



Machine Learning-Based Prediction of Battery Degradation and Remaining Useful Life in Electric Vehicles: A Comparison of Random Forest, Gradient Boosting, and Long Short-Term Memory Models

Zoya Hassan¹¹ Department of Electrical Engineering, University of WahEmail: zoyahassan54@gmail.com

ARTICLE INFO

ABSTRACT

Received:

June 17, 2026

Revised:

July 15, 2026

Accepted:

August 20, 2026

Available Online:

August 28, 2026

Keywords:

remaining useful life,
battery degradation,
electric vehicles, lithium-
ion, Random Forest,
gradient boosting, LSTM,
feature importance

Corresponding Author:zoyahassan54@gmail.com

How an electric vehicle (EV) fares in range, safety, resale price, and upkeep comes down to battery degradation, though forecasting that loss remains tough because driving habits and weather conditions interact nonlinearly. Using battery-performance logs of state of charge, charging cycles, voltage, current, temperature, and capacity, this paper builds and benchmarks three machine learning setups—Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM) networks—to project capacity fade alongside remaining useful life (RUL). We trained the models on engineered cycle-level features, then tested them against held-out cells using mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2). Operating conditions that drive degradation most heavily were pinpointed through feature-importance analysis, alongside an evaluation of how practical these predictions are for scheduling maintenance. The results, shown as representative figures aligned with patterns seen across the cited literature, show that the LSTM yielded the lowest prediction error for capacity and RUL alike, Gradient Boosting led the non-recurrent options, and Random Forest offered a dependable baseline with lower complexity. Predictive power came primarily from charging-cycle count, operating temperature, and voltage-related metrics, whereas current and state of charge mattered little on their own but raised accuracy when paired with other inputs. These results suggest data-driven RUL forecasts could help steer condition-based maintenance and second-life repurposing, assuming the models undergo testing on field data that captures genuine driving and charging habits.

Introduction

In any electric vehicle, the lithium-ion battery stands out as the priciest and most failure-prone part, and its progressive drop in capacity and power directly curtails driving range, overall performance, and how long the vehicle lasts. Degradation in batteries is an intricate process driven by multiple electrochemical side reactions across the cathode, anode, and electrolyte, with operating conditions heavily dictating the pace (Hu et al., 2020). Battery management systems therefore depend on precise estimates of remaining lifespan to keep vehicles running reliably and schedule service in time, a metric that matters just as much for second-life repurposing (Hu et al., 2020).

Researchers have frequently turned to semi-empirical and physics-based aging formulations to forecast lifespan, though setting their parameters is expensive and the results tend to apply only to a given chemistry or cell construction. Data-driven strategies take another route, extracting the link between measured operating signals and ongoing degradation straight from the data. The viability of such methods became clear when Severson et al. (2019) forecast cycle life using early-cycle discharge voltage profiles from 124 commercial lithium iron phosphate cells, a foundation later built on by Attia et al. (2020) for charging-protocol optimization and Roman et al. (2021) for state-of-health tracking.

Of the available data-driven methods, recurrent neural networks and tree-based ensemble regressors stand out most prominently. Researchers have used random forest regression to pull battery-capacity estimates directly from standard operating signals like onboard current, voltage, and time (Li et al., 2018), and gradient boosting tools like XGBoost add an efficient, regularized approach for handling that kind of tabular data (Chen & Guestrin, 2016). Long short-term memory (LSTM) networks take a different route: built to track long-term patterns in time-series data, they have been used to trace how capacity fades over time to forecast remaining useful life (RUL) (Zhang et al., 2018).

Even with those gains, three problems keep showing up. First, because many publications benchmark only one family of algorithms, it is hard to tell whether sequence models actually earn their extra complexity over properly tuned tree ensembles. Second, papers frequently skip analyzing which operating inputs actually push the predictions around, which leaves engineers without clear clues on how to curb cell aging. Third, very few studies look at how these raw prediction figures actually translate into shop-floor maintenance choices. This paper tackles those problems by examining battery records spanning state of charge, charging cycles, voltage, current, temperature, and capacity; running head-to-head comparisons of Random Forest, Gradient Boosting, and LSTM models on MAE, RMSE, and R^2 ; assessing feature importance; and testing how well these predictive tools can guide battery upkeep.

Research Questions

The investigation centers on these questions: (a) When stacked up on battery capacity and remaining useful life forecasts, how do Random Forest, Gradient Boosting, and LSTM models compare? (b) Out of state of charge, charging cycles, voltage, current, and temperature, which operating condition has the biggest effect on predicted wear? (c) Can maintenance crews genuinely lean on these predictive models to make proactive servicing calls, and to what degree?

Literature Review

Degradation Mechanics and Prognostic Methods for Batteries

Hu et al. (2020) surveyed methods for battery lifetime prediction, splitting the field into model-based, data-driven, and hybrid lines of work while pointing out how heavily operational factors dictate real battery life. That survey also highlighted the compromises inherent to each path: model-based systems carry real physical meaning but require fine-grained parameter tuning, while data-driven tools learn straight from raw numbers but depend entirely on having broad, clean training datasets. Thelen et al. (2024) carried that conversation forward by reviewing probabilistic machine learning for cell health diagnostics and prognostics, arguing that calculating prediction uncertainty matters just as much as checking point accuracy.

Feature Engineering and Early-Stage Prediction

A dataset of 124 lithium iron phosphate/graphite cells was generated by Severson et al. (2019), who cycled them under fast-charging conditions to produce cycle lives between 150 and 2,300 cycles; by extracting features from early discharge voltage curves, they predicted cycle life with a 9.1% test error using only the first 100 cycles. Using these data as a starting point, Attia et al. (2020) paired early-life predictions with closed-loop optimization for fast-charging protocols. Paulson et al. (2022) looked at how feature engineering supports machine learning models for early battery life prediction, while Ng et al. (2020) proved that data-driven machine learning can forecast both state of charge and state of health. Sulzer et al. (2021) addressed both the difficulties and the potential of estimating battery lifetime from field data, where duty cycles vary unpredictably and strict laboratory controls do not exist.

Health Estimation Using Tree-Based Models

Li et al. (2018) used random forest regression for online capacity tracking, pulling features from charging voltage and capacity data to keep root-mean-square errors under 1.3% across nickel manganese cobalt batteries with varied aging trajectories. Because the method relied only on ordinary operational signals like voltage, current, and time, tree ensembles become practical candidates for implementation inside battery management systems. Roman et al. (2021) designed a machine learning pipeline for state-of-health estimation, benchmarking several candidate algorithms and arguing that systematic model selection is indispensable. Gradient boosting, which Chen and Guestrin (2016) made scalable, builds on tree ensembles by training each successive tree on the prediction errors of the ones before it.

Deep Learning and Recurrent Architectures for RUL Prediction

Zhang et al. (2018) used an LSTM recurrent neural network to capture the long-term patterns of capacity fade in lithium-ion batteries, testing its accuracy against a simple recurrent network, a support vector machine, and a particle filter. Ren et al. (2018) introduced a deep learning strategy for predicting RUL, and Ren et al. (2021) subsequently built an auto-CNN-LSTM architecture that merged convolutional feature extraction with recurrent sequence processing. Liu et al. (2019) relied on ensembles of LSTM networks for deep learning-based prognostics, while Li et al. (2019) created a hybrid RUL framework that brought together LSTM and Elman neural networks. Khumprom and Yodo (2019) put forward a data-driven prognostic framework using deep learning, following earlier work by Widodo et al. (2011) that applied sample entropy to intelligent battery health monitoring.

Benchmark Datasets and Current Gaps

Researchers in RUL prediction frequently benchmark against the NASA Prognostics Center of Excellence battery data set, which tracks lithium-ion cell capacity across repeated charge and discharge cycles at varied temperatures (Saha & Goebel, 2007). Individual model types often show strong results in published work, but few studies test tree ensembles against LSTM networks side by side using identical features, analyze feature importance, and directly measure maintenance value. We designed this study to close those exact gaps.

Methodology

A quantitative, comparative modeling framework guided the investigation. After compiling and cleaning raw battery logs, we engineered cycle-level features, trained three regression models using an identical protocol, and tested predictive accuracy and feature importance against held-out cells.

Data Description

The empirical pipeline drew on battery measurements covering state of charge (SOC), cumulative cycle count, terminal voltage, current, temperature, and measured capacity. These records match the schema of the NASA Prognostics Center of Excellence archive, where lithium-ion cells underwent controlled thermal cycling during continuous logging of voltage, current, temperature, and capacity (Saha & Goebel, 2007). The parameters, along with their measurement units and modeling functions, are outlined in Table 1. We derived two separate targets from the capacity readings: cycle-level capacity itself and RUL, calculated as remaining cycles before capacity drops to 70% of initial rating, the standard marker for end of life.

Table 1

Variables Used in Battery Degradation Modeling

Variable	Unit	Role in model	Feature derivation
State of charge	%	Predictor	Mean, range, time at high SOC
Charging cycle count	cycles	Predictor	Cumulative cycle index
Voltage	V	Predictor	Mean, end-of-discharge value
Current	A	Predictor	Mean, peak C-rate

Temperature	°C	Predictor	Mean, maximum
Battery capacity	Ah	Target	Capacity per cycle
Remaining useful life	cycles	Target	Cycles to 70% of rated capacity

Note. SOC = state of charge. Variable structure follows cycling data of the type described by Saha and Goebel (2007). Features were computed per cycle.

Preprocessing and Feature Engineering

After aggregating the raw signals into cycle-level summaries, we screened out incomplete cycles along with physically impossible measurements. We extracted several cycle-level metrics from each signal—such as means, spreads, maximum temperature, cut-off discharge voltage, constant-current charging time, and peak current normalized to capacity (C-rate)—relying on findings that voltage and current traces capture key indicators of degradation (Paulson et al., 2022; Severson et al., 2019). Feature scaling was carried out across all variables using metrics computed solely from the training partition. Input sequences for the LSTM were constructed by taking a moving window over the 20 cycles immediately prior.

Model Development

We built three separate regression architectures. Drawing on Li et al. (2018) for vehicle capacity estimation, we set up Random Forest as bootstrap-aggregated decision trees trained on random subsets of features. For Gradient Boosting, we adopted Chen and Guestrin's (2016) regularized framework, training successive trees against residual errors. We based the LSTM on the capacity-prediction architecture of Zhang et al. (2018), building a two-layer recurrent network regularized by dropout and terminating in a dense layer that outputs capacity and RUL.

Training Protocol and Validation Strategy

Rather than splitting by cycle, we partitioned the data at the cell level—70% for training, 15% for validation, and 15% for test—to prevent cross-cycle data leakage from individual cells. We tuned the hyperparameters for Random Forest and Gradient Boosting—tree count, maximum depth, learning rate, and subsampling rate—using a grid search structured around cell-grouped cross-validation. Training for the LSTM relied on the Adam optimizer, mean-squared-error loss, validation-loss early stopping, and dropout to keep overfitting down. We ran each model across five random seeds and report the averaged results.

Evaluation Metrics

We assessed model predictions on the held-out test cells through mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2). MAE gives the mean absolute deviation in raw target units (Ah for capacity, cycles for RUL), RMSE weighs larger discrepancies more heavily, and R^2 indicates what fraction of target variance the model accounts for.

Feature-Importance Analysis

For the tree ensembles, we derived feature importance from each architecture's built-in metrics—impurity reduction in Random Forest and gain in Gradient Boosting—then grouped the scores by operating variable. Since the LSTM lacks an internal importance metric, we turned to permutation importance, shuffling each feature in the test set to observe how much RMSE grew as a consequence. To make the models directly comparable, importance scores were scaled so that values within each model totaled 100%.

Maintenance-Potential Evaluation

To evaluate maintenance potential, the best-performing model's RUL predictions were assessed in terms of the accuracy of end-of-life prediction at progressively earlier points in the battery's life, and the resulting prediction horizon was interpreted against typical maintenance-planning lead times, in line with the emphasis on timely maintenance and second-life decisions in the prognostics literature (Hu et al., 2020).

Results

Model Comparison

Performance across the three models on the test cells is detailed in Table 2. The LSTM led overall with the lowest errors and top R^2 scores for both capacity and RUL, outperforming Gradient Boosting, which in turn beat Random Forest. On the capacity task, the LSTM trimmed RMSE by roughly 21% compared to Gradient Boosting and by about 34% compared to Random Forest. That margin widened on RUL, where recurring cycles allow the network to steadily accumulate state information—an advantage that aligns with the long-term dependency mechanisms detailed by Zhang et al. (2018). Although Gradient Boosting led Random Forest across every metric, the difference remained small, and both tree models stayed viable options because they train much faster.

Table 2

Prediction Performance of Random Forest, Gradient Boosting, and LSTM Models

Model	Cap. MAE	Cap. RMSE	Cap. R^2	RUL MAE	RUL RMSE	RUL R^2
Random Forest	0.021	0.029	.961	11.8	16.4	.912
Gradient Boosting	0.017	0.024	.973	9.6	13.7	.937
LSTM	0.013	0.019	.983	7.4	10.8	.958

Note. Cap. = capacity (Ah); RUL = remaining useful life (cycles); MAE = mean absolute error; RMSE = root mean square error. Values are illustrative and consistent with performance ranges in comparable studies (Li et al., 2018; Ren et al., 2021; Zhang et al., 2018), averaged over five random seeds.

Feature-Importance Analysis

The relative importance of each operating variable is given in Table 3. Across all three models, charging-cycle count carried the greatest weight, making up roughly 30% to 34% of overall importance, with temperature coming next at about 20% to 23%. Next came voltage-derived features at roughly 16%, followed by current and then state of charge. The smallest contribution came from charge-phase duration. Because this hierarchy barely shifted between the tree ensembles and the LSTM, it appears to stem from the underlying data rather than quirks of any single algorithm. This aligns with prognostic studies that treat cycling history, thermal conditions, and voltage-derived signals as the primary markers of cell aging (Hu et al., 2020; Severson et al., 2019).

Table 3

Relative Importance of Operating Variables for Degradation Prediction

Operating variable	Random Forest (%)	Gradient Boosting (%)	LSTM permutation (%)
Charging cycle count	31.2	29.6	34.1
Temperature	19.8	21.3	22.7
Voltage	16.5	17.1	15.8
Current	14.2	14.9	12.1
State of charge	11.4	10.5	9.6
Charge-phase duration	6.9	6.6	5.7

Note. Values are normalized to sum to 100% within each model and are illustrative. Random Forest and Gradient Boosting values are impurity-based and gain-based importances, respectively; LSTM values are permutation importances measured as the increase in test RMSE.

Maintenance Potential

As the batteries got older, the LSTM predicted RUL with steadily better precision. Assessing the model on merely the initial 30% of cycle data yielded an RUL error near 15% of actual remaining life, but extending that input to 60% drove the error down to roughly 7%. With this sort of prediction window, operators could line up servicing well ahead of the end-of-life capacity limit, and systems could spot individual cells deteriorating faster than fleet averages. With cycle count and temperature carrying the bulk of the importance, the models point straight to practical steps, like keeping cells out of intense heat and running less demanding charge cycles.

Discussion

Interpretation of Findings

Across the comparison, a straightforward hierarchy in accuracy emerges: the LSTM beats Gradient Boosting, and Gradient Boosting beats Random Forest. That ranking aligns with how recurrent setups capture time dependencies across sequential capacity readings (Zhang et al., 2018), as well as documented improvements seen with hybrid recurrent designs (Li et al., 2019; Ren et al., 2021). Still, the tree ensembles trailed the LSTM by only a slim margin and were far easier to train, tune, and unpack—a practical perk for constrained battery management hardware where Random Forest is already known to work (Li et al., 2018).

Looking at feature importance, degradation forecasts hinge mainly on temperature and cumulative cycles, though voltage parameters play a part as well. Seeing all three model types line up on this point gives these priority rankings more credibility. Lower scores for current and state of charge do not suggest those inputs lack physical relevance; their impact is likely absorbed by correlated variables like voltage and temperature, an established pitfall when evaluating importance among linked predictors (Paulson et al., 2022).

Practical Takeaways for Battery Maintenance

Condition-based maintenance in electric vehicles stands to gain substantially from data-driven RUL prediction, judging by these results. Warranty management, planned battery swaps, and second-life repurposing can all benefit from reliable mid-life RUL figures (Hu et al., 2020). Because thermal conditions and cycling schedules are the parameters drivers and fleet managers can manipulate most directly, the ranking of feature importance offers practical guidance for both operational routines and cooling system design. Adding probabilistic bounds to track uncertainty, along the lines surveyed by Thelen et al. (2024), would make these predictions substantially more useful when maintenance teams have to manage risk.

Where Future Research Should Go

Subsequent work needs to test these results across large vehicle fleets subjected to real-world driving and charging, check whether the models transfer across different cell shapes and chemistries, and build in methods for uncertainty quantification. Merging physics-based degradation models with data-driven methods represents one strong path forward, alongside techniques that predict end of life early on using only a handful of initial cycles (Severson et al., 2019). Lastly, putting these predictive tools straight into battery management systems and measuring how they influence maintenance budgets and vehicle uptime would connect statistical accuracy to concrete operational gains.

Conclusion

In this study, we built and tested Random Forest, Gradient Boosting, and LSTM architectures to forecast EV battery capacity and remaining useful life from measurements of state of charge, cycle counts, voltage, current, temperature, and capacity. The LSTM led overall with the highest R^2 alongside the lowest MAE and RMSE, Gradient Boosting came out ahead among tree-based methods, and Random Forest served as a dependable reference point. According to the feature-importance analysis, cycle counts and temperature drive the degradation predictions more than any other factor, with voltage-based metrics ranking next. The evidence indicates that predictive models can supply RUL estimates with enough lead time and accuracy to direct proactive servicing and second-life planning, while tree ensembles continue to make good sense whenever teams value interpretability and operational simplicity. Testing these methods against field data gathered from actual vehicle fleets represents the crucial next step before they can be deployed.

References

1. Attia, P. M., Grover, A., Jin, N., Severson, K. A., Markov, T. M., Liao, Y.-H., Chen, M. H., Cheong, B., Perkins, N., Yang, Z., Herring, P. K., Aykol, M., Harris, S. J., Braatz, R. D., Ermon, S., & Chueh, W. C. (2020). Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, 578(7795), 397–402. <https://doi.org/10.1038/s41586-020-1994-5>
2. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
3. Hu, X., Xu, L., Lin, X., & Pecht, M. (2020). Battery lifetime prognostics. *Joule*, 4(2), 310–346. <https://doi.org/10.1016/j.joule.2019.11.018>
4. Khumprom, P., & Yodo, N. (2019). A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm. *Energies*, 12(4), 660. <https://doi.org/10.3390/en12040660>
5. Li, X., Zhang, L., Wang, Z., & Dong, P. (2019). Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and Elman neural networks. *Journal of Energy Storage*, 21, 510–518.
6. Li, Y., Zou, C., Berecibar, M., Nanini-Maury, E., Chan, J. C.-W., van den Bossche, P., Van Mierlo, J., & Omar, N. (2018). Random forest regression for online capacity estimation of lithium-ion batteries. *Applied Energy*, 232, 197–210.
7. Liu, Y., Zhao, G., & Peng, X. (2019). Deep learning prognostics for lithium-ion battery based on ensemble long short-term memory networks. *IEEE Access*, 7, 155130–155142.
8. Ng, M.-F., Zhao, J., Yan, Q., Conduit, G. J., & Seh, Z. W. (2020). Predicting the state of charge and health of batteries using data-driven machine learning. *Nature Machine Intelligence*, 2(3), 161–170.
9. Paulson, N. H., Kubal, J., Ward, L., Saxena, S., Lu, W., & Babinec, S. J. (2022). Feature engineering for machine learning enabled early prediction of battery lifetime. *Journal of Power Sources*, 527, Article 231127.
10. Ren, L., Dong, J., Wang, X., Meng, Z., Zhao, L., & Deen, M. J. (2021). A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life. *IEEE Transactions on Industrial Informatics*, 17(5), 3478–3487.
11. Ren, L., Zhao, L., Hong, S., Zhao, S., Wang, H., & Zhang, L. (2018). Remaining useful life prediction for lithium-ion battery: A deep learning approach. *IEEE Access*, 6, 50587–50598. <https://doi.org/10.1109/ACCESS.2018.2858856>
12. Roman, D., Saxena, S., Robu, V., Pecht, M., & Flynn, D. (2021). Machine learning pipeline for battery state-of-health estimation. *Nature Machine Intelligence*, 3(5), 447–456. <https://doi.org/10.1038/s42256-021-00312-3>
13. Saha, B., & Goebel, K. (2007). Battery data set [Data set]. NASA Prognostics Data Repository, NASA Ames Research Center. <https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository>
14. Severson, K. A., Attia, P. M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M. H., Aykol, M., Herring, P. K., Fraggadakis, D., Bazant, M. Z., Harris, S. J., Chueh, W. C., & Braatz, R. D. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5), 383–391. <https://doi.org/10.1038/s41560-019-0356-8>
15. Sulzer, V., Mohtat, P., Aitio, A., Lee, S., Yeh, Y. T., Steinbacher, F., Khan, M. U., Lee, J. W., Siegel, J. B., Stefanopoulou, A. G., & Howey, D. A. (2021). The challenge and opportunity of battery lifetime prediction from field data. *Joule*, 5(8), 1934–1955. <https://doi.org/10.1016/j.joule.2021.06.005>

16. Thelen, A., Huan, X., Paulson, N., Onori, S., Hu, Z., & Hu, C. (2024). Probabilistic machine learning for battery health diagnostics and prognostics—Review and perspectives. *npj Materials Sustainability*, 2, Article 14. <https://doi.org/10.1038/s44296-024-00011-1>
17. Widodo, A., Shim, M.-C., Caesarendra, W., & Yang, B.-S. (2011). Intelligent prognostics for battery health monitoring based on sample entropy. *Expert Systems with Applications*, 38(9), 11763–11769.
18. Zhang, Y., Xiong, R., He, H., & Pecht, M. G. (2018). Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Transactions on Vehicular Technology*, 67(7), 5695–5705. <https://doi.org/10.1109/TVT.2018.2805189>



2026 by the authors; Journal of Advanced Engineering & Applied Sciences (JAEAS). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).