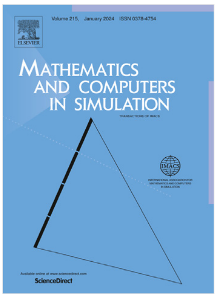

Capacity estimation of Lithium-ion batteries through a Machine Learning approach

Simone Barcellona, Lorenzo Codecasa, Silvia Colnago,
Loris Cannelli, Christian Laurano, Gabriele Maroni

PII:	S0378-4754(25)00215-0
DOI:	https://doi.org/10.1016/j.matcom.2025.05.022
Reference:	MATCOM 6874
To appear in:	*Mathematics and Computers in Simulation*
Received date :	3 February 2025
Revised date :	18 April 2025
Accepted date :	20 May 2025

Highlights

Capacity Estimation of Lithium-Ion batteries Through a Machine Learning Approach

Simone Barcellona, Lorenzo Codecasa, Silvia Colnago, Loris Cannelli, Christian Laurano, Gabriele Maroni

- Creation and validation of a machine learning algorithm to estimate the actual capacity of a single lithium-ion battery using only two experimental points.
- Training the algorithm on data from a second battery to accurately predict the performance of the first battery, excluded from the training process.

Capacity Estimation of Lithium-Ion batteries Through a Machine Learning Approach

Simone Barcellona[a,*], Lorenzo Codecasa[a], Silvia Colnago[b], Loris Cannelli[c], Christian Laurano[a], Gabriele Maroni[c]

[a]DEIB – Politecnico di Milano, Piazza Lonardo da Vinci, 32, Milano, 20133, Italy,
[b]Ricerca sul Sistema Energetico RSE SpA, Via Rubattino 54, Milano, 20134, Italy,
[c]IDSIA Dalle Molle Institute for Artificial Intelligence, SUPSI, Via la Santa 1, Lugano, 6962, Switzerland,

Abstract

Lithium-ion Batteries (LiBs) have become of paramount importance due to their employment in application fields, including renewable energy sources and electric vehicles (EVs), which heavily rely on them. This has spurred research efforts to develop battery models capable of predicting and estimating battery behavior to optimize usage and reduce degradation. To this end, key state parameters, including State Of Charge (SOC) and State of Health (SOH), should be accurately estimated. In the literature, many estimation methods are based on the knowledge of the relationship between Open-Circuit Voltage (OCV) and SOC. The latter can be modeled in different ways, organized into three main approaches: table-based, analytical, and artificial intelligence approaches. Among these, Machine Learning approaches have gained popularity and have shown great promise for this purpose. However, previous studies typically require many OCV-SOC data points or entire fragments of the OCV curve, which makes them unsuitable for EV applica-

*Corresponding author

Email addresses: simone.barcellona@polimi.it (Simone Barcellona), lorenzo.codecasa@polimi.it (Lorenzo Codecasa), silvia.colnagoa@rse-web.it (Silvia Colnago), loris.cannelli@idsia.ch (Loris Cannelli), christian.laurano@polimi.it (Christian Laurano), gabriele.maroni@idsia.ch (Gabriele Maroni)

[1]This work has been partially financed by the Research Fund for the Italian Electrical System under the Three-Year Research Plan 2025-2027 (DM MASE n. 388, 06.11.2024), in compliance with the Decree of April 12th, 2024.

tions. To address this limitation, the present paper develops and validates an ML algorithm to estimate the battery capacity of a LiB using only two experimental OCV points, accounting for different levels of cycle aging. The results demonstrate that the model, when trained on an accelerated-aged battery, can accurately predict the actual capacity of other batteries with similar characteristics but different aging levels.

Keywords: Lithium-ion battery, Cycle aging, Machine learning, Capacity estimation
PACS: 0000, 1111
2000 MSC: 0000, 1111

1. Introduction

Nowadays, the implementation of Renewable Energy Sources (RESs) has become of key importance in sustainable development and addressing global warming. Unfortunately, the intrinsic intermittent and unpredictable nature of RESs makes it difficult to match energy supply with demand. Equally important for the environment is the ongoing electrification of the transportation sector, spearheaded by the rise of Electric Vehicles (EVs). Lithium-ion Batteries (LiBs) play a crucial role in both sectors due to their excellent performance, characterized by high energy and power density as well as low self-discharge rates [1].

On the other hand, to properly use LiBs and increase their efficiency and lifetime, it is important to predict and estimate some of their parameters and states. In particular, the State Of Charge (SOC) and the State Of Health (SOH) are useful for ensuring that LiBs operate within their safe limits and under optimal conditions. As a consequence, battery modeling becomes essential to fulfill these goals. In the literature, numerous battery models address three key aspects: electrochemical behavior, thermal behavior, and aging mechanisms. These aspects can be represented using physical, circuit-based, data-driven approaches, or a combination of these methods [2].

It is well known that LiBs degrade due to calendar aging or cycle aging. While calendar aging refers to the battery being stored under certain conditions (such as SOC and temperature), cycle aging refers to the battery operation under specific conditions (such as SOC range, current rate, and temperature) [3, 4]. In any case, battery aging results in either a decrease in capacity (energy fade) or an increase in internal resistance (power fade).

While temperature is indeed a factor contributing to irreversible battery aging, it also influences battery performance in a reversible manner by affecting both internal resistance and capacity [5].

Similarly, this degradation can cause an irreversible change in the Open-Circuit Voltage (OCV) characteristic, while temperature affects it in a reversible manner. Specifically, the literature contains a lot of research that aims to model OCV as a function of SOC or the absolute state of discharge (q) in different ways and how it changes with aging [6, 7, 8] or temperature [9, 10, 11]. In particular, we can recognize three main categories: table-based approaches, analytical approaches, and artificial intelligence approaches. The former are straightforward to implement but require the knowledge of many experimental points to be stored in lookup tables if high accuracy is required [12]. The analytical approach uses mathematical functions, such as polynomials [13], logarithms [14], exponentials [15], or combinations thereof [16], to fit the OCV-SOC or OCV-q relationship. This approach is relatively simple and does not require a huge amount of experimental data. However, achieving high accuracy depends on the type of function employed and the number of parameters. In recent years, artificial intelligence approaches based on Machine Learning (ML) algorithms have become popular and interesting for this goal, [17, 18]. Various features have been extracted to build different ML strategies. They are commonly derived from voltage, current, and temperature measurements during the charging or discharging processes. For example, in [19], features were built from these quantities during the constant current and constant voltage charging process, while in [20] considered the same quantities along with elapsed time of the discharge process. In [21, 22] various unique features were constructed based on discharge capacity curves and the integral of battery temperature.

Considering the complexity of battery research, it is essential to separately analyze the irreversible effects of cycle aging and the reversible effects of battery temperature. As an initial step, this study investigates how the OCV curve evolves with cycle aging while maintaining a constant battery temperature. The ultimate aim is to estimate the actual battery capacity, which is derived based on the observed variations in the OCV curve, specifically targeting the SOH in terms of capacity fade for EV applications under fixed temperature conditions.

In the literature, SOH estimation methods can be broadly classified into two main groups: model-based and non-model-based approaches. Model-based approaches rely on modeling the degradation mechanisms occurring within the battery or on modeling the variation of certain electrical parameters with aging. These methods are typically implemented using approaches such as Kalman filters or observers [23, 24, 25]. These estimation methods are usually very accurate but computationally intensive. Conversely, the non-model-based approaches are simpler and demand less computational effort. A straightforward method for estimating the actual battery capacity involves integrating the current during battery discharge after fully charging the battery to 100% of SOC. However, this approach is unsuitable for EVs, as only the charging phase can be controlled. Other methods are based on the reconstruction of the OCV-SOC curve. Nevertheless, estimating the OCV-SOC curve requires the acquisition of various OCV points. Since the OCV can only be measured once the relaxation effect from diffusion processes are extinguished, the battery must remain inactive for a significant period to allow it to reach equilibrium. To address this, in [26] the OCV curve of a LiB was reconstructed from fragments, combining them based on their monotonicity and shape correlation. The authors, in [27], employed a multi-population genetic algorithm to reconstruct the whole OCV curve from fragments. Nonetheless, collecting these OCV curve fragments remains a challenging task for EV applications.

In light of the above, the present study focuses on developing and validating an ML algorithm to estimate the actual capacity of a single LiB using only two experimental OCV points, while considering various levels of cycle aging at a fixed room temperature. An undeniable advantage of ML is its ability to create models that generalize and scale effectively, achieving strong performance even on unseen data during training. The proposed method proves to be suitable for EVs, as the two experimental points can be obtained while the EV is idle at various times, such as overnight or when parked for extended periods of over an hour without power exchange.

Furthermore, the analysis was expanded by training the algorithm using data from a second battery to predict the performance of the first battery, which was excluded from the training process. Significantly, the second battery was cycled at an increased current rate to expedite the cycle aging test. This strategy offers the benefit of quicker data acquisition from the second battery, which can subsequently be utilized for predictions across different current rates.

Starting from the results of this work, future research can focus on building a more diverse training set by incorporating OCV-q curves from multiple batteries of the same type to better account for inter-battery variability. Additionally, future studies can extend these results to develop a more versatile battery capacity estimation approach that can accommodate varying operating conditions (as temperature) and different battery chemistries.

2. Battery Model and OCV extraction

The electric circuit model and the OCV extraction procedure used to obtain the OCV-q experimental curves are detailed in [15] and are briefly summarized here for completeness. Since our focus is on evaluating the discharge OCV curves, accounting for the various dynamic behaviors of the battery is not required. Instead, only the steady-state behavior of the battery is of interest. Therefore, we used a straightforward zero-order Thevenin electric circuit model. It is worth noting that this battery model was employed solely to process the experimental data and was not used in the proposed OCV estimation algorithm.

Figure 1 shows the battery model composed of an ideal voltage source representing the open circuit voltage of the battery, which varies with the absolute state of discharge (q) and cycling level (Q), and a series resistor (R_b). This resistor accounts for the battery's high-frequency resistance, the resistance related to the solid electrolyte interface, the resistance related to the charge transfer process, and finally, the resistance related to the diffusion process within the electrodes and electrolyte. This electric model is described in mathematical terms by the following equation, which relates the output battery voltage to the battery current:

$$v_b(q, Q) = V_o(q, Q) - R_b \cdot i_b. \tag{1}$$

The value of the absolute state of discharge, expressed in Ah, can be calculated using the equation:

$$q(t) = \frac{1}{3600} \int_0^t i_b \cdot d\tau + q(0), \tag{2}$$

where $q(0)$, is the initial absolute state of discharge. Through the knowledge of q and of the actual battery capacity (C_a), the SOC can be determined as:

$$SOC = \left(1 - \frac{q(t)}{C_a}\right) \times 100. \tag{3}$$

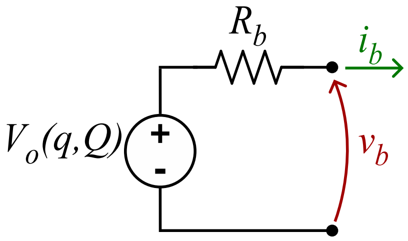

Figure 1: Battery equivalent circuit model.

The level of cycle aging (cycling level) was quantified through the total moved charge of the LiB defined as:

$$Q(t) = \frac{1}{3600}\int_0^t |i_b| \cdot d\tau. \tag{4}$$

From an experimental standpoint, the OCV of a battery can be evaluated by measuring the output voltage removing the voltage across the internal resistance. To achieve this, the Galvanostatic Intermittent Titration Technique (GITT) or pseudo-OCV tests can be employed [28, 29]. Both, however, are time-consuming. The GITT involves a sequence of partial discharges, each followed by a prolonged resting period to allow the battery to reach a steady state before measuring the OCV. The pseudo-OCV test requires fully discharging the battery at a low currant rate, usually lower than 0.1C, neglecting the resistance voltage drop. Alternatively, in [15], the OCV of the battery was derived by discharging it at a 1C rate, which significantly expedited the test. This was done by measuring the output battery voltage and compensating for the voltage drop across its internal resistance using the model reported in Fig 1. Furthermore, although the internal resistance can vary with SOC, according to the assumption done in [30], it was assumed to remain constant as a function of q. Thus, the resistance R_b was estimated at the beginning of the discharge following the procedure outlined in [31]. Essentially, when the battery initiates discharge, an electric transient occurs. By selecting a sufficiently long time interval to consider the electric transient extinguished, the internal resistance can be calculated as the ratio between voltage and current variations. Finally, the OCV interval corresponding to the electric transient is eliminated.

3. Experimental Procedure

The experimental procedure used to obtain the OCV-q curves was conducted on two 10 Ah pouch $LiCoO_2$ cells (model 8773160K). These batteries, manufactured by General Electronics Battery Co. Ltd., have a voltage range of 2.75 V – 4.2 V and a maximum discharge current of 100 A (10C).

3.1. Test procedure

The test procedure consisted of two phases: the battery characterization phase and the aging phase. During the characterization phase, the OCV curves were obtained. This phase was conducted at the beginning of the test and repeated after each aging phase.

The characterization phase was identical for both batteries and performed at 20°C. The batteries were fully charged using a Constant Current–Constant Voltage (CC-CV) protocol, with the CC phase carried out at 10 A (1C) and the CV phase at 4.2 V until the current dropped below 100 mA (0.01C). At this point, the batteries were considered fully charged. Subsequently, the batteries were discharged at a constant current of 10 A until the minimum voltage of 2.75 V was reached. The OCV curve was derived from the discharge data by removing the voltage drop across the internal resistance, as previously described.

The aging phase consisted of a series of charging and discharging current steps performed at 30°C, cycling the batteries for approximately 400-700 Ah. The current steps had amplitudes of 25 A (2.5C) and 50 A (5C) for the two batteries, respectively. Charging and discharging were constrained by both SOC (20–80%) and voltage (3.45 V – 4.05 V) boundaries. The aging phase was repeated until a total moved charge of approximately 20 to 30 kAh was reached.

It is worth noting that, if the voltage boundaries are not reached during either the charging or discharging phase, each cycle consistently moves the same amount of charge. However, if at least one of the voltage limits is reached, the voltage drop caused by the battery internal resistance results in a different amount of charge being moved depending on the current rate. However, this variation is not of concern because, as previously reported in [32], between 20% and 80% of the SOC range and at a constant temperature, capacity fade depends solely on the moved charge. Moreover, as analyzed in [33], under the same assumptions, even the current rate does not affect capacity fade.

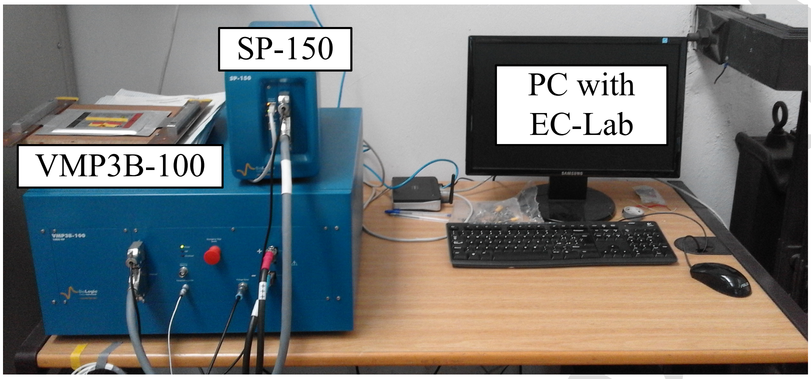

Figure 2: SP-150, VMP3B-100, and PC, [15].

3.2. Experimental setup

To charge and discharge the batteries and to collect the data, a potentiostat (SP-150) and a booster (VMP3B-100) manufactured by Biologic Science Instrument were employed. These devices are controlled via ethernet by a PC with EC-Lab software. The battery temperature was kept constant by three Peltier cells placed below the LiB under test and over a heatsink. Throughout the aging process, the current in the Peltier cells was managed to maintain the battery temperature at approximately 30°C. The battery temperature was measured using a temperature probe placed on the top face of the battery under test. However, it is possible to neglect the temperature gradient between the top and bottom faces of the battery, as the pouch battery is very thin.

As the battery current was reduced to 10 A for the characterization phase, the battery temperature naturally began to decrease. Upon reaching 20°C, the thermostat deactivated the Peltier cells to facilitate the OCV measurement at approximately 20°C. The figures of the test setup and temperature control system are the same of [15] and reported in Figure 2 and 3.

4. Machine Learning Model

This study aims to assess the efficacy of employing an ML model to capture and learn the OCV-q relationship from data. Unlike mathematical functions, which rely on predefined equations and assumptions about the underlying relationships within the data, ML models possess the capability

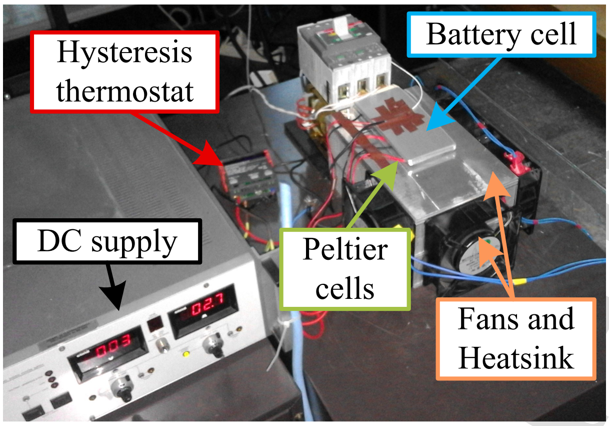

Figure 3: Battery control system, [15].

to autonomously discern complex patterns and structures (see, e.g., [34] and references therein).

4.1. Dataset creation

The experimental campaign described in section 3 resulted in 62 and 43 OCV-q curves at different cycling levels, for the battery cycled at 25 A and 50 A, respectively, as illustrated in the Figures 4 and 5. These curves are characterized by a nearly linear behavior with almost no curvature, followed by a point of maximum curvature known as the *knee point*, where the curve visibly bends, transitioning from a low to high curvature behavior. This study introduces an ML-based methodology to estimate the actual capacity of the battery, considering a realistic scenario where only two observations from an OCV curve are used. Specifically, during a rest period of an EV exceeding 1 hour, it is feasible to measure an OCV point ($V_o^{(n)}$) and its corresponding absolute state of discharge value ($q^{(n)}$) assessed using the Coulomb counting method. Given two pairs of such values, $q^{(1)}, V_o^{(1)}$, and $q^{(2)}, V_o^{(2)}$, observed at significant intervals of absolute state of discharge and without the knowledge of cycling level (Q), it is possible to use a predictive model to determine when the OCV-q curve reaches the cutoff voltage of 2.75 V, which represents the actual battery capacity. Since the voltage drop across the internal resistance was subtracted, the experimental OCV points do not actually reach the cutoff voltage of 2.75 V. For simplicity, the abscissa of the last experimental point $(q(V_{min}), V_{min})$ of the OCV-q curve, being closest to the cutoff voltage, was chosen as the target variable for predictions with the ML model.

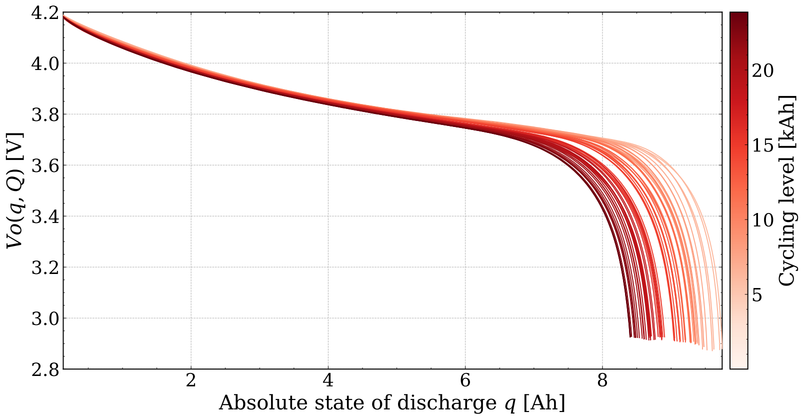

Figure 4: 62 OCV-q experimental curves at different cycling levels, for the battery cycled at 25 A.

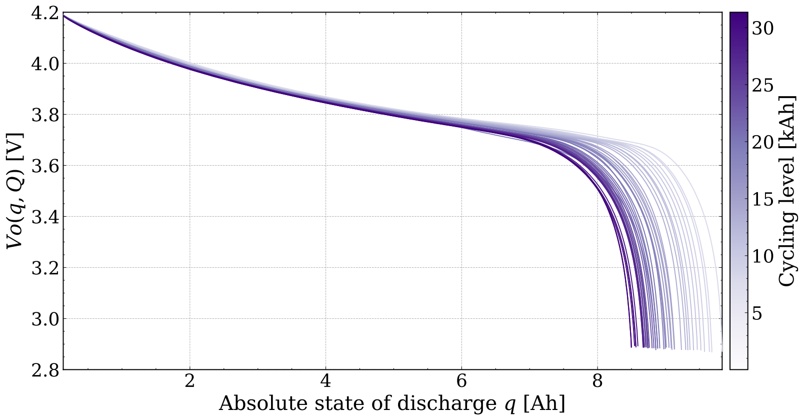

Figure 5: 43 OCV-q experimental curves at different cycling levels, for the battery cycled at 50 A.

As indicated in [35], the optimal operational SOC range for LiBs spans from 20% to 80%. In practical terms, this suggests that capturing OCV-q curve observations near or beyond the knee point, although very informative, is improbable under typical usage conditions. Thus, to generate a training dataset for an ML model that closely mimics reality, the sampling domain for the absolute state of discharge values q was constrained according to the nominal capacity of the LiB under test (10 Ah) and its reduction due to cycle aging. Therefore, it was opted to restrict the absolute state of discharge values between 2 Ah and 6 Ah. The abscissa values of the first point $q^{(1)}$ in each pair were uniformly sampled within this range, while the subsequent points $q^{(2)}$ were uniformly sampled at a minimum 1Ah distance from the first abscissa point, up to 6 Ah. Any point pairs where $q^{(2)} = q^{(1)} + 1$ exceeded 6 Ah were systematically excluded from the dataset. The resulting dataset was structured such that each row represents a pair of sampled points from the OCV curves, encapsulating four coordinates in the OCV-q plane $\left(q^{(1)}, V_o^{(1)}, q^{(2)}, V_o^{(2)}\right)$, which served as the input features for the ML model. Additionally, an extra column in each row was dedicated to the target variable, i.e., the actual battery capacity, namely the absolute state of discharge value at the minimum voltage ($q(V_{min})$) of the curve from which the pair was sampled. This design highlights the methodology's intent to model the battery capacity estimation based on limited observations from the discharge curve.

4.2. Algorithm

The ML model that was used in this work is the Light Gradient Boosting Machine (LightGBM) framework. LightGBM represents an advanced implementation of the renowned Gradient Boosting (GB) framework [36] and is known for its exceptional efficiency and performance on tabular regression tasks involving small-to-medium datasets, minimal data preprocessing requirements, and native compatibility with interpretability tools such as SHAP [37]. At its core, LightGBM operates by constructing an ensemble of weak prediction models, typically simple decision trees, in a sequential manner. Each successive tree within the sequence tries to rectify the residuals or errors of its predecessors, ultimately culminating in a predictive model that aggregates all the weak decision trees. LightGBM, unlike traditional GB methods, utilizes Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). GOSS prioritizes instances with larger gradients during training, enhancing efficiency without sacrificing accuracy. EFB combines exclusive features to reduce computational complexity effectively by reducing the feature space. LightGBM also employs a leaf-wise tree growth algorithm, contrasting with the level-wise approach in standard GB, resulting in faster convergence and improved efficiency [38]. Its suitability for this work is supported by recent benchmarking results [39], which show that tree-based ensemble models like LightGBM and XGBoost consistently outperform other approaches, including Random Forests, Support Vector Machines, and deep learning models, on structured datasets. In our application, these strengths enabled us to build a model that is not only accurate and fast, but also interpretable, which is crucial for real-world deployment in safety-critical systems like battery management.

4.3. Feature Importance and Feature Engineering

Although the primary goal of ML is often to maximize the accuracy of the predictions of a target variable, based on the available informative features, the identification of the most relevant features can be of equal importance. Knowing which predictors significantly affect the response improves the comprehension of causal relationships and can inform future research. In addition, it assists in making informed decisions in various practical applications. ML models, including LightGBM, are highly effective in predicting the variable of interest. Yet, unlike physical models that transparently reveal the mathematical functions mapping inputs to predictions, these ML models are often viewed as black boxes due to their complexity and to the number of parameters involved. However, it is still feasible to decode their internal prediction mechanisms through ad hoc explanation techniques, such as *feature importance* methods, which not only shed light on how predictions are made, but also serve to validate a priori domain knowledge. Additionally, leveraging this knowledge enables improvement of the performances through *feature engineering.*

Feature importance. In this work, it was decided to measure feature importance in terms of Shapley values. The concept of Shapley values traces back to cooperative game theory [40]. The application of Shapley values to ML for explaining model predictions was proposed in [41]. Roughly speaking, in the ML context each feature is considered as an individual player in a coalition, collectively contributing to the formation of a prediction. More specifically, given an ML model $f(\cdot)$ and an instance x of a dataset X, possessing, e.g., K features $x = [x_1, x_2, \ldots, x_K]$, the sum of the Shapley values for all the features must be equal to the deviation between the model prediction $f(x)$ and an estimated baseline model output y_{base}. This can be written as: $f(x) = y_{base} + \phi(x_1) + \ldots + \phi(x_K)$, where $y_{base} \triangleq \mathbf{E}[f(X)]$ denotes the average prediction of the ML model f across all training data, and $\phi(x_j)$ denotes the Shapley value for the jth feature [40, 41]. In the context of this work, a positive Shapley value $\phi(\cdot)$ indicates a positive deviation in the absolute state of discharge compared to the expected value y_{base}, while a negative Shapley value indicates the opposite. By computing the average magnitude of Shapley values for each feature: $\langle|\phi(x_j)|\rangle \triangleq \frac{1}{n}\sum_{i=1}^{n}|\phi_i(x_j)|$, where n is the size of the dataset, we can assess the overall importance of each feature on the model predictions.

The theory of Shapley values computes the importance to assign to each feature, trying to weigh its contribution to the final prediction.

In this work, Shapley values were computed with the Tree SHAP algorithm, which provides exact Shapley values for tree ensembles in polynomial time, as introduced by Lundberg et al. [42]. This method is implemented efficiently in the `lightgbm` package through the `pred_contrib=True` option, which we used during prediction. The implementation relies on optimized C++ backends and supports fast, scalable computation.

Feature engineering (FE). FE is the process of using domain knowledge to select, modify, or create new features from raw data in order to enhance the performance of ML models. In this work, the following features were derived and evaluated:

- `Slope`: the slope of the line segment connecting a couple of sampled points $\frac{V_o^{(2)}-V_o^{(1)}}{q^{(2)}-q^{(1)}}$.
- `Midpoint`: midpoint of the ordinate coordinates $\frac{V_o^{(1)}+V_o^{(2)}}{2}$.
- `Harmonic Mean`: harmonic mean of the ordinate coordinates: $\frac{2}{\frac{1}{V_o^{(1)}}+\frac{1}{V_o^{(2)}}}$.
- $\texttt{dist}^{(2)}((\bar{q},\bar{V}_o))$: Euclidean distance of the second point from a fixed point $(\bar{q},\bar{V}_o)$: $\sqrt{(\bar{q}-q^{(2)})^2+\left(\bar{V}_o-V_o^{(2)}\right)^2}$. In this work: $(\bar{q},\bar{V}_o) = (6,4)$.

5. Results

All computations were performed on a server with two 64-core AMD EPYC 7742 processors, 256 GB of RAM, and 4 Nvidia RTX 3090 GPUs. The hardware resources of the server were limited to 16 CPU threads and 0 GPUs. Python was used as a programming language, with the LightGBM library for the ML training. The code to reproduce the results is available at the GitHub repository: https://github.com/gabribg88/LiB_Capacity_Estimation.

5.1. Experiment with a single battery dataset

Starting from the 62 OCV-q curves gathered during the experimental campaign from the first battery cycled at 25 A, a dataset comprising a total of approximately $N = 10^4$ observation pairs was generated. These pairs were evenly distributed across each OCV-q curve, resulting in $\lfloor N/62 \rfloor$ pairs per curve, in accordance with the sampling methodology detailed in section 4.1. For each pair, the target variable, which corresponds to the battery capacity, is defined as the discharge value at the minimum voltage point of the corresponding OCV-q curve, denoted as $(q(V_{min}))$. To rigorously assess the predictive accuracy on unseen data, 12 out of the 62 curves (approximately 20%) were randomly selected as the test set. The observation pairs corresponding to these selected curves were excluded from the training set to form the test dataset. This separation guarantees an unbiased evaluation of the model performance against data not encountered during its training. For model tuning, a standard group k-fold cross-validation strategy was employed. This strategy involved dividing the dataset into five folds (k=5), with the constraint that samples from each OCV-q curve can belong to one fold only. This approach simulates a realistic scenario where the model is challenged with predicting outcomes for OCV-q curves unseen during training. This cross-validation technique tries to mirror real-world applications where the model must generalize well to new batteries or conditions.

The performances of three approaches were compared:

1. *Naive prediction*: this benchmark method involves predicting the battery capacity as the average value of the target variable distribution within the training dataset. This approach does not leverage any ML algorithms or feature inputs, and it is shown as a benchmark.
2. *Model trained without FE*: this approach involves training a LightGBM model directly on the raw data, without applying any FE technique.

3. *Model trained with FE*: contrary to the previous method, this approach incorporates FE, as detailed in section 4.3, to enhance the model performance.

To quantify the performances of the described approaches, we used the following three metrics, where $q_i(V_{min})$ and $\hat{q}_i(V_{min})$ represent the target value and the corresponding prediction of the ith observation, respectively. The index i ranges from 1 to n, with n being the total number of observations in the test set:

- Absolute Percentage Error (APE):

$$\text{APE}_i = \left| \frac{q_i(V_{min}) - \hat{q}_i(V_{min})}{q_i(V_{min})} \right| \times 100.$$

- Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \left(\frac{1}{n} \sum_{i=1}^{n} \left| \frac{q_i(V_{min}) - \hat{q}_i(V_{min})}{q_i(V_{min})} \right| \right) \times 100.$$

- Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| q_i(V_{min}) - \hat{q}_i(V_{min}) \right|.$$

- Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (q_i(V_{min}) - \hat{q}_i(V_{min}))^2}.$$

5.1.1. Prediction results

The numerical results are presented in Table 1. The naive approach, which serves as a baseline, shows the highest errors across all metrics. The improvement is dramatic when moving from the naive approach to an approach that use an ML model. Regarding the model without FE approach, the MAPE drops significantly from 3.97% to 0.59%, and both the MAE and RMSE experience substantial reductions. Interestingly, incorporating

FE into the training process yields the best results across all metrics. The MAPE further decreases to 0.48%, and the MAE and RMSE are reduced to their lowest values in the comparison. Training time of the model without FE on the described hardware is approximately 5 s, while it increases to 8.5 s with FE. Both models have comparable inference times, around 0.3 s. Thus, model training can be done offline on suitable hardware, while inference can be easily performed on embedded devices too.

Table 1: Comparative performance analysis of three predictive approaches. Bold numbers indicate the best results across all approaches.

Approach	**MAPE (%)**	**MAE (Ah)**	**RMSE (Ah)**
Naive prediction	3.97	0.35	0.38
Model without FE	0.59	0.05	0.07
Model with FE	**0.48**	**0.04**	**0.06**

Figure 6 provides a diagnostic of the test predictions made by the best performing model, LightGBM with FE. Left panel of the figure displays a scatter plot comparing the actual absolute state of discharge values at minimum voltage $q(V_{min})$ versus the model predictions, where each point denotes the median value and the bars indicate the 5th and 95th percentiles. The tight alignment of these median predictions with the bisector line denotes accurate model performance. The right panel illustrates the APE distribution across the test dataset, showing the median APE at 0.33%.

Finally, to evaluate the impact of training set size on performance, a massive Monte Carlo simulation was conducted, the results of which are shown in Figure 7. This figure, in fact, illustrates the sensitivity of the performance in terms of MAPE of the best model as the number of OCV curves in the training set varies, as well as the total number N of observation pairs. To enhance the statistical significance of the results, we replicated each condition 100 times, utilizing different random seeds for each replication. The experimental results 1) confirm that expanding the dataset improves performance (as expected), but more importantly, 2) can be used as a guideline for estimating the effort required for data acquisition campaigns and the dataset size needed during training to achieve the desired performance threshold.

We acknowledge the slight discrepancy between the MAPE value reported in Table 1 (0.48%) and the median MAPE shown in Figure 7 for 50 training curves and 10k samples ($\sim$ 0.58%). This difference arises from the exper-

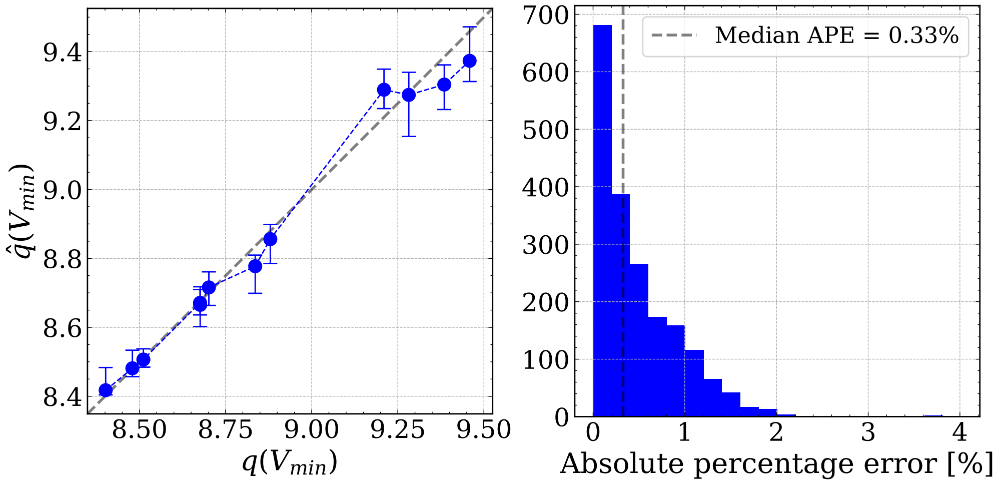

Figure 6: Prediction accuracy of the best model: left panel shows a scatter plot of actual vs. predicted values, with median predictions and percentile ranges. Right panel illustrates APE distribution.

imental protocol: in the experiment reported in Table 1, the model was evaluated on a fixed set of 12 test curves, sampled once at random and residing mostly within the interpolating regime of the model, i.e., at moderate aging levels, well-represented in the training data. In contrast, the massive Monte Carlo experiment in Figure 7 involved resampling the test curves for each run, which occasionally included curves from extreme aging conditions, where the model's performance naturally degrades due to its limited extrapolation capability. As a result, the median error in Figure 7 is slightly higher, but the originally reported result remains within the empirical distribution of outcomes. We therefore consider it a plausible, though favorable, instance, and we now clarify this to avoid misinterpretation.

5.1.2. Feature analysis

Feature importance results are shown in Figure 8. The ranking of the features, ordered by their global importance $\langle|\phi(x_j)|\rangle$, as calculated on the test dataset with the approach described in section 4.3, is shown in the left panel. Starting from a baseline value (y_{base}) of approximately 9 Ah, the chart highlights that the most influential feature is $V_o^{(2)}$, with an average contribution of 0.24 Ah to the final prediction. This remarks the critical role of the second voltage observation in assessing the actual battery capacity. The

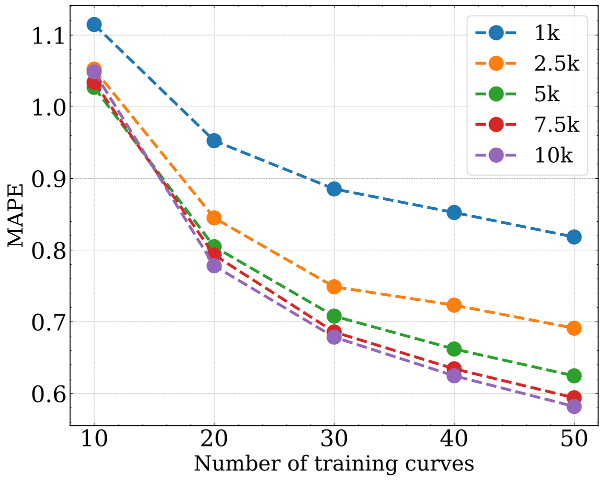

Figure 7: Performance of our proposed model trained with FE, in terms of MAPE as a function of dataset size. Median over 100 repetitions (dots) are depicted.

engineered feature $\texttt{dist}^{(2)}((6,4))$ also plays a significant role, adding on average 0.15 Ah to the prediction. The panel on the right side further elucidates the direction of these contributions: high values of $V_o^{(2)}$ positively affect the prediction (resulting in a higher predicted absolute state of discharge than the base value), whereas low values negatively influence the prediction (leading to a lower predicted absolute state of discharge than the base value). Conversely, for $\texttt{dist}^{(2)}((6,4))$, the impact is opposite. This model behavior aligns with expectations, thereby affirming pre-existing domain knowledge. The total SHAP value computation time over the entire test set was negligible in comparison to training times and comparable to inference durations. Consequently, the computational overhead of this feature attribution technique is minimal, and it remains suitable even for real-time or embedded use cases.

5.2. Generalization experiment with a second battery dataset

In practical scenarios, deploying an ML-based approach for battery capacity prediction across multiple batteries (potentially from different manufacturers or subjected to varying operating conditions) requires robust generalization. To assess this property, a dedicated experiment was performed using a new dataset consisting of 43 OCV-q curves gathered from a second battery, aged under a more aggressive current of 50 A (shown in Figure 5);

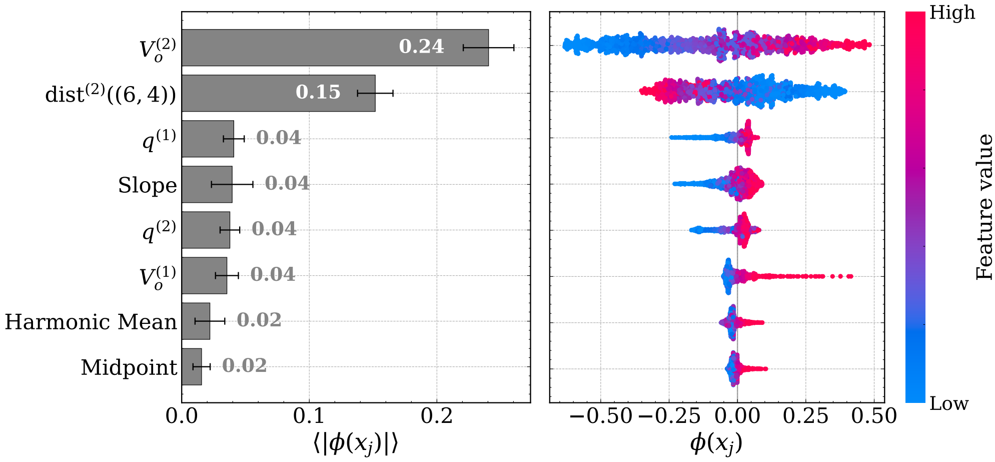

Figure 8: Feature importance analysis measured as Shapley values. Left: bars represent average feature contribution across the dataset, computed on the test set using the models trained over five different folds. Whiskers indicate variability across folds. Right: feature contribution shown for each data point. The color indicates the feature value.

see section 3 for details on data acquisition. All 43 OCV-q curves were used as the training dataset to train the LightGBM model with FE (the best model identified previously), and the resulting model was then tested on a set of 12 OCV-q curves from the previously studied battery cycled at 25 A. These 12 test curves were identical to those used in Section 5.1.1 (the original experiment), allowing for a direct comparison of performance metrics.

The results of this generalization test are shown in Figure 9.

Table 2 reports the corresponding MAPE, MAE, and RMSE. Compared to the results obtained when training and testing on curves taken from the same battery, the error levels are somewhat higher (e.g., MAPE rising to 0.84% from 0.48%), indicating the natural performance drop when a model is trained on data from one specific battery and deployed on a different one. Nonetheless, the errors remain reasonably low, underscoring the capacity of the proposed approach, and especially the feature engineering step, to generalize beyond its original training conditions.

While these results are encouraging, further improvements can be pursued to enhance model robustness. For instance, constructing the training set from multiple batteries rather than just one (even if aged under varied conditions) would likely provide more comprehensive exposure to the

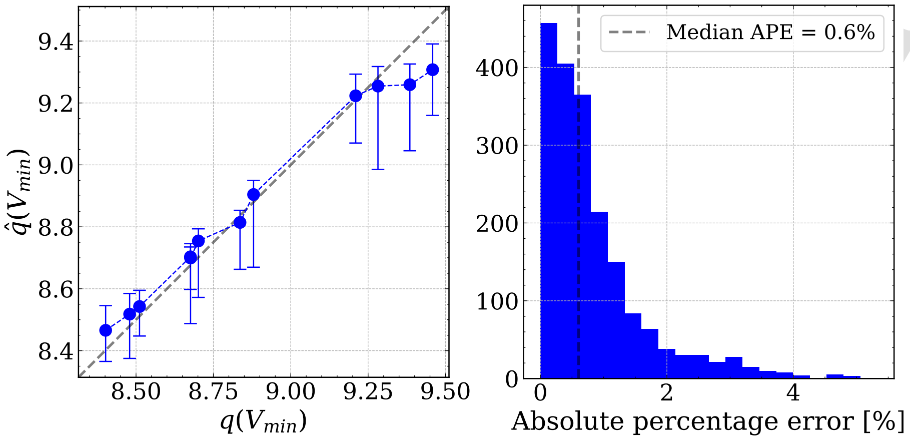

Figure 9: Prediction accuracy of the best model, tested on a second battery: left panel shows a scatter plot of actual vs. predicted values, with median predictions and percentile ranges. Right panel illustrates APE distribution.

Table 2: Performance results of the best model for the generalization test.

Approach	**MAPE (%)**	**MAE (Ah)**	**RMSE (Ah)**
Model with FE	0.84	0.08	0.11

range of possible battery behaviors. This broader training data could improve the model ability to learn generalizable patterns, ultimately reducing performance degradation when inferring on new battery datasets.

6. Conclusion

Battery capacity estimation is of paramount importance in determining the SOH of LiBs. Consequently, several techniques have been developed in recent years to achieve this goal.

In this work, we presented an ML approach based on LightGBM to estimate the capacity of LiBs. The method relies on only two observed OCV points and their corresponding absolute state of discharge values, making it practical for on-board implementation. Our first set of experiments, conducted on a dataset of 62 OCV-q curves from a battery cycled at 25 A, demonstrated that the proposed feature-engineering strategy significantly boosts performance, achieving a MAPE of 0.48%, and MAE and RMSE of 0.04 Ah and 0.06 Ah, respectively.

Subsequently, to test the model capacity to generalize, a new battery, aged at a more aggressive current of 50 A, was used to gather training data, and the model was then evaluated on 12 OCV-q curves taken from the original battery. Although the estimation errors rose slightly (e.g., MAPE to 0.84%), the results still indicate strong predictive performance across different datasets. These findings confirm the promise of using lightweight ML algorithms and carefully designed features to build generalized models of battery capacity, even when trained on data from different aging conditions.

Moving forward, our goal is to enhance the robustness and applicability of this technique. Specifically, future work will involve constructing a more diverse training set, incorporating OCV-q curves from multiple batteries (of the same type) to better capture inter-battery variability. Additionally, the impact of temperature on battery capacity will be thoroughly investigated and integrated into the model. By broadening the data collection scope and explicitly modeling temperature effects, we anticipate further improvements in reliability, paving the way toward more accurate and widely deployable battery SOH estimators.

References

[1] A. M. Divakaran, M. Minakshi, P. A. Bahri, S. Paul, P. Kumari, A. M. Divakaran, K. N. Manjunatha, Rational design on materials for developing next generation lithium-ion secondary battery, Progress in Solid State Chemistry 62 (2021) 100298.

[2] S. Barcellona, L. Piegari, Lithium Ion Battery Models and Parameter Identification Techniques, Energies 10 (2017) 2007.

[3] I. Bloom, B. Cole, J. Sohn, S. Jones, E. Polzin, V. Battaglia, G. Henriksen, C. Motloch, R. Richardson, T. Unkelhaeuser, D. Ingersoll, H. Case, An accelerated calendar and cycle life study of Li-ion cells, Journal of Power Sources 101 (2001) 238–247.

[4] R. Wright, C. Motloch, J. Belt, J. Christophersen, C. Ho, R. Richardson, I. Bloom, S. Jones, V. Battaglia, G. Henriksen, T. Unkelhaeuser, D. Ingersoll, H. Case, S. Rogers, R. Sutula, Calendar- and cycle-life studies of advanced technology development program generation 1 lithium-ion batteries, Journal of Power Sources 110 (2002) 445–470.

[5] S. Ma, M. Jiang, P. Tao, C. Song, J. Wu, J. Wang, T. Deng, W. Shang, Temperature effect and thermal impact in lithium-ion batteries: A review, Progress in Natural Science: Materials International 28 (2018) 653–666.

[6] L. Wang, D. Lu, Q. Liu, L. Liu, X. Zhao, State of charge estimation for lifepo4 battery via dual extended kalman filter and charging voltage curve, Electrochimica Acta 296 (2019) 1009–1017.

[7] A. Klintberg, E. Klintberg, B. Fridholm, H. Kuusisto, T. Wik, Statistical modeling of ocv-curves for aged battery cells, IFAC-PapersOnLine 50 (2017) 2164–2168. 20th IFAC World Congress.

[8] S. Tong, M. P. Klein, J. W. Park, On-line optimization of battery open circuit voltage for improved state-of-charge and state-of-health estimation, Journal of Power Sources 293 (2015) 416–428.

[9] Y. Xing, W. He, M. Pecht, K. L. Tsui, State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures, Applied Energy 113 (2014) 106–115.

[10] A. Farmann, D. U. Sauer, A study on the dependency of the open-circuit voltage on temperature and actual aging state of lithium-ion batteries, Journal of Power Sources 347 (2017) 1–13.

[11] V. Knap, D.-I. Stroe, Effects of open-circuit voltage tests and models on state-of-charge estimation for batteries in highly variable temperature environments: Study case nano-satellites, Journal of Power Sources 498 (2021) 229913.

[12] S. Sundaresan, B. Devabattini, P. Kumar, K. Pattipati, B. Balasingam, Tabular Open Circuit Voltage Modelling of Li-Ion Batteries for Robust SOC Estimation, Energies 15 (2022) 9142.

[13] H. He, R. Xiong, X. Zhang, F. Sun, J. Fan, State-of-Charge Estimation of the Lithium-Ion Battery Using an Adaptive Extended Kalman Filter Based on an Improved Thevenin Model, IEEE Transactions on Vehicular Technology 60 (2011) 1461–1469.

[14] R. Xiong, H. He, H. Guo, Y. Ding, Modeling for Lithium-Ion Battery used in Electric Vehicles, Procedia Engineering 15 (2011) 2869–2874.

[15] S. Barcellona, L. Codecasa, S. Colnago, L. Piegari, Cycle Aging Effect on the Open Circuit Voltage of Lithium-Ion Battery, International Conference on Electrical Systems for Aircraft, Railway, Ship Propulsion and Road Vehicles and International Transportation Electrification Conference (ESARS-ITEC) (2023).

[16] C. Zhang, J. Jiang, L. Zhang, S. Liu, L. Wang, P. Loh, A Generalized SOC-OCV Model for Lithium-Ion Batteries and the SOC Estimation for LNMCO Battery, Energies 9 (2016) 900.

[17] H. Rauf, M. Khalid, N. Arshad, Machine learning in state of health and remaining useful life estimation: Theoretical and technological development in battery degradation modelling, Renewable and Sustainable Energy Reviews 156 (2022) 111903.

[18] D. Roman, S. Saxena, V. Robu, M. Pecht, D. Flynn, Machine learning pipeline for battery state-of-health estimation, Nature Machine Intelligence 3 (2021) 447–456.

[19] M. Cao, T. Zhang, W. Jia, Y. Liu, A deep belief network approach to remaining capacity estimation for lithium-ion batteries based on charging process features, Journal of Energy Storage 48 (2022) 103825.

[20] Z. Cui, C. Wang, X. Gao, S. Tian, State of health estimation for lithium-ion battery based on the coupling-loop nonlinear autoregressive with exogenous inputs neural network, Electrochimica Acta 393 (2021) 139047.

[21] E. Petkovski, I. Marri, L. Cristaldi, M. Faifer, State of health estimation procedure for lithium-ion batteries using partial discharge data and support vector regression, Energies 17 (2023) 206.

[22] K. Severson, P. Attia, N. Jin, et al., Data-driven prediction of battery cycle life before capacity degradation, Nature Energy 4 (2019) 383–391.

[23] R. Bustos, S. A. Gadsden, P. Malysz, M. Al-Shabi, S. Mahmud, Health monitoring of lithium-ion batteries using dual filters, Energies 15 (2022).

[24] J. Kim, B. H. Cho, State-of-charge estimation and state-of-health prediction of a li-ion degraded battery based on an ekf combined with a per-unit system, IEEE Transactions on Vehicular Technology 60 (2011) 4249–4260.

[25] I.-S. Kim, A technique for estimating the state of health of lithium batteries through a dual-sliding-mode observer, IEEE Transactions on Power Electronics 25 (2010) 1013–1022.

[26] X. Xu, Z. Xu, T. Wang, J. Xu, L. Pei, Open-circuit voltage curve reconstruction for degrading lithium-ion batteries utilizing discrete curve fragments from an online dataset, Journal of Energy Storage 56 (2022) 106003.

[27] J. Sun, Y. Tang, J. Ye, T. Jiang, S. Chen, S. Qiu, A novel capacity and initial discharge electric quantity estimation method for lifepo4 battery pack based on ocv curve partial reconstruction, Energy 243 (2022) 122882.

[28] X. Qiao, Z. Wang, E. Hou, G. Liu, Y. Cai, Online estimation of open circuit voltage based on extended kalman filter with self-evaluation criterion, Energies 15 (2022).

[29] K. Zhang, R. Xiong, Q. Li, C. Chen, J. Tian, W. Shen, A novel pseudo-open-circuit voltage modeling method for accurate state-of-charge estimation of lifepo4 batteries, Applied Energy 347 (2023) 121406.

[30] D. Andre, M. Meiler, K. Steiner, C. Wimmer, T. Soczka-Guth, D. Sauer, Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. i. experimental investigation, Journal of Power Sources 196 (2011) 5334–5341.

[31] S. Barcellona, S. Grillo, L. Piegari, A simple battery model for EV range prediction: Theory and experimental validation, 2016 International Conference on Electrical Systems for Aircraft, Railway, Ship Propulsion and Road Vehicles and International Transportation Electrification Conference (ESARS-ITEC) (2016) 1–7.

[32] S. Barcellona, M. Brenna, F. Foiadelli, M. Longo, L. Piegari, Analysis of Ageing Effect on Li-Polymer Batteries, The Scientific World Journal 2015 (2015) 1–8.

[33] S. Barcellona, L. Piegari, Effect of current on cycle aging of lithium ion batteries, Journal of Energy Storage 29 (2020).

[34] C. Bishop, Pattern recognition and machine learning, Springer 2 (2006) 531–537.

[35] E. D. Kostopoulos, G. C. Spyropoulos, J. K. Kaldellis, Real-world study for the optimal charging of electric vehicles, Energy Reports 6 (2020) 418–426.

[36] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, A comparative analysis of gradient boosting algorithms, Artificial Intelligence Review 54 (2021) 1937–1967.

[37] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[38] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, Lightgbm: A highly efficient gradient boosting decision tree, Advances in neural information processing systems 30 (2017).

[39] L. Grinsztajn, E. Oyallon, G. Varoquaux, Why do tree-based models still outperform deep learning on typical tabular data?, Advances in neural information processing systems 35 (2022) 507–520.

[40] L. S. Shapley, A value for n-person games, Contribution to the Theory of Games 2 (1953).

[41] E. Štrumbelj, I. Kononenko, Explaining prediction models and individual predictions with feature contributions, Knowledge and information systems 41 (2014) 647–665.

[42] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable ai for trees, Nature machine intelligence 2 (2020) 56–67.