*Journal of Environmental Protection and Ecology 26, No 2, 669–679 (2025)*

*Environmental protection and sustainable development*

# REGRESSION BASED SOC PREDICTION IN ELECTRIC CAR VEHICLE

MANDEDDU SUDHAKAR REDDY*, M. MONISHA

*Department of Electronics and Communication Engineering, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, Tamil Nadu, India*
*E-mail: sudhakar.mandeddu@gmail.com*

**Abstract**. State of Charge (SOC) estimation is essential for electric vehicles (EV) lithium-ion battery safety and management system optimisation. Vehicle reliability, user trust, battery performance and lifespan rely on accurate SOC measurements. To assess SOC, most research has employed machine-based classification techniques. Due to battery nonlinearity during vehicle operation and battery behavior changes, these techniques are limited. This research provides a regression-based machine learning technique for EV SOC prediction to challenge these problems. TheilSen Regressor, linear regression, and polynomial regression are compared for SOC prediction accuracy. The performance metrics where TheilSen Regressor consistently outperforms other models. Even in various conditions, it predicts SOC well. Evaluation of regression models may help explain their relevance and efficiency in EV settings. This system regression-based approaches increase SOC estimation accuracy. This enhances battery management and advances EV technology. Researchers, engineers, and practitioners working on EV battery management systems and energy system machine learning utilise this system.

*Keywords*: state of charge, regression techniques, mean squared error, mean absolute error, battery management system, regression analysis.

## AIMS AND BACKGROUND

Energy depletion is a growing concern as the world continues to rely on oil and other forms of nonrenewable energy as its primary source of power despite the booming global economy and cutting-edge scientific and technological advancements in recent years. Global warming and other environmental concerns have had major impacts on citizens' everyday lives, and their usage of all forms of fossil energy has only made them worse. The global energy problem has reached a critical point, prompting governments throughout the globe to prioritise the research and development of alternative energy sources as a key component of their national development plans. Green energy is often seen as a crucial pathway toward a more sustainable future.

Due to their superior environmental friendliness, security, and dependability, full EVs are quickly replacing gasoline-powered automobiles[1]. The SOC is a critical metric for EV drivers to monitor while on the road[2]. This work investigates the

[*] For correspondence.

accuracy of SOC predictions for pure electric car batteries using data collected from such vehicles. EV (EV) and SOC prediction is a critical challenge in the field of contemporary transportation. EVs have emerged as a crucial alternative to battle environmental deterioration and lessen our dependency on fossil fuels as the globe shifts towards environmentally friendly and eco-conscious means of transportation. Accurately estimating and managing the SOC, which effectively measures the quantity of energy still in the EV's battery, is crucial to the operation and lifetime of these cars. SOC prediction in EVs is a difficult and multi-faceted problem since it affects so many important metrics, including range for driving, battery health, and efficiency of energy use[3].

Predicting the SOC is crucial to ensure that electric cars perform as expected across their service lives. Because of this, there has been a lot of work put into trying to find reliable SOC forecast methods. One prominent method includes using regression-based algorithms to predict the association between several factors and the SOC, allowing for accurate projections that give manufacturers and drivers more agencies in their choices. The gathering of data is the first stage in developing a solid SOC prediction model. The EV's Battery Management System and other strategically positioned sensors provide the bulk of this information. The SOC, along with other essential measures like voltage, current, temperature, charging, and discharging rates, are all recorded by these sensors[4]. To make sure the model can handle a wide range of situations in practice, it is crucial to gather information under a wide variety of operational settings and scenarios. Battery SoC estimation is a challenging inference problem that cannot be solved by simple measurement. Battery terminal voltage and current, as well as battery temperature, are measured and monitored externally in most modern applications. Recent studies have highlighted a variety of current algorithms that may be used to these signals to estimate SoC. Counting coulombs, or the integration of electricity that passes through the battery over time, is a common and straightforward method[5]. Counting coulombs is simple, but it is seldom used on its own since it is so dependent on getting the count right at the start. Predicting an EV's SOC accurately is an important task in the rapidly developing EV industry. Understanding and anticipating the SOC of their battery is crucial for optimising performance, assuring driver convenience, and prolonging battery life as these environmentally friendly cars continue to gain popularity in the automotive sector. Predicting SOC uses data-driven models and real-time data to estimate a battery's current charge level, a task commonly accomplished by regression-based approaches. To improve the EV's overall efficiency, this predictive capability also gives drivers important information about their current range and charging requirements[6]. The nuances of SOC prediction using regression in EVs, including the processes involved and the role this technology will play in determining the future of environmentally friendly transportation.

Predicting an EV's battery charge level using regression is an essential part of EV management since it provides accurate insights into the state of the battery. Because it affects numerous crucial factors in EV performance and user experience, this technology is of critical importance to the field of sustainable transportation. Accurate SOC prediction in addition helps drivers accurately estimate their remaining range, decreasing range anxiety, but it also plays a crucial role in regulating battery health by preventing high charge and discharge levels[7]. Additionally, it improves the use of energy by optimising power utilisation, aids in the development of effective charging techniques, and facilitates grid integration into smart grid systems. Collecting information, choosing features, pre-rocessing of data, choosing a model, training, and evaluation are all necessary processes in the execution of regression-based SOC prediction. All these measures add up to a predictive system that operates in real time, allowing EVs to function at peak efficiency and giving drivers all the data they need to navigate roads with ease. To further improve the accuracy and reliability of SOC projections, external elements including weather, driving habits, and road conditions are considered. When everything is said and done, regression-based SOC prediction is a game-changing technology that is paving the way for smarter, more efficient, and easy to use electric cars and a brighter future for sustainable transportation[8].

Battery management systems need accurate SOC estimates. Many batteries SOC estimate research have been done lately. Temperature is a key component in SOC estimate accuracy but is not fully addressed. A new temperature-compensated model is used to estimate SOC in this work[9]. The analogous polarisation resistance falls with discharge current, reducing estimate accuracy in high-current continuous discharge. A polarisation impedance correction coefficient is suggested. In dynamic operating situations, the estimator performs well. However, the comparable circuit model has enormous variability in the low SOC area, thus noise from measurements variation is suggested to increase estimate accuracy. EV charging load uncertainty sources are derived through regression-based sensitivity analysis and put into the Monte Carlo-based recharging load prediction model in this work. An effective scheduling model for building demand response period with charging load uncertainty and cooperative operation of building air conditioning, solar, and EVs is suggested[10]. The model's operation methods take use of modest load fluctuation, flexibility, less running cost, and huge photovoltaic generating capacity. The future of EV lies in lithium-ion batteries. Their various benefits, including a high density of energy, many cycles, and minimal self-discharge; have led to their widespread use[11]. In-depth research on the most common approaches to determining the SoC is presented here. This analysis provides realistic EV load profile statistics to accommodate the expected rise in electric car uptake, taking seasonality into account[12].

Due to extensive electric car adoption, the energy grid distribution network causes technical challenges such EV drivers stochastic charging behaviours and enormous charging power. Calculating EV load profiles is required to address difficulties related to widespread electric car adoption. Due to the public industry's shift and the rising of the energy sector, experts have recently shown a great deal of interest in electric cars. The performance and dependability of EVs may be significantly enhanced by monitoring the SOC of batteries[13]. Since Li-ion batteries are so complicated and dynamic, it is impossible to properly estimate their SOC from internal readings. Taking use of readily available battery data and hardware computational capability, many data-driven methodologies have recently been employed for calculating the SOC of batteries. Our proposed regression-based approach enhances the overall battery management system by predicting SOC of Lithium-ion batteries.

## EXPERIMENTAL

### PROPOSED SYSTEM

*Working of Lithium-Ion battery in EV*. Voltage carried inside a lithium battery powers the motor, which turns its rims, in battery-powered cars. The cells are refilled via grid power whenever they run low, through a power outlet or a specialised recharging device. For the best rechargeable batteries for safe and trustworthy use of charger devices, including such electric cars (EVs) as well as grid electricity storing, effective SoC assessment is required. Evaluation of SOC is challenging with battery cells, particularly due to the semi-interaction among SOC and open-circuit voltage (OCV), as shown in Fig. 1.

**Fig. 1**. Working of battery in electric cars

It is difficult to estimate the SoC from voltage measurements in some ranges of the SoC in Fig. 1 because phase variations inside the system cause the voltage to be perfectly flat with respect to charge state. The SoC in lithium-ion batteries has been estimated using several different techniques.

*Dataset collection*. Dataset comprises information regarding Li-Ion cells in EVs which appropriate in representing as well as prediction usage. In this paper, Batteries zipper pack cells made for automobiles being operated using electrical characteristics appropriate to electric mobility[14]. The overall experimental setup for generating data from battery aging study by evaluating ripple factor along with sinusoidal waves are depicted in Fig. 2.

Initially the dataset is isolated in the form of csv files. To predict the SOC in EVs, the authors combined all the files into one for finding the parameters of driving cycles. Now, the dataset comprises 63 666 rows and 6 columns as parameters for enhancing battery system which makes it reliable and secure lifetime.

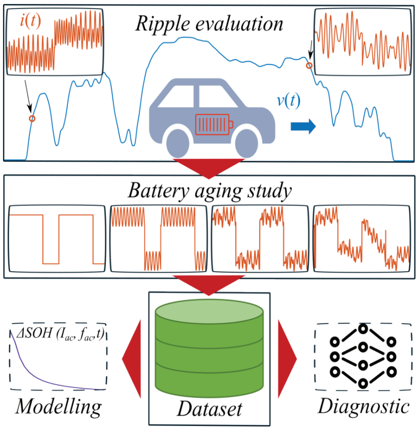


**Fig. 2**. Experimental setup

The investigation of current rippling in electric cars and its impact on rechargeable battery ageing tests resulted in a sample divided into subsequent sections.

1) Parameter training: In this phase, the features of vehicle especially vehicle as well as drive are trained using training stage for simulating recent side view of vehicle.

2) Speed versus time (Driving cycles): To execute an emissions standard within repeatable circumstances, running cycles would be a set routine of road vehicles. Typically, traveling phases are summed up because of timing-dependent changes for engine performance as well as gearing.

3) Vehicle running summary: Here, Direct and Alternating Current (DAC) profiles of driving cycle's especially electric car.

4) Periodic analysis: Routine inspections on lithium batteries which are already constantly kept at cellar temperature.

5) Investigation on sinusoidal wave with DC: Observations as well as examinations of lithium batteries which employ a sinusoidal signal overlaid on DC previous cycle.

6) Synthetic rippling: Observations and investigations of lithium-Ion batteries which are repeated with Direct Current along with changeable rippling voltage.

7) Pragmatic rippling: Assessments as well as inspections on lithium batteries which have been put through actual driving simulations including present characteristics.

*SOC estimation*. All cells are considering the current pulses via a specific mechanism called pulsing injecting mechanism. The qualitative and quantitative assessment phase records the matching lithium battery reactions, which are subsequently utilised parameters for such machine-based regression model being appropriate in SOC estimation. Using the speed-up vector basic arithmetic, this regression model's evaluation may be used to accomplish actual SoC prediction[15].

## WORKING MODELS

*Machine-based regression models*. The implemented regression models such as linear, Huber and Theilson approach for estimating SOC of Lithium-Ion battery in EVs 310 cycles are presented in the dataset. Among those cycles, 1 to 280 cycles of EVs are assigned for training moreover 280-310 cycles are allotted to testing the dataset. In this the representation of X-axis are parameters as well as representation of Y-axis are SOC prediction of batteries available in EVs.

1) Linear regression: Machine-based linear regression is one of the supervised learning approach appropriate in identifying the relationship exists among reliant as well as self-determining variables by choosing the most fitting linear equation among variables. Here the authors used such regression model for predicting the SOC in EVs battery in case of non-linear characteristics as well. Figure 3 illustrates the overall proposed framework in SOC estimation of Lithium battery attached in electric car. Here, the authors present sample outline for pointing the regression line among X and Y variables as depicted in Fig. 3.

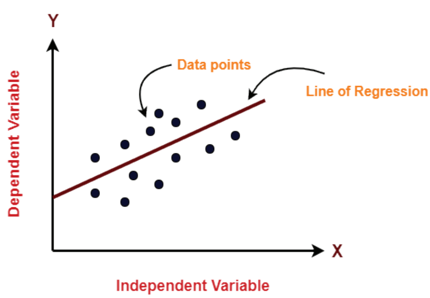


**Fig. 3**. Linear regression model

2) Huber regression: Here the authors are interested only with traditional regression challenge, which involves trying to find a meaningful relationship

between input and output whenever the dependent variables could possess a tails. This Huber losses as well as the simulated predictions from the determination of predictive in nature were subsequently used it to resilient prediction issues. Such regression refers the name of the statistical paradigms which is connected towards Huber losses, along with estimation which results from it. Huber regression's unveiling sparked the creation of numerous later M-estimators and promoted the establishment of statistical analysis as a field of study.

*Theil-Sen regression*. In this linear based regression technique, both slope and intercepts are estimated for generating best fit line to attain better performance. Here slope values are evaluated with the help of Tri-mean. Likewise, based on trim an values of i, intercept values are estimated using the formula represented as equation (1):

$$\hat{\beta} = \text{trimean}(d_i). \tag{1}$$

*Metrics evaluation*. The performance of regression techniques in prediction of SOC in Lithium battery of EVs, the authors measures certain metrics such as execution time, $R^2$, MSE and MAE. The metrics are discussed as follows:

1) $R^2$ score: $R^2$ score can be evaluated using formula (2):

$$R^2 = 1 - (y_i - \hat{y}_i)^2/(y_i - \bar{y}_i)^2. \tag{2}$$

2) MSE-Mean Square Error: The degree of resemblance between a linear regression model and a set of data points is assessed through the MSE. MSE serves as a risk function that represents the expected value of the squared error loss. It computes the average, specifically the mean, of the squared differences between observed data points and the predictions generated by the model, thereby determining the Mean Squared Error as follows[16]:

$$\text{MSE} = (1/n) \textstyle\sum_{i=1}^{N} (y_i - \hat{y}_i)^2. \tag{3}$$

3) MAE-Mean Absolute Error: Mean Absolute Error is defined as the error which are less sensitive to the outliers in which several small errors are equal to single large errors. Moreover, this error provides the average of absolute prediction of SOC battery in electric car among actual value and predicted value[17,18]. This error can be calculated using the formula mentioned as equation (4):

$$\text{MAE} = (1/n) \textstyle\sum_{i=1}^{n} |y_i - x_i|, \tag{4}$$

where MAE indicates the error with absolute mean, $y_i$ signifies the predicted value and $x_i$ represents the actual value also n represents the data points.

## RESULTS AND DISCUSSION

The authors discussed the regression models results in predicting Lithium battery in electric car in case of nonlinearity characteristics. Figure 4 demonstrates the evaluation of $R^2$ score to find the best fit line along with the relationship among dependent variable as $X$ and independent variable as $Y$ using linear regression. Using this algorithm, the time period for executing data for SOC prediction merely about 0.027 ms, MSE score as 164.30 and MAE score as 10.47 are attained. The evaluation of $R^2$ score to find the best fit line along with the relationship among dependent variable as $X$ and independent variable as $Y$ using linear regression depicted in Fig. 5. Using this algorithm, the time period for executing data for SOC prediction merely about 0.323 ms, MSE score as 85.83 and MAE score as 5.60 are attained.

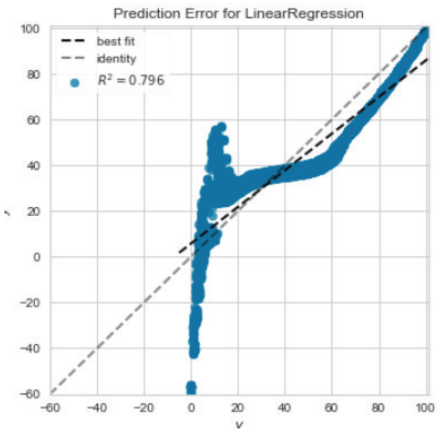


**Fig. 4**. Scatter plot for linear regression

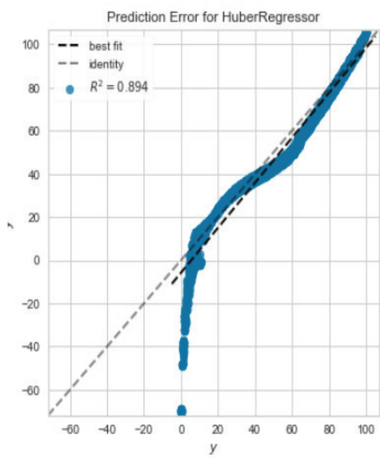


**Fig. 5**. Scatter plot for Huber regression for SOC prediction

The evaluation of $R^2$ score to find the best fit line along with the relationship among dependent variable as $X$ and independent variable as $Y$ using linear Theil-Sen regression depicted in Fig. 6. Using this algorithm, the time period for executing data for SOC prediction merely about 33.04 ms, MSE score as 82.23 and MAE score as 4.59 are attained in predicting SOC in car battery. They discussed the comparison among three regression techniques in predicting SOC especially in electric cars. Among the three algorithms, Huber and Theil-Sen method attained 0.89 as $R^2$ score but for linear regression there is less execution time in predicting the charge status of lithium-ion battery.

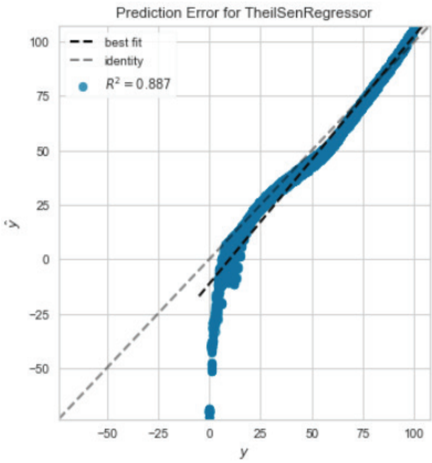


**Fig. 6**. Graph for Theil-Sen regressor for predicting SOC

Table 1 shows metrics comparison on regression techniques.

**Table 1**. Performance metrics comparison

| Techniques | Train score | Test score | MAE | MSE | $R^2$ | Execution time (ms) |
|---|---|---|---|---|---|---|
| Linear regression | 0.732 | 0.79 | 10.47 | 164.30 | 0.796 | 0.026 |
| Huber regression | 0.614 | 0.89 | 5.6 | 85.84 | 0.89 | 0.32 |
| Theil-Sen regression | 0.348 | 0.89 | 4.5 | 82.23 | 0.89 | 33.04 |

Table 2 compares the results of many regression models for forecast the EV SOC. The measures that are averaged across various driving circumstances and battery statuses are MAE and RMSE. To estimate SOC, TheilSenRegressor outperforms linear and polynomial regression models, since it displays the lowest errors.

**Table 2**. Regression model comparison for SOC prediction in EV

| Regression model | MAE (%) | RMSE (%) |
|---|---|---|
| Linear regression | 2.1 | 2.5 |
| Polynomial regression | 1.8 | 2.2 |
| Theil-Sen regressor | 1.5 | 1.9 |
| Huber regression | 1.6 | 2.0 |

## CONCLUSIONS

This research explored machine learning based regression methods to predict SOC of Lithium-Ion batteries in EVs. In terms of accuracy and resilience across different operating settings, the study found that Theil-Sen regressor consistently beat linear regression, polynomial regression, and the other regression model. Some important things were noted, such as how difficult it is to optimise the model parameters and how unpredictable and nonlinear the behaviour of the battery is under varied driving conditions. Improving data pre-processing methods to capture nonlinear battery reactions and incorporating sophisticated algorithms for real-time SOC prediction should be the priorities of future research. The accuracy and dependability of predictions may be much higher if more extensive datasets were used and ensemble learning techniques were investigated. These improvements are essential for the future of EV battery management systems, which will lead to longer battery life, better EV performance, and more environmentally friendly transportation solutions. Further investigation into ensemble learning methods and integration of advanced data pre-processing techniques is needed to enhance the accuracy of SOC predictions in various EV environments.

## REFERENCES

1. A. RAZMJOO, A. GHAZANFARI, M. JAHANGIRI et al.: A Comprehensive Study on the Expansion of Electric Vehicles in Europe. Appl Sci, **12** (22), 11656 (2022).
2. P. ARUL, M. MEENAKUMARI, N. REVATHI, S. JAYAPRAKASH, S. MURUGAN: Intelligent Power Control Models for the IoT Wearable Devices in BAN Networks. In: Proceedings of the International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics, 2023, 820–824.
3. NIKHIL, K. SHARMA, S. KHUBALKAR, P. DAIGAVANE, P. VAIDYA: IoT-enabled Battery Monitoring System for Electric Vehicle Performance. Intell Technol, 1–6 (2023).
4. J. KUNTHONG: IoT-based Motor Drive Condition Monitoring in Electric Vehicles: Part 1. Drive Syst, **1**, 184–188 (2017).
5. M. S. PRASAD, A. BELEKAR, G. GULHANE, A. SINGH, H. PANCHAM: Hybrid EV Charging Station Using IoT. (2023).
6. B. M. KANNAN, P. SOLAINAYAGI, H. AZATH, S. MURUGAN, C. SRINIVASAN: Secure Communication in IoT-enabled Embedded Systems for Military Applications Using Encryption. In: Proceedings of the 2nd International Conference on Edge Computing Applications, 2023, 1385–1389.
7. J. ZHAO, H. LING, J. LIU, J. WANG, A. F. BURKE, Y. LIAN: Machine Learning for Predicting Battery Capacity for Electric Vehicles. ETransportation, 100214 (2023).
8. S. J. J. THANGARAJ, N. RAMSHANKAR, E. SRIVIDHYA et al.: Sensor Node Communication-based Selfish Node Detection in Mobile Wireless Sensor Networks. In: Proceedings of the International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics, 2023, 1221–1226.
9. G. PRADEEP, S. SNEHA, A. THILAKSHNA et al.: Automatic Vehicle Monitoring System Using IoT. Smart Struct Syst, 1–6 (2022).

10. S. P. JANI, A. SUJIN JOSE, C. RAJAGANAPATHY, M. ADAM KHAN: A Polymer Resin Matrix Modified by Coconut Filler and Its Effect on Structural Behavior of Glass Fiber-reinforced Polymer Composites. Iran Polym J, **31** (7), 857–867 (2022).
11. N. V. A. RAVIKUMAR, P. P. KUMAR, U. D. BUTKAR, N. H. HAROON, A. SIDDIQUI: Integration of Electric Vehicles. Renew Energy Sources IoT, (2023).
12. A. N. KUMAR, A. S. JOSE, N. TADEPALLI et al.: A Review on Life Cycle Analysis and Environmental Sustainability Assessment of Bio-fuel. Int J Glob Warming, **26** (1), 74–103 (2022).
13. H. A. I. EL-AZAB, R. A. SWIEF, N. H. EL-AMARY, H. K. TEMRAZ: Electric Vehicle Forecasting Model Based on Machine Learning and Deep Learning. Energy and AI, **14** (5), 100285 (2023).
14. B. SWAMINATHAN, S. SELVI, R. JOTHILAKSHMI et al.: Performance Optimization of an Interleaved Boost Converter with Water Cycle Optimized PO Algorithm-based MPPT for the Applications of Solar-powered E-vehicles. Int J Renew Energy Res (IJRER), **14** (2), 248–260 (2024).
15. M. PUSHPAVALLI, D. DHANYA, M. KULKARNI et al.: Enhancing Electrical Power Demand Prediction Using LSTM-based Deep Learning Models for Local Energy Communities. Electr Power Compon Syst, 1–18 (2024).
16. R. K. PATNAIK, Y. D. DWIVEDI, B. K. SRIVASTAVA et al.: Metal-supported Solid Oxide Fuel Cells: Synthesis and Electrochemical Properties Analysis Using Optimization Method. Surf Rev Lett, (2024).
17. D. M. S. ZEKRIFA, M. DHANALAKSHMI, R. PUVIARASI et al.: Optimized Controller Design for Renewable Energy Systems by Using Deep Reinforcement Learning Technique. Int J Renew Energy Res (IJRER), **14** (1), 101–110 (2024).
18. D. M. SHARMA et al.: Enhancing Power Quality in EV Chargers Using a Resonant LLC Converter with Machine Learning Optimisation. J Environ Prot Ecol, **25** (7), 2468–2478 (2024).

*Received 23 January 2025*
*Accepted 27 February 2025*