# Maximizing Electric Vehicle Battery Efficiency: A Multi-Model Machine Learning Approach for RUL Prediction of NMC-LCO Batteries

S. Sivakumar[1], Moazzam Haidari[2], Dr. A. M. Arun Mohan[3], Dr. M. Bheemalingaiah[4], Dr. R. Jayasudha[5], Senthilkumar C[6]

[1]*Department of Electrical and Electronics Engineering, Vel Tech Rangarajan Dr.Sagunthala R and D Institute of Science and Technology, Avadi, Chennai, Tamilnadu, India. siva.carthick@gmail.com*
[2]*Department of Electrical Engineering, Saharsa College of Engineering, Saharsa, Bihar, India. moazzam53211@gmail.com*
[3]*Department of Civil Engineering, Sethu institute of Technology, Pulloor, Kariapatti Tamilnadu, India. arunmohan@sethu.ac.in*
[4]*Department of CSE, J.B. Institute of Engineering and Technology, Hyderabad, India. bheemasiva2019@gmail.com*
[5]*Department of Mathematics, Dr.N.G.P.Institute of Technology, Coimbatore, India. jayasudha@drngpit.ac.in*
[6]*Department of Mathematics, Jeppiaar Institute of Technology (Autonomous), Sriperumbhudhur, Chennai, India. csenthilkumarmaths@gmail.com*

***Abstract:*** Electric vehicles (EV) are becoming more prevalent because they are good for the environment and don't cost much to run. One big problem with EVs, though, is that their batteries don't last long. There is a complete way to figure out how long Nickel Manganese Cobalt-Lithium Cobalt Oxide (NMC-LCO) batteries will still work after this study. The information used in this study comes via the Hawaii Natural Energy Institute consist of 15 different batteries that were put through over 1000 rounds of controlled settings. A method with several steps is used, starting with collecting data and preparing it, then choosing features and getting rid of outliers. The RUL forecast method is made with machine learning (ML) methods like Bagging Regressor, XG Boost, Cat Boost, Light GBM and Extra Trees Regressor. Feature value analysis helps find important factors that affect the health and lives of a battery. Statistical tests show that there are no missing as well as duplicate data points and getting rid of outliers makes the method more accurate. Not surprisingly, XG Boost turned out to be the best algorithm, making predictions that were very close to being correct. This study shows how important RUL forecast is for improving battery lifetime management, especially in electric car uses, to make sure that resources are used in the best way possible, costs are kept low, and the environment is protected.

Keywords: Electric Vehicle, ML, Regression, Prediction Method, Battery, Electric Vehicle.

## 1. Introduction

Electric vehicles (EV) mostly use lithium-ion (Li-ion) batteries, which are selected for their unique qualities, such as their high energy density, lack of memory effect, long lifespan, and ability to be charged and discharged in a variety of ways [1-3]. Even with these benefits, the auto business sees problems like changing weather, more pollution in the air from cars [4], and uncertainty in the supply lines for green energy [5-7]. The energy contained in electric vehicle batteries offers a viable resolution to environmental issues and ambiguities [8]. The decarbonization of the transportation industry relies on developments in and the widespread acceptance of electric vehicles including improved range, safety, and dependability [9-11]. However, the usage of Li-ion batteries has hurdles [12], such as capacity deterioration, environmental consequences, and difficulties in end-of-life organization.

The idea of RUL is very important in prognostic upkeep and dependability manufacturing [13]. It shows how long or how much use a part, device, as well as system is likely to have before it fails or stops working properly [14-16]. In the case of EV batteries, predicting RUL requires using machine learning methods that take into account a number of factors [17]. When an EV battery is used continuously for about several years, its capacity usually drops by about 12%, so this is a big problem [18]. It's hard to predict RUL and keep an eye on capacity loss [19], especially since Li-ion batteries lose capacity slowly over time as they are charged and discharged [20]. Battery management systems (BMSs, which) are in charge of these jobs.

Improving the lifetime of electric vehicle (EV) batteries requires the ability to predict the non-linear degradation of battery capacity [21]. When it comes to making longer-lasting batteries and charging procedures, machine learning (ML) provides useful tools for predicting battery life, helping owners plan their trips, and more [22]. To solve the complicated problems with battery performance, ML offers precise, scalable, and non-invasive solutions. Powering vehicles with electricity promotes efficient use of renewable energy sources and lowers overall travel costs [23-25]. Benefiting producers, users, and global sustainability, this research seeks to provide a dependable approach for forecasting the lifetime of electric vehicle batteries.

The main goal of this research is to guess the RUL for Li-ion batteries, which is a necessary job that has important real-world uses. Predicting RUL is very important for businesses that use Li-ion batteries a lot because it helps with planned repair and making good use of resources. The cycle index, discharge duration, and peak voltage discharge are crucial pieces of information that we have on hand to do this. With these characteristics, a robust and reliable prediction model may be constructed using the objective variable, RUL, which indicates the remaining battery life.

## 2. Methods and Materials

The proposed method for RUL prediction is shown in Figure 1 as a block diagram. Data gathering and preparation come first, followed by the selection of features and elimination of outliers. When optimizing hyperparameters using five-fold cross validation, the Grid Search CV technique is used. In order to create a model that can predict the RUL of a battery, we used the following machine learning algorithms: XGBoost, Bagging Regressor, LightGBM, CatBoost, and Extra Trees Regressor. After that, the model's efficacy is assessed using regression performance indicators. Lastly, the most important characteristics in the dataset are identified by doing a feature importance analysis.

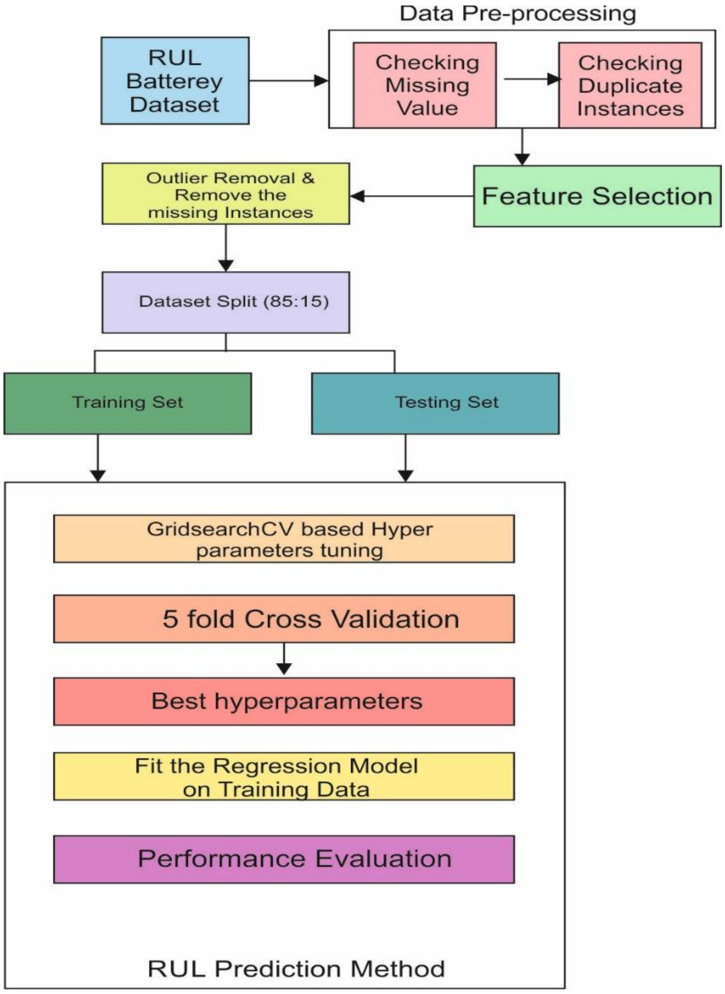

Figure 1 Proposed method for RUL prediction

2.1 Data

This data set is based on research conducted at the Hawaii Natural Energy Institute on fifteen NMC-LCO 18,700 batteries. Over a thousand cycles of conventional charging and discharging were performed on these batteries at 25°C. Features in the dataset monitor the voltage and current patterns during each cycle; these patterns are used to estimate the RUL of the batteries. Due to the fact that the cycling routine is identical to that of EV batteries, this data is particularly useful for comprehending the deterioration and lifetime of EV batteries.

2.2 Data Pre-Processing

Duplicate and missing value checks were performed on the data. Following preprocessing, it was noted that the dataset did not include any instances that were either missing or duplicated. The reliability of the dataset may be guaranteed by checking for duplicate or missing entries.

In particular, machine learning models are vulnerable to biassed studies and erroneous predictions caused by missing values. Every entry in the dataset will be distinct once duplicates are removed. Because duplicate data may impact the quality and efficiency of statistical analyses and machine learning method training, this is especially crucial to keep in mind. There were no occurrences that were identified to be missing or duplicated in this collection. There were 16,000 occurrences and 7 features in the dataset after preprocessing.

2.3 Selection of Feature

Figure 2 shows the RUL dataset's heat map. The correlation coefficient between the RUL and cycling index is -1, as seen by the heatmap. when discharge, the maximum voltage correlates with RUL at 0.78, whereas when charging, the lowest voltage correlates with RUL at −0.76. When compared to the dependent variable, RUL, other traits show weak connections. It is worth noting that the cycle index and RUL values are inversely related. As an example, the RUL is

1112 when the cycle index is 1, and vice versa when the cycle index is 1. The cycle index also becomes zero when the RUL equals 1113.

Figure 2 RUL Dataset's Heatmap

The negative connection between the two variables implies that the cycle index may be deceiving and not provide useful information about the battery's real health or remaining life. Overfitting may occur if the model learns this inverse connection too strongly when trained using the cycle indices as a feature. A model is said to have overfit when it becomes so used to the training data that it becomes less effective at generalizing to new, unknown data.

The method’s performance on real-world data could be affected if it places too much emphasis on the cycle index. This inverse connection might not be applicable or other variables might play a bigger impact. They decided to remove the periodic index feature from the framework to help with this.

The decision-making process takes into account any external influences or material deterioration that may impact RUL. Predictions might be off kilter if the model starts to rely too much on the cycle index and ignores these other relevant factors or outside influences. There were 16,000 occurrences and 7 features in the dataset after feature selection.

2.4 Outlier Removal

Outliers that damage batteries and health. Model predictions may be erroneous due to outliers skewing mean and standard deviation. Remove these to enhance model accuracy and RUL predictions. Noise from outliers makes it hard to spot battery lifespan and performance trends. Table 1 compares battery performance feature skewness and kurtosis before and after outlier reduction. The same goes for the discharge period, Time Constant Current, Charging Times, Decrement 3.6-3.4 V, and Duration at 4.15 V all showed extreme kurtosis and skewness prior to the elimination of outliers. The distributions of these measurements were right-skewed, characterized by big tails and abrupt peaks. After removing outliers, skewness values became negative or approached zero. More symmetrical distributions have skewness values around zero.

The feature distributions improved their balance and symmetry as a result of the removal, as shown by the lowered or negative skewness. Also, compared to their original values, the kurtosis values in the “After Removal of the Outliers” column were much lower. It may be inferred from this drop that the data is less peaked or tail heavy.

The normal distribution has larger tails when the kurtosis value is positive, and a distribution with smaller tails when the value is negative. The datasets that were subject to outlier reduction clearly exhibit flatter distributions with less extreme values or are more closely aligned with a normal distribution, as seen by these lower kurtosis values. Researchers developed an ML method using 14,500 examples with 7 characteristics after eliminating the outliers.

Table 1 Compares Battery Performance Feature Skewness and Kurtosis Before and After Outlier Reduction

| Feature | Before Outlier Removal | | After Removal of the Outlier | |
|---|---|---|---|---|
| | Skew | Kurtosis | Skew | Kurtosis |
| Discharge Time | 16.3001 | 340.994 | -0.155 | -1.207 |
| Decrement | 9.987 | 254.345 | 0.242 | -0.900 |
| Max. Voltage Discharge | -0.531 | 12.566 | -0.080 | -0.967 |
| Min. Voltage charging | 0.330 | 1.146 | 0.214 | -0.236 |
| Time | 16.239 | 341.629 | -0.107 | -1.207 |
| Time Constant Current | 24.771 | 697.545 | -0.139 | -1.172 |
| Charging Time | 22.771 | 588.791 | -0.126 | -0.655 |
| RUL (Cycles) | 0.007 | -1.209 | -0.013 | -1.203 |

## 2.5 ML Method Development

### 2.5.1 Splitting Data

There was an 85/15 split in the RUL dataset between the training and testing sets. One thousand two hundred and fifty examples were reserved for testing, whereas twelve thousand were used for instruction. The following characteristics were taken into account while developing the RUL prediction model: The following variables are measured in s: voltage drop from 3.6 to 3.4 V, discharge time, minimum voltage charging, maximum voltage discharge, time constant current, time at 4.15 V, and charge time. Improving the model's dependability and practicality, this information split is essential for testing it in different real-world contexts.

### 2.5.2 ML Method Selection

An important part of any machine learning process is picking the correct model architecture or algorithm. Try out a few various methods and see how they do on a test set; that way you can make an educated choice. Important steps include feeding the model training data and tweaking its parameters to make better predictions. A number of state-of-the-art ML algorithms were evaluated for the RUL prediction method. These included CatBoost, XGBoost, Extra Trees Regressor, LightGBM and Bagging Regressor.

### 2.5.3 K-Fold Cross Validation Hyper Parameter Optimization

Hyperparameter tuning is crucial for improving a machine learning method for RUL prediction by assuring the best possible settings for the provided battery dataset. Hyperparameters are constant, but parameters are trained with new data and adjusted accordingly. A well-known method called Grid Search CV was used to refine the RUL prediction method. In order to find the optimal model configuration, this approach use 5-fold cross-validation to methodically test different hyperparameter values.

### 2.5.4 Analysing RUL Regression Method Performance

Many metrics may be used to assess the efficacy of a regression approach. These include R-squared test the mean squared error (MSE), the root mean squared error (RMSE), and the mean absolute error (MAE). Given a collection of ‘n' instances, the formulae for computing these metrices are given, with $y_i$ denoting real values and $y_p$ in lieu of anticipated values. A measure of how much of the variation in the dependent variable can be predicted from the independent

variables is the R-squared value, which is obtained from the value of the coefficient of determination.
Equation (1) is used to calculate the MAE.

$$\mathrm{MAE} = \frac{|(y_i - y_p)|}{n} \qquad (1)$$

Equation (2) is used to compute the MSE.

$$\mathrm{MSE} = \frac{\sum(y_i - y_p)^2}{n} \qquad (2)$$

Equation (3) is used to compute the RMSE.

$$\mathrm{RMSE} = \sqrt{\frac{\sum(y_i - y_p)^2}{n}} \qquad (3)$$

The coefficient of determination, or R-Squared, is determined using Equation (4).

$$R^2 = 1 - \frac{\sum(y_i - y_p)^2}{\sum(y_i - \bar{y}_i)^2} \qquad (4)$$

In this case, $y_i$ is the average of all the real numbers.

## 3. Results and Discussion

Figure 3 shows the results of comparing the performance metrics of several ML procedures for predicting the RUL values of batteries. R-Squared, MSE, MAE and RMSE are among of the metrics that are assessed for the test and training datasets. Both training and testing sets show that the XG Boost algorithm predicts RUL more accurately than the others, with the lowermost possible RMSE and MAE values. Furthermore, the R-Squared values are quite high, indicating that the model fits the data almost well. Even though XG Boost has lower MAE and RMSE values, the Bagging Regressor method does a little better, particularly on the training set.
Nevertheless, it constantly displays high R-Squared values, which suggests a great capacity for prediction. In comparison to earlier models, the Light GBM approach achieves much higher MSE, MAE and RMSE values on both the testing and training sets. Still, with R-Squared values greater than 0.996, we have a solid predictive method. The Cat Boost method outperforms the prior models in terms of error measures, including MAE, MSE, and RMSE. A little drop in R-Squared values suggests a less accurate forecast when compared with other methods.
The error metrics for the testing and training sets are the highest for Extra Trees Regressor compared to the other methods that were assessed. There is less of an ideal fit between the method and the data, since the R-Squared values are marginally lesser than the another methods.

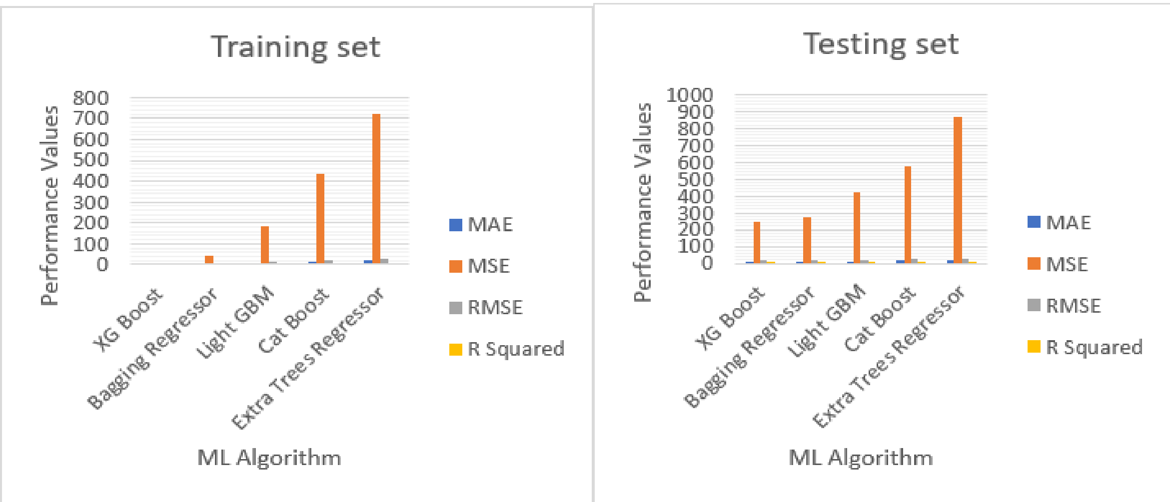


Figure 3 Comparing Performance Metrics of Several ML for Predicting RUL Values

The order of effectiveness for RUL prediction algorithms is as follows: Bagging Regressor, XGBoost, Cat Boost, Light GBM and Extra Trees Regressor. A comprehensive visual and analytical representation of the XGBoost ML algorithm's predicted RUL for batteries is shown in Figure 4.

Figure 4a shows a visual comparison of the batteries' actual RUL values with the values based on RUL anticipated by the XG Boost algorithm, in the form of an Actual Vs. anticipated Plot. If the predictions were flawless, every data point would fall on a diagonal line, showing that the expected and actual values were exactly same. If the data doesn't follow this line, it means the model's predictions are off. The discrepancies between the anticipated and real RUL values are seen in Figure 4b's residual plot.

Typically, the horizontal references line is located at y= 0, and the horizontal distance between each data point and this line shows the number and direction of the errors. A well-performing method's residuals should seem like they're scattered about this line, with no discernible trends. The RUL prediction model that is powered by the XGBoost algorithm is shown in

Figure 4c as the residual histogram. In order to better comprehend the distribution and frequency of the prediction mistakes, this histogram is provided. If the methods mistakes are uniformly dispersed and have no bias, then the data should follow a bell curve with zero as its center.

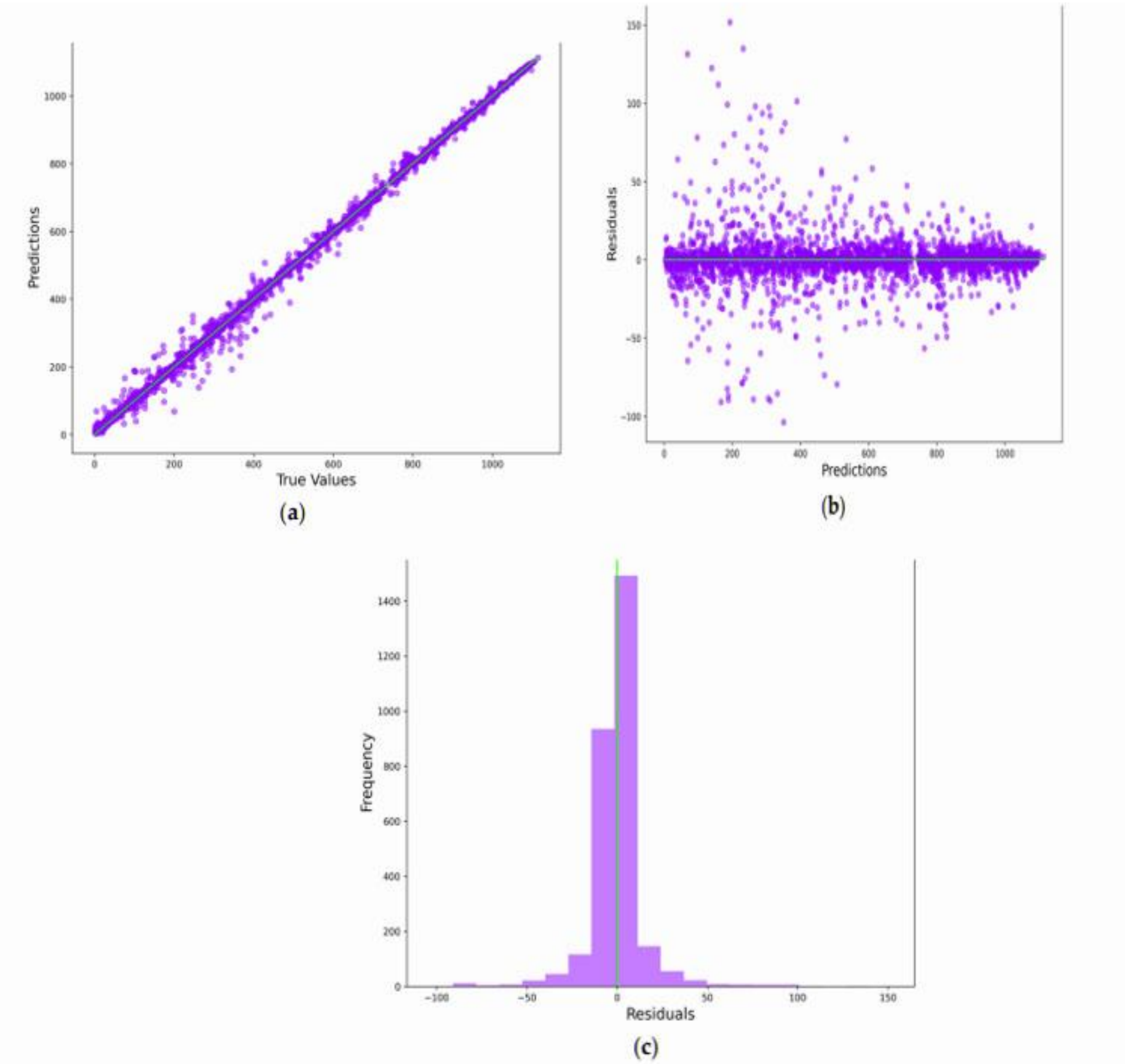


Figure 4 (a) The real vs. projected plot, (b) Residual plot, and (c) Residual histogram illustrating RUL prediction using the XGBoost ML method

A feature importance plot is shown in Figure 5. It demonstrates how several characteristics are significant in determining battery RUL levels. High relevance is indicated by a score near 1, whilst lesser significance is indicated by a value closer to 0. To make sure the method works effectively with new data, features with high significance ratings are important.
The feature with the best score of 0.8220 for RUL prediction between all of the provided characteristics is the "Time at 4.15 V" feature, which stands out. The "Time Constant Current" characteristic is very important, with a score of 0.0980. With a score of 0.0014, the "Decrement 3.6-3.4 V" characteristic is the least important of all the features given and makes very little contribution to predicting RUL.
There must be no biases or overfitting caused by high-importance characteristics for the model to be considered useful. It is possible to reduce the model's robustness and generalizability by placing too much emphasis on a particular feature, regardless of how predictive it is. That is why it is essential to take into account all aspects fairly, including the ones with lower ratings.
Improving electric vehicle infrastructure in the long run is one of the goals of this project. Businesses that rely on Li-ion batteries may save expenses and improve efficiency via proactive maintenance and improved resource management if RUL predictions are spot on.
The electric car industry is one that relies heavily on lithium-ion batteries. A precise prediction of the RUL is crucial for the implementation of appropriate preventative maintenance for these batteries. With accurate RUL projections, applications and companies can maximize resource use, plan maintenance at the optimal time, and save money and run more efficiently. Researching and predicting RUL also aids in making Li-ion batteries last longer, which is crucial for eco-friendly development because of the toll that disposing and replacing batteries has on the environment. Owners of electric vehicles are able to better organize their travels with the help of accurate RUL projections.
The field is well-suited to studies that examine how technology and business interact with one another. The efficiency of different machine learning methods in RUL prediction may be better understood by comparing them. In order to facilitate future research and development in relevant domains, it is necessary to identify top-performing algorithms such as XGBoost. This study suggests that reliable RUL projections might alleviate worries over the disposal of Li-ion batteries, taking into account their effects on the environment. Innovations in technology must be accompanied with sustainable habits if we are to reduce environmental damage.

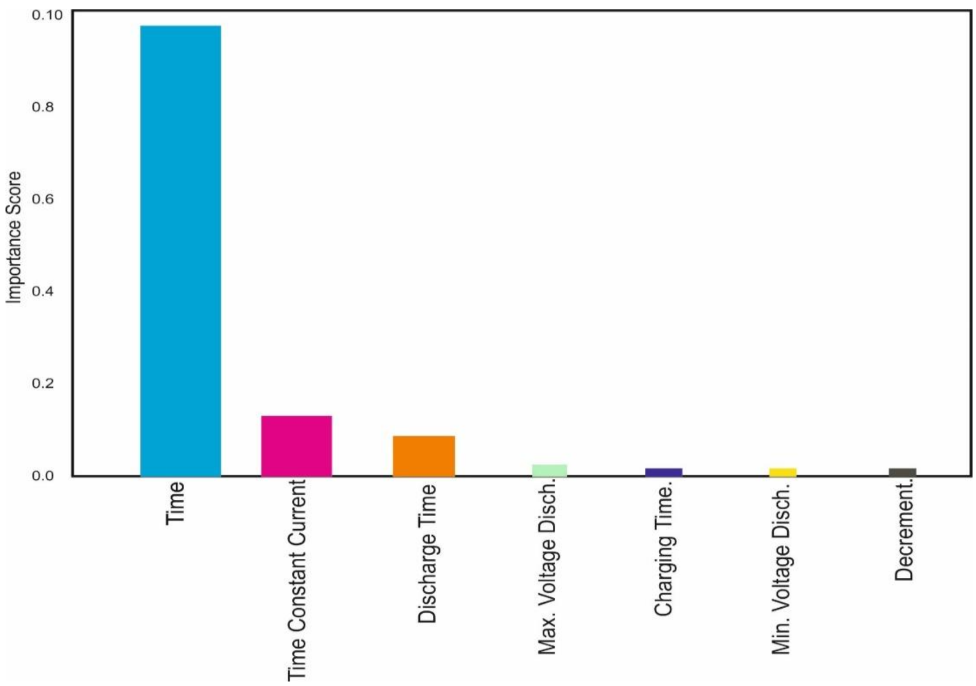

Figure 5 Feature Importance Plot

## 4. Conclusion

The significance of machine learning approaches is highlighted in this paper, which thoroughly investigates the difficulties of RUL value prediction for Li-ion batteries. The results showed that XGBoost was the best ML algorithm for RUL prediction out of all the ones evaluated. The results prove that XGBoost makes very precise and error-free predictions of RUL. Improving travel planning and contributing to the advancement of batteries having longer lifespans are two areas where these discoveries may be useful for both manufactures and users. This study helps to enhance electric car infrastructure in a sustainable fashion, which is a game-changer at the junction of technology and industry. It is reasonable to assume that precise RUL prediction helps to prolong battery longevity, which in turn reduces environmental problems related to battery disposal, given the emphasis on bearable growth and the environmental significances of electric vehicles. Businesses that use Li-ion batteries may be able to save money and run more efficiently if their RUL predictions are precise enough to allow for proactive maintenance and optimizing resources. Since data availability and quality may have a major effect on how well machine learning models perform, these factors are one of the research's limits. Another possible limitation of this study is that it does not consider how different environmental factors affect battery performance. Temperature, humidity, and use patterns are factors that may impact RUL, but they were not specifically mentioned.

## References


1. Karthick, K., S. Ravivarman, and R. Priyanka. "Optimizing Electric Vehicle Battery Life: A Machine Learning Approach for Sustainable Transportation." World Electric Vehicle Journal 15.2 (2024): 60.
2. Zhang, Xinfang, et al. "A review of machine learning approaches for electric vehicle energy consumption modelling in urban transportation." Renewable Energy 234 (2024): 121243.
3. de la Torre, Rocio, et al. "Simulation, optimization, and machine learning in sustainable transportation systems: models and applications." Sustainability 13.3 (2021): 1551.
4. Suanpang, Pannee, and Pitchaya Jamjuntr. "Optimal Electric Vehicle Battery Management Using Q-learning for Sustainability." Sustainability 16.16 (2024): 7180.
5. Naresh, Vankamamidi S., Guduru VNSR Ratnakara Rao, and D. V. N. Prabhakar. "Predictive machine learning in optimizing the performance of electric vehicle batteries: Techniques, challenges, and solutions." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery (2024): e1539.
6. Ullah, Irfan, et al. "A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards sustainability." Energy & Environment 33.8 (2022): 1583-1612.
7. Zhao, Jingyuan, et al. "Machine learning for predicting battery capacity for electric vehicles." ETransportation 15 (2023): 100214.
8. Sultan, Yara A., et al. "Enhancing electric vehicle battery lifespan: integrating active balancing and machine learning for precise RUL estimation." Scientific Reports 15.1 (2025): 777.
9. Ermon, Stefano, et al. "Learning policies for battery usage optimization in electric vehicles." Machine learning 92 (2013): 177-194.
10. Das, Kaushik, Roushan Kumar, and Anurup Krishna. "Analyzing electric vehicle battery health performance using supervised machine learning." Renewable and Sustainable Energy Reviews 189 (2024): 113967.

11. Punyavathi, R., et al. "Sustainable power management in light electric vehicles with hybrid energy storage and machine learning control." Scientific Reports 14.1 (2024): 5661.
12. Sabzi, Shahab, and László Vajta. "Optimizing Electric Vehicle Charging Considering Driver Satisfaction through Machine Learning." IEEE Access (2024).
13. Chen, Yukai, et al. "Electric vehicles plug-in duration forecasting using machine learning for battery optimization." Energies 13.16 (2020): 4208.
14. Manuguerra, Luca, Federica Cappelletti, and Michele Germani. "A machine learning based method for parametric environmental impact model for electric vehicles." Journal of Cleaner Production 454 (2024): 142308.
15. Singh, Chaitanya, et al. "Applied machine tool data condition to predictive smart maintenance by using artificial intelligence." International Conference on Emerging Technologies in Computer Engineering. Cham: Springer International Publishing, 2022.
16. Mousaei, Arash, Yahya Naderi, and I. Safak Bayram. "Advancing state of charge management in electric vehicles with machine learning: a technological review." IEEE Access 12 (2024): 43255-43283.
17. Mazhar, Tehseen, et al. "Electric vehicle charging system in the smart grid using different machine learning methods." Sustainability 15.3 (2023): 2603.
18. Pathmanaban, P., et al. "Optimizing electric bike battery management: Machine learning predictions of LiFePO4 temperature under varied conditions." Journal of Energy Storage 99 (2024): 113217.
19. Rajesh, P. K., T. Soundarya, and K. V. Jithin. "Driving sustainability-The role of digital twin in enhancing battery performance for electric vehicles." Journal of Power Sources 604 (2024): 234464.
20. Pandey, Purnendu Shekhar, et al. "A smart parking system based on IoT technologies." Linguistica Antverpiensia (2021): 59-64.
21. Lin, Shih-Lin. "Deep learning-based state of charge estimation for electric vehicle batteries: Overcoming technological bottlenecks." Heliyon 10.16 (2024).
22. Jui, Julakha Jahan, et al. "Optimal Energy Management Strategies for Hybrid Electric Vehicles: A Recent Survey of Machine Learning Approaches." Journal of Engineering Research (2024).
23. Recalde, Angel, et al. "Machine Learning and Optimization in Energy Management Systems for Plug-In Hybrid Electric Vehicles: A Comprehensive Review." Energies 17.13 (2024): 3059.
24. Priya, S., D. Gunapriya, and M. Batumalay. "Novel Battery Management with Fuzzy Tuned Low Voltage Chopper and Machine Learning Controlled Drive for Electric Vehicle Battery Management: A Pathway Towards SDG." Journal of Applied Data Sciences 5.3 (2024): 936-947.
25. Pujari, Pradyumna, Lakshya Rubi Gnanasambanthan, and Prabhav Sharma. "Data-Driven Battery Health Optimization in Electric Vehicles: A Machine Learning Approach." 2024 IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET). IEEE, 2024.