KFUPM
Journal of Undergraduate Research International

Diamond Open Access 1

Article Aerospace Engineering

Development of a Data-Driven Model or Predicting the State of Health of the Li-ion Battery in an Unmanned Aerial Vehicle

Jana Eid Aljohani and Mohammad Irfan Alam*

Cite https://doi.org/10.64589/juri/231542

Submitted: April 23, 2026 **Revised:** May 05, 2026 **Accepted:** May 12, 2026

ABSTRACT

Lithium-ion batteries are key enablers of sustainable aviation; however, gradual battery degradation poses critical safety and reliability challenges, particularly for unmanned aerial vehicles (UAVs) that operate in high-temperature environments. This study developed a data-driven model for predicting the state-of-health (SOH) of Li-ion batteries using machine learning. The NASA Prognostics Center of Excellence Li-ion battery aging dataset (battery #5) was used to train and test a random forest regression model with features selected to capture the effects of high-temperature operation and voltage behavior. The proposed random forest model was benchmarked against a linear regression baseline and achieved highly accurate SOH estimation with a coefficient of determination of 0.9077 and low root mean square error (0.2578) and mean absolute error (0.2185). These results provide a reliable foundation for estimating the remaining useful life of a battery. The proposed approach supports UAV reliability and safe flight operations in high-temperature environments.

Keywords: lithium-ion batteries; state of health (SOH); remaining useful life (RUL); machine learning; random forest; UAV reliability; data-driven model

1. INTRODUCTION

In recent years, the aviation industry has significantly shifted toward electric propulsion as part of broader efforts to reduce carbon emissions and noise pollution [1,2]. Within this transition, unmanned aerial vehicles (UAVs) and electric aircraft are emerging as cleaner and more efficient alternatives that reduce environmental impact [3]. In the Middle East, UAVs face heightened challenges because the high ambient temperatures accelerate internal chemical reactions inside their batteries, shorten their service lives, and increase the likelihood of sudden failure. As the core enabler of the transition to electric aircraft, lithium (Li)-ion batteries serve as the primary energy-storage systems for electric UAVs because of their high energy density and power capability, which are essential for satisfying aviation safety and reliability requirements [4]. Throughout their operational lifespan, lithium-ion batteries undergo capacity fading because of the complex electrochemical processes that occur during repeated charge/discharge cycles [5]. The operational safety and reliability of battery systems critically depend on the accurate monitoring of their state of health (SOH) to prevent unexpected failures during flights. The SOH is defined as the ratio of the battery's current usable capacity to its initial nominal capacity and is the main parameter used to quantify battery degradation. Accurately estimating the SOH is challenging because it depends nonlinearly on operating variables, such as temperature and voltage [6]. Traditional model-based monitoring methods often struggle to meet the precision and safety requirements of aerospace applications because they rely on detailed knowledge of internal electrochemical mechanisms and require continuous calibration, which is difficult to maintain under operating conditions [7]. Zhu et al. showed that localized hotspots inside lithium batteries caused by fast charging and high temperatures promote lithium dendrite growth and trigger internal short circuits, increasing the risk of thermal runaway [8]. However, their work relied on micro-Raman spectroscopy, which is difficult to deploy for onboard battery monitoring in operational UAVs.

To overcome these limitations, recent research has shifted toward data-driven approaches that use machine learning (ML) algorithms to estimate the SOH directly from the measured operating data [9]. Algorithms such as random forest (RF) regression perform well in capturing complex degradation patterns without requiring explicit knowledge of the internal electrochemical reactions. When trained on established benchmark datasets, such as the NASA Prognostics Center of Excellence (PCoE) prognostic data, these models provide accurate SOH predictions suitable for the safety-critical requirements of electric aviation [10]. A review of ML for battery-life prediction emphasized that prediction accuracy strongly depends on the data quality and treatment of high-dimensional feature spaces [11]. Although this review provided a useful overview of dimensionality reduction and data processing, it did not address ML strategies for the real-time safety monitoring of UAVs operating at high temperatures. The effectiveness

Published under an exclusive license by KFUPM

Table 1. *Description of Dataset Features*

Variable	Description	Unit
Start time	Timestamp marking the beginning of each cycle	Date / Time
Time	Elapsed time within the cycle	Seconds (s)
Mode	Operating mode: charge (1), discharge (−1), or rest (0)	-
Voltage charger	Charger-side voltage	Volts (V)
Temperature battery	Battery surface temperature	°C
Voltage load	Load-side voltage during discharge	Volts (V)
Current load	Load current during discharge	Amperes (A)
Temperature MOSFET	Temperature of the MOSFET component	°C
Temperature resistor	Temperature of the resistor component	°C
Mission type	Usage condition	-

of an RF model for the online capacity estimation of lithium-ion batteries was demonstrated, but the evaluation was limited to controlled laboratory conditions and did not consider the high-temperature stress characteristics of UAV operation [12].

Building on this body of work, the present study develops and evaluates a RF regression model for the SOH estimation of Li-ion batteries used in UAVs, with an explicit focus on high-temperature operation. The model was trained and tested on the NASA PCoE Li-ion Battery Aging Dataset and benchmarked against a linear regression (LR) baseline to quantify the benefits of the nonlinear, ensemble-based approach. The estimated SOH was then used to project the remaining useful life (RUL) so that maintenance actions can be planned before the battery reaches its end of life (EOL). Although RF models have previously been used for battery-health monitoring, these studies have largely focused on electric vehicles and standard laboratory conditions. The present study has three main novel contributions: (i) the model is trained and evaluated on cycles in which the maximum cell temperature reaches up to 92.8 °C (mean 84.5 °C), as these conditions are representative of UAV operation in hot climates; (ii) a less conservative EOL threshold of 70% of nominal capacity (rather than the conventional 80%) was adopted to reflect aerospace safety margins, and the model's ability to track degradation up to this threshold was explicitly assessed; and (iii) the predicted SOH trajectory was translated into a quantitative RUL estimate, providing an end-to-end framework that links data-driven SOH prediction to actionable maintenance decisions for UAV battery management.

2. METHODOLOGY

2.1. Dataset Description.

The dataset used in this study was obtained from the NASA Ames PCoE Li-ion battery aging dataset, specifically battery #5 [13]. This dataset is a widely used benchmark for battery-health management and contains cycling data for 18650-sized lithium-ion cells with a nominal capacity of 2.0 Ah, from the beginning to the end of their life. The dataset includes comprehensive time-series observations of the battery collected over multiple charge–discharge cycles under carefully controlled experimental conditions.

The data were organized into cycles, each of which included three operating modes: charging (mode 1), discharging (mode −1), and resting (mode 0). Each row of the dataset represents a single measurement point within a cycle and records the electrical and thermal quantities, including time, voltage, current, temperature, and capacity-related parameters. These features are essential for characterizing the changes in battery behavior during cycling and tracking degradation trends over the battery lifetime. In addition to measuring capacity fading, temperature measurements are important for understanding accelerated aging and potential safety risks, such as overheating and thermal instability.

Table 1 lists each dataset variable along with its physical unit. The overall degradation trend is illustrated in Figure 1, which plots the battery capacity against the cycle number and shows the steady capacity fade that motivates the use of a data-driven prognostic approach.

2.2. Data Preprocessing.

2.2.1. Capacity estimation and SOH labeling. To analyze battery degradation, the dataset was first filtered by operating mode, and only the discharge cycles (mode −1) were retained. Because the discharge capacity was not directly recorded in the raw dataset, it was computed for each cycle using Coulomb counting. The per-cycle capacity and corresponding SOH were calculated using Eq. (1) and (2), respectively.

$$\text{Capacity} = \int I\,dt \tag{1}$$

$$\text{SOH} = \frac{C_n}{C_0} \times 100 \tag{2}$$

Here, C_n is the measured discharge capacity at cycle n (Ah) and C_0 is the initial (nominal) capacity at the beginning of life (2.0 Ah).

2.2.2. End of life thresholding. The EOL of the battery is defined as an SOH threshold of 70% of its nominal capacity. The data were then labeled into two regions: a safe zone (SOH $\geq$ 70%) and a failure zone (SOH <70%). The safe zone, consisting of 257 complete discharge cycles, was retained as the master dataset for model development. Table 2 presents the descriptive statistics of the processed features, confirming that they fell within the expected operational ranges.

The choice of 70% (rather than the more common 80%) as the EOL threshold was motivated by the UAV application.

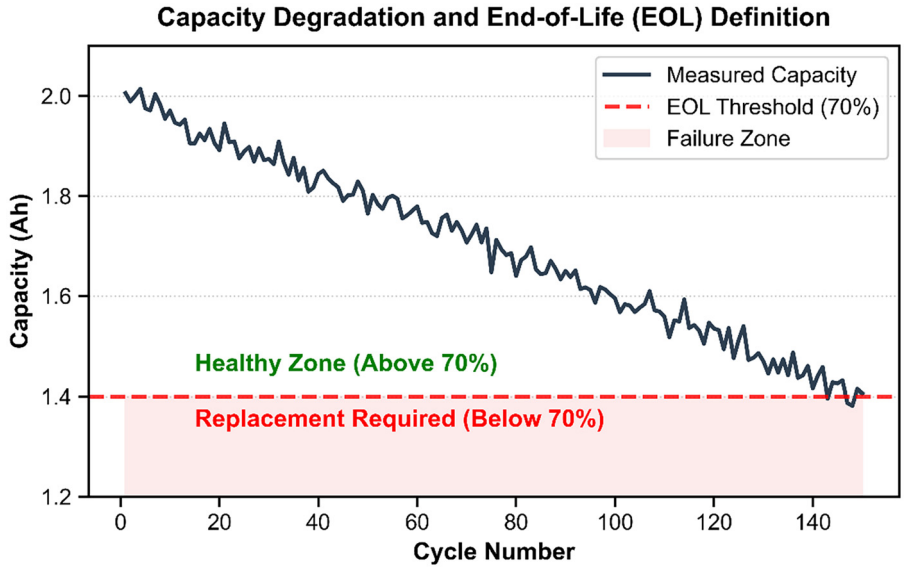

Figure 1. Lithium battery capacity degradation curve taken from the test dataset with the EOL threshold defined at 1.4 Ah (70% of the nominal capacity)

In aerospace applications, battery mass directly affects payload capacity and endurance; therefore, retiring a cell at an 80% SOH discards a significant portion of its useful life and unnecessarily increases operating costs. Because the present study focuses on high-temperature operation, the 70% threshold also exposes the RF model to more aggressive late-life degradation behavior, which is the regime in which accurate prediction is most critical for safety.

2.3. Exploratory Data Analysis.

2.3.1. Statistical distribution and thermal variability. The statistical distribution of the SOH was examined using a histogram. As shown in Figure 2a, the SOH values are broadly distributed but predominantly concentrated between 80% and 100%, indicating that the dataset captures multiple stages of battery degradation under realistic battery operating conditions. Figure 2b presents a boxplot of the maximum cycle temperature (T_{max}), indicating that most discharge cycles occurred under elevated temperature conditions, which is directly relevant to UAV operation in hot environments.

2.3.2. Feature correlation and degradation behavior. Figure 3 shows a correlation heatmap of the main variables (SOH, calculated capacity, T_{avg}, T_{max}, and V_{avg}). The SOH and calculated capacity show a perfect positive correlation (1.00), which is expected because the SOH is calculated by dividing the measured capacity by the nominal capacity. Therefore, including the calculated capacity as a model input introduces target leakage, and the model simply predicts a rescaled version of one of its inputs rather than learning the underlying degradation physics. Therefore, the calculated capacity was deliberately excluded from the model inputs, and only T_{max} and V_{avg} were retained as the predictors. V_{avg} is strongly negatively correlated withT_{max} (-0.91), indicating that a drop in the average voltage is accompanied by a substantial increase in T_{max}. The SOH shows a moderate positive correlation with V_{avg} (0.61) and a weaker direct correlation with temperature, suggesting that the voltage behavior is a particularly informative non-leaky feature for estimating the SOH and, by extension, the RUL in this dataset.

2.4. Model Development.

2.4.1. Algorithm selection: linear regression vs. random forest. LR provides a least-squares fitting of a hyperplane that best describes the relationship between the independent features (T_{max}, V_{avg}) and the dependent variable (SOH). The value of LR lies in its mathematical simplicity and interpretability, and it serves as a baseline that indicates whether battery degradation follows a consistent linear trend with respect to the thermal stress and voltage.

Because lithium-ion degradation is a strongly nonlinear process, an ensemble-based nonlinear model was used in this study.

Table 2. *Descriptive Statistics of the Refined Battery Dataset (n = 257)*

Feature	Mean	Std. Deviation	Minimum	Maximum	Unit
Average voltage (V_{avg})	6.39	0.22	6.13	7.42	V
Maximum temperature (T_{max})	84.53	11.80	28.67	92.80	°C
Average temperature (T_{avg})	58.78	7.23	25.32	63.44	°C
Average current (I_{avg})	-1.00	0.00	-1.00	-1.00	A
Calculated capacity	2.21	0.13	1.97	2.47	Ah
SOH	89.66	5.22	79.96	100.16	%

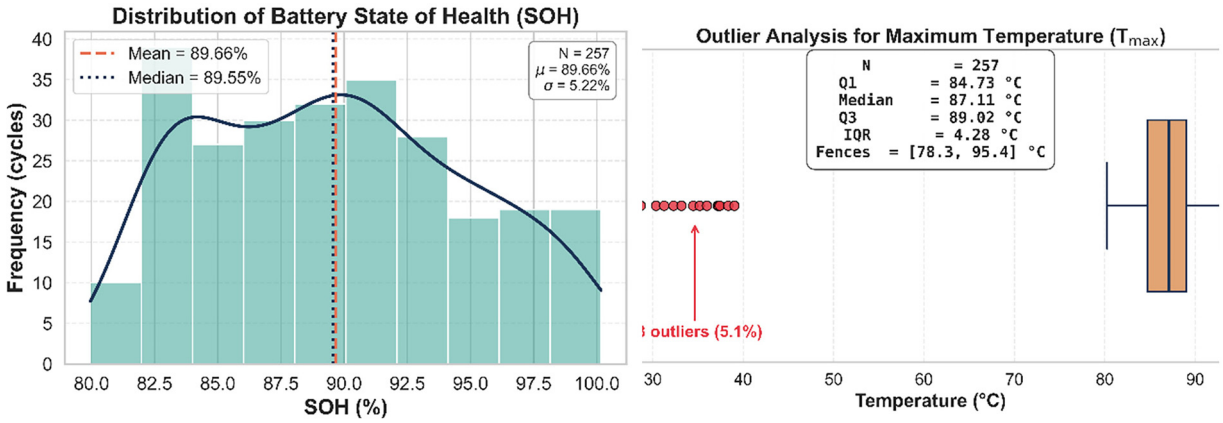

Figure 2. (a) SOH distribution. (b) Boxplot of T_{max} showing elevated temperatures

The RF algorithm was selected over other nonlinear alternatives, such as support vector regression or neural networks, for several reasons, particularly those relevant to small, undergraduate-scale studies. First, RF performs well with relatively small datasets (257 cycles after preprocessing), whereas neural networks typically require much larger training sets to prevent overfitting. Second, RF is robust to outliers and differences in feature scaling, which is advantageous for raw sensor data, such as cell temperature and voltage. However, RF is highly sensitive to its kernel and hyperparameter choices and benefits from careful feature normalization. Third, RF provides built-in feature importance scores that support the physical interpretation of the resulting model, which is an important requirement for safety-critical aerospace applications. RF is an ensemble learning method that constructs many independent decision trees from bootstrapped samples of training data and averages their predictions. This bagging mechanism allows the model to capture complex interactions between the high-temperature stress and voltage behavior, whereas the aggregation step reduces the variance and lowers the risk of overfitting to outliers, such as those commonly observed during exploratory data analysis.

2.4.2. Data splitting and training strategy. To ensure an unbiased evaluation of the two models on the 257 cycles, a holdout strategy was adopted, in which the cycles were split into two disjointed subsets:

1. The training set (80%, 205 cycles) was used to fit both the LR and RF models and learn the mapping from the input features (T_{max}, V_{avg}) to the target SOH.
2. The held-out test set (20%, 51 cycles) was used exclusively as an independent benchmark for evaluating the final predictive accuracy of each model.

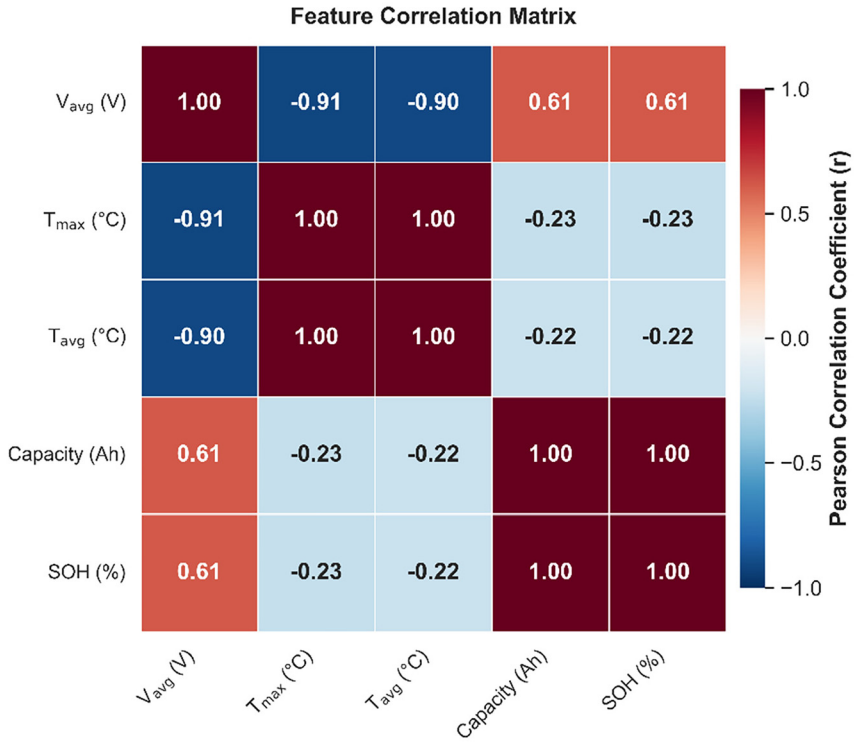

Figure 3. Feature correlation matrix

The Random Forest (RF) regressor's hyperparameters were specifically created based on this partitioning. In order to simultaneously predict the computed capacity and the State of Health (SOH), the ensemble architecture consists of 100 separate decision trees that are trained in parallel using a multi-output framework. To allow the model to properly represent the non-linear deterioration trajectory of the Li-ion cells, the maximum depth of the trees was left unconstrained. To reduce overfitting, the minimum number of samples needed to separate an internal node was set at two. A multiple Linear Regression (LR) model was trained under the same conditions and data-splitting approach to serve as a baseline for comparison and to assess performance differences in predictive accuracy in order to test the efficiency and generalization capabilities of the suggested model.

2.4.3. Feature engineering and input selection. Feature engineering is one of the most important steps in achieving accurate predictions, as it identifies the variables that most strongly influence battery degradation. Based on the statistical analysis presented above and the established knowledge of the electrochemistry of lithium-ion batteries[5], T_{max} and V_{avg} were selected as basic inputs for the models:

1. T_{max}: An elevated temperature is the dominant accelerator of Li-ion degradation. T_{max} was selected because it correlates with the growth of a solid electrolyte interphase (SEI) layer on the anode. This film thickens with increasing temperature and its continuous growth consumes active lithium ions and increases the internal resistance, resulting in a decrease in the SOH. Therefore, T_{max} is an accessible external proxy for these internal chemical changes.
2. V_{avg}: The V_{avg} during the discharge cycle was selected as a representative metric because it reflects the internal resistance of the cell. As the SEI layer thickens and the cell ages, the internal impedance increases, which manifests as a decrease in the average discharge voltage. Therefore, V_{avg} provides a complementary, non-thermal predictor of the SOH that is sensitive to the same underlying degradation mechanism as T_{max}.

2.4.4. Target variables (SOH and RUL). The predictive models in this study were designed to estimate two complementary quantities (SOH and RUL) that provide a comprehensive assessment of battery health.

1. SOH: is ratio of the present capacity to the nominal capacity (in %) and is a measure of the present condition of the battery relative to its as-manufactured state. Predicting the SOH enables the continuous monitoring of the degradation process.
2. RUL: is the number of remaining charge–discharge cycles before the battery reaches its EOL, defined in this study as SOH = 70% of the nominal capacity.

Together, these quantities provide a both present-state diagnostic (SOH) and a future estimate (RUL), which are both essential for ensuring UAV safety, scheduling maintenance, and enabling comprehensive battery health management solutions.

3. RESULTS

3.1. Visualization of Prediction Accuracy. Figure 4 presents a detailed scatter plot of the SOH predicted by the RF model versus the measured SOH. The predicted values are tightly clustered around the identity line ($y = x$), confirming the high accuracy reported for the metrics above. The close agreement across the full range of degradation levels indicates that the modeled results can be extrapolated to the cycles that were not observed during training.

Figure 5 compares the measured SOH degradation trajectory with that predicted by the RF model across all discharge cycles. The predicted curve closely tracks the true degradation path, confirming the ability of the model to capture both gradual capacity fading and a more rapid transition region near the EOL. This good agreement enables the accurate estimation of the cycle at which the battery reaches the EOL threshold of 1.4 Ah (70% of its nominal capacity).

3.2. Statistical Metrics. Table 3 summarizes the quantitative performances of the RF and LR models on the held-out

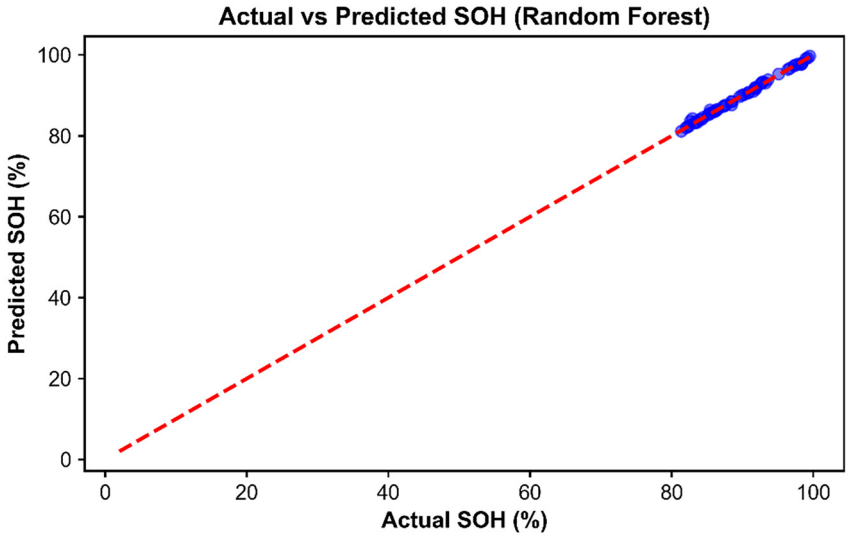

Figure 4. Scatter plot of the SOH predicted by RF versus the measured SOH

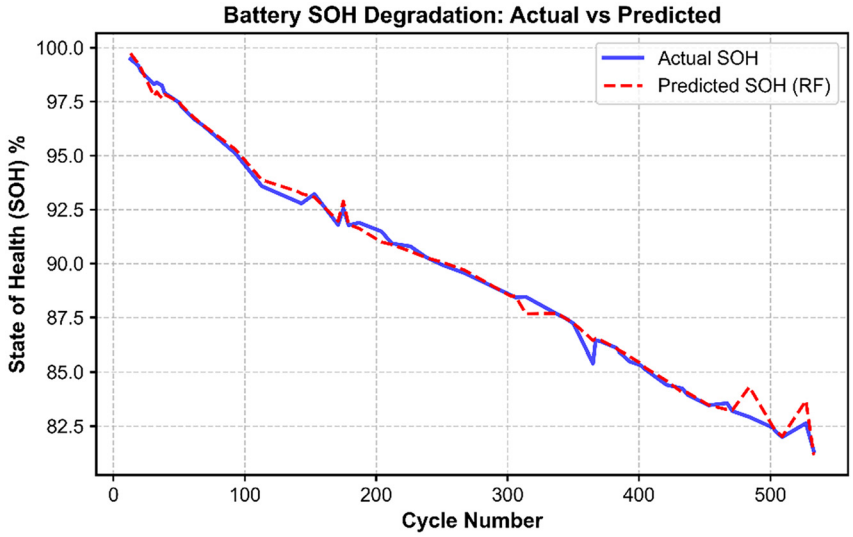

Figure 5. Measured and RF-predicted SOH curves as a function of cycle number

Table 3. *Error metrics and R^2 scores for the SOH prediction models on the held-out test set*

Metrics / Model	RF	LR
MAE	0.2185	2.4192
RMSE	0.2578	4.1303
R^2	0.9077	-22.6865

test dataset. The RF model achieved a high coefficient of determination (R^2) of 0.9077, demonstrating its ability to explain the observed variations in battery degradation. It also yielded the lowest error parameters (RMSE and MAE), which is consistent with the tight clustering shown in Figure 4. By contrast, the LR model exhibits a negative R^2 and substantially higher error metrics, confirming that a purely linear model does not effectively capture the nonlinear degradation dynamics of this dataset.

4. DISCUSSION

The RF model clearly outperformed the LR baseline in predicting the SOH and, by extension, RUL. This is attributed to the bagging mechanism underlying the RF, which trains each tree using a different bootstrap sample of data. Bagging reduces the variance and makes the model robust to small perturbations in the training set and noisy outliers[14]. In addition, because each individual tree partitions the feature space recursively, the ensemble captures the complex nonlinear interactions between the temperature and voltage that drive capacity fading. LR cannot represent these nonlinear interactions, resulting in predictions that deviate strongly from the measured SOH values. The large difference between the two models, particularly the negative R^2 obtained by LR, confirms that battery degradation under elevated thermal stress is strongly nonlinear and cannot be adequately modeled using a linear relationship.

The high R^2 score of the RF model (0.9077) indicates that it accounted for approximately 91% of the observed variation in the SOH. The corresponding low RMSE confirms that the RF predictions do not vary far from the true values at any point during cycling, while the low MAE indicates a precise cycle-by-cycle estimation. Both of these properties are essential in aerospace applications, where a single large prediction error indicates conditions that may result in an unsafe mission. The accurate prediction of the SOH degradation curve is a prerequisite for estimating the RUL. In the battery-aging literature, the EOL is typically defined as the point at which the usable capacity drops to 80% of the nominal capacity; however, the exact threshold varies with the cell chemistry, test conditions, and intended application[15]. In this study, a lower EOL threshold of 70% of the nominal capacity (1.4 Ah) was adopted to evaluate the ability of the model to track degradation beyond the standard failure limit. This is particularly relevant for aerospace safety margins, in which a lower threshold provides additional headroom for emergency landing scenarios or extended missions. Once the SOH trajectory is known, the RUL can be estimated by Eq. (3) to extrapolate the predicted curve until it reaches the EOL threshold.

$$\mathrm{RUL} = C_{\mathrm{EOL}} - C_n \tag{3}$$

where C_{EOL} is the predicted cycle number at which the SOH reaches the EOL threshold (1.4 Ah) and C_n is the present operational cycle.

As shown in Figure 5, the RF model accurately predicts the SOH across the cycle range. As a representative example, in cycle 52, the measured SOH is 97.3%, whereas that predicted by the RF model is 97.4%, corresponding to an absolute error of only 0.1 percentage points. The aggregated test-set metrics (MAE = 0.2185 and RMSE = 0.2578) further confirm good cycle-by-cycle stability across the entire prediction range, indicating a meaningful predicted RUL. In aerospace applications, translating an SOH prediction into an actionable RUL estimate is crucial, because it enables the conversion of raw numerical outputs into practical maintenance decisions and enables timely battery replacement before in-flight failure. In addition, because the predicted SOH trajectory does not vary significantly from the measured curve, the cycle number at which the trajectory crosses the 70% EOL threshold can be estimated with high confidence.

Despite the promising results, this study has limitations. For example, the model was trained and tested only on the battery

#5 data from the NASA PCoE dataset, which was used as a baseline case study. Therefore, the validation of additional cells from the same dataset (e.g., batteries #6, #7, and #18, which were tested under different operating profiles) is a necessary next step to confirm the generalizability of the proposed approach before it can be deployed in real UAV battery health management systems.

5. CONCLUSIONS

The study clearly demonstrated the utility of using a multi-output random forest regression model with a robust data-driven framework to predict simultaneously the computed capacity and health status of lithium-ion batteries. The key value of this work is not in considering battery degradation as a standalone criterion but in the possibility of mapping complex, non-linear electromagnetic and thermal degradation pathways based only on non-invasive, real-time measurable operational features (voltage, current and peak temperature). This framework provides a reliable and easy to deploy solution for flight safety monitoring that fully satisfies the stringent weight, computational and safety requirements of drone applications by providing robust structural constraints that prevent data leakage. The significant performance gap between the linear reference line model and the optimized cluster model further emphasizes the necessity of nonlinear machine learning architectures to capture dynamic capacity loss.

Despite the great predictive accuracy and the strong generalization capabilities reached, it is important to point out some limits of this work. The model was trained and validated using constant current charge-discharge cycle profiles from well-established laboratory databases, which may not capture the transient aerodynamic loads and changing weather conditions that drones experience during real-world flight maneuvers such as hovering, vertical climbs, or sudden wind gusts. Moreover, the ability to instantaneously transfer the model to other chemical compositions (e.g., lithium-sulfur cells or solid-state batteries) without retraining is limited by its dependence on statistical features learned from a single battery chemical composition.

Overall, the study findings demonstrated that data-driven battery health management can meaningfully contribute to the safety and reliability of electric aviation by anticipating failures before they occur. Based on these results, the following directions are recommended for future research.

- Explore sequence-based deep learning architectures such as long short-term memory networks to predict the RUL directly from time-series data, rather than via an analytical projection of the SOH curve.
- Validate the proposed model on additional cells under harsh thermal and operational conditions to better simulate real-world UAV flight environments in hot-climate regions.
- Extend the comparative analysis by benchmarking the RF model against additional nonlinear methods, such as support vector regression, gradient boosting, or long short-term memory networks. This will provide a broader and more rigorous benchmark for evaluating the model performance across different levels of model complexity under various thermal conditions.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Jana Eid Aljohani – *Department of Aerospace Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*;
Email: s202258620@kfupm.edu.sa

Corresponding Author

Mohammad Irfan Alam – *Advisor, Assistant Professor, Department of Aerospace Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*;
0000-0003-0296-3364
Email: mohammad.alam.1@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the Department of Aerospace Engineering at the King Fahd University of Petroleum and Minerals (KFUPM).

REFERENCES

(1) Agarwal, R. K. Sustainable (green) aviation: challenges and opportunities. *SAE Int. J. Aerosp.* **2**, 1–20 (2009). https://doi.org/10.4271/2009-01-3085

(2) Cheng, X. *et al.* A perspective on sustainable energy materials for lithium batteries. *SusMat* **1**, 38–50 (2021). https://doi.org/10.1002/sus2.4

(3) Matlock, J., Warwick, S., Sharikov, P., Richards, J. & Suleman, A. Evaluation of energy efficient propulsion technologies for unmanned aerial vehicles. *Trans. Can. Soc. Mech. Eng.* **43**, 481–489 (2019). https://doi.org/10.1139/tcsme-2018-0231

(4) Barrera, T. P. *et al.* Next-generation aviation Li-ion battery technologies—enabling electrified aircraft. *Electrochem. Soc. Interface* **31**, 69–74 (2022). https://doi.org/10.1149/2.f10223if

(5) Menye, J. S., Camara, M. & Dakyo, B. Lithium battery degradation and failure mechanisms: a state-of-the-art review. *Energies* **18**, 342 (2025). https://doi.org/10.3390/en18020342

(6) Lin, C., Tang, A. & Wang, W. A review of SOH estimation methods in lithium-ion batteries for electric vehicle applications. *Energy Procedia* **75**, 1920–1925 (2015). https://doi.org/10.1016/j.egypro.2015.07.199

(7) Lv, X., Li, Q. & Wang, K. State monitoring of lithium-ion batteries based on in situ magnetic techniques: a review. *Ionics* **31**, 7595–7613 (2025). https://doi.org/10.1007/s11581-025-06489-6

(8) Zhu, Y. *et al.* Fast lithium growth and short circuit induced by localized-temperature hotspots in lithium batteries. *Nat. Commun.* **10**, 2067 (2019). https://doi.org/10.1038/s41467-019-09924-1

(9) Madani, S. S., Ziebert, C., Vahdatkhah, P. & Sadrnezhaad, S. K. Recent progress of deep learning methods for health monitoring of lithium-ion batteries. *Batteries* **10**, 204 (2024). https://doi.org/10.3390/batteries10060204

(10) Kim, E. *et al.* Data-driven methods for predicting the state of health, state of charge, and remaining useful life of Li-ion batteries: a comprehensive review. *Int. J. Precis. Eng. Manuf.* **24**, 1281–1304 (2023). https://doi.org/10.1007/s12541-023-00832-5

(11) Xue, P. *et al.* Solutions for lithium battery materials data issues in machine learning: overview and future outlook. *Adv. Sci.* **11**, e2410065 (2024). https://doi.org/10.1002/advs.202410065

(12) Li, Y. *et al.* Random forest regression for online capacity estimation of lithium-ion batteries. *Appl. Energy* **232**, 197–210 (2018). https://doi.org/10.1016/j.apenergy.2018.09.182

(13) Saha, B. & Goebel, K. Battery data set. *NASA Prognostics Data Repository* https://data.nasa.gov/dataset/li-ion-battery-aging-datasets (2007).

(14) Wang, G., Lyu, Z. & Li, X. An optimized random forest regression model for Li-ion battery prognostics and health management. *Batteries* **9**, 332 (2023). https://doi.org/10.3390/batteries9060332

(15) Xing, Y., Ma, E. W. M., Tsui, K. L. & Pecht, M. Battery management systems in electric and hybrid vehicles. *Energies* **4**, 1840–1857 (2011). https://doi.org/10.3390/en4111840