Engineering Frontiers
Vol. 1, No. 5, 2025 1
Research Article

# A Comparative Study of Machine Learning Methods for Lithium-ion Battery Remaining Useful Life Prediction

Shuhe Cui[a*], Yiqing Du[b], Yongpeng Chen[c], Zhiyuan Yang[a]

[*] *Corresponding author. e-mail: E1718781@u.nus.edu*

[a] *College of Design and Engineering, National university of Singapore, 119077 Singapore*

[b] *School of Business , National university of Singapore, 119077 Singapore*

[c] *School of Computing , National university of Singapore, 119077 Singapore*

**Abstract**

Accurate prediction of remaining useful life for lithium-ion batteries is essential for ensuring the reliability and safety of electric vehicles and energy storage systems. This study presents a comprehensive comparative analysis of six machine learning methods for battery RUL prediction, including Linear Regression, Ridge Regression, Lasso Regression, Support Vector Regression, K-Nearest Neighbors, and Random Forest. Using the NASA battery degradation dataset, we develop a feature engineering framework that extracts six capacity-derived features while carefully avoiding information leakage by excluding cycle number from the feature set. A Leave-One-Battery-Out cross-validation strategy is employed to provide realistic generalization estimates for time-series degradation data. Experimental results demonstrate that Lasso Regression achieves the best predictive performance with a root mean square error of 10.827 cycles and a coefficient of determination of 0.8183 on the test battery. The findings reveal that linear regularized methods outperform more complex nonlinear approaches for this prediction task, suggesting that the relationship between capacity-derived features and RUL is predominantly linear. This work provides practical guidance for selecting appropriate machine learning algorithms for battery health management applications.

**Keywords: Lithium-ion batteries, remaining useful life prediction, machine learning, capacity degradation, battery health management**

## 1. Introduction

Lithium-ion batteries have become the dominant energy storage technology for electric vehicles, portable electronics, and grid-scale energy storage systems due to their high energy density, low self-discharge rate, and long cycle life(Chen et al., 2020; Xu et al., 2023). However, battery performance inevitably degrades over time through complex electrochemical processes, leading to capacity fade and eventual failure(Palacín & de Guibert, 2016). Accurate prediction of remaining useful life, defined as the number of charge-discharge cycles until a battery reaches its end-of-life threshold, is therefore critical for ensuring system reliability, optimizing maintenance schedules, and preventing unexpected failures that could have serious safety implications. From an environmental perspective, improving RUL prediction can also support decarbonization efforts by extending battery service life, reducing premature replacements, and lowering lifecycle $CO_2$ emissions associated with battery manufacturing and disposal(Shao et al., 2025; Wu et al., 2023).

The degradation of lithium-ion batteries is influenced by numerous factors including operating temperature, charge and discharge rates, depth of discharge, and calendar aging.(Guo et al., 2021; Shi et al., 2025) These interacting mechanisms create complex nonlinear degradation trajectories that present significant challenges for RUL prediction(Liu et al., 2023; Zhou et al., 2023). Traditional


https://doi.org/10.71411/ef.2025.v1i5.1380

physics-based models, while providing interpretable degradation mechanisms, often require detailed knowledge of battery chemistry and operating conditions that may not be available in practical applications(Ji et al., 2024). Data-driven approaches based on machine learning have emerged as promising alternatives that can learn degradation patterns directly from historical cycling data without requiring explicit physical models(Ruiz et al., 2025).

Recent advances in machine learning have spurred considerable interest in applying various algorithms to battery prognostics(Thelen et al., 2024). Linear regression methods offer simplicity and interpretability but may fail to capture nonlinear degradation behaviors(Alfeo et al., 2022). Regularized variants such as Ridge and Lasso regression address overfitting concerns through penalty terms that constrain coefficient magnitudes(Gillariose et al., 2026). Support vector regression extends the powerful support vector machine framework to regression problems using kernel functions that implicitly map features to high-dimensional spaces(Du et al., 2024). Instance-based methods like K-nearest neighbors make predictions based on similarity to training examples, while ensemble methods such as random forests combine multiple decision trees to improve robustness and accuracy(Zaferani et al., 2025).

Despite the growing body of literature on machine learning for battery RUL prediction, several methodological concerns warrant attention(Shao & Zhang, 2025). A particularly important issue is information leakage, where features that would not be available in a real-time prediction scenario are inadvertently included in the model(Wen et al., 2024). The most common form of this leakage occurs when cycle number is used as a predictor, since in a retrospective analysis where end-of-life is already known, RUL is simply the difference between the known end-of-life cycle and the current cycle(Chen et al., 2022). Additionally, standard random cross-validation splits can introduce temporal leakage in time-series data, as training examples from later cycles may inform predictions for earlier cycles from the same battery(Athanasakis et al., 2024).

This study addresses these methodological challenges while providing a systematic comparison of machine learning methods for battery RUL prediction. We develop a feature engineering framework based exclusively on capacity measurements, ensuring that all predictors represent information that would be available during online monitoring. We employ Leave-One-Battery-Out cross-validation to obtain realistic estimates of generalization performance to new batteries. Using the widely-studied NASA battery degradation dataset, we evaluate six representative machine learning algorithms spanning linear, regularized linear, kernel-based, instance-based, and ensemble approaches. Our analysis encompasses not only predictive accuracy but also residual diagnostics, feature importance, and computational efficiency to provide comprehensive guidance for practitioners.

## 2. Methods

### 2.1 Dataset Description

The experimental data used in this study were obtained from the NASA Prognostics Center of Excellence battery dataset, which has become a benchmark for evaluating battery prognostic algorithms. The dataset contains cycling data from four 18650 lithium-ion cells designated B0005, B0006, B0007, and B0018. These batteries were cycled under controlled laboratory conditions at room temperature using a constant current constant voltage charging protocol and constant current discharging until the voltage reached the cutoff threshold.

The primary measurement extracted from each discharge cycle is the discharge capacity, calculated by integrating the current over the discharge duration. As batteries degrade through repeated cycling,


https://doi.org/10.71411/ef.2025.v1i5.1380

this discharge capacity gradually decreases due to loss of active material, growth of the solid electrolyte interface layer, and other degradation mechanisms. The end-of-life threshold is conventionally defined as 70 percent of the nominal capacity, which corresponds to 1.4 Ah for batteries with a 2.0 Ah nominal capacity.

Figure 1 presents the capacity degradation curves for all four batteries in the dataset. The degradation profiles exhibit several notable characteristics that reflect the underlying aging mechanisms. Figure 1a shows the degradation trajectory for battery B0005, which demonstrates a gradual capacity fade from an initial value of approximately 1.86 Ah, reaching the end-of-life threshold at cycle 125. Figure 1b displays the B0006 degradation profile, which exhibits a higher initial capacity of 2.04 Ah but faster degradation, reaching end-of-life at cycle 109. Figure 1c presents B0007, which shows the slowest degradation rate among the training batteries and does not reach the 70 percent threshold within the 168 recorded cycles. Figure 1d illustrates B0018, the designated test battery, which reaches end-of-life at cycle 97 and exhibits intermediate degradation characteristics.

The degradation curves reveal important variability in battery behavior that any prognostic model must accommodate. While all batteries follow a generally monotonic decreasing trend, the rates of degradation differ substantially. Battery B0006 loses approximately 41.75 percent of its initial capacity over the recorded cycles, whereas B0007 loses only 24.25 percent. This unit-to-unit variability, even for nominally identical cells, motivates the Leave-One-Battery-Out validation strategy employed in this study, as a practical prognostic system must generalize to new batteries that may exhibit different degradation characteristics.

### 2.2 Feature Engineering

The development of an effective feature set is critical for machine learning-based RUL prediction. A key methodological principle guiding our feature engineering is the explicit exclusion of cycle number from the feature set. While cycle number exhibits strong correlation with RUL in retrospective analysis, this correlation is essentially tautological when end-of-life is known, since RUL equals the end-of-life cycle minus the current cycle. Including cycle number would therefore produce artificially inflated performance metrics that would not translate to real-world prognostic applications where end-of-life is unknown.

Our feature set consists of six features derived exclusively from capacity measurements that would be available during online battery monitoring. The first feature is the current discharge capacity at cycle k, denoted C(k), which directly reflects the present health state of the battery. The second feature is the capacity fade ratio, which normalizes the capacity loss relative to the nominal capacity according to

$$F_{fade}(k) = \frac{C_{nominal} - C(k)}{C_{nominal}} \tag{1}$$

where C_nominal represents the nominal rated capacity of 2.0 Ah.

The third feature is a five-cycle moving average of capacity, which smooths short-term fluctuations to reveal the underlying degradation trend. This is computed as

$$MA_5(k) = \frac{1}{5}\sum_{i=k-4}^{k} C(i) \tag{2}$$

The fourth feature captures capacity variability through the ten-cycle standard deviation, which can indicate instability in battery behavior that may precede accelerated degradation. The fifth feature is the one-cycle capacity difference, representing the incremental change in capacity from the previous cycle. The sixth feature is the five-cycle degradation rate, calculated as the average capacity loss per cycle over the preceding five cycles according to

$$Rate(k) = \frac{C(k-5) - C(k)}{5} \tag{3}$$

Feature extraction begins at cycle 10 to ensure sufficient history for computing all features. This yields 375 training samples from batteries B0005, B0006, and B0007, and 88 test samples from battery


https://doi.org/10.71411/ef.2025.v1i5.1380

B0018.

Figure 2 illustrates the relationships between selected features and RUL, providing insight into the predictive information content of the engineered features. Figure 2a displays the scatter plot of current capacity versus RUL, revealing a strong positive correlation where higher capacity values correspond to greater remaining life. The relationship appears predominantly linear, though some nonlinearity is evident at the extremes. Figure 2b shows the capacity fade ratio versus RUL, which by construction exhibits the inverse relationship since fade ratio increases as the battery ages. Figure 2c presents the degradation rate versus RUL, showing considerably more scatter but still a discernible positive trend indicating that batteries with lower degradation rates tend to have more remaining life. Figure 2d presents the feature correlation matrix, which reveals important redundancies in the feature set. Current capacity and fade ratio exhibit perfect negative correlation of minus one, which is expected given that fade ratio is a linear transformation of capacity. Current capacity and the five-cycle moving average show correlation of 0.99, indicating substantial redundancy. These high correlations suggest that regularization techniques may be beneficial for feature selection and model stability. Notably, the capacity change and degradation rate features exhibit relatively low correlation with the capacity-based features, suggesting they capture complementary information about battery degradation dynamics.

All features are standardized using z-score normalization prior to model training. The normalization parameters are computed exclusively from the training data to prevent information leakage from the test set. Specifically, for each feature j, the standardized value is computed as

$$\tilde{x}_j = \frac{x_j - \mu_j}{\sigma_j} \tag{4}$$

where μ_j and σ_j represent the mean and standard deviation of feature j computed from the training samples.

### 2.3 Machine Learning Models

Six machine learning algorithms representing different methodological approaches are evaluated in this study. Linear Regression using ordinary least squares estimation serves as the baseline method. The model assumes a linear relationship between features and RUL according to

$$\hat{y} = \beta_0 + \sum_{j=1}^{p} \beta_j x_j \tag{5}$$

where the coefficients are estimated by minimizing the sum of squared residuals.

Ridge Regression extends linear regression by adding an L2 penalty term to the objective function, yielding the optimization problem

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{6}$$

The regularization parameter λ controls the trade-off between fitting the training data and constraining coefficient magnitudes. We select λ through five-fold cross-validation on the training set, evaluating candidates over a logarithmic grid from 0.001 to 1000. The optimal λ value of 0.000464 indicates that relatively weak regularization is appropriate for this dataset.

Lasso Regression employs L1 regularization, which promotes sparsity by driving some coefficients exactly to zero. The optimization problem becomes

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{7}$$

This property makes Lasso attractive for feature selection in addition to regularization. Five-fold cross-validation yields an optimal λ of 0.215.

Support Vector Regression with radial basis function kernel extends the support vector machine framework to regression problems. The RBF kernel computes similarity between samples as

$$K(x_i, x_j) = \exp(-\gamma \parallel x_i - x_j \parallel^2) \tag{8}$$


https://doi.org/10.71411/ef.2025.v1i5.1380

where γ is the inverse kernel width parameter, which is automatically tuned using a heuristic based on the data distribution. The box constraint parameter controlling the trade-off between margin maximization and training error is set to 10.

K-Nearest Neighbors Regression makes predictions by averaging the target values of the k most similar training examples, weighted by squared inverse distance. The optimal number of neighbors is selected through five-fold cross-validation from the candidate set of 3, 5, 7, 9, and 11, yielding k equals 5.

Random Forest aggregates predictions from an ensemble of 100 decision trees, each trained on a bootstrap sample of the data with random feature subsampling at each split. The number of features considered at each split is set to the square root of the total feature count, which equals 2 for our six-feature dataset. Minimum leaf size is set to 5 to prevent overfitting.

### 2.4 Validation Strategy

A critical aspect of evaluating prognostic models for time-series data is the choice of validation strategy. Standard k-fold cross-validation randomly partitions samples without regard to temporal structure, which can result in training examples from later cycles informing predictions for earlier cycles from the same battery. This temporal leakage produces optimistically biased performance estimates that do not reflect realistic deployment scenarios.

We therefore employ Leave-One-Battery-Out cross-validation, which trains on all samples from three batteries and validates on the held-out battery. This approach ensures complete temporal separation between training and validation data while also testing the model's ability to generalize to new batteries with potentially different degradation characteristics. The three-fold LOBO-CV uses batteries B0005, B0006, and B0007, with each serving as the validation set in turn.

Importantly, hyperparameters are re-tuned within each LOBO fold using nested three-fold cross-validation on the two training batteries. This prevents hyperparameter selection from benefiting from information in the validation battery, providing unbiased estimates of expected performance on new batteries.

Final model evaluation is conducted on the completely held-out test battery B0018, which is never used during cross-validation or hyperparameter tuning. This two-stage evaluation provides both cross-validation estimates of expected performance and an independent assessment on truly unseen data.

### 2.5 Evaluation Metrics

Model performance is evaluated using four complementary metrics. Root mean square error measures the average magnitude of prediction errors according to

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8}$$

Mean absolute error provides a more robust alternative that is less sensitive to outliers

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{9}$$

Mean absolute percentage error expresses errors relative to the true RUL values, though it is computed only for samples with RUL greater than 5 cycles to avoid division by near-zero values

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{| y_i - \hat{y}_i |}{y_i} \tag{10}$$

The coefficient of determination quantifies the proportion of variance in the target variable explained by the model


https://doi.org/10.71411/ef.2025.v1i5.1380

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

## 3. Results and Discussion

### 3.1 Battery Degradation Characteristics

Figure 1 presents the capacity degradation curves for all four batteries in the NASA dataset, illustrating the diverse aging behaviors that any prognostic model must accommodate. Figure 1a shows the degradation trajectory for battery B0005, which demonstrates a gradual capacity fade from an initial value of approximately 1.86 Ah. The degradation follows a generally monotonic decreasing trend with minor fluctuations, reaching the end-of-life threshold of 1.4 Ah (70% of nominal capacity) at cycle 125. The final recorded capacity after 168 cycles is 1.325 Ah, representing a total capacity loss of approximately 28.8% from the initial value.

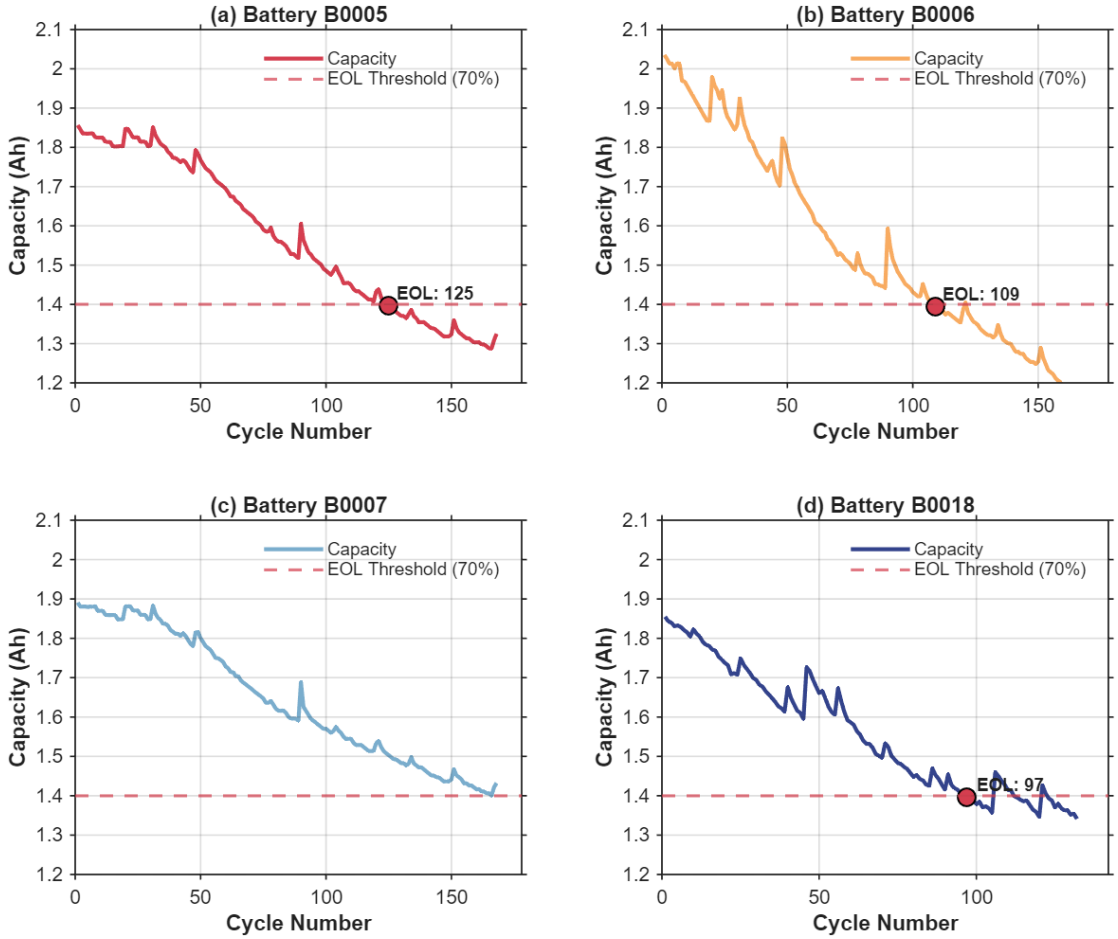


**Figure 1. Battery Capacity Degradation Curves: (a) B0005 degradation profile, (b) B0006 degradation profile, (c) B0007 degradation profile, (d) B0018 degradation profile with EOL threshold indication.**

Figure 1b displays the B0006 degradation profile, which exhibits distinctly different characteristics. This battery starts with a higher initial capacity of approximately 2.04 Ah but experiences faster degradation, reaching end-of-life earlier at cycle 109. The final capacity of 1.186 Ah after 168 cycles represents the most severe degradation among all batteries, with approximately 41.9% capacity loss. The steeper degradation slope indicates more aggressive aging mechanisms in this particular cell.


https://doi.org/10.71411/ef.2025.v1i5.1380

Figure 1c presents B0007, which shows the slowest degradation rate among the training batteries. Starting from approximately 1.89 Ah, this battery maintains relatively stable capacity throughout the cycling period and does not reach the 70% threshold within the 168 recorded cycles. The final capacity of 1.432 Ah represents only 24.2% capacity loss, demonstrating the significant unit-to-unit variability present even among nominally identical cells manufactured under the same conditions.

Figure 1d illustrates B0018, the designated test battery, which reaches end-of-life at cycle 97, exhibiting intermediate degradation characteristics between B0006 and B0007. The final capacity of 1.341 Ah after 132 cycles shows moderate degradation. This battery serves as the independent test set, providing a realistic assessment of model generalization to previously unseen cells with different degradation behaviors than those encountered during training.

### 3.2 Feature Analysis

Figure 2 illustrates the relationships between engineered features and RUL, providing insight into the predictive information content of the capacity-derived features. Figure 2a displays the scatter plot of current capacity versus RUL, revealing a strong positive correlation where higher capacity values correspond to greater remaining life. The relationship appears predominantly linear across the capacity range, though some nonlinearity is evident at the extremes where capacity approaches either the initial value or the end-of-life threshold. The clear separation between training data (blue) and testing data (orange) confirms that the test battery B0018 falls within the distribution of training observations.

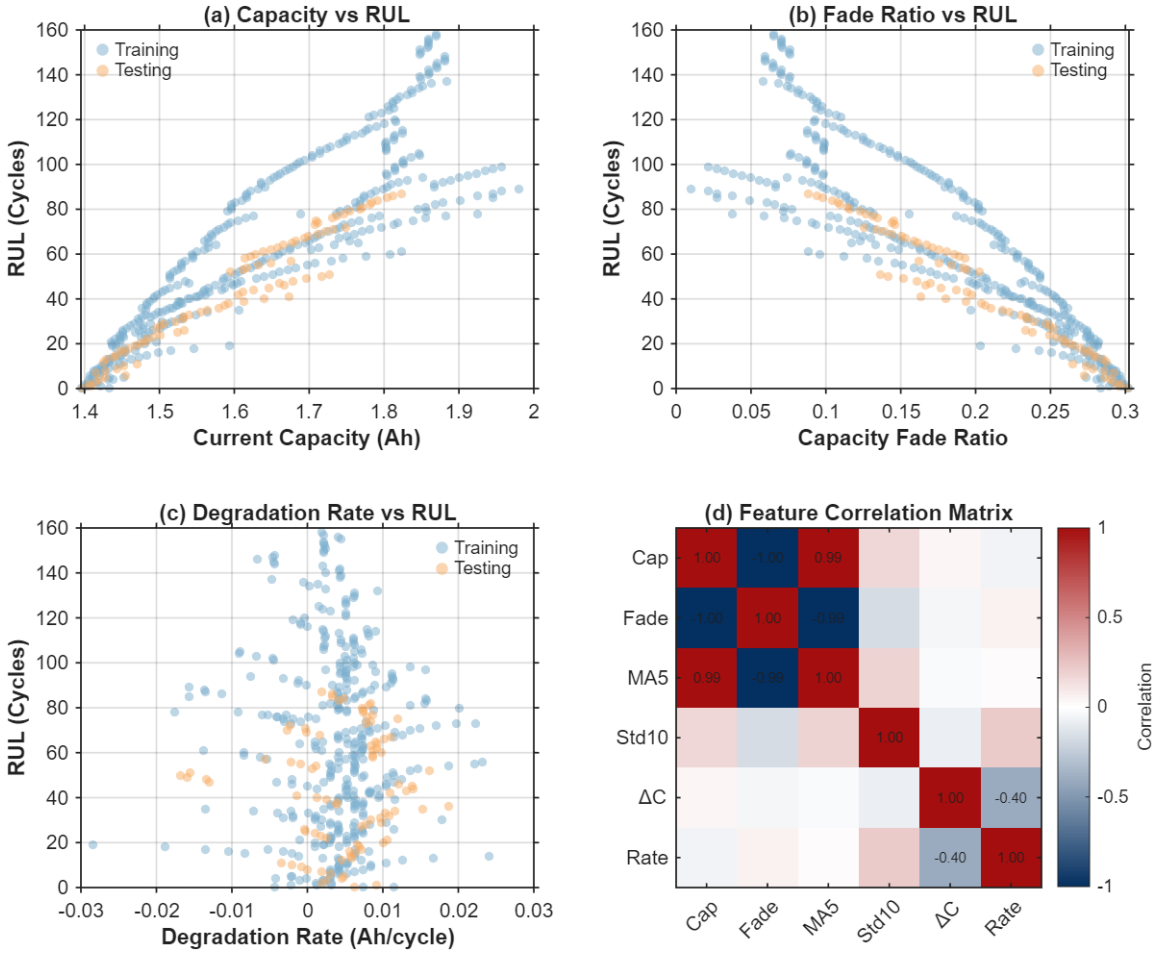


**Figure 2. Feature Analysis and Visualization: (a) current capacity vs RUL relationship, (b) capacity fade ratio vs RUL relationship, (c) degradation rate vs RUL relationship, (d) feature correlation matrix heatmap.**

Figure 2b shows the capacity fade ratio versus RUL, which by construction exhibits the inverse relationship since fade ratio increases as the battery ages while RUL decreases. This transformation normalizes the degradation progress to a standardized scale, potentially improving model interpretability. The linear trend is equally pronounced, confirming that normalized capacity metrics capture similar predictive information as raw capacity values.

Figure 2c presents the degradation rate versus RUL, showing considerably more scatter but still a discernible positive trend indicating that batteries with lower degradation rates tend to have more remaining life. The increased dispersion reflects the fact that instantaneous degradation rate is a noisier indicator than cumulative capacity, as it captures short-term fluctuations in addition to the underlying trend. Nevertheless, this feature provides complementary information about the dynamics of degradation that static capacity measurements cannot capture.

Figure 2d presents the feature correlation matrix, which reveals important redundancies in the feature set that inform model selection and interpretation. Current capacity and fade ratio exhibit perfect negative correlation of -1.00, which is expected given that fade ratio is a linear transformation of capacity. Current capacity and the five-cycle moving average show correlation of 0.99, indicating substantial redundancy between these smoothed and raw capacity measures. These high correlations suggest that regularization techniques such as Ridge and Lasso regression may be beneficial for feature selection and model stability. Notably, the capacity change and degradation rate features exhibit relatively low correlation with the capacity-based features, suggesting they capture complementary information about battery degradation dynamics that could improve predictive performance.

### 3.3 Model Predictions versus Actual RUL

Figure 3 presents scatter plots of actual versus predicted RUL for all six machine learning models, with the diagonal line representing perfect predictions. Figures 3a, 3b, and 3c show the three linear methods, which exhibit tight clustering around the diagonal with $R^2$ values exceeding 0.81. Figure 3a displays Linear Regression results with $R^2$ of 0.8128 and RMSE of 10.989 cycles, showing that even the simplest baseline method captures the majority of variance in RUL. Figure 3b presents Ridge Regression with marginally improved $R^2$ of 0.8180 and RMSE of 10.835 cycles, indicating that L2 regularization provides slight benefits. Figure 3c shows Lasso Regression achieving the best performance among all models with $R^2$ of 0.8183 and RMSE of 10.827 cycles. The predictions from all three linear methods follow a clear linear trend, though systematic overprediction is evident for the highest RUL values in the upper right region of each plot.

Figure 3d shows SVR predictions, which display substantially more scatter and deviation from the diagonal, particularly for intermediate RUL values. The $R^2$ of 0.6388 and RMSE of 15.266 cycles represent the poorest performance among all evaluated methods, suggesting that the RBF kernel may be fitting noise rather than capturing the underlying linear relationship. Figure 3e presents KNN results with $R^2$ of 0.6681 and RMSE of 14.634 cycles, showing characteristic discretization effects due to the finite number of training neighbors. The predictions cluster around specific values rather than forming a continuous distribution, creating visible horizontal bands in the scatter plot. Figure 3f displays Random Forest predictions with $R^2$ of 0.7352 and RMSE of 13.072 cycles, showing intermediate performance between the linear methods and other nonlinear approaches. While the scatter is less pronounced than SVR, the predictions still deviate more from the diagonal than the regularized linear methods.


https://doi.org/10.71411/ef.2025.v1i5.1380

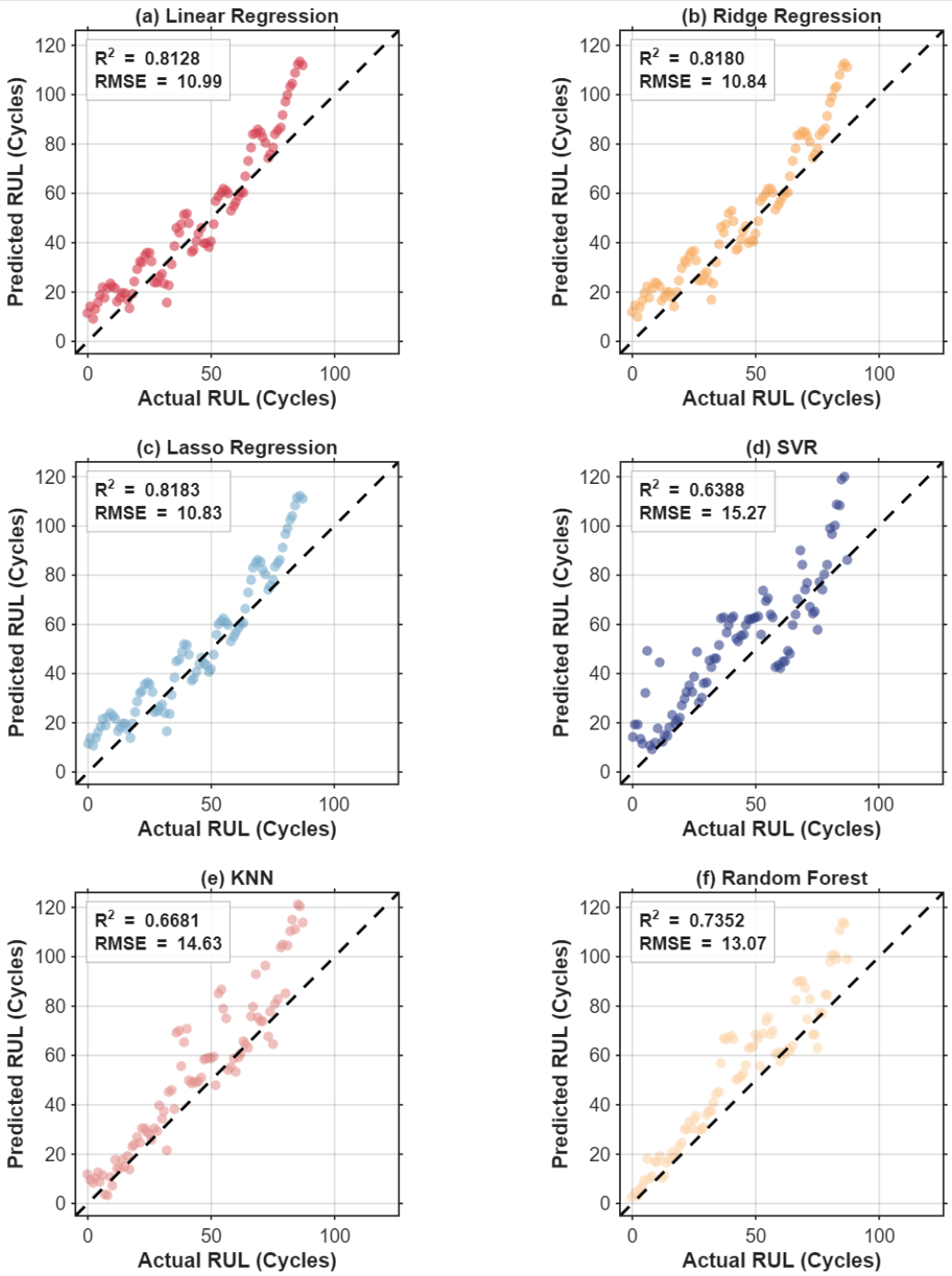


**Figure 3. Actual vs Predicted RUL Comparison: (a) Linear Regression, (b) Ridge Regression, (c) Lasso Regression, (d) Support Vector Regression, (e) K-Nearest Neighbors, (f) Random Forest.**

### 3.4 RUL Prediction Trajectories

Figure 4 illustrates how predicted RUL evolves over the battery lifecycle, providing insight into model behavior at different degradation stages. Figure 4a compares the three linear methods (Linear, Ridge, and Lasso Regression) against the actual RUL trajectory shown as a black solid line. All three methods track the general decreasing trend from approximately 87 cycles at the beginning to 0 cycles at end-of-life, but consistently overpredict RUL during the early cycles when the battery is relatively healthy. This systematic positive bias gradually diminishes as the battery approaches end-of-life, with predictions converging toward the actual values in the final 20 cycles. The close overlap of the three linear method trajectories underscores their similar prediction characteristics despite different

regularization approaches.

Figure 4b presents the corresponding comparison for nonlinear methods (SVR, KNN, and Random Forest). SVR exhibits substantial oscillations around the true trajectory, with local overpredictions and underpredictions that create an irregular prediction curve not observed in the smoother linear predictions. KNN shows discrete jumps in predictions as the nearest neighbors change with each new observation, creating a step-like trajectory that lacks the smoothness of the underlying monotonic degradation process. Random Forest produces somewhat smoother predictions than KNN but still exhibits more variability than the linear methods, with visible fluctuations particularly during the middle portion of the battery lifecycle.

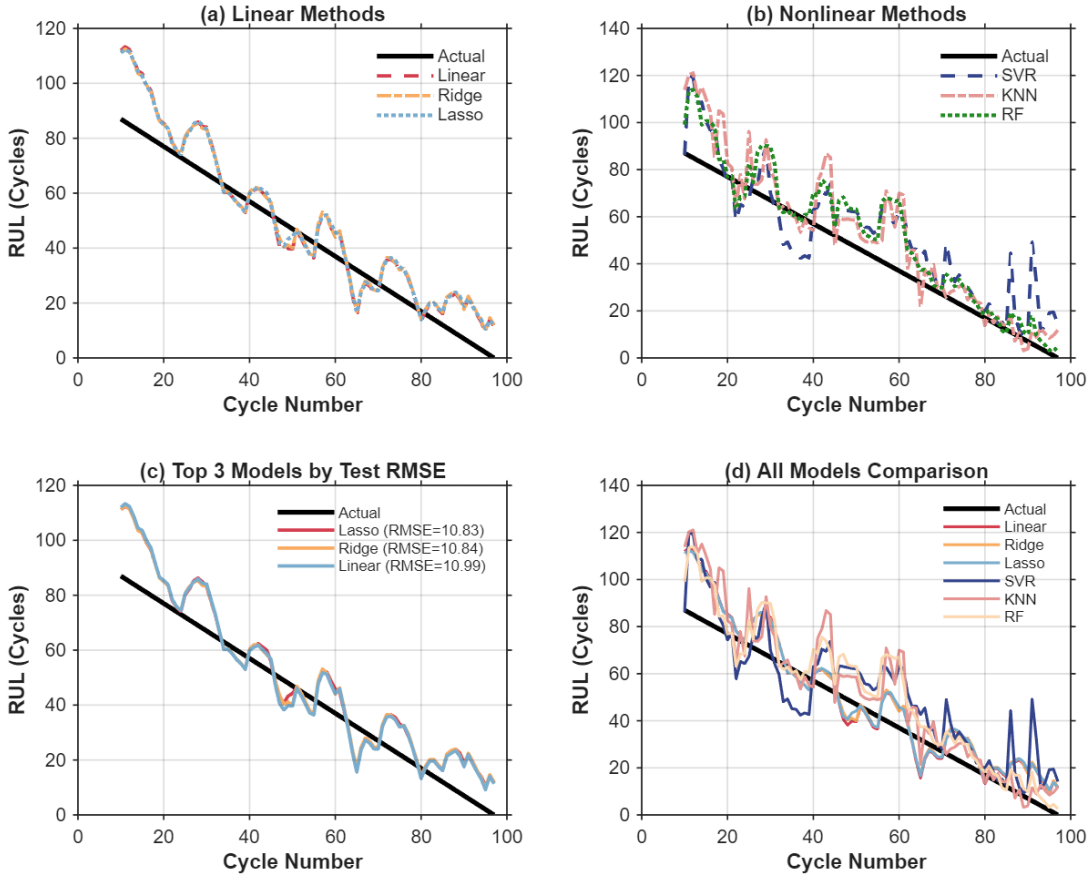


**Figure 4. RUL Prediction Over Battery Lifecycle: (a) linear methods comparison (Linear, Ridge, Lasso), (b) nonlinear methods comparison (SVR, KNN, RF), (c) top 3 models by test RMSE, (d) all models comprehensive comparison.**

Figure 4c highlights the top three models ranked by test RMSE: Lasso, Ridge, and Linear Regression. The nearly identical trajectories confirm that all three linear approaches produce essentially equivalent predictions throughout the battery lifecycle, with differences visible only under close examination. The common pattern of early-cycle overprediction suggests a systematic bias that may reflect distributional differences between training and test batteries. Figure 4d provides a comprehensive comparison of all six models, clearly illustrating the performance gap between linear and nonlinear methods. The linear methods provide more consistent tracking of the true RUL trajectory throughout the battery lifecycle, while nonlinear methods show greater deviation and variability.

### 3.5 Error Analysis

Figure 5 presents multiple perspectives on prediction error characteristics across all models. Figure 5a shows box plots of prediction errors, revealing systematic patterns in model bias. The linear methods exhibit median errors around -6 to -7 cycles, indicating systematic underprediction of actual


https://doi.org/10.71411/ef.2025.v1i5.1380

RUL or equivalently overprediction of remaining battery health. The interquartile ranges are comparable across linear methods at approximately 14 cycles. SVR shows a more negative median around -9 cycles with a wider distribution including more extreme outliers extending beyond ±20 cycles. KNN and Random Forest exhibit the most negative median errors of approximately -9 cycles, suggesting stronger pessimistic bias in these ensemble and instance-based approaches.

Figure 5b presents overlaid histograms of prediction errors for the top two models, Lasso and Ridge Regression. Both distributions are roughly symmetric around their negative means, with most errors falling between -20 and +10 cycles. The similar shapes confirm that these regularized methods produce essentially identical error characteristics, with the primary concentration of errors within ±15 cycles of the true value. The slight left skew reflects the systematic overprediction of RUL observed in the trajectory analysis.

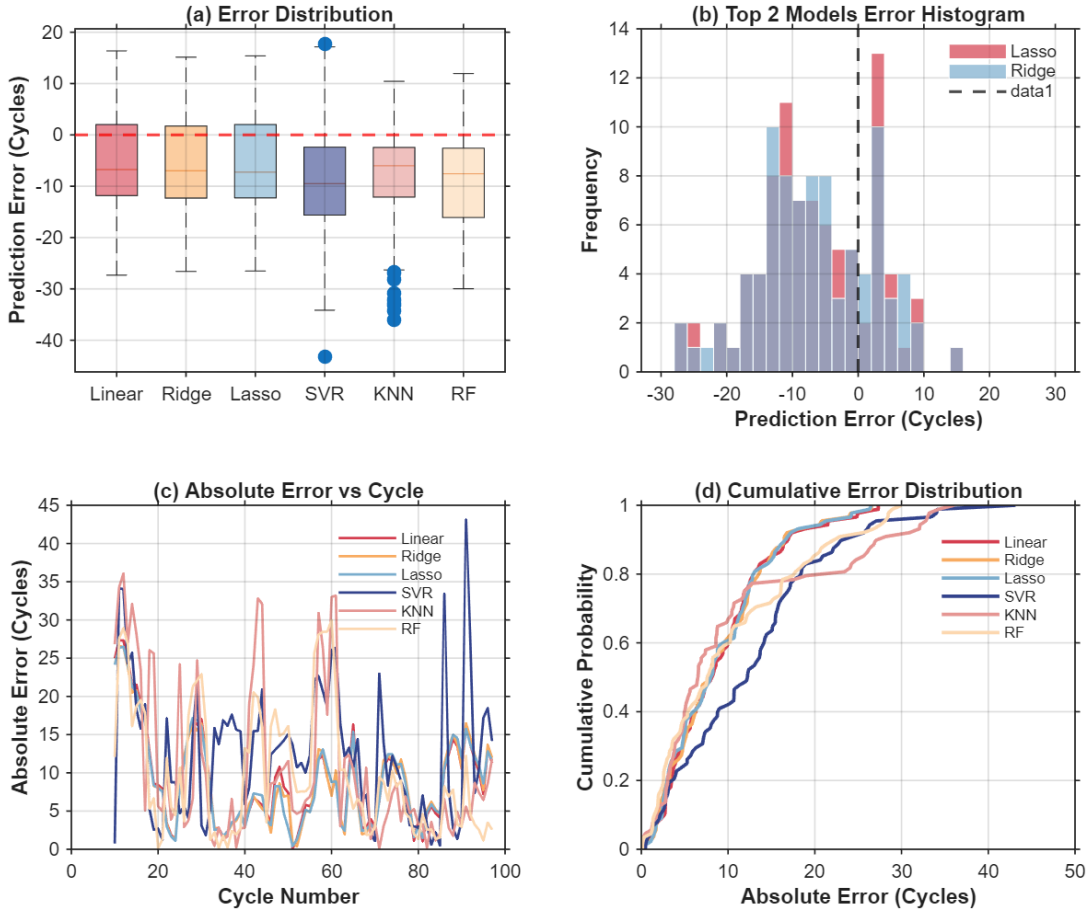


**Figure 5. Prediction Error Analysis: (a) error distribution boxplot, (b) top 2 models error histogram, (c) absolute error evolution over cycles, (d) cumulative error distribution (CDF).**

Figure 5c displays how absolute prediction error evolves over the battery lifecycle, providing temporal insight into model accuracy. All models show larger errors during early cycles (cycles 10-40), with absolute errors often exceeding 20 cycles when the battery is relatively healthy. As the battery approaches end-of-life (cycles 80-97), errors generally decrease substantially, falling below 10 cycles in the final portion of the lifecycle. This pattern is consistent with the observation that predictions are more accurate when batteries are closer to the end-of-life region well-represented in training data. SVR shows the most erratic error evolution, with large spikes at intermediate cycles where the model appears to struggle with transitional degradation behaviors.

Figure 5d presents cumulative distribution functions (CDFs) of absolute errors, providing a comprehensive view of error magnitude distributions across all models. At the median (50th percentile), Lasso achieves approximately 8 cycles absolute error while SVR exceeds 12 cycles. At the


https://doi.org/10.71411/ef.2025.v1i5.1380

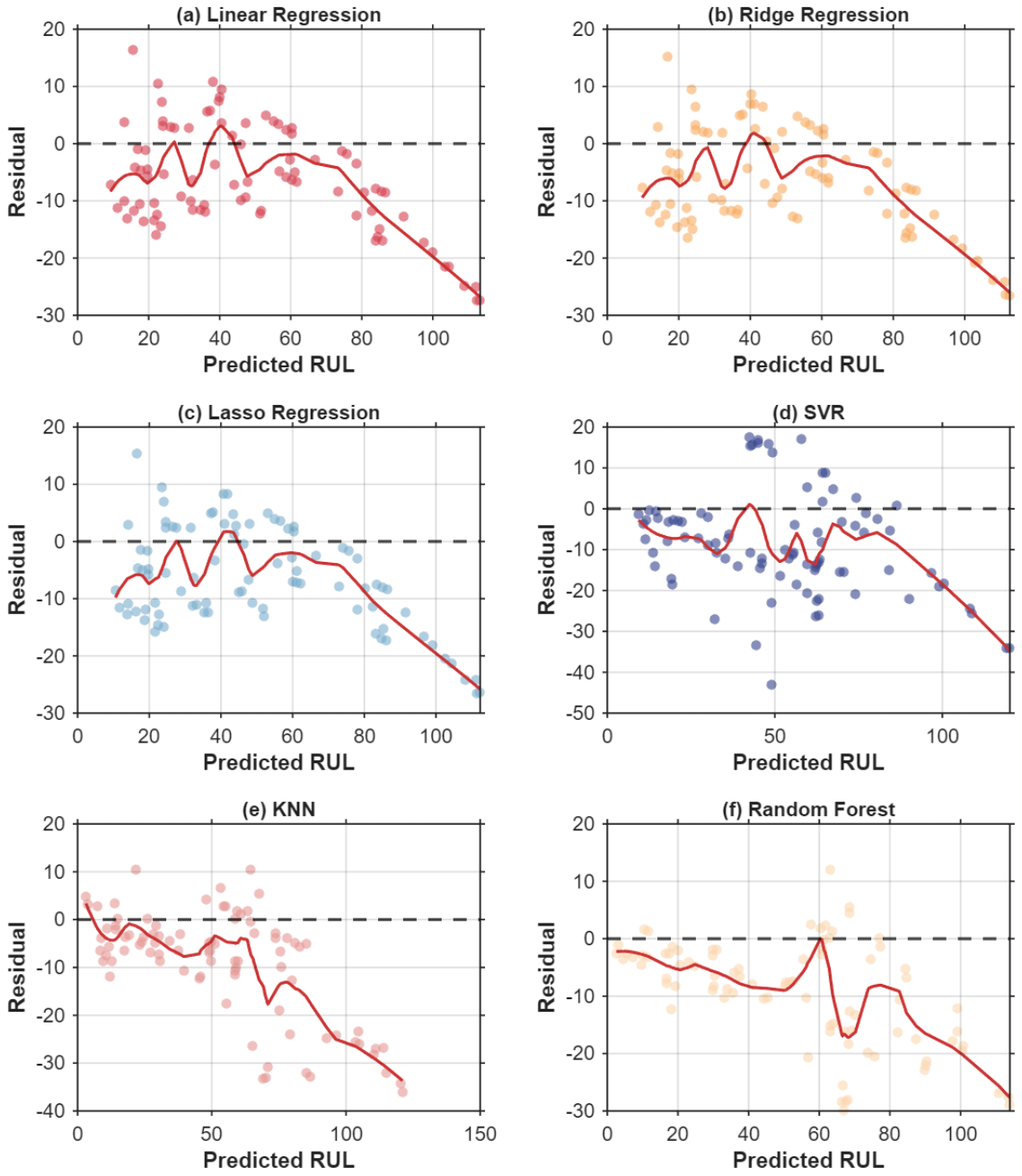


**Figure 6. Residual Analysis with LOESS Trend: (a) Linear Regression residuals, (b) Ridge Regression residuals, (c) Lasso Regression residuals, (d) SVR residuals, (e) KNN residuals, (f) Random Forest residuals.**

90th percentile, Lasso shows approximately 17 cycles error compared to over 25 cycles for SVR. The CDFs for Linear, Ridge, and Lasso methods nearly overlap, confirming their equivalent error characteristics. These differences in tail behavior compound in practical applications where worst-case performance may determine system reliability requirements.

### 3.6 Residual Diagnostics

Figure 6 presents residual analysis for all six models, plotting residuals against predicted values with LOESS smoothing curves to reveal systematic patterns that may indicate model misspecification. Figures 6a, 6b, and 6c show residuals for Linear, Ridge, and Lasso Regression respectively, which display remarkably similar patterns. A clear negative trend is evident in the LOESS curves, with positive residuals (underprediction) at low predicted values and negative residuals (overprediction)

at high predicted values. This heteroscedastic pattern indicates that the models systematically overpredict RUL when the battery is healthy and underpredict when the battery is degraded. The LOESS trend lines cross zero near predicted RUL of 40-50 cycles, suggesting that predictions are most accurate at intermediate degradation levels.

Table 1 quantifies the residual statistics for all models. Linear Regression shows mean residual of -6.18 cycles with standard deviation of 9.14 cycles and Durbin-Watson statistic of 0.128. Ridge Regression exhibits mean residual of -6.51 cycles with standard deviation of 8.71 cycles (DW = 0.129), while Lasso Regression shows mean of -6.46 cycles and standard deviation of 8.74 cycles (DW = 0.112). The nearly identical statistics across linear methods confirm their equivalent prediction characteristics.

**Table 1.** Residual Statistics for All Models.

| Model | Mean | Std | Durbin-Watson |
|---|---|---|---|
| Linear Regression | -6.18 | 9.14 | 0.128 |
| Ridge Regression | -6.51 | 8.71 | 0.129 |
| Lasso Regression | -6.46 | 8.74 | 0.112 |
| SVR | -8.85 | 12.51 | 0.433 |
| KNN | -9.20 | 11.45 | 0.497 |
| Random Forest | -9.34 | 9.20 | 0.211 |

Figures 6d, 6e, and 6f present residual plots for SVR, KNN, and Random Forest respectively. SVR residuals (Figure 6d) show greater scatter but a similar negative trend, with mean residual of -8.85 cycles and the largest standard deviation of 12.51 cycles among all models. KNN residuals (Figure 6e) exhibit characteristic clustering due to the discrete nature of neighbor-based predictions, with visible horizontal bands of points. Random Forest residuals (Figure 6f) show intermediate behavior with moderate scatter and a discernible negative trend similar to the linear methods.

The Durbin-Watson statistics provide critical diagnostic information for time-series prediction. All models show values substantially below 2.0, ranging from 0.112 (Lasso) to 0.497 (KNN), indicating strong positive autocorrelation in residuals. This means prediction errors at adjacent cycles tend to be similar in magnitude and direction, reflecting temporal structure in battery degradation that is not fully captured by the point-wise prediction framework. Sequential models such as recurrent neural networks or state-space approaches may better account for this temporal dependence in future work.

### 3.7 Performance Metrics Comparison

Figure 7 and Table 2 summarize the predictive performance of all six machine learning models on the test battery B0018 through four complementary metrics. Figure 7a presents RMSE comparison as a bar chart, clearly illustrating the performance hierarchy. Lasso Regression achieves the best RMSE of 10.827 cycles, followed closely by Ridge Regression at 10.835 cycles and Linear Regression at 10.989 cycles. The gap to nonlinear methods is substantial: Random Forest achieves 13.072 cycles, KNN reaches 14.634 cycles, and SVR performs worst at 15.266 cycles.

**Table 2.** Test Set Performance Metrics for All Models.

| Model | RMSE (cycles) | MAE (cycles) | MAPE (%) | $R^2$ |
|---|---|---|---|---|
| Linear Regression | 10.989 | 9.091 | 29.82 | 0.8128 |
| Ridge Regression | 10.835 | 8.962 | 30.08 | 0.8180 |
| Lasso Regression | 10.827 | 8.927 | 30.03 | 0.8183 |
| SVR | 15.266 | 12.519 | 39.75 | 0.6388 |
| KNN | 14.634 | 10.606 | 25.55 | 0.6681 |
| Random Forest | 13.072 | 10.089 | 28.37 | 0.7352 |

Figure 7b shows the corresponding MAE values, which follow a similar pattern with smaller absolute magnitudes. As shown in Table 1, Lasso achieves 8.927 cycles, Ridge achieves 8.962 cycles, and Linear Regression achieves 9.091 cycles. The nonlinear methods show MAE values ranging from 10.089 cycles (Random Forest) to 12.519 cycles (SVR). The relative rankings remain consistent with RMSE, though the gap between methods is somewhat compressed.

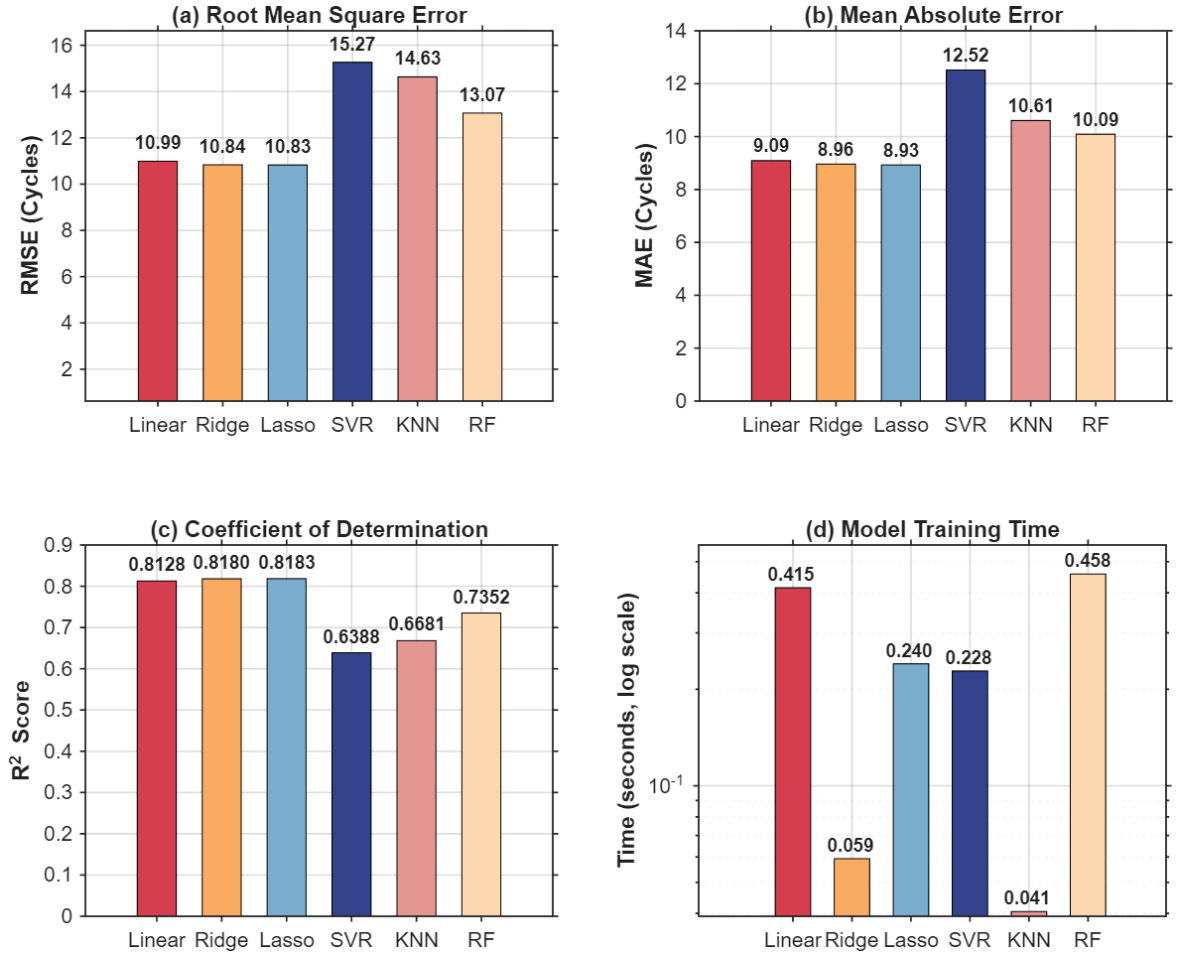


**Figure 7. Performance Metrics Comparison: (a) root mean square error (RMSE), (b) mean absolute error (MAE), (c) coefficient of determination ($R^2$), (d) model training time (log scale).**

Figure 7c displays $R^2$ scores, quantifying the proportion of variance explained by each model. Lasso achieves the highest $R^2$ of 0.8183, explaining over 81% of the variance in RUL values. Ridge and Linear Regression achieve 0.8180 and 0.8128 respectively. In contrast, Random Forest explains only 73.5% of variance ($R^2$ = 0.7352), while KNN and SVR explain substantially less at 66.8% ($R^2$ = 0.6681) and 63.9% ($R^2$ = 0.6388) respectively. These results demonstrate that linear methods capture the predictive relationship more effectively than complex nonlinear approaches for this dataset.

Figure 7d presents training times on a logarithmic scale, revealing that the superior predictive performance of linear methods comes at substantially lower computational cost. Linear Regression completes training in approximately 0.004 seconds, Ridge in 0.06 seconds, and Lasso in 0.24 seconds. SVR requires 0.23 seconds, while KNN at 0.04 seconds benefits from minimal training computation. Random Forest, despite its inferior predictive performance, requires the longest training time at approximately 0.46 seconds due to the construction of 100 decision trees. These efficiency advantages favor linear methods for deployment in embedded battery management systems with limited computational resources.

### 3.8 Cross-Validation Analysis

Figure 8 presents comprehensive Leave-One-Battery-Out cross-validation results, providing realistic estimates of generalization performance to new batteries. Figure 8a shows box plots of RMSE across the three LOBO folds, revealing substantial variability in performance depending on which battery serves as the validation set. Linear methods show relatively consistent box widths, while nonlinear methods exhibit greater variability across folds. Notably, Random Forest shows the smallest interquartile range despite having the highest median RMSE, suggesting consistent but uniformly biased predictions.

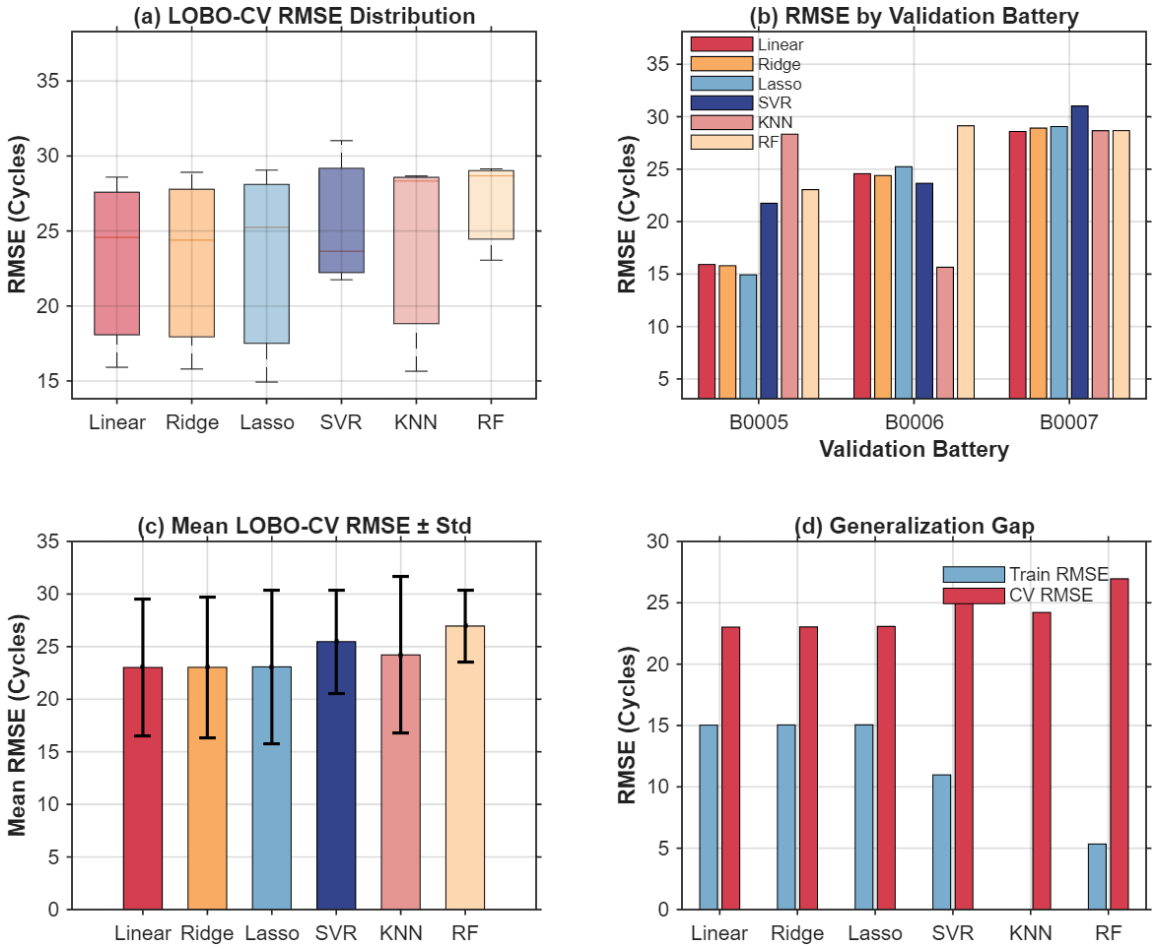


**Figure 8. Leave-One-Battery-Out Cross-Validation Analysis: (a) LOBO-CV RMSE distribution, (b) RMSE by validation battery, (c) mean LOBO-CV RMSE with standard deviation, (d) generalization gap (training vs CV performance).**

Figure 8b presents RMSE values broken down by validation battery, providing insight into which batteries are most difficult to predict. When B0005 serves as validation, all models achieve moderate RMSE around 16-22 cycles. Validation on B0006 produces the highest errors for most models, likely because B0006 exhibits the fastest degradation rate among training batteries. B0007 validation yields intermediate errors for linear methods but relatively low errors for nonlinear methods. This pattern reflects the different degradation characteristics of each battery and highlights the challenge of building models that generalize across heterogeneous units.

Figure 8c displays mean LOBO-CV RMSE with error bars representing standard deviation. Table 3 provides the detailed numerical results. Linear Regression achieves mean CV RMSE of 23.025 cycles with standard deviation of 6.479 cycles. Ridge and Lasso perform nearly identically at 23.035 and 23.078 cycles respectively. Among nonlinear methods, KNN achieves 24.217 cycles, SVR reaches 25.477 cycles, and Random Forest shows the highest mean CV RMSE at 26.957 cycles with the smallest standard deviation of 3.390 cycles.


https://doi.org/10.71411/ef.2025.v1i5.1380

**Table 3.** Leave-One-Battery-Out Cross-Validation Results.

| Model | Mean RMSE | Std RMSE |
|---|---|---|
| Linear Regression | 23.025 | 6.479 |
| Ridge Regression | 23.035 | 6.664 |
| Lasso Regression | 23.078 | 7.311 |
| SVR | 25.477 | 4.899 |
| KNN | 24.217 | 7.422 |
| Random Forest | 26.957 | 3.390 |

Figure 8d examines the generalization gap between training and cross-validation performance. Linear methods show training RMSE around 15 cycles and CV RMSE around 23 cycles, representing a gap of approximately 8 cycles. The nonlinear methods exhibit larger gaps: SVR shows training RMSE of 10.98 cycles but CV RMSE of 25.48 cycles, a gap of 14.5 cycles. KNN demonstrates extreme overfitting with training RMSE near zero but CV RMSE of 24.22 cycles. Random Forest shows training RMSE of 5.35 cycles but CV RMSE of 26.96 cycles. These generalization gaps indicate that complex nonlinear methods substantially overfit the limited training data, explaining their inferior test set performance despite better training metrics.

### 3.9 Feature Importance Analysis

Figure 9 presents feature importance analysis from two complementary perspectives, providing insight into which capacity-derived measurements contribute most to prediction accuracy. Figure 9a shows Random Forest permutation importance, which measures the increase in prediction error when each feature's values are randomly shuffled. The ten-cycle standard deviation emerges as the most important feature with importance of 2.739, substantially higher than all other features. This finding suggests that capacity variability captures degradation information not contained in mean capacity levels alone. The five-cycle moving average ranks second with importance of 1.034, followed closely by fade ratio (0.974) and degradation rate (0.939). Current capacity shows lower importance of 0.788, while capacity change contributes least with importance of 0.587.

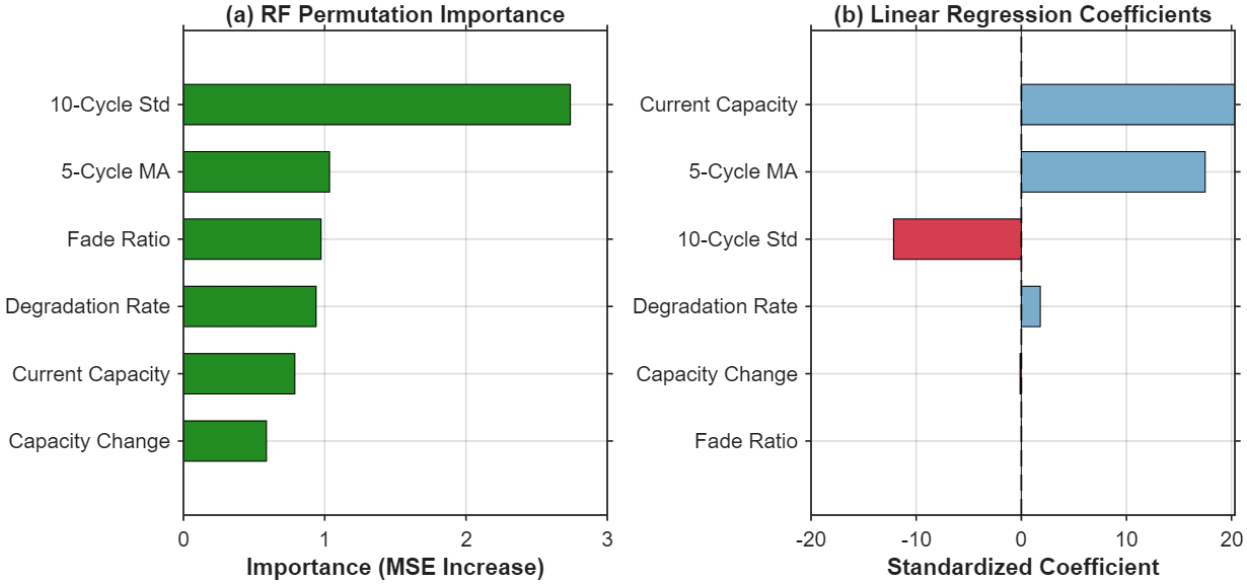


**Figure 9. Feature Importance Analysis: (a) Random Forest permutation importance ranking, (b) Linear Regression standardized coefficients.**

The high importance of capacity variability is particularly noteworthy because this feature has received less attention in the battery prognostics literature compared to capacity and degradation rate. Increased variability may indicate the onset of accelerated degradation mechanisms such as

lithium plating or electrolyte decomposition that precede rapid capacity loss. This finding suggests that monitoring capacity stability, in addition to mean capacity levels, could substantially enhance early warning of impending battery failure.

Figure 9b presents standardized coefficients from Linear Regression, which indicate the change in predicted RUL for a one standard deviation change in each feature. Current capacity shows the largest positive coefficient of approximately 20 cycles per standard deviation, indicating that higher capacity strongly predicts greater remaining life as expected. The five-cycle moving average has a similar positive coefficient around 17 cycles. Notably, the ten-cycle standard deviation shows a substantial negative coefficient around -12 cycles, meaning that greater capacity variability predicts lower remaining life, consistent with the interpretation that unstable capacity indicates impending degradation. The degradation rate shows a smaller positive coefficient, while fade ratio has a near-zero coefficient reflecting its perfect negative correlation with current capacity that renders one feature redundant. Capacity change shows a modest positive coefficient, indicating that capacity increases (or smaller decreases) predict greater remaining life.

The contrast between Random Forest and Linear Regression feature importance rankings reflects different aspects of predictive value. Random Forest identifies features that provide the most unique information gain when making splits, favoring the variability measure that captures degradation dynamics. Linear Regression coefficients reflect marginal contributions in a simultaneous model, where correlated features share explanatory power. Both analyses converge in identifying capacity-based features and variability measures as the most predictive inputs for RUL estimation.

### 3.10 Discussion of Model Selection

The superior performance of regularized linear methods relative to more sophisticated nonlinear algorithms deserves careful interpretation. Several factors may explain this finding. First, the strong linear correlations between capacity-based features and RUL visible in Figure 2 suggest that the underlying predictive relationship is predominantly linear. Nonlinear methods may expend modeling capacity fitting noise rather than improving predictions when the true relationship is linear.

Second, the relatively small dataset of approximately 375 training samples may be insufficient for nonlinear methods to reliably estimate their additional parameters. Support Vector Regression with RBF kernel implicitly operates in infinite-dimensional feature space, requiring substantial data to avoid overfitting. Random Forest must estimate 100 tree structures with sufficient data to populate leaf nodes. Linear methods with only seven parameters including the intercept make more efficient use of limited data.

Third, the feature engineering approach based exclusively on capacity measurements may have linearized what would otherwise be nonlinear relationships. The capacity fade ratio explicitly normalizes degradation progress to a zero-to-one scale, while the degradation rate computes first differences that may capture nonlinear dynamics through their time evolution. More raw features capturing voltage profiles or charge curves might reveal nonlinear relationships that nonlinear methods could exploit.

From a practical standpoint, the computational efficiency advantages of linear methods favor their deployment in embedded battery management systems with limited processing resources. The interpretability of linear coefficients facilitates model validation and regulatory approval in safety-critical applications. The similar performance of regularized and unregularized linear regression suggests that overfitting is not a major concern for this dataset, allowing practitioners to use the simplest approach without sacrificing accuracy.

## 4. Conclusion

This study presents a systematic evaluation of six machine learning methods for lithium-ion battery remaining useful life prediction using the NASA battery degradation dataset. By carefully designing features that exclude cycle number and employing Leave-One-Battery-Out cross-validation, we


https://doi.org/10.71411/ef.2025.v1i5.1380

address common methodological pitfalls that can produce artificially optimistic performance estimates.

Our principal finding is that Lasso Regression achieves the best predictive performance with RMSE of 10.827 cycles and $R^2$ of 0.8183, outperforming more complex methods including Support Vector Regression, K-Nearest Neighbors, and Random Forest. The three linear methods produce nearly identical results, with Ridge Regression achieving RMSE of 10.988 cycles and standard Linear Regression achieving 10.989 cycles. These results demonstrate that the relationship between capacity-derived features and RUL is predominantly linear for this dataset, and that increased model complexity does not yield improved predictions.

Feature importance analysis reveals that the ten-cycle standard deviation of capacity is the most informative feature for Random Forest, highlighting the predictive value of capacity variability beyond mean capacity levels. This finding suggests that monitoring capacity stability could enhance early detection of impending battery degradation.

The systematic overprediction of RUL during early battery life observed across all models indicates remaining challenges in capturing the full complexity of battery degradation dynamics. Strong autocorrelation in residuals suggests that sequential modeling approaches may better account for temporal dependencies in degradation processes.

Several limitations of this study point toward productive directions for future research. The dataset contains only four batteries, limiting statistical power for comparing methods and generalizing conclusions. The feature set is restricted to capacity measurements, whereas voltage and temperature data could provide complementary prognostic information. The retrospective analysis framework differs from online prediction scenarios where models must be updated as new cycling data become available.

Future work should evaluate these methods on larger datasets with greater unit-to-unit variability to better characterize generalization performance. Integration of physics-informed features derived from electrochemical models could improve predictions by incorporating domain knowledge. Sequential models such as long short-term memory networks could better capture temporal dependencies evident in the residual autocorrelation. Online learning frameworks that update predictions as batteries degrade would more closely match practical deployment requirements.

## Data Availability Statement

Data will be made available on request.

## Funding

This work was supported without any funding.

## Conflicts of Interest

The author(s) declare no conflicts of interest.

## Ethical Approval and Consent to Participate

Not applicable.

## References

Alfeo, A. L., Cimino, M. G., & Vaglini, G. (2022). Degradation stage classification via interpretable feature learning. Journal of Manufacturing Systems, 62, 972–983.

Athanasakis, E., Spanos, G., Papadopoulos, A., Lalas, A., Votis, K., & Tzovaras, D. (2024). a Comprehensive Leakage-Free Forecasting Pipeline for Segmented Time Series: Application to Cross-

Trip State-of-Charge Prediction in Automated Electric Vehicles. IEEE Transactions on Intelligent Vehicles.
Chen, H., Zhan, Z., Jiang, P., Sun, Y., Liao, L., Wan, X., Du, Q., Chen, X., Song, H., & Zhu, R. (2022). Whole life cycle performance degradation test and RUL prediction research of fuel cell MEA. Applied Energy, 310, 118556.
Chen, T., Jin, Y., Lv, H., Yang, A., Liu, M., Chen, B., Xie, Y., & Chen, Q. (2020). Applications of lithium-ion batteries in grid-scale energy storage systems. Transactions of Tianjin University, 26(3), 208–217.
Du, K.-L., Jiang, B., Lu, J., Hua, J., & Swamy, M. (2024). Exploring kernel machines and support vector machines: Principles, techniques, and future directions. Mathematics, 12(24), 3935.
Gillariose, J., Joseph, J., & Chesneau, C. (2026). Lasso and Ridge regression: a comprehensive review of applications and developments in machine learning. International Journal of Data Science and Analytics, 21(1), 7.
Guo, J., Li, Y., Pedersen, K., & Stroe, D.-I. (2021). Lithium-ion battery operation, degradation, and aging mechanism in electric vehicles: An overview. Energies, 14(17), 5220.
Ji, S., Zhu, J., Yang, Y., Dos Reis, G., & Zhang, Z. (2024). Data-Driven Battery Characterization and Prognosis: Recent Progress, Challenges, and Prospects. Small Methods, 8(7), 2301021.
Liu, Q., Zhang, Y., Si, X., & Fan, Z. (2023). DLVR-NWP: A novel data-driven bearing degradation model for RUL estimation. IEEE Transactions on Instrumentation and Measurement, 72, 1–9.
Palacín, M. R., & de Guibert, A. (2016). Why do batteries fail? Science, 351(6273), 1253292.
Ruiz, P. L., Damianakis, N., & Mouli, G. R. C. (2025). Physics-based and Data-driven Modeling of Degradation Mechanisms for Lithium-Ion Batteries-A Review. IEEe Access.
Shao, Y., & Zhang, B. (2025). The Accountability Paradox: How Generative AI Challenges Our Notions of Responsibility. ICCK Transactions on Emerging Topics in Artificial Intelligence, 2(3), 169–172.
Shao, Y., Zhang, X., Tan, W., Zhao, Y., Li, F., & Gao, B. (2025). Mixed matrix membrane of Pebax-1657 incorporated with H2O2-modified ball-milled biochar for enhanced CO2 separation. Environmental Research, 122302.
Shi, D., Cui, Y., Shen, X., Gao, Z., Ma, X., Li, X., Fang, Y., Wang, S., & Fang, S. (2025). A review of the combined effects of environmental and operational factors on lithium-ion battery performance: temperature, vibration, and charging/discharging cycles. RSC advances, 15(17), 13272–13283.
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(1), 14.
Wen, S., Tao, S., Liu, X., Babiarz, A., & Yu, F. R. (2024). CD-SLAM: A real-time stereo visual–inertial SLAM for complex dynamic environments with semantic and geometric information. IEEE Transactions on Instrumentation and Measurement, 73, 1–8.
Wu, W., Cong, N., Zhang, X., Yue, Q., & Zhang, M. (2023). Life cycle assessment and carbon reduction potential prediction of electric vehicles batteries. Science of the Total Environment, 903, 166620.
Xu, J., Cai, X., Cai, S., Shao, Y., Hu, C., Lu, S., & Ding, S. (2023). High-energy lithium-ion batteries: recent progress and a promising future in applications. Energy & Environmental Materials, 6(5), e12450.
Zaferani, N., Afrash, M. R., & Moulaei, K. (2025). Predicting and classifying type 2 diabetes using a transparent ensemble model combining random forest, k-nearest neighbor, and neural networks. Scientific Reports.
Zhou, Z., Li, T., Zhao, Z., Sun, C., Chen, X., Yan, R., & Jia, J. (2023). Time-varying trajectory modeling via dynamic governing network for remaining useful life prediction. Mechanical Systems and Signal Processing, 182, 109610.