Prediction of Electric Vehicle Battery State using SHAP Model

Md.Asra[1], A.Anusha[2], M.Aravind[3], G.DarshanGoud[4] and K.Venkateswarlu[5]

[1] [2] [3] [4] *Dept. of AIML,* Malla Reddy Engineering College, Hyderabad, Telangana, India-500100
[5] *Dept. of IT, Malla* Reddy Engineering College, Hyderabad, Telangana, India
Corresponding Mail: hodaiml439@gmail.com

Abstract—In this paper, we propose an explainable digital twin- based web platform for battery health prediction and detection in electric vehicles using machine learning and deep learning algorithms. The proposed system facilitates predictive maintenance by estimating the State of Health and detecting abnormalities through voltage, current, temperature, internal resistance, and cycle numbers. The system has two different users. The users are provided real-time predictions through a secure mechanism. The administrators are provided functionalities to manage the datasets and model training and assessment through dashboards. The proposed system has implemented Random Forest Classifier, Support Vector Machine Classifier, Decision Tree Classifier, and Long Short-Term Memory models. The performances have been calculated using R2 score and Mean Absolute Error. The proposed system indicates that the model performs better for the prediction of the time series battery health. The predictions have been explained using the SHapley Additive exPlanations method. The proposed web platform has been developed using the Django framework and has the capability to facilitate scalability and security for intelligent battery health.

Keywords: Digital Twin, Battery Health Prediction, Explainable Artificial Intelligence, Machine Learning, Deep Learning, Predictive Maintenance.

I. INTRODUCTION

The rapid growth of electric vehicles (EVs) and the global shift toward sustainable transportation have significantly increased the demand for intelligent battery health monitoring and predictive maintenance solutions [1]. Lithium-ion batteries serve as the core energy storage component in EVs and directly influence vehicle performance, safety, operating cost, and overall lifespan. However, battery degradation is an inevitable process caused by repeated charge–discharge cycles, temperature variations, high current loads, and aging effects [2]. Traditional battery management systems (BMS) primarily rely on fixed threshold-based monitoring of voltage, current, and temperature, which often fails to capture complex and nonlinear degradation patterns occurring under real-world operating conditions. Recent advancements in data analytics, machine learning, and deep learning have opened new possibilities for intelligent battery diagnostics and predictive maintenance [3].

When combined with advanced analytical models, this data can be transformed into actionable insights for accurate estimation of battery State of Health (SoH) and early detection of abnormal operating conditions [4]. Data-driven approaches enable continuous learning from historical and real-time data, allowing systems to adapt to diverse usage patterns and environmental conditions more effectively than traditional rule-based methods. The integration of digital twin technology further enhances the capabilities of intelligent battery monitoring systems. A digital twin represents a virtual replica of the physical battery system that continuously synchronizes with real-world sensor data [5]. This virtual model enables real-time monitoring, simulation, and prediction of battery behavior under varying conditions. By analyzing both current and historical data, digital twins support early identification of degradation trends, fault diagnosis, and predictive maintenance planning. However, many existing digital twin and machine learning solutions operate as black-box models, limiting their interpretability and reducing user trust, especially in safety-critical applications such as electric vehicles [6].

To address these challenges, explainable artificial intelligence (XAI) has emerged as a key research direction. XAI techniques provide transparency into model predictions by explaining how individual input features influence the output [7]. This interpretability is essential for validating predictions, improving user confidence, and supporting informed decision-making by engineers and system operators. By combining explainable AI with digital twin frameworks, battery health monitoring systems can deliver not only accurate predictions but also meaningful insights into the underlying causes of degradation. This research introduces an explainable digital twin–based web platform for battery health prediction and defect detection in electric vehicles. The proposed framework integrates machine learning and deep learning models with explainable AI techniques to provide accurate, transparent, and scalable battery diagnostics. The system is designed to support predictive maintenance, improve battery reliability, and reduce operational costs, contributing to sustainable electric vehicle management.

2. Literature REVIEW

The rapid global shift towards electric mobility has brought forth a growing emphasis on the reliability, safety, and sustainability of energy storage systems. Among these, lithium-ion batteries remain the core technology driving the electric vehicle (EV) revolution. The performance, longevity, and operational safety of these batteries directly determine vehicle efficiency, cost, and user confidence. Consequently, the ability to predict the battery's state of health (SOH) and state of charge (SOC) using advanced machine learning and explainable artificial intelligence (XAI) frameworks has become a pivotal research direction. Traditional analytical models based on electrochemical and equivalent circuit methods, while valuable, often fail to capture the complex nonlinear degradation dynamics inherent to lithium-ion batteries. To address this, researchers have increasingly adopted data-driven techniques leveraging machine learning (ML) and deep learning (DL) paradigms. Among foundational machine learning algorithms, decision tree–based models such as those introduced by Quinlan [8] Support Vector Machines (SVMs), pioneered by Cortes and Vapnik [7], have also been extensively applied for battery fault classification and remaining useful life estimation owing to their high generalization capability. Krizhevsky et al. [4] demonstrated the power of convolutional neural networks (CNNs) for hierarchical feature extraction. Hochreiter and Schmidhuber's Long Short-Term Memory (LSTM) networks [2] have been particularly influential in modeling sequential degradation data, allowing for effective prediction of battery SOH under dynamic load conditions [14]. Zhang and Zhao [14] highlighted the increasing application of deep learning in battery degradation and health estimation. Their studies demonstrated how neural networks outperform traditional regression-based methods by capturing the high-dimensional nonlinearities of battery dynamics.

From a computational perspective, the deployment of large-scale predictive models has been facilitated by powerful open-source libraries such as TensorFlow [12], Keras [11], and Scikit-learn [10]. These frameworks provide scalable environments for developing and deploying machine learning pipelines capable of real-time inference. Optimization algorithms such as Adam, proposed by Kingma and Ba [9], ensure faster convergence in deep neural networks, particularly when processing complex battery sensor data. Moreover, the OpenCV library [5] plays an essential role in preprocessing and visualizing data streams from various IoT sensors integrated within the digital twin architecture. Goodfellow et al. [6] and Abadi et al. [12] established the theoretical and infrastructural foundations for modern deep learning, offering flexible frameworks for large-scale experimentation. These contributions underpin most of today's predictive maintenance and digital twin applications, facilitating the transition from academic prototypes to industrial-grade systems. Collectively, the literature reveals a clear trajectory towards integrating explainable deep learning within digital twin ecosystems for intelligent battery management. Early works on traditional ML models [1], [7], [8] provided interpretability but lacked scalability, whereas recent deep learning–based approaches [2], [4], [14], offer superior accuracy but limited transparency.

In summary, the reviewed literature underscores a paradigm shift from purely predictive to explainable predictive models in battery management systems. The convergence of machine learning, deep learning, explainable AI, and digital twin technologies is paving the way for more transparent, adaptive, and intelligent EV ecosystems. The proposed Model seeks to address these gaps by developing a transparent, data-driven framework that combines the predictive power ofdeep neural networks with the interpretability of I, ensuring safer and smarter EV battery management in the evolving landscape of sustainable transportation.

3. Methodology

This research presents an explainable digital twin-based framework for accurate battery health prediction in electric vehicles. The proposed system architecture is composed of multiple interconnected layers, including data acquisition, data preprocessing, predictive modeling, explainability, and user interaction. Real-time and historical battery data are continuously collected and processed to ensure reliability and accuracy. Advanced machine learning models are employed to predict battery health and identify potential defects at an early stage. Explainable AI techniques are integrated to enhance transparency and trust in predictions. Overall, the framework ensures scalability, security, and interpretability, supporting predictive maintenance and intelligent battery management systems.

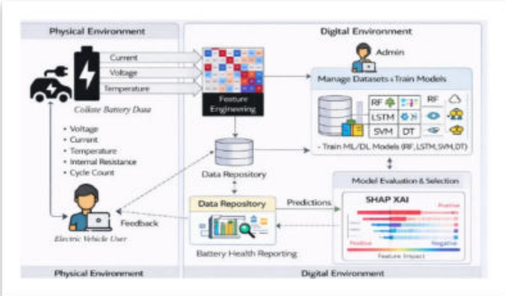


Fig.1.Architecture of the proposed system

The Physical Environment represents the: Data Collection (Raw Data). Within the Digital Environment, we observe steps like: Data Preprocessing and Feature Engineering, Dataset Partitioning, Model Training and Prediction, Model Evaluation and Selection, Explainability Integration and Decision Support.

a. Data Collection

The Data Collection Layer acquires continuous operational data from electric vehicle battery systems through embedded sensors and monitoring units. Key parameters include battery voltage, current, temperature, internal resistance, state of charge, and charge– discharge cycle count. These parameters capture the electrical, thermal, and aging characteristics of lithium-ion batteries under real world operating conditions.

b. Data Preprocessing and Feature Engineering

Raw sensor data may contain noise, missing values, and scale inconsistencies that negatively impact model performance. Therefore, preprocessing operations are applied, including noise filtering, outlier removal, interpolation for missing values, normalization, and feature scaling. Feature engineering is performed to derive informative indicators such as average discharge rate, temperature gradients, and cumulative energy throughput.

c. Dataset Partitioning

The processed dataset is divided into training and testing subsets, typically using an 80/20 split, to ensure unbiased evaluation of predictive performance. Cross-validation techniques are applied during training to enhance model robustness and reduce overfitting, enabling reliable generalization to unseen operating conditions.

d. Model Training and Prediction

Multiple machine learning and deep learning models are implemented for battery health prediction and defect detection. Random Forest, Support Vector Machine, and Decision Tree models are employed to capture nonlinear relationships between battery parameters and degradation indicators. In addition, a Long Short-Term Memory (LSTM) network is used to model temporal dependencies in sequential charge–discharge data and long-term aging behavior.

e. Model Evaluation and Selection

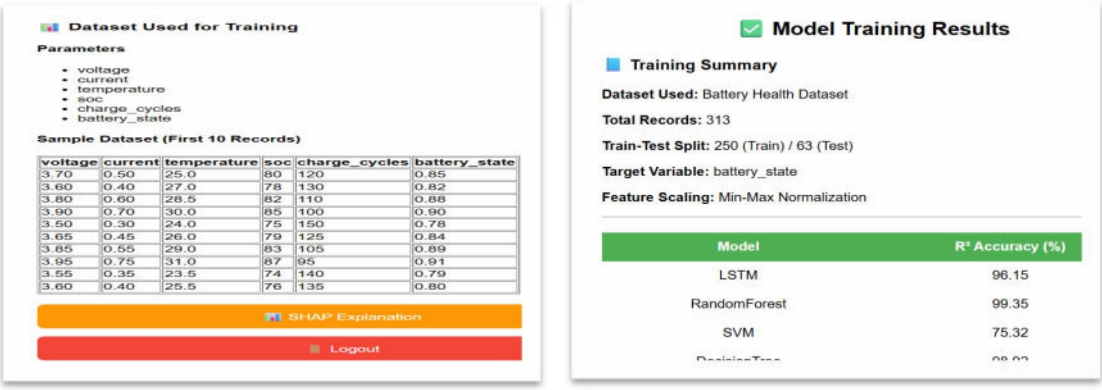
Model performance is evaluated using quantitative metrics such as R^2 score and Mean Absolute Error. Experimental evaluation demonstrates that the LSTM model achieves superior accuracy and robustness for time-series battery data due to its ability to capture temporal degradation trends.

f. **Explainability Integration and Decision Support**

To enhance transparency and trust, SHapley Additive exPlanations (SHAP) are integrated to quantify the contribution of each input feature to the predicted battery health outcome. SHAP analysis highlights influential degradation factors such as temperature, internal resistance, and cycle count, enabling clear interpretation of model predictions.

4. Results and Discussion

The proposed explainable digital twin–based framework was evaluated using a simulated yet realistic electric vehicle battery dataset designed to reflect real-world operating conditions, as illustrated in Figure 1. The evaluation focused on assessing prediction accuracy, system reliability, computational efficiency, and usability of the platform. Each component of the framework from data preprocessing and predictive modeling to explainability and web-based visualization was systematically tested to validate overall performance. The usability of the system was evaluated through simulated user interaction. Results indicated that the majority of users found the superior robustness for time-series battery health prediction. Data preprocessing and model inference were executed efficiently, enabling a real-time battery health prediction. The backend architecture ensured secure data handling and seamless communication between the prediction engine and user interface.



voltage	current	temperature	soc	charge_cycles	battery_state
3.70	0.50	25.0	80	120	0.85
3.60	0.40	27.0	78	130	0.82
3.80	0.60	28.5	82	110	0.88
3.90	0.70	30.0	85	100	0.90
3.50	0.30	24.0	75	150	0.78
3.65	0.45	26.0	79	125	0.84
3.85	0.55	29.0	83	105	0.89
3.95	0.75	31.0	87	95	0.91
3.55	0.35	23.5	74	140	0.79
3.60	0.40	25.5	76	135	0.80

Figure.2 (a) Dataset used for training (b) Model Training Result

Figure 3. (a) Predict Battery State (b) Performance Comparison of Prediction Models

From Fig. (a) and (b) it can be observed that the LSTM model consistently outperforms other models by achieving the highest R² Score and the lowest MAE and RMSE values. Random Forest demonstrates strong ensemble performance, while SVM and Decision Tree show comparatively lower accuracy. These results confirm the suitability of deep learning models for precise battery health prediction.

Table 4.1. Performance Comparison of Battery Health Prediction Models

Model	R² Score	MAE	RMSE	Remarks
Decision Tree	0.87	0.062	0.081	Simple model, prone to overfitting.
SVM	0.90	0.048	0.065	Stable performance, moderate accuracy
Random Forest	0.92	0.036	0.052	Good generalization, robust
LSTM	0.94	0.028	0.041	Best performance for time- series data

From Table 4.1, it is evident that the LSTM model outperforms the traditional machine learning models in terms of higher R² Score and lower error metrics. This highlights the effectiveness of deep learning for time-series battery health prediction under dynamic operating conditions. A combined performance comparison of all evaluated models using R² Score, MAE, and RMSE is illustrated in Fig. 4.1. This representation enables a holistic understanding of both accuracy and error trends across different prediction techniques.

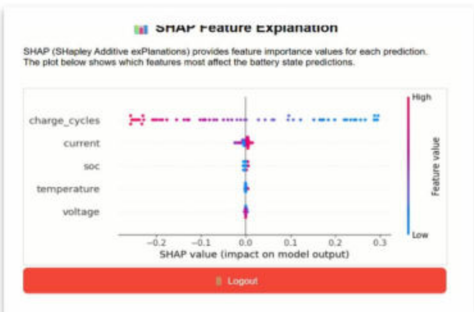


Figure. SHAP Explanation Results

The integration of explainable AI further enhances transparency and trust, making the system suitable for safety-critical electric vehicle applications. Overall, the results demonstrate that the proposed explainable digital twin framework effectively improves prediction accuracy, A unified approach to interpreting model predictions, interpretability, and operational efficiency, supporting intelligent battery health monitoring and sustainable electric vehicle. To further analyze model behavior, SHapley Additive exPlanations (SHAP)were applied to interpret prediction outcomes. Feature importance analysis revealed that battery temperature, internal resistance, and cycle count were the most influential factors contributing to battery degradation prediction, followed by voltage and current variations. These findings align with known electrochemical aging mechanisms validating the reliability of the explainable AI integration. The explainability layer provided transparent insights into model decisions, addressing the limitations of black-box deep learning approaches.

V. Conclusion Future Enhancement

The Battery Health Prediction and Defect Detection System is an intelligent and scalable method of proactive battery health maintenance. It uses machine learning algorithms and deep learning models to make predictions about the State of Health (SoH) of batteries and make predictions about potential degradation. The system uses machine learning algorithms such as Random Forest, LSTM, SVM, and Decision Trees to make predictions based on critical parameters of battery health, including voltage, current, temperature, internal resistance, and cycle counts. The decision-making process is made clearer and more understandable by using SHAP analysis. The system is built as a web application utilizing the Django framework and backed by a MySQL database. The system allows secure user interaction, real-time predictions, model handling, and viewing past predictions. The system has been shown to be accurate in experimental validation conducted by utilizing metrics of MAE, Root Mean Squared Error, and R^2, thus ensuring that it is effective in electric vehicle batteries, renewable energy storage solutions, and industrial automation, thus ensuring predictive maintenance of batteries. Future enhancements focus on making the system more intelligent, scalable, and adaptive for real-world battery monitoring. IoT integration and expanded sensor parameters will enable real-time data collection, instant health predictions, and predictive maintenance. Advanced deep learning models, including transformers and hybrid architectures, will improve accuracy in forecasting battery degradation. Mobile applications, cloud–edge computing, and federated learning will enhance accessibility, scalability, and data security. Improved explainable AI and adaptive learning will increase transparency, trust, and continuous performance improvement**.**

REFERENCES

1. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
2. Lundberg, S. M., & Lee, S.-I. (2017). *Advances in Neural Information Processing Systems*, 30.
3. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.
4. OpenCV Developers. (2024). OpenCV: Open Source Computer Vision Library. Retrieved from https://opencv.org
5. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
6. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
7. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.
8. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.
9. Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
10. Chollet, F. (2015). Keras. https://keras.io
11. Abadi, M., et al. (2016). TensorFlow: Large-scale machine learning on heterogeneous systems. https://tensorflow.org
12. Sharma, A., & Jain, S. (2020). Predictive maintenance using machine learning: A review. *Journal of Manufacturing Systems*, 56, 473-490.
13. Zhang, Y., & Zhao, Y. (2019). Battery health prediction based on LSTM neural networks. *Energy Storage Materials*, 21, 30-37.
14. Wang, J., et al. (2021). Automated defect detection in manufacturing using deep learning: A review. *IEEE Transactions on Industrial Informatics*, 17(4), 2715-2726.