38th International Electric Vehicle Symposium and Exhibition (EVS38) Göteborg, Sweden, June 15-18, 2025

A data-driven approach to map the aging of dismantled commercial high-energy NMC cells

Md Sazzad Hosen[1], Amir Farbod Samadi[1], Kashif Raza[1], Maitane Berecibar[1]

[1]*Md Sazzad Hosen (corresponding author), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium, md.sazzad.hosen@vub.be*

Executive Summary

The second-life application of vehicle batteries is getting attention as millions of battery systems, modules, or cells are going to enter the market in the coming decade. The performance uncertainty with or without the historical knowledge of the batteries' vehicle usage is a concern. In this work, such a 2nd life battery pack is dismantled, and the Nickel-Manganese-Cobalt (NMC) 141 Ah cells are extensively investigated to understand the 2nd life degradation behavior. The one and a half year long test campaign followed a suitable stationary test matrix, generating a valuable dataset. The aging dataset is then used to train different machine learning algorithms, resulting in a root-mean-square-error (RMSE) of 0.65 with ElasticNet while validating against a stationary profile.

Keywords: battery aging; second life battery; state of health; machine learning; real-life profile

1 Introduction

The State of Health (SoH) estimation of NMC battery cells is a crucial aspect of assessing the performance and longevity of batteries, particularly in the 2nd applications [1]. The millions of cars sold in the past decades are going to expire in the coming years, and the numbers will only increase over time [2]. These batteries from on-road vehicles should be either repurposed or recycled, requiring an established strategy. The vehicle batteries can be easily reused in a different application where the market is also growing, and the service demand is less demanding than the on-road usage [3]. However, there are numerous challenges of acquiring, analyzing, sorting, remanufacturing, and monitoring these batteries, just to name some out of many. On top of that, technical challenges such as a lack of first-life usage information, data log, potential mechanical impact from packaging, etc., increase the uncertainty of the batteries in the 2nd application. Within this research work, 2nd life battery cells dismantled from a commercial vehicle pack are studied extensively in terms of aging, developing an accurate estimation model to understand the SoH [4]. The 2nd life cells of NMC 141 Ah prismatic types are produced by a prominent manufacturer. The cells follow a detailed pre-check before being investigated for a long-term life campaign.

In the field of SoH modeling, there are many estimation modeling methodologies available in the literature, which can be divided into empirical [5], electrochemical [6], and data-driven techniques [7]. The latter has gained attention in recent years due to the development of numerous algorithms, especially machine learning algorithms that are frequently trained with success on the battery time-series dataset. Besides, no prior knowledge of battery history and a complex invasive characterization process make the machine learning algorithms more suitable for aging prediction work. In this research, after evaluating different trained algorithms' performance on lab- and real-scale profiles, ElasticNet is reported to be the best performer. In the following part of the paper, the test protocols are described in Section 2, after which the results are analyzed in Section 3. The model development work, including feature engineering and validation results, is reported in Section 4. In Section 5, the conclusion is outlined along with the future work.

2 Experimental setup

For the test campaign, the 2nd life cells were characterized by generating a high-quality dataset with a range of characterization and cycle life tests. The test campaign ran for close to 18 months. A total of 16 cells are studied as part of 8 operating conditions. Fig. 1 displays the physical cell under test, which is a prismatic NMC/C Li-ion battery with 141 Ah nominal capacity, 3.7V nominal voltage, 2.14kg weight and a volume of 0.94m^3.
The 2nd life aging test campaign started with a physical pre-check and preconditioning of the cells before making the final setup. The aging conditions are preselected according to the scenarios presented in Table 1. The test procedure includes the beginning or end of life (BoL/EoL) standard characterization and then cycling as per defined conditions with a frequency of 100 full equivalent cycles to track the SoH until the actual capacity drops below 60% of the nominal [8]. A full equivalent cycle is defined as one full charge event equal to the nominal capacity value.

Figure 1: Sample pouch (left) and prismatic (right) cells setup for aging tests.

Table 1: Test conditions for the 2nd life test campaign.

Condition nr.	Depth of cycling (DoD)	SoC Range	Temperature	C-rate (charge/discharge)	Nr. of cells
1	80%	10%-90%	25	1	2
2	70%	20%-90%	25	1	2
3	60%	20%-80%	25	1	2
4	40%	30%-70%	25	1	2
5	80%	10%-90%	35	1	2
6	80%	10%-90%	15	1	2
7	80%	10%-90%	25	0.33 (50A)	2
8	60%	20%-80%	25	0.33 (50A)	2
Total cells					16

As part of the characterization procedure, which has been performed at BoL, EoL, and intermediate frequency of 100 FEC includes standard capacity test, hybrid pulse power characterization (HPPC), and open circuit voltage (OCV) tests. The capacity test is designed with three full-discharge cycles at a C/3 rate, where C refers to the nominal capacity of the cell. The HPPC test is performed to calculate internal resistance (IR) at different SoC percentages (at 80%, 50%, 20%) with a 10s 0.33C pulse. The slow quasi-OCV test is done with a C/20 charge-discharge cycle at room temperature. Hence, a so-called dynamic test, referring to a real-life day-long profile optimized for the cell, is performed with two cells to understand the battery characteristics for non-static cycling. The number of FEC-scale is calculated as equivalent to 15 FEC for each dynamic cycling test in this case, before doing the characterization.

3 Results

The dismantled battery cells showed almost similar SoH compared to a fresh battery, meaning that they were used at a minimum, for which there is no information available. The battery capacity fades, internal resistance check, and OCV shift are tracked during the aging campaign via characterization tests. Fig. 2 displays the capacity fade of the studied cells, where the high DoD and temperature are found to be more prominent, impacting a sharp decay in the capacity fade [9]. Similar results are found in internal resistance growth and OCV curves as displayed in Fig. 3.

The cells have completed a maximum of 1600 FECs by the end of the campaign, with a maximum of almost 15% SoH drop, while the maximum internal resistance growth of 26%. The EoL limit was set to 60% SoH and/or 100% IR growth; however, no cell reached that limit. The SoH is calculated as the actual measured capacity value against the BoL value in percentiles, while IR calculation follows a similar approach, but only to cover the growth compared to the BoL IR value. Note that figures 2 and 3 refer to Table 1 conditions, where two consecutive numbered cells are linked to consecutive conditions.

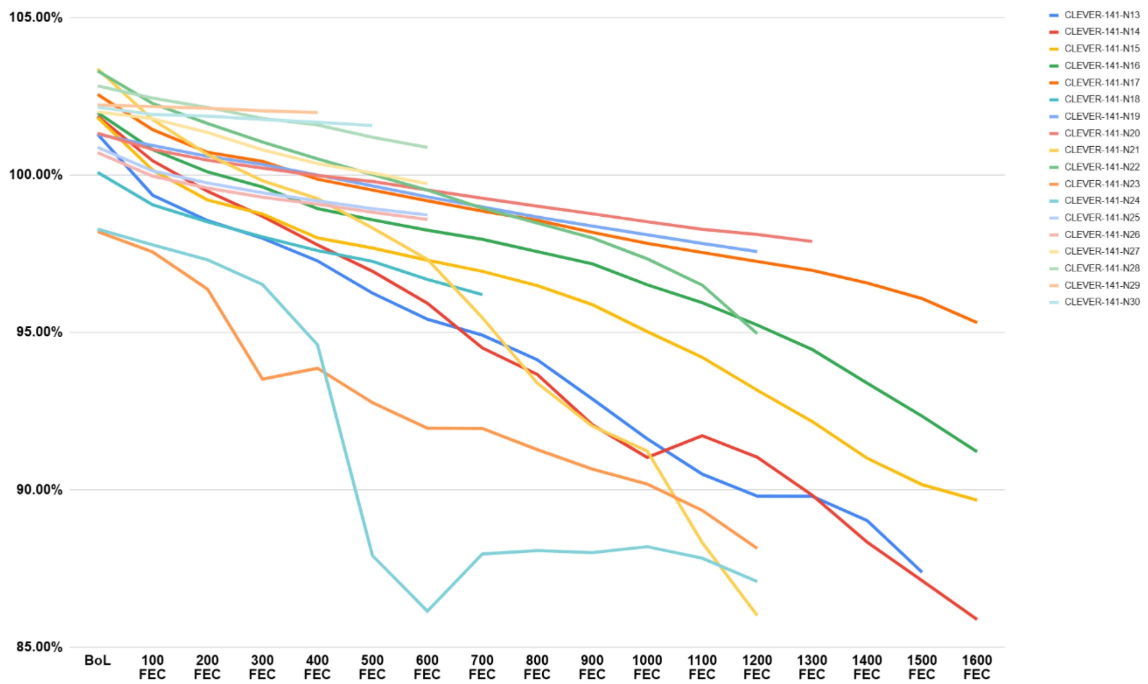

Figure 2: Capacity fade results of 18 NMC 141 Ah cells, where the vertical axis represents the SoH percentile, and the horizontal axis represents the performed FECs.

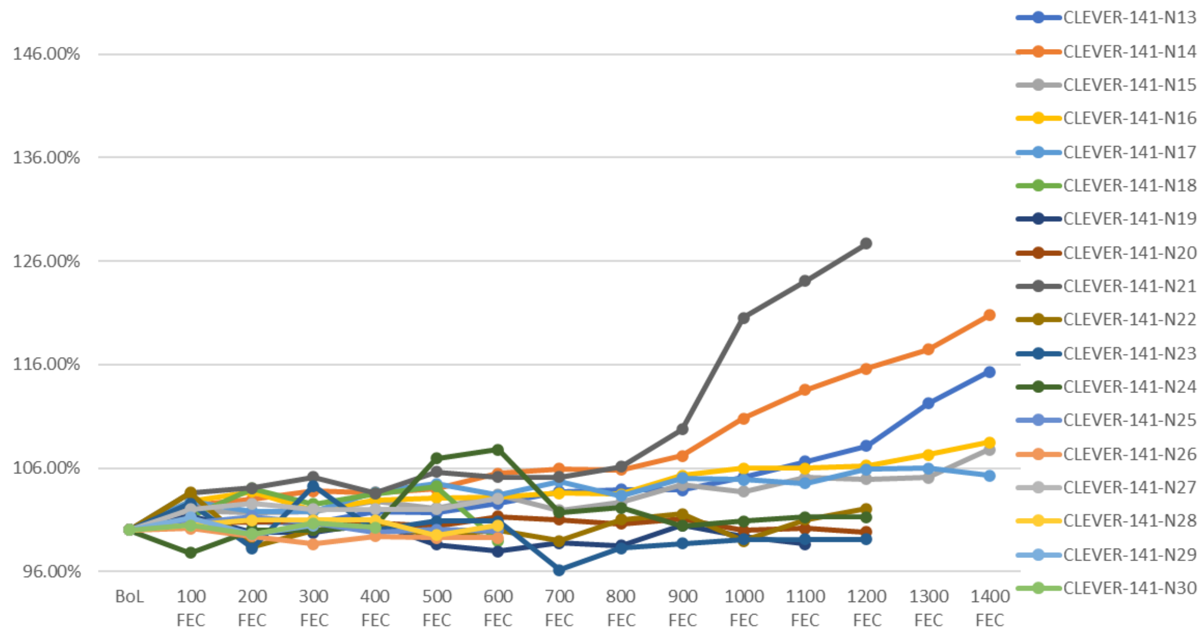

Figure 3: Internal resistance growth results of 18 NMC 141 Ah cells, where the vertical axis represents the growth percentile, and the horizontal axis represents the performed FECs.

4 Machine learning modeling

4.1 Feature Engineering

One of the challenges for the SoH estimation is to find suitable features that can capture the internal changes of the battery and correlate with capacity degradation. The battery current, voltage, and temperature data are the most commonly available signals that can be measured during battery operation and can provide useful information about the battery condition. However, these signals are affected by various factors, such as the operating conditions, the measurement noise, and the battery aging mechanisms, and thus, they are not directly related to the SoH of the battery.

In this work, critical features are selected using different techniques such as the Pearson Coefficient, Pearson component analysis, etc. As shown in Fig. 4, the Pearson coefficient is deployed to identify the 24 features out of a total of 43 features for the training because they have a score of more than 0.8. These features are extracted either directly from the tests performed during the campaign (i.e., capacity values) or from analysis (i.e., incremental capacity curves). Fig. 4 provides a comprehensive view of the interrelationships between variables, guiding the selection of independent and relevant features for the model. However, since the Correlation Matrix analysis is analogous to the Pearson Coefficient method (different representation), this method has not been used directly to retain important analysis. In this project, the Pearson coefficient finally won the competition since it revealed higher performance in comparison to other methods.

The training dataset was prepared by selecting relevant features (X), with the target variable being the SoH of the battery. After feature selection, the dataset was split into training and test sets using an 80-20 split, with 80% of the data used for model training and 20% for testing. This split ensures that the models can be evaluated on unseen data to measure their performance.

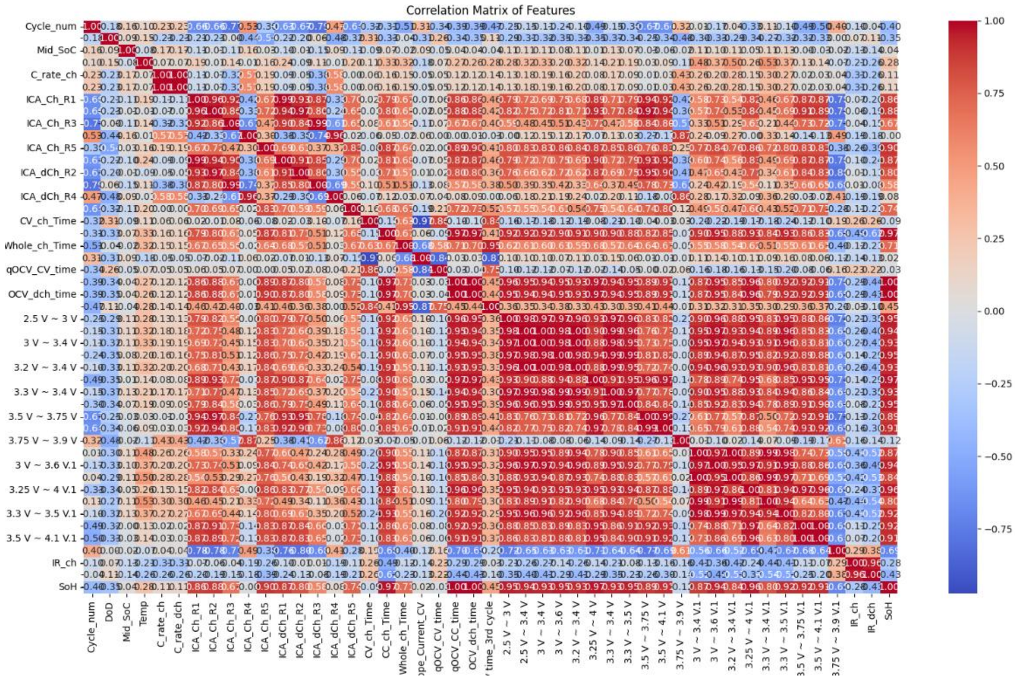

Figure 4: Correlation matrix for the dataset.

4.2 **Algorithm training**

ML techniques have been widely applied to estimate SoH for lithium-ion batteries. After training and testing different algorithms such as Gaussian Process Regression [10], ElasticNet [11], Random Forest Regression [12], etc., the best-performing algorithm is found to be the ElasticNet regression, which has been deployed to validate the dynamically cycled dataset.

The ElasticNet as expressed in equation (1), combines the penalties of Lasso (L1) and Ridge (L2) regularization. It linearly combines the two penalties to perform variable selection and regularization, preventing overfitting [13].

$$\hat{y} = \arg\min\left(\sum_{i=1}^{N}(y_i - X_i\beta)^2 + \lambda_1|\beta|_1 + \lambda_2|\beta|_2^2\right) \quad (1)$$

Where β represents the model parameters (coefficients), λ1 controls the L1 penalty (Lasso), λ2 controls the L2 penalty (Ridge).

The training process was run on a system with the following specifications:

- Processor: Intel Core i7-6820HQ @ 2.70 GHz, 8 cores
- RAM: 32 GB DDR4
- GPU: NVIDIA GeForce Quadro M2200 (4 GB GDDR5)
- Operating System: Windows 11 Pro
- Software Environment: Python 3.9, Scikit-learn 1.0.2, and TensorFlow 2.6.0 for model development and training

These system specifications ensured efficient dataset handling and expedited the training process, especially for computationally intensive models like Gradient Boosting and Gaussian Process Regression.

4.3 **Evaluation Metrics and Validation**

The trained models are validated using the assembled dataset to predict SoH. The predicted SoH values are compared with the actual SoH values to calculate the error. As the final step, the error is analyzed, and hyperparameter tuning is performed.

The performance of different models is compared using one evaluation metric, which is expressed in equation

(2).
RMSE (Root Mean Squared Error):

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

The predicted SoH values are compared with the actual SoH values to calculate the error. The performance of different models is compared using different evaluation metrics such as RMSE. The best RMSE score of the validation is found to be 0.065 in one of the validation cells, and the curves are displayed in Fig. 5. It is clear that though the algorithm could not identify the actual SoH with good accuracy at the BoL but shows a promising convergence by cycle numbers.

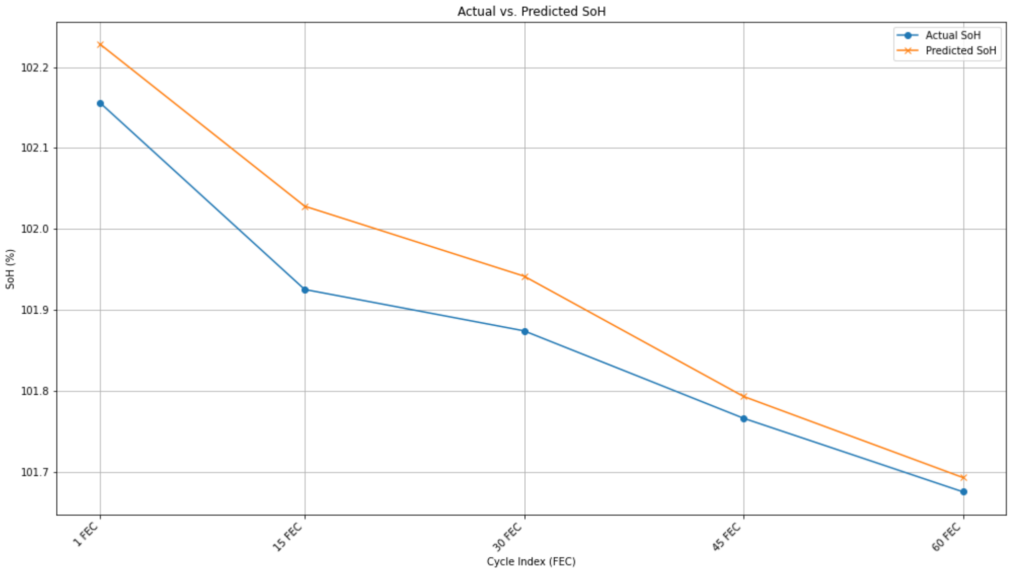

Figure 5: Validation curve of a sample cell (ID – N30).

5 Conclusion

The 2nd life NMC 141 Ah cells were characterized in this study. However, the status of the cells was close to fresh, meaning they had spent an unusual time in the vehicle. The cells were cycled under the same type of conditions. The high DoD cycling was found to have a faster degradation, and together with cold temperature, the cells experienced the most aging.
When different ML algorithms are trained with the generated dataset, the best-performing model is found as ElasticNet, with which the validation task is performed on an unknown dataset by dynamically cycling the cells. The RMSE value obtained from the validation is 0.065. The robustness of the model can also be increased by retraining with the availability of new data. The inclusion of first-life aging information could also improve prediction accuracy. Future work could be on how to reuse such a trained model for a completely new dataset of another type of NMC cell, facilitating transfer learning.

Acknowledgments

The research presented in this article is supported by the joint R&D project CLEVER, which is funded by the Brussels funding organization INNOVIRIS under grant number 2022-JRDIC-6b. The project has also received funding from the Europe Union's Horizon Europe project REBORN with the grant ID – 101138031.

References

[1] A. A. Kebede *et al.*, "Optimal sizing and lifetime investigation of second life lithium-ion battery for grid-scale stationary application," *J Energy Storage*, vol. 72, no. PC, p. 108541, 2023, doi:

10.1016/j.est.2023.108541.
[2] J. Van Mierlo *et al.*, “Beyond the State of the Art of Electric Vehicles : A Fact-Based Paper of the Current and Prospective Electric Vehicle Technologies,” *World Electric Vehicle Journal*, vol. 12, no. 1, p. 20, 2021, doi: https://doi.org/10.3390/wevj12010020.
[3] A. Kampker, H. H. Heimes, C. Offermanns, J. Vienenkötter, M. Frank, and D. Holz, “Identification of Challenges for Second-Life Battery Systems—A Literature Review,” Apr. 01, 2023, *MDPI*. doi: 10.3390/wevj14040080.
[4] M. S. Hosen, J. Jaguemont, J. Van Mierlo, and M. Berecibar, “Battery lifetime prediction and performance assessment of different modeling approaches,” *iScience*, vol. 24, no. 2, 2021, doi: 10.1016/j.isci.2021.102060.
[5] D. I. Stroe, M. Swierczynski, A. I. Stan, V. Knap, R. Teodorescu, and S. J. Andreasen, “Diagnosis of Lithium-Ion Batteries State-of-Health based on Electrochemical Impedance Spectroscopy Technique,” pp. 4576–4582, 2014, doi: 10.1109/ECCE.2014.6954027.
[6] W. Sung and C. B. Shin, “Electrochemical model of a lithium-ion battery implemented into an automotive battery management system,” *Comput Chem Eng*, vol. 76, pp. 87–97, May 2015, doi: 10.1016/j.compchemeng.2015.02.007.
[7] X. Li, C. Yuan, and Z. Wang, “State of health estimation for Li-ion battery via partial incremental capacity analysis based on support vector regression,” *Energy*, vol. 203, p. 117852, 2020, doi: 10.1016/j.energy.2020.117852.
[8] M. S. Hosen *et al.*, “Electro-aging model development of nickel-manganese-cobalt lithium-ion technology validated with light and heavy-duty real-life profiles,” *J Energy Storage*, vol. 28, 2020, doi: 10.1016/j.est.2020.101265.
[9] K. Jalkanen, J. Karppinen, L. Skogström, T. Laurila, M. Nisula, and K. Vuorilehto, “Cycle aging of commercial NMC / graphite pouch cells at different temperatures,” *Appl Energy*, vol. 154, pp. 160–172, 2015, doi: 10.1016/j.apenergy.2015.04.110.
[10] R. R. Richardson, M. A. Osborne, and D. A. Howey, “Gaussian process regression for forecasting battery state of health,” *J Power Sources*, vol. 357, pp. 209–219, 2017, doi: 10.1016/j.jpowsour.2017.05.004.
[11] C. Wang, J. Chen, and Y. Liu, “The Elastic Net Regularized Extreme Learning Machine for State of Charge Estimation,” *J Electrochem Soc*, vol. 170, p. 120502, 2023, doi: 10.1149/1945-7111/ad0ff7.
[12] Y. Li *et al.*, “Random forest regression for online capacity estimation of lithium-ion batteries,” vol. 232, no. October, pp. 197–210, 2018, doi: 10.1016/j.apenergy.2018.09.182.
[13] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *J R Stat Soc Series B Stat Methodol*, vol. 67, no. 2, pp. 301–320, 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

Presenter Biography

Md Sazzad Hosen is a research professor at the Vrije Universiteit Brussel in Belgium. He has worked in battery characterization and modeling for the last 8 years. He is a Mechanical graduate with a PhD in Engineering Sciences. His research interests include battery modeling, BMS, 2nd life battery systems, etc. He has authored more than 50 publications and has been working as an expert reviewer for reputed academic journals.