Prediction of Remaining Useful Life (RUL) of Lithium ion (Li-ion) Batteries

MSc Research Project
Data Analytics

Rashmikant T. Shukla
Student ID: x18181236

School of Computing
National College of Ireland

Supervisor: Dr. Rashmi Gupta

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Rashmikant T. Shukla
Student ID:	x18181236
Programme:	Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Dr. Rashmi Gupta
Submission Due Date:	17/08/2020
Project Title:	Prediction of Remaining Useful Life (RUL) of Lithium ion (Li-ion) Batteries
Word Count:	8095
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard speci ed in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	27th September 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not su cient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator o ce must be placed into the assignment box located outside the o ce.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Prediction of Remaining Useful Life (RUL) of Lithium ion (Li-ion) Batteries

Rashmikant T. Shukla
x18181236

Abstract

In recent time Li-ion battery gained popularity because of their high charge density, portability and longer life span. It compliments the human quest for green energy. As many of the green energy applications like Electrical Vehicle, Wind Energy and Solar Energy use Li-ion battery as their energy storage device. So, a better and intelligent Remaining Useful Life (RUL) prediction model will improve the reliability of these systems. This research states that autoencoder can be used to learn time-based battery parameter and their dimension reduction. Features from an autoencoder are fed into another neural network which predicts RUL of battery. NASA battery degradation data set is used for this analysis and target data are extracted based on the geometric features. The model evaluation is based on R-square and Mean Squared Error.

1 Introduction

We heavily rely on machines, devices and equipment in our day-to-day life. Many of these systems have an energy storage device, battery as their integral part. Various batteries are used in these systems, but Li-ion batteries are emerged as one of the best performing options because of its high charge density and long-life span. These machines are meant to falter after the continuous use over a period. To ensure the smooth functioning of battery-based equipment, maintenance is performed. It is essential to monitor the health of batteries to ensure the proper functioning of these systems. There are di erent maintenance programs like reactive maintenance, preventive maintenance and predictive maintenance which are adopted for these systems.

Reactive maintenance is performed when equipment stops working. It is the traditional way which may leads to downtime period and there is always an uncertainty about the current health of the equipment. In the preventive maintenance, periodic maintenance is scheduled to ensure that machine does not brake while in operation. In this process the selection of maintenance period is of utmost importance otherwise it may result into over maintenance or under maintenance. These both scenarios are risky and possess substantial time and economic loss. In predictive maintenance one collects the data regarding the health of the equipment in terms of various sensor and operational data, then accordingly maintenance of the device can be scheduled. This method is cost e ective, and it reduces the chances of downtime.

Predictive maintenance in batteries are used for their health monitoring, which includes prediction of battery's State of Charge (SOC) and Remaining Useful Life (RUL).

SOC indicates the current capacity of batteries in comparison to its rated capacity and RUL indicates how much longer battery is going to last under current working conditions. This research is specifically focused on the prediction of RUL of Li-ion batteries. Accurate RUL prediction helps in overall maintenance of the system and gives the clear idea about replacement of the batteries.

Predictive maintenance domain got a great boost by advancement of sensor technology, Internet of Things (IOT) and computational capabilities. Sensors are used to track performance of different parts by being integrated with the devices. This generates a substantial amount of data, which is then stored and analysed. Based on these data, predictions are made with the help of various machine learning models. Same approach is used in electronic devices, home appliances and industrial equipment. If we talk about batteries, then traditionally the preventive techniques were used for maintenance this is also due to its enclosed structure and complex electrochemical nature. This limits the usage of sensors. Still, the earlier research has shown the potential of using internal current, voltage and impedance profiles of batteries for health monitoring but these were only limited to experimental levels (Vutetakis and Viswanathan; 1995).

It took some time for the world to acknowledge the potential of Li-ion batteries, but eventually it happened. Even, 2019 Chemistry Nobel price were given to the scientists, responsible for the emergence of the Li-ion batteries.[1] Now one can see the blooming market of the Li-ion batteries as the world is exploring new alternate energy options like solar energy, wind energy, electrical vehicles etc and all these are using Li-ion batteries as energy storage devices. It is projected that valuation of Li-ion batteries will be doubled in coming five years.[2] The reliability and longevity of these systems can be ensured by enhancing the performance of the Li-ion batteries. This can be achieved by an accurate health monitoring of the batteries. Initial research, Goebel et al. (2008), has shown how data of the Li-ion batteries can be stored and different machine learning techniques can be very effective in health monitoring of the batteries. Ibid has shown the potential of using relations between capacity and impedance parameter to predict the RUL of the battery. Even simple baseline regression models were able to make reasonable predictions. Degradation profile of the batteries are non-linear in nature and accurate prediction can be made by using algorithms which are good at learning non-linear parameters.

Several research papers have used Probabilistic Algorithms, Support Vector Machine (SVM), Kalman Filters, Optimization Algorithm etc. to predict the State of Health (SOH), SOC and RUL of the batteries. These algorithms use different internal parameter profiles like Electrochemical Impedance Spectroscopy (EIS), open circuit voltage, voltage under load in order to apply machine learning models (Li et al.; 2017). However, it needs a good understanding of the internal chemistry of the Li-ion batteries based on which appropriate profile can be chosen for the final prediction.

In recent years, this domain has seen research papers which are using deep learning and neural network methods for RUL prediction (Shen et al.; 2020; Qu et al.; 2019; Zhang et al.; 2018). The performance of the deep learning methods is very promising, especially Long Short Term Memory (LSTM) model has given the state-of-the-art results. Deep learning models have an advantage over the traditional methods as the network itself learns the features, so it provides leverage against the expertise in complex electrochemical profiles of batteries which was compulsory in previous researches. Most of the research

[1] https://www.nobelprize.org/prizes/chemistry/2019/press-release/

[2] https://www.prnewswire.co.uk/news-releases/global-lithium-ion-power-battery-market-size-projecte
html

focuses on a single battery or gives di erent models for di erent kind of batteries and usually they provide only short-term RUL predictions. To address this, a combination of autoencoder and Deep Neural Network (DNN) is used in this research, in order to produce prediction on multi-battery dataset. In previous research an autoencoder was used to analyse and known to be work well with the non-linear trends which is an added advantage over the other models.

1.1 Research Question

The research question for this paper is *"How well a hybrid model with Autoencoder and Deep Neural Network can predict the RUL of Li-ion batteries based on their charging and discharging cycle?"*

1.2 Research Objective and Contribution

Objective of this research are,

1. A literature review on the SOC and RUL prediction.
2. To understand the charging, discharging cycle of battery in order to develop a data-driven model to predict remaining time before the failure of the battery.
3. Feature Extraction from raw data.
4. Implementation and evaluation of model with autoencoder and neural network.

Main contribution of the paper is the use of autoencoder with numerical data as it is mainly used with the image related problems. This research will also con rm the ability of autoencoder in fusion of time domain data. The model in this research also tries to nd the options for the LSTM models when dealing with time related data. Uses of Li-ion batteries in critical applications like aircraft, satellite, automobiles make this research very critical in order to make battery health management system more intelligent and accurate. It not only improves the overall system reliability but also prevents the mishaps during operations. This research also tries to reduce the dimension of the original signal in order to nd minimum dimension, which will be enough to give good prediction of RUL. The nal model will be evaluated on its R-square and Mean Squared Error. The robustness and accuracy of the model will be compared with the state of art LSTM models.

The remaining part of this paper is organized as follows. Section 2 gives a brief account on related works. Followed by Section 3, which discusses the used methodology through data selection, Data Preparation, Feature Selection, Data Mining, RUL Calculation, Data Normalization and Evaluation subsections. Followed by Section 4, which gives Design Speci cation of the models. Followed by Section 5, on details of Implementation. followed by section 6, Evaluation, Section 7, covers Conclusion and Future Work. Last section 8 is for acknowledgement.

2 Related Work

In literature one can see that as the importance of Li-ion batteries was realised, there were e orts to nd the methods to monitor the health of Li-ion batteries. This meant to improve the performance and reliability of batteries, in turn improving the reliability of critical systems with Li-ion batteries like Aircrafts and Satellites. As these applications needed prior and accurate RUL and SOC calculations for the batteries. There are three approaches for health monitoring of the batteries; 1) Model based, which include Empirical Model, Equivalent Circuit model and electrochemical models and Filter methods; and 2) Data driven models which includes di erent machine learning algorithm models Support Vector Machine (SVM), Neural Networks etc. 3) Hybrid Model which are a combination of model based and data driven models (see Figure 1).

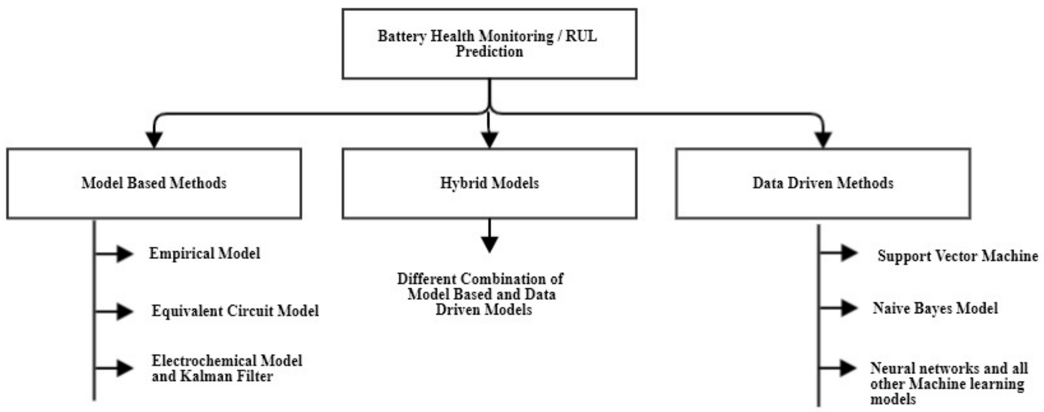

Figure 1: Di erent Methods of Battery Health Monitoring and RUL Prediction

2.1 Model Based Methods

Model based techniques create a battery model which is equivalent to real application battery and an estimation algorithm to predict the voltage or any other parameter. In this method a SOC is applied to the battery model in order to get the voltage same as the terminal voltage in real application. If one can get the exact voltage, then it can be said that SOC is of real battery. In this technique, the assumption is made that there is no noise in the voltage calculation, which is not possible in real applications. These further can be categorised in three di erent group which are explained one by one in next section.

2.1.1 Empirical Model

The basic empirical model represents voltage as a function of current and SOC. In order to get better accuracy, other parameters like ambient temperature can also be used. Samadani et al. (2015), uses Electrochemical Impedance Spectroscopy (EIS) for the creation of empirical battery model. Using this non-linear relation can be plotted up to a certain extent.

2.1.2 Equivalent Circuit Model

In equivalent circuit model battery model is created with a resistor and resistance-capacitance circuit. It is easy to set up and get quite good result with non-linear characteristics of the battery. Sangwan et al. (2016), use a modified equivalent circuit with two set of resistance{capacitance network to estimate SOC of the battery.

2.1.3 Electrochemical Model

Electrochemical Models tries to represent internal working of the battery that is transfer of charge from cathode to anode in term of mathematical equation usually it is represented by partial differential equation of high order, which makes it difficult to calculate SOC in real time but there are many papers which have given the methods to simplify these equation in order to estimate SOC of the Li-ion battery (Meng et al.; 2018).

2.1.4 Kalman Filter

In literature one can find many applications of Kalman Filter for RUL and SOC prediction of Li-ion battery, it's kind of extension to the electrochemical model. In general, Kalman filter are used to estimate unknown state of a system given the previous state data with some system noises, they are widely used in navigation type of application and for signal processing systems. They are popular because they are used to provide an estimation of unmeasurable quantity based on the nearest measurable quantities.

In batteries SOC cannot be calculated directly instead it can be estimated by other measurable parameters like internal impedance, voltage and current.Mo et al. (2016), have used a Kalman filter with particle swarm optimization to give improvement on the particle filter method of the RUL prediction. Kalman filter provide better results with the noisy data and particle swarm optimization improve the particle degeneracy problem of particle filter. The proposed model take account the internal impedance, aging rate of batteries and number of charging discharging cycle in order to estimate the capacity of Li-ion battery and uses 30% fade in rated capacity criteria as end of life criterion. For the evaluation of the model performance RMSE is used, while for accuracy comparison estimation error percentage method is used.

Yao et al. (2015), used an improved version of Kalman filter, Extended Kalman filter for SOC estimation of Li-ion batteries. They have used concept of equivalent circuit model to estimate the open circuit voltage. Based on this open circuit voltage SOC is estimated, this method provides better precision with the non-linear characteristics of the battery. To make the model more reliable under noise, Buss's adaptive rule is applied. Hu et al. (2012), also presented an extended Kalman filter model for the SOC and capacity estimation of Li-ion battery. They proposed a multi-scale framework for SOC and capacity estimation, which resulted in high efficiency and accuracy.

2.2 Hybrid Model

There are few models which uses a combination of data driven and model-based methods to predict SOC and RUL of Li-ion batteries. Majority of these models use Kalman Filter with another model. He et al. (2014), developed a neural network model for SOC prediction of the battery and usage Kalman Filter to reduce the error. Zheng and Fang (2015), also uses the unscented Kalman Filter with Relevance Vector Regression model to make short term RUL prediction. These hybrid models improved the performance of the standalone models.

2.3 Data Driven Model

As technology progressed, the computing power enhanced, and sensor technology improved which made possible the collection and storage of data and many applications of data driven methods can be seen in RUL and SOC prediction of Li-ion batteries.

2.3.1 Support Vector Machine

Support vector machine is a kernel-based machine learning technique which is predominantly used for classi cation problems, but it can be used for regression problems as well. Battery pro le is regressive in nature with linear and nonlinear part, so many papers used this concept to apply Support Vector Regressor (SVR) for the RUL and SOC prediction. SVR tries to map the non-linear feature in high dimension to treat it as linear. Patil et al. (2015), had used SVM for the RUL prediction. They have used two SVM model, rst classi es the di erent discharging cycle into four groups and the later one is used for RUL prediction. Classifying SVM helps to recognise the suitable discharge cycle for RUL prediction, based on this data for the SVR is prepared. Classifying SVM uses the radial basis function kernel while the regression one is tested on Gaussian, exponential, hyperbolic and multilayer perceptron kernel out of which perceptron kernel gave the best results. These models are compared with respect to their accuracy.

Li et al. (2017), also used SVM model for the RUL prediction of the Li-ion batteries. This model uses the relation between RUL and internal resistance, capacity, derivative voltage, terminal voltage relation for the prediction. Applied SVM model gave 95% accuracy and it is compared with the neural network with forty neurons. The results of the SVM are better than the neural network but this cannot be considered a fair comparison as used neural network is very novice and if comparison is to be made it should be with the state of art. Although SVM are performed well but they are highly sensitive to data and their performance with unseen data changes considerably, which cannot be considered as a robust model.

2.3.2 Naïve Bayes

Ng et al. (2014), presented a na•ve Bayes model for RUL prediction. Its results are compared with SVM model, Na•ve Bayes model gives competitive results and it is more robust to di erent kind of data unlike the SVM model. In this paper di erent set of data is made using Bayes probability theorem and they future used for the prediction. They have not used capacity as their target variable instead they used the number of cycles remaining as dependent variable. So, now this becomes regression problem which is predicted using Bayes approach. Result are almost comparable to SVM model with majority of the cases achieving same or better RMSE value than the SVM model.

2.3.3 Neural Networks

Neural networks have gained popularity in machine learning and data analytics domain because of its self-learning capabilities. In academia there have been many papers which present di erent kind of neural nets to predict SOH and RUL of the Li-ion batteries.

Chemali et al. (2018), presented a DNN model with three layers to estimate the SOC of the Li-ion battery. It gives quite competitive result. The focus of this paper is to develop a model which can be used in electrical vehicles, the model initially requires considerable time for training but once the model is trained, that is it has found the weights then it can be very fast in prediction of SOC and RUL with appropriate hardware and this can be used in online methods too. It has used the early stopping method to avoid the over tting. This research focused on one battery and take care of di erent ambient temperature that means model is robust to di erent temperature, other than this they have introduce noise into the data which make nal model robust to noise, which may be entered during to di erent operating conditions. This does not provide a generalize model which can be used with di erent batteries. Other than this, it shows that neural network can be a good choice for health monitoring of the batteries.

Zhang et al. (2018), used Long Short-Term Memory Recurrent Neural Network (LSTM RNN) for RUL prediction of batteries. In this they have generated data under di erent ambient temperature and as the goal is to plot degradation characteristics of battery, they have used di erent current rate for discharging. For 25° C ambient temperature discharging current rate is 1C and 2C while for 40° C current rate is 1 C and 3.5 C. Presented model have gated LSTM RNN architecture with root mean square prop backpropagation method and a Monte Carlo simulation unit to introduce noise to make model resilient to di erent type of noises. This model performed better than previous SVM models Particle Filter models. As data travels sequentially in LSTM RNN network so there is always a chances of vanishing gradients which is solved by gated architecture but still research have shown that attention-based model can outperform these networks.

Many research have used NASA battery dataset with di erent neural networks and achieved good results for the RUL and SOH of Li-ion batteries. This gives a boost to the decision of selecting NASA battery dataset for this research. Shen et al. (2020), uses Deep Convolutional Neural Network (DCNN) for health prognostics of Li-ion batteries. They used transfer learning method to ensure that even partial charging cycle data should be enough to make the prediction on unseen data. They have used data from multiple Li-ion batteries to build a DCNN model, then weight learned from this pretrained model is used for transfer learning, in this there are three di erent DCNN Transfer Learning models which are ensembled to make the nal prediction about the capacity of the battery. So, this ensemble model is trained on relatively smaller target data and based on this, the nal capacity prediction is made on NASA dataset. Finally, they compared the result from DCNN-ETL with random forest regression model, Gaussian process regression model and other DCNN models based on their RMSE values, results from DCNN-ETL models are better than others. This model emphasises on online prediction for Li-ion batteries, but it needs enormous diverse data for pre-training, which is quite di cult to get even after that the nal prediction may be very subjective.

Qu et al. (2019), presented a LSTM based model for SOH and RUL prediction, there approach is to identify shortcomings of previous LSTM models and come up with the solution for each and nally make a model in combination with these components which helps to improve the performance of LSTM model. In LSTM models sliding window approach is used to make the prediction so one can say that only a particular set of features are involved in prediction to further generalise this process attention mechanism is used which improved the LSTM performance, they also have used particle swarm optimization and to cater the noise in battery data Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) is used, which take care of noises introduce due to di erent operating environment. For this they have used NASA battery dataset and results are better than Simple Recurrent Neural Network (RNN), simple LSTM models and Relevance Vector Machine, evaluation are made based on RMSE.

Wang et al. (2018), also presented a LSTM model on the NASA dataset. This paper uses Spearman Correlation based Li-ion battery health indices to extract the features from the charging and discharging cycle. It uses a memory cell structure to address the problems of vanishing gradient and exploding gradient. Proposed Dynamic Long Short-Term Memory (DLSTM) model is evaluated based on the RMSE value.

From the literature one can say that DNN, CNN and LSTM are the common neural networks for the prediction of RUL of Li-ion batteries. However, these all methods have some limitations like DNN models are specific to a battery that limits its uses in real life applications. CNN requires large data and often suffers from the overfitting while LSTM model have critical window selection task which is very research specific. Other than this battery's nonzero relaxation time makes implementation of LSTM difficult due to capacity regeneration phenomenon. So, using Autoencoder one tackles the complex gate structure of LSTM, capacity regeneration problem in recurrent neural networks and time vector data fusion with other parameter of batteries (Qin et al.; 2016).PA-LSTM and DLSTM are the best performing model when compared with the other neural network based techniques (Qu et al.; 2019; Wang et al.; 2018).

3 Methodology

In presented research, Knowledge Discovery in Database (KDD) methodology is followed for prediction of RUL of Li-ion battery (Fayyad et al.; 1996). However, SEMMA and KDD are almost equivalent, but steps of KDD are more suitable for this research (Azevedo and Santos; 2008). Other data mining methodology is Cross Industry Standard Process for Data Mining (CRISP-DM), it's initial phase is focused on understanding the objective and requirement of the project from business point of view and in final stage full deployment in included but this project understood the objective from academic point of view while also knowing its business implication. This project does not provide an account on deployment of used models on some IT system in business setup, but it derives the knowledge which show it is useful way to predict RUL and it has applicability in real systems.

3.1 Data Selection

In order to predict RUL of any device or systems usually uses three kind of dataset

1. Dataset have reading from entire life span of the system.
2. Dataset have only failure reading.
3. Dataset have few readings and their threshold is already defined.

This research uses the NASA AMES Centre battery dataset (Goebel et al.; 2008). This dataset has four battery reading B005, B006, B007 and B0018. Each battery has charging, discharging and impedance profile of multiple cycles and given threshold condition is reduction in battery capacity by 30% of its rated capacity. Batteries in the dataset are 18650-size Li-ion cells that which are used in studies at Idaho National Laboratory. These tests were done under controlled ambient temperature. These files are in MATLAB format which are converted into json format for better readability and usability in python. The selection of this dataset is also motivated by the fact that this is one of the most used datasets for different deep learning network like LSTM, DNN for health monitoring of the Li-ion battery. That means it can be used for neural network although it is very small but other paper have achieved good result on the same data. Other than this this dataset comes from one of the renowned institution NASA and it is publicly available, so there is not much ethical issue regarding it.[3] The dataset is of batteries, there is no personal type of information, so it does not raise any question under GDPR. Description of Charging, Discharging and Impedance cycle of Li-ion battery in selected dataset.

Charging Cycle: Li-ion batteries are rechargeable so they are recharge by constant voltage or constant current source. In this dataset Charging of the batteries are done under constant current of 1.5A until the voltage reached to 4.2V (single battery cell's maximum voltage) and then it is continued under this voltage until current dropped to 20mA.

Discharging Cycle: The process of using stored energy in Li-ion battery is called discharging cycle. In this dataset discharging is done at constant current of 2A until battery voltages of B005 reached 2.7V, B006 reached 2.5V, B007 reached 2.2V and B0018 reached 2.5 V.

Impedance Cycle: Impedance measurements are taken by Electrochemical Impedance Spectroscopy (EIS) and selected frequency are from 0.1 Hz to 5kHz.

Table 1: Di erent Batteries and their Number of Cycle.

Batteries	Number of Cycles
B005	168
B006	168
B007	168
B0018	132

3.2 Data Preparation

Original dataset is in .mat format, in which data is stored in hierarchical format. For the easy usage and make it readable in python, it is converted into json le by enumerating data from MATLAB le to dictionary using loadmat from the SciPy library. For charging, discharging cycle separate json le is created which have cycle as key and inside each cycle various reading are captured regarding that particular cycle. Data is captured till the battery's capacity reached to the threshold condition.

3.3 Feature Selection

The dataset contains multiple charging and discharging cycle and each cycle have di erent datapoints. They cannot be used directly for the model creation, instead feature needs to be equal in each cycle. To address this, one option is to take some points randomly form each cycle, but it always has risk of losing the important data, so the approach needs to be backed up by the behaviour of the Li-ion battery. Feature extracted must retain the battery's behaviour so that a good prediction can be obtained from the data. Other

[3]https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/

than this, these extracted features requires to remain reliable for all operating condition and for other similar batteries, then only the batteries appropriate degradation can be mapped.

To ensure that all these conditions are met, Lu et al. (2014) described the concept of geometric metric to estimate the capacity of battery based on their voltage, current and temperature pro le. If the capacity estimation is correct one can say that the set of extracted features are accurately representing the charging and discharging of Li-ion battery over their lifetime. Ibid have used same NASA battery dataset to de ne Geometric features of the Li-ion battery. These features were successful in mapping the capacity degradation. The geometric feature were also used in other published work (Ren et al.; 2018). These features can be interpreted as time when voltage reached its maximum value, in our case it is 4.2V, time when current started to drop and time when max temperature reached under operating ambient conditions. These geometric features were able to depict the capacity degradation of Li-ion battery under various operating and aging condition. Thus, based on this concept data point is extracted from each charging and discharging cycle and collated to make the nal data for feature fusion in autoencoder.

These geometric features can be seen, tracking the time-based relationship between internal parameters of battery with each charging discharging cycle. In Figure 2(1), one can see that during charging the maximum voltage get delayed as the number of cycle increases. Similarly, for temperature as the battery get old it takes more and more time to reach same maximum temperature. During discharging similar trend is followed for voltage and current as the battery get old discharging process become quick. This is can be seen for all the parameter as shown in Figure 2.

All the feature were extracted based on below equations.
For Charging Cycle,

Batteries terminal voltage is according to equation (1):

$$(t_i, v_i), \quad s.t.\ v_i \quad 4.2V \quad i = 1, 2, 3, 4 \ldots n \tag{1}$$

In the above equation (t) is a time when the battery voltage reaches the maximum value for the rst time and (v) is the maximum voltage achieved by the battery during charging cycle, (i) is no of cycle up to (n), which represent sample size.

Batteries terminal current is according to equation (2):

$$(t_i, A_i), \quad s.t.\ A_i \quad 1.5Amp \quad i = 1, 2, 3, 4 \ldots n \tag{2}$$

In the above equation (t) represent the time when the current started to drop.(A) is the value of the current when it started to drop.(n) is total sample size.

$$(t_i, T_i) = t_i, T_i \quad at\ maxT_i \quad i = 1, 2, 3, 4 \ldots n \tag{3}$$

In the above equation, (t) is the time when the temperature reaches the maximum value. (T) is the maximum temperature of the battery during charging. max T is the maximum temperature. (n) is sample size.

Batteries current at load is according to the equation (4):

$$(t_i, A_i), \quad s.t.\ A_i \quad just\ before\ it\ drops \quad i = 1, 2, 3, 4 \ldots n \tag{4}$$

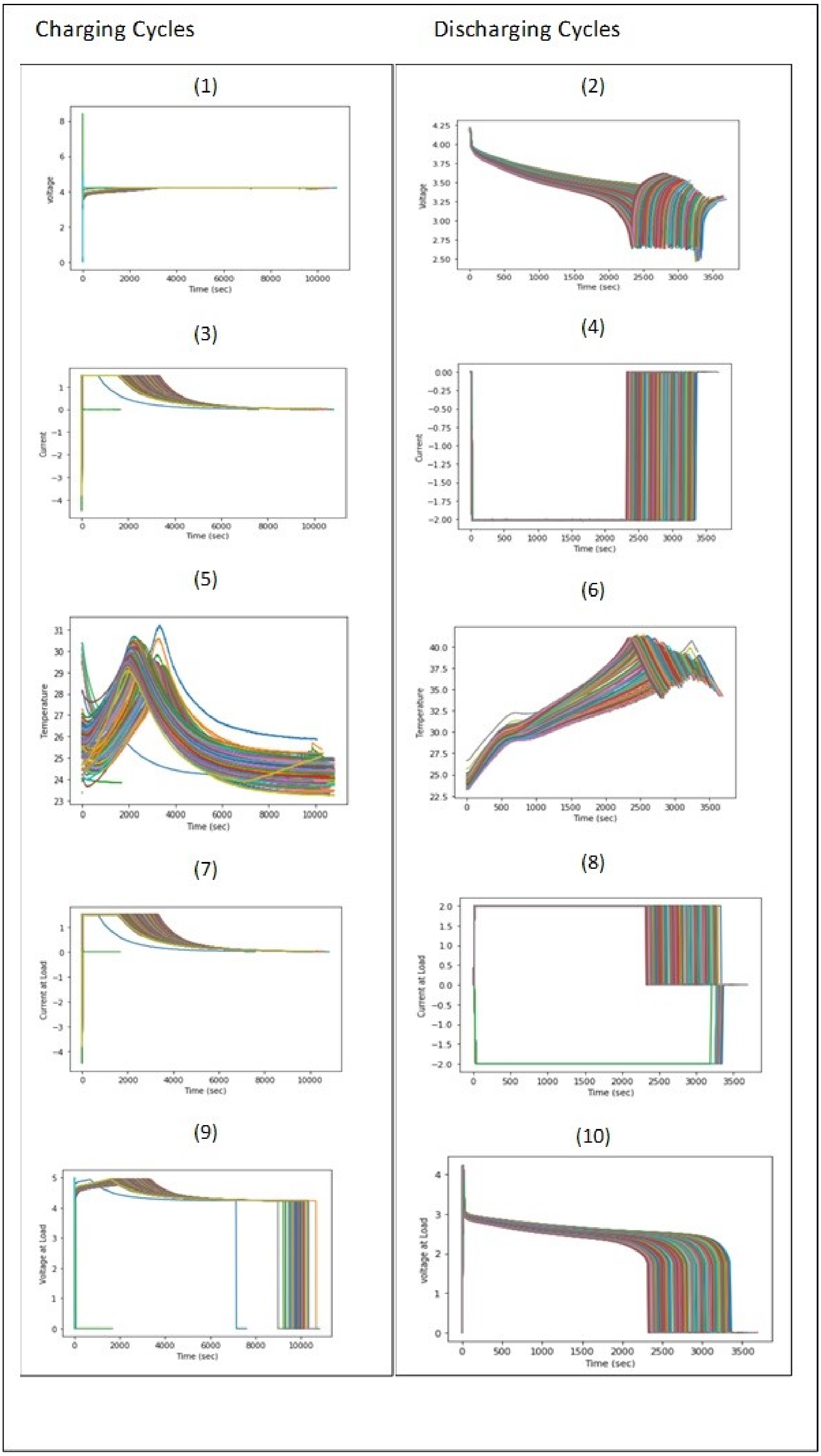

Figure 2: Changes in Di erent Battery Parameter over the di erent Cycles

Where, (t) is time just before current starts to drop. (A) is value of current in Amp when it started to drop. (n) is sample size.

$$(t_i, v_i) = (t_i, v_i) \quad at\ v_i max \quad i = 1, 2, 3, 4 \ldots n \tag{5}$$

Where, (t) is time at which voltage at load reaches maximum value. (v) is value of maximum voltage at the load. (n) is number of samples.

For Discharging cycles,

Batteries terminal voltage is according to equation (6):

$$(t_i, v_i) = (t_i, v_i) \quad at\ v_i min \quad i = 1, 2, 3, 4 \ldots n \tag{6}$$

Where,(t) is time when batteries voltage reaches its minimum value. (v) is minimum voltage of battery during discharging. (n) is sample size.

Batteries terminal current is according to equation (7):

$$(t_i, A_i), \quad s.t.\ A_i > \quad 2A \quad i = 1, 2, 3, 4 \ldots n \tag{7}$$

Where,(t) is time when the terminal current gradually start increasing. (A) is value of current when it start increasing. (n) is sample size.

Batteries Temperature is according to equation(8):

$$(t_i, T_i) = t_i, T_i \quad at\ maxT_i \quad i = 1, 2, 3, 4 \ldots n \tag{8}$$

Where, (t) is time at which temperature reaches its maximum value. (T) maximum temperature value achieved by battery during discharging. (n) is sample size.

Batteries current at load is according to equation (9):

$$(t_i, A_i), \quad s.t.\ A_i > \quad 2A \quad i = 1, 2, 3, 4 \ldots n \tag{9}$$

Where,(t) is time when the current at load gradually start increasing. (A) is value of current measured at load when it start increasing. (n) is sample size.

Batteries voltage at load during discharging is according to equation (10):

$$(t_i, v_i) = f(t_i, v_i) \quad at\ min(v_i)\ s.t.\ v_i \neq 0g \quad i = 1, 2, 3 \ldots n \tag{10}$$

(t) is time when voltage is minimum but not zero. (v) non zero minimum voltage value. (n) is sample size.

This give 20-dimensional dataset, corresponding capacity of batteries also added into this so the nal data is 21-dimensional.

3.4 Data Mining

This paper predicts the RUL of Li-ion battery in term of remaining cycle. Battery used are mostly having 168 cycle as their end of life cycle. So, it is a regression kind of problem where a number will be predicted for the RUL of battery. For this paper has used autoencoder which will rstly reduce the dimension along with the fusion of features. Data from the autoencoder is then used in another simple neural network for RUL prediction.

3.5 RUL Calculation for Training of the Prediction Model

Before passing the data to prediction model, calculation of RUL is required, this dataset have reading till the battery reaches to their end of life criteria. For example B005 has 168 cycles let's call it (n), let's say battery for prediction is currently in ith cycle, RUL can be calculated as

$$RUL_i = n + 1 - i \quad Where, (0 < i < n) \tag{11}$$

At this stage final data for prediction model is converted into supervised dataset with corresponding reading of RUL of the battery using equation (11).

3.6 Data Normalization

In process of model creation often feature have different ranges, which may introduce some error into the result. To avoid this paper uses data normalization technique, where all the data is normalized in the range of 0 to 1 using minimum maximum normalization technique (Galea and Capelo; 2018).

$$X_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{12}$$

It is widely used normalization techniques it ensures that all the data are in same scale on other hand it is less effective with the outlier but in present case outlier are not the concern.

3.7 Evaluation

As the problem in hand is regression type, so to evaluate the performance of the model R-square and Mean Square Error (MSE) are taken and based on this caparison will be done with other models. R-square indicate the variability explained by the features selected for the model. It ranges between 0 to 1. Otherwise this value is converted into percentage. MSE indicate the quality of the prediction. Smaller the MSE better the prediction.

4 Design Specification

Process flow of this paper is presented in figure 3. The entire model developing process divided into three parts viz. Feature Extraction, Feature Fusion and RUL prediction.

4.1 Feature Extraction

In this part battery data is converted into json data format. From json file based on the concept of geometric feature of the Li-ion batteries a set of 21 feature are extracted from the data. It has observations of voltage, current and temperature for each charging, discharging cycle with their corresponding time values. These time-based features are the reason for the use of LSTM and RNN model for SOC and RUL prediction, which are suffered by capacity regeneration and complex gate selection methodology. This is avoided in presented research by use of the autoencoder in the feature fusion section.

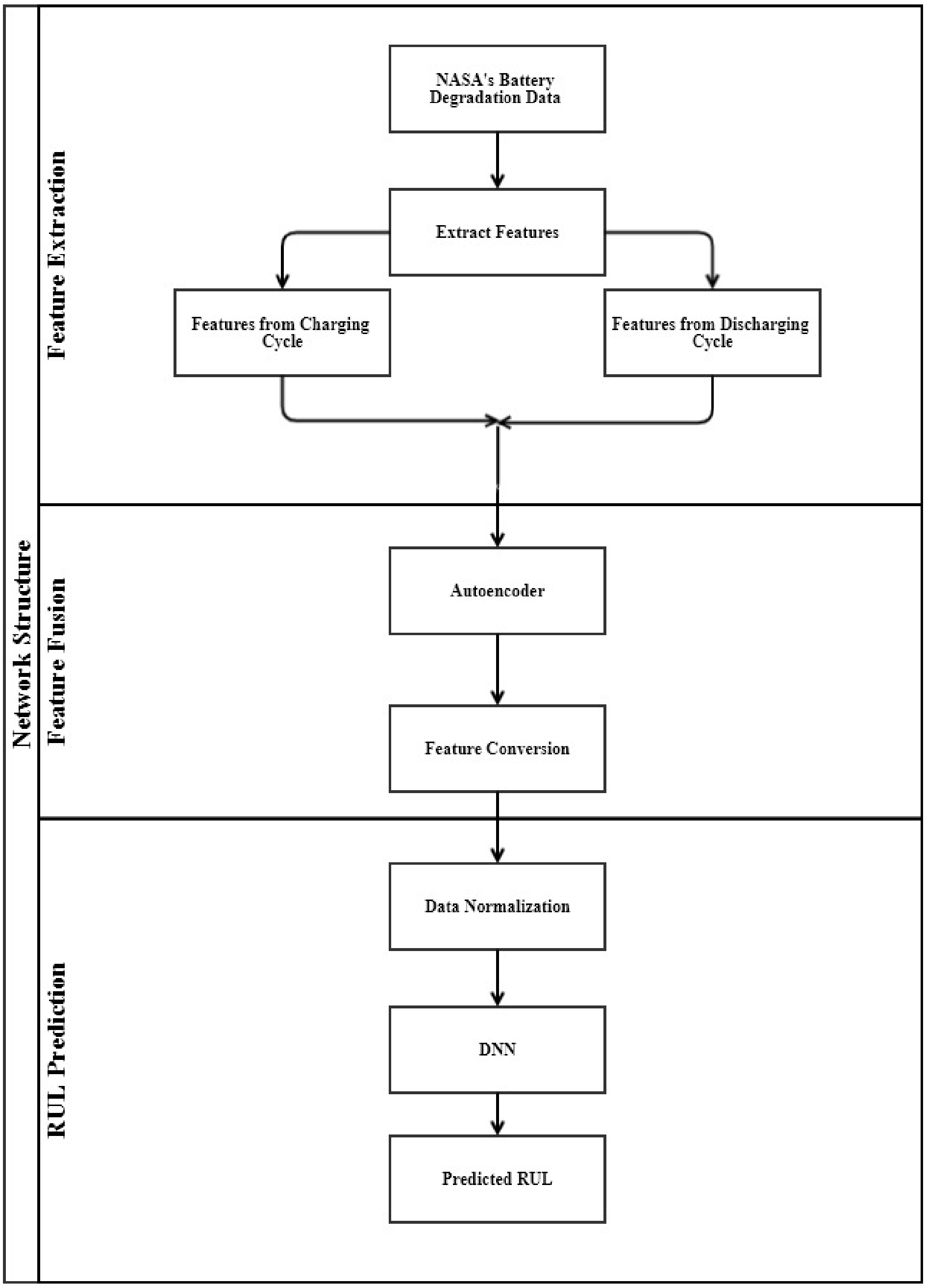

Figure 3: Process Flow for RUL Prediction

4.2 Feature Fusion

In this part paper uses Autoencoder which solves multiple complications like it reduces the dimension of data, it learns the time-based representation of the features, it does not get a ected by non-zero relaxation time of the Li-ion batteries. Autoencoder are unsupervised techniques which means they do not need labelled data to learn the feature with appropriate architecture they can learned the feature from the data itself (Goodfellow et al.; 2016). Autoencoder have symmetrical structure that is input, and output layer are similar. They are designed in such a way that their central layer is smaller than the encoder and decoder (as shown in Figure 4). When the output from the decoder is same to the input at encoder layer then the network have learned the weights and this central layer can be used for further analysis.

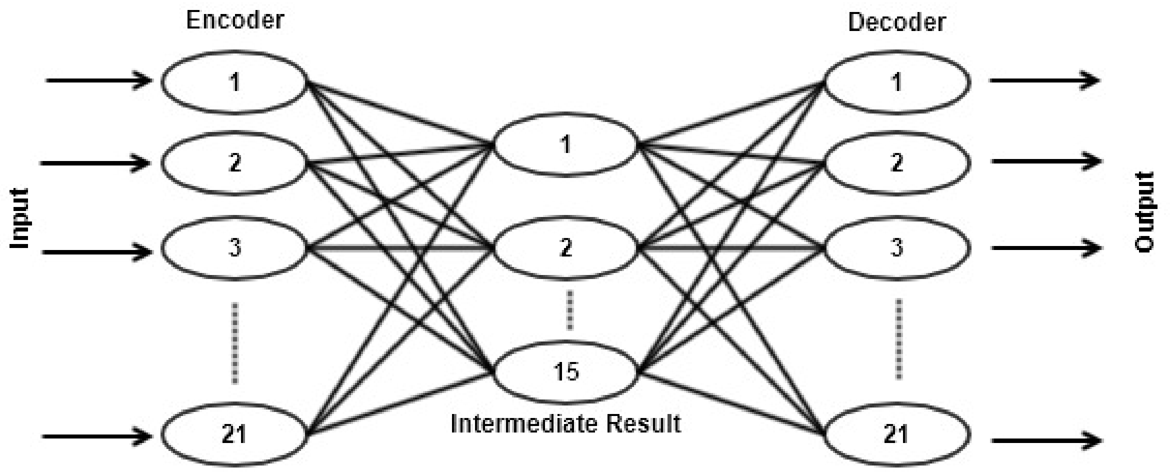

Figure 4: Autoencoder Architecture

Autoencoder used in this paper have (Recti ed Linear Unit) ReLU at encoder side and sigmoid at the output size as activation function. When autoencoder is used as dimensionality reduction it can be compared with the Principle Component Analysis (PCA) only di erence is activation function. PCA uses Gaussian distribution which extract linear features while in this case ReLU and Sigmoid extract linear as well as nonlinear characteristics. Performance of the autoencoder in dimensionality reduction with non-linear data is better than other commonly used methods (Goodfellow et al.; 2016). It t the purpose as batteries have nonlinear nature. To control the encoder and decoder layer Mean Squared Error is used as loss function and adadelta as optimizer. From the hidden layer 15-dimensional data is extracted when the networks encoder and decoder are very close. We do not want exact replica as that will mean that network has over t.

4.3 RUL Prediction

After getting 15-dimensional data from the Autoencoder which is fused with time characteristic of battery parameters. Data is converted from unsupervised to supervised by adding the RUL data. RUL is number of cycle battery going to last before it becomes dis-functional.For this a simple deep neural network is used which have three dense layer input layer has 512 neurons with ReLU as activation function, then next layer is dropout layer and output layer has one neuron with sigmoid activation function.This nal layer gives the RUL value for corresponding input features.

5 Implementation

This parts support document is mentioned in the design manual of the paper. For entire research coding is done in python using jupyter notebook and google colab.

5.1 Data Processing

First thing was to convert MATLAB file into json file so that it can be opened in Python. All the data related work is done in Jupyter Notebook.

After this feature were extracted based on the equations in section 3.3. From each cycle data was stored in different list, each equation provided two set of features, in total 20 feature were collated from the data and converted into csv file one for each battery. This step was tedious as generalizing each cycle required lots of testing with different cycle. This was done based on trial and errors and finally condition was derived for the loop coding during the feature extraction process. Last feature is derived by multiplying maximum terminal voltage and current. This represents corresponding capacity of the battery. So, paper uses these 21-dimensional data for the model creation.

5.2 Model Training and Testing

This step is performed using google colab which help in quick training and testing of model. Autoencoder and neural network is created based on architecture discussed in section 4.2. Files from all the battery joined to make a single file. At this stage data is split into training and testing set. The splitting ratio is selected as 90% for training and 10% for testing. High percentage of data was allotted to training as the used dataset is very small and target was to train the model on maximum data. Then Autoencoder is trained on the data of batch size of 12 for 1000 epoch when the loss hit the approximate constant state then training is stopped, and 15-dimensional feature are extracted from the hidden layer of autoencoder.

Figure 5 shows the model loss of autoencoder, validation loss closely follows the training loss and becomes approximately constant at the later stage of the epochs. This means data is not over fitted or under fitted.

After this prediction phase starts; before starting the prediction, data is normalized as discussed in section 3.6. RUL feature is added as target variable for the training of Neural Network as explained in section 3.5. Neural Network architecture is created as explained in section 4.3. This neural network uses 'Mean Squared Error' as loss function and 'adam' as optimizer, batch size is 16 and trained for 500 epoch. Loss function of the model becomes constant at the later stage of the operation (see Figure 6).

From this trained model final prediction is done on the test data. R-square and Mean Squared Error is calculated for the finally predicted RUL value.

6 Evaluation

As the RUL prediction is a regression problem so the evaluation can be done on the basis of the Mean Square Error (MSE) and R- Square values. MSE represent the deviation between predicted value and actual value, smaller the test MSE better is the prediction (James et al.; 2013a). R-Square explains the proportion of variation within predicted value (Y) explained by the independent set of features (James et al.; 2013b). It always lies

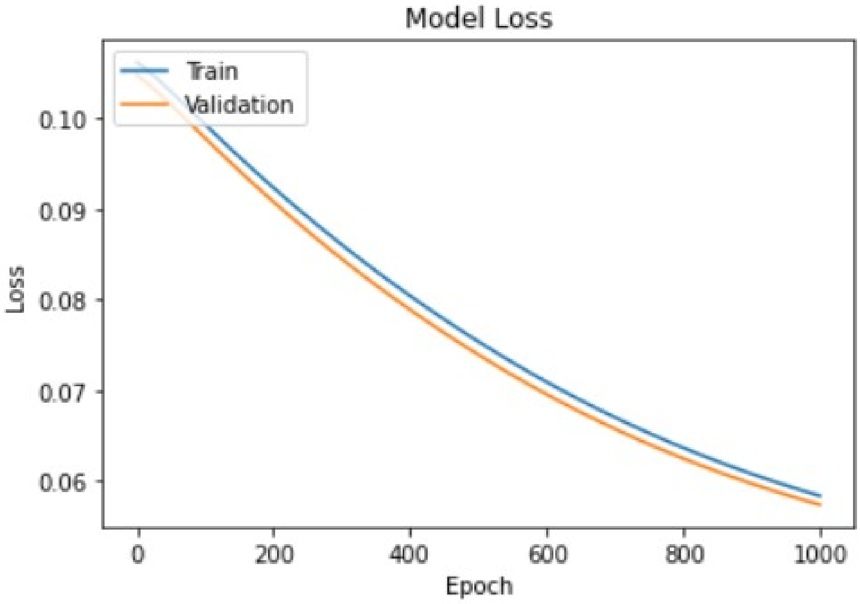

Figure 5: Training Model Loss of Autoencoder

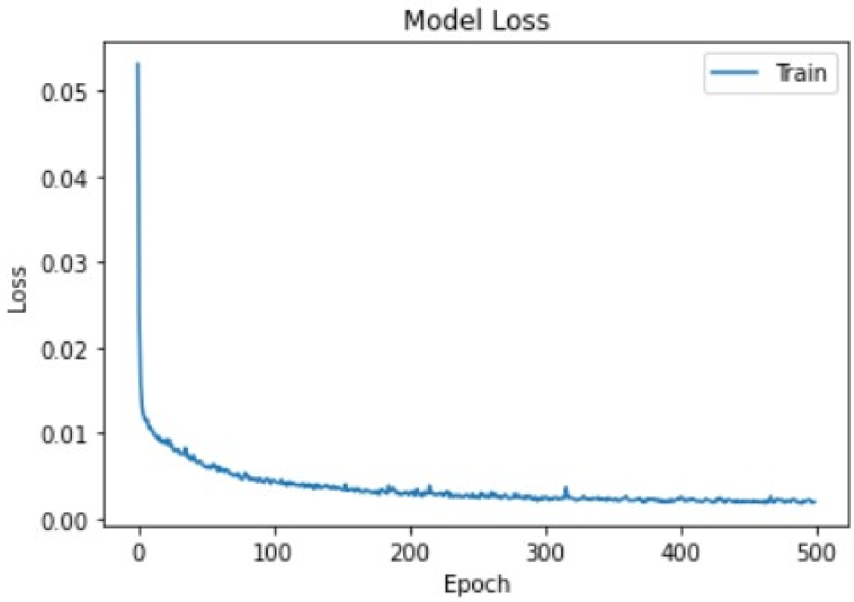

Figure 6: Prediction Model Loss

between 0 to 1, if its value is near to 1, that means maximum proportion of variability in predicted value is explained by the used model. In this paper used model gave maximum R-squared 95.7% and MSE is 0.0038 as shown in Table 2.

R-square and MSE is calculated for di erent set of test data and minimum R-squared 87.05% and MSE is 0.0105. So, based on this one can say it is quite robust to di erent set of testing data.

These results were also compared with the previous models on the same dataset.

Table 2: Evaluation of Prediction Model

Evaluation Parameter	Minimum Value	Maximum Value
R-square	87.05%	95.70%
MSE	0.0105	0.003

6.1 Linear Regression Model

After extracting 15-dimensional data from the Autoencoder paper tried to use a simple regression model for the RUL prediction it achieved competitive R-Square of around 90%, judging from the point of complexity and interpretability of the model it is quite good result.

6.2 Discussion

For RUL prediction earlier best-performing model was improved LSTM since the batteries charging, discharging cycles have time dependent features, so the choice of LSTM was obvious in order to retain the time data. This paper had argued how autoencoder can be used with the time dependent features and the results shows it has been successful up to a certain level since the results from the proposed model is comparable with the best performing models of PA-LSTM, DLSTM and better than the ANN model ((Qu et al.; 2019; Wang et al.; 2018). From Table 3 one can say that presented model is almost as good as the model with the best MSE. So, the autoencoder successfully learned the time feature and compressed them into smaller dimension. As the dataset was small, so there is always chances that autoencoder do not converge, that's why data was used in very small batches in training phase. Other than this used dataset has only one ambient temperature, so the nal model will not be robust against di erent working temperature. This issue can easily be resolved by using data from di erent ambient temperature.

Table 3: Comparison Between State of Art RUL Prediction Deep Learning Model and Autoencoder-DNN Model

Model	Mean Square Error
Autoencoder+DNN	0.003
PA-LSTM	0.003
DLSTM	0.002
ANN	0.009

This research was able to meet all objectives set during the starting phase of the research. Literature review section gives a brief historical evolution of RUL and SOC calculation in batteries. Then in later part it focuses on current methods of deep neural networks like ANN, LSTM etc. The big chunk of time was devoted to the data point selection which is discussed in 3.3 great detail with the help of various mathematical equations.

7 Conclusion and Future Work

This research was started with the aim of making an accurate prediction for RUL of Li-ion batteries, which will make the battery health monitoring more reliable and intelligent.

This will help the electrical automobile industry, renewable energy plants etc. to maintain environmental sustainability and progress towards the clean energy goal. Almost all the objectives from this research are ful lled. This research proves that it is possible to recognise, monitor and analyse the voltage, current, temperature and all other geometric parameters. In this research, the geometric parameter of Li-ion batteries was used successfully to predict RUL with the time-based feature fusion using autoencoder. It shows that autoencoder can be an e ective tool to deal with numerical data or regression problems.

The used method has the comparative result with the state of art DLSTM and PA-LSTM models. It con rms the e ectiveness of autoencoder in extracting non-linear characteristics. The extracted feature gave quite good results even with simple regression prediction model which again validate the quality of the extracted feature.

Dataset used, has only one ambient temperature, so to make model more robust it needs to be trained on dataset of other ambient temperatures. This research obtained the prediction only for the single cell of battery, while in real application usually a set of cells are used. Hence, there is a room to further extend this methodology to predict the RUL of entire battery pack.

8 Acknowledgement

I would like to thank my supervisor Dr. Rashmi Gupta for her invaluable suggestions, support and feedback throughout the research. I would also thank the researcher at NASA AMES for publishing data publicly. Lastly, I would like to thank my family and friends for keeping me motivated and focused.

References

Azevedo, A. I. R. L. and Santos, M. F. (2008). Kdd, semma and crisp-dm: a parallel overview, *IADS-DM* .

Chemali, E., Kollmeyer, P. and Preindl, M. (2018). State-of-charge estimation of li-ion batteries using deep neural networks: A machine learning approach, *Journal of Power Sources* **400**: 242{255.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. et al. (1996). Knowledge discovery and data mining: Towards a unifying framework., *KDD*, Vol. 96, pp. 82{88.

Galea, A. and Capelo, L. (2018). *Applied Deep Learning with Python: Use scikit-learn, TensorFlow, and Keras to create intelligent systems and machine learning solutions*, Packt Publishing.
URL: *https://books.google.ie/books?id=dPFsDwAAQBAJ*

Goebel, K., Saha, B., Saxena, A., Celaya, J. R. and Christophersen, J. P. (2008). Prognostics in battery health management, *IEEE Instrumentation Measurement Magazine* **11**(4): 33{40.

Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*, MIT Press. `http://www.deeplearningbook.org`.

He, W., Williard, N., Chen, C. and Pecht, M. (2014). State of charge estimation for li-ion batteries using neural network modeling and unscented kalman lter-based error cancellation, *International Journal of Electrical Power and Energy Systems* **62**: 783{791.

Hu, C., Youn, B. D. and Chung, J. (2012). A multiscale framework with extended kalman lter for lithium-ion battery soc and capacity estimation, *Applied Energy* **92**: 694{704.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013a). *An Introduction to Statistical Learning: with Applications in R*, Springer.
URL: *https://faculty.marshall.usc.edu/gareth-james/ISL/*

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013b). *An Introduction to Statistical Learning: with Applications in R*, Springer.
URL: *https://faculty.marshall.usc.edu/gareth-james/ISL/*

Li, X., Shu, X., Shen, J., Xiao, R., Yan, W. and Chen, Z. (2017). An on-board remaining useful life estimation algorithm for lithium-ion batteries of electric vehicles, *Energies* **10**(5): 691.

Lu, C., Tao, L. and Fan, H. (2014). Li-ion battery capacity estimation: A geometrical approach, *Journal of power sources* **261**: 141{147.

Meng, J., Luo, G., Ricco, M., Swierczynski, M., Stroe, D.-I. and Teodorescu, R. (2018). Overview of lithium-ion battery modeling methods for state-of-charge estimation in electrical vehicles, *Applied Sciences* **8**(5).
URL: *https://www.mdpi.com/2076-3417/8/5/659*

Mo, B., Yu, J., Tang, D., Liu, H. and Yu, J. (2016). A remaining useful life prediction approach for lithium-ion batteries using kalman lter and an improved particle lter, *2016 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 1{5.

Ng, S. S., Xing, Y. and Tsui, K. L. (2014). A naive bayes model for robust remaining useful life prediction of lithium-ion battery, *Applied Energy* **118**: 114{123.

Patil, M. A., Tagade, P., Hariharan, K. S., Kolake, S. M., Song, T., Yeo, T. and Doo, S. (2015). A novel multistage support vector machine based approach for li ion battery remaining useful life estimation, *Applied energy* **159**: 285{297.

Qin, T., Zeng, S., Guo, J. and Skaf, Z. (2016). A rest time-based prognostic framework for state of health estimation of lithium-ion batteries with regeneration phenomena, *Energies* **9**(11): 896.

Qu, J., Liu, F., Ma, Y. and Fan, J. (2019). A neural-network-based method for rul prediction and soh monitoring of lithium-ion battery, *IEEE Access* **7**: 87178{87191.

Ren, L., Zhao, L., Hong, S., Zhao, S., Wang, H. and Zhang, L. (2018). Remaining useful life prediction for lithium-ion battery: A deep learning approach, *IEEE Access* **6**: 50587{50598.

Samadani, E., Farhad, S., Scott, W., Mastali, M., Gimenez, L., Fowler, M. and Fraser, R. (2015). Empirical modeling of lithium-ion batteries based on electrochemical impedance spectroscopy tests, *Electrochimica Acta* **160**.

Sangwan, V., Sharma, A., Kumar, R. and Rathore, A. K. (2016). Equivalent circuit model parameters estimation of li-ion battery: C-rate, soc and temperature e ects, *2016 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*, pp. 1{6.

Shen, S., Sadoughi, M., Li, M., Wang, Z. and Hu, C. (2020). Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries, *Applied Energy* **260**: 114296.

Vutetakis, D. G. and Viswanathan, V. V. (1995). Determining the state-of-health of maintenance-free aircraft batteries, *Proceedings of the Tenth Annual Battery Conference on Applications and Advances*, pp. 13{18.

Wang, C., Lu, N., Wang, S., Cheng, Y. and Jiang, B. (2018). Dynamic long short-term memory neural-network-based indirect remaining-useful-life prognosis for satellite lithium-ion battery, *Applied Sciences* **8**(11): 2078.

Yao, L. W., Aziz, J. A., Idris, N. R. N. and Alsofyani, I. M. (2015). Online battery modeling for state-of-charge estimation using extended kalman lter with busse's adaptive rule, *IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society*, pp. 004742{004747.

Zhang, Y., Xiong, R., He, H. and Pecht, M. G. (2018). Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries, *IEEE Transactions on Vehicular Technology* **67**(7): 5695{5705.

Zheng, X. and Fang, H. (2015). An integrated unscented kalman lter and relevance vector regression approach for lithium-ion battery remaining useful life and short-term capacity prediction, *Reliability Engineering and System Safety* **144**.