International Journal of Computing and Digital Systems
2025, VOL. 17, NO. 1, 1–10

http://dx.doi.org/10.12785/ijcds/1571031381

Advancing Lithium-ion Battery Management: A Comprehensive Approach for Enhanced Remaining Useful Life Prediction

Sthitprajna Mishra[1], Chinmoy Kumar Panigrahi[1], Subhra Debdas[1], M Disha[2], Komal Jaiswal[2], Ramesh Chandra Khamari[3] and Bijay Kishor Shishir Sekhar Pattanaik[4]

[1] *School of Electrical Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India*
[2] *School of Computer Application, KIIT Deemed to be University, Bhubaneswar, Odisha, India*
[3] *Electrical Engineering, Government College of Engineering, Keonjhar, Odisha, India*
[4] *School of Computer Engineering,GITA, Bhubaneswar, Odisha, India*

Received 17 May 2024, Revised 11 January 2025, Accepted 14 January 2025

Abstract: Accurately predicting the Remaining Useful Life (RUL) of lithium-ion batteries is crucial for optimizing battery management systems, ensuring reliable performance, and maximizing operational efficiency. This paper presents an advanced approach using a Random Forest Regressor (RFR) combined with sophisticated feature extraction techniques to enhance the accuracy and reliability of battery lifespan predictions. The methodology involves extracting a comprehensive set of features from battery degradation data, carefully selected to capture various aspects of battery health and performance. These features provide a holistic understanding of the battery's condition. Data visualization tools are utilized to aid in the interpretation of these features, allowing stakeholders to gain actionable insights from the prediction results. By integrating RFR with these advanced feature extraction techniques, the proposed approach significantly improves battery RUL predictions. The ensemble learning capabilities of RFR, coupled with the richness of the extracted features, enable the model to capture complex relationships within the data, leading to more accurate and reliable lifespan predictions. This work has practical implications beyond academic interest, offering substantial benefits for improving battery management strategies and enhancing overall system reliability. More precise RUL predictions allow stakeholders to plan maintenance schedules proactively, optimize resource allocation, and mitigate risks associated with battery degradation. This ultimately contributes to prolonged battery lifespans, reduced downtime, and improved operational efficiency across various applications, including electric vehicles and renewable energy storage systems. This study demonstrates the effectiveness of the Random Forest Regressor (RFR) model, achieving an RMSE of 0.048, MAE of 0.032, and R^2 of 0.92, which significantly enhances operational efficiency, reduces downtime, and enables proactive maintenance across applications like electric vehicles and renewable energy storage.

Keywords: Artificial Neural Network, Random Forest Regressor, Remaining Useful Life, Root Mean Square Error

1. INTRODUCTION

In contemporary discourse, the global community is increasingly aware of the complex challenges posed by declining air quality and rising temperatures These aspects of the environment are caused mainly by human beings; therefore, intensive and communal measures have to be adopted to address the consequences of these problems. From the options provided below as the strategies to solving these concerns, increasing EV efficiency is one of the solutions viewed to solve the problem of emission of greenhouse gases, enhancing poor air quality. The highly important problem in battery management systems is the accurate estimation of Lithium-Ion Battery Remaining Useful Life. Since these batteries form the core of current electric transportation and renewable power storage systems, it is of uttermost importance to enhance their performance, specifically – their efficiency and durability. However, with the development of new methods for testing batteries, the original techniques of assessing RUL are insufficient, especially because of their inability to consider nonlinear degradation processes in the battery systems. This limitation does not only impact the reliability of predictions but also is troublesome in fields where battery failure results in system turbulence, often high risks to safety, and high costs of downtime. For instance, the accidental immolation in electric automobiles could put the passengers and fellow motorists in harm's way or a fire that brings blackouts to an electric vehicle battery from renewable systems cut in renewable energy storage appliances. Thus, the demand for precise, reliable and flexible methods for determining battery life has aroused interest in both research and industry. The value of this study involves its capacity to resolve these issues in a study by integrating RFR with improved feature extraction procedures. The objective of this research is to

E-mail address: sthitprajnamishra26@gmail.com, panigrahichinmoy@gmail.com, subhra.debdas@gmail.com, manvidisha0803@gmail.com, komaljaiswal8984@gmail.com, rameshkhamari$_f$ee@gcekjr.ac.in, bkssp@hotmail.com

improve the RUL prediction of LiBs by applying the RFR model jointly with feature extraction techniques. Because the traditional degradation models fail to give accurate predictions due to the complex relationships present in battery health data, the proposed approach in this paper is effective in handling the nonlinear degradation processes, and thus the problems faced in BMS are solved with enhanced predictive performance and real-world applications. Thus, using the ensemble learning capacity of RFR, the presented approach intends to improve battery lifespan predictions' reliability and accuracy. The methodology involves extensive data analysis of battery degradation data thereby making it possible to identify salient features mostly sign that capture different aspects of battery health and its performance. Such an approach ensures that stakeholders are in a better position to make appropriate decisions when it comes to the aspects of maintenance, resources and operations.[1]

A. *Lithium-Ion Batteries in Electric Vehicle System*

Essential to improving the efficiency of EV is Lithium-ion (Li-ion) batteries that remain popular due to its high energy density, long cycle life and comparatively lighter weight than other battery systems. It is these batteries that form the basis of the modern electric mobility. Energy density of a batter is defined as amount of energy per unit mass or volume and for the use in electric vehicles is very critical. This means that the higher the energy density the longer time the vehicle can travel on a single charge hence making it more competitive with counterpart ICEVs.[2] However, despite the fact that Li-ion batteries have become very crucial in most industrial applications, there is still a major problem in enhancing the efficiency and service life of these batteries. Temperature fluctuations, current drain profiles, DoD or the extent to which energy is employed before recharging affect the battery or cause its deterioration. These, over time, result in deterioration of battery capacity and efficiency destroying the vehicle's range and a higher tendency to develop faults. This is where estimation of the Remaining Useful Life (RUL) of batteries becomes very important. The ability to RUL predicts serves critical roles of evaluating battery quality, planning for its use, and security.[3] A precise definition of a battery's lifetime makes it possible for stakeholders to make sound decisions on battery maintenance, components' procurement, and replacement schedules, and thus augment the effectiveness and reliability of systems that require batteries. For instance, users of electric buses or delivery vehicles such as fleet operators can fix timings for maintenance hence eliminating chances of breakdowns or costly repairs. Likewise, renewable energy systems can incorporate battery replacement schedules proactively and this will help in preventing disruptions of energy supplies. Accurate predictions of RUL enable scheduling of maintenance tasks and ensures the systems do not breakdown frequently thus making the operations safer and cheaper.[4] In further fashion, traditional approaches for managing dependence betwixt vigorous and nominal covariates are also ineffective in RUL prediction. In fact, there is gradually evolving improvement in battery technology but the conventional approaches of estimating RUL proved to be incompetent in handling nonlinear form of battery degradation. CNFs of the battery models have been utilizing conventional models, where EM are normally used or simple analytical equations based on some fairly assumptions concerning the battery reaction are employed. Although these methods may be effective under laboratory like conditions, they hardly give the true picture of actual operating conditions. For example, battery used in electrical vehicles is subjected to varying conditions such as temperature, charging frequency as well as usage frequency which affect the rate of battery degradation.[5] Where these conventional approaches leave off is the fact that battery health is dependent to several factors that only these complex models can capture. The degradation process in lithium-ion batteries is therefore highly nonlinear so that even slight changes in operating conditions have drastic effects on the life of such batteries. Also excluded in these traditional models is the way the factors may interact and thus, for instance, high temperatures in conjunction with fast charging could make degradation go faster than either factor could by itself. As a result, they might not accurately predict the trend and thus they are not reliable for actual usage.[6]

B. *Recent Techniques for Estimating Remaining Useful Life*

A need arises to apply complex methods which employ more complex algorithms in the prediction of RUL with attempts at enhancing battery management strategies. Another way of enhancing battery efficiency is by developing machine learning algorithms to predict Li-ion battery's lifespan. These algorithms take into account the historical data on usage of battery and include the temperature, number of cycles undertaken through charge-discharge and internal impedance which provides a good prognosis for the battery.[7] For instance, smart sites application that use machine learning algorithm work from large amount of data on battery performance and hence give more better recommendations. With this data these models are able to predict on the longevity of a battery in terms of its ability to perform optimally as well as when it would require recharging or replacement. This analysis can be used to maintain, control or even extend battery performance and life span while also planning the most appropriate time to maintain or replaced them.[8]

C. *Random Forest Regressor (RFR) and Benefits in Our Study*

Of all the machine learning methods, the Random Forest Regressor (RFR) has proven to be the most effective for Battery RUL prediction. RFR is another method in the ensemble learning where several decision trees are created during the training process and the results of all the created trees are combined in order to provide high accuracy and diminish the overfitting issue.[9-11] This is particularly useful especially when it comes to RUL prediction of batteries because batteries degrade due to influence of many factors and these factors often have complex relationships that are very hard to model with a simple linear model. By virtue of this, the RFR model has the ability of capturing such relationships, and the model is way much more flexible and robust compared to any linear model that might be in place in the social sciences.[12] The advantage of the RFR is in the ability to combine square forecasts originating from a myriad of decision trees, making it easy to identify complex nonlinearities in the data. In the Random Forest case each tree in the forest makes a prediction on the basis of the samples and the final prediction is a mean of all of them. This combined approach reduces the overfitting problem and increases model performance when tested on unseen data. With this capability and in addition to the richness of feature extraction used to feed the model, more accurate and dependable lifespan prediction is made.[13]

D. *The Importance of Accurate RUL Predictions in Terms of the Socio-Economic and Environmental Consequences*

It is critical to remember that, besides enhancing technical value, accurate RUL predictions provide numerous other advantages. Better battery management not only increases battery durability, decreases operational expenses, but furthermore promotes consumers' EVs trust. The biggest potential downside for those who are thinking about switching to EVs is the unknown of how long the batteries are going to last and the cost of battery replacement. Nonetheless, as the consumers gain confidence in the battery capacity and reliability, the acceptance of the electric vehicles is expected to increase.[14] This change will mark a positive socio-economic impact in any nation or geographical location that it is implemented. For instance, the shift toward electric cars will help to decrease the emission of greenhouse gases, especially in urban areas where automobile emissions pose a significant threat to the environment. Environmental aspects will benefit from increased use of evs because emissions will decrease, and this will significantly affect public health since respiratory diseases and pollution-related diseases will decrease. The use of EVs also supports global strategies to mitigate the effects of climate change through the reduction of the use of fossil energy sources. Nonetheless, there are certain issues remaining even when we use more sophisticated machine learning models such as the RFR. Fluctuations in battery degradation rates caused by environment influences, different charging protocols, and usage stress are quite problematic for precise RUL determination.[15] This demonstrates the importance of having accurate and flexible predictive models to capturing the varying characteristics of battery degradation under different usage profiles.[16] Going forward, one will have to develop more sophisticated models that incorporate the latest battery technologies as well as enhance the existing models in extreme cold and hot conditions. Further, the incorporation of these models with the IoT devices is also possible where wirelessly connected batteries' health can be managed concurrently and the quality of the simulated RUL will be boosted.[17]

In our comprehensive study, we aimed to enhance the prediction accuracy of lithium-ion battery life cycles by meticulously extracting and analyzing a diverse set of features. This effort was crucial in developing a predictive model that could reliably estimate the Remaining Useful Life (RUL) of batteries, a key factor in advancing battery management systems.

Our methodology centered around the construction of a Random Forest Regressor (RFR) model, specifically tailored for predicting battery life cycles based on the features we extracted. The RFR model was chosen for its robust performance and capability to handle complex datasets, making it an ideal candidate for this application. We paid particular attention to the selection and extraction of features, ensuring that our dataset was rich and comprehensive, thereby maximizing the model's predictive accuracy.

A significant part of our study involved investigating the impact of tuning model hyperparameters on the precision of our predictions. By carefully adjusting parameters such as the number of trees in the forest, the maximum depth of each tree, and the minimum samples required to split a node, we were able to enhance the model's performance. This optimization process was critical in ensuring that our model not only predicted battery life cycles accurately but also performed consistently under varying conditions.

Furthermore, we conducted an in-depth analysis of the relative importance of different input features in forecasting battery life cycles. This analysis was essential for enhancing the interpretability of our model, allowing us to understand which features had the most significant impact on the predictions. By identifying the key predictors, we could provide valuable insights into the factors that most influence battery longevity.

Lastly, we benchmarked the performance of our RFR model against other advanced machine learning techniques that are currently recognized as state-of-the-art in this domain. This comparison was vital in validating the effectiveness of our approach and ensuring that our model stood up to the highest standards in predictive accuracy and reliability. Our results demonstrated that the RFR model, with its optimized feature set and hyperparameters, performed competitively, marking a substantial contribution to the field of battery life cycle prediction. The next section examines the current literature to present the past approaches and limitations in predicting the lithium ion battery (LIB) Remaining Useful Life (RUL) for comprehensive understanding of current methodologies.

E. *Related Work*

The estimation of Remaining Useful Life, RUL, and State of Health, SOH, of lithium-ion batteries are significant for the battery management systems and battery safety besides prolonging the lifespan of the batteries. The literature shows that this task has been addressed in many ways with different approaches and techniques from model based to use of advanced algorithms in machine learning. Nonetheless, several concerns are still present which involve uncertainty in battery behaviors, RTMS implementation and how multiple data stream can be effectively merged for better predictive models.[18] Wang et al. [9] developed a fractional-order model based approach which dealt with state estimation of hybrid power source system including Li-ion battery and ultra capacitor. Their approach included load trajectory factors in order to improve the battery state estimation. Based on this, Wang and Chen [10] proposed a framework based on unscented particle filter or UPF for predicting the SOC and the remaining discharge time of Li-ion batteries. This approach incorporated the sophisticated filtering processes into the models of the battery system with the hope of improving the prediction performance when the uncertainties in the battery are considered. These model based approach to state estimation has been shown to provide better estimates than the classic methods and their limitations include the fact that they rely heavily on system models and may not perform well in highly noisy or under conditions where some data is missing. Using the neural network approaches and based on machine learning techniques, Qu et al. [19] used methodologies for RUL estimation as well as SOH of lithium-ion batteries. Their work proves how techniques such as neural networks are useful to capture complex battery degradation profiles highlighting on the possibilities of achieving accurate representations of RUL. Models such as Neural Networks (NNs) and Gaussian Processes (GPs) are outperformed by RFR in many key areas, which make the RFR particularly effective for lithium-ion battery RUL prediction. Unlike GPs, RFR intrinsically handles nonlinearity with its ensemble of decision trees and can do so without complex preprocessing or assumptions of data distribution. Moreover, it is robust to missing data, since it uses available values for decision splits, contrasting with NNs and GPs that need imputation or are poorly behaved. Moreover, RFR avoids overfitting through its averaging output across trees, resulting in improved generalization potential thanNNs which are susceptible to overfitting without rigorous regularization. RFR offers clear feature importance, and makes NNs more interpretable and practical. Furthermore, RFR is computationally cheap and scales well, and hence it is more applicable to real time applications than GPs, whose efficiency degrades for large datasets due to their computational overhead. However, these advantages make RFR an attractive candidate for RUL prediction because it is robust and practical. Neural networks are powerful tools for learning from data but they can be slow learners depending on the nature of data and can be strongly overfitted if not properly handled. In the same vein, Ayob et al. [15] expanded on estimation techniques of SOC, SOH, and RUL in supercapacitor management systems, stressing on possible measures for implementation as well as realistic limitations of such systems [20-22]. lithium-ion batteries. Other techniques which have been widely used in battery RUL prediction include regression-based approaches. The following are the researched papers: [14], who used Gaussian process regression in order to estimate SOH and RUL through IHI(Indirect Health Indicators). They were able to demonstrate potential of different complex regression models for modelling non-linear relationships and improving the predictability than simpler linear models. However, there is an issue with scalability for the large number of data points since it has a growing time complexity proportional to the number of points in particular.In Paper [23-27] adopted the hybrid data-driven methods for SOH and RUL predictions elaborate frameworks using extreme learning machines. These methodologies sought to fuse multiple data sources and be able to handle fluctuations in batteries' degradation trends in order to produce reliable prediction results [26]. At the same time, these methods themselves have the limitations in generalization of the results for different types of battery chemistries and battery usage conditions. As for the future IoT application, battery health monitoring system based on Internet of Things have raised interest. Similarly, Patil and Kendule [26] employed of Artificial Neural Networks (ANN) for the RUL of such systems to put an instance on the capability of IoT integration toward real-time battery monitoring. Real-time data and collecting data, and incorporating superior machine learning models can enable better prediction and pro activity in terms of maintenance and management. However, with regard to issues of scalability, latency and data security they are still the problems which are to be solved to make use of IoT-based solutions widespread.[28] Nangare et al. [2] also carried out a similar vast study related to data analysis to extract features for EV battery, while focusing on the various features of battery for improved prediction accuracy and reliability [10]. Review Expansion and Knowledge Hitches Although research in prediction of RUL and SOH for lithium-ion batteries has improved in the recent past, some areas still need more research to be filled. Most of them include modeling based approaches (Wang et al. [9]) and filters (Wang and Chen [10]) which require systematic assumption and model which may not be possible for all battery types and working conditions.Additionally, purely data-driven approaches like neural networks (Qu et al. [19]) often require large training datasets and can suffer from overfitting or poor generalization to new data. Furthermore, regression-based methods (Jia et al. [22]) and hybrid models (Gou et al. [25]) face challenges in scalability and may struggle to adapt to highly variable degradation patterns across different battery applications.

The lack of real-time predictive models that can adapt to dynamic operating environments is a critical limitation in the current state of research. Although IoT-based approaches (Patil and Kendule [26]) offer potential solutions, these systems must address the challenges of real-time data processing, latency, and system integration to be effective in practical applications.

A comparative analysis of the findings and the study's major contributions Our proposed approach differ from the state-of-the-art techniques mentioned above for using a Random Forest Regressor (RFR), a machine learning model that can effectively analyze high-corespondence features and relationships. As distinguish from other types of models, such as neural network, Gaussian process regression, the RFR for example is capable of handling missing data and is less sensitive to overfitting even when using small data sets. Furthermore, the RFR model provides the feature importance that enables understanding which variables that significantly impact battery degradation; such information may be difficult to parse in a different machine learning methods, such as deep neural networks. Furthermore, by using polynomial features and interaction terms we are able to expand the features space and get a better capture of nonlinear relations between battery health indicators than the linear based methods. One of the limitations of previous work is that most studies have investigated the prognostic performance of a single battery degradation data source. The same is true for cross-validation methods and real-time data processing, placing our method at a logarithmic advantage over the usage of IoT-based models. The three identified areas of feature extraction, explainable outputs, and real-time performance form a research gap that ties high-end theoretical development to practical implementation of Battery Management System. The next section provides insights from the literature, and uses such insights to inform the methodology applied in this work, including how data was collected, preprocessed, features extracted and models developed to improve the accuracy of RUL prediction.

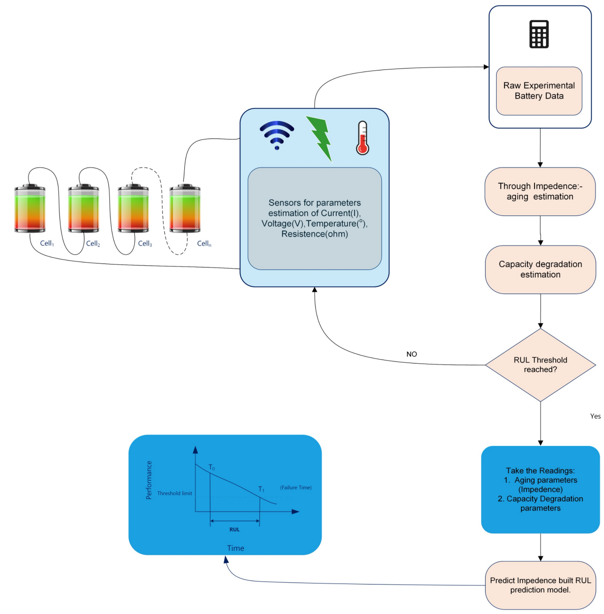

Figure 1. Proposed Model

2. PROPOSED MODEL

The method outlines how Remaining Useful Life (RUL) of lithium ion batteries is estimated through state of the art machine learning techniques. In this approach, we combine robust data acquisition, preprocessing, feature engineering, modeling and evaluation to provide accurate and trustable predictions, which contribute to optimization of decision making in battery management. While critical for modern applications, lithium ion cells present risks such as fires, explosions, functionality degradation, and high operational costs, of which some aspects arise from their cyclic life of less than 500 cycles, and rate of degradation. Challenges for addressing these challenges first require selection of suitable datasets so that effective battery management techniques can be robustly developed. Battery performance parameters were obtained from publicly available datasets that include different battery models and battery operating conditions from the organizations and research institutions. These datasets were standardized by imputing missing values, normalizing features and removing outliers with Zscore analysis. Feature engineering involved extraction of critical attributes including voltage drop, capacity fade and temperature variation; RFE and feature importance analysis were used to identify important predictors. In order to account for non-linearity, polynomial features and interaction terms were included. Statistical imputation was used to address the problems of missing data, dataset imbalance, and computational time, with techniques such as SMOTE and dimensionality reduction methods to address challenges such as dataset imbalance, and high computational time. Using the Random Forest Regressor (RFR) model to get accurate and interpretable RUL predictions that are robust in real-world battery management challenges, this comprehensive approach provides a solution of choice. The process for collecting data for the development of Li-ion battery RUL prediction models is depicted and shown in Figure 1, which involves choosing battery cells and extraction of accurate parameter data for generating health indicators for prediction.

3. METHODOLOGY

A. *Data Collection and Preprocessing*

In the first step, battery operational data pertaining to the battery under-observation are acquired in a structured way from plethora of available sources in order to have an extensive data sample depicting the nature of battery's behavior. This dataset includes the following important characteristics: discharge time, voltage and other characteristics. The battery degradation data that was used in the current work was obtained from publicly accessible datasets, and input from research institutions and other industrial entities. The dataset is composed of around less than one lakh data points in multiple types of battery models.This study utilized a dataset in the region of 100,000 data points drawn from publicly available data repositories deriving from research institutions and industry. Key battery performance parameters including voltage, current, temperature, number of cycles and capacity fade were included, over a range of operational conditions and for several battery models. For analysis, Python was used, libraries such as Pandas and NumPy for the data preprocess, Scikit-learn for the RFR implementation and feature importance analysis, Matplotlib and Seaborn for data visualization, and Imbalanced learn for SMOTE oversampling with class weighting from scikit learn. With these tools, we had a robust and efficient pipeline to make RUL predictions accurately. The preprocessing steps involved handling missing values through imputation techniques, normalizing feature scales, and encoding categorical variables as necessary. Additionally, outliers were identified and handled to improve the robustness of the model.

B. Feature Extraction

Feature extraction is important in enhancing the accuracy of RUL predictions where only useful features are chosen and transformed from the enhanced data set. The voltage drop and the capacity fade, temperature differences and discharge rate selects from the set of specific features for the given analysis. These features are known to have a direct relation to the health and performance of the battery and have been seen to cause considerable difference in the lifespan of the battery.
For instance, the voltage drop gives a measure of whether the battery is capable of maintaining voltage across the load while the capacity fade gives a clue from which extent the battery is capable of holding a charge after several cycles of charging and discharging. Temperature fluctuation is important in obtaining thermal impact on battery degradation while discharge rate over the time can assist in studying wear pattern in the battery cell. More complex approaches that used to feature engineering were applied to search for polynomial features and interaction terms of these variables, which determine the nonlinear nature to improve the accuracy of the RUL model.

C. Model Development

RFR model is built and trained based on the curated dataset with significant focus on architecture and hyperparameters of the model. The RFR algorithm selected as the best performer in the experiment due to its efficiency in case of managing complicated relationships and a high number of features is capable of addressing the problem of battery life cycle forecasting in consideration of the extracted features. For model training, the following hyperparameters were tuned: the number of estimators (trees), maximum depth of each tree, minimum samples required to split a node and criteria for split. A hyperparameter tuning was performed where the search space within the following parameters was explored. We chose our final model configuration to contain 200 trees, of maximum depth 30, and minimum samples per leaf set of 2. During the training phase, 80% of data were used for training while 20% were used for testing with 5 fold cross validation to curb on over fitting of the model.

D. Model Evaluation

The efficacy of the developed RFR model is rigorously assessed using a comprehensive set of performance metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). These metrics provide a detailed evaluation of the model's ability to accurately predict the battery's remaining life cycle. The cross-validation results indicate an RMSE of 0.048, MAE of 0.032, and an R^2 score of 0.92, suggesting that the model provides highly reliable predictions for the RUL of lithium-ion batteries.

E. Interpretation and Insights

Beyond model evaluation, the results are interpreted to derive actionable insights and recommendations for battery management and optimization strategies. Visualizations, such as feature importance plots and prediction vs. actual performance comparisons, are generated to facilitate a clear understanding of the model's predictions and highlight areas of interest. The subsequent section presents the results and analysis based on the methodological framework and highlights the performance metrics and predictive capability of the Random Forest Regressor model of capturing battery degradation patterns.

4. Result Analysis

The Random Forest Regressor (RFR) model has demonstrated promising accuracy in predicting the life cycle of lithium-ion batteries. By utilizing the provided dataset, the model effectively captures the intricate nonlinear dynamics associated with battery degradation, thereby providing reliable estimations of the remaining useful life (RUL) of the batteries.

A key aspect of this study is the analysis of the correlation matrix, as depicted in Figure 2. This matrix is derived from the battery dataset and offers valuable insights into the relationships between various variables. Each cell in the matrix represents the correlation coefficient between two features. A coefficient approaching 1 signifies a robust positive correlation, indicating that as one feature increases, the other feature also tends to increase. Conversely, a coefficient nearing -1 indicates a strong negative correlation, suggesting that as one feature increases, the other feature tends to decrease. A coefficient close to 0 suggests minimal to no correlation, implying that the features do not have a significant linear relationship.

Analyzing the correlation matrix is crucial for understanding the interdependencies among different features. This understanding aids in the selection of relevant features and the construction of a more effective model. By identifying which features are closely related, researchers can refine the feature selection process, enhancing the model's ability to accurately predict battery life cycle and health.

Overall, the integration of the RFR model with a detailed analysis of the correlation matrix facilitates a deeper understanding of battery degradation patterns, leading to more precise and reliable RUL predictions. This approach ultimately supports improved battery management strategies and enhances the overall reliability and efficiency of systems relying on lithium-ion batteries. Table 1. shows the result obtained by the RFR model.

Random Forest Regressor: The Random Forest Regressor (RFR) is an ensemble learning method that

Metric	Value
RMSE	0.048
MAE	0.032
R^2 Score	0.92

TABLE I. Evaluation Metrices

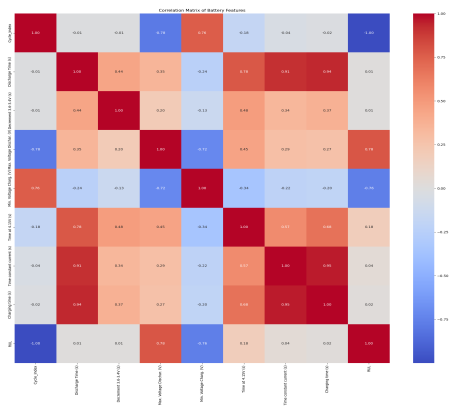

Figure 2. Correlation Matrix

operates by constructing a multitude of decision trees during training and outputting the mean prediction of the individual trees for regression tasks. The prediction $\hat{y}_i$ of the RFR for a given sample i can be represented as:

$$\hat{y}_i = \frac{1}{N} \sum_{j=1}^{N} f_j(x_i) \tag{1}$$

where:
$\hat{y}_i$ is the predicted value for sample i.
N is the number of trees in the forest.
$f_j(x_i)$ is the prediction of the j^{th} tree for sample i.

A. Model Performance Metrics

The performance of the RFR model is evaluated using the following metrics:

1) Root Mean Squared Error (RMSE):

RMSE quantifies the differences between predicted and actual values, emphasizing larger errors due to its square nature. The RMSE is computed using the formula:

$$RMSE = \sqrt{\left(\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right)} \tag{2}$$

2) Mean Absolute Error (MAE):

MAE measures the average magnitude of errors between predicted and actual values, providing a linear representation of errors. The MAE is calculated as:

$$MAE = \frac{1}{n} \sum_{i=1} |y_i - \hat{y}_i| \tag{3}$$

3) R-squared (R²) Score:

R² represents the proportion of the variance for the dependent variable that's explained by the independent variables in the model. It is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y}_i)^2} \tag{4}$$

where:
y_i is the actual value for sample i.
$\hat{y}_i$ is the predicted value for sample i.
$\overline{y}_i$ is the mean of actual values.

5. Visualization and Interpretation

To augment the interpretability and comprehensibility of the model predictions, an interactive dashboard was developed using Dash, a Python framework for building analytical web applications. The dashboard provides intuitive visualizations of the predicted cycles, actual cycles, and cycles left for each battery in the dataset. Additionally, scatter plots are incorporated to depict the relationships between specific battery parameters and cycle index, facilitating deeper insights and understanding of battery degradation mechanisms.

The Battery Cycles Overview plot as shown in Figure 3 illustrates the predicted cycle, actual cycle, and remaining cycles for each battery in the dataset over successive cycle indices. The blue line represents the predicted cycle values generated by the Random Forest Regressor model, while the green line depicts the actual cycle values observed in the dataset. The red line signifies the number of cycles remaining based on the predicted and actual cycles, providing insights into the discrepancy between predicted and actual battery life cycles.

The box plot visually compares the distribution of predicted cycle values generated by the Random Forest Regressor model (blue box) with the distribution of actual cycle values observed in the dataset (green box). The box plot as shown in Figure 4 provides summary statistics such as the median, quartiles, and outliers for both predicted and actual cycle distributions, enabling a comparison of the central tendency and variability between the two distributions.

The scatter plot as shown in Figure 5 illustrates the relationship between two voltage-related features, namely

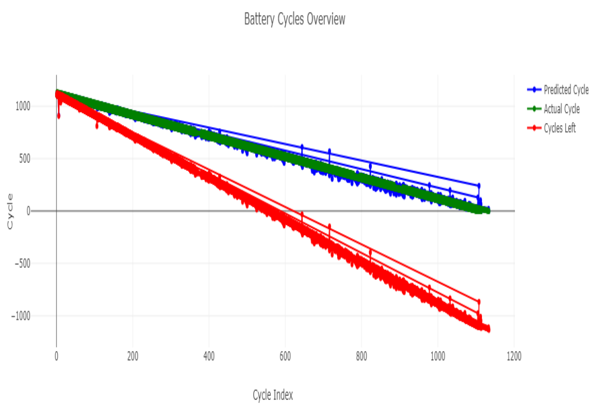

Figure 3. Battery Cycles Overview

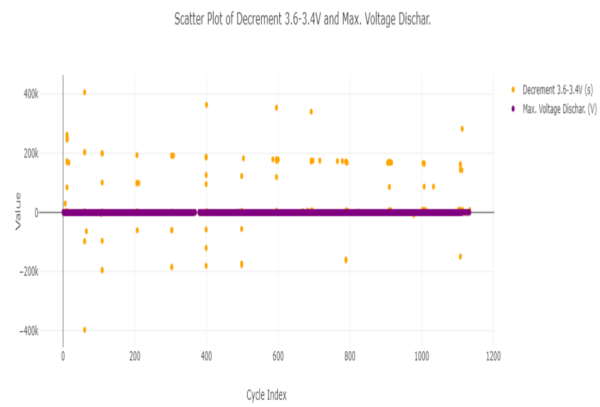

Figure 5. Scatter Plot of Decrement 3.6-3.4V and Max. Voltage Discharge

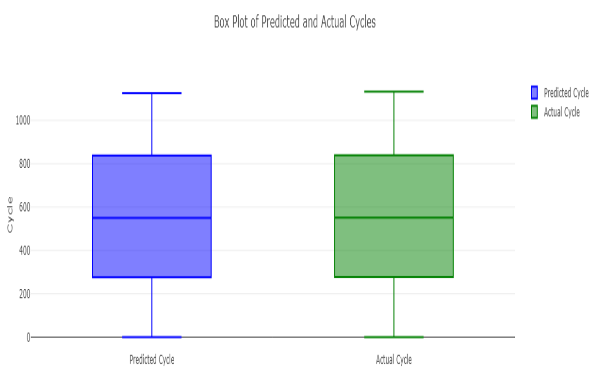

Figure 4. Box Plot of Predicted and Actual Cycles

the decrement of 3.6-3.4V (s) and the maximum voltage discharge (V), over successive cycle indices. Each point on the scatter plot represents a battery cycle, with the x-axis denoting the cycle index and the y-axis representing the values of the two voltage-related features. The scatter plot enables visualization of any patterns or trends in the voltage-related features across different battery cycles, aiding in the analysis of battery performance and degradation. The results provide a solid foundation for discussion on the practical implications, limitations, and future directions for improving lithium-ion battery management systems, as elaborated in the next section.

6. Discussion

The proposed RFR model has a strong capability of predicting the battery life cycle using obtained features. The computation of the RMSE and MAE gives the extent of the prediction error and the R^2 score offers an understanding of how well the model explains the variation in the dataset. The figures below the model further emphasise the ability of the model to identify these trends and patterns in battery degradation. Since the RFR model has rendered the feature importance, it is easier within this analysis to identify the parameters that have some significance in the battery life cycle prediction. It may also provide more directions for future research and development initiatives relating to factors that are responsible for battery degradation rate and ways of improving battery management and utilization. Considered limitations in this study include its advancements. The most obvious drawback is first, reliance on historical battery degradation data does not necessarily capture dynamic real world conditions, particularly charging protocol, temperature convulsion, or operational stress. Moreover, the Random Forest Regressor (RFR) is quite robust, although it cannot be used when working with large number of dimensions and low diversity. Feature selection and cross validation partially mitigate this issue, but we still need more work to generalize the model to new datasets and extreme scenarios. Finally, the current methodology does not take account of real time data streams that are severely limiting its ability to operate in rapidly changing operating environments. The practical implications of improved RUL predictions are quite important for stakeholders in the electric vehicle (EV) and renewable energy sectors. Accurate RUL predictions allow for proactive maintenance scheduling that minimizes downtime and avoids costly mid-operation failures for fleet operators. For instance, delivery fleets or public transportation systems can do battery replacements during between peak hours if what the service must remain intact. Likewise, in the renewable energy area, precise RUL estimates permit more plug and play of battery storage systems to energy grids. Predicting when batteries will fail allows grid operators to prevent disruptions, resource allocation, and maintain energy reliability during peak demand periods. What's more, the RUL prediction gives customers a means to have confidence in EV technology through improved certainty of avoiding sudden battery failures or high replacement costs that may otherwise impede uptake of clean energy.

Practical Implications

It is evidenced that the proposed approaches for the RUL estimation processes of LIBs, which are derived from the extensions of feature extraction and prediction by using the RFR model, have vast potentials in the battery management systems for multiple applications. It supports the improvement of the reliability of RUL predictions and hence the improvement of decision-making by the different stakeholders in terms of maintenance period, investment, and battery replacement strategies. The estimation of the RUL of batteries to the nearest possible value is very useful in planning the maintenance activities so as to operate the battery till the time it needs replacement or servicing. Not only does it assist in enhancing the battery's service life but also reduce the operational costs as compared to those that are incurred for warranty and other downtimes occasioned by early failures of batteries in the system, besides improving the efficiency of the overall system. As much as it may mean better prediction in case of battery failure consumers feel confident with the electric cars and other sustainable transport technologies. Such implications are not restricted to the gains that users make out of the system, but also embrace societal issues, for instance, the physical world. Improvement in battery efficiency is expected to lead to reduction in greenhouse gas emissions because battery efficient electric cars use less energy and have longer life expectancy thus they will be replaced more rarely. This will assist in the elimination of smog particularly in the urban areas whereby emission of cars is a major issue on environment. Furthermore, improvement in the battery of the EV has improved the reliability and efficiency of technological devices in the fight against green mobility and climate change through emission reduction. The closing section summarizes the main results of this study, illustrates the applicability of this work to Battery Management Systems, and proposes future research directions on RUL prediction methodologies.

7. CONCLUSIONS AND FUTURE WORK

This study demonstrates the effectiveness of a Random Forest Regressor (RFR) model for accurately predicting the Remaining Useful Life (RUL) of lithium-ion batteries, achieving notable performance with an RMSE of 0.048, MAE of 0.032, and an R^2 score of 0.92. By integrating advanced feature extraction techniques and optimizing hyperparameters, the model successfully captures complex degradation patterns, enhancing battery management systems across applications like electric vehicles and renewable energy storage. The results underscore the model's potential to improve operational efficiency, reduce downtime, and enable proactive maintenance. Future work could focus on incorporating real-time data and addressing environmental variables such as temperature fluctuations to further refine predictions and increase model applicability in diverse operating conditions. The integration of Random Forest Regression (RFR) with advanced feature extraction techniques indeed holds substantial promise for improving battery health assessments and refining Remaining Useful Life (RUL) predictions. By capturing the complex relationships within battery degradation data, this model can generate more accurate predictions, which are crucial for the effective management of batteries, particularly in applications such as Electric Vehicles (EVs) and renewable energy storage systems. The detailed and methodical feature extraction process plays a vital role in identifying the underlying patterns and potential degradation sources that affect battery performance. This enables the RFR model to make more informed decisions regarding maintenance schedules, resource allocation, and operational strategies. By boosting the adoption of electric vehicles, improving air quality, and reducing pollutant emissions, this research aligns with global goals for greener, more sustainable transportation and energy systems. Enhanced RUL predictions can play a role in accelerating the transition to clean energy solutions, contributing positively to both environmental and societal well-being.

However, this work does not take into account the influence of different temperature variations and failure thresholds. Focusing on temperature variations is critical since LiBs are highly sensitive to temperature fluctuations, which can significantly impact their charge/discharge cycles, capacity retention, and overall lifespan. Incorporating failure thresholds, which account for the conditions under which batteries fail or degrade beyond repair, would also allow for a more nuanced model that better reflects real-world battery behavior. The idea of integrating a digitally controlled power supply embedded within the Random Forest (RF) optimization model is promising. This would enable more accurate, real-time monitoring of battery capacity in challenging or uncontrolled environmental conditions. Developing such an estimator for the charger to provide online, in-loop capacity estimation will also be valuable for battery management systems, especially in applications where external conditions vary frequently, such as in electric vehicles or off-grid energy storage systems. This approach could help mitigate performance degradation and improve the safety and efficiency of the charging process, offering a more adaptable and resilient solution for LiBs in diverse environmental conditions. It also opens the door for more accurate predictive maintenance and real-time decision-making in battery management systems.

References

[1] E. Acar, N. Jain, P. Ramu, C. Hwang, and I. Lee, "A survey on design optimization of battery electric vehicle components, systems, and management," *Structural and Multidisciplinary Optimization*, vol. 67, no. 3, pp. 27, 2024.

[2] K. Nangare, K. Nidubrolu, and P. Gaikwad, "Study of Derived Battery Features for Real-Time Estimation of SOH and RUL of EV Battery Using Data Analysis," SAE Technical Paper, pp. 2024-26-0123, 2024.

[3] Z. Xu, N. Xie, and K. Li, "Remaining useful life prediction for lithium-ion batteries with an improved grey particle filter model," *Journal of Energy Storage*, vol. 78, pp. 110081, 2024.

[4] S. Jafari and Y. C. Byun, "Accurate remaining useful life estimation of lithium-ion batteries in electric vehicles based on a measurable feature-based approach with explainable AI," *The Journal of Supercomputing*, vol. 80, no. 4, pp. 4707–4732, 2024.

[5] M. S. H. Lipu, M. S. A. Rahman, M. Mansor, T. Rahman, S. Ansari, A. M. Fuad, and M. A. Hannan, "Data-driven health and life prognosis management of supercapacitor and lithium-ion battery storage systems: Developments, implementation aspects, limitations, and future directions," *Journal of Energy Storage*, vol. 98, p. 113172, 2024.

[6] J. Li, G. Bai, J. Yan, and J. Gu, "On-line parameter identification and SOC estimation of nonlinear model of lithium-ion battery based on Wiener structure," *Journal of Energy Storage*, vol. 92, pp. 112094, 2024.

[7] S. Wang, R. Zhou, Y. Ren, M. Jiao, H. Liu, and C. Lian, "Advanced data-driven techniques in AI for predicting lithium-ion battery remaining useful life: a comprehensive review," *Green Chemical Engineering*, 2024.

[8] A. Rastegarpanah, M. E. Asif, and R. Stolkin, "Hybrid Neural Networks for Enhanced Predictions of Remaining Useful Life in Lithium-Ion Batteries," *Batteries*, vol. 10, no. 3, pp. 106, 2024.

[9] Y. Wang, G. Gao, X. Li, and Z. Chen, "A fractional-order model-based state estimation approach for lithium-ion battery and ultra-capacitor hybrid power source system considering load trajectory," *Journal of Power Sources*, vol. 449, pp. 227543, 2020.

[10] Y. Wang and Z. Chen, "A framework for state-of-charge and remaining discharge time prediction using unscented particle filter," *Applied Energy*, vol. 260, pp. 114324, 2020.

[11] Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming, and A. A. Lee, "Identifying degradation patterns of lithium-ion batteries from impedance spectroscopy using machine learning," *Nature Communications*, vol. 11, no. 1, p. 1706, 2020.

[12] L. Tao, H. Wu, and X. Zheng, "Multifeature-based online remaining useful life prediction of lithium-ion batteries in stages using cascaded data-driven algorithm," *Quality and Reliability Engineering International*, vol. 40, no. 5, pp. 2527–2546, 2024.

[13] X. Wang, B. Hu, X. Su, L. Xu, and D. Zhu, "State of health estimation for lithium-ion batteries using random forest and gated recurrent unit," *Journal of Energy Storage*, vol. 76, pp. 109796, 2024.

[14] M. Wei, M. Ye, Q. Wang, and J. P. Twajamahoro, "Remaining useful life prediction of lithium-ion batteries based on stacked autoencoder and Gaussian mixture regression," *Journal of Energy Storage*, vol. 47, pp. 103558, 2022.

[15] A. Ayob, S. Ansari, M. S. H. Lipu, A. Hussain, and M. H. Md Saad, "SOC, SOH, and RUL estimation for supercapacitor management system: Methods, implementation factors, limitations, and future research improvements," *Batteries*, vol. 8, no. 10, pp. 189, 2022.

[16] Z. Wang, Q. Ma, and Y. Guo, "Remaining useful life prediction of lithium-ion batteries based on deep learning and soft sensing," *Actuators*, vol. 10, no. 9, pp. 234, 2021.

[17] Z. Yang, Y. Wang, and C. Kong, "Remaining useful life prediction of lithium-ion batteries based on a mixture of ensemble empirical mode decomposition and GWO-SVR model," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[18] S. Wang, Y. Fan, S. Jin, P. Takyi-Aninakwa, and C. Fernandez, "Improved anti-noise adaptive long short-term memory neural network modeling for the robust remaining useful life prediction of lithium-ion batteries," *Reliability Engineering and System Safety*, vol. 230, pp. 108920, 2023.

[19] J. Qu, F. Liu, Y. Ma, and J. Fan, "A neural-network-based method for RUL prediction and SOH monitoring of lithium-ion battery," *IEEE Access*, vol. 7, pp. 87178–87191, 2019.

[20] Y. Li, M. Abdel-Monem, R. Gopalakrishnan, M. Berecibar, E. Nanini-Maury, N. Omar, P. van den Bossche, and J. Van Mierlo, "A quick online state of health estimation method for Li-ion battery with incremental capacity curves processed by Gaussian filter," *Journal of Power Sources*, vol. 373, pp. 40–53, 2018.

[21] W. Duan, S. Song, F. Xiao, Y. Chen, S. Peng, and C. Song, "Battery SOH estimation and RUL prediction framework based on variable forgetting factor online sequential extreme learning machine and particle filter," *Journal of Energy Storage*, vol. 65, pp. 107322, 2023.

[22] J. Jia, J. Liang, Y. Shi, J. Wen, X. Pang, and J. Zeng, "SOH and RUL prediction of lithium-ion batteries based on Gaussian process regression with indirect health indicators," *Energies*, vol. 13, no. 2, pp. 375, 2020.

[23] G. Dong, F. Yang, Z. Wei, J. Wei, and K.-L. Tsui, "Data-driven battery health prognosis using adaptive Brownian motion model," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4736–4746, 2019.

[24] P. Li, Z. Zhang, Q. Xiong, B. Ding, J. Hou, D. Luo, Y. Rong, and S. Li, "State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network," *Journal of Power Sources*, vol. 459, pp. 228069, 2020.

[25] B. Gou, Y. Xu, and X. Feng, "State-of-health estimation and remaining-useful-life prediction for lithium-ion battery using a hybrid data-driven method," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 10854–10867, 2020.

[26] A. Patil and J. Kendule, "IoT-based battery health monitoring and its remaining useful life prediction using artificial neural network," in *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pp. 1–5, 2024.

[27] H. Tian, P. Qin, K. Li, and Z. Zhao, "A review of the state of health for lithium-ion batteries: Research status and suggestions," *Journal of Cleaner Production*, vol. 261, pp. 120813, 2020.

[28] S. Mishra, C. K. Panigrahi, S. Debdas, D. V. P. Varma, and M. Yadav, "IoT enabled battery status monitoring system for electric vehicles," in *2023 IEEE 3rd International Conference on Sustainable Energy and Future Electric Transportation (SEFET)*, pp. 1–5, 2023.