.

The relationship between current, voltage and previous SOC to predict SOC of

NI-MH battery pack is explored in [39]. They show that neural network can overcome initial error in SOC. They verified this by setting different initial value of the SOC in the 5 neural networks that they tested. A artificial neural network (ANN) model proposed by [40] for computing the capacity available in lead acid battery. To verify the accuracy of the model, other methods [40] including multilevel Peukert equation [41] is used to find the Peukert constants for peukert equation which describes the relationship between discharge current and available capacity. In [42], discharge current, terminal voltage and temperature are taken as inputs to estimate the SOC of the $LiFePO_4$ battery. The resulting error from the neural network was reduced using the unscented Kalman Filter. Model validation was done using the data collected from battery using both Federal driving schedule and dynamic stress testing. In [43], neural network- Extended Kalman Filter based SOC estimation model for both Lithium-ion and Ni-MH batteries is proposed. The results of simple neural network and NN-EKF are compared in [43] and the error decreased in NN-EKF by 2%.

In [44], a method to estimate the State of Actual Capacity (SOAC) in Electric Vehicles while also taking into account regenerative capacity using the neural network is proposed. SOAC denotes battery residual available capacity of lead acid batteries. NN is verified by comparing the estimated SOAC from the proposed model and calculated SOAC from the data collected. In [45], a method to calculate Battery available Capacity (BAC) for EV's by finding the relationship between discharging current, temperature of the battery surface and BAC is proposed. The resultant mapping of neural network has a correlation coefficient of 0.9986 and a low Average relative percentage error of less than 2%. In [46], current and voltage

are used as inputs to predict the SOC of Lithium ion battery with high accuracy. The model was tested with the data from the UDDS drive cycle and mixed drive cycle in which SOC was calculated by using coulomb counting to train the proposed model.

A new model making use of Radial Basis Function Neural Network (RBFNN) which takes into account the battery degradation and cycle life model is proposed by [47]. RBF was first used in neural network by [48]. The model was evaluated by [47], using data from varying age levels and different temperatures of $10°$C, $25°$C and $40°$C. Researchers compared the results from the RBFNN with conventional NN and found the performance of the RBFNN improved by 54%. In [49], RBFNN is used to estimate SOC for the Ni-MH battery. The model was trained using the data from the battery which was charged at different current rate from 0.5 C to 10 C. It was shown that the absolute error was contained within 5% when the model was tested on data from different current profiles. In [50], battery SOC is estimated by NN while the inputs for model are selected by correlation analysis. Linear correlation analysis (LCA), Nonparametric Correlation analysis (NCA) and Partial correlation analysis are used for input variable selection. The inputs were selected from 7 input variables namely discharging current (i), terminal voltage (v), Ampere hours used or battery capacity released (Ah), Time average voltage (tav), time average tav (ttav), time derivative of voltage (dvt) and second time derivative of voltage (ddvt). On comparing all the methods, [50] found that absolute errors for all the three methods are less than 5%. Some other neural network methods are presented in [51], researchers have a SOC estimation algorithm using NN which investigates pulse current loads. Researchers have also shown that the number of

neurons in the hidden layer of the NN accuracy is increased until number reaches 6 and after that degradation process starts. In [52], a three layer feed forward NN to estimate SOC which is trained using modified particle swarm optimization (PSO) is proposed. The model was trained and tested using the FUDS, hill climb profile and random drive tests. it showed that the error was less than 5% for most data points while error increased at beginning and end of SOC trajectory because of the noise present in data. In [53], researchers have proposed to design a radial basis function neural network (RBFNN) based non linear observer to calculate SOC. The estimation error was arbitrarily small which was proved using Lyapunov stability analysis. Results from this observer were compared with the EKF using FUDS cycles was compared which concluded the mean error of the proposed observer to be around 0.35% while the mean error of the EKF was found to be at 1.16%.

In [54], a back propagation neural network (BPNN) is proposed which is improved by backtracking search algorithm (BSA) that increases the accuracy, which is achieved by figuring out the number of hidden layer neurons and learning rate for the BPNN. Researchers tested the proposed model with data collected at three different temperatures by using dynamic stress testing and FUDS drive profiles. The results from BSA based BPNN are compared with other commonly used algorithms like RBFNN, Generalized Regression neural network (GRNN) and Extreme Machine Learning (ELM) algorithms. It was concluded that BPNN-BSA for FUDS cycle at 25°C reduced the error by 59%, 58% and 62% respectively when compared with RBFNN-BSA, GRNN-BSA and ELM-BSA respectively. In [55], Cascade Correlation Neural Network (CCNN) [56] which was developed in 1990 to determine the type of battery and the SOC of the battery is used. It uses current, voltage, power

and angles for current and voltage drop as inputs to determine the type of battery with 96.02% maximum average success rate but it only uses current,voltage,power and time to calculate the SOC of the battery with maximum average success rate of 99.03%.

In [57], Nonlinear Autoregressive with Exogenous input based neural network (NARXNN) algorithm to calculate SOC is presented which uses lightning search algorithm (LSA) that is used to calculate hidden layer neurons. The proposed algorithm is validated by comparing the NARXNN-LSA with the NARXNN based on particle swarm optimization (PSO) with data collected at three different temperatures. They also compare the NARXNN-LSA with BPN-LSA and RBFNN-LSA but NARXNN-LSA outperforms other two with RMS error of 0.89% while RMS for other two was more than 2 at $0°$C.

2.3.2 Support Vector based Methods

Support Vector Regression (SVR) [61] is a supervised learning algorithm which is used for classification and regression [62]. It has also been used for anomaly detection [63], fault diagnosis [64], clustering [65] and in medical field [66], [67]. SVM uses kernel function to map input data x into a higher dimensional feature space. In ϵ-Support vector regression proposed by [62], the goal is to find a $F(x)$ which stands for all the training data and at the same time is as flat as possible which means for each training data (x_i, y_i), the actually obtained target should contain error less than ϵ. Deviation larger than ϵ is not accepted. [68] describes the equations as follows: Here b is bias term while w signifies flatness. Also In (x_i, y_i), x_i is the set of features used to find the target y_i. The assumption in equation 4.1 is such

TABLE 2.1: Comparison of ML-based methods used in SOC Estimation

Reference	Inputs	Method used	Output	Error
[7]	I, V, SOC_{t-1}[1], ΔV [2]	SVR	SOC	5% (RMS)[3]
[39]	I, V,SOC_{t-1}	ANN	SOC	4.2%(Avg. error)
[42]	I, V, T	NN(with UKF)	SOC	2.5%
[43]	I, V, ΔI, ΔV, SOC_{EKF}	NN-EKF	SOC	<1%(RMS)
[44]	X_1[4], X_2,X_3,X_4,X_5,X_6[5],X_7[6]	NN	SOAC[7]	1.27%(ARPE)[8]
[45]	I, T	NN	BAC[9]	5.29%(MRE)[10]
[46]	I_t, V_t, I_{t-1},V_{t-1}	ANN	SOC	4.91e-08(MSE)[11]
[47]	I, V, C_n[12]	RBFNN	SOC	2.4%(MAE)[13]
[49]	I, V, Temp.	RBFNN	SOC	5%(AE)[14]
[52]	I, Ah-used, Tp[15],V-min[16]	NN	SOC	5%
[53]	$V_B(t)$, $V_{OC}(t)$[17],$V_i(t)$[18], t, $I_B(t)$ [19]	RBFNN	SOC	2.23%(RMS)
[54]	I, V, T	BPNN-BSA	SOC	1.74%(RMS)
[57]	I, V, T	NARXNN-LSA	SOC	0.89%
[58]	I, V, Power	$v-$SVR	SOC	1.02%(APE)[20]
[59]	I , V, T[21]	SVR	SOC	0.71% (RMS)
[60]	I, V, Temp.	LS-SVM	SOC	<5%

[1] SOC measured at last second
[2] Difference of Voltage in last Second
[3] Root mean Square Error
[4] X_{1-5}- Discharged capacity for five different current ranges
[5] Regenerative capacity for regenerative current
[6] Temperature
[7] State of available capacity
[8] Avg. relative percentage Error
[9] Battery available capacity
[10] Maximum relative error
[11] Mean Squared Error
[12] Practicable Capacity
[13] Mean Absolute Error
[14] Absolute Error
[15] Avg. temperature of battery modules
[16] Minimum voltage of battery modules
[17] OCV at measured SOC
[18] Battery voltage after instantaneous voltage rise
[19] Battery current before rest interval
[20] Avg. percentage error
[21] Cell temperature in °C

function always exists which can pair (x_i, y_i) with ϵ error but that is not the case always. To allow 4.1 to deal with in-feasible constraints, [61] introduced slack variables which were adopted from soft margin loss function explained in [69]. SVM has been explained in [70], [71], [72].

In [59] , they have used three inputs namely cell voltage (V), Cell Current (A) and cell temperature (°C) of a high-capacity lithium iron manganese phosphate ($LiFeMnPo_4$). In their research they have used Support vector Machines regression technique and the kernel chosen was radial basis function (RBF) which has shown to have more accuracy in highly non-linear problems [73]. The model was tested on two data sets, static and dynamic stress test (DST) according to United States of America battery atrium (USABC) is used [74]. For dynamic stress test and static (constant charging and discharging) profile, their results were measured by using coefficient of determination. The coefficient of determination value of 1.0 is the indication of Support Vector Machine (SVM) model being a perfect fit, while in their study they achieved the coefficient of determination value of 0.98. They have also used the same Support Vector Machine Model to predict the cell voltage using cell SOC (%), Cell current(A) and cell temperature(°C). The coefficient of determination for fitted voltage model was 0.97. In [60], a model based on least square support vector machine to estimate the SOC is proposed. They tested the model on Ni-MH battery pack and other combinations of Urban Dynamometer Driving Schedule (UDDS) and US06 in advanced Vehicle Simulator (ADVISOR) [75].

In [7], a model based on support vector machine is presented using current, voltage, SOC_{t-1} (SOC measured at the end of last second) and ΔV (Variation in

voltage in last second) to estimate the State of charge of large scale Lithium-ion-polymer (LiP) battery pack. The model with potential use for Electric Vehicles extracts support vectors from battery history. The model is validated with help of collecting data from simple SOC test and two more dynamic tests using US06 drive cycles. The first US06 drive cycle in which SOC change is 20% in 63 mins, gives the RMS error of 5.76% with the maximum positive error of up to 12% during the drive cycle. During second more aggressive US06 drive cycle in which the SOC change is 25% in 7.5 mins, yielded the RMS error of 2.5% though maximum error increased to 13%. In [58] a model based on v-SVR is proposed. The model is validated by comparing the results with neural network with data from different drive cycles using ADVISOR software. The results from this model were compared with those obtained from the NN which show that for New York city drive cycle the the average percentage error (APE) reduced from 1.64% in NN to 1.02% in the proposed model.

2.4 Summary

In this chapter, machine learning methods for SOC estimation are reviewed. The main advantage of Machine learning techniques is that it can take characteristics of battery as inputs which is not possible in conventional methods. As discussed earlier, the wide range of inputs are used in estimating the SOC. This allow us the freedom to adapt these techniques to fit the training data more accurately to predict the SOC with less error. The same model can be trained with data from different batteries without much effort. In machine learning models most of the

computation is required while training the model which can be done offline or in the lab with appropriate data and then model can be loaded onto a chip for real time use.In most of the papers, the common problem that most of them face is the estimation of the initial SOC, Although in adaptive it is rectified automatically after sometime. While the learning algorithms have reduced error and improved accuracy of the estimation process but still there is a long way to go before these can be used in real time applications. Lot of research needs to be done in order to find the correct balance between the complexity and the accuracy of the algorithms.

Chapter 3

Proposed SOC Estimation Techniques using ML Algorithms

The continuous monitoring of state of charge (SOC) of a lithium-ion battery is essential to avoid over-charging or over-discharging in order to ensure safe operation as well as to reduce its average life cycle cost. However, an accurate SOC estimation of lithium-ion battery has become a major challenge in the automotive industry. In this chapter, k-nearest neighbour (kNN) and random forest regression (RFR) concepts have been employed to estimate the SOC, based on the measured voltage, current and previous SOC. A Gaussian filter is employed to minimize the errors in SOC estimation techniques. The effectiveness of the proposed hybrid methods is verified on the experimental data of the lithium-ion battery under different standard driving schedules and temperatures.

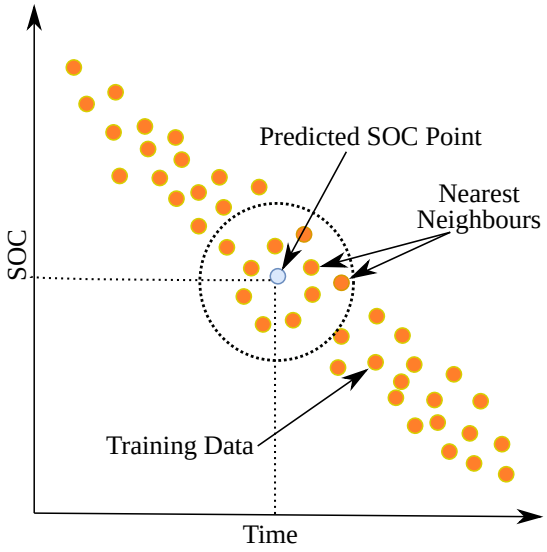


FIGURE 3.1: Working of kNN in 1- dimensional feature space.

3.1 kNearest neighbor Regression

In kNearest neighbor (kNN) Regression algorithm, "nearest neighbours" are used to predict the values of the target variable. To calculate the nearest neighbour a distance method is applied [76]. Distance method used in this study is Euclidean distance which is discussed in next section. After calculating "nearest neighbours", weighted average of nearest neighbors is used as final prediction. The value of target variable from the testing data set is predicted using from the values of target variable in the subgroup of "nearest neighbours" from training data set which is given as:

$$f(x_p) = \frac{\sum_{i=1}^{k} w_k f(x_i)}{\sum_{i=1}^{k} w_k} \tag{3.1}$$

where,

$f(x_p)$ is the predicted value for input x_p,

$f(x_i)$ is the target values from training data set,

w_k is the weighting factor of the k^{th} neighbour and

k is number of neighbours which are considered in making predictions.

The distance dependent weights [77], are calculated by:

$$w_k = \frac{\left(\frac{1}{d_{p,i}}\right)^t}{\sum_{i=1}^{k}\left(\frac{1}{d_{p,i}}\right)^t} \tag{3.2}$$

where,

$d_{p,i}$ is weighted distance between test point x_p and neighbour x_i and

t is a parameter which influences the rate of decrease in w.

3.2 Euclidean Distance

The distance metric used in this experiment was Euclidean Distance. It can be calculated by using the equation given below:

$$D = \sqrt{\sum_{i=1}^{N} w_n (x_t - x_i)^2} \tag{3.3}$$

where,

N denotes the number of features,

x_t is n^{th} feature values of the test point and

x_i is n^{th} feature values of training point and

w_n is the weight assigned to the nth feature.

3.3 Experimental Data

Every machine learning model needs training data for predicting the SOC of a battery. As the batteries are chemically different from each other and behave differently, the data used to train the model for estimating the SOC will differ with each type and the environment where the battery is supposed to be used. For data to be used for training the models, it must meet the following conditions:

1. The data that we have must cover the range the real world operation of learning model is expected to predict. If our training data is not diverse or it does not contain the data points from the whole range of SOC i.e. learning model is trained using the data only from certain range or certain point, the predicted SOC will contain errors which will make the model unfeasible.

2. The training data should be compatible with the data which will be used to estimate the output in real life application. You can not use the data from a mobile phone battery to train a model which will be used to predict the SOC of electric car.

3. Before the model can be deployed in real life applications, it needs to be tested for accuracy. For this, we will need testing data which must be different from training data. This is usually achieved by dividing the data into two parts, training data and testing data.

TABLE 3.1: Specifications of test battery

Type	Nominal Capacity (Ah)	Nominal Voltage (V)	Cut-off Voltage (V)	Maximum Current (A)
18650 $LNMC$	2.0	3.60	2.4 - 4.2	2.2

4. Before feeding the training data to machine learning models, all the data points must be within the same range. To achieve this, data must be normalized to the range of 0 to 1 increasing its computation complexity.

5. The data collected should be accurate. If it contains error then the training process of algorithm will have unexpected errors and the results may contain larger errors or the results may not be favourable at all.

In this experiment three types of simulated drive cycles namely Dynamic stress testing (DST), Federal Urban Driving Schedule (FUDS) and Federal highway driving schedule (US06) are used. The above mentioned test profiles were run on a Lithium Nickel Manganese Cobalt Oxide (LNMC) battery whose capacity is 2 ampere-hour. Cut-off voltage of the cell is from 2.4 to 4.2 volts. The specifications of test cell are written in Table 3.1.

3.3.1 Dynamic Stress Testing Profile

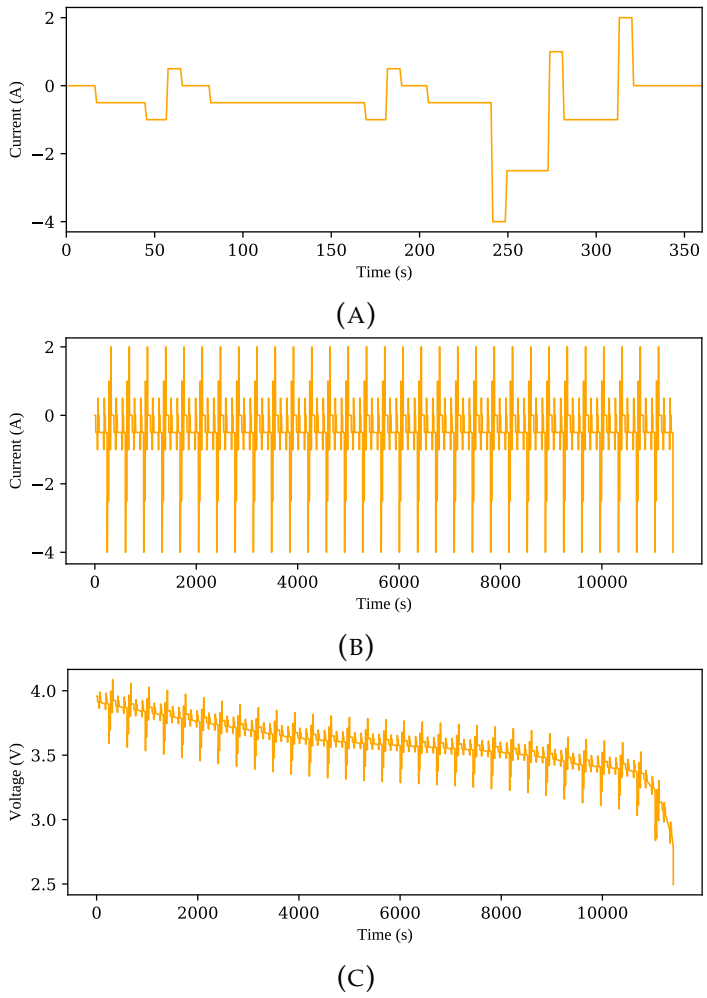
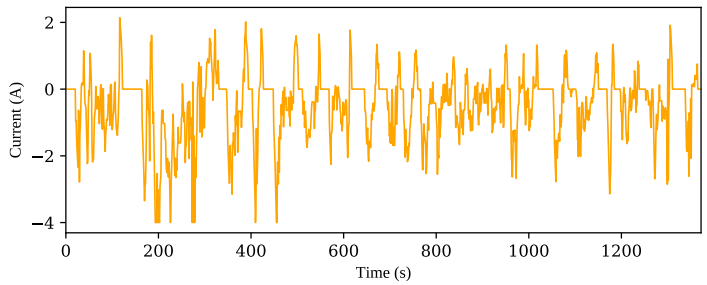


FIGURE 3.2: Dynamic Stress Testing Profile A) Dynamic Stress Testing Cycle B) Full Current cycle during whole test profile C) Full Voltage cycle during whole test profile.

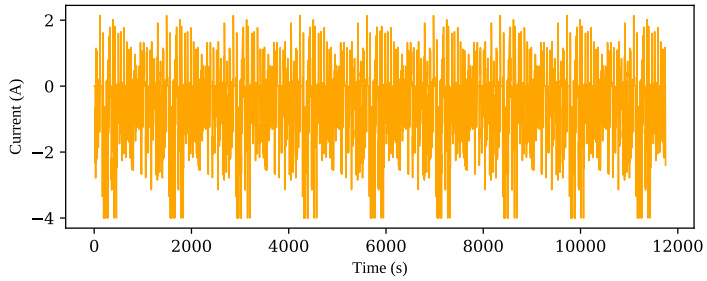
Dynamic Stress Testing (DST) test cycle is 360 seconds in duration. In DST test profile, current stays between 2 ampere to -4 ampere which is shown in Figure 3.2 (A). Figure 3.2 (A), zoomed version of Figure 3.2 (B), shows the current cycle which

is repeated in this test profile. Figure 3.2 (B) and (C) show changes in current and voltage during whole DST test profile.

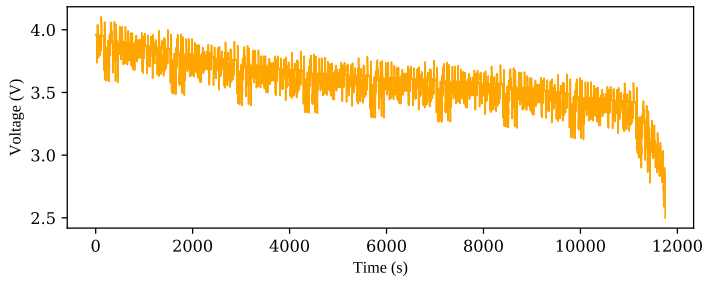
3.3.2 Federal Urban Driving Schedule Test Profile



(A)



(B)



(C)

FIGURE 3.3: FUDS data A) FUDS test cycle B) Full voltage cycle during the FUDS test profile C) Full current cycle during the FUDS test profile.

Federal Urban Driving Schedule (FUDS) is a simulated driving test profile designed by USABC [74]. Current cycle of FUDS test profile was longest of all three

used in this study at 1372 seconds. Current cycle of FUDS is shown in Figure 3.3 (A) in which current stays between 2 to -4 ampere. This current cycle was repeated until SOC of the cell dropped to 0%. Figure 3.3 (B) shows current profile during complete FUDS test profile. Figure 3.3 (C) shows the change in voltage during complete FUDS test profile.

3.3.3 United States Highway Driving Schedule Test Profile

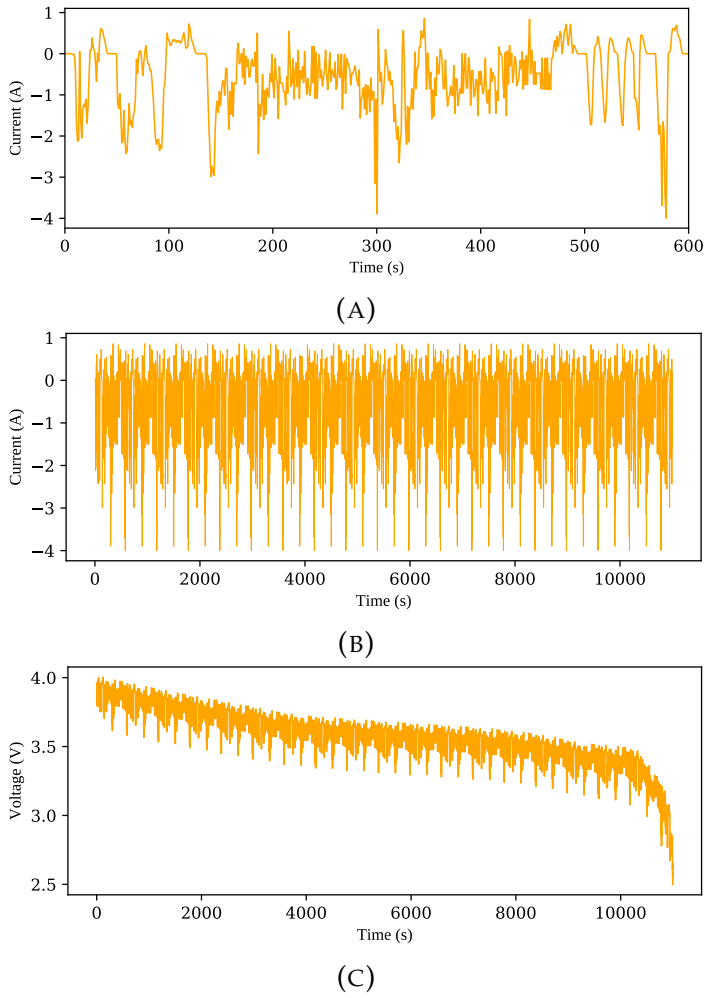


FIGURE 3.4: US06 data A) US06 test cycle B) Full voltage cycle during the US06 test profile C) Full current cycle during the US06 test profile.

United States Highway Driving Schedule (US06) Test Profile is 600 seconds in duration. In US06 test cycle, current stays between 1 to -4 ampere as shown in Figure 3.4(A). Full current and voltage profiles for US06 test are shown in Figures 3.4(B) and (C).

TABLE 3.2: Input variables used in this study.

Input Variables	Range
Current (A)	(-4) - 2
Voltage (V)	2.4 - 4.2
Δ V (V)	0 - 0.1
SOC_{t-1} (%)	0.0 - 0.8

3.4 Input Features and Output

The inputs used in this study are shown in Table 3.2. The inputs used for training the model and predicting SOC are current (A), voltage (V), Δ voltage (difference between voltage from last second) and SOC_{t-1} (SOC at the end of last second). Because SOC_{t-1} is used in predicting SOC, the SOC from the previous step was divided by 100. For example, if the SOC predicted at last step is 70 then SOC_{t-1} for current step will be 70/100 i.e. 0.70. The inputs and range of inputs are shown in Table 3.2. The features for this study were all scaled between the range of [0,1]. The features were normalized by:

$$x = 2\frac{x - x_{min}}{x_{max} - x_{min}} \tag{3.4}$$

where,

x_{min} is minimum values of input vector x and

x_{max} is maximum value of input vector x.

Both training and testing data were normalized before training and testing processes.

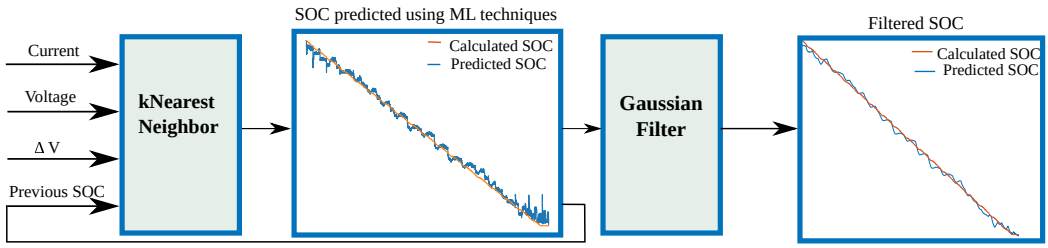


FIGURE 3.5: Working of SOC estimation algorithm using kNN algorithm

3.5 SOC Estimation Method based on kNN

The block diagram for proposed hybrid method based on kNN is shown in Fig. 3.5 [78]. At first, SOC is calculated using kNN, mentioned in Section 3.1, which is filtered by implementing Gaussian filter in the next stage. While implementing kNN, value of k is very important as it plays a very important role in making predictions. The value of k was found to be 10 for the purpose of this study. Value of k is found by trial and error method.

3.6 Gaussian filter

To reduce the variability in the results, a Gaussian filter [79, 80, 81] is introduced. Gaussian filter reduces the variability in the results produced by kNN. The equation used is as follows:

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \tag{3.5}$$

where,

x is window size,

μ is mean value,

σ is standard deviation of Gaussian Distribution.

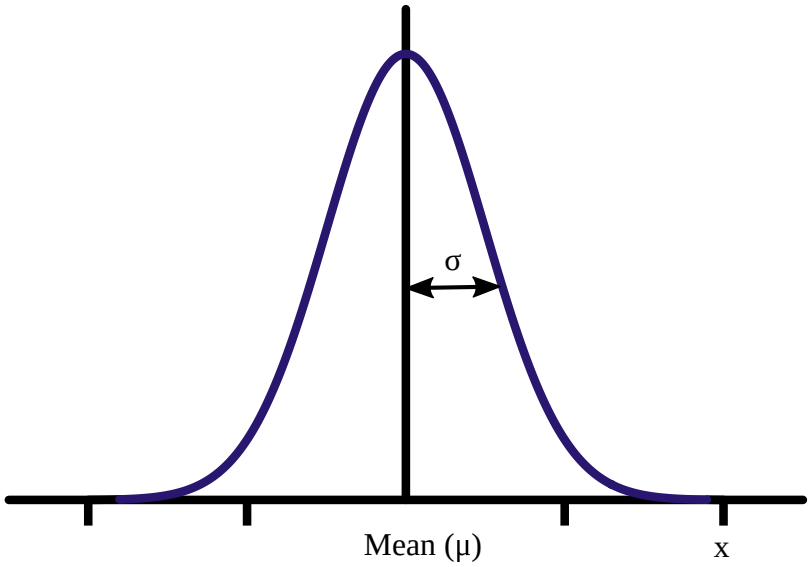


FIGURE 3.6: Working principle of Gaussian filter.

Gaussian filter calculates gaussian weighted moving average over a window of samples. The window size for this study is set at 250.

3.7 Random Forest Regression

Random forest regression (RFR) [82] generates many decision trees for regression and its output is calculated by averaging the output of all decision trees. The working principle of RF regression is shown in Fig. 3.7. The decision tree [83] is a model that does not have any prior tree structure. The structure of tree depends on the

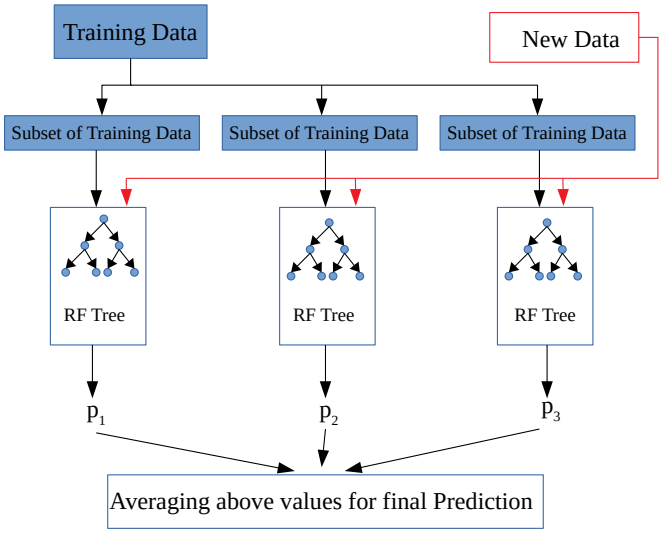


FIGURE 3.7: Working of Random Forest Regression Algorithm

complexity of the training data during the learning stage. Decision tree comprises of two nodes: decision node and leaf node. Every sample of training data is evaluated by the decision nodes and passed onto different nodes depending on the value of the features of the sample.

RF regression generates regression trees using the training data X, which is given by, $X = x_1, x_2, x_3.....x_n$ which produces forest. This method produces k outputs $T_1(x), T_2(x), ..., T_k(x)$ corresponding to each tree. To calculate the final result, all of tree predictions are averaged with the equation given below:

$$RF\left(X\right) = \frac{1}{k}\sum_{k=1}^{k}\hat{T}_k\left(x\right). \tag{3.6}$$

Steps included in RF are shown below:

1. First of all, inputs for random forest regression model are identified which in our case are Current, Voltage, ΔV and SOC_{t-1}.

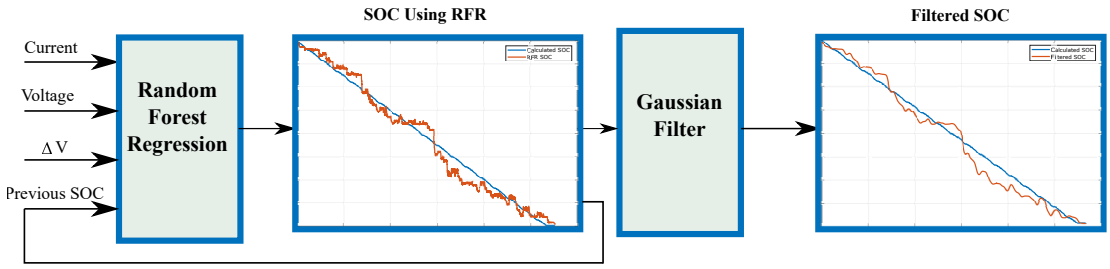


FIGURE 3.8: SOC Estimation method based on Random Forest Regression.

2. Grow each tree in the forest, while optimal number of trees (n = 200) selected, making use of new training set which is generated from original data with replacement. Finding optimal trees depends on evaluation of forest done by using Mean squared error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - Y_i) \tag{3.7}$$

where,

n represents number of samples,

P_i represents predicted values and

Y_i represents true values.

3. Predict SOC values for new data by averaging the predictions of n trees.

3.8 SOC Estimation Method based on RFR

A SOC estimation technique based on RFR is proposed in this section. Figure 3.8 shows the working of proposed SOC estimation method based on RFR. Four inputs

(shown in Table 3.2) namely current, voltage, Δ V (voltage difference between current and last measured state) and SOC_{t-1} (SOC predicted at last second) are used to predict SOC using RFR. The output produced from RF regression contained small variations. To remove the variations, Gaussian filter, as shown in Equation 3.5, is introduced.

3.9 Results

3.9.1 Performance Metrics Used

For evaluating the performance of the models, two performance evaluation statistics are used. Mean Absolute Error (MAE) and Coefficient of Determination (COD) are used to evaluate the models with data from FUDS and US06 at different temperatures. MAE is average absolute error observed in the experiment. MAE is metric of average error expected from the model. MAE is defined as difference between experimental values and true values which can be expressed as:

$$MAE = \frac{\sum_{i=1}^{n} |p_i - y_i|}{n} \tag{3.8}$$

where,

p_i is predicted value of SOC from predictive model and

y_i is true value of SOC.

COD is used to measure the accuracy of machine learning model. It shows the proportion of variation in dependent variable (SOC) from the independent variables (input variables). The value of COD lies between 0 and 1 where 0 signifies

that dependent variable cannot be predicted the set of available independent variables. COD value of 1 means model predicts new values with full accuracy. COD can be calculated using different sum of squares which can be calculated from following equations.

$$SS_{tot} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{3.9}$$

The above equation gives us total sum of squares (SS_{tot}) and mean ($\bar{y}$) of samples is calculated by:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{3.10}$$

Residual sum of squares (SS_{res}) is given by:

$$SS_{res} = \sum_{i=0}^{n} e_i^2 \tag{3.11}$$

where e is the difference between true values and the values predicted by model. The COD is given by:

$$COD = 1 - \frac{SS_{res}}{SS_{tot}} \tag{3.12}$$

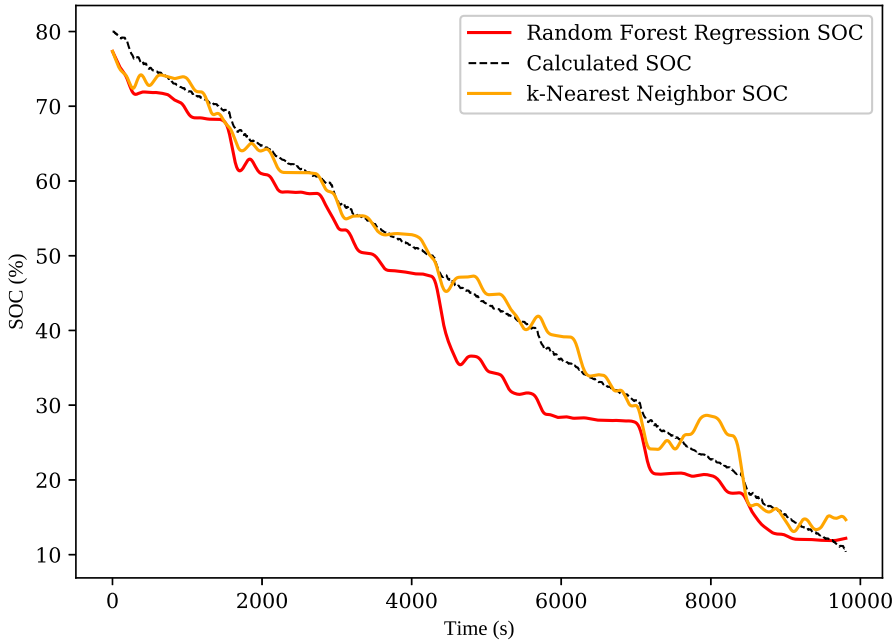


FIGURE 3.9: Results using kNN and RFR for FUDS test profile at $0°$C

3.9.2 Results at $0°$C

This section talks about results obtained from both kNN and RF regression at $0°$C for FUDS test profile. The kNN based SOC estimation based algorithm performed much better than RFR based SOC estimation method. Overall MAE produced by kNN based model was 1.56 as compared to 4.36 produced by RFR based SOC estimation algorithm. Performance of kNN is also better compare to RFR at $0°$C which yielded COD value of 0.98. Although this COD value was worst for kNN during the complete study but it was significantly better than 0.93, value produced by RFR. Figure 3.9 shows the results generated from this study at $0°$C.

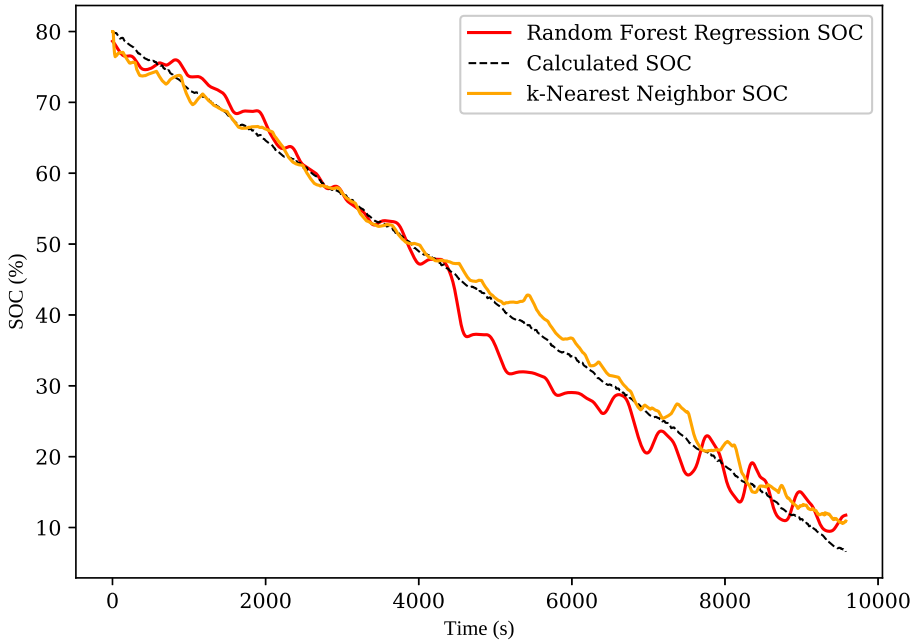


FIGURE 3.10: Results using kNN and RFR for US06 test profile at $0°$C

Figure 3.10 shows results produced by testing US06 test profile at $0°$C. RFR based algorithm performed better than it did for FUDS test profile at same temperature. RFR based algorithm produced MAE value of 2.59 while kNN based algorithm produced MAE of 1.39. kNN based algorithm produced best COD value of study for US06 test profile of 0.99 and performance of RFR also improved with COD value of 0.97 comparing to FUDS test profile at same temperature.

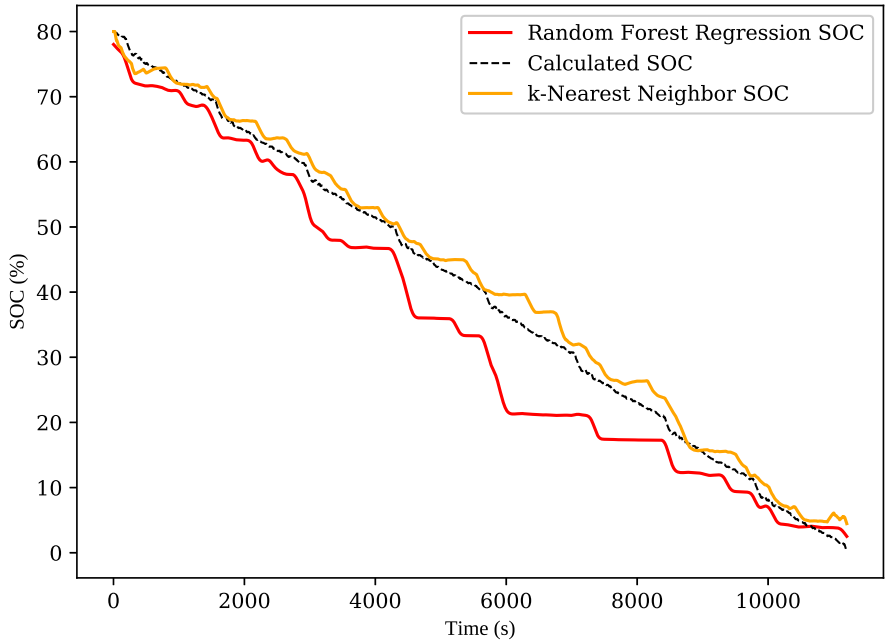


FIGURE 3.11: Results using kNN and RFR for FUDS test profile at $25°C$

3.9.3 Results at $25°C$

At $25°C$, results obtained from FUDS and US06 test profiles are shown in Figure 3.11. At $25°C$, both of the estimation algorithms produced maximum errors. kNN produced maximum MAE value of 1.83 during complete study. For RFR based estimation algorithm, this temperature produced maximum MAE of 5.17 and maximum COD value of 0.92. COD value for kNN based algorithm was 0.99.

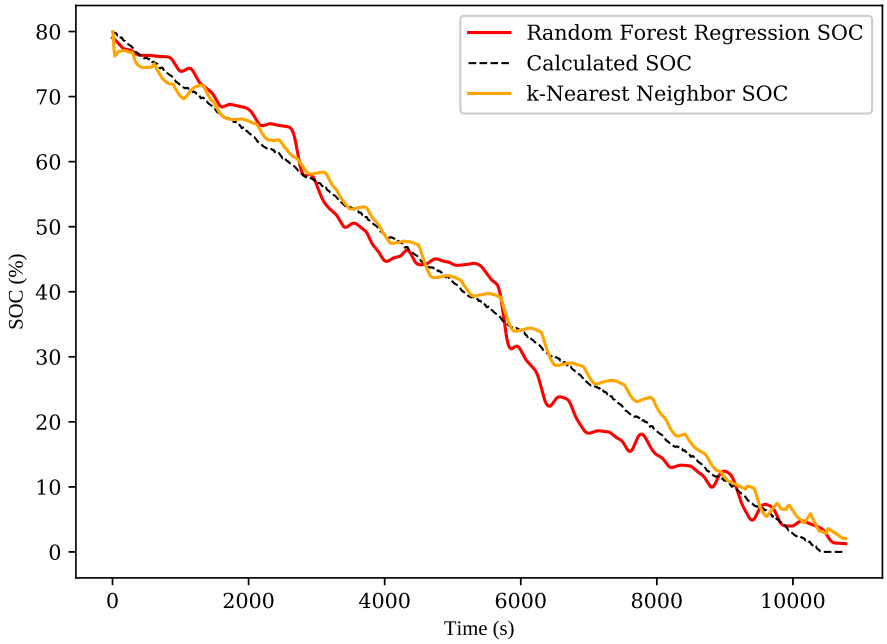


FIGURE 3.12: Results using kNN and RFR for US06 test profile at 25°C

SOC estimation results for US06 test profile are shown in Figure 3.12. kNN based algorithm performed better than RFR based algorithm at this temperature. kNN produced MAE value of 1.52 and COD value of 0.99. RFR also performed good resulting in MAE value of 2.88 and and COD value of 0.97.

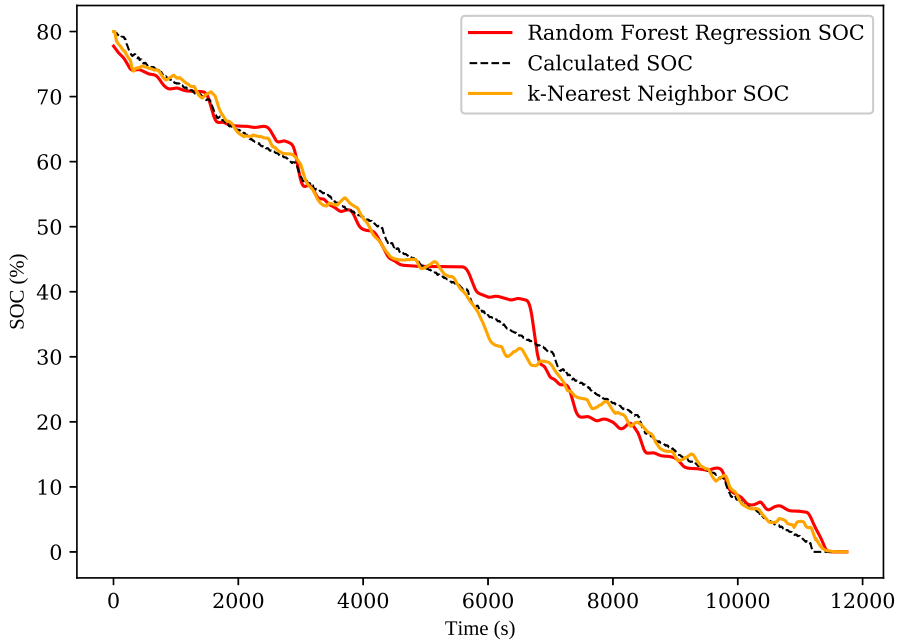


FIGURE 3.13: Results using kNN and RFR for FUDS test profile at $45°$C

3.9.4 Results at $45°$C

Results produced at $45°$C are discussed in this section. Both kNN and RFR based estimation methods produced best results with minimum errors with FUDS test profile at $45°$C which are shown in Figure 3.13. kNN based model produced MAE value of 1.14 while RFR based algorithm produced MAE of 1.97 which were lowest for both of the estimation algorithms. RFR produced best COD value of 0.98 with kNN obtaining COD value of 0.99.

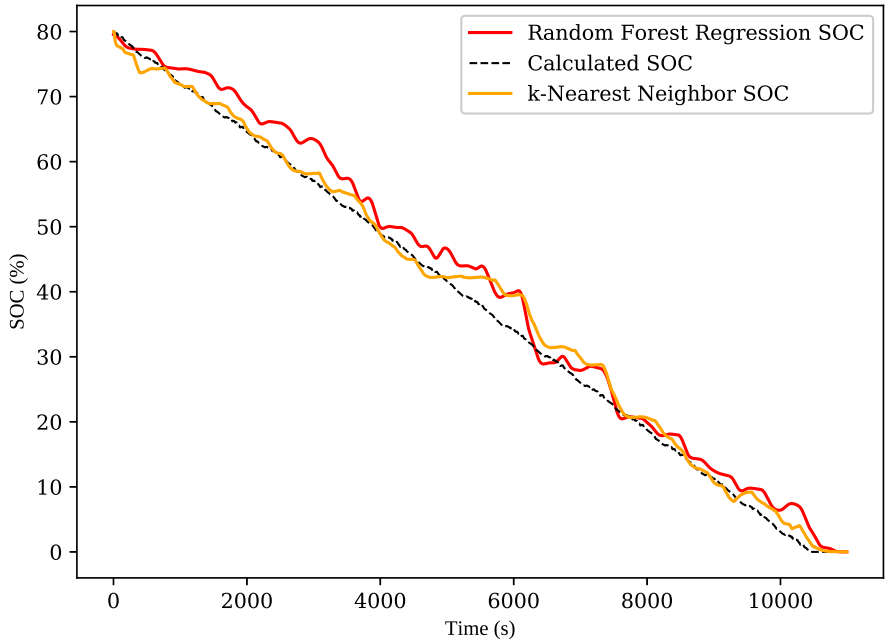


FIGURE 3.14: Results using kNN and RFR for US06 test profile at $45°$C

Results obtained from US06 test profile at $45°$C are shown in Figure 3.14. kNN based estimation estimation algorithm produced good results keeping the MAE value at 1.60 and achieving COD of 0.99. RFR also obtained good results with MAE of 2.88 and COD of 0.97.

3.10 Summary

In this chapter, results from two novel SOC estimation techniques are presented at three different temperatures. Table 3.3 shows performance evaluation of both of estimation techniques. In this chapter, MAE and COD, metrics used for performance evaluation are defined. Data used in this study for training and testing the SOC

TABLE 3.3: Performance evaluation of kNN and RFR based SOC estimation models in this study.

Test Profiles		FUDS			US06		
		0°C	25°C	45°C	0°C	25°C	45°
kNN based	COD_{kNN}	0.98	0.99	0.99	0.99	0.99	0.99
	MAE_{final}	1.56	1.83	1.14	1.39	1.53	1.60
RFR based	COD_{RFR}	0.93	0.92	0.98	0.97	0.97	0.97
	MAE_{final}	4.36	5.17	1.97	2.59	2.88	2.88

estimation models are shown in this chapter. In next chapter, newly proposed SOC estimation techniques are compared with estimation techniques based on SVR and NN.

Chapter 4

Comaprison of Various SOC Estimation Methods

In this chapter, proposed SOC methods are compared with existing SOC estimation techniques. Four SOC estimation techniques based on kNN, RFR, SVR and NN are compared on basis of their accuracy while keeping same parameters. To compare SOC estimation techniques, training and testing data discussed in Section 3.3 and input features shown in Table 3.2 are kept same and result are summarised in Table 4.1.

4.1 Support Vector Regression

Support Vector Machine [61] is a supervised learning algorithm which is used for classification and regression [62]. It has also been used for anomaly detection [63], fault diagnosis [64], clustering [65] and in medical field [66], [67]. SVM uses kernel function to map input data x into a higher dimensional feature space. In ϵ-Support vector regression proposed by [62], the goal is to find a $F(x)$ which stands for all the training data and at the same time is as flat as possible which means for each training data (x_i, y_i), the actually obtained target should contain error less than ϵ. Deviation larger than ϵ is not accepted. [68] describes the equations as follows:

$$\text{minimize:} \quad \frac{1}{2}|w|^2$$

$$\text{subject to:} \quad \begin{cases} \mathrm{y}_i \text{ - } \langle w, \mathrm{x}_i \rangle - b \leq \epsilon \\ \langle w, \mathrm{x}_i \rangle + b - y_i \leq \epsilon \end{cases} \tag{4.1}$$

Here b is bias term while w signifies flatness. Also In (x_i, y_i), x_i is the set of features used to find the target y_i. The assumption in equation 4.1 is such function always exists which can pair (x_i, y_i) with ϵ error but that is not the case always.

4.2 Neural Network

Neural network (NN) is a powerful and intelligent algorithm to map nonlinear inputs to a target output. A neural network consists of three main parts: input layer, output layer and hidden layer(s) as shown in Figure 4.1. Inputs go from input

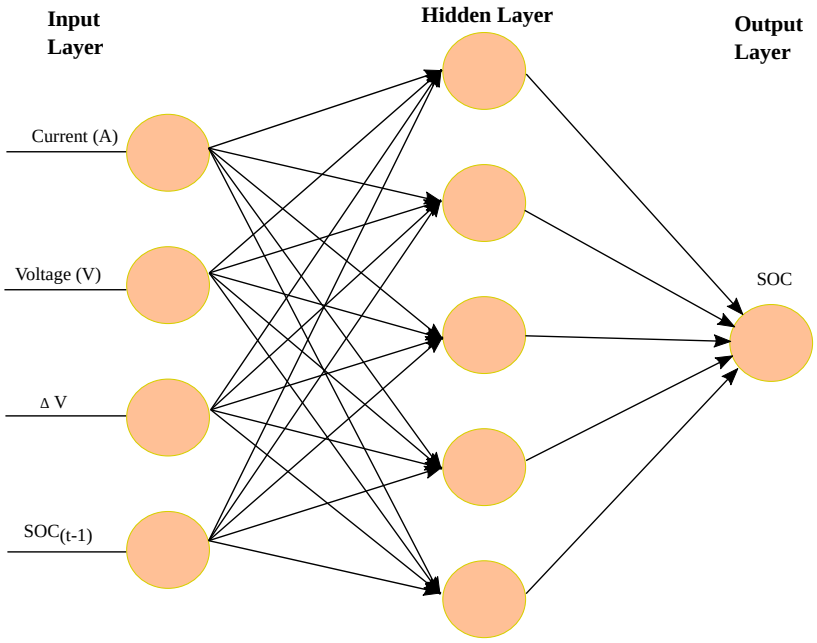


FIGURE 4.1: Working of NN.

layer to output layer from which outputs are produced, which is then compared to the target output which may contain an error. To reduce the error, weights and biases are added in the hidden layer. During initialization, weights and biases are set to random variable with a range and we can also define maximum number of iterations. During each iteration, outputs of neurons are calculated in hidden layer and in output layer. Information about the errors in output layer and hidden layer is calculated using which the weights and bias in the hidden layer are updated to reflect the errors. This process is repeated until we reach the maximum number of iterations or the value of error becomes negligible.

For the purposes of our study, four inputs mentioned in Table 3.2 are selected for the input layer and at output layer SOC is obtained.

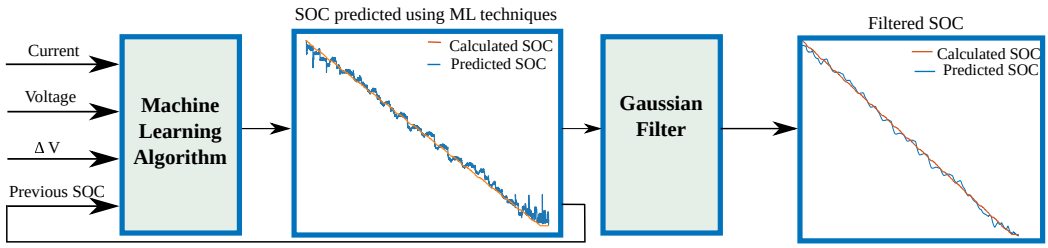


FIGURE 4.2: Working of generalised SOC estimation used for comparison

4.3 SOC Estimation technique used for comparison

There are lot of ML-based SOC estimation techniques proposed in literature but no proper studies are performed under dynamic conditions such as Federal test driving schedules. For comparing SOC estimation algorithms, a generalised SOC estimation algorithm is developed. Figure 4.2 show the generalised SOC algorithm. ML algorithm is used to estimate SOC by using four inputs mentioned in Table 3.2. Input features used in this study are kept unchanged for all of SOC estimation methods. Data used in this study is also unchanged. DST test profile is used as training data for all four of the SOC estimation techniques namely kNN based, RFR based, SVR based and NN based. All four of SOC estimation techniques are tested using two test profiles. FUDS and US06 test profiles are used for performance evaluation of all estimation techniques. Results obtained from these are filtered by using a Gaussian filter. Small variations in results are filtered out by Gaussian filter.

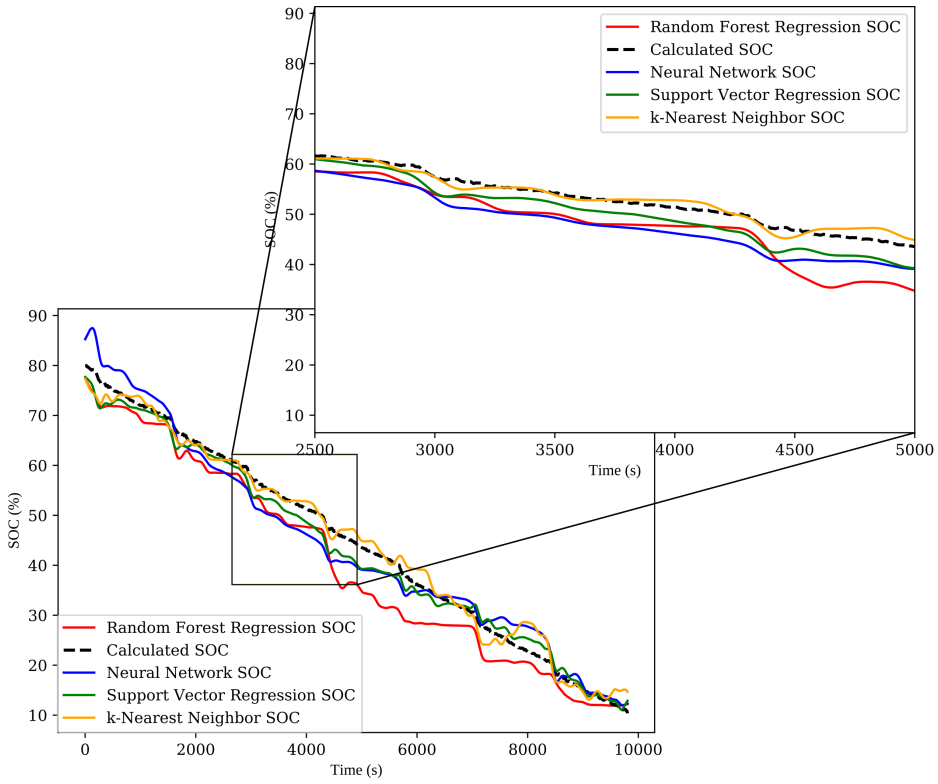


FIGURE 4.3: Comparison of ML-based SOC estimation models for FUDS test profile at $0°C$

4.4 Results

4.4.1 Comparison results at $0°C$

This section presents comparison of four ML-based techniques at $0°C$. Figure 4.3 shows the SOC predictions made by different SOC estimation techniques for FUDS test profile. Best performance of FUDS test profile at this temperature is achieved by kNN with MAE value of 1.56. kNN and SVR based SOC estimation method produced best COD value of 0.98 at this temperature meaning 98% variation in SOC is explained by the input parameters selected. COD of 0.98 shows this is worst performance of kNN when looking at all other temperatures but it still outperformed

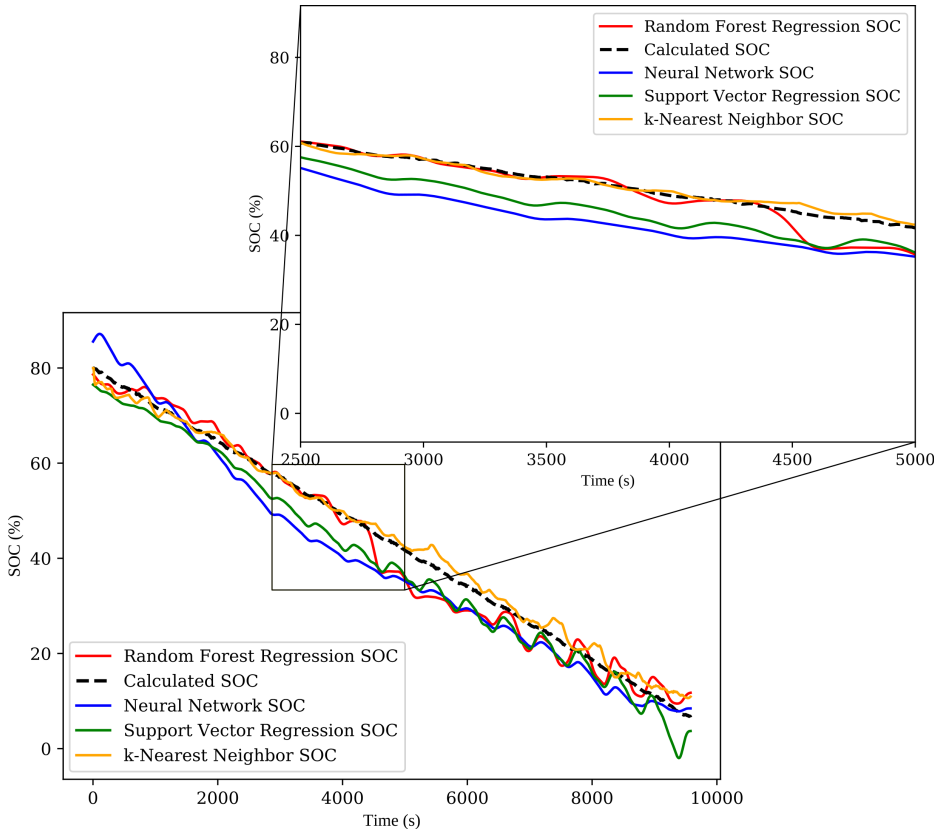


FIGURE 4.4: Comparison of ML-based SOC estimation models for US06 test profile at 0°C

other three estimation techniques which can be seen. RFR produced worst results out of four SOC estimation techniques compared in this study with MAE value of 4.36 and COD value of 0.93. Results from NN were slightly better than RFR with MAE of 3.10 and COD of 0.96.

Figure 4.4 shows predictions of four ML-based estimation techniques using US06 test profile. kNN based estimation technique is best performer out of all four with US06 test profile at 0°C. MAE calculated for kNN based estimation technique is 1.39 with COD of 0.99. RFR based technique performed good achieving COD of 0.97 and MAE of 2.59. NN performed worst for this profile with COD of 0.93 and

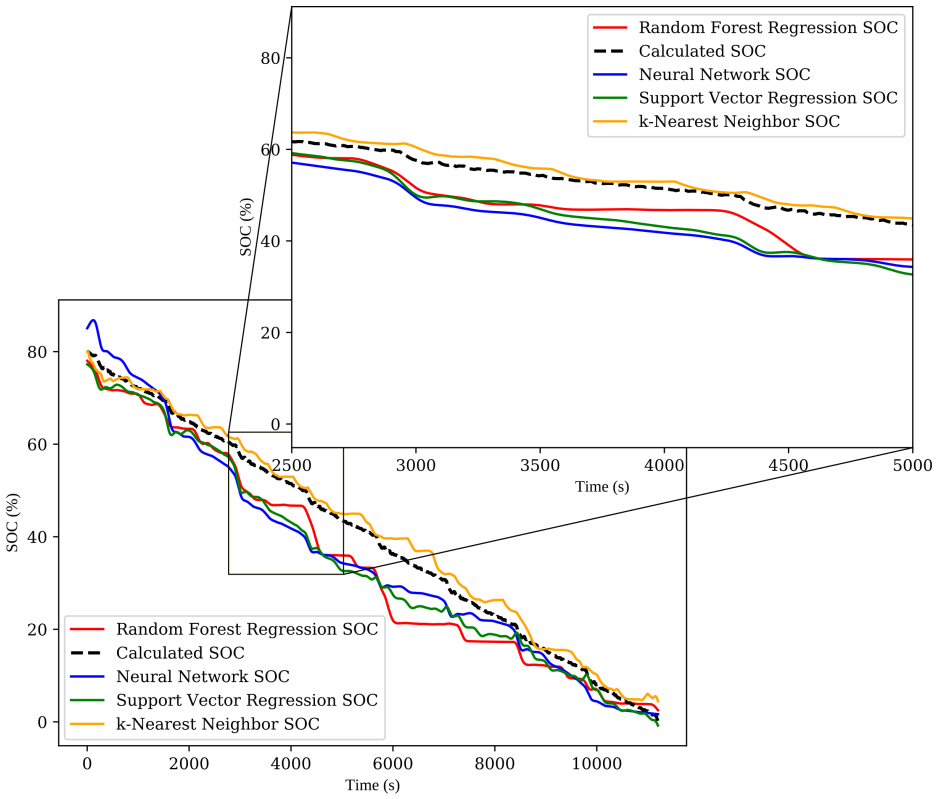


FIGURE 4.5: Comparison of ML-based SOC estimation models for FUDS test profile at 25°C

AME of 4.87. SVR performed better than NN with MAE of 3.89 and COD of 0.95.

4.4.2 Comparison results at 25°C

This section discusses results from both test profiles at 25°C. Figure 4.5 shows prediction results for FUDS test profile. At this temperature, kNN based estimation technique obtained best results with MAE of 1.83 and COD of 0.99. FUDS test profile at 25°C produced highest errors for kNN and RFR based estimation techniques when compared with results of same technique from other temperatures. RFR based estimation method also produced highest error for 25°C with MAE of

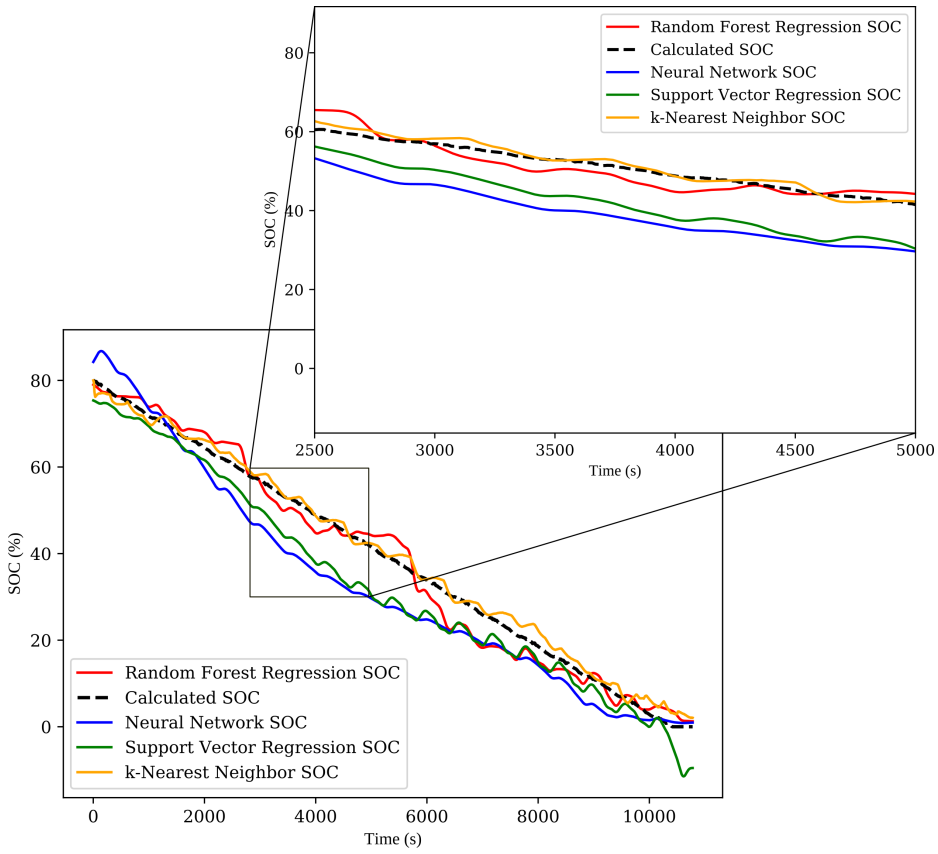


FIGURE 4.6: Comparison of ML-based SOC estimation models for US06 test profile at 25°C

5.17 and COD 0.92. Both NN and SVR produced same COD of 0.93 and estimation techniques based on NN and SVR produced MAE of 4.83 and 4.93 respectively. Figure 4.6 shows prediction results from US06 test profile at 25°C. NN and SVR produced biggest errors of the respective techniques at all temperature. NN based estimation technique produced MAE of 6.90 while COD for NN is calculated to be 0.88 which is worst for complete study. SVR based estimation techniques also produced big MAE value of 5.70 with SVR producing COD of 0.92. kNN based estimation technique produced best results with MAE of 1.53 and COD for kNN is 0.99. RFR based estimation technique performed better than both NN and SVR

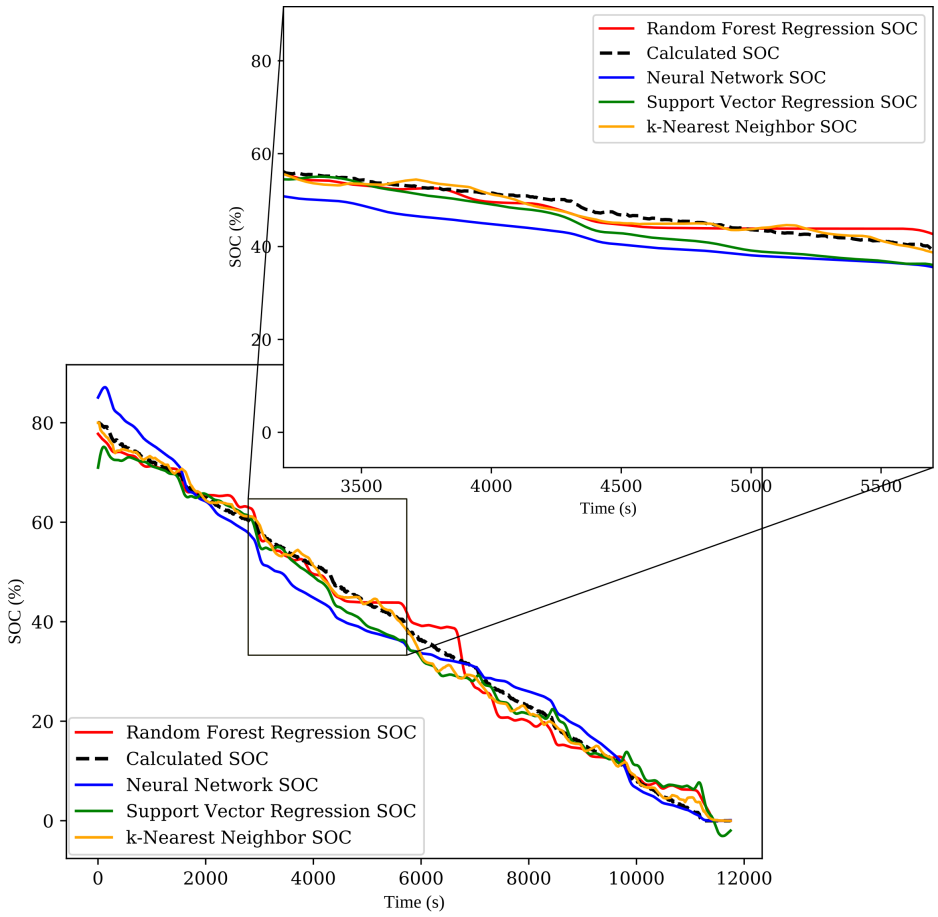


FIGURE 4.7: Comparison of ML-based SOC estimation models for FUDS test profile at $45°C$

based estimation techniques with MAE of 2.88 and RFR obtained COD of 0.97.

4.4.3 Comparison results at $45°C$

This section discusses prediction results from FUDS test profile at $45°C$. At this temperature, all four estimation techniques produced good results with maximum MAE of 2.95 which is obtained by NN based estimation technique. COD of 0.97 for NN is lowest for this temperature. kNN based estimation technique obtained

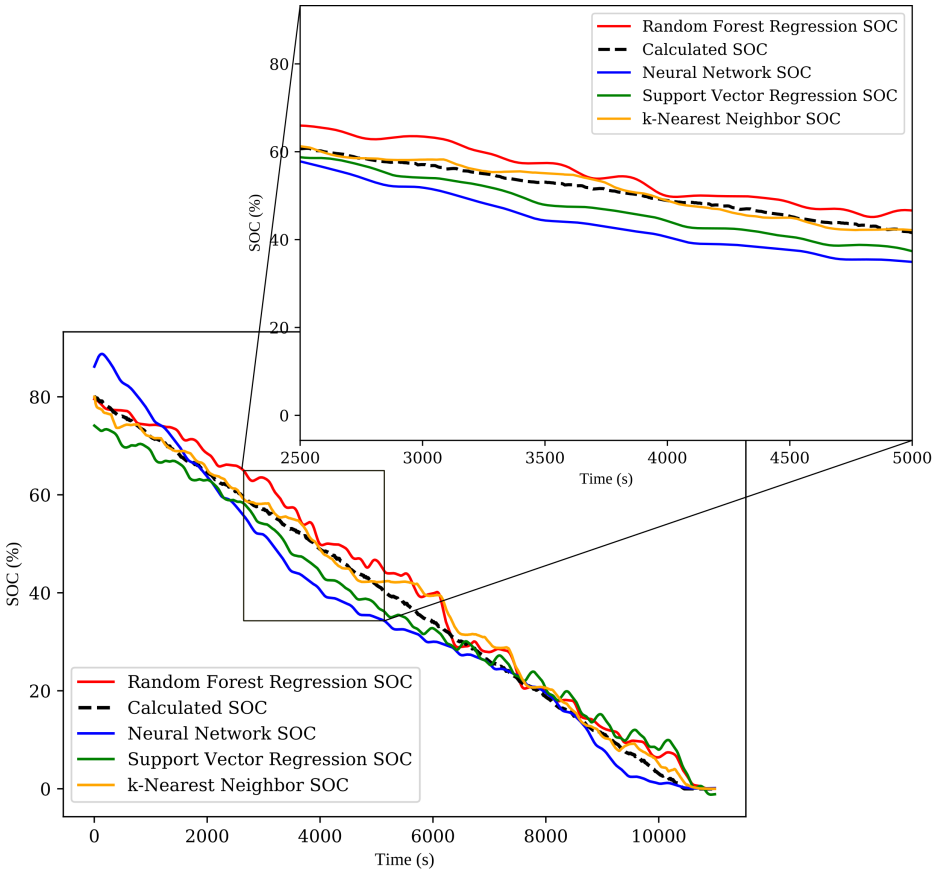


FIGURE 4.8: Comparison of ML-based SOC estimation models for US06 test profile at $45°$C

best result with MAE value of 1.14 which is lowest for our complete study and kNN produced COD of 0.99. RFR based estimation technique obtained MAE of 1.97 while RFR produced COD of 0.98. SVR produced COD of 0.98 with estimation technique based on SVR produced MAE of 2.25.

Results obtained from different ML-based estimation for US06 test profile at $45°$C is shown in Figure 4.8. Best performance of this profile is achieved by kNN based estimation technique with MAE of 1.6 with kNN showing COD of 0.99. NN

TABLE 4.1: Performance evaluation of models studied in this study.

Test Profiles		FUDS			US06		
		0°C	25°C	45°C	0°C	25°C	45°C
kNN based	COD_{kNN}	0.98	0.99	0.99	0.99	0.99	0.99
	MAE_{final}	1.56	1.83	1.14	1.39	1.53	1.60
RFR based	COD_{RFR}	0.93	0.92	0.98	0.97	0.97	0.97
	MAE_{final}	4.36	5.17	1.97	2.59	2.88	2.88
NN based	COD_{NN}	0.96	0.93	0.97	0.93	0.88	0.96
	MAE_{final}	3.10	4.82	2.95	4.87	6.90	3.75
SVR based	COD_{SVR}	0.98	0.93	0.98	0.95	0.92	0.97
	MAE_{final}	1.88	4.93	2.25	3.89	5.70	3.08

based estimation technique is the worst performer for US06 test profile at this temperature shoeing MAE of 3.75 and NN produced COD of 0.96. RFR and SVR based estimation techniques produced MAE of 2.88 and 3.08 respectively.

4.5 Summary

Table 4.1 shows results from four ML-based estimation algorithms. This chapter compared the results from SOC estimation techniques proposed in chapter 3 with NN and SVR based SOC estimation techniques. All four methods were compared using same data and same set of inputs. In this chapter SOC estimation methods based on kNN, RFR, NN and SVR were compared. Results from all methods are shown in Figures 4.3 - 4.8 and performance of all methods is compared in Table 4.1.

While kNN based SOC estimation method produced consistent and better results, it also has a downside. For kNN to work, training data needs to be stored locally. Everytime kNN makes a prediction, it accesses training data to calculate nearest neighbors which increases the computational complexity. For RFR based

SOC estimation method, since it is creating two hundred trees of bootstrap samples from data of one discharge cycle, its performance can be improved by using more training data. While NNs use neurons in hidden layer to make predictions, with addition of every neuron in hidden layer, computation complexity of neural network is increased. Out of all method tested during this study, kNN requires more computation power than other three SOC estimation algorithms. In the next chapter, contributions of this thesis are listed along with discussion on future works.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

Chapter 1 of this thesis discusses battery management systems and its importance. Advantages and disadvantages of Lithium-ion batteries were discussed. SOC estimation process of these batteries is also discussed. Traditional methods of SOC estimation like Coulomb counting method, open circuit voltage method, Kalman filter and impedance spectroscopy are also discussed briefly.

Chapter 2 discusses ML-based SOC methods proposed in literature. This chapter discusses the working of ML-based estimation methods. SOC estimation techniques are based on two main methods: SVR and NN. Performance of SOC estimation techniques using SVR and NN is shown in the form of a table.

Chpater 3 proposes two new SOC estimation techniques based on kNN and RFR algorithm. Both of their performance was verified by testing them with FUDS and US06 drive cycles. Equations used for finding nearest neighbors in kNN and for creating decision trees in RFR are shown shown. Metrics used for performance evaluation of estimation techniques are also defined.

Chapter 4 shows comparison of new proposed techniques with techniques already proposed in literature. This chapter compares four estimation techniques namely kNN based, RFR based, SVR based and NN based. Data sets used in this study are unchanged throughout all methods to give performance evaluation of estimation techniques.

5.2 Contributions

- **kNN based SOC estimation technique**

  A new SOC estimation technique based on kNN algorithm is proposed. This estimation technique reduced errors of the process and produced best results in our study.

- **RFR based SOC estimation technique**

  Hybrid SOC estimation technique based on RFR algorithm is presented. The accuracy and robustness of estimation techniques based on kNN and RFR is confirmed by comprehensive tests performed by using FUDS and US06 test cycles.

- **Comparison of ML-based SOC estimation techniques**

  New techniques (kNN based and RFR based) proposed in this thesis were compared with existing techniques such as SVR and NN using same set of inputs extracted from same US06 and FUDS data sets. Testing and training data were kept same in order to compare performance of these SOC estimation techniques.

5.3 Future Work

There are challenges for machine learning algorithms but their usefulness, accuracy and their ability to take in the battery characteristics as inputs cannot be ignored. This work can be extended to other battery chemistries. Only four inputs are used in this study, more battery characteristics can be incorporated as inputs for better estimation of SOC.

Bibliography

[1] G. L. Plett, "Extended kalman filtering for battery management systems of lipb-based hev battery packs: Part 3. state and parameter estimation," *Journal of Power sources*, vol. 134, no. 2, pp. 277–292, 2004.

[2] W. M. Budzianowski, "Negative carbon intensity of renewable energy technologies involving biomass or carbon dioxide as inputs," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 9, pp. 6507–6521, 2012.

[3] S. M. Rezvanizaniani, Z. Liu, Y. Chen, and J. Lee, "Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (ev) safety and mobility," *Journal of Power Sources*, vol. 256, pp. 110–124, 2014.

[4] K. S. Ng, C.-S. Moo, Y.-P. Chen, and Y.-C. Hsieh, "Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries," *Applied Energy*, vol. 86, no. 9, pp. 1506 –1511, 2009, ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2008.11.021.

[5] L. Zhao, M. Lin, and Y. Chen, "Least-squares based coulomb counting method and its application for state-of-charge (soc) estimation in electric vehicles," *International Journal of Energy Research*, vol. 40, no. 10, pp. 1389–1399, 2016.

[6] J. Wang, B. Cao, Q. Chen, and F. Wang, "Combined state of charge estimator for electric vehicle battery pack," *Control Engineering Practice*, vol. 15, no. 12, pp. 1569–1576, 2007.

[7] T. Hansen and C.-J. Wang, "Support vector based battery state of charge estimator," *Journal of Power Sources*, vol. 141, no. 2, pp. 351–358, 2005.

[8] X. Zhang, X. Kong, G. Li, and J. Li, "Thermodynamic assessment of active cooling/heating methods for lithium-ion batteries of electric vehicles in extreme conditions," *Energy*, vol. 64, pp. 1092–1101, 2014.

[9] M. Coleman, C. K. Lee, C. Zhu, and W. G. Hurley, "State-of-charge determination from emf voltage estimation: Using impedance, terminal voltage, and current for lead-acid and lithium-ion batteries," *IEEE Transactions on industrial electronics*, vol. 54, no. 5, pp. 2550–2557, 2007.

[10] S. Lee, J. Kim, J. Lee, and B. H. Cho, "State-of-charge and capacity estimation of lithium-ion battery using a new open-circuit voltage versus state-of-charge," *Journal of power sources*, vol. 185, no. 2, pp. 1367–1373, 2008.

[11] L Lavigne, J Sabatier, J. M. Francisco, F Guillemard, and A Noury, "Lithium-ion open circuit voltage (ocv) curve modelling and its ageing adjustment," *Journal of Power Sources*, vol. 324, pp. 694–703, 2016.

[12] X. Dang, L. Yan, K. Xu, X. Wu, H. Jiang, and H. Sun, "Open-circuit voltage-based state of charge estimation of lithium-ion battery using dual neural network fusion battery model," *Electrochimica Acta*, vol. 188, pp. 356–366, 2016.

[13] M.-W. Cheng, Y.-S. Lee, M. Liu, and C.-C. Sun, "State-of-charge estimation with aging effect and correction for lithium-ion battery," *IET Electrical Systems in Transportation*, vol. 5, no. 2, pp. 70–76, 2014.

[14] J. Chiasson and B. Vairamohan, "Estimating the state of charge of a battery," *IEEE Transactions on control systems technology*, vol. 13, no. 3, pp. 465–470, 2005.

[15] K.-S. Ng, C.-S. Moo, Y.-P. Chen, and Y.-C. Hsieh, "State-of-charge estimation for lead-acid batteries based on dynamic open-circuit voltage," in *Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International*, IEEE, 2008, pp. 972–976.

[16] J. P. Rivera-Barrera, N. Muñoz Galeano, and H. O. Sarmiento-Maldonado, "Soc estimation for lithium-ion batteries: Review and future challenges," *Electronics*, vol. 6, no. 4, p. 102, 2017.

[17] N. Watrin, B. Blunier, and A. Miraoui, "Review of adaptive systems for lithium batteries state-of-charge and state-of-health estimation," in *Transportation Electrification Conference and Expo (ITEC), 2012 IEEE*, IEEE, 2012, pp. 1–6.

[18] M Mastali, J Vazquez-Arenas, R Fraser, M Fowler, S Afshar, and M Stevens, "Battery state of the charge estimation using kalman filtering," *Journal of Power Sources*, vol. 239, pp. 294–307, 2013.

[19] M. Urbain, S. Rael, B. Davat, and P. Desprez, "State estimation of a lithium-ion battery through kalman filter," in *Power Electronics Specialists Conference, 2007. PESC 2007. IEEE*, IEEE, 2007, pp. 2804–2810.

[20] M. W. Yatsui and H. Bai, "Kalman filter based state-of-charge estimation for lithium-ion batteries in hybrid electric vehicles using pulse charging," in *Vehicle Power and Propulsion Conference (VPPC), 2011 IEEE*, IEEE, 2011, pp. 1–5.

[21] Y. Tian, B. Xia, W. Sun, Z. Xu, and W. Zheng, "A modified model based state of charge estimation of power lithium-ion batteries using unscented kalman filter," *Journal of power sources*, vol. 270, pp. 619–626, 2014.

[22] G. L. Plett, "Sigma-point kalman filtering for battery management systems of lipb-based hev battery packs: Part 2: Simultaneous state and parameter estimation," *Journal of power sources*, vol. 161, no. 2, pp. 1369–1384, 2006.

[23] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, Ieee, 2000, pp. 153–158.

[24] M. S. El Din, A. A. Hussein, and M. F. Abdel-Hafez, "Improved battery soc estimation accuracy using a modified ukf with an adaptive cell model under real ev operating conditions," *IEEE Transactions on Transportation Electrification*, 2018.

[25] Y. Zheng, L. Lu, X. Han, J. Li, and M. Ouyang, "Lifepo4 battery pack capacity estimation for electric vehicles based on charging cell voltage curve transformation," *Journal of Power Sources*, vol. 226, pp. 33–41, 2013.

[26] Q. Yan and Y. Wang, "Predicting for power battery soc based on neural network," in *Control Conference (CCC), 2017 36th Chinese*, IEEE, 2017, pp. 4140–4143.

[27] R. Yang, R. Xiong, H. He, H. Mu, and C. Wang, "A novel method on estimating the degradation and state of charge of lithium-ion batteries used for electrical vehicles," *Applied Energy*, vol. 207, pp. 336–345, 2017.

[28] L. Rozaqi and E. Rijanto, "Soc estimation for li-ion battery using optimum rls method based on genetic algorithm," in *Information Technology and Electrical Engineering (ICITEE), 2016 8th International Conference on*, IEEE, 2016, pp. 1–4.

[29] N. Khayat and N. Karami, "Adaptive techniques used for lifetime estimation of lithium-ion batteries," in *Electrical, Electronics, Computer Engineering and their Applications (EECEA), 2016 Third International Conference on*, IEEE, 2016, pp. 98–103.

[30] L. Sanchez, I. Couso, and J. C. Viera, "Online soc estimation of li-fepo4 batteries through a new fuzzy rule-based recursive filter with feedback of the heat flow rate," in *Vehicle Power and Propulsion Conference (VPPC), 2014 IEEE*, IEEE, 2014, pp. 1–6.

[31] H. Sheng and J. Xiao, "Electric vehicle state of charge estimation: Nonlinear correlation and fuzzy support vector machine," *Journal of Power sources*, vol. 281, pp. 131–137, 2015.

[32] A. Fotouhi, K. Propp, and D. J. Auger, "Electric vehicle battery model identification and state of charge estimation in real world driving cycles," in *Computer Science and Electronic Engineering Conference (CEEC), 2015 7th*, IEEE, 2015, pp. 243–248.

[33] J. Xu, B. Cao, Z. Chen, and Z. Zou, "An online state of charge estimation method with reduced prior battery testing information," *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 178–184, 2014.

[34] Y. Ma, Q. Chen, and Z. Qi, "Research on the soc definition and measurement method of batteries used in evs," *JOURNAL-TSINGHUA UNIVERSITY*, vol. 41, no. 11, pp. 95–97, 2001.

[35] J. Garche, A. Jossen, and H. Döring, "The influence of different operating conditions, especially over-discharge, on the lifetime and performance of lead/acid batteries for photovoltaic systems," *Journal of Power Sources*, vol. 67, no. 1-2, pp. 201–212, 1997.

[36] R. Chandrasekaran, W. Bi, and T. F. Fuller, "Robust design of battery/fuel cell hybrid systems—methodology for surrogate models of pt stability and mitigation through system controls," *Journal of Power Sources*, vol. 182, no. 2, pp. 546–557, 2008.

[37] R. Xiong, J. Cao, Q. Yu, H. He, and F. Sun, "Critical review on the battery state of charge estimation methods for electric vehicles," *IEEE Access*, vol. 6, pp. 1832–1843, 2018.

[38] M. U. Cuma and T. Koroglu, "A comprehensive review on estimation strategies used in hybrid and battery electric vehicles," *Renewable and Sustainable Energy Reviews*, vol. 42, pp. 517–531, 2015.

[39] C. H. Piao, W. L. Fu, Jin-Wang, Z. Y. Huang, and C. Cho, "Estimation of the state of charge of ni-mh battery pack based on artificial neural network," in

INTELEC 2009 - 31st International Telecommunications Energy Conference, 2009, pp. 1–4. DOI: 10.1109/INTLEC.2009.5351908.

[40] C. Chan, E. Lo, and S. Weixiang, "The available capacity computation model based on artificial neural network for lead–acid batteries in electric vehicles," *Journal of Power Sources*, vol. 87, no. 1, pp. 201 –204, 2000, ISSN: 0378-7753. DOI: https://doi.org/10.1016/S0378-7753(99)00502-9.

[41] S.-k. Song, "State-of-charge measuring method using multilevel peukert's equation," English, *Journal of Power Sources*, vol. 70, no. 1, p. 157, 1998.

[42] W. He, N. Williard, C. Chen, and M. Pecht, "State of charge estimation for li-ion batteries using neural network modeling and unscented kalman filter-based error cancellation," *International Journal of Electrical Power & Energy Systems*, vol. 62, pp. 783 –791, 2014, ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2014.04.059.

[43] Z. Chen, S. Qiu, M. A. Masrur, and Y. L. Murphey, "Battery state of charge estimation based on a combined model of extended kalman filter and neural networks," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*, IEEE, 2011, pp. 2156–2163.

[44] W. Shen, "State of available capacity estimation for lead-acid batteries in electric vehicles using neural network," *Energy Conversion and Management*, vol. 48, no. 2, pp. 433 –442, 2007, ISSN: 0196-8904. DOI: https://doi.org/10.1016/j.enconman.2006.06.023.

[45] W. Shen, C. Chan, E. Lo, and K. Chau, "A new battery available capacity indicator for electric vehicles using neural network," *Energy Conversion and*

Management, vol. 43, no. 6, pp. 817 –826, 2002, ISSN: 0196-8904. DOI: https://doi.org/10.1016/S0196-8904(01)00078-4.

[46] M. Ismail, R. Dlyma, A. Elrakaybi, R. Ahmed, and S. Habibi, "Battery state of charge estimation using an artificial neural network," in *Transportation Electrification Conference and Expo (ITEC), 2017 IEEE*, IEEE, 2017, pp. 342–349.

[47] L. Kang, X. Zhao, and J. Ma, "A new neural network model for the state-of-charge estimation in the battery degradation process," *Applied Energy*, vol. 121, pp. 20 –27, 2014, ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2014.01.066.

[48] D. S. Broomhead and D. Lowe, "Radial basis functions, multi-variable functional interpolation and adaptive networks," Royal Signals and Radar Establishment Malvern (United Kingdom), Tech. Rep., 1988.

[49] H. Guo, J. Jiang, and Z. Wang, "Estimating the state of charge for ni-mh battery in hev by rbf neural network," in *Intelligent Systems and Applications, 2009. ISA 2009. International Workshop on*, IEEE, 2009, pp. 1–4.

[50] C.-H. Cai, Dong-Du, Z.-Y. Liu, and H. Zhang, "Artificial neural network in estimation of battery state of-charge (soc) with nonconventional input variables selected by correlation analysis," in *Proceedings. International Conference on Machine Learning and Cybernetics*, vol. 3, 2002, 1619–1625 vol.3. DOI: 10.1109/ICMLC.2002.1167485.

[51] S Grewal and D. Grant, "A novel technique for modelling the state of charge of lithium ion batteries using artificial neural networks," 2001.

[52] J. Peng, Y. Chen, and R. Eberhart, "Battery pack state of charge estimator design using computational intelligence approaches," in *Battery Conference on Applications and Advances, 2000. The Fifteenth Annual*, IEEE, 2000, pp. 173–177.

[53] J. Chen, Q. Ouyang, C. Xu, and H. Su, "Neural network-based state of charge observer design for lithium-ion batteries," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 1, pp. 313–320, 2018.

[54] M. A. Hannan, M. S. H. Lipu, A. Hussain, M. H. Saad, and A. Ayob, "Neural network approach for estimating state of charge of lithium-ion battery using backtracking search algorithm," *IEEE Access*, vol. 6, pp. 10 069–10 079, 2018.

[55] R. Bayir and E. Soylu, "Real time determination of rechargeable batteries' type and the state of charge via cascade correlation neural network," *Elektronika ir Elektrotechnika*, vol. 24, no. 1, pp. 25–30, 2018.

[56] S. E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture," in *Advances in neural information processing systems*, 1990, pp. 524–532.

[57] M. H. Lipu, M. A. Hannan, A. Hussain, M. H. Saad, A. Ayob, and F. Blaabjerg, "State of charge estimation for lithium-ion battery using recurrent narx neural network model based lighting search algorithm," *IEEE Access*, 2018.

[58] Q.-S. Shi, C.-H. Zhang, and N.-X. Cui, "Estimation of battery state-of-charge using ν-support vector regression algorithm," *International Journal of Automotive Technology*, vol. 9, no. 6, pp. 759–764, 2008.

[59] J. C. Álvarez Antón, P. J. G. Nieto, C. B. Viejo, and J. A. V. Vilán, "Support vector machines used to estimate the battery state of charge," *IEEE Transactions on Power Electronics*, vol. 28, no. 12, pp. 5919–5926, 2013, ISSN: 0885-8993. DOI: 10.1109/TPEL.2013.2243918.

[60] X. Wu, L. Mi, W. Tan, J. L. Qin, and M. N. Zhao, "State of charge (soc) estimation of ni-mh battery based on least square support vector machines," in *Advanced Materials Research*, Trans Tech Publ, vol. 211, 2011, pp. 1204–1209.

[61] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[62] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999, ISSN: 1045-9227. DOI: 10.1109/72.788640.

[63] G. Wang, J. Yang, and R. Li, "Imbalanced svm-based anomaly detection algorithm for imbalanced training datasets," *ETRI Journal*, vol. 39, no. 5, pp. 621–631, 2017.

[64] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560 –2574, 2007, ISSN: 0888-3270. DOI: https://doi.org/10.1016/j.ymssp.2006.12.007.

[65] Z. Wang, Y.-H. Shao, L. Bai, and N.-Y. Deng, "Twin support vector machine for clustering," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 10, pp. 2583–2588, 2015.

[66] G. Orrù, W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 4, pp. 1140 –1152, 2012, ISSN: 0149-7634. DOI: https://doi.org/10.1016/j.neubiorev.2012.01.004.

[67] K. Drosou and C. Koukouvinos, "Proximal support vector machine techniques on medical prediction outcome," *Journal of Applied Statistics*, vol. 44, no. 3, pp. 533–553, 2017.

[68] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[69] K. P. Bennett and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable sets," *Optimization methods and software*, vol. 1, no. 1, pp. 23–34, 1992.

[70] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, ACM, 1992, pp. 144–152.

[71] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in neural information processing systems*, 1997, pp. 155–161.

[72] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[73] M. A. Hannan, M. Lipu, A Hussain, and A Mohamed, "A review of lithium-ion battery state of charge estimation and management system in electric

vehicle applications: Challenges and recommendations," *Renewable and Sustainable Energy Reviews*, vol. 78, pp. 834–854, 2017.

[74] U. S.A.B. C. (USABC), "Electric vehicle battery test procedures manual,"

[75] "Advisor 2.1: A user-friendly advanced powertrain simulation using a combined backward/forward approach," *IEEE Trans. on Vehicular Technology.*, vol. 48, no. 6, pp. 1751–176, 1999.

[76] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[77] G. Tutz and D. Koch, "Improved nearest neighbor classifiers by weighting and selection of predictors," *Statistics and Computing*, vol. 26, no. 5, pp. 1039–1057, 2016, ISSN: 1573-1375. DOI: 10.1007/s11222-015-9588-z.

[78] M. S. Sidhu, D. Ronanki, and S. Williamson, "Hybrid state of charge estimation approach for lithium-ion batteries using k-nearest neighbour and gaussian filter-based error cancellation," in *2019 IEEE 28th International Symposium on Industrial Electronics (ISIE)*, IEEE, 2019, pp. 1506–1511.

[79] Y. Li, M. Abdel-Monem, R. Gopalakrishnan, M. Berecibar, E. Nanini-Maury, N. Omar, P. van den Bossche, and J. Van Mierlo, "A quick on-line state of health estimation method for li-ion battery with incremental capacity curves processed by gaussian filter," *Journal of Power Sources*, vol. 373, pp. 40–53, 2018.

[80] K. Ito, "Gaussian filter for nonlinear filtering problems," in *Decision and Control, 2000. Proceedings of the 39th IEEE Conference on*, IEEE, vol. 2, 2000, pp. 1218–1223.

[81] H. H. Afshari, S. A. Gadsden, and S Habibi, "Gaussian filters for parameter and state estimation: A general review of theory and recent trends," *Signal Processing*, vol. 135, pp. 218–238, 2017.

[82] Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J. C.-W. Chan, P. van den Bossche, J. Van Mierlo, and N. Omar, "Random forest regression for online capacity estimation of lithium-ion batteries," *Applied energy*, vol. 232, pp. 197–210, 2018.

[83] L. Breiman, *Classification and regression trees*. Routledge, 2017.