REST Journal on Emerging trends in Modelling and Manufacturing
Vol: 7(3), 2021
REST Publisher
ISSN: 2455-4537
Website: http://restpublisher.com/journals/jemm/

Hist Gradient Boosting Regression for Electric Vehicle Charging Duration and Charge State Prediction

Vijesh Malappurath Paramadam
Application Architect
Corresponding Author Email: Vijesh.Paramadam@gmail.com

Abstract

This study analyses electric vehicle charging patterns using Hist Gradient Boosting Regression to predict critical parameters including charging time, energy consumption, and final battery levels. The analysis examines 100 charging sessions, revealing unique relationships between initial battery levels, charging time, energy consumed, and final charge levels. Strong positive correlations were identified between charging time and energy consumption (0.9), as well as between initial and final battery levels (0.88). The model showed exceptional performance in predicting final battery capacity, with consistent R^2 values were 0.979 for training and 0.981 for test data, indicating strong generalization capabilities. However, predictions for charging time and energy consumption exhibited significant performance degradation between training and test phases, suggesting over fitting challenges. These findings provide valuable insights for charging infrastructure optimization, grid management strategies, and the development of intelligent battery management systems to support the sustainable expansion of electric vehicle adoption.

Keywords: Electric Vehicle Charging, Hist Slope Incremental Regression, Battery Charge State, Energy Consumption Prediction, Charging Infrastructure Optimization, Machine Learning, Plot Management

Introduction

The global electric vehicle fleet is growing rapidly and is expected to exceed 130 million units by 2030. This significant expansion has led to a decade of extensive research into EV charging methods, using methods ranging from early exploratory studies to sophisticated, data-driven approaches. In this area, the current study uses the Caltech JPL site dataset to analyse specific charging behaviours of EV users. This study uses a hybrid analysis approach, combining intersection-based clustering with established methods such as k-means and Ward connectivity, to explore user charging patterns based on connection time, session duration, and energy consumption. The analysis identifies four distinct user groups, each with unique charging behaviours and energy requirements. These findings provide important insights for enhancing grid management and charging infrastructure optimization strategies [1]. The shift from fossil fuel to electric cars is contingent upon the advancement of a robust charging network. In the Korean context, while there are several public charging stations, with the growth of home-based charging solutions faces unique barriers rooted in the country's specific urban and residential landscape. Effective planning for the electric vehicle future depends on a clear understanding of driver charging patterns to guide the strategic use of charging facilities. This research addresses this need by examining a week of 2021 charging data from 297 EV drivers. The method uses latent class analysis to identify unique behavioural profiles and multi-legit analysis to identify key factors influencing these behaviours [2]. This manuscript uses a functional data analysis approach to examine electric vehicle charging behaviours utilizing actual information from Kansas City, Missouri's 455 charging stations, over five years. By analysing daily utility occupancy, daily energy usage, as well as station-level fluctuations, the study aims to develop theoretical support for strategic Planning for EV infrastructure, implementing regulations, and efficiently managing power grid load. The proposed operational data analysis method offers distinct advantages over traditional discrete-time-based models by preserving the smooth, continuous nature of charging behaviour, providing greater modelling flexibility, and accommodating time series data of varying lengths with minimal assumptions. By using this powerful approach, our research not only identifies key charging patterns, but also reveals an important finding: significant discrepancies between charging and parking times, which highlights the urgent need for improved parking regulation and enforcement at charging stations [3]. COVID-19 lockdowns have profoundly altered global electric vehicle charging patterns, as stay-at-home orders have disrupted daily life and grid operations. These behavioural changes, which include long-term changes such as the increase in remote work, are having significant impacts on grid congestion. This research examines the impact of different lockdown phases on EV charging in three specific locations. It further projects future grid congestion levels by modelling different scenarios for EV adoption rates and the potential for long-term displacement of work activities to home [4]. Research on the integration of electric vehicles and their Influence on the electricity grid started in the 1980s. Numerous later investigations have assessed regional electric vehicle adoption rates and their potential effects on electricity demand. A consistent conclusion throughout this work is that unmanaged electric vehicle charging will keep pace with and increase current peak loads on the grid [5]. While the literature has explored PHEV charging from various angles, the most widespread focus has been on simple overnight charging, which is typically initiated in the evening. This paper advances the field by introducing, for the first time, a comprehensive charging method that simultaneously optimizes overall energy cost, battery condition, electricity costs, and actual PHEV driving habits [6]. Electric vehicles offer a dual benefit: They emit fewer pollutants that exacerbate smog and climate change, and they diversify the U.S. transportation fuel mix. As electric vehicle technology has advanced, with driving ranges now exceeding 200-300 miles, their viability for mainstream adoption has increased. However, this growth depends on a well-planned charging network. Determining the location and capacity of this infrastructure scientifically is critical to promoting widespread electric vehicle adoption. In response to this need, modelling tools have been developed, such as the National Renewable Energy Laboratory's Electric Vehicle Infrastructure Project Model, which calculates the quantity of charging infrastructure needed to support a national electric vehicle fleet [7]. In 2016, the United States saw a significant increase in the number of newly registered battery and plug-in electric cars, with over 159,000. While this still represents a small overall market share, their combined penetration rate within the private transportation industry is significantly higher. This growing presence has prompted Legislators should reconsider infrastructure planning to create room for these vehicles and analyse the environmental sustainability of new infrastructure projects [8]. Replacing fossil fuel vehicles with electric vehicles is critical to changing global energy demand. However, successfully integrating the growing fleet plugging electric cars into the electrical grid is dependent upon reliable charging control. Without this, uncontrolled charging can cause significant grid load fluctuations and degrade power quality [9]. One important electrification strategy is the application of plug-in hybrid electric car mitigate environmental issues. They offer a unique dual benefit: providing the extended range of a traditional gasoline-powered car while enabling emission-free, electric-powered travel for typical daily distances [10].

Materials And Method

Materials

Initial Battery Level (%): The initial battery state of charge denotes the battery's degree of charge in an electric car pack at the time a specific process begins, typically at the start of a charging session. It is an important parameter, typically stated as a proportion of the battery's overall capacity, where 0% denotes complete discharge, and 100% denotes complete charging. This initial value is essential for calculating the required charging time, calculating the amount of energy required to achieve the desired SOC, and analysing charging behaviour. It also serves as a key input for managing battery health and optimizing grid load during charging events.
Charging Duration (hrs): Charging time is the total time required to fill an electric vehicle's battery from a specific starting charge level to a target level. It is a key metric measured from the moment the charging session is initiated until it is manually stopped or automatically stopped once the desired charge level is reached. This time period is not fixed; it is affected by many factors, including the initial battery level, the capacity of the battery, the power output of the charging station (Level 1, 2, or DC Fast), and the vehicle's on board charger limitations. Understanding charging time is essential for user convenience, infrastructure planning, and managing demand on the power grid.
Energy Consumed (kWh): The total amount of electrical energy delivered to an electric vehicle's battery during charging is the energy consumed, measured in kilowatt-hours (kWh). This represents the actual amount of electricity transferred to the vehicle, similar to the number of gallons pumped into a gasoline car. This measurement is calculated by multiplying the power consumption (in kilowatts) by the time the energy flows (in hours). It is an important measurement for determining the price of charging, evaluating vehicle performance, comprehending user charging habits, and projecting the total load and energy demand that electric vehicles place on the power grid.
Final Battery Level (%): Final battery state refers to the battery condition of an electric car at the end of a specific event, usually a charging session. Expressed as a percentage of total capacity, it represents the end point of the energy transfer process. This measurement is crucial for analysing charging efficiency and user behaviour, as it reveals the target charge level a driver chooses. It directly affects the previous charging duration and the total

energy consumed. Furthermore, the final battery state is a key determinant of the vehicle's subsequent driving range, and is essential for studies on battery longevity and grid load management.

Instructions For Machine Learning

Hist Gradient Boosting Regression: This advanced machine learning technique is based on histograms technique within the ensemble learning family, known for its high predictive performance on tasks involving continuous outcomes. Its core algorithm revolves around the sequential construction of multiple decision trees, where each new model is strategically trained to address the remaining discrepancies between the prior values' actual and anticipated values in the sequence. This iterative correction process is guided by the principles of gradient descent, which systematically minimizes a predefined loss function, such as the mean square error, which quantifies the overall error of the prediction of the model. The "histogram-based" component refers to a specific optimization that combines continuous feature values into separate histograms, dramatically speeding up the training process by reducing the computational overhead required to find optimal split points in the trees. By combining these relatively weak, individual predictors (trees), the algorithm builds a single, robust, and highly accurate ensemble model. The strength of this approach lies in this integrated combination, which often yields better results compared to a single complex model, making it a powerful tool for regression analysis in a variety of domains, from financial forecasting to scientific research.

Results And Discussions

This dataset captures electric vehicle (EV) charging sessions, describing the initial and final battery levels (%), charging duration (hours), and energy consumed (kWh). The patterns reveal key charging behaviours and vehicle characteristics. Typically, sessions start with a low or medium battery level and charge until a full or target state is reached. The data shows a direct relationship between energy consumed and the increase in battery percentage. However, performance varies; some sessions use significant energy for a small gain, which represents a potential loss of energy as heat, especially when the battery is almost full. Charging duration depends not only on the energy added, but also on the charging speed, which can vary between sessions. Fast charges with high energy transfer in a short time are visible (e.g., ~25 kWh in 2.86 hours), while slow, trickle charges also occur. Many entries show a final battery level of 100%, indicating that users are often charging to full capacity. Data is essential for analysing EV usage, planning charging infrastructure, and understanding battery performance and degradation over time.

TABLE 1. Descriptive Statistics

	Initial Battery Level (%)	Charging Duration (hrs)	Energy Consumed (kWh)	Final Battery Level (%)
count	100	100	100	100
mean	42.9125	2.4913	18.6549	64.8639
std	20.82396	1.171974	9.277037	22.18401
min	10.39	0.53	4.5	20.63
25%	23.525	1.47	10.5725	49.9
50%	42.49	2.52	16.925	63.1
75%	61.1125	3.5675	25.6025	83.0775
max	79.08	4.44	42.2	100

Based on 100 data samples, Table 1 provides descriptive statistics for key parameters including initial battery level (%), charging time (hours), energy consumed (kWh), and final battery level (%). The average initial battery level is 42.91%, while the final level is 64.86% on average, reflecting typical charging progress. The average charging time is 2.49 hours, and the average energy consumption is 18.65 kWh. The data shows moderate variability, as indicated by standard deviations ranging from 1.17 to 22.18. The values have wide ranges, suggesting a variety of charging conditions and battery levels across the dataset, which is suitable for robust model analysis.

Effect of Process Parameters

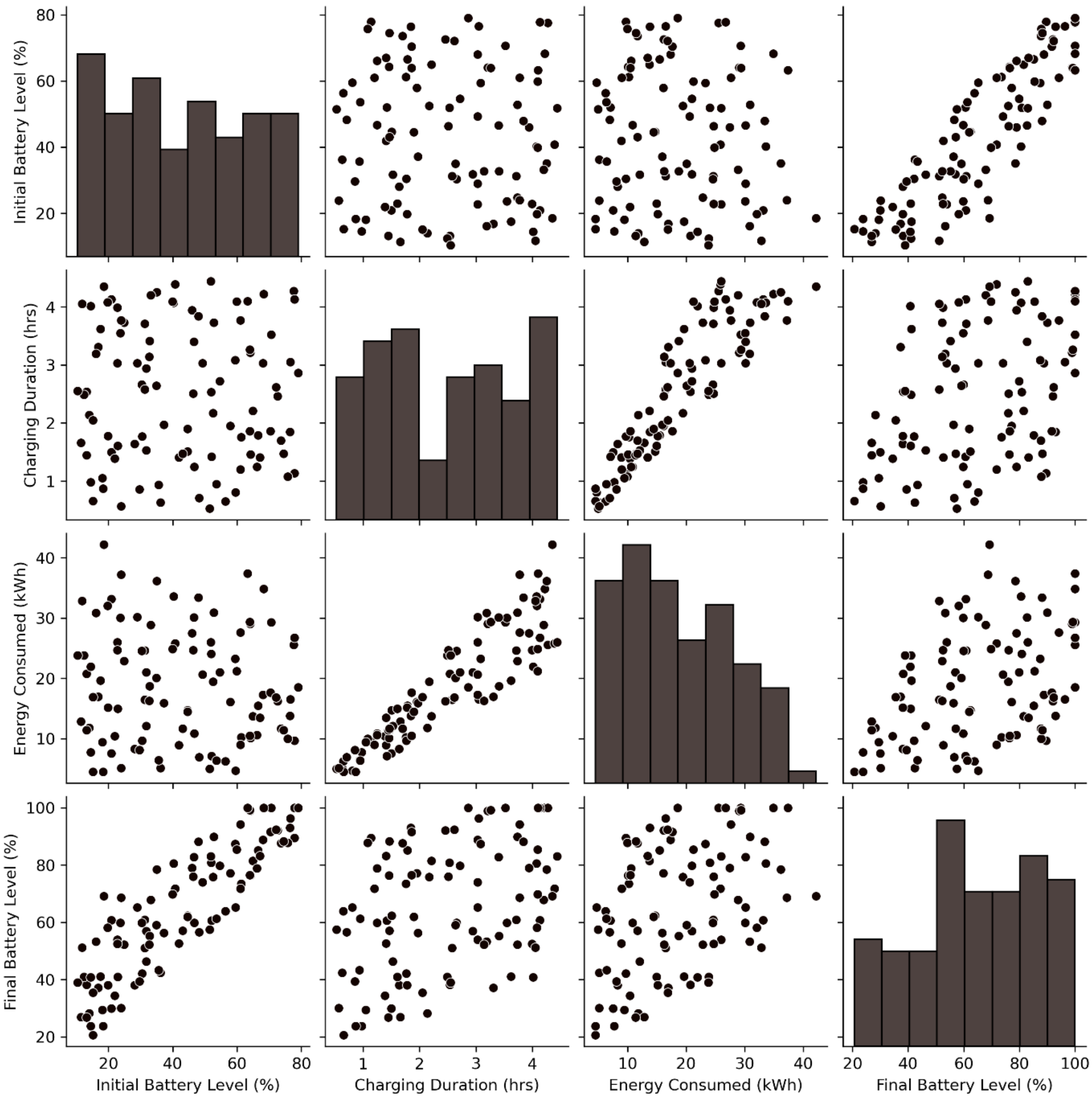

FIGURE 1. Scatter plot of the various Electric Vehicle Charging Patterns

Figure 1 illustrates scatter plots showing the relationships between electric vehicle charging parameters. A strong positive correlation is observed between charging time, energy consumed, and final battery state, indicating that longer charging times lead to higher energy consumption and increased battery sizes. The plots confirm consistent charging behaviour at different initial battery states.

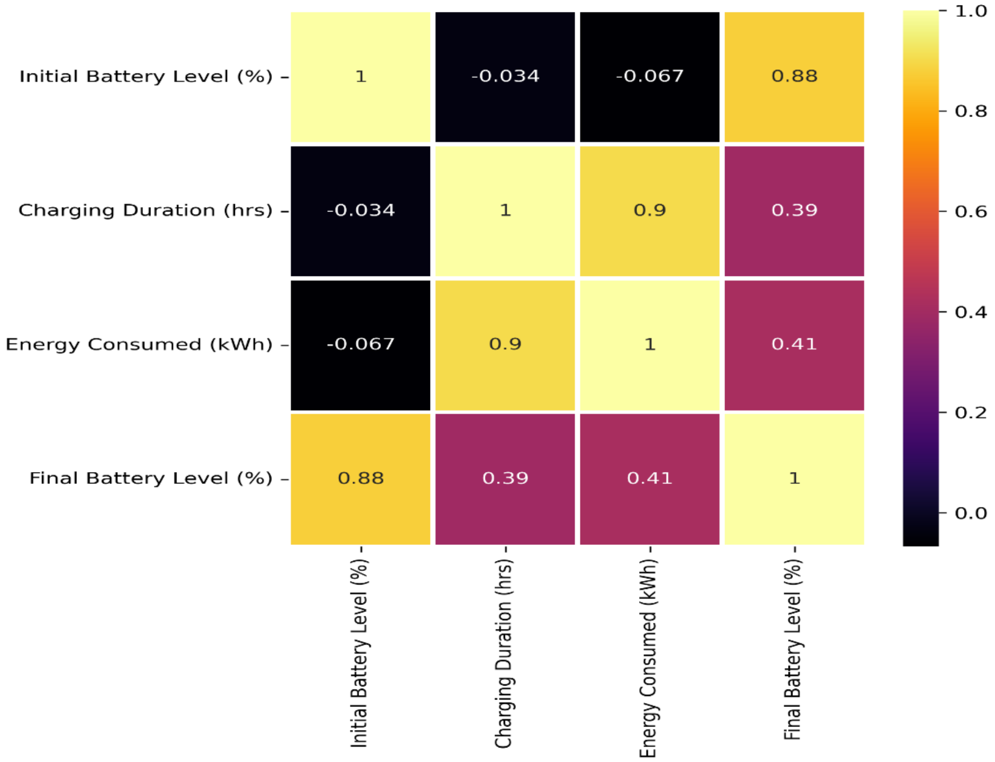

FIGURE 2. Heat map of the relationship between process parameters and outcomes

Figure 2 provides a heat map illustrating the correlations between electric vehicle charging parameters. A significant positive connection (0.9) exists between charging time and energy consumed, indicating proportional energy consumption over time. Similarly, the initial and final battery states show a high correlation (0.88), reflecting stable charging capacity and predictable battery performance trends.

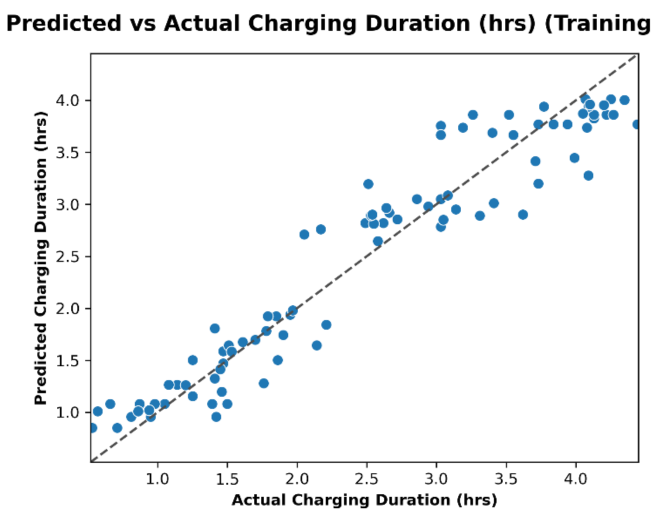

FIGURE 3. Hist Gradient Boosting Regression on Charging Duration (hrs): training data

Figure 3 illustrates the performance of the Hist Gradient Boosting Regression model in predicting charging duration (hours) utilizing training data. The scatter plot contrasts the actual and anticipated values, with points closely aligned on the diagonal reference line, showing strong prediction accuracy and minimal deviation over the charging duration range.

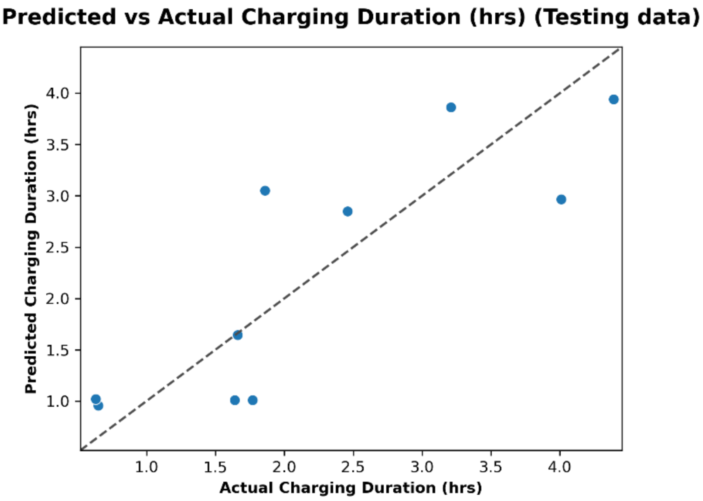

FIGURE 4. Hist Gradient Boosting Regression on Charging Duration (hrs): testing data

Figure 4 presents the performance of the Hist Gradient Boosting Regression model on the experimental data for charging duration (hours). The scatter plot compares the projected and actual values, revealing a generally stable trend along the diagonal line. Although small deviations occur, the model demonstrates reliable generalization and satisfactory prediction accuracy on unobserved data.

TABLE 2. Performance Metrics of Hist Gradient Boosting Regression on Charging Duration (hrs) (Training Data and Testing Data)

Parameter	Charging Duration (hrs)	
Data	Train	Test
Model	Hist Gradient Boosting Regression	Hist Gradient Boosting Regression
R2	0.91865	0.69785
EVS	0.91865	0.69785
MSE	0.10872	0.45097
RMSE	0.32972	0.67154
MAE	0.26248	0.58288
Max Error	0.81308	1.19005
MSLE	0.00966	0.04678
Med AE	0.22436	0.54027

Table 2 Performance metrics predicting charging duration (hours) reveal significant challenges for model interplay. The stark difference between the high training R^2 (0.919) and the significantly lower test R^2 (0.698) indicates poor generalization. This is further evidenced by the severe degradation in error metrics; for example, the test RMSE (0.672) is more than twice the train RMSE (0.330). Such a large performance drop indicates if the training data over fits the model and its learned patterns are not reliably transferred to new data. As a result, for practical use, the model's ability to interact with each other under various conditions and provide consistent, accurate charging time predictions is currently limited and requires refinement.

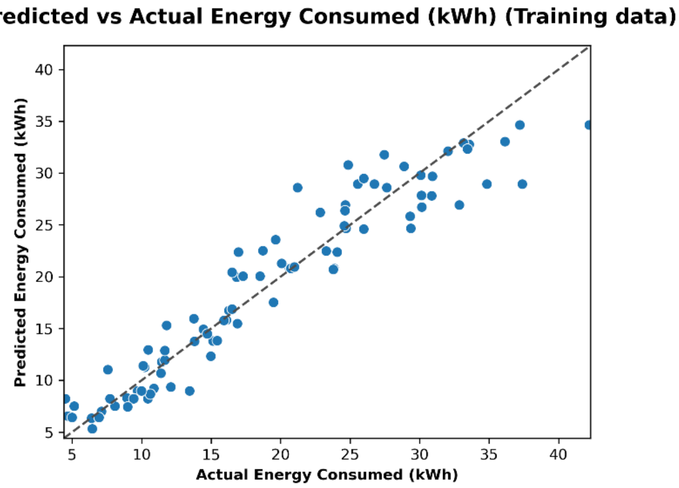

FIGURE 5. Hist Gradient Boosting Regression on Energy Consumed (kWh): training data

Figure 5 shows the performance of the Hist Gradient Boosting Regression model in predicting the energy consumed (kWh) using the data used for training. The scatterplot demonstrates a significant alignment between the actual and projected values along the diagonal reference line, indicating high model accuracy, stable learning behaviour, and minimal deviation over the entire energy consumption range.

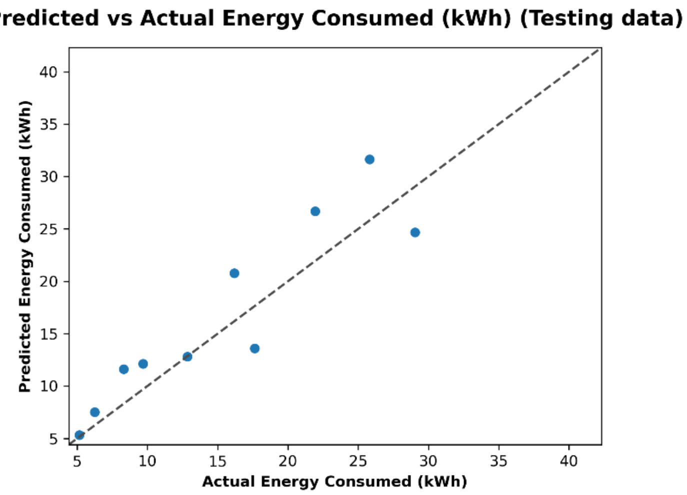

FIGURE 6. Hist Gradient Boosting Regression on Energy Consumed (kWh): testing data

Figure 6 illustrates the performance of the Hist Gradient Boosting Regression model on the test data for energy consumed (kWh). The scatterplot shows the predicted values following the general trend of the actual values, with moderate deviations. Overall, the model demonstrates acceptable generalization and maintains reasonable prediction accuracy in unobserved energy consumption patterns.

TABLE 3. Performance Metrics of Hist Gradient Boosting Regression on Energy Consumed (kWh) (Training Data and Testing Data)

Parameter	Energy Consumed (kWh)	
Data	Train	Test
Model	Hist Gradient Boosting Regression	Hist Gradient Boosting Regression
R2	0.90957	0.78802
EVS	0.90957	0.81963
MSE	7.81025	13.16567
RMSE	2.79468	3.62845
MAE	2.10014	3.08317
Max Error	8.43714	5.84027
MSLE	0.02488	0.03772
Med AE	1.62536	3.67238

Table 3 shows that the performance of the HGBR model is compromised by the significant performance drop between training (R^2=0.910) and test (R^2=0.788) for the energy consumption metrics. This difference indicates that the model's learned patterns do not fully transfer to the unobserved data, which may be due to over fitting or insufficient feature generalization. Key error metrics such as RMSE and MAE also increase significantly from train to test, further confirming this discrepancy. For reliable interoperability in a battery management system, the model must demonstrate highly consistent performance across datasets, ensuring that its predictions of energy consumption are reliable under varying, real-world operational conditions.

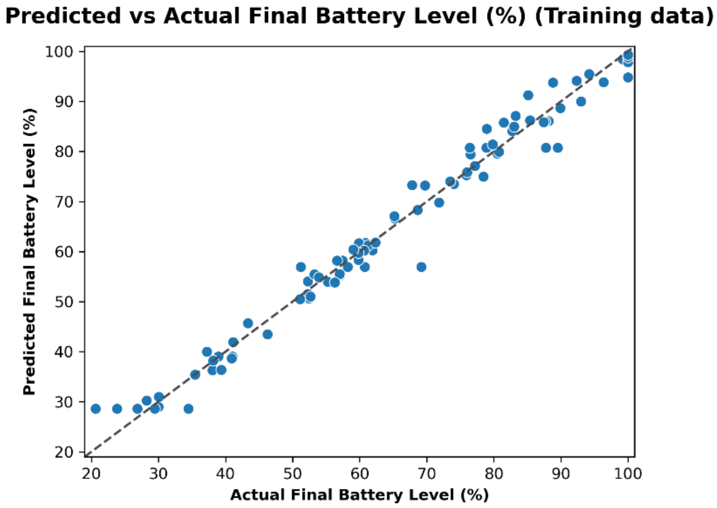

FIGURE 7. Hist Gradient Boosting Regression on Final Battery Level (%): training data

Figure 7 demonstrates the performance of the Hist Gradient Boosting Regression model in predicting the final battery level (%) using the training data. The scatterplot shows a high degree of agreement between the expected and actual values along the diagonal line, indicating better model learning, higher accuracy, and minimal error over the full range of final battery percentage levels.

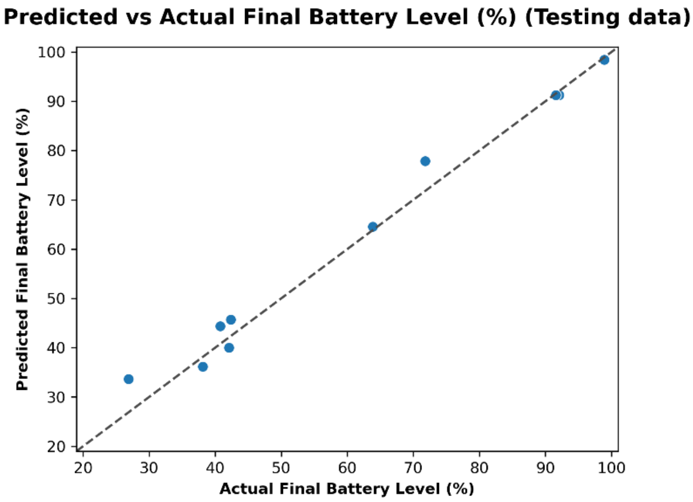

FIGURE 8. Hist Gradient Boosting Regression on Final Battery Level (%): testing data

Figure 8 presents the performance of the Hist Gradient Boosting Regression model on the test data to predict the final battery level (%). The scatterplot shows that the predicted values generally follow a diagonal trend, indicating good generalization. Despite small deviations, the model maintains reliable prediction ability at different final battery percentage levels.

TABLE 4. Performance Metrics of Hist Gradient Boosting Regression on Final Battery Level (%) (Training Data and Testing Data)

Parameter	Final Battery Level (%)	
Data	Train	Test
Model	Hist Gradient Boosting Regression	Hist Gradient Boosting Regression
R2	0.97945	0.98142
EVS	0.97945	0.98496
MSE	9.65414	11.63821
RMSE	3.10711	3.41148
MAE	2.26535	2.60109
Max Error	12.20579	6.79573
MSLE	0.00397	0.00717
Med AE	1.66798	2.02339

Based on the performance metrics in Table 4, the Hist Gradient Boosting Regression model demonstrates excellent inter-individual performance for predicting final battery level. This is demonstrated by the high and consistent R^2 scores across both the training (0.979) and test (0.981) datasets. Such parity shows that the model performs effectively when applied to fresh, untested data without significant over fitting. Key error metrics such as RMSE and MAE also align closely across datasets, demonstrating consistent predictive performance. This strong consistency is critical for a model to perform effectively in real-world applications, reliably providing accurate state-of-charge estimates for battery management systems across a variety of operational scenarios.

Conclusion

Based on a comprehensive examination of charging trends for electric vehicles using Hist Gradient Boosting Regression, this study demonstrates varying levels of predictive success across key charging parameters. The model achieved exceptional performance in predicting final battery capacity, with remarkably consistent R^2 values were 0.979 for training and 0.981 for test data, indicating strong generalization ability and reliable applicability to real-world battery management systems. This consistency, combined with low error metrics, confirms the model's readiness for practical use in state-of-charge assessment applications. However, predictions for charging duration and energy consumption presented significant challenges with significant performance degradation between the training and test phases. The charging duration model showed a significant drop from an R^2 of 0.919 to 0.698, while energy consumption predictions decreased from 0.910 to 0.788, indicating over fitting issues that limit operational reliability. These discrepancies highlight the complex, nonlinear nature of charging behaviours that are influenced by multiple factors, including charging infrastructure variability, user habits, and environmental conditions. The strong correlations identified between charging time and energy consumed (0.9), as well as initial and final battery levels (0.88), provide valuable insights for infrastructure planning and grid management optimization. Moving forward, addressing generalization limitations through improved feature engineering, the ability to include expanded datasets to capture diverse charging scenarios, and ensemble techniques can significantly improve model robustness. These findings give a major contribution to the development of intelligent charging systems, support the strategic expansion of EV infrastructure, and facilitate a broader transition toward sustainable electric transportation.

References

1. Chen, Zhong, Ziqi Zhang, Jiaqing Zhao, Bowen Wu, and Xueliang Huang. "An analysis of the charging characteristics of electric vehicles based on measured data and its application." IEEE access 6 (2018): 24475-24487.
2. Sadeghianpourhamami, Nasrin, Nazir Refa, Matthias Strobbe, and Chris Develder. "Quantitive analysis of electric vehicle flexibility: A data-driven approach." International Journal of Electrical Power & Energy Systems 95 (2018): 451-462.
3. Chen, C., Song, Y., Hu, X., & Guardiola, I. G. (2020). Analysis of Electric Vehicle Charging Behavior Patterns with Function Principal Component Analysis Approach. Journal of Advanced Transportation, 2020, 1–12. https://doi.org/10.1155/2020/8850654
4. Brinkel, N., Schram, W., AlSkaif, T., & van Sark, W. (2021). A Quantitative Analysis of the Short-Term and Structural Impact of COVID-19 Measures on Electric Vehicle Charging Patterns. https://doi.org/10.1109/SEST50973.2021.9543213
5. Azadfar, Elham, Victor Sreeram, and David Harries. "The investigation of the major factors influencing plug-in electric vehicle driving patterns and charging behaviour." Renewable and Sustainable Energy Reviews 42 (2015): 1065-1076.
6. Bashash, Saeid, Scott J. Moura, Joel C. Forman, and Hosam K. Fathy. "Plug-in hybrid electric vehicle charge pattern optimization for energy cost and battery longevity." Journal of power sources 196, no. 1 (2011): 541-549.
7. Chen, Chenxi, Yang Song, Xianbiao Hu, and Ivan G. Guardiola. "Analysis of electric vehicle charging behavior patterns with function principal component analysis approach." Journal of advanced transportation 2020, no. 1 (2020): 8850654.
8. Al-Ogaili, Ali Saadon, Tengku Juhana Tengku Hashim, Nur Azzammudin Rahmat, Agileswari K. Ramasamy, Marayati Binti Marsadek, Mohammad Faisal, and Mahammad A. Hannan. "Review on scheduling, clustering, and forecasting strategies for controlling electric vehicle charging: Challenges and recommendations." Ieee Access 7 (2019): 128353-128371.
9. Kelly, Jarod C., Jason S. MacDonald, and Gregory A. Keoleian. "Time-dependent plug-in hybrid electric vehicle charging based on national driving patterns and demographics." Applied Energy 94 (2012): 395-405.
10. Daina, Nicolò, and John W. Polak. "Hazard based modelling of electric vehicles charging patterns." In 2016 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific), pp. 479-484. IEEE, 2016.
11. Xiang, Yue, Zhuozhen Jiang, Chenghong Gu, Fei Teng, Xiangyu Wei, and Yang Wang. "Electric vehicle charging in smart grid: A spatial-temporal simulation method." Energy 189 (2019): 116221.
12. Limmer, Steffen. "Dynamic pricing for electric vehicle charging—a literature review." Energies 12, no. 18 (2019): 3574.
13. Schey, Stephen L., John G. Smart, and Don R. Scoffield. A first look at the impact of electric vehicle charging on the electric grid in the EV project. No. INL/CON-11-22696. Idaho National Lab.(INL), Idaho Falls, ID (United States), 2012.
14. Desai, Ranjit R., Roger B. Chen, and William Armington. "A pattern analysis of daily electric vehicle charging profiles: Operational efficiency and environmental impacts." Journal of Advanced Transportation 2018, no. 1 (2018): 6930932.
15. Dang, Qiyun. "Electric vehicle (EV) charging management and relieve impacts in grids." In 2018 9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG), pp. 1-5. IEEE, 2018.
16. Hill, Graeme, Phil T. Blythe, and Craig Higgins. "Deviations in Markov chain modeled electric vehicle charging patterns from real world data." In 2012 15th International IEEE Conference on Intelligent Transportation Systems, pp. 1072-1077. IEEE, 2012.
17. Cai, Hua, Xiaoping Jia, Anthony SF Chiu, Xiaojun Hu, and Ming Xu. "Siting public electric vehicle charging stations in Beijing using big-data informed travel patterns of the taxi fleet." Transportation Research Part D: Transport and Environment 33 (2014): 39-46.
18. Franke, Thomas, and Josef F. Krems. "Understanding charging behaviour of electric vehicle users." Transportation Research Part F: Traffic Psychology and Behaviour 21 (2013): 75-89.
19. Jiang, Zeming, Hao Tian, Mohammed J. Beshir, Surendra Vohra, and Ali Mazloomzadeh. Analysis of Electric Vehicle Charging Impact on the Electric Power Grid. No. DOE-USC-00192-118. City of Los Angeles Department, 2016.

20. Falchetta, Giacomo, and Michel Noussan. "Electric vehicle charging network in Europe: An accessibility and deployment trends analysis." Transportation Research Part D: Transport and Environment 94 (2021): 102813.
21. Collins, Michael, Robert E. Schapire, and Yoram Singer. "Logistic regression, AdaBoost and Bregman distances." Machine Learning 48, no. 1 (2002): 253-285.
22. Huang, Yantao, and Kara M. Kockelman. "Electric vehicle charging station locations: Elastic demand, station congestion, and network equilibrium." Transportation Research Part D: Transport and Environment 78 (2020): 102179.
23. Bryden, Thomas S., George Hilton, Andrew Cruden, and Tim Holton. "Electric vehicle fast charging station usage and power requirements." Energy 152 (2018): 322-332.
24. Ma, Hongwei, Xin Yang, Jianrong Mao, and Hao Zheng. "The energy efficiency prediction method based on gradient boosting regression tree." In
25. Ayaru, Lakshmana, Petros-Pavlos Ypsilantis, Abigail Nanapragasam, Ryan Chang-Ho Choi, Anish Thillanathan, Lee Min-Ho, and Giovanni Montana. "Prediction of outcome in acute lower gastrointestinal bleeding using gradient boosting." PLoS One 10, no. 7 (2015): e0132485.