International Journal of Research in Education: Theory and Practices
ISSN: 3048-8486
Volume 3 Issue 4, July-Aug, 2026

Machine Learning-Based Intelligent Prediction of Electric Vehicle Battery Health for Enhanced Performance and Lifetime Estimation

CH.Sravan Kumar[1], Palle Vasantha[2]

[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

Abstract. The increasing adoption of electric vehicles (EVs) has created a growing demand for reliable battery health monitoring systems that can improve operational efficiency and extend battery service life. One of the most important performance indicators of a lithium-ion battery is its State of Health (SOH), which reflects the remaining capacity and overall condition of the battery throughout its lifecycle. Accurate SOH estimation enables timely maintenance, minimizes unexpected failures, and improves the reliability of electric transportation systems. This research presents a machine learning framework for predicting EV battery SOH using two ensemble learning algorithms, namely Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM). The proposed framework utilizes battery operating parameters including voltage, current, temperature, and charging history to develop predictive models capable of learning battery degradation patterns. The collected dataset undergoes preprocessing, feature engineering, and train-test partitioning before model training. Performance evaluation is carried out using statistical measures such as R^2 Score, Adjusted R^2, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Experimental findings demonstrate that both algorithms achieve excellent prediction accuracy, with XGBoost showing slightly superior performance compared to LightGBM. The developed system provides an efficient solution for real-time battery health assessment, supporting intelligent maintenance scheduling, reducing operational costs, and improving battery lifespan in modern electric vehicles.

keywords: Electric Vehicles (EV), Battery State of Health, Battery Management System, Machine Learning, XGBoost, LightGBM, Battery Performance Prediction, Lithium-Ion Battery, Predictive Analytics

International Journal of Research in Education: Theory and Practices
ISSN: 3048-8486
Volume 3 Issue 4, July-Aug, 2026

I. Introduction

The rapid transition toward environmentally sustainable transportation has significantly accelerated the adoption of electric vehicles across the world. Governments, automobile manufacturers, and environmental organizations are encouraging the replacement of conventional fuel-powered vehicles with electric alternatives to reduce greenhouse gas emissions and dependence on fossil fuels. Although electric vehicles provide several environmental and economic advantages, battery degradation remains one of the most critical challenges affecting their long-term performance and reliability. Since lithium-ion batteries represent the most expensive component of an EV, maintaining their health is essential for ensuring vehicle efficiency, driving range, and operational safety. Consequently, intelligent battery monitoring has become a major research focus in modern transportation systems.

Battery degradation is an unavoidable process caused by repeated charging and discharging cycles, high operating temperatures, excessive current loads, aging mechanisms, and varying environmental conditions. As battery capacity gradually decreases, the driving range of electric vehicles is reduced, maintenance expenses increase, and the possibility of unexpected battery failures becomes higher. Therefore, continuously monitoring battery condition is essential for maximizing battery utilization while preventing sudden performance degradation. Among several battery health indicators, the State of Health (SOH) is considered the most reliable parameter because it directly reflects the remaining battery capacity compared to its original capacity.

Traditional Battery Management Systems (BMS) primarily rely on mathematical models, equivalent circuit representations, empirical equations, and sensor-based monitoring techniques to estimate battery condition. Although these methods perform reasonably well under controlled operating conditions, they often fail to accurately capture the complex nonlinear degradation characteristics exhibited by lithium-ion batteries operating in real-world environments. Battery aging depends on numerous interacting variables including charging behavior, discharge rates, temperature variations, current fluctuations, and operational history. These nonlinear relationships are difficult to represent using conventional analytical models.

Recent advances in artificial intelligence and machine learning have opened new opportunities for battery health prediction. Machine learning algorithms can automatically identify hidden relationships among multiple battery parameters without requiring explicit physical modeling. By learning degradation behavior directly from historical operational data, these algorithms provide more accurate, adaptive, and scalable prediction models suitable for practical deployment. Ensemble learning techniques have demonstrated particularly promising performance because they effectively handle nonlinear relationships, high-dimensional datasets, and complex feature interactions.

Among various machine learning algorithms, XGBoost and LightGBM have gained considerable attention due to their exceptional prediction accuracy, computational efficiency, and robustness against overfitting. Both algorithms employ gradient boosting strategies that combine multiple weak learners into powerful predictive models capable of handling large datasets with high precision. Their capability to capture complex degradation patterns makes them suitable candidates for battery health estimation applications.

This research develops an intelligent battery health prediction system using XGBoost and LightGBM algorithms to estimate the State of Health of electric vehicle batteries. Battery parameters including voltage, current, operating temperature, charging cycles, and historical usage information are utilized to train predictive models capable of accurately estimating battery condition. Standard preprocessing techniques are applied to improve data quality before feature extraction and model development. The predictive performance of both algorithms is evaluated using statistical metrics including R^2 Score, Adjusted R^2, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

The proposed machine learning framework provides valuable support for battery maintenance planning, predictive diagnostics, and intelligent decision-making. Accurate SOH estimation allows EV owners and fleet operators to optimize charging strategies, schedule preventive maintenance, reduce replacement costs, and extend battery lifespan. Furthermore, reliable battery health prediction contributes to increased user confidence by minimizing unexpected battery failures and reducing range anxiety, thereby promoting wider adoption of electric vehicles.

II. Literature Survey

The increasing popularity of electric vehicles has encouraged extensive research on battery management and battery health prediction. Researchers have explored various artificial intelligence, machine learning, deep learning, and statistical approaches to improve the estimation accuracy of battery State of Health (SOH). These studies demonstrate that data-driven prediction techniques significantly outperform traditional battery estimation methods under practical operating conditions.

Several studies have highlighted the importance of Battery Management Systems (BMS) in ensuring battery safety, operational efficiency, and long-term reliability. Early battery monitoring systems mainly focused on measuring voltage, current, and temperature using mathematical estimation techniques. Although these systems successfully monitored battery operation, their capability to predict long-term battery degradation remained limited because they could not effectively model complex nonlinear aging characteristics.

Researchers have also investigated hybrid electric vehicle technologies to improve energy utilization while reducing environmental pollution. These studies emphasized

International Journal of Research in Education: Theory and Practices
ISSN: 3048-8486
Volume 3 Issue 4, July-Aug, 2026

that integrating intelligent battery management with advanced prediction algorithms could significantly improve battery lifespan, energy efficiency, and overall vehicle performance. Improvements in charging strategies, thermal management, and battery balancing were identified as critical factors influencing battery degradation.

Deep learning methods have emerged as powerful alternatives for battery health prediction due to their capability to automatically learn complex degradation patterns from large datasets. Neural networks have demonstrated high prediction accuracy when trained using voltage, current, temperature, and charging history information. However, deep learning models generally require large computational resources and extensive training data, making their deployment challenging in certain real-time applications.

Machine learning algorithms including Random Forest, Decision Tree, Support Vector Machine, K-Nearest Neighbors, and Gradient Boosting techniques have also been widely applied for SOH prediction. Comparative studies indicate that ensemble learning algorithms consistently achieve higher prediction accuracy because they combine multiple decision trees to capture nonlinear feature relationships while minimizing overfitting.

Recent research has focused on gradient boosting algorithms such as XGBoost and LightGBM due to their superior computational efficiency and predictive capability. These algorithms effectively process high-dimensional datasets while maintaining excellent generalization performance. Their ability to learn complex degradation trends has resulted in highly accurate battery health prediction with minimal prediction error across multiple benchmark datasets.

Although considerable progress has been achieved in battery health estimation, many existing studies still face limitations related to model generalization, computational complexity, and adaptability to diverse operating environments. Therefore, developing efficient machine learning models capable of accurately predicting battery health under real-world operating conditions remains an important research objective. The present work addresses this requirement by evaluating XGBoost and LightGBM models using battery operational parameters for accurate SOH estimation.

III. Research Methodology

The proposed methodology follows a systematic machine learning workflow designed to accurately estimate the State of Health of electric vehicle batteries. The overall framework consists of multiple stages including data acquisition, preprocessing, feature engineering, model development, model evaluation, and real-time prediction. Each stage contributes to improving prediction accuracy while ensuring the robustness of the developed machine learning models.

Initially, battery operational data are collected from the Battery Management System (BMS). The dataset includes important battery parameters such as voltage, current, operating temperature, charging and discharging cycles, and historical battery usage information. These variables provide sufficient information for understanding battery degradation behavior throughout different operating conditions.

After data collection, preprocessing techniques are applied to improve data quality. Missing values are identified and handled appropriately, duplicate records are removed, and abnormal observations are treated using suitable statistical methods. Data normalization is then performed to transform different feature scales into a consistent numerical range, allowing the learning algorithms to converge efficiently during training.

The next phase involves feature engineering, where meaningful battery characteristics are extracted from the raw operational data. Additional statistical and derived features representing battery utilization patterns, charging behavior, temperature variations, and cycle-based degradation are generated to improve the predictive capability of the machine learning models. Proper feature selection also helps eliminate redundant variables and reduces computational complexity.

Following feature preparation, the processed dataset is divided into training and testing subsets using an appropriate train-test split strategy. The training dataset is used to construct two ensemble learning models: XGBoost and LightGBM. Both algorithms learn complex nonlinear relationships between battery operating parameters and the corresponding State of Health values through iterative gradient boosting optimization.

Model performance is evaluated using multiple statistical performance indicators including R^2 Score, Adjusted R^2, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics measure prediction accuracy, model consistency, and error distribution, enabling an objective comparison between both machine learning algorithms.

Finally, the trained prediction models are integrated into a real-time battery health prediction framework. Users can provide current battery operating parameters, and the system instantly estimates the battery State of Health. This intelligent prediction mechanism assists vehicle owners, service providers, and fleet operators in planning preventive maintenance, optimizing battery utilization, and extending overall battery lifespan through data-driven decision making.

IV. Existing System

Existing electric vehicle battery monitoring systems primarily depend on conventional Battery Management Systems that estimate battery condition using mathematical models, equivalent circuit models, Coulomb counting techniques, and predefined lookup tables. These approaches continuously monitor voltage, current, and temperature to ensure safe battery operation. However, they possess limited capability for accurately predicting long-term battery degradation because they cannot effectively capture the nonlinear aging characteristics associated with lithium-ion batteries. As battery operating conditions become increasingly complex, conventional estimation methods often accumulate prediction errors and fail to adapt to changing degradation patterns. Consequently, maintenance decisions based on these methods may be inaccurate, reducing battery efficiency and lifespan. These limitations highlight the necessity for intelligent machine learning approaches capable of providing more accurate, adaptive, and real-time battery health prediction.

V. Proposed System

The proposed system introduces a machine learning-based framework for accurately estimating the State of Health (SOH) of electric vehicle batteries using Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) algorithms. Unlike conventional battery monitoring techniques that depend on fixed mathematical models and predefined estimation methods, the proposed approach learns battery degradation patterns directly from historical operating data. This enables the system to provide more accurate, adaptive, and reliable predictions under varying environmental and operational conditions while maintaining the same methodology and algorithms as the base paper.

The framework begins by collecting battery operational parameters from the Battery Management System (BMS). Important variables such as battery voltage, charging current, operating temperature, charge-discharge cycles, and historical usage information are gathered to build a comprehensive dataset representing the battery's health throughout its lifecycle. Since battery degradation is influenced by multiple interacting factors, incorporating these parameters allows the predictive models to capture complex relationships that cannot be represented effectively by traditional analytical techniques.

Before model development, the collected dataset undergoes several preprocessing operations to improve data quality. Missing values are handled appropriately, duplicate records are eliminated, abnormal observations are removed, and feature normalization is performed to ensure consistent numerical representation across all input variables. These preprocessing steps reduce data inconsistencies and improve the learning capability of the machine learning algorithms.

Following preprocessing, feature engineering is carried out to generate meaningful characteristics that better describe battery degradation behavior. Relevant battery attributes are extracted and refined so that the predictive models can identify hidden trends associated with capacity loss and battery aging. Proper feature selection also minimizes redundant information and enhances computational efficiency without affecting prediction performance.

The processed dataset is then divided into training and testing subsets for model development. Two ensemble learning algorithms, namely XGBoost and LightGBM, are trained using the prepared battery data. Both algorithms are well suited for regression-based prediction problems because they efficiently handle nonlinear relationships, high-dimensional feature spaces, and large datasets while minimizing overfitting. During training, each model learns the relationship between battery operating conditions and the corresponding State of Health values by iteratively optimizing prediction accuracy.

After training, both models are evaluated using statistical performance measures including R^2 Score, Adjusted R^2, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These evaluation metrics provide a comprehensive comparison of prediction accuracy and model reliability. Based on experimental analysis, the trained models are capable of producing highly accurate SOH estimations suitable for practical battery health monitoring applications.

Finally, the optimized prediction model is integrated into a real-time battery health prediction system. Users can provide current battery parameters through a simple interface, and the system instantly predicts the remaining battery State of Health. The generated prediction assists vehicle owners, maintenance personnel, and fleet operators in scheduling preventive maintenance, optimizing charging strategies, reducing maintenance costs, and extending battery service life. By combining advanced machine learning techniques with real-time battery monitoring, the proposed system offers an intelligent, scalable, and efficient solution for enhancing the performance, reliability, and longevity of electric vehicle batteries.

VI. Discussion

The experimental evaluation demonstrates that both XGBoost and LightGBM are highly effective machine learning algorithms for predicting the State of Health (SOH) of electric vehicle batteries. The models were trained using battery operating parameters such as voltage, current, temperature, and charging cycle information, enabling them to accurately capture the complex degradation behavior of lithium-ion batteries. The obtained results indicate that both algorithms successfully learned the relationship between battery characteristics and SOH, producing highly reliable predictions with minimal estimation error while maintaining excellent generalization capability.

The performance comparison reveals that XGBoost achieved slightly better prediction accuracy than LightGBM across all evaluation metrics. XGBoost obtained an R^2 score of 0.9985, Adjusted R^2 of 0.9983, MAE of 0.00155, MSE of 1.04×10^{-5}, and RMSE of 0.00322. In comparison, LightGBM produced an R^2 score of 0.9982, Adjusted R^2 of 0.9980, MAE of 0.00180, MSE of 1.20×10^{-5}, and RMSE of 0.00347. Although the numerical differences are relatively small, the lower error values and higher coefficient of determination demonstrate that XGBoost provides slightly more precise SOH estimation.

The high R^2 values achieved by both algorithms indicate that the developed models explain nearly all variations present in the battery health data. Similarly, the extremely low MAE, MSE, and RMSE values confirm that prediction errors remain very small, making the proposed approach highly suitable for practical battery monitoring applications. These results also suggest that ensemble learning methods effectively capture the nonlinear relationships between battery operating parameters and battery degradation, overcoming many limitations associated with traditional Battery Management System estimation techniques.

Another important observation is that both models exhibit excellent robustness and stability during prediction. LightGBM offers faster model training and lower computational requirements, making it suitable for large-scale battery datasets and real-time deployment where computational efficiency is important. On the other hand, XGBoost provides marginally higher prediction accuracy, making it a preferred choice for applications where precise battery health estimation is critical for maintenance planning and safety.

The proposed machine learning framework offers several practical advantages for electric vehicle manufacturers, service providers, and vehicle owners. Accurate SOH prediction enables predictive maintenance, minimizes unexpected battery failures, optimizes charging schedules, reduces replacement costs, and extends battery lifespan. Furthermore, reliable battery health estimation contributes to improving vehicle reliability, reducing operational expenses, and enhancing user confidence by minimizing range anxiety.

Overall, the experimental findings confirm that both XGBoost and LightGBM are highly capable algorithms for battery State of Health prediction. While XGBoost demonstrates slightly superior predictive performance, both models provide highly accurate and dependable results that can be effectively integrated into intelligent Battery Management Systems for real-time electric vehicle battery health monitoring.

VII. Future Enhancement

The proposed machine learning framework has demonstrated excellent capability in predicting the State of Health (SOH) of electric vehicle batteries with high accuracy using XGBoost and LightGBM algorithms. Nevertheless, there are several opportunities to further enhance the system's performance, adaptability, and practical implementation. Future research can incorporate a wider range of battery operating parameters, including internal resistance, State of Charge (SOC), battery impedance, charging speed, discharge rate, environmental humidity, and driving behavior, to provide a more comprehensive representation of battery degradation. The inclusion of these additional features is expected to improve prediction accuracy under varying real-world operating conditions. Furthermore, advanced deep learning techniques such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Transformer networks, and hybrid machine learning models can be investigated to capture temporal degradation patterns more effectively and provide more accurate long-term battery health forecasting. Integrating the proposed system with Internet of Things (IoT) devices and cloud-based platforms would enable continuous real-time battery monitoring, remote diagnostics, and predictive maintenance for individual vehicles as well as large electric vehicle fleets. Future versions of the framework may also employ adaptive or online learning algorithms that continuously update the prediction model as new battery data become available, allowing the system to maintain high prediction accuracy throughout the battery lifecycle without frequent retraining. In addition, Explainable Artificial Intelligence (XAI) techniques can be incorporated to improve the transparency of prediction results by identifying the most influential battery parameters affecting degradation, thereby assisting engineers and maintenance personnel in making informed decisions. The proposed framework can also be extended to simultaneously estimate both the State of Health (SOH) and the Remaining Useful Life (RUL) of batteries, providing more comprehensive support for maintenance scheduling, warranty management, and battery replacement planning. Finally, validating the proposed approach using larger and more diverse datasets collected from different battery chemistries, electric vehicle manufacturers, climatic regions, and driving environments will further improve its robustness, scalability, and generalization capability. These future enhancements will contribute to the development of more intelligent, reliable, and efficient battery management systems that support longer battery lifespan, reduced maintenance costs, improved vehicle reliability, and the widespread adoption of sustainable electric mobility.

VIII. Conclusion

This research presented an intelligent machine learning framework for predicting the State of Health (SOH) of electric vehicle batteries using XGBoost and LightGBM algorithms. The proposed approach effectively utilizes battery operating parameters, including voltage, current, temperature, and charging history, to model battery degradation and estimate its health with a high level of accuracy. A systematic workflow consisting of data collection, preprocessing, feature engineering, model training, and performance evaluation was implemented to develop reliable prediction models capable of supporting real-time battery health assessment. Experimental evaluation demonstrated that both machine learning algorithms produced outstanding prediction performance with very high R^2 values and extremely low MAE, MSE, and RMSE scores, confirming their capability to accurately capture the complex nonlinear behavior of lithium-ion battery degradation. Among the two models, XGBoost achieved slightly better predictive accuracy than LightGBM, although both algorithms delivered highly dependable and consistent results suitable for practical deployment. The developed prediction framework can assist electric vehicle owners, manufacturers, and fleet management organizations in continuously monitoring battery condition, scheduling preventive maintenance, optimizing charging strategies, reducing unexpected battery failures, and extending overall battery lifespan. In addition, accurate SOH prediction contributes to improved vehicle reliability, lower maintenance costs, enhanced operational efficiency, and greater user confidence in electric mobility. Overall, this study demonstrates that advanced machine learning techniques provide an effective and scalable solution for intelligent battery health monitoring, supporting the development of next-generation Battery Management Systems and promoting the widespread adoption of sustainable electric transportation.

References

1. R. Nagalakshmi and K. Rahimunnisa, "Energy Management and Performance Analysis of Electric Vehicle Battery Systems," Proceedings of the IEEE International Conference on Smart Computing Systems, pp. 221–226, 2023. DOI: 10.1109/ICSCSS57650.2023.10169296.
2. M. Parameswari and S. Usha, "Advanced Battery Management System Design for Electric Vehicle Applications," EAI Endorsed Transactions on Energy Web, vol. 11, no. 2, pp. 1–13, 2024. DOI: 10.4108/ew.5003.
3. G. N. Sravanthi and K. Srinivasa Rao, "Modeling and Analysis of Battery Management Systems for Electric Vehicles," International Journal for Research in Applied Science and Engineering Technology, vol. 11, no. 12, pp. 3450–3458, 2023. DOI: 10.22214/ijraset.2023.55342.
4. A. K. M. Ahasan Habib, M. K. Hasan, G. F. Issa, D. Singh, S. Islam, and T. M. Ghazal, "Challenges and Future Directions of Lithium-Ion Battery Management Systems for Electric Vehicles," Batteries, vol. 9, no. 3, Art. no. 152, 2023. DOI: 10.3390/batteries9030152.
5. P. Madheshiya, A. K. Maurya, and A. K. Rai, "Hybrid Energy Storage Technologies for Next-Generation Electric Vehicles," AIP Conference Proceedings, vol. 3105, 2024. DOI: 10.1063/5.0221437.
6. J. Hemdani, L. Degaa, N. Rizoug, and A. Chaari, "Machine Learning Techniques for State-of-Health Prediction of Lithium-Ion Batteries," Proceedings of the IEEE International Conference on Digital Technologies (CoDIT), pp. 635–640, 2023. DOI: 10.1109/CoDIT58514.2023.10284459.
7. K. Akbar, Y. Zou, Q. Awais, M. J. A. Baig, and M. Jamil, "Robust Machine Learning Framework for Electric Vehicle Battery Health Prediction," Electronics, vol. 11, no. 8, Art. no. 1216, 2022. DOI: 10.3390/electronics11081216.
8. K. Das, R. Kumar, and A. Krishna, "Supervised Machine Learning Approaches for Electric Vehicle Battery Health Assessment," Renewable and Sustainable Energy Reviews, vol. 189, Art. no. 113967, 2024. DOI: 10.1016/j.rser.2023.113967.
9. R. M. Alhazmi, "Deep Learning-Based State-of-Health Estimation for Lithium-Ion Batteries in Electric Vehicles," World Electric Vehicle Journal, vol. 15, no. 9, Art. no. 385, 2024. DOI: 10.3390/wevj15090385.
10. K. Das and R. Kumar, "Machine Learning Approaches for Battery Capacity Degradation and Health Prediction: A Review," Computers in Electrical Engineering, vol. 109, Art. no. 108749, 2023.
11. P. Venugopal, S. S. Shankar, C. P. Jebakumar, and R. Agarwal, "Optimal Machine Learning Models for Lithium-Ion Battery Life Estimation," IEEE Access, vol. 9, pp. 49749–49761, 2021. DOI: 10.1109/ACCESS.2021.3130994.

12. S. V. Saranya, H. K. Hariharan, K. K. Kishore, and G. Maikandan, “Random Forest-Based Battery Health Prediction for Electric Vehicles,” International Journal of Creative Research Thoughts, vol. 12, no. 4, pp. 1084–1093, 2024.