South Asian Journal of Science and Technology

Open Access Full Text Article 153(2025) 49-64 DOI: 10.26524/sajet.2025.15.20

Machine Learning-Based framework for battery Remaining Useful Life Prediction

Munhamoh M.T. [1,a], ***Musiiwa P.B.*** [1,2,b*], ***Kapuya E.T.*** [2,c], ***Musademba D.*** [2,d], ***Rushambwa M. C.*** [1,e], ***Palaniappan R.*** [3,f], ***Vijean V.*** [4,g]

[1] Electronic Engineering Department, School of Engineering and Technology, Harare Institute of Technology, Zimbabwe, Ganges Road, Belvedere, Harare

[2] Mechatronics Department, School of Engineering and Technology, Chinhoyi University Technology, Zimbabwe, Ganges Road, Belvedere, Private Bag 7724, Chinhoyi, Zimbabwe.

[3] College of Engineering, Department of Mechatronics Engineering, University of Technology, Bahrain, 18041 Kingdom of Bahrain.

[4] BioSIM Group, Sport Engineering Research Centre, University Malaysia Perlis, 02600 UniMAP, Arau, Perlis, Malaysia.

E-mail: [a] h230960y@hit.ac.zw , [b] pbmusiiwa@hit.ac.zw, [c] engelbert.kapuya@gmail.com, [d]dmusademba@gmail.com, [e,*] mrushambwa@hit.ac.zw, [f]prkmect@gmail.com, [g]vicky.86max@gmail.com.

Article info

Received 10 June 2025 Received in revised form 17 August 2025 Accepted 7 September 2025

Keywords:

Machine Learning, Random Forest, Long-Term Short Memory, Correlation Coefficient

.

https://sajet.in/index.php/journal/article/view/348

Abstract

Lithium-ion batteries find great use in critical power requirement in electric vehicles, renewable energy storage systems, aerospace and aviation, medical devices and substation DC systems. These systems require reliable operation and higher safety considerations thereby preventing catastrophic failures. Real time monitoring of critical battery parameters such as capacity, voltage, current and temperature is paramount for predictive maintenance. Several Machine Learning based techniques such as Decision Tree Regression, Random Forest, Support Vector Regression, Gaussian Process Regression and Long-Term Short Memory can be used to predict the Remaining useful life (RUL) of batteries. In this study, three machine-learning models are considered. These are the Random Forest (RF), which represents the ensemble methods class of machine learning. Support Vector Regression (SVR), representing the classical regression models class and the Long-Term Short Memory (LSTM) representing the deep learning/sequence models class. The selected models are on the basis of being good representative of each class of Machine – Learning models. The methodology used include downloading and loading in MATLAB the online NASA data set. Exploratory data analysis in MATLAB, preparing the Data for Machine Learning, Implementing the three Machine

Learning Models, Comparing the Models and making Remaining Useful Life (RUL) predictions. The performance parameters such as the Root Mean Square Error (RMSE) and the Statistical Correlation Coefficient $\mathbf{R^2}$ are analysed to find the Model performance in predicting RUL. The LSTM performed better than Random Forest in accuracy and long-term prediction. The technique is complex and slower to train. However, the SVR model performs better with hyper-parameter tuning. The study contributes to the increasing body of Machine Learning techniques in predictive maintenance. Routine checking done practically usually requires work force. Predictive maintenance, allow for real time monitoring according to the model developed and corrective action taken. The study provides techniques for predictive maintenance, for batteries and other system with measurable raw operational data. Further research may be required on the integration of different Machine Learning based techniques in predicting the RUL. This improves the prediction accuracy, robustness and adaptability.

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 require [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 sources because of the lighter weight, huge voltage density, and longer charging/discharging life as well as less self-discharge [2]. 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 [3]. Model based methods have proved to be better than physics-based methods, but they have complex calculations [4] . 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. Other studies were carried out on how to improve system efficiency in the manufacturing industry using Machine-Learning techniques [5] . The study improved operational efficiency by predicting equipment failure before total breakdown thereby reducing down time and maintenance/replacement costs [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 [7] . Accurate prediction of RUL enhances predictive maintenance thereby saving costs by timely predicting equipment failures [8] . 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 [9] . Machine learning-based approaches provide better results than statistical approaches in terms of accuracy [10] . 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 [11]. The statistical parameters help to provide a comparative analysis of RF, LSTM and SVR 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 [12] . The degradation patterns influenced by factors such as, temperature, aging, and charge/discharge cycles of the cell [13] . 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, LSTM and SVR models in predicting battery RUL.

3. Model Description

3.1 Random Forest (RF)

The RF technique is an ensemble method, which combines multiple decision trees with great variability to improve on the prediction accuracy [13] . This inherent structure allows for interpretability of model input features. The technique is a good feature selector. The technique is very robust, and can is suitable for high-dimensional data [14]. The ensemble technique is suitable for noisy and complex problems in decision trees. However, for multiple models, the computational costs increase, interpretation and explanation become complex [15].

3.2 Long Short-Term Memory (LSTM)

LSTM model is an improvement of Recurrent Neural Network (RNN), which fits well with time series data capturing long dependencies [16]. This done though improving the hidden layer in RNN [13]. LSTM techniques require a large data set. The method is suitable for sequential data processing, learning from the past data values [9].

3.3 Support Vector Regression (SVR)

This model is used for regression tasks. The model finds a margin of error (support vectors) along the hyperplane. The error margin defined by the predicted values [17].

4. RUL Prediction Framework

This section outlines the methodology used in developing a machine learning-based framework for Li-ion battery RUL prediction using MATLAB simulation-based approach.

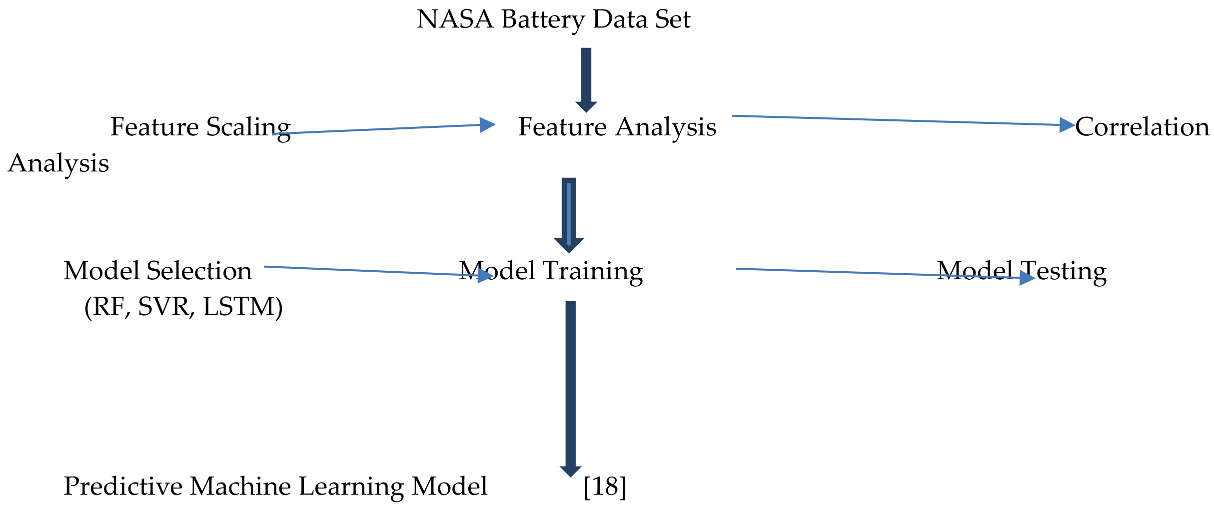

Figure 1: RUL Prediction Framework

4.1 Data Acquisition

This involves downloading and loading in MATLAB the publicly available online NASA battery data set. The data set shows how the battery Ampere –hour (AH) capacity degrades over charging and discharging cycles. The data is placed in a MATLAB folder. 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

Data pre-processing involves feature scaling to standardize the temperature, current, voltage over time. This is done to handle these greatly varying independent values. This is followed by feature analysis of the independent variables. Correlation analysis then analyses the relationship between these parameters and the battery's capacity. Data preprocessing carried out through exploratory data analysis (EDA). The battery capacity is plotted against the charging/discharging cycle number. The cycle at which 20% of the original battery capacity remains is noted. This represents the end of life (EOL) of the battery under consideration [16].

4.3 Feature Analysis

The selected features are shown by means of graphs during discharging process. The cell starts to discharge at approximately 1.9Ah and degrades gradually to about 1.35Ah. This trend is applicable to all battery cells. The cell capacity continues to decrease as the number of cycles increases. This decrease is associated with the electrochemical processes taking place within the cell forming some insulating material on the electrodes [2]. This decreases the cell capacity.

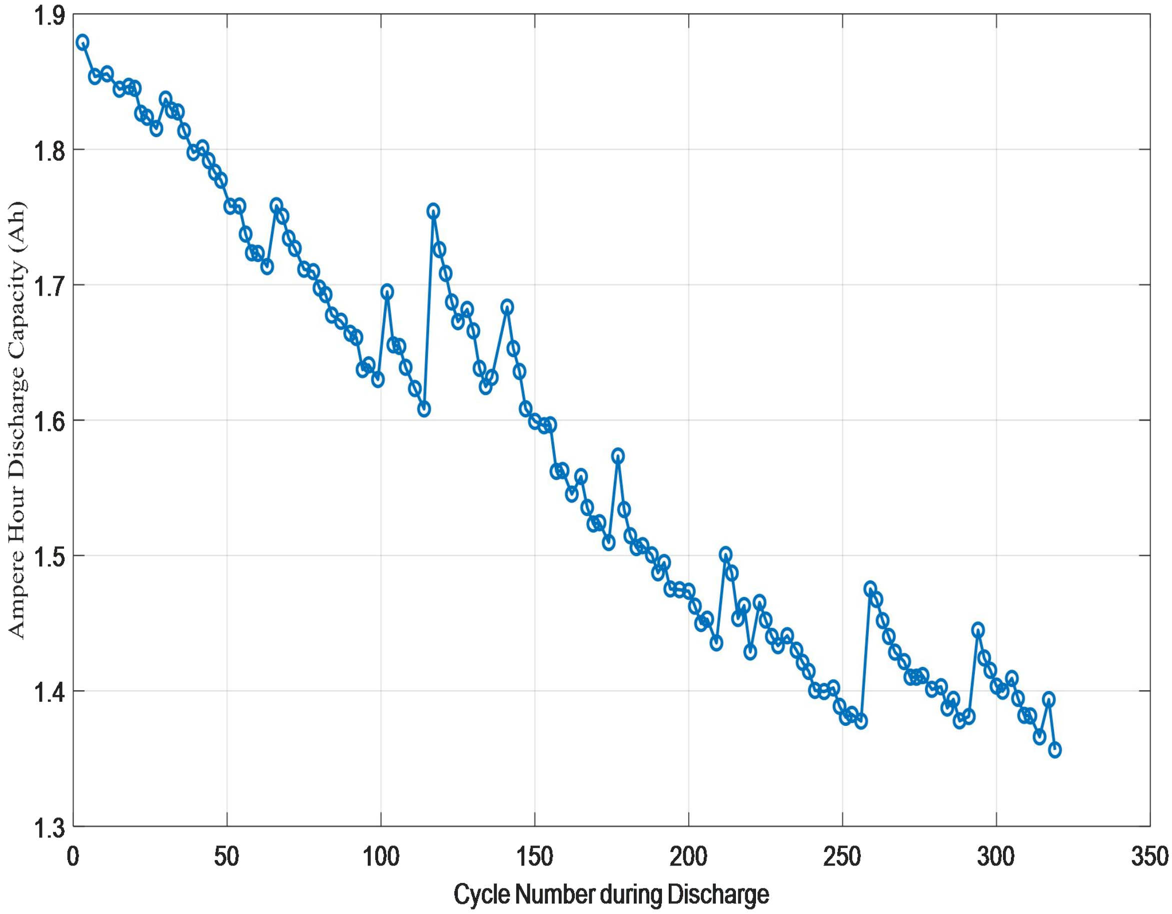

Figure 2: Discharge capacity plot for one sampled cell (B0018)

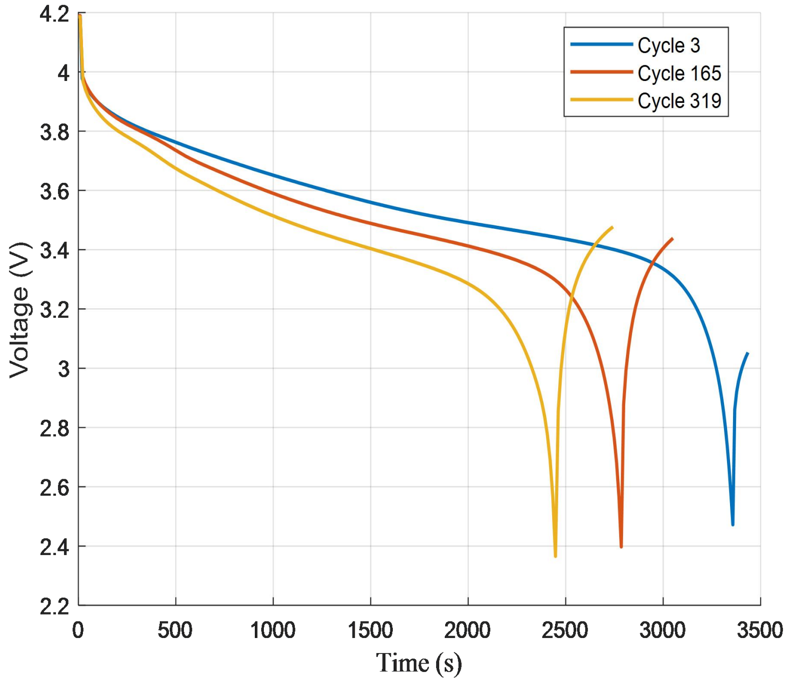

Figure 3: Voltage versus Time for sampled cycles

The voltage falls from 4.2V reaching a cut-off value of 2.2V before charging commences. The graph is a constant current discharge curve. The curves indicate that as the number of cycles increase the cell capacity also decreases. This is due to the electrochemical processes during charging and discharging [13].

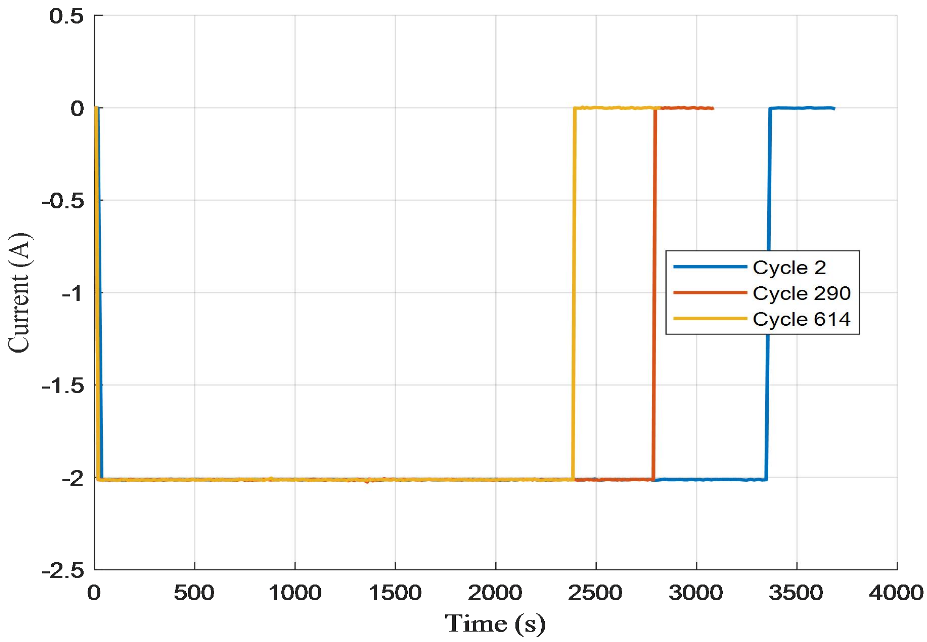

Figure 4: Current versus Time for a few sampled cycles

The current remains constant (CC) at 2A up to around 2500 sampled cycles. As the number of cycles increases, the time for which the current remains constant decreases. The graph shows constant current discharge pattern [19].

The temperature increases from around 24℃ to around 38℃ during discharge. The increase in temperature is due to the chemical activity between electrolyte and electrodes during the discharge [15] . Through visualization and calculating summary statistics, one gets to know the data characteristics. This sets up a basis for battery degradation data exploration before modelling.

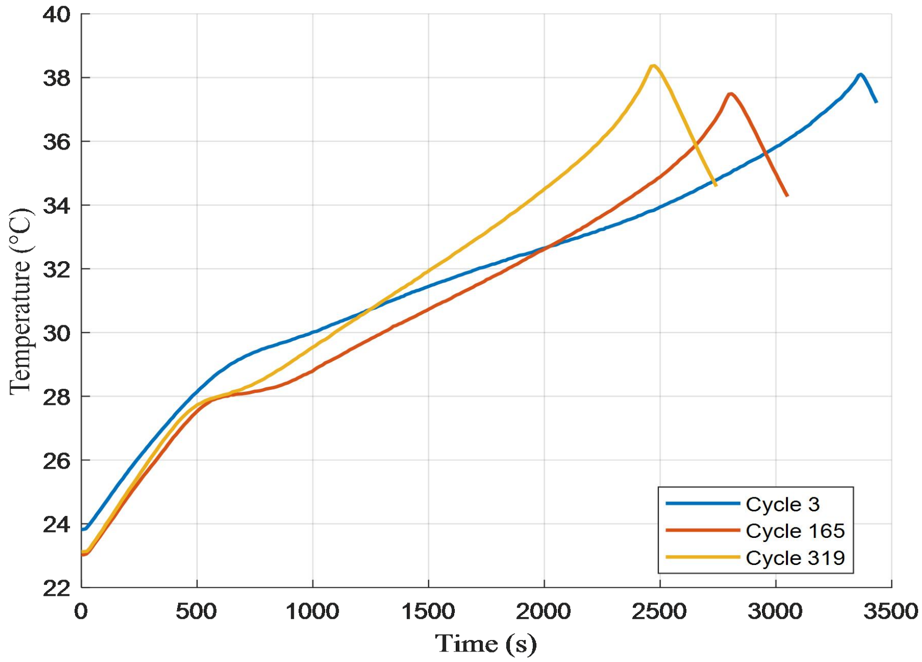

Figure 5: Temperature versus Time for sampled cycles

4.4 Features Correlation Analysis

The relationship between each independent variable voltage, current and temperature versus time is analyzed by plotting the variables in MATLAB.

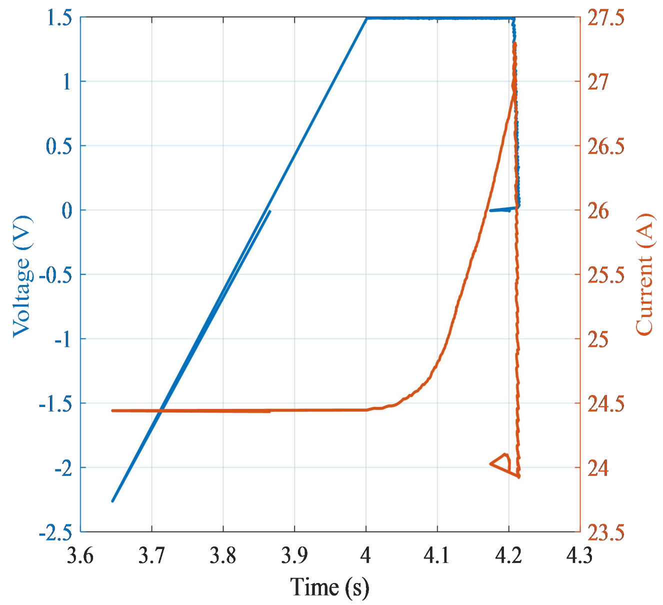

Figure 6: Voltage/Current versus time plot

Voltage rises steadily to maximum while impressing a constant current on the cell. This is constant-current (CC) charging. The voltage is then kept constant (CV) charging) at its maximum allowing the current to fall [16]. A sharp fall in both voltage and current then occurs.

The voltage and temperature steadily rise to a maximum plateau before falling rapidly. This signal falls at the start of the discharge time. A constant current is impressive on the cell causing the temperature to rise steadily to maximum limit. After the limit, a sharp fall of both current and temperature occurs.

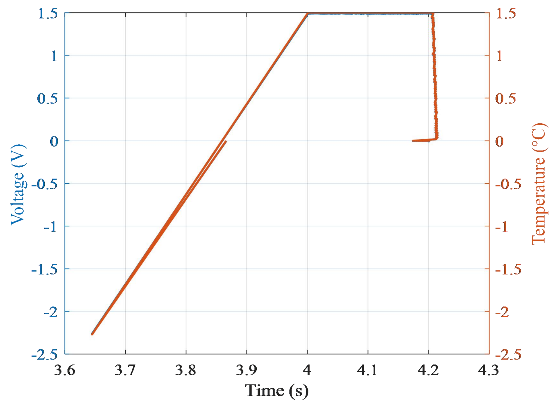

Figure 7: Voltage/Temperature versus time plot

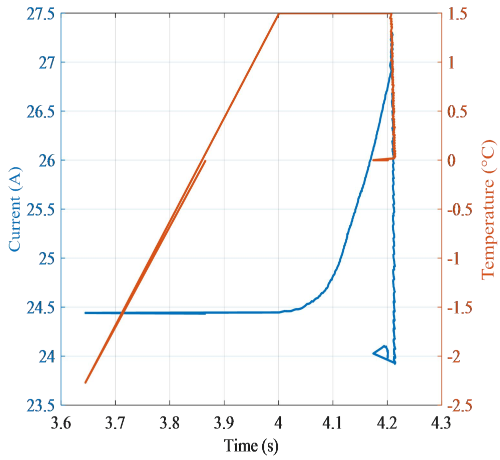

Figure 8: Current/Temperature versus time plot

5. Computing Simple Statistics for Data

The end of life (EOL) of the battery cells is the cycle at which the battery capacity degrades to 80% of its original capacity [18]. Figure 10 shows the EOL of the different battery cells in the data set.

6. Model Construction

For this study, three machine-learning models are considered. These are the Random Forest (RF), which represents the ensemble methods class of machine learning, Support Vector Regression (SVR), representing the classical regression models class and the Long-Term Short Memory (LSTM) representing the deep learning/sequence models class. The selected models are on the basis that they are a good representative of each class of Machine – Learning models. The models are also capable of capturing the nonlinear degradation tendencies in the battery data [11].

7. Preparing Data for Machine Learning

The simple features used in this study are capacity versus cycle number, voltage, and temperature and current. These are the input features. The output feature is the Remaining Useful Life (RUL). For illustration purposes, we use the Capacity feature. A table containing capacity feature and the output RUL produced, for the first cell B0005, we have:

For the cell B0005, as the number of cycles increase the RUL decreases.

7.1 Splitting the Data.

The data for Machine-learning is split into training and test data. In this study, 70% of the rows are used for training and 30% for testing purposes. This gives two files in MATLAB *splitdata* and *trainTestData.mat.*

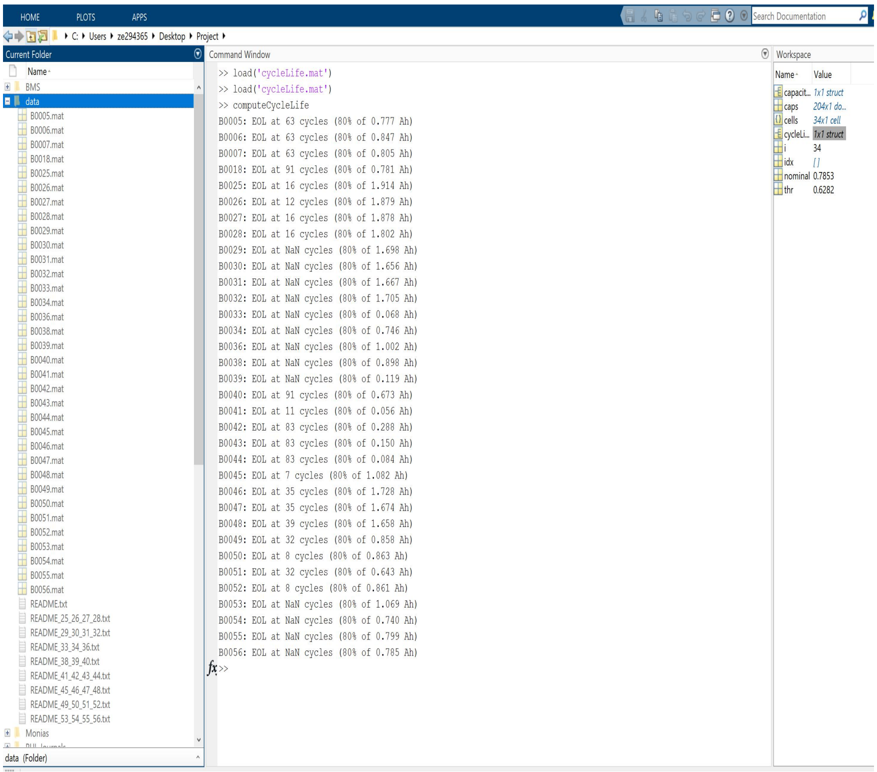

Figure 9: Battery cell End of Life (EOL)

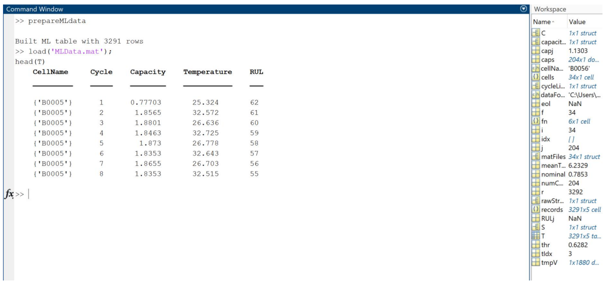

Figure 10: Capacity Feature Table

Figure 11: Split Data

The data is split into 492 training samples and 211 test samples.

8. Proposed Machine- Learning Models

8.1 Random Forest (RF)

The RF technique is an ensemble method, which combines multiple decision trees with great variability to improve on the prediction accuracy [10] . This inherent structure allows for interpretability of model input features. The technique is a good feature selector [11].

The MATLAB Machine-Learning and Statistics toolbox provides bagged decision trees for Random Forest. The idea is to train a Random Forest regression model that predicts the Remaining Useful Life (RUL) from the battery cycles and battery capacity.

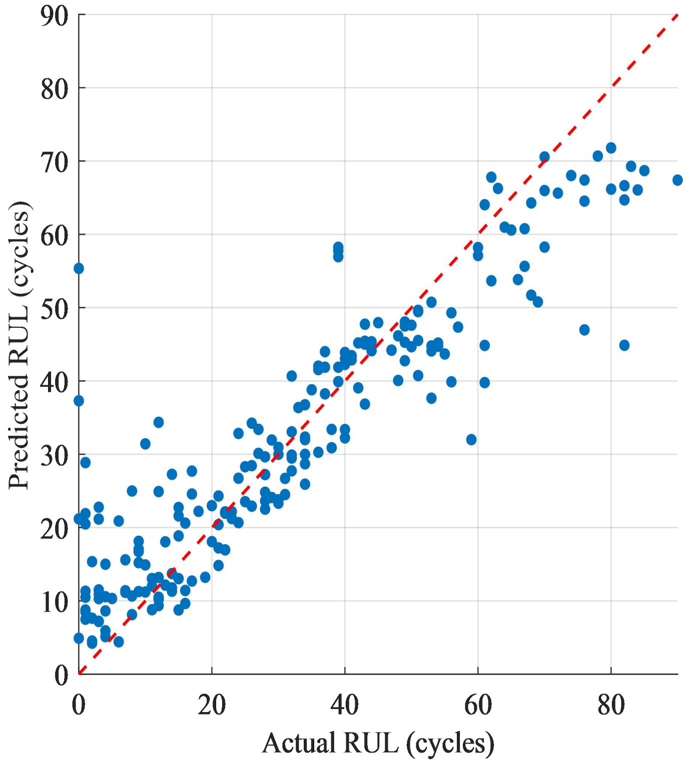

Figure 12: Random Forest Predicted versus Actual RUL

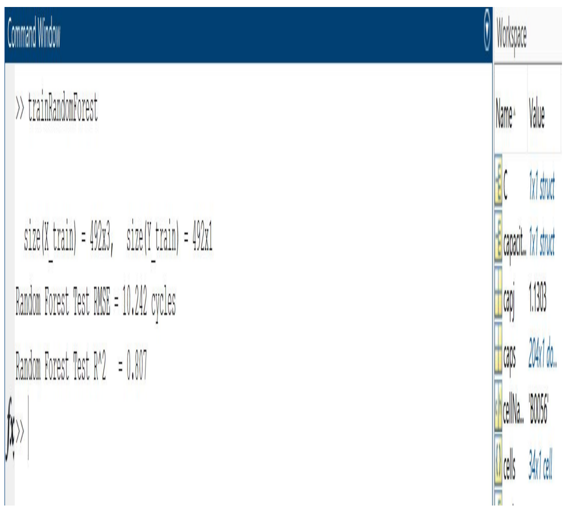

Figure 13: Random Forest Summary Statistics

For the Random Forest, the Root Mean Square Error (RMSE) is 10.242 and the Correlation Coefficient (R^2) is 0.807.

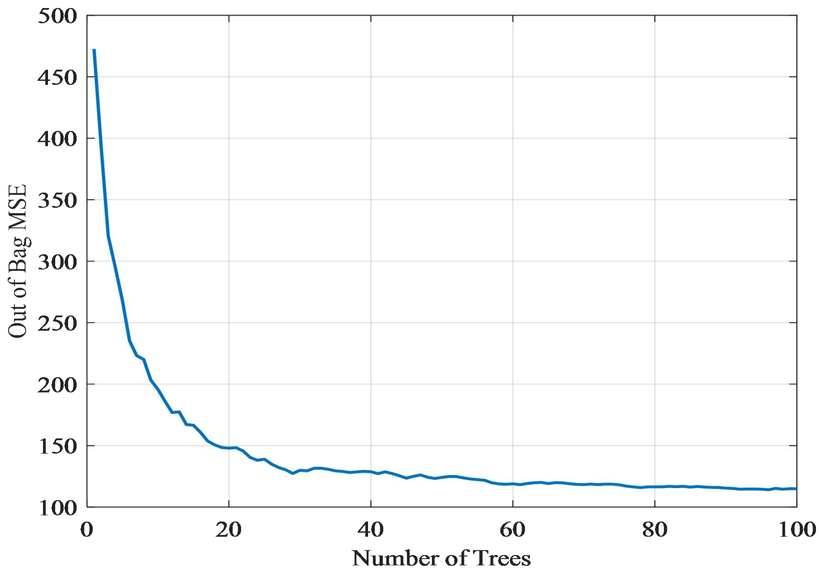

Figure 14: Out of Bag Error versus Number of Grown Trees.

The out of bag error decreases as the number of grown trees increases with error constant below 100 trees.

8.2 Long Short-Term Memory

The LSTM network is a type of recurrent neural network (RNN) designed for time series data. They are particularly suited for time-series prediction tasks [20].

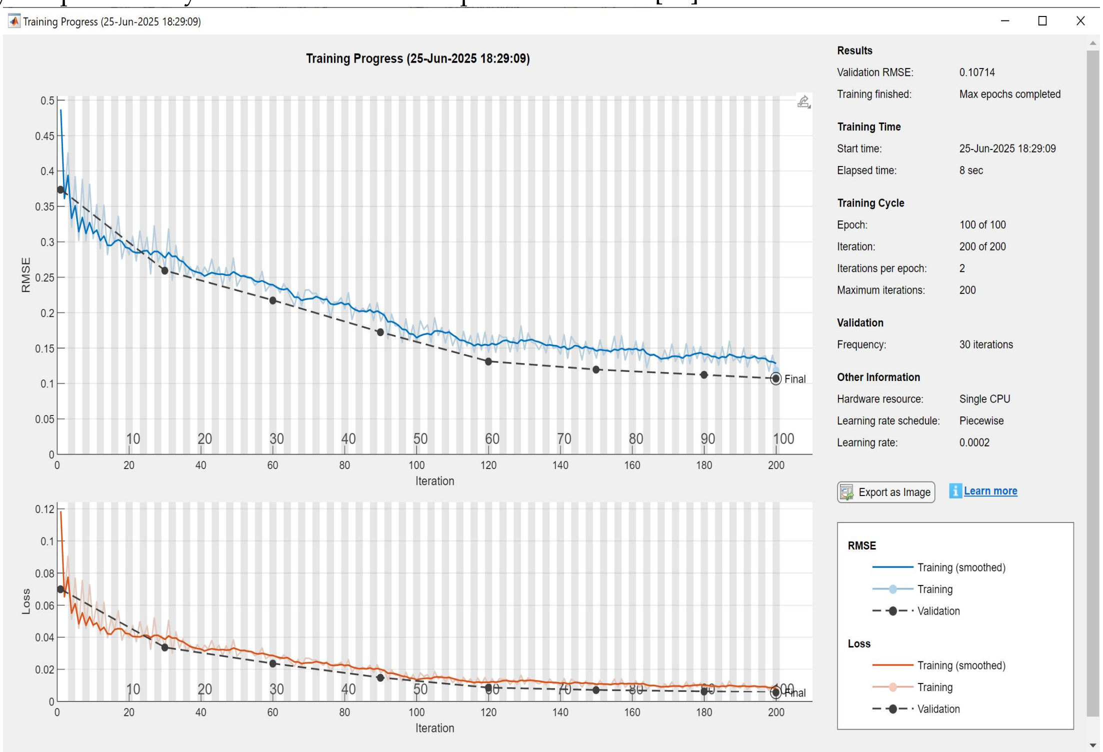

Figure 15: LSTM Training Progress

There is a general rapid drop in RSME for the initial iterations giving a rough RUL prediction process. The decreasing trend continues as iterations increase but at a less rapid manner. On the loss curve, the training and validation losses decrease with the validation curve closely following the training curve. A sign that there is no overfitting. With a lower loss function, the model prediction accuracy is higher [21].

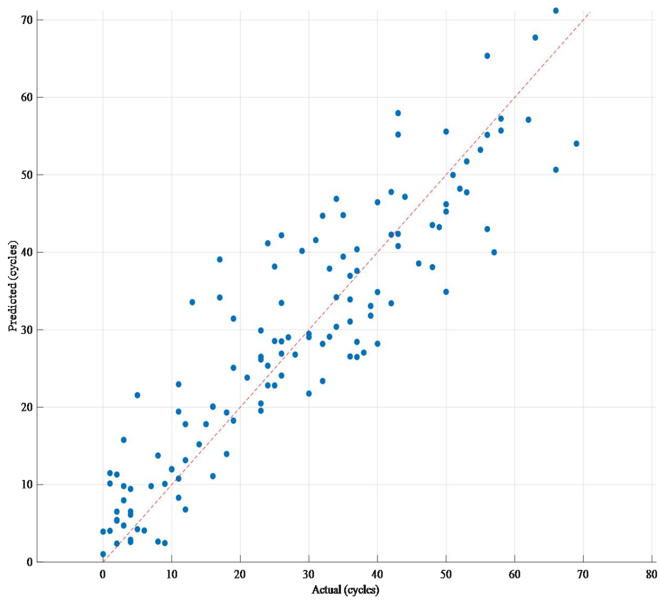

Figure 16: LSTM Predicted Cycles versus Actual Cycles

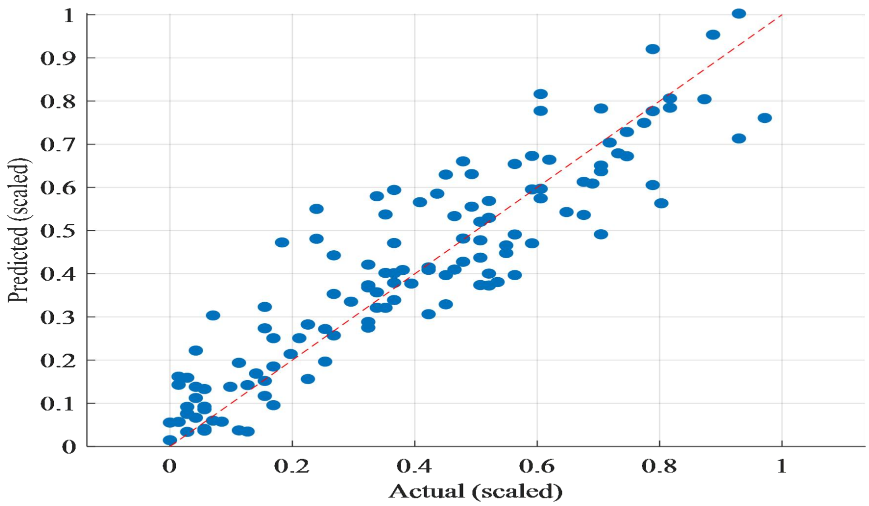

Figure 17: LSTM Predicted scaled RUL versus Actual RUL

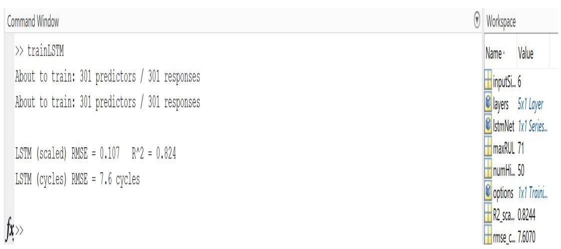

Figure 18: LSTM Summary Statistics

The LSTM RMSE is 7.6 cycles with $R^2 = 0.824$. The variable *net* in the workspace is the trained LSTM model.

9. Support Vector Regression (SVR)

This model represents the classical (shallow) regression models. The features considered in this model are cycle, capacity and mean temperature. These are the inputs to the model. The RUL being the output. The train/test split is generated in MATLAB with 492 training samples and 211 test samples.

Figure 19: SVR Split Data

We train a Support Vector Regression with inputs Cycle, Capacity, and Mean Temperature to give battery RUL.

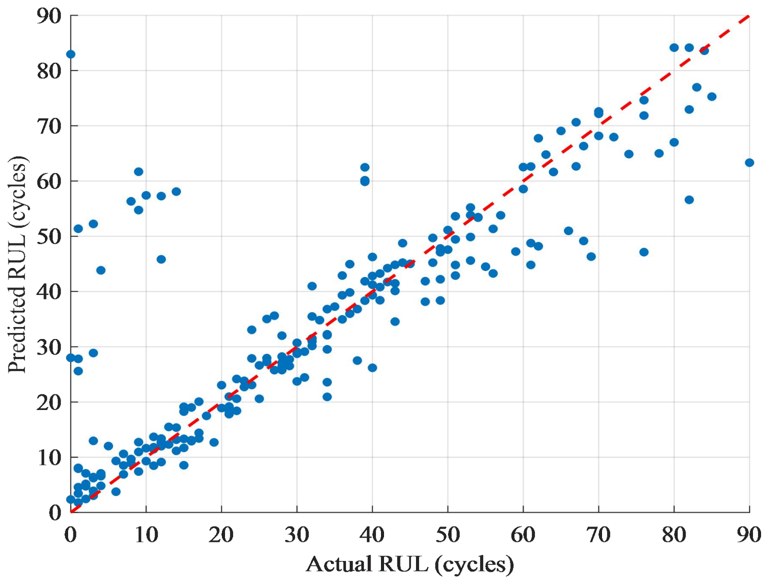

Figure 20: SVR Predicted RUL versus Actual RUL

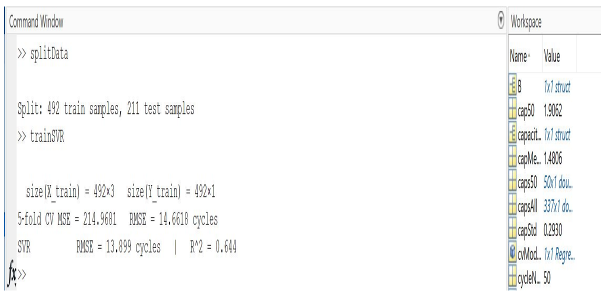

Figure 21: SVR Split data

The SVR RMSE = 13.899 cycles with $R^2 = 0.644$.

10. Hyper- Parameter SVR Tuning

In order to improve the performance of the SVR model, MATLAB can optimize the Box Constraint and Kernel Scale [22].

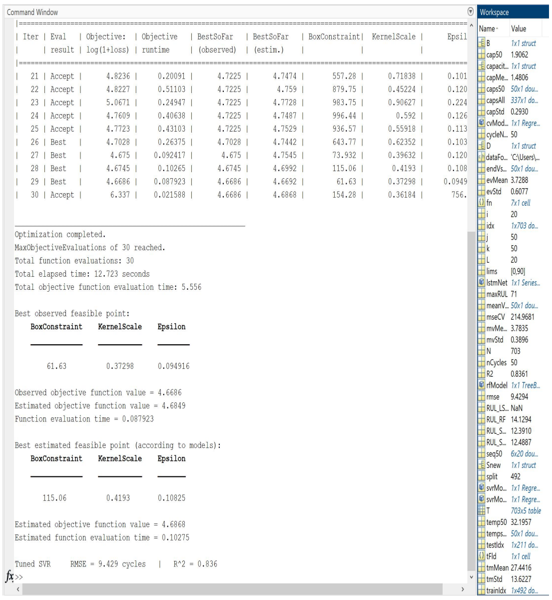

Figure 22: Tuned SVR parameters

With the tuned SVR, RSME improves to 9.429 cycles with R^2rising to 0.836.

11. Comparison of the Models

Table 1:Summary Statistics for the Models

Machine-Learning Model	RMSE (cycles)	R^2
Random Forest	10.242	0.807
LSTM	7.6	0.824
Support Vector Regression	13.899	0.644
Tuned Support Vector Regression	9.429	0.836

Before SVR hyper-parameter tuning the LSTM model proved to be the best model to predict battery RUL with RMSE = 7.6 Cycles and R^2 = 0.824. After parameter tuning, the SVR gave better results. We thus use the tuned SVR model for RUL prediction in this study.

In another study, with NASA battery data set, 70% of the data was used for training and 30% used for testing [7]. When the data for training increased to 70%, prediction improved compared to 50% training. The RSME of 0.038 was obtained being the best amongst RF (0.138), Non –Linear Autoregressive model (0.185) and Recurrent Neural Network (0.045). In this study, the RUL obtained for the LSTM was 76 compared to the real value of 84. These obtained results go hand in hand with the results obtained in this study. The Mean Absolute Error (MAE) of the battery RUL of 60% training data set, 20% selected for validation and 20% used for model testing was 62 cycles [6] . In this study, the RF model made reasonable predictions with the predicted RUL decreasing as the true value decreases. However, it is noted that some parameters of the RF model may be hard to obtain in real world applications.

In the study of predicting the remaining usage time of bearings, the Machine-Learning Algorithms Support Vector Machines (SVM), RF classifier and K-Nearest Neighbors were compared on their prediction accuracy [23] . The results indicated that the SVM achieved the highest accuracy of 96.74%, with RF at 95.95% and K-neighbor 91.77%. The SVM proved to be a better model in this case.

The results from literature suggest that in most applications, the LSTM model outperforms the other models. However, in most cases other models such as the RF and SVM may produce better results. These differences may be because different data sets are used under different circumstances. It is also important to note that the input features may be different for different applications.

12. RUL Prediction Using Models

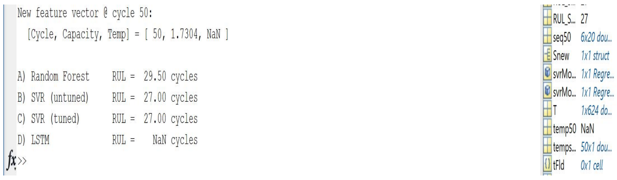

Figure 23: RUL prediction

In this case, cell B0040 from the data set is used to find RUL. The remaining RUL using the tuned SVR is 27 cycles. We can use the same approach and analysis for any other cell in the data set.

13. Conclusion

This study looked at three Machine Learning models used to determine battery remaining useful life using the NASA data set. In the paper, MATLAB is used to inspect the data, carry out feature analysis and pre-processing the data. Correlation analysis between the selected dominant features done. With each model, training and testing are done according to that model. The RUL for each model was generated. The statistical parameters RMSE and R^2 were generated to aid the model comparisons. The LSTM model performed better than any other model in predicting RUL. However, after hyper-parameter tuning with the SVR model, the accuracy of prediction increased. The tuned SVR model thus gave the best results. Although the analysis in this study determined the battery RUL, the same principles can be applied to most systems in predictive maintenance. More studies are required with other different systems.

Acknowledgement

The authors are thankful to Chinhoyi University of Technology and Harare Institute of Technology for providing support of publishing and providing facilities for conducting this research work.

REFERENCES

1. 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.
2. 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.
3. 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.
4. "Bootstrap Aggregation, Random Forests and Boosted Trees | QuantStart." Accessed: Apr. 23, 2025. [Online]. Available: https://www.quantstart.com/articles/bootstrap-aggregation-random-forests-and-boosted-trees/

5. R. Palaniappan, “Predictive manufacturing: State of the art, design standards, limitations, challenges, and future perspectives,” in *AIP Conference Proceedings*, American Institute of Physics, Nov. 2024. doi: 10.1063/5.0235999.
6. 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.
7. 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.
8. 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.
9. P. Wlodarczak, “Machine Learning and its Applications,” *Machine Learning and its Applications*, Oct. 2019, doi: 10.1201/9780429448782.
10. S. Ding, L. Wang, H. Dai, and X. He, “Prognosticating nonlinear degradation in lithium-ion batteries: operando pressure as an early indicator preceding other signals of capacity fade and safety risks,” *Energy Storage Mater*, vol. 75, Feb. 2025, doi: 10.1016/j.ensm.2024.103998.
11. 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.
12. F. Alanazi, “Electric Vehicles: Benefits, Challenges, and Potential Solutions for Widespread Adaptation,” *Applied Sciences (Switzerland)*, vol. 13, no. 10, May 2023, doi: 10.3390/APP13106016.
13. K. Karthick, S. Ravivarman, and R. Priyanka, “Optimizing Electric Vehicle Battery Life: A Machine Learning Approach for Sustainable Transportation,” *World Electric Vehicle Journal 2024, Vol. 15, Page 60*, vol. 15, no. 2, p. 60, Feb. 2024, doi: 10.3390/WEVJ15020060.
14. S. Jin, X. Sui, X. Huang, S. Wang, R. Teodorescu, and D. I. Stroe, “Overview of machine learning methods for lithium-ion battery remaining useful lifetime prediction,” *Electronics (Switzerland)*, vol. 10, no. 24, Dec. 2021, doi: 10.3390/ELECTRONICS10243126.
15. 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.
16. X. Hu, L. Xu, X. Lin, and M. Pecht, “Battery Lifetime Prognostics,” *Joule*, vol. 4, no. 2, pp. 310–346, Feb. 2020, doi: 10.1016/J.JOULE.2019.11.018.
17. 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.
18. 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.
19. 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.
20. 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.
21. X. Tang, H. Wan, W. Wang, M. Gu, L. Wang, and L. Gan, “Lithium-Ion Battery Remaining Useful Life Prediction Based on Hybrid Model,” *Sustainability (Switzerland)*, vol. 15, no. 7, Apr. 2023, doi: 10.3390/su15076261.

22. U. Farooq, M. Ademola, and A. Shaalan, "Comparative Analysis of Machine Learning Models for Predictive Maintenance of Ball Bearing Systems," *Electronics (Switzerland)*, vol. 13, no. 2, Jan. 2024, doi: 10.3390/electronics13020438.
23. R. Palaniappan, "Comparative Analysis of Support Vector Machine, Random Forest and k-Nearest Neighbor Classifiers for Predicting Remaining Usage Life of Roller Bearings," *Informatica (Slovenia)*, vol. 48, no. 7, pp. 39–52, Apr. 2024, doi: 10.31449/inf.v48i7.5726.