International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2026, 10 (1), 79-87
DOI: 10.36287/ijmsit.10.1.9
Received: July 16, 2026; Accepted: July 28, 2026

Cross-Battery State-of-Health Estimation of Lithium-Ion Batteries Using Leave-One-Battery-Out Validation and Machine Learning

M. Musab Bayat

Sivas Cumhuriyet University, Electrical-Electronics Eng. Dept. Sivas, Türkiye (muhammed@cumhuriyet.edu.tr)
(ORCID: 0000-0002-4631-6516)

Abstract – Accurate state-of-health (SOH) estimation is essential for improving the safety, reliability, and operational efficiency of lithium-ion batteries used in electric vehicles and renewable energy storage systems. Although numerous machine learning approaches have been proposed for SOH prediction, many studies evaluate their performance using randomly divided training and testing datasets, which may overestimate the generalization capability of the developed models. This study presents a cross-battery SOH estimation framework based on Leave-One-Battery-Out (LOBO) validation using the NASA battery degradation dataset. A comprehensive set of twenty-one statistical and physics-inspired features was extracted from voltage, current, temperature, discharge time, and energy measurements to characterize battery degradation behavior. Four regression models, namely Random Forest (RF), Least Squares Boosting (LSBoost), Long Short-Term Memory (LSTM), and Support Vector Regression (SVR), were trained and evaluated under the LOBO validation strategy, where each battery was used once as an unseen test battery while the remaining batteries were employed for training. Model performance was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). The experimental results demonstrated that the RF model achieved the best overall performance with an average RMSE of 0.0342, an average MAE of 0.0305, and an average R^2 of 0.8654. LSBoost and LSTM produced competitive prediction accuracies, whereas SVR exhibited poor cross-battery generalization with a negative average R^2 value. Furthermore, feature importance analysis revealed that VoltageRMS, MeanVoltage, Energy, MeanPower, and DischargeTime were the most influential variables affecting SOH estimation. The proposed framework provides a practical and reliable methodology for evaluating machine learning models under realistic operating conditions and highlights the importance of cross-battery validation for developing robust battery health prediction systems.

Keywords *– Lithium-ion battery, SOH estimation, Leave-One-Battery-Out validation, Cross-battery prediction, Machine learning, Feature importance.*

Citation: Bayat, M. M. (2026). Cross-Battery State-of-Health Estimation of Lithium-Ion Batteries Using Leave-One-Battery-Out Validation and Machine Learning. International Journal of Multidisciplinary Studies and Innovative Technologies, 10(1): 79-87.

I. INTRODUCTION

The rapid growth of electric vehicles, renewable energy systems, and large-scale battery energy storage installations has significantly increased the demand for reliable lithium-ion battery technologies. Due to their high energy density, extended cycle life, and low self-discharge rate, lithium-ion batteries have become the dominant energy storage solution for modern transportation and stationary applications. However, battery performance inevitably degrades during long-term operation because of complex electrochemical aging mechanisms, primarily manifested by irreversible capacity fading and increased internal resistance. These degradation phenomena not only shorten battery lifetime but also affect system reliability, maintenance costs, and operational safety [1].

To quantify this degradation, the state of health (SOH) has become one of the most widely accepted indicators in battery management systems (BMSs). In general, SOH refers to the ratio of the available maximum capacity of a battery to its rated capacity when new, although resistance-based and energy-based definitions have also been reported depending on the application and battery hierarchy. Accurate SOH estimation enables BMSs to optimize charging strategies, improve operational safety, predict maintenance schedules, and prevent unexpected catastrophic failures. To achieve reliable SOH monitoring, extensive research has been dedicated to developing robust and accurate estimation methodologies under diverse operating conditions [2].

In the literature, traditional SOH estimation approaches have generally revolved around direct measurement methods, electrochemical impedance spectroscopy (EIS), and equivalent circuit models (ECMs) [3-5]. Although direct measurement methods (e.g., Coulomb counting) offer high precision under specific operational processes, they require long testing times and remain insufficient under dynamic operating conditions [6, 7]. Similarly, while physics-based models and electrochemical models are successful in explaining internal battery degradation mechanisms, their real-time applications remain limited due to high computational burdens and complex parameter identification processes [8, 9]. To overcome these limitations, data-driven machine learning (ML) algorithms, which can extract meaningful Health Indicators (HIs) directly from raw battery data (voltage, current, temperature), have gained widespread popularity in recent years [10-12]. Methods such as support vector machines (SVM) and Gaussian process regression (GPR) have been shown to map battery capacity degradation with high accuracy [13-15].

Recent advances in artificial intelligence have considerably expanded the range of data-driven SOH estimation techniques. In addition to conventional machine learning algorithms, deep learning models such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Networks (CNNs), and hybrid CNN–LSTM architectures have demonstrated superior capability in capturing nonlinear temporal degradation patterns and extracting representative health features from battery cycling data [16–18]. More recently, Transformer-based architectures have attracted increasing attention owing to their self-attention mechanism, which effectively captures long-range temporal dependencies and improves estimation robustness under complex operating conditions [19–21]. Furthermore, Bayesian learning frameworks and uncertainty-aware intelligent models have been introduced to quantify prediction uncertainty, thereby enhancing the reliability of battery management systems deployed in real-world applications [22].

Despite the rapid development of these advanced prediction algorithms, ensuring their reliability under real-world conditions remains a critical challenge, primarily due to methodological flaws in model evaluation. A significant portion of current ML-based SOH estimation studies is criticized for reporting overly optimistic results (optimistic bias) due to "data leakage" arising from the random splitting of training and testing data [23, 24]. The presence of consecutive cycle data from a single cell in both the training and testing sets causes the model to memorize the aging trend of that specific cell rather than learning cell-to-cell variation [25, 26]. This situation drastically degrades the generalizability of the models when deployed to unseen cells or different battery chemistries [27, 28]. To address this bottleneck, cell-wise or group-wise cross-validation protocols have recently been proposed in the literature [29]. In particular, the Leave-One-Battery-Out (LOBO) validation strategy has recently been adopted as a realistic evaluation protocol for cross-battery SOH estimation by ensuring that the target battery remains completely unseen during model training [30,31]. Recent studies indicate that integrating LOBO validation with transfer learning and multi-source data fusion architectures enables the development of generic and robust SOH estimation models without requiring labeled historical data from the target battery [32].

Alongside the rapid evolution of prediction algorithms and the adoption of rigorous validation protocols like LOBO, increasing attention has also been devoted to feature engineering and model interpretability to further mitigate these generalizability issues. Rather than relying solely on raw voltage or capacity measurements, recent studies have demonstrated that statistical descriptors, incremental capacity analysis (ICA), differential voltage analysis (DVA), relaxation-voltage characteristics, and physics-informed health indicators can substantially improve SOH estimation accuracy [33]. Consequently, recent review papers emphasize that future research should focus not only on improving prediction accuracy but also on enhancing model interpretability, feature quality, uncertainty quantification, and rigorous validation strategies capable of evaluating model robustness under practical operating conditions [34].

To address the remaining generalizability gap and provide a realistic evaluation under unseen-battery conditions, the present study proposes a comprehensive cross-battery SOH estimation framework based on Leave-One-Battery-Out (LOBO) validation using the NASA lithium-ion battery degradation dataset. A total of twenty-one statistical and physics-inspired features are extracted from voltage, current, temperature, discharge time, energy, and power measurements to characterize battery degradation behavior. Four representative prediction models, namely Random Forest (RF), Least Squares Boosting (LSBoost), Long Short-Term Memory (LSTM), and Support Vector Regression (SVR), are comparatively evaluated under an identical LOBO validation framework to provide a fair assessment of their generalization performance. Furthermore, predictor importance analysis is performed to identify the most influential degradation indicators and to enhance the interpretability of the proposed framework. The findings of this study offer practical guidance for developing reliable and robust SOH estimation models suitable for next-generation battery management systems.

The rest of this paper is structured as follows: Section 2 describes the NASA battery degradation dataset, the proposed feature extraction procedure, and the adopted machine learning models. Section 3 discusses the comparative results, prediction performance, and feature importance analysis. Finally, Section 4 summarizes the main findings of this study and outlines potential directions for future research.

II. MATERIALS AND METHOD

A. Battery Dataset Description

The experimental data used in this study were obtained from the publicly available NASA Prognostics Center of Excellence (PCoE) lithium-ion battery dataset, which has been widely adopted as a benchmark database for battery degradation analysis, remaining useful life prediction, and state-of-health (SOH) estimation studies [35]. The dataset was generated by NASA Ames Research Center under controlled laboratory conditions and contains comprehensive aging experiments, including charge, discharge, and impedance measurements collected throughout the battery lifetime.

In this study, four lithium-ion battery cells, namely B0005, B0006, B0007, and B0018, were selected for SOH estimation analysis. These cells were subjected to repeated charge and discharge cycles until their capacities degraded below the predefined end-of-life threshold. The recorded measurements include voltage, current, temperature, time, and discharge capacity information for each operational cycle [35].

The discharge capacity obtained from each cycle was considered as the primary degradation indicator, and SOH was calculated using the capacity-based definition:

$$SOH\% = \frac{C_{current}}{C_{rated}} \times 100 \tag{1}$$

where $C_{current}$ represents the measured discharge capacity at a given cycle and C_{rated} denotes the nominal capacity of the battery.

The selected NASA cells exhibit different degradation trajectories and lifetime characteristics, providing a suitable benchmark for evaluating the generalization capability of machine learning-based SOH estimation models. Unlike conventional random data splitting strategies, this study considers each battery cell as an independent degradation domain, enabling a realistic evaluation of prediction performance on previously unseen batteries.

B. SOH Calculation and Feature Engineering

The performance of data-driven SOH estimation models strongly depends on the quality and representativeness of the extracted health indicators (HIs). Since battery degradation is reflected through variations in electrical and thermal characteristics, feature engineering has become a critical step for transforming raw operational measurements into meaningful degradation-related descriptors [25,33]. In this study, statistical and physics-inspired features were extracted from the measured voltage, current, temperature, discharge time, and energy profiles to characterize the degradation behaviour of lithium-ion batteries.

The discharge capacity degradation was first converted into the SOH indicator, which was subsequently used as the prediction target for all machine learning models. Instead of directly utilizing raw measurement sequences, representative statistical descriptors were generated to reduce data dimensionality and improve the robustness of the learning process. Previous studies have demonstrated that voltage-based, current-based, and temperature-related indicators can effectively capture battery aging trends and provide informative inputs for SOH prediction models [12,33].

A total of twenty-one features were extracted from each discharge cycle. These features were grouped into four main categories: voltage-related features, current-related features, temperature-related features, and operational/energy-related features. Voltage-based indicators included mean voltage, maximum voltage, minimum voltage, voltage range, voltage root mean square (RMS), and voltage standard deviation. Current-related indicators consisted of mean current, current RMS, and current standard deviation. Thermal characteristics were represented by mean temperature, maximum temperature, minimum temperature, and temperature variation. In addition, discharge duration, energy consumption, power-related descriptors, and voltage degradation trends were included to capture cycle-level aging behaviour.

The extracted features can be expressed as:

$$X_i = [V_{mean}, V_{max}, V_{min}, V_{std}, , I_{mean}, T_{mean}, T_{max}, T_{min}, \Delta T, t_d, E, V_{range}, V_{rms}, I_{rms}, I_{std}]$$

where X_i represents the feature vector obtained from the i^{th} discharge cycle. Here, t_d denotes discharge duration, E represents the delivered energy.

Feature engineering based on statistical descriptors provides two important advantages: (i) it reduces the computational burden compared with processing complete time-series signals, and (ii) it improves model interpretability by linking prediction behavior with physically meaningful battery degradation characteristics [26,33]. Therefore, the extracted feature set was used as the common input space for all investigated machine learning algorithms under the proposed LOBO validation framework.

C. Machine Learning Models

To assess the performance of various learning approaches for lithium-ion battery SOH estimation, four representative machine learning models were selected in this study: Random Forest (RF), Least Squares Boosting (LSBoost), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) networks. These models were selected to provide a comprehensive comparison between ensemble learning, kernel-based regression, and deep learning approaches under an identical validation framework. Previous studies have demonstrated that different machine learning architectures exhibit distinct advantages depending on feature characteristics, available degradation data, and operating conditions [12,17,24].

a) Random Forest

Random Forest (RF) is an ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy and reduce overfitting. Each tree is trained using randomly selected subsets of samples and features, while the final prediction is obtained by averaging the outputs of individual trees. Due to its robustness against noise, capability of handling nonlinear relationships, and intrinsic feature importance estimation ability, RF has been widely applied in battery degradation analysis and SOH estimation problems [26,29].

In this study, RF was employed as a baseline ensemble regression model to evaluate the relationship between extracted health indicators and battery SOH degradation. Furthermore, the predictor importance obtained from the RF model was used to identify the contribution of individual features to SOH estimation.

b) Least Squares Boosting

Least Squares Boosting (LSBoost) is a boosting-based ensemble regression method that sequentially combines weak learners to minimize prediction errors. Unlike bagging approaches such as RF, boosting algorithms focus on correcting the errors of previously generated models, enabling improved prediction performance for complex nonlinear relationships.

LSBoost has been widely adopted in regression problems where degradation trends exhibit nonlinear characteristics. In battery SOH estimation, boosting-based methods have demonstrated competitive performance due to their ability to combine multiple weak predictors and improve generalization capability [17,29].

c) Support Vector Regression

Support Vector Regression (SVR) is a regression algorithm that employs kernel-based learning capable of modelling nonlinear relationships by mapping input features into a higher-dimensional feature space through kernel functions. The algorithm achieves robust regression performance by minimizing prediction errors within an insensitive loss function while controlling model complexity.

SVR has been extensively investigated for lithium-ion battery SOH estimation because of its strong generalization ability, especially when the available degradation data are limited [10,12]. In this study, SVR was included as a representative traditional machine learning model for comparison with ensemble and deep learning approaches.

d) Long Short-Term Memory Network

Long Short-Term Memory (LSTM) networks are a class of recurrent neural networks capable of learning long-term dependencies from sequential data. Through memory cells and gating mechanisms, LSTM networks can effectively learn temporal degradation patterns from battery cycling data and have become one of the most widely used deep learning architectures for battery state estimation applications [15,16].

Although LSTM models are particularly suitable for raw time-series analysis, they were included in this study using the extracted cycle-level features to provide a fair comparison with conventional machine learning models under the same feature space. This approach enables an objective evaluation of whether deep learning architectures provide additional advantages when high-quality engineered features are available.

D. Leave-One-Battery-Out Validation Framework

A major challenge in data-driven battery SOH estimation is the reliable evaluation of model generalization capability. Many existing studies employ random training and testing data splitting strategies, where samples from the same battery cell may appear in both training and testing subsets. Although this approach can provide high prediction accuracy, it may lead to data leakage and overly optimistic performance estimation because the model can implicitly learn the degradation trajectory of the same battery instead of capturing general degradation characteristics [26].

To overcome this limitation, this study adopts a Leave-One-Battery-Out (LOBO) validation strategy, where each battery cell is completely excluded from the training process and used only as an unseen test battery. This evaluation protocol provides a more realistic representation of practical battery management system (BMS) applications, where a prediction model trained using historical battery data is expected to estimate the SOH of a new battery with unknown degradation behavior [23,31].

In the proposed framework, the NASA battery dataset containing four cells (B0005, B0006, B0007, and B0018) was divided into four independent validation scenarios. For each iteration, one battery was selected as the test battery, while the other three batteries were utilized for model training. The validation procedure can be summarized as follows:

Scenario 1:
- Training batteries: B0006, B0007, B0018
- Testing battery: B0005

Scenario 2:
- Training batteries: B0005, B0007, B0018
- Testing battery: B0006

Scenario 3:
- Training batteries: B0005, B0006, B0018
- Testing battery: B0007

Scenario 4:
- Training batteries: B0005, B0006, B0007
- Testing battery: B0018

For each validation scenario, identical feature sets, preprocessing procedures, and machine learning models were applied to ensure a fair comparison among RF, LSBoost, SVR, and LSTM algorithms. The final performance indicators were calculated by aggregating the prediction results obtained from all four independent battery-level experiments.

The overall LOBO procedure can be mathematically expressed as:

$$\mathrm{D}_{\mathrm{train}}^{(\mathrm{i})} = \mathrm{D}_{\mathrm{all}} \setminus \mathrm{D}_{\mathrm{i}} \tag{2}$$

$$\mathrm{D}_{\mathrm{test}}^{(\mathrm{i})} = \mathrm{D}_{\mathrm{i}} \tag{3}$$

where $\mathrm{D_i}$ represents the dataset belonging to the i^{th} battery cell. Therefore, the test battery data are never observed during the training phase, providing an unbiased evaluation of model robustness and transferability to unseen batteries [23,32].

E. Performance Evaluation Metrics

To quantitatively assess the prediction performance of the proposed SOH estimation models, three widely adopted regression metrics were employed: root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). These metrics have been extensively used in battery health estimation studies to assess both prediction accuracy and the capability of machine learning models to capture degradation trends [25,26].

The root mean square error (RMSE) reflects the discrepancy between the predicted and actual SOH values while giving higher importance to larger prediction errors. It is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2} \tag{4}$$

where $\mathrm{y_i}$ represents the measured SOH value, $\hat{\mathrm{y}}_\mathrm{i}$ denotes the predicted SOH value, and N is the number of testing samples.

The mean absolute error (MAE) represents the average absolute difference between predicted and experimental SOH values and provides a direct interpretation of prediction deviation:

$$MAE = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}|\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}| \tag{5}$$

The coefficient of determination (R^2) was adopted to evaluate how well the prediction models explain the variance of the measured SOH values. It is expressed as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{6}$$

where y¯ represents the mean value of the measured SOH observations. A higher R^2 value indicates stronger agreement between the predicted and actual degradation behavior.

In this study, RMSE, MAE, and R^2 values were calculated separately for each battery under the LOBO validation framework. The final model comparison was performed using the averaged performance obtained from all unseen battery prediction scenarios, ensuring an unbiased evaluation of model generalization capability.

III. RESULTS

A. Comparative Performance of Machine Learning Models

The proposed Leave-One-Battery-Out (LOBO) validation framework was applied to evaluate the generalization capability of four representative machine learning algorithms, namely Random Forest (RF), Least Squares Boosting (LSBoost), Long Short-Term Memory (LSTM), and Support Vector Regression (SVR). Unlike conventional random train–

test splitting, the adopted validation strategy ensured that the testing battery remained completely unseen during model training, thereby providing a more realistic assessment of model robustness under cross-battery prediction scenarios.

Table 1. Comparison of SOH prediction performance of different models

Model	RMSE	MAE	R²
Random Forest	**0.0342**	**0.0305**	**0.8654**
LSBoost	0.0360	0.0314	0.8567
LSTM	0.0363	0.0344	0.8400
SVR	0.1097	0.0957	−0.2222

As observed in Table 1, the Random Forest model achieved the highest prediction performance among all evaluated algorithms, producing the lowest RMSE (0.0342) and MAE (0.0305) while simultaneously yielding the highest coefficient of determination (R^2=0.8654). These results indicate that the ensemble-based learning strategy successfully captured the nonlinear relationship between the extracted health indicators and battery degradation patterns under unseen-battery conditions.

LSBoost demonstrated prediction accuracy comparable to that of RF, with only a marginal increase in RMSE (approximately 5%) and a slightly lower R^2. The relatively small performance difference between these two ensemble learning methods suggests that tree-based ensemble algorithms are particularly effective for learning the complex interactions among voltage, current, temperature, and energy-related degradation features extracted from battery cycling data.

Although LSTM has been widely reported as one of the most powerful deep learning architectures for battery health prediction, its performance in this study was slightly inferior to that of the ensemble learning models. This outcome is reasonable because the LSTM model was trained using cycle-level engineered features rather than complete sequential time-series measurements. Consequently, the temporal modeling capability of LSTM could not be fully exploited, whereas RF and LSBoost benefited from their ability to efficiently learn nonlinear feature interactions within structured tabular data.

Among the evaluated algorithms, SVR exhibited substantially lower prediction accuracy, resulting in the highest prediction errors and a negative coefficient of determination. The negative R^2 value indicates that the SVR model failed to adequately describe the degradation behaviour of previously unseen batteries under the adopted LOBO validation framework. This observation suggests that kernel-based regression methods may have limited generalization capability when degradation trajectories differ significantly between battery cells.

Overall, the obtained results demonstrate that ensemble learning algorithms provide the most reliable and robust SOH estimation performance under realistic cross-battery prediction scenarios. More importantly, the proposed LOBO validation framework effectively reveals differences in model generalization that might remain hidden when conventional random train–test splitting strategies are employed.

B. Overall Performance Comparison of the Prediction Models

To comprehensively evaluate the proposed battery SOH prediction framework, four machine learning algorithms, namely Random Forest (RF), Least Squares Boosting (LSBoost), Long Short-Term Memory (LSTM), and Support Vector Regression (SVR), were evaluated using the Leave-One-Battery-Out (LOBO) validation strategy. This evaluation protocol is considerably more challenging than conventional random train-test splitting because the testing battery is completely excluded from the training stage. Consequently, the obtained results provide a more realistic assessment of the models' capability to generalize to previously unseen batteries.

The average prediction performance of the four models is summarized in Fig. 1, where the comparison is presented using RMSE, MAE, and coefficient of determination (R^2). As illustrated in Fig. 1(a)**,** the Random Forest model achieved the lowest average RMSE among all evaluated approaches, indicating the highest prediction accuracy. LSBoost exhibited a very similar performance with only a marginal increase in prediction error, whereas LSTM produced slightly higher RMSE values. In contrast, SVR showed a substantially larger prediction error, demonstrating its limited capability to capture the nonlinear degradation characteristics of lithium-ion batteries under cross-battery validation. Similarly, the average MAE values shown in Fig. 1(b) reveal that RF consistently produced the smallest absolute prediction errors. LSBoost again ranked second, while LSTM generated moderately larger deviations from the ground truth SOH values. SVR presented the highest MAE by a considerable margin, indicating poor robustness when applied to batteries not included in the training dataset. The comparison of the coefficient of determination in Fig. 1(c) further confirms these observations. Random Forest achieved the highest average R^2 value, followed closely by LSBoost. Although LSTM maintained reasonably high predictive capability, its performance remained inferior to the ensemble-based algorithms. Notably, the SVR model produced a negative R^2 value, indicating that its predictions were worse than simply using the average SOH value of the testing battery. Such behavior clearly demonstrates the inability of conventional kernel-based regression methods to generalize under the challenging LOBO validation scenario.

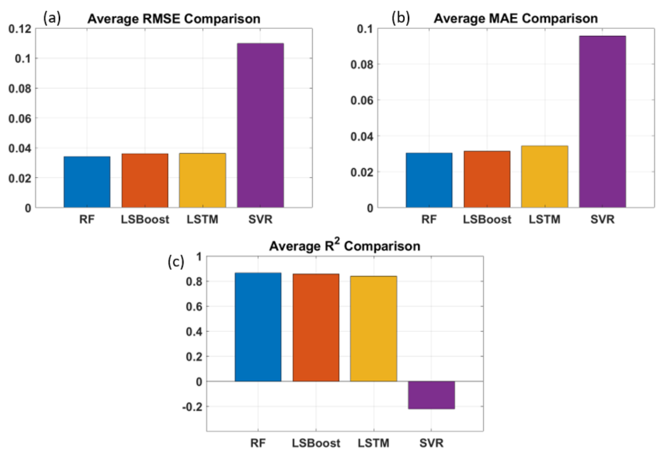

Fig. 1. Performance comparison of the evaluated machine learning models under the leave-one-battery-out (LOBO) validation strategy based on (a) root mean square error (RMSE), (b) mean absolute error (MAE), and (c) coefficient of determination (R²).

Overall, the comparative results indicate that ensemble learning approaches, particularly Random Forest, provide superior robustness and generalization capability for lithium-ion battery SOH estimation. These findings are consistent with recent studies reporting that tree-based ensemble models effectively capture nonlinear degradation characteristics while maintaining strong resistance to overfitting in practical battery applications [17], [24], [26], [34].

C. SOH Prediction under Leave-One-Battery-Out Validation

To further investigate the generalization capability of the evaluated models, SOH prediction curves obtained for each testing battery are presented in Fig. 2. Unlike conventional validation strategies, the LOBO protocol evaluates each model using a battery that has never been observed during training, thereby providing a realistic simulation of practical battery health prediction scenarios. As shown in Fig. 2, all three advanced models (RF, LSBoost, and LSTM) successfully captured the overall degradation tendency of the batteries throughout the discharge cycles. Nevertheless, noticeable differences can be observed in their ability to reproduce local degradation behaviors and abrupt SOH variations.

For Battery B0005, all models closely followed the measured degradation trajectory, although Random Forest demonstrated the closest agreement with the experimental SOH values throughout the entire lifecycle. LSBoost also achieved competitive performance, whereas LSTM exhibited slightly larger fluctuations during the later degradation stages. Battery B0006 represents the most challenging prediction scenario because of its relatively rapid degradation behavior. In this case, Random Forest maintained the closest approximation to the actual SOH evolution, while both LSBoost and LSTM tended to overestimate battery health during the later discharge cycles. This behavior suggests that deep learning models may require substantially larger training datasets to accurately capture complex degradation dynamics. For Battery B0007, the prediction performances of Random Forest and LSBoost remained highly consistent with the measured SOH values, whereas LSTM underestimated several local degradation patterns. Similar observations can be made for Battery B0018, where Random Forest consistently followed the measured SOH trajectory with relatively small deviations throughout the complete degradation process.

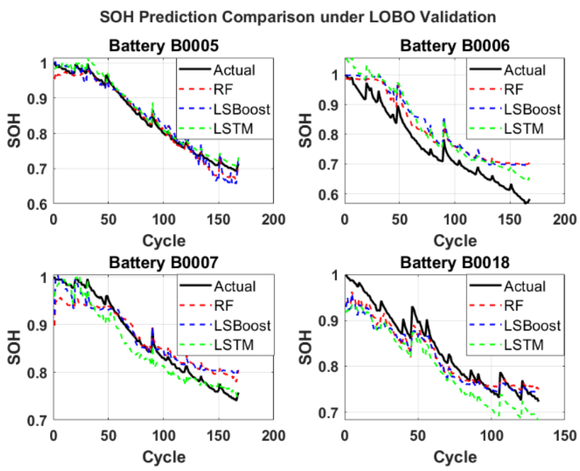

Fig. 2. Comparison of SOH prediction results obtained using Random Forest (RF), LSBoost, and LSTM models for the four NASA batteries under leave-one-battery-out (LOBO) validation.

Overall, Random Forest demonstrated the most stable prediction performance across all four batteries, indicating superior robustness against variations in battery degradation characteristics. This consistency highlights the effectiveness of ensemble learning in extracting generalized degradation information from heterogeneous battery datasets. Similar observations have recently been reported in battery SOH literature, where ensemble-based machine learning models frequently outperform deep neural networks when the available training data are relatively limited [17], [25], [31].

D. Analysis of Random Forest SOH Prediction Performance

To provide a more detailed evaluation of the proposed framework, Fig. 3 presents the measured and predicted SOH profiles of batteries B0005, B0006, B0007, and B0018 obtained using the best-performing model, namely Random Forest (RF), under the Leave-One-Battery-Out (LOBO) validation framework. Since Random Forest consistently achieved the lowest prediction errors and the highest coefficient of determination among all evaluated models, it was selected for detailed analysis. As illustrated in Fig. 3, the predicted SOH values closely follow the experimentally measured degradation trajectory throughout the entire battery lifetime. The Random Forest model successfully captures both the overall downward degradation trend and the local fluctuations that naturally occur during battery aging. This behavior indicates that the proposed feature engineering strategy effectively preserves the degradation characteristics required for accurate SOH estimation.

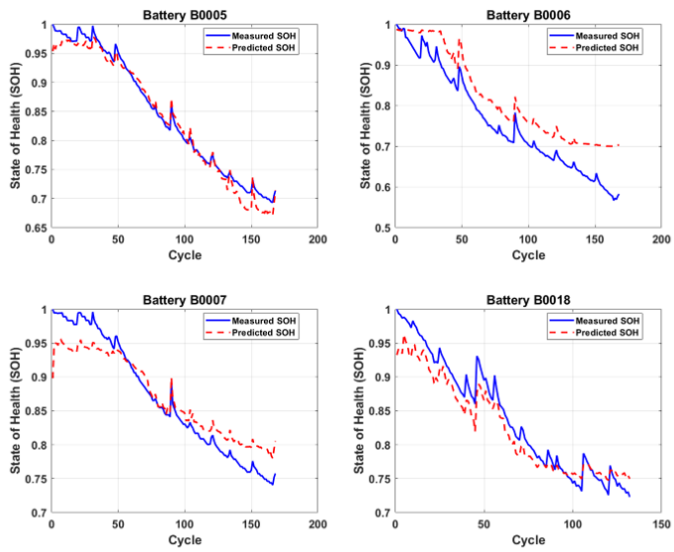

Fig. 3. Measured and predicted state of health (SOH) profiles obtained by the Random Forest (RF) model under the Leave-One-Battery-Out (LOBO) validation framework for (a) B0005, (b) B0006, (c) B0007, and (d) B0018.

A particularly important observation is that the prediction accuracy remains stable across both the early and late stages of battery degradation. During the initial cycles, where the SOH changes relatively slowly, the predicted values almost overlap with the measured data. As the battery approaches its end-of-life region, where degradation becomes more nonlinear and irregular, the prediction errors increase only slightly while the model continues to follow the overall degradation trajectory. This demonstrates the strong nonlinear approximation capability of the Random Forest algorithm.

Furthermore, the model exhibits good robustness against temporary capacity recovery phenomena, which appear as short-term increases in SOH during battery cycling. Such regeneration effects are commonly observed in lithium-ion batteries due to electrochemical relaxation processes and often degrade the performance of conventional regression models. Nevertheless, the Random Forest model accurately tracks these local variations without introducing significant prediction oscillations.

The close agreement between the measured and predicted SOH curves confirms that the proposed feature extraction procedure successfully captures the essential degradation information contained in the discharge cycles. Instead of relying on a single health indicator, the proposed framework combines voltage-, current-, temperature-, power-, and energy-related descriptors, enabling the model to learn complex nonlinear degradation relationships.

Overall, the results presented in Fig. 3 demonstrate that Random Forest provides highly reliable SOH prediction under the challenging LOBO validation protocol. The obtained performance suggests that the proposed framework possesses strong potential for practical battery management systems, where the prediction model must operate on batteries that were not included during model training.

While the prediction curves presented in Fig. 3 visually demonstrate the capability of the Random Forest model, the quantitative prediction results are summarized in Table 2. The model achieved the highest prediction accuracy for battery B0005 with an RMSE of 0.0150 and an R^2 value of 0.9783. Conversely, battery B0006 exhibited the largest prediction error (RMSE = 0.0648), indicating that its degradation trajectory is comparatively more difficult to generalize under the strict LOBO validation framework. Nevertheless, the Random Forest model consistently maintained satisfactory prediction accuracy across all four batteries, with R^2 values ranging from 0.7242 to 0.9783.

Table 2. Battery-wise prediction performance of the Random Forest model under the LOBO validation framework.

Test Battery	RMSE	MAE	R^2
B0005	0.0150	0.0122	0.9783
B0006	0.0648	0.0601	0.7242
B0007	0.0290	0.0246	0.8832
B0018	0.0238	0.0203	0.9183

The statistical summary presented in Table 3 further demonstrates the consistency of the proposed Random Forest framework across different batteries. The average RMSE and MAE values were 0.0332 and 0.0293, respectively, while the mean coefficient of determination reached 0.8760. Although some performance variation exists among the four batteries, as reflected by the standard deviations, the obtained results indicate stable prediction capability under the strict LOBO validation protocol. The relatively low standard deviations of RMSE and MAE suggest that the proposed framework maintains consistent estimation accuracy despite battery-to-battery variability, supporting its applicability for practical battery management systems.

Table 3. Statistical summary of the Random Forest prediction performance across all LOBO validation experiments.

Metric	Mean	Standard Deviation	Best	Worst
RMSE	0.0332	0.0219	0.0150	0.0648
MAE	0.0293	0.0212	0.0122	0.0601
R^2	0.8760	0.1086	0.9783	0.7242

E. Practical Implications

Fig. 4 presents the distribution of prediction errors obtained by Random Forest model for each battery under the LOBO validation framework. The prediction errors for B0005 and B0007 are centered very close to zero with relatively small dispersion, indicating highly accurate and nearly unbiased SOH estimation. Although B0018 exhibits a slight positive bias, its prediction errors remain confined within a narrow range, demonstrating stable model performance. In contrast, B0006 shows the largest error spread with a tendency toward negative prediction errors, which is consistent with its comparatively higher RMSE reported in Table 2. This observation suggests that the degradation trajectory of B0006 differs more substantially from those of the remaining batteries, making cross-battery prediction inherently more challenging. Nevertheless, the overall prediction error distributions remain sufficiently compact, confirming the robustness and generalization capability of the proposed Random Forest model when evaluated on previously unseen batteries.

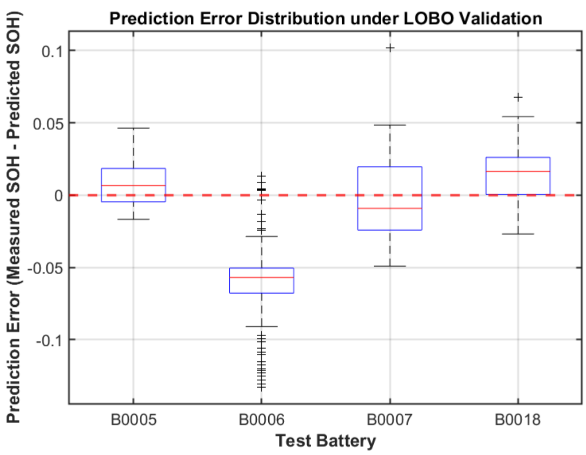

Fig. 4. Distribution of Random Forest prediction errors for each NASA battery under the Leave-One-Battery-Out (LOBO) validation framework. The dashed horizontal line represents zero prediction error.

Fig. 5 provides further insight into the decision-making process of the proposed Random Forest model through feature importance analysis. Fig. 5 presents the Top-10 most influential features identified by the Random Forest model under the LOBO validation framework. The feature importance analysis reveals that the proposed SOH estimation framework relies primarily on voltage-, energy-, and discharge-related descriptors, indicating that these variables contain the most discriminative information regarding battery degradation. Several temperature- and current-related features also contribute to the prediction process, although their relative importance is lower. This result demonstrates that the proposed feature engineering strategy successfully captures multiple degradation characteristics rather than depending on a single health indicator.

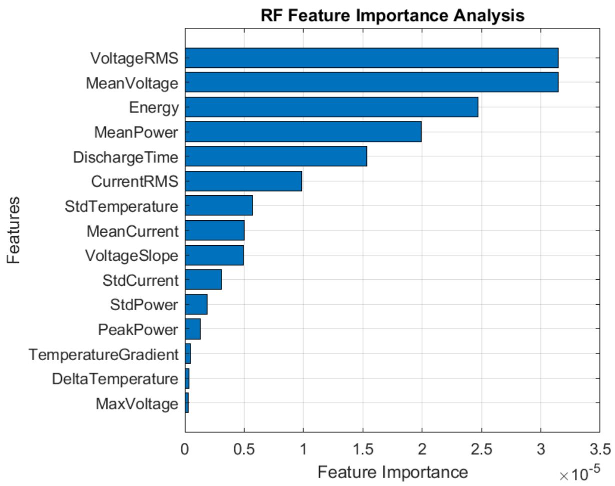

Fig. 5. Top-10 feature importance scores obtained from the Random Forest model under the Leave-One-Battery-Out (LOBO) validation framework.

Overall, the results presented in Figures 4 and 5 illustrate that the proposed framework delivers reliable prediction performance while providing valuable insights into the decision-making process of the Random Forest model. The combination of feature importance analysis and prediction error evaluation confirms that the proposed feature engineering strategy is both interpretable and reliable under the stringent LOBO validation protocol. These findings highlight the practical applicability of the proposed approach for battery management systems, where reliable SOH estimation of previously unseen batteries is essential for safe and efficient operation.

Among the evaluated machine learning algorithms, Random Forest consistently achieved the most balanced performance by combining high prediction accuracy with strong robustness across different batteries. Although LSTM is capable of modeling temporal dependencies, its performance is more sensitive to the limited amount of available training data and hyperparameter selection. Similarly, LSBoost and SVR produced competitive results but exhibited relatively larger prediction errors for some batteries under the strict LOBO validation protocol. These findings indicate that ensemble learning algorithms remain highly competitive for practical SOH estimation problems involving relatively small and heterogeneous battery datasets.

Despite these promising outcomes, several limitations remain. The proposed framework was evaluated using the NASA lithium-ion battery degradation dataset, which contains a limited number of battery cells tested under laboratory conditions. Although the LOBO protocol significantly improves evaluation realism, further validation using larger datasets with different battery chemistries, operating temperatures, charging protocols, and real-world driving conditions would provide additional evidence regarding the generalization capability of the proposed approach. Furthermore, future studies may investigate transfer learning, domain adaptation, uncertainty quantification, and explainable artificial intelligence techniques to further improve model robustness and practical applicability.

IV. CONCLUSION

This study presented a comprehensive machine learning-based framework for lithium-ion battery state-of-health (SOH) estimation using a feature engineering approach combined with a stringent Leave-One-Battery-Out (LOBO) validation strategy. Unlike conventional random train-test splitting methods, the adopted evaluation protocol provides a more realistic assessment of model generalization by evaluating the prediction capability on previously unseen batteries.

Four machine learning algorithms, including Random Forest, LSBoost, Support Vector Regression, and Long Short-Term Memory networks, were systematically evaluated using the NASA lithium-ion battery degradation dataset. The results demonstrated that the Random Forest model achieved the most consistent and balanced performance across different battery cells, providing accurate SOH estimation while maintaining strong robustness under cross-battery evaluation conditions. The obtained results confirm that ensemble learning approaches remain highly effective for SOH estimation problems involving limited and heterogeneous battery datasets.

Furthermore, the proposed feature engineering strategy successfully captured multiple degradation characteristics by integrating voltage-, current-, temperature-, power-, energy-, and discharge-related descriptors. The feature importance analysis revealed that battery degradation information is distributed among several complementary indicators rather than being represented by a single health parameter. This finding highlights the importance of comprehensive feature extraction for developing accurate and interpretable battery health monitoring systems.

The combination of accurate prediction performance, LOBO-based evaluation, and model interpretability demonstrates the practical potential of the proposed framework for battery management systems, where reliable SOH estimation of unknown batteries is essential for safe and efficient operation. However, the current study is limited by the number of available cells and the laboratory-based conditions of the NASA dataset. Future research will focus on validating the proposed framework using larger multi-source datasets, different battery chemistries, and real-world operating conditions. Additionally, advanced approaches such as transfer learning, domain adaptation, uncertainty quantification, and explainable artificial intelligence methods can be investigated to further enhance the robustness and deployment capability of battery SOH estimation models.

Statement of Research and Publication Ethics

The author declare that this study complies with Research and Publication Ethics

REFERENCES

[1] R. Xiong, L. Li, and J. Tian, “Towards a smarter battery management system: A critical review on battery state of health monitoring methods,” *Journal of Power Sources*, vol. 405, pp. 18–29, 2018.

[2] A. Barré, B. Deguilhem, S. Grolleau, M. Gérard, F. Suard, and D. Riu, “A review on lithium-ion battery ageing mechanisms and estimations for automotive applications,” *Journal of Power Sources*, vol. 241, pp. 680–689, 2013.

[3] W. Waag, C. Fleischer, and D. U. Sauer, “Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles,” *Journal of Power Sources*, vol. 258, pp. 321–339, 2014.

[4] X. Hu, H. Yuan, C. Zou, Z. Li, and L. Zhang, "Co-estimation of state of charge and state of health for lithium-ion batteries based on fractional-order calculus," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10319–10329, 2018.

[5] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche, "Critical review of state of health estimation methods of Li-ion batteries for real applications," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 572–587, 2016.

[6] K. Tang, B. Luo, D. Chen, C. Wang, L. Chen, F. Li, et al., "The state of health estimation of lithium-ion batteries: A review of health indicators, estimation methods, development trends and challenges," *World Electric Vehicle Journal*, vol. 16, no. 8, p. 429, 2025.

[7] K. Movassagh, A. Raihan, B. Balasingam, and K. Pattipati, "A critical look at coulomb counting approach for state of charge estimation in batteries," *Energies*, vol. 14, no. 14, p. 4074, 2021.

[8] C. Vidal, P. Malysz, P. Kollmeyer, and A. Emadi, "Machine learning applied to electrified vehicle battery state of charge and state of health estimation: State-of-the-art," *IEEE Access*, vol. 8, pp. 52796–52814, 2020.

[9] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation," *Journal of Power Sources*, vol. 134, no. 2, pp. 277–292, 2004.

[10] S. Patil, S. M. Havaldar, S. Mathad, and K. R. Patil, "Lithium-ion battery state of health estimation using support vector regression (SVR)," in *Proc. 2023 Int. Conf. Ambient Intell., Knowl. Informat. Ind. Electron. (AIKIIE)*, 2023, pp. 1–6.

[11] Y. Li, X. Huang, J. Meng, K. Shi, R. Teodorescu, and D. I. Stroe, "State of health estimation for lithium-ion battery based on sample transfer learning under current pulse test," *Batteries*, vol. 10, no. 5, p. 156, 2024.

[12] X. Sui, S. He, S. B. Vilsen, J. Meng, R. Teodorescu, and D. I. Stroe, "A review of non-probabilistic machine learning-based state of health estimation techniques for Lithium-ion battery," *Applied Energy*, vol. 300, p. 117346, 2021.

[13] Z. Zhou, A. Aitio, and D. Howey, "Learning Li-ion battery health and degradation modes from data with aging-aware circuit models," *Applied Energy*, vol. 397, p. 126375, 2025.

[14] R. R. Richardson, C. R. Birkl, M. A. Osborne, and D. A. Howey, "Gaussian process regression for in situ capacity estimation of lithium-ion batteries," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 127–138, 2018.

[15] E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi, "Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6730–6739, 2017.

[16] Q. Yao, X. Song, and W. Xie, "State of health estimation of lithium-ion battery based on CNN–WNN–WLSTM," *Complex & Intelligent Systems*, vol. 10, no. 2, pp. 2919–2936, 2024.

[17] Y. Wang, S. Guo, Y. Cui, L. Deng, L. Zhao, J. Li, and Z. Wang, "A comprehensive review of machine learning-based state of health estimation for lithium-ion batteries: data, features, algorithms, and future challenges," *Renewable and Sustainable Energy Reviews*, vol. 224, p. 116125, 2025.

[18] S. Wang, B. Hu, Q. Zhang, C. Fernandez, and J. M. Guerrero, "Hybrid models for predicting the state of health of lithium-ion batteries: A comprehensive review," *Journal of Energy Storage*, vol. 157, p. 121600, 2026.

[19] J. Guirguis and R. Ahmed, "Transformer-based deep learning models for state of charge and state of health estimation of li-ion batteries: A survey study," *Energies*, vol. 17, no. 14, p. 3502, 2024.

[20] Z. Li, X. Zhang, and W. Gao, "State of health estimation of lithium-ion battery during fast charging process based on BiLSTM-Transformer," *Energy*, vol. 311, p. 133418, 2024.

[21] Y. Zhang, Z. Li, L. Kong, H. Xu, H. Shen, and M. Chen, "Multi-step state of health prediction of lithium-ion batteries based on multi-feature extraction and improved Transformer," *Journal of Energy Storage*, vol. 105, p. 114538, 2025.

[22] W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method," *Journal of Power Sources*, vol. 196, no. 23, pp. 10314–10321, 2011.

[23] K. Maher and N. Yerken, "Comprehensive machine learning for lithium-ion battery state-of-health estimation using group-wise cross-validation," in *Proc. 2025 14th Int. Conf. Renew. Energy Res. Appl. (ICRERA)*, 2025, pp. 1465–1468.

[24] G. R. Sylvestrin, J. N. Maciel, M. L. M. Amorim, J. P. Carmo, J. A. Afonso, S. F. Lopes, and O. H. Ando Junior, "State of the art in electric batteries' state-of-health (soh) estimation with machine learning: A review," *Energies*, vol. 18, no. 3, p. 746, 2025.

[25] Y. Li, K. Liu, A. M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, et al., "Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review," *Renewable and Sustainable Energy Reviews*, vol. 113, p. 109254, 2019.

[26] D. Roman, S. Saxena, V. Robu, M. Pecht, and D. Flynn, "Machine learning pipeline for battery state-of-health estimation," *Nature Machine Intelligence*, vol. 3, no. 5, pp. 447–456, 2021.

[27] J. Lu, R. Xiong, J. Tian, C. Wang, and F. Sun, "Deep learning to estimate lithium-ion battery state of health without additional degradation experiments," *Nature Communications*, vol. 14, no. 1, p. 2760, 2023.

[28] K. Q. Zhou, Y. Qin, and C. Yuen, "Graph neural network-based lithium-ion battery state of health estimation using partial discharging curve," *Journal of Energy Storage*, vol. 100, p. 113502, 2024.

[29] R. Pandit and N. Ahlawat, "A standardized comparative framework for machine learning techniques in lithium-ion battery state of health estimation," *Future Batteries*, vol. 7, p. 100099, 2025.

[30] J. Wu, Z. Liu, Y. Zhang, D. Lei, and Y. Zhang, "A model to estimate the state of health of lithium-ion batteries for different charging rates," *Future Batteries*, vol. 6, p. 100074, 2025.

[31] M. F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, "Predicting the state of charge and health of batteries using data-driven machine learning," *Nature Machine Intelligence*, vol. 2, no. 3, pp. 161–170, 2020.

[32] Y. Yang, Y. Xu, Y. Nie, J. Li, S. Liu, L. Zhao, et al., "Deep transfer learning enables battery state of charge and state of health estimation," *Energy*, vol. 294, p. 130779, 2024.

[33] Q. Li and W. Xue, "A review of feature extraction toward health state estimation of lithium-ion batteries," *Journal of Energy Storage*, vol. 112, p. 115453, 2025.

[34] H. Li, X. Xie, X. Zhang, A. F. Burke, and J. Zhao, "Battery state estimation for electric vehicles: Translating AI innovations into real-world solutions," *Journal of Energy Storage*, vol. 115, p. 116000, 2025.

[35] B. Saha and K. Goebel, "Battery Data Set," NASA Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, 2007.