PREDICTING STATE OF HEALTH (SOH) OF EV BATTERIES USING MACHINE LEARNING

Mr. V. Murugan[1], G. Anjali[2], Ch. Kruthika[3], A. Pallavi Reddy[4], Ch. Sudheer[5]

[12345]*Department of Information Technology, TKR College of Engineering and Technology, Telangana, India*

Abstract - *Electric Vehicles (EVs) are becoming a key solution in the transition toward sustainable transportation. One of the most critical components of an EV is its battery, and its longevity significantly affects vehicle performance and user trust. This study presents a machine learning-based system for estimating the State of Health (SOH) of EV batteries using predictive modeling techniques. The system incorporates a user registration and authentication module, secure password reset using OTP-based verification, and a robust ML pipeline that processes an extended EV battery dataset to predict SOH. Key models trained include XGBoost Regressor, LightGBM Regressor, and Random Forest Regressor with evaluation metrics such as MAE, MSE, R^2, and RMSE. The system also generates interpretable visualizations like correlation heatmap, evaluation metric bar graphs, actual vs. predicted SOH plots, and feature importance graphs. Based on the predicted SOH, the system provides contextual feedback such as battery health status, estimated time to replacement, and health maintenance recommendations. The implementation, developed using Django and Python, offers a user-friendly web interface for battery health inference, making it applicable for battery management systems in modern electric vehicles.*

Key Words: Electric Vehicles (EVs), Battery State of Health (SOH), Machine Learning, Battery Management System (BMS), XGBoost Regressor, LightGBM Regressor, Random Forest Regressor, Feature Importance Analysis, Django Framework.

1. INTRODUCTION

Electric Vehicles (EVs) are rapidly becoming one of the most important solutions for reducing air pollution, fossil fuel dependency, and carbon emissions. The shift from conventional internal combustion engine vehicles to EVs has increased the demand for efficient and reliable battery systems. In an electric vehicle, the battery is the core component that determines the driving range, charging performance, safety, and overall user satisfaction. However, lithium-ion batteries degrade over time due to continuous charging and discharging cycles, temperature variations, and usage patterns. This degradation affects the battery's capacity, efficiency, and reliability, making battery health monitoring a critical requirement in modern EV systems.

State of Health (SOH) is one of the key parameters used to measure the condition of a battery compared to its original state. Accurate SOH estimation helps in detecting early degradation, planning preventive maintenance, and reducing the risk of sudden battery failure. Traditional methods of battery testing rely on physical inspections and electrochemical analysis, and multi-physics based modelling approaches, which are often time-consuming, costly, computationally intensive, and not feasible for real-time applications. Therefore, machine learning-based methods are gaining attention due to their ability to learn complex battery behaviour from data and generate accurate predictions.

This project proposes a machine learning-based EV battery SOH prediction system integrated into a Django web application. The system trains regression models such as XG Boost, Light GBM, and Random Forest using an extended EV battery dataset and predicts SOH based on user-input battery parameters. The system also provides battery health status, estimated replacement time, and recommendations along with visualization graphs to support better decision-making.

1.1 Importance of EV Battery SOH Prediction

Battery health monitoring is essential for improving EV reliability and ensuring safe operation. Accurate SOH prediction supports early detection of battery degradation and helps users and manufacturers take timely actions. It also reduces maintenance costs, improves battery lifespan, and enhances driving performance. Machine learning models can provide faster and more scalable solutions compared to manual testing, making them suitable for real-world battery management systems.

1.2 Motivation and Problem Overview

Electric Vehicles (EVs) are rapidly gaining popularity as a sustainable alternative to conventional fuel-based transportation. However, the performance, reliability, and user trust in EVs are highly dependent on the battery, which is the most expensive and critical component of the vehicle. Over time, lithium-ion batteries degrade due to repeated charging and discharging cycles, temperature variations, and operating conditions. This degradation reduces driving range, increases charging time, and may lead to unexpected failures, resulting in high maintenance and replacement costs. Traditional battery health evaluation methods rely on laboratory testing and electrochemical models, which are costly, time-consuming, and not suitable for real-time monitoring.

To address these limitations, this project focuses on developing a machine learning-based system to estimate the State of Health (SOH) of EV batteries using operational parameters such as voltage, current, temperature; charge cycles, discharge cycles, average charge rate, average discharge rate and time elapsed. The major challenge is to accurately predict SOH while ensuring the system is scalable, efficient, and interpretable for practical applications. The proposed solution integrates regression models such as XG Boost , Light GBM, and Random Forest with a Django-based web interface, enabling SOH prediction, battery health status, estimated time to replacement, and maintenance recommendations. This approach supports preventive maintenance, improves battery lifecycle management, reduces operational costs, and enhances the reliability of EV battery management systems.

2. PROPOSED SYSTEM

The proposed system focuses on evaluating the performance of electric vehicle (EV) batteries by analyzing the State of Health (SOH) using advanced machine learning techniques. As the demand for electric vehicles continues to rise, monitoring battery health becomes crucial for ensuring efficiency, longevity, and safety. This system is designed to provide an intelligent solution for estimating the SOH of EV batteries based on various input parameters such as battery voltage, current, temperature, and other operating conditions. By utilizing powerful regression algorithms like XG Boost, Light GBM, and Random Forest the system aims to accurately assess battery performance and provide valuable insights into battery degradation over time. The dataset for this project is prepared and used for model training, followed by evaluation and visualization.

The system not only calculates key evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) values, but also visualizes the results using various graphs like correlation heatmap, bar graphs of metrics, actual vs predicted SOH plots, and feature importance graphs. These visual aids enhance the interpretability of the model's performance and allow stakeholders to make informed decisions regarding battery maintenance and replacement. The entire process is integrated into a Django-based web application where users can interactively view results. This approach reduces reliance on traditional, time-consuming physical testing and offers a scalable, data-driven solution. Overall, the proposed system is a significant step toward smart battery management in electric vehicles, promoting sustainability and improving vehicle reliability.

2.1 System Architecture

The proposed system architecture is designed as a modular and secure web-based framework that integrates user management, data processing, and machine learning-based battery health prediction into a unified platform. The architecture begins with a user registration and authentication layer, incorporating OTP-based password recovery to ensure secure access. Once authenticated, users interact with a Django-based web interface that handles input requests and communicates with the backend server. The backend consists of a data preprocessing module that cleans, normalizes, and prepares the EV battery dataset for analysis, followed by a machine learning pipeline where trained regression models such as XGBoost, LightGBM, and Random Forest are employed to estimate the battery's State of Health (SOH). The prediction results are passed to a visualization and interpretation layer, which generates correlation heatmap, performance metric graphs, actual-versus-predicted SOH comparisons, and feature importance plots. Finally, a decision-support module interprets the predicted SOH to provide meaningful feedback, including battery health status, estimated replacement timelines, and maintenance recommendations, making the architecture suitable for battery management applications in electric vehicles.

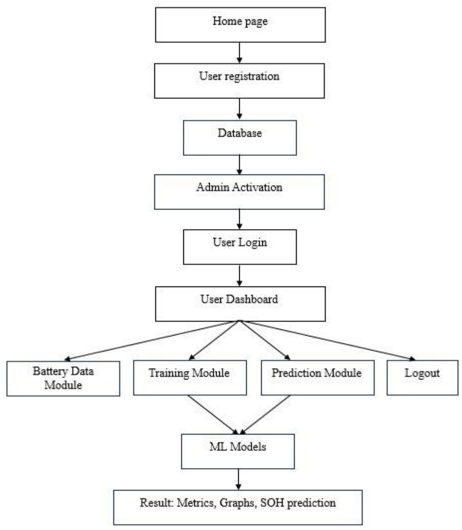

Fig -1: System Architecture

2.2 Secure Web-Based User Management and Data Handling

The proposed system incorporates a secure and user-centric web-based architecture developed using the Django framework. This module manages user registration, login, and session handling, ensuring controlled access to the battery health prediction platform. To enhance security, an OTP-based password recovery mechanism is implemented, which verifies user identity before allowing credential reset.

Once authenticated, users can submit battery-related inputs or datasets through a responsive web interface. The backend server efficiently handles data storage and retrieval while ensuring data integrity and confidentiality. This layer acts as the gateway between users and the intelligent prediction engine, enabling seamless and secure interaction with the system.

2.3 Machine Learning Prediction and Decision Support Layer

At the core of the proposed system lies the machine learning prediction layer responsible for estimating the State of Health (SOH) of EV batteries. Preprocessed battery data is fed into trained regression models, including XGBoost Regressor, LightGBM Regressor, and Random Forest Regressor, which are optimized to capture nonlinear degradation patterns. The system evaluates model performance using metrics such as MAE, MSE, RMSE, and R^2 score to ensure reliable predictions. The predicted SOH values are further analyzed to generate interpretable outputs, including correlation heatmap, bar graphs of evaluation metrics, feature importance graphs and actual versus predicted SOH plots. Based on these results, the decision-support module provides actionable insights such as battery health status, estimated replacement timelines, and health maintenance recommendations, making the system effective for EV battery management.

3. IMPLEMENTATION DETAILS

3.1 Web Application and Security Implementation

The system is implemented using the Django web framework, which follows the Model–View–Template (MVT) architecture to ensure modularity and scalability. User-related functionalities such as registration, login, logout, and session management and handled using Django's built-in authentication mechanisms. To enhance security, an OTP based password recovery module is implemented, where a one-time password is generated and sent to the registered email address for identity verification. User credentials are securely stored using hashing techniques, and role-based access control ensures that only authenticated users can access the battery health prediction features. The frontend is developed using HTML, CSS, and Bootstrap to provide a responsive and user-friendly interface.

3.2 Data Preprocessing and Machine Learning Model Implementation

The backend processing layer is developed using Python, where the EV battery dataset undergoes extensive preprocessing before model training. This includes handling missing values, removing outliers, feature scaling, and correlation analysis to improve model performance. The cleaned dataset is split into training and testing sets. Three regression-based machine learning models XGBoost Regressor, LightGBM Regressor, and Random Forest Regressor are trained to predict the battery State of Health (SOH). Hyperparameter tuning is performed to optimize model accuracy and reduce overfitting. Model performance is evaluated using standard metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score to select the most reliable prediction model.

3.3 Visualization, Prediction Output, and Decision Support

Once the SOH prediction is generated, the system presents results through an interpretability and visualization module.

Graphical outputs such as correlation heatmap, evaluation metric bar charts, feature importance graphs and actual versus predicted SOH plots are generated using Python visualization libraries. These visual insights help users understand the influence of different battery parameters on SOH prediction. Based on the predicted SOH value, the system categorizes battery health into different levels and provides contextual feedback, including battery health status, estimated time to battery replacement and maintenance recommendations. This decision-support functionality enhances the practical usability of the system, making it suitable for EV battery management and monitoring applications.

4. RESULTS AND PERFORMANCE ANALYSIS

The proposed machine learning-based EV battery SOH estimation system was evaluated using multiple regression models, including XGBoost Regressor, LightGBM Regressor, and Random Forest Regressor. The performance of each model was assessed using standard evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score. Experimental results indicate that all three models effectively capture the nonlinear degradation patterns of EV batteries; however, ensemble-based models demonstrated superior prediction accuracy compared to baseline approaches. Among the evaluated models, Random Forest achieved lower error values and higher R^2 scores, indicating strong generalization capability.

Model	MAE	MSE	R^2 Score	RMSE
XGBoost	2.076	13.485	0.910	3.672
LightGBM	2.053	12.801	0.915	3.577
Random Forest	1.987	11.225	0.925	3.350

The actual versus predicted SOH plots shows a close alignment between predicted values and ground truth data, validating the robustness of the trained models. Correlation heatmap analysis helped identify influential battery parameters, while feature importance graphs provided interpretability by highlighting key factors contributing to battery degradation. These visualizations enhance transparency and build user trust in the prediction results. The evaluation metric comparison further confirms the consistency and reliability of the proposed ML pipeline across different performance measures.

In addition to numerical accuracy, the system demonstrates strong practical performance through its decision-support functionality. Based on the predicted SOH, the system successfully categorizes battery health levels and provides meaningful insights such as battery health status, estimated replacement time to replacement and maintenance recommendations. The integration of accurate prediction models with an interactive Django-based web interface ensures inference and ease of use, making the proposed system suitable for deployment in modern EV battery management systems.

5. CONCLUSION

This project presents a machine learning-based system for predicting the State of Health (SOH) of electric vehicle (EV) batteries, aimed at enhancing battery performance and lifespan. By leveraging advanced algorithms such as XGBoost, LightGBM, and Random Forest the system provides accurate SOH predictions, evaluated using metrics like MAE, MSE, RMSE, and R^2.

The Django web application enables users to input battery parameters, receive predictions, and visualize results through correlation heatmap, evaluation metric bar graphs, feature importance graphs, and actual vs predicted SOH plots, facilitating quick and informed decision-making while reducing reliance on time-consuming physical tests.

By supporting preventive maintenance, early detection of battery degradation, and estimated replacement timelines, the system contributes to improved safety, and optimized energy usage. Additionally, it promotes environmental sustainability by extending battery lifespan and minimizing waste. Overall, the project demonstrates a scalable, efficient, and user-friendly approach to smart battery management in EVs.

6. FUTURE WORK

Future work can focus on enhancing the proposed EV battery SOH estimation system by incorporating larger and more diverse real-world datasets to improve model robustness and generalization across different battery chemistries and operating conditions. The system can be extended to support real-time data acquisition from onboard sensors and Internet of Things (IoT) platforms, enabling continuous battery health monitoring. Advanced deep learning techniques such as Long Short-Term Memory (LSTM) networks and transformer-based models may be explored to capture temporal degradation patterns more effectively. Additionally, integrating Remaining Useful Life (RUL) prediction alongside SOH estimation would provide more comprehensive battery lifecycle insights. Future enhancements may also include cloud-based deployment for scalability, edge-computing integration for low-latency inference, and tighter integration with vehicle Battery Management Systems (BMS) to support predictive maintenance and intelligent energy management in next generation electric vehicles.

REFERENCES

[1] J. Hemdani, L. Degaa, N. Rizoug, and A. Chaari, “State of health prediction of lithium-ion battery using machine learning algorithms,” in Proc. IEEE Int. Conf. on Digital Technologies (CoDIT), 2023, doi: 10.1109/CoDIT58514.2023.10284459.

[2] J. Hemdani, M. Soltani, A. J. Telmoudi, and A. Chaari, “Prediction of aging of battery for electric vehicles based on a modified version of neural networks,” in Proc. 29th Mediterranean Conf. on Control and Automation (MED), 2021.

[3] F. Heinrich and M. Pruckner, “Virtual experiments for battery state of health estimation based on neural networks and in-vehicle data,” Journal of Energy Storage, Apr. 2022.

[4] J. A. Sanguesa, V. Torres-Sanz, P. Garrido, F. J. Martinez, and J. M. Marquez-Barja, “A review on electric vehicles: Technologies and challenges,” Smart Cities, 2021.

[5] Y. Zhi, H. Wang, and L. Wang, “A state of health estimation method for electric vehicle Li-ion batteries using GA-PSO-SVR,” Springer, 2022.

[6] M. Al-Saadi, J. Olmos, A. Saez-de-Ibarra, J. Van Mierlo, and M. Berecibar, “Fast charging impact on the lithium-ion batteries’ lifetime and cost effective battery sizing in heavy-duty electric vehicles applications,” Energies, 2022.

[7] Y. Li, K. Li, X. Liu, Y. Wang, and L. Zhang, "Lithium-ion battery capacity estimation—A pruned convolutional neural network approach assisted with transfer learning," Applied Energy, 2021.

[8] R. Zahedi, M. Ghodusinejad, A. Aslani, and C. Hachem-Vermette, "Modelling community-scale renewable energy and electric vehicle management for cold-climate regions using machine learning," Energy Strategy Reviews, 2022.

[9] R. Xiong, H. He, H. Guo, and Y. Ding, "Modeling for lithium-ion battery used in electric vehicles," Procedia Engineering, 2011.

[10] L. Yao, S. Xu, A. Tang, F. Zhou, J. Hou, Y. Xiao, and Z. Fu, "A review of lithium-ion battery state of health estimation and prediction methods," World Electric Vehicle Journal, 2021.