

Article

Multi-Source Impedance and Discharge Feature Learning for Cross-Battery State-of-Health Estimation of Lithium-Ion Batteries

Syed Adil Sardar, Farhan Akhtar, Wajid Ali and Woo Young Kim *

Department of Electronic Engineering, Jeju National University, Jeju 63243, Republic of Korea; adilsardar83@stu.jejunu.ac.kr (S.A.S.)

* Correspondence: semigumi@jejunu.ac.kr

Abstract

Accurate state-of-health (SOH) estimation is essential for the reliable operation of lithium-ion batteries. However, predicting SOH for previously unseen batteries remains challenging because degradation behavior varies among cells. This study proposes a multi-source feature-learning framework that combines electrochemical impedance spectroscopy (EIS), discharge-profile, and aging-related information for cross-battery SOH estimation. EIS and discharge data from 34 batteries in the National Aeronautics and Space Administration (NASA) battery aging dataset are processed to construct 1830 matched impedance-SOH samples. A total of 101 features are extracted from raw and rectified impedance spectra, resampled impedance points, NASA-provided impedance parameters, discharge profiles, and cycle-related information. Random Forest (RF), Extra Trees (ET), Gradient Boosting (GB), Histogram Gradient Boosting (HGB), and Extreme Gradient Boosting (XGBoost) models are evaluated using random sample splitting and strict battery-wise validation. Under random validation, GB achieves the best performance, with a coefficient of determination (R^2) of 0.9788. Under strict battery-wise validation, ET achieves a mean absolute error (MAE) of 4.9131 percentage points, a root mean square error (RMSE) of 7.3664 percentage points, and an R^2 of 0.7855. The performance difference between the two validation strategies demonstrates the importance of battery-grouped evaluation when assessing generalization to unseen cells. Overall, the results indicate that combining impedance- and discharge-derived information provides a promising basis for cross-battery SOH estimation, although further leakage-controlled validation across broader operating conditions is required.

Keywords: lithium-ion battery; state-of-health estimation; electrochemical impedance spectroscopy; discharge features; cross-battery prediction; ensemble learning

Academic Editor: Odne S. Burheim

Received: 22 June 2026

Revised: 27 July 2026

Accepted: 5 August 2026

Published: 6 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Lithium-ion batteries (LIBs) have become a leading energy-storage technology for electric vehicles, renewable-energy systems, portable electronics, and smart-grid applications [1,2]. Their high energy density, long cycle life, low self-discharge rate, and rapid-charging capability make them suitable for a wide range of modern energy applications [3,4]. Nevertheless, repeated charge–discharge cycling causes electrochemical and structural degradation, resulting in gradual capacity loss and increased internal resistance [5–

7]. These degradation processes reduce power capability and energy efficiency, shorten service life, and may increase safety risks [6,8]. Consequently, accurate state-of-health (SOH) estimation is an essential function of a battery management system (BMS). Reliable SOH information can support failure prevention, predictive maintenance, operational safety, and efficient utilization of battery energy-storage systems [9–11].

SOH-estimation methods are generally classified as model-based, data-driven, or hybrid approaches [12–15]. Model-based methods, including equivalent-circuit models, electrochemical models, and Kalman filter-based estimators, can provide physically meaningful representations of battery behavior [13,16–18]. However, their accuracy depends on appropriate model structures, reliable parameter identification, and knowledge of operating conditions. These requirements are difficult to maintain because battery parameters vary with aging, temperature, state of charge, and usage history [19–21]. Data-driven methods, particularly machine learning (ML) and deep learning (DL), can identify non-linear degradation patterns directly from measured data and have therefore received considerable attention [14,22,23]. Existing studies have used voltage, current, temperature, capacity, and cycle-related indicators for SOH estimation. Nevertheless, these externally measured variables may not fully characterize the internal electrochemical processes associated with battery degradation [22,24–26].

EIS provides complementary information about the internal electrochemical behavior of LIBs. Impedance spectra reflect contributions from ohmic resistance, charge-transfer processes, interfacial polarization, and diffusion-related behavior across different frequency regions [13,25,27,28]. Because these responses evolve during degradation, EIS-derived indicators can support battery-health assessment [29,30]. However, some EIS-based studies rely primarily on a limited number of fitted parameters, such as electrolyte resistance (R_e) and charge-transfer resistance (R_{ct}), without fully utilizing changes in the impedance-spectrum shape [31–33]. Moreover, many data-driven studies employ random sample splitting, through which measurements from the same battery may appear in both the training and test sets [34–36]. This sample overlap can produce optimistic results because a model may learn battery-specific degradation patterns. Battery-wise validation, in which complete batteries are reserved for testing, therefore provides a more rigorous assessment of generalization to previously unseen cells [37–40].

To address these limitations, this study proposes a multi-source feature-learning framework for cross-battery SOH estimation. The framework integrates statistical and shape-based descriptors from raw and rectified impedance spectra, resampled impedance points, NASA-provided impedance parameters, discharge-profile indicators, and cycle-related information. A dataset comprising 1830 matched impedance–SOH samples from 34 NASA batteries is constructed, from which 101 input features are extracted. The performances of RF, ET, GB, HGB, and XGBoost are evaluated using both random sample splitting and strict battery-wise validation. Unlike random validation, the battery-wise strategy reserves complete batteries for testing and therefore provides a more rigorous assessment of generalization to previously unseen cells. Under strict battery-wise validation, ET achieves an MAE of 4.9131 percentage points, an RMSE of 7.3664 percentage points, and an R^2 of 0.7855. These findings demonstrate the potential of combining impedance- and discharge-derived information for cross-battery SOH estimation while emphasizing the importance of battery-grouped validation [41–43].

2. Dataset and Preprocessing

2.1. NASA Lithium-Ion Battery Aging Dataset

The NASA battery aging dataset was used to develop and evaluate the proposed SOH-estimation framework [44,45]. The dataset contains charge, discharge, and EIS measurements collected from LIBs under accelerated aging conditions. During charging, the

cells were operated using a constant-current–constant-voltage protocol until the terminal voltage reached 4.2 V. During discharge, the cells were subjected to constant-current loading until their specified cutoff voltages were reached. Repeated cycling resulted in progressive capacity degradation, while the EIS measurements captured corresponding changes in the cells' internal electrochemical behavior.

Data from 34 unique batteries were successfully processed, producing 1830 matched impedance–SOH samples. Because the EIS and discharge measurements were not always recorded at identical cycle positions, each EIS measurement was paired with the most recent preceding discharge-capacity measurement from the same battery. The corresponding capacity-based SOH value was then assigned as the target label for that EIS sample. This approach ensured chronological matching without using future capacity measurements. The battery IDs, experimental conditions, sample counts, and SOH ranges should be reported separately to clarify the composition of the processed dataset.

Representative charge–discharge voltage curves for selected batteries at different cycles are provided in Figure S1. The selected charge and discharge cycle numbers are summarized in Table S3. These curves show the evolution of voltage behavior during cycling and provide additional context for the discharge-profile features used in the proposed framework, including voltage statistics, voltage drop, discharge duration, and temperature-related indicators.

2.2. SOH Calculation

The SOH of each battery was calculated from its measured discharge capacity relative to the nominal cell capacity [46,47]. SOH was calculated as:

$$SOH (\%) = 100 \times \min\left(\frac{C_{\text{measured}}}{C_{\text{nominal}}}, 1\right) \quad (1)$$

where C_{measured} is the measured discharge capacity obtained from the discharge cycle, and C_{nominal} is the nominal battery capacity, set to 2.0 Ah in this study. The minimum operator limits the calculated SOH to 100%.

In experimental battery datasets, measured discharge capacity can fluctuate around the nominal capacity due to measurement uncertainty, temperature variation, rest effects, short-term capacity recovery, and cell-to-cell variability. Therefore, capacity values slightly higher than the nominal capacity do not necessarily indicate a physically meaningful SOH greater than 100%. In this study, such values were not removed only because they exceeded the nominal capacity; instead, the calculated SOH values above 100% were capped at 100% using Equation (1). Only records with missing, incomplete, or physically inconsistent capacity information were excluded during preprocessing.

This capping procedure may compress early-life samples near 100% SOH and can slightly influence the SOH-label distribution. In the processed dataset, six matched impedance–SOH samples, corresponding to approximately 0.33% of the 1830 samples, had measured discharge capacity values higher than the nominal capacity and were capped at 100% SOH. The details of these capped samples are provided in Table S2. Only records with missing, incomplete, or physically inconsistent capacity information were removed during preprocessing. Therefore, the model results should be interpreted with this preprocessing choice in mind. Representative SOH degradation trajectories for selected batteries are presented in Figure 1.

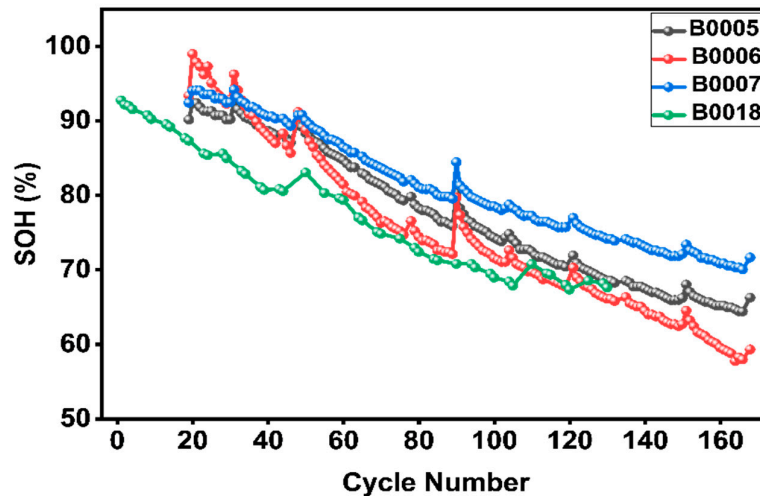


Figure 1. SOH degradation curves of representative lithium-ion batteries.

2.3. Data Cleaning and Duplicate Handling

Some battery IDs were identified in more than one MATLAB (R2026a Student Edition, MathWorks, Natick, MA, USA) data file. To prevent duplicate observations and unintended data repetition, only one source file was retained for each battery. When multiple files contained the same battery ID, the file providing the largest number of complete and usable impedance–SOH samples was selected. Records containing missing impedance arrays, invalid complex values, unavailable discharge capacities, or incomplete measurement information were excluded. After duplicate removal and data-quality screening, the final dataset contained 1830 matched samples from 34 unique batteries. A summary of the processed dataset is provided in Table 1.

Table 1. Summary of the processed dataset.

Parameter	Value
Data source	NASA battery aging dataset
Batteries processed	34
Total samples	1830
Input features	101
Target variable	SOH (%)
Nominal capacity	2.0 Ah
Sample-matching method	Each EIS measurement matched to the most recent preceding discharge measurement
Validation methods	Random split, battery-wise split, repeated battery-wise split

3. Proposed Methodology

3.1. Overall Framework

The proposed framework combines impedance-, discharge-, and aging-related information for cross-battery SOH estimation. As illustrated in Figure 2, the framework consists of six main stages: dataset acquisition, measurement extraction, SOH calculation and chronological label matching, multi-source feature extraction, feature integration, and ML-based prediction and validation. First, discharge and EIS measurements are extracted from the NASA MATLAB files. Each EIS record is then matched with the most recent preceding discharge-capacity measurement to assign its SOH label. Statistical and shape-based impedance descriptors, resampled impedance points, NASA-provided impedance

parameters, discharge-profile indicators, and cycle-related variables are subsequently extracted and integrated into a unified feature matrix. Finally, the developed models are evaluated using random sample splitting, strict battery-wise validation, and repeated battery-wise validation. Prediction performance is assessed using MAE, RMSE, and R^2 .

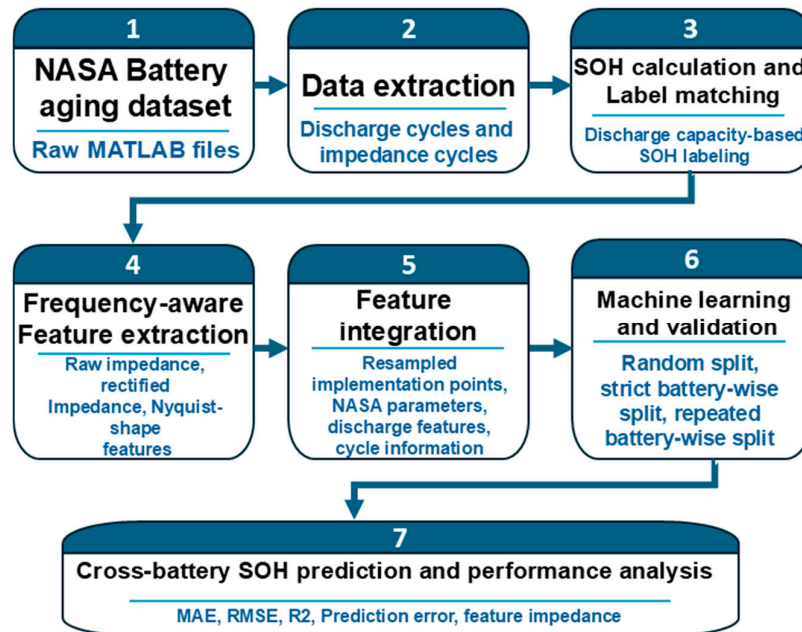


Figure 2. Overall workflow of the proposed multi-source feature-learning framework for cross-battery SOH estimation.

3.2. Impedance Feature Extraction

Each EIS measurement is represented as a complex-valued impedance spectrum containing real and imaginary components. The real component reflects resistive contributions, whereas the imaginary component represents frequency-dependent capacitive, polarization, diffusion, and possible inductive behavior. Because these impedance responses evolve during battery aging, both the magnitude and geometry of the Nyquist response can provide degradation-sensitive information for SOH estimation.

A representative Nyquist plot obtained from the NASA dataset is shown in Figure 3. The plot was generated from the original raw Battery impedance data of battery B0005 using a stable selected EIS record. The early, middle, and late stages were selected according to the sequential order of EIS measurements in the NASA MATLAB file. For B0005, the impedance records were first indexed sequentially from the beginning to the end of the aging test. A stable EIS record was then selected from the early, middle, and late portions of the impedance sequence. The first EIS record was not used for the representative Nyquist plot because it contained noisy and non-ideal points and did not clearly illustrate the semicircle-like impedance response. Therefore, sequential EIS numbers 54, 140, and 258, corresponding to overall cycle indices 153, 329, and 573, were selected as representative early, middle, and late aging stages, respectively.

According to the NASA dataset description, the EIS measurement was performed over a frequency range of 0.1 Hz–5 kHz. The high-frequency region is mainly associated with ohmic resistance, the middle-frequency arc reflects charge-transfer and interfacial polarization behavior, and the low-frequency region is related to diffusion and concentration-polarization processes. However, the MATLAB files do not provide a separate point-

wise frequency vector. Therefore, exact frequency values were not assigned to individual impedance points, and the points were connected according to their recorded sequence.

In the NASA MATLAB files, the raw impedance spectrum is stored as Battery impedance, while the fitted impedance parameters, including R_e and R_{ct} , are stored separately for each EIS record. Therefore, R_e and R_{ct} were used as numerical features in this study and were not manually extracted from the visual intercepts of the plotted Nyquist curve. To avoid ambiguity, the metadata of the first B0005 EIS record associated with the original NASA-provided fitted R_e and R_{ct} values are summarized separately in Table S1.

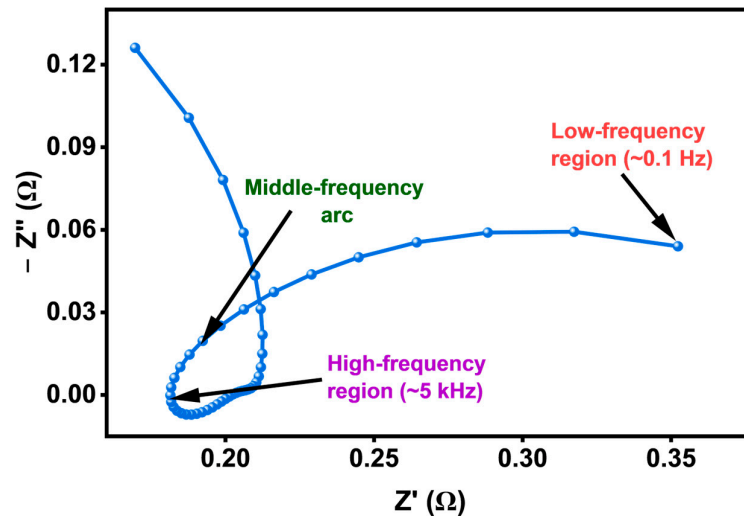


Figure 3. Representative Nyquist plot of battery B0005 generated from a stable selected early-stage raw EIS record. The curve was plotted using the original NASA Battery impedance data from sequential EIS number 54, corresponding to overall cycle index 153. The high-, middle-, and low-frequency regions are indicated according to the NASA EIS sweep range of 0.1 Hz–5 kHz, and the points are connected according to their recorded sequence because no point-wise frequency vector is provided in the MATLAB file.

The Nyquist response shown in Figure 3 does not form a perfectly ideal semicircle. This behavior is expected because the NASA impedance data are experimentally measured raw spectra rather than idealized equivalent-circuit responses. The non-ideal shape may arise from high-frequency inductive behavior, measurement noise, interfacial polarization, diffusion-related low-frequency response, and aging-induced deformation of the impedance curve. Therefore, Figure 3 is used as a representative experimental impedance response for feature extraction rather than as an ideal fitted electrochemical model. To further support the representativeness of the EIS data, additional Nyquist plots for batteries B0005, B0006, B0007, and B0018 at selected aging stages are provided in Figure S2.

The impedance at a given frequency can be expressed as:

$$Z(f) = Z'(f) + jZ''(f) \quad (2)$$

where $Z'(f)$ and $Z''(f)$ represent the real and imaginary impedance components at frequency f , respectively [48,49].

In addition to statistical descriptors, geometric features were extracted to characterize aging-related changes in the Nyquist response [25,50]. The real-impedance span was calculated as:

$$Z'_{span} = Z'_{max} - Z'_{min} \quad (3)$$

$$A_{Nyquist} = \left| \sum_{k=1}^{m-1} \frac{-Z''_k - Z''_{k+1}}{2} (Z'_{k+1} - Z'_k) \right| \quad (4)$$

The relative arc height was expressed as:

$$R_{arc} = \frac{\max_k(-Z''_k)}{Z'_{span}} \quad (5)$$

where m is the number of impedance points, Z'_{span} represents the range of the real component, $A_{Nyquist}$ describes the integrated area under the Nyquist response, and R_{arc} represents the maximum imaginary-impedance magnitude normalized by the real-impedance span. These descriptors summarize changes in the magnitude and geometry of the impedance response during battery aging.

In the NASA dataset, Battery impedance represents the raw complex impedance spectrum, whereas Rectified Impedance represents the NASA-provided rectified complex impedance spectrum. In this study, rectified impedance was used directly as provided in the NASA dataset. It does not refer to a Kramers–Kronig transformation performed by the authors, and it is not obtained by taking the absolute value of impedance. Both raw and rectified impedance spectra were treated as complex-valued data, and their real and imaginary components were extracted separately for feature generation.

The raw and rectified impedance arrays were processed separately to preserve complementary information from the original and rectified spectra. Statistical descriptors, including the mean, standard deviation, minimum, maximum, range, and magnitude-related measures of the real and imaginary components, were extracted from each spectrum. Shape-based descriptors, including the overall, early-region, and late-region slopes, were also calculated to describe changes in the Nyquist-curve geometry.

Because the raw and rectified impedance spectra contain different numbers of points, the impedance sequences were converted into fixed-length representations before model development. In the NASA data used in this study, the raw impedance spectrum contains 48 complex points, whereas the rectified impedance spectrum contains 39 complex points. For each sequence, the real and imaginary components were separated and interpolated independently over a normalized point index. If an impedance sequence contained (m) points, the normalized point index was defined as:

$$S_k = \frac{(k-1)}{(m-1)} \quad k = 1, 2, 3, \dots, m \quad (6)$$

where s_k is the original normalized point index and m is the total number of impedance points. The sequence was then linearly interpolated onto a fixed normalized grid $\tilde{s}_l$, defined as:

$$\tilde{s}_l = \frac{(l-1)}{(q-1)} \quad l = 1, 2, 3, \dots, q \quad (7)$$

where q is the fixed number of resampled points. Since the MATLAB files do not provide explicit point-wise frequency values, this procedure was performed along the recorded measurement sequence rather than along a physical frequency axis.

This resampling procedure was used only to obtain fixed-length inputs for machine-learning models. It should not be interpreted as physical frequency normalization or frequency-aligned electrochemical modeling. Instead, the resulting variables represent sequence-normalized impedance-shape features and compressed geometric curve information along the recorded sequence. These features are useful for pattern recognition within the NASA measurement protocol, but they are not fully frequency-aligned electrochemical descriptors and may not be directly portable to EIS datasets measured using different frequency grids or instruments.

NASA-provided impedance parameters, including electrolyte resistance (R_e) and charge-transfer resistance (R_{ct}), were included when available. Discharge-derived variables, such as mean voltage, voltage drop, current statistics, temperature behavior, and discharge duration, were extracted to represent complementary operating and degradation information. Cycle number and EIS measurement index were also retained as indicators of aging progression. The resulting feature categories and their intended interpretations are summarized in Table 2.

Table 2. Extracted feature categories.

Feature Category	Examples	Physical Meaning
Raw Impedance Statistics	Means, standard deviations, minima, maxima, and ranges of Z' and Z''	Overall impedance magnitude and dispersion
Rectified impedance statistics	Statistical descriptors of rectified real and imaginary components	Corrected impedance response
Nyquist shape descriptors	Real-impedance span, integrated area, arc ratio, and regional slopes	Geometric changes in the impedance response
Resampled impedance features	Resampled real and imaginary impedance points	Detailed impedance-shape variation along the recorded sequence
NASA impedance parameters	R_e , and R_{ct}	Electrolyte and charge-transfer resistance
Discharge-profile features	Voltage statistics, voltage drop, current statistics, temperature behavior, and discharge duration	Operating behavior and capacity-related degradation
Aging-progression features	Cycle number and EIS measurement index	Relative progression of battery aging

3.3. Machine Learning Models

Five machine learning regression models were evaluated in this study: RF, ET, GB, HGB, and XGBoost. These models were selected because they are well suited for nonlinear regression problems, can handle tabular datasets with mixed feature groups, and generally provide stable performance without requiring very large training datasets. This makes them appropriate for learning the relationship between the extracted impedance-based features and battery SOH.

RF is an ensemble method that builds multiple decision trees using bootstrapped subsets of the training data and averages their outputs to improve prediction stability. ET follows a similar tree-based ensemble structure but introduces additional randomness during feature splitting, which can reduce variance and improve generalization. GB builds decision trees sequentially, where each new tree attempts to correct the errors of the previous model, allowing it to capture complex nonlinear degradation patterns. HGB improves computational efficiency by using histogram-based feature binning during training. XGBoost is a regularized gradient boosting method designed to improve predictive accuracy, control overfitting, and handle complex feature interactions effectively.

3.4. Validation Strategies

Three validation strategies were used to evaluate the prediction performance and generalization ability of the developed models. First, a random train–test split was applied, where samples from all batteries were randomly divided into training and testing subsets. This strategy provides a general measure of model performance when the training and testing data are drawn from a similar distribution. However, because samples from the same battery may appear in both subsets, the results may be optimistic.

Second, a strict battery-wise validation strategy was used. In this case, complete batteries were separated between the training and testing sets, ensuring that the models were

evaluated on batteries that were not seen during training. This approach is more representative of practical battery management applications, where a trained model is expected to estimate the SOH of new cells with different aging behavior.

Third, repeated battery-wise validation was performed using five different unseen-battery splits. This evaluation was included to examine the stability of the models and to reduce the dependence of the results on a single train–test partition.

To reduce label leakage, each EIS measurement was matched only with the most recent preceding discharge-capacity measurement from the same battery, and no future discharge-capacity information was used for SOH labeling. In addition, strict battery-wise validation was used to prevent samples from the same battery from appearing in both the training and test sets. Nevertheless, some features may still act as indirect proxy indicators of SOH. In particular, discharge duration under constant-current discharge is closely related to discharge capacity, while cycle number and EIS measurement index represent chronological aging progression. Therefore, the results are interpreted as multi-source SOH estimation rather than EIS-only prediction, and leakage-controlled feature ablation is identified as an important direction for future work. The model performance was assessed using MAE, RMSE, and R^2 as defined in Equations (8)–(10), respectively. MAE measures the average prediction error, RMSE gives greater weight to larger errors, and R^2 describes the overall agreement between the measured and predicted SOH values [51,52].

$$MAE = \frac{1}{n} \sum_{i=1}^n |SOH_i - \widehat{SOH}_i| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (SOH_i - \widehat{SOH}_i)^2} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (SOH_i - \widehat{SOH}_i)^2}{\sum_{i=1}^n (SOH_i - \bar{SOH})^2} \quad (10)$$

where SOH_i is the measured SOH, $\widehat{SOH}_i$ is the predicted SOH, $\bar{SOH}$ is the mean measured SOH, and n is the number of test samples.

4. Results

The developed models were evaluated using random sample splitting, strict battery-wise validation, and repeated battery-wise validation. The random split was used to assess sample-level prediction under similar training and test distributions, whereas the battery-wise strategies were used to examine generalization to previously unseen cells. The results obtained under these validation conditions are presented and compared in the following subsections.

4.1. Random Train–Test Performance

The random train–test results are presented in Table 3. In this validation strategy, samples from all batteries were randomly divided into training and testing subsets. All models achieved strong prediction performance, showing that the extracted frequency-aware impedance features contain meaningful information related to battery degradation. Among the evaluated models, GB achieved the best random split performance with an R^2 value of 0.9788 and an RMSE of 2.1794%. XGBoost and ET also showed high prediction accuracy, with R^2 values of 0.9740 and 0.9693, respectively. These results confirm that the proposed feature set can effectively capture the relationship between impedance behavior and battery SOH when training and testing samples come from a similar data distribution.

Table 3. SOH prediction performance under random train–test validation.

Model	MAE (%)	RMSE (%)	R ²
RF	1.0379	3.9241	0.9313
ET	0.6120	2.6246	0.9693
GB	1.0250	2.1794	0.9788
HGB	0.9304	3.0640	0.9581
XGBoost	0.9392	2.4132	0.9740

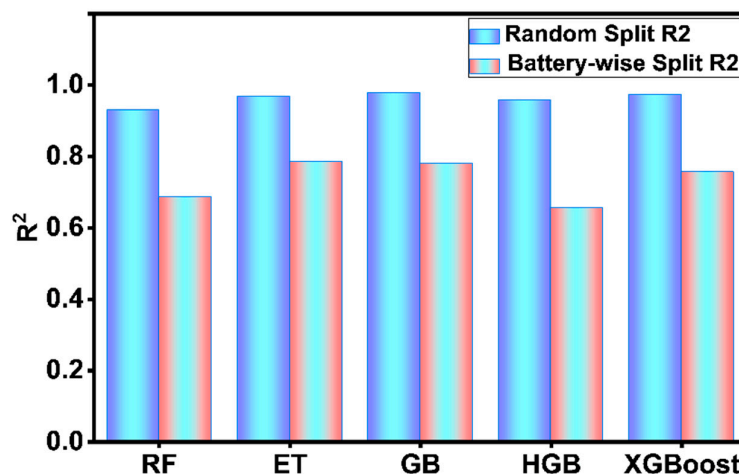
4.2. Battery-Wise Validation Performance

The battery-wise validation results are shown in Table 4. This evaluation is more challenging than random splitting because the test batteries were not included during model training. Under this strict validation strategy, ET achieved the best single-split battery-wise performance, with an MAE of 4.9131%, RMSE of 7.3664%, and R² of 0.7855. GB also showed competitive performance, achieving an R² of 0.7806. These results indicate that the proposed frequency-aware feature set can improve cross-battery generalization by capturing degradation-related information from impedance spectra and discharge profiles.

Table 4. SOH prediction performance under strict battery-wise validation.

Model	MAE (%)	RMSE (%)	R ²
RF	6.0666	8.9080	0.6864
ET	4.9131	7.3664	0.7855
GB	5.3373	7.4504	0.7806
HGB	6.6181	9.3233	0.6565
XGBoost	5.6283	7.8426	0.7569

To illustrate the effect of the validation strategy, Figure 4 compares the R² values obtained under random-split and strict battery-wise validation. All models showed lower performance when evaluated on previously unseen batteries. The largest differences were observed for RF and HGB, whereas ET and GB maintained comparatively stronger cross-battery performance. This consistent reduction confirms that random sample splitting provides optimistic estimates when samples from the same batteries are shared between the training and test sets.

**Figure 4.** Comparison of R² values obtained under random-split and strict battery-wise validation for the evaluated ML models.

4.3. Repeated Battery-Wise Validation

Repeated battery-wise validation was performed using five different battery-grouped partitions to examine model stability across different unseen-cell combinations. As summarized in Table 5, ET achieved the highest average R^2 of 0.8811 ± 0.0629 , followed by XGBoost with 0.8765 ± 0.0703 and RF with 0.8710 ± 0.0507 . RF produced the lowest average MAE of 3.2205 percentage points and the smallest variation in R^2 . In contrast, GB and HGB showed lower average R^2 values and greater variability across the five partitions. These findings demonstrate that performance depends on the particular batteries reserved for testing and should therefore be reported using both average accuracy and between-split variability.

Table 5. Repeated battery-wise validation performance.

Model	MAE (%)	RMSE (%)	R^2
RF	3.2205	0.8710	0.0507
ET	3.3041	0.8811	0.0629
GB	3.5967	0.8412	0.1067
HGB	3.3839	0.8425	0.0937
XGBoost	3.2570	0.8765	0.0703

4.4. Prediction and Error Analysis

The actual and predicted SOH values of the best battery-wise model are shown in Figure 5. This figure provides a direct visual comparison between the measured SOH values and the model predictions for unseen test batteries. Overall, the predicted values follow the same trend as the measured SOH values, showing that the model can capture the main degradation behavior from the extracted impedance and discharge-based features. The closeness of the points to the ideal prediction line indicates that the proposed frequency-aware feature set contains useful information for estimating battery health, even when the test batteries were not included during training.

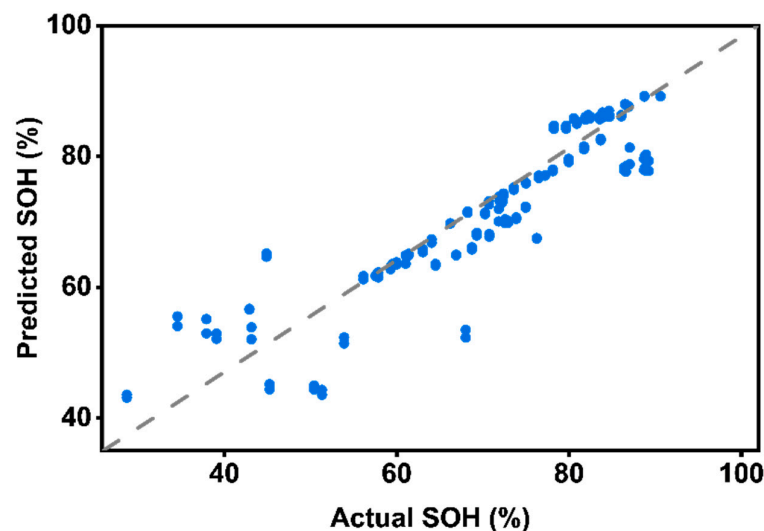


Figure 5. Measured versus predicted SOH values obtained using the ET model for seven unseen test batteries under strict battery-wise validation. The blue dots represent individual battery samples, and the dashed line indicates the ideal one-to-one agreement between the measured and predicted SOH values..

The prediction-error distribution for ET under strict battery-wise validation is shown in Figure 6. Prediction error was calculated as:

$$e_i = \widehat{SOH}_i - SOH_i \quad (11)$$

where negative values indicate SOH underestimation and positive values indicate overestimation. Most errors are concentrated near zero, indicating relatively accurate predictions for a substantial proportion of the test samples. However, the distribution contains both negative and positive tails, showing that larger errors occur for certain measurements from unseen batteries. These deviations are examined further using battery-level error analysis.

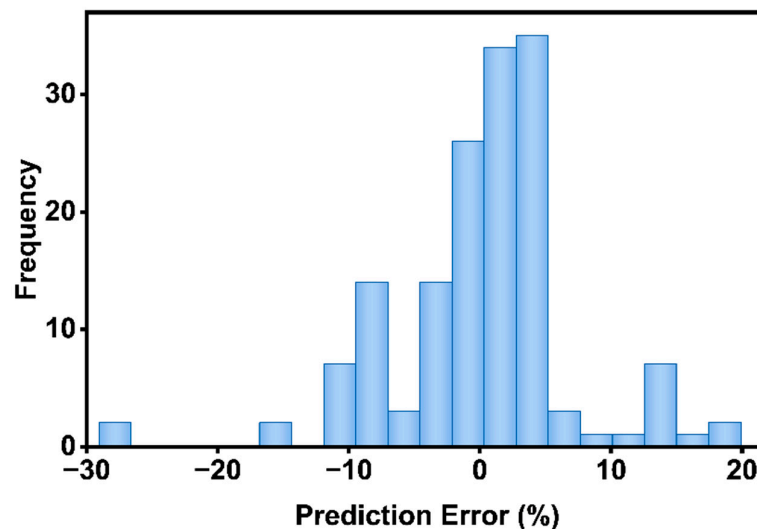


Figure 6. Distribution of ET prediction errors for seven unseen batteries under strict battery-wise validation. Negative and positive values represent SOH underestimation and overestimation, respectively.

Battery-level MAE values for the seven unseen test batteries are compared in Figure 7. Prediction performance varied substantially among the batteries. B0049 showed the largest error, with an MAE of approximately 16.5 percentage points, followed by B0050 and B0038 with approximately 12.8 and 8.1 percentage points, respectively. In contrast, B0029, B0042, B0044, and B0047 produced MAE values below approximately 3 percentage points. This variation indicates that the model generalizes effectively to some batteries but performs poorly when the degradation trajectory or operating characteristics differ from those represented in the training set. The result highlights the importance of reporting battery-level performance in addition to aggregate metrics.

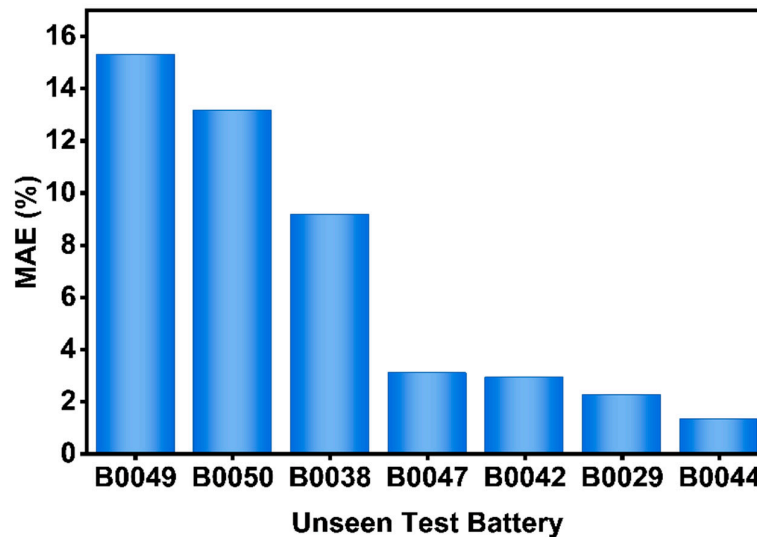


Figure 7. Battery-level MAE values obtained using ET for seven unseen batteries under strict battery-wise validation.

4.5. Feature Importance Analysis

Feature importance was calculated from ET to examine the relative contribution of the input variables under strict battery-wise validation. As shown in Figure 8, mean temperature was the most influential feature, followed by voltage standard deviation, maximum temperature, voltage drop, discharge duration, and mean voltage. Cycle number and EIS measurement index also contributed to the predictions. In comparison, the displayed impedance-derived features, including the raw-spectrum slopes, arc ratio, and R_e , showed lower individual importance.

These results indicate that the model relies strongly on discharge-profile, thermal, and aging-progression information rather than predominantly on impedance descriptors. However, tree-based importance values measure predictive contribution and do not establish a causal relationship with degradation. Moreover, discharge duration, cycle number, and EIS measurement index may act as direct or indirect proxies for capacity loss. Their influence should therefore be examined using feature-group ablation before conclusions are drawn regarding the independent contribution of impedance information.

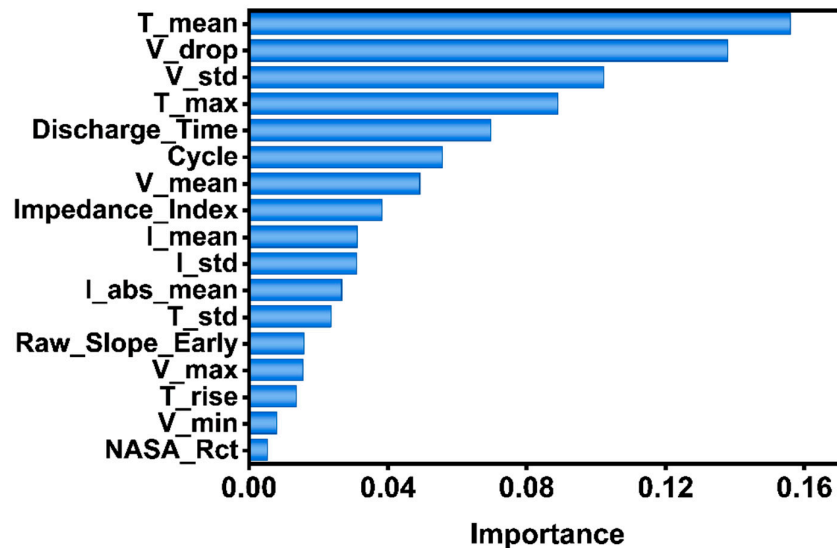


Figure 8. ET feature-importance values obtained under strict battery-wise validation.

5. Discussion

The results demonstrate that combining impedance-, discharge-, thermal-, and aging-related information can support cross-battery SOH estimation. All models achieved high accuracy under random-split validation, with GB producing an R^2 of 0.9788. However, performance declined under strict battery-wise validation, where ET achieved the best result with an R^2 of 0.7855. This reduction shows that random sample splitting does not fully represent the difficulty of predicting SOH for previously unseen batteries. The battery-wise evaluation therefore provides a more realistic assessment of model generalization and reveals the influence of cell-to-cell differences in degradation behavior.

Repeated battery-wise validation further showed that model performance depends on the batteries selected for testing. ET achieved the highest average R^2 of 0.8811 ± 0.0629 , whereas RF produced the lowest average MAE and the smallest variation in R^2 . Nevertheless, the battery-level analysis revealed substantial differences among unseen cells. Although four test batteries produced MAE values below approximately 3 percentage points, B0049, B0050, and B0038 showed considerably larger errors. These findings indicate that aggregate performance metrics can conceal poor generalization for individual batteries and that both average and battery-specific results should be reported.

The feature-importance results show that the predictions were driven mainly by thermal, voltage, discharge-duration, and aging-progression variables, while the displayed impedance-derived descriptors had lower individual importance. This finding supports the multi-source nature of the framework but also requires careful interpretation. Because SOH was calculated from discharge capacity, discharge duration under constant-current loading may be strongly related to the target value. Similarly, cycle number and EIS measurement index may act as proxies for chronological aging. Consequently, the reported performance cannot be attributed solely to impedance information. Feature-group ablation, particularly the removal of discharge duration, cycle number, and EIS measurement index, is required to quantify the independent contribution of the impedance features.

The sequence-normalized resampling procedure also has an important limitation. Because point-wise frequency vectors are not available in the NASA MATLAB files, the resampled impedance features are based on the recorded order of impedance points rather than exact physical frequency positions. Therefore, these features should be interpreted as geometric curve-shape descriptors and compressed pattern-recognition inputs

rather than fully frequency-aligned electrochemical descriptors. The reported accuracy therefore reflects generalization across unseen batteries within the NASA measurement protocol, but it does not by itself verify portability to impedance spectra measured using different frequency grids, instruments, or EIS protocols. Future work should use datasets with explicit frequency vectors and perform interpolation on a common logarithmic frequency grid before model training.

Several limitations should also be considered. The framework was developed using a laboratory dataset collected under controlled protocols, whereas practical battery systems experience wider variations in temperature, load profile, state of charge, and measurement noise. The temporal separation between an EIS measurement and its matched discharge-capacity label may introduce additional uncertainty. Moreover, EIS acquisition requires specialized excitation and measurement procedures that may limit direct online implementation. Future work should therefore include leakage-controlled feature ablation, uncertainty quantification, external validation using independent datasets, and evaluation under broader operating conditions. Domain adaptation and transfer-learning approaches may also improve generalization across batteries with different chemistries, aging histories, and usage patterns.

6. Disclaimer on Method Transferability

The proposed framework is currently validated within the NASA battery aging dataset and its specific EIS measurement protocol. Because the MATLAB files do not provide point-wise frequency vectors, the resampled impedance features are sequence-normalized rather than fully frequency-aligned electrochemical descriptors. Therefore, the present method should be interpreted as a NASA-dataset-based multi-source SOH estimation framework rather than a directly transferable universal EIS diagnostic model. Application to other datasets, instruments, battery chemistries, or practical BMS environments requires explicit frequency-vector information, alignment on a common physical frequency grid, recalibration using target-cell data, and external validation under representative operating conditions.

7. Conclusions

This study developed a multi-source feature-learning framework for cross-battery SOH estimation using EIS, discharge-profile, thermal, and aging-related information. Data from 34 NASA batteries were processed to construct 1830 matched impedance–SOH samples with 101 input features, and five tree-based models were evaluated using random-split, strict battery-wise, and repeated battery-wise validation. Although GB achieved a high random-split R^2 of 0.9788, performance decreased under strict battery-wise validation, where ET achieved an MAE of 4.9131 percentage points, an RMSE of 7.3664 percentage points, and an R^2 of 0.7855. Repeated battery-wise validation further showed that performance varied across unseen batteries. Feature-importance analysis indicated strong contributions from thermal, voltage, discharge-duration, and aging-progression variables, highlighting the need for future feature-group ablation. Because the impedance features are sequence-normalized rather than frequency-aligned, the present framework should be interpreted within the NASA measurement protocol. Direct transfer to other datasets or practical BMS environments requires explicit frequency-vector information, frequency-grid alignment, recalibration, and external validation.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/batteries12080291/s1>, **Figure S1:** Representative charge–discharge voltage curves of selected NASA lithium-ion batteries at different aging stages. **Figure S2:** Nyquist plots of batteries (a) B0005, (b) B0006, (c) B0007, and (d) B0018 at selected aging stages.

Table S1: Selected NASA batteries and aging cycles used for representative charge–discharge voltage curves. **Table S2:** Matched impedance–state-of-health (SOH) samples capped at 100% SOH. **Table S3:** Metadata of the first B0005 EIS measurement used to clarify the NASA-provided R_e and R_{ct} values, with the representative Fig. 3 record also noted.

Author Contributions: Conceptualization, S.A.S. and W.Y.K.; methodology, S.A.S. and W.Y.K.; software, S.A.S.; validation, F.A. and W.A.; formal analysis, S.A.S.; investigation, S.A.S.; data curation, S.A.S.; visualization, S.A.S.; writing—original draft preparation, S.A.S.; writing—review and editing, F.A., W.A., and W.Y.K.; supervision, W.Y.K.; project administration, W.Y.K.; conceptual guidance, W.Y.K.; methodology review, W.Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw battery-aging data used in this study are publicly available from the NASA Ames Prognostics Center of Excellence data repository.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (OpenAI, GPT-5.5) solely for language editing, including grammar, spelling, and sentence clarity. The tool was not used to generate scientific results, analyze data, or formulate conclusions. Following its use, the authors reviewed and edited the manuscript and take full responsibility for its content.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Khan, F.N.U.; Rasul, M.G.; Sayem, A.S.M.; Mandal, N.K. Design and optimization of lithium-ion battery as an efficient energy storage device for electric vehicles: A comprehensive review. *J. Energy Storage* **2023**, *71*, 108033.
2. Xing, J.; Bliznakov, S.; Bonville, L.; Oljaca, M.; Maric, R. A Review of Nonaqueous Electrolytes, Binders, and Separators for Lithium-Ion Batteries. *Electrochem. Energy Rev.* **2022**, *5*, 14.
3. Wang, C.Y.; Liu, T.; Yang, X.G.; Ge, S.; Stanley, N.V.; Rountree, E.S.; Leng, Y.; McCarthy, B.D. Fast charging of energy-dense lithium-ion batteries. *Nature* **2022**, *611*, 485–490.
4. Tanim, T.R.; Kim, S.; Colclasure, A.M.; Yang, Z.; Gering, K.; Weddle, P.J.; Evans, M.; Dufek, E.J.; Lin, Y.; Wen, J.; et al. Rational designs to enable 10-min fast charging and long cycle life in lithium-ion batteries. *J. Power Sources* **2023**, *582*, 233519.
5. Ibraheem, R.; Strange, C.; Dos Reis, G. Capacity and Internal Resistance of lithium-ion batteries: Full degradation curve prediction from Voltage response at constant Current at discharge. *J. Power Sources* **2023**, *556*, 232477.
6. Wildfeuer, L.; Karger, A.; Aygöl, D.; Wassiliadis, N.; Jossen, A.; Lienkamp, M. Experimental degradation study of a commercial lithium-ion battery. *J. Power Sources* **2023**, *560*, 232498.
7. Li, S.; Zhang, C.; Zhao, Y.; Offer, G.J.; Marinescu, M. Effect of thermal gradients on inhomogeneous degradation in lithium-ion batteries. *Commun. Eng.* **2023**, *2*, 74.
8. Lee, D.; Kim, B.; Shin, C.B.; Oh, S.M.; Song, J.; Jang, I.C.; Woo, J.J. Modeling the Combined Effects of Cyclable Lithium Loss and Electrolyte Depletion on the Capacity and Power Fades of a Lithium-Ion Battery. *Energies* **2022**, *15*, 7056. <https://doi.org/10.3390/en15197056>.
9. Toughzaoui, Y.; Toosi, S.B.; Chaoui, H.; Louahlia, H.; Petrone, R.; Le Masson, S.; Gualous, H. State of health estimation and remaining useful life assessment of lithium-ion batteries: A comparative study. *J. Energy Storage* **2022**, *51*, 104520.
10. Wang, F.; Zhai, Z.; Liu, B.; Zheng, S.; Zhao, Z.; Chen, X. Open access dataset, code library and benchmarking deep learning approaches for state-of-health estimation of lithium-ion batteries. *J. Energy Storage* **2024**, *77*, 109884.
11. Neunzling, J.; Winter, H.; Henriques, D.; Fleckenstein, M.; Markus, T. Online State-of-Health Estimation for NMC Lithium-Ion Batteries Using an Observer Structure. *Batteries* **2023**, *9*, 494. <https://doi.org/10.3390/batteries9100494>.
12. Zhang, J.; Li, K. State-of-Health Estimation for Lithium-Ion Batteries in Hybrid Electric Vehicles—A Review. *Energies* **2024**, *17*, 5753. <https://doi.org/10.3390/en17225753>.
13. Chen, L.; Bao, X.; Lopes, A.M.; Xu, C.; Wu, X.; Kong, H.; Ge, S.; Huang, J. State of health estimation of lithium-ion batteries based on equivalent circuit model and data-driven method. *J. Energy Storage* **2023**, *73*, 109195.

14. Li, X.; Ju, L.; Geng, G.; Jiang, Q. Data-driven state-of-health estimation for lithium-ion battery based on aging features. *Energy* **2023**, *274*, 127378.
15. Lanubile, A.; Bosoni, P.; Pozzato, G.; Allam, A.; Acquarone, M.; Onori, S. Domain knowledge-guided machine learning framework for state of health estimation in Lithium-ion batteries. *Commun. Eng.* **2024**, *3*, 168.
16. Sardar, S.A.; Iqbal, S.; Park, J.; Han, S.; Kim, W.Y. Field Data Retrieval for Electric Vehicles and Estimating Equivalent Circuit Model Parameters via Particle Swarm Optimization. *Technologies* **2025**, *13*, 91. <https://doi.org/10.3390/technologies13030091>.
17. Xu, Z.; Wang, J.; Lund, P.D.; Zhang, Y. Co-estimating the state of charge and health of lithium batteries through combining a minimalist electrochemical model and an equivalent circuit model. *Energy* **2022**, *240*, 122815.
18. Feng, Y.; Xue, C.; Han, F.; Cao, Z.; Yang, R.J. State-of-Charge and State-of-Health Estimation in Li-Ion Batteries Using Cascade Electrochemical Model-Based Sliding-Mode Observers *Batteries* **2024**, *10*, 290. <https://doi.org/10.3390/batteries10080290>.
19. Lucaferri, V.; Quercio, M.; Laudani, A.; Riganti Fulginei, F. A Review on Battery Model-Based and Data-Driven Methods for Battery Management Systems *Energies* **2023**, *16*, 7807. <https://doi.org/10.3390/en16237807>.
20. Yang, Z.; Wang, X. An improved parameter identification method considering multi-timescale characteristics of lithium-ion batteries *J. Energy Storage* **2023**, *59*, 106462.
21. Van, C.N.; Quang, D.T. State of charge estimation for Lithium-Ion battery cell considering the influence of aging parameters and operating temperature *Int. J. Electrochem. Sci.* **2023**, *18*, 100379.
22. Lu, J.; Xiong, R.; Tian, J.; Wang, C.; Sun, F. Deep learning to estimate lithium-ion battery state of health without additional degradation experiments *Nat. Commun.* **2023**, *14*, 2760.
23. Yao, Q.; Song, X.; Xie, X.; W. State of health estimation of lithium-ion battery based on CNN–WNN–WLSTM. *Complex Intell. Syst.* **2024**, *10*, 2919–2936.
24. Li, Z.; Zhang, F.; Cai, Z.; Xu, L.; Shen, S.; Yu, P. State of health estimation of lithium-ion batteries based on interval voltage features *J. Energy Storage* **2024**, *102*, 114112.
25. Liu, Y.; Wang, L.; Li, D.; Wang, K. State-of-health estimation of lithium-ion batteries based on electrochemical impedance spectroscopy: A review. *Prot. Control Mod. Power Syst.* **2023**, *8*, 41.
26. Kong, X.; Bonakdarpour, A.; Wetton, B.T.; Wilkinson, D.P.; Gopaluni, B. State of Health Estimation for Lithium-Ion Batteries. *IFAC-Pap.* **2018**, *51*, 667–671.
27. Shao, B.; Zhong, J.; Tian, J.; Li, Y.; Chen, X.; Dou, W.; Liao, Q.; Lai, C.; Lu, T.; Xie, J. State-of-Health Estimation of Lithium-Ion Batteries Based on Electrochemical Impedance Spectroscopy Features and Fusion Interpretable Deep Learning Framework. *Energies*, **2025**, *18*, 1385. <https://doi.org/10.3390/en18061385>.
28. Zhao, X.; Liu, S.; Li, E.; Wang, Z.; Gu, F.; Ball, A.D. A hybrid intelligent model using the distribution of relaxation time analysis of electrochemical impedance spectroscopy for lithium-ion battery state of health estimation *J. Energy Storage* **2024**, *84*, 110814.
29. Abu Qahouq, J. An Electrochemical Impedance Spectrum-Based State of Health Differential Indicator with Reduced Sensitivity to Measurement Errors for Lithium–Ion Batteries *Batteries* **2024**, *10*, 368. <https://doi.org/10.3390/batteries10100368>.
30. Chen, B.R.; Police, Y.R.; Li, M.; Chinnam, P.R.; Tanim, T.R.; Dufek, E.J. A mathematical approach to survey electrochemical impedance spectroscopy for aging in lithium-ion batteries *Front. Energy Res* **2023**, *11*, 1132876. (In English)
31. Xia, B.; Qin, Z.; Fu, H. Rapid estimation of battery state of health using partial electrochemical impedance spectra and interpretable machine learning *J. Power Sources* **2024**, *603*, 234413.
32. Du, X.; Meng, J.; Amirat, Y.; Gao, F.; Benbouzid, M. Exploring impedance spectrum for lithium-ion batteries diagnosis and prognosis: A comprehensive review *J. Energy Chem.* **2024**, *95*, 464–483.
33. Xiao, Y.; Huang, X.; Meng, J.; Zhang, Y.; Knap, V.; Stroe, D.I. Electrochemical Impedance Spectroscopy-Based Characterization and Modeling of Lithium-Ion Batteries Based on Frequency Selection *Batteries* **2024**, *11*, 11. <https://doi.org/10.3390/batteries11010011>.
34. Yenioglu, M.; Aycicek, E.; Erdinc, O. A Data-Driven Framework for Lithium-Ion Battery Remaining Useful Life Prediction Using CNN and Machine Learning Models *Batteries* **2026**, *12*, 135. <https://doi.org/10.3390/batteries12040135>.
35. Tang, P.; Qiu, Z.; Yao, Z.; Pan, J.; Cheng, D.; Gu, X.; Sun, C. Lithium-ion battery RUL prediction based on optimized VMD-SSA-PatchTST algorithm *Sci. Rep.* **2025**, *15*, 26824.
36. Amiri, M.N.; Spitthoff, L.; Varagnolo, D.; Burheim, O.S.; Lamb, J.J. Data-driven state of health estimation of lithium-ion batteries: Impact of voltage, time, and derivative features *Int. J. Electrochem. Sci.* **2026**, *21*, 101372.
37. Zhou, J.; Rong, J.; Zhang, J.; Liu, C.; Yi, F.; Jiao, Z.; Zhang, C. Deep learning estimation of state of health for lithium-ion batteries using multi-level fusion features of discharge curves *J. Power Sources* **2025**, *653*, 237781.

38. Seo, Y.; Kim, T.; Stephane, B. Robust SOH Prediction for Lithium-Ion Batteries via ProbSparse Informer Architecture. In Proceedings of the 2025 Global Reliability and Prognostics and Health Management Conference (PHM-Xian), Xi'an, China, 10–12 October 2025; pp. 1–8.
39. Xu, Q.; Wang, X.; Ye, H.; Gong, L.; Tan, P.; Pan, T. An accurate state of health estimation method for lithium-ion batteries based on expansion force analysis *Energy* **2025**, *325*, 136155.
40. Han, D.; Zhou, N.; Chen, Z. Enhancing State of Health Prediction Accuracy in Lithium-Ion Batteries through a Simplified Health Indicator Method *Batteries* **2024**, *10*, 342. <https://doi.org/10.3390/batteries10100342>.
41. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016. <https://doi.org/10.1145/2939672.2939785>.
42. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized tree *Mach. Learn.* **2006**, *63*, 3–42.
43. Jerome, H. F. Greedy function approximation: A gradient boosting machine *Ann. Stat.* **2001**, *29*, 1189–1232.
44. National Aeronautics and Space Administration (NASA). *Li-ion Battery Aging Datasets*; NASA Open Data Portal: Washington, DC, USA, 2010.
45. Saha, B.; Goebel, K. *Battery Data Set, NASA Ames Prognostics Data Repository*; NASA Ames Research Center: Moffett Field, CA, USA, 2007.
46. Dannier, A.; Brando, G.; Ribera, M.; Spina, I. Li-Ion Batteries for Electric Vehicle Applications: An Overview of Accurate State of Charge/State of Health Estimation Methods. *Energies* **2025**, *18*, 786. <https://doi.org/10.3390/en18040786>.
47. Zou, Z.; Xia, X.; Zhang, C.; Yue, J.; Xia, B.; Zhou, C. State of Health Estimation of Lithium-Ion Batteries Based on Voltage Data Segment Hybrid Model *Batteries* **2026**, *12*, 83. <https://doi.org/10.3390/batteries12030083>.
48. Lazanas, A.C.; Prodromidis, M.I. Electrochemical Impedance Spectroscopy — A Tutorial. *ACS Meas. Sci. AU* **2023**, *3*, 162–193.
49. Piccardi, A.; Pettinato, S.; Rossi, M.C.; Girolami, M.; Salvatori, S. Low-Frequency Measurements in Electrochemical Impedance Spectroscopy: A Brief Review. *Electronics* **2025**, *14*, 4048. <https://doi.org/10.3390/electronics14204048>.
50. Zhou, Y.; Dong, G.; Tan, Q.; Han, X.; Chen, C.; Wei, J. State of health estimation for lithium-ion batteries using geometric impedance spectrum features and recurrent Gaussian process regression *Energy* **2023**, *262*, 125514.
51. Pohlmann, S.; Mashayekh, A.; Buberger, J.; Estaller, J.; Wiedenmann, A.; Kuder, M.; Neve, A.; Weyh, T. Correlation analysis and feature extraction using impedance spectroscopy over aging of lithium ion batteries. *J. Energy Storage* **2025**, *105*, 114715.
52. Jiang, B.; Zhu, J.; Wang, X.; Wei, X.; Shang, W.; Dai, H. A comparative study of different features extracted from electrochemical impedance spectroscopy in state of health estimation for lithium-ion batteries *Appl. Energy* **2022**, *322*, 119502.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.