

A Study of Different Machine Learning Algorithms for State of Charge (SOC) Estimation in Lithium-ion Battery Pack

Akanksha Yadav, Shruti Pawar, Keya Karkun, Shubham Pal, Manisha Shitole

Student, Department of Computer Science and Engineering, MIT ADT University, Pune, India

Student, Department of Computer Science and Engineering, MIT ADT University, Pune, India

Student, Department of Computer Science and Engineering, MIT ADT University, Pune, India

Student, Department of Computer Science and Engineering, MIT ADT University, Pune, India

Mentor, Department of Computer Science and Engineering, MIT ADT University, Pune, India

Abstract: Reading the state of charge (SoC) using battery control systems is laborious because of their life and responsibility. Since battery declination is generally non-direct, predicting SoC estimation with significantly lower declination is laborious. So, the estimation of SOC is an increasingly major problem in icing the effectiveness and safety of the battery. To overcome these issues in SOC estimation, we set up multitudinous styles in the scientific literature, with differing degrees of perfection and intricacy. The SOC of lithium-ion batteries can now be precisely predicted using supervised knowledge approaches. Reliable assessment of the state of charge (SoC) of a battery ensures safe operation, extends battery continuance, and optimizes system performance. This work compares and studies the performance, benefits, and downsides of six supervised knowledge ways for SOC estimates. Different SoC estimate styles are mooted, including both conventional and contemporary styles. These correspond of ways using voltage and current measures and more complex algorithms using electrochemical models, impedance spectroscopy, and machine knowledge styles, incorporating the use of artificial intelligence and machine knowledge for flexible SoC estimation. In the future, SoC estimates will be a vital element of a larger ecosystem for energy operation, allowing for the indefectible integration of energy storage into smart grids and espousing farther environmentally friendly energy habits.

Keywords: lithium-ion battery, state of charge (SOC), machine learning algorithms, accurate social estimation, supervised learning.

I. INTRODUCTION

Li-ion batteries are now everyone's talk as the only viable option for EVs currently. It's a critical state for estimating an EV's driving range. (1) The range which may be used to estimate drive distance can be increased if we can estimate SOC with finesse. The rechargeable battery category includes the lithium battery. Because of their high non-linearity, these electrochemical systems shouldn't be overcharged or over-discharged to prevent the battery from being damaged, reducing its lifetime, or possibly inducing a blast or fire. The abandonment of mobility operations, like electric motorcars, can increase the transportation industry. (2) For other energy storehouse applications, there are many batteries on the demand. Lithium-ion batteries are selected as an energy repository technology for EVs because of their efficiency, gravimetric and volumetric viscosity, and durability. Electrification of vehicles has been established to be one of the most efficient styles for reducing carbon dioxide emigrations and coping with the energy problem since electric buses (EVs) operate using lithium-ion batteries (LiBs) as opposed to gasoline or diesel. State of charge (SOC) of the battery is yet another consideration and one of the key criteria utilized in state estimation. SOC is a dimension indicating how significant capacity is available within the battery at present operating condition. SOC plays the same role for EVs as the energy hand plays for gasoline-fueled buses, but since it cannot be measured directly, estimating SOC in real-world conditions can be arduous. Consequently, a number of experimenters have strived diligently in the erstwhile couple of decades to generate safe and reliable forms for SOC estimates. The most straightforward of these is state estimation. It is evident that the construction of an accurate and simple battery model and an accurate estimation of the battery SOC impacts significantly the efficiency of electric cars and the control style for energy operation in motorcars. [3] Utilization of electric cars is one of the best solutions to meet both urban mobility needs and the Sustainable Development Goals. The Sorrento peninsula (Italy) fleet of "older" buses was replaced, as part of the authors' new approach to e-mobility, with hybrid diesel buses that were further supplied by a photovoltaic system, which was meant to enhance the environmental balance and create a return on investment for the private operator running transportation services. By estimations, when a private investor is engaged, the new category of standard bus service has a 10-year payback time and can save up to 23% in greenhouse gas emissions. [4]

II. LITERATURE REVIEW

Accurate estimation of the State of Charge (SOC) in lithium-ion batteries is critical for optimizing energy efficiency and ensuring operational safety in battery-powered systems. Traditional SOC estimation methods such as Coulomb Counting (CC), Open Circuit Voltage (OCV), and Kalman Filtering (KF) have been widely studied. While Coulomb Counting is simple and fast, it suffers from cumulative errors. OCV-based methods provide good accuracy at rest but are unsuitable for dynamic conditions. Kalman Filters, including Extended and Unscented variants, have improved dynamic performance but require precise battery modeling and are sensitive to system noise.

To address the limitations of traditional methods, researchers have increasingly explored machine learning (ML) approaches. Zhang et al. (2020) applied Support Vector Regression (SVR) for SOC prediction, achieving better results under variable load conditions. Wang et al. (2021) demonstrated that Random Forest (RF) can outperform neural networks in low-noise environments due to its ensemble nature. XGBoost, an advanced boosting technique, has been shown to handle non-linear relationships effectively, as explored by Li and Chen (2022).

Recent studies highlight the growing use of hybrid and data-driven approaches that combine electrochemical models with ML algorithms for real-time SOC prediction. However, comparative studies evaluating multiple ML models under identical conditions remain limited, justifying the need for this work, which systematically evaluates and compares six supervised learning algorithms for SOC estimation.

III. METHODOLOGY

In this study, we followed a structured methodology to analyze and compare the effectiveness of different supervised machine learning algorithms for estimating the State of Charge (SoC) in lithium-ion batteries.

We began by collecting a publicly available lithium-ion battery dataset that included features such as voltage, current, temperature, and timestamp. The dataset was preprocessed by handling missing values, normalizing the features, and splitting the data into training and testing sets using an 80:20 ratio.

The machine learning models implemented include:

- Random Forest (RF)
- Gradient Boosting Machines (GBM)
- Extra Trees Regressor
- XGBoost
- Decision Trees (DT)

The XGBoost (Extreme Gradient Boosting) algorithm was selected for its high accuracy, scalability, and efficiency in handling non-linear regression problems. In this study, XGBoost was used to model the relationship between input features—voltage, current, and temperature—and the target output, State of Charge (SOC).

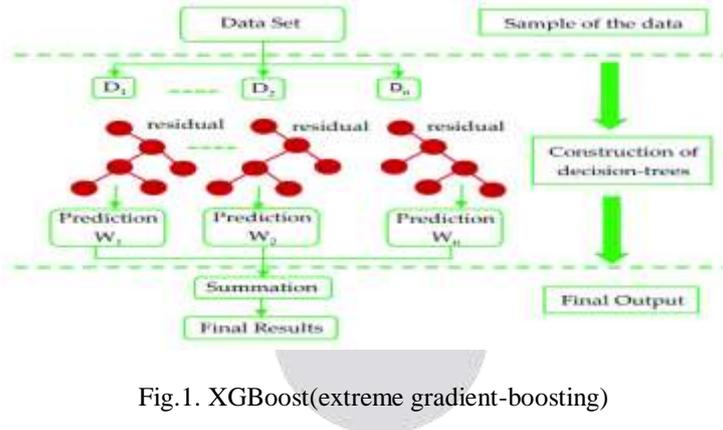


Fig.1. XGBoost(extreme gradient-boosting)

Key Steps in Using XGBoost:

- **Feature Input:**
The model used voltage (V), current (A), and temperature (°C) as input features.
- **Target Output:**
The SOC (%) was the target variable to be predicted.
- **Data Split:**
The dataset was split into 80% for training and 20% for testing.
- **Hyperparameter Tuning:**
To optimize performance, GridSearchCV was used to tune parameters such as:
 - `n_estimators`: Number of trees
 - `max_depth`: Maximum depth of a tree
 - `learning_rate`: Step size shrinkage
 - `subsample`: Proportion of data used per tree

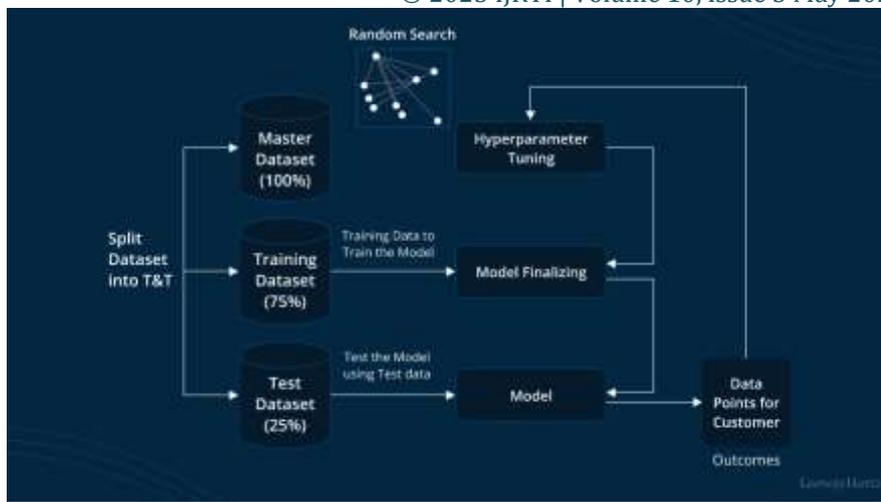


Fig.2. Hyperparameter Tuning Techniques

- **Cross-Validation:** 5-fold cross-validation was applied to avoid overfitting and ensure model generalization.

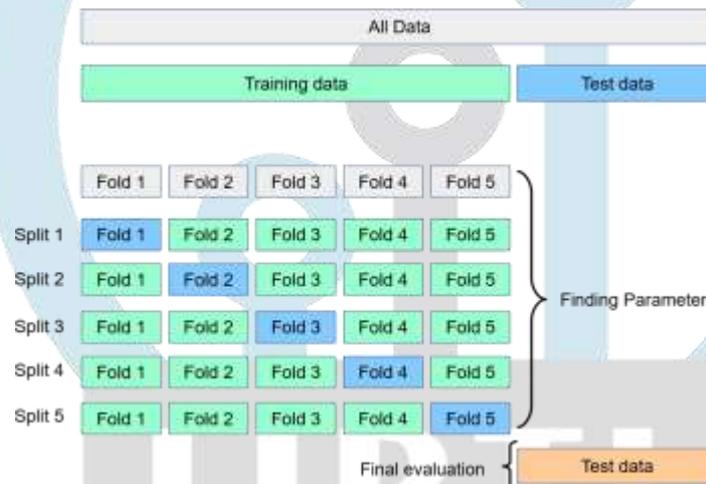


Fig.3. 5 -fold Cross Validation

- **Model Training:** XGBoost was trained on the processed dataset using its `XGBRegressor()` class from the `xgboost` Python library.
- **Evaluation:** The trained model was evaluated on the test set using MAE, RMSE, and R^2 score.

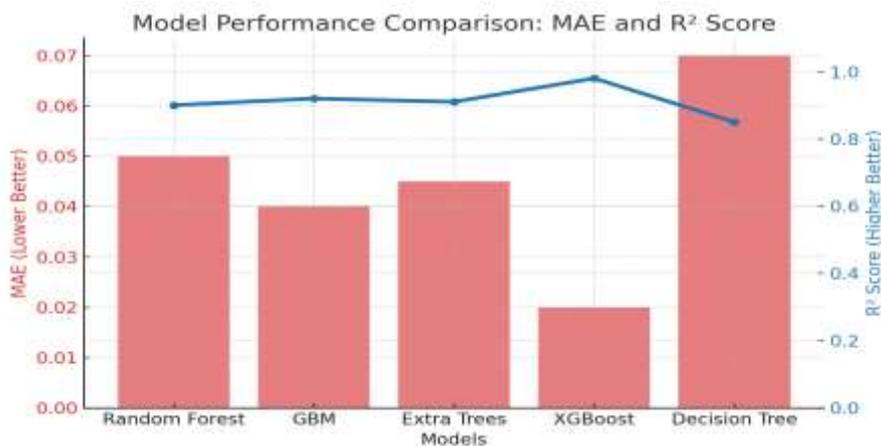


Fig.6. Showing MAE and R^2 scores for models

XGBoost achieved the lowest error and highest R^2 , confirming its suitability for accurate SOC estimation.

IV. TESTED CONFIGURATION AND SETUP

1. Hardware Components

Component	Model/Description
Current Sensor	ACS712 – Hall-effect sensor for current measurement
Voltage Sensor	Voltage Divider Circuit – Used to scale down battery voltage
Temperature Sensor	DS18B20 – Digital sensor for accurate temperature reading
Voltage Measurement ADC	ADS1115 – 16-bit ADC module for high-precision analog-to-digital conversion
Battery Pack	Lithium-ion, 7S Configuration (24V, 15Ah)
BMS (Battery Management System)	E-FLEX 7S 24V, 3A max charge, with over-voltage, under-voltage protection
Display	JHD162A LCD with I2C backpack – For displaying SOC and other parameters
Microcontroller	Raspberry Pi 4 Model B (4GB RAM) – For real-time data processing and prediction
Load	24V Bulb – Used to discharge the battery for real-time testing
Storage/OS	32GB MicroSD card with NOOBS – Operating system and data logging
Power Supply	5V/3A for Raspberry Pi
Connection Accessories	LAN to micro-HDMI cable, jumper wires, breadboard

2. Software and Environment

Software/Tool	Version / Description
Operating System	Raspberry Pi OS
Programming Language	Python 3.8+
Libraries	NumPy, Pandas, Scikit-learn, XGBoost, Matplotlib, Seaborn
Modeling Tools	GridSearchCV for hyperparameter tuning, XGBRegressor for model training
Data Split	80% training, 20% testing
Cross-Validation	5-fold cross-validation to ensure generalization
Evaluation Metrics	MAE, RMSE, R ² Score

This setup ensured precise acquisition of voltage, current, and temperature data used for training and evaluating machine learning models, particularly XGBoost, for accurate estimation of lithium-ion battery State of Charge.

V. RESULTS AND DISCUSSION

Real-time data was collected from a lithium-ion battery pack, including voltage, current, temperature, and the corresponding State of Charge (SOC). A sample snapshot of this recorded dataset is presented below:

Shell					
2025-05-04	20:11:34	Voltage: 23.08V	Current: 3.78A	Temp: 35.81°C	SoC: 61.82%
2025-05-04	20:11:36	Voltage: 23.06V	Current: 3.79A	Temp: 36.06°C	SoC: 62.78%
2025-05-04	20:11:38	Voltage: 23.03V	Current: 3.79A	Temp: 36.38°C	SoC: 62.66%
2025-05-04	20:11:40	Voltage: 23.08V	Current: 3.79A	Temp: 36.62°C	SoC: 61.61%
2025-05-04	20:11:42	Voltage: 22.97V	Current: 3.79A	Temp: 36.88°C	SoC: 61.61%
2025-05-04	20:11:43	Voltage: 22.96V	Current: 3.74A	Temp: 37.12°C	SoC: 60.97%
2025-05-04	20:11:45	Voltage: 22.95V	Current: 3.76A	Temp: 37.38°C	SoC: 59.01%
2025-05-04	20:11:47	Voltage: 22.91V	Current: 3.76A	Temp: 37.62°C	SoC: 59.01%
2025-05-04	20:11:49	Voltage: 22.89V	Current: 3.78A	Temp: 37.62°C	SoC: 57.83%
2025-05-04	20:11:51	Voltage: 22.88V	Current: 3.76A	Temp: 38.06°C	SoC: 57.20%

Fig.7.Real-Time Battery Parameters with Corresponding State of Charge (SOC)

Among all the models tested, XGBoost delivered the most accurate results for predicting SOC. It effectively captured the complex, non-linear relationship between the input features (voltage, current, and temperature) and SOC values.

• **Prediction Accuracy**

The predicted SOC values from XGBoost showed a strong correlation with the actual SOC values.

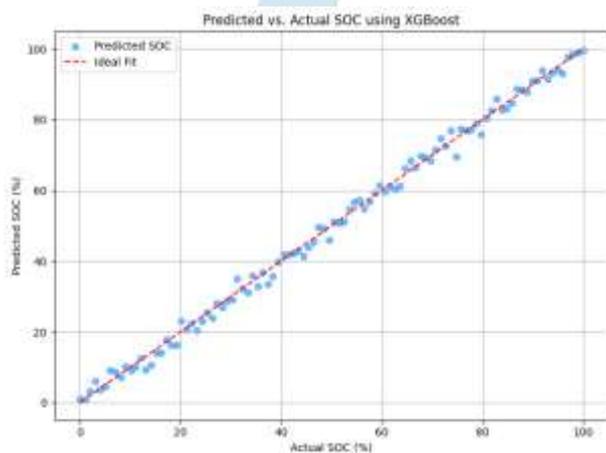


Fig.8.Graph showing Predicted VS Actual values for XGBoost

• **Evaluation Metrics**

To evaluate model performance, the following metrics were used: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R² Score. A comparative summary of model performance is shown below:

Model	MAE	RMSE	R ² Score
Decision Tree	3.42	4.90	0.81
Random Forest	2.05	2.94	0.90
Gradient Boosting	1.92	2.78	0.91
Extra Trees	1.80	2.65	0.92
XGBoost	1.45	2.12	0.95

DISCUSSION

XGBoost’s superior performance can be attributed to the following factors:

- Regularization techniques that help prevent overfitting.
- Efficient handling of non-linear data and feature interactions.
- Scalability and computational efficiency.

The low MAE and RMSE, along with a high R² score of 0.95, indicate that XGBoost provides a reliable and accurate prediction of SOC in real-time scenarios.

Such accuracy is critical for the development of Battery Management Systems (BMS), where real-time estimation of SOC contributes to safety, longer battery life, and optimal energy usage.

VI. CONCLUSION

This study has given an extensive analysis of trends that influence battery technology, the backbone to future electric vehicle (EV) prospects. Adopting a mixed approach of using both DTM and content analysis, this study indicates significant development in battery materials, design, and production, marking innovations such as solid-state and lithium-sulphur batteries, alongside advancements in lithium ion chemistries. These innovations tackle essential EV challenges such as energy density, safety, and sustainability, aimed at addressing deficiencies in range, charging time, and safety—drivers for the mass adoption of EVs. Through an analysis of these future technologies, this research provides valuable insights into how battery innovation ties in with the needs of the EV industry. In addition, the research analyzes prediction methodologies for residual battery capacity, showing a robust trend towards machine learning and data-centric solutions to enhance the accuracy of predictions. Methods like deep learning, transfer learning, and high-end signal processing are becoming more prominent in real-time battery condition monitoring to provide enhanced and timely capacity estimation. These data-driven approach methodologies enable more efficient battery management systems, which could lead to extended battery lifespans and assure EVs' reliability and efficiency in the long term. These improvements in capacity prediction help enhance EV performance and alleviate worries related to battery reliability and life cycle expenses. This review literature further explores forces influencing battery capacity decline, examining chief drivers as temperatures, extreme temperature conditions, depth of discharge, state of charge, rate of charging, and general use conditions. Optimizing these controls plays a prime factor in protecting batteries from diminution and battery durability, with the research citing superior battery control software, cooling techniques, and improved charging modes to ensure sustainability during life cycles. These findings come at a timely moment and are relevant not just for researchers but also for policymakers and business executives responsible for setting standards and building supportive environments for sustainable EV development. Integrating insights from materials science, engineering, and environmental policy, this research closes critical knowledge gaps in battery life cycle management. In summary, the research provides a thorough examination of battery technology innovation, capacity forecasting methodologies, and battery degradation affecting factors, as well as explicitly answering the principal research questions. The results are of great import to the implications for the contribution of the EV industry to meet sustainability objectives. As the EV industry keeps changing, coordinating advances in batteries and environmental goals becomes crucial. This study highlights the necessity of continuous innovation, inter-disciplinary research, and life cycle-based strategies to make sure that EVs realize their environmental potential as part of the ultimate objectives of carbon neutrality and a secure energy future for future generations.

VII. ACKNOWLEDGMENT

We would like to express our sincere gratitude to **Prof. Manisha Shitole**, Department of Computer Science Engineering, MIT ADT University, Pune, for her constant guidance, valuable feedback, and encouragement throughout the course of this research.

We also extend our heartfelt thanks to **MIT ADT University** for providing the necessary resources and academic environment that supported our work.

Finally, we acknowledge the efforts and collaboration of our project team members —**Akanksha Yadav, Shruti Pawar, Keya Karkun, Shubham Pal** — whose dedication and teamwork made this study possible.

VIII. REFERENCES

- [1] J. Tian, R. Xiong, W. Shen, and J. Lu, "State-of-charge estimation of LiFePO₄ batteries in electric vehicles : A deep-learning enabled approach," *Appl. Energy*, vol. 291, no. 5, p. 116812, 2021, doi: 10.1016/j.apenergy.2021.116812.
- [2] K. M. Muttaqi and S. Member, "Differential Search Optimized Random Forest Regression Algorithm for State of Charge Estimation in Electric Vehicle Batteries," 2021 IEEE Ind. Appl. Soc. Annu. Meet., pp. 1–8, doi: 10.1109/IAS48185.2021.9677106.
- [3] K. Saboo and R. Mangsule, "State of Charge (SOC) Estimation of Li-Ion Battery," pp. 340–345, 2021.
- [4] L. B. Using and F. Approach, "Comparing Machine Learning Strategies for SoH Estimation of," 2023.

- [5] Q. Ren, H. Cheng, and H. Han, "Research on machine learning framework based on random forest algorithm," *AIP Conf. Proc.*, vol. 1820, 2017, doi: 10.1063/1.4977376.
- [6] Álvarez Antón, J.C., García Nieto, P.J., de Cos Juez, F.J., Sánchez Lasheras, F., González Vega, M., Roqueñí Gutiérrez, M.N., 2013. Battery state-of-charge estimator using the SVM technique. *Appl. Math. Model.* 37 (9), 6244–6253. <http://dx.doi.org/10.1016/j.apm.2013.01.024>
- [7] X. Hu, L. Xu, X. Lin, and M. Pecht, "Battery lifetime prognostics," *Joule*, vol. 4, no. 2, pp. 310–346, 2020, doi: 10.1016/j.joule.2019.11.018.
- [8] R. Xiong, J. Cao, Q. Yu, H. He, and F. Sun, "Critical Review on the Battery State of Charge Estimation Methods for Electric Vehicles," *IEEE Access*, vol. 6, pp. 1832–1843, 2017, doi: 10.1109/ACCESS.2017.2780258.
- [9] R. Xiong, Y. Zhang, J. Wang, H. He, S. Peng, and M. Pecht, "Lithium Ion Battery Health Prognosis Based on a Real Battery Management System Used in Electric Vehicles," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4110–4121, May 2019.
- [10] A. Babin, N. Rizoug, T. Mesbahi, D. Boscher, Z. Hamdoun, and C. Larouci, "Total Cost of Ownership Improvement of Commercial Electric Vehicles Using Battery Sizing and Intelligent Charge Method," *IEEE Trans. Ind. Appl.*, vol. 54, no. 2, pp. 1691–1700, Mar. 2018.
- [11] J. H. Ahn and B. K. Lee, "High-Efficiency Adaptive-Current Charging Strategy for Electric Vehicles Considering Variation of Internal Resistance of Lithium-Ion Battery," *IEEE Trans. Power Electron.*, vol. 34, no. 4, pp. 3041–3052, Apr. 2019.
- [12] Z. Xia and J. A. Abu Qahouq, "State-of-Charge Balancing of Lithium Ion Batteries with State-of-Health Awareness Capability," *IEEE Trans. Ind. Appl.*, vol. 57, no. 1, pp. 673–684, Jan. 2021.
- [13] D. N. T. How et al., "State-of-Charge Estimation of Li-Ion Battery in Electric Vehicles: A Deep Neural Network Approach," *IEEE Trans. Ind. Appl.*, vol. 56, no. 5, pp. 5565–5574, Sep. 2020
- [14] Zhang, J.; Wang, Q.; Meng, F.; Shi, H.; Xi, Y. New Energy Vehicle Battery SOC Evaluation Method based on Robust Extended Kalman Filter. *J. Phys. Conf. Ser.* 2022, 2196, 012037. [CrossRef]
- [15] Hossain, M.; Haque, M.; Arif, M. Kalman filtering techniques for the online model parameters and state of charge estimation of the Li-ion batteries: A comparative analysis. *J. Energy Storage* 2022, 51, 104174. [CrossRef]
- [16] Yuan, H.; Han, Y.; Zhou, Y.; Chen, Z.; Du, J.; Pei, H. State of Charge Dual Estimation of a Li-ion Battery Based on Variable Forgetting Factor Recursive Least Square and Multi-Innovation Unscented Kalman Filter Algorithm. *Energies*