SoH Estimation of ASPILSAN INR18650A28 Cells Using Real-World Factory Data

Nisanur YILDIRAN[1*] and Teoman KARADAĞ[1,2]

[1] Inonu University, Electrical Electronics Engineering, Malatya, TÜRKİYE
[2] Istinye University, Electrical Electronics Engineering, Istanbul, TÜRKİYE
nisanur.yildiran@inonu.edu.tr

Abstract. The estimation of the State of Health (SoH) of lithium-ion batteries is essential for ensuring the reliability and lifetime optimization of battery management systems (BMS). In this study, the long-term aging behavior of nickel-rich $LiNiMnCoO_2$ (NMC) INR18650A28 cells was analyzed using real-world production data from ASPILSAN Energy. Unlike previous works based on laboratory datasets, this study utilizes factory-origin, field-realistic data. Aging tests were performed with a WONIK PNE CC05-15 battery cycler at 23 °C under a constant current/constant voltage (CC/CV) protocol. The cells, with a nominal capacity of 2800 mAh, were charged at 0.5C (1.4 A) to 4.2 V and discharged at 1C (2.8 A) to 1.5 V for 1200 cycles. The raw voltage–capacity data were filtered and used for differential capacity (dC/dV) analysis, from which health-related electrochemical features were extracted as model inputs. These features and corresponding SoH values were employed to train regression-based models in MATLAB's Regression Learner. Linear Regression, Robust Linear Regression, and Bagged Trees achieved the highest prediction accuracy. The Linear Regression model performed best with RMSE = 0.325 and MAE = 0.250, followed by the Robust model (RMSE = 0.378, MAE = 0.269) and Bagged Trees (RMSE ≈ 6.52, MAE ≈ 1.99). Results demonstrate that dC/dV-based features effectively represent battery degradation, enabling accurate SoH prediction with regression algorithms.

Keywords: SoH Estimation, Machine Learning Algorithms, Regression, Lithium Ion Battery.

1 Introduction

Lithium-ion batteries (LIBs) are widely used in numerous application domains such as energy storage systems, electric vehicles, and portable electronic devices. The performance, reliability, and long-term usability of LIBs are largely evaluated based on their State of Health (SoH) value [1]. SoH represents the change in a battery's ability to retain capacity as its internal structure evolves during use. It can be defined with respect to capacity, internal resistance, power, or cycle count, and is regarded as a key indicator that quantitatively reflects the aging process occurring within the cell. Accurate

R. Rzayev et al. (eds.), *Proceedings of the International Conference on Current Problems in Engineering and Applied Sciences (ICCPEAS 2025)*, Advances in Engineering Research 299,
https://doi.org/10.2991/978-94-6239-668-5_43

estimation of SoH is essential both for Battery Management Systems (BMS) and for enabling users to make informed predictions about the system's operational status [2].

There exists an extensive body of literature on SoH estimation, covering various cell chemistries and diverse methodological approaches used for SoH diagnosis [1–9].

Zhang et al. [3] integrated features extracted through Incremental Capacity Analysis (ICA) into an Artificial Neural Network (ANN), proposing a new online approach capable of simultaneously estimating SoH and Remaining Useful Life (RUL). The method was validated using NASA datasets and achieved <3% error for SoH estimation, along with <4-cycle mean absolute error (MAE) and <6-cycle Root Mean Square Error (RMSE) for RUL prediction. Amir et al. [4] proposed a dynamic 2-RC equivalent circuit model that incorporates temperature effects. The accuracy of the model was verified using CALCE and NASA datasets, yielding RMSE values of 0.05–0.07 and 0.59–0.80, respectively. Bustos et al. [5] developed a new SoH estimation strategy by combining the sliding innovation filter (SIF) with the interacting multiple model (IMM) algorithm. Tests performed on the NASA B005 dataset showed that the proposed approach reduced the voltage RMSE from 0.0377 to 0.0257 V compared with the conventional Kalman Filter (KF)-IMM. Ni et al. [6] designed a hybrid method for SoH prediction and knee-point detection, integrating an Improved Newton–Raphson-based optimizer for support vector regression with an adaptive boosting algorithm (INRBO-SVR-AdaBoost), and combining it with a second-order polynomial method (SOHEM). This approach achieved SoH estimation performance below 0.89% RMSE and 0.75% MAE. Much of the current literature relies on publicly available datasets, most of which contain short-term cyclic aging tests that do not fully capture the diversity of real-world usage profiles.

In this study, these limitations are addressed by performing SoH estimation using field data obtained directly from the real production process of INR18650A28 cells manufactured by ASPILSAN Energy. The cells employ a nickel-rich $LiNiMnCoO_2$ (NMC) chemistry [7]. Aging tests were conducted using a WONIK PNE CC05-15 Battery Cycler at a constant temperature of 23 $°C$, and the performance evolution of the cells was monitored over 1000 cycles [8]. The collected test data were processed in Python using voltage–capacity information and subjected to a preprocessing stage. The cycle-based voltage–capacity data corresponding to the charging process were then transferred to MATLAB, where dC/dV analysis was performed. This analysis provided insights into the aging characteristics and electrochemical behavior of the cells. The resulting features were used as inputs to regression models, and SoH estimation was carried out through regression analysis. MATLAB's Regression Learner tool was employed for this purpose. All regression models were tested on the dataset, enabling an assessment of the effectiveness of both linear and nonlinear modeling approaches. The results indicated that the models achieving the lowest estimation errors were Linear Regression (LR), Robust Linear Regression (RLR), and Bagged Trees (BT), respectively.

The prominent aspect of this work is the use of real factory-level field data—rarely encountered in the literature—for SoH estimation, and the demonstration that high-accuracy prediction can still be achieved. The findings show that dC/dV-based features capture battery aging with high fidelity, and that even relatively simple regression algorithms can deliver strong SoH prediction performance when combined with appropriate data-processing steps.göstermektedir.

2 Battery Dataset

2.1. Definition of SoH

SoH is a crucial parameter that is crucial to determine in battery cells and battery systems. In the literature, it is defined as the percentage of the capacity that the battery can hold at a given moment relative to the capacity value of a fresh battery [10]. Its mathematical definition is given in Equation 1.

$$SoH = \frac{C_{current}}{C_{new}} \; x \; 100\% \tag{1}$$

2.2. Dataset

In this study, real battery aging data provided by ASPILSAN Energy were used. Aging tests were conducted using a WONIK PNE CC05-15 Battery Cycler to examine the cycle life and performance degradation of the cells. The cells tested were INR18650A28-type cells with a nickel-rich NMC chemistry, and an image of the cell is presented in Fig. 1.

All tests were carried out at a constant ambient temperature of 23 $°C$. The nominal capacity of the cells is 2800 mAh. Charging was performed using a CC/CV protocol.

First, the cells were charged to 4.2 V with a constant-current stage at 0.5C (1.4 A). Subsequently, the charging process was completed once the current dropped below 20 mA. A resting period of 10 minutes was applied after both the charging and discharging steps. During discharge, the cells were discharged at a constant current of 1C (2.8 A) down to a cutoff voltage of 1.5 V. This charge–rest–discharge sequence was defined as one cycle and repeated for 1000 cycles.

Information regarding the cell and the aging process is provided in Table 1.

Fig. 1. INR18650A28 battery cells

Table 1. Aging Process & Battery Cell Informations.

Model	Technical Specifications
Battery Chemistry	$LiNiMnCoO_2$ (NMC)
Nominal Capacity (mAh)	2800
Max. Charge Voltage (V)	4.2
Min. Discharge Voltage (V)	2.8
Charge-Rate (C)	0.5
Discharge-rate (C)	1
Ambient Temperature ℃	23

In Fig. 2, the capacity fade curve of the battery cell is presented.

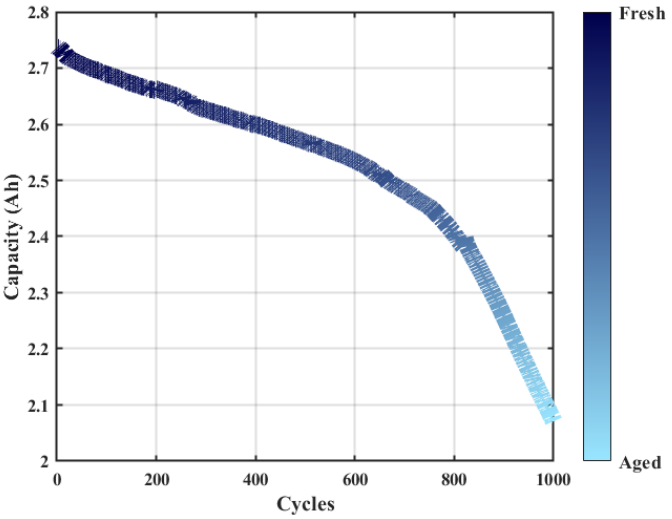

Fig. 2. Capacity fade characteristic of INR18650A28

As observed in Fig. 2, the initial capacity of the cell exhibits a decreasing trend as the cycle count increases. The cell reaches the commonly accepted usability threshold of 80% capacity at approximately 900 cycles. The aging behavior obtained from this cyclic degradation process is consistent with the expected characteristics of NMC chemistry and aligns well with the corresponding datasheet specifications.

Capacity Analysis

To accurately estimate the SoH of battery cells, various methods that provide comprehensive and reliable information about their electrochemical characteristics are commonly employed [11–13]. The method used in this study enables a detailed and in-depth analysis of the electrochemical processes occurring within the internal structure of the

cell. By generating the dC/dV curve and examining its peak and valley points, the aging behavior of the battery can be evaluated. Changes over time in the amount of capacity variation per unit voltage, as well as the reduction in the area under the curve with increasing cycle count—indicating a decrease in the retainable capacity—allow meaningful inferences to be drawn. The mathematical expression of this approach is given in Equation 2.

$$\frac{dC}{dV} \approx \frac{\Delta C}{\Delta V} \tag{2}$$

The specifications of the system used for the analyses are presented in Table 2.

Table 2. Device Configuration

Item	Specifications
Operating System	Windows 64-bit
CPU	Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz
GPU	NVIDIA GeForce GTX 1050 Ti
Memory (RAM)	16 GB

The three-dimensional dC/dV–Voltage–Cycles graph obtained from the dC/dV analysis performed in Matlab 2023b is presented in Fig. 3.

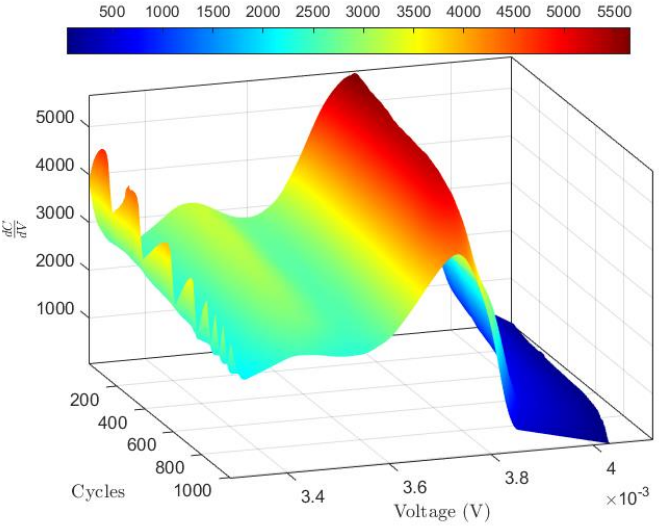

Fig. 3. 3D ~ dC/dV characteristic of INR18650A28

As observed in Fig. 3, the reactions arising from lithium-ion transport and mobility at the initial voltage region lead to the formation of a small peak in the dC/dV curve within the 3.4–3.6 V interval. In the 3.6–3.9 V range, the main peak of the dC/dV curve appears, corresponding to the electrochemical reactions occurring between the anode and cathode within the cell structure. The magnitudes of both peaks decrease as the

cycle count increases, indicating an inverse relationship between the number of cycles and the amount of capacity the cell is able to retain.

In Fig. 4, the dC/dV–Voltage curves corresponding to cycles 50, 100, 150 … 1000 are presented. It is clearly observed that both the magnitude of the peak and its position decrease over cycling. When the region around the peak is examined more closely in Fig. 4, it becomes evident that the aging behavior follows the expected pattern and direction.

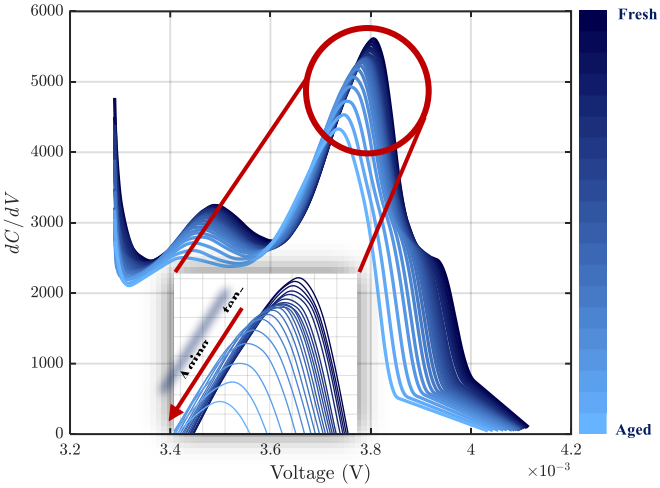

Fig. 4. 2D ~ dC/dV – Voltage Graph

When the aging tendency in Fig. 4 is examined, it is observed that the overall structure of the dC/dV–Voltage curves changes progressively with increasing cycle number. This evolution clearly reflects the degradation processes occurring within the battery cell.

The sharp and high peak observed in the fresh state corresponds to the dominant intercalation plateau and indicates rapid Li^+ insertion and extraction, with no significant kinetic limitations in charge-transfer behavior. As the number of cycles increases, the Solid Electrolyte Interphase (SEI) layer thickens, and the amount of Li^+ available to interact with the electrode decreases. These phenomena lead to a noticeable reduction in the peak magnitude of the dC/dV curve. In addition, the peak positions shift toward lower voltages, which is attributed to increased internal resistance causing the electrochemical reaction to occur under higher polarization.

Overall, the decrease in peak height, the shift of the corresponding voltage toward lower values, and the narrowing of the curve demonstrate that degradation in the cell is occurring not only in a kinetic sense but also in a thermodynamic sense. These observations constitute the characteristic electrochemical signature of SoH deterioration.

Data from the ~3.4 and ~3.8V voltage range, which provides information about the main aging characteristics for the analysis to be performed in the regression learner, has been extracted. 471 observation data points were used for each model.

Linear Regression

LR is one of the oldest and most fundamental methods in the field of statistical modeling [14]. The underlying assumption is that the relationship between the dependent variable y and the independent variables x_i can be expressed as a linear function. The model can be formulated as presented in Equation 3.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \tag{3}$$

Here, β_0 represents the intercept of the model, while $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients corresponding to the independent variables that influence the SoH. ε represents the error term and is typically assumed to follow a normal distribution with zero mean.

Owing to its simplicity, interpretability, and fast implementation, this method is widely used as a reference point in many regression analyses.

Robust Linear Regression

RLR is a variant developed to reduce the sensitivity of ordinary least squares (OLS) to outliers. The concept was introduced by Peter Huber in the 1960s, laying the foundations of robust statistics [15]. The main idea is to replace the squared error loss with a robust loss function that limits the influence of large residuals.

The model is expressed in the following general form:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum \sum_{i}^{n} \rho(y_i - x_i^T \beta) \tag{4}$$

Here, $\rho(\cdot)$ is a robust loss function that applies weighting based on the magnitude of the residuals.

The term $e_i = y_i - x_i^T \beta$ denotes the error, and β corresponds to the regression coefficients. For example, the Huber loss is defined as follows:

$$\rho_\delta(u) = \begin{cases} \dfrac{1}{2}u^2, & |\, u \,| \leq \delta, \\ \delta(|\, u \,| - \dfrac{1}{2}\delta), & |\, u \,| > \delta \end{cases} \tag{5}$$

This loss function behaves quadratically for small errors but transitions to a linear form for large errors, thereby preventing extreme outliers from disproportionately affecting the model coefficients. In general, loss functions used in RLR behave similarly to ordinary regression for small residuals, while for large residuals they reduce their influence on the model by assigning lower weights. This structure prevents outliers from exerting a disproportionate impact on the regression coefficients.

Bagged Trees (Bootstrap Aggregated Trees)

The Bagged Trees method is based on the principle of "bootstrap aggregation," introduced by Leo Breiman [16]. It aims to reduce variance and improve prediction stability by combining multiple decision trees (DT) trained on different bootstrap samples of the dataset.

The method begins by generating B new samples from the original dataset using the bootstrap technique. An independent decision tree $f_b(x)$ is trained on each subset.

The final prediction is then obtained by averaging the outputs of all trees:

$$\hat{f}(x) = \frac{1}{B}\sum_{b=1}^{B} f_b\,(x) \tag{6}$$

This averaging process reduces the high variance associated with a single decision tree and makes the predictions more stable.

Here, $\hat{f}(x)$ represents the final prediction. B is the total number of DT models, $f_b(x)$ denotes the prediction of the $b-th$ tree.

Bagging is particularly effective in reducing the tendency of models to overfit in noisy datasets and represents one of the earliest examples of ensemble learning.

3 Results and Discussion

In this section, the analysis results and prediction plots of the three models that exhibited the best estimation performance are presented in order.

3.1 Linear Regression Model

In the Regression Learner interface, the "Linear Regression" model was trained using all 471 features, with Principal Component Analysis (PCA) disabled. On the validation set, the model achieved RMSE = 0.32526, MAE = 0.25044, and $R^2 = 1.00$. The training time was measured as 2.04 s. The resulting prediction model has a compact structure of approximately 5 MB.

Fig. 5 presents the prediction plot of the LR model alongside the actual observation data. The near-complete overlap between the observation curve and the model's predicted curve demonstrates that the LR model is capable of estimating SoH with high accuracy.

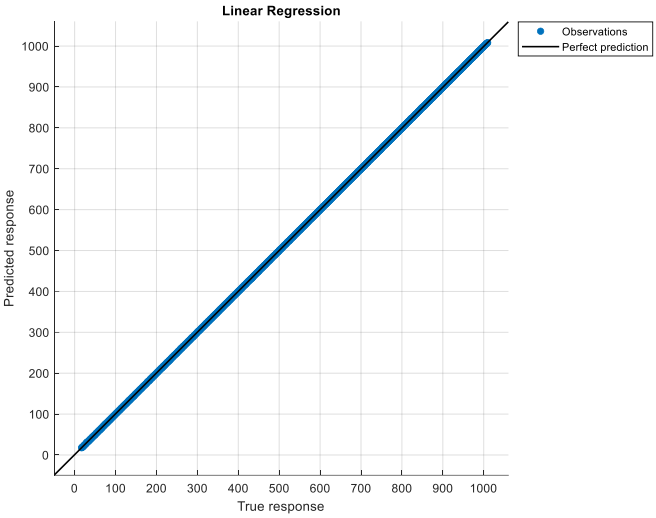

Fig. 5. LR- Estimation Graph

3.2 Robust Linear Model

The RLR model was trained with the "Robust" option enabled. Using all 471 features, the validation error metrics were observed as RMSE = 0.3779, MAE = 0.26907, and $R^2 = 1.00$. The training process took 7.75 s, and the resulting model size was approximately 5 MB. This model produced more stable and balanced results in datasets containing outliers. Fig. 6 shows that the RLR model aligns well with the corresponding dataset in estimating SoH and exhibits strong prediction performance with low error.

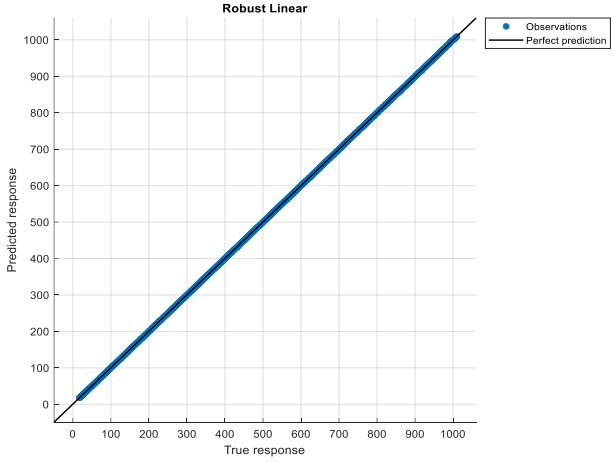

Fig. 6. RLR- Estimation Graph

3.3 Bagged Trees

The Bagged Trees algorithm, which ranked as the third-best performing model, was constructed using 30 individual decision trees (learners). The minimum leaf size for each tree was set to 8. All 471 features were used, and PCA was disabled in this configuration.

The validation results were RMSE = 6.5205, MAE = 1.9959, and $R^2 = 1.00$, and the training time was recorded as 9.80 s. Although this model has a higher computational cost compared to the other methods, it provides a more balanced overall performance by reducing prediction variance. The prediction graph is presented in Fig. 7.

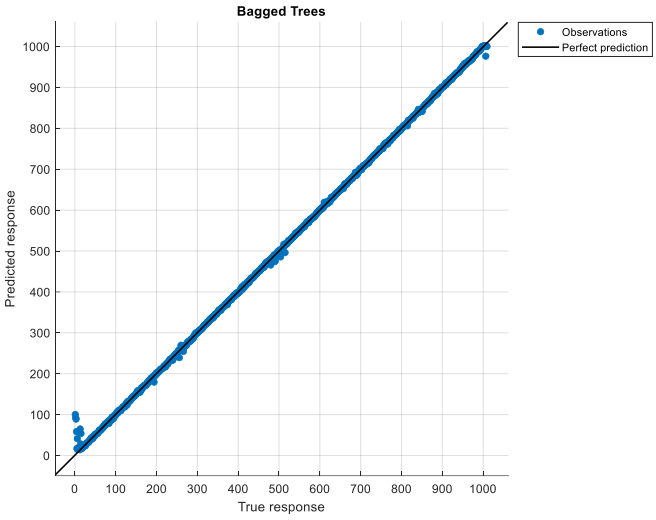

Fig. 7. BT- Estimation Graph

Table 3presents the values of four different error metrics, along with the prediction times and model sizes, for the three estimation models.

Table 3. Estimation Model Results

Model	RMSE	R^2	MSE	MAE	Prediction Speed	Training Time (s)	Model Size
LR	0.32526	1.00	0.10579	0.25044	~9600 obs/sec	2.0399	~5 MB
RL	0.3779	1.00	0.14281	0.26907	~5800 obs/sec	7.7457	~5 MB
BT	6.5205	1.00	42.517	1.9959	~4700 obs/sec	9.8026	~3 MB

4 Conclusion

In this study, State of Health (SoH) estimation was performed using real production and long-term aging data from INR18650A28 NMC cells manufactured by ASPILSAN Energy.

Unlike open-access standard datasets, the analysis performed on these 1000 cycle data points obtained from the factory and reflecting field conditions revealed that the electrochemical properties obtained from dC/dV curves represent cell aging with high accuracy. Among the three different models trained in the Regression Learner environment, LR showed the most successful performance in SoH estimation with values of RMSE = 0.325 and MAE = 0.250. While the RLR model exhibited a more balanced structure in regions containing outliers, the BT model lagged behind other methods in terms of accuracy despite its high variance reduction capability. The results obtained demonstrate that high accuracy in SoH estimation can be achieved by combining physically meaningful features and appropriate regression algorithms without using complex deep learning structures.

Furthermore, the use of real world-factory data has strengthened the model's adaptability to industrial applications and made the study unique and innovative. Future work aims to increase the generalizability of the models by analyzing larger cell groups that include different temperature conditions and testing optimization-based prediction algorithms.

Acknowledgments. The authors would like to thank ASPILSAN ENERGY INC. for providing the battery test data used in this study.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Teng, J.H., Chen, R.J., Lee, P.T., Hsu, C.W.: Accurate and Efficient SOH Estimation for Retired Batteries. Energies (Basel). 16, (2023). https://doi.org/10.3390/en16031240.
2. Yu, Z., Liu, N., Zhang, Y., Qi, L., Li, R.: Battery SOH Prediction Based on Multi-Dimensional Health Indicators. BATTERIES-BASEL. 9, (2023). https://doi.org/10.3390/batteries9020080.
3. Seol, S., Lee, J., Yoon, J., Kim, B.: Improving SOH estimation for lithium-ion batteries using TimeGAN. MACHINE LEARNING-SCIENCE AND TECHNOLOGY. 4, (2023). https://doi.org/10.1088/2632-2153/acfd08.
4. Iurilli, P., Brivio, C., Carrillo, R.E., Wood, V.: Physics-Based SoH Estimation for Li-Ion Cells. Batteries. 8, (2022). https://doi.org/10.3390/batteries8110204.
5. Dikmen, İ.C., Yildiran, N., Karadag, T.: Machine Learning Approaches for Enhancing the SoH Estimation of LTO Batteries. International Journal of Automotive Science And Technology. 9, 48–59 (2024). https://doi.org/10.30939/IJASTECH..1522403.

6. Dikmen, İ.C., Yıldıran, N., Karadağ, T.: Efficient State of Health Estimation for LTO Batteries Using Liquid Time-Constant Neural Networks. In: 3rd INTERNATIONAL CONGRESS of ELECTRICAL and COMPUTER ENGINEERING.
7. Yildiran, N., Dikmen, I.C., Karadag, T.: State of Health Estimation of Lithium Titanate Oxide Batteries Through Data-Driven Techniques and Machine Learning. 8th International Artificial Intelligence and Data Processing Symposium, IDAP 2024. (2024). https://doi.org/10.1109/IDAP64064.2024.10711165.
8. Dikmen, I.C., Yıldıran, N., Karadağ, T.: Multi-Chemistry Battery Management System for Electric Vehicles. The European Journal of Research and Development. (2022).
9. Yıldıran, N.: ELEKTRİKLİ ARAÇLARDA KULLANILAN FARKLI KİMYALARA SAHİP LİTYUM TABANLI PİLLERİN SAĞLIK DURUMLARININ ANALİZİ VE MODELLENMESİ, https://tez.yok.gov.tr/UlusalTezMerkezi/giris.jsp, (2025).
10. Su, L., Xu, Y., Dong, Z.: State-of-health estimation of lithium-ion batteries: A comprehensive literature review from cell to pack levels. Energy Conversion and Economics. 5, 224–242 (2024). https://doi.org/10.1049/ENC2.12125.
11. Lee, C.-J., Kim, B.-K., Kwon, M.-K., Nam Kanghyun and Kang, S.-W.: Real-Time Prediction of Capacity Fade and Remaining Useful Life of Lithium-Ion Batteries Based on Charge/Discharge Characteristics. Electronics (Basel). 10, (2021). https://doi.org/10.3390/electronics10070846.
12. Vennam, G., Sahoo, A., Ahmed, S.: A survey on lithium-ion battery internal and external degradation modeling and state of health estimation. J. Energy Storage. 52, 104720 (2022). https://doi.org/https://doi.org/10.1016/j.est.2022.104720.
13. Ye, M., Wei, M., Wang, Q., Lian, G., Ma, Y.: State of health estimation for lithium-ion batteries based on incremental capacity analysis under slight overcharge voltage. Front. Energy Res. 10, (2022). https://doi.org/10.3389/fenrg.2022.1001505.
14. Vilsen, S.B., Stroe, D.I.: Battery state-of-health modelling by multiple linear regression. J. Clean. Prod. 290, (2021). https://doi.org/10.1016/j.jclepro.2020.125700.
15. Huber, P.J.: Robust Estimation of a Location Parameter. https://doi.org/10.1214/aoms/1177703732. 35, 73–101 (1964). https://doi.org/10.1214/AOMS/1177703732.
16. Breiman, L.: Bagging predictors. Mach. Learn. 24, 123–140 (1996). https://doi.org/10.1007/BF00058655/METRICS.