Gaussian Process Regression for State of Charge Prediction: Overcoming Temperature Variability in Li-ion Battery Applications

Md Ismail Hossain, KhandokerMainul Islam, Tarifuzzaman Riyad, Md Rafsan Jany, Al Shahriar Zishan, Mohammad Nasir Uddin

Department of Electrical and Electronic Engineering, American International University-Bangladesh, Dhaka-1229, Bangladesh

Corresponding author: Al Shahriar Zishan, Email: 21-45125-2@student.aiub.edu

Accurately predicting the State of Charge (SOC) in lithium-ion batteries is essential for efficient energy management in electric vehicles. Due to the nonlinear characteristics of battery behavior, which depend heavily on temperature and operating conditions, SOC estimation presents a significant engineering challenge. This study proposes a data-driven approach using Gaussian Process Regression (GPR) to predict SOC, leveraging a comprehensive dataset with voltage, current, temperature, and average historical measurements. The model was trained and optimized through Bayesian hyperparameter tuning, and itsperformance was evaluated across four temperature conditions (-10°C, 0°C, 10°C, and 25°C) to assess robustness. The results demonstrate high prediction accuracy, with an overall Root Mean Square Error (RMSE) below 0.02 and Maximum Absolute Error (MAE) below 0.1 across all temperature settings. These metrics confirm the model's reliability and adaptability to varied thermal environments, highlighting the potential of GPR for practical SOC estimation in automotive applications. This approach provides a practical alternative to traditional electrochemical models, supporting advancements in battery health monitoring and energy management.

Keywords: Battery, SOC, Unscented Kalman Filter, Kalman Filter, Nonlinear state, Degrading system.

2025. In Sandeep Kumar & Kavita Sharma (eds.), *Computational Intelligence and Machine Learning*, 121–130. Computing & Intelligent Systems, SCRS, India. DOI: https://doi.org/10.56155/978-81-975670-5-6-10

1. Introduction

As the global demand grows for electric vehicles (EVs) and renewable energy storage systems, efficient battery management systems (BMS) are becoming increasingly crucial [1]. At the core of BMS lies the accurate estimation of the battery's State of Charge (SOC), a key parameter for optimizing performance, extending battery life, and avoiding operational failures [2]. Accurate SOC prediction is essential for managing energy usage and preventing battery issues like overcharging or deep discharging, which can accelerate degradation [3]. However, due to the nonlinear and dynamic behavior of batteries, influenced by factors such as temperature, aging, and operational conditions, SOC estimation presents significant challenges [4]. Traditional SOC estimation techniques, such as Coulomb Counting and Extended Kalman Filter (EKF) models, have shown effectiveness in controlled environments but often fail to maintain accuracy in real-world settings where conditions are highly variable [5]. These methods struggle with sensor noise, environmental fluctuations, and cumulative errors. To overcome these limitations, machine learning (ML) has emerged as a promising alternative. By leveraging large datasets, ML techniques such as neural networks, support vector machines, and deep learning models can capture complex battery behaviors, offering more adaptive and accurate SOC predictions [6]. Despite these advancements, several challenges remain, particularly in ensuring real-time predictions and computational efficiency in resource-constrained environments like EVs. Furthermore, the ability of existing ML models to generalize across different battery chemistries and operating conditions is still limited. There is also a critical need for models that can account for long-term battery degradation without requiring frequent retraining, making this an ongoing research focus.

The main objective of this research study is to establish a comprehensive machinelearning model for accurately predicting the State of Charge (SOC) of lithium-ion batteries. This model will represent the dynamic behavior of the battery system under various operating conditions, such as temperature fluctuations, aging, and load variations. For Bangladesh, this study will contribute to the development of efficient battery management systems in electric vehicles (EVs) and renewable energy storage setups, ensuring optimal performance and extended battery life. The proposed model aims to integrate seamlessly into existing energy infrastructures, enabling real-time SOC predictions for both domestic and commercial applications. Additionally, by improving battery efficiency and reducing energy waste, this research will help promote sustainable energy practices and reduce carbon emissions, making it a key contributor to the green energy transition in the region.

Neural networks effectively predict SOC under controlled conditions, but their real-time applicability across varying battery chemistries is limited. There is a need for more robust, generalizable models [7]. Support Vector Machines (SVM) provide high accuracy for SOC estimation in hybrid vehicles but struggle with variable operational conditions. Adaptive models are needed for dynamic environments [8]. Combining Extended Kalman Filters (EKF) with machine learning improves SOC estimation accuracy, but scalability across different battery systems is still an issue. Further optimization is required [9]. Particle filters address non-Gaussian errors in SOC prediction but have high computational costs, limiting their real-time usage. More computationally efficient solutions are needed [10]. Deep learning models effectively predict SOC and account for battery degradation, but high computational complexity limits their practical deployment. More efficient models are necessary [11]. Machine learning combined with physical models enhances SOC estimation, but adaptability to varying operational conditions remains a challenge. Further research is required to improve adaptability [12]. Recurrent Neural Networks (RNNs) show promise in SOC prediction for sequential data but require extensive training data, limiting generalization. There is a need for more flexible models [13]. Gaussian Process Regression provides uncertainty quantification for SOC prediction, but scalability and real-time application remain challenges. More scalable approaches are needed for dynamic systems [14].

2. Battery State of Charge Prediction

The dynamic equivalent circuit model of a lithium-ion battery, shown in Figure 1 plays an integral role in simulating the battery's behavior for SOC prediction. The model consists of an open-circuit voltage (V_{OC}) source, representing the battery's inherent voltage, alongside a series resistor (R_s) that accounts for the internal resistance, resulting in an immediate voltage drop during current flow. The parallel RC network captures the slower dynamic effects, such as voltage relaxation, due to electrochemical processes within the battery.

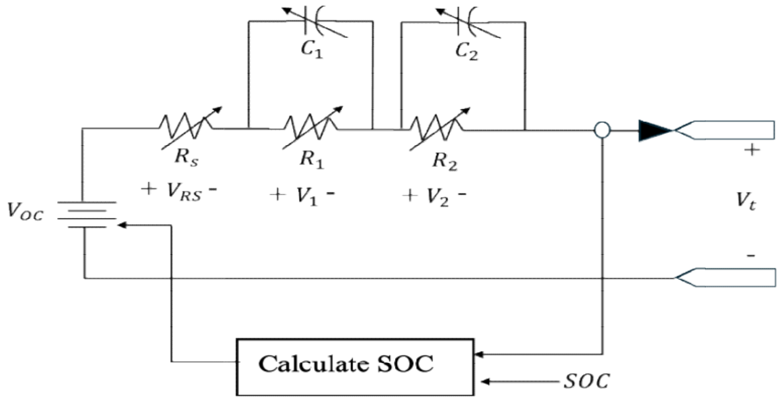

Figure 1. Dynamic equivalent circuit of Li-ion battery.

The combination of the V_{OC}, internal resistance, and RC network enables the model to predict the transient and steady-state behavior of the battery during charge and discharge cycles. The current flowing through the circuit is integrated over time to estimate the SOC, a key metric in life cycle prediction. By simulating both instantaneous and long-term battery dynamics, the model provides accurate insights into battery degradation, which, integrated with the machine learning framework (GPR) presented in this study, significantly enhances the precision of battery SOC predictions.

The SOC of an electric vehicle (EV) battery refers to the percentage of energy remaining in the battery relative to its total capacity. This metric is crucial for both the driver and the vehicle's energy management system, as it directly impacts the vehicle's range, performance, and safety. As shown in Figure 2, the power output of a battery is not constant across all SOC levels. At lower SOC, the maximum available power typically decreases, limiting the vehicle's acceleration and performance. Maintaining an optimal SOC range ensures efficient power delivery and a smooth driving experience. The accuracy of SOC estimation also helps drivers plan their trips confidently, reducing the likelihood of unexpected battery depletion.

SOC also plays a significant role in traction batteries long-term health and safety. Operating the battery at extremely low or high SOC levels can accelerate degradation, reducing lifespan. Overcharging or profoundly discharging the battery, especially in extreme temperatures, can cause irreversible damage to the battery cells and lead to capacity loss. To ensure longevity, keeping the battery within moderate SOC levels during operation is advisable. Furthermore, maintaining awareness of SOC is critical for safety, as running the battery at deficient levels can lead to complete discharge, potentially resulting in battery failure. By carefully managing SOC, drivers and vehicle systems can maximize both the performance and durability of the battery, ensuring safer and more reliable operation.

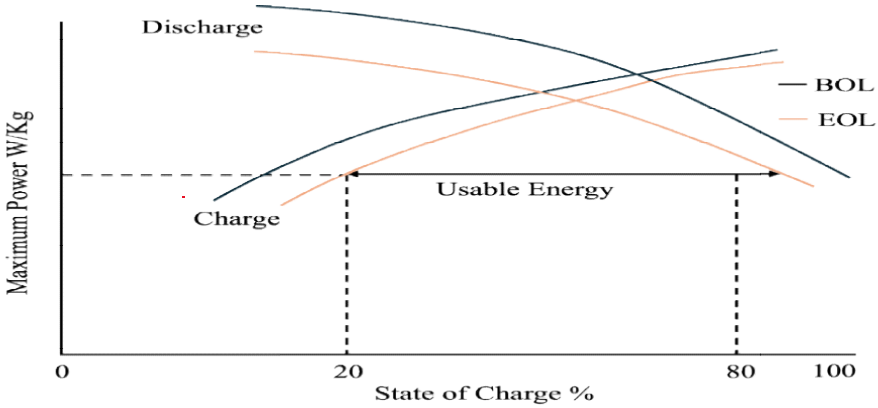

Figure 2. Lithium Battery Charge-Discharge Characteristics: Ideally, Lithium batteries are used between 20 – 80% of their capacity. If the limits exceed, the battery is prone to permanent damage.

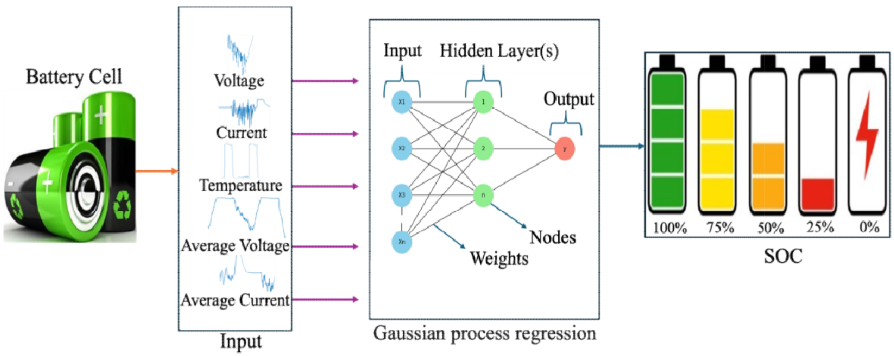

Figure 3. Proposed Framework for SOC Estimation Using Gaussian Process Regression.

Figure 3 shows the proposed framework for predicting a battery's SOC utilizes a GPR model to estimate SOC based on critical operational data such as current, voltage, temperature, and average current. This data is gathered from the battery, preprocessed, and as input of GPR model, where the model's internal computations use these variables to predict the SOC accurately. This approach harnesses the strength of GPR in handling nonlinear relationships and uncertainties, making it particularly effective for SOC estimation in electric vehicle batteries. By combining data acquisition, advanced machine learning-based prediction, and efficient output of SOC information, the framework provides a reliable solution for real-time monitoring, contributing to better battery management in terms of performance, safety, and longevity in electric vehicles.

3. Materials and Methods

3.1 Dataset Description

The dataset utilized in this study is a curated subset of the LG 18650HG2 Li-ion Battery Data [15]. This comprehensive dataset includes time-series data representing a Li-ion battery's behavior during multiple driving cycles conducted under four distinct ambient temperatures: -10°C, 0°C, 10°C, and 25°C. Each data point, sampled at 100-second intervals, provides key measurements, voltage, current, and temperature, alongside calculated moving averages for voltage and current over the preceding 500 seconds, which capture the battery's recent operational trends and mitigate transient fluctuations. The target variable, SOC, is normalized within a range of [0, 1], where 0 indicates a fully discharged state and 1 a fully charged state. The difference in the number of time steps between the training and test data is due to variations in data sampling frequency. The training data is collected more frequently to capture more granular battery dynamics. The test data is sampled at a lower frequency to reflect real-world operational conditions, where less frequent monitoring is typically used.

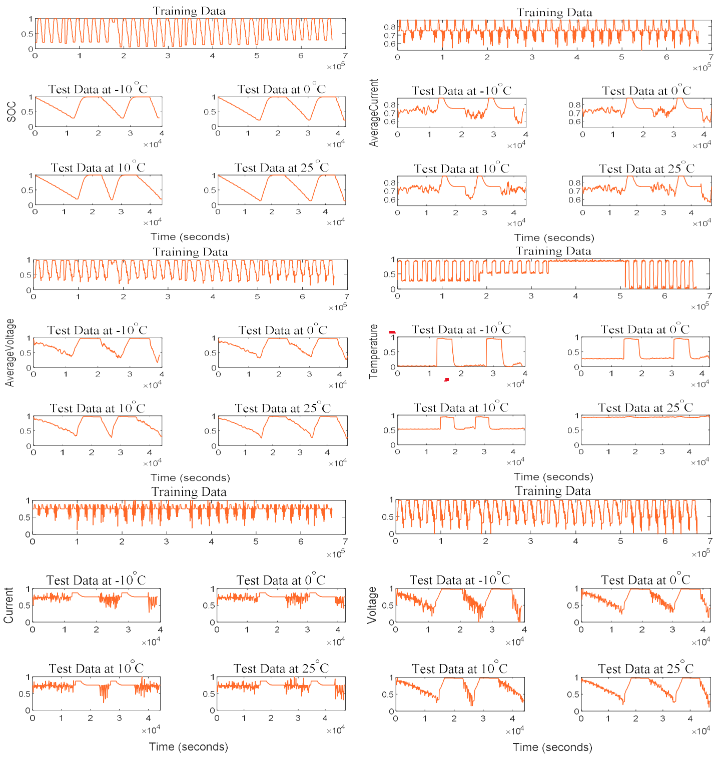

Figure 4. Training and Test data

3.2 Gaussian Process Regression

Modeling non-linear interactions in data is a specialty of Gaussian Process Regression (GPR), a non-parametric Bayesian regression approach. It offers a probabilistic framework that supports estimations of prediction uncertainty in addition to predictions themselves. GPR is an efficient way to capture the intricate link between the SOC and many input factors, including temperature, voltage, and current, for estimating SOC in batteries.

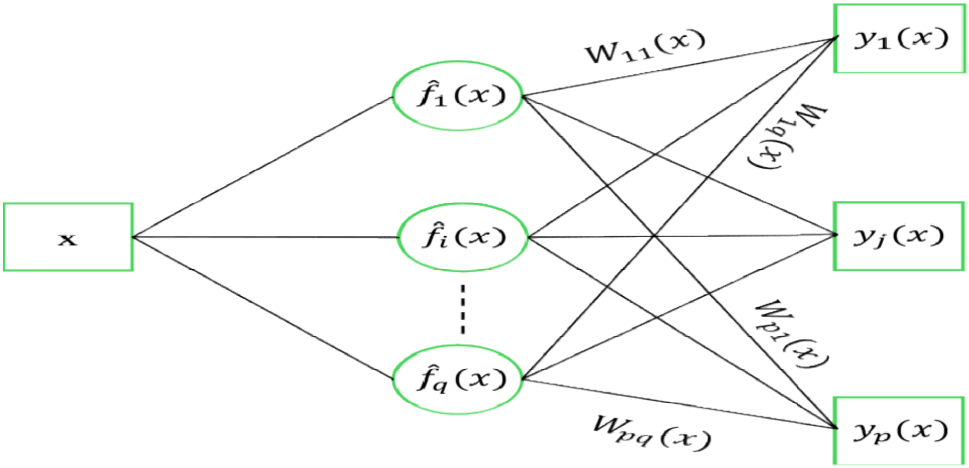

Figure 5. Gaussian process regression network structure.

The objective of GPR is to acquire the function f(x) that maps input characteristics "x" to output targets "y". A Gaussian process prior is applied over the function space, which is completely characterized by the covariance function (kernel) $k(x, x')$ and the mean function m(x):

$$f(x) \sim GP(m(x), k(x, x')) \quad (1)$$

The covariance function $k(x, x')$,often chosen as the Radial Basis Function (RBF) kernel, controls the smoothness and variability of the function, and is given by:

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{(x-x')^2}{2l^2}\right) \quad (2)$$

Whereσ_f^2 is the signal variance and l are the length scale. Given training data $\{x, y\}$, the GPR model predicts the SOC at a new test pointx_* by calculating the posterior mean μ_* and covarianceΣ_*, which are given by:

$$\mu_* = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} y \quad (3)$$

$$\Sigma_* = K(X_* X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*) \quad (4)$$

Here, $K(X, X)$ is the covariance matrix of the training inputs, $K(X_*, X)$ is the covariance between the test points and training pointsand σ_n^2 represents noise in the observations. The posterior means provides the predicted SOC, while the covariancegives the uncertainty in the prediction. When estimating SOC, the GPR model predicts the SOC using input factors such as temperature, voltage, and current while also giving estimates of the degree of uncertainty in each prediction. The accompanying diagram shows this procedure. The GPR framework learns the correlations between the parameters and the SOC from the battery data supplied to the model. The comparison graphic on the right illustrates how well GPR can predict complicated battery behavior by showing how well the anticipated SOC values match the actual values.

3.3 Training GPR Model

The GPR model was trained to predict the battery's SOC by optimizing key hyperparameters to capture the nonlinear relationships in the data. The model was configured to find the optimal settings for the Basis Function, Kernel Function, and Standardize options. The optimization process utilized Bayesian Optimization, which sequentially evaluated different hyperparameter configurations to minimize the error between the predicted and true SOC values. Fig.6 shows that the objective values quickly converged to a minimum, indicating efficient hyperparameter tuning. The plot displays both the minimum observed objective and estimated minimum objective over a series of 30 evaluations, with a sharp reduction in objective value occurring within the first few evaluations.

Parallel processing was employed to further enhance computational efficiency, allowing multiple configurations to be tested simultaneously. The final model configuration selected for GPR included an Exponential kernel function with no basis function, as this setup provided the lowest objective value, signifying optimal performance for this dataset. The convergence to a minimum objective value demonstrates the GPR model's capacity to learn the underlying patterns in the SOC data effectively, achieving a reliable predictive accuracy that reflects the nonlinear behaviors of SOC across varying battery conditions.

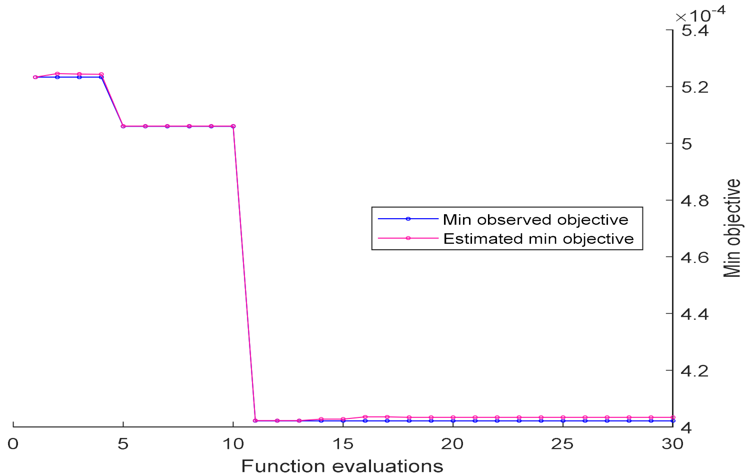

Figure 6. Min objective vs number of function evaluations.

The working flowchart (Fig.7) for predicting battery SOC using a GPR model begins with data acquisition, followed by preprocessing steps resampling, moving averages, and normalization. The data is then split into training and test sets across different temperatures to assess robustness. The GPR model is trained to capture nonlinear relationships, with Bayesian optimization tuning key hyperparameters for accuracy. After training, SOC predictions are generated and constrained within [0, 1]. Finally, performance is evaluated using metrics like RMSE and MAE to confirm prediction reliability under varied conditions.

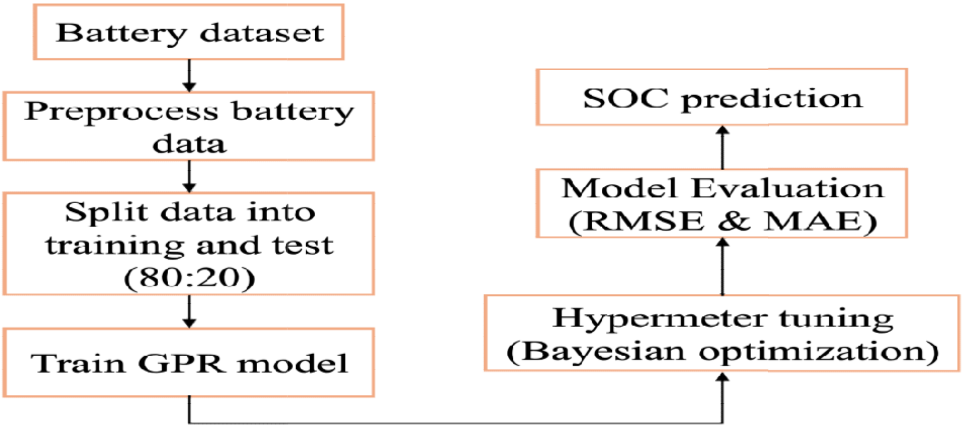

Figure 7. Working flowchart of the model.

4. Result and Discussion

The accuracy of the GPR model in predicting the State of Charge is assessed through two key metrics: Root Mean Squared Error (RMSE) and Maximum Absolute Error (MAE). These metrics quantify the difference between actual SOC values and predicted values, providing insight into the model's predictive performance. This performance, particularly across various ambient temperatures, has significant implications for the real-world application of the model in battery management systems. Equation 5 and 6 5 delineate the RMSE and MAE. Wherey_i represents the actual SOC values, $\hat{y}_i$ denotes the predicted SOC values, and n is the total number of data points. RMSE emphasizes more significant errors due to the squaring of differences, making it particularly sensitive to large deviations between predicted and actual values.

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{5}$$

$$\text{MAE}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i) \tag{6}$$

The RMSE and MAE metrics assess the performance of the GPR model in predicting the SOC across different ambient temperatures (-10°C, 0°C, ten °C, and 25°C), with lower values indicating higher predictive accuracy. RMSE and MAE values are reported for each temperature condition, demonstrating the effectiveness of the GPR model in accurately estimating SOC under varying operational environments.

Figure 8 shows the comparison between actual and predicted State of Charge (SOC) values using the Gaussian Process Regression model across different ambient temperatures (-10°C, 0°C, 10°C, and 25°C). The results show that the predicted SOC aligns closely with the actual SOC values, indicating accurate model performance on training and test data. For each temperature condition, the predicted SOC values effectively follow the accurate SOC patterns, demonstrating the model's robustness and reliability across various thermal environments. This close alignment across all tested temperatures suggests that the GPR model successfully captures the nonlinear dynamics of battery SOC with minimal deviation between predicted and actual values, highlighting its suitability for practical applications in battery health monitoring and energy management across diverse conditions.

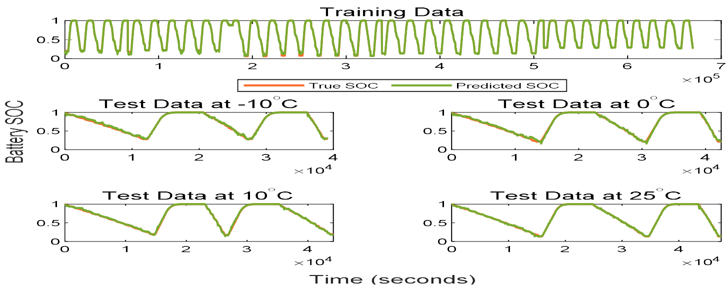

Figure 8. True SOC vs predicted SOC.

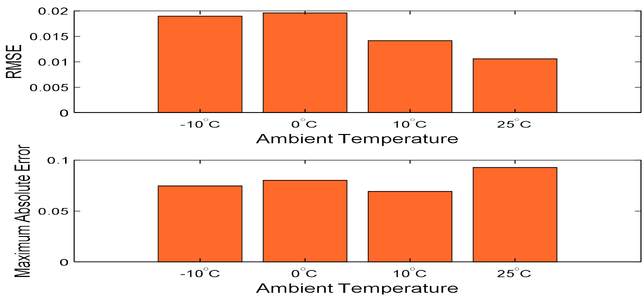

Figure 9. RMSE and MAE value.

Figure 9 presents the Root Mean Square Error (RMSE) and Maximum Absolute Error values of the Gaussian Process Regression (GPR) model's predictions for battery State of Charge (SOC) across different ambient temperatures (-10°C, 0°C, 10°C, and 25°C). Lower RMSE and Maximum Absolute Error values indicate more accurate SOC predictions, while higher values suggest reduced accuracy for the corresponding temperatures. The variations in error values across temperatures highlight the model's sensitivity to different thermal conditions. These results demonstrate the model's robustness and reliability in SOC estimation. Notably, the model shows improved predictive accuracy at certain temperatures, which is essential for real-world battery management systems applications.

The GPR model demonstrated high accuracy in predicting the SOC, as evidenced by the close alignment between actual and predicted SOC values across different ambient temperatures. This alignment reflects the model's capacity to capture the complex, nonlinear relationships inherent in battery behavior across varied thermal conditions. Additionally, the RMSE and MAE values remained consistently low, with overall RMSE below 0.02 and MAE below 0.1, underscoring the model's reliability in SOC estimation. These metrics affirm that the GPR model performs well regarding overall prediction accuracy and maintains robustness under different temperature settings. It is well-suited for practical applications in battery management systems where precise SOC monitoring is essential.

5. Conclusion

This study demonstrates the effectiveness of Gaussian Process Regression (GPR) in accurately predicting the State of Charge (SOC) for lithium-ion batteries in electric vehicles, addressing the challenges posed by temperature-dependent and nonlinear battery behavior. The model achieved an overall RMSE of less than 0.02 and an MAE of less than 0.1 across different ambient temperatures, indicating vital predictive accuracy and generalizability. Notably, the model performed consistently across temperature settings, with slightly lower errors observed at moderate and warmer conditions (10°C and 25°C), underscoring its robustness. Using Bayesian optimization for hyperparameter tuning enhanced model accuracy, providing an efficient, data-driven solution for SOC estimation. This approach offers a viable alternative to traditional methods requiring extensive battery-specific knowledge, making it well-suited for scalable deployment in automotive battery management systems. Future work could explore extending this approach to other battery chemistries and real-time SOC monitoring to further enhance the reliability and applicability of data-driven SOC estimation methods.

6. Acknowledgements

Mohammad Nasir Uddin greatly acknowledges the financial support of American International University-Bangladesh in the form of travel grant.

References

[1] Lipu, Molla Shahadat Hossain, Abdullah Al Mamun, Shaheer Ansari, Md Sazal Miah, Kamrul Hasan, Sheikh T. Meraj, Maher GM Abdolrasol et al. "Battery management, key technologies, methods, issues, and future trends of electric vehicles: A pathway toward achieving sustainable development goals." Batteries 8, no. 9 (2022): 119.
[2] Zhou, Long, Xin Lai, Bin Li, Yi Yao, Ming Yuan, Jiahui Weng, and Yuejiu Zheng. "State estimation models of lithium-ion batteries for battery management system: status, challenges, and future trends." Batteries 9, no. 2 (2023): 131.
[3] Park, Seongyun, Jeongho Ahn, Taewoo Kang, Sungbeak Park, Youngmi Kim, Inho Cho, and Jonghoon Kim. "Review of state-of-the-art state estimation technologies for battery management systems of stationary energy storage systems." Journal of Power Electronics 20 (2020): 1526-1540.
[4] Wang, Zuolu, Guojin Feng, Dong Zhen, Fengshou Gu, and Andrew Ball. "A review on online state of charge and state of health estimation for lithium-ion batteries in electric vehicles." Energy Reports 7 (2021): 5141-5161.
[5] Yang, Shichun, Sida Zhou, Yang Hua, Xinan Zhou, Xinhua Liu, Yuwei Pan, Heping Ling, and Billy Wu. "A parameter adaptive method for state of charge estimation of lithium-ion batteries with an improved extended Kalman filter." Scientific reports 11, no. 1 (2021): 5805.
[6] Rastegarpanah, Alireza, Mohammed Eesa Asif, and Rustam Stolkin. "Hybrid Neural Networks for Enhanced Predictions of Remaining Useful Life in Lithium-Ion Batteries." Batteries 10, no. 3 (2024): 106.
[7] Vennam, G., A. Sahoo, and S. Ahmed. "A survey on lithium-ion battery internal and external degradation modeling and state of health estimation." Journal of Energy Storage 52 (2022): 104720.
[8] Naresh, Vankamamidi S., Guduru VNSR Ratnakara Rao, and D. V. N. Prabhakar. "Predictive machine learning in optimizing the performance of electric vehicle batteries: Techniques, challenges, and solutions." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery (2024): e1539.
[9] Kaleem, M. Bin, Wei He, and Heng Li. "Machine learning driven digital twin model of Li-ion batteries in electric vehicles: a review." Artif. Intell. Auton. Syst 1, no. 1 (2023): 0003.
[10] Ouyang, Tiancheng, Chengchao Wang, Peihang Xu, Jinlu Ye, and Benlong Liu. "Prognostics and health management of lithium-ion batteries based on modeling techniques and Bayesian approaches: A review." Sustainable Energy Technologies and Assessments 55 (2023): 102915.
[11] Ng, Man-Fai, Jin Zhao, Qingyu Yan, Gareth J. Conduit, and Zhi Wei Seh. "Predicting the state of charge and health of batteries using data-driven machine learning." Nature Machine Intelligence 2, no. 3 (2020): 161-170.
[12] Meng, Huixing, Mengyao Geng, and Te Han. "Long short-term memory network with Bayesian optimization for health prognostics of lithium-ion batteries based on partial incremental capacity analysis." Reliability Engineering & System Safety 236 (2023): 109288.
[13] Mienye, IbomoiyeDomor, Theo G. Swart, and George Obaido. "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications." Information 15, no. 9 (2024): 517.
[14] Farid, Maor. "Data-driven method for real-time prediction and uncertainty quantification of fatigue failure under stochastic loading using artificial neural networks and Gaussian process regression." *International Journal of Fatigue* 155 (2022): 106415.
[15] Kollmeyer, Phillip, Carlos Vidal, Mina Naguib, and Michael Skells. "LG 18650HG2 Li-ion battery data and example deep neural network xEV SOC estimator script." *Mendeley Data* 3 (2020): 2020.