National Institute of Technology Karnataka, Surathkal

Battery SOH prediction using Machine Learning Techniques

Submitted by:

Aditya S Gourishetty (171EE103)

Dwijesh Athrey K (171EE116)

M Nikhil Bharath (171EE223)

Supervisor: Dr. Debashisha Jena

Department of Electrical and Electronics Engineering

Contents

1 Declaration

We hereby declare that the project entitled "Battery SOH prediction using Machine Learning Techniques" was carried out by us during the VI Semester of the academic year 2019-2020. We declare that this is our original work and has been completed successfully according to the direction of our guide Dr.Debashisha Jena and as per the specifications of NITK Surathkal.

Place: NITK,Surathkal
Date: 09-06-2020

Aditya S Gourishetty (171EE103)

Dwijesh Athrey K (171EE116)

M Nikhil Bharath (171EE223)

2 Acknowledgement

Firstly we would like to thank Dr. Debashisha Jena, Dept. of Electrical and Electronics for his prolonged support which was crucial for driving this project to its completion. We are also immensely grateful to Dr.Debashisha Jena for giving his time for providing his profound insights on our initial and intermediate results and accordingly suggesting changes to our methods to get better results without which our project could not have arrived at this stage.

We would also like to thank Dr. Jena's teaching assistants for providing us with appropriate guidance in acquiring the required knowledge on Battery charge-discharge characteristics which was essential for this project. We are also thankful to Electrical and Electronics Department, National Institute of Technology Karnataka for providing us the platform to learn and acquire the relevant skills and for giving the opportunity to be a part of such projects.
Lastly, we thank our families and friends for providing us with constant support throughout the course of the project.

3 Abstract

In today's world, batteries are indispensable and hence, ensuring that the batteries are safe to use at any point is of paramount importance. Prognostics and health management is one methodology that has been applied successfully to many fields, one of which is rechargeable battery systems. One of the most important aspects of this is to accurately estimate the state of health (SOH) of the battery. In this project, we aim to use different machine learning techniques such as probablistic neural networks (PNN), K-nearest neighbours (KNN) and support vector machine regression on battery charging and discharging cycle data to estimate the SOH of the battery.

4 Introduction

Electrical energy is the most commonly used form of energy in the world, given the ease and convenience with which it can be converted into other different forms of energy as per requirements. The conventional sources of power for electricity production are coal, oil and natural gas, which even today account for about two-thirds of the total production. Newer, eco-friendly sources of power for electricity generation such as wind, water and solar energy have gained a lot of traction of late due to the emphasis on reducing global warming.

All the above mentioned sources are mainly used in case of large scale generation. However, in the case of small-scale requirements, the most popular source of electricity is batteries. A battery is a device consisting of one or more electrochemical cells, which convert chemical energy into electrical energy. First invented by Alessandro Volta in 1800, batteries have evolved into an ubiquitous energy source that powers numerous devices we use in our daily lives.

4.1 Types of batteries

Batteries are mainly classified into two types:

1. **Primary batteries:**
   They can produce current immediately after assembly and cannot be recharged. They are generally used in low-current applications and places where other electrical power sources are not readily available. They generally have high energy densities. Examples : Zinc-carbon battery, alkaline battery

Figure 1: Alkaline batteries

2. **Secondary batteries:**
   Also known as rechargeable batteries, they are manufactured with the active chemicals in the discharged state and have to be charged before first use. These batteries can be recharged after use by passing an electric current through its electrodes to reverse the chemical reactions that occur during discharge. They generally have lower energy densities compared to primary batteries. Examples: Lead-acid battery, Nickel-Cadmium (NiCd) battery, Nickel metal hydride(NiMH) battery, Lithium-ion(Li-ion) battery.

Figure 2: Lead Acid batteries

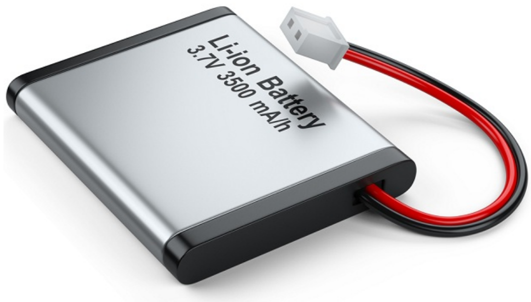

Figure 3: Lithium-Ion batteries

4.2 Prognostics and Health Management of rechargeable batteries

Prognostics and Health Management (PHM) is a technique used to predict potential failures of various devices and systems used in multiple fields such as manufacturing, aerospace etc. Given that rechargeable batteries are used today in a large number of portable devices as well as electric vehicles, ensuring that the health of the battery is within safe limits and can be continued to be used is of paramount importance.
The most important characteristic of a battery for estimating its health is known as the State of Health (SOH) of the battery. It is generally expressed as a percentage and is defined to be the maximum charge that the battery can release after it has been fully charged as a fraction of the original rated charge capacity of the battery.

$$\mathrm{SOH} = \frac{Q_{\mathrm{max}}}{Q_{\mathrm{rated}}} \times 100\%$$

Figure 4

Where:
Q_{max} = Maximum charge that the battery can release
Q_{rated} = Rated charge capacity of the battery

Many studies have been performed to estimate the SOH of a battery accurately. Methods used commonly include electrochemical impedance spectroscopy, equivalent circuit models and statistical data-driven models. Of late, data driven models have been gaining a lot of traction, given the exponential developments in the field of machine learning.

In this project, we have used three different machine learning techniques to estimate SOH of the battery. They are:

1. Probabilistic Neural Network (PNN)
2. K - Nearest Neighbours (KNN)
3. Support Vector Machine (SVM)

The dataset used is obtained from NASA's Prognostics Center of Excellence website. It contains information about the current and voltage profiles in both constant current and constant voltage modes, as well as discharge cycles and impedance mode.
The indicators that represent the health of the battery and hence can be used for SOH estimating using the various machine learning techniques mentioned above are chosen following Lin, Liang, Chen;2013. [1]

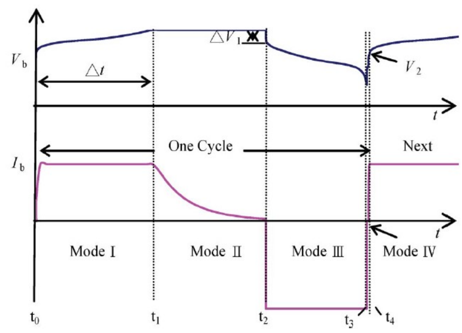

Figure 5: Voltage and current profiles during charging and discharging cycles

The different modes are as follows:
Mode I : Charging at constant current
Mode II : Charging at constant voltage
Mode III : Discharging at constant current
Mode IV : Open circuit
The indicators that are chosen to estimate SOH of the battery are Δt, ΔV1, V2, which are defined respectively as the charging time at constant current, instantaneous drop in voltage at the start of the discharge cycle and the open circuit voltage after the battery is fully discharged.

These three indicators are the features that the machine learning algorithm has to be trained on and be used to estimate the SOH of the battery.

5 Data preprocessing

The Data taken from NASA's Prognostics Center of Excellence contains four .mat data files which contain data as mentioned on the website is regarding a set of four Li-ion batteries (5, 6, 7 and 18) which were run through 3 different operational profiles (charge, discharge and impedance which are similar to the cycles mentioned in [1]) at room temperature where charging was carried out in a constant current (CC) mode at 1.5A until the battery voltage reached 4.2V. Charging was then resumed in a constant voltage (CV) mode, until the charge current dropped to 20mA. Then discharge was carried out at a constant current (CC) level of 2A until the battery voltage fell to 2.7V, 2.5V, 2.2V and 2.5V for batteries 5, 6, 7 and 18 respectively.
The structure of the data in these files is given in figure

Data Structure:

cycle: top level structure array containing the charge, discharge and impedance operations
- type: operation type, can be charge, discharge or impedance
- ambient_temperature: ambient temperature (degree C)
- time: the date and time of the start of the cycle, in MATLAB date vector format
- data: data structure containing the measurements
  - for charge the fields are:
    - Voltage_measured: Battery terminal voltage (Volts)
    - Current_measured: Battery output current (Amps)
    - Temperature_measured: Battery temperature (degree C)
    - Current_charge: Current measured at charger (Amps)
    - Voltage_charge: Voltage measured at charger (Volts)
    - Time: Time vector for the cycle (secs)
  - for discharge the fields are:
    - Voltage_measured: Battery terminal voltage (Volts)
    - Current_measured: Battery output current (Amps)
    - Temperature_measured: Battery temperature (degree C)
    - Current_charge: Current measured at load (Amps)
    - Voltage_charge: Voltage measured at load (Volts)
    - Time: Time vector for the cycle (secs)
    - Capacity: Battery capacity (Ahr) for discharge till 2.7V
  - for impedance the fields are:
    - Sense_current: Current in sense branch (Amps)
    - Battery_current: Current in battery branch (Amps)
    - Current_ratio: Ratio of the above currents
    - Battery_impedance: Battery impedance (Ohms) computed from raw data
    - Rectified_impedance: Calibrated and smoothed battery impedance (Ohms)
    - Re: Estimated electrolyte resistance (Ohms)
    - Rct: Estimated charge transfer resistance (Ohms)

Figure 6: NASA Battery Data Structure

5.1 Pre-processing

To obtain the features of charging time delta T, instantaneous drop in voltage delta V1 and open circuit voltage V2 from the continuous current, voltage and time vector data as given in figure , the matlab scripts with the following code snippets were used:

Overall data collection: Consecutive cycles of charge and discharge in the data are treated as part of being one complete cycle and contribute to a single data point.

```
load('B0005.mat');
BData=B0018;
Data_del_V=[];
Data_V=[];
Data_T=[];
SOH=[];
for i=1:length(BData.cycle)
    if(strcmp(BData.cycle(i).type,'charge'))
        [x , y] = charge_data2(BData.cycle(i).data.Time,BData.cycle(i).data.Current_measured,
    BData.cycle(i).data.Voltage_measured);
        Data_del_V=[Data_del_V;y];
```

```
        Data_T=[Data_T;x];
    end

    if(strcmp(BData.cycle(i).type,'discharge'))
        Data_V=[Data_V;discharge_data2(BData.cycle(i).data.Time,BData.cycle(i).data.
    Voltage_measured)];
        SOH=[SOH;BData.cycle(i).data.Capacity];
    end
end
m=min(length(Data_del_V),length(Data_V));
Data=[Data_del_V(1:m) Data_V(1:m) Data_T(1:m) SOH(1:m)];
```

Function for delta V1 and delta V2: Here the charge cycle data for charging current is passed through low pass filter to remove noise while calculating delta T using which delta V1 is calculated through voltage data of the same cycle.

```
function [del_t,del_V] = charge_data2(charge_t,charge_I,charge_V)
% lowpass filter
DI=diff(lowpass(rmmissing(charge_I),0.1));
del_t=0;

% for finding the sudden peak in the graph for del_t
for i=100:length(DI)
    if(DI(i)<-0.005)
        del_t=charge_t(i);
        break
    end
end

%for finding del_V1
peak_V=charge_V(i);
if(peak_V~=0)
    del_V=peak_V-charge_V(length(charge_V));
else del_V=0;
end

end
```

Functions for V2: Here for V2, directly the end of discharge cycle is taken.

```
function V2= discharge_data2(discharge_t,discharge_V)
V2=discharge_V(end);
end
```

6 Prediction Models and their results:

6.1 Probabilistic neural network:

Probabilistic neural network is a form of artificial neural network wherein the information moves in only forward direction starting from the input node. Generally used in classification and pattern recognition problems, the algorithm involves calculating the probability density function (PDF) for each class. Using the PDF of all the classes obtained while testing, the algorithm then classifies the new input data to a particular class based on the highest posterior probability computed. The PNN network consists of four layers. These include input layer, pattern layer, summation layer and the output layer.

1. The input layer consists of nodes containing the values obtained from the measurements made.
2. The pattern layer consists of gaussian functions formed using the given data points as centers.
3. The summation layer performs average operation of the outputs from the pattern layer for each class.
4. The output layer selects the maximum of the probabilistic density function. The associated class is then determined as the output result.

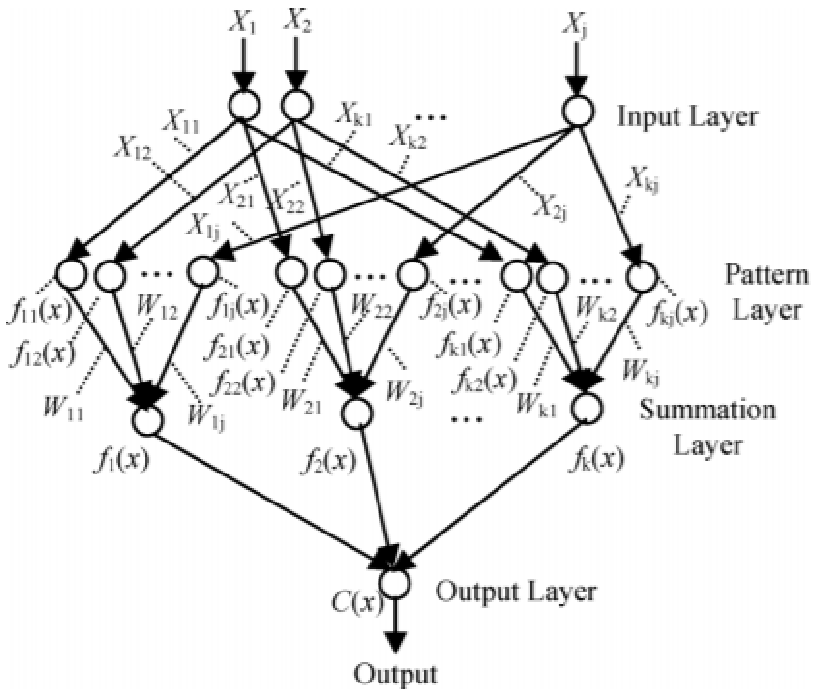

Figure 7

The above network is governed by these following equations:

$$\begin{aligned} \|X - X_{kj}\| &= (X - X_{kj})' \cdot (X - X_{kj}) \\ &= \sum_i (X_i - X_{kj})^2 \end{aligned}$$

$$f_{kj}(X) = \exp\left(\frac{-\|X - X_{kj}\|}{2\sigma^2}\right)$$

$$f_k(X) = \left(\frac{1}{n_k}\right) \sum_{j=1}^{n_k} W_{kj} f_{kj}(X)$$

Figure 8

The input parameters taken for our estimation include constant current charging time (Delta(t)), the instantaneous voltage drop at the start of discharge (Delta(V1)) and the open circuit voltage of a fully charged battery (Delta(V2)). The output is to determine the State of Health (SoH) of the battery. There are a total of 636 samples of measurements extracted from the dataset. The training and testing data are segregated using 'train_test_split' function (imported from scikit-learn) with a ratio of 0.95:0.05. Below is the complete network architecture pertaining to the model used for our estimation. The variable 'k' (=604) represents the number of training examples and 'j' (=3) represents the number of parameters.

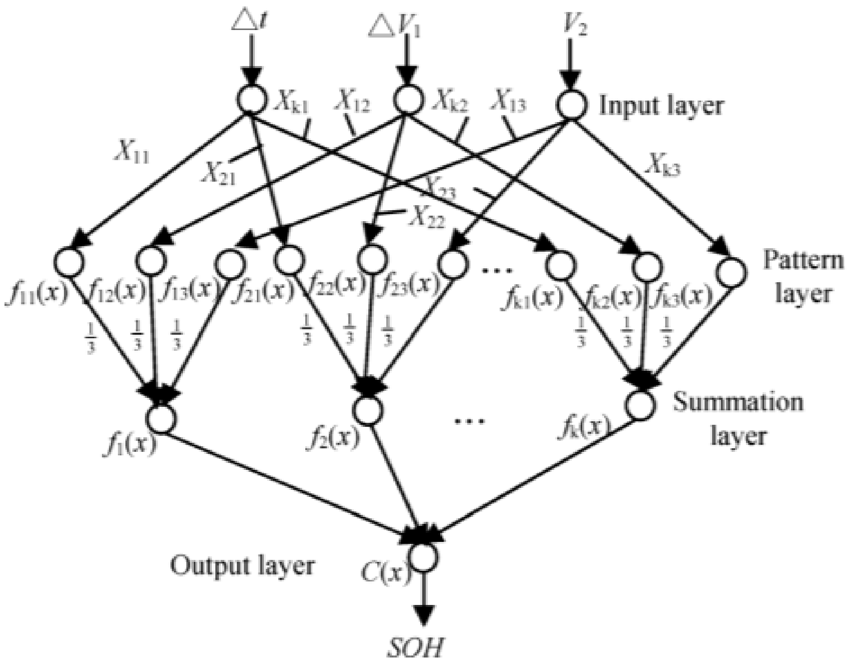

Figure 9

Training: The training process merely consists of forming the weight matrix to represent the training data (Delta(t), Delta(V1) and Delta(V2)) and setting up the smoothing parameter, sigma = 0.1. There are a total of 604 measurement samples as part of the matrix. Below is a visual representation of the weight matrix that is to be formed.

$$X_{kj} = \begin{array}{cc} & \begin{array}{ccc} \Delta t_k & \Delta V_{1(k)} & \Delta V_{2(k)} \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ k \end{array} & \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots \\ x_{k1} & x_{k2} & x_{k3} \end{bmatrix} \end{array}$$

Figure 10

Testing: There are a total of 32 samples for testing which are then given as inputs to the network. The pattern layer computes the euclidean distance between the test inputs and all the existing samples in the weight matrix as per the equation (1) in the figure (6). The gaussian function is then applied to this as per the equation (2) in figure (6) and the result is stored in f(X)kj. Now, the averaging of the pattern layer takes place as per the equation (3) in the figure (6) where the weights are applied such that the sum of weights (Wk1 + Wk2 + Wk3 = 1). In this case it is taken to be each for the 3 parameters. The result is then stored in fj(x). Finally, the output layer takes the maximum value from the summation layer and its corresponding label is determined as the State of Health (SoH) of the battery for the given sample of test input.

Result: The model is implemented by hard coding through python mainly using numpy and pandas libraries for optimized data manipulation and pyplot for data visualisation.
The results obtained are shown in figure

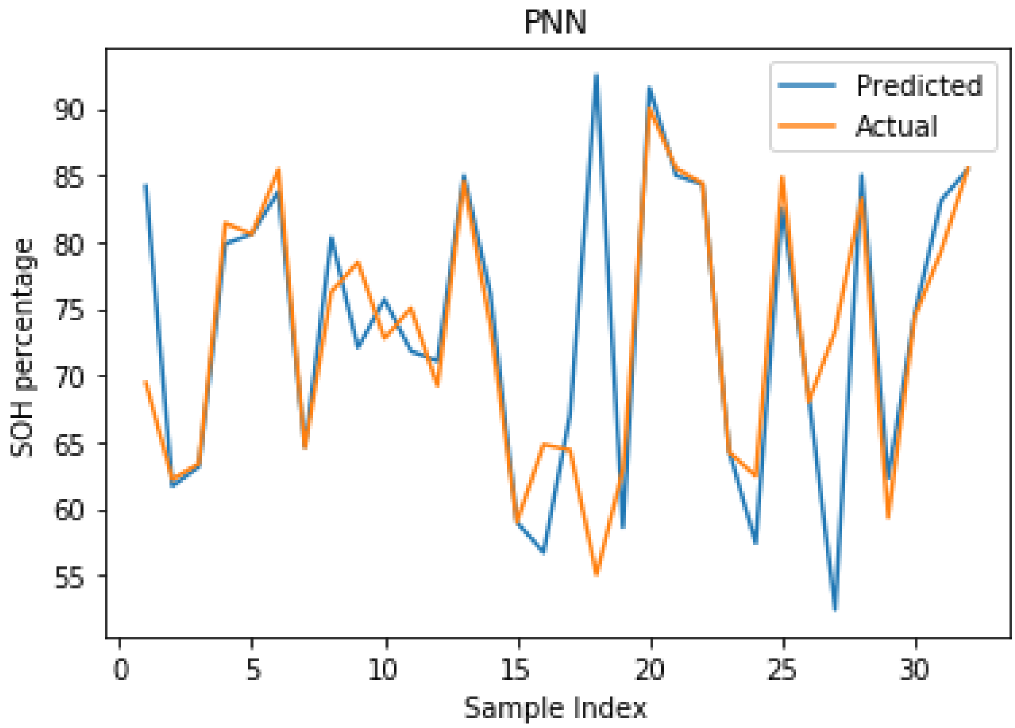

Figure 11: PNN Predictions and Actual Values of SOH

In figure one can see that that the predictions obtained on the testing data fairly coincide with that of actual values. Here we obtained a mean squared error of about 0.5% signifying a satisfactory result.

6.2 KNN:

K-Nearest Neighbours is one of the simplest machine learning algorithms. The implementation was done using scikit-learn python library. For this application we have chosen this algorithm with values of K ranging from 1 to 5 and predicting with the closest K neighbours and recorded the results where we obtained a mean squared error of about 0.3% which is a better result than PNN. Result are shown in figure.

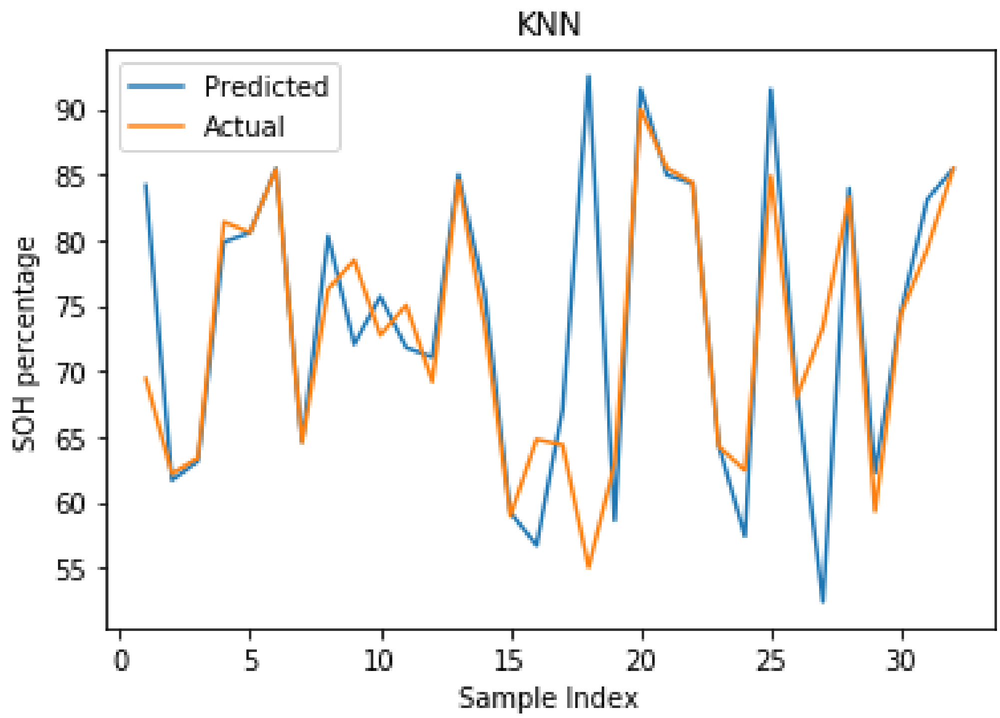

Figure 12: KNN Predictions and Actual Values of SOH

6.3 Support Vector Machine regressor:

Similar to KNN for we used scikit-learn python library for implementing this model by also tuning C and epsilon values for best results. For this model too we got a similar result of about 6% of mean squared error. The results are shown in figure

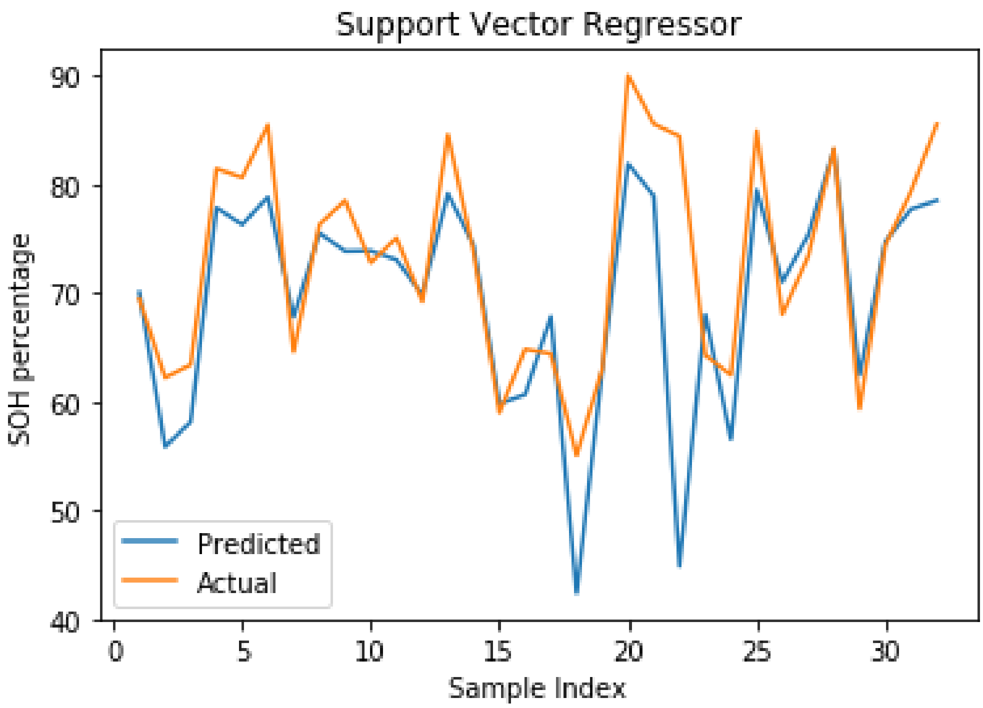

Figure 13: SVGR Predictions and Actual Values of SOH

7 Conclusion:

In summary we have concluded that KNN produces significantly best results for predicting SOH given a similar kind of dataset. Also, the accuracy of features extracted can be further improved by reduction of noise in the dataset which in turn will improve the accuracy of the model. A comparison of various models is shown in figure.

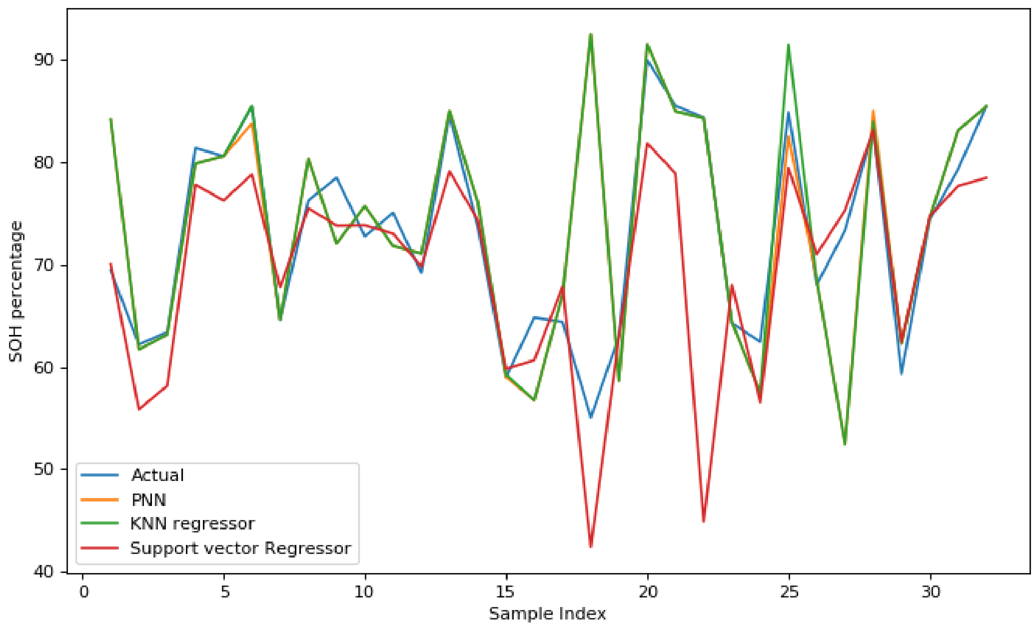

Figure 14: Comparison of predictions by various models and Actual Values of SOH for test data

8 References

Research citations

1. Ho-Ta Lin, Tsorng-Juu Liang, Senior Member, IEEE, and Shih-Ming Chen, Student Member, IEEE "Estimation of Battery State of Health Using Probabilistic Neural Network"

2. Phattara Khumprom and Nita Yodo "A Data-Driven Predictive Prognostic Model for Lithium-Ion Batteries based on a Deep Learning Algorithm"

3. Yohwan Choi, Seunghyoung Ryu , Kyungnam Park, AND Hongseok Kim , (Senior Member, IEEE). "Machine Learning-Based Lithium-Ion Battery Capacity Estimation Exploiting Multi-Channel Charging Profiles"

Other Sources

1. Figure.2 Lead acid battery - https://4.imimg.com/data4/YH/IQ/MY-...-batteries-500x500.jpg

2. Figure.3 Lithium ion Battery - https://www.eletimes.com/wp-content...ithium-Ion-Battery.jpg

3. Battery SOH information - https://en.wikipedia.org/wiki/State$_of_health$