International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

A Comparative Study of Random Forest and LSTM Models for Battery Remaining Useful Life Prediction

Monias Tapiwanashe Munhamoh [1], Peter Bukelani Musiiwa [2]

[1] Student, Electronic Engineering Department, Harare Institute of Technology
[2] Lecturer, Electronic Engineering Department, Harare Institute of Technology

Abstract

In critical areas requiring reliable power such as electric vehicles, renewable energy storage systems, aerospace and aviation, medical devices and substation DC systems Lithium-ion batteries find many applications. These critical and sophisticated systems require reliable operation and higher safety considerations thereby preventing unwarranted failures, which may be catastrophic. The uninterrupted power requirement is very critical for the safety operation of these systems. Real time monitoring of the critical battery parameters such as capacity, voltage, current and temperature becomes very important. This is crucial for predictive maintenance resulting in planning for the necessary routine maintenance and replacement at the end of life of a system. Failures are thus detected before total system collapse.

Routine inspections and checking of critical parameters done in most cases requires a lot of human intervention and the regular maintenance does not suit the unexpected failures which in most cases occur suddenly. On the other hand, Machine Learning models offer predictive maintenance techniques according to the model built from the model features.

Machine Learning based techniques such as Decision Tree Regression, Random Forest, Support Vector Regression, Gaussian Process Regression and Long Term Short Memory are used to predict the Remaining useful life (RUL) of Lithium-ion batteries. This paper looks at two machine-learning models used to predict the remaining battery useful life. The Random Forest (RF), representing ensemble methods class of machine learning and the Long Term Short Memory (LSTM) representing the deep learning/sequence models class are discussed. The selected models chosen on the basis that they are a good representative of the respective class of Machine – Learning models.

The methodology used in this study include downloading and loading in MATLAB the publicly available online NASA data set. Preparation of the data for modelling is done through exploratory Data Analysis in MATLAB. The model features such as battery capacity, voltage, current and temperature are considered in this study. These parameters chosen on the basis of their great influence in the determination of battery remaining useful life.

The two Machine –learning models are implemented in MATLAB. The performance parameters Root Mean Square Error (RMSE) and the Statistical Correlation Coefficient $\boldsymbol{R^2}$ are obtained to find the Model performance in predicting RUL.

The simulated results in this paper proved that Random Forest is a better model than the LSTM when used with NASA data set for RUL prediction. The LSTM is more complex and slower to train, although the accuracy of the model increases with continuous training. This conclusion is based on the comparison of

simulation results of RMSE and $\boldsymbol{R^2}$obtained. It is noted that with continuous training the performance parameters of the LSTM model do improve greatly. This may imply that the model can be a better RUL predictor.
The study provides simulation techniques in the use of Machine-Learning Models in predictive maintenance. This is necessary to avoid unwarranted system failures, minimising maintenance costs and reducing plant/system downtime. Similar simulations and analysis can be done for all systems where predictive maintenance is required. Such system which require very high reliability and system security. The availability of accurate measurable data (affecting system deterioration) is crucial for such simulations if the results are to be generalised. More studies are therefore required for other such systems.

Keywords: Regression, Random forest, Long-term short memory

1. Introduction

Lithium-ion batteries find great use in power sensitive applications such as medical devices, electric vehicles, aerospace and aviation, industrial robotics and control, renewable energy storage systems and military and defence tactical equipment. In these critical systems the safety, reliability and accuracy of the system is required [1]. Safety concerns of the battery leading to overheating and explosions resulting in fire risk are at the core of the need for real time monitoring [2].
The lithium-ion battery ranks among the best power source because of the lighter weight, huge voltage density, and longer charging/discharging life as well as less self-discharge [3]. These technical parameters however tend to degrade with continuous use of the battery. There is therefore need to predict the remaining useful life of the battery to ensure safety and reliability of the system.
Traditional physical methods for predicting RUL have shortcomings in capturing the non-linear degradation pattern of the Lithium- ion battery [4]. Model based methods have proved to be better than physics based methods but they have complex calculations. These methods require large amount of data for accurate predictions. They are therefore not suited for most practical situations and real time determination of battery RUL [5].
Machine learning techniques have proved to use external battery parameters [6]. The change in the battery technical parameters such as capacity and internal resistance among other parameters is at the core of Machine Learning algorithms to analyse the degradation pattern and predict battery RUL [4]. Accurate prediction of RUL enhances predictive maintenance thereby saving costs by timely predicting equipment failures.
The Random Forest (RF), Long-Short Term Memory (LSTM) and The Support Vector Regression (SVR) models are the Machine- Learning models used for predicting battery RUL in this paper. Each of the chosen model is representative of a class of Machine –Learning models. Machine learning-based approaches provide better results than statistical approaches in terms of accuracy [7]. However, more computational power is required. A description of the two Machine Learning techniques is given. The model performance Root Mean Square Error (RMSE) and R^2 examines the three methods' performance in predicting RUL. The statistical parameters help to provide a comparative analysis of RF and LSTM techniques in battery RUL prediction.
The challenges encountered in the use of the models include the nonlinear degradation tendency of the charge/discharge process of the batteries [8]. The degradation patterns influenced by factors such as,

temperature, aging, and charge/discharge cycles of the cell [9]. The operating conditions also change with varying battery applications presenting a challenge of model generalization.

2. Objective of the Study

This paper aims to provide a Machine Learning-Based framework for battery Remaining Useful Life Prediction. The paper compares the RF and LSTM models in predicting battery RUL.

3. Model Description

3.1 Random Forest (RF)

The RF technique is an ensemble method. This implies that, the model does not consider a single tree but rather considers a number of trees forming a forest. The weaknesses of a given tree are “covered” up by other trees in the forest. This gives great variability in the forest to improve the model prediction accuracy [10]. The “forest” structure of the RF model enables interpretability of model input features such as voltage, capacity fade, current and temperature. The RF technique is a good selector of these input features required for modelling [11]. The RF technique allows for variability in the data accommodating noisy measurements and outliers in the data set without greatly affecting model performance [12].

3.2 Long Short-Term Memory (LSTM)

The LSTM model captures long-term tendencies in the data that repeats its characteristics over a given time interval extending the ordinary recurrent neural network (RNN) framework to capture long-term dependencies in the sequential data [13]. The enhancement is achieved by redesigning the RNN’s hidden layer to maintain information over expanded sequences. Training an LSTM model requires a large data set. The accuracy of prediction of the model however increases with increased training [14].

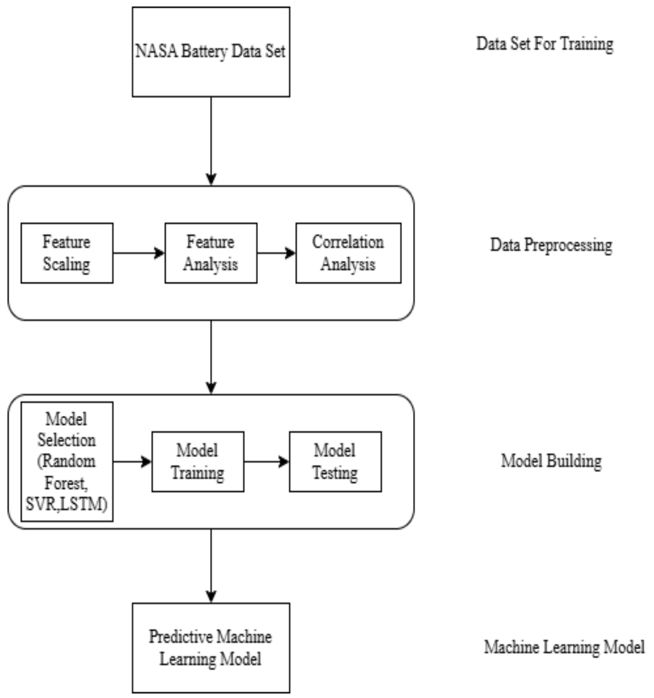
4. RUL Prediction Framework

The roadmap for developing a machine learning-based framework using Li-ion battery to predict battery RUL using MATLAB simulation based approach is illustrated below.

International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Figure 1: Framework for predicting battery RUL



[13]

4.1 Data Acquisition

In this paper, the data set is the publicly available online NASA battery data set. The data acquisition involves downloading the data and loading the data in MATLAB folder. The features under consideration being the battery capacity, voltage, current and temperature. These are the quantities that affect battery degradation pattern as the battery undergoes charging and discharging cycles. The data set shows how the battery Ampere –hour (AH) capacity degrades over charging and discharging cycles. The data for 1 cycle has variables temperature, time, voltage and current. These parameters affect the battery degradation profile over each charging and discharging cycle.

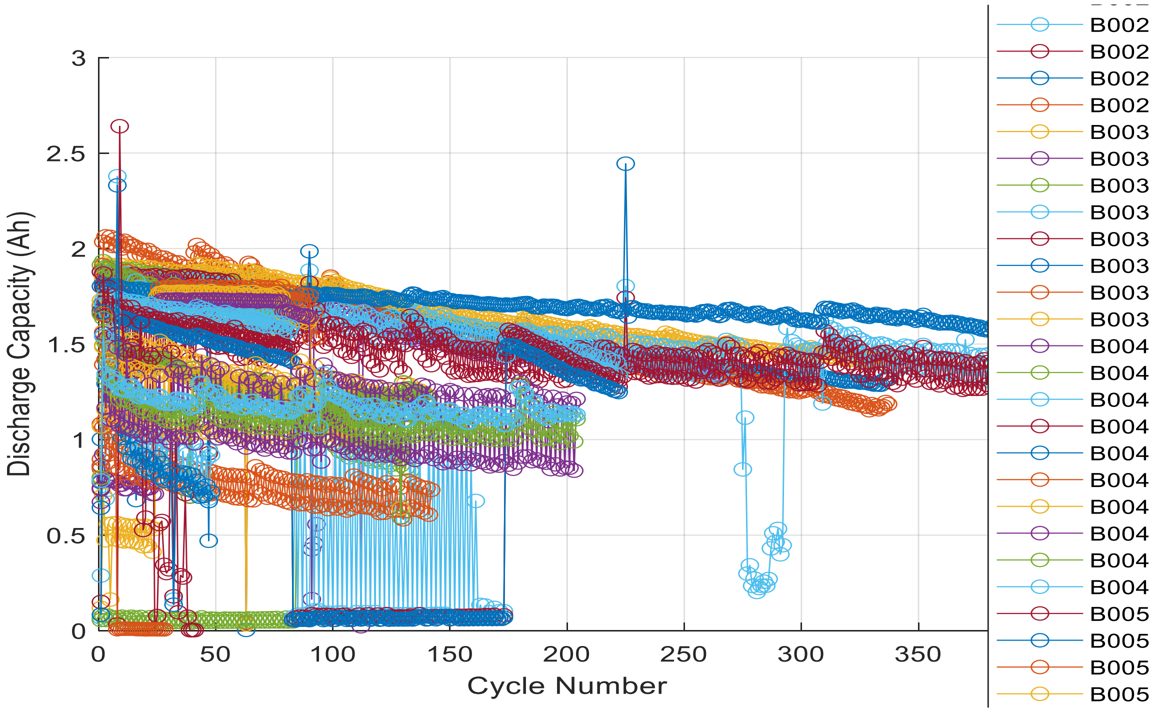
4.2 Data Pre-processing

Feature scaling to standardize the input temperature, current, voltage over time marks the beginning of data pre-processing. Since these variables vary greatly in time space, standardising thus helps to handle the features [14]. The analysis of the features is obtained by simulation in MATLAB to obtain the various feature plots. The plots also show the relationship between these variables over battery cycles. Exploratory data analysis (EDA) is also carried out as part of data pre-processing with battery capacity is plotted against the cycle number. The cycle at which 80% of the original battery capacity is used is known as the End of Life (EOL) of the respective cell [11].

4.3 Feature Analysis

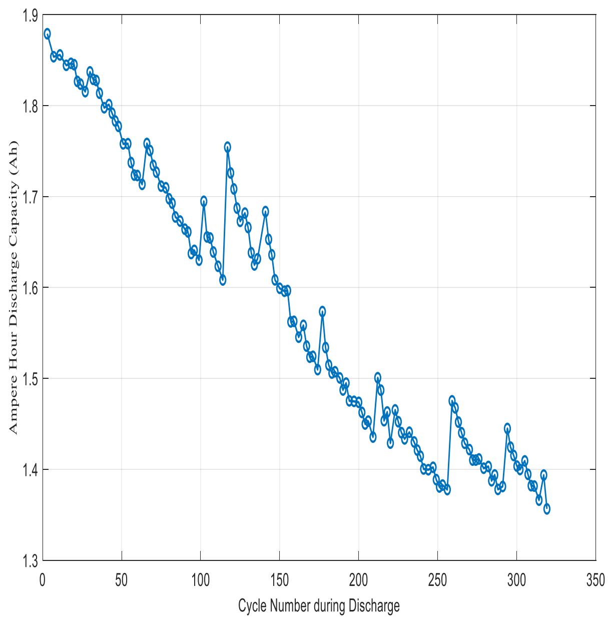
The graphical approach is used to show the variation of the different features as the battery undergoes charging and discharging cycles.

Figure 2: Discharge Capacity of Cells



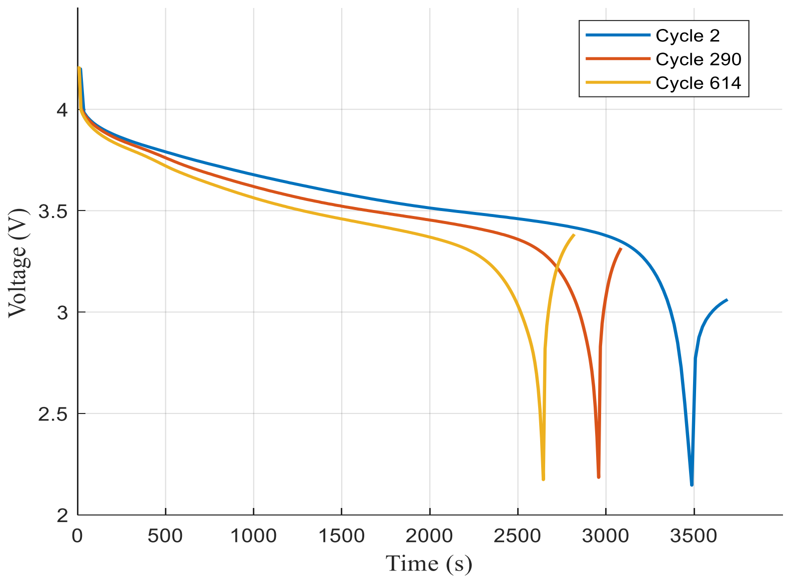
The figure shows the variation of the different cell capacities for the indicated battery cells. There is a general decrease in capacity over the number of cycles.

Figure 3: Discharge capacity for a cell



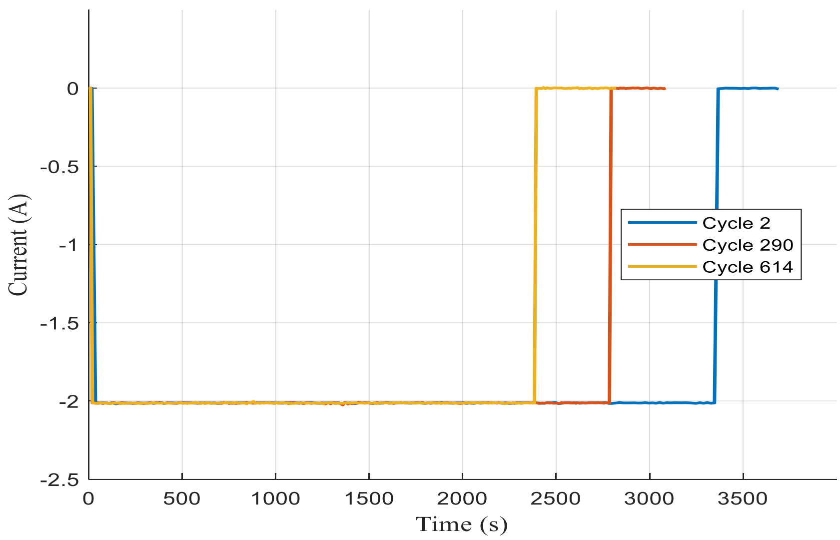
The graph shows that the cell capacity generally decreases over the number of discharge cycles. Discharging starts at approximately 1.87Ah capacity and degrades to about 1.35Ah. The trend of discharge applies to all the battery cells in the data set. The decrease in capacity is as a result of the chemical action taking place between electrolyte and electrodes [9]. This generally weakens the electrolyte and decreases the capacity of the cells.

Figure 4: Voltage vs Time for Cell B0007



Form a value of around 4.2v the voltage falls to around 3.3 v and falling rapidly to a cut-off voltage of 2.2v. The cell then begins to charge. The starting voltage is around 4.19V falling to a cut off value of around 2.3V. Thereafter the cell charging commences. With increase in the number of cycles the cell capacity decreases. Thus the battery capacity decrease with continuous usage.

Figure 5: Current vs Time for cell B0005

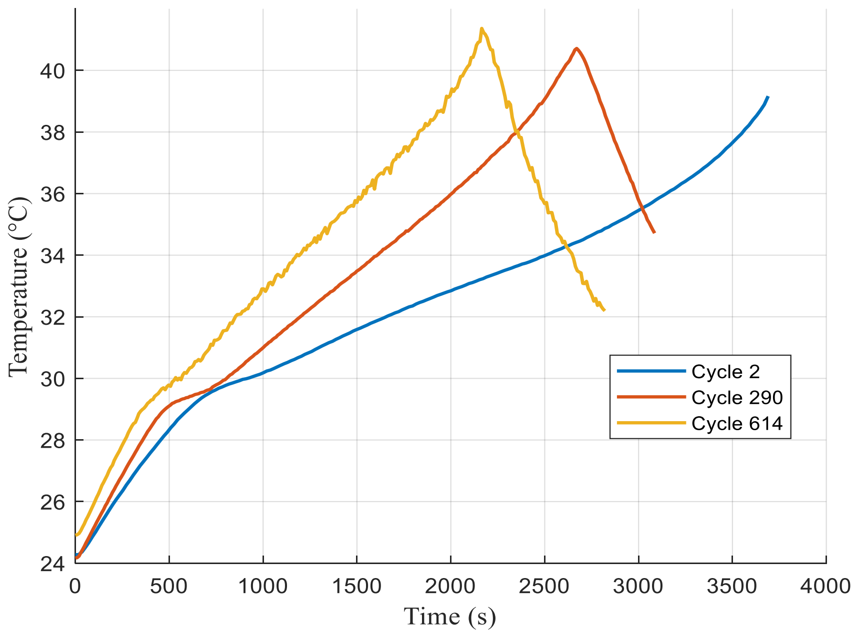


The constant current curves depict a decrease in battery capacity as the charging cycles increase.

International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

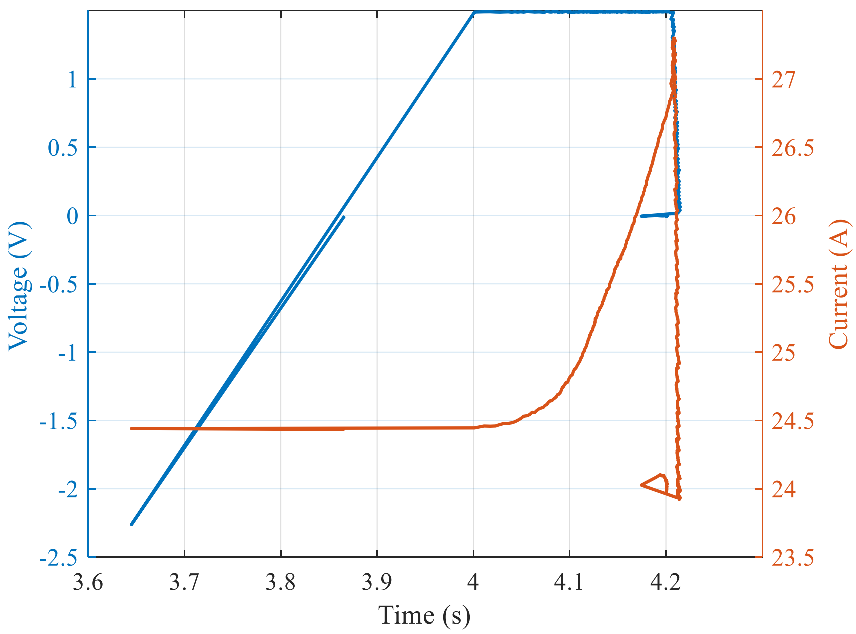
Figure 6: Temperature vs time for cell B0006



The chemical activity within the cell as charges move between electrodes and electrolyte results in an insulating layer on the electrodes. This increases temperature from around from around 24℃ to around 38℃ during the discharge process.

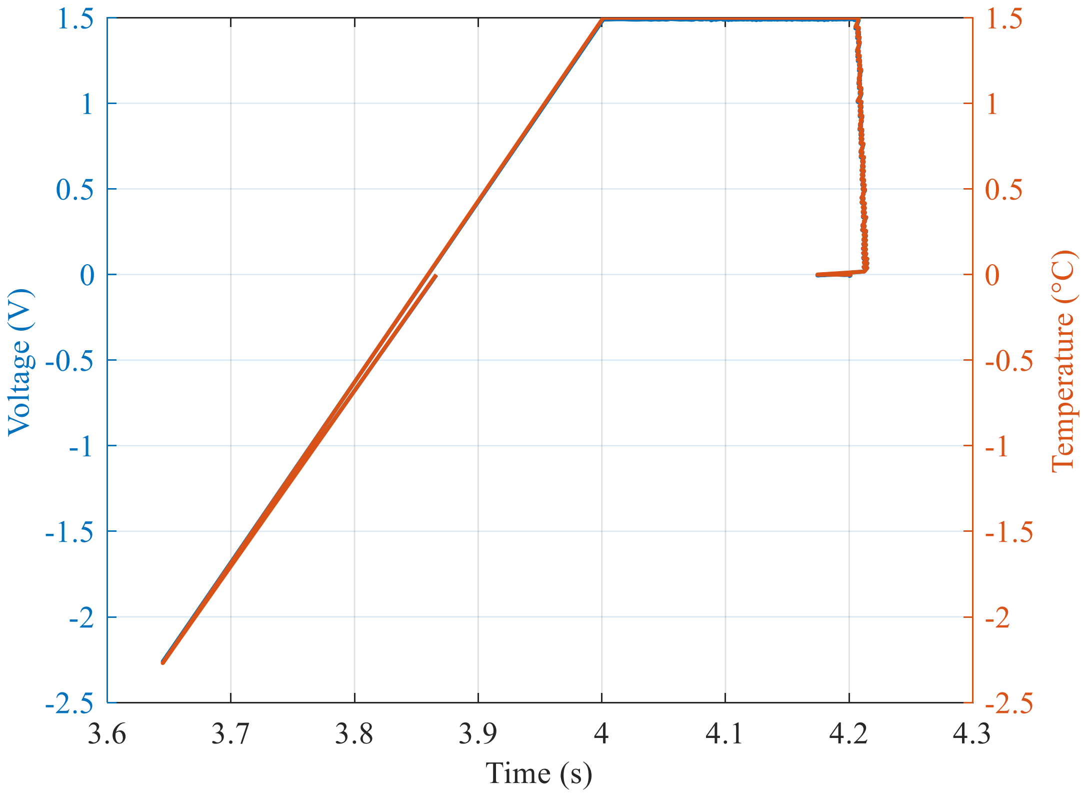
4.5 Correlation Analysis between Features

Figure 7: Voltage and Current vs time plot



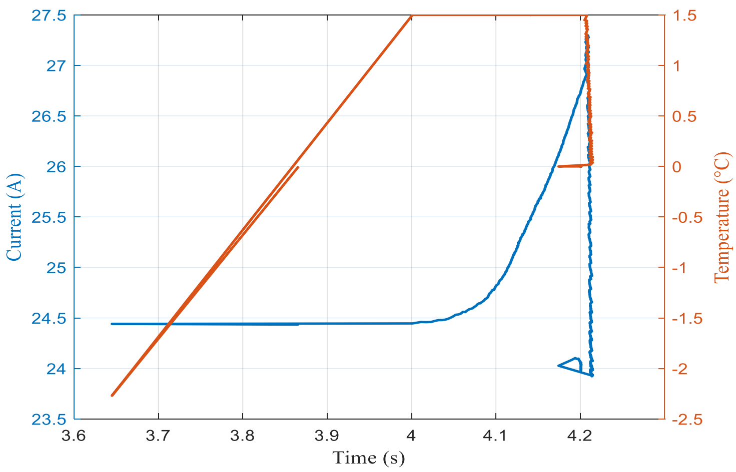
The voltage steadily rises when a constant current is impressed on a cell. This is a CC–CV charge/discharge characteristic.

Figure 8: Voltage and Temperature vs time



The graph is a linear ramp of both voltage and current reaching a plateau and then dropping sharply during discharge. It should be noted that in practice the zero lag between voltage and temperature is not real. In this case the code ordered temperature and voltage as equal quantities.

Figure 9: Current/Temperature versus Time



The constant current results in heat loss within the cell. As a result temperature rises steadily reaching a new steady state even if the current rises. As the charger switches polarity to discharge, the current reverses direction and temperature drops cooling the cell.

4.6 Cell End of Life (EOL)

The EOL of the cell is taken as the cycle at which the battery capacity degrades to 80% of its original capacity. The figure below shows cell EOL of a few cells in the data set.

Figure 10: End of Life of Cells (EOL)

```
Command Window
B0049: EOL at 32 cycles (80% of 0.858 Ah)
B0050: EOL at 8 cycles (80% of 0.863 Ah)
B0051: EOL at 32 cycles (80% of 0.643 Ah)
B0052: EOL at 8 cycles (80% of 0.861 Ah)
B0053: EOL at NaN cycles (80% of 1.069 Ah)
B0054: EOL at NaN cycles (80% of 0.740 Ah)
B0055: EOL at NaN cycles (80% of 0.799 Ah)
B0056: EOL at NaN cycles (80% of 0.785 Ah)
>> computeCycleLife
B0005: EOL at 63 cycles (80% of 0.777 Ah)
B0006: EOL at 63 cycles (80% of 0.847 Ah)
```

Workspace

Name	Value
ans	1x4 cell
capacit...	1x1 struct
caps	204x1 do...
cellFile	'C:/.../B00...
cells	34x1 cell
cycleIdx	1
cycleLi...	1x1 struct
i	34
idx	[]
nominal	0.7853
thr	0.6282

5. Model Construction

Two Machine-Learning Models are analyzed in this study, that is, the Random Forest (RF), representing the ensemble methods and the Long Term Short Memory (LSTM) representing the deep learning class. These models are selected being good representatives of each class of Machine – Learning techniques.

6. Preparing the Data for Machine Learning

The data below shows the cycle number, capacity and temperature as input features and the RUL as the output.

Figure 10: Capacity Feature

```
Command Window
>> prepareMLdata

Built ML table with 3291 rows
>> load('MLData.mat');head(T)
    CellName     Cycle    Capacity    Temperature    RUL
    ________     _____    ________    ___________    ___

    {'B0005'}      1      0.77703       25.324       62
    {'B0005'}      2       1.8565       32.572       61
    {'B0005'}      3       1.8801       26.636       60
    {'B0005'}      4       1.8463       32.725       59
    {'B0005'}      5        1.873       26.778       58
    {'B0005'}      6       1.8353       32.643       57
```

As the number of cycles increase the RUL decreases.

6.1 Splitting the Data.

The data is split into training data and test data with 70% of the rows are used for training and 30% used for testing.

International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Figure 11: Split Data

```
Command Window
>> splitData

Split: 492 train samples, 211 test samples
fx >>
```

7. Proposed Machine- Learning Models

7.1 Random Forest (RF)

The scatter plot for the predicted RUL versus the actual RUL is shown in the MATLAB plot below.

Figure 12: Random Forest Scatter Plot

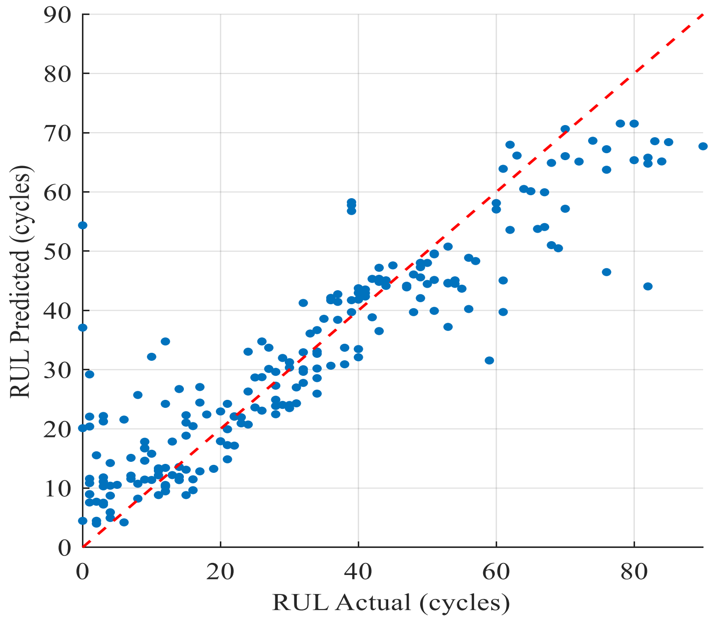


Figure 13: Random forest statistics

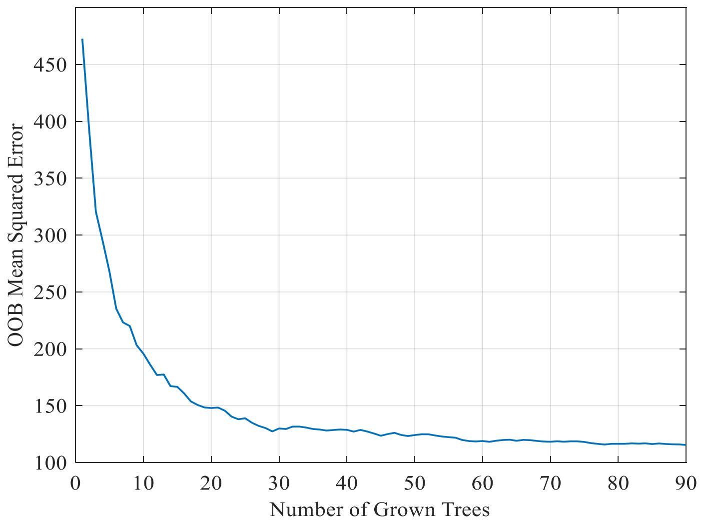
```
Command Window
>> trainRandomForest

  size(X_train) = 492x3,   size(Y_train) = 492x1
Random Forest Test RMSE = 10.242 cycles
Random Forest Test R^2   = 0.807
>> trainRandomForest

  size(X_train) = 492x3,   size(Y_train) = 492x1
Random Forest Test RMSE = 10.292 cycles
Random Forest Test R^2   = 0.805
fx >>
```

RMSE) = 10.242 Cycles and R^2 = 0.807

Figure 14: Out-of- bag error

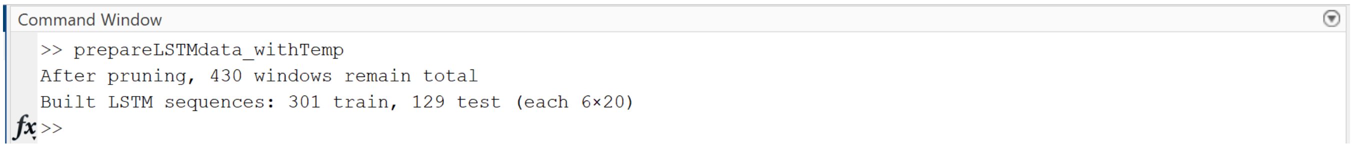


The figure shows a decreasing error as number of grown trees increase reaching a constant value of around 40. This happens below 90 trees.

7.2 Long Short-Term Memory

We prepare LSTM data for training purposes.

Figure 16: Prepare LSTM Data



```
Command Window
>> prepareLSTMdata_withTemp
After pruning, 430 windows remain total
Built LSTM sequences: 301 train, 129 test (each 6×20)
fx >>
```

The LSTM sequence has: 301 train and 129 test (6 x 20 features). After pruning 430 windows remain.

LSTM's training progress is shown below.

Figure: 17 LSTM Training Progress

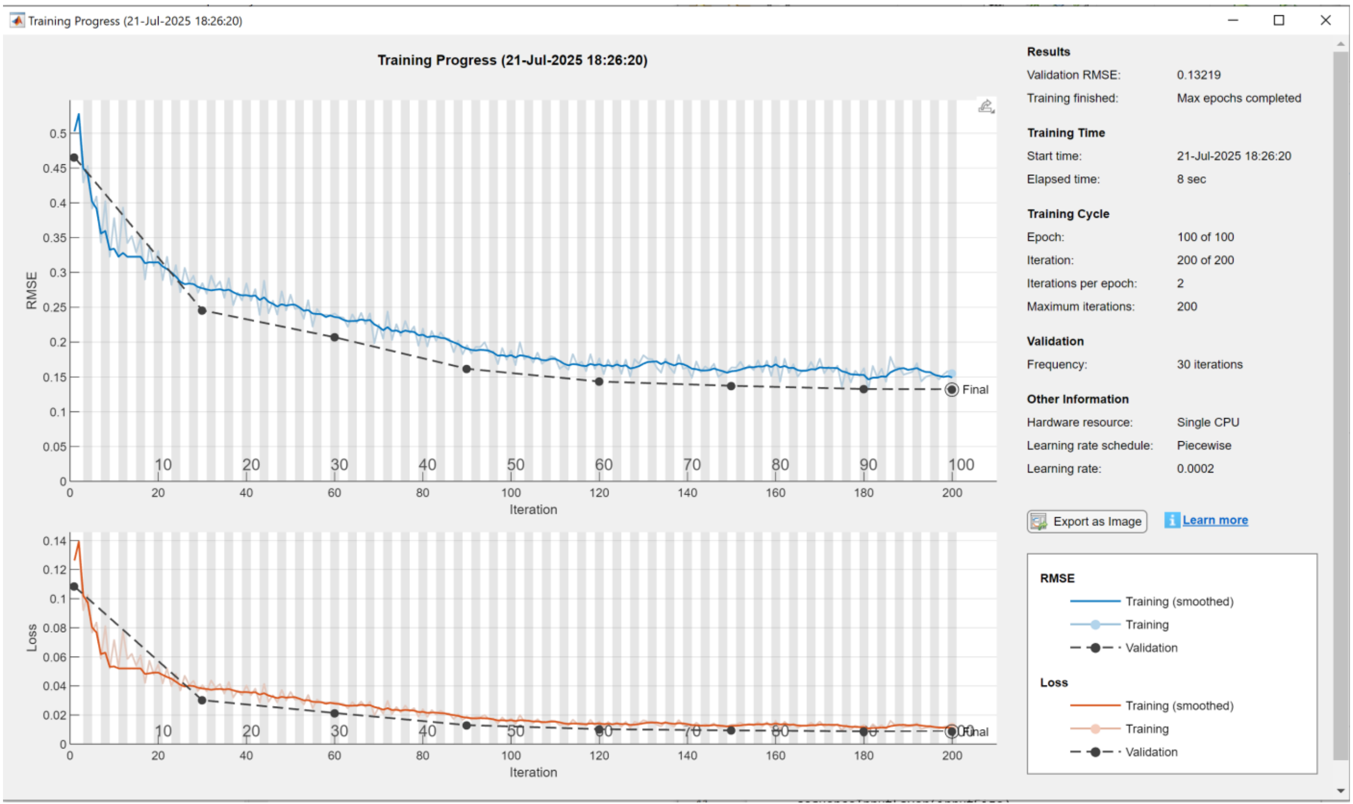


Figure 18: LSTM Scatter Plot

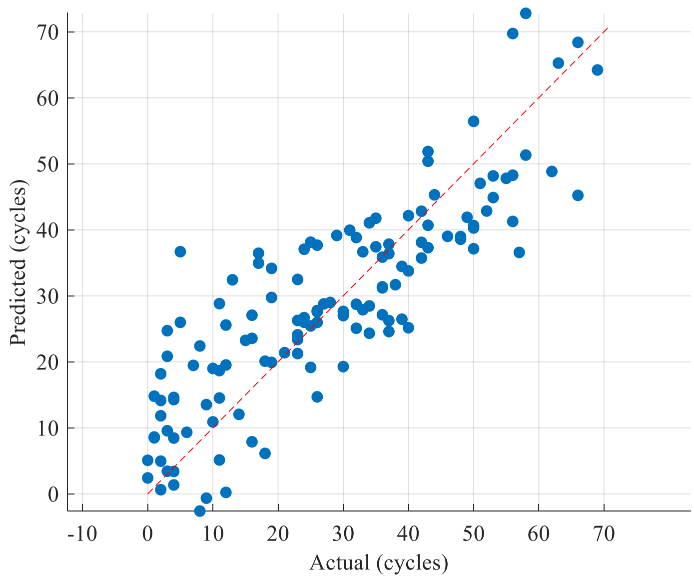


Figure 19: LSTM Scaled Scatter Plot

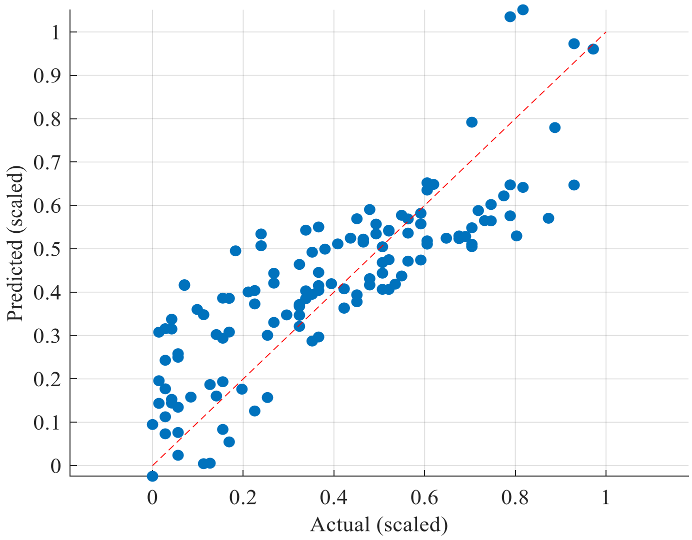


Figure 20: LSTM Summary Statistics

```
Command Window
>> prepareLSTMdata_withTemp
After pruning, 430 windows remain total
Built LSTM sequences: 301 train, 129 test (each 6×20)
>> trainLSTM
About to train: 301 predictors / 301 responses
About to train: 301 predictors / 301 responses

LSTM (scaled) RMSE = 0.132   R^2 = 0.733
LSTM (cycles) RMSE = 9.4 cycles

fx >>
```

RMSE (scaled) = 0.132, RMSE (cycles) = 9.4 cycles and R^2 = 0.733.

7.3 Comparison of the Two Models

Table 1Summary Statistics for the Models

Machine-Learning Model	RMSE (cycles)	R^2
Random Forest	10.242	0.807
LSTM	9.4	0.733

The results presented in this study show that the Random Forest is a better predictor of battery RUL compared to the LSTM model. It should however be noted that, the prediction accuracy of the LSTM improves with increasing training [6]. This might mean that with continuous training the LSTM model might produce better results. Continuous training is thus required when using the LSTM model for prediction purposes.

8. Conclusion

In this study, the NASA battery data set was used to compare the Random Forest model and the Long Short-Term Memory model for battery remaining useful life prediction. The exploratory data analysis was carried out in MATLAB. The input features analyzed are the battery capacity, voltage, current and temperature with the battery RUL as the output. The data for each of the models was prepared for Machine-Learning. The Machine learning models were implemented in MTLAB and the Root Mean Square Error (RMSE) and the Correlation coefficient, R^2 were produced to aid the model comparison. The graphical correlation analysis was carried out in MATLAB to depict the relationship between these features as the cells charge and discharge. The generated results in this paper show that the Random Forest performs better than the LSTM model in battery RUL prediction using the NASA battery data set. However, it has been noted that with continuous model training the LSTM model produces better results.

The study depicts the power of Machine-Learning models in predicting battery Remaining Useful Life. This is very important for predictive maintenance. Safety and reliability of the batteries is required for most practical applications. Although the study focused on the NASA battery data set, it should be noted that the same analysis can be done for all other maintenance systems where safety, security and reliability are required. In these safety critical systems, real time status of the system is required. This helps to detect system defects before total system failure occurs thereby preventing total failure. Total system collapse results in increased down time and increased replacement costs of the system. More studies are however required on the use of Machine-Learning models in such systems for predictive maintenance.

9. Acknowledgement

We would like to acknowledge the Harare Institute of Technology (HIT) for affording the opportunity to carry out this study and fostering a research driven environment in their degree programmes. Special thanks goes to the Electronic Engineering department Lecturers for providing the necessary theoretical and practical knowledge which contributed greatly to this study.

10. Authors' Biography

Monias Tapiwanashe Munhamoh is a Master of Technology candidate at Harare Institute of Technology studying Industrial Automation. His Research interests focuses on Machine-Learning algorithms in predictive maintenance in safety critical systems. He holds a Bachelor of Technology degree in Electrical and Electronic Engineering. Munhamoh has wide experience in Electrical power systems and control. The current study focuses on comparison of Machine-Learning algorithms in battery predictive maintenance. Peter Bukelani Musiiwa is a Lecturer at Harare Institute of Technology. He holds a Master in Technology degree in VLSI design among other Engineering degrees.

REFERENCES

1. K. Park, Y. Choi, W. J. Choi, H. Y. Ryu, and H. Kim, 'LSTM-Based Battery Remaining Useful Life Prediction with Multi-Channel Charging Profiles', IEEE Access, vol. 8, pp. 20786–20798, 2020, doi: 10.1109/ACCESS.2020.2968939.

2. S. Oyucu, F. Doğan, A. Aksöz, and E. Biçer, 'Comparative Analysis of Commonly Used Machine Learning Approaches for Li-Ion Battery Performance Prediction and Management in Electric Vehicles', Applied Sciences (Switzerland), vol. 14, no. 6, Mar. 2024, doi: 10.3390/app14062306.
3. M. S. Reza, M. Mannan, M. Mansor, P. J. Ker, T. M. I. Mahlia, and M. A. Hannan, 'Recent advancement of remaining useful life prediction of lithium-ion battery in electric vehicle applications: A review of modelling mechanisms, network configurations, factors, and outstanding issues', Jun. 01, 2024, Elsevier Ltd. doi: 10.1016/j.egyr.2024.04.039.
4. N. Yang, H. Hofmann, J. Sun, and Z. Song, 'Remaining Useful Life Prediction of Lithium-ion Batteries with Limited Degradation History Using Random Forest', IEEE Transactions on Transportation Electrification, 2023, doi: 10.1109/TTE.2023.3323976.
5. W. Xu et al., 'Long-Short-Term-Memory-Based Deep Stacked Sequence-to-Sequence Autoencoder for Health Prediction of Industrial Workers in Closed Environments Based on Wearable Devices', Sensors 2023, Vol. 23, Page 7874, vol. 23, no. 18, p. 7874, Sep. 2023, doi: 10.3390/S23187874.
6. V. Safavi, A. Mohammadi Vaniar, N. Bazmohammadi, J. C. Vasquez, and J. M. Guerrero, 'Battery Remaining Useful Life Prediction Using Machine Learning Models: A Comparative Study', Information 2024, Vol. 15, Page 124, vol. 15, no. 3, p. 124, Feb. 2024, doi: 10.3390/INFO15030124.
7. Y. H. Chang, Y. C. Hsieh, Y. H. Chai, and H. W. Lin, 'Remaining-Useful-Life Prediction for Li-Ion Batteries', Energies (Basel), vol. 16, no. 7, Apr. 2023, doi: 10.3390/en16073096.
8. H. You et al., 'Investigation of lithium-ion battery nonlinear degradation by experiments and model-based simulation', Energy Storage Mater, vol. 65, p. 103083, Feb. 2024, doi: 10.1016/J.ENSM.2023.103083.
9. J. Fan, J. Fan, F. Liu, J. Qu, and R. Li, 'A Novel Machine Learning Method Based Approach for Li-Ion Battery Prognostic and Health Management', IEEE Access, vol. 7, pp. 160043–160061, 2019, doi: 10.1109/ACCESS.2019.2947843.
10. Y. Che, X. Hu, X. Lin, J. Guo, and R. Teodorescu, 'Health prognostics for lithium-ion batteries: mechanisms, methods, and prospects', Energy Environ Sci, vol. 16, no. 2, pp. 338–371, Jan. 2023, doi: 10.1039/D2EE03019E.
11. X. Li, D. Yu, V. Søren Byg, and S. Daniel Ioan, 'The development of machine learning-based remaining useful life prediction for lithium-ion batteries', Journal of Energy Chemistry, vol. 82, pp. 103–121, Jul. 2023, doi: 10.1016/J.JECHEM.2023.03.026.
12. K. Luo, X. Chen, H. Zheng, and Z. Shi, 'A review of deep learning approach to predicting the state of health and state of charge of lithium-ion batteries', Journal of Energy Chemistry, vol. 74, pp. 159–173, Nov. 2022, doi: 10.1016/j.jechem.2022.06.049.
13. Y. Yang, 'A machine-learning prediction method of lithium-ion battery life based on charge process for different applications', Appl Energy, vol. 292, p. 116897, Jun. 2021, doi: 10.1016/J.APENERGY.2021.116897.
14. Z. Ren and C. Du, 'A review of machine learning state-of-charge and state-of-health estimation algorithms for lithium-ion batteries', Energy Reports, vol. 9, pp. 2993–3021, Dec. 2023, doi: 10.1016/j.egyr.2023.01.108.