Utilising Machine Learning for Precise State-of-Charge Prediction in Li-ion Batteries

Avinash Khatri
School of Architecture, Computing and Engineering
University of East London
London, United Kingdom
avinashkhatrikc@gmail.com

Jaswinder Lota
School of Architecture, Computing and Engineering
University of East London
London, United Kingdom
j.lota@uel.ac.uk

Prasanna Nepal
Department of Electrical and Electronics Engineering
Nepal Engineering College
Bhaktapur, Nepal
nepalp189@gmail.com

Mohammad Hossein Amirhosseini
School of Architecture, Computing and Engineering
University of East London
London, United Kingdom
m.h.amirhosseini@uel.ac.uk

***Abstract*—As the popularity of electric propulsion using batteries rises alongside the demand for renewable energy, effective battery management and monitoring are crucial for sustainability and efficiency in electric vehicles (EVs). The battery monitoring system (BMS) utilizes IoT and sensor networks to assess crucial battery metrics such as remaining useful life (RUL), state-of-health (SoH), and state-of-charge (SoC), based on measurements of current status, temperature, and voltage. Machine learning (ML) and artificial intelligence (AI) are increasingly utilized to enhance BMS accuracy, addressing challenges like real-time data processing and the accuracy of estimations. This paper investigates the effectiveness of Linear Regression and Random Forest models in estimating SoC. During the hyperparameter tuning phase, the models were optimized using the Grid Search method, and their performance was evaluated at various temperatures: -10°C, 0°C, 10°C, and 25°C. The findings indicate that the models' effectiveness enhances as the temperature increases. Random Forest model demonstrated the best performance at 25°C with R^2 score of 0.99646, and the RMSE score of 0.000264. This paper not only contributes to advancing Li-ion battery monitoring system, but also empowers professionals in this field to harness machine learning capabilities effectively.**

Keywords—Li-ion batteries, battery management system, state-of-charge, machine learning

I. Introduction

The automobile industry is swiftly advancing with the growing popularity of EVs, which are quickly becoming a pivotal element of the global energy economy [1,2]. One of the primary challenges in designing and operating EVs is the efficient management of their batteries [3,4]. To enhance both functionality and safety in electric vehicles, Battery Management Systems (BMS) are utilized. These systems monitor battery conditions, offer protection, predict operational statuses, and optimize overall performance [1]. To achieve this, the BMS evaluates the RUL, SoH, and SoC by analysing measurements of current status, temperature, and battery voltage [2,5,6]. Nevertheless, accurately forecasting battery performance in real-world conditions, including fluctuating environments and aging, continues to be a significant challenge. Models developed for estimating battery parameters are typically calibrated under controlled conditions, which may not fully capture the real-time variations experienced in practical settings. Therefore, the reliability of methods such as employing a Kalman filter is constrained in dynamic real-world environments. Recently, the integration of AI and ML into BMS has provided promising solutions to address these limitations [5,7].

With the growing use of lithium batteries across various applications, the need for developing efficient and scalable recycling processes has become increasingly critical [22]. Conventional recycling techniques frequently encounter obstacles such as the effective recovery of valuable materials, economic viability, and greenhouse gas emissions [23]. Consequently, the use of ML in lithium battery recycling has garnered attention as a data-driven method to forecast recycling potential [24].

This study contributes to enhancing Li-ion battery monitoring systems by evaluating the effectiveness of specific machine learning models in predicting the SoC across varying temperatures in two different datasets.

II. Previous works

Li-ion batteries are favoured for their extended lifespan and high energy density, making them a popular choice for use in portable devices and EVs [25]. For batteries operating in highly dynamic environments, accurately estimating the battery state is essential. Of the different battery states that need to be estimated, the SoC is essential for preserving the battery's efficiency and safety [25].

The trend in battery management leverages AI/ML to analyze extensive data accumulated over a battery's operational life, combined with cloud computing for enhanced manufacturing and digital twins. For instance, convolutional neural networks (CNNs) have been utilized to predict the SoC from battery data [9]. CNN models and memory correlation techniques have also been used to enhance fault prediction for Li-ion batteries, leading to more reliable performance assessments [12]. Support Vector Machine (SVM) has also been employed to predict the SoC, outperforming traditional methods and thereby extending battery life [10]. Moreover, Deep Learning (DL) models have been used to optimize charge/discharge rates, adapting to varying driving conditions to boost performance and longevity [11]. AI can also be used to integrate data from multiple sensors, significantly improving the accuracy of the SoC predictions for batteries [13].

Additional machine learning approaches, including feed-forward neural networks [17], deep neural networks [18], long-short-term-memory (LSTM) [19], and Gaussian process regression frameworks [20], have been investigated for the SoC estimation.

While AI/ML enhances battery state prediction in BMS, challenges persist in real-time data processing, computational power, and refining the accuracy of estimated parameters. Moreover, these techniques typically require substantial real-world datasets to attain satisfactory generalization capabilities [21].

Linear Regression (LR) and Random Forest (RF) models are particularly effective for handling large sets of data variables, enhancing efficiency. The accuracy of these models could be further optimized through hyperparameter tuning, which identifies the best settings to minimize the loss function and improve prediction accuracy. Our literature review revealed that these models are not commonly used in battery management systems for predicting the SoC, and their efficiency has not been extensively studied. We also noted that many previous studies did not evaluate the models' performance across different temperatures to assess the impact of temperature variations.

In this research, we employ RF and LR models to predict the SoC in batteries and evaluate their performance across different temperatures. The prediction accuracy will be enhanced through hyperparameter tuning, using the Grid Search Cross-Validation (GSCV) method. For comparative analysis, the outcomes from the Random Forest model will be benchmarked against those from a LR model to select the best performing model.

III. Methodology

A. Data Collection

The research utilizes two specific datasets including the LG 18650HG2 Li-ion Dataset [14] and the dataset from the Tesla Model 3 2170 Li-ion cells [15].

Figure 1 depicts the data distribution in the LG 18650HG2 Li-ion battery dataset, showcasing metrics such as SoC, temperature, current status, voltage, average current status, and average voltage. Similarly, Figure 2 presents the corresponding data plots for the dataset from the Tesla Model 3 2170 Li-ion cells. In both figures, the x-axis indicates the number of data points.

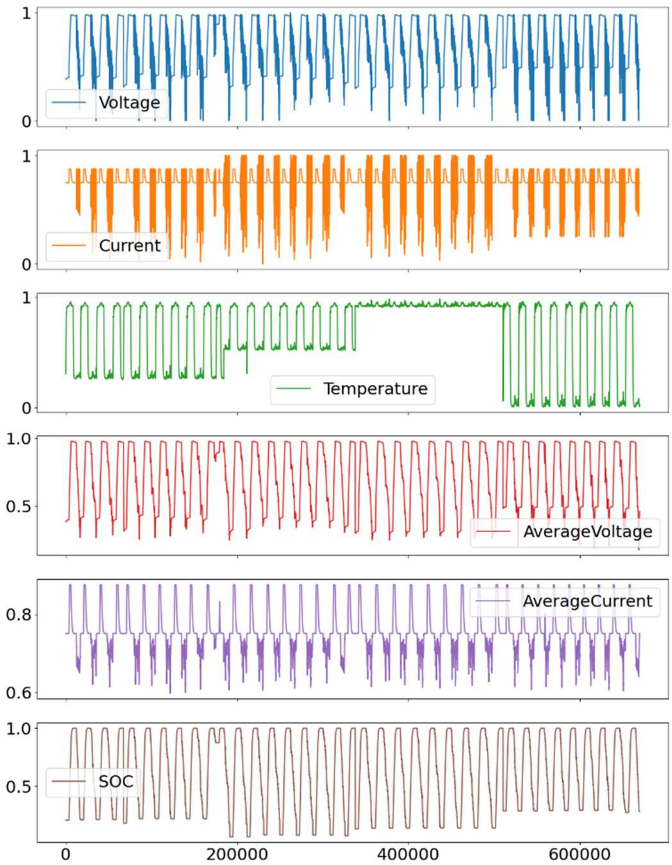

Fig. 1. Data distribution within the LG 18650HG2 Li-ion battery dataset

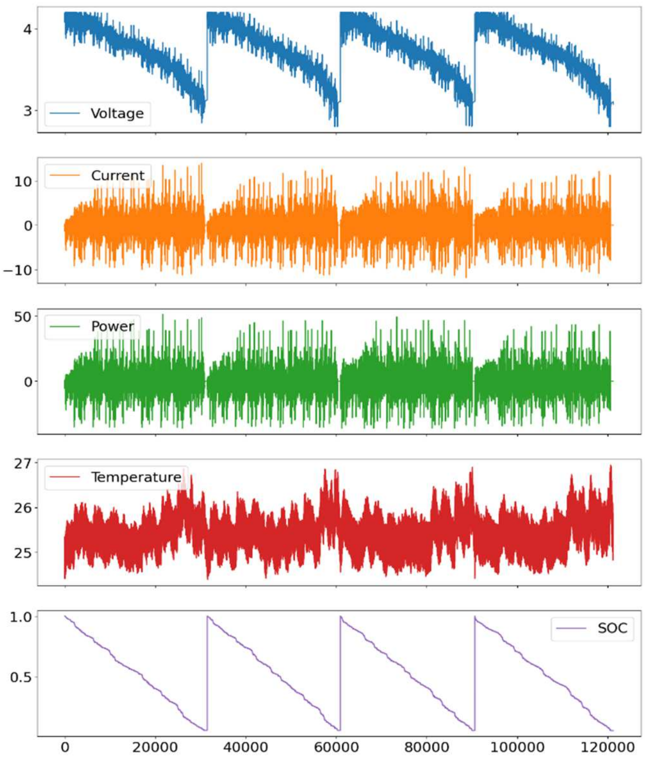

Fig. 2. Data distribution within the dataset from the Tesla Model 3 2170 Li-ion cells

B. Graphical User Interface

Randomly selected data from the datasets were transmitted via the ESP8266 microcontroller to a web server, where they were used to populate a MySQL database through phpMyAdmin. A graphical user interface was created using PHP to display real-time data from the database. This database was then used to predict the SoC. Figure 3 shows the developed user interface.

Fig. 3. Graphical User Interface to display data in real-time

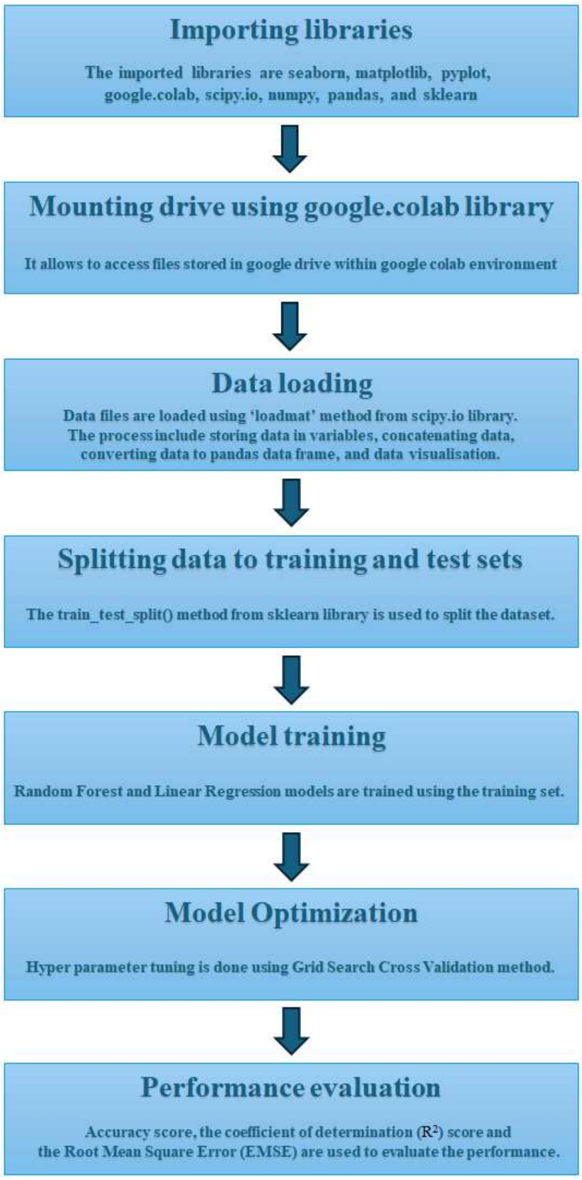

Fig. 4. Demonstration of the methodology for SoC prediction

C. Model Implementation

Figure 4 demonstrates our methodology for SoC prediction, which starts with importing necessary libraries for data analysis and visualization. The google drive is mounted to access the data files and the data is loaded using 'loadmat' function. The data is then concatenated and converted into a Pandas data frame and divided into training and testing datasets. These datasets are employed to train the ML models, and the results are evaluated using the root mean square error (RMSE) and the coefficient of determination (R^2) score.

D. Model Evaluation

As previously mentioned, the models' performance is assessed using the RMSE and R^2 score. The R^2 score is a crucial metric for evaluating the performance of ML models. It offers insight into how precisely the actual data align with the predictions. The R^2 score measures the proportion of the variance in the dependent variable that is predictable from the independent variables. A score of 1 means the model accounts for all the variance in the target variable, while a score of 0 indicates it accounts for none of it.

A higher R^2 score generally suggests that the model fits the data more effectively. This means that an R^2 score close to 1 signifies that the actual values match the predicted values, while A score near 0 signifies that the model is ineffective at capturing the underlying trend in the data. R^2 is also useful for comparing the performance of different models. A model with a higher R^2 score is generally considered to be better at predicting the target variable than one with a lower R^2 score.

Additionally, the RMSE measures the average size of prediction inaccuracies by taking the square root of the mean of the squared variances between the observed values and the forecasted values. Essentially, it measures the extent to which the predicted values diverge from the actual values. Since RMSE is expressed in the same units as the target variable, it is more easily interpretable than some other metrics. RMSE penalizes larger errors more than smaller ones due to the squaring of differences. This makes it sensitive to outliers, as larger errors have a disproportionately large impact on the RMSE value.

RMSE is useful for comparing model performance, with a lower RMSE indicating a better fit to the data. When evaluating models, the one with the lowest RMSE is typically regarded as the best performing in terms of prediction accuracy.

E. Data Integration and model Deployment

The ESP8266 microcontroller utilizes HTTP GET requests to transmit data to the web server, connecting to Wi-Fi through an ESP8266 Wi-Fi module for seamless data transfer. Sensor data is centrally managed in a MySQL database, simplifying the retrieval and analysis of historical information and facilitating integration with other data analysis systems.

IV. RESULTS AND DISCUSSION

A. SoC prediction for LG 18650HG2 Li-ion data

Table 1 shows the R^2 scores for the LG 18650HG2 Li-ion Cell dataset across various temperatures using LR and RF models. At -10°C, the LR model attained an R^2 score of 0.96605, while the RF model outperformed it with an R^2 score of 0.99472. At 0°C, the LR model's performance improved significantly, reaching an R^2 score of 0.99116, with the RF model attaining a similar R^2 score. At 10°C, the LR model achieved an R^2 score of 0.98917, and the RF model reached 0.99570. At 25°C, the LR model maintained strong accuracy with an R^2 score of 0.98557, while the RF model attained its highest accuracy with an R^2 score of 0.99895.

TABLE I. R^2 SCORE FOR 18650HG2 LI-ION DATA AT VARIOUS TEMPRATURES.

Model	Temperature			
	-10°C	0°C	10°C	25°C
LR	0.96605	0.99116	0.98917	0.98557
RF	0.99472	0.99407	0.99570	0.99895

The comparison reveals notable patterns. The RF model consistently outperformed the LR model across all temperatures, suggesting its superior ability to capture complex relationships within the dataset. Notably, all models showed improved accuracy as temperatures rose from -10°C to 25°C, indicating better predictive performance at higher temperatures.

In addition, TABLE 2 shows the RMSE scores for both models, revealing a clear trend of decreasing error with increased temperature. At -10°C, the LR model registered an RMSE of 0.002074, and the RF model demonstrated a lower RMSE of 0.000330. At 0°C, both models showed reduced RMSE values. The LR model recorded an RMSE of 0.000635, while the RF model achieved a lower RMSE of 0.000428. At 10°C, the RMSE values decreased further, with the LR model achieving an RMSE of 0.000816 and the RF model performing better with an RMSE of 0.000316. At 25°C, the RMSE values reached their lowest levels. The LR model recorded an RMSE of 0.0012147, while the RF model attained the lowest RMSE of 0.0000871.

TABLE II. RMSE SCORES FOR 18650HG2 LI-ION DATA AT VARIOUS TEMPERATURES.

Model	Temperature			
	-10°C	0°C	10°C	25°C
LR	0.002074	0.000635	0.000816	0.0012147
RF	0.000330	0.000428	0.000316	0.0000871

B. SoC prediction for Tesla 3 2170 Li-ion Cell data

Based on the experiments on 18650HG2 Li-ion data, we were convinced that the best performance will be achieved at higher temperatures. As a result, the R^2 scores for the Tesla Model 3 2170 Li-ion Cell data, were calculated for both LR and RF models at 25°C temperature and are presented in TABLE 3. The LR model achieved an R^2 score of 0.96681, indicating a reasonable fit and explaining approximately 96.68% of the variance in the dependent variable. In comparison, the RF model outperformed the LR model with an R^2 score of 0.99646, signifying a strong fit and accounting for approximately 99.64% of the variance.

TABLE III. R^2 SCORE FOR TESLA 3 2170 LI-ION CELL DATA

Model	25 °C
LR	0.96681
RF	0.99646

Additionally, Table 4 presents the RMSE scores for the LR and RF models at a temperature of 25°C. The LR model recorded an RMSE score of 0.002492, reflecting an average discrepancy of 0.2492% between predicted and actual values. The RF model achieved a slightly improved RMSE score of 0.000264. This corresponds to an average discrepancy of 0.0264%, representing a noticeable enhancement in predictive accuracy over the LR model.

TABLE IV. RMSE SCORE FOR TESLA 3 2170 LI-ION CELL DATA

Model	25 °C
LR	0.002492
RF	0.000264

V. CONCLUSION

This paper outlines a comprehensive framework for real-time data transmission, web-based monitoring, and precise SoC prediction using machine learning models. It describes how voltage, current state, and temperature data are generated and stored in a MySQL database to facilitate easy retrieval and analysis of data. Two specific datasets, the LG 18650HG2 Li-ion and the Tesla Model 3 2170 Li-ion Cell, were utilized for SoC prediction. The models employed—Linear Regression (LR) and Random Forest (RF)—were trained using these datasets.

In comparative analysis, the RF model consistently outperformed the LR model in terms of accuracy and predictive performance. It demonstrated superior results in both R^2 scores and RMSE values, achieving the highest accuracy and the smallest average discrepancy between actual values and predicted values.

Additionally, it was observed that the accuracy of all models improved with increases in temperature, suggesting more reliable predictions of battery behavior under elevated temperatures. Exploring the reasons for this temperature-related enhancement in model performance could be a fruitful direction for future research.

VI. FUTURE WORK

We plan to extend this research by implementing and evaluating additional machine learning models. In future work, we will also consider incorporating datasets related to different battery types to enhance the generalizability of our results.

REFERENCES

[1] H. Gabbar, A. Othman, and M. Abdussami, "Review of Battery Management Systems (BMS) Development and Industrial Standards," *Technologies*, vol. 9, no. 2, pp. 28, 2021.

[2] M. K. Tran and M. Fowler, "A review of lithium-ion battery fault diagnostic algorithms: Current progress and future challenges," *Algorithms*, vol. 13, no. 3, 2020.

[3] J. Cao, X. Chen, R. Qiu, and S. Hou, "Electric vehicle industry sustainable development with a stakeholder engagement system," *Technology in Society*, vol. 67, pp. 101771, 2021.

[4] W. Liu, T. Placke, and K. T. Chau, "Overview of batteries and battery management for electric vehicles," *Energy Reports*, vol. 8, pp. 4058–4084, 2022.

[5] M. S. H. Lipu *et al.*, "Battery Management, Key Technologies, Methods, Issues, and Future Trends of Electric Vehicles: A Pathway toward Achieving Sustainable Development Goals," *Batteries*, vol. 8, no. 9, pp. 119, 2022.

[6] R. Li, H. Wang, H. Dai, J. Hong, G. Tong, and X. Chen, "Accurate state of charge prediction for real-world battery systems using a novel dual-dropout-based neural network," *Energy*, vol. 250, pp. 123853, 2022.

[7] M. Radfar, A. Nakhlestani, H. Le Viet, and A. Desai, "Battery Management Technique to Reduce Standby Energy Consumption in Ultra-Low Power IoT and Sensory Applications," *IEEE Transactions on Circuits and Systems*, vol. 67, no. 1, pp. 336–345, 2020.

[8] J. X. Li, S. W. Sin, U. F. Chio, Y. J. Wu, C. S. Lam, and R. P. Martins, "Digital Battery Management Unit with Built-In Resistance Compensation, Modulated Frequency Detection and Multi-Mode Protection for Fast, Efficient and Safe Charging," *IEEE Transactions on Circuits and Systems*, vol. 67, no. 11, pp. 4063–4074, 2020.

[9] T. W. Sun and T. H. Tsai, "A Battery Management System Using Interleaved Pulse Charging with Charge and Temperature Balancing Based on NARX Network," *IEEE Transactions on Circuits and Systems,* vol. 69, no. 4, pp. 1811–1819, 2022.

[10] X. Fan, W. Zhang, C. Zhang, A. Chen, and F. An, "SOC estimation of Li-ion battery using convolutional neural network with U-Net architecture," *Energy*, vol. 256, pp. 124612, 2022.

[11] G.A. Rathy, P. Sivasankar, P. Karthikeyan, and K. Gunasekaran, "Estimation of SoC using SVM regression technique for an Efficient Electric Vehicle Battery Management System using c-RIO DAQ," *American Journal of Research*, pp. 89–103, 2020.

[12] S. Wang, S. Jin, D. Bai, Y. Fan, H. Shi, and C. Fernandez, "A critical review of improved deep learning methods for the remaining useful life prediction of lithium-ion batteries," *Energy Reports*, vol. 7, pp. 5562–5574, 2021.

[13] J. Sun, S. Ren, Y. Shang, X. Zhang, Y. Liu, and D. Wang, "A novel fault prediction method based on convolutional neural network and long short-term memory with correlation coefficient for lithium-ion battery," *Energy Storage*, vol. 62, no. January, pp. 106811, 2023.

[14] E. Galiounas, T. G. Tranter, R. E. Owen, J. B. Robinson, P. R. Shearing, and D. J. L. Brett, "Battery state-of-charge estimation using machine learning analysis of ultrasonic signatures," *Energy and AI*, vol. 10, no. July, pp. 100188, 2022.

[15] M. Kollmeyer, Philip; Vidal, Carlos; Naguib, Mina; Skells, "LG 18650HG2 Li-ion Battery Data and Example Deep Neural Network xEV SOC Estimator Script," *Mendeley Data*, no. v3, 2020.

[16] P. J. K. F. K. M. N. A. Emadi, "Tesla Model 3 2170 Li-ion Cell Dataset and Battery SOC Estimation Blind Modeling Tool," *Borealis*, no. V1, 2023.

[17] W. He, N. Williard, C. Chen, M. Pecht, "State of charge estimation for Li-ion batteries using neural network modeling and unscented Kalman filter-based error cancellation," *Electrical Power & Energy Systems*, 62, pp. 783-791, 2014.

[18] E. Chemali, P.J. Kollmeyer, M. Preindl, A. Emadi, "State-of-charge estimation of Li-ion batteries using deep neural networks: a machine learning approach," *Power Sources*, 400, pp. 242-255, 2018.

[19] E. Chemali, P.J. Kollmeyer, M. Preindl, R. Ahmed, A. Emadi, "Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries," *IEEE Transactions on Industrial Electronics*, 65, pp. 6730-6739, 2017.

[20] G.O. Sahinoglu, M. Pajovic, Z. Sahinoglu, Y. Wang, P.V. Orlik, T. W ada, "Battery state-of-charge estimation based on regular/recurrent Gaussian process regression," *IEEE Transactions on Industrial Electronics*, 65, pp. 4311-4321, 2017.

[21] F. Feng, X. Hu, J. Liu, X. Lin, B. Liu, "A review of equalization strategies for series battery packs: variables, objectives, and algorithms Renew." *Renewable and Sustainable Energy Reviews*, 116, pp. 109464, 2019.

[22] H. Helander, M. Ljunggren, "Battery as a service: analysing multiple reuse and recycling loops." *Resources, Conservation and Recycling*, Vol. 197, 107091, 2023.

[23] M. Pagliaro, F. Meneguzzo, "Lithium battery reusing and recycling: a circular economy insight." *Heliyon 5*, 01866, 2019.

[24] A. Valizadeh, M.H. Amirhosseini, Y. Ghorbani, "Predictive precision in battery recycling: unveiling lithium battery recycling potential through machine learning," *Computers & Chemical Engineering*, vol. 183, 108623, 2024.

[25] Fei Feng, Sangli Teng, Kailong Liu, Jiale Xie, Yi Xie, Bo Liu, Kexin Li, "Co-estimation of lithium-ion battery state of charge and state of temperature based on a hybrid electrochemical-thermal-neural-network model," *Power Sources*, vol. 455, pp. 227935, 2020.