Development of a Global Model for Estimating the SOH of Lithium-Ion Batteries from Cycling Data

1st Mohamed Atrani
National Higher School for Computer Science and Systems Analysis (ENSIAS)
Mohammed V University in Rabat
Rabat, Morocco
email: mohamedatrani7@gmail.com

2nd Nasser Tamou
National Higher School for Computer Science and Systems Analysis (ENSIAS)
Mohammed V University in Rabat
Rabat, Morocco
email: tamounasser@gmail.com

***Abstract*—This article presents the development of a predictive model for estimating the State of Health (SOH) of lithium-ion batteries. The model uses key parameters such as cell voltage, nominal capacity, charge/discharge rate, ambient temperature, and the number of charge/discharge cycles. A testing procedure has been established, and aging tests were conducted on three battery chemistries: Lithium Cobalt Oxide (LiCoO), Nickel Manganese Cobalt (NMC), and Lithium Iron Phosphate (LFP), with a particular focus on the LFP battery to collect data and train three machine learning algorithms: Linear Regression (LR), Random Forest (RF), and Neural Networks (NN). The objective is to identify the most accurate model for predicting the SOH of different lithium-ion batteries. Ultimately, the RF model can estimate SOH by inputting these parameters and the number of charge/discharge cycles, providing a reliable approximation.**

***Index Terms*—State of Health (SOH), lithium-ion batteries, aging test procedure, machine learning.**

I. Introduction

The assessment of the State of Health (SOH) of lithium-ion batteries is a critical area of research, particularly as their use in electric vehicles and energy storage systems continues to grow [1], [2]. Numerous studies have investigated various methods for estimating SOH, including physical modeling approaches and machine learning techniques [3]. However, these methods often have limitations, such as dependence on specific data sets.

The primary objective of this study is to develop a predictive model capable of estimating the SOH of lithium-ion batteries using key parameters, including cell voltage, nominal capacity, charge/discharge rate, ambient temperature, and the number of charge/discharge cycles. The goal is to determine which machine learning model provides the highest accuracy in predicting the SOH for different lithium-ion chemistries, specifically Lithium Cobalt Oxide (LiCoO), Nickel Manganese Cobalt (NMC), and Lithium Iron Phosphate (LFP) [4].

To achieve this, a series of aging tests were conducted on the three chemistries mentioned, with a particular emphasis on LFP. The data collected were used to train three machine learning algorithms: Linear Regression (LR), Random Forest (RF), and Neural Networks (NN). This process aims to identify the most accurate model for delivering reliable real-time SOH estimates.

In practice, the model provides real-time estimates of the SOH, enabling efficient monitoring of battery degradation and adjustment of testing strategies. It also enhances prediction accuracy by correcting discrepancies, making it a versatile tool for improving SOH estimation. This model can be integrated into battery management systems (BMS) for electric vehicles and energy storage systems.

An introduction is presented in Section 1. The study scope is outlined in Section 2. The test procedure for the development of the SOH model is detailed in Section 3. Section 4 focuses on tuning data and the learning algorithms for estimating the SOH. Finally, a conclusion is provided in Section 5.

II. Study Scope

This section provides an overview of the different types of batteries studied, the basic battery model used, and the mechanisms that lead to battery aging, along with methods for calculating the State of Health (SOH) in this study [2].

A. Battery Types and Simple Battery Model

The battery types used and examined in this study include Lithium-Cobalt-Oxide, which offers high energy density but is expensive and more susceptible to safety concerns [3]. Nickel-Manganese-Cobalt provides a good balance between cost, safety, and performance, making it popular in electric vehicles [3]. Lastly, Lithium-Iron-Phosphate is known for high safety and longevity, though it has slightly lower energy density [3].

The simple battery model used in this study to express the relationship between voltage and current is represented by the following equation:

$$V = E - R \cdot I$$

Where V represents the measured voltage (in volts), E the electromotive force or open-circuit voltage (in volts), R the internal resistance (in ohms), and I the current flow (in amperes). This equation helps describe the behavior of the battery during operation by accounting for voltage drops due to internal resistance.

B. Battery Aging

Batteries degrade over time due to several factors. And the aging process can be divided into different phases and types. The aging phases include the Beginning of Life (BOL), when the battery is at 100% of its nominal capacity; the Middle of Life (MOL), which occurs when the battery's capacity has dropped to 80% of its nominal value; and the End of Life (EOL), reached when the battery's capacity ranges between 70% and 80% of its original capacity.

Battery aging factors considered in this study include both storage and operation. These factors are calendar aging, which refers to the natural degradation of the battery over time, even when it is not in use, and cycle aging, which is the gradual degradation caused by repeated charge and discharge cycles [2]; high charge/discharge currents, which accelerate aging by increasing the battery's internal resistance; and temperature effects, where elevated temperatures accelerate unwanted chemical reactions, leading to faster degradation [2].

C. Methods for estimating State of Health (SOH)

Two common methods for estimating State of Health (SOH) are used:

Capacity-Based Estimation

This method estimates the SOH based on the battery's current capacity compared to its nominal capacity:

$$\text{SOH}_{\text{capacity}} = \left(\frac{\text{Current Capacity}}{\text{Nominal Capacity}}\right) \times 100$$

Internal Resistance-Based Estimation

This method evaluates the SOH based on changes in the battery's internal resistance:

$$\text{SOH}_{\text{resistance}} = \left(\frac{R_{\text{reference}}}{R_{\text{current}}}\right) \times 100$$

Both methods provide an indication of the battery's health, though they rely on different measurable parameters.

III. Test Procedure for the Development of the SOH Model

In this section, we focus on establishing a rigorous testing procedure to evaluate the aging of lithium iron phosphate batteries specifically designed for hybrid vehicles. This procedure is also applicable to other types of lithium-ion batteries. It involves subjecting the battery to 100 charge and discharge cycles. After each cycle, we perform a check of the battery's remaining capacity as well as its internal resistance. These measurements are conducted under different temperature conditions to assess the impact of thermal variations on the battery's performance. This approach aligns with the standards established in the literature, such as ISO 12405 and IEC 62660, thereby ensuring the reliability and validity of the obtained results. [4] The key battery's characteristics used in this study are summarized in Table I :

TABLE I
KEY PARAMETERS OF THE BATTERY USED IN THIS STUDY

Parameter	Value
Nominal cell voltage	3.65 V
Nominal pack voltage	350 V
Cell voltage limit @25°C	2.8 V / 4.25 V
Pack voltage limit @25°C	269 V / 408 V
Cell temperature limit	-30°C / +55°C
Storage temperature limit	-20°C / +35°C
Nominal integrated cell capacity	60 Ah
Nominal usable cell capacity	51 Ah
Nominal integrated cell energy	219 Wh
Nominal usable cell energy	186 Wh
Nominal integrated pack capacity	60 Ah
Nominal usable pack capacity	51 Ah
Nominal integrated pack energy	21.0 kWh
Nominal usable pack energy	17.8 kWh
Usable SOC range	10% - 95%

A. Battery Check-up

To monitor the evolution of battery aging, it is essential to conduct regular check-ups by measuring electrical capacity and internal resistance. These tests are referred to as Reference Performance Test (RPT).

a) Capacity Check-up: The capacity check-up consists of performing a full nominal discharge at a rate of C/2 to measure the current electrical capacity of the battery. The Capacity (C) can be calculated using the following equation:

$$C = \int I(t)\,\mathrm{d}t$$

where:

- C is the capacity in ampere-hours (Ah), - $I(t)$ is the current in amperes (A), - $\mathrm{d}t$ is the time element.

Steps for measuring the capacity of the battery in the test bench are detailed in Table II.

TABLE II
STEPS FOR MEASURING BATTERY CAPACITY

Step	Phase	Conditions
1	Preparation	Balance the pack
2	Discharge	Discharge at 1C, cutoff voltage 2.8 V
3	Rest	Temperature at 25°C
4	Recharge	Charge at C/2 with BMS limitations
5	Rest	Temperature at 25°C
6	Discharge	Discharge at 1C down to 2.8 V
7	Rest	Temperature at 25°C
8	Discharge	Discharge at C/20
9	Rest	Temperature at 25°C
10	Repetition	Repeat steps 5 to 7 twice

b) Direct Current Internal Resistanc (DCIR) Check-up: The internal resistance check-up uses the Hybrid Pulse Power Characterization (HPPC) profile to measure the battery's internal resistance. The cell internal resistance can be calculated using the following equation:

$$R_{\text{DCIR}} = \frac{\Delta V}{\Delta I}$$

where:

- R_{DCIR} is the internal resistance in ohms (Ω), - ΔV is the change in voltage (V) during the pulse, - ΔI is the change in current (A) during the pulse.

The steps for measuring internal resistance are detailed in Table III.

TABLE III
STEPS FOR MEASURING INTERNAL RESISTANCE

Step	Phase	Conditions
1	Temperature set	Temperature 25°C
2	Charge	C/2 with BMS limitations, Temperature 25°C
3	Discharge	1C-CC
4	Rest (Relay Open)	Temperature 25°C
5	Discharge	3C-CC
6	Rest (Relay Open)	Temperature 25°C
7	Repetition	Repeat steps 1 to 6, 3 times

B. Cycling Test

Battery cycling involves subjecting the battery to multiple charge and discharge cycles under strict conditions. The detailed procedure for battery cycling is summarized in Table IV.

TABLE IV
STEPS FOR BATTERY CYCLING

Step	Phase	Conditions
1	Temperature set	Set test temperature to 25°C
2	Rest	-
3	Recharge	Charge at C/2 with BMS limitations
4	Rest	-
5	Discharge	Discharge at 1C-CC, down to 3.35 V
5bis	Memory	Relay open, BMS ON/OFF
6	Check-up	Capacity and DCIR measurement (RPT)

C. Test Sequence

- **RPT0 (Initial Test)**: Identify the initial capacity and internal resistance.
- **Cycling 1**: 100 full charge/discharge cycles.
- **RPT1**: Check-up after 100 cycles.
- **Cycling 2**: Another 100 cycles.
- **RPT2**: Check-up after 200 cycles.
- Test continues until 75% of initial capacity is reached (EOL).

D. Experimental Data from Battery Aging Test

The results obtained from the battery aging tests performed on a test bench allow for the plotting of aging curves that highlight the evolution of capacity and internal resistance over cycles.

The Open Circuit Voltage (OCV) curve is a critical metric that illustrates the relationship between the battery's voltage and its state of charge (SOC) when no current is flowing in the circuit [9]. This curve is essential for estimating the SOC without directly measuring the current. To obtain the OCV curve, the battery is allowed to rest after each charge or discharge step, giving the voltage time to stabilize. This curve is presented in Fig 1.

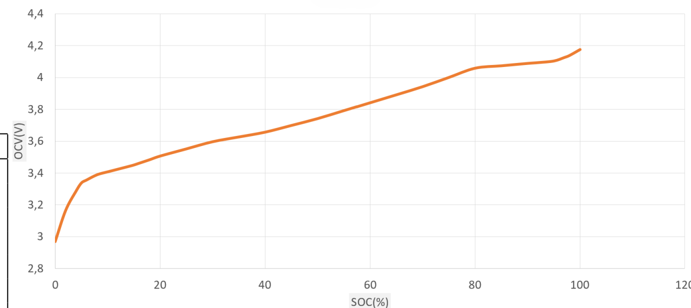

Fig. 1. OCV curve of the studied battery

The capacity curve, shown in Fig 2, represents the evolution of the battery's capacity as a function of charge/discharge cycles. Over time, as the battery ages and is cycled more, its ability to store energy decreases. This capacity loss reflects the degradation of active materials inside the cells, providing important information about the battery's life expectancy, degradation rate, and its capacity to meet energy storage needs during its operational lifecycle.

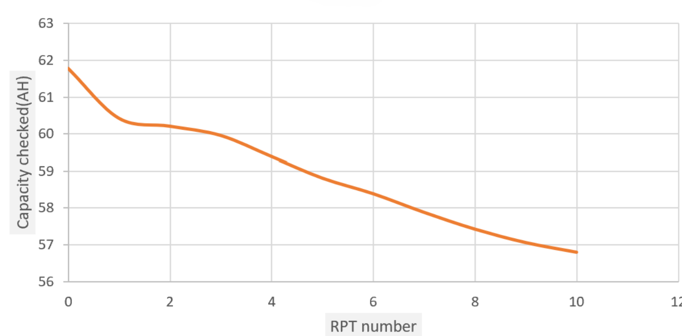

Fig. 2. Capacity evolution as a function of cycling

Similarly, the internal resistance curve, depicted in Fig 3, traces the battery's internal resistance over cycles. As the battery ages, its internal resistance typically increases, which negatively affects performance and efficiency. Higher internal resistance leads to increased heat generation and power loss during charge/discharge cycles, ultimately reducing the battery's ability to supply power efficiently.

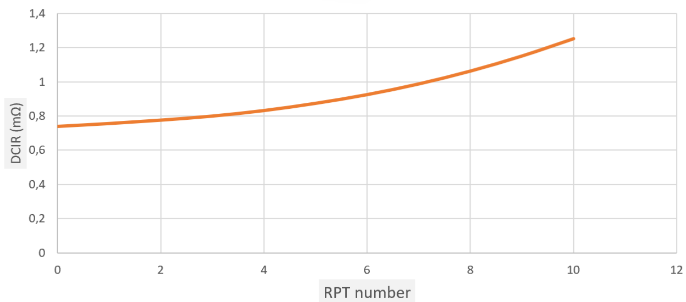

Fig. 3. Evolution of internal resistance as a function of cycling

Finally, the aging graph of the LFP battery, displayed in Fig 4, illustrates the capacity degradation of an LFP battery dedicated to a Plug-in Hybrid Electric Vehicle (PHEV) during charge and discharge cycles. This graph shows the typical aging profile for this battery type [10], where initial capacity loss occurs more rapidly before stabilizing and slowing down. After numerous charge/discharge cycles, the battery retains around 92% of its original capacity, indicating a satisfactory lifespan over several years of use.

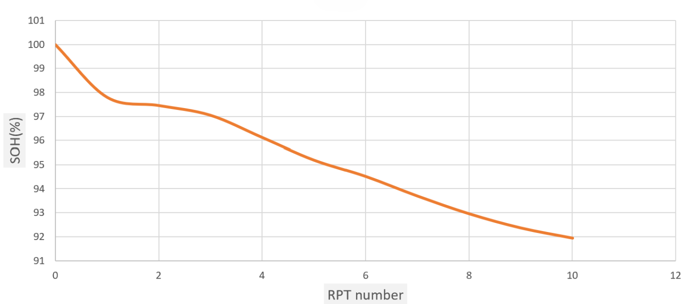

Fig. 4. Capacity degradation of an LFP battery over cycles

IV. Data, Learning Algorithm Implementation, and SOH Estimation Results

The objective of this study is to develop a model capable of estimating the State of Health of lithium-ion batteries. The SOH measures the health status of a battery by comparing its current energy storage capacity to its initial nominal capacity. This model is based on cycle data collected from tests conducted in specialized centers, following the procedure outlined in the previous section. These data were used to train three machine learning algorithms to build, test, and validate the model, in order to determine the best among them. This model provides a better understanding of the gradual degradation of batteries over charge and discharge cycles.

A. Input and Output Data

The main input data used for training and testing the model include the nominal voltage of the battery cell (in volts), the nominal capacity of the battery (in Ah), the charge/discharge rate (C-rate), which is the ratio of the charge/discharge current to the nominal capacity of the battery, the ambient temperature (in °C) during the test, and the number of complete charge and discharge cycles performed on the battery.

The main output data is the State of Health, expressed as a percentage. The SOH is a critical indicator of a battery's performance, as it quantifies the degradation of energy capacity over time. An SOH of 100% means that the battery is in perfect condition, while a lower value indicates some degradation.

Table V below presents a sample of the training data used to develop the model.

B. Learning Models and Algorithm

To estimate the State of Health of batteries, several machine learning algorithms were evaluated to determine which one offers the best accuracy and reliability. The three main models explored include Linear Regression (LR), which identifies linear relationships between input data and output (SOH) [11]. While it is simple and fast, it may lack accuracy in cases where relationships are nonlinear. Random Forest (RF) is another model based on an ensemble of decision trees, which excels at capturing nonlinear and complex relationships, making it an excellent choice for SOH prediction [12]. Additionally, it is robust against variations in input data. Neural Networks (NN) were also explored, as they can learn complex structures in the data and are particularly useful when nonlinear and hidden relationships between variables are crucial [13]. However, they may require more data and fine-tuning to avoid overfitting issues.

The algorithm for the program consists of several key steps. First, the battery data is loaded from an Excel file. Then, relevant columns are selected for analysis, such as nominal voltage, nominal capacity, C-rate, ambient temperature, number of charge-discharge cycles, and SOH. The independent features (X) are separated from the target variable (y). Two new nonlinear features are added to the dataset: the square of the number of charge-discharge cycles (cycles_squared) and the ratio between nominal capacity and C-rate (capacity_rate_ratio).

Next, the data is split into training (80%) and testing (20%) sets. The three machine learning models, Random Forest, Linear Regression, and Neural Network, are then created. Each model is trained on the training data and subsequently evaluated on the test data by calculating performance metrics such as Mean Squared Error (MSE), Coefficient of Determination (R^2), and Cross-Validation.

The prediction curves of each model are plotted and compared to the actual SOH values. Lastly, predictions are made for two specific battery examples not present in the training data, and the SOH for these examples is predicted using the three models. The final results, including prediction errors and scores, are displayed for comparison.

TABLE V
SAMPLE OF INPUT AND OUTPUT DATA FOR MODEL DEVELOPMENT

Nominal Voltage (V)	Nominal Capacity (Ah)	C-rate	Ambient Temperature (°C)	Cycles	State of Health (%)
3.65	60	60	35	0	100
3.65	60	60	35	1	97.80
3.65	60	60	35	2	97.46
3.65	60	60	35	3	97.05
3.65	60	60	35	4	96.13
3.65	60	60	35	5	95.20
3.65	60	60	35	6	94.52
3.65	60	60	35	7	93.70
3.65	60	60	35	8	92.96
3.65	60	60	35	9	92.37
3.65	60	60	35	10	91.95
3.65	60	60	25	0	100
3.65	60	60	25	1	98.44
3.65	60	60	25	2	98.13
3.65	60	60	25	3	97.38
3.65	60	60	25	4	96.78
3.65	242	121	35	0	100
3.65	242	121	35	1	97.69
3.65	242	121	35	2	96.49
3.65	242	121	35	3	95.81
3.65	139	69.5	35	0	100
3.65	139	69.5	35	1	98.84
3.65	139	69.5	35	2	97.19
3.65	139	69.5	35	3	97.19

C. Results and Discussion

The three models Linear Regression, Random Forest, and Neural Networks were tested on two distinct battery examples not included in the training data. Table VI and Table VII present the characteristics of the two batteries and the corresponding predictive State of Health results obtained from each algorithm, respectively. Additionally, the three models were tested on 9 test samples randomly taken from the batteries in the training data. Table VIII shows the predicted State of Health results for these samples from the three models

TABLE VI
BATTERY CHARACTERISTICS FOR EXAMPLE 1 AND EXAMPLE 2

Characteristic	Example 1	Example 2
Nominal Voltage (V)	3.65	3.65
Nominal Capacity (Ah)	139	139
C-rate	69.5	69.5
Ambient Temperature (°C)	35	35
Charge/Discharge Cycles	0	3

TABLE VII
SOH PREDICTIONS FOR EXAMPLE 1 AND EXAMPLE 2

Model	SOH (%) for Example 1	SOH (%) for Example 2
RF	99.35	96.95
LR	99.22	97.37
NN	103.81	90.43

TABLE VIII
COMPARISON OF MODEL PREDICTIONS

Sample	Actual Values	RF	LR	NN
1	93.5	93.8	93.5	93.0
2	95.8	95.9	95.5	99.0
3	100.0	99.2	98.2	105.0
4	100.0	99.1	98.7	102.0
5	100.0	99.2	99.0	97.2
6	98.5	98.5	98.3	97.2
7	97.0	97.5	97.5	97.5
8	98.8	98.5	98.8	97.3
9	100.0	99.3	99.8	97.2

The following is an analysis of the results obtained by the three algorithms tested on the battery examples and samples.

Random Forest (RF): The predictions for Example 1 (99.35%) and Example 2 (96.95%) demonstrate that the Random Forest model effectively captures non-linear patterns within the battery data. As shown in Table VIII, the RF model remains consistently close to the actual values, particularly in samples 1, 3, 5, and 9. This alignment highlights the model's robustness, with only minimal deviations from the actual SOH values, reinforcing its reliability for predicting SOH in both cases.

Linear Regression (LR): Linear Regression yielded predictions of 99.22% for Example 1 and 97.37% for Example 2, which indicate solid but slightly less accurate performance compared to Random Forest. As seen in Table VIII, LR follows the general trend of the actual values but struggles with higher deviations, particularly in samples 3 and 4. This suggests that while LR performs well in simpler cases, it may not fully capture the complexities present in the battery's behavior, especially when non-linearity plays a significant role.

Neural Network (NN): The Neural Network predictions exhibit larger variability, with an overestimation of 103.81% for Example 1 and an underestimation of 90.43% for Example 2. Table VIII reflects this inconsistency, particularly in sample 3, where the NN model predicts a significant overestimation (105%), and sample 9, where it falls short of the actual value. This fluctuation highlights the potential need for further optimization of hyperparameters and model training to reduce overfitting and improve accuracy across diverse datasets.

Overall, the results indicate that Random Forest outperforms both Linear Regression and Neural Networks in terms of stability and accuracy across different battery characteristics.

We can estimate the SOH using the Random Forest equation as follows:

$$\text{SOH} \approx k_1 \cdot C_1 + k_2 \cdot C_2 + k_3 \cdot C_3 + k_4 \cdot C_4 + k_5 \cdot C_5 + k_6 \cdot C_6 + k_7 \cdot C_7$$

where the coefficients and characteristics are defined in the Table IX.

TABLE IX
COEFFICIENTS AND CHARACTERISTICS

Symbol	Value	Description
k_1	0.4934	Coefficient for charge_discharge_cycles
k_2	0.4358	Coefficient for cycles_squared
k_3	0.0367	Coefficient for ambient_temperature
k_4	0.0167	Coefficient for C-rate
k_5	0.0134	Coefficient for nominal_capacity
k_6	0.0039	Coefficient for capacity_rate_ratio
k_7	0.0000	Coefficient for nominal_voltage
C_1	-	Charge_discharge_cycles
C_2	-	Cycles_squared
C_3	-	Ambient_temperature
C_4	-	C-rate
C_5	-	Nominal_capacity
C_6	-	Capacity_rate_ratio
C_7	-	Nominal_voltage

V. CONCLUSION

In this study, we developed a robust model for estimating the State of Health of lithium-ion batteries, with a particular focus on Lithium Iron Phosphate batteries intended for hybrid vehicles. By utilizing key parameters such as cell voltage, nominal capacity, charge/discharge rate, ambient temperature, and the number of charge/discharge cycles, we established an effective predictive model.

Aging tests conducted on three types of batteries (LiCoO, NMC, and LFP) provided valuable data for training three algorithms: Linear Regression, Random Forest, and Neural Networks. The results demonstrate that the Random Forest model stands out for its ability to capture complex and nonlinear relationships in the data, making it the preferred option for accurate SOH predictions.

This model is not only applicable to any lithium-ion battery, but it can also be integrated into battery management systems for real-time predictions. Additionally, it can help correct convergent algorithms, improve battery usage strategies, and optimize preventive maintenance schedules. In summary, this work paves the way for significant advancements in battery management and sustainability, contributing to the efficiency of modern energy systems.

REFERENCES

[1] J. M. Tarascon and M. Armand, "Issues and challenges facing rechargeable lithium batteries," Nature, vol. 414, no. 6861, pp. 359–367, 2001.

[2] J. Vetter, P. Novák, M. R. Wagner, C. Veit, K. C. Möller, J. O. Besenhard, M. Winter, M. Wohlfahrt-Mehrens, C. Vogler, and A. Hammouche, "Ageing mechanisms in lithium-ion batteries," Journal of Power Sources, vol. 147, no. 1-2, pp. 269–281, 2005.

[3] N. Nitta, F. Wu, J. T. Lee, and G. Yushin, "Li-ion battery materials: present and future," Materials Today, vol. 18, no. 5, pp. 252–264, 2015.

[4] Xu, G., et al. (2019). "Thermal analysis and solutions for electric vehicle batteries: A review." *Journal of Energy Storage*, 24, 100788.

[5] Birkl, C. R., et al. (2017). "Degradation diagnostics for lithium ion cells." *Journal of Power Sources*, 341, 373-386.

[6] Miller, T. A., et al. (2018). "Lithium-ion battery hybrid pulse power characterization and its application to predict battery performance." *IEEE Transactions on Energy Conversion*, 33(1), 109-117.

[7] Hwang, J., et al. (2020). "Cycle life and performance of lithium-ion batteries under different charging conditions." *Journal of Power Sources*, 448, 227512.

[8] Zhang, Y., et al. (2021). "Performance evaluation of lithium-ion batteries using advanced testing protocols." *Electrochimica Acta*, 367, 137364.

[9] Lin, Y., et al. (2020). "Open Circuit Voltage Characterization of Lithium-Ion Batteries: Effects of Temperature and State of Charge." *IEEE Transactions on Power Electronics*, 35(7), 7167-7176.

[10] Li, H., et al. (2020). "Degradation Mechanisms of Lithium Iron Phosphate Batteries: A Review." *Journal of Power Sources*, 448, 227-245.

[11] Zhang, Y., et al. (2020). "A Review of Linear Regression for the Analysis of Battery State of Health." *Energies*, 13(2), 250.

[12] Breiman, L. (2001). "Random Forests." *Machine Learning*, 45(1), 5-32.

[13] Heaton, J. (2016). *Introduction to Neural Networks with Python*. Heaton Research, Inc.