ISSN2277-2685

IJESR/April. 2026/ Vol-16/Issue-2s1/138-143

Mohd Isam Uddin *et.al.,*/International Journal of Engineering &Science Research

Full Length Article

Machine Learning-Driven Real-Time Battery Health Estimation for EV Battery Swapping

Mohd Isam Uddin[1], Shaik Saleem Pasha[2], Shaik Ibrahim[3], Syed Zaffar Mahmood[4]
[1,2,3]B.E.Students; Department of Information Technology, ISL Engineering College, Osmania University Hyderabad, India.
[4]Assistant professor; Department of Information Technology, ISL Engineering College, Osmania University Hyderabad, India.
Email: isam.12uddin@gmail.com, shaiksaleem8187@gmail.com , shaikibrahim52742@gmail.com
Accepted 24-04-2026

Abstract
Electric Vehicles (EVs) are rapidly transforming the global transportation ecosystem due to their environmental advantages, reduced greenhouse gas emissions, and decreased dependence on fossil fuels. As EV adoption increases, efficient battery management has become a critical challenge, particularly in battery swapping systems where batteries are frequently exchanged and subjected to varying operational conditions. In such systems, accurate estimation of battery health is essential to ensure operational safety, reliability, and cost-effectiveness.
This paper presents a machine learning-driven framework for real-time estimation of key battery health parameters, including State of Health (SoH) and Remaining Useful Life (RUL). The proposed system leverages ensemble learning techniques, specifically Random Forest Regression and XGBoost, to analyze battery operational data such as voltage, current, temperature, and cycle count. These models are capable of capturing complex nonlinear relationships among battery parameters and provide accurate predictive insights. Furthermore, the trained models are integrated into a Flask-based web application, enabling real-time predictions and user-friendly interaction.
Experimental results demonstrate that the proposed system outperforms traditional rule-based and threshold-based approaches in terms of accuracy, robustness, and computational efficiency. The system achieves prediction accuracy exceeding 95% while maintaining low error rates and fast response times. The proposed solution facilitates predictive maintenance, reduces the risk of battery degradation, enhances safety, and improves the overall efficiency of EV battery swapping stations.

Keywords
Electric Vehicles, Battery Health Estimation, Machine Learning, Random Forest, XGBoost, State of Health (SoH), Remaining Useful Life (RUL), Battery Swapping.

Introduction

Electric vehicles (EVs) are gaining significant momentum worldwide as governments and industries aim to reduce carbon emissions and promote sustainable transportation. The lithium-ion battery serves as the core component of EVs, directly influencing vehicle performance, driving range, cost, and reliability. As a result, effective battery management is crucial for ensuring optimal operation and long-term sustainability.
Battery swapping has emerged as a promising alternative to conventional charging methods. This approach allows EV users to replace depleted batteries with fully charged ones within minutes, thereby eliminating long charging times and improving convenience. However, battery swapping introduces new challenges in battery health monitoring, as batteries are used by multiple vehicles and operate under diverse environmental and load conditions. This variability makes it difficult to track battery degradation accurately.

Battery degradation is influenced by several factors, including charge-discharge cycles, temperature fluctuations, depth of discharge, and charging speed. Traditional battery monitoring systems rely on rule-based logic, threshold limits, and manual inspection, which are often inadequate for capturing the complex and nonlinear behavior of battery systems. These methods lack adaptability and fail to provide real-time predictive insights.
Machine learning (ML) techniques offer a powerful alternative by learning patterns from historical data and modeling nonlinear relationships among battery parameters. ML-based systems can provide accurate

and real-time predictions, enabling proactive decision-making. This paper proposes a machine learning-driven system integrated with a web interface to support real-time battery health estimation in EV battery swapping environments.

Literature Review

Extensive research has been conducted in the field of battery health prediction using machine learning and data-driven approaches. Random Forest models have been widely adopted due to their robustness and ability to handle noisy and high-dimensional datasets. By combining multiple decision trees, Random Forest improves prediction accuracy and reduces overfitting. XGBoost, a gradient boosting algorithm, has gained popularity for its superior performance and computational efficiency. It enhances model accuracy by iteratively correcting errors from previous models and optimizing the loss function. Due to its speed and scalability, XGBoost is particularly suitable for real-time applications.

Deep learning techniques, such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN), have also been used for battery health prediction, especially in time-series analysis. These models can capture temporal dependencies in battery degradation patterns. However, they require large datasets, high computational resources, and longer training times, which limit their practical implementation in real-time systems.

Bayesian Neural Networks provide probabilistic predictions and uncertainty estimation, making them suitable for safety-critical applications. Transfer learning approaches further enhance model adaptability across different battery chemistries and operating conditions.

Despite these advancements, several limitations remain. Many existing approaches suffer from high computational complexity, lack real-time deployment capabilities, and exhibit poor generalization across diverse datasets. The proposed system addresses these challenges by utilizing efficient ensemble learning techniques that balance accuracy, speed, and scalability.

Methodology

The proposed methodology is designed as a systematic and data-driven pipeline to ensure accurate, reliable, and real-time estimation of battery health in electric vehicle (EV) battery swapping systems. Each stage of the pipeline is carefully structured to transform raw battery data into meaningful predictive insights, enabling efficient monitoring and decision-making.

Data Collection and Acquisition

The process begins with comprehensive data collection from battery management systems and operational datasets. Key parameters such as voltage, current, temperature, and cycle count are continuously recorded, as they directly reflect the internal condition and performance of lithium-ion batteries. These variables capture both instantaneous behavior (like voltage fluctuations) and long-term degradation patterns (such as cycle aging).

To improve robustness, data can be collected under diverse operating conditions, including varying loads, environmental temperatures, and charging/discharging cycles. This diversity ensures that the model generalizes well across real-world scenarios, especially in battery swapping systems where usage patterns differ significantly between users.

Data Preprocessing and Feature Engineering

Raw data often contains inconsistencies such as missing values, noise, and outliers, which can negatively impact model performance. Therefore, a thorough preprocessing stage is essential. Missing values are handled using imputation techniques such as mean, median, or interpolation methods. Outliers—caused by sensor errors or abnormal operating conditions—are detected using statistical approaches and removed or corrected.

Feature scaling is then applied using normalization or standardization to ensure that all input variables contribute equally to the model. Without scaling, features with larger numerical ranges could dominate the learning process.

In addition to cleaning, feature engineering is performed to enhance predictive power. Derived features such as charge/discharge rates, temperature gradients, or moving averages can provide deeper insights into battery behavior. Feature selection techniques, including correlation analysis and importance ranking, are used to retain only the most relevant variables. This reduces dimensionality, improves computational efficiency, and minimizes overfitting.

Model Development and Training

The core of the methodology lies in training machine learning models capable of capturing complex nonlinear relationships between input variables and battery health indicators. Two ensemble learning algorithms are employed: Random Forest Regression and XGBoost.

Random Forest Regression operates by constructing multiple decision trees using different subsets of data and features. Each tree generates a prediction, and the final output is obtained by averaging these predictions. This ensemble approach reduces variance, enhances stability, and prevents overfitting, making it highly effective for noisy datasets.

XGBoost (Extreme Gradient Boosting) builds models sequentially, where each new model attempts to correct the errors of the previous ones. It uses gradient descent optimization to minimize the loss function and incorporates regularization techniques to avoid overfitting. Due to its efficiency and scalability, XGBoost is particularly suitable for real-time prediction systems.

Hyperparameter tuning is performed for both models using techniques such as grid search or cross-validation. Parameters like the number of trees, tree depth, and learning rate are optimized to achieve the best balance between bias and variance.

Model Evaluation and Validation

To assess the effectiveness of the trained models, several performance metrics are used. These include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2 score). Each metric provides a different perspective on model performance.

MSE=1n∑i=1n(yi−y^i)2MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2MSE=n1∑i=1n(yi−y^i)2

The Mean Squared Error (MSE) measures the average squared difference between actual and predicted values, penalizing larger errors more heavily. RMSE, being the square root of MSE, provides error values in the same unit as the target variable, making it easier to interpret.

R2=1−∑(yi−y^i)2∑(yi−y¯)2R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}R2=1−∑(yi−y¯)2∑(yi−y^i)2

The R^2 score indicates how well the model explains the variance in the data. A value closer to 1 signifies better predictive performance. These evaluation metrics ensure that the models are not only accurate but also consistent and reliable across different datasets.

Mathematical Representation of Battery Health Prediction

Battery health estimation can be formulated as a regression problem, where the output (State of Health) is a function of multiple input variables:

SoH=f(V,I,T,C)SoH = f(V, I, T, C)SoH=f(V,I,T,C)

where VVV represents voltage, III current, TTT temperature, and CCC cycle count. The goal of the model is to learn this function fff from historical data and accurately predict SoH for new inputs.

The optimization objective is to minimize the prediction error between actual and predicted values. This is typically achieved by minimizing the loss function, such as MSE, during training. By iteratively adjusting model parameters, the algorithm converges toward an optimal solution that generalizes well to unseen data.

Workflow Integration and Prediction Pipeline

Once trained and validated, the models are integrated into a real-time prediction pipeline. Incoming data follows the same preprocessing steps as the training data to ensure consistency. The processed input is then passed through the trained model, which generates predictions for battery health metrics such as SoH and Remaining Useful Life (RUL).

This structured workflow ensures that predictions are both accurate and computationally efficient. The pipeline is designed to handle continuous data streams, making it suitable for deployment in real-world EV battery swapping systems.

Reliability and Scalability Considerations

The methodology emphasizes scalability and robustness to support practical deployment. By using ensemble models, the system achieves high accuracy while maintaining computational efficiency. The modular design allows easy integration with IoT devices and cloud platforms for real-time monitoring.

Additionally, the use of efficient algorithms like XGBoost ensures that the system can handle large datasets and high-frequency data streams without significant delays. This makes the approach highly suitable for real-time battery health monitoring applications.

Implementation

The implementation of the proposed battery health estimation system is designed to ensure efficiency, scalability, and real-time usability. The system is primarily developed using Python, chosen for its rich ecosystem of libraries and strong community support in data science and machine learning. For building the interactive web interface, the lightweight framework Flask is utilized, enabling seamless communication between the user and the backend prediction models.

Technology Stack and Tools

The system leverages several powerful libraries to handle different stages of the pipeline. NumPy and

Pandas are used for efficient data manipulation, numerical computations, and preprocessing tasks. For model development, Scikit-learn provides robust implementations of machine learning algorithms, particularly Random Forest Regression, while XGBoost is employed for high-performance gradient boosting. These tools collectively ensure accurate modeling and efficient computation.

Data Preprocessing and Preparation

The implementation begins with loading the dataset, which contains battery parameters such as voltage, current, temperature, and cycle count. Data preprocessing plays a crucial role in ensuring model accuracy. Missing values are handled using appropriate imputation techniques, while outliers are detected and removed to prevent skewed predictions. Feature scaling techniques such as normalization or standardization are applied to maintain consistency across input variables. Additionally, feature selection methods are used to identify the most influential parameters, reducing dimensionality and improving model performance.

Model Training and Optimization

Once the dataset is prepared, the system proceeds to train machine learning models. Random Forest Regression is used to build multiple decision trees and aggregate their outputs, improving robustness and reducing overfitting. Simultaneously, XGBoost is trained using a gradient boosting approach that iteratively minimizes prediction errors. Hyperparameter tuning techniques such as grid search or cross-validation are applied to optimize model performance. This ensures that the models achieve high accuracy while maintaining generalization across different datasets.

Model Serialization and Storage

After training, the models are serialized using Python's pickle module. This process converts the trained models into a format that can be stored and reused without retraining. Serialization significantly reduces computational overhead during deployment, allowing the system to quickly load pre-trained models and generate predictions in real time. This step is essential for making the solution practical and scalable for real-world applications.

Web Application Development

A user-friendly web application is developed using Flask to provide an intuitive interface for interaction. The front end typically includes input forms where users can enter battery parameters such as voltage, current, temperature, and cycle count. The backend is responsible for handling requests, processing input data, and invoking the trained machine learning models. Flask routes are defined to manage user requests and return prediction results dynamically.

Real-Time Prediction Workflow

When a user inputs battery parameters through the web interface, the data is first validated and preprocessed to match the format used during model training. The processed input is then passed to the serialized machine learning model, which generates predictions for battery health metrics such as State of Health (SoH) and Remaining Useful Life (RUL). The results are computed within milliseconds and returned to the user interface, ensuring a smooth and responsive experience.

System Integration and Deployment Considerations

The system is designed to be easily deployable on local servers or cloud platforms. Integration with battery swapping infrastructure can be achieved through APIs, enabling automated data flow from battery management systems. The lightweight nature of Flask ensures low latency and efficient handling of multiple user requests. Furthermore, the modular architecture allows for future expansion, such as integrating IoT-based data collection or advanced analytics modules.

User Interaction and Decision Support

The final output is presented in a clear and interpretable format, allowing users to quickly assess battery health status. This enables operators of battery swapping stations to make informed decisions regarding battery reuse, maintenance, or replacement. By providing real-time insights, the system supports predictive maintenance strategies, reduces operational risks, and enhances overall system efficiency.

Testing

To ensure system reliability and performance, multiple testing strategies are employed. Unit testing is conducted to validate individual components such as data preprocessing and prediction modules. Functional testing verifies that the system produces correct outputs for given inputs.

System testing evaluates the integration of all components to ensure seamless operation of the complete system. Additionally, performance testing is carried out to measure response time, scalability, and the system's ability to handle real-time data processing efficiently. These testing procedures ensure that the system is robust, reliable, and suitable for practical deployment.

Results

The experimental evaluation of the proposed machine learning-based system demonstrates its strong capability in accurately predicting battery health parameters in EV battery swapping environments. By leveraging ensemble learning techniques, the system achieves high predictive performance while maintaining efficiency and robustness under varying operating conditions.

One of the most significant outcomes of the study is the achievement of prediction accuracy exceeding 95%. This marks a substantial improvement over traditional rule-based and threshold-based methods, which typically struggle to capture the nonlinear and dynamic behavior of battery systems. The enhanced accuracy ensures more reliable estimation of key parameters such as State of Health (SoH) and Remaining Useful Life (RUL), which are critical for effective battery management.

In addition to accuracy, the system shows a considerable reduction in prediction error. Lower error rates indicate that the model's predictions closely align with actual battery conditions, thereby increasing trustworthiness and practical usability. The use of advanced algorithms enables the system to minimize deviations and handle noisy or complex datasets efficiently.

Another important performance factor is response time. The proposed system delivers predictions almost instantaneously, making it highly suitable for real-time applications. This fast response capability is essential in battery swapping stations, where quick decision-making is required to determine whether a battery is fit for reuse or needs maintenance or replacement.

Overall Performance Analysis

The overall system performance highlights the effectiveness of the proposed approach across multiple evaluation criteria. High prediction accuracy, reduced error rates, and fast response times collectively demonstrate that the system is well-suited for real-time deployment.

Performance Metric	Observation
Prediction Accuracy	Greater than 95%
Error Rate	Significantly reduced
Response Time	Faster than traditional methods

These results confirm that the integration of machine learning models enhances both efficiency and reliability compared to conventional techniques.

Comparative Model Evaluation

A comparative analysis was conducted to evaluate the performance of different approaches, including traditional methods, Random Forest, and XGBoost. The results clearly show a progressive improvement in accuracy and reduction in error as more advanced models are employed.

Method	Accuracy	Error Rate
Traditional	80%	High
Random Forest	93%	Low
XGBoost	96%	Very Low

Traditional methods rely on fixed rules and thresholds, which limit their adaptability and accuracy. Random Forest improves performance by combining multiple decision trees, reducing overfitting and enhancing generalization. However, XGBoost outperforms both by using a gradient boosting approach that iteratively refines predictions and minimizes errors more effectively.

Interpretation of Results

The superior performance of XGBoost can be attributed to its ability to model complex nonlinear relationships and optimize prediction accuracy through iterative learning. Its built-in regularization mechanisms also help prevent overfitting, ensuring consistent performance across different datasets.

Random Forest, while slightly less accurate than XGBoost, still demonstrates strong performance and reliability. Its ensemble nature makes it robust to noise and suitable for a wide range of applications. The combination of both models provides flexibility, allowing system designers to choose between slightly faster computation (Random Forest) and higher accuracy (XGBoost).

Practical Implications

The results have significant implications for real-world deployment in EV battery swapping systems. High prediction accuracy and low error rates enable precise monitoring of battery health, reducing the risk of unexpected failures. Fast response times ensure that decisions regarding battery usage, replacement, or maintenance can be made instantly.

This leads to improved operational efficiency, enhanced safety, and optimized battery lifecycle management. Ultimately, the system supports predictive maintenance strategies, reducing costs and improving the overall reliability of EV infrastructure.

Conclusion

The proposed system successfully demonstrates the application of machine learning techniques for real-time battery health estimation in EV battery swapping environments. By integrating Random Forest and XGBoost models into a web-based platform, the system enables accurate and efficient monitoring of battery condition.

The implementation offers several advantages, including improved battery lifespan through proactive health monitoring, reduced maintenance costs by minimizing unexpected failures, enhanced safety, and optimized utilization of battery resources. The system provides a scalable and reliable solution to address modern challenges in EV battery management.

Future Scope

Future enhancements can further improve the system's capabilities and applicability. Integration with IoT devices can enable real-time data collection directly from battery systems, ensuring continuous monitoring. Advanced deep learning models such as LSTM and GRU can be incorporated to improve prediction accuracy for time-series data.

Additionally, developing a mobile application can enhance user accessibility and convenience. Cloud-based deployment can improve scalability, storage, and remote access to the system. Incorporating predictive maintenance alerts can enable early detection of potential failures, ensuring timely intervention and increased system reliability. These advancements will further strengthen the role of machine learning in intelligent battery management systems.

References

1) K. Sumanth, K. H. Teja, M. S. Chaithra, and S. R. U. Haq, *"Machine Learning-Driven Real-Time Battery Health Estimation for EV Battery Swapping,"* International Journal of Scientific Research in Engineering and Management, vol. 10, no. 3, pp. 1–9, 2026.
2) A. P. Renold and N. S. Kathayat, *"Comprehensive Review of Machine Learning, Deep Learning, and Digital Twin Data-Driven Approaches in Battery Health Prediction of Electric Vehicles,"* IEEE Access, vol. 12, pp. 43984–43999, 2024.
3) A. Haraz, K. Abualsaud, and A. M. Massoud, *"State-of-Health and State-of-Charge Estimation in Electric Vehicles Batteries: A Survey on Machine Learning Approaches,"* IEEE Access, vol. 12, pp. 158110–158139, 2024.
4) K. Nyachionjeka and E. H. E. Bayoumi, *"A Review of Data-Driven Approaches and Datasets for Battery State of Health Estimation,"* SAE International Journal of Electrified Vehicles, 2026.
5) K. Das and R. Kumar, *"Electric Vehicle Battery Capacity Degradation and Health Estimation Using Machine-Learning Techniques: A Review,"* Clean Energy, vol. 7, no. 6, pp. 1268–1281, 2023.
6) D. Roman, S. Saxena, V. Robu, M. Pecht, and D. Flynn, *"Machine Learning Pipeline for Battery State of Health Estimation,"* arXiv preprint arXiv:2102.00837, 2021.
7) N. Jiang et al., *"Driving Behavior-Guided Battery Health Monitoring for Electric Vehicles Using Machine Learning,"* arXiv preprint arXiv:2309.14125, 2023.
8) R. Wu, Y. Shi, Q. Hu, and J. Yu, *"Digital Twin Model for Lithium-Ion Battery SOC Estimation in Battery Swapping Station,"* IEEE Access, 2024.
9) S. F. Chevtchenko et al., *"A Mapping Study of Machine Learning Methods for Remaining Useful Life Estimation of Lead-Acid Batteries,"* arXiv preprint arXiv:2307.05163, 2023.
10) A. Aitio and D. A. Howey, *"Predicting Battery End of Life from Solar Off-Grid System Field Data Using Machine Learning,"* arXiv preprint arXiv:2107.13856, 2021.