INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt26116

Ensemble Learning-Based Battery Health Estimation: A Comparative Study of RUL, SOC, and SOH Prediction Using Machine Learning Models

C.L. Sravanthi [a, *], J.N. Chandra Sekhar [a]

[a] Department of Electrical and Electronics Engineering, S V U College of Engineering, Sri Venkateswara University, Tirupati-517502, Andhra Pradesh India

* Corresponding Author Email: sravanthi.cl@gmail.com

DOI: https://doi.org/10.54392/irjmt26116

Received: 30-07-2025; Revised: 24-12-2025; Accepted: 07-01-2026; Published: 30-01-2026

Abstract: Lithium-ion batteries are paramount energy storage devices used in modern electric vehicles because of its favourable energy density, light weight and environmental benefits. However, issues related to ageing, degrade mechanism, long term reliability and safety remain critical, particularly under variable operating conditions. Reliable estimation of Remaining Useful Life (RUL), State of Charge (SOC) and State of Health (SOH) is therefore required to support efficient energy management and prevent premature battery failure in electric vehicle applications. Recent years, machine learning algorithms have been increasingly adopted for battery state estimation. Nevertheless, systematic comparison of commonly used algorithms for simultaneously predicting RUL, SOC and SOH are still limited. In this work five machine learning models-Decision Tree (DT), Random Forest (RF), Bagging Regressor (BR), XG Boost (XGB) and an ensemble voting regressor combining Random Forest and XG Boost are implemented and evaluated using Hawaii Natural Energy Institute (HNEI) data set comprising capacity, voltage, temperature and cycle life features. Model performance is assessed using Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE) and R-Square (R^2). Ensemble model consistently achieves lower prediction error and improved stability with less computational time compared to individual learners across all estimation tasks. The results indicate that combining tree-based models with gradient boosting improves generalization performance while maintain computational feasibility, making the proposed approach suitable for practical battery management systems deployment in electric vehicles.

Keywords: Battery Management Systems, Lithium-Ion, Ensemble Machine Learning, Grid Search Cross Validation, Voting Regressor, Performance Evaluation Metrics.

1. Introduction

Electric vehicles adoption has witnessed substantial growth in recent years due to demand for sustainable environment and energy efficient transportation. Lithium-ion batteries are primary energy storage technology used in electric vehicles because their compact size, high energy density and power. However, their performance gradually declines with ageing and charge-discharge cycles, which impacts reliability and operational efficiency [1]. Accurate estimation of battery predictions like RUL, SOH and SOC of lithium-ion batteries have become crucial for effective battery management systems. Widely, electric vehicles provide practical means for reducing greenhouse gas emissions [2]. The performance EVs is highly influenced by battery management systems, which also impact safety and reliable operation. It is anticipated that nanotechnology would keep propelling developments in battery technology, especially in LIB's. The function of nanotechnology in Li-ion batteries for electric vehicle application has been comprehensively investigated in [3]. The ageing and deterioration of electrochemical components within Li-ion can reduce capacity and output due to increased internal impedance. Mechanical stress, operational temperature, and electrochemical processes all contribute to performance loss [4]. This provides simple information for battery community researchers to improve EV battery safety and achieve beneficial and practical effects. The complex and nonlinear nature of battery capacity deterioration forms it ambitious to anticipate capacities and RUL accurately [5]. The RUL prediction by data-dependent models using Machine learning algorithms suggested in last few years and numerous methods been suggested in the literature as a means of achieving accurate prediction of battery future capacities and RUL [6]. These methods can generally be enumerated as: model-based and data-driven approaches. A comprehensive evaluation of advantages and limitations of machine learning-based approaches for estimating state of charge and state of health was presented in [7]. The input characteristics, metrics, hyperparameters, datasets and techniques for SOH and SOC estimation using ML models for EV battery forecasting is presented [8]. For electric vehicles to operate safely and optimally, Battery State of Charge (SOC) estimation must be done accurately [9]. In order to improve battery management, a trustworthy data-driven architecture that makes use of sophisticated analytics for accurate real-time SOC monitoring was presented in [10].

Techniques and importance of SOH and RUL estimates are thoroughly explained and additionally, a variety of machine learning algorithms are carefully and independently examined according to each category and subcategory used for SOH and RUL assessment [11]. Lastly benefits in terms of computation and accuracy is given with goal of offering insight for future advancement. The continuous or abrupt deterioration of the battery is assessed when integrating the precise RUL prediction and SOH estimation; for this reason, it is essential to model the battery degradation [12, 13]. To guarantee battery safety and dependability, the battery deterioration model must be developed using SOH and RUL estimations [14].

A battery model with UKF based SOC estimate are used to offer an equalization strategy based on passive balancing control. A circuit with a 2nd order Thevenin equivalent is used to mimic LIB's [15]. Most of contemporary findings on widely used ML approaches in forecasting SOC and SOH and overview of both BMSs and ML was reviewed in [16, 17]. Recent data-dependent models to forecast SOH of LIBs and suggests global forecasting procedure that involves gathering datasets for LIB charging and discharging, processing data and features, and choosing algorithms was explained in [18]. A novel LIB's RUL prediction technique, termed Auto-CNNLSTM. This approach is based on deep convolutional neural networks (CNN) and long short-term memory (LSTM) to extract further insights from finite input. This approach employs an autoencoder to enhance data dimensionality for improved training of CNN and LSTM. A filter is employed to provide continuous and steady output by smoothing the projected value. Findings on real-world dataset illustrate efficacy of the suggested strategy in comparison to other widely-used techniques [19].

Precise estimation of battery SOC is necessary for ensuring reliable operation and energy efficiency in EV applications. A comparative study reported in the literature evaluates machine-learning regression approaches, including support vector machines, neural networks, ensemble techniques, and Gaussian process regression, to capture the connection between real-time data and battery State of Charge [9].

This study focuses on estimating the remaining useful life, state of charge and state of health of lithium-ion batteries with five machine-learning approaches using fine parameter tunning. Model performance is evaluated using standard prediction error metrics and actual-versus-predicted plots are used to examine the estimation behaviour of each algorithm. The results show that hybrid model combining RF and XG Boost provides more accurate and consistent predictions for RUL, SOC, and SOH when compared with the individual learning algorithms.

The remainder of paper is structured as follows: Section 2 contains related research observed between existing and proposed models. Section 3 gives brief description about proposed prediction methods RUL, SOC, SOH and machine learning algorithms. Section 4 presents Proposed methodology work Flow with data set description and performance error indices. Section 5 provides results and discussions for individual Battery estimations and plots related to estimation and comparative analysis is also provided. Lastly, conclusion is outlined in Section 6.

2. Related Research

Over the last 10 years, estimation of RUL, SOC, SOH has become a crucial for Predictive Maintenance (PdM) task and a significant topic of research. This prediction is crucial to guarantee overall reliability and safety of the system in order to develop an effective maintenance plan. RUL, SOC, SOH can be predicted from historical data. ML techniques may be used to anticipate their results with greater accuracy. This section reviews significant research work and remarkable approaches that have laid groundwork for our exploration.

An aged Li-Ion battery based on cobalt was cycled under constant frequency control profile for grid application, and impedance development over time was modelled using ML algorithms. Accuracy of various machine learning approaches was compared in order to ascertain the battery's health and charge level as it ages [20]. Results from experiments demonstrated that machine learning predicated on Random Forest algorithm employed for this intention is in a profitable way.

Authors introduce an innovative approach to improve precision and reliability of battery performance analysis by integrating multiple machine learning techniques. Their hybrid model, which combines k-NN, Random Forest, and Extreme Gradient Boosting algorithms, demonstrates outstanding results in [21, 22].

For electric vehicles to operate safely and optimally, Battery State of Charge (SOC) estimation must be done accurately. In order to analyse the intricate link between driving data and battery SOC many ML egression techniques, such as Support vectors, Gaussian Process, Neural Network and Ensemble Method were compared in [9]

A ML based battery model was developed, validated, and compared in study reported in [23]. According to the experiment's findings, developed model is accurate to anticipate how capacity and SOH would change while electric car batteries operated. Additionally, the comparison research demonstrated that the battery model's final accuracy was 99.98%. A SOH estimate technique for LIB's based on XG Boost model in order to increase estimation's accuracy was suggested approach which exhibits a 10%–20% increase in accuracy when compared to XGB, RFR, KNN, SVM, and LR approaches in [24]. A novel preprocessing technique for artificial intelligence models to enhance battery SOH estimates. The relative SOC, which could be readily computed during charging using present consolidation approach, served as foundation for suggested preprocessing [25].

The similarity coefficient between datasets and useable capabilities computed in order to compare suggested approach with standard preprocessing method based on constant time intervals. Random Forest and Light GBM technique were suggested for accurate RUL prediction of lithium-ion batteries. It gets excellent R2 values, indicating model is quite good at predicting battery RUL for the given dataset [26].

Table 1 provides a comparative analysis of machine learning algorithms used for battery predictions. This summarizes the methodology used by various researchers, highlighting their contributions and limitations. Above comparison clearly indicates need for balanced approach to improve accuracy, generalization and practical applicability for use of ensemble models. From the literature review, existing works focus on single health metrics leaving for RUL, SOC, SOH prediction simultaneously. Therefore, this work focuses on developing Grid Search–Optimized Ensemble Learning Framework that integrates Random Forest (RF) and XGBoost (XGB) models to improve predictive accuracy and robustness of lithium-ion battery health estimations. The optimization of hyperparameters through Grid Search 3-fold Cross-Validation (GSCV) ensures consistent with lower computational for dataset of HNEI.

Table 1. Comparative analysis of ML based approaches for predictions

Authors, Year [Ref]	Methodology	Contributions	Remarks
Jafari *et al,* 2023 [21]	KNN+RF+XGB for RUL prognostics	Hybrid machine learning framework with dual-target prediction strategy	Model complexity is higher which may increase training time and memory requirements.
Jafari *et al*, 2023 [26]	RUL Estimation using RF and LGBM with HHO	Hyperparameter optimization is used to achieve high R-square and low error metrics	Increase in Training time with optimization algorithm and model complexity
Sekhar *et al.*, 2023 [27]	RUL prediction using RF, SVR and LR	Compared multiple ML algorithms for estimation	Lacks ensemble strategies for accuracy and robustness
Davide Aloisio *et al.*, 2021 [20]	RF, SVM, ANN and LR using EIS data for SOC calculation	Multiple ML algorithms are compared Individually for LIB estimation of SOC	Data acquisition is difficult for real time implementation and conducted for single battery chemistry
Jafari, S.; Shahbazi, Z *et al.*, 2022 [22]	XGB method was used for SOC estimation	SOC online estimation frame work using partial CC charging	Calculated under fixed temperature conditions and generalization limiting.
Pranav *et al.*, 2024 [9]	SVM and GPR are used for estimating SOC	Comparative analysis was done using real time EV driving data for predictions	High Computational complexity.
Shuxiang Song, et al, 2020 [24]	XG Boost regression for SOH prediction	XGB based SOH estimation model was developed to predict the performance metrics	Performance is sensitive to noise and hyperparameter tunning
Khalid Akbar *et al,* 2022 [23]	LR, SVR & RF are used for SOH evaluation	Performed comparative analysis of ML based regression models	LR and SVR fail to capture nonlinear degradation behaviour

In this work accurate prediction of Remaining Useful Life (RUL), State of Charge (SOC), and State of Health (SOH), is estimated with five different machine learning algorithms such as Decision Tree (DT), Random Forest (RF), Bagging Regressor, XG Boost (XGB) and ensemble model (RF + XGB).

3. Materials and Methods

For reliable operation and improved energy efficiency in battery management systems accurate estimation of battery parameters like RUL, SOH and SOC are vital for safe operation and battery second-life applications.

3.1. Definitions

a) ***Remaining Useful Life (RUL):*** Indicates the remaining time a battery can be used effectively before it starts to fail and it is calculated using equation (1) below.

$$\mathrm{RUL} = N_{EOL} - \mathrm{N} \quad (1)$$

Here, N_{EOL}: maximum number of battery cycles N: quantity of battery cycles or Cycle Index

b) ***State of Charge:*** It indicates the amount of energy remaining in a battery and is usually expressed as a percentage. In this study, SOC is computed from the available dataset using normalized linear interpolation, as defined in equation (2).

$$SOC = \left(\frac{V_{measured} - V_{min}}{V_{max} - V_{min}}\right) \times 100 \quad (2)$$

Where $V_{measured}$ is battery voltage, V_{max}voltage at 100%SOC, $V_{min}$is voltage at 0% SOC.

c) ***State of Health:*** Indicates how healthy a battery is, based on its remaining useful life and how many cycles it has already undergone and is measured using equation (3).

$$SOH = \left(\frac{RUL}{RUL + Cycle\_Index}\right) \times 100 \quad (3)$$

Here Cycle Index: number of cycles already used

RUL: number of cycles remaining

3.2. Machine Learning Algorithms

Selecting an appropriate machine learning model is an significant step in the paradigm development process. This requires training multiple models and comparing their performance on test data to make an informed choice. From the historical data during training ML models adjust their parameters to improve prediction accuracy. The machine learning algorithms used for prediction of SOC, RUL and SOH are explained briefly below.

3.2.1. Decision Tree (DT)

It is supervised learning model used for regression problems. Based on the dataset given test are performed on basis of features. Here decision nodes make decision where as leaf node are output of those decisions [27]. The best attribute is measured using Attribute selection measure. Here it compares the values of attribute with nodes and generates the decision tree node which contains the best attribute. For estimation of the decision tree node gain index is used to select the attribute as shown in equation (4) below.

$$\text{Gain} = \text{Entropy(i)} - \sum_{v}^{a} \left|\frac{i_v}{i}\right| . Entropy(i_v) \quad (4)$$

Were 'i' is instances at attribute 'a', i_v is subset of i, 'v' individual value that attribute a can take and 'a' is feasible values.

3.2.2. Random Forest (RF)

It is an ensemble learning technique that builds numerous decision trees during training. Each tree is trained on a randomly elected sample drawn from original dataset, and at every split, only a subset of input characteristics is considered [28]. It will reduce overfitting and improves the model's ability to infer to unseen data. For classification tasks, node splitting is commonly based on the Gini Index or Information Gain, whereas for regression problems, criteria such as mean squared error (MSE) or variance reduction are used. The final prediction in regression is obtained by averaging the outputs of all individual trees. This algorithm can manage complex data, reduce overfitting, and generate precise forecasts. Mathematical RF regression approach is shown in below in equation (5).

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^{N} P_i(x) \quad (5)$$

Where$\hat{Y}$expected value, N RF model decision, and $P_i(x)$ forecast of i[th] decision tree.

3.2.3. Bagging Regressor (BR)

The Bagging Regressor is a hybrid learning approach designed to improve regression performance by aggregating predictions from multiple base learners, typically decision trees. By training each learner on different bootstrap sample of the dataset, this method effectively reduces model variance and mitigates the risk of overfitting, thereby improving overall predictive accuracy. It is necessary to first define bootstrapping in order to understand bagging [29]. This process of producing bootstrapped samples from the designated dataset is known as bootstrapping. The Bagging procedure first generates bootstrapped samples. Mathematically, Bagging regressor is calculated as shown in equation (6) below.

$$\widehat{Bag}_{pre} = \hat{B}_1(X) + \hat{B}_2(X) + \cdots + \hat{B}_n(X) \quad (6)$$

Where bagged prediction are in Left hand side and the individual learners on right hand side.

3.2.4. Extreme Gradient Boost Regressor (XG Boost)

Extreme Gradient Boosting [30] or XG Boost, is a potent machine learning technique that excels in regression problems. Its capacity to manage intricate datasets and minimize overfitting is its main advantage. Accurate predictions may be made at every stage of the decision-making process by XG Boost by adding a regularization component to its loss function. This ensures optimal execution in RUL prediction, where accuracy is essential. Objective function is given below in equation (7).

$$\hat{Y} = \emptyset(x_i) = \sum_{m=1}^{M} P_m(x_i) \quad (7)$$

The final prediction $\hat{Y}$ obtained by aggregating outputs from a hybrid M decision tree, where each individual $P_m(x_i)$ provides prediction for the inputx_i.

3.2.5. Ensemble Model (Random Forest and XG Boost Combination)

To improve robustness and generalization of SOH prediction, we employed an ensemble learning strategy by combining Random Forest and Extreme Gradient Boosting using a Voting Regressor framework. For regression problems hybrid algorithms are effective as they combine multiple models to improve accuracy and stability. In this work, a Voting Regressor is used for RF and XGB, which simply averages the outputs of both models. Here RF provides robustness against noise and overfitting, while XG Boost efficiently captures complex nonlinear patterns through boosting and regularization. The output is obtained is calculated as shown in equation (8).

$$\hat{Y}_{ensemble}(x) = \frac{1}{2}(\hat{Y}RF\ (x) + \hat{Y}XGB(x)) \quad (8)$$

Where $\hat{Y}RF\ (x)$ is RF model and $\hat{Y}XGB(x)$ is XG Boost model.

Voting Regressor was selected as it reduces the overfitting risk for real-time battery management systems in EV's. Whereas stacking methods, require an additional meta-learner which increase model complexity and risk of overfitting, whereas Furthermore, optimized weighting strategies were evaluated during evaluation. For the XG Boost model, parameters including max_depth were selected to control tree complexity, learning rate to regulate the step size, subsample to introduce randomness, and num_estimators to define number of boosting iterations. In addition, gamma was tuned for regularization, and colsample_bytree was used to specify the proportion of features used during tree construction. Table 2 below gives the best hyper parameters used for training the algorithms to get accurate results.

The ensemble model benefits from combining the strengths of Random Forest, which reduces variance through bagging, and XGBoost, which lowers bias using boosting techniques. Since the two models learn patterns in different ways, their integration leads to more accurate and stable predictions.

Algorithm 1. Proposed RF+XGB Ensemble Model for Battery Health Estimation

Start

Input

- Battery Dataset

X= Extracted Battery Features

Y= Target variables (RUL, SOC & SOH)

- Output

Performance Metrics: MAE, MSE, RMSE and R-square

Step 1: Data Preprocessing

Battery dataset is cleaning, normalization, feature extraction is done in preprocessing stage and split into training and testing sets.

Step 2: Hyperparameter Tunning

Models are individually tunned using Grid Search with 3-fold cross-validation to obtain optimal models.

Step 3: Hybrid Model Construction

The optimized RF and XGB models were combined using Voting Regressor.

Step 4: Model Training and Prediction

Voting Regressor is trained on training dataset and RUL, SOC and SOH are predicted on the test dataset.

Step 5: Performance Evaluation

Predictions are evaluated using performance error indices like MAE, MSE, RMSE and R-Square.

End Algorithm

Table 2. Parameters used for tunning ML models

Models	Parameters
Decision Tree	n _ estimators:100, max_depth:10, min_samples: 4,
Random Forest	num_estimators:100, maximum_depth:20, min_samples leaf:1, min_samples split: 2
Bagging Regressor	n _ estimators:100, max_features:1.0, max_samples:0.5
XG Boost	N_estimators:100, max_depth:5, subsample:0.8, colsample_bytree:1.0
Ensemble (RF + XGB)	weights = [0.45, 0.55], voting = 'soft'

4. Methodology

4.1. Proposed Work Flow

The methodology for proposed work is shown in figure 1. Process begins with data gathering and preparation, continued by feature selection. The dataset from Hawaii Natural Energy Institute (HNEI) is utilized in modelling lithium-ion batteries. Operative techniques and tools used for implementing the strategy consists of Windows 11 utilizing 2.10 GHz, Ram of 16 GB with Python 3.10.12 programming language [31].

4.2. Data set Description

The dataset used consists of experimental measurements collected from 14 lithium-ion batteries, each cycled for more than 1000 charge–discharge cycles under a controlled ambient temperature of 25 °C. The batteries were charged at a C/2 rate and discharged at 1.5C, as reported in [32]. For every battery, the cycling data contains 15,064 records, where each record corresponds to a single operational instance. The recorded parameters include the cycle number, overall discharge time, discharge time within voltage range of 3.6 V to 3.4 V, initial voltage values at the beginning of both discharge and charge processes, charging duration until 4.15 V is reached, constant-current charging time, and total charging time. These parameters are considered key indicators of battery behaviour and are used as input features for assessing battery performance and estimating the remaining useful life.

4.3. Data Processing

In preprocessing stage data Cleaning is done by removing noise, missing or inconsistent values. feature extraction selects relevant parameters. Feature engineering was used to extract relevant information and generate new characteristics from raw data. Normalization, which can improve efficiency of ML algorithms, was done to make sure that every feature was on an equal scale suitable for RUL prediction. In contrast, datasets related to State of Charge and State of Health were not on consistent scale. To ensure uniformity and enhance performance of machine learning models, SOC dataset was normalized using min–max normalization. For SOH prediction we used Cycle index and RUL values.

4.4. Model Training

Dataset was divided it into three categories for 70% training data,15% testing data and 15% validation. These are conducted using dataset throughout the model training phase. The validation set was used for model selection and parameter tuning, test set was used to assess final model's performance, and training set was used to build machine learning models. Grid search CV hyperparameter tuning is used for model optimization with a 3- fold cross validation (cv=3, verbose=1). The GSCV hyperparameter Used are param_grid includes N_estimators, maximum_depth, learning_rate, subsample, colsample_bytree and criterion.

4.5. Performance Metrics

We employed four measures to evaluate efficacy and precision of Models techniques in forecasting the batteries' RUL, SOC, and SOH [33].

Mean Absolute Error (MAE): Computes absolute deviation between expected and observed as shown in equation (9).

$$\text{MAE}=\frac{1}{N}\sum\left|Y-\hat{Y}\right| \tag{9}$$

Mean Square Error (MSE): Estimation of average squared variance between values that model predicts and actual value that it provides as given by equation (10).

$$\text{MSE}=\frac{1}{N}\sum\left|Y-\hat{Y}\right|^2 \tag{10}$$

Root Mean squared Error (RMSE): It is MSE square root and measured using equation (11).

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum\left|Y-\hat{Y}\right|^2} \tag{11}$$

R-Square Error (R^2): Indicates how well predicted values match the original values as shown in equation (12).

$$R^2 = 1-\frac{MSE(Model)}{MSE(Base\ line)} \tag{12}$$

Here Y is actual value and $\hat{Y}$ Predicted value.

5. Results and Discussion

Analysis of study is thoroughly presented in this section. Effectiveness of Machine learning algorithms were estimated and information about the precision and accuracy prediction of battery RUL, SOC and SOH by calculating the error-index values were assessed.

5.1. RUL Estimation Results

RUL estimation input features include Cycle index, voltage, Current, discharge time, charging time, etc. Output features are predicted number of remaining cycles (RUL). This prediction shows many cycles remain before the battery reaches end-of-life. We employed four parameters to evaluate the precision and efficacy of our approach in forecasting the batteries' RUL. The MAE, MSE and RMSE calculate discrepancies between actual and expected RUL values. Table 3 summarizes the performance models for RUL estimation.

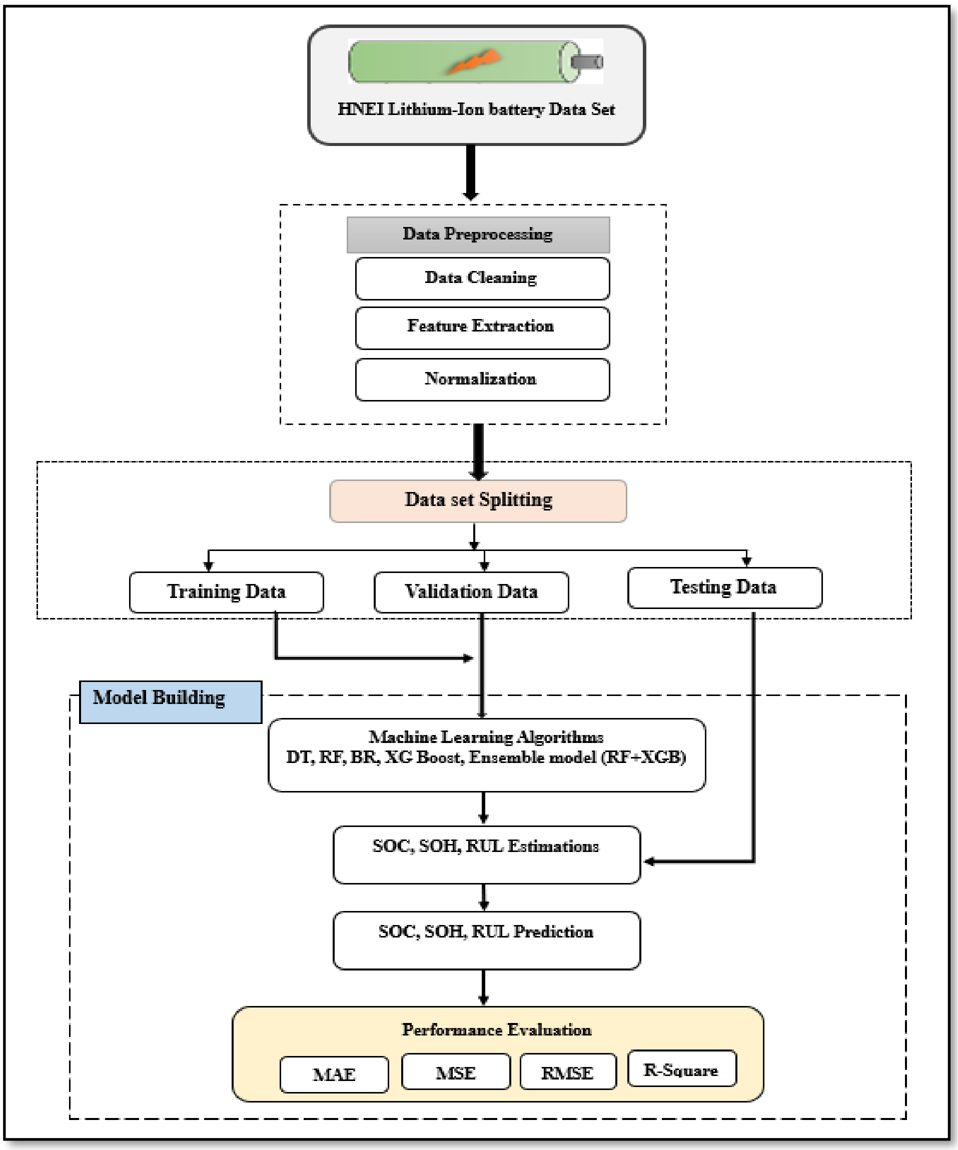

Figure 1. Methodology for Proposed Work

Table 3. Remaining useful life (RUL) estimation results

Algorithm	MAE	MSE	RMSE	R-Squared
Decision tree Regressor	2.3278	2.222	3.4960	0.9979
Random Forest Regressor	0.2218	1.4976	1.2222	0.9985
Bagging Regressor	1.3716	5.5807	2.3623	0.9997
XG Boost Regressor	0.2465	1.0689	1.1189	0.9981
Ensemble (RF+XG Boost)	0.2201	1.0411	1.0821	0.9999

Figure 2 results show plot against actual vs predicted and perfect prediction. Here DTR Overestimation and underestimation both visible. RFR Indicates high accuracy, low error.BR generally has good fit, but more deviation than Random Forest and XG Boost. Whereas XGBR shows accurate results. Finally, the Ensemble prediction RF with XG Boost gives best performance visually and predicted values lie almost exactly on the red line which indicates minimal error, likely highest R^2.

RUL estimation, comparing them across metrics of MAE, MSE, RMSE, and R^2 is shown in figure 3. Performance of R^2 especially with RF + XGB Ensemble, Bagging, and XG Boost hit close to 1.0 which is an Excellent fit.

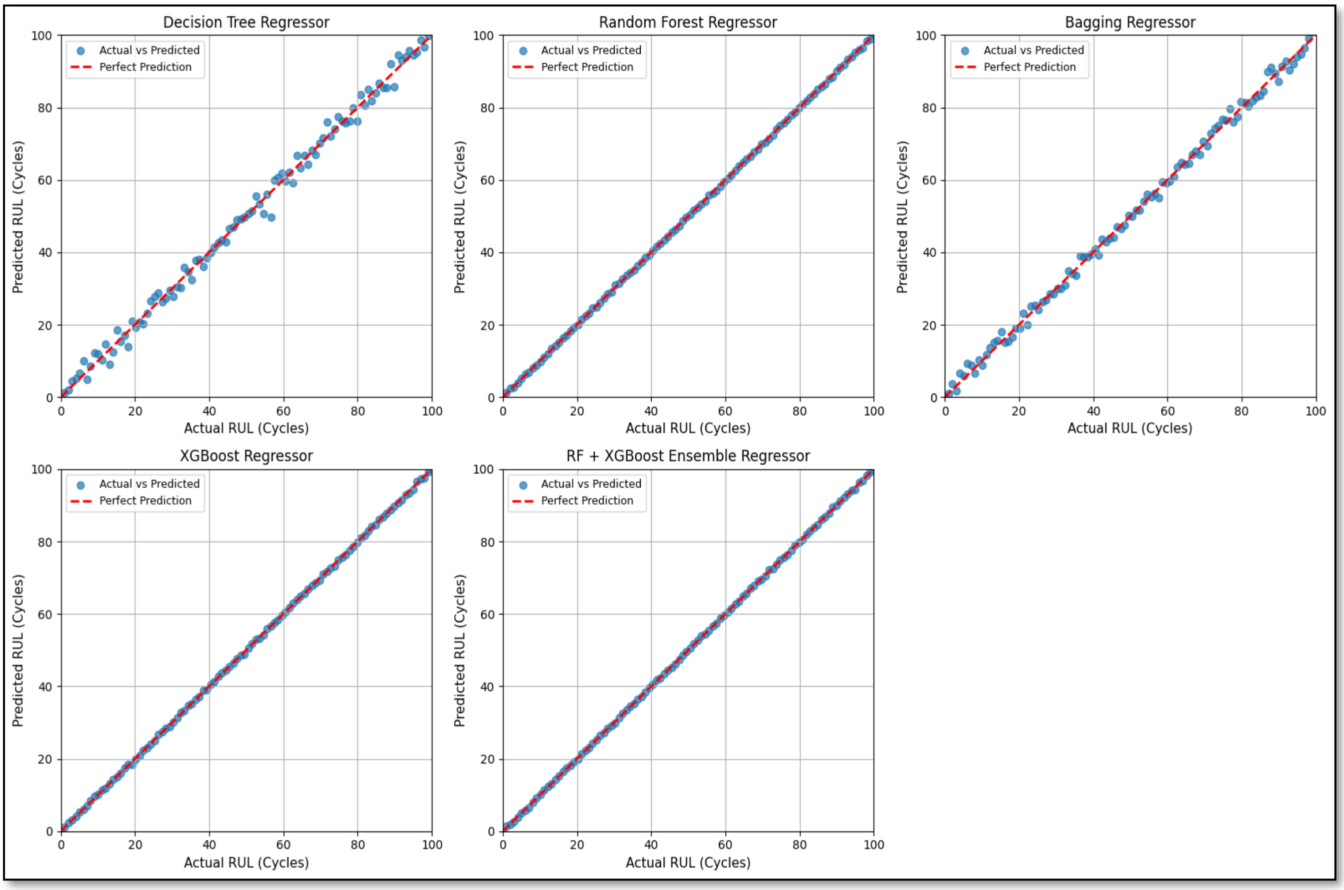

Figure 2. Plots for actual Vs predicted RUL for ML algorithms

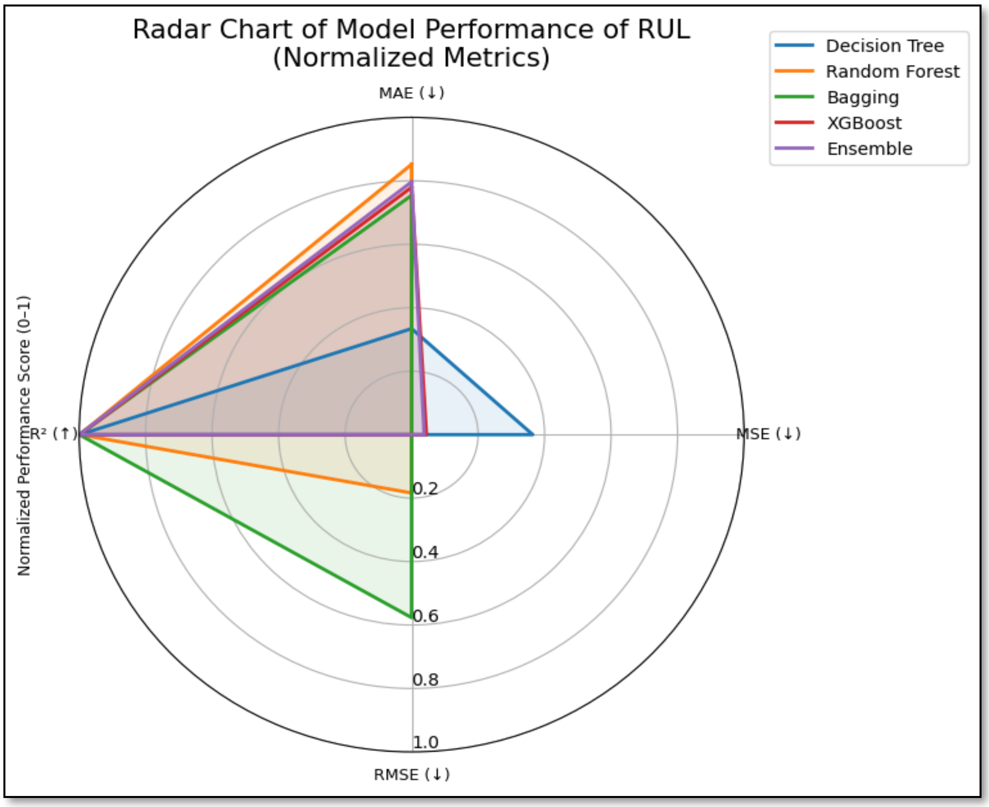

Figure 3. Radar chart

Whereas decision Tree lags behind here slightly. RF + XGB Ensemble performs best across all error metrics, shown by its large coverage toward the edge. The ensemble model achieved lowest RMSE and highest R^2, closely tracking actual RUL values with minimal deviation, indicating superior generalization performance.

5.2. SOC Estimation Results

As SOC estimate is critical for determining a battery's remaining capacity. Normalized Linear interpolation is one of the most basic ways for calculating SOC between known data points. SOC estimation input features include voltage, minimum voltage and maximum voltage. Output features are State of Charge (%) at a given moment. It measures charge level in real time. We have extracted the minimum and maximum voltages from data and calculated the SOC. In this method it keeps values within a common scale and prevents numerical overflow or underflow. It can also be fed directly into ML models that expect normalized input which makes models more robust across different battery chemistries. Table 4 below gives the SOC generated values for the given machine learning algorithms.

From the predictions it is clear that R2 values of all the ML methods are very reasonable. Apart to all the metrics performance ensemble method of random forest with XG boost has achieved Best overall performance, combining strengths of both models. Whereas bagging regressor very good and competitive on MAE/MSE, but slightly behind the ensemble. Random forest performs well, especially on error metrics, but slightly lower R^2. Decision tree performs moderate and decent error but significantly lower R^2. XG Boost gives highest error and lowest R^2 shown in figure 4. The radar distribution plot for all the four metrics is shown in the figure 5 below.

Finally, conclusion drawn is that the ensemble approach significantly outperformed all individual models in SOC estimation. XGBoost alone showed higher variance in predictions, likely due to overfitting on localized pattern

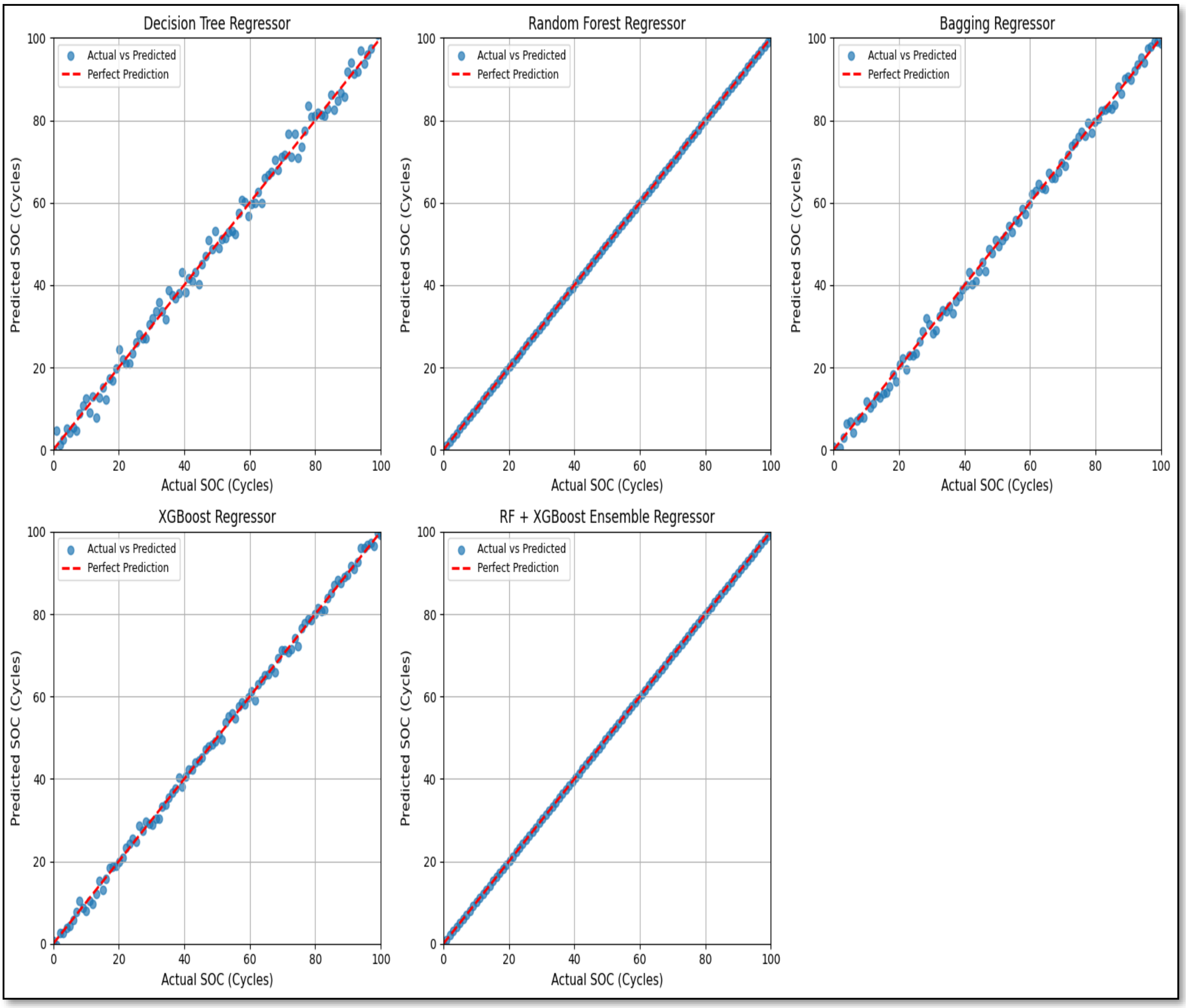

Figure 4. SOC estimation plots

Table 4. State of Charge (SOC) prediction results

Algorithm	MAE	MSE	RMSE	R-Squared
Decision tree Regressor	0.0086	0.0201	0.1419	0.9988
Random Forest Regressor	0.0073	0.0171	0.1307	0.9987
Bagging Regressor	0.0087	0.0215	0.1465	0.9998
XG Boost Regressor	0.0739	0.2074	0.4554	0.9977
Ensemble (RF+XG Boost)	0.0032	0.0157	0.1267	0.9999

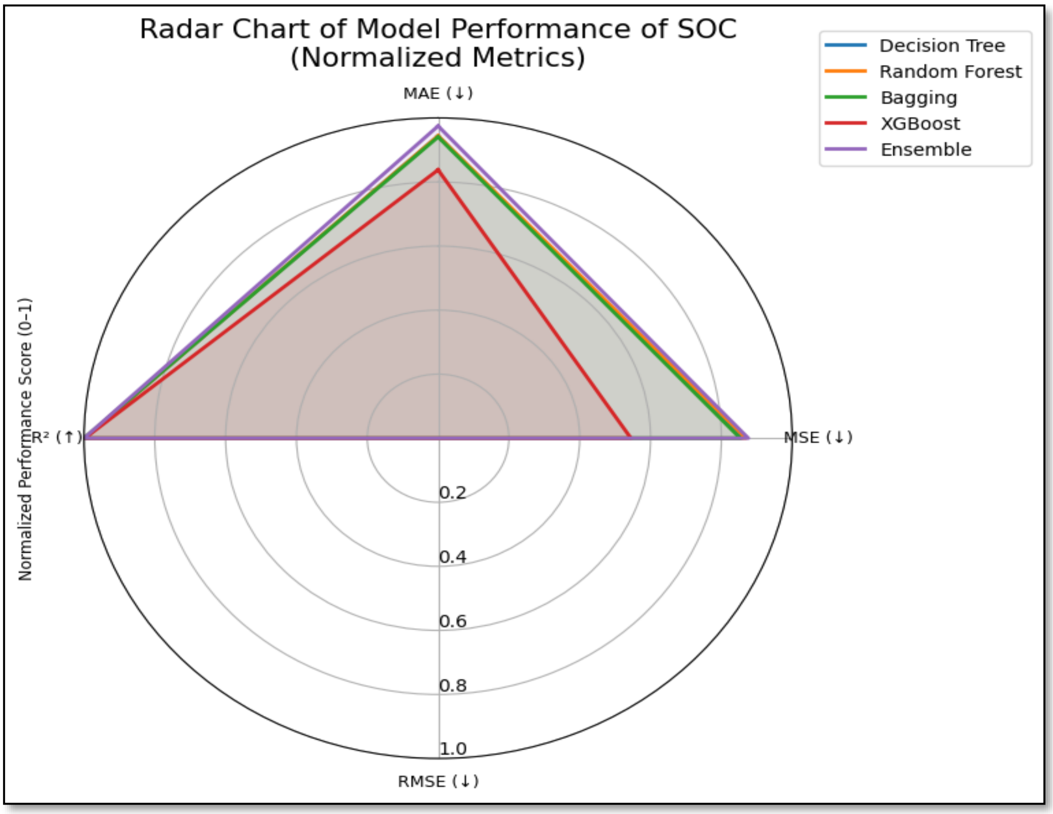

Figure 5. Radar chart of Model Performance

4.3. SOH Estimation Results

In this research State of Health is calculated by finding ratio of the RUL to total expected life based on availability of the data set features. Here it links the Cycle Index and future prediction (RUL) in one metric. SOH estimation input features include Cycle index, voltage, discharge time, charging time, etc. Output features are SOH (%) computed using the available RUL and cycle index. This indicates current battery health relative to its initial capacity. This method of estimation is simple, interpretable, and doesn't rely on voltage or current sensor data. Hence in machine learning, where RUL is predicted and SOH is derived. Table 5 provides the results obtained in prediction of SOH.

Figure 6 displays scatter plots of Actual vs. Predicted State of Health (SOH) for regression models like Decision Tree, Random Forest, Bagging, XG Boost, and a RF + XG Boost Ensemble. The RF + XG Boost Ensemble Regressor shows a nearly perfect alignment of the predicted points along the ideal line, signifying exceptional predictive accuracy. Random Forest and XG Boost models demonstrate high prediction fidelity, with minimal dispersion around the ideal line.

As illustrated in Figure 7, RF+ XGBoost ensemble demonstrates superior performance compared to the other models. It yields lowest MAE, MSE, and RMSE values while achieving highest R^2, approaching unity. These results indicate that the proposed ensemble delivers more accurate and stable SOH predictions. Table 6 below gives comparison of analysis of predicted proposed ensemble method with other techniques.

Table 5. State of Health (SOH) prediction results

Algorithm	MAE	MSE	RMSE	R-Squared
Decision tree Regressor	0.1273	0.0804	0.2835	0.9999
Random Forest Regressor	0.1298	0.0467	0.2168	0.9999
Bagging Regressor	0.1303	0.0552	0.2349	0.9999
XG Boost Regressor	0.1577	0.0513	0.2266	0.9999
Ensemble (RF+XG Boost)	0.1224	0.0391	0.1977	1.0

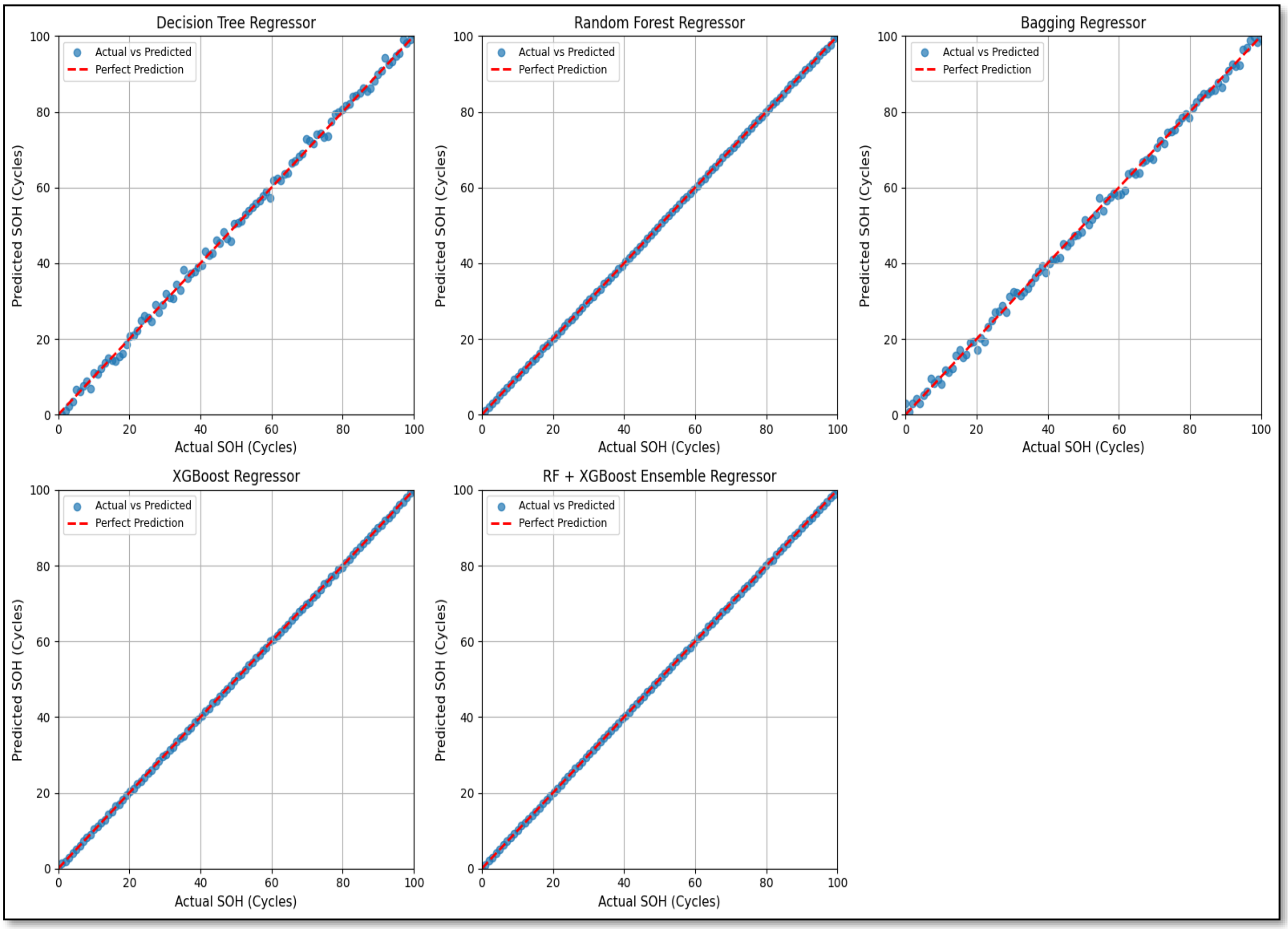

Figure 6. State of Health scatter plots

Table 6. Results comparison between existing techniques and proposed model

Ref	Prediction	Algorithm	MAE	MSE	RMSE	R-Square
[22]	RUL	Hybrid Model (KNN+XGB+RF)	1.00895	-	-	0.9964
[28]	RUL	RF+HHO	36.883	-	46.35	0.979
[28]	RUL	**Proposed Model (RF+XGB)**	**0.2201**	**1.0411**	**1.0821**	**0.9999**
[21]	SOC	RF	1.87	-	-	0.98
[23]	SOC	XGBR		10.03	2.56	-
[24]	SOC	RF	0.15	0.036	0.9	0.72
[24]	SOC	**Proposed Model (RF+XGB)**	**0.0032**	**0.0157**	**0.1267**	**0.9999**
[25]	SOH	CART	-	0.03	-	0.9995
[26]	SOH	XGB	0.1615	-	0.2132	-
[26]	SOH	**Proposed Model (RF+XGB)**	**0.1224**	**0.0391**	**0.1977**	**1.0**

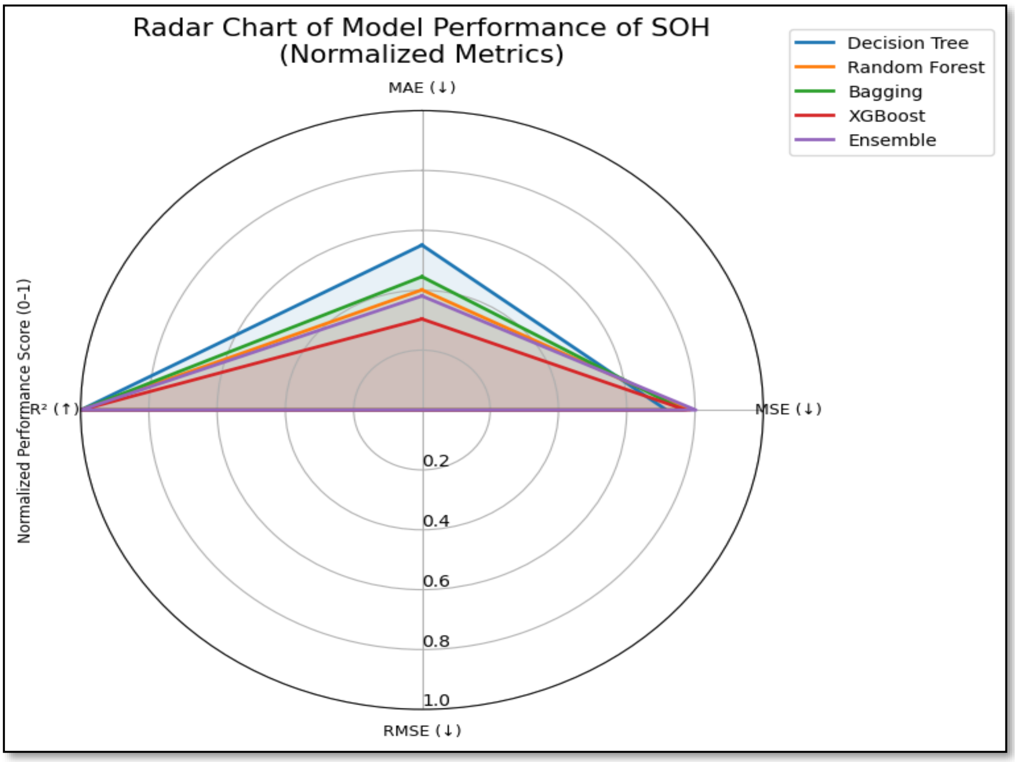

Figure 7. Radar Distribution Plot

6. Conclusion

This study evaluates five machine learning techniques namely Decision Tree, Random Forest, Bagging Regressor, XG Boost, and ensemble Random Forest and XG Boost using Grid search cross validation parameter tunning for predicting battery health indicators like Remaining Useful Life (RUL), State of Charge (SOC), and State of Health (SOH), using the HNEI battery dataset and experiments were carried in google co-lab. The results show that proposed ensemble algorithm systematically outperforms individual models across all prediction tasks with less computational time. The ensemble achieved a minimum RMSE of 1.08 with an R^2 value of 0.9999, reflecting high accuracy and strong reliability in RUL estimation. It was observed that for SOC prediction an RMSE of 0.1267 and an R^2 of 0.9999 with good accuracy. In the case of SOH estimation, lowest RMSE of 0.1977 and a near-perfect goodness of fit ($R^2 = 1.0$), highlighting its robustness and generalization capability.

Overall, Findings confirm that ensemble learning provides a reliable and accurate solution for battery health prognostics. This strong predictive performance of proposed model highlights its potential for real-time deployment in battery management systems for electric vehicle applications.

Future work will focus on incorporating deep learning models to better capture long-term degradation behaviour, as well as integrating interpretability techniques such as SHAP to enhance model transparency and trustworthiness. In addition, the proposed framework will be extended to multiple battery chemistries and a wider range of operating conditions to further improve robustness and real-world applicability.

References

[1] M.A. Hannan, M.M. Hoque, A. Hussain, Y. Yusof, P.J. Ker, State-of-the-art and energy management system of lithium-ion batteries in electric vehicle applications: Issues and recommendations. IEEE Access, 6, (2018) 19362-19378. https://doi.org/10.1109/ACCESS.2018.2817655

[2] S. Pan, L.M. Fulton, A. Roy, J. Jung, Y. Choi, H.O. Gao, Shared use of electric autonomous vehicles: Air quality and health impacts of future mobility in the United States. Renewable and Sustainable Energy Reviews, 149, (2021) 111380. https://doi.org/10.1016/j.rser.2021.111380

[3] N.H. Paulson, J. Kubal, L. Ward, S. Saxena, W. Lu, S.J. Babinec, Feature engineering for machine learning enabled early prediction of battery lifetime. Journal of Power Sources, 527, (2022) 231127. https://doi.org/10.1016/j.jpowsour.2022.231127

[4] C.L. Sravanthi, J.N. Chandra Sekhar, Advances in Nano Metal Oxides and Nanotechnology for Lithium-Ion Batteries in EV: A Comprehensive Review. Journal of Modern Technology, 2(01), (2025) 254-263. https://doi.org/10.71426/jmt.v2.i1.pp254-263

[5] M. Elmahallawy, T. Elfouly, A. Alouani, A.M. Massoud, A comprehensive review of lithium-ion

batteries modeling, and state of health and remaining useful lifetime prediction. Ieee Access, 10, (2022) 119040-119070. https://doi.org/10.1109/ACCESS.2022.3221137

[6] C.L. Sravanthi, J.N. Chandra Sekhar, N.C. Alluraiah, C. Dhanamjayulu, H. Kumar Pujari, B. Khan, An Overview of Remaining Useful Life Prediction of Battery Using Deep Learning and Ensemble Learning Algorithms on Data-Dependent Models, International Transactions on Electrical Energy Systems, 2025(10), (2025) 2242749. https://doi.org/10.1155/etep/2242749

[7] A. Haraz, K. Abualsaud, A. Massoud, State-of-health and state-of-charge estimation in electric vehicles batteries: a survey on machine learning approaches. IEEE Access, 12, (2024) 158110 – 158139. https://doi.org/10.1109/ACCESS.2024.3486989

[8] M.F. Ng, J. Zhao, Q. Yan, G.J. Conduit, Z.W. Seh, Predicting the state of charge and health of batteries using data-driven machine learning. Nature Machine Intelligence, 2(3), (2020) 161-170. https://doi.org/10.1038/s42256-020-0156-7

[9] D. Obuli Pranav, P.S. Babu, V. Indragandhi, B. Ashok, S. Vedhanayaki, C. Kavitha, Enhanced SOC estimation of lithium-ion batteries with RealTime data using machine learning algorithms. Scientefic Reports 14, (2024) 16036. https://doi.org/10.1038/s41598-024-66997-9

[10] V. Chandran, C.K. Patil, A. Karthick, D. Ganeshaperumal, R. Rahim, A. Ghosh, State of Charge Estimation of Lithium-Ion Battery for Electric Vehicles Using Machine Learning Algorithms. World Electric Vehicle Journal, 12(1), (2021) 38. https://doi.org/10.3390/wevj12010038

[11] H. Rauf, M. Khalid, N. Arshad, Machine learning in state of health and remaining useful life estimation: Theoretical and technological development in battery degradation modelling. Renewable and Sustainable Energy Reviews, 156, (2022) 111903. https://doi.org/10.1016/j.rser.2021.111903

[12] L.B. Chilakapati, G.T. Manohar, Power Quality Improvement in a Grid Connected Solar-PV System With Adaptive Neuro-Fuzzy Controller Based UPQC”. International Research Journal of Multidisciplinary Technovation 7(1), (2025) 264-279. https://doi.org/10.54392/irjmt25118

[13] C. Zhang, S. Zhao, Z. Yang, Y. Chen, A reliable data-driven state-of-health estimation model for lithium-ion batteries in electric vehicles. Frontiers in Energy Research, 10, (2022) 1013800. https://doi.org/10.3389/fenrg.2022.1013800

[14] H. Meng, Y.F. Li, A review on prognostics and health management (PHM) methods of lithium-ion batteries. Renewable and Sustainable Energy Reviews, 116, (2019) 109405. https://doi.org/10.1016/j.rser.2019.109405

[15] Y. Muratoglu, A. Alkaya, Unscented Kalman Filter based State of Charge Estimation for the Equalization of Lithium-ion Batteries on Electrical Vehicles. Engineering, Technology & Applied Science Research. 9, (2019) 4876–4882. https://doi.org/10.48084/etasr.3111

[16] C. Shan, C.S. Chin, V. Mohan, C. Zhang, Review of Various Machine Learning Approaches for Predicting Parameters of Lithium-Ion Batteries in Electric Vehicles. Batteries 10, (2024) 181. https://doi.org/10.3390/batteries10060181

[17] X. Li, D. Yu, V.S. Byg, S.D. Ioan, The development of machine learning-based remaining useful life prediction for lithium-ion batteries. Journal of Energy Chemistry, 82, (2023) 103-121. https://doi.org/10.1016/j.jechem.2023.03.026

[18] M. Zhang, D. Yang, J. Du, H. Sun, L. Li, L. Wang, K. Wang, A review of SOH prediction of Li-ion batteries based on data-driven algorithms. Energies, 16(7), (2023) 3167. https://doi.org/10.3390/en16073167

[19] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, M.J. Deen, A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life. IEEE transactions on industrial informatics, 17(5), (2020) 3478-3487. https://doi.org/10.1109/TII.2020.3008223

[20] D. Aloisio, G. Campobello, S.G. Leonardi, F. Sergi, G. Brunaccini, M. Ferraro, V. Antonucci, A. Segreto, N. Donato, Comparison of machine learning techniques for SoC and SoH evaluation from impedance data of an aged lithium ion battery. *Acta IMEKO*, 10(2), (2021) 80-87. https://doi.org/10.21014/acta_imeko.v10i2.1043

[21] S. Jafari, Y.C. Byun, S. Ko, A novel approach for predicting remaining useful life and capacity fade in lithium-ion batteries using hybrid machine learning. IEEE Access, 11, (2023) 131950-131963. https://doi.org/10.1109/ACCESS.2023.3329508

[22] S. Jafari, Z. Shahbazi, Y.C. Byun, S.J. Lee, Lithium-ion battery estimation in online framework using extreme gradient boosting machine learning approach. Mathematics, 10(6), (2022) 888. https://doi.org/10.3390/math10060888

[23] K. Akbar, Y. Zou, Q. Awais, M.J.A. Baig, M. Jamil, A machine learning-based robust state of health (SOH) prediction model for electric vehicle

batteries. Electronics, 11(8), (2022) 1216. https://doi.org/10.3390/electronics11081216

[24] S. Song, C. Fei, H. Xia, Lithium-ion battery SOH estimation based on XGBoost algorithm with accuracy correction. Energies, 13(4), (2020) 812. https://doi.org/10.3390/en13040812

[25] S. Jo, S. Jung, T. Roh, Battery State-of-Health Estimation Using Machine Learning and Preprocessing with Relative State-of-Charge. Energies 14, (2021) 7206. https://doi.org/10.3390/en14217206

[26] S. Jafari, Y.C. Byun, Optimizing battery RUL prediction of lithium-ion batteries based on Harris hawk optimization approach using random forest and LightGBM. IEEE Access, 11, (2023) 87034-87046. https://doi.org/10.1109/ACCESS.2023.3304699

[27] J.C. Sekhar, B. Domathoti, E.D. Santibanez Gonzalez, Prediction of battery remaining useful life using machine learning algorithms. Sustainability, 15(21), (2023) 15283. https://doi.org/10.3390/su152115283

[28] V.T. Ha, P.T. Giang, Experimental study on remaining useful life prediction of lithium-ion batteries based on three regression models for electric vehicle application. Applied Sciences, 13(13), (2023) 7660. https://doi.org/10.3390/app13137660

[29] C.L. Sravanthi, J.N. Chandra sekhar, Predicting Remaining Useful Life of Lithium-Ion Batteries for Electric Vehicles Using Machine Learning Regression Models. ITEGAM-JETIA. 11(51), (2025) 143-150. https://doi.org/10.5935/jetia.v11i51.1267

[30] K. Karthick, S. Ravivarman, R. Priyanka, Optimizing Electric Vehicle Battery Life: A Machine Learning Approach for Sustainable Transportation. World Electric Vehicle Journal, 15(2), (2024) 60. https://doi.org/10.3390/wevj15020060

[31] G. dos Reis, C. Strange, M. Yadav, S. Li, Lithium-ion battery data and where to find it. Energy and AI, 5, (2021) 100081. https://doi.org/10.1016/j.egyai.2021.100081

[32] Dataset Link. Available online: https://www.kaggle.com/datasets/ignaciovinuales/battery-remaining-useful-life-rul.

[33] M.Z. Naser, A.H. Alavi, Error Metrics and Performance Fitness Indicators for Artifcial Intelligence and Machine Learning in Engineering and Sciences. Architecture, Structures and Construction, 3, (2021) 499–517. https://doi.org/10.1007/s44150-021-00015-8

Authors Contribution Statement

C.L. Sravanthi: Conceptualization, Investigation, Data collection, Data analysis, Writing original manuscript. J.N. Chandra Sekhar: Supervision, Writing, review and editing. Both the authors read and approved the final version of the manuscript.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?

Yes

About the License