



ARTICLE

Toward Reliable Battery Life Prediction: A Hybrid Data-Driven Framework with Uncertainty Quantification

Mingqi Liu, Ying Wang*, Wujiang Li, Juyong Cao and Fuyong Yang

School of Mechanical and Automobile Engineering, Shanghai University of Engineering Science, No. 333, LongTeng Rd., Shanghai, 201620, China

*Corresponding Author: Ying Wang. Email: wangyingcae@sues.edu.cn

Received: 17 October 2025; Accepted: 22 December 2025; Published: 12 July 2026

ABSTRACT: Accurately predicting battery life is essential for performance management and system safety. Due to the complexity and diversity of internal mechanisms in lithium-ion batteries, their nonlinear characteristics directly give rise to uncertainty in the battery degradation process. However, most existing prediction methods do not fully account for the uncertainty caused by various factors and only provide a point estimate finally. To address this issue, this paper proposes a new framework that combines Random Forest and Conformal Prediction to predict battery life and quantify the uncertainty of the results. This approach leverages the efficiency of Random Forest while enhancing computational robustness and reliability through conformal prediction. The method utilizes early degradation data to select relevant features. Based on this, high-importance feature combinations are selected, and a Random Forest model is used to obtain point estimates. Then, the Conformal Prediction method is introduced to quantify uncertainty and generate prediction intervals with confidence levels and sample-specific bounds. Furthermore, the proposed method is compared against existing uncertainty quantification approaches, with coverage evaluation conducted to enhance the credibility of the prediction results. This method offers a new perspective for the practical application of battery lifetime prediction. Integrating uncertainty quantification into lithium-ion battery research can improve the reliability of the results and support decision-making in practical applications.

KEYWORDS: Lithium-ion battery; uncertainty quantification; conformal prediction; random forest

1 Introduction

Under the combined pressures of the energy crisis and environmental pollution, more countries are investing substantial resources in developing clean and efficient energy technologies. Among various energy storage technologies, lithium-ion batteries have emerged as one of the most promising options due to their superior performance in energy efficiency and environmental sustainability. At the same time, lithium-ion batteries offer high energy density, long cycle life, and low self-discharge, while remaining relatively environmentally friendly [1–3]. These advantages have driven their widespread adoption in portable electronics, electric vehicles, and grid-scale energy storage systems [4–7]. However, increasing charge-discharge cycles and operational frequencies lead to irreversible internal degradation [8], resulting in aging phenomena such as capacity fade [9,10], power loss, and thermal instability [11–13]. These effects can ultimately lead to complex battery degradation, safety failures, and the end of battery life. Therefore, accurately assessing battery lifespan and instability is critically important.



Current approaches for predicting lithium-ion battery life are primarily divided into model-based and data-driven methodologies [14]. Model-based methods emphasize establishing physical or chemical models that capture the intrinsic degradation mechanisms, material properties, and state-of-health deterioration patterns of batteries [15]. These models combine electrochemical kinetics, thermodynamics, and aging dynamics. They simulate key processes such as electrochemical reactions, material degradation, temperature variation, and capacity loss. In 2001, Bloom et al. [16] proposed a model based on time and Arrhenius kinetics to analyze power loss in batteries. Lian et al. [17] used a semi-empirical model for the capacity degradation. To improve accuracy, Li et al. [18] proposed a single-particle model based on the relationship between SEI growth rate and cycle number to describe lithium battery health status. Subsequent studies have employed more advanced models and methods for battery analysis. These include health assessment methods based on Electrochemical Impedance Spectroscopy [19] and degradation models that incorporate internal resistance and lithium inventory [20]. Techniques such as the H-Infinity Filter have been applied to address model errors and mitigate the effects of unknown external noise [21]. Fractional-order models combined with Unscented Kalman Filtering have also been used to estimate the State of Health, along with Gaussian linear models [22].

In recent years, the development of big data and machine learning technologies has made data-driven methods a prominent research focus. Data-driven methods do not require modeling the complex internal chemical reactions and physical degradation mechanisms of lithium-ion batteries. Instead, it focuses on capturing the relationships between performance degradation data and battery health status, as well as the temporal correlations within degradation data across different time points. This method still offers excellent predictive accuracy and has significant advantages in terms of efficiency and flexibility. The earliest data-driven methods mainly relied on classical machine learning models, such as support vector machines (SVM) and decision trees. These methods extract features from degradation data, such as voltage and capacity, to estimate health status and predict battery life. Representative examples include the ensemble learning model developed by Wang et al. [23], which improves the accuracy of remaining useful life prediction. Another example is the weighted least squares support vector machine method proposed by Xiong et al. [24]. This method uses health indicators extracted specifically from the battery voltage–capacity curve as input. With the improvement of computational power and the increase in data volume, deep learning methods have gradually become the mainstream technology in battery life prediction. The nonlinear degradation point prediction method proposed by Fan et al. [25] is one such approach. Čivilis et al. [26] proposed a few-shot learning technique to create different prediction models. Researchers have gradually realized the limitations of single methods and have begun to explore the integration of multiple methods. For example, a hybrid method combining grey neural networks and particle filtering [27]. By incorporating feature selection, Severson et al. [28] proposed a data-driven method for the early stages of battery use by analyzing the electrochemical features. Methods that combine time series analysis, such as LSTM, can capture long-term dependencies and complex nonlinear relationships in time series data. LSTM can be combined with multi-layer perceptron [29] to capture the trend of changes in battery health status and improve prediction accuracy. When combined with transfer learning methods [30], it is suitable for data-scarce scenarios and demonstrates enhanced applicability in practical applications.

Feature selection is generally the final step in feature engineering. Currently, feature selection methods can be broadly classified into two types: univariate-based feature selection and model-based feature selection. However, these methods may overlook the relationships and redundancy among features, and often rely on specific model assumptions [31]. In contrast, Random Forest (RF) naturally accounts for nonlinear and interaction relationships between features. It also provides a built-in feature importance metric, which allows

for intuitive and stable feature selection [32]. Therefore, using RF for feature selection is both convenient and effective.

At the same time, differences between individual lithium-ion batteries are inevitable due to manufacturing and environmental factors. These differences lead to uncertainty in battery degradation processes, modeling accuracy, and life prediction. Currently, most studies provide only point estimates for battery life and mainly focus on improving prediction accuracy, often overlooking system uncertainty [33,34].

Existing research generally classifies uncertainty into two types: epistemic uncertainty and aleatoric uncertainty [35]. Epistemic uncertainty, also known as model uncertainty, arises from the limitations of the model itself. Aleatoric uncertainty, or data uncertainty, results from measurement errors, noise in the training data, and the uncertainty contained in battery degradation [36]. These uncertainties can affect the accurate estimation of the battery state, thereby impacting the reliability of the system. Therefore, how to quantify and evaluate such uncertainties has become one of the most pressing challenges.

Lin et al. [37] proposed a Gaussian Process model to predict the remaining useful life and its associated uncertainty. Lu et al. [38] proposed a combined random forest and probability model for predicting residual stiffness and evaluating structural fatigue reliability. Xiao and Qiao [39] proposed an outlier-detection-based hybrid scenario generation approach. And this approach is transferable to broader decision-making under uncertainty. Zhu et al. [40] proposed a fusion method known as Informer-Gaussian Process Regression (Informer-GPR). It quantifies the uncertainty in performance degradation prediction and enhances the extraction of key information during system degradation. In recent years, the Conformal Prediction method has also made significant progress in addressing uncertainty quantification problems across various fields. Iwakin and Moazeni [41] combined convolutional neural networks and bidirectional long short-term memory networks to predict water demand patterns using CP. This approach outperforms traditional models in prediction performance and handling uncertainty. Hajibabae et al. [42] used CP to address and estimate uncertainty issues in concrete systems. These studies collectively demonstrate the versatility of the CP method across different fields, from resource demand forecasting to complex engineering applications, highlighting the importance of quantifying uncertainty.

In summary, uncertainty evaluation is crucial for improving the reliability of prediction results. To address this issue, this paper proposes a novel hybrid data-driven method for early prediction of lithium-ion battery lifetime and for quantifying the uncertainty of the prediction results. The key contributions of this work are as follows.

- (1) We propose a unified framework that jointly performs battery life prediction and uncertainty quantification, aiming to improve the reliability and decision usefulness of life prediction results.
- (2) By integrating Random Forest (RF) with Conformal Prediction (CP), we generate model-agnostic prediction intervals based on conformity scores, and the proposed estimation method does not rely on strong distributional assumptions.
- (3) We employ a machine-learning-based feature screening technique to remove redundant or irrelevant inputs and identify early-cycle features that are highly correlated with battery lifetime. Notably, by using only the first 100 cycles, the proposed approach enables earlier and more efficient prognostic diagnosis.

The rest of this paper is as follows. Section 2 describes the relevant theories involved in the proposed method. Section 3 introduces the lithium battery datasets used in this paper. In the Section 4, the performance of the proposed method is analyzed. Section 5 discusses the limitations of the current study and provides suggestions for future work. Finally, the Section 6 summarizes the whole paper.

2 Methods

2.1 Random Forest

Random Forest is an ensemble learning algorithm based on decision trees, which is widely used for regression problems. In Random Forest, multiple decision trees are trained by randomly selecting subsamples from the original dataset under the principle of independent premise [43]. The structural diagram shown in Fig. 1. Different decision trees are trained on different sample sets, effectively reducing the risk of overfitting. For a new input, all trained trees produce their individual outputs (Result 1, Result 2, . . . , Result n), which are then aggregated in the result synthesis block, for example by weighted averaging (for regression) or majority voting (for classification), to obtain the final prediction result.

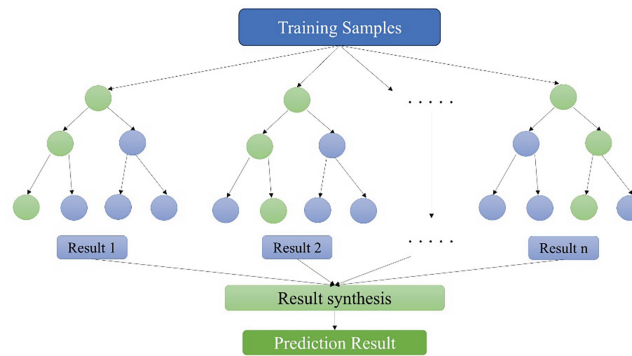


Figure 1: Random forest model

2.2 Variable Importance Measure

Variable importance measure (VIM) is used to calculate the importance of sample features, quantitatively describing their contribution to classification or regression. Random Forest can be employed for feature importance assessment using the Out-of-Bag (OOB) error method. Since OOB data is unbiased, it enables reliable evaluation of a feature's influence on dependent variables [44]. The main procedure includes:

1. Calculate the prediction error rate using Out-of-Bag data.
2. Randomly permute the observed values of variable X_j and recalculate the OOB prediction error rate.
3. Calculate the average value among all the trees of the standardized difference between the two OOB error rates, which is the permutation importance of the variable X_j .

2.3 Conformal Prediction

Conformal prediction is a confidence-based prediction method that typically generates predictions with defined confidence levels. And it is implemented by outputting a prediction interval instead of a single predicted value. All generated prediction intervals will contain true labels or values at a given confidence level, for both classification and regression problems. A critical characteristic of CP lies in its model-agnostic and distribution-free properties [45]. Therefore, it is a flexible framework compatible with any pre-trained machine learning model. This capability enables the transformation of prediction intervals into statistically rigorous uncertainty quantification. The general procedure of conformal prediction can be summarized as follows [46]:

1. The dataset is partitioned into training, calibration, and test sets. A predictive model is trained on the training set to generate predicted values $\hat{y}$ for the target label y based on input features x .

2. According to the significance of the selected model, define a scoring function $s(x, y) \in R$, the larger the value of this function, the worse the consistency between the predicted result $\hat{y}$ of the feature x by the model and the true label y .
3. For a given confidence level $1 - \alpha$, compute $\hat{q}$ as the $\frac{(n+1)(1-\alpha)}{n}$ quantile of the calibration scores $s_1 = s(X_1, Y_1)$, $s_2 = s(X_2, Y_2)$, $\dots$, $s_n = s(X_n, Y_n)$.
4. Use this quantile to form the prediction sets for new points (x_{test}, y_{test}) to form a new prediction set:

$$C(X_{test}) = \{y: s(X_{test}, y) \leq \hat{q}\}. \quad (1)$$

This set is the result range given in conformal prediction, and it can be proven that:

Suppose all the data (x, y) and (x_{test}, y_{test}) are i.i.d. and define $\hat{q}$ as in step 3 above and $C(X_{test})$ as in step 4 above. Then the following holds:

$$P(y_{test} \in C(X_{test})) \geq 1 - \alpha. \quad (2)$$

Under fairly general conditions, it also holds:

$$P(y_{test} \in C(X_{test})) \leq 1 - \alpha + \frac{1}{n+1}. \quad (3)$$

This demonstrates that the true label is contained in the set $C(X_{test})$ with an approximate confidence level $1 - \alpha$.

3 Description of Experimental Data

The dataset used in this study was provided by Reference [28], consisting of 124 commercial LFP lithium iron phosphate batteries cycled under fast-charging conditions. These lithium batteries were manufactured by A123 Systems, model APR18650M1A, with a rated capacity of 1.1 Ah and a nominal voltage of 3.3 V, divided into three batches based on test dates. The experiment was conducted at an ambient temperature of 30°C using a 48-channel Arbin LBT potentiostat. All cells were charged via one-step or two-step fast-charging strategies fixed at 10 min. Discharge employed identical 4C-rate current. Testing ceased when the capacity degraded to 80% of nominal capacity. This dataset is a commercial battery dataset covering cycle lives ranging from 150 to 2300 cycles, including cycle counts, capacity, voltage, current, and temperature per cycle. Fig. 2 shows capacity degradation curves and cycle life distribution. During the testing process, cross-batch evaluation was performed on the batteries in the dataset. The three batches of batteries in the dataset were separately used for training and testing.

During the data preprocessing phase, we removed batteries that did not achieve 80% capacity retention. Additionally, we cleaned the data from each batch, removing obviously noisy or anomalous data points to avoid their interference with the analysis results. To ensure comparability across cells and cycles, each curve was resampled onto a uniform voltage grid from 3.3 V to 2.8 V with a 0.001 V step (500 points), and missing values on this grid were filled via linear interpolation. We ensured that the training, calibration, and test datasets were strictly separated at the battery level to prevent any data leakage between the sets.

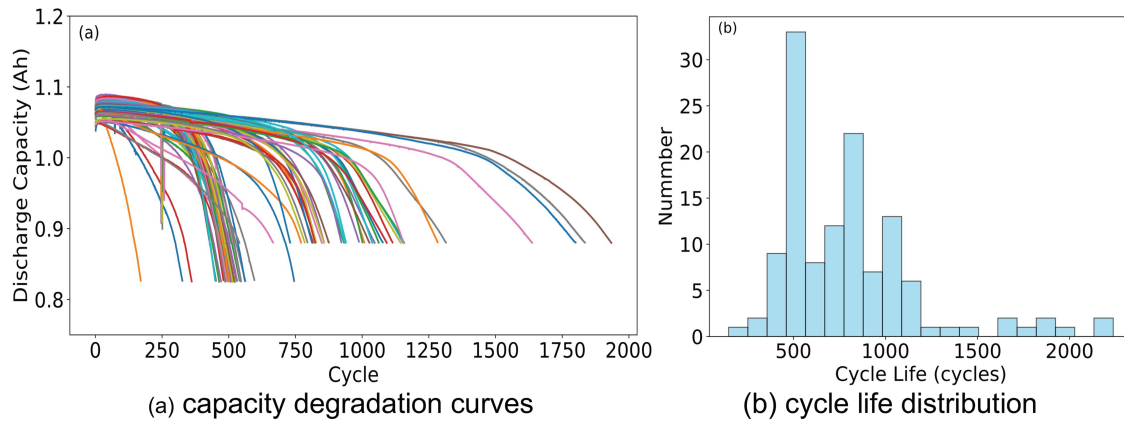


Figure 2: Dataset description (a) The relationship between the discharge capacity and cycle count of 124 batteries in the dataset (b) Lifetime distribution of 124 batteries

4 Experimental Comparison

4.1 Variable Importance Measure

Feature selection in data-driven battery life prediction presents significant challenges. During the selection process, the feature space may have high dimensionality, but not all features significantly contribute to prediction results and accuracy. Managing high-dimensional features involves balancing quantity, correlations, and redundancy, which increases computational complexity and reduces model generalizability. Therefore, it is necessary to screen for an optimal feature set before prediction.

Random Forest and other tree-based models inherently perform feature selection during training, significantly reducing the need for manual feature engineering in high-dimensional spaces. By variable importance measure, we can visualize each feature's contribution to determine an optimal subset.

First, we used data from the first 100 cycles of the training set and evaluated 14 candidate features with an RF model. Feature importance scores were computed and ranked in descending order, as shown in Fig. 3. The vertical axis lists the candidate features, and the horizontal axis shows their corresponding importance scores, where a larger value indicates a greater contribution to the prediction of the target variable. All features are first ranked in descending order of importance. Second, based on the feature ranking, the feature dimensions were progressively increased for prediction to evaluate accuracy. In summary, a comprehensive evaluation has been conducted based on the two factors previously discussed. Features with importance values exceeding 0.25 were classified as highly correlated features, significantly influencing prediction results. As shown in Fig. 3, selecting the top 5 features as a high-importance combination significantly influences the final prediction results. Table 1 presents the types of features and their detailed descriptions. Based on these results, the final feature set selected includes: Variance, Discharge Capacity slope, Kurtosis, Minimum and Internal resistance difference.

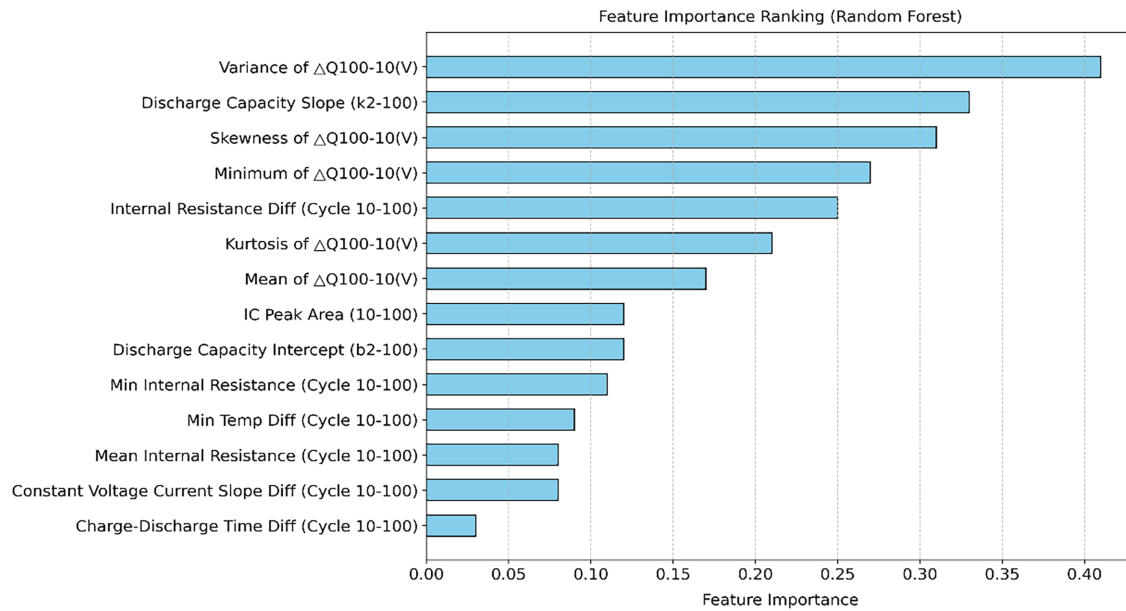


Figure 3: Feature importance ranking

Table 1: Feature types and descriptions

	Features	Descriptions
$\Delta Q_{100-10(V)}$	Variance	Variance of $\Delta Q_{100-10(V)}$
	Skewness	Skewness of $\Delta Q_{100-10(V)}$
	Minimum	Minimum of $\Delta Q_{100-10(V)}$
	Mean	Mean of $\Delta Q_{100-10(V)}$
	Kurtosis	Kurtosis of $\Delta Q_{100-10(V)}$
	IC peak area	The difference in the peak area of the IC curve, cycles 2 to 100
Discharge capacity fade curve features	Discharge capacity slope	Slope of the linear fit to the capacity fade curve, cycles 2 to 100
	Discharge capacity intercept	Intercept of the linear fit to capacity fade curve, cycles 2 to 100
Other features	Temperature	Min temperature difference, cycles 10 to 100
	Internal resistance	Min internal resistance, cycles 10 to 100
		Mean internal resistance, cycles 10 to 100
		Internal resistance difference, cycles 10 to 100
	Time	Charge-discharge time difference, cycles 10 to 100
	Slope	Constant voltage current slope difference, cycles 10 to 100

4.2 Evaluation Function

The following are used as evaluation functions, defined as follows [47]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

$$NMSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

$$MPIW = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i^U - \hat{y}_i^L) \quad (9)$$

$$CWC = MPIW \times \left(1 + \max \left(0, 1 - \frac{PICP}{\alpha} \right) \right) \quad (10)$$

In this equation, $p(y_i|x_i, \theta)$ is the predicted probability of y_i given the input x_i and model parameters θ .

Mean Predictive Interval Width (MPIW) quantifies the uncertainty of model predictions by evaluating the width of the prediction interval. Calibration-Weighted Coverage (CWC) integrates both the width of the prediction interval and a penalty mechanism associated with the coverage rate. PICP denotes the coverage probability of the prediction interval, and α represents the confidence level. The NMSE is employed to quantify the discrepancy between the model predictions and the observed values.

4.3 Experiment Results Analysis

Given that machine learning models influence prediction accuracy, the selected feature set is fed into multiple algorithms for comparative analysis. These include Artificial Neural Network (ANN), Elastic Net, K-Nearest Neighbors Regression (KNN), Random Forest, Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (AdaBoost) among others. All models are trained using the same methodology. Finally, the evaluation metrics defined in Section 4.2 are applied to quantify the computational outcomes of all models. The results are summarized in Table 2.

Table 2: Comparison of six prediction algorithms

	RMSE	MAE	MAPE	R ²	Time/s
ANN	132.70	96	9.42	0.895	12.527
Elastic net	172.43	129	13.7	0.793	8.788
RF	168	103	11.60	0.853	0.031
XGBoost	185	110	11.76	0.833	0.078
KNN	371	237	27.7	0.433	0.004
AdaBoost	168	112	12.60	0.843	0.056

Table 2 presents the specific RMSE, R^2 , computation time, and other details for six prediction algorithms. Fig. 4 shows the training and prediction results of the dataset used in the six models. A comparison with Table 2 reveals that RF, AdaBoost, ANN, and XGBoost exhibit strong performance across multiple metrics, including RMSE, MAE, MAPE, and R^2 . Among them, ANN achieves the highest predictive accuracy, whereas RF provides a favorable trade-off between accuracy and computational efficiency. Specifically, RF yields a lower RMSE than XGBoost and better-controlled MAPE than AdaBoost. In contrast, KNN and Elastic Net exhibit poor performance. Moreover, ANN and Elastic Net require substantially longer computation time than other methods. Overall, RF achieves strong predictive performance and is robust to noisy data. It also offers high computational efficiency and stable performance on relatively small datasets. Therefore, we select RF for the subsequent uncertainty quantification.

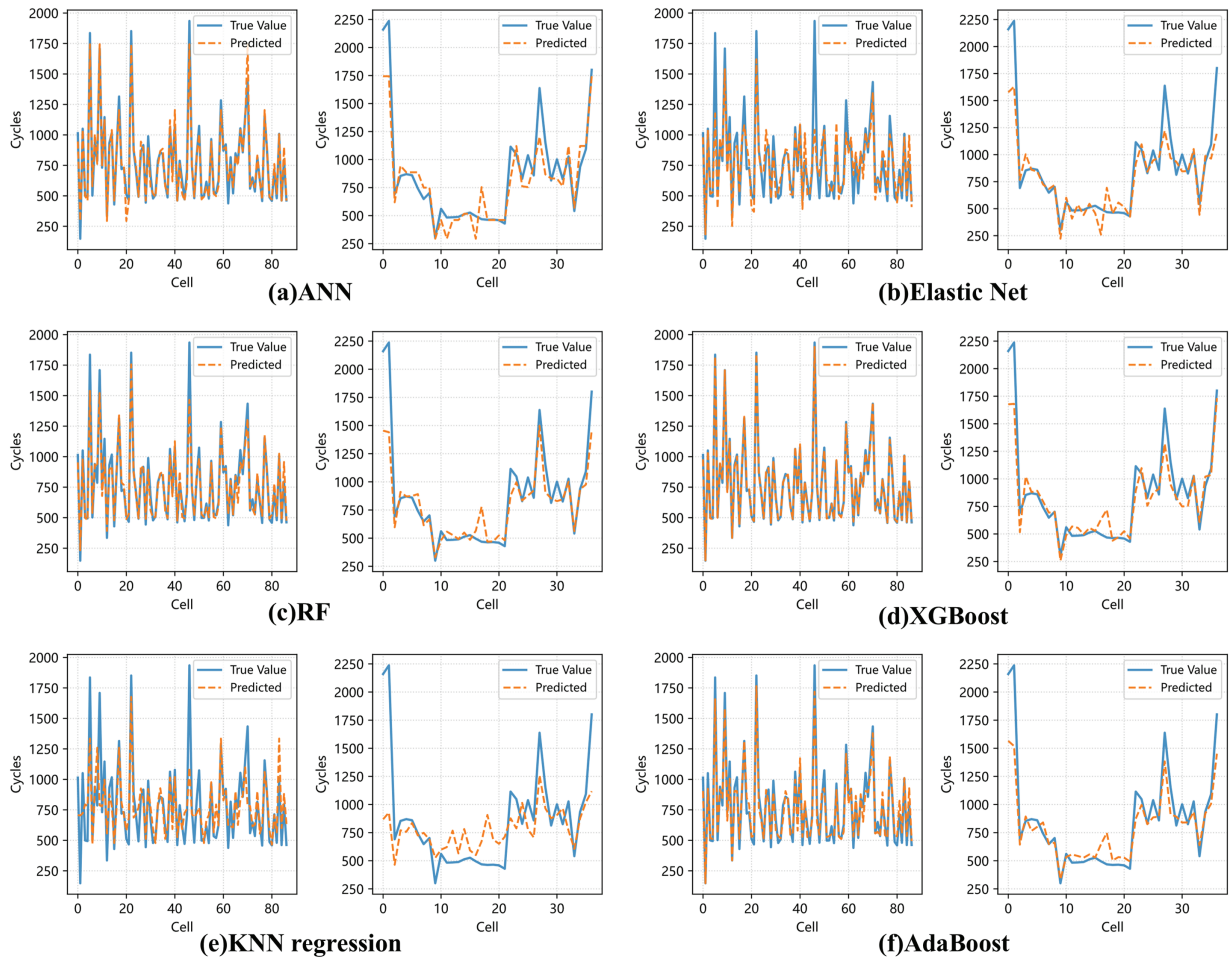


Figure 4: Training and testing performance of six machine learning models. For each model, the left panel shows results on the training dataset, while the right panel shows performance on the testing dataset

Due to the complexity and diversity of internal failure mechanisms in lithium-ion batteries, their nonlinear characteristics directly give rise to uncertainty in the data. Additionally, the degradation features of lithium-ion batteries exhibit significant variation as the state of health changes. This variation makes it difficult to achieve high-quality training for degradation models, resulting in substantial model uncertainty. These two types of uncertainty are not independent. They are interrelated to some extent. In order to investigate the impact of these two uncertainties on the lifetime of lithium-ion batteries, uncertainty

quantification was applied to the preprocessed battery cycling dataset using conformal prediction. The uncertainty region of the new model incorporating conformal prediction is visualized using an error bar plot. Quantitative validation, conducted using the evaluation metrics defined in Eqs. (4)–(10) and summarized in Table 3, assesses the numerical performance metrics of the method based on the obtained results.

Table 3: Comparison of experimental results at different confidence levels

Confidence level	90%	95%
MPIW	329	405
RMSE	126	126
CWC	329	405
NMSE	0.19	0.19
R^2	0.828	0.828

Fig. 5 presents the error bar plot for the predicted lifetimes at two different confidence levels. The observed battery lifetimes are marked with circular data points, while the corresponding predicted mean values are shown with a different marker style. The vertical error bars indicate the prediction intervals associated with each prediction. As the confidence level increases from 90% to 95%, the prediction intervals become wider, reflecting greater uncertainty coverage. The predicted values closely follow the trend of the observed data, and the majority of observed values fall within the predicted intervals.

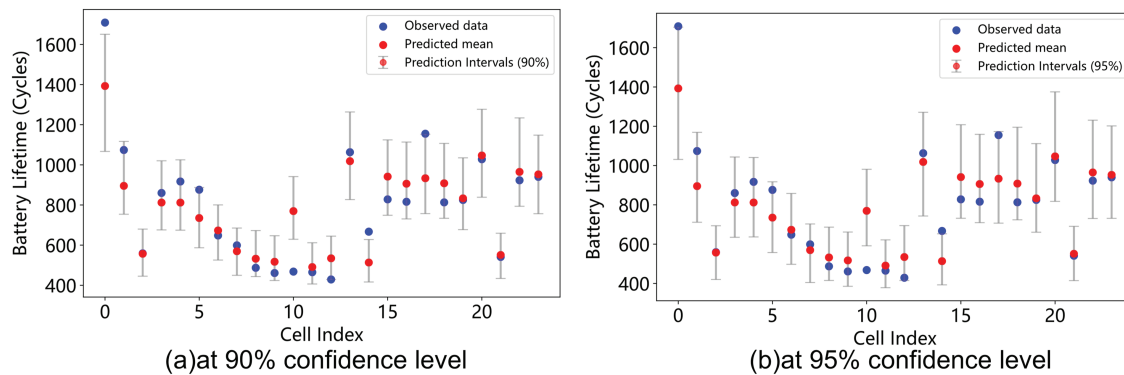


Figure 5: Uncertainty quantification results for 24 batteries at different confidence levels (a) At 90% confidence level (b) At 95% confidence level

Fig. 6 similarly shows the corresponding uncertainty regions at both confidence levels. To ensure the robustness of the results, cross-validation was used to evaluate the model performance. The circular data point represents the actual battery lifespan, while the shaded region represents the given uncertainty interval. The results of both figures indicate that conformal prediction effectively quantifies the uncertainty intervals of cycle life. According to the data in Table 3, with the confidence level rising from 90% to 95%, the MPIW increases from 329 to 405 cycles, indicating a total increase of 76 cycles in the average uncertainty interval width. At the 95% confidence level, the MPIW is 405 cycles, accounting for 19.3% of the battery lifetime range observed in the dataset. In addition, the differences in metrics such as NLL and NMSE are minimal, indicating that the variation in the discrepancy between the predicted probabilities and actual values is limited. Moreover, the prediction interval width increases systematically with higher confidence levels: compared to the 90% confidence level, the uncertainty width at 95% increases by an average of 76

cycles, representing a relative growth of 23%. This trend reflects the conservative nature of the method, which ensures statistical robustness by expanding the interval under stricter confidence requirements. Wider prediction intervals may indicate higher uncertainty. In practical applications, the system could adopt more conservative strategies, such as scheduling maintenance earlier or reducing the battery load.

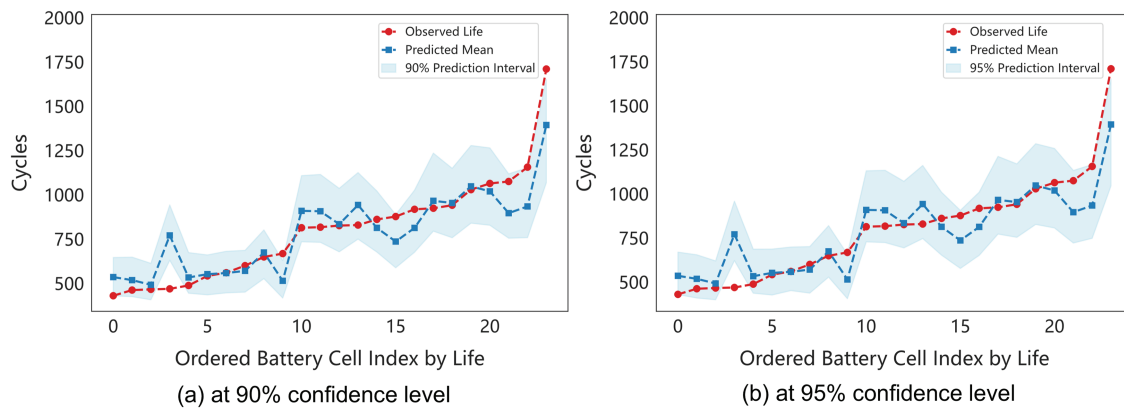


Figure 6: Uncertainty bands of the sorted batteries at different confidence levels

Next, compare three uncertainty quantification methods: Quantile Random Forest (QRF), Gaussian Process Regression, and Conformal Prediction. These three methods can be used to estimate the uncertainty of model predictions and generate prediction intervals. To evaluate the performance of these methods in prediction, we will use QRF as the baseline model and compare it with GPR and the proposed model.

QRF is a method based on the extension of Random Forest, which can provide predictions at different quantiles. Therefore, it not only provides point predictions but also quantifies prediction intervals. GPR is a non-parametric Bayesian regression method that provides a full predictive distribution. Thus, both its point predictions and interval widths can more accurately reflect the underlying uncertainty of the data. These two models typically demonstrate strong predictive performance in nonlinear and complex regression tasks.

Fig. 7 presents the prediction results of the QRF and GPR models under the same training conditions, at different confidence levels. Fig. 7a,b shows dark and light shaded regions representing the 90% and 95% confidence prediction intervals, respectively. The comparison reveals that the MPIW of QRF is smaller than that of GPR in both cases. Since the generated prediction intervals lack theoretical guarantees, a coverage evaluation of the uncertainty intervals for all three models is performed. In the coverage evaluation shown in Fig. 8, the empirical coverage of the QRF test set is lower than the nominal coverage, with 87.5% and 91.7% not meeting the baseline target coverages of 90% and 95%. GPR generally performs more stably in cases of small sample sizes and high noise levels. The uniformity of the uncertainty region in Fig. 7b further supports this observation. Compared to QRF, the region shows less fluctuation and is more evenly distributed. As the confidence level increases, the interval width increases more uniformly. In the coverage evaluation, the empirical coverages are 91.7% and 96.3%, meeting the baseline requirements of 0.9 and 0.95, respectively, indicating good calibration performance. The results show that the actual coverage of the RF-CP model meets the nominal coverage requirement, indicating that the model does not exhibit underconfidence. Furthermore, compared to other models, the coverage difference is smaller, indicating that our model does not exhibit significant overconfidence.

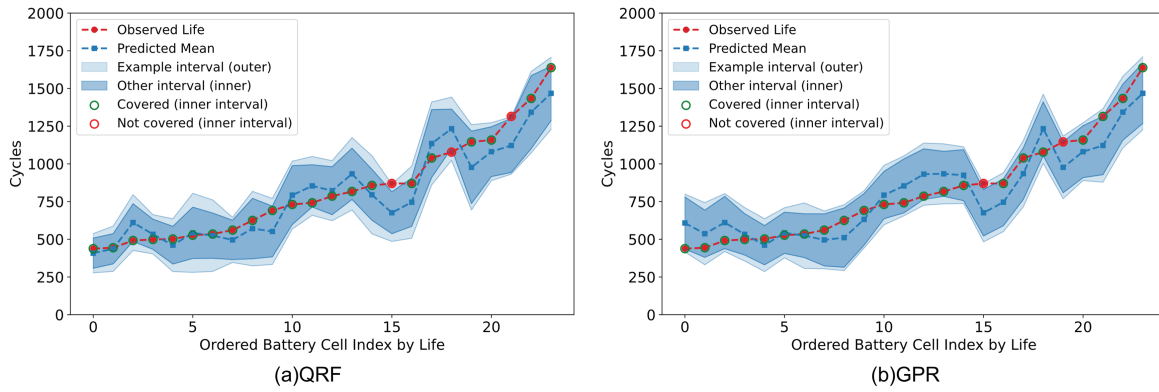


Figure 7: Uncertainty regions of other models

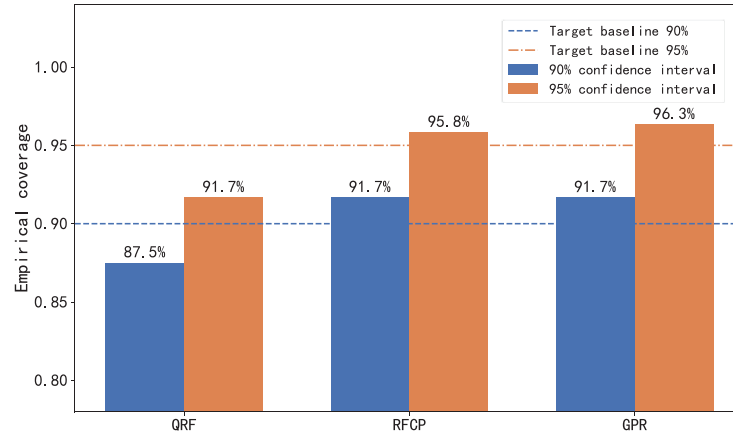


Figure 8: Comparison of the prediction interval coverage for three different models at 90% and 95% confidence targets

Table 4 compares the uncertainty metrics of three models, including MPIW, RMSE, and CWC. Table 4 shows that both the QRF model and the proposed model are based on the RF method. However, the proposed model combines the efficiency of RF with the theoretical guarantees offered by CP. The proposed model has an MPIW of 405 cycles, which is about 20 cycles smaller. The proposed model achieves smaller interval width while better meeting coverage requirements, resulting in a more reliable interval. In the quantification process, coverage should not be sacrificed for smaller prediction intervals. Both the proposed model and GPR satisfy the theoretical coverage guarantees, with the proposed model having a slightly smaller interval width than GPR. Additionally, GPR is computationally more complex, leading to longer computation times in large datasets.

Table 4: Comparison of uncertainty metrics across different models

Methods	90% confidence level			95% confidence level		
	MPIW	RMSE	CWC	MPIW	RMSE	CWC
QRF	321	129	330	424	132	438
RFCP	329	126	329	405	126	405
GPR	342	116	342	439	116	439

5 Discussion and Outlook

The findings above raise several important considerations for application. This section discusses future research directions and potential applications related to uncertainty.

Uncertainty quantification has important practical value. Given a prediction interval from early-cycle data, the interval width can be used to prioritize testing resources: cells with consistently wide intervals are flagged early for additional cycles or measurements. Meanwhile, cells with narrow intervals can be tested with reduced effort, thereby lowering overall time and cost. When integrated into a battery management system, the lower bound provides a conservative lifetime indicator. If lower bound falls below a predefined threshold, the controller can switch to a more conservative operating policy by reducing the charge rate and strictly controlling the temperature. This helps maintain safety margins while still exploiting performance.

Although current studies and the RF-CP method proposed in this work quantify prediction uncertainty, it still has several limitations. First, the types of cells and the range of operating conditions in the dataset are relatively limited, so the applicability of the resulting prediction intervals to different chemistries and usage scenarios remains to be verified. Second, this study uses residual-based nonconformity scores, which provide an aggregated measure of predictive uncertainty. However, this method does not distinguish between aleatoric and epistemic uncertainty within the current framework.

Based on the above limitations, future research can proceed in several directions. First, uncertainty quantification should be extended to more battery datasets and operating conditions, enabling cross-dataset and cross-platform validation. If necessary, transfer learning can also be incorporated. Second, by combining the current framework with Bayesian models and deep ensembles, methods can be developed to explicitly decompose aleatoric and epistemic uncertainty. This will provide a more detailed description of the different sources of uncertainty. Third, richer evaluation metrics and assessment schemes should be developed to more comprehensively evaluate the quality of the uncertainty estimates and their suitability for practical applications.

6 Conclusions

Considering the multiple sources of uncertainty in lithium-ion battery cycle life prediction, this paper proposes a hybrid model that combines Random Forest with Conformal Prediction. Rather than providing point predictions, this method offers interval predictions, using Conformal Prediction to quantitatively define the uncertainty range of battery life. The model can reliably predict the full life cycle of a battery using degradation data from the first 100 cycles. This has been confirmed through comparisons with six machine learning models. Furthermore, the prediction intervals were quantified at different confidence levels. To evaluate the predictive accuracy and the ability to characterize uncertainty, we compared the proposed approach with two existing UQ methods, QRF and GPR. Metrics such as MPIW were calculated to assess the reliability and width of the prediction intervals, and the coverage of the results was also examined. The final results show that the RF-CP method provides stable uncertainty quantification across varying confidence levels. It covers approximately 19.3% of the observed battery lifespan range in the dataset and meets the baseline coverage requirement.

In summary, this method provides a new perspective for practical applications in battery lifetime prediction. Integrating uncertainty quantification into lithium-ion battery research can lead to more accurate, reliable, and interpretable predictions of battery degradation trends, supporting trustworthy decision-making in practical applications.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Study conception and design: Mingqi Liu, Wujiang Li; Data collection: Mingqi Liu; Analysis and interpretation of results: Wujiang Li, Mingqi Liu; Writing—review and editing: Mingqi Liu, Ying Wang; Visualization: Juyong Cao; Supervision: Fuyong Yang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, Ying Wang, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Zhang G, Wei X, Tang X, Zhu J, Chen S, Dai H. Internal short circuit mechanisms, experimental approaches and detection methods of lithium-ion batteries for electric vehicles: a review. *Renew Sustain Energy Rev.* 2021;141(9):110790. doi:10.1016/j.rser.2021.110790.
2. Feng X, Ouyang M, Liu X, Lu L, Xia Y, He X. Thermal runaway mechanism of lithium ion battery for electric vehicles: a review. *Energy Storage Mater.* 2018;10(3):246–67. doi:10.1016/j.ensm.2017.05.013.
3. Su L, Zhang S, McGaughey AJH, Reeja-Jayan B, Manthiram A. Battery charge curve prediction via feature extraction and supervised machine learning. *Adv Sci.* 2023;10(26):2301737. doi:10.1002/advs.202301737.
4. Wang M, Wu S, Chen Y, Luan W. The snowball effect in electrochemical degradation and safety evolution of lithium-ion batteries during long-term cycling. *Appl Energy.* 2025;378:124909. doi:10.1016/j.apenergy.2024.124909.
5. Zhang X, Li Z, Luo L, Fan Y, Du Z. A review on thermal management of lithium-ion batteries for electric vehicles. *Energy.* 2022;238:121652. doi:10.1016/j.energy.2021.121652.
6. Zeng X, Li M, Abd El-Hady D, Alshitari W, Al-Bogami AS, Lu J, et al. Commercialization of lithium battery technologies for electric vehicles. *Adv Energy Mater.* 2019;9(27):1900161. doi:10.1002/aenm.201900161.
7. Huang L, Liu L, Lu L, Feng X, Han X, Li W, et al. A review of the internal short circuit mechanism in lithium-ion batteries: inducement, detection and prevention. *Int J Energy Res.* 2021;45(11):15797–831. doi:10.1002/er.6920.
8. Keil J, Jossen A. Electrochemical modeling of linear and nonlinear aging of lithium-ion cells. *J Electrochem Soc.* 2020;167(11):110535. doi:10.1149/1945-7111/aba44f.
9. Wu K, Wu X, Lin Z, Sun H, Wang M, Li W. Identifying the calendar aging boundary and high temperature capacity fading mechanism of Li ion battery with Ni-rich cathode. *J Power Sources.* 2024;589(9):233736. doi:10.1016/j.jpowsour.2023.233736.
10. Ma S, Zhang X, Wu S, Fan E, Lin J, Chen R, et al. Unraveling the nonlinear capacity fading mechanisms of Ni-rich layered oxide cathode. *Energy Storage Mater.* 2023;55(4):556–65. doi:10.1016/j.ensm.2022.12.009.
11. Roth EP, Doughty DH. Thermal abuse performance of high-power 18650 Li-ion cells. *J Power Sources.* 2004;128(2):308–18. doi:10.1016/j.jpowsour.2003.09.068.
12. Liu J, Zhang Y, Bai J, Zhou L, Wang Z. Influence of lithium plating on lithium-ion battery aging at high temperature. *Electrochim Acta.* 2023;454:142362. doi:10.1016/j.electacta.2023.142362.
13. Jung SK, Hwang I, Chang D, Park KY, Kim SJ, Seong WM, et al. Nanoscale phenomena in lithium-ion batteries. *Chem Rev.* 2020;120(14):6684–737. doi:10.1021/acs.chemrev.9b00405.
14. Hu X, Xu L, Lin X, Pecht M. Battery lifetime prognostics. *Joule.* 2020;4(2):310–46. doi:10.1016/j.joule.2019.11.018.
15. Zhu ZW, Miao JW, Zhu XY, Wang XX, Qiu JY, Zhang H. Research progress in lithium-ion battery remaining useful life prediction based on machine learning. *Energy Storage Sci Technol.* 2024;13(9):3134–49. (In Chinese). doi:10.19799/j.cnki.2095-4239.2024.0713.

16. Bloom I, Cole BW, Sohn JJ, Jones SA, Polzin EG, Battaglia VS, et al. An accelerated calendar and cycle life study of Li-ion cells. *J Power Sources*. 2001;101(2):238–47. doi:10.1016/S0378-7753(01)00783-2.
17. Lian Y, Wang JV, Deng X, Kang J, Zhu G, Xiang K. Remaining useful life prediction of lithium-ion batteries using semi-empirical model and bat-based particle filter. In: *Proceedings of the 2020 IEEE International Symposium on Circuits and Systems (ISCAS)*; 2020 Oct 12–14; Seville, Spain. p. 1–5.
18. Li J, Adewuyi K, Lotfi N, Landers RG, Park J. A single particle model with chemical/mechanical degradation physics for lithium ion battery state of health (SOH) estimation. *Appl Energy*. 2018;212(4):1178–90. doi:10.1016/j.apenergy.2018.01.011.
19. Luo F, Huang H, Ni L, Li T. Rapid prediction of the state of health of retired power batteries based on electrochemical impedance spectroscopy. *J Energy Storage*. 2021;41(3):102866. doi:10.1016/j.est.2021.102866.
20. Khodadadi Sadabadi K, Jin X, Rizzoni G. Prediction of remaining useful life for a composite electrode lithium ion battery cell using an electrochemical model to estimate the state of health. *J Power Sources*. 2021;481(39):228861. doi:10.1016/j.jpowsour.2020.228861.
21. Li L, Hu M, Xu Y, Fu C, Jin G, Li Z. State of charge estimation for lithium-ion power battery based on H-infinity filter algorithm. *Appl Sci*. 2020;10(18):6371. doi:10.3390/app10186371.
22. Ouyang T, Xu P, Lu J, Hu X, Liu B, Chen N. Coestimation of state-of-charge and state-of-health for power batteries based on multithread dynamic optimization method. *IEEE Trans Ind Electron*. 2022;69(2):1157–66. doi:10.1109/TIE.2021.3062266.
23. Wang S, Ma H, Zhang Y, Li S, He W. Remaining useful life prediction method of lithium-ion batteries is based on variational modal decomposition and deep learning integrated approach. *Energy*. 2023;282(5):128984. doi:10.1016/j.energy.2023.128984.
24. Xiong W, Xu G, Li Y, Zhang F, Ye P, Li B. Early prediction of lithium-ion battery cycle life based on voltage-capacity discharge curves. *J Energy Storage*. 2023;62:106790. doi:10.1016/j.est.2023.106790.
25. Fan W, Zhu J, Qiao D, Jiang B, Wang X, Wei X, et al. Prediction of nonlinear degradation knee-point and remaining useful life for lithium-ion batteries using relaxation voltage. *Energy*. 2024;294(6):130900. doi:10.1016/j.energy.2024.130900.
26. Čivilis A, Petkevičius L, Šaltenis S, Torp K, Markucevičiūtė-Vinckė I. Few-shot learning for triplet-based EV energy consumption estimation. *Appl Artif Intell*. 2025;39(1):2474785. doi:10.1080/08839514.2025.2474785.
27. Chen L, Ding Y, Liu B, Wu S, Wang Y, Pan H. Remaining useful life prediction of lithium-ion battery using a novel particle filter framework with grey neural network. *Energy*. 2022;244(1):122581. doi:10.1016/j.energy.2021.122581.
28. Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Z, et al. Data-driven prediction of battery cycle life before capacity degradation. *Nat Energy*. 2019;4(5):383–91. doi:10.1038/s41560-019-0356-8.
29. Li P, Wu X, Grosu R, Hou J, Ilolov M, Xiang S. Applying neural network to health estimation and lifetime prediction of lithium-ion batteries. *IEEE Trans Transp Electr*. 2025;11(1):4224–48. doi:10.1109/TTE.2024.3457621.
30. Wang Y, Zhu J, Cao L, Gopaluni B, Cao Y. Long short-term memory network with transfer learning for lithium-ion battery capacity fade and cycle life prediction. *Appl Energy*. 2023;350(3):121660. doi:10.1016/j.apenergy.2023.121660.
31. Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications. *Comput Biol Med*. 2019;112(4):103375. doi:10.1016/j.combiomed.2019.103375.
32. Gregorutti B, Michel B, Saint-Pierre P. Correlation and variable importance in random forests. *Stat Comput*. 2017;27(3):659–78. doi:10.1007/s11222-016-9646-1.
33. Zhou R, Wang S, Ren Y, Liu K, Peng X, He L, et al. Remaining useful life prediction of lithium-ion batteries based on informer model with multi-feature extraction and Bayesian optimization. *Chem Eng Sci*. 2025;316:121878. doi:10.1016/j.ces.2025.121878.
34. Liu L, Huang J, Zhao H, Li T, Li B. PatchFormer: a novel patch-based transformer for accurate remaining useful life prediction of lithium-ion batteries. *J Power Sources*. 2025;631:236187. doi:10.1016/j.jpowsour.2025.236187.
35. Ghahramani Z. Probabilistic machine learning and artificial intelligence. *Nature*. 2015;521(7553):452–9. doi:10.1038/nature14541.

36. Kong JZ, Cui D, Hou B, Liu X, Wang D. New short-long-term degradation model for precise battery health prognostics. *IEEE Trans Ind Electron.* 2023;70(9):9527–37. doi:10.1109/TIE.2022.3212365.
37. Lin YH, Ding ZQ, Li YF. Similarity based remaining useful life prediction based on Gaussian process with active learning. *Reliab Eng Syst Saf.* 2023;238:109461. doi:10.1016/j.ress.2023.109461.
38. Lu H, Zheng G, Hua Y, Talemi R, Gryllias K, Chronopoulos D. Prediction of the residual stiffness of composite materials under random vibration loading using a combined probabilistic random forest and probabilistic stiffness model. *ASCE-ASME J Risk Uncertainty Eng Syst Part A Civ Eng.* 2025;11(2):4025017. doi:10.1061/ajrua6.rueng-1556.
39. Xiao D, Qiao W. Hybrid scenario generation method for stochastic virtual bidding in electricity market. *CSEE J Power Energy Syst.* 2021;7(6):1312–21. doi:10.17775/CSEEJPES.2021.00890.
40. Zhu W, Li C, Xu Y, Yang W, Xie C. High accuracy and adaptability of PEMFC degradation interval prediction with Informer-GPR under dynamic conditions. *Energy.* 2024;307:132781. doi:10.1016/j.energy.2024.132781.
41. Iwakin O, Moazeni F. Improving urban water demand forecast using conformal prediction-based hybrid machine learning models. *J Water Process Eng.* 2024;58:104721. doi:10.1016/j.jwpe.2023.104721.
42. Hajibabae P, Behnood A, Ngo T, Mohammadi Golafshani E. Carbonation depth assessment of recycled aggregate concrete: an application of conformal prediction intervals. *Expert Syst Appl.* 2025;268(8):126231. doi:10.1016/j.eswa.2024.126231.
43. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32. doi:10.1023/A:1010933404324.
44. Raschka S. *Python machine learning*. Birmingham, UK: Packt Publishing; 2015.
45. Auer A, Gauch M, Klotz D, Hochreiter S. Conformal prediction for time series with modern Hopfield networks. *arXiv:2303.12783*. 2023.
46. Angelopoulos AN, Bates S. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv:2107.07511*. 2021.
47. Nam NB, Ogliari E, Leva S, Pafumi E, Alberti D, Duong MQ. Comparative analysis of conformal prediction techniques and machine learning models for very short-term solar power forecasting. *Energy AI.* 2025;21(12):100573. doi:10.1016/j.egyai.2025.100573.