IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁒ Peer-reviewed / Refereed journal ⁒ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12216

Review on Optimization and Artificial Intelligence Algorithms for Effective Battery Management in EVs

Kalika Patil1, Mangal Patil2

Department of Electronics and Communication Engineering,

Bharati Vidyapeeth (Deemed to be University) College of Engineering Pune, India[1]

Department of Electronics and Communication Engineering,

Bharati Vidyapeeth (Deemed to be University) College of Engineering Pune, India[2]

Abstract: Battery management systems (BMS) play a vital role in the safety, efficiency, and longevity of electric vehicles (EVs). As electric mobility increases, the BMS has a critical impact on enhancing the overall vehicle performance and energy management, making its optimization vital. In this regard, this paper presents how advanced artificial intelligence (AI) algorithms involving machine learning (ML), deep learning (DL), and reinforcement learning (RL) can overcome the major issues confronting BMS: precise state-of-charge (SOC) estimation, state-of-health (SOH) prediction, thermal management, and charge-discharge efficiency. AI approaches such as XGBoost and CatBoost achieve high performance for SOC and SOH predictions, with metrics like MAE, RMSE, and R^2 reaching values of 2.243, 3.2, 0.99, and 17.1, 23.97, 0.99, respectively, showcasing the potential for superior accuracy and robustness. The integration of AI systems facilitates improved adaptability and intelligent energy distribution, propelling the journey toward a sustainable and efficient electric vehicle landscape.

Keywords: BMS, optimization EV , AI, ML, SOC, XGBoost

1 INTRODUCTION

The development of energy supply on a worldwide basis is key to achieving, maintaining, and even improving the quality of modern life. Currently, fossil fuel combustion dominates power grids. Renewable energy is mostly not part of the energy supply chain, where conventional fossil fuels still prevail, even though there is an ever-increasing global demand for clean energy. However, the natural resources of fossil fuels are limited, and increasing energy demands aggravate pollution. Environmental degradation is aggravated by inefficient centralized power generation systems, and structural adjustments need to be undertaken to facilitate the transition from conventional to renewable energy sources [2].

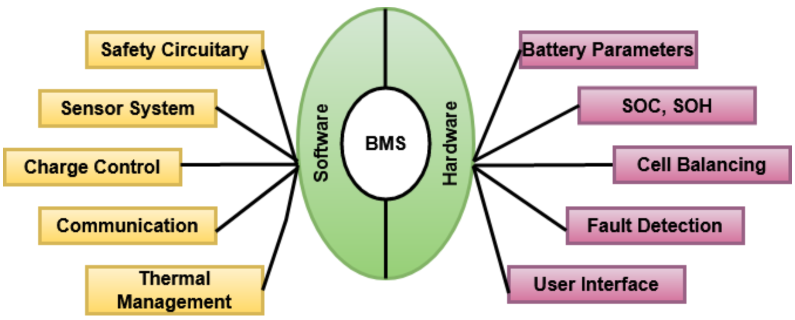

Fig. 1. Different parameters of Battery Management System [1]

As figure 1 shows the different parameters for battery management system in software and hardware scenario that affects the battery life span. To increase the use of renewable power generation, technological advancements offer the opportunity to tackle sustainability challenges; however, renewable energy sources need to be embedded into the energy supply system. Owing to the exhaustion of fossil fuels and the need to combat pollution, the application of renewable energy in industries has attracted extensive research in recent years. Solar and wind energy are among the most heavily discussed renewable energy resources owing to their sustainability and availability. Throughout history, these resources have been exploited by large corporations that require large-scale investment in infrastructure. However, in response to climate change and sustainability issues, consumers are actively involved in the generation of renewable energy [3].

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12216

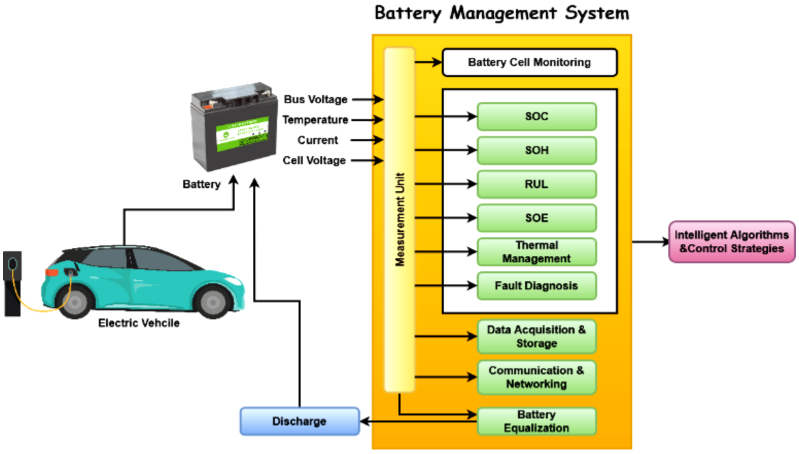

Fig. 2. Factors affected for battery Management System [1]

An additional aspect is the threshold in battery technology in power systems, which also makes the energy supply more stable and reliable. Emerging energy storage technologies, along with solar and wind power, have led to the creation of distributed generation systems as clean and renewable energy solutions. However, challenges remain, such as the high cost of energy production, optimal power efficiency, and consistent energy supply. This implies that the system includes unit size, which plays a significant role in the overall performance and scalability of the power system. This transition is best exemplified by electric vehicles (EV), which utilize electricity and one or more electric motors as their sources of energy and propulsion. EVs include bicycles, scooters, cars, trains, and vans. This study highlights the role of intelligent algorithms for accurate battery state estimation, such as SOC, SOH, and RUL. Additionally, several controller architectures have been investigated for battery balancing, fault detection, and thermal management. The intelligent controllers and algorithms used for BMS are also discussed, and limitations in implementations across all domains are explained with recommendations and future directions for improving accuracy, adaptability, and reliability [4].

Research gaps, such as the limited exploration of regression models for charging duration estimation and the absence of real-time energy optimization strategies, emphasize the need for more robust and scalable solutions. Furthermore, the integration of ML with physical models, while promising, still struggles to generalize across diverse battery chemistries and operational conditions.

2 LITREATURE REVIEW

In recent years, advanced battery management system (BMS) technology for electric vehicles (EVs) has received increasing research attention because of the increasing interest in the integration of artificial intelligence (AI) techniques for battery performance optimization in various studies. The primary subjects of these systems include SOC estimation, SOH prediction, charge-discharge efficiency, and thermal management. Figure 2 shows the different factors for BMS which affects the lifespan of battery. All these factors influence the battery longevity and safety of EV. ML has been widely used in BMS to forecast the status and performance of the battery cells. Some studies have used SVR, and random forest regression models are used for SOC and SOH estimation. Accurate prediction of battery conditions using these models would help in efficient energy management, leading to better EV performance. Recent studies have highlighted the fusion of multiple AI approaches to produce robust and accurate BMSs. Hybrid methods that integrate ML, DL, and RL techniques are emerging as effective solutions that significantly enhance battery health prediction, energy optimization, and fault detection. A hybrid model using fuzzy logic and network identifiers has been devised for the nonlinear optimization of the battery health of EVs, which has been shown to result in higher accuracy for SOH prediction and fault diagnosis.

In addition, people are interested in using AI for the thermal management of BMS. Good thermal management is necessary for battery safety and crucial for high-performance EVs. Neural network and decision tree machine learning models have been utilized to anticipate temperature variations, optimize the cooling system of battery packs, avoid overheating, and extend battery lifespans. Table 1 and table 2 shows the latest research from year 2023 and 2024.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12216

Table 1. Review from 2024

Ref	Challenges	Methods	Dataset	Research Gap
[5]	- Evaluating model accuracy in predicting SOC and temperature. - Handling outliers during data analysis and model training.	- Gradient descent, stochastic gradient descent, simulated annealing for	- Battery Initial Temperature influences charging rate and efficiency. - Room Temperature affects charging dynamics and performance.	- Lack of real-world implementation data for model validation. - Limited exploration of alternative optimization algorithms and techniques.
[6]	- Computational complexity for large-scale battery systems. - Generalization to different battery chemistries and aging profiles.	- Deep Reinforcement Learning for battery cycle optimization.	- Simplified lithium-ion battery model for simulations. - Performance metrics include cycle life and energy efficiency.	- Generalization of DRL policy across different battery types needed. - Computational efficiency for large-scale systems requires further investigation.
[7]	- Ensuring safety constraints during reinforcement learning actions. - Adapting to changes in battery dynamics effectively.	- Adaptive Gaussian process models - Twin-delayed DDPG algorithm	- Datasets include temperature and voltage data for battery charging. - Data used for training dynamic Gaussian process models.	- Ensuring safety constraints in RL-based battery charging optimization. - Addressing system safety to prevent irreversible battery damages.
[8]	- Limited understanding of modeling process visualization and theory. - Importance of data balance for optimal classification results.	- Five classification algorithms for crystal system classification.	- 339 types of lithium silicate cathode materials. - Data includes 11 columns of physical quantities.	- Limited exploration of additional classification algorithms. - Need for broader dataset diversity and validation.

Table 2. Literature from 2023

Ref	Methods	Challenges	Dataset	Research Gap
[9]	KNN algorithm; dataset divided into subsets for supervised learning.	Overfitting due to limited data; complexity from additional features.	Real-world dataset from 100+ users' charging sessions.	Limited focus on regression; emphasis on duration prediction for charging stations and DSOs.
[10]	SNN, NARX, and hybrid ANFIS architectures implemented in MATLAB.	Robustness in real-world conditions remains unproven.	Li-ion ECM RC battery model with input-output measurement data.	No explicit gaps mentioned; focus is on SOC estimator development and performance.
[11]	Feature screening, fusion, and acquisition probability evaluation.	Feature selection and fusion; adapting to varied driving scenarios.	NASA, MIT, and CALCE battery degradation datasets.	Real-world driving behavior impact and redundancy in feature sets.
[12]	Physics-informed neural networks (PINN) integrating ML and physics-based models.	Complexity of dynamic cell parameters; dataset generalization issues.	Lithium-ion Power Cell LP2 51Ah-BEV cells (NMC-622 cathode, graphite anode).	Generalization across scenarios and complexity of parameter estimation.
[13]	Machine learning; historical driving data for optimization.	Real-time energy management and computational resource demands.	EV data on battery capacity, routes, and weather.	Real-time energy optimization limitations and scalability issues.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⋇ Peer-reviewed / Refereed journal ⋇ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12216

[14]	RNNs with various optimizers; cross-validation for accuracy.	Balancing voltage-current weighting; custom model development.	Open-source battery cycling data from the Battery Research Group.	Custom models to improve middle charge estimations.
[15]	Convex optimization and dual loop DP algorithms.	Addressing performance and consumer attitudes toward EVs.	Comparative algorithm evaluation; no dataset specifics.	Impact of optimization strategies on consumer adoption.
[16]	Cloud computing and ML for pattern extraction.	Hardware constraints and predictive modeling at various timescales.	No dataset specifics mentioned; onboard chip data used.	Scalability and predictive modeling challenges across timescales.

Unified modeling frameworks: State-of-the-art AI techniques do not generalize, rather focus on SOC or SOH estimation. To date, there is no comprehensive framework to address all the BMS challenges including thermal management, charge-discharge, and energy distribution. Although AI approaches have demonstrated success, their integration with physics-based models to describe complex electrochemical dynamic continues to be underexplored.

Generalization across battery types: These models are mostly specific to lithium-ion battery types, as opposed to other, emerging chemistries (e.g. solid-state, sodium-ion). EV battery conditions (e.g., temperature-usage patterns) change over time. Existing AI solutions cannot dynamically adapt to such changes.

Data scarcity for edge cases: Training datasets may not fully reflect extreme conditions like overcharging, deep discharging, or thermal runaway, making it challenging to accurately predict results using AI. Due to hardware constraints, high-end AI models are still primarily applied at low power, edge-computing for electric vehicles.

Interdisciplinary collaboration: Using our proprietary AI technology, we can develop Decentralized BMS capable of addressing new technologies at the top of the stack, which the existing models are often too leaky to handle, and therefore do not integrate or 129ptimize: However, there have been no studies on how AI in the BMS and the overall optimization of the BMS would improve battery recycling, second life applications and overall sustainability.

3 MACHINE LEARNING FOR EVBO

The integration of machine learning techniques into electric vehicle (EV) charging systems has opened new avenues for optimizing battery performance and energy management. A study on EV charging session classification using ML and provided insights into forecasting charging durations but remained limited to classification-based approaches without exploring regression models. Similarly, machine learning-based energy optimization systems for EVs have focused on historical driving data to predict energy requirements. However, challenges like frequent recharging and high computational resource demands underscore the need for scalable solutions. These studies highlight the growing role of ML in addressing energy efficiency and operational challenges in EV ecosystems.

3.1 Deep Learning for Battery State-of-Charge and Health Monitoring

Recent advancements in deep learning have significantly enhanced battery state-of-charge (SOC) estimation and health monitoring capabilities. Approaches such as the integration of shallow neural networks (SNN) and NARX architectures have demonstrated robustness against parameter variations. Additionally, hybrid adaptive neural fuzzy inference systems (ANFIS) have shown promise in estimating SOC under harsh conditions. Despite these advancements, practical challenges like the inability to accurately capture middle SOC ranges and reliance on specific datasets remain unresolved. These limitations call for further exploration of customized models that generalize well across diverse battery chemistries and conditions.

3.2 Feature Engineering and Fusion Techniques in Battery Health Prediction

The use of feature engineering and fusion techniques has improved the accuracy of battery health predictions. Studies employing scenario-based feature fusion and acquisition probability evaluation have demonstrated enhanced performance in monitoring battery health under dynamic conditions. However, the redundancy of features and inaccuracies stemming from real-world driving behaviors present significant challenges. Research leveraging NASA, MIT, and CALCE battery degradation datasets highlights the importance of robust feature screening techniques to mitigate these challenges, paving the way for more accurate and reliable battery health monitoring frameworks.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁘ Peer-reviewed / Refereed journal ⁘ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12216

3.3 Hybrid Modeling Approaches for Lithium-Ion Batteries

Hybrid modeling approaches that combine physics-based and data-driven techniques are gaining traction in lithium-ion battery research. Physics-informed neural networks (PINNs) have been successfully applied to state estimation tasks, offering a balance between complexity and accuracy. Despite their potential, these models require large datasets for effective training and often struggle with generalization to unseen scenarios. The integration of domain knowledge from physics-based models with the predictive power of machine learning has been proposed as a promising avenue to overcome these limitations, especially in the context of dynamic cell parameters.

4 PREDICTION METHOD

4.1 Dataset

Dataset 1, used in the first two case studies, was generated based on the battery degradation simulation model in the MATLAB Simulink environment 42. The dataset contains 945 battery aging tests with different parameters including the state of charge (SOC), depth of discharge (DOD), temperature, and circulating current (C-rate). Our model-based simulation is flexible in that researchers can define the battery types according to the manufacturer's long-sheet and can configure it to simulate the effects of the ambient temperature, dynamics of the internal resistance, and aging characteristics. This active model, while realistically penalizing linear amp-hour capacity degradation, may ultimately require more sophisticated predictive algorithms to account for real-world nonlinear capacity fading. While singular output models remain dominant in predicting remaining useful life (RUL) in linear degradation scenarios, advanced techniques are best suited for state representation in nonlinear scenarios because they can easily capture complex behavior.

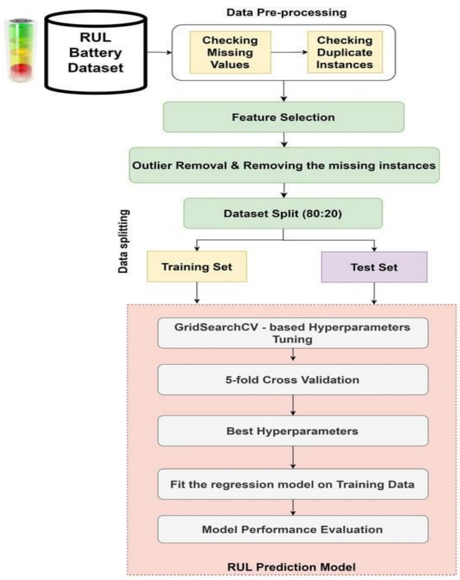

Fig. 3. RUL prediction model [19]

Dataset 2: Data from 124 commercial Li-ion phosphate/graphite cells (A123 system, model: APR18650M1A, nominal capacity: 1.1 Ah) subjected to fast-charging protocols until end-of-life (EOL, defined as 80% of their initial capacity) comprise Dataset 2. Using a standard 30 °C controlled environment, different fast-charging policies (at various C rates) were implemented over the course of up to three separate stages. Charging started with a C-rate (C1), moved to an intermediate C-rate (C2), and ended with discharge at a rate of 4 C until the voltage decreased to a cutoff range of 3.6 V–2. 0 V The dataset was used in case 3 and contains valuable information about the influence of fast-charging strategies on the performance and degradation of a battery.

The third dataset is based on a study by the Hawaii Natural Energy Institute on nickel manganese cobalt cobalt oxide (NMC-LCO) batteries. In particular, this dataset consists of 14 standalone 18,650 cells (nominal 2.8 Ah) that were cycled beyond 1000 charge-discharge cycles under standardized conditions.

The batteries were maintained at a temperature of 25 °C and charged with a constant current–constant voltage (CC-CV) at a rate of C/2 (i.e., one half of the battery capacity per hour). Discharge was conducted at 1.5 C, or 1.5-times the battery capacity in one hour. This dataset provides in-depth insights into the cycling behavior and degradation kinetics of NMC-LCO batteries under controlled laboratory conditions.

DOI: 10.17148/IARJSET.2025.12216

4.2 Data Preprocessing

As Figure 3 shows first part of system is data processing. The dataset was thoroughly checked for any missing and duplicate values. The absence of missing or duplicate instances in the data was confirmed after thorough preprocessing. This is an important step to preserve the reliability of the dataset, as removing missing values avoids biased analyses or incorrect predictions (especially in machine learning models). Moreover, this step also guarantees that each record is unique, an essential factor for statistical analyses and machine learning model training. This preprocessing is important as redundant data can affect the efficiency and accuracy of the model.

Feature Selection

By performing feature selection, the predictors best suited to the target variable, Remaining Useful Life (RUL), were selected. The analysis identified the following correlations. In contrast, the cycle index showed a perfect negative correlation with the RUL (correlation coefficient = −1.00). A positive correlation was observed between the maximum discharge voltage and RUL with a correlation coefficient of 0.78.

During the charging process, the minimum voltage showed a negative relationship with the RUL (R = −0.76).

Outlier Detection

The negative correlation between the cycle index and RUL was so strong that one could argue for its use as a feature. However, it was decided not to include this within the model. This decision was motivated by the risk of overfitting because the model could rely on this inverse relationship, which would reduce its generalizability to unseen data. Overfitting occurs when the model maps noise or spurious patterns that are peculiar to the training data and not generalizable relations. Beyond the cycle index, this also captures external factors such as material degradation and environmental conditions that could affect RUL and are otherwise not captured by this feature. The model was trained on only the most relevant features, significantly contributing to its robustness and overall predictive capabilities. Outlier detection was performed to assess extremely high or extremely low values that could bias the analysis or negatively influence the model training. Based on the study of outliers in this dataset, the following can be obtained:

Z-Score Analysis: For numerical features, we used the Z-score method, in which we flagged data points with Z-scores beyond a certain threshold (for example, $|Z| > 3$) as outliers. This method is useful for detecting outlier values that are redundant in a distribution.

Interquartile Range (IQR) method: The I The IQR method based on the spread of the middle 50% of the data was used to detect outliers. Data points that fell below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$ were considered outliers.
Visualization Techniques: Data visualization of RUL, voltage, and cycle index using box and scatter plots. Visual inspection was performed to identify patterns or groups of outliers.

Model Selection.

Several machine-learning algorithms have been considered for predicting the Remaining Useful Life (RUL) of batteries. The selection process was guided by the characteristics of the dataset, relationships between features, and the need to capture both linear and nonlinear patterns. The following models are shortlisted:

Linear Regression: Establish baseline performance and understand linear relationships.
Random forest regressors: Capture nonlinear relationships and assess feature importance owing to its ensemble-based structure.
XGBoost (extreme-gradient boosting): It was selected for its ability to handle structured datasets efficiently, capturing complex interactions while reducing overfitting through built-in regularization techniques.
Support Vector Regression (SVR): Model complex relationships using kernel functions.
Neural Networks: To leverage their capacity to model highly nonlinear interactions and dependencies in a dataset.
XGBoost and CatBoost are chosen for their high accuracy, efficiency, and ability to handle complex battery datasets. XGBoost excels in scalability and managing missing data, while CatBoost effectively handles categorical features and minimizes overfitting, making both ideal for SOC and SOH predictions in Battery Management Systems.

4.3 XGBoost Algorithm

XGBoost is a highly efficient and scalable implementation of gradient boosting. It is particularly well suited for predictive tasks that require both accuracy and speed, such as battery health monitoring. The key features of XGBoost include the following.
Gradient Boosting Framework: XGBoost builds an ensemble of decision trees sequentially, where each tree corrects the errors of the previous trees, minimizing a specified loss function.

DOI: 10.17148/IARJSET.2025.12216

Regularization: XGBoost incorporates L1 (lasso) and L2 (ridge) regularization to prevent overfitting, which is crucial when working with datasets prone to noise or redundant features.
Handling Missing Data: XGBoost handles missing values by learning the optimal splitting directions during the training.
Parallelization: Unlike traditional gradient boosting methods, XGBoost supports parallelized computations, thereby significantly reducing the training time.
Customizable Objective Functions: This allows the use of user-defined loss functions, making it versatile for diverse applications.

Model Training with XGBoost.
XGBoost was trained on a pre-processed training dataset. The hyperparameters were tuned using a grid search to optimize the performance of the model. The key hyperparameters tuned include the following:
Learning Rate (η): The step size is controlled at each boosting step. A smaller learning rate with more boosting rounds generally improves the performance.
Max Depth: The maximum depth of each decision tree is determined, allowing the model to capture intricate data patterns.
Subsample: The fraction of training samples used for each tree helps to prevent overfitting.
Gamma: The minimum loss reduction required to make a further partition on a leaf node controls tree complexity.
Colsample_bytree: The fraction of features used to construct each tree improves the robustness of the model.

4.4 CATBoost Algorithm.
CatBoost is a gradient-boosting algorithm that optimizes the loss function by iteratively adding decision trees to minimize the errors. The CatBoost algorithm is explained step-by-step using the following mathematical notations:
Given a dataset$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, where $\mathbf{x}_i \in \mathbb{R}^d$ represents the features and $y_i \in \mathbb{R}$ (for regression) or $y_i \in \{0,1\}$ (for binary classification) represents the target. · The goal is to find a model $F(\mathbf{x})$ that minimizes the loss function $\mathcal{L}$,which is defined as:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \ell\left(y_i, F(\mathbf{x}_i)\right) \quad (1)$$

Where ℓ is the loss function, such as Mean Squared Error (MSE) for regression

$$\ell\left(y_i, F(\mathbf{x}_i)\right) = \frac{1}{2}\left(y_i - F(\mathbf{x}_i)\right)^2, \quad (2)$$

Log-Loss for classification:

$$\ell\left(y_i, F(\mathbf{x}_i)\right) = -\left(y_i \log\hat{p}_i + (1 - y_i)\log(1 - \hat{p}_i)\right), \quad (3)$$

With $\hat{p}_i = \sigma\left(F(\mathbf{x}_i)\right)$ and $\sigma(z) = \frac{1}{1+e^{-z}}$ (sigmoid function).

Gradient Boosting Framework
The model $F(\mathbf{x})$ is represented as an additive ensemble of decision trees:

$$F_t(\mathbf{x}) = F_{t-1}(\mathbf{x}) + \eta h_t(\mathbf{x}), \quad (4)$$

Where $F_t(\mathbf{x})$ is the model at iteration t, $\eta$is the learning rate, $h_t(\mathbf{x})$ is the base learner (decision tree) added at iteration t At each iteration, $h_t(\mathbf{x})$ is trained to minimize the residuals (the gradient of the loss function):

$$r_i^{(t)} = -\frac{\partial \ell\left(y_i, F_{t-1}(\mathbf{x}_i)\right)}{\partial F_{t-1}(\mathbf{x}_i)}. \quad (5)$$

For categorical features x_j, CatBoost uses an ordered target statistics approach: A random permutation π of the dataset indices is applied. For each sample i,the mean target value of x_j is computed only using preceding samples in the permutation:

$$\text{Encoded}\left(x_{ij}\right) = \frac{\sum_{k<\pi(i)} y_k \cdot \mathbb{I}\left(x_{kj} = x_{ij}\right)}{\sum_{k<\pi(i)} \mathbb{I}\left(x_{kj} = x_{ij}\right) + \alpha}, \quad (6)$$

where ll is the indicator function, and $\alpha > 0$ is a smoothing parameter to avoid overfitting

CatBoost constructs trees symmetrically:

· For each split, both branches have the same structure · Fora node n,the split is chosen to maximize the reduction in the loss function:

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12216

$$\Delta\mathcal{L} = \mathcal{L}_{\text{parent}} - (\mathcal{L}_{\text{left}} + \mathcal{L}_{\text{right}}), \quad (7)$$

where $\mathcal{L}_{\text{left}}$ and $\mathcal{L}_{\text{right}}$ are the losses for the left and right child nodes.
The final prediction is an ensemble of all trees:

$$F_T(\mathbf{x}) = \sum_{t=1}^{T} \eta\, h_t(\mathbf{x}), \quad (8)$$

where T is the total number of iterations. For regression, the prediction is:

$$\hat{y}_i = F_T(\mathbf{x}_i),$$

and for classification:

$$\hat{p}_i = \sigma(F_T(\mathbf{x}_i)) = \frac{1}{1+e^{-F_T(\mathbf{x}_i)}}. \quad (9)$$

L2 Regularization: Adds a penalty term to the loss:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \ell(y_i, F(\mathbf{x}_i)) + \lambda \parallel F \parallel^2 \quad (10)$$

where λ controls the regularization strength.
Learning Rate (η) : Controls the contribution of each tree to the final ensemble.

Evaluation Metrics

The MAE is determined using $$MAE = \frac{|(y_t - y_p)|}{n} \quad (11)$$

The MSE is determined using $$MSE = \frac{\Sigma\,(y_t - y_p)^2}{n} \quad (12)$$

The RMSE is determined using $$RMSE = \sqrt{\frac{\Sigma\,(y_t - y_p)^2}{n}} \quad (13)$$

The R-Squared, or coefficient of determination, is calculated by

$$R^2 = 1 - \frac{\Sigma\,(y_t - y_p)^2}{\Sigma\,(y_t - \bar{y}_t)^2} \quad (14)$$

Here, $\bar{y}_\iota$ is the mean of all of the actual values.

5 RESULTS AND DISCUSSION

The performance of the XGBoost and CatBoost models on the Remaining Useful Life (RUL) prediction task was evaluated using multiple metrics, including (Mean Absolute Error (MAE), MSE (Mean Squared Error (MSE), RMSE (Root Mean Squared Error (RMSE), and Coefficient of Determination (R ^{2}) across both the training and test datasets. A detailed analysis is provided in table 3.

Table 3. Comparison OF XGBOOST and CatBoost Algorithm

	Train Set				**Test Set**			
ML	**MAE**	**MSE**	**RMSE**	**RSquared**	**MAE**	**MSE`**	**RMSE**	**RSquared**
XGBoost	2.243	10.62	3.2	0.99	8.1	245.9	15.8	0.99
CatBoost	15.15	431	20.78	0.995	17.1	574.95	23.97	0.99

XGBoost demonstrated exceptional predictive accuracy and generalization ability as shown in figure 4(a) and figure 5, as reflected by its consistently low error metrics and high R^2 values in both the training and test sets. The MAE was 2.243 and the RMSE was 3.260, indicating a precise prediction with minimal deviation from the true values. An R^2 value of 0.999 confirmed a nearly perfect fit to the training data. The model maintained strong performance, with an MAE of 8.191 and an RMSE of 15.684. The high R^2 value of 0.997 highlights the robustness of the generalization capability of the model for unseen data. The performance of XGBoost can be attributed to its gradient-boosting framework, which efficiently handles complex interactions between features and prevents overfitting through regularization.

CatBoost also shows strong performance shown in figure 4 (b) , but slightly lags behind XGBoost in terms of predictive accuracy. MAE and RMSE were 15.15 and 20.782, respectively, with an R^2 value of 0.995. While the model fits the

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁘ Peer-reviewed / Refereed journal ⁘ Vol. 12, Issue 2, February 2025

DOI: 10.17148/IARJSET.2025.12216

training data well, the error metrics are higher than those of XGBoost, suggesting less precise predictions. The test results revealed an MAE of 17.122 and RMSE of 23.978, with an R^2 value of 0.994. These metrics, although competitive, indicate a marginal drop in accuracy when compared to XGBoost.

The performance of CatBoost is notable for its ability to handle categorical data effectively and its gradient-boosting-based approach. However, additional fine-tuning may be required to match the performance of XGBoost for this specific application.

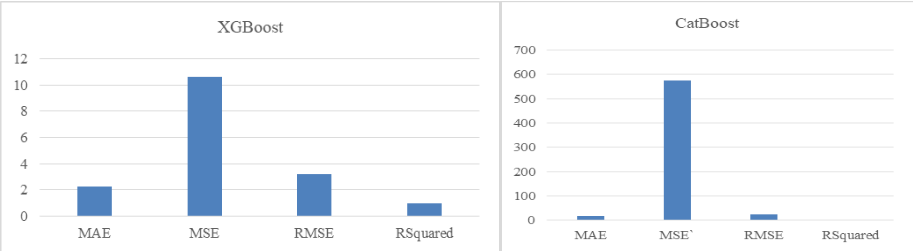

Fig. 4. Evaluation of Train data a) XGBoost b) CatBoost

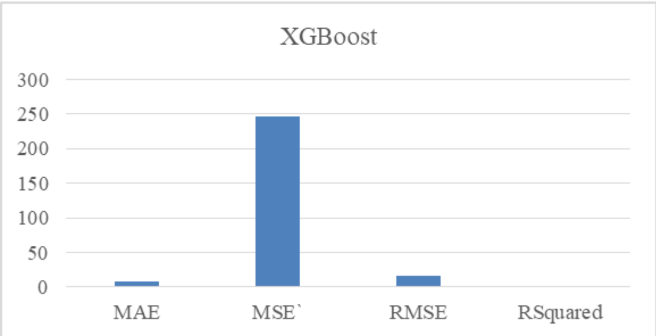

Fig. 5. Evaluation of XGBoost for Test data

The superiority of CatBoost over XGBoost in certain contexts lies in its ability to handle categorical features more effectively, thanks to its unique ordered boosting and preprocessing techniques. CatBoost also reduces overfitting through novel approaches like oblivious decision trees, which ensures stability and consistency in performance across datasets. On the other hand, XGBoost outperforms CatBoost in scenarios with larger datasets and numerical features, leveraging its parallel processing capabilities and efficient gradient boosting mechanism. The choice between the two often depends on the nature of the dataset (e.g., categorical vs. numerical dominance), computational constraints, and specific application requirements.

6 CONCLUSION

This paper presents AI algorithms and comparision, including machine learning (ML), deep learning (DL) for battery management system (BMS) optimization in electric vehicles (EVs). Thermal management of EV batteries is essential, and these AI techniques play a vital role in addressing significant roadblocks faced by Evs, including the accurate estimation of state-of-charge (SOC) and state-of-health (SOH) prediction. Incorporating artificial intelligence techniques can lead to more accurate information gathering, intelligent computation, and data-driven decisions, resulting in better battery management systems to ensure longer battery lifetime, efficient performance, and safety. The continuous improvement of these technologies will also lead to smarter and more efficient BMS systems designed for Evs, which will further accelerate EV adoption, resulting in a cleaner and more sustainable future. In this study, an ensemble of different machine learning algorithms based on datasets of EV performance was used to enhance prediction accuracy and reduce mean absolute errors, which indicates that future studies can benefit from developing more sophisticated new AI algorithms or hybrid models and applying them to other EV domains, such as energy management and autonomous driving. This study is limited by the reliance on simulated data and specific AI models, which may not fully capture real-world battery behavior and scalability across diverse EV systems.

DOI: 10.17148/IARJSET.2025.12216

REFERENCES

[1]. Manoj,& Josh, F. T. (2023). A comprehensive review on optimization and artificial intelligence algorithms for effective battery management in EVs. International Journal of Electrical and Electronic Engineering & Telecommunications, 12(5), 334-341. https://doi.org/10.18178/ijeetc.12.5.334-341

[2]. Boddapati, & Daniel, S. A. (2022). Design and feasibility analysis of hybrid energy-based electric vehicle charging station. Distributed Generation & Alternative Energy Journal, 37(1), 41-72.

[3]. Bhargavi, & Jayalakshmi, N. (2019). A new control strategy for plug-in electric vehicle of DC microgrid with PV and wind power integration. Journal of Electrical Engineering & Technology, 14, 13-25.

[4]. Hu, X., Deng, Z., Lin, X., Xie, Y., & Teodorescu, R. (2021). Research directions for next-generation battery management solutions in automotive applications. Renewable and Sustainable Energy Reviews, 152, 111695.

[5]. Gupta, S. K. (2024). Mathematical analysis and non-linear optimization of EV battery health using fuzzy logic and neural computing techniques. Communications on Applied Nonlinear Analysis, 31(2s), 84-100.

[6]. Baccari, Tipaldi, & Mariani, V. (2024). Deep reinforcement learning for cell balancing in electric vehicles with dynamic reconfigurable batteries. IEEE Transactions on Intelligent Vehicles. https://doi.org/10.1109/TIV.2024.3369107

[7]. Chowdhury, M. A., Al-Wahaibi, S. S. (2024). Adaptive safe reinforcement learning-enabled optimization of battery fast-charging protocols. AIChE Journal, e18605.

[8]. Yin, Y., Xu, G., Xie, Y., Luo, Y., Wei, Z., & Li, Z. (2024). Utilizing deep learning for crystal system classification in lithium-ion batteries. Journal of Theory and Practice of Engineering Science, 4(3), 199-206.

[9]. Matrone, S., Ogliari, E. G. C., Nespoli, A., Gruosso, G., & Gandelli, A. (2023). Electric vehicles charging sessions classification technique for optimized battery charge based on machine learning. IEEE Access, 11, 52444-52451.

[10]. Tudoroiu, N., Zaheeruddin, M., Tudoroiu, R. E., Radu, M. S., & Chammas, H. (2023). Intelligent deep learning estimators of a lithium-ion battery state of charge design and MATLAB implementation—A case study. Vehicles, 5(2), 535-564.

[11]. Jiang, Zhang, J., Jiang, W., Ren, Y., Lin, J., Khoo, E., & Song, Z. (2023). Driving behavior-guided battery health monitoring for electric vehicles using machine learning. arXiv preprint arXiv:2309.14125.

[12]. Singh, Ebongue, Y. E., Rezaei, S., & Birke, K. P. (2023). Hybrid modeling of lithium-ion battery: Physics-informed neural network for battery state estimation. Batteries, 9(6), 301.

[13]. Padmavathy, R., Greeta, T., & Divya, K. (2023). A machine learning-based energy optimization system for electric vehicles. E3S Web of Conferences, 387, 04008.

[14]. Sadykov, M., Haines, S., Broadmeadow, M., Walker, G., & Holmes, D. W. (2023). Practical evaluation of lithium-ion battery state-of-charge estimation using time-series machine learning for electric vehicles. Energies, 16(4), 1628.

[15]. Wang, X., Ji, W., & Gao, Y. (2023). Optimization strategy of the electric vehicle power battery based on the convex optimization algorithm. Processes, 11(5), 1416. https://doi.org/10.3390/pr11051416

[16]. Shi, Zhao, Eze, Wang, Z., Wang, Lian & Burke, A. F. (2023). Cloud-based artificial intelligence framework for battery management system. Energies, 16(11), 4403. https://doi.org/10.3390/en16114403